

مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل المميز الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

أ.م.د. صباح منفي رضا / كلية الادارة والاقتصاد / جامعة بغداد
أ.م.د. رباب عبد الرضا صالح / كلية الادارة والاقتصاد / جامعة بغداد
م. عادل عبد اللطيف / كلية الادارة والاقتصاد / جامعة بغداد

المستخلص

ان الهدف من الدراسة هو بيان القدرة التنبؤية الافضل بين أنموذج الانحدار اللوجستي والدالة المميزة الخطية بأستعمال البيانات الاصلية اولا ثم المركبات الرئيسية لتقليل الابعاد بين المتغيرات لبيانات المسح الاجتماعي والاقتصادي للاسرة لمحافظة بغداد لعام 2012 وتضمنت عينة البحث 615 مفردة لـ 13 متغير، 12 منها متغير توضيحي والمتغير المعتمد شمل العاملين والعاطلين عن العمل، تم اجراء المقارنة بين الطريقتين اعلاه واتضح من خلال المقارنة ان أنموذج الانحدار اللوجستي افضل من أنموذج الدالة المميزة الخطية بأستعمال البيانات الاصلية، اما بأستعمال المركبات الرئيسية تم تقليل المتغيرات الى 5 عوامل رئيسية بنسبة 62.875% من التباين الكلي وكانت النتائج متساوية للأنموذجي الانحدار اللوجستي والدالة المميزة الخطية. وان اداء أنموذج الانحدار اللوجستي للبيانات الاصلية تقريبا متساوي من استعمال المركبات الرئيسية بينما اداء أنموذج الدالة المميزة الخطية بأستعمال المركبات الرئيسية كان افضل من البيانات الاصلية.

المصطلحات الرئيسية للبحث: تحليل المركبات الرئيسية، الدالة المميزة الخطية، الانحدار اللوجستي، اختبار

هوتلنك - T^2 ، اختبار لاحصاءة Bartielts & Keiser .



مجلة العلوم
الاقتصادية والإدارية
العدد 95 المجلد 23
الصفحات 367 - 386

1-1 المقدمة

ان تقدم البلدان وتطورها يقاس من خلال بعض الظواهر التي تؤثر على المجتمع اذ تعد ظاهرة البطالة من اهم هذه الظواهر التي تقاس مدى التقدم والتطور الحاصل في البلدان وحظيت هذه الظاهرة باهتمام كبير ودراسات واسعة لمعرفة العوامل التي تؤثر على هذه الظاهرة والحلول التي تساعد على الحد من هذه الظاهرة اذ لا يوجد بلد خالي من هذه الظاهرة بصورة مطلقة ولكن بنسب متفاوتة حسب درجات التحدي والوقوف امام هذه الظاهرة وهذا يؤدي الى الاختلاف في وجهات النظر والتقييم وتعرف البطالة وفقا لتعريف المنظمة العمل الدولية " العاطل عن العمل هو كل من قادر عن العمل وراغب فيه ويبحث عنه ويقبله عند مستوى الأجر السائد ولكن دون جدوى " [1].

ومن اجل الوقوف على اهم العوامل التي تؤثر على البطالة والتنبؤ بالعاطلين عن العمل والعاملين وتصنيفهم تم استعمال بعض الاساليب الاحصائية المستعملة في البيانات الفئوية وهي التحليل العنبر وأنموذج الانحدار اللوجستي الثنائي.

2-1 مشكلة البحث

ان مشكلة البحث تكمن في تحليل البيانات التي يكون فيها متغير الاستجابة من النوع الثنائي او الرتبي اذ ان استعمال طرق الانحدار الاعتيادية تؤثر على عدم دقة النتائج في تقدير المعالم لذا يلجأ الباحثين الى استعمال التحليل العنبر وأنموذج الانحدار اللوجستي وتقليل الابعاد للمتغيرات التوضيحية تم استعمال المركبات الرئيسية كمدخلات للتنبؤ.

3-1 هدف البحث

ان الهدف من البحث هو تحسين القدرة التنبؤية للاساليب الاحصائية المستعملة في البيانات الفئوية وهي التحليل العنبر وأنموذج الانحدار اللوجستي الثنائي تم استعمال البيانات الاصلية اولا ومن ثم استعمال المكونات الرئيسية كمدخلات للتنبؤ والمقارنة بين الاسلوبين على اساس التصنيف الصحيح لمشاهدات المتغير المعتمد.

1-2 التحليل العنبر Factor Analysis [4,6,8]

وهو اسلوب احصائي الغرض منه تبسيط الارتباطات بين مختلف المتغيرات الداخلة في التحليل من اجل الوصول الى العوامل المشتركة التي تصف وتفسر العلاقة بين المتغيرات الداخلة في التحليل، والهدف منه وصف علاقات التباين بين عدد كبير من المتغيرات عن طريق عدد قليل من العوامل حيث يعتمد النموذج العنبر على اساس افتراض امكانية تجميع المتغيرات على اساس معاملات الارتباط بينهم وهذا يعني ان جميع المتغيرات الداخلة في مجموعة معينة ذو ارتباط عالي مع بعضها وقليل مع المجموعات الاخرى حيث ان كل مجموعة من المتغيرات تمثل عامل وهناك عدة طرائق للتحليل العنبر نذكر منها طريقة المركبات الرئيسية.

1-1-2 المركبات الرئيسية Principal Components [4,6,8]

ان تحليل المركبات الرئيسية يستعمل لاجاد مجموعة صغيرة من التراكيب الخطية للتباينات وتكون غير مرتبطة مع بعضها وهذه التراكيب الخطية تختار اعلى التباينات وان تحليل المركبات في حالة وحدات القياس متشابهة للمتغيرات تستعمل القيم الاصلية وعندما تختلف تحول الى القيم المعيارية وبعدها يتم حساب مصفوفة الارتباط تم ايجاد القيم الذاتية والمتجهات المميزة.
والمركبات الرئيسية تعتبر تراكيب خطية لـ P من المتغيرات

$$\begin{aligned}
 Y_1 &= a_1 X = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p \\
 Y_2 &= a_2 X = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p \\
 &\vdots \\
 Y_p &= a_p X = a_{p1} X_1 + a_{p2} X_2 + \dots + a_{pp} X_p
 \end{aligned}
 \quad (1)$$

Y : تمثل مصفوفة المكونات الرئيسية $Y_1 \dots Y_p$

a : تمثل مصفوفة المتجهات المميزة $a_1 \dots a_p$

X : تمثل مصفوفة المتغيرات التوضيحية $X_1 \dots X_p$

2-2 التحليل التمييزي [5,7,8] Discriminant Analysis

من الاساليب المهمة في متعدد المتغيرات والذي يهتم بتفريق مجموعة مختلفة من المفردات الى مجموعتين او اكثر من خلال ايجاد توافق خطية للمتغيرات التوضيحية تعرف بدالة التمييز التي من خلالها يتم تمييز المفردات ثم بعدها تتم عملية تصنيف المفردات الجديدة الى إحدى المجاميع بأقل خطأ ممكن وهناك عدة انواع لنماذج التحليل التمييزي منها الدالة المميزة الخطية.

1-2-2 الدالة المميزة الخطية لمجموعتين Linear Discriminant Function for Two Groups

في حالة التمييز الخطي لمجموعتين يتطلب تساوي مصفوفتي التباين للمجموعتين اما متجهات المتوسطات الحسابية تكون مختلفة وان الدالة المميزه الخطية هي عبارة عن تراكيب خطية يضم P من المتغيرات التي تعظم الفروق بين المتوسطات وتكتب الدالة المميزة الخطية ليفشر بالصيغة الاتية:-

$$Z = A'X$$

....(2)

$$A = S^{-1}(\bar{X}_1 + \bar{X}_2)$$

$$DL(Z) = Z - 1/2 A'(\bar{X}_1 + \bar{X}_2)$$

حيث ان :

Z : الدالة المميزه الخطية ليفشر

A' : متجه المعالم للدالة المميزة الخطية

\bar{X}_1 : موجه متوسطات المتغيرات للمجتمع الاول

\bar{X}_2 : موجه متوسطات المتغيرات للمجتمع الثاني

S^{-1} : معكوس مصفوفة التباين والتباين المشترك لجميع المتغيرات وللمجاميع كافة

$DL(Z)$: الحد الفاصل المقدر للدالة المميزة الخطية

حيث ان الموجه A يعظم الفروق بين متوسطي المجموعتين ثم بعدها تبدأ عملية تصنيف المفردات الجديدة للدالة المميزة أذ تصنف المفردة الجديدة الى المجتمع الاول اذا كانت $Z > K$ والى المجموعة الثانية اذا كانت $Z < K$ حيث ان

... (3)

$$K = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$$

K : كمية ثابتة تقلل من احتمال خطأ التصنيف وهي المنتصف ما بين متوسطي المجموعتين.

3-2 التحليل اللوجستي [5,7,8] Logistical Analysis

وهو اسلوب احصائي يهتم بتحليل البيانات التي يكون فيها متغير الاستجابة من النوع المتقطع ويستعمل أنموذج الانحدار اللوجستي لوصف العلاقة بين متغير الاستجابة والمتغيرات التوضيحية ويكون على نوعين أنموذج الانحدار اللوجستي ثنائي الاستجابة ومتعدد الاستجابة وسنركز على النوع الاول.

1-3-2 أنموذج الانحدار اللوجستي الثنائي Binary logistic regression Model

وهو نوع من تحليل الانحدار يكون فيها متغير الاستجابة مؤلف من مستويين فقط ويمكن كتابة أنموذج الانحدار اللوجستي بصيغة خطية بدلالة ما يعرف ب logit بالصيغة الاتية :

$$\log it(p_i) = \log\left(\frac{P_1}{1-p_1}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K$$

.... (4) OR

$$P_1 = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_K X_K)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_K X_K)}$$

ويسمى $\log \frac{P_1}{1-p_1}$ لوغارتم نسبة الافضلية وتسمى النسبة $\frac{P_1}{1-p_1}$ بنسبة الافضلية او افضلية النجاح.

4-2 بعض الاختبارات المهمة Some important tests

1-4-2 اختبار ^[8] Keiser Meyer Olijks & Bartielts

وهي من الاختبارات التي تستعمل في التحليل العاملي للتأكد من ان البيانات ملائمة للدراسة حيث تستعمل على مدى كفاية العينة واختبار مصفوفة الارتباط هل هي مصفوفة احادية ام لا حيث يشترط في التحليل العاملي بان مصفوفة الارتباط ليست مصفوفة احادية ويكون الاختبار حسب الفرضية الاتية :

H_0 : مصفوفة الارتباطات هي مصفوفة أحادية

H_1 : مصفوفة الارتباطات ليست مصفوفة احادية

فاذا كانت P-Value اصغر من مستوى الدلالة 0.05 يعني رفض فرضية العدم وقبول بالفرضية البديلة .

2-4-2 اختبار T^2 - hotelling ^[1]

لاختبار معنوية الدالة المميزة فاننا نختبر الفرضية الاتية:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

اي اختبار ان المجموعتين تعودان لنفس المجتمع ام لمجموعات مختلفة بافتراض تساوي مصفوفة التباين والتباين المشترك والمقاس المستعمل هو T^2 - hotelling

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2)$$

(5) ...

$$F = \frac{(n_1 + n_2 - p - 1)}{(n_1 + n_2 - 2)p} T^2$$

فاذا كانت F الحسابية اكبر من F الجدولية دل على ان الدالة المميزة لها القابلية على التميز والعكس صحيح اي انه لا يمكن تكوين الدالة المميزة لعدم وجود فروق بين متجهات المتوسطات للمجموعتين.

3-4-2 اختبار ^[8] Wilks Lambda

لاختبار وجود علاقة خطية بين مجموعتين من المتغيرات فاننا نختبر الفرضية التالية :

H_0 : لا توجد علاقة خطية بين مجموعتين من المتغيرات

H_1 : توجد علاقة خطية بين مجموعتين من المتغيرات

وأحصاء الاختبار هي :



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل المميز الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

$$(6) \dots \lambda = \frac{|W|}{|W + H|}$$

حيث ان:

W : مصفوفة التباين والتباين المشترك داخل المجاميع

$W+H$: مصفوفة التباين والتباين المشترك الكلي

والصيغة اعلاه تتوزع تقريبا توزيع χ^2 بدرجة حرية $P(K-1)$ فاذا كانت P -value اقل من 0.05 اي معنوية نرفض H_0 اي توجد علاقة خطية بين المتغيرات اي ان الدالة التمييزية لها القدرة على التمييز.

4-4-2 اختبار Omnibus chi-square^[8]

هذا الاختبار يوضح معنوية الانموذج اي يوجد علاقة بين المتغير المعتمد والمتغيرات التوضيحية وحسب الفرضية الاتية:

H_0 : النموذج غير معنوي

H_1 : النموذج معنوي

فاذا كانت P -value اقل من 0.05 اي معنوية نرفض H_0 اي ان النموذج معنوي .

5-4-2 اختبار Wald^[8]

لبيان اهمية ومعنوية المعلمات المقدرة بطريقة الامكان الاعظم بالنسبة للانموذج اللوجستي نختبر احصاءة $Wald$ وحسب الفرضية الاتية:

$$H_0 = \beta_K = 0$$

$$H_1 = \beta_K \neq 0$$

واحصاءة الاختبار هي:

$$(7) \dots W = \left(\frac{\beta}{S(\beta)} \right)^2$$

حيث β تمثل معلمة المتغير

$S(\beta)$ الخطأ القياس لمعلمة المتغير

فاذا كانت P -value اقل من 0.05 اي معنوية نرفض H_0 اي ان معاملات المتغير التوضيحي معنوية.

3- الجانب التطبيقي^[2] Application Part

1-3 مقدمة

من اجل تحقيق الغرض من البحث تم استعمال بيانات المسح الاجتماعي والاقتصادي للاسرة في العراق عام 2012 لمحافظة بغداد و باستعمال البرنامج الاحصائي spss20 وقد تم اختبار البيانات من الافتراضات الخاصة بالتحليل العاملي والانحدار اللوجستي والتحليل المميز وبعدها تم تطبيق الأساليب المذكورة اعلاه لاجل المقارنة بين الانحدار اللوجستي مع الدالة المميزة الخطية اولا بالبيانات الاصلية ثم تم حساب المركبات الرئيسية واجريت بعدها المقارنة.



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العمير الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

2-3 بيانات البحث

شملت عينة البحث من بيانات المسح الاجتماعي والاقتصادي للأسرة في العراق عام 2012 محافظة بغداد حيث طرحت عدة اسئلة من استمارة الاستبيان وتم تحديد المتغيرات الآتية :-

- X_1 : يمثل العمر حيث تم تحديد العمر من الفئة 15 سنة - 60 سنة
 X_2 : يمثل الحالة الاجتماعية حيث تم تحديد (متزوج=1، اعزب=2، مطلق=3، منفصل=4)
 X_3 : يمثل أعلى شهادة اكملها والدك
 X_4 : يمثل أعلى شهادة اكملتها والدتك
 X_5 : يمثل أعلى شهادة حصلت عليها
 X_6 : يمثل هل تستخدم الانترنت حيث (نعم=1، كلا=2)
 X_7 : يمثل هل تعاني من عجز مستمر مشخص طبياً حيث (نعم=1، كلا=2)
 X_8 : يمثل هل تعاني من مرض مزمن مشخص طبياً حيث (نعم=1، كلا=2)
 X_9 : يمثل عدد الاولاد حيث (واحد=1، لا يوجد=2، أكثر من واحد=3)
 X_{10} : يمثل نوع البيئة حيث (حضر=1، ريف=2)
 X_{11} : يمثل حجم الأسرة الذي يتمثل بعدد افراد الأسرة
 X_{12} : يمثل الجنس حيث (ذكر=1، انثى=2)
Y: يمثل المتغير المعتمد حيث (يعمل بدوام كامل=0، لا يعمل=1)
- لا شهادة=1، فقط يقرأ=2، يقرأ ويكتب=3، ابتدائية=4، متوسطة=5، اعدادية=6، مهني=7، دبلوم=8، بكالوريوس=9، دبلوم عالي=10، ماجستير=11، دكتوراه=12

3-3 نتائج التحليل العاملي

1- في البداية تم اختبار (KMO and Bartlett's) للكم على مدى كفاية العينة واختبار مصفوفة الارتباط هل هي مصفوفة الدية ام لا حيث يشترط في التحليل العاملي بان مصفوفة الارتباطات لا تكون مصفوفة أحادية حيث من الجدول رقم (1) نلاحظ بان قيمة (Kaiser-Meyer) بلغت (0.574) وهي اكبر من قيمة الحد الأدنى لهذا الاختبار والبالغة (0.50) وهذا يعني بإمكاننا الحكم بكفاية حجم العينة في هذا التحليل.

جدول رقم (1) KMO and Bartlett's test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.574
Bartlett's Test of Sphericity	Approx. Chi-Square	1304.083
	Df	66
	Sig.	.000

اما بالنسبة الى اختبار بارتليت (لاثبات بان مصفوفة الارتباطات هي مصفوفة غير احادية) فيجب اختبار الفرضية التالية:

فرضية العدم : مصفوفة الارتباطات هي مصفوفة احادية

الفرضية البديلة : مصفوفة الارتباطات ليست مصفوفة احادية

ومن خلال قيمة بارتليت والبالغة (1304.083) وبمستوى دلالة (0.00) وهي اصغر بكثير من مستوى المعنوية (0.05) وهذا يعني رفض فرضية العدم وقبول الفرضية البديلة والقائلة بان مصفوفة الارتباطات ليست أحادية.

من الجدول رقم (2) يمثل القيم الأولية والمستخلصة للاشتركاكات والقيم الأولية للاشتركاكات تؤخذ مساوية للواحد في طريقة المكونات الرئيسية في حالة اعتماد مصفوفة الارتباط اذ ان القيمة المستخلصة للمتغير الاول تشير الى 0.855 من التباينات في قيم المتغير الاول (شهادة الام) تفسرها العوامل المشتركة (تم استخلاص خمسة عوامل) وان قيمة الاشتركاكية تتراوح ما بين الصفر والواحد وهي تعبر عن مربع الارتباط المتعدد للمتغير الاول مع العوامل وهكذا لبقية المتغيرات ونلاحظ ان جميع قيم المستخلصة لاشتركاكية المتغيرات تفسر نسبة عالية من التباين ماعدا المتغير (العجز المشخص) حيث يعاني من نسبة قليلة في حصوله لاشتركاكية احد المتغيرات وهذا يشير الى عدم اهمية هذا المتغير.



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنبر الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

جدول رقم (2) Communalities

	Initial	Extraction
العمر	1.000	.755
الحالة الزوجية	1.000	.745
شهادة الاب	1.000	.733
شهادة الام	1.000	.855
الشهادة	1.000	.643
الانترنت	1.000	.528
العجز المشخص	1.000	.271
المرض	1.000	.570
الاولاد	1.000	.691
البيئه	1.000	.588
عدد الافراد	1.000	.598
الجنس	1.000	.567

2- جدول التباين الكلي المفسر : من الجدول رقم (3)
جدول رقم (3) التباينات الكلية المفسره

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.462	20.514	20.514	2.462	20.514	20.514	2.458	20.48	20.48
2	1.489	12.406	32.92	1.489	12.406	32.92	1.42	11.837	32.317
3	1.381	11.51	44.43	1.381	11.51	44.43	1.373	11.438	43.754
4	1.193	9.939	54.369	1.193	9.939	54.369	1.21	10.084	53.838
5	1.021	8.507	62.875	1.021	8.507	62.875	1.084	9.037	62.875
6	0.951	7.923	70.798						
7	0.86	7.166	77.964						
8	0.788	6.567	84.532						
9	0.697	5.81	90.341						
10	0.619	5.158	95.5						
11	0.336	2.796	98.296						
12	0.204	1.704	100						

Extraction Method: Principal Component Analysis.

نلاحظ ان المكون الرئيسي الاول الذي له اكبر جذر كامن يساوي 2.462 يفسر تقريبا 20.514 من التباينات الكلية لمتغيرات البطالة و المكون الثاني له جذر كامن 1.489 ومقدار ما يفسره من التباينات اكثر من 12 من التباينات وبذلك فان المكون الاول والثاني يفسران معا اكثر من 32 هيكل التباينات التراكمي للمتغيرات التي عددها 12 متغير ، والمكون الثالث له جذر كامن 1.381 وهو يفسر اكثر من 11% من التباينات وبذلك فان المكونات الثلاثة تفسر نسبة 44.43 من هيكل التباينات للمتغيرات و المكون الرابع له جذر كامن 1.193 وهو يفسر اكثر من 9% من التباينات وبذلك فان المكونات الاربعه تفسر نسبة 54.369 من هيكل التباينات للمتغيرات و المكون الخامس له جذر كامن 1.021 وهو يفسر اكثر من 8% من التباينات وبذلك فان المكونات الخمسة تفسر نسبة 62.875 من هيكل التباينات للمتغيرات، تم اهمال بقية المكونات ذات الجذور المميزة التي تقل عن الواحد وفقا لمعيار كيسر Kaiser اما النسبة المتبقية والبالغة (37.125) تعود لمتغيرات خارجية او متغيرات غير مسيطر عليها من قبل الباحث



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنبر الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

3- جدول مصفوفة المكونات: هنا سوف يتم عرض جدولين أحدهما نتائج قبل التدوير والآخر نتائج بعد التدوير وسيكون عرض المتغيرات المهمة فقط والتي نسب تشبعها أكبر من (0.35) أي المتغيرات المؤثرة في الدراسة وحسب أهميتها في العامل ثم بعد ذلك تسمية هذه العوامل وكانت النتائج كما يلي :

ا- نتائج قبل التدوير

جدول (4) يبين مصفوفة المركبات قبل التدوير

Component Matrix

	Component				
	1	2	3	4	5
العمر	-.739-	.446	.047	.053	.072
الحالة الزوجية	.151	.319	-.281-	.187	-.712-
شهادة الاب	.835	.142	.040	-.111-	-.042-
شهادة الام	.898	.210	.008	.055	.032
الشهادة	-.032-	.127	.780	-.108-	-.070-
الانترنت	-.117-	-.169-	-.639-	.237	.144
العجز المشخص	.010	-.411-	.200	-.055-	-.244-
المرض	.411	-.599-	.182	.092	.030
الاولاد	.352	.692	.020	.238	.178
البيئة	.113	-.267-	-.052-	.708	-.024-
عدد الافراد	.261	-.014-	-.353-	-.480-	.419
الجنس	.004	.046	.283	.521	.461

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

من الجدول أعلاه والذي يبين نتائج قبل التدوير نلاحظ بأن أغلب المتغيرات توزعت على العامل الاول والعامل الثاني وبقية العوامل توزعت بشكل مبعثر على بقية العوامل حيث ظهرت في بعض العوامل متغير واحد او متغيرين في حين يجب ان يحتوي العامل على الاقل ثلاثة متغيرات ليؤخذ كعامل مؤثر ، لذا سوف نعتد على نتائج بعد التدوير الذي توزعت على العوامل بشكل أفضل.

ب- نتائج بعد التدوير :

من جدول رقم (5) نلاحظ من مصفوفة المركبات المدورة ان اقوى المتغيرات ارتباطا بالمكون الاول كانت لشهادة الام اذ تشبعه 0.892 يليه شهادة الاب بتشبعه 0.836 ثم متغير العمر بتشبع 0.744- واقل المتغيرات هو هل تعاني من العجز المشخص بتشبع 0.009 وتعكس التشبعات السالبة لبعض المتغيرات الى اتجاه تأثير هذه المتغيرات وتفاعلاتها تشير بالاتجاه المعاكس حيث ان زيادة في تشبع متغير شهادة الام يؤدي الى زيادة تشبع في نسب كل من المتغيرات ذات التشبعات الموجبة وانخفاض التشبع في نسب المتغيرات ذات التشبعات السالبة وزيادة تشبع متغير العمر يرافقه نقصان في نسب المتغيرات ذات التشبعات الموجبة وعليه نستطيع تسمية المركبة الاولى بشهادة الوالدين والعمر.

المركبة الثانية اقوى المتغيرات ارتباطا بها كان متغير عدد الاولاد اذ تشبعه 0.727 ومتغير هل تعاني من مرض مزمن بتشبع 0.46- واقل المتغيرات هو استخدامك للانترنت بتشبع 0.028. ويمكن ان تسمى بمركبة عدد الاولاد والمرض المزمن.

المركبة الثالثة اقوى المتغيرات ارتباطا بها هي الشهادة الحاصل عليها بتشبع موجب 0.794 واستعمال الانترنت بتشبع سالب 0.708- واقل المتغيرات ارتباطا بها هو هل تعاني من مرض مزمن بتشبع 0.006. ويمكن ان تسمى بمركبة الشهادة الحاصل عليها.

المركبة الرابعة اقوى المتغيرات ارتباطا بها هي البيئة بتشبع 0.702 والجنس بتشبع 0.616 واقل المتغيرات ارتباطا هو الحالة الزوجية بتشبع 0.03- ويمكن ان تسمى بمركبة البيئة .

المركبة الخامسة اقوى المتغيرات ارتباطا بها هي الحالة الزوجية بتشبع 0.842 واقل المركبات ارتباطا بها هي استعمالك للانترنت بتشبع 0.004- ويمكن ان تسمى بمركبة الحالة الزوجية.



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنبر الخطي
بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

جدول رقم (5) Rotated Component Matrix^a

	Component				
	1	2	3	4	5
العمر	-.744-	.428	.127	-.031-	.039
الحالة الزوجية	.122	.107	-.096-	-.030-	.842
شهادة الاب	.836	.098	.108	-.089-	.077
شهادة الام	.892	.215	.055	.062	.074
الشهادة	-.040-	-.044-	.794	.053	-.079-
الانترنت	-.115-	.028	-.708-	.114	-.004-
العجز المشخص	.009	-.499-	.135	.029	.052
المرض	.412	-.546-	.006	.248	-.200-
الاولاد	.336	.727	.124	.158	.100
البيئه	.083	-.143-	-.231-	.702	.119
عدد الافراد	.299	.125	-.319-	-.470-	-.414-
الجنس	-.011-	.238	.111	.616	-.344-

اما المتغيرات التي لم تدخل في اي مركبة فكانت متغير العجز المشخص ومتغير عدد الافراد حيث تعتبر هذه المتغيرات غير مؤثره (غير مهمة).

3-4 التحليل المميز

قبل البدء بالتحليل المميز هناك بعض الاختبارات التي ينبغي التأكد من تحقيقها منها معنوية الدالة المميزة وان العلاقة بين المتغيرات التوضيحية تكون خطيه ولا يوجد تعدد خطي بين المتغيرات التوضيحية وانها تتوزع توزيعا طبيعيا وفيما يأتي بعض هذه الاختبارات

1- لمعرفة معنوية الدالة المميزة الخطية ومن خلال اختبار (hotalling - T^2) وحسب المعادلة رقم (5) نجد ان قيمة T^2 كالاتي :

$$T^2 = 181.19$$

اما قيمة F المحسوبة فكانت $F = 13.66$ وهي اكبر من قيمتها الجدولية (13,601) F والبالغة (1.75) لذا نرفض فرضية العدم وهذا يعني ان الدالة المميزه لها امكانية جيدة للتمييز
2- من خلال اختبار VIF حول وجود التعدد الخطي اتضح انه لا يوجد تعدد خطي حيث ان قيمة VIF لكل المتغيرات المستقلة اقل من 5.0 وكما في الجدول رقم (6)



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنبر الخطي
بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

جدول رقم (6) يوضح اختبار VIF

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	2.973	.317		9.391	.000		
الجنس	.014	.029	.019	.505	.614	.968	1.033
العمر	.004	.002	.145	2.547	.011	.439	2.280
الحالة الزوجية	.017	.013	.050	1.262	.207	.912	1.097
شهادة الاب	-.001	.004	-.017	-.313	.754	.482	2.073
شهادة الام	.007	.005	.110	1.650	.099	.323	3.092
الشهادة	-.018	.008	-.086	-2.076	.038	.841	1.190
الانترنت	.135	.120	.044	1.121	.263	.918	1.089
العجز المشخص	-.159	.082	-.075	-1.940	.053	.956	1.046
المرض	-.054	.040	-.057	-1.357	.175	.797	1.255
الاولاد	-.144	.019	-.374	-7.642	.000	.596	1.677
البينه	.105	.032	.127	3.246	.001	.928	1.077
عدد الافراد	.001	.003	.017	.428	.669	.895	1.117

1-4-3 الدالة المميزة الخطية للبيانات الاصلية

من خلال الجدول رقم (7) ومن خلال اختبار Wilks' Lambda نجد ان المتغيرات (شهادة الام الاولاد ، شهادة الاب، الشهادة ، البينه ، الانترنت) تمتاز بمعنوية عالية ولها تاثير كبير من حيث التفرقة بين المجموعتين اما باقي المتغيرات ليس لها تاثير معنوي .

جدول رقم (7) Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
العمر	.997	2.065	1	613	.151
الحالة الزوجية	.997	1.822	1	613	.178
شهادة الاب	.983	10.292	1	613	.001
شهادة الام	.980	12.255	1	613	.000
الشهادة	.988	7.520	1	613	.006
الانترنت	.991	5.574	1	613	.019
العجز المشخص	.997	1.666	1	613	.197
المرض	.998	1.205	1	613	.273
الاولاد	.916	56.384	1	613	.000
البينه	.990	6.420	1	613	.012
عدد الافراد	1.000	.071	1	613	.789
الجنس	1.000	.050	1	613	.822

من الجدول رقم (8) تم اختبار معنوية الدالة المميزة الخطية حيث توجد قوه للتمييز من خلال التباين بين المجموعتين ولكافة المتغيرات قد فسرت 100% من التباين ومعامل الارتباط القانوني 0.374 .



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنبر الخطي
بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

جدول رقم (8) Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.163 ^a	100.0	100.0	.374

a. First 1 canonical discriminant functions were used in the analysis.

من خلال الجدول رقم (9) وللتأكد من وجود علاقة خطية بين المتغيرات تم استعمال مقياس wilks lambda حيث نجد قيمتها 0.86 ونجد معنوية Chi-square فالدالة التمييزية تكفي لتفسير التباين بين المجموعتين اي توجد علاقة خطية بين المتغيرات .

جدول رقم (9) Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.860	91.493	12	.000

من الجدول رقم (10) نجد معاملات الدالة المميزة الخطية فالدالة المميزة الخطية هي :

$$Y = -1.038 - 0.034 X_1 - 0.132 X_2 + 0.011 X_3 - \dots - 0.144 X_{12}$$

جدول رقم (10) معاملات الدالة المميزة

	Function
	1
العمر	-.034-
الحالة الزوجيه	-.132-
شهادة الاب	.011
شهادة الام	-.059-
الشهادة	.138
الانترنت	-1.060-
العجز المشخص	1.247
المرض	.428
الاولاد	1.130
البينه	-.829-
عدد الافراد	-.010-
الجنس	-.114-
(Constant)	-1.038-



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنبر الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

جدول رقم (11) دالة تصنيف المجاميع

العمل	Function
	1
2.00	.925
3.00	-.175-

من الجدول رقم (11) نجد مايلي

$$\mu^{\wedge} = 1/2(0.925 - 0.175)$$

حيث ان قاعدة التصنيف

IF $Y \geq \mu^{\wedge}$ تصنف الى المجتمع الاول واذا كانت $Y \leq \mu^{\wedge}$ تصنف الى المجتمع الثاني
من الجدول رقم (12) يوضح النسبة المئوية لتصنيف مشاهدات المتغير المعتمد
جدول رقم (12) يوضح النسبة المئوية لتصنيف مشاهدات المتغير المعتمد

	العمل	Predicted Group Membership		Total
		2	3	
Original	Count	2	94	98
		3	511	517
	%	2	95.9	100
		3	98.8	100

a. 83.7% of original grouped cases correctly classified.

من الجدول اعلاه نجد ان هناك $515=511+4$ مشاهدة صنف بشكل صحيح وان احتمال التصنيف الصحيح كان $83.7\% = 515/615$ وان 100 مشاهدة فقط صنف بشكل خاطئ وان احتمال التصنيف الخاطئ كان 16.3% .

4-3-2 الدالة المميزة بأستعمال المركبات الرئيسية

تم استعمال المركبات الرئيسية لتقليص الابعاد لبيانات البطالة في محافظة بغداد والبالغ عددها 5 مركبات وسيتم تطبيق الاختبارات الاحصائية المتعلقة بالدالة المميزة على المركبات الرئيسية، من الجدول رقم 13 ومن خلال اختبار wilks lambda نجد ان المركبة الاولى والثالثة تمتاز بمعنوية عالية في التميز ولها تاثير كبير في التفرقة بين المجموعتين

جدول رقم(13) Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
pc1	.988	7.741	1	613	.006
pc2	1.000	.232	1	613	.630
pc3	.991	5.280	1	613	.022
pc4	.999	.423	1	613	.516
pc5	.996	2.641	1	613	.105



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العنقري الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

من الجدول رقم 14 تم اختبار معنوية الدالة المميزة الخطية حيث توجد قوه للتمييز من خلال التباين بين المجموعتين ولكافة المتغيرات قد فسرت 100% من التباين ومعامل ارتباط القانون 233.

جدول رقم (14) Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.058 ^a	100.0	100.0	.233

من الجدول رقم (15) تم اختبار وجود علاقة خطية بين المتغيرات من خلال اختبار wilks lambda حيث نجد قيمتها 0.945 ونجد معنوية chi- square اي توجد علاقة خطية بين المجموعتين من المركبات
جدول رقم Wilks' Lambda (15)

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.945	34.215	5	.000

من خلال الجدول رقم (16) نجد معاملات الدالة المميزة الخطية للمركبات.
الدالة المميزة الخطية للمركبات هي:

$$Y = -1.829 + 0.077 * pc1 + \dots - 0.527 * pc5$$

جدول رقم (16) معاملات الدالة المميزة

	Function
	1
pc1	.077
pc2	.147
pc3	.216
pc4	.117
pc5	-.527-
(Constant)	-1.829-

Unstandardized coefficients

من الجدول رقم (17) نجد مايلي

$$\mu^{\wedge} = 1/2(0.551 - 0.104)$$

جدول رقم (17) مجاميع الدالة

Y	Function
	1
2.00	.551
3.00	-.104-

من الجدول رقم (18) يوضح النسبة المئوية لتصنيف مشاهدات المتغير المعتمد

جدول رقم (18) يوضح النسبة المنوية لتصنيف مشاهدات المتغير المعتمد

	Y	Predicted Group Membership		Total
		2	3	
Original	Count	2	3	
	%	2	3	
		1	97	98
		2	515	517
		1	99	100
		0.4	99.6	100

a. 83.9% of original grouped cases correctly classified.

من الجدول اعلاه نجد ان هناك $516=515+1$ مشاهدة صنف بشكل صحيح وان احتمال التصنيف الصحيح كان 83.9% وان 99 مشاهدة فقط صنف بشكل خاطئ وان احتمال التصنيف الخاطئ كان 16.1%.

5-3 الانحدار اللوجستي

1-5-3 الانحدار اللوجستي بأستعمال البيانات الاصلية

ان اول خطوة في التحليل اللوجستي هو ان نقوم بتضمين الحد الثابت فقط في الانموذج في الخطوة الصفرية ثم بعدها يتم اضافة المتغيرات التوضيحية للاستدلال على اثرها مجتمعه في متغير الاستجابة لتحديد كفاءة الانموذج ككل، من الجدول رقم (19) تم استعمال طريقة الامكان الاعظم بطريقة الدورات التكرارية لمشتقات دالة الامكان من خلال الحصول على اقل قيمة لسالب ضعف لوغارتم دالة الامكان وقد حصلنا في الدورة الرابعة لمشتق سالب ضعف دالة الامكان وهي مساوية الى 539.465 وتوقفنا في هذه الدورة لان التغير في معاملات الحد الثابت اصبح اقل من 0.001 واعتبرنا معالمها افضل

جدول رقم (19) الدورات التكرارية بالخطوة صفر

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	547.420	1.363
	2	539.523	1.637
	3	539.465	1.663
	4	539.465	1.663

من الجدول رقم (20) يظهر النسبة المنوية للتصنيف الصحيح لمشاهدات المتغير المعتمد حيث ان النسبة المنوية للتصنيف الصحيح بلغت 84.1%.

جدول رقم (20) يوضح النسبة المنوية للتصنيف لمشاهدات المتغير المعتمد

Observed	العمل	Predicted		Percentage Correct
		العمل		
		2.00	3.00	
Step 0	2.00	0	98	.0
	3.00	0	517	100.0
Overall Percentage				84.1



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل المميز الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

من الجدول رقم (21) يوضح اختبار معنوية الانموذج بوجود علاقة بين المتغير المعتمد (البطالة) والمتغيرات التوضيحية الأخرى حيث كانت نسبة الاختبار معنوية وهذا يدل على جودة التوفيق للانموذج الكامل اي ان النموذج ملائم عند ادخال المتغيرات التوضيحية والحد الثابت

جدول رقم (21) اختبار Test Omnibus chi-square

	Chi-square	Df	Sig.
Step 1 Step	145.994	38	.000
Block	145.994	38	.000
Model	145.994	38	.000

من الجدول رقم 22 يوضح قيمة سالب ضعف لو غارتم دالة الامكان والتي بلغت 393.472 وهي اقل من القيمة للانموذج الذي يتضمن الحد الثابت فقط مما يدل على جودة الانموذج اما قيمة R-Square تشكل تقريبا 0.362.

جدول رقم(22) يوضح قيمة سالب ضعف لو غارتم دالة الامكان

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	393.472 ^a	.211	.362

بينما الجدول رقم (23) يوضح النسبة المنوية للتصنيف الصحيح بعد دخول المتغيرات التوضيحية بالانموذج وهي اعلى من الخطوة الصفرية حيث تبلغ قيمتها 86.8%.

جدول رقم (23) يوضح النسبة المنوية للتصنيف لمشاهدات المتغير المعتمد

Observed	Predicted		
	العمل		Percentage Correct
	2.00	3.00	
Step 1 العمل 2.00	29	69	29.6
3.00	12	505	97.7
Overall Percentage			86.8

ومن الجدول 24 يوضح اهم المتغيرات الداخلة بالنموذج وهي (العمر، الحالة الزوجية لشهادة، عدد الاولاد، البيئة) اما باقي المتغيرات فهي غير معنوية.



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل العميز الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

جدول رقم 24 يوضح المتغيرات الداخلة في أنموذج

جدول رقم 23 يوضح المتغيرات الداخلة بأنموذج								
Variables in the Equation								
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
العمر	0.063	0.018	11.918	1	0.001	1.065	1.028	1.104
الحالة الزوجية			12.629	5	0.027			
(1) الحالة الزوجية	-16.998-	40194.067	0	1	1	0	0	
(2) الحالة الزوجية	-15.416-	40194.067	0	1	1	0	0	
(3) الحالة الزوجية	1.027	43235.33	0	1	1	2.794	0	
(4) الحالة الزوجية	-527-	56842.219	0	1	1	0.59	0	
(5) الحالة الزوجية	-17.162-	40194.067	0	1	1	0	0	
شهادة الاب			5.123	9	0.824			
(1) شهادة الاب	0.405	0.471	0.74	1	0.39	1.5	0.596	3.774
(2) شهادة الاب	0.088	1.348	0.004	1	0.948	1.092	0.078	15.326
(3) شهادة الاب	-242-	0.521	0.216	1	0.642	0.785	0.283	2.179
(4) شهادة الاب	-042-	0.687	0.004	1	0.951	0.959	0.249	3.685
(5) شهادة الاب	20.386	8582.66	0	1	0.998	7.14E+08	0	
(6) شهادة الاب	1.459	1.208	1.46	1	0.227	4.304	0.403	45.914
(7) شهادة الاب	-1.293-	1.335	0.938	1	0.333	0.275	0.02	3.756
(8) شهادة الاب	-391-	2.1	0.035	1	0.852	0.677	0.011	41.493
(9) شهادة الاب	16.651	40192.97	0	1	1	17032347	0	
شهادة الام			4.571	8	0.802			
(1) شهادة الام	-215-	0.479	0.202	1	0.653	0.806	0.316	2.06
(2) شهادة الام	0.427	1.703	0.063	1	0.802	1.533	0.054	43.185
(3) شهادة الام	-1.158-	0.719	2.597	1	0.107	0.314	0.077	1.284
(4) شهادة الام	0.939	1.177	0.636	1	0.425	2.557	0.254	25.699
Step 1 ^a								
(5) شهادة الام	17.348	17477.72	0	1	0.999	34203950	0	
(6) شهادة الام	-1.836-	41099.111	0	1	1	0.159	0	
(7) شهادة الام	18.275	27356.609	0	1	0.999	86451037	0	
(8) شهادة الام	-092-	2.352	0.002	1	0.969	0.912	0.009	91.568
الشهادة			10.799	7	0.148			
(1) الشهادة	1.666	0.73	5.214	1	0.022	5.289	1.266	22.098
(2) الشهادة	0.852	0.721	1.398	1	0.237	2.345	0.571	9.629
(3) الشهادة	0.852	0.76	1.256	1	0.262	2.344	0.529	10.394
(4) الشهادة	-19.698-	40192.97	0	1	1	0	0	
(5) الشهادة	1.741	0.993	3.071	1	0.08	5.702	0.814	39.955
(6) الشهادة	1.529	1.487	1.057	1	0.304	4.613	0.25	85.033
(7) الشهادة	0.921	0.881	1.092	1	0.296	2.512	0.446	14.132
الانترنت	0.761	1.07	0.506	1	0.477	2.141	0.263	17.431
العجز المشخص	-1.286-	1.08	1.417	1	0.234	0.276	0.033	2.297
المرض	-346-	0.393	0.775	1	0.379	0.708	0.328	1.528
الاولاد			41.456	2	0			
(1) الاولاد	3.552	0.567	39.28	1	0	34.866	11.483	105.864
(2) الاولاد	3.161	1.169	7.306	1	0.007	23.589	2.384	233.401
اليينه	0.962	0.337	8.139	1	0.004	2.617	1.351	5.068
عدد الافراد	0.02	0.029	0.483	1	0.487	1.02	0.964	1.08
الجنس	0.301	0.277	1.182	1	0.277	1.352	0.785	2.327
Constant	14.003	40194.067	0	1	1	1206469.7		

a. Variable(s) entered on step 1: العمر، الحالة الزوجية، شهادة الاب، شهادة الام، الشهادة، الانترنت، العجز المشخص، المرض، الاولاد، اليينه، عدد الافراد، الجنس.



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل المميز الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

2-5-3 الانحدار اللوجستي بأستعمال المركبات الرئيسية

من الجدول رقم (25) يوضح اختبار معنوية الانموذج بوجود علاقة بين المتغير المعتمد (البطالة) والمركبات حيث كانت نسبة الاختبار معنوية وهذا يدل على جودة التوفيق للانموذج الكامل اي ان النموذج ملائم عند ادخال المركبات والحد الثابت

جدول رقم (25) Test Omnibus chi-square

	Chi-square	Df	Sig.
Step 1 Step	37.420	5	.000
Block	37.420	5	.000
Model	37.420	5	.000

من الجدول رقم 26 يوضح قيمة سالب ضعف لو غارتم دالة الامكان والتي بلغت 502.046 وهي اقل من القيمة للانموذج الذي يتضمن الحد الثابت فقط مما يدل على جودة الانموذج اما قيمة R-Square تشكل تقريبا 0.101 .

جدول رقم (26) يوضح قيمة سالب ضعف لو غارتم دالة الامكان

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	502.046 ^a	.059	.101

الجدول رقم (27) يوضح النسبة المنوية للتصنيف الصحيح بعد دخول المركبات بالانموذج وهي اقل بقليل من الخطوة الصفرية حيث تبلغ قيمتها 83.9% وهي اقل من النسبة المنوية بعد دخول المتغيرات التوضيحية للانموذج.

جدول رقم (27) يوضح النسبة المنوية للتصنيف لمشاهدات المتغير المعتمد

Observed	Y	Predicted		Percentage Correct
		Y		
		2.00	3.00	
Step 1 Y	2.00	1	97	1.0
	3.00	2	515	99.6
Overall Percentage				83.9

جدول رقم 28 يوضح جميع المركبات داخلة بالانموذج

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
pc1	-.053	0.01	25.873	1	0	0.948	0.929	0.968
pc2	-.117	0.044	7.114	1	0.008	0.89	0.817	0.97
pc3	-.119	0.062	3.63	1	0.057	0.888	0.786	1.003
pc4	-.129	0.07	3.347	1	0.067	0.879	0.766	1.009
pc5	0.656	0.222	8.738	1	0.003	1.927	1.247	2.977
Constant	3.054	0.548	31.056	1	0	21.201		



مقارنة بين أنموذج الانحدار اللوجستي وأنموذج التحليل المميز الخطي بأستعمال المركبات الرئيسية لبيانات البطالة لمحافظة بغداد

4- الاستنتاجات

- 1- من خلال اختبار KMO وجود علاقة معنوية بين المتغيرات ومن ثم فإن البيانات المستعملة في البحث مناسبة لإجراء التحليل العاملي.
- 2- تم تقليص المتغيرات والبالغ عددها 12 الى 5 مركبات رئيسية وتم تسمية كل مركبة (المركبة الاولى الشهادة للوالدين والعمر، المركبة الثانية عدد الاولاد والمرضى المزمن، المركبة الثالثة الشهادة والانترنت، المركبة الرابعة البيئة والجنس، المركبة الخامسة الحالة الزوجية).
- 3- تم اجراء بعض الاختبارات الخاصة بالتحليل المميز وتم التاكيد من معنوية الدالة المميزة باختبار hotelling وتم التاكيد من عدم وجود التعدد الخطي وتم أستعمال اختبار wilk للتأكد من وجود علاقة خطية بين المتغيرات التوضيحية.
- 4- نتيجة التصنيف الصحيح للمتغيرات الاصلية بلغت 83.7% بشكل عام .
- 5- تم استعمال المركبات الرئيسية واتضح ايضا وجود علاقة خطية بين مجموعتين من المركبات.
- 6- نتيجة التصنيف الصحيح للمركبات ارتفعت بشكل قليل عن البيانات الاصلية حيث بلغت 3.9%.
- 7- تم استعمال طريقة الامكان الاعظم بطريقة الدورات التكرارية حيث بلغت في الخطوة صفر بوجود الحد الثابت فقط بـ 539.465 في حين بلغت بالخطوة الاولى بوجود المتغيرات التوضيحية بـ 393.472 وهذا يدل على جودة الانموذج بوجود المتغيرات التوضيحية .
- 8- نسبة التصنيف الصحيح للخطوة صفر 84.1% بينما ارتفعت للخطوة الاولى وكانت 86.8%.
- 9- تم استعمال المركبات الرئيسية بدلا من المتغيرات واتضح ان النموذج ملائم عند استعمال المركبات حيث بلغت سالب ضعف لو غارت دالة الامكان 502.049 وهي اقل من الخطوة صفر لكنها اكبر من استعمال المتغيرات بالخطوة الاولى مما يدل على ان جودة الانموذج باستعمال المتغيرات الاصلية افضل من استعمال المركبات للانحدار اللوجستي.
- 10- نسبة التصنيف الصحيح للخطوة صفر 84.1% بينما انخفضت بقليل للخطوة الاولى وكانت 83.9%.
- 11- نسبة التصنيف الصحيح باستعمال المركبات الرئيسية متساوية للدالة المميزة والانحدار اللوجستي.
- 12- اهم المتغيرات المعنوية الداخلة بالانموذج للمتغيرات الاصلية بالنسبة للانحدار اللوجستي هي العمر والحالة الزوجية والشهادة وعدد الاولاد والبيئة اما باقي المتغيرات فهي غير معنوية اما بالنسبة للمركبات فجميعها داخلة بالانموذج .

5- التوصيات

- 1- استعمال الطرق الحصينة للمركبات الرئيسية والمقارنة بين الدالة المميزة الخطية والانحدار اللوجستي باستعمال المركبات الرئيسية .
- 2- اجراء دراسة شاملة حول البطالة في العراق وباستعمال الاساليب الاحصائية المتقدمة لمعرفة اهم المتغيرات التي تسبب البطالة حيث اقتصرنا الدراسة على محافظة بغداد فقط.

المصادر

- 1- الطائي، خالد ضاري عباس وعتاب، محمد رمضان "استخدام الدالة المميزة التربيعية في تمييز انماط الارقام العربية " جامعة النهرين، المجلد 14، العدد 1، 2011
- 2- عبيد، عهود جبارة "البطالة ومشكلاتها بين النظرية والتطبيق في مدينة بغداد " جامعة بغداد، مجلة البحوث التربوية والنفسية العدد (40) / 2014
- 3- وزارة التخطيط، الجهاز المركزي للاحصاء "المسح الاجتماعي والاقتصادي لاسره في العراق " 2012
- 4- Aguilera, A., Escabias, M., . Valderrama, M.(2006)' Using principal components for estimating logistic regression with high- dimensional multicollinear data Computational Statistics & Data Analysis 50 (2006) 1905–1924 www.elsevier.com/locate/csda



- 5- Kong-lai,Z., jing,L. (2010)' Studies of Discriminant Analysis and Logistic Regression Model Application in Credit Risk for China's Listed Companies' MANAGEMENT SCIENCE AND ENGINEERING Vol. 4, No. 4, 2010, pp. 24-32.
- 6- Milewska,A. , Jankowska,D. Citko,D. Więsak,T. AcacioM((2014) 'The Use of Principal Component Analysis and Logistic Regression in Prediction of Infertility Treatment Outcome'STU DIES IN LOGIC, GRAMMAR AND RHETORIC 39 (52) 2014
- 7- Ramtohol, P., N. Mamode Khan and M. Hosenally (2014) 'Factors Influencing Online Shopping in Mauritius: An Application of Principal Component Analysis and Binary Logistic Regression' Journal of Social Sciences 2014, 10 (4): 179.184.
- 8- Suleiman,S., Suleman.,I. Usman,U., Salami, Y. O,(2014)' Predicting an Applicant Status Using Principal Component, Discriminant and Logistic Regression Analysis'*International Journal Of Mathematics And Statistics Invention (IJMSI), Volume 2 Issue 10.*



A comparison between the logistic regression model and Linear Discriminant analysis using Principal Component unemployment data for the province of Baghdad

Abstract

The objective of the study is to demonstrate the predictive ability is better between the logistic regression model and Linear Discriminant function using the original data first and then the Home vehicles to reduce the dimensions of the variables for data and socio-economic survey of the family to the province of Baghdad in 2012 and included a sample of 615 observation with 13 variable, 12 of them is an explanatory variable and the depended variable is number of workers and the unemployed.

Was conducted to compare the two methods above and it became clear by comparing the logistic regression model best of a Linear Discriminant function written using the original data, either using Principal Component was reduced variables to 5 key factors by 62.875% of the total variance and the results were equal . That the performance of a logistic regression equal to using the original data and Principal Component, while performing a Linear Discriminant function using Principal Component was better than the original data.

Keywords: Principal Components analysis, Linear Discriminant Function, Logistic Regression, test T^2 – hoteling, test Keiser Meyer Oljkins & Bartielts.