تشخيص النموذج الملائم للانحدار بطريقتي التقليدية والمقترحة لدراسة تأثير بعض المتغيرات علي خلايا الدم البيضاء

محمد محمود فقى حسين & كاوه محمد جمال رشيد
كلية الإدارة والاقتصاد - جامعة السليمانية - العراق

الملخص: تم في هذا البحث تشخيص النموذج الملائم للانحدار بطريقتي التقليدية والمقترحة لدراسة تأثير كل من تركيب (Lymphocyte (LYM وتركيب Granulocytes (GRA) وتركيب ("Eosinophils + Basophils + Monocyte = MIC") على White (WBC) Blood Cell لان خلايا الدم البيضاء هي خلايا الجهاز المناعي وظيفتها الدفاع عن الجسم ضد كل الأمراض المعدية والمواد المثيرة للجهاز المناعي ومن خلال هذه البحث يمكن الكشف عن العوامل الذي تؤثر على خلايا دم البيضاء وكذلك تشخيص النموذج الملائم له بهدف استخدامها لأغراض التنبؤ في المستقبل. أظهرت نتائج الدراسة أن تركيبتين (MID) و(GRA) يعتبران من تراكيب مهمة في هذه الدراسة ويؤثران على خلايا دم البيضاء بشكل مباشر، وكذلك من خلال المقارنة بين الطريقتين بالاعتماد على مقاييس (MSE, R2, AIC, Schwarz's BIC) حصلنا على نتائج متطابقة ولكن الطريقة المقترحة قد تميزت بسهولة تطبيق ودقة النتائج. خلصت الدراسة إلى مجموعة من التوصيات من أهمها: ضرورة قيام الجهات العليا بتشجيع الباحثين في إجراء بحوث ودراسات أكثر في هذا المجال لان هذا الموضوع يعتبر من المواضيع المهمة له علاقة مباشرة بصحة الإنسان وكذلك استخدام طرائق الجديدة لاختيار أفضل نموذج للانحدار، ومن ثم مقارنتها مع الطرائق السابقة لمعرفة الطريقة الأكفأ لاختيار أفضل نموذج انحدار.

الكلمات المفتاحية: الاختيار الأمامي، الاختيار الخلفي، الاختيار الخطوي، الارتباط، الجذر التربيعي لمتوسط مربعات الأخطاء، معامل التحديد، متوسط مربع الخطأ، معيار اكاكي ومعيار بيز، معيار شوارز البيزي

المقدمة:

١. الجانب النظري

١.١ الطرق التقليدية لاختيار المتغيرات التتابعي Sequential Variable Selection Procedures[1][2][4][7][9]

الاختيار التتابعي للمتغيرات: طرق اختيار أفضل مجموعة جزئية تعتمد على تقييم آل المجموعات الجزئية من النموذج الكامل والتعرف على أفضل نماذج مختزلة حسب معايير معينة. حساب جميع النماذج الممكنة هي أنسب طريقة لاختيار المتغيرات. فمثلا إذا كان لدينا 8 متغيرات مستقلة $x_1, x_2, ..., x_8$ فإنه يوجد ٢٥٦= 2^8 لكي تؤخذ في الاعتبار.

١.١.١ الاختيار الأمامي: Forward Selection

١. ابدأ بنموذج يحوي الحد الثابت (القطع) فقط.
٢. اعتبر كل النماذج التي تحوي متغير واحد واختار النموذج الذي له اكبر t وذلك بمقارنتها مع قيمة معينة مسبقا t_c تسمى القطع Cutoff Value.
٣. اعتبر كل النماذج التي تحوي متغيرين وذلك بإضافة متغير مستقل جديد للنموذج الذي يحوي متغير واحد اختير حسب الخطوة السابقة واختار النموذج بمتغيرين الذي فيه وذلك طالما t المتغير الثاني له اكبر t وذلك طالما $t \geq t_c$.
٤. نستمر بهذه الطريقة (حيث الخطوة التالية اعتبار نماذج من ثلاثة متغيرات) وتتوقف العملية عندما لا يوجد متغير للإضافة له $t \geq tc$.

باستخدام هذا الخوارزم فإن عدد النماذج التي يمكن حسابها تكون على الأكثر $k(k+1)/2$ وهي اقل بكثير من 2^k إذا كانت k كبيرة. قيمة القطع التي تؤخذ عادة (t_c =٢) لأنها تعطي قاعدة شاملة لاختبار له حجم تقريبي ($\alpha = 0.05$) لأهمية متغير الانحدار وبمجرد اختيار متغير بهذه الطريقة فيجب أن يظل في النموذج.

٢.١.١ الاختيار الخلفي Backward Selection:

١. ابدأ بالنموذج الكامل الذي يحوي كل المتغيرات k المستقلة.
٢. لنعتبر إزالة متغير واحد وذلك بإقصاء المتغير الذي له أصغر t وذلك بمقارنتها مع قيمة معينة مسبقا t_c.
٣. ثم أعيد تطبيق نموذج يحوي $(k-1)$ متغير مستقل المتبقية ومن ثم نحذف المتغير الذي له أصغر t بحيث $t < t_c$.
٤. ونستمر على هذه الطريقة حتى لا يتبقى متغير يمكن حذفه له $t < t_c$

باستخدام هذا الخوارزم عدد النماذج التي يمكن تطبيقها على الأكثر k-1 نموذج وهي أقل بكثير من 2^k إذا كانت k كبيرة. مرة أخرى قيمة نقطة القطع $t_c = 2$ لنفس الأسباب المذكورة أعلاه.

٣.١.١ الاختيار الخطوي Stepwise Selection:

١. نفس الثلاثة خطوات في الاختيار الأمامي.
٢. الآن لننظر في حذف متغير له $t < t_c$ حتى لو كان أول متغير أضيف. وهذا الذي يجعل الاختيار الخطوي مختلف عن الاختيار الأمامي حيث يمكن حذف متغير من النموذج.
٣. بالنموذج الناتج في الخطوة السابقة والذي قد يحوي متغير أو اثنين نضيف متغير له اكبر $t \geq t_c$.
٤. بالنموذج الناتج في الخطوة السابقة والذي قد يحوي متغيرين أو ثلاثة نحذف أي متغير له أصغر $t < t_c$.
٥. ونستمر حتى تنتهي جميع المتغيرات المقترحة.

٢.١ الطريقة المقترحة Suggestion Method[3]

من اجل التعرف على هذه الطريقة لابد لنا من توضيح الأساس الذي تم الاعتماد عليه لاقتراح هذه
الطريقة، تعتمد هذه الطريقة على استخدام معامل الارتباط الجزئي بين المتغير المعتمد والمتغيرات
التوضيحية، ويمكن توضيح هذه الطريقة باتجاهين:

١.٢.١ الأسلوب العكسي ويتم فيه:

١. إيجاد مصفوفة الارتباط R بين المتغير المعتمد والمتغيرات التوضيحية

$$R = \begin{bmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_m} \\ . & r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_m} \\ . & . & r_{x_2x_2} & \cdots & r_{x_2x_m} \\ . & . & . & . & . \\ . & . & . & . & r_{x_mx_m} \end{bmatrix}$$

٢. إيجاد معكوس مصفوفة الارتباط R^{-1}

$$R^{-1} = \begin{bmatrix} c_{yy} & c_{yx_1} & c_{yx_2} & \cdots & c_{yx_m} \\ . & c_{x_1x_1} & c_{x_1x_2} & \cdots & c_{x_1x_m} \\ . & . & c_{x_2x_2} & \cdots & c_{x_2x_m} \\ . & . & . & . & . \\ . & . & . & . & c_{x_mx_m} \end{bmatrix}$$

٣. احتساب معامل الارتباط الجزئي للمتغير المعتمد مع كل متغير من المتغيرات التوضيحية بثبوت المتغيرات الأخرى. وبصورة عامة فان معامل الارتباط الجزئي بين المتغيرين (i و j) بعد جعل تأثير جميع المتغيرات الأخرى ثابتًا هو:

$$r_{ij}.(all\ over\ \text{var}iable) = \frac{-c_{ij}}{\sqrt{c_{ii}c_{jj}}}$$

٤. ترتيب القيم المطلقة لمعامل الارتباط الجزئي المحسوبة في الخطوة (3) ترتيبًا تنازليًا.

٥. إيجاد معادلة الانحدار باستخدام جميعا لمتغيرات وحساب قيمة متوسط مربعات الخطأ (MSE).

٦. إزالة أو الاستبعاد المتغير الذي يمتلك أقل قيمة معامل ارتباط جزئي مطلقة ومن ثم احتساب (MSE) جديدة للمعادلة التي تضم المتغيرات التوضيحية المتبقية.

٧. نقارن قيمة (MSE) للمعادلة في الخطوة (٥) مع قيمة (MSE) في الخطوة (٦) فإذا كانت قيمة متوسط مربعات الخطأ لجميع المتغيرات أقل من قيمة متوسط مربعات الخطأ بعد حذف المتغير التوضيحي الذي يمتلك أقل معامل ارتباط جزئي، فهذا دليل على معنوية جميع المتغيرات ونتوقف عن الحذف. أما إذا كانت أكبر فهذا دليل على عدم معنوية (أهمية) المتغير المحذوف. ثم نحذف المتغير التوضيحي الذي يمتلك ثاني اقل معامل ارتباط جزئي ونقارن قيمة متوسط مربعات الخطأ بعد حذف المتغير التوضيحي الثاني مع قيمة متوسط مربعات الخطأ بعد حذف المتغير التوضيحي الذي يمتلك أقل معامل ارتباط جزئي وهكذا.....

٢.٢.١ الأسلوب الأمامي وفيه يتم:

إتباع نفس خطوات الخوارزمية السابقة (الأسلوب العكسي) عدا الخطوات الآتية:

٥- تم إيجاد معادلة الانحدار لكل من المتغير المعتمد والمتغير التوضيحي الذي يمتلك اكبر معامل ارتباط جزئي مع حساب قيمة (MSE).

٦- نضيف المتغير التوضيحي والذي يمتلك ثاني اكبر معامل ارتباط جزئي مع حساب قيمة (MSE).

٧- نقارن قيمة (MSE) للمعادلة في الخطوة (5) مع قيمة (MSE) في الخطوة (٦) فإذا كانت قيمة متوسط مربعات الخطأ في الخطوة (٥) أقل من قيمة متوسط مربعات الخطأ في الخطوة (٦)، فهذا دليل على معنوية المتغير الأول وعدم معنوية المتغير المضاف ومن ثم يمكننا الجزم بعدم معنوية باقي المتغيرات ونتوقف عن إضافة المتغيرات . أما إذا كانت أكبر فهذا دليل على معنوية المتغير المضاف. ثم نضيف المتغير التوضيحي الذي يمتلك ثالث اكبر معامل ارتباط جزئي ونقارن قيمة متوسط مربعات الخطأ بعد إضافة المتغير التوضيحي الذي يمتلك ثاني اكبر معامل ارتباط جزئي مع قيمة متوسط مربعات الخطأ بعد إضافة المتغير التوضيحي الذي يمتلك ثالث اكبر معامل ارتباط جزئي وهكذا

٢. معيار اختيار أفضل نموذج: Criteria for Choice of Best Model

هناك معايير عدة لاختيار أفضل نموذج من بين عدة نماذج مقترحة لوصف بيانات معطاة في حالة الانحدار الخطي البسيط أو المتعدد وهي:

٢.١ الجذر التربيعي لمتوسط مربعات الأخطاء (RMSE)[5]

إن الجذر التربيعي لمتوسط مربعات الأخطاء هي دالة لمجموع مربعات الأخطاء SSE وعدد المشاهدات (n) وعدد المتغيرات المستقلة (k ≤p +1) حيث k تشتمل على المعلمة الثابتة في النموذج (Intercept) و يمكن كتابتها كما يلي:-

$$\mathbf{RMSE} = \sqrt{\frac{\mathbf{SSE}}{\mathbf{n-k}}} \quad \text{..... (2)}$$

حيث يتم حساب RMSE لكل النماذج، والنموذج الذي تكون فيه قيمة (RMSE) قليلة هو النموذج الأفضل. إن هذه الطريقة تعتمد على عدد المعممات في النموذج (k) لذلك فان إضافة معلمات أخرى للنموذج سوف يقل لكل من البسط والمقام انظر(Beal 2007)

٢.٢ معامل التحديد Coefficient of determination [10]

معامل التحديد ويعطي النسبة منا لتغير الكلي في البيانات والذي تم تفسيره بواسطة النموذج ويعطى بالعلاقة:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \text{..... (3)}$$

ومن البديهي القيم الكبيرة لـ R^2 توحي أن النموذج قد أسترد كثير من التغير في البيانات عن طريق الانحدار.

٣.٢ متوسط مربع الخطأ Mean Square Error [10]

وهو المقدر غير الحيازي للتباين الخطأ σ_e^2 وطبعا يختار النموذج الذي يعطي اقل متوسط مربع الخطأ.

٤.٢ معيار اكاكي و معيار بيز AIC and BIC [9][8][6]

لغرض اختيار أفضل النماذج، وكلا المعيارين يستلزمان حاصل جمع جزئين رئيسيين هما اللوغارتم الطبيعي لدالة الامكان الخاصة بالنموذج تحت الاختيار وعدد معلمات النموذج وبالشكل الآتي:

AIC= -2Ln (likelihood) + 2(P+1) ..... (4)

BIC= -2Ln (likelihood)+ (P+1) Ln (n) .....(5)

حيث n تمثل حجم العينة قيد الدراسة وأن P عدد المعلمات المقدرة.

إن القرار حول اختيار أفضل نموذج باستخدام معياري أكاكي وبيز يعتمد على الحصول على أقل قيمة لهذين المعيارين (Anderson and Burnham ٢٠٠٢).

٥.٢ معيار شوارز البيزي Schwarz's Bayesian Criterion [9][8][6]

في إطار الكلام عن بيز، فان Schwarz (1978) وصف معيار آخر سمي باسمه وهو معيار شوارز البيزي للمعلومات لاختيار النموذج الأفضل والذي يرمز لوب SBC ويمكن كتابته كما يلي:

$$\mathbf{SBC} = \mathbf{n}\ln\left(\frac{\mathbf{SSE}}{\mathbf{n}}\right)\mathbf{K}\ln\mathbf{n} \text{..... (6)}$$

إن البسط في الحد الأول لهذا المعيار هو (n lnSSE) و الذي يقل كلما ازداد عدد المتغيرات المستقلة أما مقام الحد الأول فهو ثابت (بحجمعينة n) والحد الثاني يزداد مع ازدياد عدد المعلمات K في النموذج، إذا كانت (n≥8) فان جزء المعيار ل SBC (K ln n) هي اكبر من جزء المعيار ل AIC (2K) لذا فان معيار SBC هو الأفضل لمعظم النماذج قيد الدراسة أنظر (Kutner *et al.*, ٢٠٠٤)

٣ الجانب التطبيقي:

تناول الجانب التطبيقي من البحث المقارنة بين الطريقة التقليدية وطريقة المقترحة لدراسة تأثير كل من (Lymphocyte، granulocytes، Monocyte = MIC" على "Eosinophils+ Basophils+ White Blood Cells). وقد استخدم البرنامجين الجاهزين (SYSTAT12, XLSTAT) في تطبيق هذين طريقتين (التقليدية، والمقترحة)

٣-١ تعاريف

١.١.٣ خلية الدم البيضاء أو الكرية البيضاء أو كرية الدم البيضاء[11](White Blood Cells)

هي إحدى خلايا الدم الرئيسية بالإضافة للخلية الحمراء والصفائح الدموية. الوظيفة الرئيسية للخلايا البيضاء هي الدفاع عن الجسم ضد الأمراض المعدية، وهي جزء من الجهاز المناعي. وهي خلايا الجهاز المناعي وظيفتها الدفاع عن الجسم ضد كل الأمراض المعدية والمواد المثيرة للجهاز المناعي هناك عدة أنواع مختلفة ومتنوعة من الكريات البيضاء، لكنها جميعا تتشكل من خلية جذعية متعددة القدرات في نقي العظام المعروفة باسم خلية جذعية مكونة للدم.

عدد الكريات البيضاء في الدم غالبا ما تكون مؤشرا على المرض. وهناك عادة بين ٦٠٠٠ - ٨٠٠٠ خلية دم بيضاء في ملم^3 من الدم، أي ما يقارب ١٪ من الدم عند البالغين الأصحاء. أثناء تعرض الجسم لهجوم من الأنتيجينات يرتفع هذا العدد قليلًا. في حالات مثل ابيضاض الدم (اللوكيميا) يكون عدد الكريات البيضاء أعلى من طبيعي، وفي نقص الكريات البيضاء يكون هذا العدد أقل من ذلك بكثير. الخصائص الفيزيائية للكريات البيضاء، مثل الحجم، والموصلية، والحبوبية، قد تتغير بسبب التفعيل أو بسبب وجود خلايا غير ناضجة أو خبيثة كما في إبيضاض الدم.

٢.١.٣ اللمفاويات (Lymphocyte) هي أكثر الكريات البيضاء شيوعا في الجهاز اللمفاوي. اللمفاويات تتميز بأن نواتها غامقة غير مركزية، وتحتوي كمية قليلة نسبيا من الهيولى.
الدم يحتوي ثلاثة أنماط من اللمفاويات:

- الخلايا البائية B cells: تنتج الخلايا البائية الأضداد التي ترتبط بمسببات الأمراض لتدميرها. وإضافة لوظيفة الربط بعد هجوم لمسبب مرض فإن بعض الخلايا البائية يصبح لها القدرة على إنتاج الأجسام المضادة النوعية لمسبب المرض لتكون بمثابة الذاكرة لجهاز المناعة وتسمى خلايا الذاكرة
- الخلايا التائية T cells :
- خلايا CD4 والخلايا المساعدة لها دور تنسيق الاستجابة المناعية وأهميتها في الدفاع ضد الجراثيم داخل الخلوية.
- الخلايا السمية و CD8 قادرة على قتل الخلايا المصابة بفيروس والخلايا الورمية.
- الخلايا القاتلة الطبيعية Natural killer cells: هي قادرة على قتل خلايا الجسم التي ترسل إشارات عند إصابتها من قبل فيروس أو عندما تصبح سرطانية.

٣.١.٣ كريات الدم البيضاء المحببة Granulocytes-WBC
(سيتوبلازمها يحتوي على حبيبات من الأجسام الحالة) أو المتعددة أشكال النواة (أي تحتوى على عدة فلقات التي تفصلها الاختناقات عن بعضها البعض) Granulocytes-WBCs or Polymorph nuclear-WBCs وهي:-

- خلايا متعادلة Neutrophils تحتوي من ٣-٥ فلقات
- خلايا حامضية Eosinophils تحتوي ٢-٣ فلقات
- خلايا قاعدية Basophiles نواتها تشغل نصف حجم كرية وهي تشبه بتشكلها حرفي s و M.

٤.١.٣ الخلية وحيدة النواة (Monocyte) نواتها لها شكل الكلية وتملك هيولى غزيرة. وظيفتها تشبه وظيفة "المكنسة الكهربائية" للخلايا العدلات، ولكن عمرها أطول بكثير لأن لها دور إضافي. حيث تقدم مسببات الأمراض إلى الخلايا ت ليتم تشكيل الجسم المضاد أو حتى يتم تذكر مسبب الأمراض هذا مرة أخرى عند تعرض الجسم له. كما يمكن للخلايا وحيدة النواة مغادرة مجرى الدم إلى الأنسجة لتتحول إلى خلية بالعة لتزيل حطام الخلايا الميتة فضلا عن مهاجمة الأحياء الدقيقة الممرضة.

- إن أيا من هذين الوظيفتين لا يمكن القيام بها من قبل العدلات بنفس كفاءة وحيدات النواة. وخلافا للعدلات، فإن وحيدات النواة قادرة على تعويض جسيماتها الحالة يعد استعمالها ولهذا السبب يعتقد أن عمرها أطول من عمر العدلات.

٥.١.٣ الأسسة (Basophil) هي المسئولة بالدرجة الأولى للاستجابة للحساسية وللمستضد عن طريق إفراز مواد كيميائية كالهستامين

٦.١.٣ الحمضة أو الخلية الحمضة (Eosinophil) أحد أنواع الكريضات. يبرز دورها الدفاعي عند الإصابات بالطفيليات.

١- الطريقة التقليدية (الانحدار المتدرج) Stepwise Method

باستخدام البيانات والخطوات الخمسة في الجانب النظري من الممكن بناء النموذج الانحدار المتدرج (التقليدي) وكما هو موضح أدناه:-

Multiple Regression Analysis

Dependent variable: WBC

Parameter	Estimate	Error	Standard Statistic	T P-Value
CONSTANT	4.29966	1.36957	3.13942	0.0030
LYM	-0.324816	0.262393	-1.2379	0.2220
GRA	-1.38992	0.542533	-2.56192	0.0138
MID	-0.598107	0.184503	-3.24172	0.0022

The Model:

$$WBC = 4.29966 - 1.38992GRA - 0.598107MID$$

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	0.195168	3	0.0650559	4.30	0.0094
Residual	0.696681	46	0.0151452		
Total (Corr.)	0.891849	49			

R-squared = 21.8835 percent
R-squared (adjusted for d.f.) = 16.7889 percent
Standard Error of Est. = 0.123066
Mean absolute error = 0.0923097
Durbin-Watson statistic = 1.6678

Stepwise regression

Method: forward selection
F-to-enter: 1.0 : F-to-remove: 1.0
Step 0:
---
0 variables in the model. 49 d.f. for error.
R-squared = 0.00% Adjusted R-squared = 0.00% MSE = 0.018201
Step 1:

Adding variable MID with F-to-enter = 2.27815
1 variables in the model. 48 d.f. for error.
R-squared = 4.53% Adjusted R-squared = 2.54% MSE = 0.0177383
Step 2:

Adding variable GRA with F-to-enter = 8.58851
2 variables in the model. 47 d.f. for error.
R-squared = 19.28% Adjusted R-squared = 15.85% MSE = 0.0153168
Step 3:

Adding variable LYM with F-to-enter = 1.5324
3 variables in the model. 46 d.f. for error.
R-squared = 21.88% Adjusted R-squared = 16.79% MSE = 0.0151452

ويمكن تلخيص خطوات الانحدار التقليدي (المتدرج) Stepwise في الجدول أدناه:-

جدول (١): يوضح تلخيص الطريقة التقليدية (الانحدار المتدرج) Method Stepwise

المقاييس				P-Value	F-ratio	المتغيرات
Schwarz's BIC	AIC	R^2	MSE			
-50.013	-55.749	4.53	0.0177	0.138	2.278	MID
-54.492	-62.140	19.3	0.015	0.007	5.613	MID,GRA

نلاحظ من خلال الجدول أعلاه و بالاعتماد على تحليل خطوات الانحدار التقليدي(الانحدار المتدرج) نرى بان المتغير (LYM) غير معنوي من ناحية الإحصائية ويمكن حذفه من النموذج.

٢- الطريقة المقترحة Suggestion Method

١. نقوم بإيجاد مصفوفة الارتباط بين متغيرات الدراسة Pearson Correlation Matrix و كما هو موضح في الجدول أدناه:-

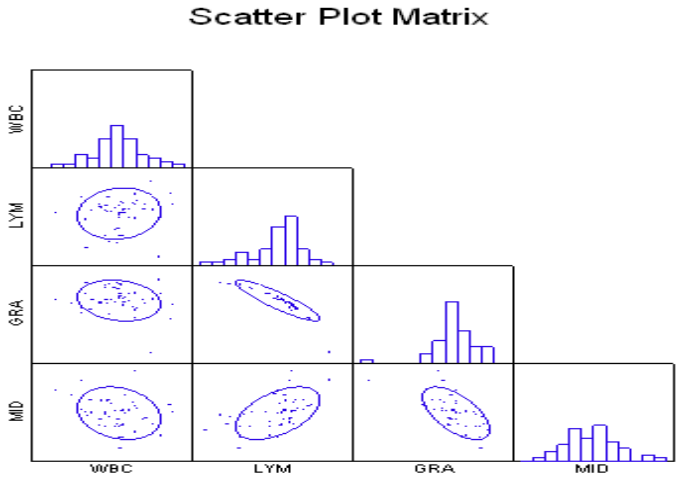
جدول (٢): يوضح مصفوفة الارتباط بين متغيرات الدراسة

	WBC	LYM	GRA	MID
WBC	1			
LYM	0.1044	1		
GRA	-0.1691	-0.8986	1	
MID	-0.2129	0.5138	-0.6204	1

٢. نقوم بإيجاد معكوس مصفوفة الارتباط بين متغيرات الدراسة Inverse of Person Correlation Matrix وكما هو موضح في الجدول أدناه:-

جدول (٣): يوضح معكوس مصفوفة الارتباط بين متغيرات الدراسة

	WBC	LYM	GRA	MID
WBC	1.2800	0.4739	1.0735	0.695
LYM		5.4549	5.3742	0.6323
GRA			7.2174	1.945
MID				2.0298



شكل (١): يوضح معاملات الارتباط بين متغيرات الدراسة

٣. احتساب معامل الارتباط الجزئي للمتغير المعتمد مع كل متغير من المتغيرات المستقلة كما هو موضح في الجدول أدناه:-

جدول (٤): يوضح قيم معامل الارتباط الجزئي بين متغيرات الدراسة

العلاقة	معامل الارتباط الجزئي	قيم معامل الارتباط الجزئي
$r_{WBC,LYM}$	$(0.44739)/\sqrt{(1.2800)\times(5.4549)}$	0.179345=
$r_{WBC,GRA}$	$(1.0735)/\sqrt{(1.2800)\times(7.2174)}$	0.353189=
$r_{WBC,MID}$	$(0.695)/\sqrt{(1.2800)\times(2.0298)}$	0.431175=

٤. نرتب قيم معامل الارتباط الجزئي ترتيبا تنازليا بعد اخذ القيمة المطلقة فيكون لدينا التسلسل الآتي:

جدول (٥): يوضح تسلسل قيم معامل الارتباط الجزئي للمتغيرات الدراسة

متغيرات الدراسة	قيم المتغيرات
MID	0.431175
GRA	0.353189
LYM	0.179345

٥. نوجد نموذج الانحدار للمتغير المعتمد و جميع المتغيرات التوضيحة ثم نحسب قيم المقاييس (MSE, R^2, AIC, Schwarz's BIC)، استبعاد المتغير الذي يمتلك اقل معامل ارتباط جزئي ثم نوجد قيم المقاييس (MSE, R^2, AIC, Schwarz's BIC) إذا كان قيم (MSE, R^2, AIC, BIC) هذا المتغير اقل من قيم (MSE, R^2, AIC, Schwarz's BIC) لجميع المتغيرات هذا يعني أن المتغير غير معنوي و يمكن حذفه من النموذج و هكذا نستمر، الجدول أدناه يوضح طريقة المقترحة

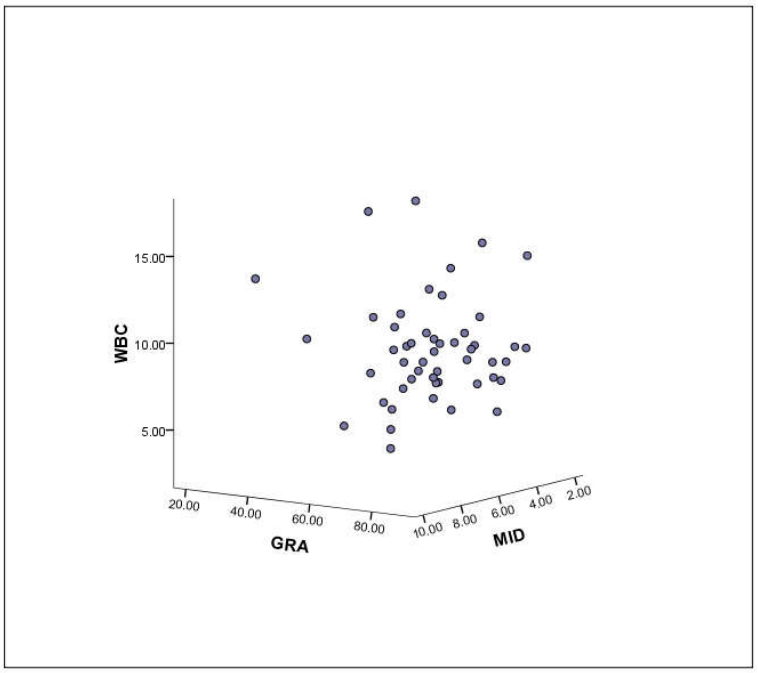
جدول (٦): يوضح الطريقة المقترحة Suggestion Method

المتغيرات	F-ratio	P-Value	المقاييس			
			MSE	Squared Multiple R	AIC	Schwarz's BIC
LYM,GRA,MID	4.295	0.009	0.015	21.9	-61.779	-52.219
GRA,MID	5.613	0.007	0.015	19.3	-62.140	-54.492

جدول (٧): يوضح تقدير معلمات النموذج

Regression Coefficients B = $(X'X)^{-1}X'Y$

التأثير Effect	المعامل Coefficients	الخطأ المعياري Standard Error	معامل المعياري Standard Coefficient	القبول Tolerance	قيمة اختبار t-test	القيمة الاحتمالية P-Value
CONSTANT	2.761	0.578	0.000	.	4.777	0.000
GRA	-0.811	0.277	-0.490	0.615	-2.931	0.005
MID	-0.569	0.184	-0.517	0.615	-3.092	0.003



شكل (٢): يوضح العلاقة بين متغير (GRA) و (MID) مع (WBC)

٤. الاستنتاجات والتوصيات:

١.٤ الاستنتاجات

١- من خلال استخدام الطريقتين المقترحة و الطريقة التقليدية حصلنا على نفس النتائج بالنسبة للمقاييس المستخدمة.

٢- من خلال هذه الدراسة أن تركيبتين (GRA)و (MID) هما من تراكيب المهمة يؤثران بشكل مباشر على (WBC)

٣- إن طريقة مقترحة لاختيار أفضل نموذج للانحدار ذات كفاءة عالية.

٤- إن طريقة مقترحة تعتبر طريقة مفضلة لان استخدامها سهل و بسيط حتى اذا كانت عدد متغيرات كثيرة.

٥- إمكانية استخدام الطريقة المقترحة لاختيار أفضل نموذج للانحدار بغض النظر عن طبيعة البيانات

٢.٤ التوصيات

١- ضرورة إجراء دراسات و بحوث أكثر في هذا المجال.

٢- استخدام طرائق الجديدة لاختيار أفضل نموذج الانحدار، ومن ثم مقارنتها مع الطرائق السابقة لمعرفة الطريقة الأكفأ لاختيار أفضل نموذج انحدار.

المصادر والمراجع

المصادر العربية:

[1] بري ، د.عدنان ماجد عبدالرحمن: (٢٠١١)،"تحليل الانحدار الخطي"، جامعة الملك سعود، قسم الإحصاء وبحوث العمليات

[2] كاظم، د. اموري هادي :(٢٠٠٢) ،" تحليل الانحدار "،وزارة التعليم العالي والبحث العلمي "، بغداد.

[٣] محمود ، غزوان هاني :(٢٠١١) ، "طريقة مقترحة لاختيار أفضل معادلة انحدار بالاعتماد على معامل الارتباط الجزئي"، المجلة العراقية للعلوم الإحصائية (٢٠)، عدد خاص بوقائع المؤتمر العلمي الرابع كلية علوم الحاسوب والرياضيات، ص ص [362-372]

[4] شيراز ، د. محمد صالح: (٢٠٠٩)، " التحليل الإحصائي للبيانات باستخدام برنامج المجموعة الإحصائية للعلوم الاجتماعية SPSS"

المصادر الانجليزية:

[5] Beal, D.J., (2007). Information Criteria Methods in SAS for Multiple Linear Regression Models. Proceedings of the Fifteenth Annual Conference of the South East SAS Users Group, Hilton Head, SC.

[6] Burnham, K.P., and Anderson, D.R. (2002). Model selection and multimodel inference: A practical information-theoretic approach, 2nd.-ed. Springer-Verlag, New York.

[7] Draper, N.R., and Smith, H. (1981). Applied Regression Analysis.2nd.-ed. Wiley, New York.

[8] Kundu, D., and Murali, G., (1996). Model selection in linear regression. Computational Statist & Data Analysis 22, pp. 461-469.

[9] Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2004). Applied linear statistical models. 4th.-ed. McGraw Hill, Chicago, IL.

[10] Stauffer, H.B., (2008). Contemporary Bayesian and Frequentist Statistical research methods for natural resource scientists. Wiley, New York.

مصادر من الانترنيت:

[11]http://ar.wikipedia.org/wiki/%D8%AE%D9%84%D9%8A%D8%A9_%D8%A7%D9%84%D8%AF%D9%85_%D8%A7%D9%84%D8%A8%D9%8A%D8%B6%D8%A7%D8%A1

Diagnosing the Appropriate Model for Regression Using the "Traditional" and "Suggested" Methods to Study the Effect of some Variables on White Blood Cells

Mohamed Mahmoud Faki Hussein and Kaveh Mohamed Gamal Rashid
College of Administration and Economics - University of Sulaymaniyah - Iraq

Received: *17/12/2021*

Abstract: In this paper diagnosis appropriate model for the regression by using two models: traditional and suggested to study the effect of the compositions of each Lymphocyte (LYM), Granulocytes (GRA) and ("Monocyte + Basophils + Eosinophils = MIC") on White Blood Cell (WBC) because white blood cells are cells of the immune system function defend the body against both infectious diseases and exciting material of the immune system and through this search can detect the factor that affect the white blood cells, as well as his diagnosis appropriate model to be used for forecasting purposes in the future. The results showed that the two structures (MID) and (GRA) are considered to be the two structures that are important in this study and affect the white blood cells directly, as well as through the comparison between the two methods based on metrics (MSE, R2, AIC, Schwarz's BIC) got the results identical, but the suggested method could easily be characterized by the application and the accuracy of the results. The study concluded a series of recommendations including the need for the higher authorities to encourage researchers to conduct research and studies more in this area because this topic is one of the important topics directly related to human health, as well as the use of methods new to choose the best model of the slope, and then compare it with the previous methods to learn the most efficient way to choose the best regression model.

Keywords: Forward Selection, Backward Selection, Stepwise Selection, Correlation, RMSE, Coefficient of determination, Mean Square Error, AIC and BIC, Schwarz's Bayesian Criterion