تطبيق دالة التوزيع اللوجستية (احادية البعد)

في دراسة تأثير الفوارق العمرية ما بين الزوجين على استمرار حياتهم الزوجية[+]

Application of the logistic dist. Function –univariate- for studying the effect of age-variations among couples on continuity of their lifes

صباح فرج عبد الحسين[*]

المستخلص :-

تعد الدالة اللوجستية أكثر الدوال ملاءمة لوصف البيانات التي تكون الاستجابة فيها ثنائية ويكون فيها المتغير التوضيحي ذا مستويات متعددة إلا أن تقدير المربعات الصغرى لمعلمات هذه الدالة يصطدم بعقبتين رئيسيتين هما : لا خطية هذه المعلمات وعدم تجانس تباينات الاستجابات المشاهدة مما يؤدي في حالة عدم تجاوزهما إلى الحصول على تقديرات متحيزة ولا تتوزع توزيعا طبيعيا وليس لها أقل تباين ممكن ، وبذلك تنتفي الفائدة من استخدام هذه الدالة . ولحسن الحظ فأن تجاوز هاتين العقبتين ليس عسيرا وذلك لسهولة التحويل الخطي للدالة اللوجستية وسهولة تحقيق التجانس في تباينات الاستجابات المشاهدة ، ولذلك استخدمت هذه الدالة بكثرة في الدراسات البيولوجية ، وهذا البحث هو محاولة لاستخدام دالة الانحدار اللوجستية في مجال اخر هو المجال الاجتماعي بالتصدي لمشكلة ازلية كبرى يعيشها الانسان في حياته الاجتماعية ، وهي مشكلة الطلاق ، حيث تم فيه دراسة معنوية واحد من أهم عوامل عدم توافق الزوجين والتي تؤدي إلى انفصالهما ، ألا وهو – الفارق في عمر الزوجين – ومن ثم تحديد المدى المقبول (غير المعنوي) لهذا الفارق من أجل سلامة الحياة الزوجية.

Abstract :-

Logistic Function is regarded as the best one to fit the data with binary responses and multiple levels of the explanatory variable. But there are two obstacles on the way of least squares estimation for the parameters : nonlinearity of the parameters and the hetero scedasiticity of the responses – variances . If the estimation is done without overcoming the two obstacles , the estimators will be biased , non – normally distributed and having no minimum variances . In this case there is no advantage of using the logistic function . Luckily , the overcoming of the two obstacles is not difficult because the parameters of the logistic function can be easily transformed to be linear and the responses – variances to be homoscedasitic. So the logistic function is used many times in biological studies. This research is a trial to use the logistic regression function in another field, that is, social life. The researcher studied the significance of the differences in couples- age as a divorce–factor and he could determine acceptable difference range for the safety of married life .

المقدمة

ان الزواج هو شريعة الله سبحانه وتعالى للإنسان خليفته في الأرض ، لكي يبني ويعمر ويديم الحياة حتى قيام الساعة ، لذلك فإن أي خلل يصيب الحياة الزوجية سينعكس سلباً على بنائها وازدهارها ، وسيتسبب في تخريب الأسرة – نواة المجتمع – وتشريد أبنائها . وهذا يعني ان اختيار الشريك الملائم هو حجر الزاوية في استمرارية ونجاح الحياة الزوجية ، ولذلك فقد أدلى الجميع بدلوهم لتوصيف هذا الشريك ، من علماء الدين وعلماء الاجتماع إلى علماء النفس والفلسفة ، واعتقد انه قد آن الأوان لكي يدلي الباحثون الاحصائيون بدلوهم في هذا المعترك ويحددوا بالأرقام معنوية عوامل عدم توافق الزوجين والمدى المقبول (غير المعنوي) لكل منها ، وبالرغم من ان الباحث سبق له وان درس

[+] تاريخ استلام البحث ٢٠٠٩/١٠/١٥ ، تاريخ قبول النشر ٢٠١٠/٤/٢٠

[*] مدرس /معهد الإدارة / الرصافة

العوامل المؤثرة في الحياة الزوجية مستخدماً التحليل المميز ، فأنه في هذا البحث يدرس عاملاً واحداً ومهماً في توافق الزوجين ألا وهو الفارق في عمر الزوجين " مستخدماً في هذه المرة النموذج اللوجستي ، وذلك للأسباب التالية :

١. أن متغير الاستجابة في الحياة الزوجية هو متغير وهمي ثنائي Dummy Binary Response ، وعادة ما يشار إليه بالرقم (1) عند تحقق الاستجابة ، وبالرقم (0) عند عدم تحققها ، وان النموذج اللوجستي Logistic Model هو النموذج الأكثر ملاءمة لوصف وتحليل البيانات الثنائية Binary Data .

٢. إن البحث قد تناول احتمال الاستجابة للانفصال عن الحياة الزوجية المشتركة عند تسعة مستويات للمتغير التوضيحي (الفارق في عمر الزوجين) . وقد لوحظ من خلال البحوث السابقة ان النموذج اللوجستي يكون أكثر تمثيلاً للبيانات عندما يكون عدد المشاهدات كبيراً أو عندما يكون عدد مستويات المتغير التوضيحي اكبر من ثلاثة مستويات ، بخلاف نموذج المنحى الطبيعي الذي تقل كفاءته إذا زاد عدد مستويات المتغير التوضيحي عن ثلاثة مستويات ويعد الباحث verhulst أول من استخدم الدالة اللوجستية Logistic function وذلك في العام ١٨٣٨ لوصف نمو المجتمع واسماها دالة النمو growth function ، وفي العام ١٩٢٠ قام الباحثان (Pearl & Reed) [1] باستخدام الدالة لحساب نمو السكان ، وقد أطلق الباحثان فيما بعد عليها اسم الدالة اللوجستية Logistic Function بدلاً من دالة النمو ، وفي العام ١٩٤٤ قام الباحث (Berkson) باختبار الفرضية التالية [2] :

(ان استجابة مجموعة من الأشخاص لجرعة معينة (X) من دواء ما تتبع التوزيع الطبيعي) وطبق دالة المنحني الطبيعي والدالة اللوجستية بعد ان حولها إلى دالة خطية على مجموعة من البيانات واختبر حسن مطابقة كل منهما باستخدام مربع كاي (X^2) لحسن المطابقة، واستنتج الباحث من نتيجة الاختبار ان الأفضلية للدالة اللوجستية ، ولاسيما عندما يكون عدد المشاهدات كبيراً جداً وعندما تكون البيانات ثنائية وأعاد الباحث نفسه في العام ١٩٥١[3] دراسته للمقارنة ما بين المنحني اللوجستي وتكامل المنحنى الطبيعي وعلى أساس ان دالة المنحنيين هي (p) التي تمثل احتمال الاستجابة عند الجرعة (X) فتوصل إلى قناعة تامة بأفضلية استخدام المنحنى اللوجستي بالرغم من تقارب النتائج مع المنحنى الطبيعي وذلك لسببين أساسيين :

١. سهولة تطبيق المنحنى اللوجستي مقارنة بتكامل المنحى الطبيعي خاصة عندما تتوزع البيانات حسب توزيع ثنائي الحدين Binomial distribution.

٢. ان تقديرات معالم الدالة اللوجستية أفضل من تقديرات معالم تكامل المنحى الطبيعي وذلك من حيث الكفاءة efficiency والكفاية sufficiency .

وفي العام ١٩٦٧ طبق الباحثان (Chamber, D. R. Cox) نموذج اللوجستيك والمنحنى الطبيعي على مجموعة من البيانات الثنائية لثلاثة مستويات من الجرعة لدواء معين فلاحظ ان الدالة اللوجستية تمثل البيانات بصورة أفضل ، وذلك باستخدام اختبار مربع كاي ، واثبتا أيضاً انه عندما تكون هناك أكثر من ثلاث مستويات للجرعة فان كفاءة المنحنى الطبيعي تقل / وفي العام ١٩٧٠ استخدم الباحث (Dr. R. Cox)[4] نموذج انحدار اللوجستيك لتحليل البيانات الثنائية واجرى اختباراً لحسن المطابقة للنموذج باستخدام اختبار مربع كاي ، ومن خلال النتائج التي توصل اليها تبين له ان تقدير النموذج اللوجستي في حالة البيانات الثنائية جيد.وفي العام ١٩٧٢ قدم (Ashton)[5] كتاباً عرض فيه أغلب الأفكار التي نشرت حول الحالة اللوجستية وتحويلها الخطي ، كما عرض بعض المنحنيات المستخدمة في الابحاث الحياتية مثل منحنى التكامل الطبيعي الذي قارن مع المنحنى اللوجستي وتوصل الى ان المنحنيين متقاربان مع الافضلية للمنحني اللوجستي .

وفي العام ١٩٨٣ قام الباحثان (Mccullagh. P. , Nelder. J. A.)[6] باستخدام اختبارين لحسن المطابقة وهما اختبار مربع كاي واختبار (Deviance) الذي هو لوغارتم نسبة الارجحية لبعض النماذج المستخدمة في التجارب

الحياتية ومنها النموذج اللوجستي ، ولاحظ الباحثان ان كلا الاختبارين يقتربان من توزيع مربع كاي ، كما استخدم الباحثان طريقة المربعات الصغرى الموزونة نظراً لعدم ثبات تباين المشاهدات . وفي العام ١٩٨٧ قدم الباحثان (Richard Ray , Little)[7] عرضاً لبعض النتائج التي أظهرت ملاءمة النموذج اللوجستي للبيانات الثنائية ، وذلك باستخدام التحويل الخطي لمتجه المتغيرات التوضيحية Explanatory variables . وفي العام ١٩٨٩ قدم الباحثان (Lemestom , Hosmer)[8] طريقة لاختيار المجموعة الجزئية المناسبة للنموذج اللوجستي وذلك باستخدام الاحصاءة التالية

$$c(q) = \frac{SSE\ (q)}{SSE\ (p)/(n - p - 1)} + 2(q + 1) - n$$

حيث ان :

SSE(q) : مجموعة مربعات البواقي للنموذج الملائم المحتوي على q من المتغيرات.

SSE(p) : مجموعة مربعات الاخطاء التي تظهر من نظام الانحدار الخطي .

فكلما كانت c(q) ذات قيمة صغيرة كلما كان النموذج المستخدم لـ q من المتغيرات هو الافضل .وفي العام ١٩٩٥ صدر للباحث Minard Scott. كتابه الموسوم Applied Logistic Regression Analysis [٩] والمنشور على الانترنت والذي تضمن تطبيقات كمية عديدة في العلوم الاجتماعية. وفي العام ١٩٩٩ اجرى الباحثان (Jereny M. G. Taylor, A. L. Siqueira)[10] تحويراً على النموذج اللوجستي المستخدم لبيانات الاستجابة الثنائية، وذلك لدراسة تأثير المعالجة treatment effect وذلك بإضافة متغير ثنائي يدل على المعالجة هو المتغير I بمعامل γوبمتغير X مستمر يخضع لتحويل القوة المعروف لـ Box –cox فأصبح النموذج :

$$Log\left[\frac{p}{1-p}\right] = \propto + Bx^{(2)} + \gamma I$$

حيث $\lambda, \gamma, \beta, \propto$ ثوابت مطلوب تقديرها

وان $\left[\frac{p}{1-p}\right]$ Log هو التحويل الخطي لنسبة الاستجابة في النموذج اللوجستي. وفي العام ٢٠٠٠ صدرت الطبعة الثانية من الكتاب الموسوم Applied Logistic Regression[11] لمولفيه David W.Hosmer& Stanley Lemeshow حيث نشر على الانترنت وتضمن العديد من التطبيقات للدالة اللوجستية في مجالات مختلفة ، فلم يقتصر على البيانات الحيوية والطبية وانما تعداها الى البيانات الاجتماعية في مجال التعليم والصحة وغيرها.وفي العام ٢٠٠٢ وما بعده نشرت على الانترنت العديد من البحوث الاحصائية حول تطبيق الدالة اللوجستية في المجال الاجتماعي منها البحث الموسوم An introduction to logistic Regression Analysis and Reporting [12] للباحثين Ping chao–Ying Joanne; kuk lida; Ingersoll,Gray M. والذي كان تطبيقا للتحليل اللوجستي في الحقل التعليمي . وفي العام ٢٠٠٦ نشر على الانترنت الباحثان التركيان Sansh Senol & Gozde Ulutagay بحثهما الموسوم [13]logistic Regression Analysis to determine the factors that affect "Green Card" usage for health services وهو تطبيق لتحليل الانحدار اللوجستي في حقل اخر من الحقول الاجتماعية وهو حقل الخدمات الصحية .

وهكذا استمرت البحوث التي تستخدم الدالة اللوجستية في اقتحام مختلف الحقول الاجتماعية وياتي هذا البحث حلقة في هذه السلسلة حيث يتصدى لابرز مشكلة اجتماعية الا وهي مشكلة الطلاق .

١-٢ المشكلة وأهمية البحث :

لعل الطلاق هو اكبر مشكلة يواجهها الإنسان في مسيرة حياته الاجتماعية ، ذلك لان ضرره لا يقتصر على الزوجين وحسب ، وإنما يمتد إلى الأسرة كلها ومن ثم المجتمع بأسره لذلك فأن التصدي له بمعرفة أسبابه وتحديد درجة معنويتها وشروط عملها ، يكتسب أهمية قصوى تستحق البحث والدراسة من حين لآخر .

وإذا كانت أسباب الطلاق قد أشبعها الباحثون الاجتماعيون دراسة بحيث أصبحت معروفة للجميع ، فأن تحديد درجة معنوية كل منها وشروط عملها تبقى مهمة الباحثين الإحصائيين .وفي هذا البحث ندرس معنوية احد هذه الأسباب وهو – الفارق العمري بين الزوجين – ونضع نتائج البحث في متناول أيدي وحدات البحث الاجتماعي في محاكم الأحوال الشخصية ، للاستفادة منها في مجال عملها .

١-٣ هدف البحث :

يهدف هذا البحث إلى تحديد المدى (غير المعنوي) لتأثير الفارق العمري بين الزوجين على مستقبل حياتهما الزوجية من خلال دراسة معنوية العلاقة بينه وبين احتمال الاستجابة المشاهدة – موضوع البحث – وهي نسبة حالات الانفصال بين الزوجين عند كل مستوى من مستويات الفارق العمري بينهما ، بغية الوصول إلى التوصيات العلمية اللازمة للحد من حالات الطلاق الناجمة عن الفارق العمري بين الزوجين .

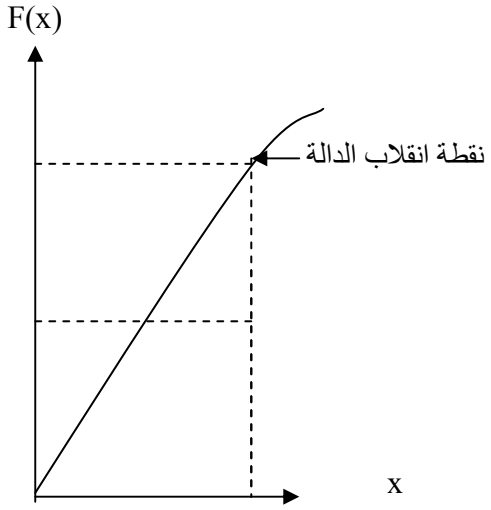
٢- الجانب النظري

٢-١ طبيعة النموذج اللوجستي Logistic Model

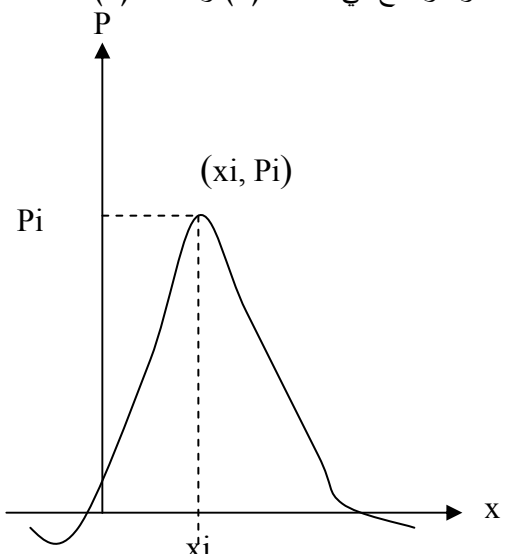
يستخدم النموذج اللوجستي بشكل واسع في الأبحاث الحياتية وذلك لتمثيل العلاقة بين متغير نسبة الاستجابات (P) حيث الاستجابة ثنائية Binary Response في مثل هذه الأبحاث – ومقياس الجرعة (X) الذي يمثل المتغيرات التوضيحية Explanatory variables إذ يعتقد بوجود تأثير كبير لهذه المتغيرات في احتمال الاستجابة P . وهو نموذج انحدار لا خطي يلاءم البيانات الثنائية Binary Data التي يكون فيها المتغير المعتمد هو متغير وهمي يمثل الاستجابة التي أما أن تكون موجودة (إيجابية) وتأخذ القيمة (١) أو غير موجودة (سلبية) وتأخذ القيمة (٥) ، ولكون الباحث الإحصائي يهتم بوجود (تحقق) الاستجابات عند كل مستوى من مستويات المتغير التوضيحي الذي يعتقد بعلاقته بها ، لذلك فان متغير نسبة الاستجابة المشاهدة هو نسبة الاستجابات المتحققة (الايجابية) عند كل مستوى من مستويات المتغير التوضيحي ، وهي نسبة متغيره من مستوى إلى آخر ويقترب توزيعها كثيرا من التوزيع الطبيعي عندما يكون حجم العينة كبيرا وذلك استناداً الى خاصية توزيع ذي الحدين الذي يؤول إلى التوزيع الطبيعي بزيادة حجم العينة ، وهذه الخاصية تعرف بأنها (الاتجاه نحو الاعتدال بزيادة حجم العينة) .

Tendency towards Normality as a sample size increases [١٤]

وكما هو موضح في الشكل (١) والشكل (٢)



الشكل (١) المنحنى اللوجستي



شكل (٢) توزيع P

فلو فرضنا ان

xi : المستوى i للتغير التوضيحي X

ni : حجم الاستجابات الكلية (الايجابية والسلبية) المشاهدة عند المستوى i

ri : عدد الاستجابات الموجودة (الايجابية) المشاهدة عند المستوى i

فأن :

$$pi = pr(y = 1 | x) = \frac{ri}{ni}$$

هو نسبة الاستجابة المتحققة (الايجابية) المشاهدة

وان :

$$1 - pi = pr(y = 0 | x) = \frac{ni - ri}{ni}$$

هو نسبة (احتمال) الاستجابة السلبية المشاهدة

وحيث ان pi , pi −1 هما دالة الى X فيمكن كتابتهما بدلالة X وفق النموذج الاتي المسمى (النموذج اللوجستي Logistic Modle) :

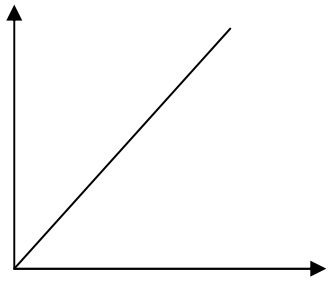
$$p = \frac{\exp(B_0 + B_1 X_1)}{1 + \exp(B_0 + B_1 X_1)} \quad \text{..... ....}(1-2)$$

$$\therefore 1 - p = \frac{1}{1 + \exp(B_0 + B_1 X_1)} \text{..... ....}(2-2)$$

٢-٢ التحويل الخطي للنموذج اللوجستي وأسبابه The linear Transformation

يصنف النموذج اللوجستي ضمن النماذج اللاخطية التي يمكن تحويلها إلى نماذج خطية وتسمى هذه النماذج بالنماذج الخطية ضمنيا (جوهريا) Intrinsically linear Model [15]. ويميل الاحصائيون عادة إلى التحويل الخطي لهذه النماذج لإزالة إنحناءات معلماتها وذلك لتأثير هذه الانحناءات السلبي في حالة وجودها على خصائص مقدرات المربعات الصغرى لها L.S estimators ومن ثم قيم الاستجابة التي يتم التنبؤ بها ، حيث تكون هذه المقدرات في الغالب متحيزة ولا تتوزع توزيعا طبيعيا وتبايناتها لا تكون أصغر ما يمكن مما يجعل نتائج الاختبارات مضللة[16] .

وقد قام الباحث Berkson منذ العام ١٩٤٤ بتحويل علاقة الانحدار اللاخطية بين المتغير التوضيحي X ومتغير نسبة الاستجابة P في النموذج اللوجستي إلى علاقة إنحدار خطية ، برسم لوغاريتم النسبة $\frac{p}{1-p}$ مقابل قياسات المتغير التوضيحي X ، كما هو واضح في الشكل (٣) .



شكل (٣)

العلاقة الخطية بين المتغير التوضيحي X

ونسبة الاستجابة المحولة Z

حيث :

$Ln\frac{p}{1-p}=z$ متغير الاستجابة الخطية

X : المتغير التوضيحي

ولبيان ذلك رياضيا نعود إلى الدالتين(٢-١) و(٢-٢)فبقسمة الأولى على الثانية :

$$\frac{p}{1-p}=\frac{\frac{\exp(B_0+B_1X_1)}{1+\exp(B_0+B_1X_1)}}{\frac{1}{1+\exp(B_0+B_1X_1)}}=\exp(B_0+B_1X_1)$$

وبأخذ لوغاريتم الطرفين يكون :

$$Ln\frac{p}{1-p}=Z=B_0+B_1X_1 \quad .....(3-2)$$

وهذه دالة انحدار خطي لـــ Z على X .

٢-٣ متوسط وتباين متغير الاستجابة Z :

سبق وأن أشار الباحث إلى أن توزيع P في النموذج اللوجستي هو قريب من التوزيع الطبيعي – لاحظ الشكل ٢-٢ وبعد تحويل P إلى Z بالتحويل اللوغاريتمي لإزالة تأثير انحناءات معلمات النموذج اللوجستي في تقدير قيم الاستجابة المتوقعة ، لابد من تحديد القيمة المتوقعة لـــ Z وتباينه في التوزيع الطبيعي .بالرجوع إلى الدالتين (٢-١) و (٢-٢) فأن :

$$p=\frac{\exp(B_0+B_1X_1)}{1+\exp(B_0+B_1X_1)} \quad \& \quad 1-p=\frac{1}{1+\exp(B_0+B_1X_1)}$$

نلاحظ ان p =g (x) ، أي ان p هي دالة الى x

ولكن $Ln\frac{p}{1-p}=z \Leftarrow$ z = f(p) ، أي ان z هو دالة لـــ P

$$\therefore z = f[g(x)]$$

(قاعدة السلسلة) chain Rule

$$\& \ \frac{\partial z}{\partial x} = \frac{\partial z}{\partial p} \cdot \frac{\partial p}{\partial x}$$

$$\because \frac{\partial z}{\partial p} = \frac{1-p}{p} \cdot \frac{1-p+p}{(1-p)^2} = \frac{1}{p(1-p)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4-2)$$

$$\& \ \frac{\partial p}{\partial x} = \frac{B_1 \exp(B_0 + B_1 X_1)}{[1+\exp(B_0 + B_1 X_1)]^2} = B_1 \frac{1}{1+\exp(B_0 + B_1 X_1)} * \frac{\exp(B_0 + B_1 X_1)}{1+\exp(B_0 + B_1 X_1)}$$

$$\therefore \frac{\partial p}{\partial x} = B_1 p(1-p) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5-2)$$

وبتعويض (٢-٤) و (٢-٥) في الطرف الايمن من قاعدة السلسلة يكون :

$$\frac{\partial z}{\partial x} = \frac{B_1 p(1-p)}{p(1-p)} = B_1 \quad \ldots\ldots\ldots\ldots\ldots\ldots (6-2) \qquad ، \qquad \frac{\partial^2 z}{\partial x^2} = 0$$

وهذا يعني أن انحدار Z على X_1 هو انحدار خطي بمعامل خطي ثابت مقداره (B_1) وبذلك تكون القيمة المتوقعة لـــ Z وتباين Z هي كما يلي :

$$E(z) = B_0 + B_1 X_1 \qquad \& \quad v(z) = \frac{1}{nipi(1-pi)} = \delta^2 i$$

٢-٤ **طريقة المربعات الصغرى الموزونة** (wls) يتضح من توزيع z ان متوسطه دالة الى (Bi, x) كما ان تباين Z هو دالة الـــى متوسطه لذلك فأنه لن يكون ثابتاً بتغير المتوسط أو تغير X وهذا يعني عدم تجانس تباين الخطأ العشوائي أي ان :

$$v(ei \mid xi) \neq \delta^2$$

ان عدم تجانس تباين الخطأ هو خرق لواحد من الفرضيات القياسية لنظرية المربعات الصغرى والتي غالباً مـــا يـــشار إليها بفرضية التجانس assumption of homoscedasticity وهي تعني ان تباين الخطأ لا يكون ثابتاً عند جميع المشاهدات وعند ذاك يسمى ذلك الخطأ بأنه غير متجانس hetreoscedastrcity وان من شأن ذلك الحـــصول علـــى مقدرات غير دقيقة ولا تكون تبايناتها اقل ما يمكن مما يجعل نتائج جميع الاختبارات مضللة [17] .

ولمعالجة هذه المشكلة هناك طريقتان : أحداهما باستخدام تحويل مناسب لدالة النموذج الخطي ولاسيما التـــي تكون تبايناتها دوالاً الى متوساطتها كما أوردها بالتفصيل Stuart & Kendal [18] ، والأخـــرى باســـتخدام طريقـــة المربعات الصغرى الموزونة وهي التي اتبعها الباحث ، وذلك كما يلي :

يتم اختيار الاوزان بحيث تتناسب عكسيا مع التباين أي اختيار :

$$Wi = \frac{1}{\delta^2 i} = nipi(1-pi) \quad \ldots\ldots\ldots\ldots\ldots (7-2)$$

حيث Wi هي الأوزان المختارة

فيكون تقدير المعلمات B_0 , B_1 حسب طريقة المربعات الصغرى تعني إيجاد قيمتيهما اللتين تجعلان الفـــارق بين الاستجابة المشاهدة والاستجابة المقدرة (المتوقعة) اقل ما يمكن ، أي التي تصغر مجموع مربعات البواقي (الخطأ) sse ، فيكون [19] .

$$SSei = \sum Wi(Zi - \hat{Zi})^2 = \sum Wi(Zi - B_0 - B_1 x_1 i)^2$$

وباستخدام التفاضل الجزئي لهذه المعادلة بالنسبة الى B_0 , B_1 ثم جعل الناتج يساوي صفر يكون

$$B_0 \sum Wi + B_1 \sum WiX_1 i = \sum WiZi$$

$$\& \quad B_0 \sum WiX_1 i + B_1 \sum WiX_1^2 i = \sum WiXi_1 Zi$$

وباستخدام المصفوفات يمكن كتابة هاتين المعادلتين الطبيعيتين كما يلي :

$$\begin{bmatrix} \sum Wi & \sum WiXi_1 \\ \sum WiXi_1 & \sum WiX_1^2 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \begin{bmatrix} \sum WiZi \\ \sum WiXi_1 Zi \end{bmatrix}$$

وهذا يعني ان :

$$X'WXB = X'W\underline{Z}$$

حيث ان

$$X = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ 1 & X_{31} \\ \vdots & \vdots \\ 1 & X_{n1} \end{bmatrix} \quad , \quad \underline{Z} = \begin{bmatrix} \ln \frac{P_1}{1-P_1} \\ \ln \frac{P_2}{1-P_2} \\ \ln \frac{P_3}{1-P_3} \\ \vdots \\ \ln \frac{P_n}{1-P_n} \end{bmatrix}$$

$$B = \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} \quad , \quad W = \begin{bmatrix} W_1 & 0 \dots\dots\dots & \dots 0 \\ o & W_2 \dots\dots\dots & .0 \\ 0 & 0 \dots\dots\dots & \dots . w_n \end{bmatrix}$$

$$X'WX = \begin{bmatrix} \sum Wi & \sum WiXi_1 \\ \sum WiXi_1 & \sum WiXi_1^2 \end{bmatrix} \quad , \quad X'W\underline{Z} = \begin{bmatrix} \sum WiZi \\ \sum WiXiZi \end{bmatrix}$$

وبذلك تكون :

$$\underline{\beta} = (X'WX)^{-1} X'W\underline{Z}$$

$$= \begin{bmatrix} \sum Wi & \sum WiXi_1 \\ \sum WiXi_1 & \sum WiXi_1^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum WiZi \\ \sum WiXiZi \end{bmatrix} \quad \dots\dots\dots\dots\dots (8-2)$$

كما يمكن تقدير قيم معلمات نموذج الانحدار اللوجستي باستخدام التباينات ، كما يلي : من المعادلة الطبيعية الاولـــى

$$\hat{\beta}_0 \sum Wi + \hat{\beta}_1 \sum WiX_1 i = \sum WiZi$$

وبقسمة الطرفين على $\sum wi$ يكون

$$\hat{\beta}_0 + \hat{\beta}_1 \frac{\sum WiXi}{\sum Wi} = \frac{\sum WiZi}{\sum Wi}$$

$$\Rightarrow \hat{\beta}_0 = \frac{\sum WiZi}{\sum Wi} - \hat{\beta}_1 \frac{\sum WiX_1 i}{\sum Wi}$$

$$\because \frac{\sum WiZi}{\sum Wi} = \bar{Z} \quad , \quad \frac{\sum WiXi}{\sum Wi} = \bar{X}_1$$

$$\therefore \hat{\beta}_0 = \bar{Z} - \hat{\beta}_1 \bar{X}_1 \quad \dots\dots\dots\dots\dots\dots\dots\dots (9-2)$$

وبتعويض (٢-٩) في المعادلة الطبيعية الثانية :

$$\hat{\beta}_0 \sum Wi\, X_1 i + \hat{\beta}_1 \sum WiX_1^2 i = \sum WiX_1 iZi$$

$$\Rightarrow \sum WiX_1 i\left[\bar{Z} - \hat{\beta}_1 \bar{X} i\right] + \hat{\beta}_1 \sum WiX^2{}_1 i = \sum WiX_1 iZi$$

$$\Rightarrow \sum Wi \bar{X}_1 \bar{Z} - \hat{\beta}_1 \sum Wi \bar{X}_1^2 + \hat{\beta}_1 \sum WiX^2{}_1 i = \sum WiX_1 iZi$$

$$\Rightarrow \hat{\beta}_1 \sum Wi \bar{X}_1^2 i - \hat{\beta}_1 \sum Wi \bar{X}_1^2 = \sum WiX_1 iZi - \sum Wi \bar{X}_1 \bar{Z}$$

$$\Rightarrow \hat{\beta}_1\left[\sum Wi \bar{X}_1^2 i - \sum Wi \bar{X}_1^2\right] = \sum WiX_1 iZi - \sum Wi \bar{X}_1 \bar{Z}$$

SX_1X_1 $\quad$ $SX_1 Z$

$$\hat{\beta}_1 = \frac{SX_1Z}{SX_1X_1} \text{..}(10-2)$$

كما ويمكن إيجاد قانون حساب قيمة مجموع مربع الانحرافات في المشاهدات العائدة لانحدار Z على X_1 والذي نرمز له $S\hat{z}\hat{z}$ كما يلي :

$$S\hat{z}\hat{z} = \sum wi\left(\hat{z} - \bar{z}\right)^2 = \sum wi\hat{z}^2 - \sum wi\bar{z}^2$$

$$\because \hat{z}i = \hat{\beta}_0 + \hat{\beta}_1 x_1$$

$$\therefore S\hat{z}\hat{z} = \sum Wi\left[\hat{\beta}_0 + \hat{\beta}_1 x_1\right]^2 - \sum wi\bar{z}^2$$

$$= \sum Wi\left[\hat{\beta}_0^2 + 2\hat{\beta}_0 \hat{\beta}_1 X_1 + \hat{\beta}_1^2 X_1^2\right] - \sum Wi\bar{Z}^2$$

$$= \hat{\beta}_0^2 \sum Wi + 2\hat{\beta}_0 \hat{\beta}_1 \sum WiX_1 + \hat{\beta}_1^2 \sum WiX_1^2 - \sum Wi\bar{Z}^2$$

$$= \left[\hat{\beta}_0^2 \sum Wi + \hat{\beta}_0 \hat{\beta}_1 \sum WiX_1\right] + \left[\hat{\beta}_0 \hat{\beta}_1 \sum WiX_1 + \hat{\beta}_1^2 \sum WiX_1^2\right] - \sum Wi\bar{Z}^2$$

$$= \hat{\beta}_0\left[\hat{\beta}_0 \sum Wi + \hat{\beta}_1 \sum WiX_1\right] + \hat{\beta}_1\left[\hat{\beta}_1 \sum WiX_1 + \hat{\beta}_1 \sum WiX_1^2\right] - \sum Wi\bar{Z}^2$$

$$= \hat{\beta}_0 \sum WiZi + \hat{\beta}_1 \sum WiX_1 iZi - \sum Wi\bar{Z}^2$$

Where :

$$\hat{\beta}_0 \sum Wi + \hat{\beta}_1 \sum WiX_1 i = \sum WiZi$$

$$\hat{\beta}_0 \sum WiX_1 + \hat{\beta}_1 \sum WiX_1^2 = \sum WiX_1 iZi$$

معادلتان طبيعيتان

$$\therefore S\hat{z}\hat{z} = \hat{\beta}_0 \sum Wi\bar{Z} + \hat{\beta}_1 \sum WiX_1 iZi = \sum Wi\bar{Z}^2$$

$$\therefore \hat{\beta}_0 = \bar{Z} - \hat{\beta}_1 \bar{X}_1 \text{ ....}(9-2)$$

$$\Rightarrow \therefore S\hat{z}\hat{z} = \sum Wi\bar{Z}\left[\bar{Z} - \hat{\beta}_1 \bar{X}_1\right] + \hat{\beta}_1 \sum WiX_1 iZi - \sum Wi\bar{Z}^2$$

$$= \sum Wi\bar{Z}^2 - \hat{\beta}_1 \sum Wi \bar{X}_1 \bar{Z} + \hat{\beta}_1 \sum WiX_1 iZi - \sum Wi\bar{Z}^2$$

$$= \hat{\beta}_1\left[\sum WiX_1 iZi - \sum Wi \bar{X}_1 \bar{Z}\right]$$

Sx_1z

$$\therefore S_{\hat{z}\hat{z}} = SSR = \hat{\beta}_1 SX_1Z \quad \text{.....(11-2)}$$

$$= \hat{\beta}_1 \left[\frac{SX_1Z}{Sx_1x_1} . Sx_1x_1 \right] = \hat{\beta}_1^{\,2} Sx_1x_1 \quad \text{.....(12-2)}$$

$$\therefore SSE = S_{(z-\hat{z})(z-\hat{z})} = S_{zz} - S_{\hat{z}\hat{z}} = S_{zz} - \hat{\beta}_1^{\,2} S_{X_1X_1} \quad \text{.....(13-2)}$$

حيث :

S_{zz} هو مجموع مربع الانحرافات الكلية في الاستجابات المشاهدة .

SSE هو مجموع مربع الانحرافات المتبقية بعد طرح مجموع مربع الانحرافات العائدة للانحدار من مجموع الانحرافات الكلية .

٢-٥ اختبار معنوية B_0 , B_1 : ان توزيع المعاينة لـ B_0 , B_1 في الانحدار الخطي هو التوزيع الطبيعي بوسط حسابي وقدره على التوالي :

$$E\left(\hat{\beta}_0\right) = \beta_0 \qquad , \quad E\left(\hat{\beta}_1\right) = \beta_1$$

وبتباين وقدره على التوالي :

$$S^2_{\hat{\beta}_0} = MSE \ (C_{00}) \qquad , \quad S^2_{\hat{\beta}_1} = MSE \ (C_{11})$$

حيث C_{00} ، C_{11} هي عناصر قطر معكوس المصفوفة :

$$\begin{bmatrix} \sum Wi & \sum WiXi \\ \sum WiXi & \sum WiXi^{\,2} \end{bmatrix}$$

ولكون MSE = $\hat{\delta}^2$ = ١ عند استخدام مقلوب تباين استجابة ثنائية باعتباره الوزن للازم لتحقيق تجانس تباين حد الخطأ العشوائي في نموذج الانحدار ولذلك يكون :

$$S^2_{\hat{\beta}_0} = C_{00} \quad \& \quad S^2_{\hat{\beta}_1} = C_{11} \quad \text{.....(14-2)}$$

كما يمكن استخدام التباينات بدل المصفوفات لحساب $S^2_{\hat{\beta}_0}$ $S^2_{\hat{\beta}_1}$ كما يلي:

$$S^2_{\hat{\beta}_1} = \frac{MSE}{Sx_1x_1} = \frac{1}{Sx_1x_1}$$

$$S^2_{\hat{\beta}_0} = \frac{1}{\sum Wi} + \bar{X}^{\,2} S^2_{\hat{\beta}_1}$$

ولاختبار معنوية B_0 , B_1 فانه يجري عادة اختبار الفرضية القائلة بأن Bi تساوي صفر ، وتكتب مع الفرضية البديلة كالاتي :

H_0 : Bi = 0

V.s H_1 : Bi $\neq$ 0 where i = 0 , 1

ونستخدم احصاءة الاختبار

$$t = \frac{\hat{\beta} i - \beta i}{S_{\hat{\beta} i}} \quad \text{.....(15-2)}$$

اذ ان احصاءة الاختبار تتوزع توزيع t-student بدرجة حرية قدرها (n-2) عندما تكون فرضية العـــدم [H_0:Bi=0] صحيحة ، أي عندما تكون احصاءة الاختبار $t = \frac{/\hat{\beta} i/}{S_{\hat{\beta} i}}$ وتقارن مع قيمة t- الجدولية بنفس درجة الحرية وبمستوى معنوية معينة = $\propto \frac{1}{2}$ فاذا كانت :

$/t/ = \frac{/\hat{\beta} i/}{S_{\hat{\beta} i}} \geq t_{\frac{1}{2}\propto, n-1} \Rightarrow$ نرفض فرضية العدم

$/t/ = \frac{/\hat{\beta} i/}{S_{\hat{\beta} i}} < t_{\frac{1}{2}\propto, n-1} \Rightarrow$ لا نرفض فرضية العدم

٢-٦ تحليل الانحرافات ومؤشرات ملاءمة النموذج :

ان تحليل الانحرافات في الاستجابات المشاهدة وفقاً لخاصية الانحدار يعني تقسيم هذه الانحرافات الى قـــسمين احدهما يعود لخاصية الانحدار (انحدار الاستجابات على المتغير التوضيحي – الفارق في عمر الزوجين) والآخر يعود لعوامل أخرى لم يدخلها الباحث في نموذجه وهي ما تبقى من الانحرافات الكلية ولذلك تسمى انحرافات(البواقي) ويمكن الإشارة إلى هذا التحليل بالرموز الآتية:

Szz : مجموع مربع الانحرافات الكلية في الاستجابات المشاهدة ويحسب وفق العلاقة :

$$Szz = \sum Wi\left(Zi - \bar{Z}\right)^2 = \sum WiZ^2 i - \frac{\left(\sum WiZi\right)^2}{\sum Wi} \quad \text{.....(16-2)}$$

$S\hat{z}\hat{z}$: مجموع مربع الانحرافات في الاستجابات المشاهدة العائدة للانحدار ويحسب وفق العلاقة (٢-١١) أو (٢-١٢) التي سبق اشتقاقها .

$SSE = S_{(z-\hat{z})(z-\hat{z})}$: مجموع مربع الانحرافات المتبقية (البواقي) بعد طرح مجموع مربع الانحرافـــات العائـــدة للانحدار من مجموع مربع الانحرافات الكلية وبحسب وفق العلاقة (٢-١٣) التي سبق اشتقاقها ومن خلال هذا التحليل للانحرافات الكلية في الاستجابات المشاهدة يمكن تحديد ملاءمة نموذج الانحدار المستخدم للبيانـــات موضـــوع البحـــث والدراسة وذلك بعدة مؤشرات منها ما يلي :

٢-٦-١ النسبة R^2 ومجموع مربعات الانحدار $S\hat{z}\hat{z}$

ان R^2 هو نسبة الاختلافات في متغير الاستجابة المشاهدة Z التي تعود لعلاقة الانحدار الخطية، أو هـــو النـــسبة التـــي يفسرها المتغير التوضيحي X_1 .

ويسمى ايضاً (معامل التحديد coefficient of determination) ويحسب وفق العلاقة :

$$R^2 = \frac{S\hat{z}\hat{z}}{Szz} \quad \text{.....(17-2)}$$

ويلاحظ ان هذا المؤشر يعتمد على قيمة مجموع مربعات الانحرافات العائدة للانحدار إذ كلما اقتربت من مجموع مربعات الانحرافات الكلية كلما اقترب هذا المؤشر من الرقم ١ أو ١٠٠% إذا استخدم كنسبة مئوية ، مما يعني قوة علاقة الانحدار الخطي وملاءمة النموذج للبيانات .ومن زاوية أخرى للنظر فأن كبر قيمة مجموع مربعات الانحرافات العائدة للانحدار تعني صغر قيمة مجموع مربعات الانحرافات المتبقية (SSE) واذا حسب متوسط هذه الانحرافات (MSE) بدرجات حرية كافية فان قيمته ستكون صغيرة جداً وهي مؤشر على معنوية علاقة الانحدار وملاءمة النموذج للبيانات .

٢-٦-٢ تحليل البواقي Residuals – Analysis

لكون البواقي حسب نظرية المربعات الصغرى تتوزع توزيعاً طبيعياً بوسط حسابي يساوي صفر وتباين يساوي δ^2 فان قيم البواقي المقدرة بواسطة نموذج الانحدار الملائم للبيانات سوف تتمحور حول لوسط الحسابي لها أي حول (الصفر) أي انها لا تأخذ شكلاً معيناً وتنحصر عادة بين $(2\pm)$ اذا ما رسمت كنقاط ازاء القيم المتوقعة لمتغير الاستجابة أو ازاء المتغير المستقل (لتوضيحي) في حالة نموذج الانحدار الخطي البسيط ، حيث تكون على شكل خطين متوازيين فاذا تبين وجود علاقة محددة بين قيم البواقي وقيم متغير الاستجابة المتوقعة أو قيم المتغير التوضيحي فمعنى ذلك ان هناك في الاقل خرقاً لواحد من فرضيات نظرية المربعات الصغرى ، وان النموذج المستخدم غير ملائم للبيانات ، هذا فضلاً عن ان رسم نقاط انتشار البواقي يكشف عن النقاط المتطرفة التي قد يكون لها أهمية في التحليل الإحصائي للبيانات اكبر من أهمية بقية البيانات .

٣ – الجانب التطبيقي

٣-١ أسلوب جمع البيانات وتصنيفها :

تم جمع البيانات عن حالات الزواج المسجلة في محاكم الاحوال الشخصية لرئاسة استئناف بغداد/الكرخ والرصافة لعام ٢٠٠٤ والبالغ عددها ٨٠٠٨ حالة زواج بأسلوب العينة العشوائية النظامية Systematic Random Sample وذلك لسهولة تطبيقه وشموليته حيث اختيرت عينة عشوائية بحجم ٢٠٠ حالة زواج وتم حساب المدى المنتظم بين كل حالة زواج مختارة واخرى ضمن العينة باستخدام العلاقة : $K = \frac{N}{n} = \frac{8008}{200} \approx 40$

واختيرت البداية العشوائية (حالة الزواج الاول في العينة) من بين تسلسلات حالات الزواج المرقمة من ١ الى ٤٠ فكانت البداية هي التسلسل ٣٧ ، وبذلك يكون تصميم العينة العشوائية النظامية بإضافة المدى ٤٠ بصورة متتالية بدءاً من التسلسل رقم ٣٧ وكما يلي :

((37 77 117 157 197 237 277 317 357 397...٧٩٩٧)) ثم جرى تتبع حالات الزواج هذه (الواردة تسلسلاتها في العينة) على مدى خمس سنوات [من ٢٠٠٤ – ٢٠٠٨] في سجلات هذه المحاكم ، فتبين ان (٤٥) حالة زواج منها انتهت بالطلاق بينما استمرت ١٥٥ حالة زواج . وبتصنيف حالات الزواج في العينة حسب الفارق العمري بين الزوجين ، حصل الباحث على البيانات الاتية كما هي موضحة في الجدول (١) .

جدول (١) تصنيف حالات الزواج حسب الفارق العمري بين الزوجين

Xi الفارق في عمر الزوجين (بالسنين)	Yi حالات الزواج		المجموع (ni)
	فاشلة (ri)	ناجحة (ni–ri)	
٣	٨	٣٧	٤٥
٥	٥	٣٣	٣٨
٩	٥	٢١	٢٦
١٥	١٠	٦	١٦
٧	٤	١٣	١٧
١٠	٣	١٥	١٨
١٢	٤	٨	١٢
١	٤	٨	١٢
٢	٢	١٤	١٦
المجموع	٤٥	١٥٥	٢٠٠

وبحساب نسبة الحالات التي تأثرت الفارق العمري عند كل مستوى من مستوياته (نسبة الاستجابة المشاهدة) من العلاقة:

$$pi = pr(y = 1|x) = \frac{ri}{ni}$$

وحساب نسبة الحالات التي لم تتأثر بالفارق العمري وعند كل مستوى من مستوياته ((نسبة عدم الاستجابة المـــشاهدة)) من العلاقة :

$$1 - pi = pr(y = 01\ x) = \frac{ni - ri}{ni}$$

حصل الباحث على البيانات الآتية كما هي موضحة في الجدول (٢)

جدول (٢) تصنيف نسب حالات الزواج (الاستجابة بالمشاهدة) حسب مستويات الفارق العمري بين الزوجين .

Xi مستويات الفارق في عمر الزوجين (بالسنين)	Pi احتمال الاستجابة المشاهدة	1−pi احتمال عدم الاستجابة المشاهدة
١	$\frac{4}{12}=0.3333$	$\frac{8}{12}=0.6667$
٢	$\frac{2}{10}=0.125$	$\frac{14}{16}=0.875$
٣	$\frac{8}{45}=0.1778$	$\frac{37}{45}=0.8222$
٥	$\frac{5}{38}=0.1316$	$\frac{33}{38}=0.8648$
٧	$\frac{4}{17}=0.2353$	$\frac{13}{17}=0.7647$
٩	$\frac{5}{26}=0.1923$	$\frac{21}{26}=0.8077$
١٠	$\frac{3}{18}=0.1667$	$\frac{15}{18}=0.8333$
١٢	$\frac{4}{12}=0.3333$	$\frac{8}{12}=0.6667$
١٥	$\frac{10}{16}=0.625$	$\frac{6}{16}=0.375$

ويشير الجدول (٢) الى ان الاستجابات تبدأ بالكبر عندما يبدأ الفارق العمري بالزيادة على ١٢ سنة لصالح الزوج أو بالنقصان عن سنة واحدة أي كلما اقترب عمر الزوجة من عمر الزوج أو تجاوزه لصالح الزوجة .

٣-٢ حساب التحويل الخطي لنسب الاستجابة المشاهدة وأوزانها : باستخدام العلاقة (٢-٣) يكون :

$$Z_1 = Ln\frac{P_1}{1-P_1} = Ln\frac{1}{2} = -0.6931$$

$$Z_2 = Ln\frac{P_2}{1-P_2} = Ln\frac{1}{7} = -1.9459$$

$$Z_9 = Ln\frac{P_9}{1-P_{19}} = Ln\frac{5}{3} = +0.5108$$

وباستخدام العلاقة (٢-٧) يكون :

$$W_1 = n_1 p_1(1-p_1) = 12\left(\frac{4}{12}\right)\left(\frac{8}{12}\right) \sim 2.6667$$

$$W_2 = n_2 p_2(1-p_2) = 16\left(\frac{2}{16}\right)\left(\frac{14}{16}\right) = 1.75$$

$$W_9 = n_9 p_9(1-p_9) = 16\left(\frac{10}{16}\right)\left(\frac{6}{16}\right) = 3.75$$

لاحظ جميع النتائج في الجدول (٣) .

(٣-٣) تقدير معادلة الانحدار الخطي للاستجابات المتوقعة واختبار معلمات نموذج الانحدار: لتقدير معادلة الانحدار الخطي للاستجابات المتوقعة ($\hat{Z}$) لا بد اولاً من تقدير معلمات نموذج الانحدار الخطي البسيط (B_0 , B_1) وذلك باستخدام العلاقة(٨-٢) فتكون :

$$\hat{\beta} = \begin{bmatrix} \sum Wi & \sum WiXi \\ \sum WiXi & \sum WiXi^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum WiZi \\ \sum WiXiZi \end{bmatrix}$$

$$\begin{bmatrix} \beta_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 31.3506 & 218.6191 \\ 218.6191 & 2132.1739 \end{bmatrix}^{-1} \begin{bmatrix} -36.8797 \\ -190.9334 \end{bmatrix}$$

$$= \begin{bmatrix} 0.111921495 & -0.011475694 \\ -0.011475644 & 0.001645647 \end{bmatrix} \begin{bmatrix} -36.8797 \\ -190.9334 \end{bmatrix}$$

$$= \begin{bmatrix} -1.9365 \\ 0.1090 \end{bmatrix}$$

$$\hat{\beta}_0 = -1.9365 \quad , \quad \hat{\beta}_1 = 0.10990$$

وبذلك تكون معادلة الانحدار الخطية للاستجابات المتوقعة :

$$\hat{Z}i = -1.9365 + 0.1090Xi \qquad i = 1,2,3,.....,9$$

$$\hat{Z}_1 = -1.9365 + 0.1090(1) = 1.8275$$

$$\hat{Z}_2 = -1.9365 + 0.1090\ (2) = -1.7185$$

$$\hat{Z}_9 = -1.9365 + 0.1090(15) = -0.3015$$

ويلاحظ ان قيمة $\hat{\beta}_0$ (سالبة) في معادلة الانحدار الخطي للاستجابات المتوقعة وهي تعني ان الزوجين في بداية حياتهما الزوجية بعيدان عن التفكير في الانفصال حيث ان العوامل المؤدية الى الانفصال ومنها الفارق في عمر الزوجين لا يظهر اثرها في بداية الحياة الزوجية وانما عندما بتقدم العمر بها .ومن الجدير بالذكر انه يمكن حساب قيم احتمالات الاستجابات المتوقعة بدلالة الدالة الاصلية (اللوجستية) بالتعويض بالعلاقة (٢-١) وقيم الاستجابات الخطية المتوقعة ($\hat{Z}$) – لاحظ الجدول رقم (٤) – كما يلي :

$$\hat{p}_1 = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{11}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{11}}} = \frac{e^{\hat{Z}_1}}{1 + e^{\hat{Z}_1}} = \frac{e^{-1.8275}}{1 + e^{-1.8275}} = 0.1385$$

$$\hat{p}_9 = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X_{91}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{91}}} = \frac{e^{\hat{Z}_9}}{1 + e^{\hat{Z}_9}} = \frac{e^{-0.3015}}{1 + e^{-0.3015}} = 0.4252$$

الجدول (٣) الاستجابات الخطية المشاهدة وأوزانها عند كل مستوى من مستويات الفارق في عمر الزوجين

Xi	X_i^2	Wi	Zi	Z_i^2	WiXiZi	WiXi	WiZi	Wi X_i^2	Wi Z_i^2
1	1	2.6667	−0.6931	0.4804	−1.8483	2.6667	−1.8483	2.6667	1.28108268
2	4	1.75	−1.9459	3.7865	−6.8107	3.5	−3.4053	7	6.626375
3	9	6.5778	−1.5316	2.3458	−30.2237	19.7334	−10.0646	59.2002	15.43020324
5	25	4.3421	−1.8872	3.5615	−40.9721	21.7105	−8.1944	108.5525	15.46438915
7	49	3.0588	−1.1786	1.3891	−25.2357	21.4116	−3.6051	149.8812	4.24897908
9	81	4.0385	−1.4351	2.0595	−52.1609	36.3465	−5.7957	327.1185	8.31729076
10	100	2.5	−1.6094	2.5902	−40.235	25	−4.0235	250	6.4755
12	44	2.6667	−0.6931	0.4804	22.1795	32.0004	−1.8483	384.0048	1.28108268
15	225	3.75	0.5108	0.2609	28.7325	56.25	+1.9155	843.75	0.978375
$\sum$		31.3506			−190.9334	218.6191	−36.8797	2132.1739	60.1033

جدول (٤) قيم الاستجابة المشاهدة والمتوقعة وقيم البواقي الموزونة

$\hat{P}i$	Zi	$\hat{Z}i$	$\overset{*}{e}$ $(Zi-\hat{Z})$	Wi
o.1385	−0.6931	−1.8275	+1.1344	2.6667
0.1521	−1.9459	−1.7185	− 0.2274	1.75
0.1667	−1.5316	−1.6095	+0.0779	6.5778
0.1992	−1.8872	−1.3915	− 0.4959	4.3421
0.2362	−1.1786	−1.1735	−0.0051	3.0588
0.2778	−1.4351	−0.9555	− 0.4796	4.0385
0.3002	−1.6094	−0.8465	− 0.7629	2.5
0.3479	−0.6931	−0.6285	− 0.0646	2.6667
0.4252	0.5108	− 0.3015	+ 0.8123	3.75

جدول (٥) القيم الموزونة للاستجابات ومستويات الفارق العمري والبواقي المتوقعة

$\sqrt{Wi}$	Xi_1	$\hat{Z}i$	$\overset{*}{e}_i$	$\sqrt{Wi}Xi_1$	$\sqrt{Wi}\ \hat{Z}i$	$\sqrt{Wi}\ \overset{*}{e}i$
1.6330	1	− 1.8275	+1.1344	1.6330	−2.9843	+1.8525
1.3229	2	−1.7185	−0.2274	2.6458	−2.2734	−0.3008
2.5647	3	−1.6095	+0.0779	7.6941	−4.1279	+0.1998
2.0838	5	−1.3915	−0.4959	10.419	−2.8996	−1.0339
1.7489	7	−1.1735	−0.0051	12.2423	−2.0523	−0.0089
2.0069	9	−0.9555	−0.4769	18.0864	−1.9202	−0.9638
1.5811	10	−0.8465	−0.7629	15.811	−1.3384	−1.2062
1.6330	12	−0.6285	− 0.0646	19.596	−1.0263	−0.1055
1.9365	15	−0.3015	+0.8123	29.0475	−0.5839	+1.5730

ولاختبار معلمات نموذج الانحدار الخطي لابد من حساب الخطأ المعياري لتقديرات هذه المعلمات ، باستخدام العلاقــة (٢-١٤) :

$$S^2_{\hat{\beta}_0} = Coo \Rightarrow S_{\hat{\beta}_0} = \sqrt{C_{00}} = \sqrt{0.111921495} = 0.3345467$$

$$S^2_{\hat{\beta}_1} = C_{11} \Rightarrow S_{\hat{\beta}_1} = \sqrt{C_{11}} = \sqrt{0.001645647} = 0.040566574$$

ويلاحظ ان تباين تقديرات هذه المعلمات ذات قيم صغيرة (جداً) وبخاصة تباين معامل انحدار Z على X_1 مما يشير الى تجانس تباين الخطأ وهو ما سيتم التأكد منه عند تحليل البواقي .ويوضع فرضية العدم (Ho) والفرضية البديلة (H_1) والتعويض باحصاءة الاختبار في العلاقة (٢-١٥) وذلك لكل من B_0 , B_1 يكون :

$$Ho : B_0 = 0 \quad Vs \quad H_1 ; \; B_0 \neq 0$$

$$\therefore / t / = \frac{|-1.9365|}{0.3345467} = 5.7884$$

وبمقارنتها مع قيمة (٠,٠٢٥ و ٧) t الجدولية والتي تساوي ٢,٣٦٥ ويكون / t/ > t (7, 0.025)

وهذا يعني ان قيمة /t/ تقع في منطقة رفض فرضية العدم لذلك نرفض فرضية العدم عند مستوى معنوية ٠,٠٥ للطرفين ، وبذلك تكون $\beta_0 \neq 0$. ومن الجدير بالذكر ان B_0 تبقى لا تساوي صفر حتى عندما نجري الاختبار عند مستوى معنوية ٠,٠١ وهذا يعني ان B_0 معنوية جداً وهو امر طبيعي لان B_0 تمثل الاستجابة المتوقعة للانفصال بين الزوجين عند بدء حياتهما الزوجية . وكذلك عند اختبار معامل الانحدار B_1 يكون :

$$H_0 : \beta_1 = 0 \quad VS \qquad H_1 : \beta_1 \neq 0$$

$$\therefore |t| = \frac{0.1090}{0.040566574} = 2.6869 > t(7,\ 0.025)$$

لذلك نرفض فرضية العدم عند مستوى معنوية ٠,٠٥ حيث ان معامل الانحدار في المجتمع لا يساوي صفر ، أي ان الفارق في عمر الزوجين يؤثر تأثيراً معنوياً (سلبياً) على مستقبل الحياة الزوجية ، وذلك في حدود الفارق العمري الذي ظهر في العينة والمحدد بالفترة المغلقة [١ ، ١٥] .

٣-٤ تحليل الانحرافات ومؤشرات ملائمة النموذج للبيانات .

باستخدام العلاقة (٢-١٦) نحصل على قيمة مجموع مربعات الانحرافات الكلية S_{zz}

$$S_{zz} = 60.10333998 - \frac{1360.112272}{31.3506} = 16.7194$$

باستخدام العلاقة(٢-١١)نحصل على قيمة مجموع مربعات الانحرافات العائدة للانحدار $S_{\hat{Z}\hat{Z}}$:

$$S_{\hat{Z}\hat{Z}} = 0.1090\ (66.24211892) = 7.22$$

وباستخدام العلاقة (٢-١٧) نحصل على المؤشر R^2 :

$$R^2 = \frac{7.22}{16.7194} = 0.43$$

وهذا يعني ان (الفارق في عمر الزوجين) لوحده يفسر (٤٣%) من الاختلافات في الاستجابات المشاهدة لانفصال الزوجين وهي بلا شك نسبة كبيرة تشير إلى أهمية هذا العامل في تحديد مستقبل الحياة الزوجية إلاّ انه من جانب آخر فأن هذا العامل يعجز عن تفسير حالات الانفصال في (٥٧%) من حالات الزواج، مما يعني ان نموذج الانحدار بصيغته الحالية المعتمدة على تأثير عامل واحد لا يمكنه ان يفسر كل أو معظم حالات الانفصال مما يجعله قاصراً عن تمثيل البيانات تمثيلاً كافياً . وللتأكد من عدم وجود خرق لفروض نظرية المربعات الصغرى للمتغير الجديد (Z) نرسم القيم المتوقعة له ($\hat{Z}$)مقابل قيم البواقي الموزونة ، أي نرسم $\sqrt{Wi}\,\hat{Z}i$ مقابل $\sqrt{Wi}\,\hat{e}i$ أو نرسم

$\sqrt{Wi}\,Xi$ مقابل $\sqrt{Wi}\,\hat{e}i$ وذلك باستخدام نتائج الجدول (٥) وكما في الشكل (٤). ويلاحظ ان النقاط تنحصر تقريباً بين (± 2) وتنتشر وان لم تكن جميعها بصورة طبيعية حول الصفر ، وترسم مستقيمين متوازيين مما يؤكد عدم وجود أي خرق لفروض المربعات الصغرى ، وان النموذج المستخدم ملائم وان كان بصورة غير كافية لتمثيل البيانات ، ولكن يلاحظ وجود نقطتين متطرفتين هما النقطتان اللتان احداثيهما السيني يمثل الفارق العمري (١و ١٥) ، ولفحص اهميتها في علاقة الانحدار الخطية ، نحسب قيمة معامل الانحدار $\hat{\beta}_1$ ونختبر معنويته بعد استبعاد هاتين النقطتين أي بعد استبعاد الاستجابات عند الفارق العمري (١ و ١٥) فيكون :

$$(X'WX)^{-1} = \begin{bmatrix} 0.196176464 & -0.024344849 \\ -0.024366849 & 0.0033804329 \end{bmatrix}$$

$$\sum WiZi = -36.9469 \quad , \quad \sum WiXiZi = 217.8175$$

$$\therefore \hat{\beta} = \begin{bmatrix} 0.196176469 & -0.02436849 \\ -0.024366849 & 0.003804329 \end{bmatrix} \begin{bmatrix} -36.9469 \\ -217.8175 \end{bmatrix}$$

$$= \begin{bmatrix} -1.940586066 \\ 0.071630101 \end{bmatrix}$$

$$\therefore \hat{Z}i = -1.940586066 + 0.071630101 X_1$$

$$S_{\hat{\beta}_0} = \sqrt{0.196176464} = 0.442918123$$

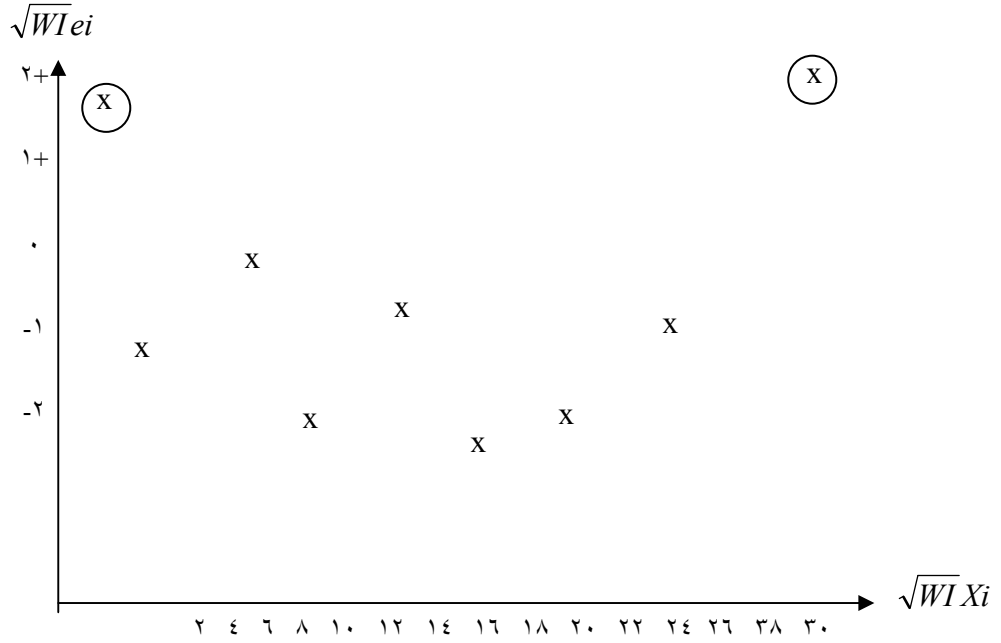
$$S_{\hat{\beta}_1} = \sqrt{0.003804329} = 0.061679242$$

$$|t| = \frac{\beta_1}{S_{\hat{\beta}_1}} = \frac{0.071630101}{0.061679242} = 1.1613$$

$$\because t(5\ ,\ 0.025) = 2.571$$

$$|t| < t(5\ ,\ 0.025) \Rightarrow \quad \therefore \beta_1 = 0$$

وهذا يعني ان الفارق العمري في الفترة المغلقة [٢ ، ١٢) غير مؤثر تأثيراً معنوياً (سلبياً) على مستقبل الحياة الزوجية ، أي ان الفارق العمري المؤثر معنوياً هو خارج هذه الفترة المغلقة مما يؤكد أهمية هاتين النقطتين المتطرفتين ، وكما يقول (Anscombe)[20] " نحن سعداء بالدفاع عن علاقة الانحدار إذا كانت هذه العلاقة لا تزال ملاءمة بعد حذف مشاهدات قليلة ، أي إننا سعداء إذا كانت علاقة الانحدار نافذة في جميع المشاهدات ولا تنشأ غالباً عن واحدة أو اثنتين من المشاهدات " .



شكل (٤) : الشكل البياني للبواقي الموزونة مقابل قيم X الموزونة

٤- الاستنتاجات والتوصيات

٤-١ الاستنتاجات

١. يلاحظ ان $\hat{\beta}_0$ الذي يمثل متوسط الاستجابة (للانفصال) في بداية الحياة الزوجية ذو قيمة سالبة ومعنوية جدا ، مما يدفع للاستنتاج أن تأثير الفارق في عمر الزوجين والعوامل الاخرى لا يظهر قطعاً عند بداية الحياة الزوجية ، إذ عندها لا يفكر الزوجان إلا في استمراريتها ورسوخها ، وإنما يظهر بتقادم الحياة الزوجيةمما يجعل ضرره أكبر .

٢. يلاحظ ان قيم الاستجابات الخطية المتوقعة هي قيم سالبة حتى عند مستوى الفارق العمري "١٥ سنة" بين الزوجين ، مما يدفع إلى الأسستنتاج انه حتى عند هذا الفارق العمري الكبير فان الحياة الزوجية لن تكون مهددة اذا كان هو العامل الوحيد المؤثر في الحياة الزوجية ، وهذا يعني ان من شروط اشتغال (فعالية) هذا العامل هو وجوب مرافقته لعامل أو أكثر من العوامل الأخرى المؤثرة سلبا في الحياة الزوجية والتي هي عادة ما يكون بعضها موجودا في أية حياة زوجية .

٣. يلاحظ من رسم البواقي أنها جميعا تنتشر قريبة من الصفر عدا نقطتين فهما متطرفتان Extreme وهما اللذان يمثلان الفارق في عمر الزوجين المؤثر بصورة معنوية على مستقبل حياتهما الزوجية وهو الفارق العمري (أكبر أو يساوي ١٥ سنة) ، والفارق العمري (أقل أو يساوي سنة واحدة) لصالح الزوج ، وحيث أنه لم يكن هناك خطأ في قياسهما لذلك نستنتج أهمية هاتين النقطتين ، وبذلك يكون المدى المقبول (غير المعنوي) للفارق العمري بين الزوجين هو $[2 \geq Xi \geq 12]$.

٤. يلاحظ أن قيمة معامل التحديد R^2 يساوي ٤٣% وهي نسبة الاختلاف في الاستجابة التي يفسرها الفارق في عمر الزوجين في معادلة الانحدار الخطي المستخدمة في هذا البحث ، وهذا يعني ان ٥٧% من الاستجابات

المختلفة تعود اسباب اختلافها لعوامل أخرى لم تتضمنها معادلة الانحدار ، مما يدفع إلى الاستنتاج أن اختيار نموذج الانحدار اللوجستي (المتعدد) لدراسة مستقبل الحياة الزوجية هو الاختيار الأفضل .

٤-٢ التوصيات

١. يوصي الباحث بوضع نتائج البحث في متناول ايدي وحدات البحث الاجتماعي في محاكم الأحوال الشخصية لغرض التعرف على المدى المقبول (غير المعنوي) لتأثير الفارق العمري بين الزوجين $[2 \geq Xi \geq 12]$ وشروط فعاليته والعمل على تسجيل الفارق العمري بين الزوجين عند مراجعتهما لغرض عقد القران وليس عند رفع دعوى الطلاق ، لكي يتسنى لهذه الوحدات توجيه الزوجين اللذين يتجاوز الفارق العمري بينهما المدى المقبول بمخاطره المستقبلية .

٢. يوصي الباحث باستخدام وسائل الاعلام وبخاصة "التلفيزيون" لتثقيف الاباء والامهات وتوعيتهم بمخاطر الزواج عندما يكون الفارق العمري بين الزوجين خارج المدى المقبول (غير المعنوي) .

٣. يوصي الباحث باستخدام النموذج اللوجستي في البحوث الاجتماعية التي تكون بياناتها ثنائية ، وليس فقط في البحوث الحياتية .

٤. يوصي الباحث باستخدام النموذج اللوجستي متعدد المتغيرات في مثل هذه البحوث الاجتماعية لكونه أكثر ملاءمة لوصف البيانات ، ومن ثم الحصول على استجابات متوقعة قريبة جدا من الاستجابات المشاهدة .

References

1. Pearl , R. & Reed , L. J. "On The rate of Growth of the population of the United states since 1790 and Mathematical Representation " *National Academy of Sciences* ,NO 6, p. 275 , 1920 .

2. Berkson , J. "Application of the Logistic Function to Bio – Assay" *JASA* , Vol. 39, pp. 357 – 365, 1944.

3. Berkson , J. "Why I prefer Logits to probits" *Biometrics* , Vol. 7 , No. 4 , pp. 327 – 339, 1951.

4. Cox . D. R. "*Analysis of Binary Data*" Chapman F Hall , London , 1970 .

5. Ashton , W.D. "*The logistic transformation with special References to it use in Bio – Assay*" ,Charles Griffin , London , 1972 .

6. Mc Cullagh , P. & Nelder , J. A. "*Generalized Linear Models*" Chapman and Hall , London, 1983 .

7. Richard Kay & Sarah little "Transformations of the explanatory variables in the logistic Regression Model for Binary Data"*Biometrika*, No.3, pp.495-501, 1987 .

8. Hosmer , R.V. & Jovanovic , B. of Lemeshow , S. "BestSubset Logistic Regression" *Biometrics*, Vol. 45, PP. 1265 – 1270, (1989) .

9. Minard scott. "Applied Logistic Regression Analysis " sage puplication series Quantitive Applications in the social scoeunces , No .106 ,1995.

10. Siqueira , A. L. & Jeremy , M. G. , Taylor, "Treatment Effects in a logisticModel involving the Box – Cox Transformation" *Jasa* , Vol. 94 , No. 445 , Theory and Methods , pp. 240 – 246, 1999 .

11. David W. Hosmer F Stanley lemeshw " Applied Logistic Regression" 2nd Ed . John wily & sons, INC ., 2000.

12. ping chao – ying; Joanne Kuk Lida; inger sall, Gray M. "An introduction to logistic Regression Analysis and Reporting " the journal of Education Research articles, September , 2002 .

13. sansh senol & Gozde Ulutagay ,"Logistic Regression Analysis to determine the factors that affect "Green Card " usage for health services ", JFS, vol 29 , pp.18–26 , 2006.

14. Seber, G.A.F. & wild, C.J. "*Nonlinear Regression*", John Wiley and Sons, New York, 1989 .

15. Draper , N. R. & F. Smith , H."*Applied Regression Analysis*" 2nd edition , John Wiley, New York, 1981 .

16. Rat Kowsky, David, A."*Non Linear Regression Modeling*"Marcel Dekker, New York, 1983 .

١٧. سامبرت جاتيرجي وبيرام برايس " تحليل الانحدار بالامثلة " ترجمة محمد مناجد الدليمي، مطابع التعليم العالي في الموصل، جامعة الموصل، ١٩٩٠ .

18. Kendall , M. G.& Stuart, A. "*The advanced Theory of statistics*" Vol. 3, Charles Griffin , London, 1968 .

١٩. الراوي خاشع محمود، "المدخل الى تحليل الانحدار" دار الكتب للطباعة والنشر، جامعة الموصل، ١٩٨٧ .

20. Anscombe , F. J. "Graphs in statistical Analysis" *The American statistician* , Vol. 2, No. 1 , pp. 17 –21, 1973 .