مقارنة بين أسلوبي انحدار الترجيح اللوغاريتمي الثنائي وتحليل الدالة التمييزية في تحديد المتغيرات التمييزية وقوة التنبؤ بالأداء الأكاديمي بجامعة أم القرى

إعـــــداد

د/ محمد موسى الشمراني

أستاذ مشارك – كلية التربية جامعة الطائف

مقارنة بين أسلوبي انحدار الترجيح اللوغاريتمي الثنائي وتحليل الدالة التمييزية في تحديد المتغيرات التمييزية وقوة التنبؤ بالأداء الأكاديمي بجامعة أم القرى

د/ محمد موسى الشمراني *

المقدمة:

تهتم الكثير من الدراسات البحثية التطبيقية عند دراسة ظاهرة معينة بتحديد المتغيرات المؤثرة في حدوثها والتنبؤ بها وبخاصة عندما يكون المتغير التابع المتنبأ به ثنائياً Binary ففي مثل هذه الحالات نسعى إلى استخدام أساليب تنبؤية لها القدرة على التنبؤ بتصنيف الحالات وتمييزها. ويشير Wilson & Hardgrave (1995) إلى أن تصنيف الطلبة في برامج الدراسات الجامعية وتوقع لنجاح الأكاديمي من المشكلات المهمة التي تتطلب استخدام أساليب إحصائية متقدمة. فقد استخدم أسلوب تحليل الانحدار الخطي التقليدي في قياس الأداء الأكاديمي ولكن لم تكن نتائج التنبؤ بالأداء في المستوى المطلوب من الدقة.

ويعد أسلوب انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية من الأساليب الإحصائية متعددة المتغيرات واسعة الاستخدام عند تحليل البيانات للمتغيرات التابعة التصنيفية، وكلاهما مناسب لنماذج التصنيف الخطية، وتختلف الطريقتان في فكرتهما الأساسية، فالأسلوب الأول اللوجستي لا يتطلب افتراضات تتعلق بتوزيع البيانات، بينما يتطلب الأسلوب الثاني تحقق افتراضات أساسية مثل افتراض التوزيع الاعتدالي، ولكنهما يتوصلان إلى نتائج لتصنيف المشاهدات (Pohar et al., 2004).

وعلى الرغم من وجود العديد من الدراسات السابقة التي اهتمت بالمقارنة بين الأسلوبين إلا أنه مازال الخلاف قائماً بين أفضلية كل منهما، فمثلاً نجد أن دراسة (Pohar et al. 2004)، و(Ebrahimmzadeh et al ., 2015) وكذلك دراسة (Santos et al. ,2014) أوصت بأفضلية أسلوب انحدار الترجيح اللوغاريتمي، بينما أكدت دراسات أخرى أفضلية تحليل الدالة التمييزية مثل دراسة

* **د/ محمد موسى الشمراني:** أستاذ مشارك – كلية التربية جامعة الطائف.

(Kardiyen & Olmus ,2016)، ودراسة (Maroco et al., 2011). وفي حين أن هناك دراسات أخرى أوضحت عدم اختلافهما من حيث الكفاءة Ruiz, (1991) ودراسة (Adler et al., 2011).

لذلك تهدف الدراسة لتناول هذه المشكلة وتوضيح كيفية استخدام كل من هذين الأسلوبين في تحليل البيانات، وإظهار الفروق في طريقة اختيار أيهما أفضل، والافتراضات التي ينبغي تحققها في البيانات، وفي تقدير الأوزان، وتفسير النتائج، وفحص ملائمتها لتعرف أيهما أفضل في التصنيف والتنبؤ بالأداء الأكاديمي ودقة ذلك.

مشكلة الدراسة وتساؤلاتها:

تكمن مشكلة الدراسة الحالية في الإجابة على التساؤلات التالية:

1. ما نمط مجموعة المنبئات التي تميز بدرجة أفضل بين الأداء الأكاديمي المرتفع والمنخفض باستخدام كل من أسلوب انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية؟
2. ما مدى دقة كل من أسلوب انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية في تصنيف المجموعتين؟

أهداف الدراسة:

تهدف الدراسة الحالية إلى المقارنة بين استخدام أسلوب انحدار الترجيح اللوغارتيمي وتحليل الدالة التمييزية في تحديد مجموعة المنبئات التي تميز بدرجة أفضل بين الأداء الأكاديمي المرتفع والمنخفض ودقة تصنيف المجموعتين.

أهمية الدراسة:

تتبع أهمية الدراسة الحالية من تناولها للمقارنة بين أسلوبي انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية في تحديد المتغيرات التمييزية والتنبؤ بالأداء الأكاديمي؛ كما يأمل الباحث أن تفيد نتائج هذه الدراسة في المجالات التالية:

1. **الأهمية النظرية:** من خلال توضيح المقارنة بين انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية.
2. **الأهمية التطبيقية:** تحسين الممارسة من خلال تقديم المساعدة للباحثين في تحديد الأسلوب الأمثل للتصنيف.

مصطلحات الدراسة وتعريفاتها الإجرائية:

الدالة التمييزية Discriminant Function: هي دوال إحصائية عددها يساوي عدد مستويات المتغير مطروحاً منه الواحد الصحيح أو عدد المتغيرات المستقلة، أيهما أقل (Hair et al, 2006)

الدرجة التمييزية The Discriminant Score: هي القيمة الناتجة من التعـويض في الدالة التمييزية باستخدام بيانات حالة ما (Huberty , 1994).

انحدار الترجيح اللوغارتيمي Logistic Regression: أسلوب إحصائي يستخدم لتحديد العلاقة بين متغير تابع من المستوى الاسمي ومتغير أو أكثر من المتغيرات المستقلة (Hosmer et al, 2013).

الأداء الأكاديمي Academic Performance: معدل الطالب أو الطالبة في الفصل الأول بعد السنة التحضيرية.

اختبار القدرات العامة: هو اختبار يقدم لطلبة المرحلة الثانوية لقياس القدرة على الاستدلال والتحليل، ويتكون من جزئين، أحدهما كمي والآخر لفظي (الزامل، 2012).

الاختبار التحصيلي Achievement Test: هو اختبار يقدم لطلبة الثانوية يغطي المفاهيم العامة في مواد الرياضيات، والفيزياء، والأحياء، والكيمياء، في مقررات المرحلة الثانوية (الزامل، 2012).

معدل السنة التحضيرية: مجموع النقاط التي حصل عليها الطالب، مقسوماً على مجموع الوحدات المقررة التي درسها الطالب في برنامج السنة التحضيرية.

حدود الدراسة:

اقتصر البحث الحالي على دراسة استخدام انحدار الترجيح اللوغارتيمي وتحليل الدالة التمييزية في تحليل بيانات طلبة السنة التحضيرية الجامعية.

الحدود المكانية: طلاب وطالبات المرحلة الجامعية الذين اجتازوا السنة التحضيرية والذين تم تسكينهم في تخصصات علمية في جامعة أم القرى بمكة المكرمة.

الحدود الزمانية: الطلاب والطالبات الذين اجتازوا السنة التحضيرية وتم تسكينهم فيما بعد في تخصصاتهم والانتهاء من دراسة الفصل الأول في التخصص للعام الدراسي ١٤٣٤/١٤٣٥هـ.

أدبيات الدراسة:

أولاً- الإطار النظري:

انحدار الترجيح اللوغاريتمي Logistic Regression: يهدف لإيجاد أفضل نموذج يصنف العلاقة بين متغير تابع (dependent or response variable) ومجموعة منبئات (predictor or explanatory) ويتميز هذا الأسلوب بالقوية والمرونة النسبية وسهولة الاستخدام، ولا يتطلب افتراضات تتعلق بالبيانات Pohar et al. (2004)، ويماثل الانحدار المتعدد، ويعد نوعاً خاصاً من أنواع الانحدار للتنبؤ بمتغير ثنائي لمجموعتين Hair et al.(2006).

صيغة انحدار الترجيح اللوغاريتمي: بافتراض x متغير مستقل كمي وY متغير تابع ثنائي، $p(x)$ احتمال النجاح للقيمة x، فإن صيغة انحدار الترجيح اللوغاريتمي تكون كالآتي:

$$\log(\frac{p(x)}{1-p(x)}) = a + bx$$

حيث $p(x) = \frac{e^{a+bx}}{1+e^{a+bx}}$ ، a ، b معالم يتم تقديرها، x متغير مستقل، على هذا الأسلوب أو النموذج الترجيح اللوغارتيمي (Agrestc, 2007).

فروض انحدار الترجيح اللوغاريتمي: يذكر Narth & Jackson (1992) أن انحدار الترجيح اللوغارتيمي أسلوب أو طريقة تصنيفية تستخدم على نطاق واسع ولا تتطلب أن يكون توزيع درجات العينة اعتدالياً ولا يحقق تجانس مصفوفة التباين والتغاير للمجموعات. ويؤكد Tabachnick & Fidell (2013) أن انحدار الترجيح اللوغاريتمي أكثر مرونة من غيره من الأساليب، مثل: تحليل الدالة التمييزية، فإن انحدار الترجيح اللوغارتيمي لا يتطلب افتراضات تتعلق بتوزيع المتغيرات.

نسبة الترجيح The Odds Ratio: إذا كان احتمال النجاح p تكون قيم الترجيح $odds = \frac{p}{1-p}$ ، فإذا كان $p = 0.75$ فإن قيمة ترجيح النجاح تساوي

$Odds = \frac{0.75}{1-0.75} = 3$ ، كما يتم تقدير احتمال النجاح بمعرفة دالة الترجيح $p = \frac{odds}{1+odds}$ ، فإذا كان الترجيح $odds = 4$ فإن $p = \frac{4}{1+4} = 0.80$.

وبالتالي إذا كان جدول البيانات من النوع 2×2، وكان ترجيح النجاح $odds_1 = \frac{p_1}{1-p_1}$ للصف الأول، وترجيح النجاح $odds_2 = \frac{p_2}{1-p_2}$ للصف الثاني، فإن نسبة الترجيح تساوي $OddsRatio = \frac{odds_1}{odds_2} = \frac{p_1/1-p_1}{p_2/1-p_2}$ ، وتسمى نسبة الترجيح odds ratio، وهي لا تتخذ قيماً سالبة (Agresti, 2007).

طرائق انحدار الترجيح اللوغارتيمي: هناك ثلاثة طرائق لانحدار الترجيح اللوغارتيمي Tabachnick & Fidell (2013) كالآتي:

١. **انحدار الترجيح اللوغاريتمي المباشر** Direct Logistic Regression يتم في هذه الطريقة إدخال جميع المتغيرات في المعادلة في وقت واحد، كما في الانحدار المتعدد وتحليل الدالة التمييزية.

٢. **انحدار الترجيح اللوغاريتمي التسلسلي** Sequential Logistic Regression يتم إدخال المتنبأت واحداً أو أكثر تلو الآخر.

٣. **انحدار الترجيح اللوغاريتمي التدريجي** Stepwise Logistic Regression يتم إدخال المنبئات وتحريكها أو حذفها لتكوين معادلة الانحدار وفقاً لمحكات إحصائية ولهذه الطريقة نفس مشكلات تحليل الانحدار التدرجي وتحليل الدالة التمييزية التدرجي.

تقييم مطابقة النموذج: يمكن التحقق من مطابقة النموذج من خلال ما يأتي:

1) **البواقي** Residual يتم التأكد من وجود دلالة إحصائية للمتغيرات المستقلة في النموذج ومدى ارتباطها بالمتغير التابع باستخدام الصيغ.
(Kutner et al., 2005)

$$Pearson\ \mathrm{Re}sidual = e_i = \frac{y_i - n_i\hat{p}_i}{\sqrt{(n_i\hat{p}_i(1-\hat{p}_i)}}$$

$$Stndardized\ \text{Re}sidual = e_i = \frac{y_i - n_i\hat{p}_i}{SE}$$

حيث $SE = \sqrt{(n_i\hat{p}_i(1-\hat{p}_i)(1-h_{ii})}$، n_i عدد الحالات، y_i قيم الخطأ المقدرة ، $\hat{p}_i$ احتمال القيمة المتوقعة، h_{ii} العناصر القطرية للمصفوفة $n \times n$، $i = 1,2,3,\ldots\ldots,n$

2) **جدول التصنيف** Classification Tables:

من جدول التصنيف يمكن تقدير الحساسية والدقة ونسبة التصنيف.

1- **الحساسية** Sensitivity: وهي احتمال أن يكون التصنيف المتوقع موجباً للمشاهدة الموجبة ويمكن حسابها بالصيغة التالية:

$$Sensitivity = \frac{TP}{TP + FN} = \frac{TP}{P}$$

حيث p المشاهدات الموجبة في التصنيف، TP المشاهدات المتوقعة الصحيحة، FN المشاهدات المتوقعة السالبة الخاطئة.

2- **الدقة** Specificity: وهي احتمال ان يكون التصنيف المتوقع سالباً للمشاهدة التي تكون فعلاً سالبة وتحسب بالصيغة الآتية:

$$Specify = \frac{TN}{FP + TN} = \frac{TP}{P'}$$

TN المشاهدات المتوقعة السالبة الصحيحة، P' المشاهدات السالبة في التصنيف

3- **نسبة التصنيف الصحيح** Hit Ratio وهي عبارة عن قيمة احتمال التصنيف الصحيح، فإذا كانت الكفاءة Efficiency والتي تأخذ الصيغة $Ef = TP + TN$.

$$HitRatio = \frac{Ef}{Total} = \frac{TP + TN}{P + P'}$$ فإن نسبة التصنيف الصحيح تساوي

3) إحصاء وولد Wald Statistic: يستخدم اختبار Wald لاختبار الدلالة الإحصائية لمعاملات انحدار الترجيح اللوغارتيمي ويتم حسابه بالصيغة

$Wald = (\frac{\hat{b}}{SE_b})^2$ ، حيث $\hat{b}$ هي قيمة معامل انحدار الترجيح الوغارتيمي، SE هي قيمة الخطأ المعياري (Osborne, 2015).

4) اختبار الدرجة Score Test: يشير هوزمر وليمشو Hosmer at el.(2013) أ ن هذا الاختبار لا يتطلب الاعتماد على نسبة الأرجحية العظمى وإنما باستخدام الدرجات الخام ويتم حسابه بالصيغة الآتية:

$$ST = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sqrt{\bar{y} - (1 - \bar{y}) \sum_{i=1}^{n} (x_i - \bar{x})}}$$

حيث x_i ترميز لقيم المشاهدات في المتغير المستقل، y_i ترميز لقيم المتغير التابع.

اختبار Hosmer & Lemeshow: يستخدم هذا الاختبار كما يذكر Osborne (2015) لتحديد إذا كان النموذج يطابق البيانات أم لا. ويستخدم اختبار مربع كاي c^2 لتقييم الفرق بين القيم المشاهدة والقيم المتوقعة.

تحليل الدالة التمييزية Discriminant Function:

هو تركيبه خطيه من متغير أو أكثر من المتغيرات المستقلة التي من شأنها أن تميز بين مجموعتين من المشاهدات او الموضوعات (أفراد، شركات، ..) في مجموعات معرفة مسبقاً. يستخدم تحليل الدالة التمييزية لتقييم مدى مطابقة تصنيف ما، وذلك بمعرفة عضوية عناصر عينة الدراسة في كل مجموعة، وكذلك التنبؤ بالموقع التصنيفي لحالات جديدة لم تصنف بعد. (Huberty,1994).

ويتم التحقق من التمييز كما يشير Hair et al.(2006) بحساب الأوزان لكل متغير مستقل للتباين بين المجموعات بالنسبة إلى التباين داخل المجموعة وتحليل الدالة التمييزية يعرف بدالة التمييز Discrimininat function. وقد اقترحه فيشر في عام 1936م، ويستخدم للفصل بين مجموعتين واستخلاص نموذج تصنيف لانتماء مجموعة من المشاهدات الجديدة،. ويأخذ الصيغة الآتية:

$$Z_{ik} = a + w_1 x_{1k} + w_2 x_{2k} + \ldots\ldots + w_i x_{ik}$$

حيث z_{ik} الدرجة التمييزية، a مقدار ثابت، w_i الوزن التمييزي، x_{ik} المتغير المستقل iفي k. ويحدد عدد الدوال التمييزية بالعلاقة $(k-1)$ حيث k عدد المجموعات أو مستويات المتغير التابع، فإذا كانت البيانات تتضمن متغيراً تابعاً له ثلاثة مستويات ($k=3$) فان تحليل الدالة التمييزية سيعطي دالتين تمييزيتين يمكن استخدامهما في تصنيف الحالات الجديدة (Hair,et al,2006). ويعود استخدام تحليل الدالة التمييزية في العلوم التربوية إلى جامعة هارفارد Harvard خلال الفترة 1950 – 1960 كما أشار Huberty (2006) إلى وجود العديد من الأبحاث بهذا الشأن من جامعة هارفارد. كما أن هناك مجالات كثيرة ومتعددة يمكن توظيف تحليل الدالة التمييزية فيها ويكون ذا فائدة كبيرة مثل تصميم الاختبارات والتنبؤ بالسلوك، وتأثيرات المعالجات، الاختلافات والفروق الاقتصادية.

فروض تحليل الدالة التمييزية: يتطلب أسلوب تحليل الدالة التمييزية كما يذكر Klecka (1980) و kardiyn & Olmus (2016) تحقق عدداً من الفروض الأساسية، منها وجود مجموعتين أو أكثر،ووجود حالتين على الأقل لكل مجموعة. أن تكون المتغيرات المستقلة متغيرات كمية تتبع توزيعاً طبيعياً وتحقق تجانس مصفوفة التباين والتغاير لكل المجموعات.

ويوجد العديد من أوجه التشابه كما ذكر (Hair et al. 2006) بين أسلوبي انحدار الترجيح الوغارتيمي وتحليل الدالة التمييزية، فكل منهما يعطي نتائج تنبؤية، ويتميز الطريقة الأسلوب الأول بكونه أقل تأثراً من الأسلوب الثاني بعدم تحقق الافتراضات الأساسية وبخاصة التوزيع الhعتدالي للمتغيرات.

الدراسات السابقة:

أجرى Kardiyen and Olmus (2016) دراسة أشارت الى أن مشكلة تصنيف المجموعتين لها تداعيات في عدد من المجالات ، مثل الطب، والاقتصاد، والتمويل، وقد تم المقارنة بين تصنيف مجموعتين باستخدام تحليل الدالة التمييزية وانحدار الترجيح اللوغاريتمي وتحليل التباين المتعدد. وتم استخدام عينات محاكاة لمختلف التوزيعات وأحجام العينات وأظهرت نتائج الدراسة أن تحليل الدالة التمييزية كان ذو أفضلية في تصنيف البيانات.

وأجرى Ebrahimzadeh et al. (2015) دراسة تنبؤية تتعلق بالحمل المرغوب وغير المرغوب. وقد قارنت الدراسة بين استخدام انحدار الترجيح اللوغاريتمي والانحدار الاحتمالي وتحليل الدالة التمييزية. استخدمت الدراسة ثلاثة نماذج للتصنيف، تكونت العينة من 887 من النساء الحوامل في عام 2012 وتم استخدام عينة طبقية وعنقودية، واستخدمت محكات الحساسية، ونسبة التوقعات. وتوصلت الدراسة إلى نتائج من أهمها أن انتشار حالات الحمل غير المرغوب بلغ 25.3% وقد كان أسلوب انحدار الترجيح اللوغاريتمي أفضل من تحليل الدالة التمييزية في دقة التصنيف.

كما أجرى Santos et al. (2014) دراسة تتعلق بتحديد نوع الجنين باستخدام تحليل الدالة التمييزية وانحدار الترجيح اللوغاريتمي واستخدمت الدراسة عينات مختلفة للمقارنة بين هذه الطرق الإحصائية وأجريت المقارنة باستخدام محكات مختلفة وأشارت النتائج إلى أن انحدار الترجيح اللوغاريتمي أفضل من تحليل الدالة التمييزية ويقدم أفضل تنبؤ.

وأجرى Li & Sun (2011) دراسة استخدمت أسلوب انحدار الترجيح اللوغاريتمي، وتحليل المكونات الرئيسية، وتحليل الدالة التمييزية للتنبؤ بالإخفاق التجاري للشركات بسبب حالة الطوارئ من الأزمة المالية في البلدان المتقدمة. واتضح أن التنبؤ باستخدام تحليل المكونات الرئيسية كان منخفضاً في الأداء مقارنة بأسلوب انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية. وقد حاولت هذه الدراسة بناء نموذج مهجن من تحليل المكونات الرئيسية وانحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية وقد توصلت الدراسة إلى أن هذا الأسلوب المختلط كان الأفضل مقارنة بالطرق الأخرى.

في دراسة قام بها Lidi-Anne & Gaston (2011) للمقارنة بين أسلوب تحليل الدالة التمييزية وتحليل الانحدار اللوغاريتمي الثنائي في الكشف عن قدرة بعض العوامل بالتنبؤ بعدم تعاطي الكحول خلال فترة الحمل والتي تمثلت في معتقدات الحامل نحو إضرار تعاطي الكحول والعمر للحامل أثناء فترة الحمل وقد تم أخذ بيانات لعينة عشوائية من الحوامل، تم تحليل البيانات باستخدام أسلوب تحليل الدالة التمييزية وتحليل انحدار الترجيح اللوغاريتمي وقد توصلت الدراسة إلى أن أسلوب تحليل الدالة التمييزية كان أكثر دقة في تصنيف متعاطي الكحول وغير المتعاطين مقارنة بأسلوب انحدار الترجيح اللوغاريتمي.

وفي دراسة أجراها Gregori et al. (2011) بهدف تحديد العوامل التنبؤية بموت مرضى السكر وقد استخدمت الدراسة عدة اساليب إحصائية تنبؤية منها تحليل الدالة التمييزية وتحليل انحدار الترجيح اللوغاريتمي وقد تم اخذ عينة من سجلات بيانات مرضى السكر بأحد المستشفيات بايطاليا، شملت عمر المريض وتاريخه المرضي، ونوع الغذاء، وطبيعة العمل وقد توصلت الدراسة إلى أن أسلوب انحدار الترجيح الوغاريتمي كان أكثر دقة في تحديد العوامل المنبئة بموت مرض السكر وذلك مقارنة بتحليل الدالة التمييزية.

كما أجرى Maroco et al. (2011) دراسة تنبؤية استخدمت عدة طرق تنبؤية ومنها تحليل الدالة التمييزية وانحدار الترجيح اللوغاريتمي. وقد استخدمت الدراسة بيانات حقيقية تكونت العينة من سجلات (92) حالة مريض لتصنيف الحالات إلى ضعف إدراكي معرفي وغير ذلك، واستخدمت محكات الدقة والحساسية للمقارنة وتوصلت إلى أن تحليل الدالة التمييزية يتمتع بدرجة عالية من الدقة والحساسية وقوة التمييز مقارنة بانحدار الترجيح اللوغاريتمي.

كما أجرى Adler et al. (2011) دراسة للمقارنة بين أسلوبي التحليل التمييزي وتحليل انحدار الترجيح اللوغاريتمي في التنبؤ بالإصابة بمرض الجلاكوما وذلك اعتماداً على بعض المنبئات مثل ضغط العين والتاريخ المرضي والعوامل الوراثية، اختيرت عينة تشمل العاديين والمصابين، وقد توصلت الدراسة الى التشابه في دقة التصنيف بين التحليل التمييزي وتحليل انحدار الترجيح اللوغاريتمي.

وفي دراسة أجراها Pohar et al. , (2004) للمقارنة بين انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية وقد استخدمت الدراسة مجموعة من المحكات لمقارنة دقة التنبؤ وتوصلت إلى أن القدرة التنبؤية متقاربة بين التحليلين إذا تحققت الافتراضات. كما أكدت أن تحليل الدالة التمييزية الأكثر ملاءمة عند تحقق التوزيع الاعتدالي، ولكن إذا لم يتم استيفاء افتراضات تحليل الدالة التمييزية الخطي فإن استخدامه ليس له ما يبرره، في حين يعطي انحدار الترجيح اللوغاريتمي نتائج جيدة بغض النظر عن شكل التوزيع. كما يتم الحصول على تقديرات انحدار الترجيح اللوغاريتمي بطريقة دالة الأرجحية القصوى والتي تتميز بمزايا أفضل.

وفي دراسة قام بها Frass & Newman (2003) للمقارنة بين تحليل الدالة التمييزية وتحليل انحدار الترجيح اللوغاريتمي وبعض الأساليب الإحصائية، هدفت الدراسة إلى تحديد كيفية اختيار الأسلوب المناسب من بين الأساليب التحليلية المتعددة بناءً على افتراضات كل أسلوب والتي يتم تصميمها لتحليل المتغير التابع مع مجموعة من المتغيرات المستقلة، وترتبط المناقشة من خلال تحليل البيانات بواسطة الإجراءات الشروط الخاصة بكل أسلوب واستخدام برامج الكمبيوتر الإحصائية، إضافة إلى ذلك توضح الدراسة أن الباحثين في حاجة إلى فهم ومعرفة مدى ملائمة كل طريقة أو أسلوب، وقد قامت الدراسة بتحليل عدد من البيانات لتوضيح ومناقشة كيفية عملية تحديد اختيار أسلوب التحليل المناسب وفق الافتراضات المتحققة في البيانات، وقد أكدت الدراسة على أفضلية تحليل انحدار الترجيح اللوغاريتمي في تصنيف البيانات مقارنة بتحليل الدالة التمييزية.

وفي دراسة أخرى أجراها بها Hwang (2001) تشير هذه الدراسة إلى أن دقة تصنيف المجموعة بالتركيز على تحليل الدالة التمييزية، إضافة إلى ذلك فان الهدف والغرض الأساسي لهذه الدراسة هو تقديم نظرة عامة لكيفية عمل تحليل الدالة التمييزية وكيف يمكن إن يساعد في الإجابة على العديد من أسئلة الباحثين المتنوعة، إضافة إلى ما سبق توضح هذه الدراسة وتشرح ما هو تحليل الدالة التمييزية، ولماذا هو على قدر من الأهمية، وكذلك تقدم الدراسة بيانات توضيحية لاستخدام تحليل الدالة التمييزية.كما تؤكد الدراسة على أفضلية تحليل الدالة التمييزية في تصنيف المشاهدات.

وفي دراسة أجراها Ruiz (1991) تم حساب كفاءة انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية لاختبار الفرضيات حول المعالم في حالة التوزيع الاعتدالي للمتغيرات وقد توصلت الدراسة إلى أن الكفاءة النسبية في التصنيف كانت متقاربة بين انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية.

الطريقة والإجراءات:

منهج الدراسة: استخدمت الدراسة المنهج الوصفي:

عينة الدراسة: تكونت عينة الدراسة من (1722) طالباً وطالبة بجامعة أم القرى المقبولين في المسارات المختلفة في السنة التحضيرية والذين تم تسكينهم في التخصصات العلمية أو الأدبية وانتهوا من دراسة فصل دراسي في التخصص بعد السنة التحضيرية للعام الجامعي ١٤٣٤/١٤٣٥هـ.

أدوات الدراسة: تم الاعتماد على بيانات عينة الطلاب والطالبات لمعدل السنة التحضيرية والمجموع الكلي لدرجات الثانوية العامة واختبار التحصيل للمرحلة الثانوية واختبار القدرات للمرحلة الثانوية والمعدل التراكمي للفصل الدراسي الأول في التخصص للسنة التالية للسنة التحضيرية.

متغيرات الدراسة:

المتغيرات المستقلة: X_1 متغير درجات الثانوية، X_2 درجات اختبار القدرات، X_3 درجات اختبار التحصيل، X_4 المعدل التراكمي للسنة التحضيرية.

المتغير التابع: Y متغير تصنيفي للمعدل التراكمي في التخصص للسنة التالية للسنة التحضيرية، يأخذ القيمة (1) إذا كان المعدل التراكمي مرتفعاً يساوي 2.75أو أعلى، والقيمة (2) إذا كان المعدل التراكمي منخفضاً اقل من 2.75).

المعالجة الإحصائية: تم استخدام حزمة البرامج الإحصائية لمعالجة البيانات SPSS، لبرنامج انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية للإجابة على تساؤلات الدراسة.

نتائج الدراسة ومناقشتها:

قام الباحث بالتحقق من مدى مطابقة البيانات باستخدام أسلوب انحدار الترجيح اللوغاريتمي، تم اختبار الدلالة الإحصائية لأسلوب انحدار الترجيح اللوغاريتمي الذي تم توفيقه، ويوضح جدول (١) نتائج المطابقة، حيث قيمة مربع كاي تساوي (869.415) بدرجات حرية 4 ومستوى دلالة إحصائية (0.000) وهذا يعني أن النموذج ملائم للتنبؤ، حيث تؤكد هذه النتيجة على أن هناك على الأقل عاملاً واحداً من معاملات النموذج لا يساوي صفراً، وهذا يعني أن هناك على الأقل متغيراً مستقلاً له أهمية ومساهمة في التصنيف والتنبؤ.

جدول (١) اختبار الدلالة الإحصائية لانحدار الترجيح اللوغاريتمي

قيمة مربع كاي	درجات الحرية	قيمة الدلالة الإحصائية
869.415	4	0.000

وللتأكد من وجود فروق ذات دلالة إحصائية بين القيم المشاهدة والقيم المتوقعة للبيانات، تم استخدام اختار Hosmer and Lemeshow لجودة المطابقة كما في جدول (٢) .

جدول (٢) اختبار Hosmer and Lemeshow لجودة المطابقة

قيمة مربع كاي	درجات الحرية	مستوى الدلالة
863.664	8	0.000

ويتضح من جدول (٢) أن قيمة اختبار مربع كاي لاختبار Hosmer and Lemeshow تساوي (863.664) وبدرجات حرية تساوي 8، ومستوى دلالة (0.000)، وهذا يعني أن قيمة اختبار مربع كاي دالة إحصائياً، مما يدل أن البيانات المتوقعة تطابق البيانات المشاهدة.

جدول (٣) جدول التصنيف الناتج من أسلوب انحدار الترجيح اللوغاريتمي

التصنيف المشاهد	التصنيف المتوقع			النسبة المئوية للتصنيف الصحيح
	منخفض	مرتفع	المجموع	
منخفض	692	128	820	84.4
مرتفع	114	785	899	87.3
المجموع	806	913	1719	85.9

من الجدول (٤) نجد أن قيمة حساسية النموذج Sensitivity للتصنيف والتي تعبر عن التنبؤات الصحيحة لمجموعة الطلبة منخفضي المعدل التراكمي تساوي (84.4 %). وأن قيمة دقة النموذج Specificity والتي تمثل نسبة التنبؤات الصحيحة لمجموعة الطلبة مرتفعي المعدل التراكمي تساوي (87.3 %) ،و أن نسبة التصنيف الصحيح بعامة، أو ما تسمى نسبة كفاءة النموذج Efficiency Ratio تساوي (85.9 %)، وهي قيمة تدل على أن القدرة التنبؤية للنموذج مرتفعة.

وقد تم حساب تقديرات المعالم في معادلة انحدار الترجيح اللوغاريتمي للمتغيرات المستقلة، كما في جدول (٤) التالي:

جدول (٤) تقديرات معالم أسلوب انحدار الترجيح اللوغاريتمي

المتغيرات	قيم معاملات الانحدار	الخطأ المعياري	إحصائية Wald	درجات الحرية	مستوى الدلالة	الدالة الاسية للمعاملات
X1	0.147	0.021	49.682	1	0.000	1.158
X2	0.009	0.014	0.000	1	0.993	1.000
X3	0.041	0.013	10.161	1	0.001	1.016
X4	1.516	0.105	208.937	1	0.000	4.555
C	−20.867	1.975	111.597	1	0.000	0.000

يلاحظ من جدول (5) أن قيمة معاملات الانحدار لجميع المتغيرات موجبة مما يعني أن لوغاريتم معامل الترجيح يزداد بمقدار قيمة كل معامل. قيمة معامل الانحدار للمتغير X_4 (معدل السنة التحضيرية) يساوي (1.516) وتفسر على أن

لوغاريتم معامل ترجيح تصنيف الطالب مرتفع التحصيل يزداد بمقدار القيمة (1.516) كلما زادت قيمة المتغير X_4 (معدل السنة التحضيرية) بمقدار درجة واحدة وذلك بعد ضبط أثر المتغيرات المستقلة الأخرى. كما يتم تفسير بقية المتغيرات بالطريقة نفسها، قيمة معامل الانحدار للمتغير X_1 (متغير درجات الثانوية) تساوي (0.147) وهذا يعني أن قيمة معامل الانحدار تزداد بمقدار (0.147) كلما زادت قيم متغير X_1 بمقدار درجة واحدة، كذلك قيمة معامل الانحدار للمتغير X_3 (اختبار التحصيل) تساوي (0.041) وهذا يعني أن قيمة معامل الانحدار تزداد بنفس القيمة كلما ارتفعت قيمة متغير X_3 (درجات التحصيل) بمقدار درجة واحدة، وأخيرا قيمة معامل الانحدار لمتغير X_2 (درجات القدرات) تساوي (0.009) ويزداد بالقيمة نفسها كلما زادت قيمة المتغير X_2 (درجات القدرات) بمقدار درجة واحدة. ويلاحظ في التفسير السابق أنه كلما ارتفعت قيمة معامل الانحدار دل ذلك على أهمية المتغير المستقل في التنبؤ بالمتغير التابع، والعكس كلما اقترب من الصفر دل على عدم أهمية المتغير المستقل في التنبؤ بالمتغير التابع.

وللتأكد من القدرة التنبؤية للنموذج تم حساب القيم لاختبار Wald كما يوضح ذلك جدول (٤) وبملاحظة قيم هذا الاختبار ومستوى الدلالة لكل متغير من المتغيرات السابقة، نجد أن قيمة اختبار Wald للمتغير X_4 (معدل السنة التحضيرية) تساوي (208.937) عند مستوى دلالة (0.000) وهذا يعني ان المتغير X_4 يسهم في تصنيف الطلبة الى مرتفعي ومنخفضي التحصيل. كما نجد أن قيمة اختبار Wald للمتغير X_1 (متغير درجات الثانوية) تساوي (49.682) عند مستوى دلالة (0.000) وهي قيمة ذات دلالة إحصائية وتعني ان المتغير X_1 يسهم في تصنيف الطلبة الى منخفضي ومرتفعي التحصيل، كما كانت قيمة اختبار Wald للمتغير X_2 (درجات القدرات) تساوي (0.000) عند مستوى دلالة (0.993) وهي قيمة ليست دالة إحصائياً، مما يعني أن المتغير X_2 ليس له قدرة تنبؤية. كما كانت قيمة اختبار Wald للمتغير X_3 (الاختبار التحصيلي) تساوي (10.161) عند مستوى دلالة (0.001) وهذا يعني أن قيمة معامل الانحدار للمتغير X_3 ذات دلالة إحصائية وبالتالي له قدرة تنبؤية .

ويمكن أيضا الحكم على القدرة التنبؤية باستخدام إحصاءات الدرجة Score كما في الجدول (٥).

جدول (٥) إحصاءات الدرجة Score للمتغيرات المستقلة

المتغيرات	إحصاء الدرجة Score	درجات الحرية	مستوى الدلالة
X_1	358.624	1	0.000
X_2	171.736	1	0.000
X_3	343.560	1	0.000
X_4	626.856	1	0.000

يتضح من جدول (٥) أن إحصائية الدرجة Score للمتغير X_4 (معدل السنة التحضيرية) كانت تساوي (626.856) عند درجة حرية واحدة ومستوى دلالة (0.000)، وهذا يعني أن معامل الانحدار للمتغيرX_4 له دلالة إحصائية في القدرة التنبؤية بالمتغير التابع. وكذلك قيمة معامل المتغير X_1 (درجات الثانوية) تساوي (358.624) عند درجة حرية واحدة ومستوى دلالة (0.000) وهذا يعني أن معامل الانحدار للمتغير X_1 له دلالة إحصائية في التنبؤ بالمتغير التابع في أسلوب انحدار الترجيح اللوغاريتمي. كذلك معامل الانحدار للمتغير X_2 (درجات القدرات) تساوي (171.736) عند درجة حرية واحدة ومستوى دلالة (0.000) وهذا يعني أن معامل الانحدار للمتغير X_2 له دلالة إحصائية وبالتالي له قدرة تنبؤية بالمتغير التابع. وأخيرا معامل الانحدار للمتغير X_3 (الاختبار التحصيلي) يساوي (343.560) عند درجة حرية واحدة ومستوى دلالة إحصائية تساوي (0.000) وهذا يعني أن معامل الانحدار للمتغير X_3 ذو دلالة إحصائية وبالتالي يسهم في القدرة التنبؤية بالمتغير التابع.

ولتحديد العوامل المسؤولة والقدرة على التنبؤ بالأداء الأكاديمي باستخدام تحليل الدالة التمييزية، تم التأكد من تحقق افتراضات تحليل الدالة التمييزية، حيث يلاحظ أن المتغيرات المستقلة هي متغيرات كمية، وأن الدراسة تعتمد على عدد 1722 حالة، وبالتالي وفقا لنظرية النهاية المركزية فإن البيانات تتبع التوزيع الاعتدالي، ولمزيد من التأكيد تم استخدام اختبار كلمجروف –سمير نوف وجاءت قيم الاختبار غير دالة إحصائياً، بما يؤكد على اعتدالية توزيع البيانات. وللتأكد من تحقق فرض تجانس التباين والتغاير تم حساب قيمة Boxs' M فمن جدول (7) نجد أن قيمة هذا الاختبار بلغت (355.452) وهي دالة إحصائياً عند مستوى دلالة (0.000) مما يعني عدم تحقق افتراض تجانس التباين والتغاير،

ولكن الحالات التي يكون عدد أفراد عينتها كبيراً تظهر الاختلافات البسيطة فروقاً في النتائج وتشير الدراسات الى أن تحليل الدالة التمييزية يمكن أن يصمد في مواجهة هذا الافتراض.

جدول (٦) اختبار Boxs' M لتجانس التباين والتغاير

Boxs' M		
355.452		
35.456	القيمة	F
10	df1	
13828831.226	df2	
0.000	Sig	

ولتحديد العوامل التمييزية والأهمية النسبية للمتغيرات المستقلة في القدرة التمييزية، تم حساب قيمة Wilks' Labda واختبار النسبة الفائية كما في جدول (٦).

جدول (٧) اختبار Wilks' Labda واختبار النسبة الفائية (F)

المتغيرات	ولكس لمدا	F	Df1	Df2	مستوى الدلالة
X_1	0.791	452.637	1	1717	0.000
X_2	0.900	190.576	1	1717	0.000
X_3	0.800	428.875	1	1717	0.000
X_4	0.635	985.502	1	1717	0.000

يتضح من جدول (٧) أن قيمة Wilks' Labda للمتغير المستقل X_4 (معدل السنة التحضيرية) تساوي (0.635) وهي قيمة دالة إحصائياً عند مستوى دلالة (0.000) وكلما كانت قيمة Wilks' Labda اقل للمتغير المستقل ودالة إحصائياً كلما كانت أهمية المتغير المستقل أقوى في القدرة التمييزية والتنبؤية. كما يلاحظ أن قيمة Wilks' Labda للمتغير X_1 (درجات الثانوية) تساوي (0.791) وهي قيمة دالة إحصائياً عند مستوى دلالة (0.000) وهذا يعني ان المتغير X_4 يعتبر أكثر أهمية في النموذج، كما كانت قيمة Wilks' Labda للمتغير X_3 (درجات الاختبار التحصيلي) تساوي (0.800) وهي قيمة دالة إحصائياً عند مستوى دلالة (0.000)، وأخيرا نجد أن قيمة Wilks' Labda للمتغير X_2 القدرات تساوي (0.900) وهي قيمة دالة إحصائيا عند مستوى دلالة (0.000).

كما يفسر اختبار النسبة الفائية (F) الأهمية النسبية للمتغيرات المستقلة المكونة للنموذج، فكلما كانت قيمة هذه النسبة مرتفعة للمتغير المستقل ودالة إحصائياً دل ذلك على أهمية المتغير. وبالنظر الى جدول (8) نجد أ جميع المتغيرات دالة إحصائيا بالنسبة لقيمة اختبار النسبة الفائية للمتغيرات ويمكن ترتيبها أهميتها النسبية استنادا إلى ذلك كالآتي:

المتغير X_4 ويعتبر الأكثر أهمية حيث قيمة اختبار النسبة الفائية يساوي (985.502)، يلي ذلك المتغير X_1 حيث بلغت قيمة اختبار النسبة الفائية له (452.637)، ثم يلي ذلك المتغير X_3 حيث كانت قيمة اختبار النسبة الفائية تساوي (428.875)، وأخيرا المتغير X_2 حيث كانت قيمة اختبار النسبة الفائية تساوي (190.567) ويعتبر أقل المتغيرات أهمية في النموذج وهي تتفق مع قيم Wilks' Labda في تحديد المتغيرات المستقلة ذات الأهمية في النموذج.

ولتحديد الدوال التمييزية ونسبة التباين الذي يمكن تفسيره، يوضح الجدول (٨) قيم الجذور الكامنة والنسبة المئوية للتباين.

جدول (٨) نتائج قيم الجذور الكامنة

الدالة التمييزية	قيم الجذور الكامنة	النسبة المئوية للتباين	الارتباط المركب
1	0.690	100	0.639

يلاحظ من جدول (٨) أن قيمة الجذر الكامن تساوي (0.690) وهي تفسر نسبة تباين (100%)، والارتباط المركب يساوي(0.639) يعني أن النسبة المئوية للتباين في المتغير التابع الذي تم تمييزه بالمتغيرات المستقلة تساوي (63.9%).

ويمكن استخدام اختبار Wilks' Lambda لاختبار الدالة التمييزية كما في جدول (٩)

جدول (٩) اختبار Wilks' Lambda للدوال التمييزية

اختبار الدالة	Wilks' Lambda	مربع كاي	درجات الحرية	مستوى الدلالة
1	0.592	899.557	4	0.000

يتضح من الجدول (٩) أن قيمة Wilks' Lambda المستخدم لاختبار الدالة التمييزية يساوي (0.597) دالة إحصائياً عند مستوى (0.000)، وكذلك قيمة مربع كاي تساوي (899.557) بدرجات حرية (4) دالة عند مستوى (0.000).

تم حساب قيم معاملات الدالة التمييزية غير المعيارية لتحديد مساهمة المتغيرات المستقلة في الدالة التمييزية كما في جدول (١٠).

جدول (١٠) قيم معاملات الدالة التمييزية غير المعيارية

المتغيرات	قيمة معامل الدالة
X_1	0.097
X_2	0.004
X_3	0.027
X_4	0.917
C	−13.895

من خلال قيم معاملات الدالة التمييزية غير المعيارية كما في جدول (11) يلاحظ أن المتغير X_1 (معدل السنة التحضيرية) الأكثر أهمية حيث بلغت قيمة المعامل له (0.917)، ثم المتغير X_1 (درجات الثانوية) حيث تساوي قيمة المعامل له (0.097)، ثم المتغير X_3 (درجات الاختبار التحصيلي) حيث تساوي قيمة المعامل له (0.027) وأخيراً المتغير X_2 (اختبار القدرات) حيث كانت قيمة المعامل له تساوي (0.004) .

تم تحديد قيم المعاملات المعيارية للدالة التمييزية كما في جدول (١١)

جدول (١١) معاملات الدالة التمييزية المعيارية

المتغيرات	قيمة معامل الدالة
X_1	0.348
X_2	0.023
X_3	0.177
X_4	0.733

ويتضح من جدول (١١) المساهمة الجزئية لكل متغير مستقل في الدالة التمييزية، حيث يلاحظ أن المتغير X_1 الأكثر أهمية ومساهمة حيث بلغت قيمة المعامل له (0.773)، ويليه من حيث الأهمية المتغير X_1 حيث تساوي قيمة المعامل له (0.348)، ثم المتغير X_3 حيث تساوي قيمة المعامل له (0.177) وأخيراً المتغير X_2 حيث كانت قيمة المعامل له تساوي (0.025) وهي تتفق مع القيم غير المعيارية في ترتيب المتغيرات من حيث الأهمية .

تقييم كفاءة النموذج:

1. **قيمة الجذور الكامنة:** كانت قيمة الجذر الكامن تساوي (0.690) وهي قيمة دالة إحصائياً وبتباين (100%).

2. **معامل الارتباط المركب:** وهو يقيس قوة العلاقة بين المتغيرات المستقلة والمتغير التابع وقد بلغت قيمته (0.639) وهي قيمة موجبة ومرتفعة تدل على ارتباط المتغيرات المستقلة لدالة التصنيف بالمتغير التابع.
3. **اختبار** Wilks' Lambda: كانت قيمته للدالة التمييزية تساوي (0.592) وهي قيمة دالة إحصائياً.
4. **مربع كاي:** كانت قيمة مربع كاي للنموذج تساوي (899.557) وهي قيمة دالة إحصائيا مما يعني أن الاختلافات لا تعود إلى الصدفة وإنما هي اختلافات جوهرية مما يدل على كفاءة النموذج.

اختبار القدرة التمييزية والتصنيفية للنموذج:

لاختبار قدرة الدالة التمييزية على التمييز والتصنيف، يوضح الجدول (١٢) النسبة المئوية للتصنيف الصحيح المشاهد والمتوقع.

جدول (١٢) نسب التصنيف الصحيح

التصنيف المشاهد	التصنيف المتوقع			النسبة المئوية للتصنيف الصحيح
	منخفض	مرتفع	المجموع	
منخفض	687	134	821	%83.7
مرتفع	117	784	901	% 87.0
المجموع	804	918	1722	% 85.4

يتضح من جدول (١٢) أن نسبة التصنيف الصحيح لمجموعة منخفضي التحصيل تساوي 83% ولمجموعة مرتفعي التحصيل كانت نسبة التصنيف الصحيح تساوي 87%، وبشكل عام فإن الدالة تصنف بشكل صحيح ما نسبته 85.4% وهي تدل على جودة التنبؤ بانتماء المشاهدات باستخدام تحليل الدالة التمييزية.

مدى الدقة في تحليل البيانات باستخدام أسلوب انحدار الترجيح اللوغارتيمي وتحليل الدالة التمييزية.

جدول (١٣) قيم حساسية التصنيف ودقة التصنيف ونسبة التصنيف الصحيح لتحليل انحدار الترجيح اللوغارتيمي وتحليل الدالة التمييزية

المحكات	نوع التحليل		قيمة الفرق بين نوعي التحليل
	انحدار الترجيح اللوغارتيمي	تحليل الدالة التمييزية	
حساسية التصنيف	84.4	83.7	0.7
دقة التصنيف	87.3	87	0.3
نسبة التصنيف الصحيح	85.9	85.4	0.5

ويتضح من جدول (١٣) أن قيمة حساسية التصنيف لتحليل انحدار الترجيح اللوغاريتمي تساوي (84.4)، بينما في تحليل الدالة التمييزية تساوي (83.7) وبفرق قدره (0.7) لصالح تحليل انحدار الترجيح اللوغاريتمي، كما نجد أن دقة التصنيف بلغت (87.3) عند استخدام انحدار الترجيح اللوغاريتمي، بينما كانت تساوي (87.0) عند استخدام تحليل الدالة التمييزية وبفرق قدره (0.3) لصالح انحدار الترجيح اللوغاريتمي، كما نجد أن قيمة نسبة التصنيف الصحيح عند استخدام انحدار الترجيح اللوغاريتمي تساوي (85.9)، بينما كانت قيمتها عند استخدام تحليل الدالة التمييزية تساوي (85.4) وبفرق قدره (0.5) لصالح تحليل انحدار الترجيح اللوغاريتمي، والملاحظ من خلال المقارنة السابقة أن قيم انحدار الترجيح اللوغاريتمي وفقاً لمحكات حساسية التصنيف، ودقة التصنيف، ونسبة التصنيف الصحيح، أعلى من تحليل الدالة التمييزية وبالتالي يكون أكثر دقة من تحليل الدالة التمييزية.

مناقشة النتائج:

للتأكد من مدى جودة مطابقة البيانات لنموذج انحدار الترجيح اللوغاريتمي، أظهرت قيمة مربع كاي على ملائمة أسلوب انحدار الترجيح اللوغاريتمي لتوفيق البيانات وأن النموذج ملائم لتنبؤ. كما أظهرت نتيجة قيمة Hosmer and Lemshow لجودة المطابقة أن البيانات المتوقعة وفق نموذج انحدار الترجيح اللوغاريتمي تطابق البيانات المشاهدة. تدل كفاءة نموذج انحدار الترجيح اللوغاريتمي والتي تساوي (85.9%) على أن القدرة التنبؤية عالية. وهي تتفق مع ما توصلت إليه دراسة (Ebrahimzadeh et al. ,2015) والتي أكدت على أفضلية مطابقة انحدار الترجيح اللوغاريتمي للبيانات في تصننيف المشاهدات بالحمل المرغوب وغير المرغوب وكذلك دراسة (Santos et al. ,2014) التي أكدت على ملائمة هذا الأسلوب للبيانات المتعلقة بتحديد نوع الجنين. كما دلت نتائج اختبار Wald وإحصاءات الدرجة في انحدار الترجيح اللوغاريتمي على أهمية المتغير X_4 (معدل السنة التحضيرية) في القدرة التمييزية والتصنيفية، ثم المتغير X_1 (درجات الثانوية)، ومن ثم المتغير X_3 (درجات الاختبار التحصيلي) وأخيرا المتغير X_2 (درجات اختبار القدرات). وهي تتفق مع ما توصلت اليه (Gregori et al. ,2011) في تحديد العوامل التنبؤية بموت مرضى السكر والتي

أكدت إلى أن أسلوب انحدار الترجيح الوغاريتمي كان أكثر دقة في تحديد العوامل المنبئة بموت مرض السكر، ودراسة (Ebrahimzadeh et al. ,2015) في تحديد العوامل المنبئة بالحمل المرغوب وغير المرغوب. أما تحليل البيانات باستخدام الدالة التمييزية فقد دلت قيمة Wilks' Lambda واختبار النسبة الفائيةF في تحليل الدالة التمييزية على دلالة جميع المتغيرات المستقلة وأهميتها في القدرة التمييزية، حيث كان المتغير X_4 (معدل السنة التحضيرية) الأكثر أهمية، ومن ثم المتغير X_1 (درجات الثانوية)، ومن ثم المتغيرX_3 التحصيل، وأخيراً المتغير X_2 (درجات اختبار القدرات). وهذه النتيجة تتفق مع ما توصلت اليه في دراسة قام بها Lidi-Anne & Gaston (2011) في تحديد العوامل المنبئة بعدم تعاطي الكحول خلال فترة الحمل والتي تمثلت في معتقدات الحامل نحو اضرار تعاطي الكحول والعمر للحامل أثناء فترة الحمل وأفضلية أسلوب تحليل الدالة التمييزية في تحديد هذه العوامل. وكذلك دراسة (Adler et al., 2011) في تحديد العوامل المنبئة بالإصابة بمرض الجلاكوما وذلك اعتماداً على بعض المنبئات مثل ضغط العين والتاريخ المرضي والعوامل الوراثية. كما أظهرت الدراسة أن النموذجان يختلفان في الاختبار المتعلقة باختبار المعالم واختبار النموذج ككل فا انحدار الترجيح اللوغاريتمي يستخدم اختبار Wald لاختبار دلالة المعالم واختبار Hosmer and Lemshow للنموذج ككل، بينما يستخدم تحليل الدالة التمييزية اختبار Wilks' Lambda لاختبار المعالم واختبار والنموذج ككل. كما تتشابه نتائج تحليل انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية في تحديد العوامل التمييزية وقوتها التنبؤية حيث كانت من حيث الأهمية والمساهمة، المتغير X_4 (معدل السنة التحضيرية)، ثم المتغير X_1 (درجات الثانوية)، ومن ثم المتغير X_3 (متغير اختبار التحصيل) والأقل المتغير X_2 (درجات اختبار القدرات). كما أشارات نتائج دقة تحليل البيانات إلى أفضلية انحدار الترجيح اللوغاريتمي مقارنة بتحليل الدالة التمييزية، وذلك وفق محكات حساسية ودقة التصنيف ونسبة التصنيف الصحيح وتتفق هذه النتيجة مع دراسة (Ebrahimmzadeh et al., 2015) والتي توصلت إلى أفضلية انحدار الترجيح اللوغاريتمي في دقة التصنيف وكذلك دراسة (Santos et al., 2014) ودراسة (Pohar et al., 2004)، ودراسة (Gregori et al. ,2011) لتحديد العوامل التنبؤية بموت مرضى السكر والتي وأكدت على أن أسلوب انحدار الترجيح

اللوغاريتمي كان أكثر دقة في تحديد العوامل المنبئة بموت مرض السكر وذلك مقارنة بتحليل الدالة التمييزية. بينما تختلف مع ما توصلت له دراسة (Kardiyen and Olmus, 2016) عند المقارنة بين تصنيف مجموعتين باستخدام تحليل الدالة التمييزية وانحدار الترجيح اللوغاريتمي والتي أظهرت أن تحليل الدالة التمييزية كان ذو أفضلية في تصنيف البيانات. كذلك دراسة (Maroco et al., 2011) والتي أكدت على أفضلية تحليل الدالة التمييزية في تصنيف المشاهدات. ودراسة (Lidi-Anne & Gaston ,2011) والتي تؤكد أن أسلوب تحليل الدالة التمييزية كان أكثر دقة في تصنيف متعاطي الكحول وغير المتعاطين مقارنة بأسلوب انحدار الترجيح اللوغاريتمي. كما تختلف مع دراسة (Maroco et al. 2011,) والتي أكدت أن تحليل الدالة التمييزية يتمتع بدرجة عالية من الدقة والحساسية وقوة التمييز مقارنة بانحدار الترجيح اللوغاريتمي.

التوصيات والمقترحات:

بناءً على نتائج الدراسة الحالية، فإن الباحث يوصي باستخدام انحدار الترجيح اللوغاريتمي الثنائي في تحليل البيانات الأكاديمية ذات المتغيرات التابعة الثنائية لتحديد نمط المتغيرات التي تميز الأداء والتصنيف والتنبؤ بالنجاح والإخفاق الأكاديمي.

كما يقترح الباحث ما يلي:

- إجراء دراسة مقارنة بين أسلوبي انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية باختلاف أحجام العينات.
- إجراء دراسة مقارنة بين أسلوبي انحدار الترجيح الوغاريتمي وتحليل الدالة التمييزية باستخدام متغيرات مستقلة مختلفة.
- إجراء دراسة في دقة التصنيف بين اسلوب تحليل انحدار الترجيح اللوغاريتمي وتحليل الدالة التمييزية بمتغيرات تابعة تحتوي على عدد من الفئات التصنيفية.

المراجـــع

الزامل، محمد عبدالله (2012). قدرة معايير القبول على التنبؤ بالتقدم اللأكاديمي لطلاب السنة التحضيرية بجامعة الملك سعود. مجلة رسالة الخليج العربي، العدد 126 ، ص **157–214**، السعودية.

Adler, W.; Brenning, A.; Potapov, S.; Schmid, M. (2011). Ensemble Classification of Paired Data. Computational Statistics & Data Analysis. 55(5), pp.1933-1941.

Ebrahimzadeh, F. ; Hajizadeh , E.; Vahabi, N.; Almasian, M. ; Bakhteyar, K. (2015). Prediction of Unwanted Pregnancies Using Logistic Regression, Probit Regression and Discriminant Analysis. Medical Journal, Islam Repub Iran, V (29), 264.

Fras, J.; Newman, I.(2003). Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique. Paper Presented at the Annual Meeting of the Eastern Educational Research Association.

Gregori, D.; Petrinco, M.; Simona, B.(2011).Using Data Mining Techniques in Monitoring Diabetes Care. The Simpler the Better, Journal of Medical Systems, 35 (2), pp.277-281.

Hair , J.; Black, W.; Babin , B.; Anderson, R.; Tatham, R. (2006). Multivariate Data Analysis. Sixth Edition. New Jersey: Upper Saddle River.

Hawng, D.(2001).Issues in Predictive Discriminant Analysis: Using and Interpreting the Leave-One-Out Jackknife Method and the Improvement-Over-Change " I"

Index Effect Size. Paper Presented at the Annual Meeting of the Southwest Educational Research Association.

Hosmer, D.; Lemeshow, S.; Sturdivant, R. (2013). Applied Logistic Regression.(3rd ed). John Wiley &Sons, U.S.

Huberty, C. (1994). Applied Discriminant Analysis. New York: John Willy & Sons.

Huberty, C.; Olejnik, S. (2006).Applied MANOVA and Discriminant Analysis. 2nd. New Jersey: John Wiley & Sons.

Inoue, Y.(2001). Educational Research and Statistics: Examples of Questions and Answers. ERIC; ED 451242.

Kardiyen,F.; Olmus, H. (2016). A Comparison of Two Group Classification Approaches to Fat tailed and Skewed Data. Communication in Statistics-Simulation and Computation, 45, pp 17-32. Taylor & Francis, ISSN: 0361-0918.

Klecka, W. (1980). Discriminant Analysis. California: Sage Publications.

Kutner, M.; Neter, J. and Li, W.; Nachtsheim, C. (2005).Applied Linear Statistical Models.(5nd ed).Irwin, McGraw-Hill, New York, U.S.A.

Li, H.; Sun, j. (2011). Empirical Research of Hybridizing Principal Component Analysis with Multivariate Discriminant Analysis and Logistic Regression for Business Failure Prediction. Expert Systems with Application, (38).6244-6253. www.elsevier.com/locate/eswa.

Lydi-Anne, V. & Gaston, G.(2011). Psychosocial determinants of intention to abstain from drinking alcohol while pregnant among a sample of women of childbearing age. Addiction Research and Theory, 19 (2) pp128-137.

Manly, B. (1994). Multivariate Statistical Methods A Primer. 2nd, Chapman & Hall.

Maroco, J. ; Silva,D.; Rodrigues, A.; Guerreiro,M ; Santana,L. ; Mendonca, A. (2011).Data Mining Methods in the Prediction of Dementia : A Real – Data Comparison of the Accuracy , Sensitivity and Specificity of Linear Discriminant Analysis, Logistic Regression , Neural Networks, Support vector machines, Classification Trees and Random forests. BMC Research Notes, 4 ; 299. Portugal.

Moor, J. (1996). Stepwise Methods are as in Discriminant Analysis as They are Anywhere Else. Paper Presented at the Annual Meeting of the Southwest Educational Research Association.

Nath, R.; Jackson, W.(1992).A Comparison of the Classical and the Linear Programming approaches to the Classification Problem in Discriminate. Journal of Statistical Computation and Simulation, 41, pp73-93.

Osborne, J. (2015). Best Practices in Logistic Regression. SAG Publications, U.S.

Pluker, A. (1995). Application of Discriminant Analysis in Research with Gifted Students. Paper Presented at the Annual Meeting of the American Educational Research Association.

Pohar, M.; Blas, M.; Turk, S. (2004). Comparison of Logistic Regression and linear Discriminant Analysis: A simulation study. Metodoloski, V(1), N(1), pp143-161.

Rencher, C. (2002). Methods of Multivariate Analysis. 2 nd ed. Canada: Wiley Interscience.

Ruiz, S. (1991). Asymptotic Efficiency of Logistic Regression Relative to Linear Discriminant Analysis. Biometrika, V (78), 2,pp.235-243.

Santos, F.; Guyomarch, P.; Bruzek, J.(2014). Statistical Sex Determination from Craniometrics: Comparison of Linear Discriminant Analysis, Logistic Regression , and Support Vector Machines. Forensic Science International ,V245,(205) .www.elsevier.com/locate/forscint.

Tabachnick, B. & Fidell, L. (2013).Using Multivariate Statistics.(6 nd ed). Pearson Education, U.S.A.

Thompson, B.(1994). Why Multivariate Methods are Usually Vital in Research: Some Basic Concepts. Paper Presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, Jun 27-29)

Wilson, L. & Hardgrave, C. (1995). Predicting Graduate Student Success in an MBA Program: Regression Versus Classification. Educational and Psychological Measurement, V 55, No.2, April 1995 , pp186-195.