مجلة جامعة بنغازي العلمية (2019) مجلد 32 (2): صفحة 133-153
Scientific Journal of University of Benghazi (2019) vol. 32 (2):pp 133-153
http://sjuob.uob.edu.ly

قياس الاختلاف في التحصيل العلمي لطلبة جامعة بنغازي – فرع الأبيار حسب الأقسام

[a] نجية حامد الزوي و [b]ياسمينة بوزيد الفقيه
[a]قسم الرياضيات – كلية العلوم والآداب الأبيار– جامعة بنغازي– بنغازي– ليبيا
[b]قسم الإحصاء – كلية العلوم – جامعة بنغازي– بنغازي– ليبيا

الملخص

تهدف هذه الدراسة إلى تحديد العوامل التي تؤدي إلى اختلاف التحصيل العلمي لدى طلاب فرع الأبيار ولتحقيق هذا الهدف استُخْدم أسلوب الانحدار الخطي المتعدد _ باعتبار أن المعدل الدراسي هو المتغير التابع وبعض المتغيرات وهي العمر ،والجنس ، وسنة التخرج ، ونظام الدراسة ،والقسم ، ومكان الإقامة هي المتغيرات التوضيحية _ وقد تبين أن نموذج الانحدار الخطي المتعدد معنوي . وأظهرت الدراسة أن العمر والجنس يؤثران على المعدل الدراسي للخريجين بحيث يكون المعدل الدراسي الأعلى عند الخريجين صغار السن ، ويكون معدل الإناث أعلى من معدل الذكور .

كما تهدف الدراسة إلى تصنيف الطلبة الخريجين إلى مجموعات حسب المعدل الدراسي ولتحقيق هذا الهدف استُخْدم أسلوب التحليل التمييزي. ووجد أن العمر والجنس هما المتغيران الأكثر أهمية في دالة التمييز والأكثر مساهمة في التصنيف . من ثم إيجاد العلاقة بين معاملات الانحدار المتعدد ومعاملات فيشر المتحصل عليها من التحليل التمييزي وذلك باستخدام البيانات التي حصلنا عليها من نموذج فيشر لبناء نموذج خطي متعدد. ثم اختبرت معنوية هذا النموذج ومدى ملاءمته لمتغيرات الدراسة.

الكلمات المفتاحية : نموذج الانحدار المتعدد- التحليل التمييزي- معاملات فشر- الدالة التمييزية

Abstract

The first objective of this study is to identify the factors that may affect the differences in the educational achievement of the students of Al-Abyar University .To achieve this goal, the multiple linear regression method was used, considering that the academic rate is the dependent variable. Some variables, age, sex, year of graduation, study system, section, place of residence are explanatory variables. The linear regression model is significant. The study showed that age and sex affect the rate For graduates, where the highest academic rate is among young graduates, and the female rate is higher than that of males. The second objective is to classify the graduate students according to the academic average. To achieve the second objective, we used the method of discriminate analysis. We found that age and gender are the variables that are important in the degree of discrimination and most contribute to the classification. And then find the relationship between multiple regression coefficients and Fisher coefficients obtained from the discriminate analysis .We used the data we obtained from the Fisher model to construct a multiple linear model and then tested the significance of this model.

Keywords: Multiple Regression Model- Discriminate analysis - Fisher's Coefficients - Discriminate Function

1. مقدمة

يمنح فرع الأبيار _ الذي تم أُنشئ سنة 2002_ درجتي البكالوريوس والليسانس للطلبة المسجلين في مختلف الأقسام التابعة لكليتي الآداب والعلوم/ فرع الأبيار(الحاسوب، الرياضيات، الفيزياء، الجغرافيا، اللغة الإنجليزية، علم النفس، اللغة العربية). ويقاس التحصيل العلمي لطلبة فرع الأبيار بالنسبة المئوية وقد سلطت هذه الدراسة الضوء على العوامل التي قد تؤثر في معدل الطلبة وتؤدي إلى اختلاف نتائج التحصيل العلمي .

1.1. أهداف الدراسة

تحديد أهم العوامل المؤثرة في المعدل الدراسي لخريجي الأقسام المختلفة بفرع الأبيار وذلك بتقدير نموذج الانحدار الخطي المتعدد. ثم تصنيف الخريجين إلى مجموعتين وفق المعدل الدراسي (منخفض أقل من 65% – مرتفع 65% فأكثر) وذلك لمعرفة تأثير بعض العوامل على المعدلات الدراسية للخريجين باستخدام التحليل التمييزي ومن ثم توضيح كيفية تحليل الدوال التمييزية وتكوينها، كما تسلط هذه الدراسة الضوء على جوانب التشابه والاختلاف بين استخدام التحليل التمييزي وتحليل الانحدار الخطي المتعدد, وذلك بتوضيح إمكانية استخدام نموذج فيشر لإيجاد معاملات الانحدار الخطي المتعدد.

مجتمع الدراسة وعينتها2.1

تكون مجتمع الدراسة من خريجي كليتي الآداب والعلوم/ فرع الأبيار البالغ عددهم 1654 حيث بلغ عدد خريجي الآداب 1238 خريج وخريجة وعدد خريجي العلوم 416خريج وخريجة موزعين على 12 عاماً دراسياً (من 2006 الى2018) . وقد تم سُحبَت عينة طبقية وذلك بتقسيم المجتمع على أساس التوزيع النسبي وسحبت العينة بنسبة 0.30 من العدد الكلي لخريجي كل قسم وتم وسُحبَ بالطريقة العشوائية . اشتملت العينة على 500 خريج وخريجة منهم 375 من خريجي الآداب موزعة على الأقسام كالتالي (علم النفس 126, اللغة الإنجليزية121, اللغة العربية 45, الجغرافيا 83) و 125 من خريجي العلوم موزعة على الأقسام كالتالي (الحاسوب 50 ,الرياضيات 50, الفيزياء 25)

2 المواد والطرق

في هذه الدراسة سيُسْتَخْدَم اثنان من الأساليب الإحصائية هما تحليل الانحدار المتعدد والتحليل التمييزي.

1.2 تحليل الانحدار المتعدد

يعد تحليل الانحدار أسلوباً إحصائياً للتنبؤ بقيم متغير تابع باستخدام قيم مجموعة من المتغيرات التوضيحية ، كما أنه يمكن استخدامه لتقويم أثر المتغيرات التوضيحية على المتغير التابع .المتغيرات التي اسْتُخْدِمَتْ في هذه الدراسة هي معدلات الخريجين (المتغير التابع) هذا المتغير يمثل بقيم من 50 إلى 100 في تحليل الانحدار المتعدد أما العوامل وهي المتغيرات التوضيحية فهي(العمر، ومكان الإقامة، والجنس، والقسم، ونظام الدراسة، وسنة التخرج)

نموذج الانحدار المتعدد يعبر عن علاقة خطية بين متغير كمي تابع Y وبعض المتغيرات التوضيحية X1,X2,...Xk ويعبر عن تلك العلاقة بالنموذج التالي :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \qquad i = 1, 2, \ldots, n \qquad (1)$$

β_i :معاملات الانحدار n :عدد المشاهدات K : عدد المتغيرات التوضيحية

الثابت β_0 يمثل قيم Y عندما تنعدم كل المتغيرات التوضيحية X_i .

المعامل β_i هو مقدار التغير في المتغير Y عند تغير المتغير التوضيحي X_i بوحدة واحدة مع ثبات المتغيرات الأخرى

($E(u_i)$.**1**حد الخطء **العشوائي** u_i يمثل الفرق بين القيمة المقدرة $\hat{Y}_i$ والقيمة الحقيقية Y_i، الفروض الأساسية لنموذج الانحدار المتعدد

$= 0$ $i = 1, 2, \ldots, n.$ 2. $cov(x_i, x_j) = 0$ عندما $i \neq j$

3. $cov(u_i, u_j) = 0$ عندما $i \neq j$ $cov(u_i, u_j) = \sigma^2$ عندما $i = j$

تعتمد طريقة المربعات الصغرى الاعتيادية(OLS) على الحصول على مقدرات بحيث تُصَغر مجموع مربعات البواقي إلى أدنى قيمه له , طريقة المربعات الصغرى تعطينا مقدرات الانحدار $\hat{\beta}_i$

بالرجوع إلى المعادلة (1) يمكننا التعبير عن نموذج الانحدار المتعدد بالشكل التالي :

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \ldots + \beta_k X_{1k} + U_1$$

$$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + ... + \beta_k X_{2k} + U_2$$

$$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + ... + \beta_k X_{nk} + \mathrm{u}_n$$

يمكن كتابة المعادلات السابقة بطريقة المصفوفات كما يلي :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

ويعرف مقدر المربعات الصغرى OLS كالتالي :

$$\hat{\beta} = (X'X)^{-1} X'Y \qquad \text{(Rencher، 2008)}$$

من الجدير بالذكر أن أشهر القياسات المستخدمة لتقدير مدى ملاءمة خط الانحدار للمعلومات المستخدمة هو ما يعرف بمعامل تحديد الجودة ويشار إليه بـ R2 . وتتراوح قيمة هذا المعامل ما بين صفر و 1 . وهو يمثل نسبة تباين y المفسر بالنموذج ، أي بالمتغيرات التوضيحية. يقسم التباين الكلي للمتغير التابع إلى تباين مفسر وتباين غير مفسر

$$SST = SSE + SSR$$

معامل التحديد المتعدد R^2 يحسب كما يلي : $$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

عيب معامل التحديد المتعدد مؤشراً على قوة العلاقة أن قيمته تتأثر بعدد المتغيرات المفسرة k وعدد المشاهدات n فكلما زاد عدد المتغيرات في النموذج زادت قيمة هذا المعامل حتى لو لم يكن لها علاقة بالمتغير التابع, لذلك استخدمنا معامل التحديد

المتعدد المعدل R^2_{adj} حيث $$\overline{R}^2 = 1 - \left[\frac{n-1}{n-k} \left(1 - R^2\right) \right]$$

وهي قيمة قريبة المعنى من R^2 وأقل منها, ميزتها أنها خالية من تأثير k , n .

تظهر فائدة معامل التحديد المتعدد المعدل بالدرجة الأولى في المقارنة بين نماذج متعددة عند البحث عن أفضل نموذج من خلال تغيير المتغيرات التوضيحية التي تدخل فيه. تُخْتَبر معنوية النموذج إي الاستدلال على وجود علاقة في المجتمع يستند

إليها النموذج وذلك باختبار الفرضية التالية $H_0 : \beta_1 = \beta_2 ... = \beta_k = 0$ الفرضية البديلة في هذه الحالة هي اختلاف أحد المعاملات على الأقل عن الصفر. ويكون ذلك باستخدام اختبار F (الفقيه ،2016)

$$F = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/n-k-1} \rightarrow F_{k,n-k-1}$$

التحليل التمييزي2.2

يستخدم التحليل التمييزي في الدراسات التي تهدف إلى تصنيف الأفراد في مجموعات على أساس متغيرات كمية منبئة. (أبوعلام، 2003) . في عام (1936)أوجد Fisher دالة خطية لتصنيف المفردة إلى إحدى المجموعتين مع تساوي التباينات وأطلق عليها اسم الدالة المميزة الخطية ومن هنا جاءت فكرة استخدام الدالة المميزة الخطية للمجتمعات متعددة المتغيَرات فالفكرة الأساسية من التمييز هي التفرقة بين مجتمعات متداخلة أو متشابكة لها الخصائص نفسها. ويعد التمييز الذي يسمى بدالة فيشرFisher طريقة فعالة للفصل بين مجتمعين في المستقبل، وأيضا للحصول على أعلى تمييز بين المجموعات على أن تكون نسبة التباين بين المجموعات بالنسبة إلى التباين داخل المجموعات كبيرا ، وذلك من خلال إيجاد تجمعات خطية لمجموعة من المتغيرات (Johnson and Wichern,2007)

ويهدف التحليل التمييزي إلى تصميم التوليفات الخطية للمتغيرات الأفضل في موضوع الدراسة. والتحقق من مدى وجود فروق ذات دلالة بين المجموعات. ثم تحديد المتغيرات التي تسهم بأكبر قدر من الاختلاف بين مجموعات المتغير التابع. وتقسيم الحالات بين مجموعات المتغير التابع بناءً على قيم المتغيرات التوضيحية. وأخيرا تقويم دقة التصنيف (نسبةً مئويةً) (جودة، 2008).

و يشترط التحليل التمييزي أن تكون المجتمعات موضوع الدراسة منفصلة وقابلة للتحديد حتى إن كانت هذه المجتمعات متداخلة فيما بينها بدرجات معينة. أن تكون كل مفردة في كل مجتمع قابلة للوصف والتحديد بمجموعة من المقاييس أو المتغيرات المستقلة، وأن تكون جميع متغيرات دالة التمييز مقاسه بقيم محدودة وأن تختلف المجتمعات موضوع الدراسات بالنظر إلى أوساطها، أي أن تكون متجهات أوساط المتغيرات للمجتمعات غير متساوية. وكذلك تساوي مصفوفة التباين المشترك في المجتمعات الإحصائية محل الدراسة (Rencher, 2002) .وأن تُصَنف المشاهدات إلى مجموعتين أو أكثر للمتغير التابع. وأن تكون المتغيرات المستقلة موزعة توزيعاً طبيعياً، فإن لم يكن كذلك يمكن أن تكون قريبة من التوزيع

الطبيعي بحيث لا يوجد في هذه المتغيرات قيم شاذة تؤثر على نتائج التحليل. كما يشترط عدم وجود ارتباط بين المتغيرات التوضيحية الداخلة في تكوين دالة التمايز حتى يمكن تفسير النتائج , وتحديد المساهمة النسبية لكل متغير في القوة التمييزية الكمية. (نجيب والرفاعي ،2006) .

1.2.2 التمييز أو التصنيف في حالة مجموعتين:

في عام 1936 قام الباحث فيشر *fisher* باقتراح دالة التمييز الخطية للوصول إلى أكبر قدر من الفصل بين مجتمعين تابعين لنفس المصدر وذلك باستعمال عدد من المتغيرات المرتبطة .

تتلخص طريقة فيشر في تحويل المشاهدات المتعددة المتغيرات X إلى مشاهدات وحيدة التغير Y بحيث تُفْصَل قيم Y التي نحصل عليها من المجتمعين بقدر المستطاع وذلك باختيار توليفة خطية تعظم مربع المسافة بين μ_1,μ_2 منسوبة إلى تباين y . حيث تشير μ_1 إلى متوسط قيم Y التي نحصل عليها من قيم x التي تنتمي إلى المجتمع الأول و تشير μ_2 إلى متوسط قيم Y التي نحصل عليها من قيم X التي تنتمي إلى المجتمع الثاني. وتكون مصفوفة تباين وتباين مشترك متساوية لكلا المجتمعين وتساوي Σ .

ومن النادر طبعا أن نعرف قيم معالم المجتمع μ_2 , μ_1 , Σ . لذلك نفرض أن لدينا n_1 من المشاهدات المأخوذة من المجتمع π_1 عن المتجه العشوائي X´ = [X_1, X_2, ... , X_k] وأن لدينا n_2 من المشاهدات المأخوذة من المجتمع π_2 عن المجتمع نفسه ، يتكون كل مجتمع من k من المتغيرات المستقلة. (Rencher ،2002)

هذه التوليفة هي : $$Y = a'X$$

نجد أن الوسط الحسابي للعينة الأول والثاني على التوالي هو :

$$\overline{Y_2} = \sum_{i=1}^{n_2} \frac{Y_{2i}}{n_i} = a'\overline{X}_2 \quad \overline{Y_1} = \sum_{i=1}^{n_1} \frac{Y_{1i}}{n_i} = a'\overline{X}_1$$

والتباين المشترك للعينة هو : $$S_y^2 = a'S_{PL}\ a$$

ويُحْصَل على أفضل توليفة بتعظيم المسافة بين متوسطيY نسبة إلى تباينها $\frac{(\overline{Y}_1 - \overline{Y}_2)}{S_y}$ ويكون ذلك بتربيعها $\frac{(\overline{Y}_1 - \overline{Y}_2)^2}{S_y^2}$

$$\frac{(\overline{Y}_1 - \overline{Y}_2)^2}{S_Y^2} = \frac{[a'(\overline{X}_1 - \overline{X}_2)]^2}{a'S_{pl}a}$$

وتُعَظم المسافة بوضع $a = S_{pl}^{-1}(\overline{X}_1 - \overline{X}_2)$ وتكون $a' = (\overline{X}_1 - \overline{X}_2)' S_{pl}^{-1}$ وبذلك تصبح التوليفة $\frac{(\overline{Y}_1 - \overline{Y}_2)^2}{S_Y^2}$

$$\frac{(\overline{Y}_1-\overline{Y}_2)^2}{S_Y^2}=\frac{[a'(\overline{X}_1-\overline{X}_2)]^2}{a'S_{pl}a}=\frac{a'(\overline{X}_1-\overline{X}_2)(\overline{X}_1-\overline{X}_2)'a}{a'S_{pl}a}$$

$$\frac{(\overline{Y}_1-\overline{Y}_2)^2}{S_Y^2}=(\overline{X}_1-\overline{X}_2)'S_{pl}^{-1}(\overline{X}_1-\overline{X}_2)$$

S_{pl}^{-1} : تمثل معكوس مصفوفة التباين المشترك المقدر لمصفوفة المعالم المجهولة

معاملات دالة التمييز الخطية هي : $a=S_{pl}^{-1}(\overline{X}_1-\overline{X}_2)$

وعليه يمكن كتابة دالة التمييز الخطية وهي ما يعرف بدالة فيشر للتمييز بالشكل التالي(Rencher،2002) :

$$Y=a'X=(\overline{X}_1-\overline{X}_2)'S_{pl}^{-1}X$$

بما انه لدينا مجموعتان يكون لدينا دالة تمييز واحدة (حيدر وحسين، 2013)

2.2.2 التصنيف باستخدام دالة فيشر

يُحْصَلُ على نقطة المنتصف $\hat{m}$ بين متوسطي عينتي المتغير الواحد $(\bar{y}_1+\bar{y}_2)$ باستخدام العلاقة

$$\hat{m}=\frac{1}{2}(\bar{y}_1+\bar{y}_2)=\frac{1}{2}(\overline{X}_1-\overline{X}_2)'S_{pl}^{-1}(\overline{X}_1-\overline{X}_2)$$

إذا كانت $Y_0=(\overline{X}_1-\overline{X}_2)'S_{pl}^{-1}X_0\geq\hat{m}$ ضع X_0 في π_1 (Rencher,2002)

و إذا كانت $Y_0=(\overline{X}_1-\overline{X}_2)'S_{pl}^{-1}X_0<\hat{m}$ ضع X0 في π_2

3.2.2 اختبار التوزيع الطبيعي

اختبار الطبيعية اختبار مهم لمعرفة معنوية متغيرات التمييز ودالة التمييز

لمعرفة إذا كانت البيانات تتبع التوزيع الطبيعي أم لا نستخدم اختبار Kolmogrove–Smirnov لاختبار الفرضية

H_0 : البيانات لا تتبع التوزيع الطبيعي H_1 : البيانات تتبع التوزيع الطبيعي

ولكن في حالة المتغيرات العشوائية المتصلة اختبار الطبيعية يمكن إهماله أكثر من ذلك لأن الاختبار ليس مطلوباً أساسياً لاشتقاق دالة التمييز (Huberty,1994)

4.2.2 اختبار تساوي مصفوفة التباين والتباين المشترك للمجموعتين

التحليل التمييزي يفترض تساوي مصفوفة التباين والتباين المشترك للمجموعتين ولاختبار ذلك نختبر الفرضية التالية :

$H_0: S_1=S_2=.....S_k$ H_1: اثنان منهما على الأقل غير متساوية

نختبر هذه الفرضية باستخدام (Box's M-test statistical)

$$M = \ln|S_i|\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} n_i \ln|S_i|$$

وقد أثبت Box أنه إذا ضرب M في ثابت C^{-1} الذي يساوي :

$$C^{-1} = 1 - \frac{2p^2 + 3(p-1)}{6(p+1)(k-1)}\left[\sum_{i=1}^{k}\frac{1}{n_1} - \frac{1}{\sum_{i=k}^{k} n_i}\right]$$

نتوصل إلى مقياس يتوزع توزيع χ^2 بدرجة حرية $\frac{1}{2}(k-1)(p-1)$ عندما تكون n_i كبيرة

أي إن $MC^{-1} \sim \chi^2_{\frac{1}{2}(k-1)(p-1)}$

ومن عيوب هذا الاختبار أنه شديد الحساسية لعدم إتباع المتغيرات للتوزيع الطبيعي ولحجم العينة الكبير

(Krzanowski and Mariott,1994)

5.2.2 اختبار عدم وجود ارتباط خطي بين المتغيرات التوضيحية

ويتم ذلك بإيجاد قيم معامل تضخم التباين VIF وقيمة Tolernec لكل متغير من المتغيرات التوضيحية, فإذا كانت قيمةVIF أقل من5 هذا يعني عدم وجود ارتباط خطي بين المتغيرات المستقلة كذلك بإيجاد قيمة Tolerance وهو مقلوب VIF فإذا كانت قيمته أكبر من 0.2 إذا لا يوجد ارتباط خطي بين المتغيرات المستقلة.وتستخدم الصيغة التالية لإيجاد قيمة VIF (Gujarati,1988)

$$VIF = \frac{1}{(1-R_j^2)} \qquad j = 1,2,3...p$$

حيث إن p: تمثل عدد المتغيرات التوضيحية R_j^2 : تمثل معامل التحديد للمتغيرات التوضيحية

6.2.2 اختبار معنوية الدالة التمييزية الخطية

تُخْتَبَر معنوية الدالة التمييزية الخطية بعدة طرق منها

1.6.2.2 اختبار F

لاختبار ما إذا كان متوسطي المجتمعين مختلفين معنويا أي لاختبار الفرضية نستخدم اختبار F ويأخذ الصيغة التالية

$$F = \frac{(n_1 + n_2 - p - 1)}{(n_1 + n_2 - 2)p} T^2 \qquad \text{حيث} \qquad T^2 = \frac{n_1 n}{n_1 + n_2} D^2$$

D هو أكبر فاصل نسبي يمكن الحصول عليه باستخدام توليفة خطية من مشاهدات متعددة المتغيرات

n_1, n_2 حجم العينتين المأخوذتين والقيمة الجدولية لهذا الاختبار هي $F_{p, n_1+n_2-p-1,}$

ففي حالة قبول الفرضية الصفرية التي تنص على تساوي متوسطي المجموعتين هذا يعني أنه لا يوجد دالة تصنيف وعند قبول الفرضية البديلة التي تنص على عدم تساوي متوسطي المجموعتين هذا يعني أن الفصل بين المجتمعين معنوي (Rencher,2002)

1.6.2.2 مقياس ولكس لمبدا – Wilks–criteria

يحسب هذا المقياس وفق التالية الصيغة :

$$\Lambda = \frac{|W|}{|T|}$$

T: هي مصفوفة التباين والتباين المشترك الكلي للمجموعات W : هي مصفوفة التباين والتباين المشترك داخل المجموعات

تتراوح قيمة(Λ) بين الصفر والواحد الصحيح ، فإذا ا كانت قيمتها قريبة من الصفر فذلك يدل على قوة التمييز

أما إذا كانت قريبة أو مساوية للواحد فإن ذلك يشير إلى أن متوسطات المجموعة متساوية ولا يوجد دالة تمييز بين المجموعات،

2.6.2.2 مقياس χ^2

يعد هذا المقياس أكثر دقة من مقياس (Λ) :وتكون صيغته الرياضية كالتالي $\chi^2 = -N\ln(\wedge)$

ويكون مقارب χ^2 بدرجة حرية $P(K-1)$ *و قد طورت صيغته بواسطة* Barttlete فأصبحت بالشكل التالي :

$$\chi^2 = \left[N - 1 - \frac{1}{2}(p+k)\right]\ln(\wedge)$$ بدرجة حرية $P(K-1)$

7.2. 2 اختبار معنوية المتغيرات التمييزية

إن تقليل عدد المتغيرات في دالة التمييز يفيد في قياس المتغيرات ذات العلاقة المعنوية وذات التأثير الأكبر على موضوع الدراسة وللتقليل من صعوبة موضوع الدراسة وتعقيدها والتعرف على المتغيرات المستقلة ذات القوة التمييزية المعنوية سنختار المتغيرات المختلفة معنوية بين المجموعتين أي إيجاد المتغيرات التي وسطها يختلف في المجموعتين ويكون ذلك باختبار الفرضية التالية :

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

ستُخْتَبَرُ الفرضية باستخدام اختبار ويلكس لمبدا ولتقدير معنوية ويلكس لمبدا يمكن استخدام اختبار F بالصيغة التالية :

$$F = \left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{n_1 + n_2 \text{ - p -1}}{P}\right)$$ $F_{P,n_1+n_2\text{-p-1}}$

Λ تمثل نسبة التباين داخل المجموعات بالنسبة للتباين الكلي

حيث إن التباين داخل المجموعات هو مجموع مربعات الفرق بين الإحداثيات التمييزية الفردية ومراكز المجموعات.

8.2.2 إيجاد العلاقة بين معامل نموذج الانحدار المتعدد وتابع فيشر

العلاقة بين التحليل التمييزي لمجموعتين وتحليل الانحدار المتعدد لوحظت لأول مرة بواسطة

Fisher (1936) ثم Flury and Riedwyl (1985) (Rencher, 2002)

نوجد معامل نموذج لانحدار المتعدد $\widehat{\beta}_{px1}$ بدلالة معامل فيشر a_{px1} باستخدام العلاقة بين مربع معامل الارتباط R^2و $\widehat{\beta}_{px1}$

$$R^2 = \frac{T^2}{(n_1+n_2-2)+T^2} \qquad \text{(Rencher ، 2002)}$$

$$R^2 = \frac{D^2 n_1 n_2 / n_1+n_2}{(n_1+n_2-2)+D^2 n_1 n_2 / n_1+n_2}$$

$$R^2 = \frac{n_1 n_2 D^2}{n_1+n_2(n_1+n_2-2)+n_1 n_2 D^2}$$

$$R^2 = \frac{n_1 n_2 D^2}{n_1 n_2 D^2 + n_1^2 + n_2^2 + 2n_1 n_2 - 2n_1 - 2n_2}$$

$$R^2 = \frac{n_1 n_2 D^2}{n_1 n_2 D^2 + (n_1+n_2)^2 - 2(n_1+n_2)} = \frac{n_1 n_2 D^2}{n_1 n_2 D^2 + (n_1+n_2)(n_1+n_2-2)}$$

$$R^2 = \frac{n_1 n / n_1+n_2\, D^2}{n_1 n / n_1+n_2\, D^2 + (n_1+n_2-2)} = \frac{n_1 n / n_1+n_2\, D^2}{T^2 + (n_1+n_2-2)}$$

$$D^2 = (\overline{X}_1 - \overline{X}_2)' S_p^{-1} (\overline{X}_1 - \overline{X}_2)$$

$$R^2 = \frac{n_1 n / n_1+n_2\, (\overline{X}_1 - \overline{X}_2)' S_p^{-1} (\overline{X}_1 - \overline{X}_2)}{T^2 + (n_1+n_2-2)} \qquad (1)$$

$$R^2 = (\overline{X}_1 - \overline{X}_2)' \hat{\beta} \qquad (2)$$

وبمساواة (1) و(2)

$$(\overline{X}_1 - \overline{X}_2)' \hat{\beta} = \frac{n_1 n / n_1+n_2\, (\overline{X}_1 - \overline{X}_2)' S_p^{-1} (\overline{X}_1 - \overline{X}_2)}{T^2 + (n_1+n_2-2)}$$

$$\hat{\beta} = \frac{n_1 n / n_1+n_2\, S_p^{-1} (\overline{X}_1 - \overline{X}_2)}{T^2 + (n_1+n_2-2)} = \frac{n_1 n / n_1+n_2}{T^2 + (n_1+n_2-2)} a$$

$$\hat{\beta} = \frac{n_1 n}{n_1+n_2(n_1+n_2-2+T^2)} a$$

3 النتائج والمناقشة

طُبقَ نموذج الانحدار الخطي المتعدد لإيجاد العوامل المؤثرة في معدلات خريجي فرع الأبيار، واستخدام التحليل التمييزي لتصنيف الخريجين حسب المعدل الدراسي ، ثم إيجاد معامل الانحدار باستخدام معامل فيشر النتائج كالتالي

جدول (1) تصنيف بيانات العينة حسب المتغيرات التوضيحية والمتغير التابع

المتغيرات الوصفية		معدل التخرج		المتغيرات الكمية	معدل التخرج		
		أقل من 65	فأكثر 65			أقل من65	فأكثر 65
X_2 مكان الإقامة	داخل الأبيار-1	185	215	X_1 العمر	22-21	24	72
	خارج الأبيار-0	55	45		24-23	70	107
X_3 الجنس	ذكر-0	81	53		26-25	47	53
	انثى-1	159	207		28-27	48	21
X_4 القسم	حاسوب-1	26	24		30-29	28	4
	فيزياء-2	6	19		32-31	11	2
	رياضيات-3	9	41		34-33	2	1
	جغرافيا-4	55	28		36-35	2	0
	لغةعربية-5	23	22	X_6 سنة التخرج	2008-2006	54	59
	إنجليزي-6	57	64		2011-2009	59	52
	علم نفس-7	64	62		2014-2012	74	82
X_5 نظام الدراسة	نظام سنة-1	204	226		2017-2015	42	49
	نظام فصل-2	36	34		2018	11	18

جدول (2) نتائج نموذج الانحدار المتعدد

النموذج	معاملات غير معيارية		معاملات معيارية	t	مستوى المعنوية	الارتباط الخطي	
	B	Std.Error	Beta			Tolerance	VIF
الثابت	-1081.8	233.05		-4.642	.000		
X_1	-1.247	.136	-.388	-9.167	.000	.901	1.11
X_2	.039	.858	.002	.045	.964	.962	1.03
X_3	2.892	.684	.180	4.230	.000	.896	1.11
X_4	-.281	.178	-.066	-1.582	.114	.936	1.06
X_5	-1.342	1.171	-.056	-1.146	.252	.686	1.45
X_6	.587	.116	.250	5.040	.000	.655	1.52

X_6 ، سنة التخرجX_5 ، نظام الدراسة X_4 ، القسم X_3، الجنس X_2 ، مكان الإقامة X_1 العمر

- من الجدول (2) نلاحظ أن معادلة الانحدار هي $Y=-1081.886-1.247\ X_1+2.892\ X_3+.587\ X_6$
- من الجدول (2) نستدل على أن العمر والجنس وسنة التخرج كانت معنوية ($sig. \leq 0.05$)
- نلاحظ من الجدول (2) أن قيمة VIF أقل من5 هذا يعني عدم وجود ارتباط خطي بين المتغيرات التوضيحية

كما وجدنا أن معامل تحديد الجودة بين معدلات الخريجين والعوامل المؤثرة فيه (العمر ,الجنس , سنة التخرج) R^2 هو 0.20 ، وهو يدل على أن هناك ملاءمةً ضئيلة للغاية بين نموذج الانحدار والبيانات .

جدول (3) تحليل التباين

النموذج		مجموع المربعات	درجة الحرية	متوسط المربعات	F	Sig.
1	Regression	7110.591	6	1185.098	20.920	.000[b]
	Residual	27928.095	493	56.649		
	Total	35038.685	499			

من خلال جدول تحليل التباين يتضح أن نموذج الانحدار ذو دلالة معنوية وكانت نتائج التحليل التمييزي كالتالي :

جدول (4) اختبار إتباع البيانات للتوزيع الطبيعي

	العمر	مكان الإقامة	الجنس	سنة التخرج	القسم	نظام الدراسة
N	500	500	500	500	500	500
Kolmogorov-Smirnov Z	3.577	10.985	8.817	2.444	4.889	11.550
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

من خلال جدول (4) يتبينُ أن مستوى المعنوية لكل المتغيرَات أصغر من0.05 وعليه نقبل بالفرضية H0 أي إن البيانات لا تتبع التوزيع الطبيعي ولكن كما ذكرنا سابقا فإن حجم العينة الكبير والمتغيرات المتصلة لا تتطلب إتباع التوزيع الطبيعي

نلاحظ من الجدول(2) المستخدم في تحليل الانحدار أن قيمة VIF أقل من5 هذا يعني عدم وجود ارتباط خطي بين المتغيرات التوضيحية كذلك نجد أن قيمة Tolerance أكبر من 0.20 إذا لا توجد مشكلة ارتباط خطي

جدول (5) اختبار تساوي مصفوفة التباين

نتيجة الاختبار		
Box's M		57.611
F	Approx.	2.708
	df1	21
	df2	900001.280
	Sig.	.000

من الجدول(5) نلاحظ أن مستوى المعنوية لهذا الاختبار 0.000 وهي أقل من 0.05 وعليه هذا الاختبار له دلالة إحصائية أي إن مصفوفات التباين والتباين المشترك متساوية

جدول (6) اختبار معنوية المتغيرات التمييزية

	Wilks' Lambda	F	درجة الحرية1	درجة الحرية 2	مستوى المعنوية
X_1	.875	71.366	1	498	.000
X_2	.995	2.456	1	498	.118
X_3	.970	15.655	1	498	.000
X_4	.996	1.981	1	498	.160
X_5	.999	.382	1	498	.537
X_6	1.000	.117	1	498	.733

نلاحظ أن مستوى الدلالة للمتغيرات العمر والجنس هي 000 أقل من0.05 وبالتالي نرفض الفرضية الصفرية أي إن مُتَغيري العمر والجنس يمثلان أساس الفصل بين المجموعتين وهذا ما تؤكده قيمة لمبداويلكس التي قيمتها اقل من واحد صحيح

جدول (7) اختبار معنوية الدالة باستخدام χ^2

	Wilks' Lambda	Chi-square	درجة الحرية	مستوى المعنوية
1	.835	88.966	6	.000

يلاحظ من الجدول(7) أن قيمة لمبدا ويلكس تساوي 0.834 ، أي إن وجود المجموعات مبرر وهذا ما يؤكده اختبار χ^2

بمستوى دلالة يساوي 0.000 إذاً نرفض الفرضية الصفرية ونقبل الفرضية البديلة أي إن المجموعتين مختلفتان .

جدول (8) الإحصاءات المتعلقة بالمجموعات

المتغيرات	الوسط الحسابي	الانحراف المعياري	المتغيرات	الوسط الحسابي	الانحراف المعياري	المتغيرات	الوسط الحسابي	الانحراف المعياري
1			2			المجموع		
X_1	25.72	2.825	X_1	23.87	2.028	X_1	24.76	2.610
X_2	.77	.421	X_2	.83	.379	X_2	.80	.400
X_3	.69	.516	X_3	.87	.511	X_3	.78	.521
X_4	4.96	1.912	X_4	4.71	2.001	X_4	4.83	1.961
X_5	1.15	.358	X_5	1.13	.338	X_5	1.14	.347
X_6	2011.93	3.446	X_6	2012.03	3.701	X_6	2011.98	3.578

من خلال الجدول (8) الذي يمثل الإحصاءات المتعلقة بالمجموعات المستعملة في الدراسة المتمثلة في المتوسط الحسابي والانحراف المعياري للمتغيرات المستعملة في العينة لكل مجموعة على حدا نلاحظ أن الخريجين بمعدل ضعيف متوسط أعمارهم 25.72 بينما متوسط أعمار الخريجين بمعدل مرتفع87 .23 .

جدول (9) المعاملات

المتغيرات	المعاملات المعيارية لدالة التمييز القانونية	مصفوفة البنية	معاملات الدالة المعيارية
X_1	.898	.849	.368
X_2	-.088	-.398	-.221
X_3	-.366	.194	-.714
X_4	.220	.062	.112
X_5	.196	-.034	.563
X_6	-.447	.141	-.125
الثابت			240.346

من قيمة المعاملات المعيارية بالجدول (9) نلاحظ أن العمر أكثر أهميه من الجنس في تكوين دالة التمييز أي إنه المؤثر الأكبر في معدل تخرج الطالب كذلك من الجدول نلاحظ أن العمر يساهم بدرجة كبيرة في التمييز بين الطلبة ذوي المعدلات المرتفعة والمنخفضة، كما نلاحظ من قيمة المعاملات المعيارية للمتغيرات التوضيحية في الدالة التمييزية القانونية ومن معامل الارتباط بين المتغيرات التوضيحية والدالة داخل المجموعة (معاملات مصفوفة البنية)، أن دالة التمييز هنا هي العمر، وقد بلغت قيمتها (0.898) في الدالة المعيارية وفي مصفوفة بنية الدالة (0.849)، وعلى هذا الأساس نطلق على دالة التمييز هنا العمر يليها الجنس الذي بلغت قيمة دالته المعيارية (0.366-) ومعامل ارتباطه

مع الدالة التمييزية في مصفوفة بنية الدالة (0.194). من خلال الجدول يمكن كتابة الدالة التمييزية مع مراعاة المتغيرات التي تم أُقْصيتْ $Z= 240.34 + 0.368\ X_1 - 0.714\ X_3$

جدول (10) نسبة التباين المفسر

الدالة	الجذر الكامن	التباين %	% النسبة التراكمية	الارتباط القانوني
1	.197[a]	100.0	100.0	.406

من الجدول نلاحظ (10) أن نسبة التباين لدالة التمييز يقدر ب 100% كما نلاحظ أن قيمة الجذر الكامن تساوي 0.197 ، وهي تشير إلى نسبة التباين المفسر بين مجموعتي المعدلات المرتفعة والمنخفضة التي تعود إلى الفروق بينهما في دالة التمييز الوحيدة، وبلغ معامل الارتباط القانوني 0.406 ، وهو يشير إلى الارتباط بين الدالة التمييزية الوحيدة والمتغيرات المستقلة، وكان مربع الارتباط 0.164 ، وبالتالي يعود 16 % من التباين إلى الفروق بين المجموعتين في دالة التمييز.

جدول (11) دوال تمركز المجموعات

المعدل	الدالة
	1
منخفض	.463
مرتفع	-.427

يتبين من دوال تمركز المجموعة أن المجموعتين تقعان موقعاً معاكساً من بعضهما بعضاً . أي إن دالة التمييز "العمر"، هي الدالة التمييزية الأولى التي يمكن أن تميز بشكل جوهري بين مجموعة المعدلات المنخفضة والمعدلات المرتفعة . كما أن زيادة عمر الطلبة تؤدي إلى ارتفاع انضمامهم إلى مجموعة المعدلات المنخفضة .

جدول (12) التصنيف

		المعدل	المجموعة المتنبئ بها		المجموع
			منخفض	مرتفع	
المجموعة الأصلية	العدد	منخفض	148	92	240
		مرتفع	78	182	260
	%	منخفض	61.7	38.3	100.0
		مرتفع	30.0	70.0	100.0

التصنيف الجيد بنسبة 66.0%

يدلنا الجدول (12) على جودة التنبؤ بعضوية الجماعة باستخدام تقنية التحليل التمييزي. أي إن الجدول يقارن التصنيف الأصلي بالتصنيف المتحصل عليه باستعمال الدالة التمييزية . نلاحظ أنه من أصل 240 طالب أصابت الدالة في إعادة التصنيف بشكل صحيح ل 148 طالب بنسبة 30% بالنسبة للمجموعة الثانية أصابت في 182 طالب من أصل 260 بنسبة 70% إجمالا أصابت الدالة في إعادة التصنيف بشكل جيد بنسبة 66%

إيجاد العلاقة بين معامل نموذج الانحدار المتعدد وتابع فيشر

$$a = S_p^{-1}(\mu^1 - \mu^2)$$

$$a = \begin{pmatrix} 0.186 & 0.157 & 0.072 & -.027 & 0.031 & -.063 \\ 0.157 & 6.465 & 0.328 & 0.052 & -.046 & -.805 \\ 0.072 & 0.328 & 4.197 & 0.148 & 0.168 & -1.33 \\ -.027 & 0.052 & 0.148 & 0.122 & -.027 & -.653 \\ 0.031 & -.046 & 0.168 & -.027 & 0.280 & 0.116 \\ -.063 & -.805 & -1.33 & -.653 & 0.116 & 12.07 \end{pmatrix} \begin{pmatrix} 1.85 \\ -0.06 \\ -0.18 \\ -0.1 \\ 0.25 \\ 0.02 \end{pmatrix} = \begin{pmatrix} 0.331 \\ -0.189 \\ -0.641 \\ -0.112 \\ 0.105 \\ 0.508 \end{pmatrix}$$

وتكون التوليفة الخطية للدالة المميزة $y = a'x$

$$y = .331X_1 - .189X_2 - .641X_3 - .112X_4 + .105X_5 + .508X_6$$

ولإيجاد قيمة (Mahalanobis Distance) نتبع الخطوات التالية :

$$D^2 = (\mu^1 - \mu^2)' S_p^{-1}(\mu^1 - \mu^2) = 0.787$$

$$T^2 = \frac{n_1 n}{n_1 + n_2} D^2 \qquad T^2 = \frac{(226)(274)}{500} 0.787 = 97.468$$

$$\hat{\beta} = \frac{n_1 n}{n_1 + n_2(n_1 + n_2 - 2 + T^2)} a = \frac{(226)(274)}{500(500 - 2 + 97.468)} \begin{pmatrix} 0.331 \\ -0.189 \\ -0.641 \\ -0.112 \\ 0.105 \\ 0.508 \end{pmatrix} = \begin{pmatrix} 0.069 \\ -0.039 \\ -0.133 \\ -0.023 \\ 0.022 \\ 0.106 \end{pmatrix}$$

يمكننا كتابة دالة الانحدار الخطي بالشكل التالي :

$$y = 0.069x_1 - 0.039x_2 - 0.133x_3 - 0.023x_4 + 0.022x_5 + 0.106x_6$$

يقيس معامل التحديد نسبة إجمالي التغيير في المتغير التابع الممكن تفسيره بواسطة معادلة الانحدار وتدل القيمة التي حُصلَ عليها أن نسبة التغير في المتغير التابع المتحصل عليها من معادلة فيشر هي 16% وهي أقل من القيمة المتحصل عليها من معادلة الانحدار السابقة :

$$R^2 = \frac{T^2}{(n_1 + n_2 - 2 + T^2)} \qquad R^2 = \frac{97.468}{(500 - 2 + 97.468)} = 0.164$$

كذلك يمكن إيجاد قيمة معامل التحديد باستخدام القانون التالي : $R^2 = (\mu^1 - \mu^2)' \hat{\beta}$

$$R^2 = (1.85 \quad -0.06 \quad -0.18 \quad -0.1 \quad 0.25 \quad 0.02) \begin{pmatrix} 0.069 \\ -0.039 \\ -0.133 \\ -0.023 \\ 0.022 \\ 0.106 \end{pmatrix} = 0.164$$

4 الخلاصة

بعد تحليل البيانات بأسلوب الانحدار المتعدد وأسلوب التحليل التمييزي حصلنا على النتائج التالية

- أن نموذج الانحدار المتعدد معنوي وأن المتغيرات التوضيحية (العمر والجنس وسنة التخرج) لها أثر معنوي على المعدل الدراسي وأن المعدل الدراسي ينقص بزيادة العمر ويكون المعدل الدراسي أعلى عند الطالبات وكذلك يزداد المعدل بزيادة السنوات أي إن الخريجين في السنوات الأخيرة معدلهم أعلى من الخريجين من بداية افتتاح الجامعة .
- أن التحليل التمييزي لبيانات الدراسة أوضح أن العمر والجنس متغيران لهما تأثير معنوي على المعدل الدراسي وذلك عند تصنيف عينة الدراسة حسب المعدل الدراسي إلى مجموعتين. كذلك اخْتُبِرَتْ معنوية الدالة التمييزية الخطية ووجد أن المجموعتين مختلفتان. من قيمة المعاملات المعيارية وجد أن العمر أكثر أهميه من الجنس في تكوين داله التمييز أي إنه المؤثر الأكبر في المعدل الدراسي للخريجين . وتبين أن العمر يساهم بدرجة كبيرة في التمييز بين الطلبة ذوي المعدلات المرتفعة والمنخفضة أي إن زيادة عمر الطلبة تؤدي إلى ارتفاع انضمامهم إلى مجموعة المعدلات المنخفضة.

كذلك حُسبَ نموذج الانحدار ومعامل الارتباط المتعدد باستخدام دالة فيشر وكان النموذج معنوياً

المراجع

مصر ، القاهرة : دار النشر للجامعات [2]SPS"أبوعلام ، رجاء (2003) ،"التحليل الإحصائي للبيانات باستخدام

الفقيه ياسمينة بوزيد، (2016) " تحليل الانحدار والاقتصاد القياسي" الطبعة الأولى, مركز بحوث العلوم الاقتصادية ليبيا، بنغازي .

الطبعة الأولى الأردن ،عمان: دار وائل للنشر SPSS".التحليل الإحصائي المتقدم باستخدام) . "جودة ،صلاح (2008

حسين علي نجيب ، غالب عوض صالح الرفاعي (2006) "تحليل ونمذجة البيانات بأسلوب الحاسوب (تطبيق شامل للحزمة الطبعة الأولى ،الأهلية للنشر والتوزيع spss)

حيدر عزالدين , حسين يسف , استخدام التحليل التمييزي في تصنيف العملاء المصرفيين الاعتباريين ,دراسة ميدانية عن المصرف السوري ,مجلة تشرين للبحوث والدراسات العلمية , اللاذقية , سوريه , المجلد35, العدد4 ,(2013) ,

References

Fisher, R. A. (1936), "The Use of Multiple Measurement in Taxonomic Problems," Annals of Eugenics, 7, 179–188.

Flury, B., and Riedwyl, H. (1985), "*T* 2 Tests, the Linear Two-Group Discriminant Function and Their Computation by Linear Regression," *American Statistician*, **39**, 20–25.

Gujarati, D. N .(1988), "Basic Econometrics", McGraw-Hill Book Company ,New York

Huberty, .C.J. (1994), "Applied Discriminate Analysis" John Wiley and Sons , New York

Johnson, R. A., Wichern, D. W. (2007), "Applied Multivariate statistical analysis", sixth edition, Pearson, prentice Hall

Krzanowski, W.J., Marriott, F.C. ,(1994), "Multivariate Analysis" John Wiley and Sons , New York

Rencher, C. A. , (2002), "Methods of Multivariate Analysis 2nd Edition" , John Wiley and Sons – Brigham Young University

Rencher, C. A.,(2008), "Linear Models in Statistics" 2nd edition. John Wiley and Sons , New York