AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

النمذجة الإحصائية لتقدير تأثير العوامل البيولوجية والوراثية على مرضى فقر الدم باستخدام الانحدار الخطي المتعدد

Statistical modeling to estimate the influence of biological and genetic factors on anemia patients using multiple linear regression

م .د اروى جاسم محمد حسن

وزارة التربية

Arowjassim564@gmail.com

المستخلص

تناول هذا البحث دراسة نوع من أنواع نماذج الانحدار الخطي، وهو نموذج الانحدار الخطي المتعدد الذي يعد أحد الأساليب الإحصائية الفعالة في دراسة وتحليل تأثير المتغيرات التوضيحية على المتغير التابع، اما فيما يخص مجال هذا البحث فهو مجال الأبحاث الطبية. فقد استخدمنا هذا الانموذج لتحديد العوامل التي تؤثر في تطور وشدة مرض فقر الدم حيث تمت دراسة وتحليل المتغيرات التوضيحية المتمثلة بـ (فقر الدم المنجلي، الثلاسيميا، الجنس) على المتغير التابع (متوسط حجم الهيموغلوبين) .

ومن خلال هذا الانموذج يمكن التعرف على المتغيرات التي لها تأثير فعلي وذات دلالة إحصائية ، مما يساعد الباحثين في هذا المجال لفهم الية المرض .

في الجانب التطبيقي أُستعمل البرنامج الاحصائي ssps لتحليل بيانات لـ (75) مريض مصاب بفقر الدم ماخوذة من مستشفى مدينة الطب لشهر نيسان للعام 2024 فقد تم سحب العينة بشكل عشوائي ، وقد استخدمنا اختبار (F) لمعرفة مدى صلاحية الانموذج كلاً فأظهرت النتائج ان الأنموذج قوي ومعنوي ، ومن ثم اختبار (t) لمعرفة معنوية كل معامل فقد أظهرت النتائج ان اذا كان المريض انثى وتكون غير مصابة بفقر الدم المنجلي والثلاسيميا فإن متوسط الهيموغلوبين المتوقع هو (29.80) وان الهيموغلوبين المتوقع يقل بمقدار (1.50) اذا كان المريض مصاب بفقر الدم المنجلي. أما في حال كان المريض مصاب بالثلاسيميا فإن الهيموغلوبين المتوقع يقل بمقدار (2.10) أيضا اذا كان المصاب ذكر يقل الهيموغلوبين المتوقع بمقدار (0.85)، وهذا يعني ان جميع المتغيرات التوضيحية مؤثرة معنوياً، كما واختيرت مشكلة التعدد الخطي ومشكلة الارتباط الذاتي وتوصل إلى ان الانموذج لا يعاني من مشكلتي التعدد الخطي والارتباط الذاتي.

الكلمات المفتاحية: النمذجة الإحصائية، العوامل الوراثية، العوامل البيولوجية، فقر الدم.

Abstract

the multiple linear regression ،This research examines a type of linear regression model which is an effective statistical method for studying and analyzing the impact of ،model explanatory variables on the dependent variable. The scope of this research is medical research. We used this model to identify the factors that influence the development and ،severity of anemia. We studied and analyzed the explanatory variables (sickle cell anemia and gender) on the dependent variable (mean hemoglobin level). Using this ،thalassemia helping ، we can identify variables that have a real and statistically significant impact،model the ،researchers in this field understand the mechanism of the disease. In the applied aspect statistical program SSPSS was used to analyze data for (75) patients with anemia taken from the Medical City Hospital for the month of April 2024. The sample was randomly drawn. and the results showed ،We used the F test to determine the validity of the model as a whole we used the t test to determine the ،that the model was strong and significant. Then significance of each coefficient. The results showed that if the patient was female and did the expected mean hemoglobin was ،not suffer from sickle cell anemia or thalassemia and that the expected hemoglobin decreased by (1.50) if the patient had sickle cell ،(29.80) the expected hemoglobin decreased by ، if the patient had thalassemia،anemia. However the expected hemoglobin decreased by (0.85). This ، if the patient was male،(2.10). Also means that all explanatory variables had a significant effect. We also tested the problem of and it was concluded that the model did not suffer ،multicollinearity and autocorrelation from the problems of multicollinearity and autocorrelation

anemia.، biological factors، genetic factors،**Keywords**: Statistical modeling

المقدمة

يؤدي الإحصاء دوراً حيوياً في المجال الطبي لكون الإحصائيات الطبية هي تدرس صحة الأنسان والأمراض وهناك الكثير من تطبيقاتها في البحوث الطبية الحيوية والطب السريري وعلم الأوبئة وغيرها. ويعتبر الإحصاء الأداة الأساسية لفهم وتحليل البيانات بشكل دقيق وعلمي حيث يساعد علم الإحصاء في تحديد العوامل التي تساهم في الإصابة بهذا المرض وكذلك تقييم مدى فعالية العلاجات وأيضا يساعد بالتنبؤ باحتمالية حدوث مضاعفات صحية بالاستناد إلى العوامل المختلفة بتالي وضع استراتيجيات وقائية، فالإحصاء ليس مجرد عد أو أرقام، بل هو الأداة الأساسية لتحليل وفهم المعلومات الصحية وتفسيرها بطريقة تساهم في تحسين وتعزيز صحة الأفراد والمجتمعات.

مشكلة البحث: إن التغيرات التي تحدث بمرض فقر الدم تؤثر بشكل مباشر على صحة الأنسان. فقد أشارت الأبحاث والدراسات الصادرة عن المنظمات الدولية إلى أن التغيرات التي تطرأ على مكونات الدم تلعب دوراً في زيادة أو

انخفاض متوسط حجم الهيموغلوبين، مما يؤدي بدوره إلى ظهور الكثير من الأمراض لذا من الضروري التعرف على مدى تأثير هذه العوامل على مكونات الدم.

هدف البحث : يسعى هذا البحث إلى تحديد العوامل المؤثرة على أمراض فقر الدم من خلال متوسط حجم الهيموغلوبين ، تم جمع البيانات وتحليلها إحصائيا باستعمال أنموذج الانحدار الخطي المتعدد مع اختبار معنوية الأنموذج من خلال معامل التحديد (R^2) وقيمة معنوية اختبار F .

فرضيات البحث:

1- متوسط حجم الهيموغلوبين بالدم يوثر بشكل معنوي على الإصابة بفقر الدم.
2- فقر الدم المنجلي لهُ تأثير معنوي في الإصابة بفقر الدم.
3- الثلاسيميا لها تأثير واضح مهم على الإصابة بفقر الدم.
4- الجنس يعتبر عاملاً مؤثراً في احتمالية الإصابة بفقر الدم.

الجانب النظري: سيتم في هذا الفصل شرح وتوضيح تحليل الانحدار كونه أحد الأساليب المهمة في التحليل الإحصائي، سوف نتناول مفهوم الانحدار وأنواعه، بالإضافة إلى معامل التحديد كما سيتم التطرق في هذا الفصل إلى مشكلة التعدد الخطي ومشكلة الارتباط الخطي مع مناقشة الاختبارات الإحصائية المرتبطة بهذه المشكلات.

1- مفهوم الانحدار : يشير الانحدار إلى دراسة العلاقة بين متغير تابع ومتغير أو اكثر توضيحي ، يعد تحليل الانحدار احد الموضوعات الأساسية في النظرية الإحصائية وله أهمية خاصة بسبب تعدد تطبيقاته في مختلف المجالات الإدارية والاقتصادية والطبية وغيرها . كما انه يستعمل في التعبير عن العلاقات بين المتغيرات من خلال نماذج رياضيه تعرف بنماذج الانحدار. الانحدار هو أسلوب استعمل لتحليل العلاقة بين المتغير المعتمد (التابع) على متغير أو أكثر من المتغيرات التوضيحية. يساعد هذا التحليل في التنبؤ بقيم المتغير التابع بناءً على قيم المتغيرات التوضيحية، يتم ذلك من خلال الاعتماد على معادلات إحصائية توفر معلومات دقيقة حول هذه العلاقة. وتعدُّ طريقة الانحدار أداة فعالة للكشف عن العلاقات بين المتغيرات، واستعملت بشكل واسع في مختلف العلوم لوصف هذه العلاقات رياضياً (الراوي، 1987)(Lee et al 2006).

2- أنواع نماذج الانحدار: يمكن تصنيف أنواع نماذج الانحدار بناءً على عدد المتغيرات التوضيحية. فالانحدار الذي يحتوي على متغير تابع ومتغير توضيحي واحد، يُطلق عليه (الانحدار الخطي البسيط) وهو ابسط أنواع الانحدار. أما إذا كان الانحدار يتضمن أكثر من متغير توضيحي، فيُعرف بـ (الانحدار الخطي المتعدد) (كاظم 1988،).

- **نموذج الانحدار الخطي البسيط:** نموذج الانحدار الخطي البسيط هو نموذج إحصائي يتم فيه التنبؤ بقيمة متغير تابع اعتماداً على متغير توضيحي واحد فقط. يقوم هذا النموذج بتمثيل العلاقة بين المتغيرين عن طريق خط مستقيم يتم تحديده بحيث يقلل مجموع مربعات الفروق بين القيم الفعلية والنقاط المقدرة. يتم التعبير عن النموذج الرياضي للانحدار الخطي البسيط بالمعادلة (كاظم و عيفإن ،1988):

$$Y_i = \beta_0 + \beta_1 X_i + u \qquad \dots (1)$$

حيث:

β_0: معلمة الحد الثابت (التقاطع مع المحور Y).

β_1: معلمة الميل الذي يعكس تأثير المتغير X على Y.

u: الخطأ العشوائي.

n :تمثل حجم العينة ،

$$i = 1،2، \dots ، n$$

لتقدير قيم (β_1: ،β_0) يتم استخدام طريقة المربعات الصغرى التي تعتمد على المعادلات الاتية (Lane & David، 2019) :

$$\hat{\beta}_1 = \frac{\sum XY - n(\bar{X})(\bar{Y})}{\sum X^2 - n(\bar{X})^2} \quad \dots (2)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1(\bar{X}) \quad \dots (3)$$

- **نموذج الانحدار الخطي المتعدد :** الانحدار الخطي المتعدد هو الأداة الإحصائية المتقدمة تُستخدم لدراسة العلاقة بين متغير تابع وعدة متغيرات توضيحية (X_1،X_2، ... ،X_k) . يساعد هذا النموذج في تحسين التفسيرات والنتائج المستخلصة من البيانات، خاصة عند وجود علاقات معقدة بين المتغيرات.

يُكتب النموذج الرياضي للانحدار الخطي المتعدد بالمعادلة الأتية (النصيراوي ، 2020):

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u_i \quad \dots (4)$$

حيث أن :

Y_i : المتغير التابع.

(β_0 ،β_1، .... β_k) : معلمات الانحدار.

β_0: معلمة الحد الثابت.

β_1: معلمة الميل الذي يعكس تأثير المتغير التوضيحي الأول (X_1) على Y.

β_2: معلمة الميل الذي يعكس تأثير المتغير التوضيحي الثاني (X_2) على Y.

β_k: معلمة الميل الذي يعكس تأثير المتغير التوضيحي k (X_k) على Y.

(X_1،X_2، ... ،X_k): المتغيرات التوضيحية .

u_i: الخطأ العشوائي(Madsen & Thyregod،2010) .

تُستعمل هذه الأداة لتحليل البيانات في مجالات متعددة مثل دراسة تأثير العمر أو المهارات على الأداء، أو لفهم العلاقة بين مجموعة من العوامل والمتغيرات. يسمح الانحدار المتعدد باختبار الفرضيات الإحصائية واستخلاص استنتاجات دقيقة بناءً على البيانات، ولكي نختبر معنوية الانموذج نستخدم اختبار (F) وذلك من خلال مقارنة القيمة المحسوبة المستخرجة مع القيمة المجدولة بدرجة حرية ومستوى معنوية (McCulloch & Searle،2001).

3- **معامل التحديد (R^2):** يُستعمل معامل التحديد لقياس مدى قدرة الأنموذج التقديري على تفسير الانحرافات الكلية في القيم. بمعنى آخر، يُظهر نسبة التغيرات في المتغير التابع التي يمكن تفسيرها بواسطة المتغيرات التوضيحية (X_1،X_2، ... ،X_k)، وبشكل عام هو مقياس يقيس قوة الأنموذج المقدر .تُحسب قيمة معامل التحديد بالمعادلة الأتية:

$$R^2 = \frac{\hat{\beta}^2 \sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2} \quad \dots (5)$$

إذا كانت قيمة $R^2 > 0.5$ فهذا يعني أن المتغيرات التوضيحية تفسر بشكل جيد التغيرات في المتغير التابع . أما إذا كانت أقل من ذلك، فإن المتغيرات التوضيحية لا تقدم تفسيرًا كافيًا للتغيرات في المتغير التابع، حيث ان قيمة معامل التحديد تكون محصورة $0 < R^2 < 1$

4-مشكلة التعدد الخطي : مشكلة التعدد الخطي هي واحدة من التحديات المرتبطة بنماذج الانحدار الخطي المتعدد، حيث يفترض هذا النموذج أن المتغيرات التوضيحية مستقلة عن بعضها وغير مرتبطة ببعضها ، ومع ذلك قد تنشأ حالات يرتبط فيها اثنان أو أكثر من المتغيرات التوضيحية بعلاقة قوية جدًا، سواء كانت علاقة تامة أو شبه تامة، مما يجعل من الصعب فصل أثر كل متغير على متغير الاستجابة بشكل دقيق (النصراوي ، 2020).

في الواقع العملي، غالبًا ما تظهر علاقات بين المتغيرات التوضيحية، خاصة في الدراسات الاقتصادية، نتيجة التداخل والتأثير المتبادل بين هذه المتغيرات. وفي حالة عدم وجود ارتباط بين المتغيرات التوضيحية، لا تكون هناك حاجة لاستخدام نموذج انحدار خطي متعدد، حيث يمكن استبداله بنموذج خطي بسيط يعتمد على عدد أقل من المتغيرات (k-1).

5-فرضية التعدد الخطي : في مشكلة التعدد الخطي يشترط أن لا يكون هناك علاقة ترابط خطية تامة أو شبة تامة بين المتغيرات التوضيحية المدخلة في أنموذج الانحدار . بمعنى آخر، يجب أن تكون المتغيرات التوضيحية مستقلة تماماً عن بعضها الآخر، وذلك لغرض تقدير معاملات الأنموذج بدقة، وايضاً يجب أن تكون عدد المتغيرات التوضيحية اقل من حجم العينة ورتبة مصفوفة القيم الخاصة بالمتغيرات التوضيحية اقل من حجم العينة.

ويمكن التعبير عن ذلك رياضياً بالعلاقة:

$$rank(x) = k + 1 < n \quad \dots (6)$$

حيث إن :

K: تمثل عدد المتغيرات التوضيحية.

n: تمثل حجم العينة .

6-اختبار مشكلة التعدد الخطي (اختبار فراير وكلوبر): يُعتبر اختبار فراير وكلوبر أحد أبرز الأدوات المستخدمة للكشف عن مشكلة التعدد الخطي، أو ما يُعرف بالازدواج الخطي. يتألف هذا الاختبار من ثلاثة أجزاء رئيسة:

1. **اختبار مربع كأي**: يُستعمل هذا الاختبار لتحديد ما إذا كانت هناك مشكلة تعدد خطي في الأنموذج.
2. **اختبار F**: يُهدف إلى التعرف على أي من المتغيرات التفسيرية يرتبط بعلاقة خطية أدت إلى حدوث المشكلة.
3. **اختبار t**: يُساعد في تحديد المتغيرات التفسيرية التي تُسبب مشكلة التعدد الخطي.

وفيما يلي شرح تفصيلي لكل جزء من هذه الأجزاء التحليلية.

1 -اختبار مربع كاي : أساس هذا الاختبار هو إحصاءه (x^2) والفرضية التي يتم اختبارها هي:

$H_0: X_j \ \ orthogonal$ (متعامدة)

$H_1: X_j \ Not \ orthogonal$ (غير متعامدة)

الصيغة المعتمدة لاختبار (x^2) هي:

$$x^2 = -\left[n - 1 - \frac{1}{6}(2k+5)\right][\ln|D|] \qquad \dots (7)$$

إذ أن :

n : تمثل حجم العينة.

k : تمثل عدد المتغيرات التوضيحية في الأنموذج .

D : تمثل مصفوفة معاملات الارتباط الأتية:

$$D = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \vdots & r_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix} \qquad \dots (8)$$

تقارن قيمة (x^2) المحسوبة مع القيمة (x^2) المجدولة بمستوى معنوية معينة و بدرجة حرية $(k(k-1)/2)$ ، فأذا كانت قيمة (x^2) المحسوبة اكبر من قيمة (x^2) المجدولة ترفض فرضية العدم وهذا يدل على وجود مشكلة التعدد الخطي ، وبعكسه تقبل فرضية العدم وهذا يدل على عدم وجود مشكلة التعدد الخطي .

2- اختبار F: بعد ان يتم رفض فرضية العدم باستعمال اختبار مربع كاي الذي تم التطرق اليه سابقاً ، هذا يعني ثبوت مشكلة التعدد لذا يصبح من الضروري استخدام اختبار F لغرض تحديد أي المتغيرات التوضيحية التي تساهم في هذه

المشكلة ، يعتمد هذا الاختبار على تقدير معامل التحديد المتعدد ($R^2_{j.1،2،...،k}$) $j=1،2،...،k$ بين المتغير التوضيحية (X_j) وبقية المتغيرات التوضيحية الأخرى ،يم حساب هذا المعامل وفقاً للصيغة التالية :

بافتراض وجود ثلاث متغيرات توضيحية وعندما يكون $j=1$.

$$R^2_{1.2،3} = 1 - (1 - r^2_{12})(1 - r^2_{13.2}) \quad \dots (9)$$

إذ أن

r_{12} : معامل الارتباط البسيط والمتمثل بالمعادلة الاتية:

$$r_{12} = \frac{\sum X_1 X_2 - n\bar{X}_1\bar{X}_2}{\sqrt{(\sum X_1^2 - n\bar{X}_1^2)(\sum X_2^2 - n\bar{X}_2^2)}} \quad \dots (10)$$

$r_{13.2}$:هو معامل الارتباط الجزئي للمتغيرين (X_1) و(X_3) بثبوت المتغير (X_2) والمتمثل بالمعادلة الأتية:

$$r_{13.2} = \frac{r_{13} - r_{12}r_{32}}{\sqrt{(1 - r^2_{12})(1 - r^2_{32})}} \quad \dots (11)$$

اما معادلة احصاءة الاختبار (F) تكون كالاتي:

$$F = \frac{R^2_{j.1،2،....k} \Big/ (K-1)}{(1 - R^2_{j.1،2،....k}) \Big/ (n-K)} \quad \dots (12)$$

مطلوب اختبار الفرضية الآتية:

$$H_0: R^2_{j.1،2،....k} = 0$$

$$H_1: R^2_{j.1،2،....k} \neq 0$$

تقارن قيمة (F) المحسوبة مع قيمة (F) الجدولية بدرجة الحرية $(K-1)$و$(n-K)$، فاذا كانت قيمة (F) المحسوبة اكبر من قيمة (F) الجدولية ترفض فرضية العدم وهذا يعني المتغير التوضيحي (X_j) مرتبط بعلاقة خطية مع باقي المتغيرات ، حيث يطبق هذا الاختبار على كل متغير توضيحي داخل في نموذج الانحدار .

بالإمكان استخدام معامل تضخم التباين (VIF) للكشف عن أي المتغيرات التوضيحية المسببة لمشكلة التعدد الخطي ، بعد أن يتم حساب معامل التحديد (R^2) لكل متغير توضيحي عند ثبوت باقي المتغيرات التوضيحية من المعادلة (9)، يحسب (VIF) باستخدام العلاقة الآتية :

$$\text{VIF} = \frac{1}{1 - R_j^2} \quad \ldots (13)$$

بشكل عام اذا كان $\text{VIF} > 5$ فإن ذلك يشير إلى وجود ارتباط خطي متعدد قوي ، مما يعني ان المتغير المعني قد يكون سبباً في مشكلة التعدد الخطي ، اما اذا تجاوز $\text{VIF} > 10$ فإن العديد من الباحثين يعتبرونه مؤشراً واضحاً على وجود مشكلة التعدد الخطي في الانموذج .

3- اختبار t: يستعمل هذا الاختبار في حالة رفض فرضية العدم ب اختبار (F) الذي تم ذكره سابقاً ويستعمل من اجل تحديد المتغيرات التوضيحية التي تكون مرتبطة خطياً ، وذلك من خلال اختبار قيم معاملات الارتباط الجزئي اذا كانت معنوية بين كل زوج من المتغيرات التوضيحية بثبوت باقي المتغيرات التوضيحية والتي يتم احتسابها وفق المعادلة السابقة (11)

اما معادلة احصاءة الاختبار (t_{ij}) للمتغيرين (X_i) و (X_j) تكون كالاتي:

$$t_{ij} = \frac{r_{ij.1،2،\ldots k}(\sqrt{n-K})}{\sqrt{1-(r_{ij.1،2،\ldots k})^2}} \quad \ldots (14)$$

مطلوب اختبار الفرضية الآتية:

$$H_0: r_{ij.1،2،\ldots k} = 0$$

$$H_1: r_{ij.1،2،\ldots k} \neq 0$$

تقارن قيمة (t_{ij}) المحسوبة مع القيمة المجدولة لتوزيع (t) بدرجة الحرية $(n-K)$، فاذا كانت قيمة (t_{ij}) المحسوبة اكبر من t الجدولية) ترفض فرضية العدم وهذا يعني ان معامل الارتباط الجزئي بين متغيرين بثبوت المتغيرات الأخرى ذات قيمة معنوية وان المتغيرين يكونون مرتبطين خطياً وعلى عكس ذلك تقبل فرضية العدم

6-مشكلة الارتباط الذاتي :

اختبار Durbin – Watson

اختبار (Durbin – Watson) هو اختبار يستعمل للكشف عن مشكلة الارتباط Autocorrelation بين البواقي Residuals في انموذج الانحدار الخطي ، البواقي هو الفرق بين قيم المتغير التابع الحقيقية والقيم المتوقعة ويحسب هذا الاختبار وفق الصيغة الرياضية الاتية :

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \qquad (15)$$

حيث إن :

e_t : تمثل الباقي في المشاهدة t.

n ; عدد الشاهدات .

ان قيمة هذا الاختبار تتراوح بين 0 و 4

فأذا كان DW≈ 2 هذا يعني لا يوجد ارتباط ذاتي وهو الوضع المثالي لا توجد مشكلة

واذا كان DW< 2 هذا يعني يوجد ارتباط ذاتي موجب

واذا كان DW> 2 هذا يعني يوجد ارتباط ذاتي سلبي

الجانب التطبيقي : في هذا المبحث نتعرف على العوامل الوراثية المرتبطة بمرض فقر الدم ، حيث سنستعرض العلامات والأعراض المميزة لهذا المرض ، بالإضافة إلى أسبابه . مع تحليل بيانات لـ (75) مريض مصاب بفقر الدم ماخوذه من مستشفى مدينة الطب لشهر نيسان للعام 2024. سيجرى تحليل الانحدار لمعرفة تأثير المتغيرات التوضيحية هي فقر الدم المنجلي (غير مصاب =0، مصاب =1) و الثلاسيميا (غير مصاب =0، مصاب =1) والجنس (أنثى =0، ذكر= 1) على المتغير التابع الذي يمثل متوسط حجم الهيموغلوبين. كما ستتم اختبار البيانات للكشف عن مشكلة التعدد الخطي ومشكلة الارتباط الذاتي كما تم الذكرفي الفصل الثاني، وقد وحلل باستعمال برنامج (SPSS)

التعرف على المتغيرات :

فقر الدم (Anemia) : هو حالة تتميز بانخفاض عدد كريات الدم الحمراء في الجسم . تحتوي هذه الخلايا على بروتين يسمى الهيموغلوبين، الذي يلعب دوراً اساسياً في نقل الاوكسجين من الرئتين إلى جميع انحاء الجسم. عندما يقل عدد كريات الدم الحمراء او ينخفض مستوى الهيموغلوبين، يصبح الدم غير قادر على حمل كمية كافية من الاوكسجين، مما يؤدي إلى نقص امداد الانسجة بالأوكسجين وظهور اعراض فقر الدم، ويعد فقر الدم من أكثر الامراض انتشاراً حول العالم، حيث يمكن ان يصيب الافراد من مختلف الاعمار، بدأً من الرضع والأطفال والمراهقين إلى النساء الحوامل وكبار السن، خصوصاً أولئك الذين يعانون من سوء التغذية.

الهيموغلوبين: هو بروتين غني بالحديد يلعب دوراً أساسيا في نقل الأوكسجين إلى الرئتين، يرتبط بالهيموغلوبين، ثم ينقل عبر الدم إلى مختلف الأعضاء والأنسجة. بعد ذلك يلتقط الهيموغلوبين ثاني أوكسيد الكاربون من الخلايا لإعادته إلى الرئتين حيث يتم التخلص منه عبر عملية الزفير. هذا البروتين يعد مكوناً رئيسياً لخلايا الدم الحمراء

هناك الكثير من الأسباب والعوامل التي تؤثر على فقر الدم منها العوامل الوراثية التي تم دراستها في هذا البحث المتمثلة بـ

1- فقر الدم المنجلي: هو طفرات في جين الهيموغلوبين تسبب في تكوين الكريات الدم الحمراء المشوهة والتي تكون ذات شكل منجلي تتفتت بسهولة وتسبب انسداد الاوعية الدموية .

2- الثلاسيميا : يسبب هذا المرض انتاج الهيموغلوبين بشكل غير طبيعي بتالي يسبب تكسر كريات الدم الحمراء بشكل سريع مع فقر دم مزمن .

توصيف المتغيرات: نتعرف في هذه الفقرة على المتغيرات الداخلة في الانموذج وهي :

Y: متوسط حجم الهيموغلوبين.

X_1: فقر الدم المنجلي.

X_2: الثلاسيميا .

X_3 : الجنس .

الجدول (3-1) يمثل بيانات للمرضى المصابين بفقر الدم

ت	متوسط حجم الهيموغلوبين (y)	فقر الدم المنجلي (X_1)	الثلاسيميا (X_2)	الجنس (X_3)
1	31.49	0	0	0
2	29.92	0	0	1
3	30.7	0	0	0
4	25.42	0	1	1
5	29.75	0	0	0
6	27.34	1	0	0
7	28.31	0	0	0
8	30.88	0	0	0
9	30.64	0	1	1
10	25.17	0	1	0
11	28.12	0	1	0

0	0	0	29.65	12
1	0	1	27.29	13
0	1	0	24.77	14
1	0	1	25.8	15
0	1	0	29.34	16
1	0	0	32.72	17
1	1	0	28.26	18
0	0	0	32.09	19
1	1	0	27.76	20
0	0	0	26.05	21
1	1	0	27.83	22
1	0	0	30.45	23
1	1	0	29.1	24
0	1	0	29.21	25
0	0	0	29.66	26
1	0	1	28.93	27
1	0	1	30.97	28
0	1	0	30	29
1	1	0	28.2	30
1	1	1	24.4	31
1	1	0	30.78	32
1	1	0	32.49	33

0	0	1	27.79	34
1	1	0	24.33	35
0	1	0	25.88	36
1	0	0	31.72	37
0	0	0	30.33	38
1	0	0	29.84	39
1	1	0	28.77	40
0	1	0	26.28	41
0	1	0	27.53	42
0	0	0	28.44	43
0	0	0	30.83	44
1	0	0	30.44	45
0	0	0	27.84	46
0	0	0	31.9	47
1	0	0	31.49	48
1	0	0	29.58	49
0	0	1	24.1	50
1	1	0	28.11	51
0	1	1	25	52
1	0	0	28.28	53
1	0	0	27.64	54
0	0	0	26.85	55

0	0	0	28.8	56
1	0	0	29.98	57
0	1	1	26	58
0	0	0	27.61	59
0	1	0	24.73	60
1	0	0	29.31	61
1	0	0	28.31	62
1	0	0	31.27	63
0	0	0	28.19	64
1	0	0	30.39	65
0	0	0	30.34	66
1	0	0	27.33	67
1	0	0	31.42	68
1	0	0	29.58	69
0	0	0	27.93	70
0	0	0	29.18	71
1	0	0	28.03	72
0	0	1	26.55	73
1	0	1	28.48	74
0	0	0	29.14	75

الجدول (3-2) يوضح الإحصاء الوصفي

	N	Mean	Std.Deviation
Y	75	28.9	2.3
X_1	75	0.19	0.40
X_2	75	o.36	0.48
X_3	75	0.52	0.50

الجدول أعلاه يوضح الإحصاءات الوصفية لعينة حجمها (75) و يوضح المتوسطات والانحرافات المعيارية للمتغير التابع والمتغيرات التوضيحية .

جدول (3-3) يمثل (AN0VA Table)

Model	Sum of Squares(SS)	Df	Mean Square(MS)	F	P-Value
Regression	112.3	3	37.4	6.2	< 0.05
Residual	427.4	71	6		
Total	539.7	74			

الجدول أعلاه يمثل جدول تحليل التباين (AN0VA) لاختبار معنوية أنموذج الانحدار وبما أن قيمة اختبار $F = 6.2$ وقيمة (P-Value) اقل من (0.05) وعليه نرفض فرضية العدم ونقبل الفرضية البديلة وهذا يعني أن أنموذج الانحدار معنوي جداً وان المتغيرات التوضيحية تفسر تغيرات ذات الدلالة في المتغير التابع .

أما قيمة معامل التحديد R^2 التي يُستعمل لقياس مدى قدرة الأنموذج التقديري على تفسير الانحرافات الكلية في القيم كانت تساوي R-squared = 0.212 وهذا يعني أن الأنموذج يفسر 21.2% من التباين . ،وقيمة Adjusted R-squared =0.178

جدول (3-4) يبين نتائج تحليل الانحدار المتعدد واختبار معنوية المعاملات

	$coefficients$	Std.Error	t-Statistic	P-Value
Constant	29.80	0.47	63.4	0.000

(β_0)				
X1	-1.50	0.61	-2.5	0.015
X2	-2.10	0.52	-4.0	0.000
X3	-0.85	0.38	-2.2	0.030

الجدول أعلاه يوضح قيم معاملات الانحدار و الخطأ المعياري وقيمة اختبار (t) و أيضا القيم الاحتمالية (الدلالة الإحصائية) ، حيث نلاحظ أن قيم (P-Value) للحد الثابت ولجميع المتغيرات التوضيحية هي اقل من 0.05 ($p <$ 0.05) وهذا يعني معامل الحد الثابت وجميع المتغيرات التوضيحية مؤثرة معنوياً على المتغير التابع وبالإمكان التنبؤ بالمتغير التابع من المتغيرات التوضيحية ، وعليه يمكن كتابة معادلة أنموذج الانحدار بالاعتماد على نتائج الجدول أعلاه

$$\hat{y} = 29.80 - 1.50\,X_1 - 2.10X_2 - 0.85X_3$$

اختبار التعدد الخطي (multicollinearity)

باستخدام معامل التضخم (VIF) سيتم الكشف عن وجود مشكلة التعدد الخطي ام لا بين المتغيرات التوضيحية لهذا البحث

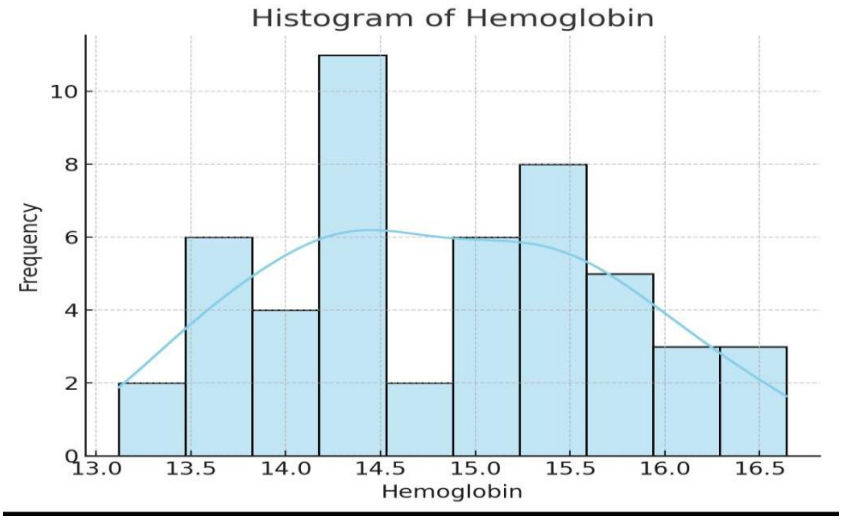
الجدول (3-5) يوضح فيم معامل التضخم (VIF) للمتعيرات التوضيحية

Coefficients	VIF
X_1	1.06
X_2	1.10
X_3	1.04

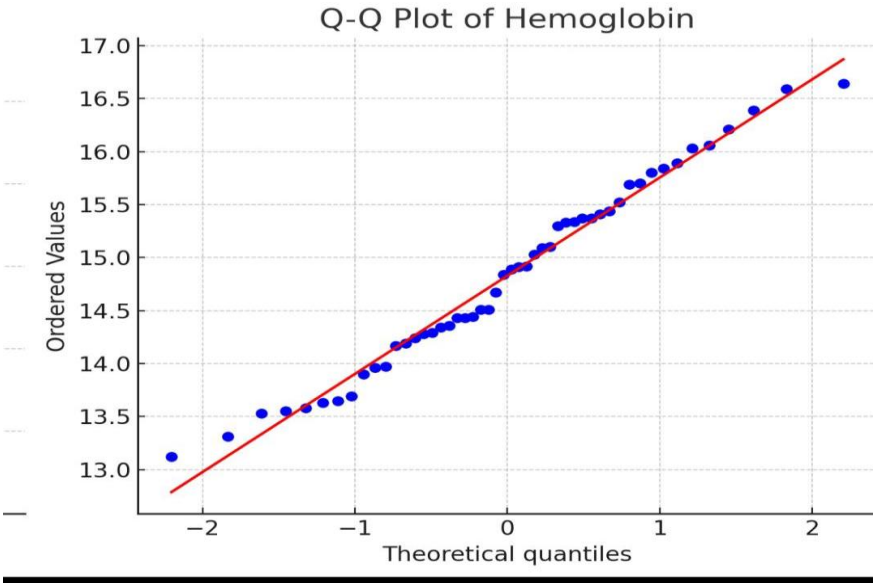
من الجدول أعلاه نلاحظ ان قيم (VIF) لجميع المتغيرات التوضيحية هي اقل من (10) وهذا يعني ان لاتوجد مشكلة التعدد الخطي

اختبار الارتباط الذاتي للبواقي (Durbin – Watson) :

قيمة الاختبار هي $2 \approx$ DW= 1.89 وهذا يعني ان البواقي مستقلة ولا توجد مشكلة الارتباط الذاتي .



الشكل (3-1) يمثل الرسم البياني التكراري للبيانات (Histogram)



الشكل (3-2) يمثل مخطط الاحتمال التجميعي

يتضح من الشكل (3-1) ان الرسم البياني التكراري للبيانات اعتدالية في توزيع البواقي، وهذا يعني ان توزيع الهيموغلوبين هو توزيع قريب جداً من التوزيع الطبيعي .

يظهر الشكل (3-2) الذي يمثل مخطط الاحتمال التجميعي توافقاً و تجمعاً جيداً للبيانات مع الخط المستقيم ، وهذا ايضاً يدل على ان البيانات تتبع التوزيع الطبيعي ، فكلما كانت النقاط قريبة من الخط المستقيم كان التوزيع طبيعي اكثر .

الاستنتاجات :-

1- ان جميع المتغيرات التوضيحية (فقر الدم المنجلي ، الثلاسيميا ، الجنس) لها تأثير معنوي وتسهم بشكل واضح في تفسير التغيرات في المتغير التابع .
2- الثلاسيميا هو العامل الأكثر أهمية في الانموذج والذي له تأثير معنوي بمقدار (2.10-) لذا يجب التركيز علىه بالتفسير التطبيقي .
3- ان جميع التأثيرات المتغيرات سالبة وهذا يعني ان كلما ارتفع احدهم يقل المتغير التابع .
4- لفقر الدم المنجلي أيضا تأثير معنوي متوسط في الانموذج وبمقدار (1.50-)، اما الجنس ايضاً له تأثير معنوي بمقدار (0.85-) لكن الأضعف من بين العوامل .
5- يشكل عام نستنتج أن أنموذج الانحدار الخطي المتعدد الذي استخدم في هذا البحث هو أنموذج قوي ومعنوي ، فقد أظهرت النتائج أن فقر الدم المنجلي والثلاسيميا والجنس هي متغيرات مؤثرة بشكل جداً واضح.

التوصيات :-

1- الأنموذج قوي لذا يمكن الاعتماد عليه مع مراقبة (VIF) .
2- نوصي بدراسة مستقبلية تضم متغيرات أخرى إلى المتغيرات التي استخدمت في هذا البحث من اجل تحسين الأنموذج أكثر.
3- نوصي بالتركيز على عنصر الثلاسيميا في التقييم والعلاج كما و نوصي بتعزيز الفحص الدوري لأنه مؤشراً مهم في التقييم لحالات فقر الدم.
4- بالإمكان استعمال هذا الأنموذج الإحصائي كأداة لمساعدة الأطباء في التنبؤات المستقبلية بالاعتماد على قراءات الثلاسيميا و فقر الدم المنجلي والجنس ، مما يزيد من دقة التشخيص وسرعته.

قائمة المصادر

1- الراوي ، خاشع محمود ، (1987) ، المدخل إلى تحليل الانحدار ، وزارة التعليم العالي والبحث العلمي ، ط1، جامعة الموصل ، العراق .
2- النصيراوي ، احمد عبد الرزاق عبد ،(2020)، تشخيص ومعالجة مشكلتي التعدد الخطي والفصل في انموذج الانحدار اللوجستي للمرضى المصابين بفقر الدم ، المجلد 12، العدد 35، مجلة الكوت للعلوم الاقتصادية والإدارية .
3- كاظم ، اموري هادي،(1988) ، طرق القياس الاقتصادي ، ط1، بغداد .
4- كاظم ، اموري هادي و عيفإن، محمد مناجد ،(1988) ، مقدمة في تحليل الانحدار الخطي ، المكتبة المركزية جامعة الموصل .
5- دليل MSDالارشادي للمستهلك لمحة عامة عن فقر الدم ، https://msdmnls.co/3xqiNxT
6- منظمة الصحة العالمية ، فقر الدم ، https://bit.ly/3VO7Vnx

7-Lane ،David M ،(2019). introduction to statistic .

Generalized Linear models with random ‹Y.(2006)‹ J.A. and Pawitan ‹ Nelder ‹J.‹ Y.‹8-Lee Chapman & Hall/CRC.‹effects

introduction to general and generalized Linear ‹ P.(2010) ‹ & Thyregod ‹H.‹9-Madsen USA.‹CRC Press‹models

John ‹ generalized Linear and Mixed models ‹S.R. (2001) ‹ and Searle ‹C.E.‹10-McCulloch Wiley & Sone /Inc.