



ISSN: -2663 -3264



مجلة اتحاد الإحصائيين العرب

مجلة علمية تصدر عن اتحاد الإحصائيين العرب في مواضيع العلوم الإحصائية والمعلوماتية

JASU



رئيس التحرير
الدكتور / حيدر محمد فرحات

مدير التحرير
الأستاذ الدكتور جاسم محمد التميمي



المجلد السابع
العدد الثاني
٢٠٢٥



ISSN: -2663 -3264

A Scientific judged journal issued By the Arab Statistician Union

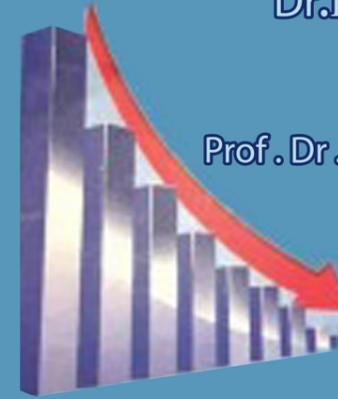
JASU

Eoditorial Chief

Dr.Haider Mohammad Farhat

Eoditorial Secretary

Prof . Dr . Jasim Muhammad Al-Tameemi



Volume 7
Issue No.2
2025

مجلة علمية محكمة تصدرها جامعة ميسان بالتعاون مع اتحاد الإحصائيين العرب في مواضيع العلوم الإحصائية والمعلوماتية



مجلة اتحاد الإحصائيين العرب
مجلة علمية محكمة يصدرها اتحاد الإحصائيين العرب
Journal of Arab Statisticians Union (JASU)

رئيس التحرير

معالي الأستاذ الدكتور حيدر محمد فريجات

مدير التحرير

الأستاذ الدكتور جاسم محمد التميمي

الناشر

اتحاد الإحصائيين العرب

عمان – الاردن

المجلد السابع / العدد الثاني لسنة 2025

رقم الايداع في دار الكتب والوثائق الوطنية ببغداد (2438) لسنة 2020م

اتحاد الإحصائيين العرب , مبنى دائرة الإحصاءات العامة , عمان المملكة الاردنية الهاشمية
Union of Arab Statisticians , Building , Amman , Jordan
P.O. Box 2892 Amman 11941 Jordan, Tel: +962 6 5300700 ,
Fax: +962 6 5300710 Emails: info@uarabs.org , jasimtimimi@yahoo.com
secgen@uarabs.org , Website: www.uarabs.org

المطبعة المركزية / جامعة ديالى

العراق – ديالى طريق بغداد / بعقوبة القديم



اسم الكتاب: مجلة اتحاد الاحصائيين العرب

سنة الطبع: ٢٠٢٥ م

الايمل: center.printer2009@gmail.com

جميع حقوق الملكية الأدبية والفنية محفوظة للمؤلف ويحظر طبع أو
تصوير أو ترجمة أو إعادة تنضيد الكتاب كاملاً أو مجزئاً أو تسجيله
على أشرطة كاسيت أو إدخاله على الحاسوب أو برمجته على اسطوانات
ضوئية إلا بموافقة المؤلف خطياً

رقم الإيداع في دار الكتب والوثائق الوطنية ببغداد (٢٤٣٨) لسنة (٢٠٢٠م)

هيئة تحرير المجلة العلمية لاتحاد الإحصائيين العرب

العمل في المجلة	الاسم ومكان العمل	ت
رئيسا	أ.د. حيدر محمد فريحات - مدير عام دائرة الإحصاءات الأردنية - الأردن	1
نائبا للرئيس	أ.د. غازي إبراهيم رحو - الأمين العام لاتحاد الإحصائيين العرب	2
مدير التحرير	أ.د. جاسم محمد التميمي - جامعة ديالى - العراق	3
عضوا	أ.د. مهدي العلق - رئيس الجمعية العراقية للعلوم الإحصائية - العراق	4
عضوا	أ.د. علا فرح عوض - رئيس الجهاز المركزي للإحصاء الفلسطيني - دولة فلسطين	5
عضوا	أ.د. عادل مانع داخل - رئيس جامعة ميسان - العراق	6
عضوا	أ.د. مزهر شعبان العاني - الأمين العام المساعد لاتحاد الإحصائيين العرب	7
عضوا	أ.د. عوض حاج علي - جامعة النيلين - جمهورية السودان	8
عضوا	أ.د. حسين عبد العزيز السيد - جمهورية مصر العربية	9
عضوا	أ.د. اسماعيل بن قانة - الجزائر - يعمل في جامعة الملك فيصل	10
عضوا	أ.د. محمد ابو صالح - جامعة عمان العربية - الاردن -	11
عضوا	د. سالم محمد القزيزي - دولة ليبيا	12
عضوا	د. وصفي طاهر صالح - جامعة تيشك الدولية - العراق	13
عضوا	د. عبد اللطيف فراخ - المغرب	14

هيئة مستشارو المجلة العلمية لاتحاد الإحصائيين العرب

ت	الاسم ومكان العمل	العمل في المجلة
1	أ.د. مزهر شعبان العاني - الأمين العام المساعد لاتحاد الإحصائيين العرب	رئيسا
2	أ.د. محمود أبو شعير - جمهورية العراق عميد كلية الرافدين الجامعة	عضوا
3	أ.د. أحسن طيار - جامعة 20 أوت 1955 - سكيكدة - سكيكدة - الجزائر	عضوا
4	أ.د. عبد الوهاب محمد جواد الموسوي - جامعة الكوفة - جمهورية العراق	عضوا
5	أ.د. مناف يوسف حمود - جمهورية العراق	عضوا
6	أ.د. أحلام أحمد جمعة - كلية الإدارة والاقتصاد - جامعة بغداد - جمهورية العراق	عضوا
7	د. يحيى بن خميس الحسيني - رئيس الجمعية العمانية للإحصاء - سلطنة عمان	عضوا
8	د. علاء عبد السلام مصطفى - جامعة ميسان - جمهورية العراق	عضوا
9	د. نبيل جورج ناسي - جامعة صلاح الدين - أربيل - العراق	عضوا
10	أ.د. محمد خليل إبراهيم - عميد كلية الإدارة والاقتصاد - جامعة ميسان - جمهورية العراق	عضوا
11	د. عامر المقدسي رئيس المركز الدولي للتنمية الإدارية (IDS) - (الولايات المتحدة الأمريكية)	عضوا
12	د. علي خالد عبد الله جامعة ميسان	عضوا
13	د. ولاء جودت الجاف - جمهورية العراق	عضوا
14	د. سلوى محمد نجرس علاوي - وزارة التعليم العالي والبحث العلمي - جمهورية العراق	عضوا
15	د. زرزار العياشي - الجمهورية الجزائرية الديمقراطية الشعبية	عضوا
16	د. محمود الصروي - الجهاز المركزي للتعبئة والإحصاء المصري - مصر	عضوا
17	د. مكين الدين أحمد عبد الله - جمهورية السودان	عضوا
18	السيد أيوي أيوب - دولة فلسطين	عضوا

شروط النشر

- 1- تنشر المجلة البحوث والدراسات العلمية في المجالات الإحصائية والمعلوماتية المكتوبة باللغة العربية أو الانكليزية أو الفرنسية على أن لا يكون البحث المقدم للنشر قد نشر أو قدم للنشر في مجلات أو دوريات أخرى أو قدم ونشر في دوريات لمؤتمرات أو ندوات.
- 2- ترسل نسخة إلكترونية من البحوث والدراسات الى مدير التحرير على أن تتضمن اسم الباحث أو الباحثين وألقابهم العلمية وأماكن عملهم مع ذكر عنوان المراسلة وأرقام الهواتف والفاكس والبريد الإلكتروني علماً أن أجور النشر اي بحث يكلف مبلغ (١٠٠) دولار. او ما يعادله بالدينار العراقي.
- 3- يرسل البحث المراد نشره بالمجلة مطبوعاً إلكترونياً على أما على دسك او بالايميل على عنوان سكرتارية التحرير وفق المواصفات أدناه:
أ. باللغة العربية باستخدام حرف نوع (Simplified Arabic)
ب. الانكليزية (Times New Roman) وبخط حجم (14) وباستخدام نظام Microsoft Word
ج. أن يتم ترك مسافة 2.5 سم لكافة أبعاد الصفحة.
د. يرفق الباحث ملخصاً عن بحثه باللغة الانكليزية أو العربية أو الفرنسية حسب لغة البحث وبما لا يزيد عن صفحة واحدة .
هـ. يتم الإشارة الى المصادر العلمية في متن البحث وفي نهايته ، مع مراعاة أن لا تتضمن الصفحة الاخيرة سوى المصادر التي تم الإشارة إليها في المتن ووفق الاصول المعتمدة في ذلك (اسم المؤلف ، سنة النشر ، عنوان المصدر ، دار النشر ،البلد)
و. ترقم الجداول والرسوم التوضيحية وغيرها حسب ورودها في البحث ، كما توثق المستعارة منها بالمصادر الأصلية.
ز. أن لا يزيد عدد صفحات البحث او الدراسة عن (20) صفحة A4. وضرورة إرسال النسخة الإلكترونية على عنوان سكرتارية التحرير.
ح. أن يكون البحث مرتب على شكل عمودين في الصفحة.
- 4- سيتم إشعار الباحث باستلام بحثه خلال مدة لا تتجاوز اسبوعين من تاريخ الاستلام.
- 5- تخضع كل البحوث المرسلة للنشر في المجلة للتقويم العلمي و الموضوعي و يبلغ الباحث بنتيجة التقويم والتعديلات المقترحة إن وجدت خلال مدة لا تتجاوز اسبوعين من تاريخ استلام الردود من كل المقيمين.
- 6- لهيئة تحرير المجلة الحق في قبول او رفض البحث ولها الحق في إجراء أي تعديل او إعادة صياغة جزئية للمواد المقدمة للنشر بما يتماشى و النسق المعتمد في النشر لديها بعد موافقة الباحث.
- 7- يصبح البحث المنشور ملكاً للمجلة ولا يجوز إعادة نشره في أماكن أخرى.
- 8- تعبر المواد المنشورة بالمجلة عن آراء أصحابها ، ولا تعكس بالضرورة وجهة نظر المجلة او اتحاد الاحصائيين العرب.
- 9- ترسل البحوث على العناوين أدناه:
العنوان البريدي للاتحاد : ص.ب. 851104 عمان 11185 الأردن.
البريد الإلكتروني : uarabs@gmail.com
أو مدير التحرير : jasimtimi@yahoo.com

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

كلمة العدد

انه من دواعي السروران تتواصل مجلة اتحاد الاحصائيين العرب بإصدار اعدادها بشكل مستمر وكما هو مخطط لها وقد بذل الاتحاد والامانة العامة وهيئة التحرير والمختصين جهودا كبيرا لإظهار هذه المجلة من اجل ان تكون نبراسا للباحثين وطريقا لإيصال بحوثهم الى اكبر عدد من الاختصاصيين حيث يتم توزيع هذه المجلة على الاجهزة الاحصائية والمؤسسات المنضوية ضمن الاتحاد بالإضافة الى اعضاء الاتحاد ومن خلال دار النشر التي تقوم بعرضها من خلال المعارض والمؤتمرات العلمية .و أقدم شكري وتقديري للأستاذ الدكتور حيدر محمد فريحات رئيس الاتحاد الإحصائيين العرب و للأمين العام لاتحاد الاحصائيين العرب الاستاذ الدكتور غازي ابراهيم رحو و الامين العام المساعد الاستاذ الدكتور مزهر العاني لجهودهم المثمرة ومتابعتهم المستمرة وتوجيهاتهم القيمة في انجاز العدد لمجلة الاتحاد مع التقدير و الاحترام.

نتقدم لكم جميعا بالتهنئة ونبارك لكم صدور المجلد السابع / العدد الثاني لسنة 2025 الخاص بوقائع المؤتمر الثامن لاتحاد الإحصائيين العرب الذي انعقد في اربيل 16-17-نيسان-2025 من مجلتنا الغراء JASU وهي تصدر بحلتها الجديدة وقد حصلت المجلة على الترقيم الدولي (ISSN 2663-3264) ومما يزيد من رصانة المجلة العلمية وتحتوي المجلة على بحوث محكمة وحديثة وتعالج مشاكل احصائية جديدة وتقدم افكار بحثية متطورة نتمنى ان يكون العدد القادم غزيرا بالبحوث القيمة وان يصدر في موعده المحدد بفضل دعمكم ومشاركتم الفعالة مدعما بالبحوث المتطورة وخاصة ان الاحصاء الان يعتبر العنصر المهم والاساسي بتداخله مع كافة العلوم مما يفسح المجال لتطبيقات عديدة ومساهمات متنوعة .
ولكم الموفقية والنجاح للجميع شاكرين كل من وضع لبنة في بناء ونشر هذا العدد من المجلة وتقبلوا وافر التقدير والاحترام

مدير التحرير

الاستاذ الدكتور جاسم محمد التميمي

الرؤية والرسالة والاهداف

الرؤية Vision

الريادة في نشر البحوث الاحصائية والسعي للوصول لتصنيف عالمي متقدم بين المجالات العلمية المحكمة , وان تكون مجلة اتحاد الاحصائيين العرب نبراسا للعلم والمعرفة وواجهة علمية وثقافية واكاديمية مشرقة لاتحادنا الموقر ومركز علم خلاق يجمع بين الاصاله والحدائنه.

الرسالة Mission

اثراء العلوم الاحصائية بأجود انواع البحوث والدراسات الاحصائية التي تربط بين الاصاله والحدائنه ضمن اطار علمي بناء باستثارة همم الباحثين وتنمية قدراتهم في النشر العلمي الاصيل باللغتين العربية والانكليزية وبما يسهم في ايصال الاحصاء بكل انواعه لكل شعوب العالم واتاحة الفرصة للباحثين لتقديم الصورة الحقيقية الناصعة للمجتمع .

الاهداف Goals

- تسعى مجلة اتحاد الاحصائيين العرب الى تحقيق الاهداف الاتية :
- 1- تشجيع البحوث والدراسات الاحصائية التي تربط الاصاله بالحدائنه وصولا الى تنمية الاعتزاز بماضيها الجميل والاختيار الواعي لما في الحدائنه من توجيهات تنفع الجيل الجديد.
 - 2- تنشيط البحث العلمي التخصصي في الاحصاء وتكنولوجيا المعلومات.
 - 3- تنمية الوعي الاحصائي والبحث العلمي لدى الجيل الجديد من خلال استعراض البحوث الاحصائية وتكنولوجيا المعلومات التي تساهم في انماء روح الاحترام للأصاله .
 - 4- التواصل العلمي والبحثي الهادف مع المراكز العلمية والعلماء والباحثين للإبراز دور الاحصاء في رفد المجتمع وتطويره.
 - 5- المساهمة في حل بعض الاشكاليات الاحصاء والمجتمع من خلال البحوث الاحصائية التي تساهم في مساعدة المجتمع لحل اشكالياته ورفد المؤسسات الاحصائية بالبحوث العلمية الرصينة.

المحتويات

الصفحة	عنوان البحث	الاسم	ت
الابحاث باللغة العربية			
20-1	واقع التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل	أ.م.د محمد بن سعيد الغافري أحمد فؤاد علوان أ.د. جاسم محمد علي التميمي	1.
43-21	مقارنة بعض مؤشرات زيادة الوزن والسمنة في المسحين العنقوديين متعدد (MICS4) و (MICS3) المؤشرات و 2011 للعامين 2006	أ.د. سجي محمد حسين ظافر طارق	2.
66-44	المفاضلة بين نتائج البرمجة الخطية ومشاكل النقل في تخفيض التكاليف الى ادنى حد ممكن (دراسة تطبيقية على شركة دهوك الزراعية لانتاج الدجاج المحلي في محافظة دهوك)	أ.م. سماهر طارق ابراهيم	3.
84-67	استعمال الانحدار اللوجستي للتنبؤ بتأثير على تخطيط كهربائية الدماغ Berger	رائد فاضل محمد الحسني حسنين جليل نعمه الساعدي	4.
106-85	التحول الرقمي ومساهمته في إدارة ملف السكان تجارب بعض الدول العربية	نيفين ناجي أنيس	5.
الابحاث باللغة الانكليزية			
12-1	Detecting and estimation outliers in multiple linear regression models	أ.د. شلال حبيب الجبوري أ.م. د. نبيل جورج ناسي	6.
32-13	Using Panel Data to Build an Efficient Statistical Model for Predicting Student Graduation Estimates	أ.د. عبيد محمد محسن مصطفى عثمان الصايغ	7.
57-33	Forecasting future sales prices and showing changes in Erbil Diamond Zone using time series models	د. محمد عبد المجيد بدليناز ديدار يوسف	8.
97-58	Prediction by Normalized Radial Basis Function Artificial Neural Network and model Selection for the Risk Factors Most Affected on Kidney Transplantation	شيماء محمد شاكر أحمد أ.م.د. نظيرة صديق برزنجي	9.

واقع التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل

أ.م.د. محمد بن سعيد الغافري
جامعة السلطان قابوس - سلطنة عمان
mghafri@squ.edu.om

أحمد فؤاد علوان
جامعة نزوى - سلطنة عمان
Ahmadfouad@unizwa.edu.om

أ.د. جاسم محمد علي التميمي
جامعة ديالى - كلية التربية للعلوم الصرفة - العراق
dr-jasim1956@uodiyala.edu.iq

المخلص :

تهدف الدراسة إلى استكشاف واقع التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل عن طريق التعرف على الممارسة الفعلية للتحول الرقمي في التعليم الجامعي ، لم يعد استخدام تقنية المعلومات وتوظيفها في خدمة الأعمال مجرد خيار لتحسين القدرات ورفع الكفاءات وزيادة الإنتاجية، بل أصبحت أمراً ضرورياً، . تعتمد الدراسة على المنهج الوصفي التحليلي، واستخدام الاستبيان بصفقتها أداة رئيسة لجمع البيانات، ويتم استطلاع وجهات نظر عينة الدراسة من الجنسين وتحدد نتائج هذه الدراسة جزئياً او كلياً بالأدوات و الإجراءات المستخدمة لجمع البيانات وتحليل الموضوعات المتعلقة بقضايا التحول الرقمي في التعليم الجامعي ويتم تصميم استبانته لدراسة التحول الرقمي في التحليل الإحصائي باستخدام برنامج الاكسل وللتحقق من صدق أداة الدراسة المتمثلة بالاستبيان يتم

استخدام صدق المحتوى وللتأكد من ثباتها يستخدم إلفا كرونباخ فضلا عن تحليل المحتوى للوثائق والمستندات في هذا الجانب التي يتم الحصول عليها من عينة الدراسة، لدعم النتائج ووضع المقترحات والتوصيات. وتم تطبيق الدراسة على جامعة بلاد الرافدين.

الكلمات المفتاحية: التحول الرقمي، جامعة بلاد الرافدين، التعليم الرقمي ، مهارات ، التحليل الإحصائي

Abstract

The study aims to explore the reality of digital transformation in statistical analysis skills using Excel by identifying the actual practice of digital transformation in university education. The use of information technology and its employment in business service is no longer just an option to improve capabilities, raise efficiencies and increase productivity, but has become a necessity. The study relies on the descriptive analytical approach, and the use of the questionnaire as a primary tool for collecting data. The views of the study sample of both genders are surveyed and the results of this study are determined partially or completely by the tools and procedures used to collect data and analyze topics related to digital transformation issues in university education. A questionnaire is designed to study digital transformation in statistical analysis using Excel. To verify the validity of the study tool represented by the questionnaire, content validity is used and to ensure its stability, Cronbach's alpha is used in addition to content analysis of documents and papers in this aspect obtained from the study sample, to support the results and develop proposals and recommendations. The study was applied to the University of Mesopotamia.

Keywords: Digital transformation, University of Mesopotamia, digital education, skills, statistical analysis

الفصل الأول

خلفية الدراسة وأهميتها

المقدمة:

القرن الحادي والعشرين" المعروف باسم العصر الرقمي "الذي يتشكل نسيجه في كل منطقة من العالم بسرعة من خلال التطورات و التغيرات المتسارعة في تكنولوجيا المعلومات والاتصالات . وهذا ما يجعلنا نعيش في عالم ذو خصائص دائمة التغير. و يعد التحول الرقمي أحد أهم هذه الابتكارات، ويعد من أهم محفزات النمو للشركات والحكومات التي تتطلع إلى البقاء في الطليعة ، وذلك من خلال تطوير حلول مبتكرة لا تضمن لها البقاء فحسب بل المنافسة أيضا. ونظرا لهذا التطور السريع، لا يمكن للتعليم أن يبقى جامد فانتشرت البرامج الإحصائية ومنها برنامج الإكسل وتزايد استعمالها في البيئات التعليمية المختلفة.

مشكلة الدراسة:

لقد أصبحت التكنولوجيا تلعب دورا أساسياً في الكيفية التي يتعلم بها الناس، وطريقة تفاعلهم مع المعلومات أيضا ، فأصبح التحول الرقمي في التعليم العالي ضرورة في أيامنا الحالية .

تتمثل مشكلة الدراسة التعرف على واقع التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل . وعليه، فإن التحول الرقمي، وما يمر به من تحديات، وخاصة في ظل عمليات التسارع والتراجع المشار إليها، إلى جانب على أن التعليم التقليدي لم يعد كافيا لتحقيق أهداف التعليم خصوصا في عصر التكنولوجيا وقلّة الدراسات في حدود علم الباحثون التي تتحدث عن التحول الرقمي في الجامعات العراقية مما جعل الباحثون البحث " عن مدى ما حققته الحرص على المضي قدماً في مشروع التحول الرقمي في العراق ، فإن العديد من الدراسات والأبحاث في مجال التحول الرقمي في المؤسسات التعليمية تؤكد جميعها مؤسسات التعليم العالي من التحول الرقمي المطلوب؟

هدف الدراسة وأسئلتها:

التعرف على واقع التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل ومن هنا جاءت مشكلة الدراسة للإجابة عن الأسئلة الآتية: السؤال الأول : ما درجة التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل في جامعة بلاد الرافدين من وجهة نظر أعضاء هيئة التدريس ؟ السؤال الثاني : هل هناك فروق ذات دلالة إحصائية عند مستوى الدلالة = 0.05 (a) في تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين تعزى إلى متغيرات (الجنس, الخبرة، اللقب العلمي)؟

أهمية الدراسة:

تتمثل أهمية هذه الدراسة من خلال معرفة واقع التحول الرقمي في جامعة بلاد الرافدين من وجهة نظر أعضاء هيئة التدريس، من خلال الجانبين النظري والعملية وفيما يلي توضيح ذلك:

أولاً : الأهمية النظرية

تنبثق أهمية هذه الدراسة من كونها تتناول موضوع أ مواكب أ للثورة التكنولوجية ومستجدياتها في الميدان التربوي وهو التحول الرقمي في المؤسسات التعليمية كما أنها توفر إطار نظري لواقع التحول في جامعة بلاد الرافدين باستخدام المهارات الإحصائية في برنامج الإكسل وتفتح المجال للباحثين والمهتمين في موضوع البرامج الإحصائية هذا بالإضافة إلى قلة الدراسات - في حدود علم التحول الباحثون - ذات الصلة بموضوع الدراسة مما سيساعد على فهم آلية وواقع الرقمي في قطاع مؤسسات التعليم العالي (الجامعات) في العراق.

ثانياً : الأهمية التطبيقية

ظهرت مجموعة من البرامج الإحصائية للتحول الرقمي التي انعكست على الأدوار الجديدة للإداريين والأكاديميين على حد سواء ومنها البرامج:

ومن MINITAB, SHAZAM, STATA, R, SAS, EXCEL, EViews, هنا جاءت أهمية الوقوف على واقع تطبيقها في مؤسسات التعليم العالي لذا يؤمل من هذه الدراسة أن تساعد أصحاب القرار في مختلف القطاعات في اتخاذ القرارات المناسبة في إجراء تحولات عميقة ومنسقة للقوى العاملة والثقافة والتقنية في إنتاج نموذج التحول الرقمي يحتذي بها.

مصطلحات الدراسة:

يعرف الباحثون المصطلحات الإجرائية حسب الدراسة الحالية كالتالي:

التحول الرقمي في التعليم :

يعرف الباحثون بأنه عملية التخلص من الطرق التقليدية لعملية التدريس واستخدام تطبيق الكلاس روم عبر شبكة الانترنت في العملية التعليمية. (ابراهيم والحداد 2018) ويعرف إجرائياً بأنه : واقع استخدام التقنيات الرقمية الحديثة كالحاسوب والذكاء الاصطناعي والحوسبة السحابية بشكل تفاعلي في الجامعات العراقية وتم قياس واقع استخدامه بقياس الدرجة الكلية لاستجابة عينة الدراسة عن فقرات الاستبيان.

برنامج الإكسل :

هو برنامج إحصائي يستخدم في معالجة البيانات والتحقق من الفرضيات التي يضعها الباحث للاختبارات المعلمية و اللامعلمية (Data Analysis) ويعرف إجرائياً بأنه برنامج إحصائي يساعد أعضاء هيئة التدريس بإدخال البيانات في البرنامج والحصول على نتائج للوصول لاتخاذ القرار في جامعة بلاد الرافدين مع كتاب الإحصاء المقرر للطلبة

الجامعات العراقية :

وتعرف الجامعات بشكل عام بأنها هي المؤسسات التعليمية المكرسة للتعليم العالي والبحث في موضوع معين، وتعد الأساس الذي يصقل معارف ومهارات الطلبة في مجالات مختلفة، وهي تمثل عقل المجتمع لأنها الجدير بالثقة أن تكون المسئول الأول عن تشكيل الهوية الثقافية للطلبة. ويعرف إجرائياً بأنه هي الجامعات الرسمية

والخاصة في العراق (مكان إجراء الدراسة التابعة لوزارة التعليم العالي والبحث العلمي، والتي تشترك بتحقيق أهداف تعليمية تعليمية معينة.
أعضاء هيئة التدريس :

يعرف عضو هيئة التدريس الجامعي بأنه العنصر الأساسي في العملية التعليمية كونه يتعامل مع الطلبة بشكل مباشر مؤثرا في ذلك في تحصيلهم العلمي، وهو بالتالي

يعمل على تطور مؤسسات التعليم بما يحمله من أعباء في رسالته العلمية والعملية مما يؤدي إلى تحقيق أهداف المجتمع (السعيدة (2015) .

ويعرفون أجراءي بأنهم : الأساتذة الذين يمارسون مهنة التدريس في الجامعات العراقية الخاصة والرسمية، ممن يحملون درجات علمية دكتوراه أو ماجستير، في أحد مجالات العلوم وذلك بمختلف الألقاب العلمية (أستاذ، أستاذ مساعد، مدرس ، مدرس مساعد)

حدود الدراسة :

تمثلت حدود الدراسة الحالية في الآتي:

الحدود المكانية : طبقت هذه الدراسة في جامعة بلاد الرافدين

الحدود الزمنية : طبقت هذه الدراسة في الفصل الدراسي الأول للعام الدراسي

2024 \ 2025

الحدود البشرية : اقتصرت هذه الدراسة على أعضاء هيئة التدريس في جامعة بلاد الرافدين.

الحدود الموضوعية : تمثلت في تقديرات أعضاء هيئة التدريس لواقع التحول الرقمي

محددات الدراسة :

تمثلت محددات الدراسة في درجة دقة استجابة أفراد عينة الدراسة عن فقرات أداة الدراسة. وبطبيعة أدواتها ومتغيراتها، إذ يمكن تعميم نتائج الدراسة الحالية على المجتمعات المشابهة، في ضوء مؤشرات صدق أداة الدراسة وثباتها.

أولاً: الإطار النظري:

وقد كان لزاماً على جامعة بلاد الرافدين ان تتخذ جميع الإجراءات اللازمة لاستكمال العملية التعليمية مستغلة التحول الرقمي في التعليم واستخدام برنامج الإكسل في تدريس مادة الإحصاء, وما أتاحه هذا التحول من فرصة للتغلب على ما لحق بالعملية التعليمية من شلل.

التحول الرقمي في التعليم:

إن عملية التحول الرقمي في التعليم بشكل عام يمكن تعريفها على أنها عملية التخلص من الطرق التقليدية لعملية التدريس واستخدام أحدث الصور والوسائل التي ظهرت مع تطور التكنولوجيا والتي تساعد الطالب على تفتيح آفاق تفكيره وخضوعه للتعلم والتجربة بعد أن كان يتلقى دروسه من المعلم ويعتمد عليه اعتماداً كلياً وكانت طريقة التدريس حينها تعتمد على التلقين أما في عملية التحول الرقمي فإنها تعتمد على الفهم والبحث والتجربة والابتكار ويكون هذا وفق إستراتيجية محددة تضعها المؤسسة التعليمية لتسهيل عملية التعليم وفي نفس الوقت الوصول بها إلى مستوى متقدم وحديث مختلف تماماً عن الطرق التقليدية القديمة في التعليم.

وقد بدا جلياً ان التحول الرقمي قد غزا جميع مناحي الحياة خصوصاً بعد الثورة الصناعية والانفجار التكنولوجي ، ويعد قطاع التعليم من اهم القطاعات التي بدأت تتحول نحو التعلم الرقمي حيث بدا من السهل انشاء بيئة تعليمية افتراضية تجمع بين المعلم والمتعلم سواء بالتزامن او بدون تزامن عبر شبكات الانترنت وباستخدام العديد من التطبيقات التي تسهل هذا التواصل. ويمكن توظيف الانترنت كما ذكر (الحناوي (2010) ، عن طريق استخدامه كأداة للتواصل بين المعلمين والادارة، وبين الطلبة والمعلمين، وبين الطلبة والادارة وبين الادارة والمجتمع ، وبين الطالب وزملائه،

وايضا توظيفها واستخدامها كوسيلة واداة تعليمية وايضا كأداة تجمع بين جميع العناصر التعليمية، وكأداة تقييم.

ثانيا: الدراسات السابقة:

- 1- دراسة ماضي وأبو حجير (2020)
- 2 - دراسة البلوشية وآخرون (2020)
- 3- دراسة أمين (2018)
- 4- دراسة (Matas , 2014)
- 5- دراسة (محمود (2018)
- 6- دراسة الغديان (2012)

الفصل الثاني

الطريقة والإجراءات

منهج الدراسة:

استخدم الباحثون المنهج الوصفي المسحي ، لتحقيق أهداف هذه الدراسة حيث تم إجراء مسح لعينة كافية من أعضاء هيئة التدريس لجمع آرائهم وتقديراتهم حول واقع التحول الرقمي في جامعة بلاد الرافدين.

عينة الدراسة:

تم تحديد حجم عينة الدراسة من أعضاء هيئة التدريس وفق المعادلة التالية ". وذلك على فرض أن مستوى الثقة المطلوب توافره في البيانات هو 95 ونسبة خطأ

$$[N = N / [1+N+ (\alpha)^2] (0.05)$$

بحيث: (0.05) تمثل حجم العينة N تمثل مجتمع الدراسة. نسبة الخطأ المسموح به حيث تم بعدها تحديد عدد أعضاء هيئة التدريس المشاركين بالطريقة العشوائية وذلك بناء على عددهم في مجتمع الدراسة بلغ (316) وفيما يلي عرض للخصائص

الديموغرافية للمشاركين في الدراسة تبعاً للمتغيرات الدراسة كما موضح في جدول (2).

الجدول (2)

توزيع عينة الدراسة من أعضاء هيئة التدريس في جامعة بلاد الرافدين تبعا
لمتغيرات الدراسة

المتغير	أبعاد المتغير	التكرار	النسبة المئوية%
	ذكور	190	60%
الجنس	إناث	126	40%
	أقل من 5 سنوات	71	23%
الخبرة	5 - أقل من 10 سنوات	80	25%
	10 - أقل من 15 سنوات	115	36%
	15 سنة - فأكثر	50	16%
	مدرس مساعد	61	19%
اللقب العلمي	مدرس	90	28%
	أستاذ مساعد	105	33%
	أستاذ	60	20%
	المجموع	316	100%

أدوات الدراسة:

حيث تكون الاستبيان في صورتها الأولية من جزئيين، شمل الجزء الأول على البيانات الديموغرافية للمشاركين من خلال الإجابة باختيار من متعدد . وشمل الجزء الثاني على محورين التحول الرقمي (14) فقرة و المهارات الإحصائية باستخدام برنامج الإكسل(6) فقرات حيث كانت الإجابات على الفقرات من خلال مقياس ليكرت الرباعي (أوافق تماما أوافق إلى حد ما لا أوافق إلى حد ما- لا أوافق تماما).

أولاً: أن قيم معامل الارتباط بين التطبيقين الأول والثاني لفقرات استبيان أعضاء هيئة التدريس تراوحت ما بين (0.809**) و (0,856**) في حين كانت قيمة معامل الارتباط الكلي للاستبيان بين التطبيقين الأول والثاني. (**0.851) كما موضح في جدول(3)

جدول (3)

معامل ارتباط بيرسون بين التطبيقين الأول والثاني لأداة الدراسة

المتغيرات	معامل الارتباط	مستوى الدلالة
التحول الرقمي	0.809**	0.000
المهارات الإحصائية	0.856**	0.000
الدرجة الكلية	0.851**	0.000

ثانياً : طريقة الاتساق الداخلي باستخدام معادلة كرونباخ إلفا بعد تطبيق الأدوات على العينة الاستطلاعية تم حساب الثبات بطريقة الاتساق الداخلي وفق معادلة كرونباخ إلفا والجدول (4) يوضح معاملات الثبات والدرجة الكلية.

جدول (4)

معاملات الثبات لأداة الدراسة وفق طريقة الاتساق الداخلي (كرونباخ إلفا)

المتغيرات	معامل الارتباط	مستوى الدلالة
التحول الرقمي	0.756**	0.000
المهارات الإحصائية	0.748**	0.000
الدرجة الكلية	0.816**	0.000

بين جدول (4) أن قيم معاملات الثبات وفق كرونباخ إلفا لمتغيرات الدراسة والدرجة الكلية تراوحت بين (0.748**) و (0.756) في حين كانت قيمة الثبات إلفا للدرجة الكلية (0.816) .

وقد تم تصميم الاستجابة على الاستبيان وفق التدرج الرباعي حسب نموذج ليكرت وكما يلي:

- (أوافق تماما) ولها (4) درجات.
 - أوافق إلى حد ما (ولها (3) درجات. لا أوافق إلى حد ما (ولها (2) درجات.
 - (لا أوافق تماما) ولها (1) درجات.
- وتم حساب طول الفئة من خلال قسمة الفرق بين الحد الأعلى للبدائل (4) والحد الأدنى للبدائل (1) على (3) مستويات (كبيرة, متوسطة, قليلة) كما هو مبين في المعادلة: $1=3/(1-4)$
- وهكذا تصبح أوزان الفقرات كما يلي:
- من (1) الى اقل من (2) درجة قليلة
 - من (2) الى اقل من (3) درجة متوسطة
 - من (3) الى (4) درجة كبيرة

متغيرات الدراسة:

- .المتغيرات المستقلة: لأعضاء هيئة التدريس
- . اللقب العلمي : (مدرس مساعد, مدرس, أستاذ مساعد, أستاذ)
- . الخبرة : أقل من 5 سنوات من 5 أقل من 10 سنوات، من 10 أقل من 15, من 15 فأكثر
- . التحول الرقمي
- .المتغير التابع : المهارات الإحصائية باستخدام برنامج الإكسل

الوسائل الإحصائية:

- للإجابة عن أسئلة الدراسة تم استخدام الأساليب الإحصائية الآتية:
- 1- معامل ارتباط بيرسون ومعادلة كرونباخ ألفا: لحساب الثبات والاتساق الداخلي
- 2- المتوسطات الحسابية والانحرافات المعيارية والرتب : للكشف عن درجة التحول الرقمي في جامعة بلاد الرافدين من وجهة نظر أعضاء هيئة التدريس
- 3- اختبار " ت " للعينات المستقلة: للكشف عن الفروق تبعا لمتغير الجنس

4- تحليل التباين الأحادي : للكشف عن الفروق في تقديرات أعضاء هيئة التدريس تبعاً لمتغيرات (اللقب العلمي , الخبرة) .

إجراءات تطبيق الدراسة:

التحقيق أهداف الدراسة ، تم إتباع الخطوات و الإجراءات الآتية: الاطلاع على الإطار النظري والدراسات السابقة ذات العلاقة بمشكلة البحث. . تطوير أداة الدراسة بصورتها النهائية بعد التحقق من صدقهما وثباتهما، وتظهر في الملحق (3) تحديد مجتمع الدراسة، وعينته (عينة عشوائية) من أعضاء الهيئة التدريسية الحصول على كتاب تسهيل مهمة من جامعة بلاد الرافدين إلى الجهات المعنية، ويظهر في الملحق (4) . توزيع أداة الدراسة الكترونياً على أفراد العينة. جمع البيانات ثم تحليلها باستخدام برنامج الإكسل لتحليل البيانات و استخراج النتائج. رصدها ومناقشتها وتقديم التوصيات.

الفصل الثالث

عرض النتائج وتفسيرها

النتائج المتعلقة بالسؤال الأول:

ما درجة التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل في جامعة بلاد الرافدين من وجهة نظر أعضاء هيئة التدريس ؟ للإجابة عن هذا السؤال تم احتساب المتوسط الحسابي الانحراف المعياري والترتيب لتقديرات أعضاء هيئة التدريس على فقرات الاستبيان لدرجة التحول الرقمي في مهارات التحليل الإحصائي باستخدام برنامج الإكسل في جامعة بلاد الرافدين من وجهة نظر أعضاء هيئة التدريس وكانت النتائج في جدول (5).

الجدول (5)

المتوسط الحسابي للانحراف المعياري والترتيب لتقديرات أعضاء هيئة التدريس
 على فقرات الاستبيان لدرجة التحول الرقمي

رقم الفقرة	الفقرات	المتوسط	الانحراف المعياري	الترتيب	الدرجة
3	تشارك الجامعة في ندوات ودورات تدريبية تتعلق بالبرامج الإحصائية	2.99	0.791	1	متوسط
13	تعمل الجامعة على تحقيق التربية الرقمية داخل حرمها	2.82	0.963	2	متوسط
17	توفر الجامعة نظام الحوافز عند استخدام برنامج الاكسل في التحليل الإحصائي	2.81	0.887	3	متوسط
8	توفر الجامعة الخدمات التقنية والفنية لمواجهة أي طارئ في برامجها	2.71	0.907	4	متوسط
19	يتم استخدام برنامج الاكسل للتحليل الإحصائي بالعمل داخل الجامعة	2.71	0.924	5	متوسط
5	تدعم الجامعة التعاون والإبداع الرقمي عند أعضاء هيئة التدريس	2.6	0.901	6	متوسط
4	تعمل الجامعة على تحسين وتطوير عملية التعليم الرقمي	2.57	0.876	7	متوسط
16	توفر الجامعة قادة متميزين بقدرتهم على وضع برنامج الاكسل مناسب للتحول الرقمي	2.55	0.95	8	متوسط
6	توفر الجامعة ميزانية مناسبة لتدريب أعضاء هيئة التدريس في التحول الرقمي	2.54	0.977	9	متوسط
1	تعمل الجامعة على إنشاء ونشر ثقافة التحول الرقمي.	2.51	0.941	10	متوسط

متوسط	11	0.994	2.48	توفر الجامعة البنية التحتية التكنولوجية المناسبة للتحويل الرقمي	7
متوسط	12	0.986	2.45	تعمل الجامعة على أشراك أعضاء هيئة التدريس بأهدافها نحو التحويل الرقمي	14
متوسط	13	0.962	2.43	تقوم الجامعة بتوفير الخطط البديلة لمتابعة العملية التعليمية أثناء الأزمات	9
متوسط	14	0.95	2.41	تقوم الجامعة باستمرار بتقييم الفوائد العائدة لها من برنامج الاكسل	15
متوسط	15	0.954	2.39	تقوم الجامعة بتقديم الإرشادات الخاصة بالتحويل الرقمي لأعضاء هيئة التدريس	2
متوسط	16	0.926	2.3	توفر الجامعة منصة تعليمية جيدة تمكن أعضاء هيئة التدريس من استخدامها	10
متوسط	17	0.969	2.27	تعمل الجامعة على تطوير القدرات والمهارات الرقمية باستخدام الاكسل	20
متوسط	18	1.028	2.25	تدعم الجامعة أمن وسلامة البيانات الرقمية الخاصة بأعضاء هيئة التدريس	11
متوسط	19	1.005	2.2	تحفز الجامعة أعضاء هيئة التدريس على استخدام برنامج الاكسل	18
متوسط	20	0.993	2.13	تستفيد الجامعة من التغذية الراجعة في تطوير تجربتها الرقمية	12
متوسط		0.446	2.51	المتوسط العام	

كما تبين أن واقع التحويل الرقمي كان متوسطا حيث تراوحت المتوسطات الحسابية لواقع التحويل الرقمي بين (2-13-2.99) بمتوسط كلي مقداره (2.51) كما هو موضح بالجدول (5). إذ جاءت في المرتبة الأولى فقرة (3) "تشارك الجامعة في ندوات ودورات تدريبية تتعلق بالبرامج الإحصائية " بمتوسط حسابي بلغ (2.99)

وهو أعلى من المتوسط الحسابي العام البالغ (2.51)، وانحراف معياري بلغ (0.791)، فيما حصلت الفقرة (12) "تستفيد الجامعة من التغذية الراجعة في تطوير تجربتها الرقمية" على المرتبة العشرين والأخيرة بمتوسط حسابي (2.13) وهو أدنى من المتوسط الحسابي الكلي والبالغ (2.51) وانحراف معياري (0.993). وبشكل عام يتبين أن واقع التحول الرقمي من وجهة نظر عينة الدراسة كان متوسطاً.

السؤال الثاني : هل هناك فروق ذات دلالة إحصائية عند مستوى الدلالة - (0.05) في تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين

تعزى إلى متغيرات (الجنس الخبرة، اللقب العلمي) ؟

أ- الجنس للكشف عن دلالة الفروق في تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين تبعاً لمتغير ذكور (إناث) تم استخراج المتوسطات الحسابية والانحرافات المعيارية لتقديرات أعضاء هيئة التدريس على استبانة الدراسة ، كما تم استخدام اختبار ت للعينات المستقلة وكانت النتائج كما في الجدول (6).

الجدول (6)

نتائج اختبار "ت" للكشف عن دلالة الفروق في تقديرات أعضاء هيئة التدريس

الدرجة التحول الرقمي في جامعة بلاد الرافدين تبعاً لمتغير الجنس

الجنس	العدد	المتوسط الحسابي	الانحراف المعياري	قيمة ت المحسوبة	مستوى الدلالة Sig
ذكور	190	50.21	8.897	0.213	0.832
إناث	126	49.99	9.004		

يبين جدول (6) ان مستوى الدلالة (0.832 اكبر من (0.05) هذا يعني لا يوجد فرق معنوي عند (0.05) تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين تبعا لمتغير الجنس.

ب متغير الخبرة: (أقل من 5 سنوات 5 أقل من 10 سنوات، 10 أقل من 15 سنة، 15 سنة فأكثر).

للكشف عن دلالة الفروق في تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة ديالى تبعا لمتغير الخبرة (أقل من 5 سنوات، 5 أقل من 10 سنوات، 10 أقل من 15 سنة، 15 سنة فأكثر تم استخراج المتوسطات الحسابية و الانحرافات المعيارية، لتقديرات أعضاء هيئة التدريس على استبانته الدراسة، في ضوء متغير الخبرة، وكانت النتائج كما في الجدول (7).

جدول (7)

المتوسطات الحسابية و الانحرافات المعيارية، لتقديرات أعضاء هيئة التدريس تبعا لمتغير الخبرة

الانحراف المعياري	المتوسط الحسابي	العدد	متغير الخبرة
8.47	47.83	71	أقل من 5 سنوات
7.166	49.64	80	5-أقل من 10 سنوات
10.398	50.81	115	10-أقل من 15 سنة
7.754	52.58	50	15 سنة فأكثر
8.926	50.12	316	المجموع

تشير المتوسطات الحسابية في الجدول (7) إلى وجود فروق ظاهرية في المتوسطات الحسابية لتقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين تبعا لمتغير الخبرة، والمعرفة مستوى الدلالة الإحصائية للفروق في المتوسطات الحسابية، تم استخدام تحليل التباين الأحادي (One Way ANOVA) وكانت النتائج كما هو موضح في الجدول (8)

جدول (8)

نتائج تحليل التباين الأحادي للكشف عن دلالة الفروق في تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين، تبعا لمتغير الخبرة

المتغير	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "ف" المحسوبة	مستوى الدلالة Sig
الخبرة	بين المجموعات	747.8	3	249.3	3.194	0.024
	داخل المجموعات	24350	312	78.05		
	المجموع	25098	315			

تشير النتائج في الجدول (8) الى وجود فروق ذات دلالة إحصائية في تقديرات أعضاء هيئة التدريس في جامعة بلاد الرافدين لدرجة التحول الرقمي تبعا لمتغير الخبرة لان مستوى الدلالة بلغت (0.024) أقل من (0.05).

ج- متغير اللقب العلمي: (مدرس مساعد مدرس أستاذ مساعد, أستاذ) للكشف عن دلالة الفروق في تقديرات أعضاء هيئة التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين تبعا لمتغير اللقب العلمي مدرس مساعد, مدرس, أستاذ مساعد, أستاذ) تم استخراج المتوسطات الحسابية والانحرافات المعيارية، لتقديرات أعضاء هيئة التدريس على استبانته الدراسة في ضوء متغير اللقب العلمي، وكانت النتائج كما في الجدول (9).

الجدول (9)

المتوسطات الحسابية والانحرافات المعيارية، لتقديرات أعضاء هيئة التدريس
 لدرجة التحول الرقمي في جامعة بلاد الرافدين تبعا لمتغير اللقب العلمي

الانحراف المعياري	المتوسط الحسابي	العدد	متغير اللقب العلمي
7.927	51.89	61	مدرس مساعد
8.765	47.94	90	مدرس
9.79	51.86	105	أستاذ مساعد
7.619	48.52	60	أستاذ
8.926	50.12	316	المجموع

جدول (10)

نتائج تحليل التباين الأحادي للكشف عن دلالة الفروق في تقديرات أعضاء هيئة
 التدريس لدرجة التحول الرقمي في جامعة بلاد الرافدين، تبعا لمتغير اللقب العلمي

المتغير	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة "ف" المحسوبة	مستوى الدلالة Sig
اللقب العلمي	بين المجموعات	1097.7	3	365.89	4.756	0.003
	داخل المجموعات	24001	312	76.925		
	المجموع	25098	315			

تشير النتائج في الجدول (10) الى وجود فروق ذات دلالة إحصائية في تقديرات
 أعضاء هيئة التدريس في جامعة بلاد الرافدين لدرجة التحول الرقمي تبعا لمتغير
 اللقب العلمي لأن مستوى الدلالة بلغت (0.003) أقل من (0.05).

التوصيات:

- في ضوء نتائج الدراسة يوصي الباحثون بما يلي:
- 1- الاهتمام بالدعم المؤسسي للتحول الرقمي في الجامعات من خلال الاهتمام بالتطوير والإبداع المستمر
 - 2- نشر الثقافة الرقمية من خلال نشر ميثاق أخلاقي للتعامل مع التحول الرقمي في الإدارة.
 - 3- تطوير الشراكات الداخلية والخارجية وخاصة في الجامعات الحكومية.
 - 4- الاستفادة من التجارب العالمية في مجال التحول الرقمي في المؤسسات التعليمية ونقل خبراتها بما يتناسب مع مجتمعاتنا.
 - 5- إجراء بحوث علمية في مجال التحول الرقمي في قطاع المدارس في العراق.

مراجع الدراسة:

1. أمين مصطفى أحمد . (2018) التحول الرقمي في الجامعات المصرية كمتطلب لتحقيق مجتمع المعرفة مجلة الإدارة التربوية، (19) 1165 - 11
2. البلوشية نوال بنت علي الحرا صياء نبهان بن حارث، العوفي، علي بن سيف (2020) واقع التحول الرقمي في المؤسسات العمانية مجلة دراسات المعلومات والتكنولوجيا عدد (1).
3. التميمي جاسم محمد، (2021) المهارات الإحصائية باستخدام برنامج SPSS
4. , مؤسسات الوراق للطباعة والنشر , عمان, الأردن.
5. الحناوي، حامد (2010 دور كتاب التكنولوجيا للصف الثاني عشر في اكساب الطلبة بعض المعايير العالمية لتكنولوجيا المعلومات. رسالة ماجستير غير منشورة، كلية التربية الجامعة الإسلامية، غزة، فلسطين.
6. السعيدة , محمد جلال (2015) مهارات التدريس الجامعي التي ينبغي توافرها لدى أعضاء هيئة التدريس في جامعة البلقاء التطبيقية من وجهة نظر طلبتهم, رسالة ماجستير غير منشورة جامعة الشرق الأوسط.

7. ماضي , خليل إسماعيل وابو حجير طارق مفلح (2020) مدى جاهزية الجامعات الفلسطينية نحو التحول الرقمي [بحث مقدم .] المؤتمر الدولي الأول في تكنولوجيا المعلومات والأعمال (ICITB2020).
8. محمود ، ولاء (2018) مقومات تنمية الموارد البشرية الأكاديمية جامعة بنها في العصر الرقمي الواقع وسيناريوهات المستقبل، بحث منشور في مجلة كلية التربية، جامعة كفر الشيخ العدد (90) العدد الأول 19 ، المجلد الثاني، السنة(2018م).
9. الغديان ,عبد المحسن عبد الرازق (2012) : تصور مقترح للتعليم المتنقل (M-Leaming) في مؤسسات التعليم العالي، مجلة مستقبل التربية العربية المركز العربي للتعليم والتنمية المجلد (20)، العدد (82).
10. Matas J. (2014): The Impact of Digital Education on Learning and Teaching, Doctor Degree, The School of Education College of Professional Studies, Northeastern University, Boston, Massachusetts, USA

مقارنة بعض مؤشرات زيادة الوزن والسمنة في المسحين العنقوديين متعدد المؤشرات (MICS3) و(MICS4) للعامين 2006 و 2011

أ.د. سجي محمد حسين

ظافر طارق

قسم الاحصاء /كلية الادارة والاقتصاد /جامعة بغداد/العراق

باحث احصائي/ العراق

المستخلص

هناك الكثير من المشاكل التي تعاني منها البلدان وهي سوء التغذية مع ارتفاع معدلات السمنة لدى الاطفال حيث نجد بان عدد الاطفال الذين يعانون من زيادة الوزن وفرط السمنة في البلدان المنخفضة والمتوسطة الدخل قد اخذ بالازدياد . وبسبب التغيرات التي طرأت على بلدنا العراق في النواحي المعيشية والتغذوية والتي ساهمت في الزيادة على الاقبال على تناول الطعام والاعذية دون الانتباه الى مضارها .لقد كان الاهتمام في هذا البحث على مقارنة مؤشري زيادة الوزن والسمنة لاطفال العراق (دون سن الخامسة) في المسح العنقودي متعدد المؤشرات (MICS3) الذي نفذ في عام 2006 و المسح العنقودي متعدد المؤشرات (MICS4) الذي نفذ في عام 2011 على مستوى المحافظات والفئات العمرية والمستوى التعليمي للام والجنس (ذكر وانثى) والمنطقة (حضر وريف) باستخدام اختبار T لمقارنة متوسطين وتحليل التباين باتجاه واحد لمقارنة اكثر من متوسطين . وتبين من خلال النتائج بانه هناك فروق معنوية ذات دلالة احصائية بين متوسطي هذين المسحين للعامين 2006 و 2011 من ناحية الوزن الى الطول , وان هناك فروقا معنويا بين نسبتي مؤشري زيادة الوزن والسمنة للعامين 2006 و 2011.

Comparison of some indicators of overweight and obesity in the Multiple Indicator Cluster Surveys (MICS3) and (MICS4) for the years 2006 and 2011

**Saja mohammed Hussein ,prof. P.HD
Zafer Tariq**

**Dept. of statistics /college of Administration and Economics
Statistical Researcher /Iraq
University of Baghdad**

Abstract

There are many problems that countries are experiencing, namely malnutrition with high rates of childhood obesity. The number of overweight and obese children in low- and middle-income countries is increasing. Because of the changes that have taken place in our country, Iraq in the aspects of living and nutrition, which contributed to the increase in the demand for food without paying attention to its harm. The interest in this research was to compare the indicators of overweight and obesity of Iraqi children (under five years) in the MICS3, which was carried out in 2006 and the MICS4, which was carried out in 2011 at the governorate level and age groups. The educational level of the mother, sex (male and female) and region (urban and rural) using the T test to compare two mean And one-way variance analysis to compare more than two averages. The results showed that there were significant differences between the average of these two surveys for the years 2006 and 2011 in terms of weight to height.

1-المقدمة وهدف البحث

1-1 المقدمة [13,11,3,2,1]

ان ظاهرة زيادة الوزن والسمنة لدى الاطفال اصبحت ظاهرة ملحوظة وتثير القلق في كثير من البلدان ومنهم العراق حيث تعتبر زيادة الوزن بشكل كبير والسمنة وخصوصا عند الاطفال خطرا على صحتهم , وبما ان من اهم اهداف التنمية

المستدامة التي وضعتها الأمم المتحدة هي القضاء على الأمراض الغير سارية وبما ان السمنة هي احدى العوامل للاصابة ببعض الامراض ولما لها من نتائج مستقبلية على صحة الانسان لذا كان من الضروري الحد من هذه الظاهرة وبدا من مرحلة الطفولة لانها تؤثر على صحة الطفل بشكل مباشر وعلى تحصيله الدراسي وعلى حياته ويمكن ان تعرضه للاصابة ببعض الامراض المزمنة مستقبلا.

الاطفال في مرحلة النمو يزيد طولهم وطول الطفل مقارنة الى طول الطفل المثالي في نفس العمر يعطي مؤشرا على كون النمو طبيعيا أم لا. هذا ويسمى مؤشر النمو الطول بالنسبة للعمر. في الحالة التي يزداد فيها طول الطفل ووزنه في الوقت نفسه فانه يعد طبيعيا إذا فان نسب الجسم تبقى طبيعية اما الطفل البدين فان وزنه يزداد وطوله يقل عن الاطفال الاخرين على نفس مستوى الطول . مقياس الوزن بالنسبة للطول هو مقياس لمدى نحف او بدانه الطفل لأن زيادة الوزن أو فقدانه هي مؤشر على استجابة الطفل للوضع الراهن وعادة ما تؤخذ قياسات الوزن بالنسبة للطول لأنها تعكس الظروف الغذائية الأخيرة.

فمن واجب العوائل والعاملين في السلك الصحي الى ممارسة السلوك الذي يدعم النمو الناجح (في كسب الطول والوزن) لدى الاطفال دون سن الخامسة ويشمل هذا السلوك مفهوم المغذيات الدقيقة قبل الولادة للامهات لبناء أساس قوي لنمو الطفل الرضيع، والأغذية والرضاعة التكميلية، والقيام بعناية صحية رصينة وتحديد التوقيت المناسب لأعطاء الوجبات ومراقبة الأمراض بصورة جيدة واي أشياء أخرى. ان عملية مراقبة النمو هو من أجل قياس الوزن والطول للاطفال بصورة دورية (مثلا" شهريا") لرؤية فيما اذا نمو الاطفال بصورة جيدة. يعد هذا ممارسة سريرية جيدة لمراقبة نمو الاطفال ويمكن ان تساعد في تشخيص المشاكل الطبية ذات العلاقة قبل ان تصبح جدية وكذلك تعزيز ممارسات العناية الجيدة.

1-2 الاسباب المؤدية الى زيادة الوزن عند الاطفال [14,4]

ان زيادة وزن الطفل في مراحله الاولى تكون جدا مقلقة ويمكن ان نعزو اسباب زيادة الوزن والسمنة عند الاطفال الى :

- 1- الاسباب الاجتماعية والاقتصادية حيث ان العوائل المحدودة الدخل والفقيرة لاتستطيع توفير الغذاء الصحي لاطفالهم، او يعتمدون على بعض انواع الاغذية التي تسبب البدانة او ان هناك بعض العوائل معتادة على تناول الاطعمة ذات السعرات الحرارية العالية وبالتالي فهي تشجع اطفالها على تناول مثل هذا النوع من الاغذية
- 2-عوامل وراثية : قد ترتبط السمنة بمرض وراثي له علاقة بالغدد الصماء، او اذا كان احد الوالدين او كلاهما يعاني من السمنة فهناك فرصة بان اطفالهم سوف يعانون منها ايضا.
- 3- قلة الحركة والنشاط البدني

1-3 كيفية الوقاية من زيادة وزن الاطفال [14.4]

تتمكن العائلة من الحفاظ على الوزن الملائم للطفل من خلال اتباع مايلي:

- 1 -السماح للطفل بالتوقف عن تناول الطعام عند احساسه بالشبع وعدم اجباره على انهاء طعامه كله
 - 2 -عدم تعويد الطفل على مكافئته باعطائه الحلوى والطعام وعدم منعه من تناول الحلويات بشكل قاطع مما يولد لديه الرغبة الدائمة في تناول الحلويات.
 - 3 -تعليم الطفل وتعييده على تناول الاطعمة الصحية والاكثر من الفواكه والخضروات وتجنب المشروبات والعصائر المحلاة.
 - 4-توجيه الطفل نحو اللعب والتي تؤدي الى الحركة وزيادة النشاط البدني
- هناك بعض الدراسات التي تناولت موضوع زيادة الوزن والسمنة منها الدراسة التي قام بها العرجان^[5] فكان هدفها التعرف على نسبة انتشار البدانة والوزن الزائد لدى طلبة جامعة البلقاء التطبيقية وقد استخدم الباحث اختبار الاستقلالية (مربع كاي) بين متغيرات الجنس والاقليم وعلاقتها ببعض المتغيرات الاخرى .
- ودراسة قام بها السالم^[6] حول رصد السلوكيات الحياتية لدى المراهقين لطلبة المرحلة الثانوية بمدينة الرياض وعلاقته بزيادة الوزن والسمنة من خلال استخدام اختبار الاستقلالية (مربع كاي) حيث تبين بان هناك فروقا جوهرية بين عدة متغيرات من

نمط الحياة ومؤشر زيادة الوزن والسمنة. وكذلك استنتج الباحث بان ارتفاع المستوى الاجتماعي والاقتصادي له علاقة ارتباط بزيادة الوزن والسمنة .
اما الدراسة التي قام بها الاميري^[7] واخرون حول السمنة عند رياض الاطفال وعلاقتها ببعض المتغيرات حيث استعمل تحليل التباين الاحادي واختبار العينتين المستقلتين ومعامل الارتباط وتوصلت الدراسة الى الاستنتاجات وهي ان الاناث يعانين من السمنة اكثر من الذكور ولا توجد فروق احصائية بين متوسط درجات مؤشر كتلة الجسم للاطفال الذين يعانون من السمنة والذي يعزى الى متغير تسلسل الطفل ضمن اخوته وان هناك علاقة بين مؤشر كتلة الجسم للاطفال المصابين بالسمنة ومتغير العادات الغذائية .

1-4 الهدف: يهدف البحث الى المقارنة بين بعض مؤشرات زيادة الوزن والسمنة وذلك من خلال المقارنة بين متوسطات المحافظات ومتوسطات الفئات العمرية ومتوسطات المستوى العلمي للام والجنس والمنطقة والمقارنة بين متوسطي الوزن الى الطول وبين نسبي زيادة الوزن والسمنة للعامين 2006 و2011 بالاعتماد على بيانات المسح العنقودي متعدد المؤشرات الثالث والرابع (MICS-3 و MICS-4).
2- الجانب النظري: يتم في هذا المبحث عرض اهم المقاييس الاحصائية المستعملة وهي:

1-2 اختبار T (Independent Samples T-Test) [9,8]

الهدف من إجراء هذا الاختبار هو اختبار معنوية الفرق بين متوسطي العينتين وفق الفرضية الاتية:

$$H_0 : \mu_1 = \mu_2$$

$$\mu_1 \neq \mu_2 \quad \dots(1)$$

$$H_1 :$$

ان احصاء الاختبار هي:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} \quad \dots(2)$$

حيث \bar{X} هي متوسط العينة: $\bar{X} = \frac{\sum X}{n}$

الانحراف القياسي للفرق بين المتوسطين ويحسب كالاتي: $S_{\bar{X}_1 - \bar{X}_2}$

$$S_{\bar{X}_1 - \bar{X}_2} = Sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S^2 p = \left[\sum X_1^2 - \frac{(\sum X_1)^2}{n_1} \right] + \left[\sum X_2^2 - \frac{(\sum X_2)^2}{n_2} \right]$$

ترفض فرضية العدم اذا كانت t المستخرجة اكبر من t الجدولية وبدرجة حرية n-2
وبمستوى معنوية $\alpha/2$

2-2 تحليل التباين في اتجاه واحد One Way Analysis of Variance

[16,10,8]

يستعمل تحليل التباين لاختبار الفرضية التالية :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \text{يوجد متوسطان على الاقل غير متساويين} \quad \dots(3)$$

وصيغة الاختبار :

$$F = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{X}_{i.} - \bar{X}_{..})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2 / (N-k)} \quad \dots(4)$$

حيث ان X_{ij} : تمثل قيمة المفردة z الواقعة ضمن العينة او المجموعة i

$\bar{X}_{i.}$: تمثل متوسط قيم المفردات الواقعة ضمن المجموعة i

$\bar{X}_{..}$: المتوسط العام لجميع قيم المفردات، n_i : تمثل عدد المفردات ضمن المجموعة i
إذا كانت قيمة F المحسوبة اكبر من القيمة الجدولية بدرجة حرية $(k-1, N-k)$
(k بمستوى معنوية α اذن نرفض فرضية العدم أي يوجد فروق معنوية متوسطات المعالجات).

3-2 طريقة أقل فرق معنوي Fisher's Least Significant Difference

[10.8]

ان السؤال الذي يبقى قائماً هو " أي مستوى من مستويات المعالجة يختلف عن الآخرين؟ أو بمعنى آخر أين توجد الفروق الحقيقية؟ إذا أشارت نتائج تحليل التباين (اختبار F) إلى أن هناك فرقاً ذا دلالة يعزى إلى مستويات المعالجة.
للإجابة على هذا السؤال يكون من خلال تطبيق احد طرائق المقارنات البعدية، وقد تم استعمال طريقة أقل فرق معنوي (LSD) وتعرف بـ ($Fisher's LSD$) فإذا كان الفرق بين متوسطي مجموعتين (A و B) مثلاً أكبر اويساوي (LSD) المحسوبة فترفض فرض العدم واذا كان الفرق بين متوسطين اصغر من (LSD) المحسوبة فلا نستطيع رفض فرضية العدم، والصيغة الرياضية لحساب LSD هي:

$$LSD_{A,B} = t_{(\alpha/2, (N-K))} \sqrt{MSE(1/n_1 + 1/n_2)} \quad \dots(5)$$

حيث MSE متوسط مجموع المربعات (التباين) داخل المجموعات

4-2 اختبار الفرق بين نسبتين Testing the difference between two Proportion [15]

يستعمل هذا الاختبار لاختبار الفرضية الاتية:

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2 \quad \dots(6)$$

وان احصاءة الاختبار هي:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}\hat{q}[1/n_1 + 1/n_2]}} \quad \dots(7)$$

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2}, \quad \hat{P}_2 = \frac{x_2}{n_2}, \quad \hat{P}_1 = \frac{x_1}{n_1} \quad \text{حيث}$$

حيث $x_1, x_2 =$ عدد قرارات النجاح في العينتين اللتان حجمهما n_1, n_2 ،
فاذا كانت Z المحسوبة اقل من قيمة Z الجدولية تقبل فرضية العدم عند مستوى دلالة معين والعكس صحيح.

3- الجانب التطبيقي

3-1 عينة البحث: تم الحصول على البيانات من المسح العنقودي المتعدد المؤشرات (MICS3) الذي نفذ في عام 2006 والمسح العنقودي المتعدد المؤشرات (MICS4) الذي نفذ في عام 2011 ويعد المسحان من المسوح التي قام به الجهاز المركزي للإحصاء وتكنولوجيا المعلومات وهيئة إحصاء إقليم كردستان بالتعاون مع وزارة الصحة وبدعم مالي وفني من منظمة الصحة العالمية للطفولة (اليونيسيف).ونفذت وعلى نطاق واسع وتمثل تمثيلاً حقيقياً للمجتمع لاسيما شمول عينة كبيرة من الاسر جرى اختيارها عشوائياً من كل محافظات العراق بما في ذلك إقليم كردستان وعليه فان كل طفل في أسرة تقيم في العراق ضمن الحدود الجغرافية

له احتمال الظهور ضمن الاطفال المختارة في المسح فكانت حجم العينة لعام 2006 (18144) اسرة وفي عام 2011 تم مضاعفة العدد فكانت حجم العينة (36850) اسرة. وتم استخدام البرنامج الإحصائي (SPSS version 20.0) [12] في تحليل البيانات .

3-2 تحليل النتائج

1. تشير نتائج جدول رقم (1) جدول الزيادة في الوزن والسمنة الى ان محافظتي (ذي قار و القادسية) كانت نسبة الزيادة في الوزن فيهما الاعلى حيث بلغتا %10.1 و%9.9 على التوالي في عام 2006 وقد انخفضتا الى %3.5 و %5.3 على التوالي في عام 2011. وكانت محافظة السليمانية اقل نسبة زيادة في الوزن بمقدار %1.5 وقد ازدادت الى %4.3 في عام 2011. اما نتائج السمنة عند اطفال العراق دون خمس سنوات فقد بلغت محافظة الانبار اعلى نسبة %9 في عام 2006 و ايضا انخفضت في عام 2011 الى %5.1. ونلاحظ المحافظات (دهوك والسليمانية وديالى وبابل والانبار و نجف) الى ارتفاع نسب الزيادة في الوزن في عام 2011 عن عام 2006 وشهدت نينوى ثبات النسبة للعامين المذكورين اما باقي المحافظات الاخرى فنلاحظ انخفاض النسب في عام 2011 عن عام 2006. اما بالنسبة لمؤشر السمنة فكانت المحافظات (دهوك والسليمانية وديالى ونجف) ايضا الى ارتفاع في النسب في العام 2011 عن عام 2006، وسائر المحافظات على العكس من ذلك.

جدول (1) يوضح اعداد ونسب الزيادة في الوزن والسمنة حسب المحافظات
 للعامين 2006 و 2011

المحافظة	الزيادة في الوزن 2006		الزيادة في الوزن 2011		السمنة 2006		السمنة 2011	
	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال
Dohuk	2.6%	25	7.4%	158	1.6%	16	2.9%	62
Nainwa	5.5%	58	5.5%	145	2.9%	30	2.5%	65
Sulimania	1.5%	7	4.3%	107	.9%	4	1.3%	33
Karkuk	5.8%	39	4.0%	54	1.3%	9	.8%	11
Erbil	4.1%	30	3.3%	62	2.3%	17	1.7%	32
Diala	7.1%	45	11.5%	182	5.0%	32	7.4%	117
Anbar	8.4%	60	9.6%	264	9.0%	63	5.1%	142
Babil	5.6%	45	8.8%	212	5.4%	43	3.6%	87
Karbala	4.2%	41	3.2%	44	2.8%	27	1.6%	22
Wasit	5.2%	47	3.4%	32	4.0%	36	1.1%	10
Salahaladin	4.6%	46	3.8%	72	2.8%	28	1.5%	29
Najaf	4.8%	44	7.3%	204	2.5%	23	3.5%	99
Qadisya	9.9%	91	5.3%	52	5.5%	51	3.3%	32
Almuthana	6.2%	68	2.4%	38	4.3%	47	1.6%	26
Thiqar	10.1%	84	3.5%	60	6.9%	57	2.0%	35
Misan	6.2%	61	2.3%	45	3.5%	34	.2%	3
Albasrah	6.2%	57	2.2%	52	4.8%	44	.7%	17
Baghdad	7.7%	92	3.5%	88	6.8%	80	2.4%	59

2. تشير نتائج جدول رقم (2) للفئات العمرية لأطفال العراق بان اعلى نسب زيادة في الوزن والسمنة كانت في الفئة الاولى (5-0) شهر. ونلاحظ ان جميع الفئات

العمرية بالنسبة لزيادة في الوزن والسمنة شهدت انخفاضا في عام 2011 عن النسب للفئات العمرية في عام 2006 .

جدول(2) يوضح اعداد ونسب الزيادة في الوزن والسمنة حسب الفئات العمرية للعامين 2006 و 2011

الفئات العمرية	الزيادة في الوزن 2006		الزيادة في الوزن 2011		السمنة 2006		السمنة 2011	
	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال
< 6 months	9.4%	131	8.4%	315	7.5%	105	5.2%	195
6-11 months	6.3%	104	6.1%	224	6.3%	105	3.1%	113
12-23 months	7.5%	261	6.4%	465	5.5%	191	2.7%	192
24-35 months	4.8%	155	4.1%	296	3.1%	99	1.6%	112
36-47 months	4.9%	150	4.4%	315	3.3%	102	1.7%	122
48-59 months	4.7%	139	4.0%	256	2.5%	76	2.3%	147

3. نتائج جدول رقم(3) تشير الى ان جميع الفئات للمستوى التعليمي للام لمؤشر الزيادة في الوزن قد انخفضت في عام 2011 عن عام 2006 باستثناء ثبات النسبة لمستوى التعليمي للام في فئة المتوسطة والثانوية وللعامين المذكورين. وعلى النسب كانت في مستوى الام الابتدائي والأقل كان في تعليم الام غير نظامي

للعاميين. وكذلك بالنسبة للسمنة نلاحظ بان جميع الفئات قد انخفضت في عام 2011 عن عام 2006 عدا الفئة (الام المتعلمة تعليم غير نظامي) فقد ارتفعت .

جدول (3) يوضح اعداد ونسب الزيادة في الوزن والسمنة حسب المستوى التعليمي للام للعامين 2006 و 2011

المستوى التعليمي للام	الزيادة في الوزن 2006		الزيادة في الوزن 2011		السمنة 2006		السمنة 2011	
	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال
غير متعلمة	5.8%	196	4.5%	392	3.7%	123	2.2%	191
ابتدائية	6.2%	463	5.4%	964	4.6%	344	2.4%	429
متوسطة وثانوية	5.8%	274	5.8%	513	4.3%	207	2.9%	257
تعليم غير نظامي	4.8%	7	2.3%	2	2.7%	4	4.6%	4

نتائج جدول (4) تشير الى ان نسبة الزيادة في الوزن والسمنة للذكور والاناث قد انخفضت في عام 2011 عن عام 2006. وان نسبة الزيادة في الوزن عند الاناث اعلى من الذكور في العاميين اما في السمنة فكانت ايضا النسبة للاناث اعلى في عام 2006 ولكن في عام 2011 النسبتين متساوية.

**جدول (4) يوضح اعداد ونسب الزيادة في الوزن والسمنة حسب الجنس للعامين
 2011 و 2006**

الجنس	الزيادة في الوزن 2006		الزيادة في الوزن 2011		السمنة 2006		السمنة 2011	
	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال
ذكر	5.5%	438	4.9%	888	3.9%	314	2.5%	456
انثى	6.5%	502	5.7%	983	4.7%	364	2.5%	425

5. نتائج جدول (5) تشير الى ان نسبة الزيادة في الوزن والسمنة قد انخفضت بالنسبة الى الحضر والريف في عام 2011 عن عام 2006. وان النسب في الريف اعلى من الحضر في الزيادة في الوزن والسمنة للعامين.

**جدول (5) يوضح اعداد ونسب الزيادة في الوزن والسمنة حسب المنطقة للعامين
 2011 و 2006**

مكان العيش	الزيادة في الوزن 2006		الزيادة في الوزن 2011		السمنة 2006		السمنة 2011	
	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال	النسبة	عدد الاطفال
الحضر	5.8%	559	5.3%	1022	3.9%	383	2.6%	505
الريف	6.3%	381	5.2%	849	4.9%	295	2.3%	376

3-3 تحليل نتائج الزيادة في الوزن:

ويصنف الزيادة في الوزن (WHZ) عند الاطفال الذين يزيد وزنهم نسبة الى طولهم عن اكثر من انحرافين معيارين عن وسيط المجتمع المرجعي .

وكان عدد الاطفال ضمن فئة الزيادة في الوزن (940) طفل أي بنسبة 6.0% في عام 2006.

وكان عدد الاطفال ضمن فئة الزيادة في الوزن (1871) طفل أي بنسبة 5.3% في عام 2011.

وكانت النتائج كما يظهر في الجدول (6)

مقارنة الزيادة في الوزن بين متوسطات المحافظات :-

لاختبار الفرضية بان متوسطات الزيادة في الوزن متساوية بين المحافظات

$$H_0: \mu_1 = \mu_2 \dots \dots \dots = \mu_{18}$$

$$H_1: \text{هناك اختلاف في محافظتين على الاقل} \dots (8)$$

كانت قيمة p-value المستخرجة من جدول (ANOVA) تساوي 0.006 وهي أصغر من مستوى الدلالة 0.05 في عام 2006 لذا نقبل الفرضية البديلة أي يوجد فروق معنوية ذات دلالة احصائية بين متوسطات المحافظات لحالة الزيادة في الوزن. ولغرض معرفة أي من هذه المحافظات ادى الى هذه الاختلافات تم اجراء اختبار المقارنات المتعددة في تحليل التباين لمعيار واحد باستعمال طريقة LSD وقد ظهرت فروق معنوية بمستوى دلالة 0.05 بين متوسطات المحافظات التالية :

- محافظة نجف مع المحافظات(ديالى والانبار والبصرة وواسط وبابل والقادسية والمثنى وذي قار

- محافظة الانبار مع المحافظات(نينوى وسليمانية وصلاح الدين).

- محافظة ذي قار مع المحافظات(نينوى وسليمانية وميسان وكركوك).

- محافظة سلیمانیه مع المحافظات(واسط والقادسية والبصرة وبابل).

- محافظة بابل مع محافظة صلاح الدين .

اما في عام 2011 فقد كانت قيمة p-value تساوي 0.129 وهي أكبر من مستوى الدلالة 0.05 لذا نقبل فرضية العدم أي لا يوجد فروق معنوية ذات دلالة احصائية بين متوسطات المحافظات.

مقارنة الزيادة في الوزن بين متوسطات الفئات العمرية للاطفال

لبيان هل ان متوسطات الفئات العمرية متساوية يكون وفق الفرضية الأتية :

$$H_0: \mu_1 = \mu_2 \dots \dots = \mu_6$$

H_1 : هناك اختلاف في فئتين عمرية على الاقل

.....(9)

فقد كانت قيمة p-value المستخرجة من جدول (ANOVA) تساوي 0.232 في عام 2006 وفي عام 2011 تساوي 0.795 وهي أكبر من مستوى الدلالة 0.05 لذا نقبل فرضية العدم أي لا يوجد فروق معنوية ذات دلالة احصائية بين متوسطات الفئات العمرية لحالة الزيادة في الوزن للعامين 2006 و 2011.

مقارنة الزيادة في الوزن بين متوسطات المستوى التعليمي للام

لاختبار الفرضية هل ان متوسطات المستوى التعليمي للام متساوية اي:

$$H_0: \mu_1 = \mu_2 \dots = \mu_4$$

H_1 : هناك اختلاف في مستويين لتعليم الام على الاقل: ... (10)

كانت النتائج المستخرجة من جدول (ANOVA) تشير الى ان في عام 2006 كانت قيمة p-value تساوي 0.576 و في عام 2011 كانت تساوي 0.337 وهي أكبر من مستوى الدلالة 0.05 لذا نقبل فرضية العدم اي لا يوجد فروق معنوية ذات دلالة احصائية بين متوسطات المستوى التعليمي للام لحالة الزيادة في الوزن للعامين 2006 و 2011.

مقارنة الزيادة في الوزن بين متوسطي الذكور والاناث

لبيان هل ان متوسطي (الذكور والاناث) متساوية وفق الفرضية (2) نجد بان قيمة p-value المقابلة لاختبار t هي (0.05 < 0.047) لذا نقبل الفرضية البديلة أي توجد فروق ذات دلالة احصائية لحالة الزيادة في الوزن بين متوسطين الذكور

والاناث لانحرافين معيارين، اما في عام 2011 فقد كانت قيمة p-value المقابلة لاختبار t(-0.408) هي 0.683 اكبر من 0.05 لذا نقبل الفرضية العدم أي لا توجد فروق معنوية ذات دلالة احصائية بين متوسطي الذكور والاناث في حالة الزيادة في الوزن .

مقارنة الزيادة في الوزن بين متوسطي الحضر والريف

لبيان هل ان متوسطي(الحضروالريف) متساوية وفق الفرضية (2) نجد بان في عام 2006 كانت قيمة p-value المقابلة لاختبار t تساوي 0.698 و في عام 2011 كانت تساوي 0.806 وفي العامين اكبر من 0.05 لذا نقبل فرضية العدم أي لا توجد فروق معنوية ذات دلالة احصائية بين متوسطين الحضروالريف لحالة الزيادة في الوزن للعامين 2006 و 2011.

جدول (6) يوضح قيم P-value للاختبار بين متوسطات مؤشرات الزيادة في

الوزن للعامين 2006 و 2011

ت	الحالات	الزيادة في الوزن	
		2006	2011
	المؤشرات		
1	المحافظات	0.006	0.129
2	الفئات العمرية	0.232	0.795
3	المستوى التعليمي للام	0.576	0.337
4	الذكور والاناث	0.047	0.844
5	الحضر والريف	0.698	0.806

3-4 السمنة :

تصنف السمنة (WHZ) عند الاطفال الذين يزيد وزنهم نسبة الى طولهم عن اكثر من ثلاثة انحرافات معيارية عن وسيط المجتمع المرجعي .

وكان عدد الاطفال ضمن فئة السمنة(678) طفل أي بنسبة 4.3% في عام 2006.

وكان عدد الاطفال ضمن فئة السمنة(881) طفل أي بنسبة 2.5% في عام 2011.

وكانت النتائج كما يظهر في الجدول (7)

مقارنة السمنة بين متوسطات المحافظات في العامين

لاختبار الفرضية (8) بين متوسطات المحافظات كانت نتائج جدول ANOVA تشير الى ان قيمة p-value في عام 2006 تساوي 0.180 وفي عام 2011 كانت تساوي 0.093 وهي أكبر من مستوى الدلالة 0.05 لذا نقبل فرضية العدم أي لا توجد فروق معنوية ذات دلالة احصائية بين متوسطات المحافظات لحالة السمنة في العامين 2006 و 2011.

مقارنة السمنة بين متوسطات الفئات العمرية للاطفال

في عام 2006 كانت نتائج جدول ANOVA لاختبار الفرضية(9)بين متوسطات الفئات العمرية للاطفال تشير الى ان قيمة p-value تساوي 0.966 و تساوي 0.247 في عام 2011 وهي أكبر من مستوى الدلالة 0.05 لذا نقبل فرضية العدم لا توجد فروق معنوية ذات دلالة احصائية بين متوسطات الفئات العمرية لحالة السمنة في العامين 2006 و 2011

مقارنة السمنة بين متوسطات المستوى التعليمي للام لعامي

لاختبار الفرضية(10) بين متوسطات المستوى التعليمي للام كانت نتائج جدول ANOVA تشير الى ان قيمة p-value في عام 2006 تساوي 0.659 وفي عام 2011 كانت قيمة p-value تساوي 0.994 وهي أكبر من مستوى الدلالة

0.05 لذا نقبل فرضية العدم أي لا توجد فروق معنوية ذات دلالة احصائية بين متوسطات المستوى التعليمي للام لحالة السمنة في العامين 2006 و 2011.

مقارنة السمنة بين متوسطي الذكور والاناث

في عام 2006 كانت قيمة p-value المقابلة لاختبار t لاختبار الفرضية(2) بين متوسطي الذكور والاناث هي 0.673 و في عام 2011 كانت تساوي 0.709 وهي اكبر من 0.05 لذا نقبل فرضية العدم أي لا توجد فروق معنوية ذات دلالة احصائية بين متوسطين الذكور والاناث لحالة السمنة في العامين 2006 و 2011.

مقارنة السمنة بين متوسطي الحضر والريف

لاختبار الفرضية(2) بين متوسطي الحضر والريف كانت قيمة p-value المقابلة لاختبار t في عام 2006 هي 0.941 اما في عام 2011 فقد كانت 0.736 لذا نقبل فرضية العدم أي لا توجد فروق معنوية ذات دلالة احصائية بين متوسطين الحضر والريف لحالة السمنة في العامين 2006 و 2011.

جدول(7) يوضح قيم P-value للاختبار بين متوسطات مؤشرات السمنة للعامين

2011 و 2006

السمنة		الحالات المؤشرات	ت
2011	2006		
0.093	0.180	المحافظات	1
0.247	0.966	الفئات العمرية	2
0.944	0.659	المستوى التعليمي للام	3
0.709	0.673	الذكور والاناث	4
0.736	0.941	الحضر والريف	5

3-5 المقارنة بين عامي 2006 و2011

لمعرفة الفروق بين العامين 2006 و2011 للمسحيين MICS3 وMICS4 من خلال ماياتي:

مقارنة الوزن الى الطول WHZ بين العامين

لاختبار الفرضية(2) بين متوسطي الوزن الى الطول (WHZ) كانت قيمة p -value المقابلة لاختبار $t(11.837)$ هي 0.00 اصغر من 0.05 لذا نقبل الفرضية البديلة أي توجد فروق ذات دلالة احصائية بين متوسطي العامين 2006 و2011 في حالة الوزن الى الطول.

مقارنة نسبة الزيادة في الوزن بين العامين

لاختبار النسبة لبيان هل توجد فروق معنوية ذات دلالة احصائية بين نسبة الزيادة في الوزن لعامي 2006 و2011 وفق الفرضية (6)بتطبيق الصيغة (7) فكانت النتائج تشير الى ان عدد المصابين في الزيادة في الوزن (940) من العدد الكلي (15777) في عام 2006 و كان عدد المصابين في الزيادة في الوزن (1871) من العدد الكلي (35413) في عام 2011 وكانت النتائج : قيمة Z المحتسبة تساوي (3.094) وقيمة (P- value) المقابله لها تساوي 0.00 وهي أصغر من مستوى الدلالة 0.05 اذن يوجد فروق معنوية ذات دلالة احصائية بين نسبة الزيادة في الوزن بين عامي 2006 و2011.

مقارنة نسبة السمنة بين العامين

كان عدد المصابين بالسمنة(678) من العدد الكلي(15777) في عام 2006 وكان عدد المصابين بالسمنة(881) من العدد الكلي (35413) في عام 2011 ولاختبار الفرضية (6) بين نسبة السمنة للعامين بتطبيق الصيغة (7) كانت النتائج بان قيمة Z المحتسبة تساوي(11.002) وقيمة (P- value) المقابله لها تساوي

0.00 وهي أصغر من مستوى الدلالة 0.05 اذن يوجد فروق معنوية ذات دلالة احصائية بين السمنة بين عامي 2006 و 2011 .

4- الأستنتاجات والتوصيات

1-4 الأستنتاجات

من تطبيق الصيغ الواردة في الجانب النظري والنتائج التي أفرزها الجانب التطبيقي تم التوصل الى الأستنتاجات في عامي 2006 و 2011:

1.ارتفعت نسبة الزيادة في الوزن في ستة محافظات (دهوك - سليمانية - ديالى - الانبار - بابل - النجف) و بقيت ثابتة في محافظة نينوى وانخفضت في المحافظات الباقية. اما في السمنة ارتفعت النسبة في اربعة محافظات (دهوك - سليمانية - ديالى - نجف) وارتفعت في المحافظات الباقية وهذا يدل على وجود تفاوت في النظام والسلوك الغذائي عند الاسر العراقية

2.انخفضت نسب الزيادة في الوزن والسمنة لجميع الفئات العمرية والجنس (للذكور والإناث) والمنطقة (للحضر والريف) بين العامين وهذا يشير الى الوعي والاهتمام من قبل الاسر بصحة اطفالهم .

3.كانت نتائج المقارنة بين متوسطات المحافظات بانه لا يوجد فروق معنوية ذات دلالة احصائية في السمنة بينما يوجد فروق معنوية ذات دلالة احصائية في الزيادة في الوزن في عام 2006. اما في عام 2011 لا يوجد فروق معنوية ذات دلالة احصائية في الزيادة في الوزن والسمنة.

4. كانت نتائج المقارنة بين متوسطات الفئات العمرية و متوسطات المستوى العلمي للام و متوسطي الحضر والريف بانه لا يوجد فروق معنوية ذات دلالة احصائية لمؤشري الزيادة في الوزن والسمنة للعامين 2006 و 2011 .

5. كانت نتائج المقارنة بين متوسطي الذكور والاناث بانه لاتوجد فروق معنوية ذات دلالة احصائية في السمنة بينما توجد فروق معنوية في الزيادة في الوزن في عام 2006. اما في عام 2011 لا يوجد فروق معنوية ذات دلالة احصائية في الزيادة في الوزن والسمنة.

6. توجد فروق معنوية ذات دلالة احصائية بين متوسطي عامي 2006 و 2011 في الوزن الى الطول (whz) .

7. توجد فروق معنوية ذات دلالة احصائية في نسبة الزيادة في الوزن ونسبة السمنة بين عامي 2006 و 2011 وهذا يشير الى ان الدولة نجحت في وضع برامج واليات لتحسن الوضع التغذوي للأطفال.

4-2 التوصيات

1. العناية بمراقبة منحنى نمو الطفل من خلال قياس وزنه وطوله.
2. اهتمام وزارة الصحة بدقة البيانات واحصائيات الاطفال لمساعدة التخطيط الناجح في وضع البرامج والخطط الكفيلة في الوصول الى النتائج المرجوه .
3. ان تربية الطفل على استعمال الاسلوب التغذوي والبدني السليم يؤدي الى الوقاية من السمنة وزيادة الوزن في الاعمار المبكر

5. المصادر

1. الجهاز المركزي للإحصاء وتكنولوجيا المعلومات، وزارة التخطيط والتعاون الإنمائي(2006) "المسح العنقودي متعدد المؤشرات (MICS-3) العراق" .
2. الجهاز المركزي للإحصاء وتكنولوجيا المعلومات، وزارة التخطيط والتعاون الإنمائي(2011) "المسح العنقودي متعدد المؤشرات (MICS-4) العراق" .
3. وزارة الصحة، دائرة الصحة العامة، معهد بحوث التغذية، (2013) " برنامج الغذاء العالمي، دليل وحدات التغذية في مراكز الرعاية الصحية الاولية".
4. كيال، قاسم (2013) " السمنة الزائدة :اطفال بدينون" مجلة كلاليت.

https://www.clalit.co.il/arb/your_health/childrens/Pages/fatness_children.aspx

5. العرجان، جعفر فارس(2011) "انتشار السمنة والوزن الزائد والنقص في الوزن لدى طلبة جامعة البلقاء التطبيقية في الاردن" العلوم التربوية ، المجلد 38 ، ملحق (6).

6. السالم، علي بن احمد بن علي(2016) " نمط الحياة وعلاقته بزيادة الوزن والسمنة لدى المراهقين : دراسة ميدانية على طلاب المرحلة الثانوية بمدينة الرياض " اطروحة دكتوراه في علم الاجتماع ، كلية الاداب ، جامعة الملك سعود ، المملكة العربية السعودية

7. الاميري عامر محمد علي، العيثاوي امل داود سليم ، الشمري سمر عدنان عبد الامير(2016) " السمنة عند اطفال الرياض وعلاقتها ببعض المتغيرات الاسرية " مجلة كلية التربية للبنات ، المجلد 27 (2).

8. ريتشارد جونسون ، دين وشرن "التحليل الاحصائي للمتغيرات المتعددة من الواجهة التطبيقية"، دار المريخ للنشر

9. المشهداني، كمال علوان خلف، عبودي، عماد حازم، عبدالله، سهيل نجم،(2012) "الاختبارات الاحصائية تطبيقات محوسبة باستخدام برنامج spss " بابل للطباعة.

10. المشهداني، محمود حسن، المشهداني، كمال علوان خلف (1989)"تصميم وتحليل التجارب " مطبعة جامعة الموصل.

11. سعيد ،ظافر طارق محمد ،(2014)،"مقارنة الحالة التغذوية لاطفال العراق لعامي 2006 و 2011 دراسة احصائية " بحث للدبلوم العالي للاحصاء التطبيقي مقدم الى مجلس كلية الادارة والاقتصاد جامعة بغداد.

12. بشير، سعد زغلول، (2003)، " دليلك الى البرنامج الاحصائي spss " المعهد العربي للتدريب والبحوث الاحصائية. الاصدار العاشر .

13. Food security and vulnerability analysis in Iraq, cosit and wfp,(2006)

14.Sarah E. Barlow(2007)"Expert Committee Recommendations Regarding the Prevention, Assessment and

Treatment of Child and Adolescent Overweight and Obesity:
Summary report, Pediatrics 120;S164

15. Z Test Calculator for 2 Population Proportions
<http://www.socscistatistics.com/tests/ztest/>

16. TIMM, Neil H;(2002)"Applied Multivariate Analysis",
springer, New-York.

المفاضلة بين نتائج البرمجة الخطية ومشاكل النقل في تخفيض التكاليف الى ادنى حد ممكن (دراسة تطبيقية على شركة دهوك الزراعية لا نتاج الدجاج المحلي في محافظة دهوك)

الاستاذ المساعد سماهر طارق ابراهيم

قسم الاحصاء / جامعة دهوك / كردستان العراق

baghdad_s78@yahoo.com

المستخلص

يهدف البحث ان مشكلة النقل من المشاكل الحيوية والمهمة لتحديد اقل كلفة نقل ممكنة لجميع البضائع وبالنظر لعدم توفر طريقة محددة تعطي حلا امثل لذلك حاولت الباحثة ان تقوم بدراسة لاهم الطرائق لحل مشاكل النقل التي تعطي حل امثل او حلا قريبا من الامثل ومفاضلة نتائج هذه الطرائق مع نتيجة البرمجة الخطية لحل مشكلة النقل المتعددة المراحل لمعرفة اي طريقة مطابقة او قريبة الى نتيجة البرمجة الخطية وذلك من خلال تطبيقها على ارض الواقع بأخذ احدى الشركات الزراعية المتمثلة بمنتوج (دجاج المائدة) الذي ينقل منحقول الشركة الثلاثة الى المحافظات (دهوك ، اربيل ، سلیمانیه) من اجل تخفيض التكاليف الاجمالية الى ادنى حد ممكن.

الكلمات المفتاحية : مشاكل النقل، طريقة البرمجة الخطية ، القيود ، دالة الهدف ، الحل الامثل.

Comparison between the results of linear programming and transportation problems in reducing costs to the minimum possible (an applied study on the Dohuk Agricultural Company for local chicken production in Dohuk Governorate)

Assistant Professor. Samaher Tariq Ibrahim
Faculty of Administration and Economics
Department of Statistics/University of Duhok
baghdad_s78@yahoo.com

Abstract

The research aims that the transportation problem is one of the vital and important problems to determine the lowest possible transportation cost for all goods, and given the lack of a specific method that gives an optimal solution, therefore, the researcher tried to study the most important methods for solving transportation problems that give an optimal solution or a solution close to the optimal, and to compare the results of these methods with the result of linear programming to solve the multi-stage transportation problem To find out which method is identical or close to the result of linear programming, by applying it on the ground by taking one of the agricultural companies represented by the product (table chicken), which is transported from the company's three fields to the governorates (Dohuk, Erbil, Sulaymaniyah) in order to reduce the total costs to the minimum possible.

key words Transportation problems, linear programming method, constraints, objective function, optimal solution.

المقدمة

نموذج النقل هو اهم نماذج البرمجة الخطية التي تقوم على اساسه النقل الاقتصادي للوحدات المنتجة باقل كلفة نقل ممكنة وباقل وقت ممكن لتلبية احتياجات تلك المراكز (Destinations) الى المواقع الطلب (sources) من مصادر الانتاج ويعد نموذج النقل مكمل للعملية الانتاجية بحسب حاجة الانتاج من

مستلزمات لهذه العملية وتعد مشكلة النقل احد الاساليب الرياضية المهمة التي ساعدت في عملية صنع القرار الاقتصادي لصناع القرار ومن اهم المشاكل الاقتصادية التي تواجه المنشآت الانتاجية اذ تبرر أهميتها من خلال توزيع السلع او البضائع الى المستهلك (2، 7). ان حل مشكلة النقل تتم من خلال تقليل الوقت اللازم لنقل البضائع من المصادر الى الطلب فهي من الامور المهمة التي تواجه المنشآت العسكرية والانتاجية على حد سواء باعتبار ان وقت النقل مستقل عن كمية السلع التي يتم شحنها من الموردين الى المستهلكين. اما كلفة النقل فتعتمد على اختلاف كمية السلع التي تم شحنها للوصول الى الحل الامثل على شرط ان تكون الكميات المطلوبة في مراكز الطلب هي نفسها في مراكز المصادر علما ان كلفة النقل لا تتغير مع الكميات المنقولة بل ثابتة ومشكلة النقل احد العناصر الاساسية في العملية الانتاجية (8، 9، 10). ان هذه الطريقة المستخدمة في البحث واجراء المفاضلة لمعرفة افضل هذه الطرائق بينها وبين طريقة البرمجة الخطية باستخدام طريقة السمبلكس التي تم تطبيقها فيشركة دهوك الزراعية لا نتاج الدجاج المحلي في محافظة دهوك على حقولها الثلاثة وهي (حقل 1 ، حقل 2 ، حقل 3)

المبحث الاول : منهجية الدراسة

مشكلة البحث:

ان زيادة تكاليف نقل الدجاج المحلي من حقول شركة دهوك الزراعية لا نتاج الدجاج المحليالى المحافظات دون وضع اي خطة للنقل يسبب ذلك هدر في الاموال مما يؤثر سلبا من الناحية الاقتصادية للشركة لذلك حاولت الباحثة في هذا البحث دراسة بعض الطرائق النقل المهمة والمتمثلة (بطريقة الركن الشمالي الغربي ، طريقة اقل كلفة ، طريقة فوجل) لحل مشكلة النقل ومفاضلتها مع البرمجة الخطية استخدام طريقة السمبلكس من اجل تخفيض التكاليف الاجمالية الى ادنى حد ممكن .

اهمية البحث

تكمن اهمية البحث ايجاد اقل كلفة ممكنة لمشكلة النقل (الدجاج المحلي) مع تحديد الوحدات المثلى التي ستنقل من الحقول الرئيسية في الشركة المذكور اعلاه الى المحافظات ضمن المطلوب والمعروض من الشركة باستخدام المفاضلة بين طرائق النقل الثلاثة و طريقة البرمجة الخطية باستخدام طريقة السمبلكس .

هدف البحث :

ان مشكلة النقل من المشاكل الحيوية والمهمة لتحديد اقل كلفة نقل ممكنة لجميع البضائع وبالنظر لعدم توفر طريقة محددة تعطي حلا امثل لذلك حاولت الباحثة ان تقوم بدراسة لاهم الطرائق لحل مشاكل النقل التي تعطي حل امثل او حلا قريبا من الامثل ومفاضلة نتائج هذه الطرائق مع نتيجة البرمجة الخطية لحل مشكلة النقل المتعددة المراحل لمعرفة اي طريقة مطابقة او قريبة الى نتيجة البرمجة الخطية وذلك من خلال تطبيقها على ارض الواقع بأخذ احدى الشركات الزراعية المتمثلة بمنتوج (دجاج المائدة) الذي ينقل من حقول الشركة الثلاثة الى المحافظات (دهوك ، اربيل ، سلیمانانية) من اجل تخفيض التكاليف الاجمالية الى ادنى حد ممكن .

فرضية البحث : نفرض ان طريقة البرمجة الخطية هي احدى افضل الطرائق المستخدمة في حل مشكلة النقل في شركة دهوك الزراعية لا نتاج الدجاج المحلي ومفاضلتها مع الطرائق الاخرى (طريقة الركن الشمالي الغربي ، طريقة اقل كلفة ، طريقة فوجل) .

المبحث الثاني :الاطارالنظري لمتغيرات الدراسة

المقدمة

تعتبر مسألة النقل احدى تطبيقات البرمجة الخطية الهامة حيث ان النماذج الرياضية المستخدمة في مشكلة النقل هي نماذج خطية والهدف من استخدامها هو ايجاد اسلوب امثل لتوزيع الوحدات او المنتوجات من عدة مصادر للعرض (معامل ،

مواني ، مراكز تسويقية) الى عدة مواقع الطلب (مراكز استهلاكية) باقل كلفة ممكنة او بأعلى ربح او بأقل وقت . ويعتمد نموذج النقل على خصائص منها ان تكون المواد التي تنقل متجانسة وان يكون مجموع الكميات المعروضة تساوي الكميات المطلوبة (14,15) .

الصياغة العامة لنموذج النقل

تمثل الصيغة الجدولية النقل منطلق ايجاد حل اولي ممكن للوصول الى الحل الامثل (النهائي) في تحقيق اقل كلفة ممكنة من مجموع تكاليف النقل ، والصيغة الجدولية لمشكلة النقل عبارة عن مصفوفة عدد صفوفها (M) تمثل المصادر (مراكز التوزيع) وعدد اعمدتها (N) وتمثل مراكز الاستلام وهو يظهر كما يلي جدول رقم (1)(14).

المراكز المصادر	N_1	N_2	-----	N_n	العرض
M_1	C_{11}	C_{12}	C_{1n}	a_1
	X_{11}	X_{12}		X_{1n}	
M_2	C_{21}	C_{22}	C_{2n}	a_2
	X_{21}	X_{22}		X_{2n}	
-	a_3
-	.	.		.	
-	.	.		.	
M_m	C_{m1}	C_{m2}		C_{mm}	a_m
	X_{m1}	X_{m2}		X_{mm}	
الطلب	b_1	b_2		b_n	$\sum a_i$ $\sum b_i$

اذ ان

مصادر العرض (M_1, M_2, \dots, M_m) مواقع الطلب (N_1, N_2, \dots, N_n)
 C_{ij} : كلفة نقل الوحدة الواحدة من المصدر (i) الى الموقع او اتجاه الطلب j
 X_{ij} : عدد الوحدات المنقولة من المصدر (i) الى الموقع او اتجاه الطلب j .
الفرضية الاساسيالنقل: نموذج النقل هو ان ما معروض في مصادر العرض اي
مجموع المعروض يساوي مجموع الطلب وفي هذه الحالة يسمى نموذج النقل
بنموذج النقل المتوازن .

في مواقع الطلب اي ($\sum b_i = \sum a_i$)

طرائق حل مشكلة النقل :

طرق ايجاد الحل الاساسي الاولي المقبول (11)

ان طرق الحل التي تعطي حلا اوليا مقبولا لمشكلة النقل الهدف منها هو الحصول
على حل ممكن لمشكلة النقل لا يتعارض مع طبيعة القيود التي تفرض على المشكلة
ومنه الحصول الى الحل الامثل ، وهناك عدة طرق شائعة تختلف من حيث النتائج
التي نتوصل اليها ، اذ كلما كانت النتائج في الحل الاولي قريب من الحل الامثل
تساعدنا من الحصول على الحل الامثل بأسرع وقت ممكن ، وكذلك تختلف من
حيث الوقت والجهد المبذل للوصول الى الحل الاولي ومن اهم الطرق واكثرها
استخداما الى الحل الاولي هي :

اولا- طريقة الركن الشمالي الغربي .

تعتبر هذه الطريقة من ابسط الاساليب الرياضية . لحل مشاكل النقل الا انها لا
تحقق في معظم الاحيان الحل الامثل لمشكلة نقل معينة اذ تبدأ عملية ايجاد الحل
من الزاوية الشمالية الغربية لجدول النقل ولذلك سميت بهذا الاسم ومن عيوب هذه
الطريقة انها لا تعتمد الكلف التي في الجدول بالاعتماد على موقع الخلية في
الجدول والهدف الاساسي لمشكلة النقل هو تقليل تكاليف النقل(12).

ثانيا - طريقة اقل كلفة.

تعتبر هذه الطريقة افضل من الطريقة السابقة لأنها تأخذ بعين الاعتبار الاقل تكلفة، وحتى نحصل على الحل الاساسي الاولي الممكن بهذه الطريقة علينا في البداية ان نتأكد من ان جدول النقل في حالة توازن ثم نتبع الخطوات التالية (12, 13).

- نبدأ بتزويد المربع ذا التكلفة الاقل بأكبر ما يمكن من عدد الوحدات من المخزون المقابل لهذا المربع .
- نتابع ملئ المربعات ذات التكلفة الاقل بالتتابع الى ان نزود جميع مراكز من المصادر المتوفر .
- نحسب التكلفة الاجمالية للمربعات المختلفة .

ثالثا - طريقة فوجل :

تعد هذه الطريقة من افضل الطرق في حل مسائل النقل وغالبا ما تعطي حلا امثلا ويمكن ان نجمل خطوات حل مسألة النقل وفقها كالآتي (14).

- ا- ايجاد الفرق بين اقل كلفتين في كل صف.
- ب- ايجاد الفرق بين اقل كلفتين في كل عمود.
- ج- تحديد اكبر فرق سواء كان في صفوف او الاعمدة .
- د- البحث عن اقل كلفة في الصف او العمود الذي يقابل اكبر فرق والبدء بتخصيص الكميات التي سترسل الى الاسواق .
- هـ - اعادة الخطوات السابقة لحين الوصول الى توزيع كامل الطاقات الانتاجية واشباع لاحتياجات الاسواق مع مراعاة استبعاد الخلايا التي تشغل.

مقدمة عن البرمجة الخطية :

ان اسلوب البرمجة الخطية يعالج المشاكل المختلفة ببناء نموذج حيث يقوم بإيجاد قيم $(x_1, x_2, x_3, \dots, x_n)$ المثلى والتي تحقق اكبر منفعة ممكنة لمتخذ القرار سواء كانت دالة (Maximization profit) او تقليل التكاليف (minimization cost) (4).

شروط البرمجة الخطية :

- 1- القدرة على تحديد موضوع البحث تحديدا رياضيا دقيقا .
- 2- محدودية الموارد البشرية والمادية الخاضعة للبرمجة مثل محدودية راس المال ... عدد العمال .. البضاعة المستورده ... الطاقة الانتاجية وغيرها .
- 3- امكانية التعبير عن الفعاليات او المتغيرات موضوع البرمجة بصورة رقمية .
- 4- ان تكون العلاقة بين المتغيرات هي علاقة خطية (3).
- 5- توفر استخدامات تنافسية للموارد البشرية والمادية موضوع البرمجة الخطية مثلا

انتاج منشأة سلعتين x_1, x_2

الضيعة الرياضية للبرمجة الخطية :

$$Max \text{ or } Min(Z) = C_1 X_1 + C_2 X_2 + C_3 X_3 + C_4 X_4 \dots \dots \dots + C_n X_n$$

حيث ان X_1 هي عدد الوحدات المنتجة من المنتج i وان C_i هي معامل رقمي يمثل ربح (كلفة) الوحدة الواحدة من المنتج X_i حيث ان $i = 1, 2, 3, \dots, n$.
اما القيود الخطية فهي تعرف كما يأتي (5) :

$$a_{11} x_1 + a_{12} x_2 + \dots \dots \dots a_{1n} x_n (\leq, =, \geq) b_1$$

$$a_{21} x_1 + a_{22} x_2 + \dots \dots \dots a_{2n} x_n (\leq, =, \geq) b_2$$

⋮
⋮
⋮
⋮

$$a_{m1} x_1 + a_{m2} x_2 + \dots \dots \dots a_{mn} x_n (\leq, =, \geq) b_m$$

شرط عدم السلبية ($x_i \geq 0 (i = 1, 2, 3, \dots, n)$)

حيث ان n هي عدد المتغيرات و m تمثل عدد القيود .

b_j قيمة المتاح من الموارد ، a_{ij} تمثل احتياجات المنتج i من المورد j حيث ان :

$$j = 1, 2, 3, \dots, n$$

$$i = 1, 2, 3, \dots, m$$

طرائق حل مشكلة البرمجة الخطية (6):

هناك ثلاثة طرق رئيسية تستخدمها بحوث العمليات لحل مشاكل البرمجة الخطية وكما يأتي:

1- طريقة الرسم البياني : وهي اداة بيانية بسيطة جدا تستخدم رغم بساطتها في معالجة مشاكل متعددة في مجال التسويق والانتاج والافراد وغيرها من المجالات الادارية حيث تشترط هذه الطريقة وجود ثلاثة متغيرات على الاكثر بسبب تعذر رسم اكثر من ثلاثة ابعاد هندسية على الورق حيث يتم رسم دالة الهدف والقيود ومن ثم ايجاد منطقة الحل الامثل.

2- طريقة الحل الجبري: وهي تمثل اسلوبا اخر من اساليب البرمجة الخطية وهذه الطريقة تتميز باتساع استخدامها في حالة زيادة عدد المتغيرات عن اثنين .

3- الطريقة الثنائية : لقد استخدمت هذه الطريقة الاول مرة عام 1947 من قبل العالم الامريكي دانتر وهي اوسع نطاقا من الطريقتين السابقتين وهذه الطريقة تتميز بكونها تتكون من عمليات ومراحل متكررة حيث تمثل كل مرحلة حلا قائما بذاته مع ملاحظة ان كل حل افضل من سابقة وهكذا حتى الوصول للحل الامثل كما ان لكل حل من هذه الحلول دالة الهدف الخاصة به . وعلى العموم تتميز هذه الطريقة بدرجة عالية من الدقة و الكفاءة قيد في معالجة مشكلات البرمجة الخطية بغض النظر عن عدد المتغيرات وسوف نستخدم هذه الطريقة في حل المشكلة قيد البحث.

المبحث الثالث:-الجانب العملي

اولا - الحالة الدراسية

شركة دهوك الزراعية لانتاج الدجاج المحلي اذ ساهمت مساهمة كبيرة في الحد من الازمة الخانقة للسوق المحلية ، وتحتوي الشركة على ثلاثة حقول هي :

1- حقل 1 .

2- حقل 2 .

3- حقل 3 .

تقوم الشركة بتوفير بيض الطعام والدجاج للاستهلاك المحلي ومن خلال عملية نقل المنتج من حقول الشركة الى المحافظات وجدنا انهم يعتمدون على تخمينهم وخبرتهم السابقة في اقل وقت امثل لنقل المنتج.

- تضمنت البيانات الحقول الثلاثة الخاصة بمادة (دجاج المائدة) كما يشير الجدول (2) الى اسم الحقول والطاقة الاستيعابية المخصصة للمحافظات .

الجدول (2) الطاقة الاستيعابية المتوقعة للحقول (دجاج المائدة)لسنة 2024

التسلسل	اسم الحقل	الطاقة الاستيعابية
1	حقل 1	250الف
2	حقل 2	300الف
3	حقل 3	300الف

الجدول من اعداد الباحثة اعتمادا على بيانات الشركة (1)

الجدول (3) كمية الطلب المتوقع للمحافظات مقاسه بالف (دجاج) لسنة 2024

التسلسل	اسم المحافظة	الطاقة الاستيعابية
1	دهوك	300الف
2	اربيل	200الف
3	سليمانية	200الف

الجدول من اعداد الباحثة اعتمادا على بيانات الشركة (1)

علما ان كلفة نقل الوحدة الواحدة للطن الواحد/كم من الحقول الى المحافظات بواسطة السيارات المبردة يكون وفقا للمعادلة الاتية

$$\text{كلفة النقل} = \text{الحمولة (الف)} \times \text{المسافة (كم)} \times \text{سعر النقل بالدينار}$$

الجدول (4) يبين الكلفة اللازمة للنقل (بالف / كم) بين الحقول والمحافظات

المحافظة حقل	سليمانية	دهوك	اربيل	العرض
A حقل 1	11	10	7	250
B حقل 2	8	9	12	300
C حقل 3	13	6	5	300
الطلب	300	200	200	850 700

جدول (5) (الكلفة اللازمة للنقل (الف / كم) بين الحقول والمحافظات بعد موازنة
 الجدول بالعمود الوهمي)

المحافظة الحقول	سليمانية	دهوك	اربيل	وهمي	35000
A حقل 1	11	10	7	0	250
B حقل 2	8	9	12	0	300
C حقل 3	13	6	5	0	300
الطلب	300	200	200	150	850 850

1- حل بطريقة الركن الشمالي الغربي

الجدول (6) الحل بطريقة الركن الشمالي الغربي

المحافظة الحقول	سليمانية	دهوك	اربيل	وهمي	العرض
Aحقول 1	11	10	7	0	250
	250	X	x	X	0
Bحقول 2	8	9	12	0	300
	50	200	50	x	250
					50
					0
Cحقول 3	13	6	5	o	300
	x	X	150	150	150
					0
الطلب	300	200	200	150	850
	50	0	150		
	0		0	0	850

مجموع العرض = مجموع الطلب

$$\sum b_i = \sum a_i$$

$$850 = 850$$

عدد الصفوف + عدد الاعمدة - 1

$$M + N - 1 =$$

$$3 + 4 - 1 = 6$$

$$Min(z) = 250 \times 11 + 50 \times 8 + 200 \times 9 + 50 \times 12 + 150 \times 5 + 150 \times 0 = 6300$$

اقل كلفة كلية ممكنة نقل دجاج المائدة بطريقة الركن الشمالي = 6300

2- الحل بطريقة اقل كلفة

الجدول (7) الحل بطريقة اقل كلفة

المحافظة الحقول	سليمانية	دهوك	اربيل	وهمي	العرض
Aحقول 1	11	10	7	0	250
	x	200	50	x	200
					0
Bحقول 2	8	9	12	0	300
	300	x	x	x	0
Cحقول 3	13	6	5	0	300
	x	0	150	150	150
					0
الطلب	300	200	200	150	850
	0	0	150	0	850

مجموع العرض = مجموع الطلب

$$\sum b_i = \sum a_i$$

$$850 = 850$$

عدد الصفوف + عدد الاعمدة - 1

$$M + N - 1 =$$

$$3 + 4 - 1 = 6$$

$$Min(z) = 300 \times 8 + 200 \times 10 + 50 \times 7 + 150 \times 5 + 150 \times 0 + 0 \times 0 = 5500$$

اقل كلفة كلية ممكنة بطريقة اقل كلفة ممكنة لنقل (دجاج المائدة) = 5500

3- الحل بطريقة فوجل

الجدول (8) الحل بطريقة فوجل

المحافظة الحقول	سليمانية	دهوك	اربيل	وهمي	العرض	الفروقات
Aحقول1	11	10	7	0	250	7,3,3,4
	150	x	100	x	150 0	
Bحقول2	8	9	12	0	300	8,1,x,x
	150	x	x	150	150 0	
Cحقول3	13	6	5	0	300	5,1,1,8
	x	200	100	X	100 0	
الطلب	300	200	200	150	850	
	150	0	100	0		
	0		0			
الفروقات	3	3	2	0		
	3	3	2	X		
	2	4	2	X		
	2	x	2	X		

مجموع العرض = مجموع الطلب

$$\sum b_i = \sum a_i$$

$$850 = 850$$

عدد الصفوف + عدد الاعمدة - 1

$$M + N - 1 =$$

$$3 + 4 - 1 = 6$$

$$\text{Min}(z) = 150 \times 11 + 150 \times 8 + 200 \times 6 + 100 \times 7 + 100 \times 5 + 150 \times 0 = 5250$$

اقل كلفة كلية ممكنة بطريقة فوجل ممكنة لنقل (دجاج المائدة) = 5250

الجدول (9) قبل توازن

المحافظة حقل	سليمانية	دهوك	اربيل	العرض
1 حقل A	11	10	7	250
2 حقل B	8	9	12	300
3 حقل C	13	6	5	300
الطلب	300	200	200	850
				700

بناء النموذج الرياضي (انموذج برمجة الخطية) لمشكلة النقل

تعريف المتغيرات

لصيغة الانموذج الرياضي (انموذج برمجة الخطية) يجب اولاً تعريف معالم

الانموذج لغرض بناء الانموذج وهذه المعالم هي كالاتي

نفرض ان عدد الالف المنقولة من الحقل (الاول) الى المحافظة (الاولى) سليمانية = X_{A1}

نفرض ان عدد الالف المنقولة من الحقل (الاول) الى المحافظة (الثانية) دهوك = X_{A2}

نفرض ان عدد الالف المنقولة من الحقل (الاول) الى المحافظة (الثالثة) اربيل = X_{A3}

نفرض ان عدد الالف المنقولة من الحقل (الثاني) الى المحافظة (الاولى) سليمانية = X_{B1}

نفرض ان عدد الالف المنقولة من الحقل (الثاني) الى المحافظة (الثانية) دهوك = X_{B2}

نفرض ان عدد الالف المنقولة من الحقل(الثاني) الى المحافظة (الثالثة) اربيل = X_{B3}

نفرض ان عدد الالف المنقولة من الحقل (الثالث) الى المحافظة (الاولى) سليمانية = X_{C1}

نفرض ان عدد الالف المنقولة من الحقل (الثالث) الى المحافظة (الثانية) دهوك = X_{C2}

نفرض ان عدد الالف المنقولة من الحقل (الثالث) الى المحافظة (الثالثة) اربيل = X_{C3}

$$Min(z) = 11x_{A1} + 10x_{A2} + 7x_{A3} + 8x_{B1} + 9x_{B2} + 12x_{B3} + 13x_{C1} + 6x_{C2} + 5x_{C3}$$

S.TO

$$x_{A1} + x_{B1} + x_{C1} = 300$$

$$x_{A2} + x_{B2} + x_{C2} = 200$$

$$x_{A3} + x_{B3} + x_{C3} = 200$$

$$x_{A1} + x_{A2} + x_{A3} \leq 250$$

$$x_{B1} + x_{B2} + x_{B3} \leq 300$$

$$x_{C1} + x_{C2} + x_{C3} \leq 300$$

$$x_{Aj} \geq 0, x_{Bj} \geq 0, x_{Cj} \geq 0$$

حل الانموذج الرياضي(انموذج البرمجة الخطية)رغبة منا باستخدام احدى البرامج المخصصة لحل مشاكل البرمجة الخطية حيث ان استخدام البرامج يحقق سرعه ودقة في حل الانموذج كما انه يستوعب اي تغير مستقبلي في المتغيرات وببساطة شديدة جدا دون الحاجة الى تغير فرضيات الانموذج ومن هذه البرامج برنامج اكسل وبرنامج WINQSB وسنعمد في حل هذا الانموذج على برنامج WINQSB باعتباره من اكثر البرامج تخصصا في حل مشاكل البرمجة الخطية . ان جدول الحل الامثل موضح بالشكل (10)حيث يوضح هذا الجدول الحل الامثل لا نموذج البرمجة الخطية.

جدول (10) استخدام البرمجة الخطية في حل مشكلة النقل المتعددة المراحل

02:51:49		Friday	December	20	2024		
Decision Variable	Solution Value	Unit Cost or Profit c(j)	Total Contribution	Reduced Cost	Basis Status	Allowable Min. c(j)	Allowable Max. c(j)
1	X1	0	11.0000	0	3.0000	at bound	8.0000 M
2	X2	0	10.0000	0	2.0000	at bound	8.0000 M
3	X3	100.0000	7.0000	700.0000	0	basic	5.0000 8.0000
4	X4	300.0000	8.0000	2,400.0000	0	basic	-M 11.0000
5	X5	0	9.0000	0	1.0000	at bound	8.0000 M
6	X6	0	12.0000	0	5.0000	at bound	7.0000 M
7	X7	0	13.0000	0	7.0000	at bound	6.0000 M
8	X8	200.0000	6.0000	1,200.0000	0	basic	-M 7.0000
9	X9	100.0000	5.0000	500.0000	0	basic	4.0000 7.0000
Objective	Function	(Min.) =	4,800.0000				
Constraint	Left Hand Side	Direction	Right Hand Side	Slack or Surplus	Shadow Price	Allowable Min. RHS	Allowable Max. RHS
1	C1	300.0000	=	300.0000	0	8.0000	0 300.0000
2	C2	200.0000	=	200.0000	0	8.0000	100.0000 300.0000
3	C3	200.0000	=	200.0000	0	7.0000	100.0000 350.0000
4	C4	100.0000	<=	250.0000	150.0000	0	100.0000 M
5	C5	300.0000	<=	300.0000	0	0	300.0000 M
6	C6	300.0000	<=	300.0000	0	-2.0000	200.0000 400.0000

تفسير النتائج جدول الحل في الشكل (10):

1- قيمة دالة الهدف = 4800

2- قيم المتغيرات :

$$X_1 = 0, X_2 = 0, X_3 = 100, X_4 = 300, X_5 = 0, X_6 = 0, X_7 = 0, X_8 = 200, X_9 = 100$$

3- المصادر المتوفرة والنادرة C_1 مصدر نادر لان $S_1 = 0$ ، C_2 مصدر نادر لان

$S_2 = 0$ ، C_3 مصدر نادر لان $S_3 = 0$ ، C_4 مصدر متوفر لان $S_4 = 150$ ، C_5

مصدر نادر لان $S_5 = 0$ ، C_6 مصدر نادر لان $S_6 = 0$.

4- اسعار الظل $S_1 = 80000$, $S_2 = 80000$, $S_3 = 70000$, $S_4 = 0$, $S_5 = 0$, $S_6 = -20000$

5- مجال التغير في معاملات المتغيرات في دالة الهدف بما لا يؤثر على امثليه الحل.

A - معامل X_1 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[8, M]$ وهذا يعني ان معامل X_1 يجب ان يكون اكبر من او يساوي 8 حتى يبقى الحل امثلا.

B - معامل X_2 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[8, M]$ وهذا يعني ان معامل X_2 يجب ان يكون اكبر من او يساوي 8 حتى يبقى الحل امثلا.

C - معامل X_3 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[5, 8]$ وهذا يعني ان معامل X_3 يجب ان يكون اكبر من او يساوي 5 حتى يبقى الحل امثلا.

D - معامل X_4 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[-M, 11]$ وهذا يعني ان معامل X_4 يجب ان يكون اقل من او يساوي 11 حتى يبقى الحل امثلا.

E - معامل X_5 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[8, M]$ وهذا يعني ان معامل X_5 يجب ان يكون اكبر من او يساوي 8 حتى يبقى الحل امثلا.

F - معامل X_6 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[7, M]$ وهذا يعني ان معامل X_6 يجب ان يكون اكبر من او يساوي 7 حتى يبقى الحل امثلا.

G - معامل X_7 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[6, M]$ وهذا يعني ان معامل X_7 يجب ان يكون اكبر من او يساوي 6 حتى يبقى الحل امثلا.

H - معامل X_8 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[-M, 7]$ وهذا يعني ان معامل X_8 يجب ان يكون اقل من او يساوي 7 حتى يبقى الحل امثلا.

H - معامل X_9 في دالة الهدف : فبا مكانه ان اخذ القيم في الفترة $[4, 7]$ وهذا يعني ان معامل X_8 يجب ان يكون اكبر من او يساوي 4 حتى يبقى الحل امثلا.

6 - مجال التغير في الطرف الايمن في القيود بما لا يؤثر على امثلية الحل .

A - القيد الاول : من الجدول نلاحظ ان الطرف الايمن للقيد يمنه اخذ القيم في الفترة $[0, 300]$ و هذا يعني ان يكون الطرف الايمن للقيد الاول يجب ان يكون اكبر من او يساوي 0 حتى يبقى الحل امثلا

B - القيد الثاني: من الجدول نلاحظ ان الطرف الايمن للقيد يمنه اخذ القيم في الفترة [100, 300] و هذا يعني ان يكون الطرف الايمن للقيد الثاني يجب ان يكون اكبر من او يساوي 100 حتى يبقى الحل امثلا

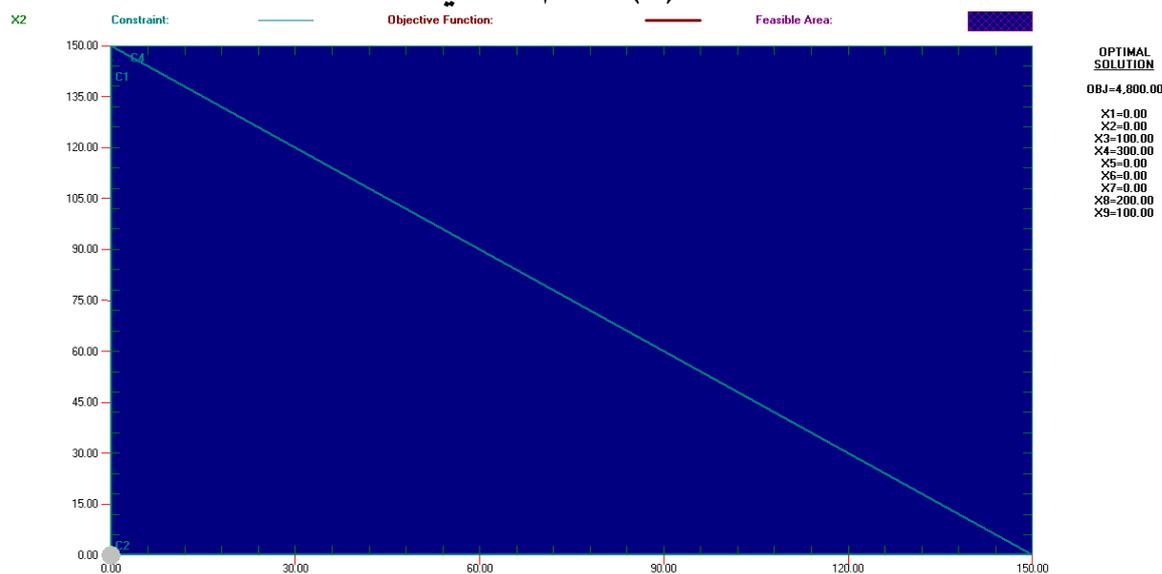
C - القيد الثالث: من الجدول نلاحظ ان الطرف الايمن للقيد يمنه اخذ القيم في الفترة [100, 350] و هذا يعني ان يكون الطرف الايمن للقيد الثالث يجب ان يكون اكبر من او يساوي 100 حتى يبقى الحل امثلا

D - القيد الرابع: من الجدول نلاحظ ان الطرف الايمن للقيد يمنه اخذ القيم في الفترة [100, M] و هذا يعني ان يكون الطرف الايمن للقيد الرابع يجب ان يكون اكبر من او يساوي 100 حتى يبقى الحل امثلا

E - القيد الخامس: من الجدول نلاحظ ان الطرف الايمن للقيد يمنه اخذ القيم في الفترة [300, M] و هذا يعني ان يكون الطرف الايمن للقيد الخامس يجب ان يكون اكبر من او يساوي 300 حتى يبقى الحل امثلا

F - القيد السادس: من الجدول نلاحظ ان الطرف الايمن للقيد يمنه اخذ القيم في الفترة [200, 400] و هذا يعني ان يكون الطرف الايمن للقيد السادس يجب ان يكون اكبر من او يساوي 200 حتى يبقى الحل امثلا

الشكل (1) الرسم البياني



جدول (11) يوضح مقارنة نتائج طرق النقل مع نتيجة البرمجة الخطية

البرمجة الخطية الحل الامثل باستخدام برنامج winQsb Optimal Solution	طريقة فوجل	طريقة اقل كلفة	طريقة الركن الشمالي الغربي	الطرق المستخدمة
4800	5250	5500	6300	النتائج

وعند مقارنة النتائج التي حصلنا عليها من طرائق النقل وطريقة البرمجة الخطية بتطبيق البرنامج Win QSB تبين ان افضل طريقة كانت هي طريقة البرمجة التي حققت الحل الامثل لا نها حققت اقل كلفة كلية ممكنة للنقل وكما في الجدول اعلاه

الاستنتاجات والتوصيات Conclusions and recommendations

الاستنتاجات conclusions:

من خلال دراسة الباحثة لطرائق النقل الموضحة اعلاه وتطبيقها على ارض الواقع فقد تبين من خلال النتائج المستخرجة ان افضل طريقة هي البرمجة الخطية في حل مشكلة النقل المتعددة المراحل لا نها حققت اقل كلفة كلية ممكنة للنقل والتي كانت قيمتها 480000 وان الطريقة التي جاءت بعد طريقة البرمجة الخطية هي الطريقة فوجل وذلك لا نها اعطت نتيجة اقرب الى الامثلية فقد كانت قيمته 525000 ومن هذا نستنتج انه يمكن الاعتماد على طريقة فوجل اعلاه لا نها اعطت النتيجة اقر بالحل الامثل في حالة البيانات المتعلقة بتوزيع مادة (دجاج المائدة) من الحقول الشركة الى المحافظات وانها افضل من الطرق الابتدائية (طريقة الركن الشمالي الغربي ، طريقة اقل كلفة) وهذه الطرق اعطت كلفة عالية جدا مقارنة بطريقتين التي ذكرت سابقا وتعتبر هذه الطريقة من الطرق الحديثة والمهمة جدا لصناع القرار الاداري والاقتصادي لأنها تسعى لتخفيض تكاليف النقل الى ادنى حد ممكن.

- 1- الاعتماد على البرمجة الخطية في حل مشاكل النقل بأنواعها حيث انها تعطي حل امثل في التعامل مع مشكلة النقل وكذلك تساهم وبشكل كبير في تقليل كلفة النقل الكلية مع تحديد الكميات المنقولة المثلى من الحقول الشركة الى المحافظات.
- 2- يجب توفير جميع البيانات من الشركة يتسنى للباحثين الاستفادة منها.
- 3- يجب ان تتبنى الشركة البحوث المميزة لتشجيع الباحثين للعمل في هذا المجال.
- 4- الاعتماد على الطريقة المقترحة في حل مشاكل النقل ذات المرحل المتعددة حيث انها تعطي مرونة كبيره جدا في التعامل مع مشكلة النقل وايجاد الحل المتوازن لها.

التوصيات

- 1- يجب ان تساعد الحكومية العراقية في دعم المنتج المحلي وعدم الاستيراد من دول الجوار مائدة الدجاج المحلي لدعم المنتج المحلي
- 2- يجب ان تساعد وزارة الزراعة العراقية الشركات المحلية في كافة المستلزمات التي تحتاجها لتطوير الحقول الانتاجية لكي تستطيع الشركة نقل المنتج للجميع محافظات العراق.
- 3- ويجب على الحكومة تشجيع الشركات الانتاجية من ما يزيد الخبرات وتوفير العمل للشباب العاطلة عن العمل.
- 4- امكانية دعم الشركة لي تصدر المنتج الى خارج البلد لتوفير العمالة الصعبة .

المصادر

- 1-المعلومات المستخدمة اعتمادا بيانات الشركة .
- 2- الكبيسي موفق محمد ، " بحوث العمليات وخوارزميات " ، الملكة الاردنية الهاشمية دار الحامد شفا بدران مقابل جامعة التطبيقية ، 1999 ، ص 84
- 3- د. احمد عبد اسماعيل الصفار ، ماجدة عبد اللطيف التميمي ، " بحوث العمليات " ، عمان / الاردن ، 2007 ، ص 148

- 4- دلال صادق الجواد ، حميد ناصر الفتال ، " بحوث العمليات " ، عمان / الاردن ، 2008 ، ص23 .
- 5- لحسن عبد الله باشيوة ، " بحوث العمليات " ، عمان / الاردن ، 2011 ، ص100 .
- 6- لحسن عبد الله بالشيوه ، " بحوث العمليات " ، عمان / الاردن ، 2011 ، ص102
- 7- محمد دباس الحميد، محمد العزاوي: " الأساليب الكمية في العلوم الإدارية"، دار اليازوري، الأردن، 2013

المصادر الانكليزية

8. Dakheel, F. I (1990), "A decision Support system for Single stage Markovian Queuing system " , Ph. D.thesis, University of Brad Ford; UK.
9. Evans, James R. (1992) "Applied production and operations Management " , 4th edition, west publishing company; USA.
10. Hiller, Frederick S.& Lieberman Gerald J., (2001), "Introduction to operations Research", 7th edition, McGraw-Hill Inc. USA.
11. sharma G & Abbs S.H , (2012) , " Optimun solution of Transpotion problem with the help of zero piont method" Intermtional journal of Engineering Research& Techology UERT ,vol,1,no,5,p.p:1-6.12. Takacs, Lagos. (1962), "Theory of Queuing", Oxford university of Press , USA.
- 12- Samuel A.E& Venkatachalathy M.,(2014) , " improving izpm for Unbalanced fuzzy transportation problems "

International journal of pure and applied Mathematics ,vol 94,
no 3,p.p :419.

13- Samuel A.E.,(2012) "improved zero piont method izpm for
Transportion problem",applied mathematical seiences,vol.6,no
.109,p.p 5426

14- Taha, Hamdy A. (1997) "*Operations Research: an
Introduction*", 6th edition, Prentice – Hall. Inc. New Jersey,
USA.

15- Takacs, Lagos. (1962), "*Theory of Queuing*", Oxford
university of Press , USA.

13- Taha H.A,(2007) ,"Operations .Resarch An Introdution" ,
8th , prantice Hall of India private Limited , New delhi.

استعمال الانحدار اللوجستي للتنبؤ بتأثير Berger على تخطيط كهربائية الدماغ

رائد فاضل محمد الحسني

تدريسي، كلية الإدارة والاقتصاد، جامعة المستنصرية، بغداد، العراق

raad@uomustansiriyah.edu.iq

أوركيد: 8702-7309-0002-0000

حسنين جليل نعمه الساعدي

تدريسي، جامعة تكنولوجيا المعلومات والاتصالات، العراق

Hasanien.1975@uoitc.edu.iq

الملخص

تستخدم تقنية تسجيل مخطط كهربائية الدماغ (Electroencephalogram) على نطاق واسع في مختلف المجالات، ومع زيادة التقدم التكنولوجي تم تحسين هذه التقنية وتطويرها، في هذا البحث تم توظيف أسلوب الانحدار اللوجستي متعدد الحدود في تحليل بيانات مخطط كهربائية الدماغ، وتوضيح فائدته في تمييز تأثير Berger بين العيون المغلقة والمفتوحة. إذ تم تحليل بيانات العينة قيد الدراسة والمكونة من 33 طالباً مراهقاً. ومن اهم الاستنتاجات المستخلصة من استخدام طريقة الانحدار اللوجستي المتعدد الحدود هو القدرة على الحصول على نتائج ذات مغزى مع تعزيز القدرة التنبؤية أيضاً. وعند مقارنة طريقة الانحدار اللوجستي ونهج التحليل العاملي، فإن الأخير يعطي نتائج أقل دقة.

الكلمات المفتاحية: الانحدار اللوجستي، التحليل العاملي، البيانات الصفية، تأثير Berger، حالة العيون.

Using Logistic Regression to Predict Berger Effect on Electroencephalography

Raaed Fadhil Mohammd

Lecturer, College of Administration and Economics,
University of Mustansiriyah, Baghdad, Iraq

raad@uomustansiriyah.edu.iq, orcid: 0000-0002-7309-8702

Hasanain Jalil NeamahAlsaedi

Lecturer, University of information technology and
communication, Baghdad,

Hasanien.1975@uoitc.edu.iq

Abstract

Electroencephalogram (EEG) recording technology is widely utilized in various sectors, and it has evolved and improved as technology has advanced. The multinomial logistic regression method was employed to analyze EEG data in this study, and its utility in distinguishing between the Berger effect in closed and open eyes was discussed. The sample's results, consisting of 33 adolescent students, were analyzed. One of the most noteworthy conclusions drawn from using the multinomial logistic regression method was the capacity to obtain meaningful results while also enhancing predictive ability. The latter produces less accurate results when comparing the logistic regression approach and the factor analysis approach.

Keywords: Logistic Regression, Factor Analysis, Class Data, Berger Effect, Eye Condition.

1- تمهيد

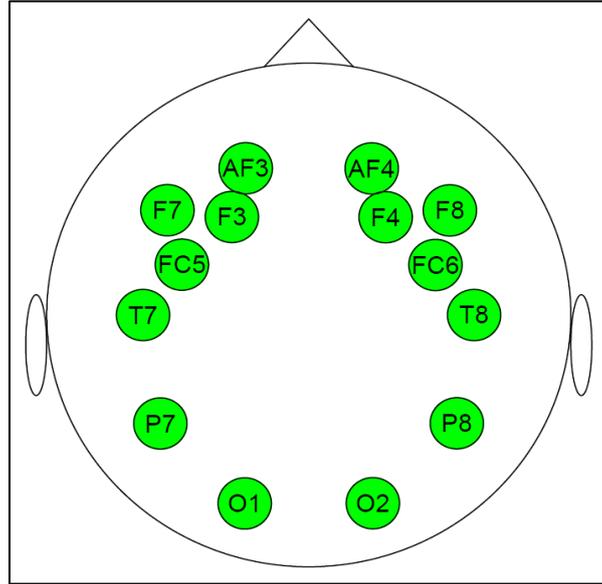
تعد تقنية تخطيط كهربائية الدماغ (Electroencephalography) من أقدم التقنيات المستخدمة لتسجيل ورسم خرائط الدماغ، والتي لا تزال تستخدم على نطاق واسع في مختلف العلوم كالطب والأعصاب وعلم النفس، كما إنها لا تزال الأفضل دقة من بين تقنيات التصوير الحديثة، ويعتبر تحليل كثافة المصدر الناتج من

تسجيلات تخطيط كهربائية الدماغ (EEG) أداة قوية وفعالة بشكل خاص في اكتشاف النشاط الكهربائي العصبي المتزامن [2]. يتم تسجيل مخططات كهربائية الدماغ (EEGs) دون أي إزعاج أو تدخل للمريض (أو العميل) وهي أرخص بكثير من أي تقنية أخرى لرسم أو تصوير خرائط الدماغ، مؤخراً تم توسيع تسجيل مخطط كهربائية الدماغ ليشمل مجالات جديدة، مثل علم النفس العصبي والهندسة وغيرها [1]. وعادة ما يتم إجراء التحليل عادةً بمساعدة البرامج التي تقدمها الشركات التي تتبع معدات تخطيط كهربائية الدماغ (EGG)، حيث تعتمد معظم هذه البرامج على تحليل فورييه (Fourier analysis)، وتقتصر بعض الدراسات الحديثة طرقاً رياضية جديدة لتحليل مخطط كهربائية الدماغ (EEG) من أجل اكتشاف أنماط جديدة أو لتكون بمثابة دعم تكنولوجي حديث [19]. في الوقت الحاضر توجد طرقاً مختلفة لتفسير البيانات المقدمة من مخطط كهربائية الدماغ (EEG)، في هذا البحث سيتم تقديم أسلوبين للتحليل يسهل التعامل معهما حسابياً من خلال إجراء عملياً ومباشراً لتحليل البيانات قابل للتكيف مع السلوكيات المختلفة أو الاستجابات المعرفية، الأسلوب الأول هو نموذج الانحدار اللوجستي الذي لا يحتاج إلى تحويلات فورييه أو ما شابه ذلك، وسيتم مقارنته بنموذج التحليل العائلي (Factor Analyze) من حيث دقة للتنبؤ والتمييز بين حالات العين (المفتوحة والمغلقة)، كما سيتم بناء نموذج رياضي يعتمد على الانحدار اللوجستي يسمح بتحديد ما إذا كان الشخص مغمض أو مفتوح العينين، يسمى تأثير (Berger) بناءً على تسجيل مخطط كهربائية الدماغ [2].

2- مخطط كهربائية الدماغ

لاتزال تقنية تسجيل مخطط كهربائية الدماغ (Electroencephalogram) تستخدم على نطاق واسع في مختلف المجالات، ومع التقدم التكنولوجي تم تحسين هذه التقنية، على سبيل المثال، تم تحديد أن الجهد القابل للاكتشاف خارج الجمجمة يتراوح بين 0.0002 مللي فولت و0.3 مللي فولت [6]. ويوضح الشكل (1) التالي مخطط تكوين الأقطاب الكهربائية في المعدات المستخدمة في تقنية (EEG)، حيث

نلاحظ أن عدد الأقطاب الكهربائية قد بلغ (14) قطباً منخفض نسبياً مقارنة بالمعدات المستخدمة في الدراسات السريرية. ومع ذلك، فإن هذا التكوين منخفض التكلفة يسمح بتسجيلات موثوقة وأعداد أكبر من الأقطاب الكهربائية [18].



الشكل (1): مخطط موضع القطب الكهربائي للدماغ

3- التحليل العاملي

يستخدم أسلوب التحليل العاملي (Factor Analyze) لتقليل مجموعة من المتغيرات المرصودة إلى عدد أقل من المتغيرات الكامنة (latent variables) والمعروفة أيضاً في الأدبيات باسم العوامل (Factors)، إذ تستخرج هذه التقنية الحد الأقصى من التباين المشترك من جميع المتغيرات وتضعها في درجة مشتركة. ويعد التحليل العاملي جزءاً من النموذج الخطي العام (GLM)، وتفترض هذه الطريقة أيضاً عدة افتراضات: هناك علاقة خطية، ولا يوجد تعدد خطي (multicollinearity)، وتتضمن المتغيرات الداخلة في التحليل، ويوجد ارتباط فعلي بين المتغيرات والعوامل [8]. بشكل عام، لنفترض أن متجهاً يضم (q) من المتغيرات، على سبيل المثال (xi)، لكل (n) من الأفراد، فإن التحليل العاملي يعبر عنه بالعلاقة (1) التالية:

$$x_i = \mu + Av_i + u_i, \quad i = 1, \dots, n \quad (1)$$

حيث إن (v_i) هو متجه المتغيرات الكامنة (latent variables) ذو الأبعاد $(s \times 1)$ ، وان (A) هي مصفوفة الثوابت المجهولة ذات الأبعاد $(q \times s)$ ، وان (u_i) هو متجه الاضطرابات غير المشاهدة ذو الأبعاد $(q \times 1)$. تم وضع الفروض الأتية:

1- متجهات المتغيرات الكامنة (v_i) مستقلة بحيث أن $v_i \sim N_s(0, I)$ ، وان $N_p(\mu, \Sigma)$ ، حيث يرمز الى التوزيع الطبيعي متعدد المتغيرات ذو الأبعاد (P) بمتوسط (μ) ومصفوفة تباين مشترك (Σ) .

2- متجهات البواقي (u_i) يتبع التوزيع الطبيعي بمتوسط صفر وتباين (ψ) ، حيث إن (ψ) هي مصفوفة قطرية. أي أن: $u_i \sim N_q(0, \psi)$

3- متجهات المتغيرات الكامنة (v_i) مستقلة عن متجهات البواقي (u_i) . لذلك، فإن (μ) يمثل متوسط المتغيرات (x_i) التي تتبع التوزيع الطبيعي، وتشير (C) الى مصفوفة التباين الخاصة بها، أي أن: $x_i \sim N_q(\mu, C)$ ، $i = 1, \dots, n$.
أن الحد الأقصى لعدد العوامل عندما لا تكون (q) كبيرة جدا (أقل من 10 تقريباً) يجب أن يكون أقل من نصف عدد المتغيرات ناقص واحد [15].

4- الانحدار اللوجستي

كما هو الحال مع جميع تحليلات الانحدار، فإن الانحدار اللوجستي (Logistic Regression) هو تحليل تنبؤي يتم إجراؤه عندما يكون المتغير التابع ثنائي الاستجابة، ويستخدم الانحدار اللوجستي لوصف البيانات وشرح العلاقة بين متغير ثنائي تابع وواحد أو أكثر من المتغيرات المستقلة الاسمية (nominal) أو الترتيبية (ordinal) أو الفترات (interval) أو النسبية. ويعتبر الانحدار اللوجستي أن متغير الاستجابة الثنائي (Y_i) يتبع توزيع برنولي، أي أن: [9]

$$P(Y_i = 1) = \pi_i$$

$$P(Y_i = 0) = 1 - \pi_i, \quad \forall i = 1, \dots, n$$

حيث إن $\pi_i \in (0, 1)$ ، وفي المشكلة قيد الدراسة بالنسبة لكل فرد (i) يتمثل متغير الاستجابة (Y_i) بالقيمتين (0 و 1) حالة العيون المغلقة والمفتوحة باحتمال (π_i) و $(1 - \pi_i)$ على التوالي. أما بالنسبة للحالة التي تكون فيها عينة البحث غير

متجانسة، فإنه يمكن تضمين مجموعة من المتغيرات المشتركة عددها $(p \geq 1)$ ، أي أن $x_i^T = (1, x_{i1}, \dots, x_{ip})$ ، ضمن الاحتمال (π_i) أي أن:

$$g(\pi_i) = \eta_i = x_i^T \beta = \beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p, \quad \forall i = 1, \dots, n \quad (2)$$

حيث يمثل $\beta = (\beta_0, \dots, \beta_p)^T$ متجه معاملات الانحدار المجهولة، وان (η_i) يمثل المتنبئات الخطية (linear predictors)، وان $g(u): \mathbb{R}(0, 1)$ هي دالة رتيبة معروفة في الأدبيات الإحصائية باسم دالة الربط (link function)، في المشكلة قيد الدراسة، يتم تحديد المتغيرات المشتركة من خلال قياسات الأقطاب الكهربائية المختلفة أو العوامل المبنية معها، كما يتم استخدام دالة التوزيع التراكمية (C.D.F) للنماذج القياسية. ويوضح الجدول (1) التالي دوال الربط التي تم استعمالها في هذه الدراسة [10].

جدول (1) دوال الربط المستخدمة في الدراسة

التوزيع	دالة الربط	الصيغة $g(u)$
logistic	Logit	$\log(u/(1-u))$
Normal	Probit	$\Phi^{-1}(u)$
Frechet	Cloglog	$\log(-\log(1-u))$
Cauchy	Cauchit	$\tan(\pi_i(u - \frac{1}{2}))$

يشير الرمز $(.) \Phi$ إلى دالة التوزيع التراكمية (C.D.F) للتوزيع الطبيعي القياسي، أما بالنسبة إلى $r \in \mathbb{N}$ ، الحالة التي يتم فيها تعريف x_i بالعناصر x_{ij}^k ، حيث إن $k = 1, \dots, r$ و $j = 1, \dots, p$ معروفة بالأدبيات الإحصائية باسم الانحدار متعدد الحدود (polynomial regression) من الرتبة (R) . بالتالي فإن دالة الإمكان (log-likelihood) لنموذج الانحدار اللوجستي تكون بالصيغة التالية: [13]

$$\ell(\beta) = \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \quad (3)$$

في ظل النهج الكلاسيكي، يمكن الحصول على مقدر الإمكان الأعظم لـ $\beta()$ ، وليكن $(\hat{\beta})$ ، كما يمكن الحصول على مثل هذه القيم عددياً باستخدام الخوارزميات، مثل خوارزمية (BFGS)، من ناحية أخرى، يمكن الحصول على الأخطاء المعيارية للمقدر $\hat{\beta}$ بناءً على المصفوفة (Hessian Matrix) المعطاة بالصيغة $[-\partial^2 \ell(\beta) / \partial \beta^T \partial \beta]$ [6].

5- منحنى (ROC)

يعد منحنى خاصية المستقبل والذي يشار اليه اختصاراً بالرمز (ROC)، تمثيلاً رسومياً مفيداً بشكل خاص لتقييم النماذج التنبؤية لتصنيف المتغيرات الثنائية. إذ نلاحظ أنه بمجرد ملاءمة النموذج $\forall c \in (0,1)$ فإنه يمكن إنشاء قاعدة تصنيف بالصيغة التالية [17]:

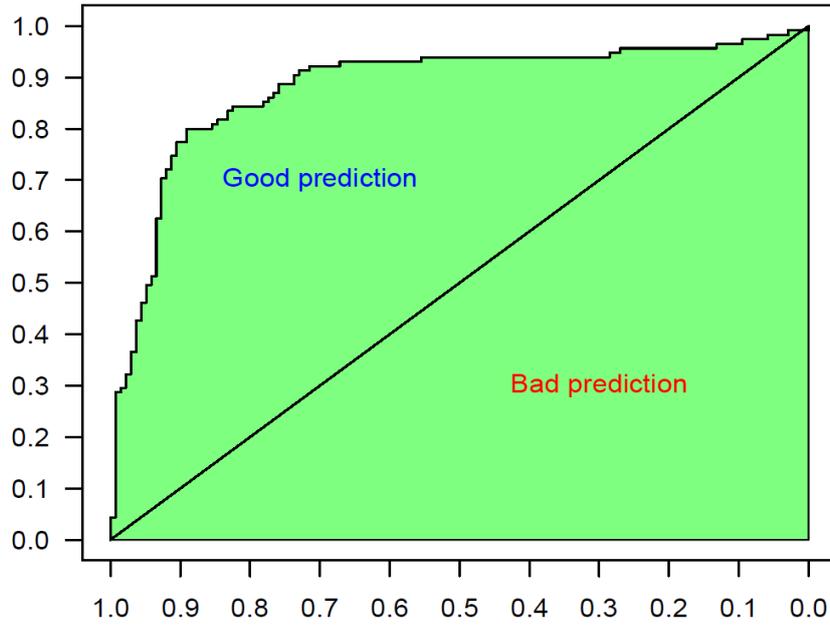
$$\hat{y}_i \begin{cases} 1, & \text{if } \hat{\pi}_i \geq c \\ 0, & \text{if } \hat{\pi}_i < c \end{cases} \quad (4)$$

كل قيمة من (c) ستنتج جدول توافقي كما في الجدول (2) التالي الذي يوضح بيانات تخطيط كهربائية الدماغ بالاعتماد على المعادلة (4).

جدول (2) تخطيط كهربائية الدماغ بناءً على المعادلة (4).

		المشاهدات		المجموع
		1	0	
القيمة التنبؤية	1	صحيح إيجابي (TP)	خطأ إيجابي (FP)	A'
	0	خطأ سلبي (FN)	صحيح سلبي (TN)	E'
	total	A	E	

وبهذه الطريقة، فإنه سوف يتم الحصول على قيم الحساسية $(TP / (TP + FN))$ والقيم النوعية $(TN / (TN + FP))$ المرتبطة بكل (c)، حيث يمكن رسم كلا المقياسين بيانياً كما في الشكل (2) التالي.



شكل (2) المخطط الرسومي لمنحنى (ROC)

إذ تعد المساحة الواقعة أسفل منحنى خاصية المستقبل (ROC) والمعروفة في الأدبيات الإحصائية باسم (AUC) مؤشراً لمقارنة النماذج المختلفة من حيث إمكانية التنبؤ بها، فكلما كانت المساحة تحت المنحنى أكبر، كلما كانت إمكانية التنبؤ بالنموذج المرتبط أفضل. أخيراً، فإن نقطة القطع (cut-off) المحددة لنموذج معين هي النقطة الموجودة داخل المنحنى الأقرب إلى النقطة التي تكون فيها الحساسية والنوعية (specificity) مساوية الى الواحد [14].

6- وصف البيانات:

تم الحصول على بيانات البحث من دراسة تقييم العلامة الحيوية كمؤشر لاضطراب نقص الانتباه / فرط النشاط، التي أجراها الباحث (Jahr) في عام 2020، حيث وضع المشاركون في الدراسة في غرفة ذات أقل قدر ممكن من الضوضاء باستثناء الضوضاء المحيطة، وطلب منهم النظر إلى جدار فارغ لمدة 5 دقائق، ثم يطلب منهم إغلاق أعينهم لمدة 5 دقائق، ومن القياسات تم اختيار البيانات بطريقة تضمن تجنب التشوهات المحتملة في بداية ونهاية التسجيلات [11].

ولتسجيل الحالات، تم استخدام نموذج (EmotivEpoC) + 14 قناة (EEG)، إذ يسمح هذا الجهاز بقياس الأقطاب الكهربائية (AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, O2)، المبينة في الشكل (1) حيث يتوافق موقع الأقطاب الكهربائية مع التكوين القياسي، وهذه الطريقة تعتبر الأكثر استخداماً والمعروفة باسم طريقة (10-20) [16][12].

يقوم الجهاز بجمع (128) قياساً في الثانية لكل قناة أو قطب، ثم يتم حساب كل قياس في القيمة الخاصة بالجهاز والتي يمكن تحليلها كقيمة أساسية تبلغ 4200 مللي فولت ويتم إضافة أو طرح الاختلافات في الجهد الكهربائي إليها. تم جمع (990) قياساً، أي ما يعادل (50%) لكل حالة (أعين مفتوحة وأعين مغلقة)، وبما أن استخدام قاعدة بيانات أولية بدون أي تصفية أمراً معقداً للغاية، لذلك تم تطبيق تقليل المشاهدات مع الأخذ بعين الاعتبار هامش القياس في نطاق 400 مللي فولت، وكما موضح بالجدول (3) التالي:

جدول (3) درجة وجنس العينة في بيانات تخطيط كهربائية الدماغ (EEG).

	الدرجة		المجموع
	الأولى	الثانية	
ذكور	7	10	17
إناث	10	6	16
المجموع	17	16	33

7- تحليل وتفسير النتائج:

7-1 التحليل العاملي: من خلال تطبيق التحليل العاملي تم الحصول على ثلاثة عوامل تفسر هذه العوامل (46.1%) من تباين البيانات، حيث يمكن من الجدول (4) ملاحظة الحمل العاملي (factorial load) للقطب الكهربائي للعوامل الثلاثة.

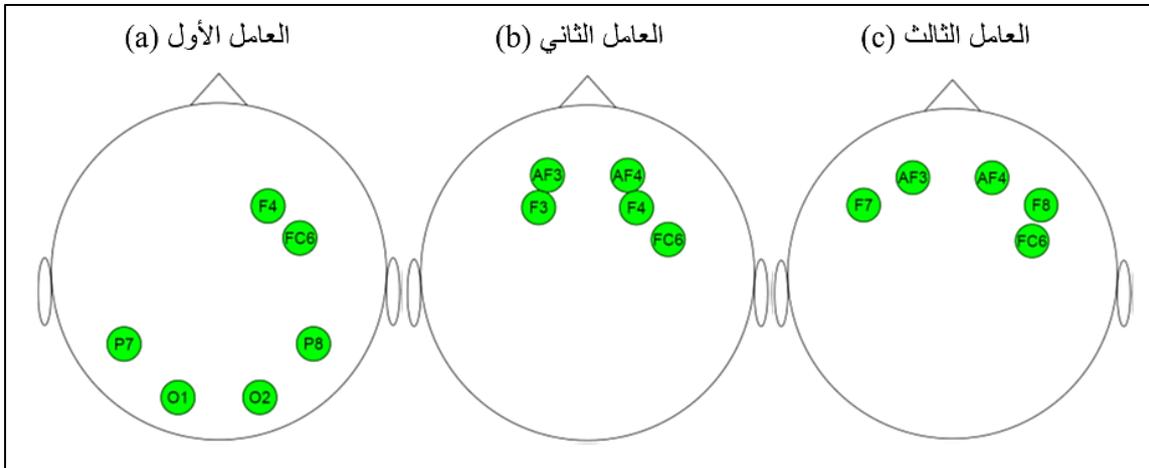
جدول (4) تركيب العوامل الثلاثة في بيانات تخطيط كهربائية الدماغ

القطب الكهربائي	العامل الأول	العامل الثاني	العامل الثالث
AF3	0.245	0.678	0.284
F7	0.330	0.363	0.358
F3	0.486	0.665	
FC5	0.222	0.243	
T7			
P7	0.654	0.344	0.104
O1	0.640	0.332	
O2	0.557	0.252	0.137
P8	0.684	0.345	0.124
T8	0.378	0.182	0.126
FC6	0.584	0.492	0.267
F4	0.554	0.657	0.172
F8			0.401
AF4	0.471	0.656	0.399
مجموع مربعات التغيرات	3.091	2.646	0.710
نسبة التباين	0.221	0.189	0.051
نسبة التباين التراكمي	0.221	0.410	0.461

في العامل الأول يتم تجميع الأقطاب الكهربائية (F4, FC6, P7, P8, O1, O2) معاً، ويرتبط هذا العامل بالجوانب البصرية الموجهة نحو شدة الضوء، والأنماط، وتمييز الألوان، والتكامل البصري، ومعالجة الحركة البصرية. أما العامل الثاني فيتكون من الأقطاب الكهربائية (AF3, AF4, F3, F4, FC6) وبذلك فإنه يرتبط بحركات العين العمودية، والوميض، وحركة العين السريعة (saccades)، والتخطيط للاستجابات المعرفية، وردود الفعل الحركية. ويمثل العامل (3) المتكون من الأقطاب

(AF3, AF4, F7, F8, FC6) المتمثلة بحركات العين، نظراً لأن المناطق المحددة مرتبطة بحركة العين العمودية، والرمش وتثبيطه، والحركات الجانبية للعينين، وكذلك الحركة السريعة للعين.

يوضح الشكل (3) الأقطاب الكهربائية التي تساهم في العوامل الثلاثة والتي تم الحصول عليها من التحليل العاملي، كما نلاحظ أن بعض الأقطاب الكهربائية قد تتكرر في العوامل، مثل (F4، FC6، AF3 وAF4). وبما أن القطب الكهربائي يمكنه الكشف عن طاقة معينة من قطب كهربائي قريب، وفي هذه الحالة يتم اقتراح دقة أفضل للإشارة. يشير العامل الأول إلى الأقطاب الكهربائية التي تقع بشكل رئيسي فوق القشرة القفوية (occipital cortex)، مما يشير إلى تنشيط القشرة البصرية. ويتكون العامل الثاني بشكل رئيسي من أقطاب كهربائية في منطقة الفص الجبهي تشير إلى الجوانب الإدراكية (cognitive aspects). أخيراً، يحتوي العامل الثالث على أقطاب كهربائية (F7 وF8) كعناصر خاصة، والتي تشير إلى الحركات الجانبية للعينين، أي الناتجة عن التنشيط العضلي.



الشكل (3) الأقطاب الكهربائية للتحليل العاملي في بيانات تخطيط كهربائية الدماغ

2-7 الانحدار اللوجستي: في هذا القسم الانحدار اللوجستي المتعلق بالتنبؤ بحالة العيون (مغلقة أو مفتوحة)، وبما أن هدفنا هو التنبؤ فإننا نركز على (AUC) كمقياس مناسب، فسيتم استخدام نموذجين: الأول استخدام العوامل التي تم الحصول

عليها في التحليل العائلي كمتغيرات مشتركة (covariates)، والثاني استخدام الأقطاب الكهربائية مباشرة كمتغيرات مشتركة.

أولاً- استخدام العوامل: في محاولة للحصول على نموذج أكثر اختزالاً وأكثر قابلية للتفسير، تم تطبيق الانحدار اللوجستي باستخدام العوامل التي تم الحصول عليها والموضحة في الشكل (3). يتم بناء النموذج بواسطة العوامل الثلاثة مع دوال الربط المختلفة (link functions) و (AUC) هي نفسها لجميع دوال الربط (0.5102)، أن نتائج التنبؤات لهذا النموذج كانت (50.15% و 51.18%) للعيون المغلقة والعيون المفتوحة، على التوالي. كما نلاحظ أن هذا النموذج ليس مفيداً جداً من الناحية التنبؤية. ولهذا السبب بحثنا عن بديل لاستخدام كل من المتغيرات بشكل مباشر (قبل إجراء التحليل العائلي)، فإننا نستخدم الانحدار متعدد الحدود باستعمال درجات مختلفة.

ثانياً- استخدام الانحدار المتعدد الحدود: تم استخدام النموذج اللوجستي بحيث يستخدم الأقطاب الكهربائية كمتغيرات مباشرة وانحدار متعدد الحدود، يوضح الجدول (5) قيم (AUC) لدرجات متعددة الحدود المختلفة (تتراوح من 1 إلى 10) ودوال الربط الأربعة المستعملة مع زيادة النسبة المئوية لكل منها فيما يتعلق بالدرجة السابقة.

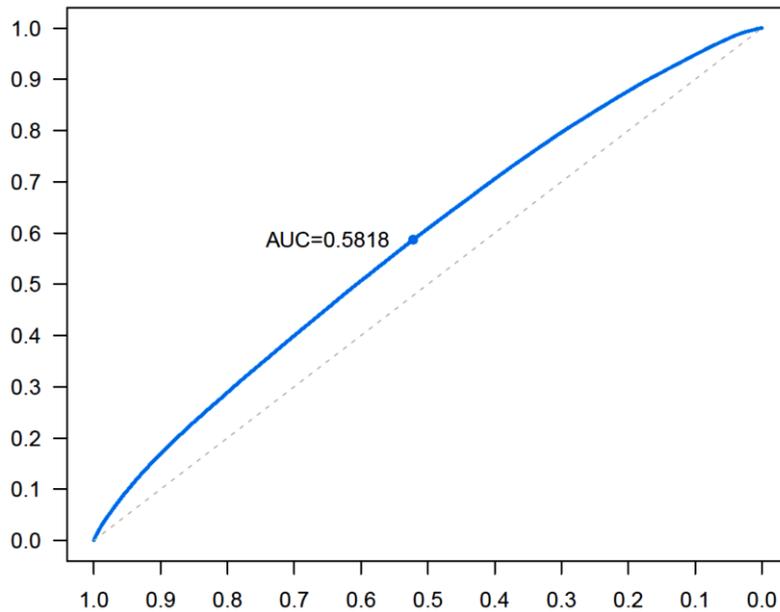
من الواضح أنه من بين دوال الربط المختلفة، توفر دالة الربط (Cauchit Link) أفضل النتائج لأن (AUC) أكبر بالنسبة لجميع درجات متعدد الحدود المدروسة. مع الأخذ في الاعتبار أن زيادة (1) في درجة متعدد الحدود تزيد من عدد المعلمات في الأقطاب الكهربائية المقاسة، تم اختيار درجة متعددة الحدود قدرها (4) لأن الزيادة المئوية للمساحة تحت المنحنى أقل من (2%) للدرجات الأعلى. علاوة على ذلك، باستخدام دالة الربط ومتعدد الحدود من الدرجة الرابعة (ولنفرض اسم هذا النموذج MG4)، فإنه يقدم (AUC) بنسبة (58.18%) وبالمقارنة مع متعدد الحدود من الدرجة العاشرة (ولنفرض اسم هذا النموذج MG10) بنسبة (1.51%) فقط. لذلك، يُفضل النموذج (MG4) على النموذج (MG10). ويوضح الشكل (4) التالي كيف يؤثر نمو متعدد الحدود بشكل أقل وأقل على (AUC) لجميع ودوال الربط.

تم تنقيح النموذج (MG4) من خلال إزالة المعاملات غير المهمة بالاعتماد على معيار معلومات كياكي (AIC)، وإزالة خمس معاملات تلك المتعلقة بـ (AF34, F33, F72, O12, F83)، حيث تشير (XX) إلى المعامل المرتبط بالقطب (XX) المرفوع للقوة (j).

كما يعرض الشكل (4) منحنى (AUC) الذي تم إنشاؤه باستخدام نموذج الانحدار متعدد الحدود (polynomial regression model) من الدرجة الرابعة، حيث تكون نقطة القطع المثالية هي (0.64847) حيث إن:

$$\hat{y}_i \begin{cases} 1 \text{ (مفتوحة عيون)}, & \text{if } \hat{\pi}_i \geq 0.64847 \\ 0 \text{ (مغلقة عيون)}, & \text{if } \hat{\pi}_i < 0.64847 \end{cases} \quad (5)$$

حيث يمثل ($\hat{\pi}_i$) الاحتمال المقدر لهذا النموذج.



الشكل (4) يوضح منحنى (ROC) لنموذج (MG4)

إذ تسمح لنا نقطة القطع (cut-off) بالتنبؤ بشكل صحيح بنسبة (56.37%) من الحالات لحالة العيون المغلقة و(54.49%) لحالة العيون المفتوحة، ويوضح الجدول (5) التالي المساحة تحت منحنى (ROC) لكل درجة متعدد الحدود لدوال الربط المختلفة في بيانات (EEG)، حيث يتم عرض الزيادة في (AUC) بين قوسين بالنسبة إلى الدرجة الأخيرة لمتعدد الحدود المستخدم.

جدول (5) قيم (AUC) لكل درجة متعدد الحدود لدوال الربط المختلفة في بيانات (EEG)

درجة متعدد الحدود	Link function دالة الربط			
	Cloglog	logit	probit	Cauchit
1	0.5322(-)	0.5325(-)	0.5325(-)	0.5329(-)
2	0.5547(↑4.2%)	0.5557(↑4.4%)	0.5552(↑4.3%)	0.5585(↑4.8%)
3	0.5661(↑2.1%)	0.5673(↑2.1%)	0.5667(↑2.1%)	0.5703(↑2.1%)
4	0.5773(↑2.0%)	0.5775(↑1.8%)	0.5767(↑1.8%)	0.5818(↑2.0%)
5	0.5799(↑0.5%)	0.5803(↑0.5%)	0.5797(↑0.5%)	0.5837(↑0.3%)
6	0.5841(↑0.7%)	0.5849(↑0.8%)	0.5844(↑0.8%)	0.5881(↑0.8%)
7	0.5846(↑0.1%)	0.5859(↑0.2%)	0.5854(↑0.2%)	0.5889(↑0.1%)
8	0.5894(↑0.8%)	0.5904(↑0.8%)	0.5899(↑0.8%)	0.5936(↑0.8%)
9	0.5906(↑0.2%)	0.5918(↑0.2%)	0.5913(↑0.2%)	0.5947(↑0.2%)
10	0.5933(↑0.5%)	0.5943(↑0.4%)	0.5938(↑0.4%)	0.5969(↑0.4%)

ويوضح الجدول (6) التالي مصفوفة التوافق (Contingency Matrix) الناتجة من استخدام دالة الربط (Cauchit Link) ونموذج متعدد الحدود من الدرجة الرابعة.

جدول (6) مصفوفة التوافق لنموذج (MG4) لبيانات (EEG)

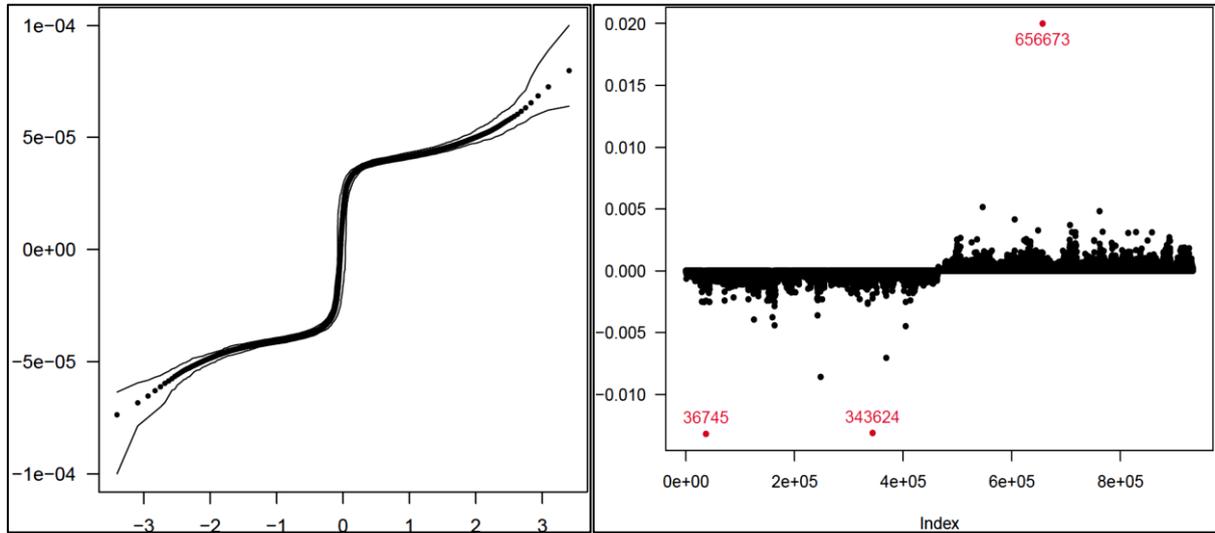
حالة العين	المشاهدات (Observations)	
	عيون مغلقة	عيون مفتوحة
عيون مغلقة التنبؤ (Prediction)	56.37%	45.51%
عيون مفتوحة	43.63%	54.49%
المجموع	100%	100%

بناءً على هذا النموذج، تم بناء الدالة (eyes) والتي تستخدم الأقطاب الـ 14 الموضحة في الشكل (1). إذ توفر الدالة تقديراً لـ (π) والتنبؤ بناءً على قاعدة التصنيف المقدمة في المعادلة (3)، مما ينتج عنه قائمة بالاحتمالات والتنبؤات. كما

نلاحظ أن نموذج متعدد الحدود من الدرجة الرابعة (G4.2) يتمتع بحساسية (sensitivity) وخصوصية (specificity) تبلغ (56.37% و 54.49%) على التوالي.

8- تحليل النموذج

تم توظيف لغة البرمجة الإحصائية (R) لتحليل البواقي (residual) وحساب مسافات (COOK) للنموذج المختار (G4.2) ويوضح الشكل (5-أ) كيف تقع بواقي مركب الانحراف في نطاقات الثقة المعنية، مع نسبة صغيرة من النقاط عند الأطراف خارج تلك الفترة. ويوضح الشكل (5-ب) مسافات (COOK) للنموذج المختار حيث نلاحظ أن هنالك ثلاث نقاط تبرز (36745، 343624، 656673). هذه النقاط هي بواقي مركب الانحراف، عند استبعاد هذه المشاهدات الثلاث خارج المجموعة لم تكن النتيجة مختلفة مما يعني أنها لم تكن مؤثرة.



الشكل (5-أ) تحليل البواقي (residual)

الشكل (5-ب) مسافات (COOK) للنموذج المختار

9- الاستنتاجات

- تم بناء نموذج رياضي يعتمد على الانحدار اللوجستي يسمح بتحديد ما إذا كان الشخص مغمض أو مفتوح العينين بناءً على تسجيل مخطط كهربائية الدماغ. إذ يمكن استخدامه لدى الأشخاص الذين يعانون من نوع ما من الإعاقة، أو عدم القدرة على استخدام اللغة، أو نقص الحافز للتعبير عن نواياهم.

- أن النموذج الناتج عن الانحدار اللوجستي متعدد الحدود يقدم قدرة تنبؤية أكبر من نموذج الانحدار الناتج عن ثلاثة عوامل، حيث يظهر زيادة تزيد عن (2.2%) في المساحة تحت المنحنى، في أي دالة ارتباط.
- نستنتج أن القدرة التنبؤية (predictive capacity) أفضل من الانحدار اللوجستي مع العوامل، ولكنها لا تزال غير مرتفعة بما يكفي، كما أن البواقي تتناسب بشكل جيد مع النموذج المختار، على الرغم من ملاحظة بعض النقاط الشاذة.

10- التوصيات

- بناءً على النتائج التي تم الحصول عليها، نوصي بإنشاء نموذج بتوزيع شبه معلمي (semiparametric) للتحقق من مدى ملاءمته للبيانات بشكل أفضل، مع إمكانية الحصول على أفضل تنبؤ ممكن.
- من الضروري إجراء مزيد من الدراسات باستخدام معايير (حالة العيون) الأكثر وضوحاً، والجمع بين الطريقة المختارة مع طرائق أخرى من أجل إنشاء برنامج جاهز للاستخدام يضم نطاقات تنبؤ عالية ومتنوعة.

References:

1. Alagia Gimeno, R. A. (2018). *Procesamiento de artefactosen EEG para aplicaciones de comunicación y control* (Doctoral dissertation, Universitat Politècnica de València).
2. Berger, H. (1929). Elektrenkephalogramm des Menschen. *Archiv f. Psychiatrie*, 87, 527-570.
3. Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*, 97(3), 210-220.

4. EMOTIV EPOC + 14 Channel Mobile EEG Headset EPOC+ User Manual Available online: www.emotiv.com/epoc/ (accessed on 01 march 2022).
5. Felten, D.; Shetty, A.; Netter, F. Netter: (2010). Atlas de neurociencia, Elsevier Masson, Barcelona, 2010.
6. Fletcher, R. (2000). *Practical methods of optimization*. John Wiley & Sons.
7. Guerrero, M. C., Parada, J. S., & Espitia, H. E. (2021). Principal components analysis of eeg signals for epileptic patient identification. *Computation*, 9(12), 133.
8. Hair, J.; Anderson, R.; Babin, B.; Black, W. (2019). *Multivariate Data Analysis*, 8a ed.; Cengage Learning.
9. Hilbe, J. M. (2009). *Logistic regression models*. CRC press.
10. Hosmer, D. W., & Lemeshow, S. (2020). *Applied Logistic Regression*, Hoboken, New Jersey, John Wiley & Sons.
11. Jahr, M. (2020). Evaluación de un biomarcador como indicador de Trastorno de Déficit de Atención/ Hiperactividad (TDAH) Un estudio correlacional. Tesis para optar al grado de Magíster en Psicología Clínica. Universidad de Santiago de Chile, Chile.
12. Klem, G. H. (1999). The ten–twenty electrode system of the international federation. The international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.*, 52, 3–6.

- 13.Koenker, R., & Yoon, J. (2009). Parametric links for binary choice models: A Fisherian–Bayesian colloquy. *Journal of Econometrics*, 152(2), 120–130.
- 14.Lieli, R. P., & Hsu, Y. C. (2019). Using the area under an estimated ROC curve to test the adequacy of binary predictors. *Journal of Nonparametric Statistics*, 31(1), 100–130.
- 15.Lloret–Segura, S., Ferreres–Traver, A., Hernández–Baeza, A., & Tomás–Marco, I. (2014). El análisis factorial exploratorio de los ítems: un guía práctica, revisada y actualizada. *Anales de psicología/annals of psychology*, 30(3), 1151–1169.
- 16.Silverman, D. (1963). The rationale and history of the 10–20 system of the International Federation. *American Journal of EEG Technology*, 3(1), 17–22.
- 17.Tharwat, A. (2021). Classification assessment methods. *Applied computing and informatics*, 17(1), 168–192.
- 18.Troy M, L., Joseph T, G., & Daniel P, F. (2012). How many electrodes are really needed for EEG–based mobile brain imaging?. *Journal of Behavioral and Brain Science*, 2012.
- 19.Xu, B., Li, W., Liu, D., Zhang, K., Miao, M., Xu, G., & Song, A. (2022). Continuous hybrid BCI control for robotic arm using noninvasive electroencephalogram, computer vision, and eye tracking. *Mathematics*, 10(4), 618.

التحول الرقمي ومساهمته في إدارة ملف السكان تجارب بعض الدول العربية

نيفين ناجي أنيس مدرس إحصاء ورياضة

معهد العبور العالي للإدارة وعلوم الحاسب ونظم المعلومات، قسم علوم الحاسب،
العلوم الأساسية جمهورية مصر العربية

nnagy@oi.edu.eg

ملخص

للرقمنة فوائد هائلة في تحسين إدارة الملف السكاني بتوفير أنظمة أكثر دقة، سرعة، ومرونة. تساعد في تحسين اتخاذ القرارات، تعزيز الشفافية، وتحقيق الاستجابة السريعة للطوارئ حيث يمكن للحكومات استخدام الرقمنة لتحسين جودة الخدمات العامة وتخصيصها بشكل أفضل بناءً على بيانات دقيقة، مما يساهم في التنمية المستدامة والتخطيط للمستقبل، تعتبر الدراسة وصفية لمدي مساهمة التحول الرقمي في ادارة الملف السكاني ببعض الدول العربية حيث تم استخدام التحليل الوصفي لبيانات استمارة وأهم النتائج أن نسبة قليلة جدا من الدول لديها نسبة 100% من عناصر هذه البنية، وتركزت نسب توافر أغلب العناصر ما بين (50% - 80%) وبينت النتائج أن عدم توفر الكوادر المدربة والمؤهلة تأتي على رأس الأسباب التي تعيق الاستفادة من التحول الرقمي بينما نسبة 63% من الدول ترى أنها قد حققت أقصى استفادة من التحول الرقمي في الملف السكاني، وكانت هذه الاستفادة في مجالات التخطيط الجيد لا مركزيا، والمتابعة والتقييم، والاستعداد للتحول الديموغرافي وبينت أن الأهداف التنموية التي ساعد التحول الرقمي على تحقيقها هي الهدف (الثالث، الرابع، الثامن، والسادس عشر) كما ساهمت في تحسين جودة الحياة بنسبة 75%، وأهم عناصرها (توفير الوقت والجهد، وتوفير التكلفة من

خلال سرعة العمل واهم التوصيات إجراء مراجعة إقليمية لمفهوم قواعد البيانات السكانية على مستوى الدول العربية التي تساعد في إدارة الملف السكاني والتوعية بها وتوحيد مؤشراتها على المستوى الإقليمي والمحلي لمتابعة الوضع السكاني وتعزيز التعاون العربي المشترك

الكلمات الافتتاحية: التحول الرقمي، التخطيط، جودة الحياة، مؤشر - المتابعة والتقييم

Abstract

Digitization offers significant benefits in improving population file management by providing more accurate, faster, and more flexible systems. It aids in decision-making, enhances transparency, and enables rapid emergency response, contributing to sustainable development and effective planning. This study employs descriptive analysis to assess the role of digital transformation in managing population files in selected Arab countries. Key findings indicate that only a limited number of countries have 100% of the required digital infrastructure elements, while most countries have availability rates ranging between 50% and 80%. The lack of trained and qualified personnel is identified as the primary challenge preventing full utilization of digital transformation. Additionally, 63% of the surveyed countries believe they have maximized the benefits of digital transformation in population management, particularly in decentralized planning, monitoring and evaluation, and preparedness for demographic transitions. Furthermore, digital transformation contributes to achieving Sustainable Development Goals (SDGs 3, 4, 8, and 16) and has improved the quality of life by 75%. It also helps in saving time and effort, reducing costs, and enhancing efficiency. The key recommendations include conducting a regional review of the concept of population databases, standardizing indicators, and strengthening Arab cooperation.

Keywords: (Digital Transformation, Planning, Quality of Life, Indicator, Monitoring and Evaluation)

1. مقدمة

الرقمنة أصبحت من الأدوات الأساسية في العصر الحديث التي تؤثر بشكل عميق في مختلف المجالات، بما في ذلك إدارة الملف السكاني. إدارة الملف السكاني تشمل جمع وتحليل البيانات السكانية للأفراد في المجتمع، مثل العمر والجنس والحالة الاجتماعية والتعليم والمهنة. الرقمنة تقدم العديد من الفوائد التي تؤثر بشكل إيجابي على هذه الإدارة، وفيما يلي أهمية الرقمنة وتأثيرها على إدارة الملف السكاني تحسين الكفاءة والدقة المتمثلة في السرعة والدقة وإدارة البيانات الضخمة وتحسين اتخاذ القرارات المتمثلة في دعم اتخاذ القرارات المبينة على البيانات وتحليل البيانات في الوقت الفعلي باستخدام أحدث التقنيات كما ترجع أهمية الرقمنة الى تيسير الوصول الى البيانات وذلك من خلال التكامل بين الجهات المختلفة المنتجة للبيانات وايضا تساعد في تحسين الخدمات العامة من خلال تخصيص الخدمات الحكومية والتفاعل مع المواطنين وتساعد ايضا في مكافحة الفساد والشفافية وتسهيل التحديثات والتحسينات المستمرة كما تدعم الرقمنة الابتكار والبحث

وبناء على ما سبق تبرز ضرورة التأكيد على أهمية توافر البيانات كمصدر قوة رئيسي لأي مجتمع قائم على المعرفة ففي حال تمت إدارتها على نحو صحيح فسوف تسهم إسهاما مؤثرا في التنمية الاقتصادية والاجتماعية، كما ستساعد على القيام بدراسات استشرافية وتيسير عملية اتخاذ القرار، ويمكن للبيانات الضخمة أن تدعم أيضا سياسات وبرامج أكثر تلبيةً للاحتياجات وأكثر دقة، وتناول عدد كبير من البحوث والدراسات السابقة موضوع التحول الرقمي في مختلف المجالات، منها دراسة (خالد الوالي 2019)¹ بعنوان "الاقتصاد الرقمي في الوطن العربي: الواقع والمستقبل" حيث ناقشت الدراسة تأثير الرقمنة على النمو الاقتصادي في الدول العربية، مع التركيز على قطاعات مثل التجارة الإلكترونية والخدمات المالية. أظهرت الدراسة أن الرقمنة تساهم في زيادة الإنتاجية وتوفير فرص عمل جديدة وتوصلت الى خلصت الدراسة إلى أن الدول العربية تحتاج إلى تبني استراتيجيات رقمية شاملة لتعزيز النمو الاقتصادي والتنافسية العالمية.

كما تناولت دراسة (UNIDO2020)² بعنوان "القطاع العام نحو الرقمنة: التحديات والفرص في الدول العربية" حيث تناولت الدراسة تحليل التحديات التي تواجه الدول العربية في تحويل خدمات القطاع العام إلى خدمات رقمية، مع التركيز على نقص البنية التحتية التكنولوجية وعدم كفاية الكوادر المؤهلة. كما أشارت إلى الفرص التي توفرها الرقمنة في تحسين كفاءة الخدمات الحكومية وتقليل الفساد. وأوصت الدراسة بضرورة زيادة الاستثمار في البنية التحتية الرقمية وتدريب الكوادر البشرية لتحقيق التحول الرقمي الفعال.

كما تناولت دراسة (سفيان قعلول وآخرون، 2020)³، بعنوان "الاقتصاد الرقمي في الدول العربية: الواقع والتحديات 2020"، قيام بعض الدول العربية بتقييم ذاتي لتجاربها واستجاباتها في التحول الرقمي في بعض المجالات وخاصة تحديد مؤشرات الاقتصاد الرقمي، في كل من (الأردن، والإمارات، والبحرين، والسودان، والصومال، وعمان، وفلسطين، وقطر، ومصر)، وفي استخدام التطبيقات الرقمية في حالات الطوارئ أثناء جائحة كورونا في بعض الدول العربية وهي (تطبيقات وزارات الصحة في كل من مصر، والإمارات، والسعودية وسلطنة عمان، وتطبيق بهيئة المعلومات والحكومة الإلكترونية في البحرين) ، أهم توصيات الدراسة إنشاء جهة متخصصة معينة بغرض تسريع وتيرة التحول الرقمي في القطاعات كافة بما يضمن زيادة مستويات الإنتاجية والتنافسية للقطاعات الاقتصادية المختلفة.

كما بينت دراسة (إيلي أحمد 2020)⁴ بعنوان "الرقمنة كأداة لتمكين المرأة في الدول العربية" حيث ناقشت الدراسة دور الرقمنة في تمكين المرأة في الدول العربية، مع التركيز على زيادة فرص العمل عن بعد وتحسين الوصول إلى التعليم والخدمات الصحية. أظهرت الدراسة أن الرقمنة يمكن أن تساهم في تقليل الفجوة بين الجنسين وأوصت الدراسة بضرورة تعزيز السياسات التي تدعم مشاركة المرأة في الاقتصاد الرقمي.

كما تناولت دراسة (عمر الفاروق 2021)⁵ بعنوان "التحديات الرئيسية للتحول الرقمي في الوطن العربي" ركزت الدراسة على تحليل التحديات الرئيسية التي تواجه الدول العربية في تحقيق التحول الرقمي، بما في ذلك نقص البنية

التحتية التكنولوجية، وضعف التشريعات الرقمية، ونقص الوعي الرقمي وخلصت الدراسة إلى أن الدول العربية تحتاج إلى تبني استراتيجيات شاملة لمواجهة هذه التحديات وتحقيق التحول الرقمي.

وبينت دراسة (فاطمة الزهراء 2021)⁶ بعنوان "التعليم الرقمي في الوطن العربي: التحديات والفرص" حيث ركزت الدراسة على تحليل تأثير الرقمنة على جودة التعليم في الدول العربية، مع التركيز على استخدام التكنولوجيا في الفصول الدراسية والتعليم عن بعد. أشارت الدراسة إلى أن الرقمنة يمكن أن تحسن جودة التعليم وتوفير فرص تعلم أكثر مرونة وأوصت الدراسة بضرورة تطوير منصات تعليمية رقمية وتدريب المعلمين على استخدام التكنولوجيا في التعليم.

ودراسة (أمل بنت عبد الله بن ابراهيم آل ابراهيم، 2022)⁷ بعنوان "تجربة التحول إلى التعليم عن بعد في التعليم العام إثر جائحة كورونا (Covid19) من وجهة نظر الطلاب 2022"، وكان الهدف من الدراسة معرفة كيفية التحول إلى التعليم عن بعد خلال جائحة كورونا في التعليم العام للفصل الدراسي الثاني من عام 2020/2019، في المملكة العربية السعودية، وطبقت على عينة بلغت 304 طالب وطالبة، وتوصلت الدراسة إلى أن حوالي 98% من الطلاب استمروا بالتعليم رغم الإغلاق المفاجئ للمدارس ولا يوجد فروق معنوية بين الطلاب والطالبات ولكن تبين أن هناك فروق معنوية بين المدارس الأهلية والحكومية في الدعم الفني والبنية التحتية، وكانت أهم توصيات الدراسة عقد ورش عمل للطلاب للتدريب، ورفع مستوى تقبلهم للتعليم عن بعد، واستغلال مهاراتهم الرقمية في التعليم، وكذلك توعية أولياء الأمور حول التعليم عن بعد وأهميته في المستقبل.

وأوضحت دراسة (محمد السيد 2022)⁸ بعنوان "التحول الرقمي في القطاع الصحي: دراسة حالة لدول الخليج العربي" حيث تناولت الدراسة تحليل تأثير الرقمنة على القطاع الصحي في الدول العربية، مع التركيز على استخدام السجلات الصحية الإلكترونية والطب عن بعد. أظهرت الدراسة أن الرقمنة يمكن أن تحسن كفاءة الخدمات الصحية وتقليل التكاليف وخلصت الدراسة إلى أن الدول العربية تحتاج إلى تبني سياسات رقمية في القطاع الصحي لتحسين جودة الخدمات الصحية.

وبينت دراسة (نيفين ناجي 2024)⁹ "فجوات تطبيق التحول الرقمي في المجالات الخدمية المقدمة للسكان: تجارب لبعض الدول العربية" أوضحت النتائج أن المؤشر العام لتطبيق الحكومات العربية للتحول الرقمي في المجالات المختلفة بلغ حوالي 70%، وكانت أعلى قيمة له في مجال التعليم بنسبة بلغت 79%، وأقل قيمة في مجال معالجة الأثر البيئي بنسبة 60%، وبلغت نسبته في مجال مكافحة الجريمة حوالي 67%. وتوصى الدراسة وضع استراتيجية إقليمية للتحول الرقمي تضمن التزام كافة الدول العربية بمسار التحول الرقمي في مجال السكان، وتعزيز التعاون العربي المشترك في مجال تبادل الخبرات بين الدول العربية الرائدة في مجال الرقمنة.

2- إشكالية الدراسة

تعتبر الرقمنة في العصر الحديث قوة كبيرة وهائلة يمكن استغلالها في مجال العمل السكاني وإدارة الملف السكاني بشكل عام لذلك تمثلت مشكلة الدراسة عدم وضوح الصورة حول مساهمة الرقمنة في إدارة الملف السكاني في الدول العربية وستحاول الدراسة الإجابة على ذلك.

3- تساؤلات الدراسة

- ما هو مدى تقدم الدول في التحول الرقمي في إدارة الملف السكاني؟

4- أهداف الدراسة

- تهدف الدراسة الى تحديد مدى التقدم بالتحول الرقمي في إدارة الملف السكاني.

5- أهمية الدراسة

تتمثل أهمية هذه الدراسة في أن مخرجاتها يمكن أن تساهم في مساعدة صانعي القرار وواضعي السياسات والبرامج في الوقوف على الوضع الحالي للمنطقة العربية في مجال الثورة الرقمية ومساهمتها في إدارة الملف السكاني.

6- أدوات الدراسة

استخدمت الدراسة بيانات تم جمعها بهدف رصد وقياس تجارب الدول في التحول الرقمي وتأثير ذلك على ادارة الملف السكاني، تم تصميم استبيان يحتوي على ثلاثة أجزاء تشمل المواضيع التالية: (أسئلة عامة، الخدمات المتاحة بها التحول الرقمي - التحول الرقمي وادواته ومساهمته على ادارة الملف السكاني السكان) بمعرفة جامعة الدول العربية.

7- منهجية الدراسة

تعتبر هذه الدراسة وصفية، وتم الاعتماد فيها على المنهج الوصفي التحليلي لأسئلة الاستمارة، لوصف مدي مساهمة التحول الرقمي بإدارة الملف السكاني وتعتبر هذه الورقة مستنبطة من تقرير "رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على الملف السكاني وذلك تنفيذاً للقرار رقم (10) الصادر عن الدورة الرابعة للمجلس العربي للسكان والتنمية لعام 2022.

8- حدود الدراسة

• الحدود المكانية والزمنية: شملت الدراسة (10) دول عربية هي: (المملكة الهاشمية الأردنية، الجمهورية التونسية، المملكة العربية السعودية، جمهورية العراق، سلطنة عمان، دولة الكويت، الجمهورية اللبنانية، جمهورية مصر العربية، المملكة المغربية، الجمهورية الاسلامية الموريتانية) عام 2023.

9- النتائج

استجابت (8) دول من بين العشر (10) المشاركة في الدراسة للأسئلة الخاصة بالتحول الرقمي والبنية التحتية، كما تبين عدم استكمال استجابات الثمان دول على كل الأسئلة الخاصة بالتحول الرقمي وتأثيره على الملف السكاني، ومن الملاحظ اختلاف الأعداد باستجابات كل سؤال في هذا القسم، وأظهرت نتائج تحليل البيانات

أن حوالي 57% فقط من العينة المستجيبة لديها بنية معلوماتية لعملية التحول الرقمي في الجهة المنوطة بالسكان، و88% من عينة الدراسة سمعت أو اطلعت على استراتيجية التحول الرقمي للدولة (المؤسسة المسؤولة عن ملف السكان)، واعتبرت أن الرؤية الخاصة بالتحول الرقمي واقعية/ عملية وواضحة ومفهومة، كما أقرت 75% من الدول المستجيبة أن (الاجراءات الحكومية تتم بين الحكومات ووزارتها - أن هناك دورات تدريبية للعاملين بالتوعية بأهمية التحول الرقمي وكيفية تطبيقه والاستفادة منه في الدولة وأيضاً بتوافر أجهزة الكترونية حديثة تتيح التحول الرقمي، (الجدول رقم (1)، (2).

جدول رقم (1): التوزيع النسبي تساؤلات البنية التحتية

لا	نعم	تساؤلات البنية التحتية
النسبة	النسبة	
42.9	57.1	هل يوجد بنية معلوماتية لعملية التحول الرقمي في الجهة المنوطة بالسكان؟
12.5	87.5	هل سمعت أو اطلعت على استراتيجية التحول الرقمي للدولة (المؤسسة المسؤولة عن ملف السكان)؟
12.5	87.5	هل تعتبر الرؤية الخاصة بالتحول الرقمي واقعية/ عملية وواضحة ومفهومة؟
25	75	هل تتم الإجراءات بين الحكومة ووزارتها بشكل رقمي؟
25	75	هل يتم تنفيذ دورات تدريبية للعاملين للتوعية بأهمية التحول الرقمي وكيفية تطبيقه والاستفادة منه في الدولة؟
25	75	هل تتوافر أجهزة الكترونية حديثة تتيح التحول الرقمي داخل الدولة؟
-	100	هل تتوافر قاعدة بيانات دقيقة ومتكاملة تخدم الملف السكاني بالدولة؟

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان.

* ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

جدول رقم (2): التوزيع العددي لقواعد البيانات المتوفرة بالجهة التي تخدم الملف السكاني

عدد الدول	قواعد البيانات المتوفرة بالجهة التي تخدم الملف السكاني
8	بيانات مواليد ووفيات
8	بيانات صحية
7	بيانات التعليم
8	بيانات المسوح الدورية
6	بيانات قيود أمنية ومعلومات جنائية والسكان المقيمين داخل الدولة وبيانات سجل مدني يحتوي على الزواج والطلاق
1	بيانات العمل
1	بيانات المنصرف من وسائل تنظيم الأسرة

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

أكدت نتائج تحليل البيانات إيجابيات وجود قواعد بيانات خاصة بالملف السكاني بالجهة المنوطة بالسكان بنسبة 100%. كما أوضحت النتائج أن الدول السبع المستجيبة بينت ايجابيات توافر قواعد البيانات التي تساعد في التخطيط الجيد والمتابعة والتقويم للتدخلات المحلية وتوفير البيانات وسرعة اتخاذ القرارات، (الجدد رقم (3)، وأن نسبة أثر التحول الرقمي على خصائص السكان بلغت حوالي 83%، (الجدول رقم (4) وشكل (1).

جدول رقم (3): التوزيع العددي ايجابيات توافر قواعد بيانات خاصة بالملف السكاني

عدد الدول	ماهي ايجابيات توافر قواعد بيانات خاصة بالملف السكاني
7	التخطيط الجيد
7	المتابعة والتقويم للتدخلات المحلية
7	توافر البيانات وسرعة اتخاذ القرارات
1	التنسيق الجيد
1	الاحتفاظ بنسخ احتياطية باستمرار للبيانات

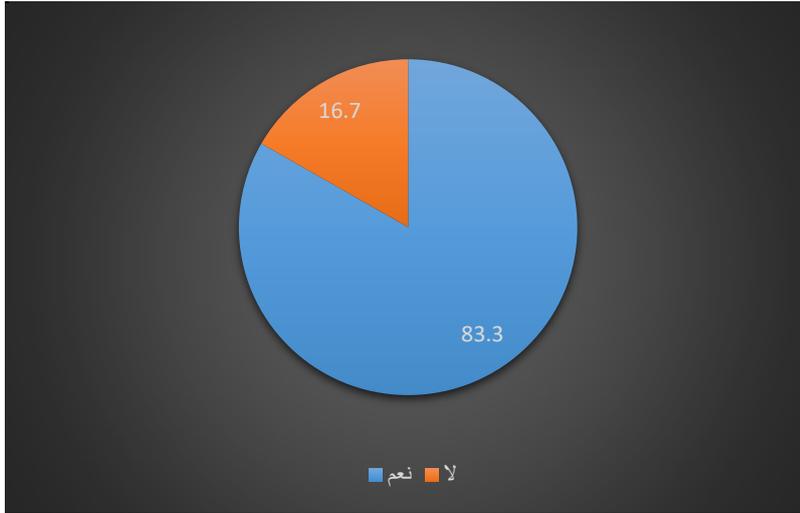
المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

جدول رقم (4): التوزيع العددي والنسبي لوجود أثر التحول الرقمي على خصائص السكان

لا		نعم		تساؤلات البنية التحتية
النسبة	العدد	النسبة	العدد	
16.7	1	83.3	5	هل يوجد أثر للتحول الرقمي على خصائص السكان؟

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

شكل (1) نسبة أثر التحول الرقمي على خصائص السكان



المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

وأظهرت نتائج تحليل البيانات أن يخصص للتحول الرقمي أثر على خصائص السكان، أظهرت النتائج أن أهم الآثار كانت في مجال التعليم في جودة فرص التعليم، والتعليم عن بعد، وتحسين الخصائص التعليمية نفسها بواقع 5 دول أي نصف عينة الدراسة وأقل الخدمات كانت في مجال تحقيق التوازن في نمط الحياة اليومية للمواطنين حيث كانت استجابة لدولة واحدة فقط، كما هو مبين بالجدول رقم(5).

جدول رقم (5): التوزيع العددي لكيفية تأثير التحول الرقمي على خصائص السكان (التعليم - الصحة..... الخ)

عدد الدول	كيف أثر التحول الرقمي على خصائص السكان (التعليم -الصحة الخ)؟
3	تقريب الخدمات من المواطنين وسهولة الحصول عليها وتقليل الوقت والجهد
5	التعليم حيث أثر ايجابيا على جودة وفرص التعليم والتعليم عن بعد وتحسن في الخصائص التعليمية
4	العمل عم طريق زيادة فرص العمل وتسهيل الوصول اليه وزيادة مستوى الانتاج
3	الصحة له أثر ايجابي في الاستقبال وقوائم الانتظار وجودة واستدامة الخدمة
1	تحقيق توازن في نمط الحياة الاجتماعية
2	ساهم في خفض معدلات الوفيات والولادات وزيادة العمر المتوقع للحياة من خلال تقديم الخدمات الصحية

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان.
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

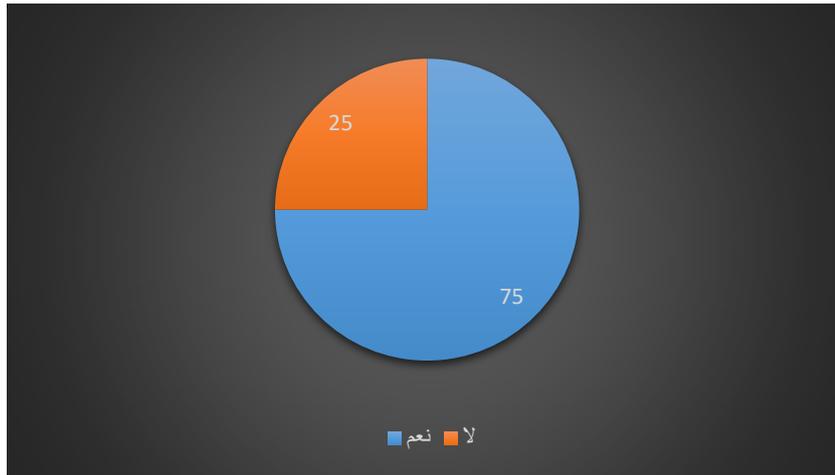
بينت نتائج تحليل بيانات الجدولين (6) و(7) وشكل (2) وجود تأثير للتحول الرقمي على تحسين جودة الحياة بنسبة 75%، وكانت أهم عناصر تأثير التحول الرقمي على تحسين جودة الحياة هي (توفير الوقت والجهد، وتوفير التكلفة من خلال سرعة العمل، وتحسين نمط الحياة اليومية للمواطنين).

جدول رقم (6): التوزيع النسبي لأثر التحول الرقمي على تحسين جودة الحياة

تساؤلات البنية التحتية		نعم	لا
		النسبة	النسبة
هل يوجد أثر للتحول الرقمي على تحسين جودة الحياة؟		75	25

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان.
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

شكل (2) نسبة أثر التحول الرقمي على جودة الحياة



المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان.
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

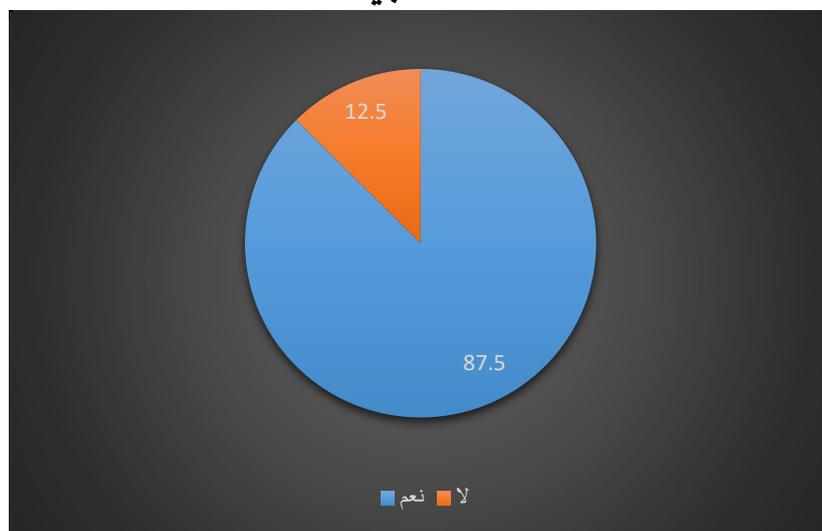
جدول رقم (7): التوزيع العددي لكيفية تأثير التحول الرقمي على تحسين جودة الحياة

عدد الدول	كيف أثر التحول الرقمي على تحسين جودة الحياة
1	تسهيل الاجراءات الحكومية
1	وضوح في التعاملات ورفع مستوى الالتزام
2	تسهيل الخدمات المقدمة للمواطنين
1	التعلم عن بعد
1	العمل عن بعد
1	اتاحة البيانات والمعلومات
4	توفير الوقت والجهد وتوفير التكلفة من خلال سرعة العمل
3	تحسين نمط الحياة اليومية للمواطنين
1	زيادة الوعي الصحي مما ادى الى تقليل الوفيات ورفع معدل توقع الحياة عند الولادة
1	تحقيق خطط التنمية المستدامة بكافة أهدافها

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

كما تشير النتائج أن للتحول الرقمي أثر على العادات والتقاليد بالوطن العربي بنسبة بلغت حوالي 88%، كما هو موضح بالشكل رقم (3)، وأن أهم أشكال هذا التأثير على العادات والتقاليد هي سرعة إنجاز العمل ومتابعة وتقييم الكوادر وتقليل التكلفة على التوالي، الجدول رقم (9).

شكل (3) التوزيع النسبي لوجود أثر التحول الرقمي على العادات والتقاليد بالوطن العربي



المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

جدول رقم (9): التوزيع العددي لكيفية تأثير التحول الرقمي على العادات والتقاليد بالوطن العربي

عدد الدول	كيف أثر التحول الرقمي على العادات والتقاليد بالوطن العربي
5	تقليل التكلفة
7	سرعة انجاز العمل
6	متابعة وتقييم الكوادر
1	توفير الجهد

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

وأوضحت النتائج أنه يوجد تأثير للتحويل الرقمي على تحقيق أهداف التنمية المستدامة، حيث أوضحت النتائج أن نسبة هذا الأثر بلغت 75%، مقابل ربع العينة (25%) ترى أنه لا يوجد أي أثر كما هو مبين بالجدول رقم (10) وكانت أهم الأهداف التنموية التي ساعد التحويل الرقمي على تحقيقها والأكثر تكرارا بين إجابات الدول هي (الهدف الثالث، والهدف الرابع، والهدف الثامن، والسادس عشر) على التوالي بينما بينت النتائج أن أقل الأهداف التي أثر التحويل الرقمي عليها في اختيارات الدول هو الهدف (الثاني عشر والهدف السابع عشر)، كما هو مبين بالجدول رقم (11).

جدول رقم (10): التوزيع العددي والنسبي لوجود أثر للتحويل الرقمي على

تحقيق أهداف التنمية المستدامة

تساؤلات البنية التحتية		لا	نعم
		النسبة	النسبة
هل يوجد أثر للتحويل الرقمي على تحقيق أهداف التنمية المستدامة؟		25	75

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحويل الرقمي وتأثير ذلك على ملف السكان.

*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

جدول رقم (11): التوزيع العددي لأهداف التنمية المستدامة التي ساعد التحول الرقمي على تحقيقها

عدد الدول	الأهداف التنموية التي ساعد التحول الرقمي على تحقيقها
1	كل الأهداف ال 17
3	الهدف الخامس (تحقيق المساواة بين الجنسين وتمكين كل النساء والفتيات)
4	الهدف الثامن العمل اللائق ونمو الاقتصاد " تعزيز النمو الاقتصادي والشامل للجميع والمستدام وتوفير العمل اللائق للجميع".
2	الهدف التاسع الصناعة والابتكار والبنية التحتية " إقامة بنية تحتية قادرة على الصمود وتحفيز التصنيع الشامل للجميع والمستدام وتشجيع الابتكار
3	الهدف (16) السلام والعدالة والمؤسسات القوية " تشجيع إقامة مجتمعات سليمة وشاملة للجميع من أجل تحقيق التنمية المستدامة، وتوفير إمكانية الوصول إلى العدالة للجميع وبناء مؤسسات فعالة وخاضعة للمساءلة وشاملة للجميع على جميع المستويات"
5	الهدف الثالث " الصحة الجيدة والرفاه"
5	الهدف الرابع " التعليم الجيد"
2	الهدف الثالث عشر المتعلق بالمناخ " من خلال استخدام مصادر الطاقة البديلة لإدامة البنية التحتية "
2	الهدف الأول " القضاء على الفقر"
2	الهدف العاشر " الحد من عدم المساواة"
1	الهدف الثاني عشر " الاستهلاك والإنتاج المسؤولان"
1	الهدف السابع عشر: عقد الشراكات لتحقيق الأهداف

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
 *ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

وأظهرت نتائج تحليل البيانات كما هو موضح بالجدول رقم (12) أن حوالي 63% من الدول ترى أنها قد حققت أقصى استفادة من التحول الرقمي في الملف السكاني، وكانت هذه الاستفادة في مجالات التخطيط الجيد لا مركزيا، والمتابعة والتقييم، والاستعداد للتحول الديموغرافي على التوالي، (الجدول رقم (13)).

جدول رقم (12): التوزيع النسبي لتحقيق أقصى استفادة من التحول الرقمي

في الملف السكاني

لا	نعم	
النسبة	النسبة	
37.5	62.5	هل تم تحقيق أقصى استفادة من التحول الرقمي في الملف السكاني؟

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
 *ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

جدول رقم (13): التوزيع العددي للاستفادة التي تحققت من التحول الرقمي في

الملف السكاني

الدول	الاستفادة التي تحققت من التحول الرقمي في الملف السكاني
5	التخطيط الجيد لا مركزيا
5	المتابعة والتقييم
4	الاستعداد للتحول الديموغرافي
1	ضبط الحالة المدنية

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان
 *ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

وتشير نتائج تحليل البيانات إلى عدم توفر الكوادر المدربة والمؤهلة والتي تعتبر السبب الرئيسي والمعوق الأول الذي يعيق الاستفادة من التحول الرقمي وتتساوي بعد ذلك باقي الأسباب حسب استجابات الدول (الجدول رقم (14)).
جدول رقم (14): التوزيع العددي للمعوقات التي تحول دون الاستفادة من التحول الرقمي في الملف السكاني

الدول	تساؤلات البنية التحتية
1	ضعف البنية التحتية
2	عدم وجود كوادر مدربة
1	عدم وجود تمويل
1	استراتيجية التحول الرقمي مستحدثة ولمم تطبق بعد
1	عدم جاهزية بعض الشركاء للخدمة

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان.
*ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

أوضحت نتائج تحليل البيانات وجود تباينات بين الدول العربية وبعضها في مجال تقنيات التحول الرقمي في إدارة الملف السكاني بشكل خاص حيث بينت النتائج أن نسبة 80% من الدول لديها قواعد بيانات سكانية محدثة ومحمية مثل الأردن ومصر والعراق هناك دولا عربية ليست لديها عناصر البنية التحتية للتحول الرقمي اللازمة لإدارة الملف السكاني مثل لبنان كما يوضح الجدول رقم (15) نسب الدول التي يتوافر لديها عناصر البنية اللازمة للتحول الرقمي، بينما تراوحت وتركزت نسب توافر أغلب العناصر ما بين (50% - 80%) بأغلب الدول.

جدول رقم (15): التوزيع النسبي لنسب توافر عناصر البنية التحتية للتحويل الرقمي اللازمة لإدارة الملف السكاني

العنصر	لا يوجد	%30	%50	%80	%100
قواعد بيانات محدثة ومدققة	14.3	14.3	28.6	42.9	-
قواعد البيانات محمية	16.7	16.7	16.7	33.3	16.7
وجود مستودعات للبيانات data warehouse	14.3	14.3	14.3	57.1	-
ربط قواعد البيانات الخاصة بالجهات الأخرى (التعليم - الصحة ...)	28.6	42.9	28.6	-	-
سرعة انترنت مناسبة للتبادل اللحظي للبيانات بين الجهات العاملة في إدارة الملف السكاني	14.3	-	28.6	57.1	-
وجود تدريب كاف للعاملين في إدارة ملف السكان في مجال الإحصاء وحساب المؤشرات.	-	-	14.3	71.4	14.3
وجود كوادرات مدربة ومؤهلة في مجال الأمن السيبراني والشبكات	14.3	28.6	14.3	28.6	14.3
وجود كوادرات مدربة لتحليل البيانات وعمل الاستقطاعات اللازمة وعمل الرسوم البيانية واستخدام الخرائط التفاعلية visualization	-	14.3	42.9	42.9	-
وجود وإتاحة البرامج اللازمة SQL,SPSS,POWER BL SPECTRUM, TABLEAU.AND OTHER	-	28.6	-	42.9	28.6
التدريب المستمر على هذه البرامج	14.3	14.3	28.6	42.9	-
الإتاحة والتدريب على الذكاء الاصطناعي R,A1,PYTHON	42.9	28.6	28.6	-	-
وضع شاشات لعرض المؤشرات السكانية لمساعدة متخذي القرار	28.6	42.9	-	14.3	14.3
وجود دعم سياسي للملف السكاني	-	-	42.9	28.9	28.9

-	57.1	28.6	-	14.3	وجود مرصد سكاني على مستوى الدولة لعرض المؤشرات الخاصة (التوزيع الديموغرافي - الصحة - العمل والبطالة - التعليم... الخ) لأدنى المستويات الإدارية بالدولة (القرى).
-	57.1	14.3	14.3	14.3	مدي الاستفادة من وجود المرصد السكاني لتحديد التدخلات وعمل الخطط المحلية
-	42.9	28.6	14.3	14.3	مدي الاستفادة من المنظومة للقيام بالمتابعة والتقييم لبرامج السكان على المستوى القطري والمحلي
14.3	14.3	42.9	14.3	14.3	وجود الدعم المادي لاستمرار واستدامة النظام
-	16.7	33.3	33.3	16.7	وجود ميزانية سنوية كافية لصيانة النظام
-	14.3	42.9	28.6	14.3	وجود ميزانية مناسبة للتدريب
14.3	14.3	42.9	28.6	-	وجود فرص لتبادل الخبرات مع الجهات الأخرى (داخلياً وخارجياً)

المصدر: محسوب من بيانات استمارة رصد وقياس تجارب الدول العربية في التحول الرقمي وتأثير ذلك على ملف السكان.
 *ملاحظة أن أعداد المستجيبين بكل مجال مختلف حيث كان هناك بيانات غير مكتملة لبعض الدول.

10- التوصيات

وتوصي الدراسة بإجراء مراجعة إقليمية لمفهوم قواعد البيانات السكانية على مستوى الدول العربية التي تساعد في إدارة الملف السكاني والتوعية بها وتوحيد مؤشراتها على المستوى الإقليمي والمحلي لمتابعة الوضع السكاني وتعزيز التعاون العربي المشترك في مجال تبادل الخبرات بين الدول العربية الرائدة في مجال الرقمنة.

المراجع

1. خالد الوالي (2019) الاقتصاد الرقمي في الوطن العربي: الواقع والمستقبل " ، مركز الدراسات الاستراتيجية في جامعة الدول العربية
2. منظمة الأمم المتحدة للتنمية الصناعية 2020 ، "التحول الرقمي في الدول العربية: التحديات والفرص [UNIDO - Report](<https://www.unido.org>)
3. سفيان قعلول، الوليد طلحة (2020)، الاقتصاد الرقمي في الدول العربية: الواقع والتحديات - دراسة اقتصادية- صندوق النقد العربي- الدائرة الاقتصادية.
4. ليلى أحمد (2020) "المرأة والتحول الرقمي في الوطن العربي" ، مركز المرأة العربية للتدريب والبحوث(CAWTAR) .
5. فاطمة الزهراء (2021)" التعليم الرقمي في الوطن العربي: التحديات والفرص" ، مجلة التربية والعلم (مجلة أكاديمية).
6. عمر الفاروق (2021) "تحديات التحول الرقمي في الدول النامية: حالة الدول العربية" ، مجلة تكنولوجيا المعلومات (Information Technology Journa
7. أمل بنت عبد الله بن ابراهيم ال ابراهيم (2022)، تجربة التحول الى التعليم عن بعد في التعليم العام إثر جائحة كورونا (Covid19) من وجهة نظر الطلاب، المجلد الثامن والثلاثون العدد الثاني عشر كلية التربية جامعة الملك سعود، ادارة البحوث والنشر العلمي (المجلة العلمية).
8. محمد السيد (2022) "التحول الرقمي في القطاع الصحي: دراسة حالة لدول الخليج العربي" ، مجلة الصحة الرقمية(Journal of Digital Health) .
9. (نيفين ناجي 2024) "فجوات تطبيق التحول الرقمي في المجالات الخدمية المقدمة للسكان: تجارب لبعض الدول العربية" ، المجلة المصرية للسكان وتنظيم الأسرة ، المجلد 56، العدد(2).2024

Detecting and estimation outliers in multiple linear regression models

Professor Dr. Shalal H. Al-Jobouri

Shalalaljobouri@yahoo.com

Assistant Prof.Dr. Nabeel G. Nancy

ngnacy_54@yahoo.com

Abstract:

In this paper, new methods were proposed to detect and estimate outliers for multiple regression models. The detection method was based on using the oblique ellipse equation whose center is (\bar{X}_j, \bar{D}_j) for each explanatory variable X_j based on the DFBETAS (D_{ij}) statistic proposed by Belsley, Kuh & Welsch. Seven methods were also proposed to find estimates for these observations, some of which were based on finding the point of intersection of the line connecting each of the outlier and the center of the ellipse; with the ellipse equation, and others were based on adding or subtracting each of the variance or the common variance multiplied by the value of the left side of the ellipse. As a practical application, the data of Chatterjee & Hadi (1988) for the Health Club were used. These methods gave good results, some of which gave better results than the deletion method, but all of them increased the value of the coefficient of determination and reduced the mean square error.

Keywords: outliers, detecting, estimating, oblique ellipse, DFBETAS statistic, Hat matrix, Catcher matrix.

اكتشاف وتقدير المشاهدات الشاذة لنماذج الانحدار الخطي المتعدد

أ.د. شلال حبيب الجبوري
أ.م. د. نبيل جورج ناسي

الملخص:

في هذا البحث أقترح طرق جديدة لاكتشاف وتقدير المشاهدات الشاذة لنماذج الانحدار المتعدد، إذ أعتمدت طريقة الأكتشاف على استخدام معادلة القطع الناقص المائل Ellipse Oblique الذي مركزه (\bar{X}_j, \bar{D}_j) لكل متغير توضيحي X_j بالأستناد الى أحصاءة (D_{ij}) المقترحة من قبل Belsley, Kuh & Welsch. كما اقترحت سبعة طرق لإيجاد تقديرات لهذه المشاهدات بعضها اعتمد على ايجاد نقطة تقاطع المستقيم الواصل بين كل من النقطة الشاذة ومركز القطع مع معادلة القطع الناقص وبعضها الآخر اعتمد على اضافة أو طرح كل من التباين أو التباين المشترك مضروباً بقيمة الطرف الأيسر للقطع الناقص. كتطبيق عملي استخدمت بيانات (Chatterjee & Hadi (1988) والخاصة بنادي الصحة Health Club ، إذ اعطت هذه الطرق نتائج جيدة بعضها أعطى نتائج أفضل من اسلوب الحذف ولكن جميعها رفعت قيمة معامل التحديد وخفضت متوسط مربعات الخطأ.

1. Introduction: Outliers are observations that appear inconsistent with the remainder of the data set. Outliers may be mistakes, or else accurate but unexpected observations which could shed new light on the phenomenon under study. In regression analysis, an outlier is an observation for which the residual is large in magnitude compared to other observations in the data set. The detection of outliers and influential points is an important step of the regression analysis. Outlier detection methods have been used to detect and remove anomalous values from data ^[9].

A common concern in empirical research is whether findings might be invalidated by a set of outlying observations. Such outliers can bias estimates and distort inference in regression models. Consequently, it is important to check for possible outlier contamination ^[10].

If the study of outliers was able to secure credit cards, by identifying suspicious transaction behavior from normal ones and could prevent anomalous network intrusions, just imagine how it could improve the medical field. Medical data analytic

will provide better and more effective results and even diseases could be detected when effective outlier detection approaches are developed. These approaches must be robust in the presence of outliers and must provide perfect decision-making towards outliers^[3].

The first beginnings of the topic of outliers dealt with such a type of observations by deleting them and analyzing the rest of the observations. As is known in statistics, any deficiency in data will lead to a deficiency in the studied information and thus affect the results of the adopted analysis. Therefore, the trend towards estimating outliers instead of deleting them aims to preserve the greatest amount of information in the sample selected for analysis.

2. Objective of the research: The main objective of the research is based on proposing a new method for detecting outliers and then estimating these observations after discovering them in the case of multiple linear regression, which pertains to the explanatory variables.

The purpose of these proposed methods of estimating is when we cannot sacrifice the size of the sample drawn from the population, especially if the sample size is very small such as medical experiments. So, we estimate these outliers and enter them into the data again.

3. Historical background: The historical roots of detecting outliers for multiple regression models go back to the second half of the last century. In 1972, Behnken & Draper^[4] studied the *Standardized Residuals* statistic $SR_i = \frac{e_i}{S_e}$ to detecting single outliers. In the following year, Ellenberg^[7] proposed the *Studentized Residuals* statistic $STR_i = \frac{SR_i}{\sqrt{1-h_{ii}}}$ where $h_{ii} = x_i(X'X)^{-1}x_i'$ and x_i represents the i^{th} row in the matrix of explanatory variables X.

In 1980, Belsley, Kuh & Welsch proposed the test statistics DFFITS, DFSTAT & COVRATIO DFBETAS, to detect outliers^[5].

In 1993, Hadi & Simonoff ^[8] proposed a method to diagnose multiple outliers in a multiple linear regression models by dividing the data into two subsets, the first of which includes M observations free of outliers and the second of which includes the doubtful observations, the extreme cases of which are tested. In 2002, Al-Jobouri & Nacy ^[1] proposed a method to detect and estimate outliers for simple linear regression models, as they relied on the equation of the oblique ellipse in detection, and they proposed eight methods to estimate them, some of which depended on finding the point of intersection of the line connecting each of the outlier point and the center of the ellipse with the equation of the ellipse, and others depended on adding or subtracting each of the variance or the covariance multiplied by the value of the left side of the ellipse. In 2017, Stephen & Senthamarai investigate the presence of outliers based on existing procedures of residuals and standardized residuals. they have used the new approach of standardized scores for detecting outliers without the use of predicted values. The performance of the new approach was verified with the real-life data.

4. Proposed methods: The researchers relied on expanding the methods proposed in their previous research for detection and estimation in the case of the simple linear model ^[1] to include multiple linear regression models, as these proposed methods work to detect and estimate outliers for the explanatory variables X_{ij} only and do not detect and estimate outliers for the response variable Y_i . The proposed method in this field is as follows:

I- Using the DFBETAS (D_{ij}) to extracting D_{ij} , considering that i and j represent the sequence of each of the observation and the explanatory variable, respectively, as follows:

$$D_{ij} = \frac{C_{ji}}{\sqrt{\sum_{i=1}^n C_{ij}^2}} * \frac{e_i}{S_{(i)}(1 - h_{ii})} < \frac{2}{\sqrt{n}}$$

where $C = (X'X)^{-1}X'$ represents the Catcher matrix.

$h_{ii} = x_i(X'X)^{-1}x_i'$ represents the diagonal of the Hat Matrix.

x_i represents the i^{th} row in explanatory matrix.

$e_i = Y_i - \hat{Y}_i$ represents the random error of observation i .

$S_{(i)}$ represents the standard deviation of the error by exclude observation i .

II- In the same manner followed in the previous research ^[1], the outliers can be diagnosed using the oblique ellipse equation whose center is (\bar{X}_j, \bar{D}_j) and for each explanatory variable separately, meaning that the number of ellipse equations will be equal to the number of explanatory variables.

$$Z_{ij} = \frac{[(X_{ij} - \bar{X}_j)\cos\theta_j + (D_{ij} - \bar{D}_j)\sin\theta_j]^2}{a_j^2} + \frac{[(D_{ij} - \bar{D}_j)\cos\theta_j - (X_{ij} - \bar{X}_j)\sin\theta_j]^2}{b_j^2}$$

where

$$\theta_j = \tan^{-1} \beta'_j, \quad b_j = \frac{\gamma_2 S_{D_j}}{\cos\theta_j}, \quad a_j = \frac{\gamma_1 S_{X_j}}{\cos\theta_j}$$

γ_1 and γ_2 are positive default values, the first determines the length of the major axis of the ellipse and the second determines the length of the minor axis.

S_{X_j} and S_{D_j} represent the standard deviation of each of the variables X_j and D_j respectively.

After substituting the points (X_{ij}, D_{ij}) in the equation j above, any point that gives $Z_{ij} > 1$ is diagnosed as an outlier.

III- Follow one of the suggested methods for estimation mentioned in the previous source to estimate the X_{ij} observations discovered as outliers, and after estimating all the outliers for all the explanatory variables, these estimates in the model replace their previous values, and then the test is repeated again (step I). If outliers appear, we move to step II, otherwise we stop.

The proposed methods for estimation are:

- The first method M1 - The estimation is represented by the point of intersection of the ellipse of the explanatory variable X_j and the response variable D_j , i.e. when $Z_{ij} = 1$; with the

equation of the line connecting the outlier point (X_{0j}, D_{0j}) and the center of the ellipse (\bar{X}_j, \bar{D}_j) , i.e.:

$$D_j = \frac{D_{0j}\bar{X}_j - X_{0j}\bar{D}_j}{\bar{X}_j - X_{0j}} + \frac{\bar{D}_j - D_{0j}}{\bar{X}_j - X_{0j}} X_j$$

- The second method M2 - The same previous method after deleting the outlier observation from the data set for each explanatory variable.

- The third method M3 - It is estimated using one of the following two equations:

$$\hat{X}_{ij} = X_{ij} \pm \sqrt{|Z_{ij} * cov(X_j, D_j)|}$$

The positive sign of the above relationship is taken if the value of the observation X_{ij} is less than its arithmetic mean \bar{X}_j , otherwise the negative sign is used.

- The fourth method M4 - Estimate using one of the following two equations:

$$\hat{X}_{ij} = X_{ij} \pm \sqrt{Z_{ij} * S_{X_j}^2}$$

- The fifth method M5 - Estimate using one of the following two equations:

$$\hat{X}_{ij} = X_{ij} \pm \sqrt{|\gamma_1 * Z_{ij} * cov(X_j, D_j)|}$$

- The sixth method M6 - Estimate using one of the following two equations:

$$\hat{X}_{ij} = X_{ij} \pm \sqrt{\gamma_1 * Z_{ij} * S_{X_j}^2}$$

- The seventh method M7 - This method depends on estimating the outlier observation by finding the point of intersection of the line connecting the outlier point (X_{0j}, D_{0j}) and the data center (\bar{X}_j, \bar{D}_j) for the explanatory variable X_j and the line $D_{ij} = \frac{2}{\sqrt{n}}$, so the observation estimate X_{ij} will be:

$$\hat{X}_{ij} = \bar{X}_j + \frac{\left(\frac{2}{\sqrt{n}} - \bar{D}_j\right) (\bar{X}_j - X_{0j})}{\bar{D}_j - D_{0j}}$$

$$\text{where } \bar{D}_j = \frac{\sum_{i=1}^n D_{ij}}{n}, \quad \text{and } \bar{X}_j = \frac{\sum_{i=1}^n X_{ij}}{n}$$

5. practical aspect: Data from Chatterjee & Hadi ^[6] for the year 1988 for Health Club were used (Table 1):

Table 1 represents the data from Health Club

i	Y	X ₁	X ₂	X ₃	X ₄
1	481	217	67	260	91
2	292	141	52	190	66
3	338	152	58	203	68
4	257	153	56	183	76
5	396	180	77	170	77
6	429	193	71	178	82
7	345	162	65	160	74
8	469	180	80	170	84
9	425	205	77	188	83
10	328	168	74	170	79
11	393	232	65	220	72
12	346	146	68	158	68
13	279	173	51	243	56
14	311	155	64	198	59
15	401	212	66	220	77
16	267	138	70	180	62
17	404	147	54	150	75
18	442	197	76	228	88
19	368	165	59	188	70
20	295	125	58	160	66
21	391	161	52	190	69
22	264	132	62	163	59
23	487	257	64	313	96
24	481	236	72	225	84
25	374	149	57	173	68
26	309	161	57	173	65
27	367	198	59	220	62
28	469	245	70	218	69
29	252	141	63	193	60
30	338	177	53	183	75

Where:

Y represents the time (seconds) taken to run one mile.

X_1 represents the weight (pounds).

X_2 represents the heart rate per minute.

X_3 represents arm and leg strength (the number of pounds a person can lift).

X_4 represents the time (seconds) taken to run a $\frac{1}{4}$ mile test.

By applying the DFBETAS statistic equation, D_{ij} can be found and by using the ellipse equation by taking the original values of X_{ij} with the extracted values of D_{ij} , we will have four ellipse equations, one for each explanatory variable X_j . By applying the same method to detect outliers, the most likely is observation no. 28 in the values of the variables X_1 and X_4 ; and observation 30 in the value of the variable X_2 , (i.e. $x_{28,1} = 245$, $x_{28,4} = 69$ and $x_{30,2} = 53$; which is very similar to other methods of detection ^[5].

As for estimating these outliers, they are done by one of the suggested methods and using the computer programs found in the source [2]. The final results of the estimates of these methods were as follows:

1- The estimates of the outliers for these methods were as follows:

Table 2 Estimates of anomalous observations

Observation no.	Variable	original value	Estimated value by							
			M1	M2	M3	M4	M5	M6	M7	
8	X_1	180	--						63.1	
11	X_1	232		192					--	
23	X_1	257							215.7	
	X_2	64							63.8	
	X_3	313							225.9	
	X_4	96							91.2	
28	X_1	245	240	191.2	237.3				166.8	192
	X_2	70	--	66.8					--	--
	X_4	69	71.5	71.5	71.9	80.2	72.1	92.7	71.2	
30	X_2	53	53.5	--	53.5	61.1	54.6	70.3	60.1	

2- The estimates of the regression parameters β_j before and after estimation by deleting outliers with a *t-test* for each parameter and an *F-test* for the model were as in table 3; By comparing the calculated *t* values for the model parameters (numbers enclosed in parentheses in Table 3) with the tabular value of the degree of freedom 25, we note that:

- The value of $\hat{\beta}_0$ is not significant for all methods at the level of $\alpha = 0.25$.
- The value of $\hat{\beta}_1$ is significant for all methods at the level of $\alpha = 0.01$.

Table 3 Estimates of regression parameters and t and F-tests

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	F
Before estimation	-3.816 (-0.067)	1.272 (4.354)	-0.526 (-0.599)	-0.506 (-2.019)	3.900 (5.123)	28.090
M1	-6.193 (-0.109)	1.226 (4.073)	-0.572 (-0.652)	-0.489 (-1.952)	4.039 (5.168)	28.636
M2	-4.022 (-0.059)	1.277 (2.720)	-0.635 (-0.605)	0.431 (-1.367)	3.835 (3.647)	20.711
M3	-7.033 (-0.123)	1.222 (3.984)	-0.561 (-0.680)	-0.486 (-1.922)	4.042 (5.112)	28.362
M4	6.238 (0.144)	1.086 (3.665)	-1.143 (-1.318)	-0.464 (-1.979)	4.697 (6.022)	34.142
M5	-3.645 (-0.065)	1.217 (4.163)	-0.671 (-0.769)	-0.492 (-2.006)	4.138 (5.383)	29.792
M6	7.594 (0.159)	0.923 (3.864)	-2.433 (-3.271)	-0.263 (-1.241)	5.590 (9.798)	49.681
M7	-24.742 (-0.364)	1.017 (2.498)	-0.322 (-0.304)	-0.320 (-1.049)	4.152 (4.203)	19.679
Deleteobserv ation no. 28	-15.234 (-0.277)	0.945 (22.835)	0.644 (-0.764)	-0.380 (-1.525)	4.589 (5.575)	28.790
Deleteobserv ationno. 30	31.593 (0.576)	1.398 (5.066)	-1.319 (-1.497)	-0.643 (-2.697)	4.225 (5.901)	33.790
Deleteobserv ations no. 28, 30	19.902 (0.381)	1.067 (3.463)	-1.422 (-1.703)	-0.519 (-2.211)	4.894 (6.429)	35.502

- The value of $\hat{\beta}_2$ is not significant for most methods at the level of $\alpha = 0.01$. The remaining methods M4 and the method of deleting observation no. 28 and the method of deleting observations no. 28 and 30 together are significant at this level. As for the method M6, the value is significant at the level of $\alpha = 0.005$.
- The value of $\hat{\beta}_3$ is significant for all methods at the level of $\alpha = 0.05$ except for the methods M6 and M7, where the value is significant at the level of $\alpha = 0.25$, and the method M2 and the method of deleting observation no. 28, where the value is significant at the level of $\alpha = 0.10$.
- The value of $\hat{\beta}_4$ is significant for all methods at the level of $\alpha = 0.0005$.
- As for the result of the F test and the two degrees of freedom (4,25), the model parameters are generally significant at the level of $\alpha = 0.005$ for all methods.

3- The result of estimation or deletion of observations was the values of the coefficient of determination for the model R^2 , the mean square error MSE, and the variance of the estimated value of the response variable $V(\hat{Y})$ as in Table 4:

From Table 4, we note that all methods increased the R^2 value and decreased the MSE value except for the M2 and M7 methods. The highest R^2 and the lowest MSE were achieved by the M6 method, thus it is more efficient than the other methods.

Table 4 Values of the coefficient of determination, the mean square error, and the variance of the estimated value of the response variable

Method	(γ_1, γ_2)	Repetition	R^2	MSE	$V(\hat{Y})$	Relative Efficiency
Before estimation	--	--	0.85406	851.935	4126.019	1.000
M1	(4.5, 5)	117	0.85644	838.036	4137.521	1.017
M2	(4.5, 5.5)	2	0.81185	1098.348	3922.090	0.776
M3	(4.5, 5.5)	10	0.85526	844.954	4131.795	1.008
M4	(4.5, 6.5)	2	0.87674	719.534	4235.592	1.184
M5	(4.5, 8)	5	0.86124	810.008	4160.716	1.052
M6	(4.5, 4.5)	20	0.91190	514.315	4405.428	1.656
M7	--	6	0.80391	1144.664	3883.825	0.744
Delete observation no. 28	--	--	0.80391	1144.664	3883.825	0.744
Delete observation no. 30	--	--	0.88018	724.004	4368.573	1.177
Delete observations no. 28,30	--	--	0.88973	647.903	4259.550	1.315

6. Conclusions:

This paper sought to propose a method to detect outliers for multiple linear regression models and then propose methods to estimate these observations instead of deleting them. In its practical aspect, a comparison was made between these proposed methods and with existing methods using the deletion method. The most important conclusions in this regard are:

- All the proposed methods increased the value of R^2 and decreased the value of MSE except for the M2 and M7 methods, which decreased the value of R^2 and increased the value of MSE.
- The highest R^2 and the lowest MSE were given by the M6 method, followed by the M4 method.
- Deleting outliers may give the highest R^2 and the lowest MSE compared to the estimation methods, but the danger lies in excluding many observations, while the estimation methods estimated only a value for one variable from one observation,

which led to increasing the value of R^2 and decreasing the value of MSE significantly and better than the deletion method.

- The estimation methods did not affect the strength of the model parameters as a whole (F -test) and at a significance level of $\alpha=0.005$. As for individually, the effect of the estimation methods was slight on the model parameters (t -test).

7. References:

- 1- الجبوري ، شلال & ناسي ، نبيل (2002) "اكتشاف المشاهدات الشاذة باستخدام معادلة القطع الناقص Ellipse في حالة الأنداد الخطي البسيط: المجلة العراقية للعلوم الإحصائية، العدد-4.
- 2- ناسي، نبيل جورج (2001) " تقييم كفاءة طرق تقدير المشاهدات الشاذة لنماذج الأنداد " رسالة دكتوراه احصاء ، كلية الإدارة والأقتصاد/ جامعة بغداد، غير منشور.
- 3- Abir Smiti (2020) "A critical overview of outlier detection methods" Elsevier; Computer Science Review; Volume 38, November 2020, 100306
- 4- Behnken, D. W. & Draper, N.R. (1972) "Residuals and their Variance Patterns" Technometrics, vol. 14 , no. 1, p.p. 101-111.
- 5- Belsley, D. A., Kuh, E. & Welsch, R.E. (1980) "Regression Diagnostics Identifying Influential data and sources of collinearity" John Wiley, New York.
- 6- Chatterjee, S. & Hadi, A. S. (1988) "Sensitivity analysis in linear regression" John Wiley, New York.
- 7- Ellenberg, J. H. (1973) "The joint distribution of the standardized least squares residuals from a general linear regression" JASA, vol. 68; p.p. 941-943.
8. Hadi, A. S. & Simonoff, J. S. (1993) "Procedures for the identification of multiple outliers in linear models" JASA, vol. 88, no. 424; p.p. 1264-1271.
9. Stephen, R.S. & Senthamarai, K. K. (2017) "Detection of Outliers in Regression Model for Medical Data" International Journal of Medical Research & Health Sciences, 6(7): 50-56; ISSN No: 2319-5886
10. Xiyu Jiao, Felix Pretis (2022) "Testing the Presence of Outliers in Regression Models" Oxford Bulletin of Economics and Statistics; Volume 84, Issue 6, p. 1452-1484

Using Panel Data to Build an Efficient Statistical Model for Predicting Student Graduation Estimates

Prof.Dr. Obaid Mahmood Mohsin Alzawbaee

obaid.mohsin808@gmail.com

Al-Siraj Private University-Fallujah-IRAQ

Mr. Mustafa Othman Alsaigh

Mustafa_alsaigh@hotmail.com

Al Mosul University

Abstract:

Background: Accurate forecasting of student graduation rates is critical for educational planning and resource allocation. Previous research has stressed the value of using panel data since it may include cross-sectional and time-series changes, resulting in a more robust analysis than single-dimensional data sets.

Objective: This research aims to construct an efficient statistical model that uses panel data to forecast students' graduation estimations and compare several statistical models to determine which one is most accurate.

Methods: We used six-panel data models: Pooled Ordinary Least Squares, one- and two-way fixed effect models, one- and two-way random effect models, and mixed models. Data was gathered from graduates from four departments over five years, with 10 students from each department every year, for 200 observations.

The results showed that the models examined had varying degrees of accuracy. The two-way fixed effect model was the most successful, with the lowest root mean square error (RMSE) and the greatest R-squared value. This model's superiority stems from its ability to manage unobserved heterogeneity across time and individual effects.

Conclusion: The research finds that the two-way fixed effect model gives the most trustworthy estimates of student graduation rates and recommends its use in educational

forecasting and planning. The model's capacity to effectively capture complicated fluctuations across both dimensions makes it an important tool in educational statistics.

Keywords: Panel Data (PD), Statistical Models, Fixed Effect Model (FEM), Random Effect Model, Time-Series Analysis..

المخلص :

هدفت هذه الورقة الى بناء نموذج للتنبؤ بتقديرات تخرج الطلاب بناء على معدل التخرج ، او متوسط السنوات الثلاث الاولى (متوسط مرجح)، ولتحقيق ذلك تم اعتماد تقنية البيانات اللوحية (البائل داتا). وتضمن البحث تطبيق ستة نماذج للبيانات اللوحية وهي نماذج (POLS) ، ونماذج التأثير الثابت احادي وثنائي الاتجاه ، ونماذج التأثير العشوائي احادي وثنائي الاتجاه ، وكذلك النماذج المختلطة . وتم اعتماد البيانات لخريجي اربعة اقسام على مدى خمس سنوات ولعشرة طلاب لكل قسم ولكل سنة ، اي ان العدد الاجمالي للمفردات هو (200). كانت دقة النماذج المعتمدة متفاوتة ، من خلال بناء ستة انواع من نماذج بيانات البائل ، وان افضل نموذج هو (نموذج التأثير الثابت ثنائي الاجاه) لانه حقق اقل قيمة (RMSE) وكبر قيمة ل (R^2) .

الكلمات المفتاحية: البيانات اللوحية ، السلاسل الزمنية ، النماذج ثابتة التأثير ، النماذج متغيرة التأثير.

1. Introduction:

In the changing world of educational research, properly anticipating student outcomes such as graduation rates is critical. Traditional techniques frequently use cross-sectional or time-series data; however, these approaches may need to reflect the intricacies inherent in educational contexts adequately. Panel data analysis emerges as a superior tool, providing a more detailed view by accounting for temporal dynamics and individual characteristics. This research aims to use the resilience of panel data to build an effective statistical model for forecasting student graduation rates. With its dual focus on time series and cross-sectional dimensions, panel data yields a rich dataset that captures both temporal progression and individual heterogeneity across pupils. This type of data has greater variability, less collinearity across variables, more degrees of freedom, and higher estimate efficiency than pure time-series or cross-sectional data. These characteristics are especially useful

in educational settings because student-specific and time-specific factors are important in determining educational results. Given the complexities of educational processes, using panel data in educational research is more than a methodological preference. Many factors impact educational results, including institutional policies, individual student actions, and external environmental conditions, which fluctuate over time. Panel data enables researchers to adjust for these characteristics, perhaps resulting in more accurate and trustworthy predictive models. Furthermore, predictive modeling of student graduation rates is of great interest to policymakers, and stakeholders involved in higher education planning. Accurate projections can guide resource allocation, policy creation, and strategic planning, resulting in better educational results. In light of this, the major goal of this research is to use panel data to create and analyze multiple statistical models to see which best predicts student graduation estimates [9],[12],[14]}

This research is especially pertinent and important given the rising availability of educational data and the growing demand for educational institutions to improve graduation rates and other academic results. With the transition to data-driven decision-making in education, the findings of this research hope to provide useful insights into the successful use of statistical models in educational forecasting. To achieve these goals, this research takes a thorough approach, examining six distinct panel data models: pooled ordinary least squares (POLS), one- and two-way fixed effect models (FEMs), one- and two-way random effect models (REMs), and mixed models. Each model has various benefits and drawbacks in dealing with unobserved variability, which is crucial for understanding and forecasting student performance and outcomes. This research's data was methodically gathered from the records of graduates from four departments at Cihan University - Sulaymaniyah over five years. Each department provided data on 10 students yearly, yielding a large dataset of 200 unique observations. The sample size and variety guarantee that the findings are statistically significant and potentially applicable to other similar

educational environments. Significant research has demonstrated the value of panel data in various disciplines, but its use in predicting educational results still needs to be explored. Previous research has mostly focused on economic and behavioral sciences, where the efficacy of panel data has been well-proven. By applying this technique to the educational sector, our research fills a significant vacuum in literature. It provides evidence on the applicability and efficacy of panel data models in projecting student graduation rates, an area ripe for investigation given the growing issues in higher education worldwide. {[2],[5],[7],[8]}. This introduction lays the groundwork for thorough research using panel data to forecast student graduation rates. This research intends to contribute to the larger area of educational planning and analysis by evaluating several statistical models to discover the most successful model. The following parts will further detail the methodology, data analysis, and the findings' significance for educational research and practice.

1.1. Research Objective:

The key objective of this research is to construct a strong statistical model that uses panel data to predict student graduation rates with high accuracy. Recognizing the complexities of educational outcomes, which are impacted by various individual and institutional variables that change over time, this research aims to capitalize on the deep insights afforded by panel data. Panel data analysis is particularly well-suited to educational contexts because it combines time-series and cross-sectional data, allowing for a more extensive examination than standard statistical approaches. In educational research, reliable forecasting of student graduation rates is critical for successful academic planning and resource allocation. It guides strategic choices such as curriculum changes, student support programs, and administrative policies. By forecasting graduation results, institutions may better match their resources to student needs, potentially boosting overall educational effectiveness and efficiency. The research used six-

panel data models to achieve this goal: Pooled ordinary least squares, one- and two-way fixed effect models, one- and two-way random effect models, and mixed models. Each model is evaluated for its ability to reliably anticipate graduation rates using a dataset that includes student records from four academic departments over five years for 200 observations. This technique assesses each model's predictive capacity and investigates their comparative efficacy in dealing with unobserved heterogeneity—variations that are not explicitly quantified yet impact student outcomes. This variation might include latent human factors like motivation or unmeasurable institutional variables like instructional quality. Finally, the research aims to select the most efficient model using metrics such as root mean square error (RMSE) and coefficient of determination (R-squared). The objective is to develop a model that makes exact predictions and contains a better knowledge of the underlying elements that influence student achievement. This research is intended to contribute substantially to educational statistics and help academic institutions implement more data-driven methods in educational and administrative planning.

1.2. Problem Statement:

Accurate projection of student graduation rates is a hard task for educational institutions worldwide. Despite the growing availability of data, many colleges fail to properly use this information to forecast outcomes and execute proactive interventions. Traditional statistical approaches frequently depend on time series or cross-sectional data, which may fail to represent the complexities of student growth and institutional dynamics. This shortcoming reveals a substantial gap in existing models' predictive capacities, as they need to account for the inherent variability and temporal fluctuations that distinguish educational contexts. Panel data analysis provides a potential answer by encompassing both aspects, although its use in the educational sector is neglected. The issue is the need for a comprehensive statistical model capable of efficiently integrating these complex data structures and providing accurate

and actionable predictions. This gap limits educational institutions' ability to provide tailored solutions and impairs their ability to engage in strategic planning and resource allocation. Furthermore, there needs to be a consensus on the best statistical model for assessing panel data in education. Various models exist, each with theoretical benefits and limits, but comparison assessments are few, leaving institutions unsure of which model best fits their data and aims. This uncertainty hinders attempts to raise graduation rates and strengthen student support systems. Furthermore, the issue of unobserved heterogeneity—variables that impact student progress but cannot be measured—presents a considerable barrier. Traditional models frequently ignore these impacts, thus resulting in skewed estimations and inferior forecasts. An urgent need is for a model that can properly adjust for these unobserved variables, resulting in a more accurate representation of student outcomes. Addressing these problems necessitates a comprehensive examination of panel data models to determine the most efficient method for forecasting student graduation rates. This research addresses a methodological gap by evaluating several panel data models and identifying which model best accounts for both observable and unobserved factors in the educational environment. The research aims to provide practical insights to help academic institutions improve their predictive analytics skills, resulting in better educational results and institutional efficiency.

2. Literature Review:

The study on educational outcomes has long recognized the difficulty of reliably forecasting student graduation rates. The intricacy stems from various elements, including individual student traits, institutional issues, and larger socioeconomic situations. Traditional statistical models, which frequently depend simply on time series or cross-sectional data, have yet to be proven inadequate in fully capturing these dynamics. As a result, panel data has gained popularity, giving a more comprehensive approach by combining temporal and cross-sectional data, resulting in a richer analytical framework. Panel

data's capacity to accept this complexity originates from its fundamental structure, which allows for modeling several dimensions, including variability across participants over time and among various organizations. This dual emphasis not only improves the robustness of the analysis but also aids in controlling unobserved variability, which is a key difficulty in educational research because many affecting factors are not readily measured or seen.

The use of panel data in educational contexts has been gradually established in the literature, with significant increases in predicted accuracy and dependability over traditional models. For example, fixed and random effect models are frequently praised for handling the quirks inherent in educational data. Fixed effect models are particularly praised for controlling time-invariant features across individuals or groups, isolating the impact of variable changes over time. Random effect models are appropriate when entity changes are considered random and uncorrelated with predictor or outcome factors. Furthermore, the research shows that mixed models, which incorporate aspects of fixed and random effects models, provide greater flexibility when dealing with data with several degrees of variability. These models effectively cope with circumstances in which some aspects are better treated as fixed and others as random, resulting in a personalized approach that closely matches the complicated structure of educational data. Despite the accepted benefits of panel data, its use in educational research has challenges. The methodological rigor necessary for model selection, the assumptions behind each model, and the computing complexity of dealing with big datasets are all common topics in academic discussions. These limitations highlight the need for continued research and methodological improvement to improve the use of panel data models in education. {[1],[6],[13],[15]}

Literature consistently supports panel data analysis as a preferable tool for educational research, particularly when predicting outcomes such as student graduation rates. It provides a sophisticated approach that traditional approaches do not,

making it a must-have tool for educational researchers looking to improve forecast accuracy and contribute significantly to educational planning and policymaking. This work lays the groundwork for the present study's investigation of data models for forecasting student graduation estimations, aiming to identify and recommend the most efficient model based on evidence [3],[4],[10],[16].

3. Methodology:

The methods component of this research uses panel data models to analyze time series and cross-sectional data concurrently, allowing for a full review of factors impacting student graduation rates across time. Different panel data models—Pooled ordinary least squares, Fixed effects, Random effects, and Mixed models—handle various assumptions and unobserved heterogeneities within the departments studied. This technique enables robust analysis by considering the data's specific properties, allowing for a more in-depth understanding of the dynamics at play.

3.1. Panel Data Models:

The methodology used in this research is panel data analysis, which offers a multidimensional approach by combining time series and cross-sectional data. This enables a more thorough assessment of the variables impacting student graduation rates across departments over time. The research employs a variety of panel data models, each designed to evaluate various assumptions about the data and the underlying mechanisms that influence student outcomes [11].

3.2. Pooled Ordinary Least Squares (POLS):

The pooled ordinary least squares model is the simplest type of panel data econometric model. It is assumed that there is no unobserved heterogeneity among cross-sections, meaning that all differences across units (e.g., departments) may be represented by observable variables [18]. The model is defined as:

$$y_{it} = \beta_0 + x_{it}\beta + \epsilon_{it} \dots \dots \dots (1)$$

Where: $\text{Var}(\epsilon_{it}) = \sigma^2$, $E(\epsilon_{it}) = 0$
 y_{it} indicates the graduation rate of the i th student in the year, x_{it} comprises independent variables like student grades, and ϵ_{it} is the error term.

3.3. Fixed Effects Model (FEM):

The fixed effects model (FEM) takes into account unobserved heterogeneity that is constant across time and correlates with independent variables. This model is critical when the difference between entities (departments) is anticipated to influence the dependent variable [16]. The model is represented as:

$$y_{it} = \alpha_i + \beta_1 x_{it} + \epsilon_{it}$$

α_i represents the unique influence of the i th entity, which does not change over time.

3.4. Random Effects Model (REM):

This model assumes that entity fluctuations are uncorrelated with independent variables across time [17]. The equation for this model is

$$y_{it} = \gamma + \beta_1 x_{it} + u_i + \epsilon_{it}$$

where u_i represents the random influence of the i th entity and ϵ_{it} represents the idiosyncratic error.

3.5. Mixed Model:

The mixed model combines the regression framework's fixed and random effects models [18]. The equation for this model is

$$Y = \beta_0 + \sum \beta_i X_i + \gamma + \epsilon$$

Y is the dependent variable, β_0 is the global intercept, β_i are the fixed effect parameters, X_i are the fixed effect variables, γ represents random effect variance, and ϵ is the residual variance.

4. Practical Applications:

This research's data was compiled from records from four departments at Cihan University-Sulaymaniyah over five years. Each department reported data on 10 students every year for 200

observations. The dependent variable (Y) reflects the students' graduation rates, whereas the independent variable (X) is a weighted average of their grades throughout the first, second, and third years.

4.1. Statistical Analysis:

The research begins with data cleaning and preparation, in which the dataset is evaluated for missing values and outliers to ensure the robustness of the next analysis. Descriptive statistics are presented to offer an overview of data distribution and central patterns.

The models are fitted using the R statistical program, allowing extensive data manipulation and testing. Model diagnostics verify normality, homoscedasticity, and multicollinearity.

4.2. Model Selection Criteria:

Several statistical factors are used for selecting models:

- Utilize Root Mean Square Error (RMSE) to assess the average size of mistakes.
- In a regression model, R-squared (R^2) represents the percentage of variation explained by an independent variable.

Table 1: Model Evaluation Criteria

Model Type	RMSE	R-Squared
Pooled Ordinary Least Squares (POLS)	2.18	0.956
One-Way Fixed Effect (1-FEM)	1.15	0.966
Two-Way Fixed Effect (2-FEM)	1.13	0.967
One-Way Random Effect (1-REM)	1.28	0.956
Two-Way Random Effect (2-REM)	1.27	0.956
Mixed Effects Model (MEM)	1.14	0.966

Each model's parameters and assumptions are compared to these criteria to identify the best match for the data. This

comprehensive approach guarantees that the chosen model fits the data well and properly captures the underlying mechanisms that influence student graduation rates.

4.3. Results:

This research analyzes alternative panel data models to accurately forecast student graduation rates, using a dataset of records from Cihan University - Sulaymaniyah's four departments for five years.

4.4. Stationarity tests:

The dataset's stationarity was initially assessed using the Augmented Dickey-Fuller (ADF) test to determine its appropriateness for panel data analysis. The ADF test was used in three distinct forms, Level, First Difference, and Second Difference, to investigate the presence of unit roots in graduation rates (Y) and weighted mean grades over three years (X1).

Table 1. Augmented Dickey-Fuller Test Results

Test Condition	ADF Statistic	P-Value	Critical Values	Conclusion
Level	-11.76509	0.0000	1%: -3.463, 5%: -2.876, 10%: -2.575	Reject H0; Stationary
First Difference	-12.98409	0.0000	1%: -3.464, 5%: -2.876, 10%: -2.575	Reject H0; Stationary
Second Difference	Data not provided in the input text for actual ADF statistic values			

The findings reveal stationarity in the Level and First Difference conditions, with significant ADF statistics (p-value < 0.05). This allows us to reject the null hypothesis of a unit root for the series.

4.5. Model Evaluation:

The stationary data was analyzed using multiple panel models to determine the best fit based on RMSE and R² values. Pooled Least Squares, One- and Two-Way Fixed Effects, One-Way Random Effects, and Mixed Effects were among the tested models.

Model Type	RMSE	R ²	Model Details
Pooled Least Square (POLS)	2.180554	0.956099	Variables: C, X1
One-Way Fixed Effect (FEM)	1.146276	0.966072	Variables: C, X1
Two-Way Fixed Effect (FEM)	1.130461	0.967002	Variables: C, X1
One-Way Random Effect (REM)	1.276203	0.955864	Variables: C, X1
Two-Way Random Effect (REM)	1.271789	0.955986	Variables: C, X1
Mixed Effect (MEM)	1.141689	0.966225	Variables: C, X1

The Two-Way Fixed Effect model is the most effective, with the lowest RMSE (1.130461) and greatest R² (0.967002), indicating higher predictive performance over other models.

4.6. Algorithmic representation of the best model: Figure 1 depicts an algorithmic representation of the specified Two-Way Fixed Effect model.

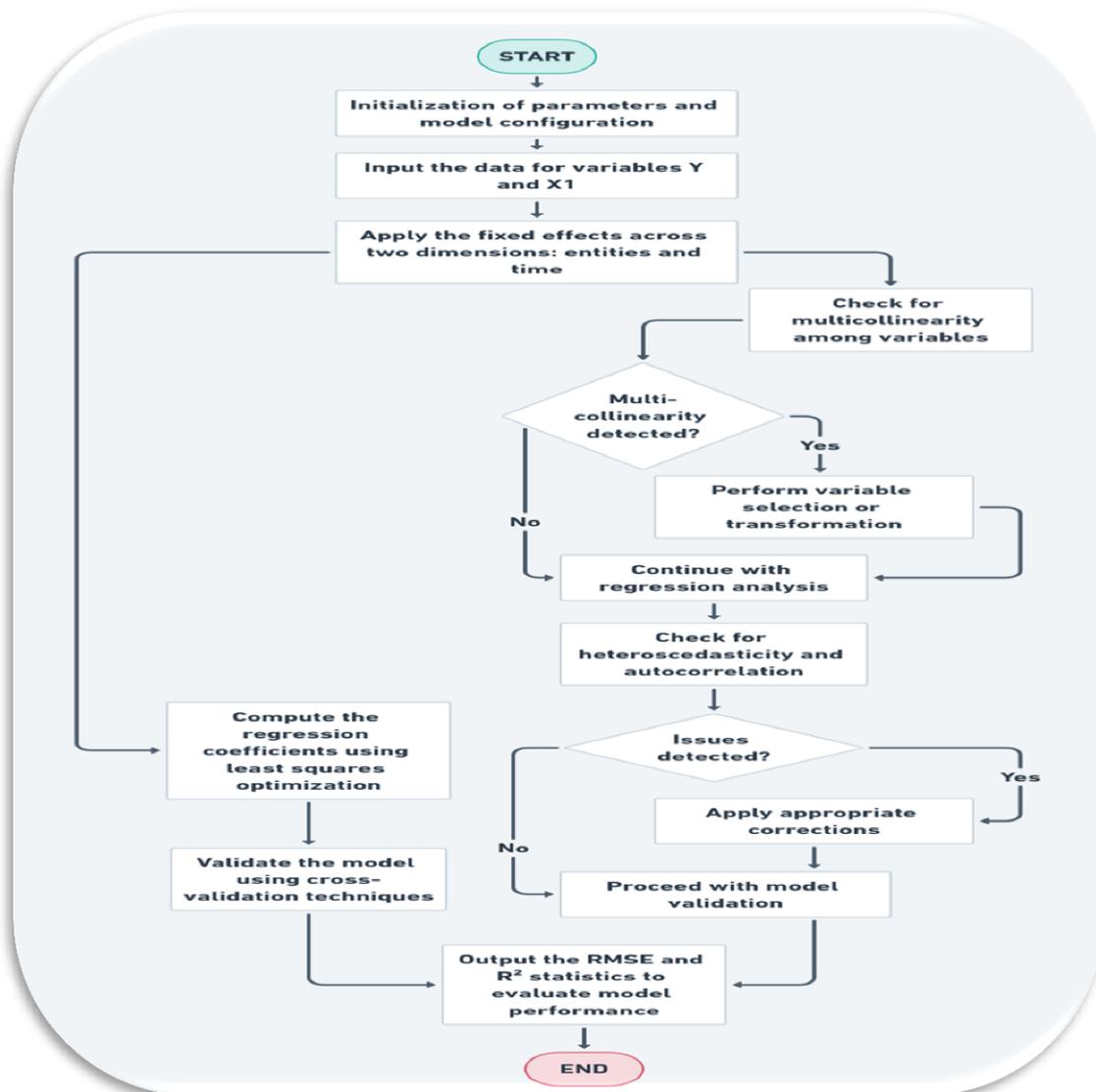


Figure 1. Enhancing Predictive Accuracy with the Two-Way Fixed Effect Model in Educational Data Analysis

This methodological approach demonstrates the resilience of applying sophisticated panel data approaches to manage complicated educational data, providing academic administrators with a trustworthy tool for efficiently forecasting student outcomes. The Two-Way Fixed Effect model's superiority demonstrates its capacity to account for differences within departments and across time, indicating a more comprehensive knowledge of the factors impacting student graduation rates.

5. Discussion:

In the scholarly conversation around the predictive modeling of student graduation rates, the introduction of panel data analysis represents a significant improvement. This research's investigation of several statistical models, notably the Two-Way Fixed Effect model, represents a significant advancement in understanding and forecasting educational results. The outcomes of this research not only confirm the value of panel data but demonstrate its superiority over traditional models, which frequently overlook the multidimensional character of educational data.

Panel data is designed to capture cross-sectional and time-series variations, allowing for a more complete investigation of changes over time and across organizations. This skill is especially useful in educational contexts, as both temporal and entity-specific effects substantially impact student results. As indicated by the results, the Two-Way Fixed Effect model displays a remarkable capacity to account for unobserved heterogeneity, which has been a significant shortcoming in prior investigations. Traditional statistical techniques, such as time series or cross-sectional analysis, have yielded limited insights since they cannot address fluctuations over time and entities at the same time. Previous research has frequently produced skewed or inconsistent results when these dimensions cross, especially when unobserved heterogeneity relates to explanatory factors. The Two-Way Fixed Effect model solves this problem by explicitly modeling time-invariant and entity-specific changes, resulting in more trustworthy and accurate predictions. Furthermore, using this model to forecast student graduation rates demonstrates its practical significance. The Two-Way Fixed Effect model surpasses its competitors by obtaining the lowest Root Mean Square Error (RMSE) and the best R-squared value among the evaluated models, and it closely aligns with the theoretical predictions mentioned in the literature on panel data effectiveness. This alignment demonstrates the model's ability to handle the intricacies inherent in educational data, including

various impacting factors, from individual student actions to institutional rules.

The findings of this research may also be compared to other studies, particularly those that used single-dimensional data analysis methodologies. It has been shown that models that only contain multidimensional data frequently pay attention to important determinants of student achievement, such as departmental features or time-related changes in educational policy. In contrast, the Two-Way Fixed Effect model's capacity to combine various characteristics leads to a more comprehensive understanding of the factors that impact graduation rates.

Furthermore, the practical ramifications of these discoveries are enormous. For policymakers, the ability to reliably anticipate student outcomes is critical. It provides better-informed decision-making about resource allocation, curricular changes, and intervention measures, all critical to increasing student success rates. Furthermore, the analytical framework presented in this research serves as a roadmap for future research, implying that comparable methodologies might be used in other areas of educational research, such as student retention rates or academic performance.

This work considerably advances the use of panel data in educational research. In particular, the Two-Way Fixed Effect model makes a convincing argument for its use in comparable circumstances. It is a methodologically sound and practically relevant tool for educational researchers and administrators. This research not only adds to academic literature by proving the efficiency of a complex statistical tool, but it also emphasizes the necessity of selecting proper analytical procedures to answer specific research questions in the field of education.

6. Conclusion & Recommendations:

The analysis of the prediction accuracy of several statistical models for projecting student graduation rates using panel data resulted in valuable insights, as evidenced by the findings of this study. The research thoroughly investigated different panel data

models, ultimately determining that the Two-way Fixed Effect Model was the better choice based on performance measures. This finding is reinforced by extensive statistical research, demonstrating the model's resilience in dealing with the intricacies of educational data.

The research was found on many hypotheses about the usefulness of various models in predicting student outcomes. The key hypothesis stated that changes in model performance will be visible when applied to real-world educational data. This prediction was supported by data, which revealed various degrees of predictive accuracy across the models, with the Two-way Fixed Effect Model beating the others in terms of both Root Mean Square Error (RMSE) and R-squared.

One of the most important discoveries of this research is the dataset's stationarity, which emphasizes the dependability of the conclusions derived via panel data analysis. The stationarity of the data indicates that the numbers have a stable mean and variation throughout time, which increases the validity of the econometric models used. This feature of the data enabled the direct use of statistical models without the need for preparatory modifications, which might complicate the research or create new sources of error.

The Two-way Fixed Effect Model, with the lowest RMSE and greatest R-squared among the examined models, has shown to be a very successful tool for assessing educational data. This model effectively addresses unobserved heterogeneity, which might skew the findings if left unaccounted for, resulting in a more accurate and trustworthy estimate of student graduation rates. This model's capacity to integrate individual and time-specific effects makes it ideal for educational settings where student attributes and temporal dynamics are important.

Besides technical validation of the Two-way Fixed Effect Model, this work emphasizes the need to choose statistical models depending on data structure and research aims, advancing educational research techniques. Complex educational data may be analyzed using the Two-way Fixed Effect Model's capacity to handle unobserved heterogeneity

across time and individual effects. This resilience is essential in educational research, where factors interact in complex ways throughout time and environment.

According to the research, advanced statistical methods may improve university and staff decision-making. The Two-way Fixed Effect Model improves graduation rate projections for successful educational planning. Correct forecasts help institutions allocate resources, create informed policies, and execute strategic plans that enhance educational results.

University administration requires resource allocation. Universities can better prepare for future needs and allocate resources where they will have the most effect with precise graduation rate estimates. To address the root causes of graduation rate reduction, a university may invest more in student support services, academic counselling, or curriculum improvement.

Another use of sophisticated statistical models is policy formation. These models may help universities create student-friendly policies. For instance, tailored interventions may help students in demographics at increased risk of not graduating. Mentorship, financial assistance, and admission modifications may create a more conducive learning environment.

High-tech statistical techniques improve strategic planning. University long-term strategies may match student outcome patterns. Foresight permits organizations to anticipate issues and seize possibilities for development. University academic offerings may be adjusted to accommodate enrollment fluctuations, for example. The methodological advances in this work affect educational research. The Two-way Fixed Effect Model proves its usefulness, setting a precedent for comparable techniques in other educational studies on student retention and academic attainment. Integrating additional granular data, such as instructor quality or student engagement levels, might improve these models' prediction value by revealing student success variables.

This research confirms the Two-way Fixed Effect Model's technical superiority and emphasizes the need for enhanced

statistical tools in educational research. The research provides a robust methodology for accurate graduation rate prediction, helping universities and university personnel enhance decision-making, resource allocation, policy formation, and strategic planning, improving educational results.

This research has larger ramifications, including recommendations for future investigations. It is proposed that comparable approaches be used in other areas of educational research, such as student retention or academic achievement, to verify the Two-way Fixed Effect Model. Additionally, investigating the integration of more granular data, such as teacher quality or student engagement levels, may reveal deeper insights into the elements impacting student performance.

This work confirms the use of panel data models in educational research and establishes a precedent for using advanced statistical analysis to understand better and improve educational results. With its strong performance, the Two-way Fixed Effect Model is a suggested tool for educational researchers and institutions looking to use data-driven insights to promote academic success. This research adds a crucial piece to the jigsaw of educational accomplishment by using methodological rigor and detailed analysis, providing a model for future studies to improve student outcomes' prediction accuracy.

References:

- [1]-Abdouli, M., & Hammami, S. (2017). Investigating the causality links between environmental quality, foreign direct investment and economic growth in MENA countries. *International Business Review*, 26(2), 264-278.
- [2]-Al-Iriani, M. A. (2006). Energy–GDP relationship revisited: an example from GCC countries using panel causality. *Energy policy*, 34(17), 3342-3350.
- [3]-Altaee, H. H. A., Saied, S. M., Esmaeel, E. S., & Adam, M. H. M. (2014). Financial development, trade openness and economic growth: Evidence from sultanate of Oman (1972-2012). *Journal of Economics and Sustainable Development*, 5(23).

- [4]-Alyasari, H. H. A., & Alzawbaee, O. M. M. (2020). Panel data evidence on the relationship between economic growth, electricity consumption and CO2 emissions in the GCC. *The Scientific Journal of Cihan University–Slemani PP*, 33, 47.
- [5]-Al-Zawbaee, O. M., & Mahmood, N. H. (2023). Building an Efficient Model by Using Panel Data. *Cihan University-Erbil Journal of Humanities and Social Sciences*, 7(1), 82-87.
- [6]-Anwar, S., & Sun, S. (2011). Financial development, foreign investment and economic growth in Malaysia. *Journal of Asian Economics*, 22(4), 335-342.
- [7]-Apergis, N., & Payne, J. E. (2009). Energy consumption and economic growth in Central America: evidence from a panel co-integration and error correction model. *Energy economics*, 31(2), 211-216.
- [8]-Balasubramanyam, V. N., Salisu, M., & Sapsford, D. (1996). Foreign direct investment and growth in EP and IS countries. *The economic journal*, 106(434), 92-105.
- [9]-Baltagi, B. H., & Baltagi, B. H. (2008). *Econometric analysis of panel data* (Vol. 4, pp. 135-145). Chichester: Wiley.
- [10]-Hassan, M. K., Sanchez, B., & Yu, J. S. (2011). Financial development and economic growth: New evidence from data panel. *The Quarterly Review of economics and finance*, 51(1), 88-104.
- [11]-Hsiao, C. (2022). *Analysis of panel data* (No. 64). Cambridge University Press.
- [12]-Husein, J. (2010). Export-led growth hypothesis in the MENA region: A multivariate cointegration, causality and stability analysis. *Applied Econometrics and International Development*, 10(2), 161-174.
- [13]-Khamfula, Y. (2007). Foreign direct investment and economic growth in EP and IS countries: The role of corruption. *World Economy*, 30(12), 1843-1854.
- [14]-Lee, C. H., & Huang, B. N. (2002). The relationship between exports and economic growth in East Asian countries: A multivariate threshold autoregressive approach. *Journal of Economic Development*, 27(2), 45-68.

[15]-Mohammed, P. A., Obed, S. A., Ali, I. M., & Kadir, D. H. (2022). US Dollar/IQ Dinar currency exchange rates time series forecasting using ARIMA model. Cihan University-Erbil Scientific Journal, 6(1), 12-19.

[16]-Ndambiri, H. K., Ritho, C., Ng'ang'a, S. I., Kubowon, P. C., Mairura, F. C., Nyangweso, P. M., ... & Cherotwo, F. H. (2012). Determinants of economic growth in Sub-Saharan Africa: A panel data approach.

[17]-Ozturk, F. (2017). Energy consumption–GDP causality in MENA countries. Energy Sources, Part B: Economics, Planning, and Policy, 12(3), 231-236.

[18]-Were, M. (2015). Differential effects of trade on economic growth and investment: A cross-country empirical investigation. Journal of African trade, 2(1), 71-85.

Forecasting future sales prices and showing changes in Erbil Diamond Zone using time series models

Asst. Prof. Dr. Mohammed Abdul Majeed Badal Banaz Didar
Yusuf

Mohammed.badal@su.edu.krd banazdyosuf@gmail.com

Sallahaddin University-Statistics Dep.

Kurdistan Region -Iraq

Abstract:

In this study, a statistical study was conducted on one of the most important residential projects, which is Diamond Residential Community in Erbil, due to its importance and prominent location in Erbil, the capital of Iraqi Kurdistan Region V.19 STATGRAPH It was found that citizens are willing to buy and invest in Diamond Complex - Erbil due to the availability of basic services and different geographical location, different climate, especially overlooking the tourist road (Masif - Shaqlawa) and connected to 120m highways 150m the purchase price of these housing units through statistical criteria that minimize the value of FPE, SBIC, HQIC, the best model that represented was the model. ARIMA (2,2,2)

Key words: ARIMA Models, Moving average, Exponential smoothing, FPE, SBIC, HQIC, R^2 , AIC.

استخدام نماذج السلسلة الزمنية للتغيرات الحاصلة على اسعار المجمعات السكنية
في مجمع الماس-مدينة أربيل للتنبؤ بأسعار المبيعات المستقبلية

د. محمد عبد المجيد بدلبناز ديدار يوسف

Mohammed.badal@su.edu.krd banazdyosuf@gmail.com

جامعة صلاح الدين-قسم الاحصاء والمعلوماتية-اقليم كورستان العراق

المخلص :

تم في هذا البحث اجراء دراسة احصائية على احد المشاريع المهمة الاستراتيجية السكنية وهو مجمع الماس السكني بمدينة اربيل لما له اهمية وموقع متميز في اربيل عاصمة اقليم كورستان العراق مستندة على بيانات لمركز البيع المباشر ومكاتب

العقارات لمجمع الماس وتحليل البيانات للمبيعات السكنية من حيث الموقع والمساحة والسعر خلال سلسلة زمنية والتنبؤ بها بعدة نماذج مختارة وباستخدام برنامج STATGRAPH V.19 تبيين من خلال التحليل توجه وإقبال المواطن على شراء واستثمار الوحدة السكنية لمجمع الماس - أربيل بسبب توفر الخدمات الأساسية والموقع الجغرافي المتميز خاصة انها مطلة على الطريق السياحي (المصيف - شقلاوة) والمرتبطة بالشوارع السريعة 120م و150م مما أدى الى ارتفاع اسعار الشراء لهذه الوحدات السكنية وعن طريق المقاييس الاحصائية التي تعطي اقل قيمة FPE, SBIC, HQIC كان افضل نموذج يمثلها هو نموذج ARIMA(2,2,2)

Introduction:

The most important sector of the economy is the real estate sector. Economists and experts say that the real estate sector is the fastest growing sector After 2005, the Kurdistan Region saw a good development in the real estate sector, due to the expansion of its cities, especially Erbil Map of Erbil between 2005 and 2024 in terms of development, the beautiful residential and commercial projects, strategic streets and surrounding circles Erbil, the expansion of Erbil in the direction of MasifShaqlawawa Street, which is considered the golden road and the heart of the diamond zone, the most important residential and commercial area There are 8 main roads that connect other parts of Iraq and the Kurdistan Region: Birmam, Shaqlawa, Koya, Sulaimani, Kirkuk. Makhmur - Tuir - Mosul - Duhok - Bahrka).

Erbil province is located in Iraq in the Kurdistan Region, one of the strategic and commercial provinces, there are several main roads in this province that connects to several other cities and provinces province 'city Erbil castle in The castle was built in a circular shape 3000 years ago. Because of this circular shape, houses and streets were slowly built around it, so Erbil took a circular shape around the castle There is a circular street, then 30 meters, then 60 meters, 40 meters and 100 meters, all built in a circle and ring around the castle with different diameters which have not yet been completed ,

From M120 Circular Street to M150 Circular Street and from KesnazanKoya to Bahrka, it is called the widest zone of Erbil,

which is the Diamond Zone The population structure of the zone has caused people from other countries to move to the zone day by day for any reason, residence, business or work, because it is a place All genes must be in the zone at any level of life , the 150-meter highway is one of the strategic projects of the Kurdistan Regional Government (KRG) that facilitates the movement of citizens. The 150-meter highway is 70 kilometers long and circles around Erbil Work is underway, one is 7.6 kilometers, the other It is 6.1 kilometers. That means the total length is 24 kilometers in120th Highway the entire 38-kilometer project consists of four lines, two of which are high-speed entrance and exit lines, and two of which are service lines on both sides of the high-speed lines.



Figure(1)Diamond Zone

Diamond Zone is strategically located between the two circular streets M120 and M150, the golden road MasifShaqlawa is considered to be the heart of the Diamond Zone, which connects Erbil to Iran directly from Erbil Castle to Haji Omeran. In the farthest part of the zone, it takes only 10 to 20 minutes to the

center of Erbil, in terms of political and international embassies of countries such as Russia The United States will give more sanctions to this zone in the future and in the future it is expected that all international embassies will visit this zone, which will be the largest economically The most important feature of this zone is that most of the businessmen, investors and capitalists of the Kurdistan Region reside in this zone City - Mastiles - Future City - Far Thark Skyscraper Towers - Safiran City - Majidi Market - Iraq's largest mall Grand Magedila Mall - Italy 2- Pocket Construction of 120 and 150 streets, QaywanMirador, Media Center, Korek Towers, which are the most important projects.

Material and method

Time Series Forecasting

utilizes data amassed over an extended period to evaluate sales velocity in accordance with the requirements of the enterprise. Drawing upon these metrics, one may develop prospective predictions employing statistical equations. A variety of methodologies exist for conducting time-series forecasting, which will assist in constructing future projections.

Forecasting methods:

Quantitative methods of forecasting

In order to generate forecasts that are both realistic and precise, quantitative methodologies encompass mathematical procedures including:

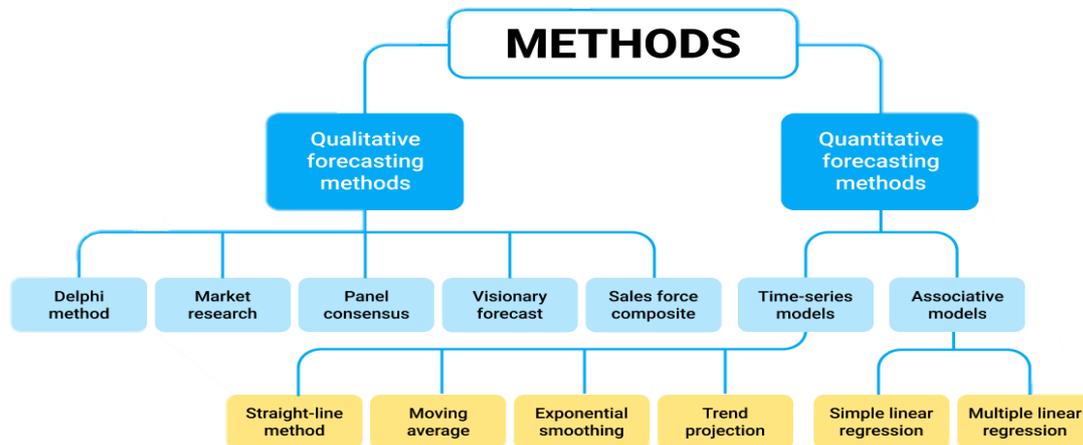
Algebra, permutations and combinations, set theory, matrix algebra, and integration.

Furthermore, mathematical methodologies such as linear programming, dynamic programming, and inventory management can additionally assist decision-makers in formulating their business strategies.

Qualitative methods of forecasting:

Qualitative approaches, in contrast to quantitative ones, are subjective in character and mostly depend on judgments, expert

opinions, intuition, emotions, or firsthand experiences. These forecasting techniques don't use any mathematical computations and are typically employed when the previous data is too limited or unlikely to be used in the future.



Figure(2) Methods of Time-series forecasting models

Time-series forecasting models :

Straight-line method

By considering historical data and trends, the straight-line technique is a time-series forecasting methodology that generates forecasts about future revenues.

Moving average model

Similar to straight-line forecasting, the moving average model is frequently used to forecast short-term trends (e.g., daily, monthly, quarterly, or half-yearly intervals).

Exponential smoothing model

Using a set of weighted averages derived from historical observations, exponential smoothing is another time-series forecasting methodology that may be used to anticipate new values, much like the moving average.

Trend projection model

When you can determine the future impact of specific variables (dependent or independent) based on their historical behavior, the trend projection model performs well.

Modelsofassociative (causal) forecasting .the predicted variable is related to other variables in the business system, according to associative or causal forecasting models.

Simple linear regression model

Simple linear regression is a kind of associative forecasting model that looks at the relationship between the independent and dependent variables to give your forecast a more thorough context.

Multiple linear regression model

As the name implies, the multiple linear regression model applies the same methodology, i.e., the same assumptions, as the simple linear regression model, but to a variety of business factors.

Delphi method

is a kind of forecasting model in which a small group of pertinent experts share their ideas and opinions about a certain issue or circumstance. Results and an accurate forecast are then produced by combining market orientation with expert viewpoints.

Market research model

Market research is a qualitative forecasting model that uses interviews with potential customers to assess how well a company's goods and services perform.

Panel consensus model

Panel consensus (also called expert opinion) is a qualitative forecasting approach where experts or employees from all levels of an organization (from low-level to top-level) discuss a product or service.

Visionary forecast model

A relevant and experienced person's personal ideas, judgments, and insights form the basis of the visionary forecasting model.

Sales force composite model

The sales force composite is another trustworthy qualitative forecasting methodology that uses sales staff input to project future sales.

Fitting the model:

The Akaike Information Criteria (AIC) can then be used to select the model that best balances these drawbacks because an over-fitted model lacks generality, while an under-fitted model might not accurately represent the variability in the outcome variable. The best model can be subjected to traditional null-hypothesis testing to determine the relationship between specific variables and the desired outcome, or the AIC criteria can be used to asymptotically overestimate the order with a positive probability:

$$AIC = 2K - 2\log(L(\hat{\theta}/y)) \quad (1)$$

The estimated model's maximum log-likelihood is $\log(L(\hat{\theta}/y))$, where K is the number of estimable parameters (or degrees of freedom). To take small data samples into consideration, this estimate has been further revised:

$$AICc = AIC + \frac{2K(K + 1)}{n - K - 1} \quad (2)$$

K and AIC are as previously mentioned if n is the sample size. If n is large in relation to K , this correction is insignificant and AIC is adequate. But $AICc$ is more widely used and often used instead of AIC . The best model is then the one with the lowest $AICc$ (or AIC) score. It's important to keep in mind that the AIC and $AICc$ scores are ordinal and meaningless on their own.

An estimate of the posterior probability of a model being true, given a specific Bayesian setup, indicates that a model is more likely to be the correct model if the Bayesian information criterion (BIC) is lower:

$$BIC = 2 \log n - 2 \log \left(L \left(\frac{\hat{\theta}}{y} \right) \right) \quad (3)$$

The Box-Ljung test is a diagnostic method for assessing the lack of fit in a time series model. Varsoc provides the final prediction error (FPE), Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), and Hannan and Quinn information criterion (HQIC) as lag order selection statistics for a sequence of vector auto-regressions of orders 1, ..., max lag(p). Under very broad circumstances, the order is consistently estimated by the BIC and HQ criteria.

Model-order statistics:

The following is the formula for the FPE (final prediction error), per Lutkepohl (2005, 147):

$$FPE = |\Sigma u| \left(\frac{T + KP + 1}{T - KP - 1} \right)^K$$

by whom wrote

$$FPE = |\Sigma u| \left(\frac{T + \bar{m}}{T - \bar{m}} \right)^K \quad (4)$$

where \bar{m} stands for the average number of parameters over the K equations. Collinearity-related variables are eliminated in this approach. According to Lutkepohl (2005), the constant term should be eliminated from the log likelihood because it has no effect on inference. the information requirements in Lutkepohl form are:

$$AIC = \ln|\Sigma u| + \frac{2PK^2}{T} \quad (5)$$

$$SBIC = \ln|\Sigma u| + \frac{\ln(T)}{T} PK^2 \quad (6)$$

$$HQIC = \ln|\Sigma u| + \frac{2\ln[\ln(T)]}{T} PK^2 \quad (7)$$

where LL is the log likelihood and PK^2 is the total number of parameters in the model.

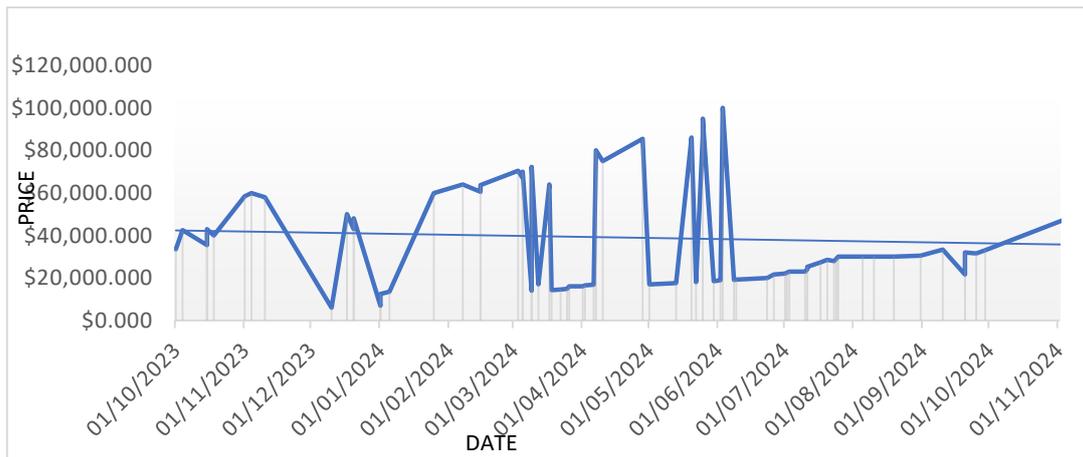
Practical Aspect

Researchers taken the data of House price change of Diamond Zone –Erbil city ,Number of observations = 219,the period (1-10-2023 to 1-11-2024) According of independent variables:

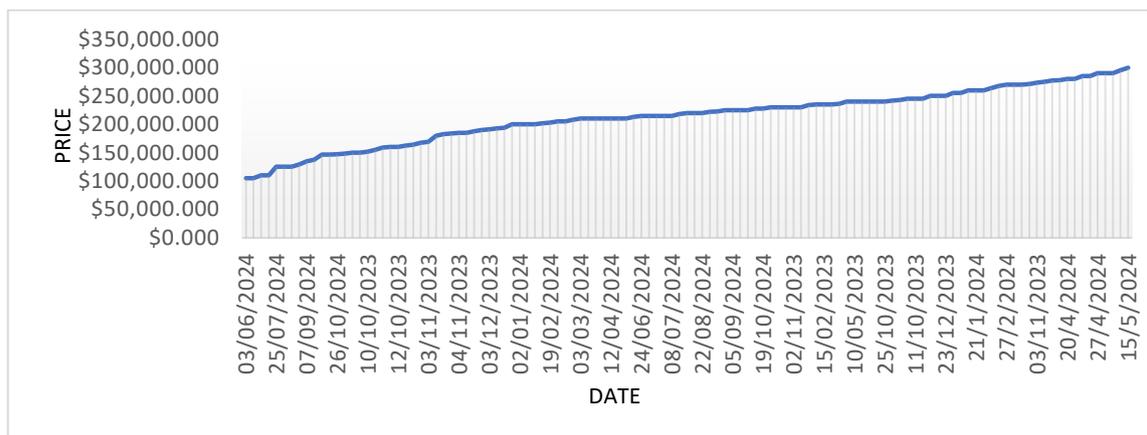
- 1-Property Type that divided by(Cadastre Land,,One - Storey House ,Two - Storey House,Land,Flat,Vella)
- 2- Area that was rang from (52 m²- 550m²)
- 3-Price that was rang from (\$6,000.000- \$850,000,000)
- 4- Geographical location.

used from analyzing Statgraph v.19 programming we draw the prices with date and analysis time series models then explain the measurements and predicting as follows in results .

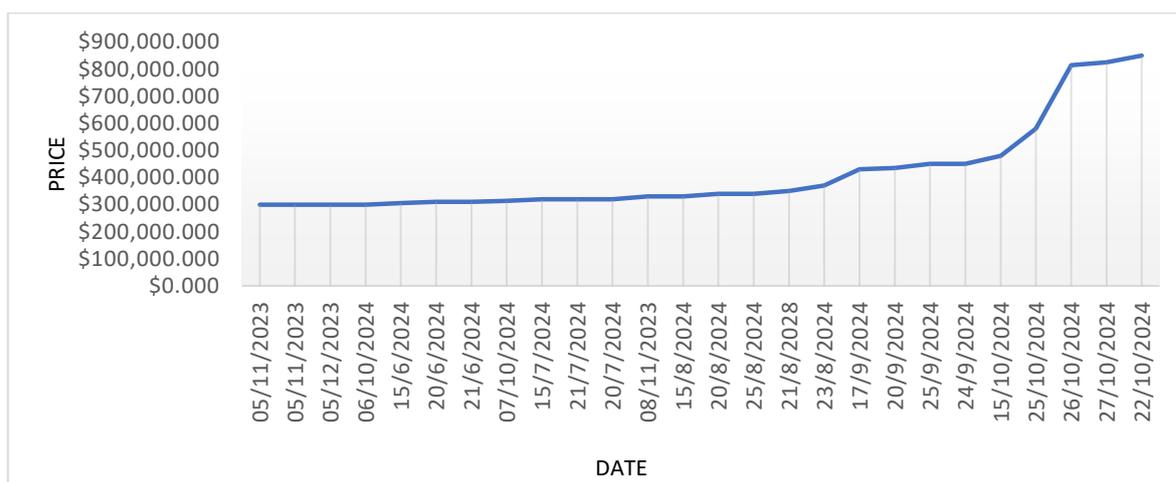
Results and discussion



**Figure(3) Prices (6.000\$ - 100.000\$) by date
Price changes of houses in this range during (10/2023 -
10/2024)**



**Figure(4) Prices in more than (100.000\$ - 300.000\$) by date
 Price changes of houses in this range during (10/2023 - 10/2024)**



**Figure(5) Prices & Date in more than (300.000 \$) by date
 Price changes of houses in this range during (10/2023 - 10/2024)**

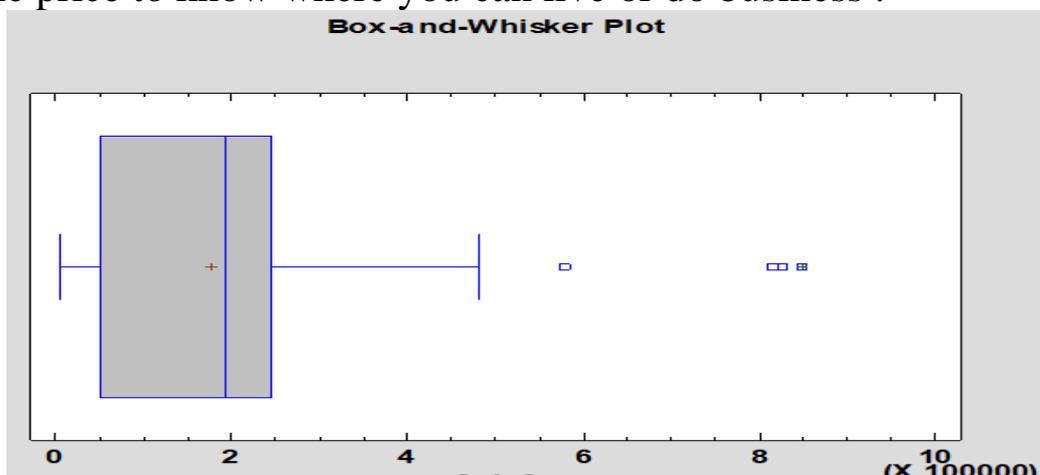
The prices of houses vary according to the area and location, but generally increased during the year .

The figures and tables show that prices vary from place to place for several reasons:

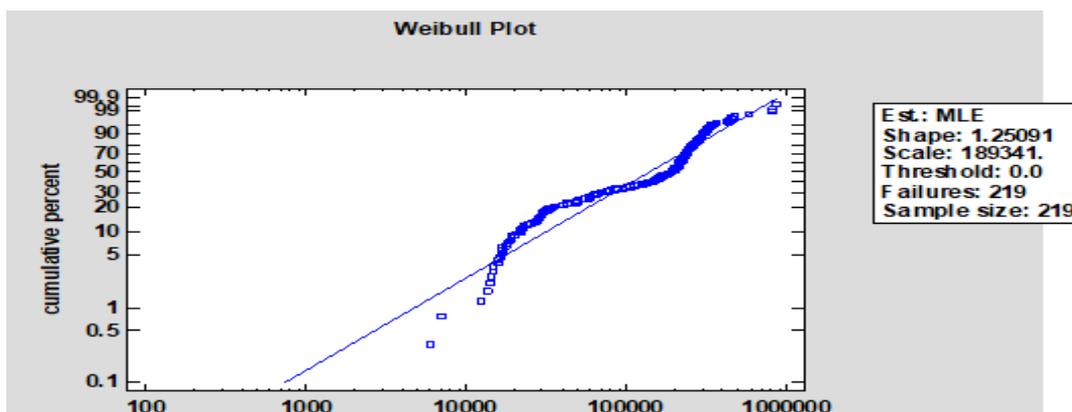
1. Geographical location
2. Area
3. service
4. Protected area
5. Natural disasters
6. renovated

7. House distribution (building map)

If you want to buy a property or a house, you must first look at the price to know where you can live or do business .



Figure(6) Stationary Mean and Variance of price
Display the average price and price difference without time



Figure(7) According to the charts above, the mean and variance stability is relatively stable .

Automatic Forecasting – Price

Data variable: Price (House prices in Diamond Zone)

Number of observations = 219

Time indices: Date (date in (10/2023) to (10/2024))

Forecast model selected: ARIMA(2,2,2)

Number of forecasts generated: 12

Table (1) Number of periods with Estimation Period

<i>Statistic</i>	<i>Estimation Period</i>
RMSE	14355.6
MAE	4047.93
MAPE	3.02162
ME	1362.21
MPE	0.438297

Table (2) ARIMA Model Summary

Parameter	Estimate	Std. Error	T	P-value
AR(1)	0.869418	0.0619978	14.0234	0.000000
AR(2)	-0.626103	0.0669427	-9.35282	0.000000
MA(1)	1.66714	0.0218535	76.287	0.000000
MA(2)	-0.963372	0.0287184	-33.5455	0.000000

Backforecasting: yes

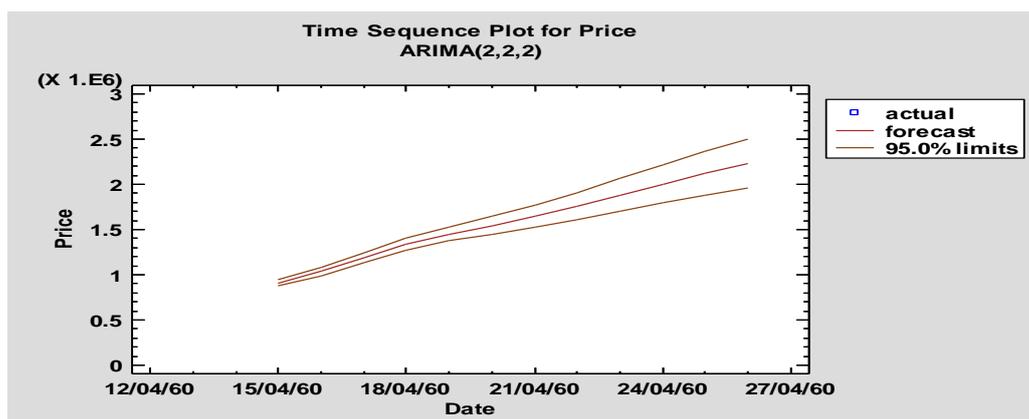
Estimated white noise variance = 2.06407E8 with 213 degrees of freedom

Estimated white noise standard deviation = 14366.9

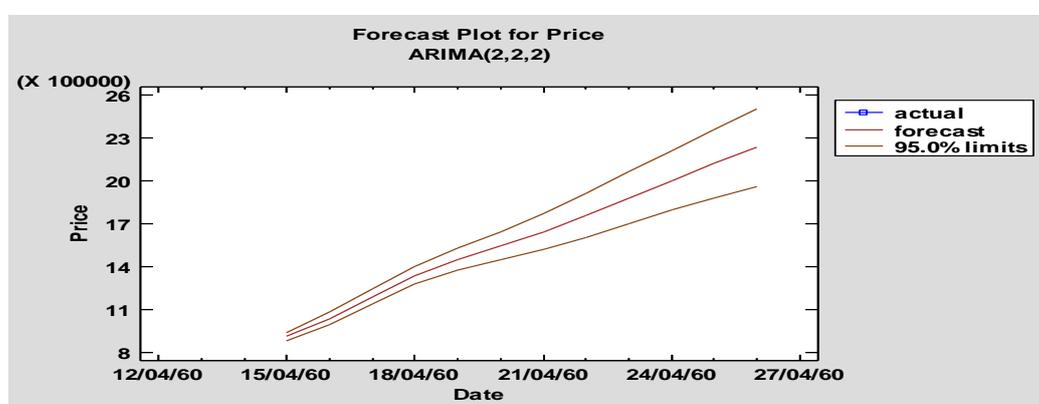
Number of iterations: 10

The Stat Advisor

This process will predict price values in the future. There are 219 time periods in the data. An autoregressive integrated moving average (ARIMA) model has been chosen as the current model. According to this concept, a parametric model that compares the most recent data value to earlier data values and noise provides the best forecast for future data. The forecasting model's words' statistical significance is compiled in the output. A 95.0% confidence level indicates that a term is statistically substantially different from zero if its P-value is less than 0.05. The AR(2) term is significantly different from 0 because its P-value is less than 0.05. The MA(2) term is significantly different from zero because its P-value is less than 0.05. The input white noise has an estimated standard deviation of 14366.9.



Figure(8) Prediction chart with ARIMA model (2,2,2)
This chart shows the forecast of future prices increasing over time without a modelmodel (2,2,2)



Figure(9) Prediction chart with ARIMA model (2,2,2)
This chart shows the forecast of future prices increasing over time without a modelmodel (2,2,2)

Models

- (A) Random walk
- (B) Random walk with drift = 3871.56
- (C) Constant mean = 176811.
- (D) Linear trend = $-36769.1 + 1941.63 t$
- (E) Quadratic trend = $7273.31 + 745.913 t + 5.4351 t^2$
- (F) Exponential trend = $\exp(9.9451 + 0.0156914 t)$
- (G) S-curve trend = $\exp(11.8488 + -6.51771 /t)$
- (H) Simple moving average of 2 terms
- (I) Simple exponential smoothing with alpha = 0.9999
- (J) Brown's linear exp. smoothing with alpha = 0.5459
- (K) Holt's linear exp. smoothing with alpha = 0.9999 and beta = 0.126

- (L) Brown's quadratic exp. smoothing with alpha = 0.2879
 (M) ARIMA(2,2,2)
 (N) ARIMA(2,2,1)
 (O) ARIMA(2,0,0) with constant
 (P) ARIMA(1,0,1)

Table (3) Estimation Period in time series models

Model	RMSE	MAE	MAPE	ME	MPE	AIC	HQC	SBIC
(A)	18278.2	3871.56	2.12735	3871.56	2.12735	19.6269	19.6269	19.6269
(B)	17904.5	4943.99	5.08453	-5.87417E-12	-3.87392	19.5948	19.601	19.6102
(C)	136896.	103875.	218.057	3.91241E-10	-186.277	23.6631	23.6693	23.6786
(D)	60172.4	28786.4	37.1329	3.18946E-12	16.4708	22.0282	22.0407	22.0592
(E)	57050.6	29767.0	18.2035	-1.39007E-10	-6.33968	21.9308	21.9495	21.9772
(F)	75026.0	52915.1	31.8928	-7099.77	-6.61367	22.4694	22.4819	22.5004
(G)	137667.	100525.	106.948	53044.0	-53.3232	23.6835	23.696	23.7144
(H)	23791.2	5772.35	3.08514	5772.35	3.08514	20.1633	20.1695	20.1788
(I)	18278.9	3854.26	2.11784	3854.25	2.11783	19.6361	19.6424	19.6516
(J)	17253.0	4144.95	2.49331	1061.52	0.724848	19.5206	19.5269	19.5361
(K)	16486.3	3718.36	1.99095	1516.33	0.256099	19.4388	19.4513	19.4698
(L)	17560.8	3946.01	2.74543	1255.05	0.702233	19.556	19.5622	19.5715
(M)	14355.6	4047.93	3.02162	1362.21	0.438297	19.1803	19.2053	19.2422
(N)	15987.1	3891.8	2.02456	1356.47	0.231188	19.3865	19.4052	19.4329
(O)	16052.5	5107.3	4.1423	-79.3611	2.18209	19.3946	19.4134	19.4411
(P)	16132.3	5307.87	2.61967	-1089.29	-0.64466	19.3954	19.4079	19.4264

Appendix-Table (3)

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	18278.2	***	OK	***	OK	***
(B)	17904.5	***	OK	***	OK	***
(C)	136896.	***	***	***	***	***
(D)	60172.4	***	***	***	OK	***
(E)	57050.6	***	***	***	OK	***
(F)	75026.0	***	***	***	***	***
(G)	137667.	***	***	***	***	***
(H)	23791.2	***	***	***	*	***
(I)	18278.9	***	OK	***	OK	***
(J)	17253.0	***	OK	***	OK	***
(K)	16486.3	**	OK	OK	OK	***
(L)	17560.8	***	**	***	OK	***
(M)	14355.6	**	**	OK	OK	***
(N)	15987.1	**	**	OK	OK	***
(O)	16052.5	OK	***	OK	OK	***
(P)	16132.3	***	***	OK	OK	***

Key:

RMSE = Root Mean Squared Error

RUNS = Test for excessive runs up and down

RUNM = Test for excessive runs above and below median

AUTO = Ljung-Box test for excessive autocorrelation

MEAN = Test for difference in mean 1st half to 2nd half

VAR = Test for difference in variance 1st half to 2nd half

OK = not significant ($p \geq 0.05$)

* = marginally significant ($0.01 < p \leq 0.05$)

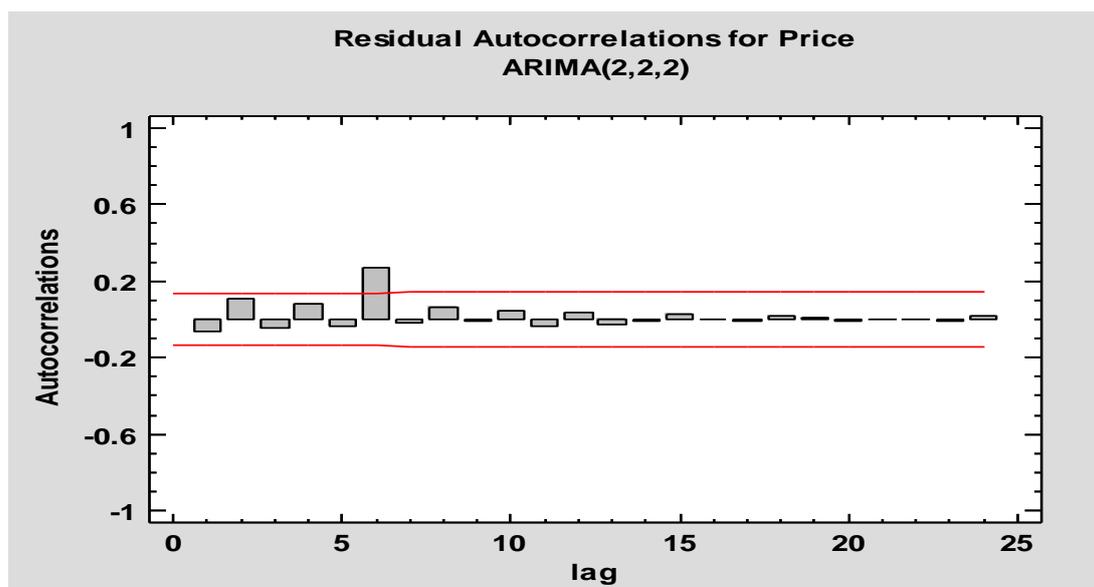
** = significant ($0.001 < p \leq 0.01$)

*** = highly significant ($p \leq 0.001$)

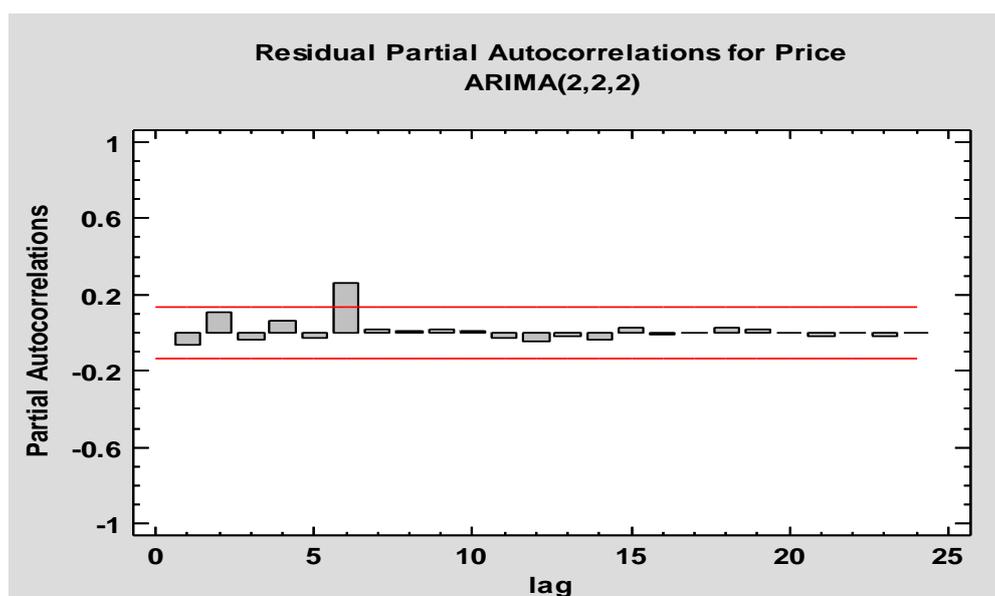
The StatAdvisor

The outcomes of fitting several models to the data are contrasted in this table. Among the models employed to make the forecasts, model M has the lowest Akaike Information Criterion (AIC) score.

The results of five tests performed on the residuals to assess each model's suitability for the data are also summarized in the table. If the model receives an OK, it has passed the test. A 95% confidence level failure is indicated by a *. At the 99% confidence level, two *'s indicate failure. Three *'s indicate a 99.9% failure rate. It should be noted that model M, which is currently chosen, passes two tests. You should give choosing a different model careful thought since one or more tests are statistically significant at the 95% or higher confidence level.



Figure(10) Autocorrelation Price Prediction

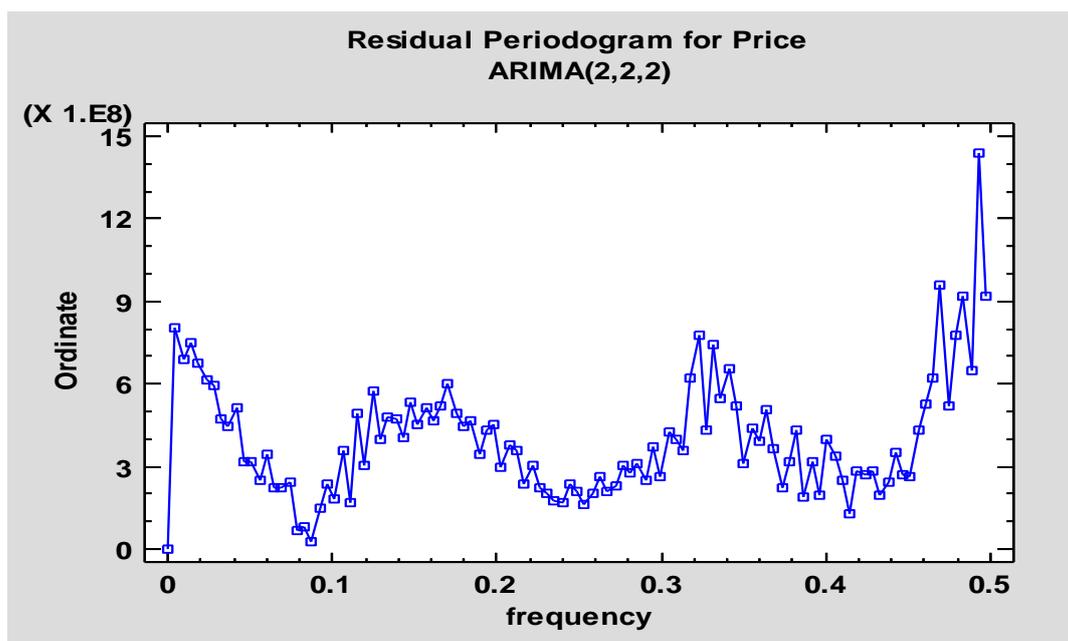


Figure(11) Autocorrelation Price Prediction

Periodogram for residuals

Data variable: Price

Model: ARIMA(2,2,2)



Figure(12) Periodogram Price Prediction

Conclusions:

In general, after the forecast for the data, it was found that over time, the price of property in the Diamond Zone will increase due to the projects, reconstruction and greenery that is set for it in the coming years)Sami Abdulrahman Park 2 - Mosque JalilKhayat 2 - Similar to BurjKhalifa) that the projects are located in this area that will have a huge impact on the increase in house prices and property in this zone, according to the results that we found a little high and low house prices , However, it is generally increasing House prices have been increasing year after year due to several main reasons such as:

1. Not granting land for housing construction.
2. Prevent the building of unapproved homes (encroachment).
3. Immigrants' arrival.
4. The replacement of rural and urban areas would result in more housing demand and higher costs since many rural residents relocate to the cities.
5. Self-reduction. People sell enormous homes in order to acquire multiple modest properties.

6. There are many excellent services in the neighborhood, including electricity, water, and roads.

7. Near public areas such as parks, shopping centers, marketplaces, governmental structures, healthcare facilities, etc.

8. The site plan, which incorporates numerous housing developments

According to the results, if the price of the house or property is above (60000\$) you can get in the areas of the diamond zone (Shawes - Korean Village - Zeen City)

The price is above (200000\$) you can buy at (Erbil Gate - Firdos City - Zen City - Aram Village 1 & 2 - Italian 1 & 2 - Seberdan - New Erbil - Mass City)

get it .

For more than \$300,000, you can purchase a property in Mountain View, Mass City, Lalav City, New Erbil, or Mass Village.

The area, floor, and location of the property in the city, village, or community, as well as the availability of song services, all affect the price. This zone is the greatest option both now and in the future if you wish to live in a strategic and tranquil area or conduct business.

References:

1. Akaike, H. (1973), “*Information theory and an extension of the maximum likelihood principle.*” In Second International Symposium on Information Theory, ed. B. N. Petrov and F. Csaki, 267–281. Budapest: Akailseoniai–Kiudo.
2. Al-Ansari, N.; AlJawad, S.; Adamo, N.; Sissakian, V.K. & Laue, J. (2018) “*Water Quality within the Tigris and Euphrates Catchments*” Journal of Earth Sciences and Geotechnical Engineering, vol. 8, no. 3, 2018, 95-121 ISSN: 1792-9040 (print version), 1792-9660 (online) Scienpress Ltd.
3. Amemiya, T. (1985), “*Advanced Econometrics*”. Cambridge, MA: Harvard University Press.

4. Badal, M. A. (2018) “*Using Multivariate Time Series Model VAR(P) to forecast Water Supply of Tigris and Euphrates Rivers in Iraq*” vol.22, no. 6, P.P. 220-228.
5. ESCWA annual report 2013: 40 years with the Arab world ESCWA, (2014); E/ESCWA/OES/2014/1
6. Franses, Ph. H. and Dijk, D.V. (2000), “*Nonlinear Time Series Models in Empirical Finance*” Cambridge University Press, ISBN 0 511 01100 8 virtual.
7. Gikungu, S.W., Waititu, A.G. (2015), “*Forecasting Inflation Rate in Kenya Using SARIMA Model*” American Journal of the oretical and Applied Statistics 4,15-18.
8. Hamilton, J. D. (1994), “*Time Series Analysis*”. Princeton: Princeton University Press.
9. Kaya, I. (1998) “*The Euphrates-Tigris basin: An overview and opportunities for cooperation under international law*” Office of Arid Lands Studies, College of Agriculture and Life Sciences, The university of Arizona; No. 44, Fall/Winter 1998; ISSN: 0277-9455 | E-ISSN: 1092-5481, Conflict Resolution and Transboundary Water Resources.
10. Luo, C.S., Zhou, L., Qingfeng, W. (2013), “*Application of SARIMA Model in Cucumber Price Forecast*”, Applied Mechanics and Materials, Vols.373-375, pp.1686-1690.
11. Lutkepohl, H. (1993), “*Introduction to Multiple Time Series Analysis*”. 2nd ed. New York: Springer.
12. Lutkepohl, H. (2005), “*New Introduction to Multiple Time Series Analysis*”. New York: Springer.
13. Michal, S. (2001), “*Time Series Analysis*” log linear publishing, Canada
14. Mira, S.K., Ahmad M.R. (2015), “*Time Series Models for Average monthly Solar radiation in Malaysia*”, Research and Education in Mathematics, International Conference Kuala Lumpur, Malaysia.
15. Nielsen, B. (2001), “*Order determination in general vector autoregressions*”. Working paper, Department of Economics, University of Oxford and Nuffield College.
16. Paulsen, J. (1984), “*Order determination of multivariate autoregressive time series with unit roots*”. Journal of Time

Series Analysis 5: 115–127. 8-Tsay, R. S. (1984), Order selection in nonstationary autoregressive models. Annals of Statistics 12: 1425–1433.

17. <https://www.facebook.com/share/19m2vpjkd4/?mibextid=w wXIfr>)

Appendix

Table 1 –House price change data is front in Diamond Zone between 1-10-2023 to 1-11-2024

Date	Property Type	Area	Price	Logic
10/12/2023	Cadastre Land	250m	\$6,000.000	Shawes
01/01/2024	Two - Storey House	125m	\$6,900.000	Shawes
01/01/2024	One - Storey House	100m	\$12,400.000	Shawes
05/01/2024	One - Storey House	52m	\$13,500.000	Shawes
09/03/2024	One - Storey House	100m	\$14,000.000	Shawes
18/03/2024	House	200m	\$14,200.000	Koryan Village
22/03/2024	One - Storey House	80m	\$14,600.000	Shawes
25/03/2024	One - Storey House	100m	\$14,800.000	Shawes
26/03/2024	Land	125m	\$16,000.000	Shawes
01/04/2024	House	170m	\$16,000.000	Shawes
02/04/2024	House	150m	\$16,500.000	Shawes
06/04/2024	House	100m	\$16,800.000	Shawes
01/05/2024	One - Storey House	100m	\$17,000.000	Shawes
12/03/2024	One - Storey House	100m	\$17,000.000	Shawes
13/05/2024	House	150m	\$17,500.000	Shawes
22/05/2024	House	100m	\$18,000.000	Shawes
30/05/2024	One - Storey House	65m	\$18,500.000	Shawes
02/06/2024	One - Storey House	100m	\$18,900.000	Shawes
08/06/2024	House	200m	\$19,000.000	Shawes
09/06/2024	One - Storey House	100m	\$19,250.000	Shawes

23/06/2024	House	140m	\$20,000.000	Shawes
26/06/2024	One - Storey House	100m	\$21,600.000	Shawes
20/09/2024	Flat	93m	\$21,600.000	Shawes
01/07/2024	Land	100m	\$22,000.000	Shawes
02/07/2024	One - Storey House	200m	\$22,500.000	Shawes
03/07/2024	House	125m	\$23,000.000	Shawes
10/07/2024	One - Storey House	70m	\$23,000.000	Shawes
11/07/2024	House	97m	\$24,000.000	Shawes
11/07/2024	One - Storey House	60m	\$25,300.000	Shawes
16/07/2024	One - Storey House	97m	\$27,000.000	Shawes
17/07/2024	Cadastre Land	250m	\$27,300.000	Shawes
23/07/2024	Land	100m	\$28,000.000	Shawes
20/07/2024	House	200m	\$28,500.000	Koryan Village
24/07/2024	One - Storey House	100m	\$29,000.000	Shawes
24/07/2024	Land	500m	\$29,000.000	Shawes
25/07/2024	One - Storey House	70m	\$30,000.000	Shawes
10/08/2024	Land	120m	\$30,000.000	Shawes
19/08/2024	House	125m	\$30,000.000	Shawes
05/08/2024	House	220m	\$30,000.000	Shawes
31/08/2024	Land	200m	\$30,500.000	Shawes
25/09/2024	One - Storey House	100m	\$31,600.000	Shawes
20/09/2024	Land	250m	\$32,000.000	Shawes
29/09/2024	One - Storey House	70m	\$33,000.000	Shawes
10/09/2024	Land	140m	\$33,300.000	Shawes
01/10/2023	Vella	400m	\$33,500.000	Ferdawx City
15/10/2023	Vella	400m	\$35,500.000	Ferdawx City
15/10/2023	Vella	400m	\$40,000.000	DarwazayHawler
18/10/2023	House	250m	\$40,000.000	Zine City
04/10/2023	House	200m	\$42,500.000	Zine City
15/10/2023	House	250m	\$43,000.000	Zine City
20/12/2023	Vella	300m	\$43,000.000	Ferdawx City
20/12/2023	House	250m	\$48,000.000	DarwazayHawler
17/12/2023	House	200m	\$50,000.000	Zine City
14/12/2023	House	250m	\$50,000.000	Zine City
10/11/2023	Vella	300m	\$50,000.000	Ferdawx City

10/11/2023	House	250m	\$53,500.000	DarwazayHawler
10/11/2023	House	250m	\$58,000.000	DarwazayHawler
01/11/2023	House	250m	\$58,500.000	Zine City
04/11/2023	House	200m	\$60,000.000	Zine City
25/01/2024	Vella	400m	\$60,000.000	Ferdawx City
15/02/2024	Vella	300m	\$60,500.000	Ferdawx City
15/02/2024	House	250m	\$63,700.000	DarwazayHawler
07/02/2024	House	250m	\$64,000.000	Zine City
17/03/2024	Vella	300m	\$64,000.000	Ferdawx City
05/03/2024	Vella	400m	\$67,000.000	Ferdawx City
05/03/2024	House	250m	\$70,000.000	DarwazayHawler
03/03/2024	House	200m	\$70,500.000	Zine City
09/03/2024	House	200m	\$72,300.000	Zine City
10/04/2024	Vella	400m	\$75,000.000	Ferdawx City
10/04/2024	House	250m	\$75,000.000	DarwazayHawler
07/04/2024	House	200m	\$80,000.000	Zine City
28/04/2024	House	200m	\$85,500.000	Zine City
20/05/2024	Vella	300m	\$86,000.000	Ferdawx City
25/05/2024	House	250m	\$87,000.000	DarwazayHawler
25/05/2024	House	250m	\$95,000.000	Zine City
03/06/2024	Vella	400m	\$100,000.000	Ferdawx City
in more than 100.000\$ - 300.000\$				
03/06/2024	House	250m	\$105,000.000	DarwazayHawler
04/06/2024	House	250m	\$105,000.000	Zine City
10/07/2024	Vella	400m	\$110,000.000	Ferdawx City
11/07/2024	House	300m	\$110,000.000	DarwazayHawler
25/07/2024	House	200m	\$125,000.000	Zine City
10/08/2024	Vella	400m	\$125,000.000	Ferdawx City
12/08/2024	House	250m	\$125,000.000	DarwazayHawler
03/08/2024	House	200m	\$129,000.000	Zine City
07/09/2024	Vella	250m	\$135,000.000	Ferdawx City
15/09/2024	House	250m	\$137,500.000	Zine City
26/10/2024	Vella	400m	\$146,500.000	Ferdawx City
26/10/2024	House	250m	\$146,500.000	DarwazayHawler
26/10/2024	House	250m	\$147,000.000	DarwazayHawler
18/10/2024	House	250m	\$148,500.000	Zine City
11/10/2024	House	250m	\$150,000.000	Zine City
10/10/2024	House	250m	\$150,000.000	Zine City
10/10/2023	Vella	200m	\$151,500.000	Etally 2
15/10/2023	Vella	240m	\$155,000.000	Etally 2
24/10/2023	Vella	320m	\$159,000.000	Etally 2
26/03/2024	Flat	85m	\$160,000.000	Var Park
12/10/2023	Vella	274M	\$160,000.000	Aram 2 Village
15/10/2023	House	242m	\$162,500.000	Aram 2 Village

20/10/2024	House	300m	\$164,000.000	Aram 1 Village
21/10/2023	House	340m	\$167,500.000	Aram 1 Village
03/11/2023	House	200m	\$169,000.000	Etally 2
10/11/2023	House	240m	\$180,000.000	Etally 2
11/11/2023	Vella	320m	\$182,500.000	Etally 2
20/11/2023	Vella	390m	\$184,000.000	Etally 2
04/11/2023	Vella	242m	\$185,000.000	Aram 2 Village
15/11/2023	Vella	240m	\$185,000.000	Aram 1 Village
03/12/2023	Vella	200m	\$187,500.000	Etally 2
11/12/2023	House	200m	\$190,000.000	Etally 2
03/12/2023	House	276m	\$191,000.000	Etally 2
06/12/2023	Vella	240m	\$192,500.000	Aram 2 Village
17/12/2023	House	242m	\$194,000.000	Aram 2 Village
08/01/2024	House	200m	\$200,000.000	Etally 2
02/01/2024	House	300m	\$200,000.000	Aram 1 Village
29/01/2024	House	242m	\$200,000.000	Aram 2 Village
14/02/2024	House	200m	\$200,000.000	Etally 2
05/02/2024	House	276m	\$202,000.000	Aram 2 Village
19/02/2024	House	253m	\$203,000.000	Sebardan
12/03/2024	Vella	320m	\$205,000.000	Etally 2
17/03/2024	Vella	200m	\$205,000.000	Etally 2
01/03/2024	Vella	345m	\$208,000.000	Aram 2 Village
03/03/2024	Vella	226m	\$210,000.000	Aram 2 Village
04/04/2024	House	200m	\$210,000.000	Etally 2
15/04/2024	Vella	240m	\$210,000.000	Etally 2
07/04/2024	House	146m	\$210,000.000	ScwerEmpair
12/04/2024	Vella	300m	\$210,000.000	Aram 1 Village
05/05/2024	Vella	200m	\$210,000.000	Etally 2
01/05/2024	Vella	242m	\$210,500.000	Aram 2 Village
27/05/2024	Vella	264m	\$213,000.000	Aram 2 Village
24/06/2024	Vella	200m	\$215,000.000	Etally 2
20/06/2024	Vella	276m	\$215,000.000	Aram 2 Village
01/07/2024	Vella	240m	\$215,000.000	Etally 2
07/07/2024	Vella	200m	\$215,000.000	Etally 2
08/07/2024	Vella	240m	\$215,000.000	Aram 1 Village
08/08/2024	Vella	320m	\$218,000.000	Etally 2
20/08/2024	Vella	200m	\$220,000.000	Etally 2
25/08/2024	Vella	242m	\$220,000.000	Aram 2 Village
22/08/2024	Vella	520m	\$220,000.000	Aram 2 Village
01/09/2024	Vella	200m	\$222,000.000	Etally 2
10/09/2024	Vella	240m	\$222,500.000	Etally 2
11/09/2024	Vella	320m	\$225,000.000	Etally 2
05/09/2024	Vella	550m	\$225,000.000	Aram 2 Village
16/09/2024	Vella	372m	\$225,000.000	Aram 2 Village

15/10/2024	Vella	200m	\$225,000.000	Etally 2
17/10/2024	Vella	240m	\$227,500.000	Aram 2 Village
19/10/2024	Vella	242m	\$227,500.000	Aram 2 Village
20/10/2024	Flat	125m	\$230,000.000	LaLav City
25/10/2024	House	450m	\$230,000.000	Ganjan City
27/10/2024	House	250m	\$230,000.000	Ganjan City
02/11/2023	House	250m	\$230,000.000	Ganjan City
23/02/2024	House	430m	\$230,000.000	Espani Village
24/02/2024	House	430m	\$234,000.000	Espani Village
02/02/2023	House	450m	\$235,000.000	Ganjan City
15/02/2023	House	450m	\$235,000.000	Ganjan City
14/05/2024	House	200m	\$235,000.000	Kurdistan
16/08/2024	House	200m	\$236,000.000	Kurdistan
17/06/2024	Flat	148m	\$240,000.000	Kurdistan
10/05/2023	House	240m	\$240,000.000	Mauntn View
10/07/2023	House	240m	\$240,000.000	Mauntn View
15/10/2023	House	300m	\$240,000.000	Mauntn View
20/10/2023	House	125m	\$240,000.000	Hawlery New
25/10/2023	House	100m	\$240,000.000	Hawlery New
10/01/2023	Vella	240m	\$241,500.000	Mass Village
10/11/2023	Vella	320m	\$243,000.000	Mass Village
11/05/2023	House	240m	\$245,000.000	Mauntn View
11/10/2023	Flat	100m	\$245,000.000	LaLav City
20/11/2023	House	300m	\$245,000.000	Mauntn View
13/11/2023	Vella	240m	\$250,000.000	Mass Village
12/03/2023	House	240m	\$250,000.000	Mauntn View
23/12/2023	House	300m	\$250,000.000	Mauntn View
22/12/2023	Vella	320m	\$255,000.000	Mass Village
01/10/2024	House	300m	\$255,000.000	Mauntn View
15/1/2024	Flat	100m	\$260,000.000	Mine Lav
21/1/2024	Flat	104m	\$260,000.000	LaLav City
11/1/2024	Vella	240m	\$260,000.000	Mass Village
15/1/2024	Vella	320m	\$263,500.000	Mass Village
25/2/2024	House	300m	\$267,500.000	Mauntn View
27/2/2024	House	240m	\$270,000.000	Mauntn View
29/2/2024	House	240m	\$270,000.000	Mauntn View
30/2/2024	House	150m	\$270,000.000	Hawlery New
3/10/2024	House	125m	\$271,000.000	Hawlery New
3/11/2023	House	240m	\$273,000.000	Mauntn View
19/3/2024	House	300m	\$275,000.000	Mauntn View
27/3/2024	House	240m	\$277,000.000	Mauntn View
3/10/2024	Vella	320m	\$277,500.000	Mass Village
20/4/2024	House	240m	\$280,000.000	Mauntn View
25/4/2024	House	300m	\$280,000.000	Mauntn View

26/4/2024	House	240m	\$285,000.000	Mauntn View
29/4/2024	House	125m	\$285,000.000	Mauntn View
27/4/2024	Vella	240m	\$290,000.000	Mauntn View
5/01/2024	House	240m	\$290,000.000	Mauntn View
5/02/2024	House	240m	\$290,000.000	Mauntn View
20/5/2024	House	240m	\$295,000.000	Mauntn View
15/5/2024	House	300m	\$300,000.000	Mauntn View
more than 300.000 \$				
5/11/2023	Flat	155m	\$300,000.000	LaLav City
5/11/2023	Vella	240m	\$300,000.000	Mass Village
5/12/2023	Vella	320m	\$300,000.000	Mass Village
6/10/2024	House	100m	\$300,000.000	Hawlery New
15/6/2024	House	150m	\$305,000.000	Hawlery New
20/6/2024	House	240m	\$310,000.000	Mauntn View
21/6/2024	House	240m	\$310,000.000	Mauntn View
7/10/2024	House	240m	\$313,500.000	Mauntn View
15/7/2024	House	125m	\$320,000.000	Hawlery New
21/7/2024	House	300m	\$320,000.000	Mauntn View
20/7/2024	Vella	240m	\$320,000.000	Mass Village
8/11/2023	House	240m	\$330,000.000	Mauntn View
15/8/2024	House	300m	\$330,000.000	Mauntn View
20/8/2024	House	300m	\$340,000.000	Mauntn View
25/8/2024	House	100m	\$340,000.000	Hawlery New
21/8/2024	Vella	240m	\$350,000.000	Mass Village
23/8/2024	Vella	300m	\$370,000.000	Mass Village
17/9/2024	House	240m	\$430,000.000	Mauntn View
20/9/2024	Flat	100m	\$435,000.000	LaLav City
25/9/2024	House	150m	\$450,000.000	Hawlery New
24/9/2024	Vella	240m	\$450,000.000	Mass Village
15/10/2024	House	100m	\$480,000.000	Hawlery New
25/10/2024	House	150m	\$580,000.000	Hawlery New
26/10/2024	House	260m	\$815,000.000	Hawlery New
27/10/2024	House	240m	\$825,000.000	Mauntn View
22/10/2024	Vella	320m	\$850,000.000	Mass Village

Prediction by Normalized Radial Basis Function Artificial Neural Network and model Selection for the Risk Factors Most Affected on Kidney Transplantation

Shaymaa Mohammed Shakir Ahmed

shaymaa.shakirahmed@su.edu.krd

Department of statistics and informatics, College of Administration and Economics, University of Salahaddin,, Kurdistan Reign, Iraq.

Assist.Prof .Dr Nazeera Sedeeq Barznji

nazeera.kareem@su.edu.krd

Department of statistics and informatics, College of Administration and Economics, University of Salahaddin, Kurdistan Reign, Iraq.

Abstract :

Kidney transplants are done to help people with chronic kidney disease or end-stage kidney failure. When the kidneys cannot filter waste properly, they need either dialysis or a kidney transplant. However, Artificial Intelligence Analysis (AI) Artificial Neural Network (ANN) refers to computer systems and advanced technologies one of it branch is a Radial basis function (RBF) method it is a modern way to approximate multivariate functions in medicine and to perform predictive tasks for this purpose A set sample was prepared based 150 patients from Hospital in Kurdistan region with kidney failed state; kidney cases are (transplant or preparing). To analyze the data of this paper, the researcher included six important risk factors that caused the greatest impact on the access of chronic kidney patients to kidney transplantation. The results show that prediction by the Normalized Importance Analysis emphasizes [Congenital kidney diseases has the most effect with100% followed it polycystic kidney disease 84.4% then, Disease of the immunity [Systemic lupus erythematosus) 80.4% then Renal artery stenosis74.4%, Diabetes71.7% and finally Chronic, uncontrolled high blood pressure(60.4%),in other hand second

way for prediction by classification analysis emphasizes that it predict the both kinds of patients affected by the risk factors. with High overall accuracy (90.4% and 85.7%) mean the model in training and testing is reliable and clinically valuable, also analysis by the ROC curve area with the ratio(0.835), indicating that the model possesses a robust capacity to accurately classify ,The test of model fit statistics represent significant performance improvements, and results show there were significant improvements in F-score. Our findings demonstrate that RBF is an effective strategy for improving the predictive power and clinical relevance of ML models used for kidney disease classification.

IBM SPSS 27with version 4.1of R programming Language were appliedfor analyze data.

Keywords: Artificial Intelligence (AI) , Radial Basis Function, Gaussian Kernel, Classification, Normalized Importance and Kidney Transplantation

التنبؤ بواسطة شبكة الأعصاب الاصطناعية باستخدام دالة الأساس الشعاعي المنظم واختيار النموذج لعوامل الخطر الأكثر تأثيراً على زراعة الكلى

شيماء محمد شاكر أحمد

shaymaa.shakirahmed@su.edu.krd

قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين، إقليم كردستان، العراق.

م.ب.د. نظيرة صديق برزنجي

nazeera.kareem@su.edu.krd

قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين، إقليم كردستان، العراق.

الملخص

تُجرى زراعة الكلى لمساعدة الأشخاص الذين يعانون من مرض الكلى المزمن أو الفشل الكلوي في مراحله النهائية. عندما لا تستطيع الكلى تصفية النفايات بشكل صحيح، فإنها تحتاج إما إلى غسيل الكلى أو زراعة كلى. ومع ذلك، تشير شبكة الأعصاب الاصطناعية لتحليل الذكاء الاصطناعي (AI) إلى أنظمة الكمبيوتر والتقنيات المتقدمة أحد فروعها هو طريقة دالة الأساس الشعاعي (RBF) ، وهي

طريقة حديثة لتقريب الدوال متعددة المتغيرات في الطب ولأداء المهام التنبؤية لهذا الغرض. تم جمع عينة مكونة من 150 مريضاً من مستشفى في منطقة كردستان بمصابين بحالات الفشل الكلوي؛ حالات الفشل الكلوي هي اما (زراعة أو تحضير لزراعة الكلية). لتحليل بيانات هذه الورقة، شمل الباحث ستة عوامل خطر مهمة تسببت في أكبر تأثير على وصول مرضى الكلى المزمنين إلى زراعة الكلى. تُظهر النتائج أن التنبؤ من خلال تحليل الأهمية المنظم يبرز [الأمراض الكلوية الخلقية لها التأثير الأكبر بنسبة 100% يليها مرض الكلى المتعدد الكيسات 84.4%، ثم مرض المناعة (الذئبة الحمراء) 80.4% ثم تضيق الشريان الكلوي 74.4%، السكري 71.7% وأخيراً ارتفاع ضغط الدم المزمن وغير المنضبط (60.4%)، من ناحية أخرى، تؤكد الطريقة الثانية للتنبؤ من خلال تحليل التصنيف أنها تتنبأ بكلا النوعين من المرضى المتأثرين بالعوامل المسببة. مع الدقة العامة العالية (90.4% و 85.7%) تعني أن النموذج في التدريب والاختبار موثوق به وقيم سريرياً، كما أن التحليل بواسطة منطقة منحنى ROC بنسبة (0.835) يشير إلى أن النموذج يمتلك قدرة قوية على التصنيف بدقة. تمثل إحصائيات اختبار ملاءمة النموذج تحسينات كبيرة في الأداء، وتظهر النتائج أن هناك تحسينات كبيرة في درجة F. تظهر نتائجنا أن RBF هي استراتيجية فعالة لتحسين القدرة التنبؤية والأهمية السريرية لنماذج التعلم الآلي المستخدمة في تصنيف أمراض الكلى. تم استخدام IBM SPSS 27 مع الإصدار 4.1 من لغة البرمجة R لتحليل البيانات

1.Introduction

The term Artificial Intelligence (AI) presents prospective methods to improve the identification and administration of several aspects within the different sectors [12]. Machine learning (ML), a subset of artificial intelligence (AI), is defined as the capacity of computers to learn autonomously without explicit programming [22] capable of analyzing extensive datasets and identifying intricate patterns suggestive of disease. Machine learning models have been extensively utilized for predicting disease, with performance evaluated using diverse measures.

The Radial Basis Function (RBF) networks are a type of Artificial Neural Network (ANN) that uses Gaussian functions was proposed by [5], the radial basis function, commonly is a widely utilized kernel function to employed across various kernel based learning algorithms. Notably, it finds frequent application in the classification tasks performed by support vector machines [7]. Radial basis functions (RBF) were first introduced by Powell to solve the real multivariate interpolation

problem it procedure produces a predictive model for one or more dependent (target) variables based on values of predictor variables. Its structures consist of three layers: input layer, hidden layer and output layer. The steps of RBF contain the training process of a RBF Neural Network involves determining the parameters required for the hidden layer and output layer, this process can be divided into three key steps: selecting centers, determining spreads, and training the output weights.

The main statistical tasks of RBF are prediction and approximation, for this purpose the researcher used RBF for this paper to predict the common risk factors caused to kidney transplant at ,which , kidney transplant is a well-established intervention for addressing end-stage chronic renal failure and is associated with enhanced patient survival rates and improved quality of life. The pioneering attempts at human-to-human kidney transplantation were conducted by [Yuri Voronoy, Jean Hamburger, and several other researchers during the 1930s to 1940s]; however, the majority of these endeavors were met with failure. This medical procedure entails a surgical operation whereby the nonfunctional kidney is supplanted by a viable kidney sourced from either a living or deceased donor[15]. In this paper we investigate the potential of applying the radial basis function (RBF) neural network architecture for the prediction by Normalized independent variables Importance(regression algorithms) , Classification and [ROC – Curve] Both classification and regression algorithms can be employed for prediction in machine analyses focusing on the six risk factors caused Kidney Transplant to emphasize significate these factors .the researcher by using those three methods of RBF analysis the results refer toIntegrating these normalized risk indicators using RBF classification improves prediction accuracy and offers critical insights for evaluating kidney transplant risks, hence facilitating clinical decision-making and targeted interventions.[Congenital kidney diseases has most effect follows it polycystic kidney disease then , Disease of the immunity then Renal artery stenosis74.4%, Diabetes71.7% and finally Chronic, uncontrolled high blood

pressure(60.4%)In a sequence,Classification Analysis emphasize that Kidney transplanted cases show overall a currency and finally ROC –Curve analysis the refers of(0. 835) of all area under the ROC –curve indicates that all risk factors have significantive effect on dependent variable.

2.Objective:

This study aims to demonstrate the benefits of RBF, as a tool for Prediction by Classification, and Normalized independent variables Importance focusing on the risk Factors caused Kidney Transplant,

3. Literature Review

3.1-Murat Kirişci (2019)[17] conducted an investigation aimed at juxtaposing the efficacy of ANN and logistic regression analyses to elucidate the determinants influencing birth weight. This inquiry encompassed a sample size of 223 neonates. The analyses derived from ANN and logistic regression methodologies, based on the gathered data, were meticulously evaluated. Both analytical frameworks possess the capability to serve as complementary tools to assist in decision-making processes. Moreover, both models hold substantial potential to aid clinicians in understanding the risk factors associated with birth weight and in estimating those risks.

3.2-Dong Chen (2021)[8] undertook a research initiative employing an innovative system for swarm optimization neural network (RBFNN-PSO), which was applied to assess the effectiveness of physical education teaching methodologies. To validate the evaluative performance of the RBFNN-PSO system, the traditional RBF neural network served as a control for comparative purposes during the training phase. The outcomes revealed that the RBFNN-PSO system achieved convergence more rapidly than the conventional RBF neural network system, with a concomitantly lower training error.

3.3-Md. Imam Hossain and Mehadi Hasan Maruf (2023)[16] have made concerted efforts to enhance the diagnosis and treatment of cardiovascular diseases, given their profound impact on societal health. With the increasing synergy between technological advancements and medical diagnostics, the utilization of data mining and the preservation of medical knowledge can significantly bolster patient management outcomes. Consequently, it is imperative to investigate the interplay of risk factors within patients' medical histories and to comprehend their respective roles in the prognostication of heart disease. This study aims to meticulously analyze various components within patient data to facilitate precise heart disease predictions. Artificial intelligence methodologies were employed and compared across two distinct heart disease datasets (comprising all features and selected features). The random plantation approach utilizing selected features attained an impressive accuracy rate of 90%, surpassing the performance of models that incorporated the entirety of input features and other artificial intelligence techniques. The proposed methodology could serve as an auxiliary framework for the early prediction of heart disease.

3.4: Zhao, A., Xing, S., Wang, X., & Sun, J. Q. (2024)[29]. presents a method for computing optimal control strategies for multi-degree-of-freedom dynamic systems using Radial Basis Function (RBF) neural networks with Gaussian activation functions. The Hamilton–Jacobi–Bellman (HJB) equation is used to define the control problem, with RBF networks approximating the value function via a policy iteration algorithm. A stabilizing initial control ensures convergence, particularly in open-loop unstable systems. A modified truncation technique reduces the system's dimensionality, making RBF application more efficient in high-dimensional spaces. Additionally, transfer learning is employed to enhance performance using experimental data. Simulations and experiments demonstrate the method's effectiveness in

achieving accurate optimal control, trajectory tracking, and system stabilization

3.5 Xin Zhang .Weihua Luo (2024)[27]In this study, aiming at the problem that the spread coverage ability determined by the trial method is limited and easy to fall into local optimum (Zhao et al., 2021), a RBF network diffusion velocity function optimization method based on the PSO algorithm is proposed to realize the large-scale and efficient search of RBF network diffusion function. On this basis, in view of the fact that it is difficult to guarantee the prediction accuracy of the data-driven model based on the displacement increment sequence constructed by a single RBF intelligent algorithm when the displacement increment mutation point exists (Li and Liu, 2005), the PSO-RBF intelligent coupling model based on the graying layer pre-processing is established to improve the prediction accuracy of the model from the perspective of improving the prediction accuracy of displacement increment mutation point.

3.6 Nguyen, V. T., Nguyen, Q. C., Van, M., Su, S. F., Garg, H., Duong, D. N., & Tan, P. X. (2025)[20]. This paper proposes a novel adaptive nonlinear PID controller based on Radial Basis Function Neural Networks (NPID-RBFNN) for controlling ballbots under external forces. Combining nonlinear PID control, RBFNN, and Balancing Composite Motion Optimization (BCMO), the controller offers lightweight computation, reduced chattering, and strong robustness against model uncertainties and disturbances. Unlike conventional PID or sliding mode control (SMC), the NPID-RBFNN ensures smoother and more stable motion. BCMO optimizes control gains, while RBFNN continuously learns during operation. Stability is verified via the Lyapunov method, and simulations confirm its effectiveness in maintaining balance under external forces.

4. Methodology (Theoretical Aspect)

4.1 Artificial Intelligence (AI)

Artificial intelligence (AI) refers to computer systems capable of performing complex tasks that historically only a human could do, such as reasoning, making decisions, or solving problems. The criterion of Artificial Intelligence (AI) is fundamentally anchored in cognitive tasks, an idea of origins spanning several centuries and is revolved in various mythological narratives, The emergence of contemporary AI, as it is recognized today, while the term "Artificial Intelligence" was first articulated by John McCarthy in 1956 The cardinal aim of AI is this technology permits intelligent systems to realized, interpret, and assimilate information from data, accelerating the scope to make informed decisions established on cognizance derived from that data.[12]

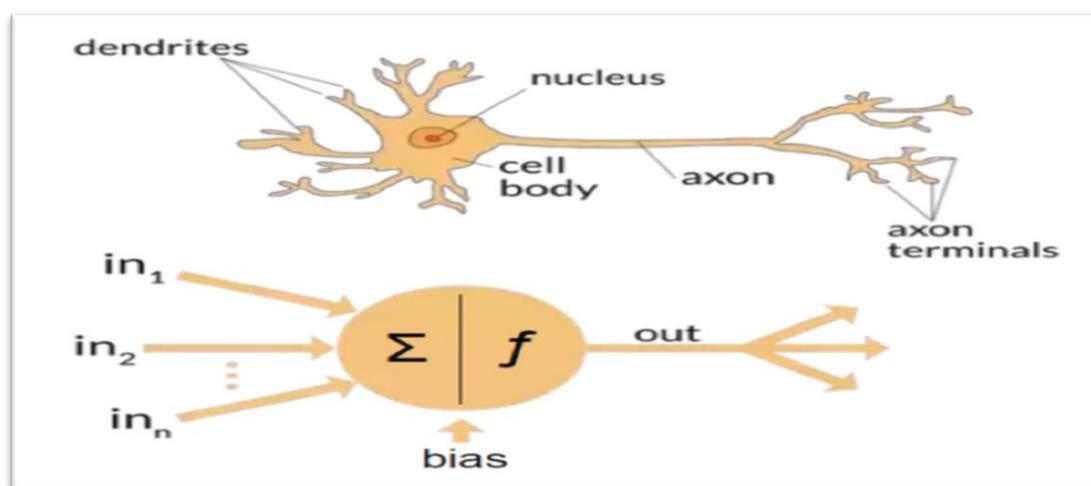


Figure (1): A biological and an artificial neuron

4.2 Machine Learning

Within the domain of machine learning, referred to an artificial neural network ANN, and abbreviated as constitutes a model that draws inspiration from the architecture and operational dynamics of biological neural networks found in the brains of animals. An (ANN) artificial comprises interconnected units or nodes termed artificial neurons, which serve as the brain's neurons. Each artificial neuron receives input signals from affiliated neurons, subsequently processes these inputs, and transmits a signal to additional connected neurons. The "signal"

is represented as a real number, and the fabrication generated by each neuron is determined through a non-linear function applied to the aggregate of its inputs, referred to as the activation function. The efficacy of the signal at each connection is governed by a weight that is subject to modification throughout the learning process. The first use of artificial neural networks (ANNs) was made by psychologist Frank Rosenblatt, who also created the perceptron. However, some of the computational applications of ANNs are related to previous mathematical findings. Research on ANNs was performed in the 1970s and 1980s. The creation of Alex Net, deep neural network with several layers, occurred in the 2010s. as it shown in Figure (1)[22]..

4.3: Model Types for Artificial Neural Nets:

Figure (2) shows Types for Artificial Neural as follows:

1. The Perceptron Single Layer.
2. Feed Forward Neural Network with Radial Basis Functional Neural Network.
3. A multilayer perceptron.
- 4- Neural Network with Recurrence.
5. Boltzmann apparatus
6. The Hopfield Network
7. Long Short-Term Memory, or LSTM

4.4 Definition of the Radial Basis Function (RBF)

In the field of mathematical modeling, a **radial basis function network** is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction, classification, and system control. They were first formulated in a 1988 paper by Broomhead and Lowe, both researchers at the Royal Signals and Radar is one of the simplest form of ANN consisting of exactly three layers, namely, input, hidden, and output layer. The restriction of only three layers

makes it simplest and somehow efficient ANN architecture. The idea of RBFNN has been derived from function approximation. An RBF network positions one or more RBF neurons in the space described by the predictor variables. This space has as many dimensions as there are predictor variables. The Euclidean distance is computed from the point being evaluated to the center of each neuron. The RBF is so named because the radius distance is the argument to the function. Output of RBFNN depends on the distance of the input from a given stored vector. The RBF neural network is described by [4].

4.5. Basic functions in Radial Basis Function (RBF)

The RBF encompasses various basis interpolation functions as outlined as table (1) in below [18].

Table (1): Basic functions in Radial Basis Function (RBF) [5]

Basic function abbreviation	Formulas $\phi(r) = \phi(\ x - \mu\)/\sigma$
Gaussian(GA)	e^{-cr^2}
Generalized Multiquadratic (GMQ)	$(C^2 + r^2)^\beta, c > 0, 1 > \beta > 0$
Inverse Multiquadratic (IMQ)	$1/\sqrt{C^2 + r^2}$
Inverse Quadratics(IQ)	$(C^2 + r^2)^{-1}$
Multi Quadratics(MQ)	$\sqrt{C^2 + r^2}$
Hyperbolic secant (sech)	$Sech(cr)$
Cubic function	r^3
Linear function	r
Monomial(MN)	r^{2k-1}
Thin-plate spline	$r^2 \log(r)$
Poly harmonic spline	$\begin{cases} \phi(r) = r^k, k = 1, 3, 5, \dots \\ \phi(r) = r^k \ln(r), k = 2, 4, 6, \dots \end{cases}$

4.5.1 Kernel: The kernel function serves the purpose of converting n-dimensional inputs into m-dimensional inputs, where m significantly exceeds n, thereby facilitating the

efficient computation of dot products in a higher-dimensional space. The fundamental principle underlying the utilization of a kernel is that a linear classifier or regression line in elevated dimensions is transformed into a non-linear classifier or regression curve within lower dimensions[7].

4.5.2 Kernel Radial Basis Function

Within the domain of machine learning, the radial basis function kernel, commonly referred to as the RBF kernel, is a widely utilized kernel function employed across various kernelled learning algorithms. Notably, it finds frequent application in the classification tasks performed by support vector machines[7].

The RBF kernel on two samples $X \in \mathbb{R}^k$ and X' , represented as feature vectors in some input space, is defined as

$$K(x, x') = \exp\left(-\frac{(\|X - X'\|)^2}{2\sigma^2}\right) \dots \dots \dots (1)$$

where, $(\|X - X'\|)^2$ the squared Euclidean distance between the two data points is denoted by the expression above. The parameter represented by σ is commonly referred to as the bandwidth or width of the kernel, which regulates the smoothness of the decision boundary.

Upon the elaboration of the aforementioned exponential expression, it shall extend to an infinite degree of X and X' consistent with the expansion of the value of the Radial Basis Function (RBF) kernel diminishes with increasing distance and spans from zero (in the limit of infinite distance) to one (when $x = x'$), it readily lends itself to interpretation as a measure of similarity The feature space associated with the kernel possesses an infinite number of dimensions; for $\sigma = 1$, its expansion via the multinomial theorem is[10]:

$$\begin{aligned} \exp\left(-\frac{1}{2}\|X - X'\|^2\right) &= \exp\left(\frac{2}{2}X^T X' - \frac{1}{2}\|X\|^2 - \frac{1}{2}\|X'\|^2\right) \\ &= \exp(X^T X') \exp\left(-\frac{1}{2}\|X\|^2\right) \exp\left(-\frac{1}{2}\|X'\|^2\right) \\ &= \sum_{j=0}^{\infty} \frac{(X^T X')^j}{j!} \exp\left(-\frac{1}{2}\|X\|^2\right) \exp\left(-\frac{1}{2}\|X'\|^2\right) \end{aligned}$$

$$= \sum_{j=0}^{\infty} \sum_{n_1+n_2+\dots+n_k=j} \exp\left(-\frac{1}{2}\|X\|^2\right) \frac{x_1^{n_1} \dots x_k^{n_k}}{\sqrt{n_1! \dots n_k!}} \exp\left(-\frac{1}{2}\|X'\|^2\right) \frac{x'_1{}^{n_1} \dots x'_k{}^{n_k}}{\sqrt{n_1! \dots n_k!}}$$

$$= \langle \phi(X) \cdot \phi(X') \rangle \dots (2)$$

$$\phi(X) = \exp\left(-\frac{1}{2}\|X\|^2\right) \left(a_{\ell_0}^{(0)} \cdot a_1^{(1)} \cdot \dots \cdot a_{\ell_1}^{(1)} \cdot \dots \cdot a_1^{(j)} \cdot \dots \cdot a_{\ell_j}^{(j)}\right)$$

Where

$$\ell_j = \binom{k+j-1}{j}, a_{\ell_j}^{(j)} = \frac{x_1^{n_1} \dots x_k^{n_k}}{\sqrt{n_1! \dots n_k!}} \mid n_1 + n_2 + \dots + n_k = j \wedge 1 \leq \ell \leq \ell_j$$

4.6 Gaussian kernel in RBF

4.6.1 Gaussian Kernels

The Gaussian kernel is widely used in probabilistic models such as Gaussian Processes [21], where it plays a key role in defining the covariance structure of the distribution. also known as the squared exponential kernel – SE kernel – or radial basis function (RBF) is defined by

$$K(X, X') = \left(\exp - \frac{1}{2} (x - x')^T \Sigma^{-1} (x - x') \right)$$

Σ the covariance of each feature across observations, is a p – dimensional matrix. When Σ is a diagonal matrix, this kernel can be written as

$$K(X, X') = \exp\left(-\frac{1}{2} \sum_{j=1}^p \frac{1}{\sigma_j^2} (x_j - x'_j)^2\right)$$

σ_j can be interpreted as defining the characteristic length scale of feature j . Furthermore, if Σ is spherical, i.e., $\sigma_i = \sigma, \forall j$,

$$K(X, X') = \exp\left(-\frac{\|X-X'\|^2}{2\sigma^2}\right) \dots (3)$$

For this kernel, the dimension of the feature space defined by $\phi(\cdot)$ is $D = \infty$. Our methods will allow us to avoid explicit computation of $\phi(\cdot)$. We can easily compute the $n \times n$ Gram matrix using this relative of the Mahalanobis Distance, even though we have implicitly projected our objects to an infinite dimensional feature space.

A Gaussian handle with RBF kernel is a effective modeling instrument since it

has expansive bolster and all-inclusive approximates. RBF kernels are the most generalized frame of kernelization and is one of the most broadly utilized parts due to its closeness to the Gaussian distribution. A kernel (covariance function) depicts the covariance of the Gaussian handle irregular factors. Together with the mean function the kernel totally characterizes a Gaussian process.

The Gaussian kernel is a normal choice for part functions

$$(\bar{X}|\lambda) = \sum_{l=1}^N a_l b_l(\bar{x}), \quad \text{where } \sum_{l=1}^N a_l = 1$$

$$b_l(\bar{X}) = \frac{1}{(2\pi)^{D/2}(\sigma)^{1/2}} \exp[-1/2 (\bar{x} - C)^T \Sigma_{l=1}^{-1}(\bar{x} - C)] \dots (4)$$

A mean vector and a covariance matrix (c and σ) together provide the most comprehensive representation of a Gaussian kernel.. The inverse of the covariance matrix serves to elucidate the interrelations among various features, endowing each kernel with an n-dimensional ellipsoidal geometry. This approach is typically more adaptable than employing the straightforward distance to the kernel centroid, which presupposes rigid independence among the variables.[1].

$$\phi_i(\bar{x}) = \frac{1}{\sqrt{\left(1 + \frac{\|\bar{x} - c_j\|}{\sigma_j^2}\right)}}$$

$$\phi_i(\bar{x}) = \frac{1}{\left(1 + \frac{\|\bar{x} - c_j\|}{\sigma_j^2}\right)^{1/2}} \dots (5)$$

Their actuation for designs distant from the centroid of the RBF which is more noteworthy than the actuation of the standard frame for these designs. In expansion, Conditions they displayed q-Gaussian RBFs as an elective to Gaussian RBF. The q-Gaussian RBF for the ith unit is characterized as:

$$\phi(\vec{x}) = \begin{cases} 1 - (1 - q) \left(\frac{\|\vec{x} - c_j\|}{\sigma_j^2} \right)^{1/1-q} \\ \text{if } 1 - (1 - q) \left(\frac{\|\vec{x} - c_j\|}{\sigma_j^2} \right) \geq 0 \\ 0 \quad \text{otherwise} \end{cases} \dots(6)$$

Where q is a genuine esteemed parameter. The q -Gaussian RBF permits diverse RBFs to be spoken to by upgrading the modern parameter q [13].

4.6.2 Supervised Learning for Gaussian RBF

For the Gaussian RBF organize, the RBF at each nodes can be relegated a different width σ_i . The RBFs can be encourage generalized to permit for self-assertive covariance networks Σ_i

$$\phi(\vec{x}) = e^{-\frac{1}{2}(\vec{x} - \vec{c}_i)^T \Sigma_i^{-1} (\vec{x} - \vec{c}_i)} \dots (7)$$

where $\Sigma_i \in R^{J_1 \times J_1}$ represents a positive definite, symmetric covariance matrix. When Σ_i^{-1} is expressed in its general form, the configuration and orientation of the axes of the hyper ellipsoid are arbitrary within the feature space. In the case that

$$\Sigma_i^{-1} = \text{diag}\left(\frac{1}{\sigma_{i,1}^2} / \dots / \frac{1}{\sigma_{i,J_1}^2}\right)$$

is a diagonal matrix characterized by non-constant diagonal entries, it is entirely defined by a vector $\vec{\sigma}_i \in R^{J_1}$, and each ϕ_i constitutes a hyperellipsoid whose axes align with the axes of the feature space, . For the J_1 -dimensional input space, each Radial Basis Function (RBF) employing a diagonal Σ_i^{-1} possesses a total of $J_1(J_1 + 3)/2$ independent adjustable parameters, whereas each RBF utilizing a uniform σ across all directions and each RBF employing a diagonal Σ_i^{-1} contains merely $(J_1 + 1)$ and $2J_1$ independent parameters, respectively. A trade-off exists between the deployment of a compact network with a multitude of adjustable parameters and the utilization of a larger network with a reduced number of adjustable parameters. When employing the RBF with a uniform σ across all directions, the gradients are derived as

$$\frac{\partial E}{\partial \vec{c}_m} = \frac{2}{N} \sum_{n=1}^N \phi_m(\vec{x}_n) \frac{\vec{c}_m - \vec{x}_n}{\sigma_m^2} \sum_{i=1}^{J_3} e_{n,i} \omega_{i,m}$$

$$\frac{\partial E}{\partial \sigma_m} = -\frac{2}{N} \sum_{n=1}^N \phi_m(\vec{x}_n) \frac{\|\vec{x}_n - \vec{c}_m\|}{\sigma_m^2} \sum_{i=1}^{J_3} e_{n,i} \omega_{i,m} \dots (8)$$

Similarly ,for the RBF using diagonal Σ_i^{-1} ,the gradients are given by

$$\frac{\partial E}{\partial c_{m,j}} = -\frac{2}{N} \sum_{n=1}^N \phi_m(\vec{x}_n) \frac{x_{n,j} - c_{m,j}}{\sigma_{m,j}^2} \sum_{i=1}^{J_3} e_{n,i} \omega_{i,m}$$

$$\frac{\partial E}{\partial \sigma_{m,j}} = -\frac{2}{N} \sum_{n=1}^N \phi_m(\vec{x}_n) \frac{(x_{n,j} - c_{m,j})^2}{\sigma_{m,j}^3} \sum_{i=1}^{J_3} e_{n,i} \omega_{i,m} \dots (9)$$

Adaptations for \vec{c}_1 and Σ_i occur along the directions of negative gradients. The \mathbf{W} parameters. To avert the imposition of unreasonable radii, the updating algorithms may also be derived by incorporating a constraint term into the Mean Squared Error (MSE) that penalizes diminutive radii.[27]

$$E_c = \sum_i 1/\sigma_i \quad \text{or} \quad E_c = \sum_{i,j} 1/\sigma_{i,j} \dots (10)$$

4.7 Training RBF Neural Networks

The training of a neural network entails the identification of appropriate values for weights and biases. In the majority of instances, training is executed via a methodology commonly referred to as the train-test paradigm. As delineated by Dhubkarya (2010), Radial Basis Function (RBF) networks are predominantly employed in supervised contexts. The training dataset comprises specific conditions where the corresponding network outputs are predetermined. The works .The training procedure is utilized to as a certain the values of weights, centers, and widths. The optimization algorithm employed for minimizing a suitable error function (e) is applied to compute the weights, as referenced by Hamadneh (2012). RBF networks are generally trained utilizing pairs of input and target values $x(t), y(t). t = 1. \dots T$ through a two-stage algorithm.

RBF networks are typically trained from pairs of input and target values $X(t). y(t). 1.2. \dots T$ by algorithm of simply fits a

linear model with coefficients w_i to the hidden layer's outputs with respect to some objective function. A common objective function, at least for regression/function estimation, is the least squares function:

$$K(w) \stackrel{\text{def}}{=} \sum_{t=1}^T K_t(w) \dots (11)$$

Where

$$K_t(w) \stackrel{\text{def}}{=} [y(t) - \varphi(X(t).w)]^2 \dots (12)$$

There are occasions in which multiple objectives, such as smoothness as well as accuracy, must be optimized. In that case it is useful to optimize a regularized objective function such as

$$H(w) \stackrel{\text{def}}{=} K(w) + \lambda S(w) \stackrel{\text{def}}{=} \sum_{t=1}^T H_t(w) \dots (13)$$

Where $S(w) \stackrel{\text{def}}{=} \sum_{t=1}^T S_t(w)$ And $H(w) \stackrel{\text{def}}{=} K(w) + \lambda S(w)$ $S(w) \stackrel{\text{def}}{=} \sum_{t=1}^T S_t(w)$

where optimization of S maximizes smoothness and λ is known as a regularization parameter [23].

4.8 Testing

Upon the construction of a machine learning model (utilizing the training dataset), it is imperative to obtain unseen data to evaluate the efficacy of the model. This dataset is referred to as the testing data, that we find it in practical aspect which can be employed to assess the performance and advancement of the algorithms' training while allowing for adjustments or optimizations for enhanced outcomes. The objective of testing is to juxtapose the outputs generated by the neural network against the targets contained within an independent set (the testing instances). It is noteworthy that the testing methodologies are contingent upon the type of project (either approximation or classification).

4.9 Architecture:

The architecture of a Neural Network - RBF delineates the specific configuration of layers and nodes within the network as.

1- layers -2 weights 3- Bias 4-Activation Function

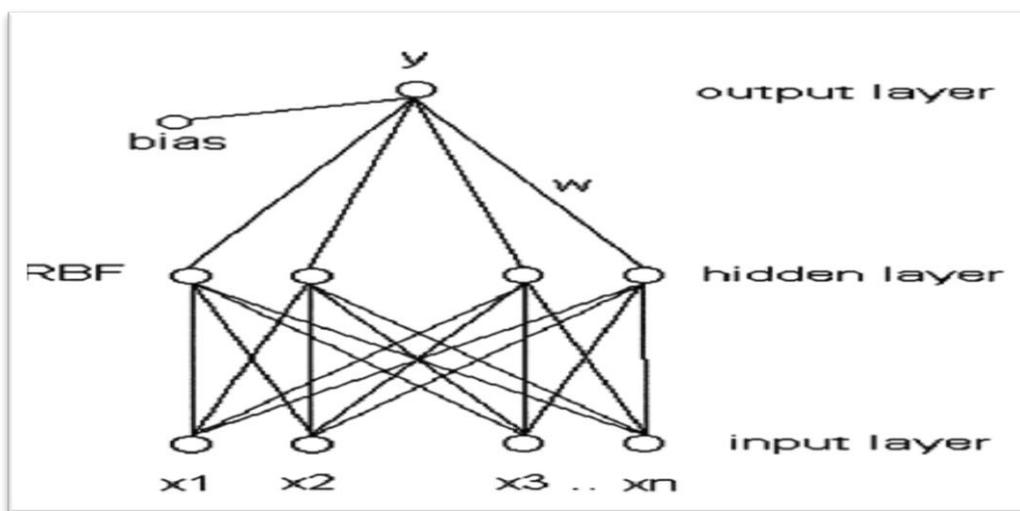
RBFN can be presented as a three-layer feed for-ward structure) see Fig. 2) The input layer serves only as input distributor to the hidden layer. Each node in the hidden layer is a radial function,

its dimensionality being the same as the dimensionality of the input data. The output from the j^{th} function for the input object x_i has the following form [23].

$$\phi_{ji} = \phi_j(x_i) = \phi(\|x_i - c_j\|) \dots (14)$$

where $\|\dots\|$ is a distance measure and c_j describes the center of the j^{th} radial basis function and $j=1, \dots, K$.

At this point the notation of the centers of RBFs has been changed to stress that they are no longer constrained to be defined by the input data vectors



Figure(2):Architecture of radial basis function network.

The output is calculated by a linear combination (i.e. a weighted sum) of the radial basis functions plus the bias, according to:

$$y(X_i) = [\sum_{j=1:k} w_j \phi(\|x_i - c_j\|) + w_0] \dots (15)$$

in matrix notation: $y = \phi w \dots (16)$

where

$$\phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{k1} & 1 \\ \phi_{12} & \phi_{22} & \dots & \phi_{k2} & 1 \\ \vdots & & & & \\ \phi_{1m} & \phi_{2m} & \dots & \phi_{km} & 1 \end{bmatrix} = [\phi_1 \quad \phi_2 \quad \dots \quad \phi_k \quad 1]$$

$$w = [w_1 \quad w_2 \quad \dots \quad w_k \quad w_0]^T$$

For multiple y the Eq. 15. has the following form:

$$y_b(X_i) = \sum_{j=1:k} w_{bj} \phi(\|x_i - c_j\|) + w_{b0} \dots (17)$$

In matrix notation:

$$Y = \Phi W \dots (18)$$

$$\text{Where } W = \begin{bmatrix} W_{11} & W_{12} \dots\dots & W_{b1} \\ W_{12} & W_{22} \dots\dots & W_{b2} \\ \dots & \dots & \dots \\ W_{1k} & W_{2k} & W_{bk} \\ W_{10} & W_{20} & W_{b0} \end{bmatrix}$$

In the RBFN, all basis functions have the same width parameter σ . In the more general case the width parameters can be tuned for each basis function individually. In the so-called hyper radial basis functions each argument is scaled by the matrix Q:

$$\| \|x_i - c_j^i\| \|^2 = [Q_{ij}^j(x_i - c_j)]^T [Q_{ij}^j(x_i - c_j)^j] \dots (19)$$

$$\text{Where } Q_j = \begin{bmatrix} Q_{11}^{(j)} & Q_{12}^{(j)} \dots\dots & Q_{1n}^{(j)} \\ Q_{21}^{(j)} & Q_{22}^{(j)} \dots\dots & Q_{2n}^{(j)} \\ \vdots & \vdots & \vdots \\ Q_{n1}^{(j)} & Q_{n2}^{(j)} & Q_{nn}^{(j)} \end{bmatrix}$$

4.9.1 Layers

a-Input Layer

The first layer is known as the input layer which accepts input for the network, The input layer are the input variables they referred to the visible layer, which includes multiple predictor variables, which accepts input for the network, with each variable corresponding to an independent neuron. At the input of hidden unit l, the input vector x is weighted by input weights W^h .

$$S_l = [X_1 W_{1l}^h \cdot X_2 W_{2l}^h \dots X_n W_{nl}^h \dots X_N W_{Nl}^h] \dots (20)$$

Where: n is the index of input; l is the index of hidden units; X_n is thenth input; $W_{n,l}^h$ is the input weight between input and hidden unit; is the input weight. RBF network with N inputs, L hidden units, and M outputs [8]

b -Hidden Layer or a Processing Layer:

comprises a specific number of nodes with RBF non-linear activation function (conventionally it is implemented as Gaussian function), and the last layer is responsible to generate the final output of the network. The layers of nodes situated between the input and output layers are referred to as hidden layers. There can exist one or multiple instances of these layers. The hidden layer is responsible for processing the outputs derived from the input layer utilizing a distance metric of

$$d_l = \sqrt{\sum_{j=1}^n k_{1l}^i [I_j - m_j^i]^2} \dots (21)$$

and the transfer functions in. The resultant output of the network is represented as a weighted summation of the outputs of $\varphi(d_i)$ emanating from the hidden layer and is articulated by the hidden layer, which comprises a Radial Basis Function (RBF) centered at a specific point, contingent upon the dimensionality of the input/output predictor variables. The activations of the hidden units are described by fundamental functions that typically encompass the activation function of the Radial Basis Function. This activation function is contingent upon the Euclidean distance between the input pattern vector and the center of the hidden neuron, which possesses a multitude of neurons. The function of clustering in the architecture of Radial Basis Function Networks (RBFN) is to establish an initial distribution of receptive fields (hidden neurons) throughout the input space of the input variables. Each neuron's output within the hidden unit l is computed by

$$\varphi_l(\bar{s}_l) = \exp\left(\frac{-\|s_l - c_l\|^2}{\sigma_l}\right) \dots (22)$$

where the activation function $\varphi_1(.)$ for the hidden unit is conventionally selected as a Gaussian function: c is the center of the hidden unit, l and σ_1 is the width of the hidden unit $||. ||$, with denoting the Euclidean distance. This indicates $\varphi_1(s_1)$ that possesses a feasible value solely when the distance $\|s_1 - c_1\|$ is less than the width [28]

c- Output Layer

The output layer is constituted of nodes that generate the output variables, which typically include a linear activation function that subsequently transmits the values to each neuron. The computation of the network output m is performed by:

$$o_m = \sum_{l=1}^L \varphi_l(s_l)W_{l,m}^o + W_{o,m}^o \dots (23)$$

Where: m represented the index of output; $W_{1,m}^o$ is the output weight between the hidden unit1 and the output unit; $W_{o,m}^o$ is the bias weight of output unit m .

One may observe from this fundamental computation the presence of four distinct types of parameters: the input weight matrix W^h , the output weight matrix W_o , the center matrix c , and the width vector σ . Typically, the input weights are uniformly initialized to the value of '1'.

4.9.2-Weights

These are the real values related with the traits. They are significant as they tell the prominence of each trait which is thread as an input to the artificial neural network.

Neurons in neighboring layers are fully related to each other. Each connection has connected weight, which decide the strength of the connection. These weights are literate during the instruction operation[24] .

4.9.3-Bias Neurons

Bias in a neural network is desired to move the activation function towards the level either towards the left or the right. We will pall it in more posterior state .In the bonus to the input and hidden neurons, each layer (except the input layer) generally includes a bias neuron that allows a constant input to the neurons in the next layer. The bias neuron has its own weight related with each connection, which is also liberated during training. bias can mention to two connected but definite ideas: bias as a common expression in machine learning and the bias neuron , bias mentions to the error proposed by approaching to a real-world problem with a clarified model. Bias computes how well the model can arrest the underlying styles in the data. A

high bias designates that the model is too naive and may under fit the data, while a low bias propose that the model is catching the underlying samples well.

4.10 Specialized algorithms normalized importance independent variables using neural networks

A higher normalized importance percentage he Gini's method. Having the higher importance percentage is meant the higher reliability and productivity, identification of significant independent variables through neural networks can be accomplished via a multitude of methodologies. One strategy involves evaluating the relative significance of each variable through an analysis of the connection weights and the sensitivity analysis inherent to the neural network model. An alternative approach entails employing an ensemble of neural networks rather than relying on a singular model, a strategy which enhances the consistency of the results derived from various methodologies Moreover, a technique that utilizes a combination of partial or wholly determined indices can be employed to ascertain the importance of specific nodes within the neural network. Independent variables serve as a crucial component of experimental design, allowing to investigate cause-and-effect relationships between variables.

Normalized architecture[6].

RBF networks can be normalized. In this case the mapping is

$$\text{where } u(\|X - c_i\|) = \stackrel{\text{def}}{=} \frac{\rho(\|X - c_i\|)}{\sum_{i=1}^N \rho(\|X - c_i\|)}$$

is known as a normalized radial basis function.

There is theoretical justification for RBF architecture in the case of stochastic data flow. Assume a stochastic

kernel approximation for the joint probability density

$$P(X \wedge y) = \frac{1}{N} \sum_{i=1}^N \rho(\|X - c_i\|) \sigma(|y - e_i|) \dots (24)$$

where the weights c_i and e_i are exemplars from the data and we require the kernels to be normalized $\int \rho(\|X - c_i\|) d^n X = 1$

And $\int \sigma(|y - e_i|) dy = 1$

The probability densities in the input and output spaces are

$$\int P(X \wedge y)dy = \frac{1}{N} \sum_{i=1}^N \rho(\|X - c_i\|)$$

And the expectation of y given an input X is

$$\varphi(x) \stackrel{\text{def}}{=} E(Y|X) = \int yP(Y|X)dy$$

where $P(y|X)$

is the conditional probability of y given X The conditional probability is related to the joint probability through Bayes' theorem

$$P(y|X) = \frac{P(X \wedge y)}{P(X)} \dots (25)$$

$$\text{which yields } \varphi(X) = \int y \frac{P(X \wedge y)}{P(X)} dy$$

This becomes

$$\varphi(X) = \frac{\sum_{i=1}^N e_i \rho(\|X - c_i\|)}{\sum_{i=1}^N \rho(\|X - c_i\|)} = \sum_{i=1}^N e_i u(\|X - c_i\|) \dots (26)$$

4.11 Performance Measurement [Classification, ROC and AUC]

In Machine Learning, performance measurement is an essential task. So when it comes to a classification problem, we can count on an AUC - ROC Curve. When we need to check or visualize the performance of the multi-class classification problem, we use the AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. It is one of the most important evaluation metrics for checking any classification model's performance. It is also written as AUROC (Area **Under the Receiver Operating Characteristics**[2].If you want to evaluate a model's quality across all possible thresholds, you need different tools

4.11.1 Classification table and Confusion Matrix

During the data training phase, a confusion matrix was employed to assess the multilayer perceptron classification model presented in Table_2. that delineates the number of observation cases that exhibit a particular attribute alongside the number of observation cases that lack that characteristic, in contrast to the anticipated instances that display the attribute and those that do not. Consequently,

The fundamental premise of the analysis is to anticipate the accurate classification of cases based on a specified criterion, as this serves as evidence that the model aligns with the observational data. Table (2) illustrates the overarching structure of the classification table.

Confusion Matrix

Confusion matrix is a matrix which maps the predicted outputs across actual outputs.[1]

Table (2) Classification Table [Confusion Matrix]

Classification	Expected		
		Predictive Positive	Predictive Negative
Observation	Actual Positive	True Positive (TP)	False Negative (FN)
	Actual Negative	False Positive (FP)	True Negative Rate (TN)

Crucial metrics were derived from the classification matrix to facilitate a precise evaluation of the classification model, thereby enabling the prediction of one variable based on the known values of other variables. This process involves mapping the data item into a continuous real-valued variable. A plethora of soft-computing algorithms are available for data mining, including neuro-fuzzy computing, genetic algorithms, neural networks, rough sets, decision trees, and their respective hybridizations[3].

In statistical analysis pertaining to binary classification and information retrieval systems, the F-score, also referred to as the F-Measure as Eq.(27), serves as an indicator of predictive efficacy. This metric is derived from the precision and recall associated with the evaluation, wherein precision is defined as the ratio of true positive results to the total number of samples predicted to be positive, encompassing those incorrectly classified as Eq.(28), while recall is characterized as the ratio of true positive results to the total number of samples that should

have been accurately classified as positives Eq.(29). Precision is alternatively termed as positive predictive value, whereas recall is commonly referred to as sensitivity within the context of diagnostic binary classification[19].

$$F_1_score = \frac{2}{recall^{-1}+precision^{-1}} = 2 \frac{precision \cdot recall}{precision+recall} =$$

$$\frac{2TP}{2TP+FP+FN} \dots (27)$$

$$Precision = \frac{TP}{TP+FP} \dots (28)$$

$$Recall = \frac{TP}{TP+FN} \dots (29)$$

Since both precision and recall are rates (ratios) between 0 and 1 . This ensures that a low value in either precision or recall has a significant impact on the overall F1 score,

The highest possible value of an F-score is 1.0, indicating perfect precision and recall, and the lowest possible value is 0, if the precision or the recall is zero.

4.11-2 Receiver-operating characteristic curve (ROC)

The ROC curve is an analytical method, represented as a graph, that is used to evaluate the performance of a binary diagnostic classification method. The diagnostic test results need to be classified into one of the clearly defined dichotomous categories, such as the presence or absence of a disease[19].

4.11.2.1 Area under the curve (AUC)

The **area under the ROC curve (AUC)** represents the probability that the model, if given a randomly chosen positive and negative example, will rank the positive higher than the negative. The perfect model above, containing a square with sides of length 1, has an area under the curve (AUC) of 1.0. This means there is a 100% probability that the model will correctly rank a randomly chosen positive example higher than a randomly chosen negative example. In other words, looking at the spread of data points below, AUC gives the probability that the model will place a randomly chosen square to the right of a randomly chosen circle, independent of where the threshold is

set. In general, an AUC of 0.5 suggests no discrimination (i.e., ability to diagnose patients with and without the disease or condition based on the test), but (0.7 to 0.8) is considered acceptable, (0.8 to 0.9) is considered excellent, and more than (0.9) is considered outstanding. A ROC curve shows the relationship between clinical sensitivity and specificity for every possible cut-off. The ROC curve is a graph with [9].

The x – axis showing (1 – specificity = false positive fraction = $\frac{FP}{(FP+TN)} \dots (30)$

The y – axis showing (sensitivity = true positive fraction = $\frac{TP}{(TP + FN)}$

The ROC curve is drawn by calculating the true positive rate (TP) and false positive rate (FP) at every possible threshold (in practice, at selected intervals), then graphing TP over FP. A perfect model, which at some threshold has a TP of 1.0 and a FP of 0.0, can be represented by either a point at (0, 1) if all other thresholds are ignored, or by the following:

4.12. Model information criterion or Model selection in machine learning

4.12.1 Model selection

is a machine learning operation utilized to select the finest model for a given task from a group of possible models? The possible models are estimated using different model selection methods. The result of model selection depends on a strong validation plan and suitable estimation metrics. Model Selection task to Maximize achievement [11].

4.12.1.1-Akaike information criterion (AIC)

The Akaike Information Criterion (AIC) counts the type of a statistical model for a given dataset. It stabilizes the trade-off between the goodness of the model's fit on the training data and the complexity of the model. It is an estimator of forecast error and just like comparative quality of statistical models for a given set of data. Given a group of models for the data, evaluates the

kind of each model, comparative to each of the other models. AIC penalizes models with more parameters, propitious the selection of easy that still illustrate the training data well(training dataset is the sample of data used to fit the model). A lower AIC value specifys a better-fitting model. specify a means for model selection. Akaike's Information Criteria (AIC) One of the most regularly used information criteria.AIC is focused to finding the finest approximating model to the unknown true data causing procedure and its applications[25].

The method of AIC (Akaike, 1973) is to chose the model that reduce the negative likelihood penalized by the number of parameters as identified in the equation

$$AIC = 2k - 2 \log(L) \dots \dots \dots (31)$$

Where L refers to the likelihood under the fitted model and k is the number of parameters in the model. Specifically, k = the number of parameters in the model. Specifically, L = the log-likelihood of the model under the fitted model, i.e., how well the model sees its training data

4.12.1.2. Bayesian Information Criterion (BIC)

Obtained from Bayesian probability and conclusion, the Bayesian Information Criterion (BIC) is a model selection statistic identical to AIC butcontains a powerful penalty for model complication. BIC is especially fitting for models trained using maximum likelihood estimation. More compound models will have a larger BIC score, which means weak modelsFormula[25]:

$$BIC = -2 \log(L) + k \log(n) \dots (32)$$

K = the number of parameters

L = the log-likelihood of the model for the training data

N = the number of data points/samples

4.12.2 Model Fitting Information -Likelihood Ratio Test

-The likelihood ratio test can be utilized to estimate the goodness of fit of a model of calculation if the sample is sufficiently large. In this case H1 identify to a 'filled' model in which the number of parameters equals the sample size n[14].

D = likelihood ratio test =

$$-2 * \ln \frac{\text{Likelihood for null model}}{\text{Likelihood for alternative model}} \dots (33)$$

D is the test statistic.

5-Practical Aspect

5.1 A brief explanation about kidney transplant

Kidney transplants are done to help people with chronic kidney disease or end-stage kidney failure. When the kidneys cannot filter waste properly, they need either dialysis (a device used to remove waste from the blood) or a kidney transplant. This medical procedure entails a surgical operation whereby the nonfunctional kidney is supplanted by a viable kidney sourced from either a living or deceased donor[15]. The inaugural successful live-related kidney transplant, performed in 1953.

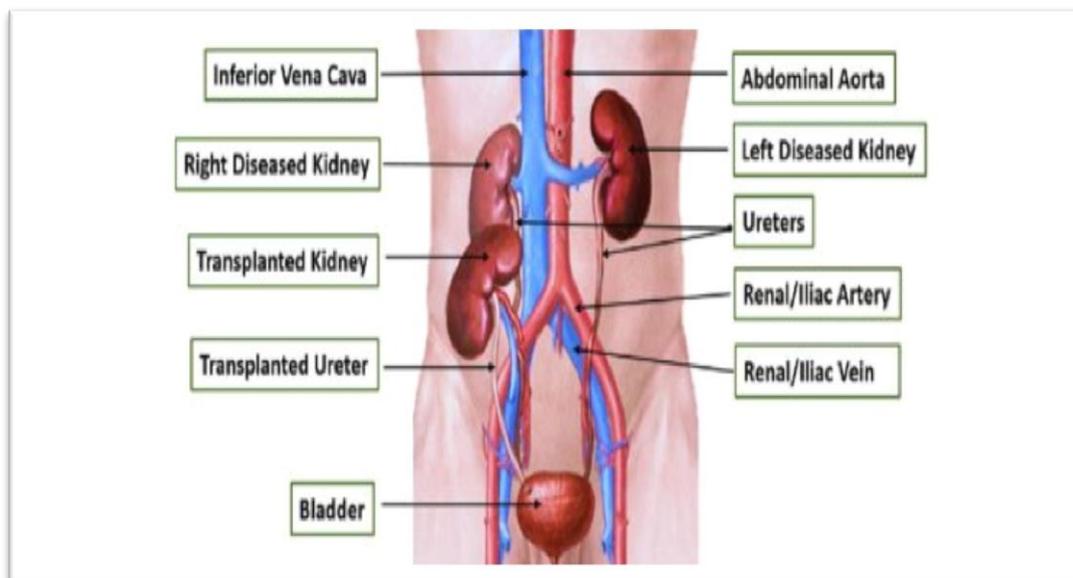


Figure (3): Kidney location after transplant[15]

The kidneys, which are two organs resembling the shape of beans, are situated bilaterally along the spine, just inferior to the ribcage. Each kidney is approximately the size of an adult fist. Their primary role is to filter and eliminate waste products, essential minerals, and excess fluid from the bloodstream via the production of urine. When the kidneys lose their capability to filter effectively, detrimental levels of fluid and waste begin to

accumulate within the body, potentially leading to elevated blood pressure and eventual kidney failure, clinically referred to as end-stage renal disease (ESRD). End-stage renal disease is characterized by the loss of approximately 90% of the kidneys' functional capacity[26].

5.2 Practical Aspect Analysis

5.2.1 Data Description

The researchers contacted several institutions specializing in kidney disease, including the Shar hospital in Sulaimania governorate in Kurdistan, 267 patients were registered with kidney disease only the data from 150 patients in that hospital enrolled between February 2015 and May 2022 were eligible for this analysis

5.2.2 Variables Description

Dependent Variable (Y): defined as binary variable [patients with Kidney transplanted , Kidney prepare for transplant patients]

Independent Variables : In this paper we have deal with common risk factors caused Kidney transplant as independent variables for this purpose of the paper, many doctors specialized in kidney diseases were consulted, and confirmed that those variables are six risk factors caused Kidney transplantation ($X_1, X_2, X_3, X_4, X_5, X_6$) they explained in table(3)

5.2.3 Statistical Analysis

5. 2.3.1: first: Descriptive Statistics of Kidney dataset

Table (3): Shows descriptive statistics of Kidney transplantation dataset

Variables		Frequency	Percent	
Dependent Y	Kidney transplanted	88	58.7%	
	Kidney prepare for transplant	62	41.3%	
Independents	X ₁ : Renal Artery Stenosis	Semi Control	54	36.0%
		Out Control	96	64.0%
	X ₂ : Polycystic Kidney Disease	Semi Control	57	38.0%
		Out Control	93	62.0%
	X ₃ : Diabetes	Semi Control	62	41.3%
		Out Control	88	58.7%
	X ₄ : Congenital Kidney Diseases	Semi Control	48	32.0%
		Out Control	102	68.7%
	X ₅ : Chronic, Uncontrolled High Blood Pressure	Semi Control	62	41.3%
		Out Control	88	58.7%
	X ₆ : Systemic Lupus Erythematosus	Semi Control	56	37.3%
		Out Control	94	62.7%

The result of descriptive statistics :

1-The dependent variable referred to the number of kidney transplanted are (88) with (58.7%) and it is higher than those number of Kidney prepare for transplant (62) with (41.3%)

2-independents variables out-control is higher than Semi Control in all variables

[X₁=64.0% ,X₂= 62.0% , X₃=58.7% , X₄= 68.7% , X₅=58.7% , X₆=62.7%]

it indicate that every risk factors has effect on the kidney transplant and the 62% of patients affected by risk factors

5.2. 3.2: Second: Advanced Statistics ANN -RBF of Kidney dataset

1 -Case Processing Summary

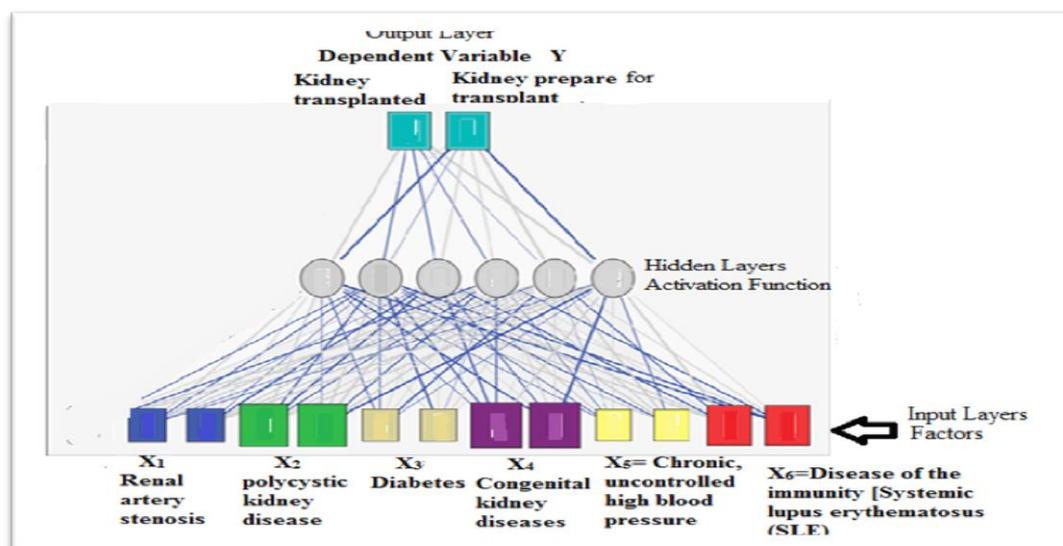
Table(4):Case Processing Summary

		Numbers	Percent
Sample	Training	115	76.7%
	Testing	35	23.3%
Excluded		0	
Total		150	100.0%

Table (4) gives information about the datasets used to build the RBF neural network model. The Case Processing Summary shows that the training sample has 115 cases, which is equal to 76.7% of the total sample. This larger portion was used to ensure that the model is designed by the effect of risk factors on Kidney Transplantation. The testing sample has 35 cases, which is equal to 23.3% of the total sample. This smaller part of the dataset was used to estimate how well the model generalizes to new hidden cases. It means that a sufficient portion was used for testing to achieve generalization. However, the excluded category has zero cases because no data records were used, meaning the full dataset was used.

2-Network Information

Figure(4) The neural network is structured to classify kidney transplant status based on 6 clinical input **variables** risk factors considered by the model to predict the kidney transplant. It has one hidden layer with 12 neurons. The output layer uses Identity activation with 2 units, reflecting binary classification output (e.g., transplanted vs. prepare for transplant).

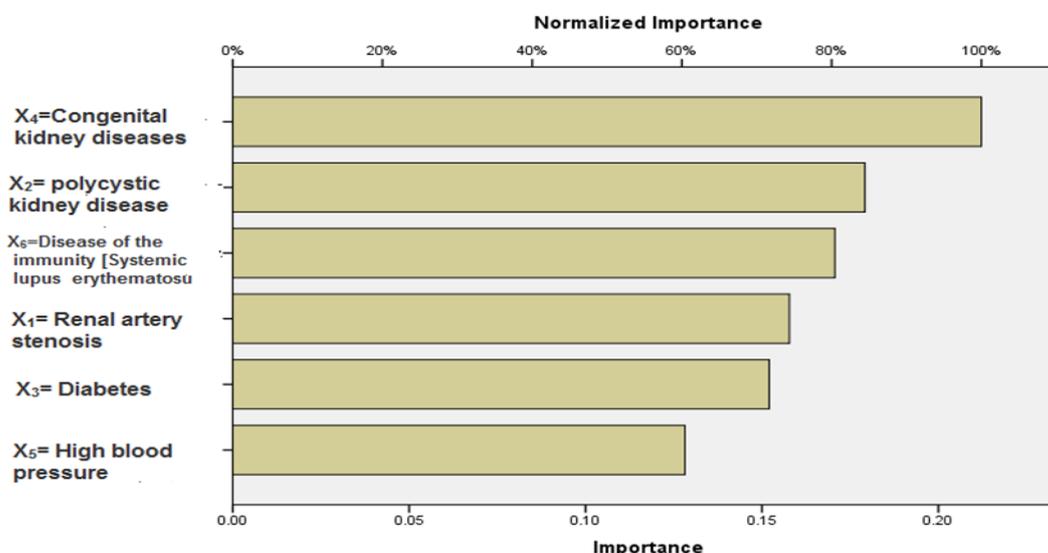


**Figure (4): Network Information Diagram
 3–The Normalized Important Independent Variables**

**Table (5): The relative importance of each factor affecting
 Kidney transplant**

Independent Variables	Importance	Normalized Importance
Renal Artery Stenosis	0.158	74.4%
Polycystic Kidney Disease	0.179	84.4%
Diabetes	0.152	71.7%
Congenital Kidney Diseases	0.212	100.0%
Chronic, Uncontrolled High Blood Pressure	0.128	60.4%
Systemic Lupus Erythematosus	0.171	80.4%

Figure (5): Feature Normalized importance of each factor affecting Kidney transplanted



From table (5) and the figure (5) and the relative normalized importance of each independent variables (risk factor affecting on the dependent variable Kidney transplant) by using the RBF model, The model identifies "Congenital kidney diseases" (X₄) as the most significant factor, with a normalized importance of 100%, the highest relative effect on the dependent Kidney transplanted variable it's more than (polycystic kidney disease) which has 84.4%, then (Systemic lupus erythematosus) with 80.4%, then (Renal artery stenosis) 74.4%, Diabetes 71.7%, (Chronic, uncontrolled high blood pressure) 60.4%. So, these values reflect how much each variable contributes to predicting the Kidney transplant in the RBF model.

4- The Performance of Diagnostic Tests a-Classification(Confusion matrix)

Table (6): Classification

sample	Observed	Predicted for the dependent variable: Y		
		Kidney transplanted	Kidney prepared for transplant	Percent Correct
Training	Kidney transplanted	67	2	97.1%
	Kidney prepare for transplant	9	37	80.4%
	Overall Percent	66.1%	33.9%	90.4%
Testing	Kidney transplanted	19	0	100.0%
	Kidney prepare for transplant	5	11	68.8%
	Overall Percent	68.6%	31.4%	85.7%

Table(6) displays a classification table or. confusion matrix for dependent variable Kidney transplant defined if the predicted probability is greater than 0.5 of the ratio of the correct prediction In the Training Sample the Kidney transplanted and Kidney prepared for transplant the results in the diagonal are (67 , 37 with **overall accuracy: 90.4%**, which is **very strong** for a medical prediction task and in the Testing sample Kidney transplanted and Kidney prepared for transplant the results in the diagonal are (19 , 11 with overall Percent 85.7%) still **very acceptable** it predict that the both kinds of patients affected by the risk factors. In general High overall accuracy (90.4% and 85.7%) suggests the model is reliable and clinically valuable.

$$F_{1_score} = 2 \left[\frac{precision * recall}{precision + recall} \right] = \frac{2TP}{2TP+FP+FN}$$

$$F_{1_training} = \frac{2(67)}{2(67) + 9 + 2} = \frac{134}{145} = 0.9241$$

$$F_{1_testing} = \frac{2(19)}{2(19) + 5 + 0} = \frac{38}{43} = 0.8837$$

So in both analyses the values are (0.9241 and 0.8837) it refer to perfect precision and to best predictive..

b-ROC curve (Receiver Operating Characteristics):

Table (7): Area Under the Curve(AUC)

Dependent Variable		Area Under the Curve(AUC)
Y	Kidney transplanted	0.835
	Kidney prepare for transplant	0.835

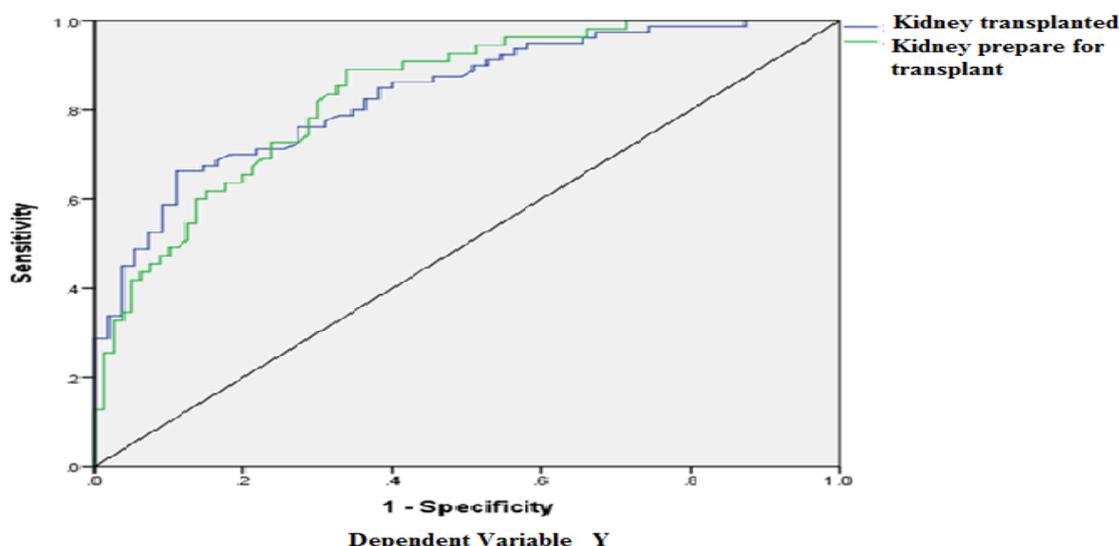


Figure (6)ROC curve

In both Table (7) and Figure (6), the area under the ROC curve (AUC) for both outcomes is 0.835, which is classified as "best" since an AUC in the range of 0.8 to 0.9 is deemed excellent, signifying that the model possesses a robust capacity to accurately classify whether a kidney is suitable for transplantation or is prepared for transplant.

5- Model Selection

a-Model Fitting Information

Table (8): Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	Log Likelihood	Chi-Square	df	Sig.
Intercept	167.097	171.312	127.666			
Final Model	162.836	196.553	99.090	28.576	6	.000

This table presents the validation plan and appropriate Model Fitting Information, indicating that the chi-square test for AIC, BIC, and the likelihood ratio yields a P-value of 0.000, which is less than 0.05. This encapsulates model fit statistics, demonstrating the final model's improvement over the simpler model and assessing its statistical significance.

b-Model Fitting Criteria

Table (9): Model Fitting Criteria Likelihood Ratio Tests of Reduced Model

Effect	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept	119.695	149.197	99.090	.000	0	.
Renal Artery Stenosis	125.803	151.091	108.579	9.489	1	0.032
Polycystic Kidney Disease	121.929	147.217	106.509	7.419	1	0.041
Diabetes	122.803	146.091	104.792	6.702	1	0.046
Congenital Kidney Diseases	123.929	149.217	111.891	12.801	1	0.000
Chronic, Uncontrolled High Blood Pressure	124.803	150.091	104.168	5.078	1	0.049
Systemic Lupus Erythematosus	120.295	144.583	105.924	5.834	1	0.043

The chi-square statistic represents the disparity in log-likelihoods between the final model and a constrained model. The reduced model is created by excluding an effect from the final model. The null hypothesis posits that all effect parameters are equal to zero. The reduced model is equivalent to the final model, as the exclusion of the effect does not augment the degrees of freedom. As the p-values for all predictor variables (six risk factors) are >0.05 , it demonstrates that the specified clinical variables significantly enhance the model fit compared to the intercept-only model. The statistical tests validate that these elements are crucial predictors in the model.

6. Conclusion

The research illustrates that employing Radial Basis Function (RBF) for predictive classification, in conjunction with the normalized significance of independent variables, proficiently identifies critical risk factors affecting kidney transplant results. The classification model demonstrates robust performance with an Area Under the Curve (AUC) of 0.835 and high accuracy in both training (90.4%) and testing (85.7%) datasets, underscoring its efficacy in differentiating between transplanted kidneys and those designated for transplant. Congenital Kidney Diseases was identified as the most significant predictor among the risk factors, with a normalized relevance of 100%, followed by Polycystic Kidney Disease at 84.4% and Systemic Lupus Erythematosus at 80.4%. All enumerated parameters substantially influence the model, as evidenced by their likelihood ratio tests with p-values below 0.05, thereby affirming their statistical relevance in forecasting transplant risk. The final fitted model markedly enhances fit compared to the intercept-only model, as evidenced by a highly significant likelihood ratio test (Chi-square = 28.576, $p < 0.001$). Integrating these normalized risk indicators using RBF classification improves prediction accuracy and offers critical insights for evaluating kidney transplant risks, hence facilitating clinical decision-making and targeted interventions. This result is derived directly from the model fitting

requirements, displayed in the tables, The decrease in Log Likelihood values with each factor indicates improved model fitting.

These results suggest all these normalized risk factors provide valuable predictive information for kidney transplant risk when used in the model, confirming the benefits of using RBF and normalized importance techniques for prediction by classification in this study .

Recommendations:

- 1-Integrate the RBF-based prediction model with normalized variable significance into clinical practice to inform pre-transplant risk evaluations and post-transplant surveillance.
- 2-Utilize recognized important risk factors (e.g., congenital kidney disorders, polycystic kidney disease, diabetes) to customize patient-specific treatment and follow-up strategies.
- 3-Investigate the integration of our predictive models with additional clinical tools and biomarkers for integrated kidney transplant risk management.
- 4-Encourage interdisciplinary collaboration among physicians, data scientists, and academics to enhance and revise predictive models as new data becomes available.
- 5-This work corresponds with current advancements in machine learning applications for predicting kidney transplant risk, demonstrating enhanced graft survival prediction and facilitating targeted therapeutic interventions for improved long-term results.

References

- [1] Al Barznji, N.S.K., Abdullah, D.F., Kareem, R. & Sedeeq, B.S., 2024. Assessing by model fitting criteria and comparing between Stirling and Aitken interpolation formulas to determine the preferred treatment of choice (surgery or surgery with chemotherapy) for lung cancer. Zanko Journal of Pure and Applied Sciences, 28(3), pp.266–281. <https://doi.org/10.21271>.

- [2] Al Barznji, N.S.K., 2018. Using logistic regression analysis and linear discrimination analysis to identify the risk factors of diabetes. *Zanko Journal*, 22(6), pp.248–268.
- [3] Amin, M.Z. & Ali, A., 2017. Application of Multilayer Perceptron (MLP) for Data Mining in Healthcare Operations. *Proceedings of the 3rd International Conference on Biotechnology*, University of Engineering and Technology, Lahore.
- [4] Broomhead, D.S. & Lowe, D., 1988. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, pp.321–355.
- [5] Broomhead, D.S. & Lowe, D., 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks. *RSRE Technical Report 4148*. Archived from original 2013.
- [6] Buhmann, M.D., 2003. *Radial Basis Functions: Theory and Implementations*. Cambridge University Press.
- [7] Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M. & Lin, C.-J., 2010. Training and Testing Low-degree Polynomial Data Mappings via Linear SVM. *Journal of Machine Learning Research*, 11, pp.1471–1490.
- [8] Chen, D., 2021. Evaluation Model of Physical Education Effect: On the Application of Radial Basis Function Particle Swarm Optimization Neural Network (RBFNN-PSO). *Computational Intelligence and Neuroscience*, 2021, Article ID 5570685.
- [9] Collins, N.A., 2005. *Final Destination: Looks Could Kill*. Black Flame. ISBN 1844163164.
- [10] Dash, C.S.K., Behera, A.K., Dehuri, S. & Cho, S.B., 2016. Radial basis function neural networks: a topical state-of-the-art survey. *Open Computer Science*, 6(1), pp.33–63.
- [11] Ding, J., Tarokh, V. & Yang, Y., 2018. Model Selection Techniques: An Overview. *IEEE Signal Processing Magazine*, 35(6), pp.16–34. <https://doi.org/10.1109/MSP.2018.2867638>.
- [12] Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., et al., 2020. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research,

practice and policy. *International Journal of Information Management*, 57, p.101994.

[13] Fernández-Navarro, F., Hervás-Martínez, C. & Martínez-Álvarez, F., 2012. Parameter estimation of q-Gaussian Radial Basis Functions Neural Networks using a hybrid evolutionary algorithm. *Neurocomputing*, 75(1), pp.123–134.

[14] Glover, S. & Dixon, P., 2004. Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, 11(5), pp.791–806. <https://doi.org/10.3758/bf03196706>.

[15] Grafals, M., Chandraker, A., et al., 2019. Noninfectious complications after kidney transplantation. In: *Chronic Kidney Disease, Dialysis, and Transplantation*, 4th ed., pp.1–12.

[16] Hossain, M.I., Maruf, M.H., Khan, M.A.R., Siddiqi Prity, F., Fatema, S., Ejaz, S. & Khan, M.A., 2023. Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran Journal of Computer Science*.

[17] Kirişci, M., 2019. Fibonacci statistical convergence on intuitionistic fuzzy normed spaces. *Journal of Intelligent & Fuzzy Systems*, 36(6), pp.5597–5604.

[18]. MontesinosLópez, O.A., MontesinosLópez, A. & Crossa, J., 2022. *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer International Publishing, Cham, pp.379–425. https://doi.org/10.1007/978-3-030-89010-0_10

[19]. Murad, S.H., Awlla, A.H. & Moahmmed, B.T., 2023. Prediction Lung Cancer based critical factors using machine learning. *Scientific Journal of University of Zakho*, 11(3), pp.447–452.

[20] Nguyen, V.T., Nguyen, Q.C., Van, M., Su, S.F., Garg, H., Duong, D.N. & Tan, P.X., 2025. Robust adaptive nonlinear PID controller using radial basis function neural networks for ballbots under external force. *Engineering Science and Technology, an International Journal*, 61, p.101914

[21] Rasmussen, C.E. & Williams, C.K.I., 2006. *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

- [22] Samuel, A.L., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3), pp.210-229.
- [23] Schwenker, F., Kestler, H.A. & Palm, G., 2001. Three learning phases for radial-basis-function networks. Neural Networks, 14(4–5), pp.439–458. [https://doi.org/10.1016/s0893-6080\(01\)00027-2](https://doi.org/10.1016/s0893-6080(01)00027-2).
- [24] Sejnowski, T.J., 2018. The Deep Learning Revolution. MIT Press.
- [25] Stoica, P. & Selen, Y., 2004. Model-order selection: A review of information criterion rules. IEEE Signal Processing Magazine, 21(4), pp.36–47. <https://doi.org/10.1109/MSP.2004.1311138>.
- [26] Syu, S.H. et al., 2019. Risk factors for complications and graft failure in kidney transplant patients with sepsis. Bosnian Journal of Basic Medical Sciences, 19(3), pp.304–309.
- [27] Wu, Y., Wang, H., Zhang, B. & Du, K.L., 2012. Using radial basis function networks for function approximation and classification. International Scholarly Research Notices, Article ID 324194.
- [28] Yadu, A.K. & Shrivastava, G., 2021. Study of Applications of Radial Basis Function Network in Forecasting. Advances in Materials and Lightweight Technologies. <https://doi.org/10.5185/amlett.2021.051628>
- [29] Zhao, A., Xing, S., Wang, X. & Sun, J.Q., 2024. Radial basis function neural networks for optimal control with model reduction and transfer learning. Engineering Applications of Artificial Intelligence, 136, p.108899.