مجلــة العلـوم الإنسانيــة والتطبيقيـــــة
Journal of Humanitarian and Applied Sciences
كلية الآداب والعلوم قصر خيار -جامعة المرقب
المجلد العاشر -العدد التاسع عشر -يونيو /2026

كفاءة نموذج الانحدار الخطي المتعدد في تحديد العوامل المؤثرة في معدل السكر التراكمي

باستخدام لغة البرمجة R

انتصار علي ارهيمة
محاضر – إحصاء
قسم الإحصاء – كلية العلوم –جامعة المرقب
EAArhema@elmergib.edu.ly

تاريخ استقبال البحث: 02 أبريل 2026م تاريخ قبول البحث: 29 أبريل 2026م

المخلص:

هدفت الدراسة إلى تحليل العلاقة بين معدل السكر التراكمي (HbA1c) وكل من مستوى السكر في الدم (FBS)، العمر، والوزن باستخدام نموذج الانحدار الخطي المتعدد باستخدام لغة R. اعتمدت على المنهج الوصفي التحليلي مع التحقق من افتراضات النموذج (الطبيعية، الاستقلالية، عدم الازدواج الخطي، وعدم وجود قيم شاذة). أظهرت الاختبارات الإحصائية (Shapiro-Wilk، VIF، Durbin–Watson) تحقق افتراضات النموذج.

بيّنت النتائج أن النموذج معنوي إحصائياً ($F = 12.55$، $p < 0.05$) ويُفسر 45% من التباين في معدل السكر التراكمي، وُجد تأثير معنوي لكل من مستوى السكر في الدم والعمر، بينما لم يكن للوزن تأثير معنوي، أوصت الدراسة بتوسيع التحليل ليشمل عوامل إضافية مع التأكيد على أهمية استخدام الانحدار الخطي ولغة R في البحوث الطبية.

الكلمات المفتاحية: اختبار التوزيع الطبيعي، اختبار دوربن – واتسون، الانحدار الخطي المتعدد، لغة البرمجة R ، معامل تضخم التباين.

Efficiency of Multiple Linear Regression in Identifying Factors Affecting Glycated Hemoglobin (HbA1c) Using R Programming.

Enttsar Ali Arhema
Lecturer - Statistics
Department of Statistics, Faculty of Science - Al-Marqab University
EAArhema@elmergib.edu.ly

Abstract

This study aimed to analyze the relationship between glycated hemoglobin (HbA1c) and fasting blood sugar (FBS), age, and weight using a multiple linear regression model implemented in R programming language. The study adopted a descriptive-analytical approach and verified the key assumptions of the regression model, including normality, independence, multicollinearity, and absence of outliers.

Statistical tests, including the Shapiro–Wilk test, Variance Inflation Factor (VIF), and Durbin–Watson test, confirmed that the model assumptions were satisfied. The results indicated that the regression model was statistically significant ($F = 12.55$, $p < 0.05$) and explained 45% of the variance in HbA1c levels.

Furthermore, the findings showed that fasting blood sugar and age had a statistically significant effect on HbA1c levels, whereas weight did not show a significant effect. The study recommends expanding future analyses to include additional explanatory variables and highlights the importance of multiple linear regression and R programming in medical research.

Keywords: Normality Test, Durbin–Watson Test, Multiple Linear Regression, R Programming Language, Variance Inflation Factor.

المقدمة

يعد مرض السكري من القضايا الصحية ذات الأهمية المتزايدة عالميًا، نظرًا لارتفاع معدلات انتشاره وما يترتب عليه من مضاعفات مزمنة تؤثر على جودة حياة الأفراد، ويُعتبر معدل السكر التراكمي (HbA1c) من أهم المؤشرات الحيوية المستخدمة لتقييم السيطرة طويلة المدى على مستوى الجلوكوز في الدم، الأمر الذي يجعله محورًا رئيسيًا في العديد من الدراسات الطبية والإحصائية، وفي هذا الإطار، برزت الحاجة إلى تحليل العوامل المؤثرة في هذا المؤشر بدقة، باستخدام نماذج إحصائية قادرة على تفسير العلاقات المعقدة بين المتغيرات، يعتمد التوصيف النظري لمشكلة البحث على افتراض وجود علاقة بين معدل السكر التراكمي وعدد من المتغيرات المستقلة، مثل مستوى السكر في الدم (FBS)، عمر المريض، ووزنه؛ حيث يُفترض أن هذه العوامل تسهم بدرجات متفاوتة في تفسير التغير في HbA1c. ويُعد نموذج الانحدار الخطي المتعدد من أكثر الأساليب الإحصائية استخدامًا في هذا السياق؛ لما يوفره من إمكانية تحليل تأثير عدة متغيرات مستقلة على متغير تابع كمي، مع التحكم في التداخل بينها، إلا أن فعالية هذا النموذج تعتمد على تحقق مجموعة من الافتراضات الإحصائية الأساسية، مثل التوزيع الطبيعي للبواقي، واستقلالية الأخطاء، وعدم وجود ازدواج خطي، وهو ما يستدعي التحقق الدقيق منها لضمان موثوقية النتائج، فقد أظهرت العديد من الدراسات وجود علاقة معنوية موجبة بين مستوى السكر في الدم ومعدل HbA1c c، كما أكدت دراسات أخرى وجود تأثير للعمر على هذا المؤشر نتيجة التغيرات الفسيولوجية المرتبطة بالتقدم في السن، في المقابل، أظهرت بعض الدراسات نتائج متباينة فيما يتعلق بتأثير الوزن؛ حيث اعتبرته بعض الأبحاث عاملاً مؤثراً، بينما لم تثبت دراسات أخرى دلالته الإحصائية، مما يعكس اختلاف النتائج تبعاً لاختلاف طبيعة العينات والأساليب المستخدمة، وانطلاقًا من هذا التباين في النتائج، تتمثل مشكلة البحث في تقييم كفاءة نموذج الانحدار الخطي المتعدد في تحديد العوامل المؤثرة في معدل السكر التراكمي، مع التركيز على التحقق من صحة افتراضاته باستخدام أدوات تحليلية متقدمة مثل لغة البرمجة R. ويتمثل الغرض الرئيسي من هذا العمل في بناء نموذج إحصائي دقيق قادر على تفسير العلاقة بين HbA1c والمتغيرات المستقلة المدروسة، وتحديد مدى إسهام كل منها في تفسير التباين في هذا المؤشر؛ كما يسعى البحث إلى اختبار الفرضيات التالية: وجود تأثير معنوي لكل من مستوى السكر في الدم، والعمر، والوزن على معدل HbA1c ، بالإضافة إلى اختبار معنوية النموذج ككل، وعليه فإن هذه الدراسة تجمع بين التحليل الإحصائي التطبيقي والمراجعة النظرية للأدبيات ذات الصلة، بهدف تقديم فهم أعمق للعوامل المؤثرة في معدل السكر التراكمي، ودعم استخدام الأساليب الإحصائية المتقدمة.

مشكلة الدراسة:

تشهد الآونة الأخيرة ازديادًا ملحوظاً في أعداد المصابين بمرض السكري، وحتى أننا لا نكاد نرى بيتًا يخلو من مرض السكري، وهذا ما دفعنا الى استخدام الأساليب الإحصائية مثل تحليل الانحدار المتعدد لنكتشف أي من هذه العوامل الأكثر تأثيرًا على مريض السكري، و كذلك على الرغم من أهمية الانحدار الخطي المتعدد في جميع الدراسات البحثية، فإنه يجب توخي الحذر الشديد عند تطبيقه؛ لأنه غالباً ما يساء فهمه ولا يتم استخدامه بالشكل الصحيح، ويمكن أن يعطي عندها نتائج غير صحيحة ومضللة، خصوصاً إذا لم تراع الشروط أو الافتراضات على البيانات عند تطبيقه ويصبح غير ذي مصداقية في النتائج والتفسير، وكذلك نظرًا لأهمية لغة البرمجة R في تكوين نموذج الانحدار الخطي وكيفية تقدير معالمه والتحقق من شروطه وافتراضاته لحصول على نتائج غير مشكوك فيه ويمكن الاعتماد عليه إلا أن هناك الكثير من الباحثين خاصة غير المتخصصين لا يستخدمون لغة البرمجة R في الدراسات و البحوث العلمية.

أهداف الدراسة:

تهدف هذه الدراسة إلى ما يلي:

1. بناء نموذج انحدار خطي متعدد وتقدير معالم هذا النموذج، والتحقق من افتراضاته باستخدام لغة البرمجة R .
2. تقدير العلاقة بين معدل السكر التراكمي وعدة عوامل منها مستوى السكر في الدم وعمر المريض ووزنه.

أهمية الدراسة:

1. تسليط الضوء على أهمية الانحدار المتعدد وكفاءة نموذج الانحدار في تحديد العوامل الأكثر تأثيراً على مريض السكري لوضع معالجات مناسبة للتقليل من الإصابة به.
2. كيفية تحليل وتكوين نموذج الانحدار لبيانات متعددة المتغيرات وكذلك كيفية تقدير معالم نموذج الانحدار المتعدد والتحقق من افتراضاته باستخدام لغة البرمجة R .

منهجية الدراسة:

اعتمد الباحثة على المنهج الوصفي التحليلي في توصيف وتحليل وتفسير الانحدار الخطي المتعدد انطلاقاً من مكونات صيغته الرياضية.

حدود الدراسة:

أجريت هذه الدراسة في مستشفى السكر بمدينة الخمس خلال الفترة الزمنية (2025) حيث تم تجميع عينات الدراسة من المرضي المترددين على المستشفى، وتم تعبئة استبيان يحتوي على مجموعة من الأسئلة حول مرض السكري لديهم وحالتهم الصحية، مع الاستدلال بالعمر والوزن ومستوى السكر في الدم ومعدل السكر التراكمي لمريض السكر، تم تجميع50 عينة من مرض السكر.

الدراسات السابقة ذات الصلة:

تعد دراسة العوامل المؤثرة في معدل السكر التراكمي (HbA1c) من الموضوعات التي حظيت باهتمام واسع في الأدبيات الإحصـائية والطبية؛ حيث سـعت العديد من الدراسـات إلى تحديد طبيعة العلاقة بين HbA1c والمتغيرات الديموغرافية والبيولوجية باستخدام أساليب تحليلية متقدمة، وعلى رأسها نموذج الانحدار الخطي المتعدد.

في هذا السياق، أجرى Badri S. A. I. (2025) دراسة هدفت إلى تحليل العلاقة بين HbA1c ومستوى السكر في الدم (FBS) باستخدام أسلوبي الارتباط والانحدار؛ حيث أظهرت النتائج وجود علاقة موجبة ومعنوية قوية بين HbA1c وFBS، مما يشير إلى أن ارتفاع مستوى السكر في الدم ينعكس بشكل مباشر على زيادة معدل السكر التراكمي، كما أكدت الدراسة ملاءمة نموذج الانحدار في تفسير هذه العلاقة، وهو ما يدعم استخدامه في الدراسات الطبية التطبيقية، ومن جانب آخر، تناولت دراســة (Yates A. P. 2002) تأثير العمر على HbA1c؛ حيث توصلت إلى وجود علاقة طردية ذات دلالة إحصائية بين التقدم في العمر وارتفاع HbA1c ، وهو ما يُعزى إلى التغيرات الفسـيولوجية المرتبطة بوظائف البنكرياس مع التقدم في السـن، وتتفق هذه النتيجة مع العديد من الدراسـات التي تؤكد أهمية العمر كعامل مفسر في نماذج الانحدار الخاصة بمرض السكري.

في المقابل، أظهرت بعض الدراسـات نتائج مغايرة فيما يتعلق بتأثير الوزن HbA1c. فقد بينت دراسـة Lin W. (2024) وجود علاقة بين مؤشـر كتلة الجسـم (BMI) و HbA1c، حيث اعتُبر الوزن عاملاً مؤثراً في بعض الحالات، خاصـة لدى الأفراد الذين يعانون من السـمنة، مما يشـير إلى أن تأثير الوزن قد يكون مرتبطاً بخصـائص العينة المدروسة، وبناءً على ما سبق، يتضح أن هناك اتفاقاً عاماً بين الدراسات على أهمية كل من العمر ومستوى السكر في الدم كعوامل مؤثرة في HbA1c، في حين يظل تأثير الوزن محل جدل؛ حيث تختلف نتائجه تبعاً لاختلاف العينات والمنهجيات المسـتخدمة، كما تؤكد هذه الدراسـات مجتمعة أهمية اسـتخدام نموذج الانحدار الخطي المتعدد كأداة فعالة في تحليل العلاقات بين المتغيرات الصحية، مع ضرورة التحقق من افتراضاته لضمان دقة النتائج.

أولاً: الإطار النظري:

1. الانحدار الخطي المتعدد (Regression Linear Multiple):

يعد أسلوب تحليل الانحدار الخطي المتعدد من أهم الأدوات البحثية المستخدمة في شتى المجالات، ويهدف إلى دراسة العلاقة البينية بين متغير متأثر أو معتمد تابع dependent variable (Y) وأكثر من متغير مستقل(X) independent variable بواسطة معادلة رياضية تسمى نموذج الانحدار الخطي المتعدد The Multiple Liner Regression Model تحليل الانحدار المتعدد هو أداة إحصائية تقوم ببناء نموذج إحصائي وذلك لتقدير العلاقة بين متغير كمي واحد، وهو المتغير التابع ومتغير كمي آخر أو عدة متغيرات كمية وهي المتغيرات المستقلة، بحيث ينتج معادلة إحصائية توضح العلاقة بين المتغيرات، ويمكن استخدام هذه المعادلة في معرفة نوع العلاقة بين المتغيرات وتقدير المتغير التابع باستخدام المتغيرات الأخرى، وانطلاقًا من هذا المفهوم حول تحليل الانحدار المتعدد سيتم في هذا البحث استخدام بيانات عينة من مرضى السكري تم أخذها من مستشفى الخمس لتقدير نماذج إحصائية مختلفة والقيام باختبار الفرضيات حول هذه النماذج، وفي هذه الدراسة نود بناء نموذج انحدار خطي متعدد، وتقدير معالم نموذج الانحدار الخطي المتعدد، والتحقق من افتراضاته

باستخدام لغة البرمجة R لتقدير العلاقة بين معدل السكر التراكمي وعدة عوامل منها: مستوى السكر في الدم، والعمر والوزن، حيث سنستخدم معدل السكر التراكمي كمتغير تابع وباقي العوامل متغيرات مستقلة.

فنموذج الانحدار الخطي المتعدد يبحث في العلاقة بين أكثر من متغيرين أحدهما متغير تابع، وبقية المتغيرات مستقلة ويمكن تمثيل النموذج الخطي العام أو المتعدد في صورته الخطية بالشكل الآتي :

$$y_i = \mu_{yi\ /\ xi1,xi2,\ldots\ldots\ldots\ldots,xip} + \varepsilon_i$$
$$= \beta_0 + \beta_2 x_{i1} + \beta_2 x_{i2} + \cdots \ldots + \beta_p x_{ip} + \varepsilon_i \ldots . (1)$$

حيث إن:

y_i : المشاهدة للمتغير التابع على المفردة رقم i

$xi1, xi2, \ldots . , xip$: مجموعة المشاهدات للمتغيرات المفسرة أو المستقلة على المفردة رقم i وعددها p

$\beta_0, \beta_1, \ldots . , \beta_p$: تعبر عن معاملات الانحدار

ε_i :الخطأ العشوائي

وبفرض أن لدينا عينة عشوائية حجمها n مفردة يصبح لدينا n من المعادلات على الصورة رقم (1) ويعبر عنها كالتالي:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1$$
$$\ldots . (2) y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \varepsilon_2$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n$$

ومن ثم يمكن وضع المعادلات الخطية (2) في صورة مصفوفات كما يلي:

$$\begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_p \end{bmatrix}$$
$$y = x\beta + \varepsilon$$

إذا أن:

Y: متجه عمودي من الرتبة (n*1)عناصره متغيرات عشوائية.

X: مصفوفه من الرتبة (n*k) عناصرها ثوابت معلومة.

β: متجه عمودي من الرتبة (p*1) عناصره معالم مجهولة.

ε: متجه عمودي من الرتبة (n*1)عناصره متغيرات عشوائية.

2. فروض نموذج الانحدار الخطي المتعدد

1- متوسط الخطأ العشوائي يساوي صفر

$$E(\varepsilon) = 0$$

أو

$$E\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ .\\ .\\ .\\ \varepsilon_n\end{pmatrix}=\begin{pmatrix}E(\varepsilon_1)\\ E\ (\varepsilon_2)\\ .\\ .\\ .\\ E(\varepsilon_n)\end{pmatrix}=\begin{pmatrix}0\\ 0\\ .\\ .\\ .\\ 0\end{pmatrix}=0$$

2- تباين الخطأ العشوائي يساوى $\sigma^2 I_n$ أي

$$E(\varepsilon\,\varepsilon`)=\sigma^2 I_n$$

$$E(\varepsilon\,\varepsilon`)=E\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ .\\ .\\ .\\ \varepsilon_n\end{pmatrix}_{nx1}\quad(\varepsilon_1\ \varepsilon_{2.....}\varepsilon_n)1_{xn}$$

$$\begin{pmatrix}E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2)\ \dots\dots\ \dots..E(\varepsilon_1\varepsilon_n)\\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2)\quad \dots\dots\quad E(\varepsilon_1\varepsilon_n)\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ E(\ \varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2)\ \dots.\dots\dots\dots\dots E(\varepsilon_n^2)\end{pmatrix}$$

$$=\begin{pmatrix}\sigma^2 & 0\ \dots\dots\dots\dots 0\\ 0 & \sigma^2\ \dots\dots\dots 0\\ & \dots\dots\dots\dots\dots\\ 0 & 0\quad \dots\quad \hat{\alpha}^2\end{pmatrix}=\sigma^2\begin{pmatrix}1 & 0\dots\dots\dots 0\\ 0 & 1\dots\dots\dots 0\\ & \dots\dots\dots\dots\dots\\ 0 & 0\ \dots\ 1\end{pmatrix}$$

$$=\sigma^2 I_n$$

3- الخطأ العشوائي: ε هو عبارة عن متغير عشوائي يتوزع توزيعًا طبيعيًا بمتوسط صفر وتباين $\sigma^2 I_n$ أي:

$$I)\varepsilon_i\ \dots\dots N(O,\sigma^2$$

4- عدم وجود ارتباط ذاتي بين الأخطاء العشوائية أي وجود استقلالية بين الأخطاء من محاولة لأخري.

$$E(\varepsilon=\varepsilon_j)=\sigma^2 I,=0 \qquad i=j,\quad i\neq j$$

$$E(\varepsilon_1\,\varepsilon_2)=0$$

$$E(\varepsilon_2\varepsilon_1)=0$$

وفي حالة وجود ارتباط بين الأخطاء العشوائية فإنه تظهر لدينا مشكلة الارتباط الذاتي بين الأخطاء العشوائية.

وهذا يعد إخلالًا بفرضية النموذج قبل تحليل الانحدار الخطي المتعدد.

5- وجود استقلالية بين المتغيرات المستقلة أي أن التباين المشترك بين (X_i, Y_i) يساوي صفر.

وفي حالة حصول ارتباط بين المتغيرات المستقلة تظهر مشكلة تسمي بمشكلة الازدواج الخطي (أو التعداد الخطي)، والذي يعد خلل بفرضية النموذج الخطي المتعدد ويجب التخلص منها قبل إجراء أي عملية، وأيضًا وجود استقلالية بين المتغيرات المستقلة والأخطاء العشوائية أي أن التغاير أو التباين المشترك كالآتي:

$$COV(X_i, \varepsilon_i) = 0$$

6- المصفوفة Xمن الرتبة K أي أن:

$$V(X) = K$$

وأن K<n

3. تقدير مُعلمات نموذج الانحدار الخطي المتعدد:

نستخدم طريقة المربعات الصغرى لتقدير معالم نموذج الانحدار الخطي المتعدد كالآتي:

بما أن نموذج الانحدار الخطي المتعدد يكتب بالصيغة التالية

$$y = X\beta + \varepsilon \ldots\ldots (3)$$

إذا أن:

$$\hat{\beta} = (x'x)^{-1}x'y$$

حيث إن:

$$x'y = \begin{bmatrix} \sum y \in_i \\ \sum x_{i1}y_i \\ \cdots \\ \sum x_{ip}y_i \end{bmatrix} x'x = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots \sum x_{i1}x_{ip} \\ \cdots & \cdots & \cdots & \cdots \cdots \cdots \\ \sum x_{ip} & \sum x_{ip}x_{i1} & \sum x_{ip}x_{i2} & \cdots \sum x_{ip}^2 \end{bmatrix}$$

حيث إن:

SSe هو مجموع مربعات الخطأ ويحسب

$$SSe = y'y - \beta'x'y$$

أما درجات الحرية لمجموع مربعات الخطأ هي $n-p$ حيث إن p هي عدد المتغيرات في النموذج، أما متوسط مربعات الخطأ فإنه يساوي

$$Mse = \hat{\sigma}^2 = \frac{SSe}{n-p}$$

4. معامل التحديد Determination Coefficient

يشير معامل التحديد المتعدد إلى النسبة التي يمكن تفسيرها من التغير الكلي في المتغير التابع بدلالة المتغيرات المستقلة المدرجة في دالة الانحدار الخطي المتعدد وهو مقياس إحصائي يحدد كفاءة نموذج الانحدار وقيمته تتروح بين الصفر والواحد الصحيح (0-1) وكلما اقترب الرقم من الواحد ازدادت المقدرة التفسيرية للنموذج ويرمز له برمز R^2 ويحسب كالآتي:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

5. معامل التحديد المعدل أو المصحح *adjusted R squared*

معامل التحديد المصحح يضبط المؤشرات الإحصائية المعتمدة على عدد المتغيرات المستقلة في النموذج؛ حيث إن معامل التحديد يعتمد على التغيرات الحاصلة في المتغير التابع Y وبالتالي فإنه لا يأخذ بعين الاعتبار عدد درجات الحرية في أي مشكل إحصائي، ولهذا الغرض يستعمل معامل آخر يسمى معامل التحديد المعدل أو المصحح ويرمز له بالرمز $\bar{R}^2$

وأن قيمة معامل التحديد تتراوح بين الصفر والواحد الموجب وأن من أهم خواص معامل التحديد المصحح أنه لا يمكنه أن يكون أكبر من معامل التحديد، وفي أقصى قيمه هو أن يساويه، كما أنه يجوز أن يكو ن معامل التحديد المصحح أصغر من معامل التحديد، ويحسب معامل التحديد المصحح كالآتي:

$$\bar{R}^2 = \left[1 - (R^2 - 1)\frac{n-1}{n-p}\right]$$

حيث إن:

n عدد أفراد العينة.

P عدد متغيرات نموذج الانحدار الخطي المتعدد.

$R^2 \geq \bar{R}^2$ إذا كانت $p > 1$

إذا كان حجم العينة n كبيرًا، فإن $R^2, \bar{R}^2$ يقتربان في قيمتهما.

6. اختبار صالحية النموذج الخطي:

الفرضية التي سيتم اختبارها هي:

الفرض العدم: النموذج الخطي المتعدد غير مناسب للتنبؤ

$$H_0 : \beta_1 = \beta_2 = \beta_3$$

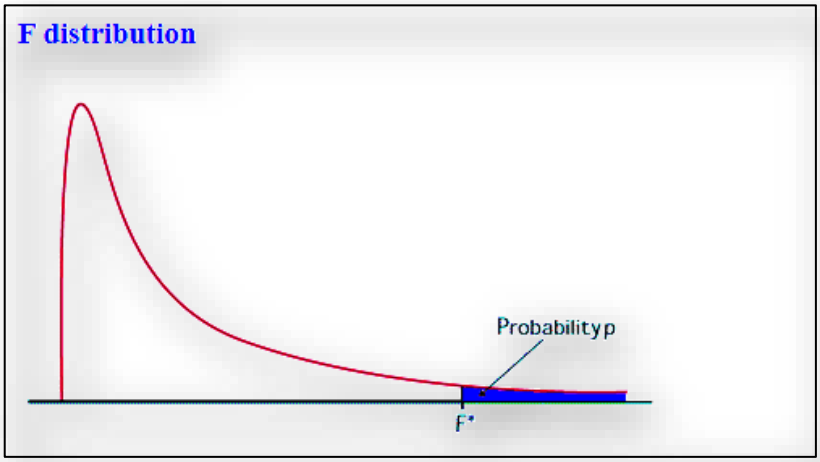
الفرض البديل: النموذج الخطي المتعدد مناسب للتنبؤ

$$H_1 : \text{at least one of } \beta_j \neq 0 \quad , j = 1,2,3$$

إحصائية الاختبار

$$F_{col} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)} \sim F_\alpha(p-1, n-p)$$

وبعد احتساب قيمة F تقارن مع قيمتها الجدولية بدرجة حرية (p-1) و (n-p) للبسط والمقام ولمستوى معنوية 0.05.



الشكل رقم (1) يوضح منطقة القبول والرفض لاختبار F

القرار: عندما $F_{col} > F_{Tab}$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 مما يدل على أنه من بين معلمات النموذج يوجد واحد على الأقل يختلف عن الصفر، أي أن هناك متغيرًا مفسرًا له تأثير جوهري على المتغير التابع بمعنى أن معادلة الانحدار المقدرة لها معنوية إحصائيًا.

عندما $F_{col} \leq F_{Tab}$ نقبل فرضية العدم H_0 أي جميع المتغيرات التفسيرية ليس له تأثير على المتغير التابع وتكون معادلة الانحدار المقدرة غير معنوية إحصائيًا.

7- اختبار صالحية مُعلمات النموذج.

لاختبار المعنويات يجب حساب مصفوفة تباين والتباين المشترك للمعالم المقدرة حتى يمكن حساب الأخطاء المعيارية لهذه المعالم المقدرة؛ حيث تشكل العناصر القطرية في المصفوفة تباين المعالم المقدرة بينما العناصر غير القطرية (أعلى وأسفل القطر) التباين المشترك والترابط بين أي اثنين من هاته المعالم المقدرة ومصفوفة التباين والتباين المشترك كالاتي:

$$V(\hat{\beta}) = \mathrm{MSe}(x'x)^{-1}$$

$$= \mathrm{MSe}\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{ip} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_{ip} & \sum x_{ip}x_{i1} & \sum x_{ip}x_{i2} & \cdots & \sum x_{ip}^2 \end{bmatrix}$$

$$V(\hat{\beta}) = \begin{bmatrix} V(\hat{\beta}_0) & cov(\hat{\beta}_0, \hat{\beta}_1) & \cdots & cov(\hat{\beta}_0, \hat{\beta}_p) \\ cov(\hat{\beta}_1, \hat{\beta}_0) & V(\hat{\beta}_1) & \cdots & cov(\hat{\beta}_1, \hat{\beta}_p) \\ \cdots & \cdots & \cdots & \cdots \\ cov(\hat{\beta}_n, \hat{\beta}_0) & cov(\hat{\beta}_n, \hat{\beta}_1) & \cdots & V(\hat{\beta}_p) \end{bmatrix}$$

وبأخذ الجذر التربيعي الموجب لكل عنصر من عناصر القطر الرئيسي نحصل على الخطأ المعياري لتقدير المعالم $S.E_{\hat{\beta}}$

$$V(\hat{\beta}) = \begin{bmatrix} S.E_{\hat{\beta}_0} & & \\ & S.E_{\hat{\beta}_1} & \\ \dots\dots\dots\dots\dots\dots & & \\ & & S.E_{\hat{\beta}_p} \end{bmatrix}$$

ولاختبار معنوية المعامل، يأخذ الفرض العدم والبديل الصورة التالية:

الفرض العدم: المتغير المستقل ليس له أثر معنوي على المتغير التابع

$$H_0: \beta_i = 0$$

الفرض البديل: المتغير المستقل ذو أثر معنوي على المتغير التابع

$$H_1: \beta_i \neq 0$$

ويحسب إحصائية اختبار معنوية المعامل بالمعادلة التالية

$$t_{col} = \hat{\beta}/S.E_{\hat{\beta}}$$

القرار: إذا كانت $|t_{col}| \leq t_{n-p,\frac{\alpha}{2}}$ نقبل فرضية العدم H_0 ونرفض الفرضية البديلة H_1 ففي هذه الحالة المعلمة ليس لها معنوية إحصائية أي يساوي معنويًا الصفر.

إذا كانت $|t_{col}| > t_{n-p,\frac{\alpha}{2}}$ نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 أي المعلمة لها معنوية إحصائية فهو يختلف معنوياً عن الصفر.

8- اختبارات افتراضات النموذج:

أ. اختبار التوزيع الطبيعي (Normality Distribution Test)

تم استخدام اختبار كولمجوروف – سمرنوف Test Smirnov- Kolmogorov لاختبار ما إذا كانت البيانات تتبع التوزيع الطبيعي من عدمه.

فروض الاختبار هي:

H_0 : البيانات تتبع التوزيع الطبيعي

H_1 : ا البيانات لا تتبع التوزيع الطبيعي

ب. حساب عامل تضخم التباين Variance Inflation Factor (VIF)

ويعرف بالعلاقة

$$\text{VIF} = \frac{1}{1 - \text{R}_\text{j}^2} = \frac{1}{\text{Tolerance}}$$

بالاعتماد على معامل التحديد R كلما أقترب من الواحد الصحيح فإن VIF يقترب من الما لانهاية، أي أنه كلما زاد الازدواج الخطي كلما زاد تباين المقدر، ونهايته تؤول إلى الما لا نهاية، فإذا كانت قيمة معامل التحديد كبيرة وهذا يعني أن المتغير المستقل X_i له ارتباط قوي بين المتغيرات المستقلة الأخرى؛ حيث إن زيادة قيمة VIFعن (10) يدل على أن هناك مشكلة الازدواج الخطي.

ت. اختبار دوربن – واتسون (Durbin - Watson Test)

لا نعلم بوجود الارتباط الذاتي قبل إجراء عملية التقدير، فقد يوجد الارتباط الذاتي وقد لا يوجد، ولذلك يمكن أن نفترض عدم وجوده في البداية ونقوم بتقدير العلاقات بالطريقة العادية دون أن نعدل بياناتها ثم نختبر الفرض القائل بوجود الارتباط الذاتي أو عدمه.

وفي الحقيقة هناك عدة اختبارات للكشف عن وجود الارتباط الذاتي ومن أهمها وأكثرها استخدامًا ما يعرف باختبار (داربن واتسون Durbin _Watson) نسبة للدين قاما باكتشافه.

وذلك لكونه مناسبًا لاختبار وجود الارتباط الذاتي من الدرجة الأولى وبموجبه نضع الفرضية الإحصائية القائلة:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

وتستخدم الأخطاء العشوائية المقدرة في حساب صيغة D.W التالية:

$$D.W = \frac{\sum_{i=2}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} \dots\dots (4)$$

حيث إن:

$(\hat{\varepsilon}_i)$ هي القيمة المقدرة لحد الخطأ العشوائي.

يجب أولاً حساب قيمة إحصائية (D.W) وذلك بموجب الصيغة (11) السابقة الذكر، ثم بعد ذلك نستخرج من الجداول الخاصة بإحصاءD.W القيمة الدنيا (dl) والقيمة العليا (du) بدرجة حرية مساوية لحجم العينة (n) وعدد المعالم المقدرة (K) ولمستوى دلالة معينة (α) وأخيرًا تقارن قيمة D.W العملية أو المحسوبة مع القيمتين الجدولتين وعلى النحو التالي:

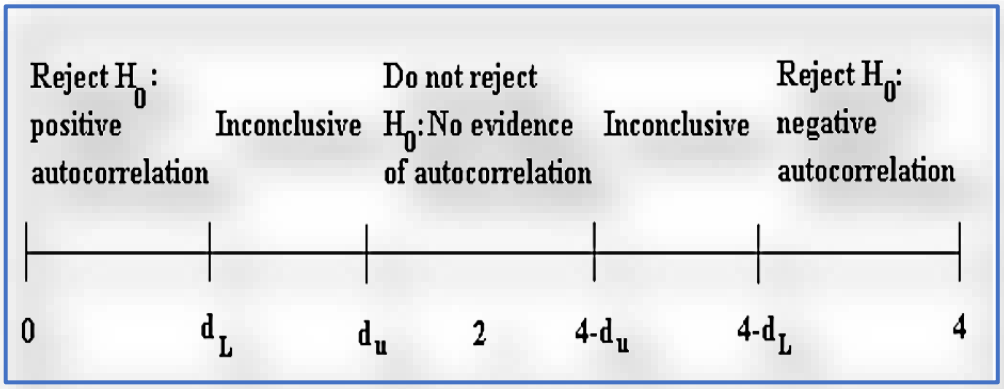
إذا كانت D.W< dl نستنتج وجود ارتباط ذاتي موجب.

إذا كانت $du < D.W < 4 - du$ نستنتج عدم وجود ارتباط ذاتي.

إذا كانت $D.W > 4 - du$ نستنتج وجود ارتباط ذاتي سالب، وفيما عدا ذلك يفشل الاختبار.

ونظرًا للخواص الإحصائية التي تتعلق بالقيمة المحسوبة d فإن الفترة التي تنحصر بين (0,4) يمكن تقسيمها إلى عدة مناطق كما موضح في الشكل رقم (3).

وتبدأ من قيمة (4−dl) إلى نهاية الفترة وهي منطقة قبول H_1 ورفض H_0 وذلك لوجود الارتباط الذاتي في الاتجاه السالب السابق لأن قيمة (ρ) سالبة.



الشكل رقم (2) يبين منطقة القبول والرفض لاختبار دوربن – واتسون

ث. القيم الشاذة: Outliers

ث/1: باستخدام الشكل الانتشاري:

تعرف القيم المتطرفة على أنها: حالات لها بواقٍ معيارية ويجب أن تكون البواقي لا تزيد عن +3 أو لا تقل عن –3 وموزعة بشكل مستطيل بحيث تكون معظم القيم مركزة في المنتصف (على امتداد نقطة الصفر) ولا تأخذ نمط الانحناء والذي يرتفع فيه طرف عن الآخر بشكل ملحوظ؛ حيث إن الانحراف عن الشكل المستطيل يدل على الاخلال بفرضية القيم المتطرف.

ث/2: استخدام معامل المسافة كوك Cook's Distance

يستخدم معامل المسافة كوك في تحليل الانحدار للعثور على القيم المتطرفة المؤثرة في مجموعة من متغيرات التنبؤ، بمعنى آخر، إنها طريقة لتحديد النقاط التي تؤثر سلبًا على نموذج الانحدار **ويتم** حساب معامل المسافة كوك عن طريق إزالة نقطة البيانات i من النموذج وإعادة حساب الانحدار، يقترح كوك فحص الحالات التي يكون فيها $Di > 0.5$، وأنه ينبغي دائمًا دراسة الحالات التي يكون فيها $Di > 1$.

صيغة معامل المسافة كوك

$$Di = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})^2}{(p+1)MSE}$$

حيث إن:

$\hat{y}_j$ هو التنبؤ من نموذج الانحدار للقيمة (j)

$\hat{y}_{j(i)}$ هو التنبؤ من نموذج الانحدار للقيمة (j) بعد حذف القيمة i من نموذج الانحدار.

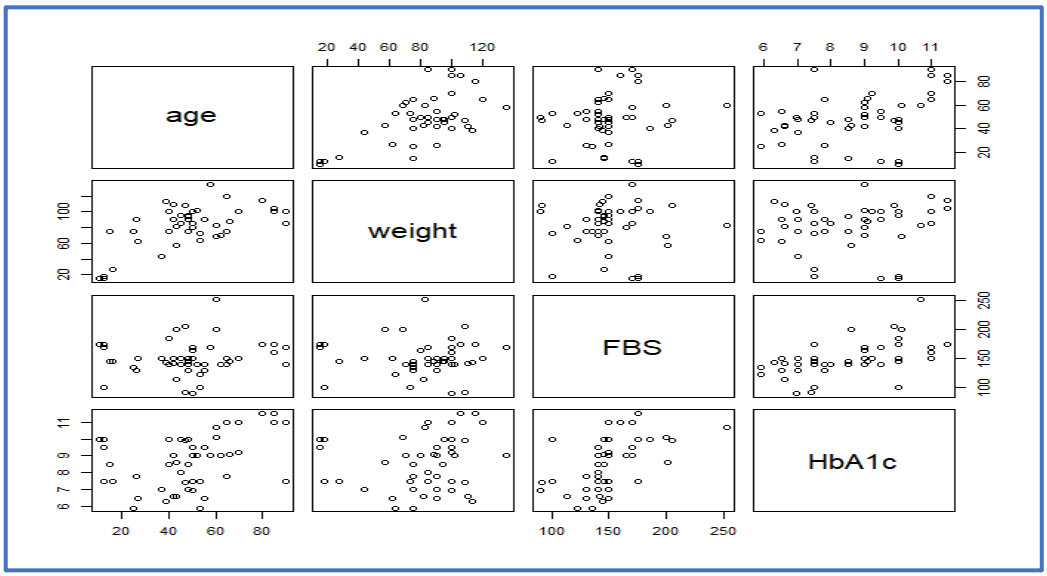
p عدد معلمات في النموذج الانحدار.

MSE هو متوسط مربع الخطأ لنموذج الانحدار.

ثانياً: الجزء العملي للدراسة:

1- التحقق من افتراضات النموذج:

أ- الشكل الانتشاري للبيانات:



الشكل رقم (3) يوضح الشكل الانتشاري للبيانات

ب- مصفوفة الارتباط

cor(suger4)

	Age	weight	FBS	HbA1c
age	1.00000000	0.6079920	0.07384877	0.4541435
weight	0.60799199	1.0000000	-0.10243324	0.2383416
FBS	0.07384877	-0.1024332	1.00000000	0.5238934
HbA1c	0.45414346	0.2383416	0.52389340	1.0000000
age	1.00000000	0.6079920	0.07384877	0.4541435

من مصفوفة الارتباط نلاحظ أنه توجد علاقة ارتباطية بين المتغيرات المستقلة طردية بين المتغير التابع والمتغيرات المستقلة.

ت- التوزيع الطبيعي للبواقي:Normality Distribution Test

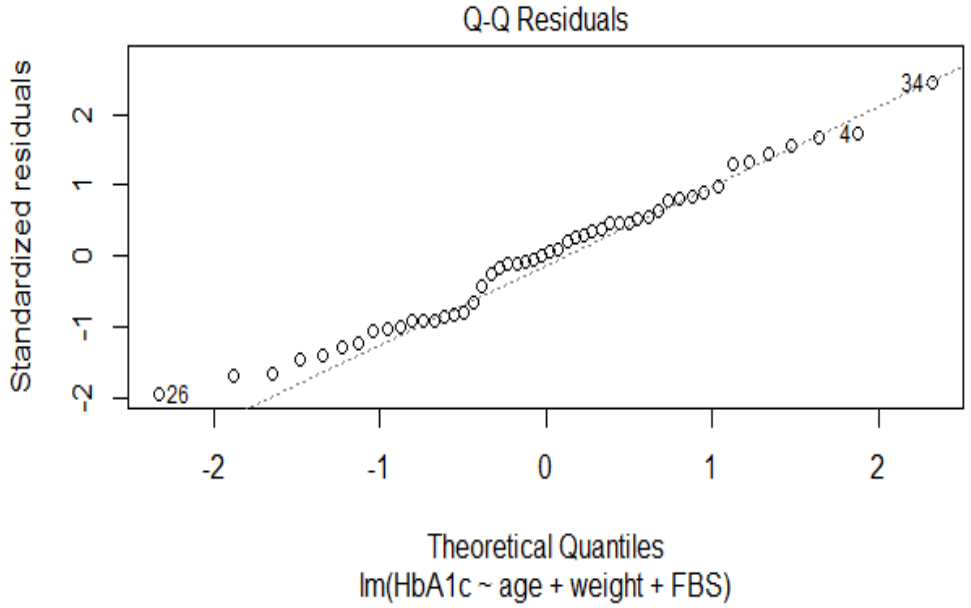
> shapiro.test(rstandard(model))

Shapiro-Wilk normality test

data: residual

W = 0.96841, p-value = 0.1992

بما أن قيمة الدلالة الإحصائية لاختبا Shapiro-Wilk تساوي (0.1992) وهي أكبر من مستوي المعنوية (0.05) فإننا نقبل فرض الصفري ونرفض الفرض البديل، أي توزيع البواقي لا يختلف عن التوزيع الطبيعي (البواقي تتبع التوزيع الطبيعي)، وكما نلاحظ من الشكل البياني رقم (5) أن البواقي قريبة من الخط المستقيم ودل ذلك على أن توزيع البواقي قريب من التوزيع الطبيعي.



الشكل رقم (4) يوضح التوزيع الطبيعي للبواقي

ث- حساب عامل تضخم التباين (VIF) *Variance Inflation Factor*

يجب أن تكون درجات VIF قريبة من 1، ولكن أقل من 5 مقبول، و10 فأكثر يشير إلى ترابط خطي مرتفع؛ لذا قد لا تكون هناك حاجة للمتغير، نلاحظ أن جميع القيم في هذا التحليل لها درجات قريبة من 1 مما يدل على عدم وجود ارتباط خطي بين المتغيرات المستقلة.

> vif(model)

age	weight	FBS
1.635006	1.643332	1.041548

ج- اختبار دوربن – واتسون Durbin - Watson Test

> dw_test <- dwtest(model)

Durbin-Watson test

data: model

DW = 2.0021, p-value = 0.5181

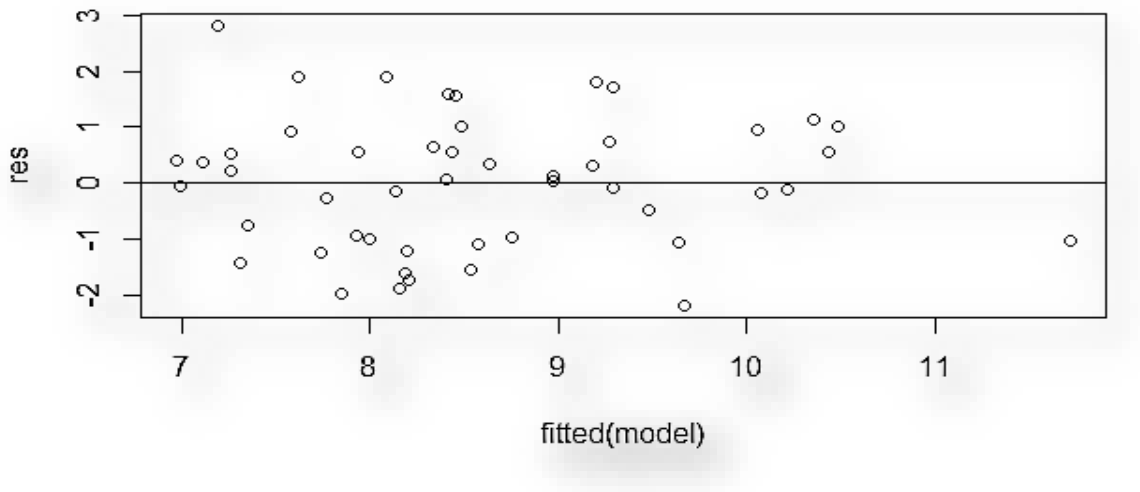
alternative hypothesis: true autocorrelation is greater than 0

إحصائية Durbin-Watson تساوي (DW = 2.0021) تشير القيمة القريبة من 2 إلى عدم وجود ارتباط ذاتي، بينما تشير القيم الأقل من 2 إلى وجود ارتباط ذاتي إيجابي وتشير القيمة القريبة من 4 إلى وجود ارتباط ذاتي سلبي.

وقيمة الدلالة الإحصائية تساوي (p-value = 0.5181) وهي أكبر من 0.05 أي أن نرفض فرض الصفري، مما يعني عدم وجود ارتباط ذاتي كبير في القيم المتبقية.

ح- القيم الشاذة (Outliers)

ح/1: باستخدام الشكل الانتشاري نلاحظ من الشكل رقم (6) أن البواقي موزعة بشكل مستطيل حيث إنها لا تزيد عن +3 أو لا تقل عن -3 أن معظم القيم مركزة في المنتصف (على امتداد نقطة الصفر) ولا تأخذ نمط الانحناء والذي يرتفع فيه طرف عن الآخر بشكل ملحوظ مما يدل عدم وجود قيم متطرف.

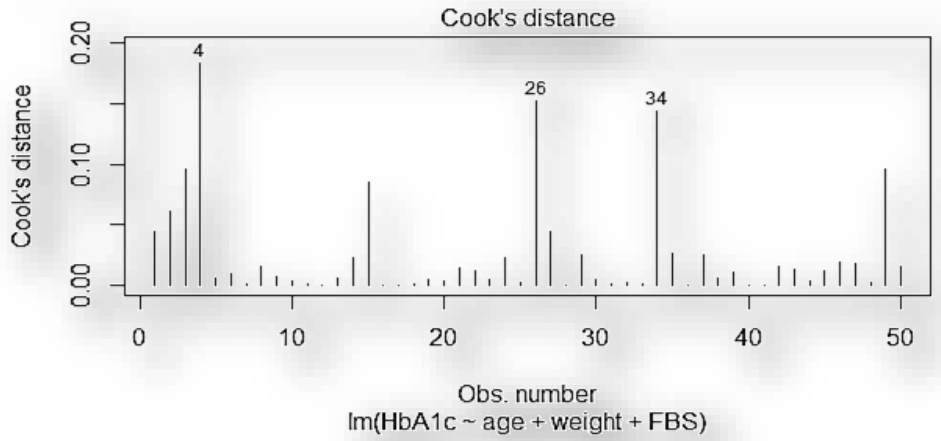


الشكل رقم (5) يوضح الشكل الانتشار للبواقي

ح/2: استخدام معامل المسافة كوك (Cook's Distance)

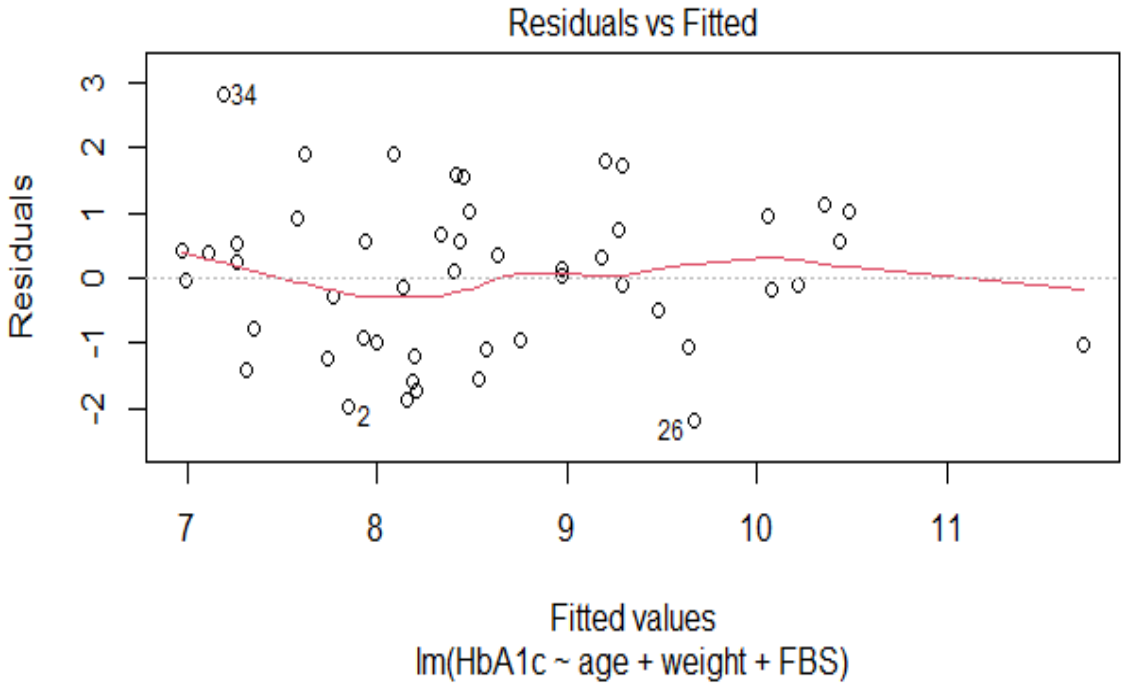
```
> cooks.distance(model)[which.max(cooks.distance(model))]
4
0.1829593
```

نلاحظ أن أكبر قيمة معامل المسافة كوك تساوي (0.1829593) لم تتجاوز الواحد الصحية حيث يمكننا القول بعدم وجود قيم متطرف وشكل رقم (7) يوضح ذلك



الشكل رقم (6) يوضح معامل المسافة كوك

مـن الشـكل (8) نلاحـظ أنـه لا يوجـد نمـط في تشـتت القـيم المقـدرة والبـواقي، عـرض التشـتت مـع زيادة القـيم المتوقعة متساوٍ تقريبًا، مما يُحقق افتراض تجانس التباين.



الشكل رقم (7) يوضح ملائمة النموذج مقابل البواقي

2- تقدير مُعلمات النموذج الانحدار الخطي المتعدد

أ. معلمات نموذج الانحدار المتعدد

$$\hat{\beta} = (x'x)^{-1}x'y$$

```
> B=solve(t(x)%*%x)%*%(t(x)%*%y)
```

	[,1]
[,1]	2.507812676
[2,]	0.032725295
[3,]	0.003798173
[4,]	0.027411779

ب. حساب مجموع مربعات الخطأ:

$$SSe = y'y - \beta' x' y$$

```
> SSE=(t(y)%*%y)-(t(B)%*%(t(x)%*%y))
```

	[,1]
[,1]	67.05372

ت. حساب مجموع مربعات الانحدار

$$SSR = \hat{\beta}' x' y - n\bar{y}^2$$

```
> SSR=(t(B)%*%(t(x)%*%y))-n*(mean(y))^2
```

	[,1]
[,1]	54.8646

ث. حساب مجموع مربعات الكلي:

$$SST = y'y - n\bar{y}^2$$

أو

$$SST = SSR + SSE$$

```
> SST=SSR+SSE
```

	[,1]
[,1]	121.9183

ج. حساب قيمة إحصاءه الاختبار F

وهي عبارة عن حاصل قسمة متوسط مجموع مربعات الانحدار MSR على متوسط مجموع مربعات الخطأ MSE.

```
> F=(SSR/(p-1))/(SSE/(n-p))
```

	[,1]
[,1]	12.54602

ح. ايجاد قيمة F الجدولية

```
> qf(p=0.05,df1=3,df2=46,lower.tail = F)
```

[,1]	] 2.806845

نلاحظ من النتائج أعلاه أن قيمة F المحسوبة تساوي 12.54602 وهي أكبر من قيمة F الجدولية عند درجتي حرية 3 و 46 بمستوى المعنوية 0.05 والتي تساوي (2.806) بناءً على ذلك نرفض فرض الصفري القائل بأن نموذج الانحدار غير معنوي وهذا يعني أن هناك تأثير معنوي من المتغيرات المستقلة على المتغير التابع

خ. حساب مصفوفة تباين وتغاير التقديرات:

> MSE*(solve(t(x)%*%x))

	[,1]	[,2]	[,3]	[,4]
[,1]	1.3283965728	5.108065e-04	-5.339495e-03	-5.849717e-03
[2,]	0.0005108065	1.432668e-04	-6.900920e-05	-1.258331e-05
[3,]	-0.0053394951	-6.900920e-05	8.633297e-05	1.054535e-05
[4,]	-0.0058497174	-1.258331e-05	1.054535e-05	3.720071e-05

أي أن تباين β_0 يساوي (1.3283965728) وبأخذ الجدر التربيعي نحصل على الخطأ المعياري لتقدير المعلمة وهو (1.152561) و β_1 يساوي (1.432668e-04) وبأخذ الجدر التربيعي نحصل على الخطأ المعياري لتقدير المعلمة وهو (0.01196941) و β_2 يساوي (8.633297e-05) وبأخذ الجدر التربيعي نحصل على الخطأ المعياري لتقدير المعلمة وهو(0.009291554) و β_3 يساوي (3.720071e-05) وبأخذ الجدر التربيعي نحصل على الخطأ المعياري لتقدير المعلمة وهو (0.006099238)

د. حساب قيمة إحصاءه الاختبار t لمعلمات النموذج.

يستخدم اختبار t في تقييم معنوية تأثير كل متغير مستقل في النموذج الخطي بشكل انفرادي عل المتغير

$$t_{col} = \hat{\beta}\imath / S.E_{\hat{\beta}}$$

> t1=(2.507813 /sqrt(1.3283965728))
[1] 2.175862
> t2=(0.032725/sqrt(1.432668e-04))
[1] 2.734053
> t3=(0.003798 /sqrt(8.633297e-05))
[1] 0.4087583
> t4=(0.027412/sqrt(3.720071e-05))
[1] 4.494332

ذ. ايجاد قيمة T الجدولية

> qt(p=0.05/2,df=46,lower.tail = F)
[1] 2.012896

نلاحـظ مـن أعـلاه أن القيمـة المطلقـة للإحصـاءات الاختبـار t4 ,t2, t1 المحسـوبة والمتمثلـة في (الحـد الثابـت وميـل متغـير العمـر وميـل متغـير مسـتوى السـكر في الـدم) علـى التـوالي أكـبر مـن القيمـة المطلقـة t الجدوليـة عنـد درجـة حريـة 48 ومسـتوى معنويـة يسـاوي 0.05 وهـي تسـاوي 2.0128 وهـذا يعـني رفـض الفـرض الصـفري وقبول الفرض البديل، أي أنه يوجد تأثير معنوي للمتغيرين عمر المريض ومستوى السكر في الدم على معدل السـكر التراكمـي بينمـا قيمـة إحصـاءه الاختبـار t3 والمتمثلـة في (ميـل وزن المـريض) أصـغر مـن قيمـة t الجدوليـة عنـد درجـة حريـة 48 ومسـتوى معنويـة يسـاوي 0.05 وهـي تسـاوي 2.0128 وهـذا يعـني قبـول الفـرض الصفري ورفض الفرض البديل أي أن لا يوجد تأثير معنوي لمتغير وزن المريض على معدل السكر التراكمي.

3- حساب معامل التحديد Determination Coefficient

نلاحـظ أن 45% مـن التغـيرات الـتي تحـدث في معـدل السـكر التراكمـي كانـت نتيجـة كـل مـن مستوى السكر في الدم والعمر والوزن للمريض أما النسبة الباقية فكانت لمتغيرات أخرى لم تدخل في نموذج الانحدار الخطي.

```
> R=SSR/SST
[1,] 0.4500112
```

4- حساب معامل التحديد المعدل أو المصحح *adjusted R squared*

```
> r=1-((1-R)*(n-1))/(n-p)
[1,] 0.4141423
```

5- اختبار مُعلمات النموذج وتفسير معنوياتها باستخدام lm function in R

```
> mode=lm(HbA1c~age+FBS+weight,data=suger4)
> summary(mode)
Call:
lm(formula = HbA1c ~ age + weight + FBS, data = suger4)
Residuals:
    Min      1Q  Median      3Q     Max
-2.17056 -1.01428  0.05881  0.71517  2.81089
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.507813   1.152561   2.176  0.03474 *
age         0.032725   0.011969   2.734  0.00885 **
weight      0.003798   0.009292   0.409  0.68460
FBS         0.027412   0.006099   4.494 4.69e-05 ***
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.207 on 46 degrees of freedom
Adjusted R-squared:  0.4141      Multiple R-squared:  0.45,
```

F-statistic: 12.55 on 3 and 46 DF, p-value: 4.034e-06

نلاحظ من النتائج أعلاه أن قيمة F المحسوبة تساوي 12.55 وقيمة الاحتمال P – value تساوي 4.034e-06 وهي أصغر من قيمة مستوى المعنوية 0.05 بناءً على ذلك نرفض الفرض الصفري القائل بأن نموذج الانحدار غير معنوي وهذا يعني أن هناك تأثير معنوي من المتغيرات المستقلة على المتغير التابع.

قيمة معامل التحديد يساوي 0.45 وهذا يعني أن مستوى السكر في الدم والعمر المريض ووزن المريض تفسر نسبة 45% من التغيرات التي تحدث على معدل السكر التراكمي للمريض، أما النسبة المتبقية والتي تساوي 55% نتيجة متغيرات أخرى لم تدخل في نموذج الانحدار الخطي المتعدد.

استنادًا إلى العوامل المؤثرة على معدل السكر التراكمي من النتائج أعلاه نلاحظ أن متغير الوزن ليس له تأثير معنوي على معدل السكر التراكمي؛ حيث إن قيمة الاحتمال P – value تساوي 0.68460 وهي أكبر من مستوى المعنوية 0.05 على ذلك نقبل الفرض الصفري القائل بأنه لا يوجد تأثير معنوي للمتغير الوزن على معدل السكر التراكمي للمرض السكري بينما يوجد تأثير معنوي للمتغير مستوى السكر في الدم على معدل السكر التراكمي؛ حيث إن قيمة الاحتمال P – value تساوي 0.0000469 وهي أصغر من مستوى المعنوية 0.05 على ذلك نرفض الفرض الصفري القائل بأنه لا يوجد تأثير معنوي للمتغير مستوى السكر في الدم على معدل السكر التراكمي للمرض السكري كما يوجد أيضا تأثير معنوي لمتغير العمر على معدل السكر التراكمي؛ حيث إن قيمة الاحتمال P – value تساوي 0.00885 وهي أصغر من مستوى المعنوية 0.05 على ذلك نرفض الفرض الصفري القائل بأنه لا يوجد تأثير معنوي للمتغير العمر على معدل السكر التراكمي للمرض السكري.

نتائج الدراسة:

1. أظهرت نتائج الدراسة عدم وجود ارتباط خطي بين المتغيرات المستقلة.
2. أظهرت نتائج الدراسة أن توزيع البواقي تتبع التوزيع الطبيعي وعدم وجود ارتباط ذاتي أي أن البواقي مستقلة ذاتيًا.
3. أظهرت نتائج الدراسة عدم وجود قيم متطرفة في البيانات الدراسة.
4. أظهرت نتائج الدراسة أن قيمة F المحسوبة تساوي 12.55وقيمة الاحتمال P – value تساوي 4.034e-06 وهي أصغر من قيمة مستوى المعنوية 0.05 في هذه الحالة نرفض فرضية العدم ونقبل الفرضية البديلة وهذا يعني أن نموذج الانحدار الخطي المتعدد معنوي.
5. أظهرت نتائج الدراسة أن 45% من التغيرات التي تحدث في معدل السكر التراكمي كانت نتيجة كل من مستوى السكر في الدم والعمر والوزن للمريض، أما النسبة للباقية فكانت لمتغيرات أخرى لم تدخل في نموذج الانحدار الخطي.
6. أظهرت نتائج الدراسة أن متغير الوزن ليس له تأثير معنوي على معدل السكر التراكمي؛ حيث إن قيمة الاحتمال P – value تساوي 0.68460 وهي أكبر من مستوى المعنوية 0.05 .

7. أظهرت نتائج الدراسة أنه يوجد تأثير معنوي للمتغير مستوى السكر في الدم على معدل السكر التراكمي؛ حيث إن قيمة الاحتمال P – value تساوي 0.0000469 وهي أصغر من مستوى المعنوية 0.05.
8. أظهرت نتائج الدراسة أنه يوجد تأثير معنوي لمتغير العمر على معدل السكر التراكمي؛ حيث إن قيمة الاحتمال P – value تساوي 0.00885 وهي أصغر من مستوى المعنوية 0.05.

المراجع:

أولاً: المراجع العربية:

1. العلي، إ. (2020). *أسس التحليل الإحصائي متعدد المتغيرات*. جامعة تشرين.
2. الراوي، ز. (2017). *طرق التحليل الإحصائي متعدد المتغيرات*. المعهد العربي للتدريب والبحوث الإحصائية.
3. صوفية، ك. (2024). كفاءة نموذج الانحدار الخطي المتعدد في تحديد العوامل المؤثرة في مرضى السكري، *مجلة العلوم الأساسية*, 37(1), 45-53. https://doi.org/10.59743/jbs.v37i1.262 .

ثانياً: المراجع الأجنبية:

1. Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics*: Use and interpretation (2nd ed.). Lawrence Erlbaum Associates.
2. Badri, S. A. I. (2025). *Correlation and regression analysis between HbA1c, FBS, and RBS levels*. Libyan Journal of Medical and Applied Sciences, 3(3), 64–71.
3. Lin, W. (2024). *Association between BMI and HbA1c*. International Journal of Environmental Research and Public Health.