وزارة التربية و التعليم العالي و البحث العلمي

جامعة سبها – كلية العلوم

قسم :الاحصاء

مشروع مقدم لاستكمال متطلبات الحصول على درجة البكالوريوس

بعنوان:-

انحدار بواسون مع تطبيق عملي باستخدام برنامج R

إعداد الطالبة: -

غريبة محمد ابوبكر

تحت إشراف:

أ. د.السعيدي المهدي الطاهر

العام الجامعي

2018-2019 ف

بسم الله الرحمن الرحيم

{ قالوا سبحانك لا علم لنا إلا ما علمتنا

إنك أنت العليم الحكيم}

صدق الله العظيم

سورة البقرة . الآية رقم (32 -33)

الإهداء

بسم الله الرحمن الرحيم

الي الأصحاب السيرة العطرة والفكر المستنير فلقد كان لهم الفضل في بلوغي التعليم العالي أطال الله في أعمارهم.

أعمامي د.عثمان ابوبكر وحافظ

الي من جعلت كل حياتي سعادة وهناء الي من أزالت عني التعب والعناء والشقاء .

أمي الغالية

الي من كد وكدح لأجلي ومن غمرني بعطفه وحنانه ومن جعل سعادتي هدفا في حياته.

أبي الغالي

الي من شاركوني أجمل اللحظات وحملوا في دواخلهم أطيب الأمنيات وأخلصوا لي في الدعوات.

إخوتي وأخواتي

الي من كان في الدرب رفيقا والي القلب قريبا وسندا ومعينا طوال دراستي.

زوجي الغالي

الي أساتذتي ممن كان لهم الدور الكبير في مساندتي الذين أجلهم واحترمهم.

أهديكم بحثي المتواضع

كلمة شكر وتقدير

بسم الله و الحمد لله و الصلاة و السلام على رسول الله نشكرُ الله تعالى أولاً وآخراً . يا من خلقت فسويت . . . أنعمت فأكثرت . . . أعطيت فأكرمت . . .لك الحمد كما ينبغي لجلال وجهك وعظيم سلطانك على ما وفقتني وأعنتني على إنجاز هذا البحث .

أتقدم بجزيل الشكر و العرفان لمن غمرونا بعلمهم و لم يبخلوا علينا بجهدهم ووقتهم في إتمام هذا البحث , فلهم منا جزيل الشكر و التقدير و الاحترام

الأستاذ الدكتور الفاضل // السعيدي المهدي الطاهر

كما يسرني أن أتوجه بالشكر لأعضاء هيئة التدريس بقسم الإحصاء

وفي الختام اتقدم بالشكر والتقدير إلى كل من كان له دور ومشاركة ولو بكلمة طيبة أو دعوة صادقة لأحظى بما فيه من نجاح وأنا أكمل عملي هذا

والحمد لله ولي التوفيق .

فهرس المحتويات

فهرس الأشكال

فهرس الجداول

الملخص

تعتبر نماذج الانحدار وسيلة هامة يلجأ اليها العديد من الباحثين من أجل دراسة تأثير المتغيرات المستقلة أو التوضيحية بشكل عام على متغير التابع أو الاستجابة حيث أخذت هذه النماذج مكانة متميزة في تطبيقات متنوعة في عدة جوانب وعلوم مختلفة . ويعتبر نموذج انحدار بواسون هي حالة خاصة من النماذج الخطية المعممة حيث يفترض متغير الاستجابة عبارة عن أعداد صحيحة موجبة. **هدفت هذه الدراسة** الي التعريف بنموذج انحدار بواسون الذي يعتبر أحد أساليب النمذجة التي تتغلب على بعض مشاكل الانحدار التقليدي والذي له العديد من التطبيقات في العديد من المجالات مثل الدراسات البيولوجيا والرعاية الصحية وعلم النفس وغيرها من الدراسات ويكون الباحث مهتما بكيفية تغير هذه الظواهر ومحاولة نمذجتها بالإضافة الي تطبيق نموذج انحدار بواسون علي بيانات حقيقية باستخدام برنامج R من حيث كيفية تقدير معالم النموذج واختبار معنوية النموذج. **بينت النتائج** بعد تقدير نموذج انحدار بواسون أن قيم احصائي الاختبار للمتغيرات الصحية كانت كبيرة بينما قيم احصائي الاختبار للمتغيرات الاجتماعية والاقتصادية صغيرة وهذا بسبب وجود قيم متطرفة كثيرة في المتغيرات الاجتماعية والاقتصادية التي أثرت علي قيم اختبار وولد.

انحدار بواسون مع تطبيق عملي باستخدام برنامج R

Poisson Regression Analysis using **R**

1.1 تمهيد

يعتبر **انحدار بواسون** هو أحد أشكال التحليل الانحداري المستخدم في نماذج العدّ وجداول الاحتمالات وله أهمية كبيرة في عدة تطبيقات وتزداد مجالات تطبيقها نظرا لطبيعة العديد من الظواهر لهذا التوزيع ويعرف أحياناً بالنموذج اللوغاريتمي الخطي. ويعتبر انحدار بواسون أحد أهم الأساليب الإحصائية التي تستخدم في حال كون قيم متغير الاستجابة أعداد صحيحة وتتبع توزيع بواسون.

2.1 أهداف البحث: تسعي الدراسة الحالية لتحقيق الأهداف التالية:

1. التعريف بنموذج انحدار بواسون وأهم فرضياته وخصائصه.
2. بيان كيفية بناء نموذج انحدار بواسون وتقدير معالمه باستخدام برنامج **R** اعتمادا على بيانات حقيقية.

3.1 أهمية البحث: تكمن أهمية هذا البحث في كونه يقدم أحد أهم نماذج تحليل الانحدار وهو نموذج انحدار بواسون كما أن أهميته تنبع من استخدام أحد البرامج الإحصائية مثل برنامج R وبيان طريقة التطبيق العملي مما يساعد المتخصصين في مجال الاحصاء وغيرهم من تحليل بياناتهم بكل يسر وسهولة.

4.1 منهجية البحث: تم استخدام المنهج الوصفي المتمثل في وصف بيانات الدارسة باستخدام الجداول والرسومات البيانية التي تساعد في التوصل إلى الخصائص العامة لبيانات الدارسة، أيضا تم استخدام المنهج التحليلي الاستنتاجي الذي يتمثل في دراسة تحليل الانحدار عن طريق بناء نموذج انحدار بواسون مع التركيز على كيفية تقدير وتفسير معالمه.

5.1 أقسام البحث: يتكون هذا البحث من أربعة فصول مقسمة علي النحو التالي:

- **الفصل الأول**: يشمل أهداف وأهمية ومنهجية وأقسام البحث.
- **الفصل الثاني**: يتضمن بعض المفاهيم الأساسية المتعلقة بتوزيع بواسون الاحتمالي ثم نموذج انحدار بواسون.
- **الفصل الثالث**: التطبيق العملي باستخدام برنامج R في تقدير معالم نموذج بواسون.
- **الفصل الرابع**: الخلاصة والتوصيات.

الفصل الثاني : نموذج انحدار بواسون

1.2 تمهيد

يتناول هذا الفصل جانبين ، الجانب الأول خصص للتعريف بتوزيع بواسون الاحتمالي وبيان اهم خصائصه الرياضية مثل الدالة الاحتمالية والوسط والوسيط والمنوال والتباين. أما الجانب فيهتم بتقديم بعض المفاهيم المتعلقة بنموذج انحدار بواسون وهو الهدف الرئيسي لهذا البحث .

2.2 توزيع بواسون The Poisson Distribution

سمى هذا التوزيع بهذا الاسم نسبة الى أحد مكتشـــفة وهو بواسون، حيث يعد هذا التوزيع واحدا من أبرز التوزيعات المتقطعة الهامة وله الكثير من التطبيقات الإحصائية ويسمي في بعض الأحيان توزيع الحوادث نادرة الحصول كحوادث الوحدات المعيبة في دفعة معينة أو الأخطاء المطبعية في كتاب معين. ويأتي توزيع بواسون كحالة تقاربية لتوزيع ثنائي الحدين كما بين ذلك العالم الرياضي الفيزيائي الفرنسي (Simeon Denis Poisson) .

الصورة العامة لدالة الكتلة الاحتمالية لتوزيع بواسون تكتب بالشكل التالي: ــ

$$f(x) = \begin{cases} \dfrac{\lambda^x \ e^{-\lambda}}{x\,!} & , \qquad x = 0,1,2,3,.......... \\ 0 & , \end{cases}$$

حيث $e = 2.718$ تمثل مقدار ثابت.

يشترط توزيع بواسون أن تكون عدد المحاولات مستقلة واحتمال النجاح ثابت من محاولة لأخري ومجموع احتمال النجاح والفشل في كل محاولة يساوي الواجد الصحيح .

ويمكن اثبات انها دالة كتلة احتمال: يلاحظ أن المتســـلســـلة

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + ... = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$f(x) \geq 0$$

$$\therefore \sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = e = 1$$

$\therefore$ الدالة f (x) تحقق شروط دالة كتلة احتمال لمتغير عشوائي منقطع.

1.2.2 دالة التوزيع لتوزيع بواســـون:

ان الدالة التوزيعية لتوزيع بواسون يمكن التعبير عنها بالشكل التالي:

$$F(x) = p(X \leq x) = \sum_{x=0}^{x} f(x) \qquad X \geq o$$

$$F(x) = \sum_{x=0}^{x} \frac{e^{-\lambda} \lambda^x}{x!} \qquad X \geq 0$$

2.2.2 خصائص توزيع بواسون Properties *of Poisson Distribution*

أولا: الوسط (العزم الأول):

$$E(x) = \mu_1'(x) = \sum_{x=o}^{\infty} x \lambda^x \frac{e^{-\lambda}}{x!}$$

$$= e^{-\lambda} \sum_{x=o}^{\infty} \frac{x \lambda \lambda^{x-1}}{x(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x=o}^{\infty} \frac{\lambda^x}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$y = x - 1$$

$$= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}$$

$$= \lambda e^{-\lambda} e^{\lambda}$$

$$E(x) = \lambda$$

ثانيا: العزم الثاني:

$$E(x^2) = \sum_{x=0}^{\infty} \; x^2 \; f(x)$$

$$= \sum_{x=0}^{\infty} \; x^2 \; \frac{\lambda^x \; e^{-\lambda}}{x\,!}$$

$$= \sum_{x=0}^{\infty} \; \left[x^2 - x + x\right] \; \frac{\lambda^x \; e^{-\lambda}}{x\,1}$$

$$= \sum_{x=0}^{\infty} \; \left[x(x-1) + x\right] \; \frac{\lambda^x \; e^{-\lambda}}{x\,!}$$

$$= \sum_{x=0}^{\infty} \; x(x-1) \; \frac{\lambda^x \; e^{-\lambda}}{x\,!} + \sum_{x=6}^{\infty} \times \frac{\lambda^x \; e^{-\lambda}}{x\,!}$$

$$= \sum_{x=0}^{\infty} \; x(x-1) \; \frac{\lambda^x \; e^{-\lambda}}{x(x-1)\,(x-2)} + E\,(x)$$

$$E(x^2) = \sum_{x=2}^{\infty} \; \frac{\lambda^x \; e^{-\lambda}}{(x-2)!} + \lambda$$

$$put \quad y = x - 2 \;\Rightarrow\; when \quad x = 2 \qquad y = 0$$
$$x = y + 2 \;\Rightarrow \qquad x = \infty \qquad y = \infty$$

$$E(x^2) = \sum_{y=0}^{\infty} \; \frac{\lambda^{y+2} \; e^{-\lambda}}{y\,!} + \lambda$$

$$= \lambda^2 \sum_{y=0}^{\infty} \; \frac{\lambda^y \; e^{-\lambda}}{y\,!}$$

$$E(x^2) = \lambda^2 + \lambda$$

$$\therefore \sigma_x^2 = E(x^2) - \left[E(x)\right]^2$$

$$= \mu_2' - \mu_1'^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$\sigma_x^2 = \lambda$$

ان الخاصية المميزة لتوزيع بواسون عن بقية التوزيعات الأخرى هي أن متوسط هذا التوزيع مساو لتباينه ويكون مساويا لقيمة المعلمة.

ثالثا: الدالة المولدة للعزوم: يمتلك توزيع بواسون دالة مولدة لعزومه وفقا للصيغة التالية:

$$M(t) = e^{\lambda(e^{t-1)})}$$

الاثبات

$$M(t) = E(e^{tx}) = \sum_{x}^{\infty} e^{tx} \; f(x)$$
$$= \sum_{x}^{\infty} e^{tx} \; \frac{\lambda^x \; e^{-\lambda}}{x!}$$
$$= e^{-\lambda} \; \sum_{x}^{\infty} \frac{\left(e^t \; \lambda\right)^x}{x!} \; = \; e^{-\lambda} \; \left(e^{et\lambda}\right)$$
$$= e^{-\lambda} \; e^{(et\lambda)} \; = e^{-\lambda + et\lambda} \; = e^{-\lambda(1-e^t)}$$

لجميع قيم t الحقيقية.

باستخدام الدالة المولدة للعزوم يمكن إيجاد الوسط الحسابي والانحراف المعياري لتوزيع بواسون كما يلي:

$$M'(t) \; = e^{\lambda(e^t-1)} \; \lambda \; e^t \; = \lambda \; e^{t+\lambda(e^t-1)}$$
$$M''(t) = e^{\lambda(e^{t-1})} \; \lambda e^t \; + \left(\lambda \; e^t\right)^2 \; e^{\lambda(e^{t-1})}$$
$$\mu \quad = M'(0) = \lambda$$
$$\sigma^{\tilde{}} = M''(0) - \mu^2 \; = \lambda + \lambda^2 \; - \lambda^2 \; = \lambda$$

∴ المعلمة λ فى توزيع بواسون هي الوسط الحسابي للتوزيع.

رابعا: الدالة المولدة للاحتمالات لتوزيع بواســون Probability Generating Function For Poisson Distribution.

$$G(t) = E\left[t^x\right] = \sum_{x=0}^{\infty} t^x \;\; f(x)$$
$$= \sum_{x=0}^{\infty} t^x \; \frac{\lambda^x \; e^{-\lambda}}{x!}$$
$$= \sum_{x=0}^{\infty} (t \; \lambda)^x \; \frac{e^{-\lambda}}{x!}$$
$$= e^{-\lambda} \; \sum_{x=0}^{\infty} \frac{(t\lambda)^x}{x!}$$
$$= e^{-\lambda} \; e^{t\lambda} \; = e^{-\lambda + t\lambda}$$

$$G(t) \; = \; e^{\lambda(t-1)} \; = e^{-\lambda(1-t)}$$

اشتقاق عزوم بواســـون من الدالة المولدة للاحتمالات

$$\therefore G'(t) = e^{\lambda(t-1)} \ \lambda$$
$$G''(t) = \lambda \ e^{\lambda(t-1)} \ \lambda$$
$$= \lambda^2 \ e^{\lambda(t-1)}$$

$$\therefore G'(1) = \lambda = \mu = E(x)$$
$$G''(1) = \lambda^2$$
$$\sigma^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

ايجاد المتوسط و التباين الدالة المولدة للاحتمالات

$$G(t) = E(t^x)$$
$$= \sum_{x=0}^{\infty} t^x \ f(x)$$
$$= \sum_{x=0}^{\infty} t^x \ \frac{\lambda^x \ e^{-\lambda}}{x!}$$
$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(t \ \lambda)^x}{x!}$$
$$= e^{-\lambda} \ e^{t\lambda}$$
$$= e^{-\lambda + t\lambda}$$
$$G(t) = e^{\lambda(t-1)}$$
$$G'(t) = e^{\lambda(t-1)} \ \lambda$$
$$\mu = G'(1) = \lambda = \mu_1^1$$
$$G''(t) = \lambda \ e^{\lambda(t-1)} \ \lambda$$
$$= \lambda^2 \ e^{\lambda(t-1)}$$
$$G'(1) = \lambda^2$$
$$\sigma_x^2 = G''(1) + G'(1) - [G'(1)]^2$$
$$\sigma_x^2 = \lambda^2 + \lambda - \lambda^2$$
$$\sigma_x^2 = \lambda$$

خامسا: الدالة المميزة لتوزيع بواســون Characteristic function of Poisson Distribution.

تعرف الدالة المميزة للمتغير العشوائي X بالشكل

$$\varphi_x(t) = E\left[e^{itx}\right]$$

بإجراء التفاضل الجزئي بالنسبة للمتغير t

$$Q'_x(t) = E\left\{\left[e^{itx}\right] i\ x\right\}$$
$$Q'_x(0) = E\left[i\ x\right]$$
$$= i\ E(x)$$
$$\Rightarrow E(x) = \frac{1}{x} Q'_x(0)$$
$$\therefore \mu = \mu'_1 = E(x) = \frac{1}{i}\ Q'_x\ (0)$$

و بالتفاضل بالنسبة الى t مرة أخرى

$$Q''_x(t) = E\left[e^{itx}\ (i\,x)\ (i\,x)\right]$$
$$Q''_x(t) = E\left[(i\,x)^2\ \ e^{itx}\right]$$
$$Q''_x(0) = E\left[i^2\ x^2\right]$$
$$Q''_x(0) = i^2\ 1 E(x^2)$$
$$\therefore E(x^2) = \frac{1}{i^2}\quad Q''_x\ (0)$$
$$\mu'_2 = E(x^2) = \frac{1}{i^2}\quad Q''_x(0)$$

و يمكن أثبات أن $\mu'_3 = E(x^3) = \frac{1}{i^3}\ \ Q'''_x\ (0)$

و بوجه عام فأن $\mu'_r = E(x^r) = \frac{1}{i^r}\ \ Q^r_x\ (0)$

سادسا: منوال توزيع بواســـون .Mode of Poisson Distribution

نلاحظ أن

$$\frac{f(x+1)}{f(x)} = \frac{\lambda^{x+1}\ e^{-\lambda}\ /\ (x+1)\,!}{\lambda^x\ e^{-\lambda}\ /\ x\,!}$$

$$\frac{f(x+1)}{f(x)} = \frac{\lambda}{x+1}$$

وعلى ذلك توجد لدينا ثلاث حالات

أ ـــ اذا كان البسط $>$ المقام أى أن

$$f(x+1) > f(x)$$
$$\lambda > x+1$$
$$\Rightarrow x > \lambda - 1$$

∴ تكون f (x) متزايدة (لها متوال وحيد)

ب ـ اذا كان البسط $\langle$ اقل المقام أى أن

$$\lambda < x+1$$
$$\therefore\ x > \lambda - 1$$

∴ تكون f (x) دالة متناقصة (لها متوال وحيد)

ج ـ اذا كان البسط = المقام

$$\lambda = x+1 \quad \Rightarrow \quad x = \lambda - 1$$
$$\therefore\ \frac{f(x+1)}{f(x)} = 1 \qquad \text{–1}$$

$$\therefore\ f(x+1) = f(x)$$

نعوض عن x من 1

$$f(\lambda - 1 + 1) = f(\lambda - 1)$$
$$\therefore\ f(\lambda) = f(\lambda - 1)$$

و يكون هناك منولان اذا كان $\lambda - 1$ عدد صحيح عند $x = \lambda - 1$, $x = \lambda$

← إما إذا كان $\lambda - 1$ عدد غير صحيح فأن المنوال يكون هو أول عدد صحيح يأتي فوراً بعد $\lambda - 1$

3.2 نموذج انحدار بواسون Poisson Regression Model

يعتبر نموذج انحدار بواسون هو أحد أنواع النماذج الخطية اللوغاريتمية (**Log-Linear Models**) وهو النموذج الملائم لتحليل البيانات التي تكون بهيئة بيانات معدودة (**Count Data**) أو معدلات (**Rate Data**) وأن هذه البيانات هي أعداد صحيحة غير سالبة وجاءت هذه التسمية للنموذج نتيجة امتلاك الخطأ العشوائي فيه توزيع بواسون وبالتالي يتوزع متغير الاستجابة وفقا لنفس التوزيع أما كونه خطيا لوغاريتميا فذلك يعني من خلال أخذ اللوغارتم الطبيعي لصيغة النموذج فإنها تتحول الي صيغة خطية ويستخدم عندما يكون متغير الاستجابة بيانات معدودة أو معدلات والتي تكون أحداث نادرة الحدوث مثل عدد تصادم السفن أو معدلات حالات تصادم السفن .. الخ.

4.2 أمثلة علي انحدار بواسون Examples of Poisson Regression

يُستخدم نموذج انحدار بواسون للتنبؤ بمتغير تابع يتكون من "عدد البيانات" مع تحديد واحد أو أكثر من المتغيرات المستقلة. ويطلق على المتغير الذي نريد التنبؤ به المتغير التابع (أو في بعض الأحيان متغير الاستجابة أو النتيجة أو الهدف) وتسمى المتغيرات التي نستخدمها للتنبؤ بقيمة المتغير التابع بالمتغيرات المستقلة أو المتغيرات التفسيرية أو التوضيحية. وفيما يلي بعض الأمثلة التي يمكن استخدام انحدار بواسون:

مثال-1: يمكن استخدام انحدار بواسون لفحص عدد الطلاب المعلقين من قبل المدارس في واشنطن في الولايات المتحدة استنادًا إلى تنبؤات مثل الجنس (ذكور واناث) ، والعرق (الأبيض والأسود والهسباني والآسيوي) ، اللغة (الإنجليزية هي لغتهم الأولى ، واللغة الإنجليزية ليست لغتهم الأولى) وحالة الإعاقة (المعاقين وغير المعاقين). هنا، "عدد حالات التعليق" هو المتغير التابع، في حين أن "النوع" ، "العرق" ، "اللغة" و "حالة الإعاقة" كلها متغيرات مستقلة اسمية.

مثال-2: يمكن استخدام انحدار يواسون لفحص عدد المرات التي يتخلف فيها الأشخاص في أستراليا عن سداد بطاقتهم الائتمانية في فترة خمس سنوات استنادًا إلى تنبؤات مثل الحالة الوظيفية (الموظف والعاطل عن العمل) والراتب السنوي (بالدولار الأسترالي) والعمر (بالسنوات) والجنس (ذكور وإناث) ومستويات البطالة في البلاد (٪ العاطلين عن العمل). هنا، "عدد مرات التخلف عن سداد بطاقات الائتمان الافتراضية" هو المتغير التابع، في حين أن "الحالة الوظيفية" و "النوع" هي متغيرات مستقلة اسمية، و "المرتب السنوي" ، "العمر" و "مستويات البطالة في البلد" هي متغيرات مستقلة مستمرة.

مثال-3: يمكن استخدام انحدار بواسون لفحص عدد الأشخاص الذين ينتظرون في طابور قسم الحوادث والطوارئ في المستشفى استنادًا إلى تنبؤات مثل وضع الوصول في الحوادث والطوارئ (سيارة الإسعاف أو تسجيل الوصول الذاتي) ، شدة الإصابة أثناء الفرز (بسيط، معتدل ، شديد) ، وقت من اليوم ويوم من الأسبوع. هنا، "عدد الأشخاص الذين ينتظرونك في قائمة الانتظار" هو المتغير التابع، في حين أن "نمط الوصول" هو متغير مستقل اسمي، "تقييم شدة الإصابة" هو متغير مستقل ترتيبي، و "الوقت من اليوم" و " يوم من الأسبوع "هي متغيرات مستقلة مستمرة.

5.2 الشكل العام لانحدار بواسون General Form For Poisson Regression Model

يعتبرانحدار بواسون (Poisson Regression) هو أحد أشكال تحليل الانحدار المناسب في نمذجة البيانات عندما تكون بيانات المتغير التابع منفصلة معدودة وتتبع توزيع بواسون. ولعرض شكل نموذج انحدار بواسون نفرض أن لدينا عينة عشوائية من المتغيرات المستقلة $Y =$ $(Y_1, Y_2, \dots Y_k)$, مستمرة أو خليط من المتغيرات المستمرة والتصنيفية كلا منها يتبع توزيع بواسون بمتوسط μ

متوسط الاستجابة لعدد الأحداث التي تحدث في فترة زمنية معينة تستبدل بدالة غير خطية (أسية) يمكن التعبير عنها بالعلاقة التالية:

$$E(y_i \mid \mathbf{x}_i) = \lambda_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$$

حيث أن $\boldsymbol{\beta}$ = تمثل متجه معلمات الانحدار وأن $X = (X_1, X_2, \dots X_k)$, هو متجه المتغيرات المستقلة

ويمكن كتابتها بشكل مكافئ:

$$\ln(\lambda_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}$$

حيث ان اللوغارتم للقيمة المتوقعة عبارة عن توليفة خطية للمتغيرات المستقلة التي تحتوي علي معالم مجهولة والتي يتم تقديرها بإحدى طرق التقدير (طريقة المربعات الصغرى أو طريقة الإمكان الأعظم) وفي بعض الأحيان يسمي نموذج انحدار بواسون بنموذج اللوغاريتمي الخطي

متحه الاستجابة ذي الرتبة n*1

مصفوفة المتغيرات المستقلة ذي الرتبة (P=1) * n .

متجه معلمات النموذج ذي الدرجة 1* (P+1) .

n: حجم العينة.

P:عدد المتغيرات المستقلة

ملاحظة: خاصية تساوي الوسط مع التباين في توزيع بواسون غير مشروطة التحقق في انحدار بواسون ففي بعض الأحيان ضمن الجوانب التطبيقية غالبا ما يكون التباين للمتغيرات المعدودة (Count Variables) أكبر من الوسط الحسابي فتعرف هذه الخاصية بفوق التشتت (Over Dispersion).

6.2 فروض نموذج انحداربواسون : Assumptions of Poisson Regression :

عند استخدام انحدار بواسون لتحليل بيانات ظاهرة معينة لابد من التحقق أو التأكد من بعض الافتراضات من أن البيانات التي تريد تحليلها يمكن تحليلها باستخدام انحدار بواسون عند تحقق تلك الافتراضات من أجل الحصول علي نتائج صحيحة يمكن الاعتماد عليها. ومن الناحية العملية أو التطبيقية يمكن التحقق من هذه الافتراضات الخمسة قبل تنفيذ انحدار بواسون. فمن الضروري أن تقوم بهذا الاجراء لأن البيانات تحت الدراسة قد لا يتحقق فرض أو أكثر من هذه الافتراضات (انتهاك البيانات لهذه الفروض) وبالتالي لا يمكن استخدام انحدار بواسون في حالة عدم تحقق واحدة أو أكثر من هذه الافتراضات أو سوف يؤدي الي عدم الوثوق في نتائج انحدار بواسون في ظل عدم تحقق الافتراضات ومن هذه الفرضيات هي:

الفرض الأول: يتكون المتغير التابع الخاص بانحدار بواسون من بيانات العد. تختلف بيانات العد للبيانات المقاسة في الأنواع الأخرى المعروفة من الانحدار (على سبيل المثال، الانحدار الخطي والانحدار المتعدد تتطلب متغيرات تابعة تقاس على مقياس "مستمر" ، يتطلب الانحدار اللوجستي ذو الحدين متغيرًا تابعًا يقاس على "ثنائي التصنيف" على نطاق ، يتطلب الانحدار الترتيبي متغيرًا تابعًا يقاس على مقياس "ترتيبي" ، ويتطلب الانحدار اللوجستي متعدد الحدود متغيرًا تابعًا يقاس على مقياس "اسمي". في المقابل، تتطلب متغيرات العد عددًا صحيحًا من البيانات التي يجب أن تكون صفرية أو أكبر. بعبارات بسيطة في "عدد صحيح" كرقم "كامل" (على سبيل المثال ، 0 ، 1 ، 5 ، 8 ، 354 ، 888 ، 23400 ، وما إلى ذلك). أيضًا ، نظرًا لأن بيانات التعداد يجب أن تكون "موجبة" (أي ، تتكون من قيم عدد صحيح غير سالب "nonnegative") ، لا يمكن أن تتكون من قيم "سالبة" (على سبيل المثال ، قيم مثل -1 ، -5 ،

-8 ، -354 ، -888 و -23400 لن تعتبر البيانات العد). علاوةً على ذلك ، يُقترح أحيانًا أن يتم تنفيذ انحدار بواسون فقط عندما يكون متوسط العدد عبارة عن قيمة صغيرة (على سبيل المثال ، أقل من 10). عندما تكون هناك أعداد كبيرة قد يكون نوعًا مختلفًا من الانحدار أكثر ملاءمة.

علي سبيل المثال تشمل أمثلة متغيرات العد عدد الرحلات التي تأخرت لأكثر من ثلاث ساعات في المطارات الأوروبية، وعدد الطلاب الذين علقتهم المدارس في واشنطن بالولايات المتحدة، وهو عدد المرات التي يعجز فيها الأشخاص في أستراليا عن سداد أقساط بطاقاتهم الائتمانية في غضون خمس سنوات الفترة، وعدد الأشخاص الذين ينتظرونك في الطابور في قسم الحوادث والطوارئ في المستشفى، وعدد الطلاب الذين حصلوا على علامة الدرجة الأولى (عادة أقل من 5) في برنامج ماجستير إدارة الأعمال

الفرضية الثانية: المتغير المستقل (واحد أو أكثر من المتغيرات المستقلة) ، والتي يمكن قياسها على نطاق مستمر أو ترتيبي أو اسمي / ثنائي التفرع. يمكن تصنيف المتغيرات الترتيبية والإسمية / ثنائية التفرع على نطاق واسع كمتغيرات فئوية.

وتتضمن أمثلة المتغيرات المستمرة وقت المراجعة (يتم قياسه بالساعات) والذكاء (يتم قياسه باستخدام درجة ذكاء IQ) وأداء الاختبار (يتم قياسه من 0 إلى 100) والوزن (يتم قياسه بالكيلو جرام). تتضمن أمثلة المتغيرات الترتيبية عناصر Likert (على سبيل المثال ، مقياس من 7 نقاط من "موافق بشدة" خلال "لا أوافق بشدة") ، من بين طرق أخرى لتصنيف الفئات (على سبيل المثال ، مقياس مكون من 3 نقاط يشرح مدى إعجاب العميل بمنتج ما تتراوح من "ليس كثيرا" إلى "نعم ، كثيرا"). تشمل أمثلة المتغيرات الاسمية النوع (على سبيل المثال ، مجموعتين - ذكور وإناث - كما يُعرف أيضًا باسم المتغير ثنائي التفرع) ، والعرق (على سبيل المثال ، ثلاث مجموعات: القوقاز ، الأمريكيين من أصل أفريقي والهسباني) والمهنة (على سبيل المثال ، خمس مجموعات: الجراح ، والطبيب وممرضة وطبيب أسنان ومعالج). تذكر أن المتغيرات الترتيبية والإسمية / ثنائية التفرع يمكن تصنيفها على نطاق واسع كمتغيرات فئوية.

الفرضية الثالثة: يجب أن يكون لديك استقلال الملاحظات. وهذا يعني أن كل ملاحظة مستقلة عن الملاحظات الأخرى؛ أي أن إحدى الملاحظات لا يمكن أن توفر أي معلومات عن ملاحظة أخرى. هذا هو افتراض مهم جدا. نقص الملاحظات المستقلة هو في الغالب قضية تصميم الدراسة. تتمثل إحدى طرق الاختبار لإمكانية استقلالية الرصد في مقارنة الأخطاء المعيارية القائمة على النماذج بالأخطاء القوية لتحديد ما إذا كانت هناك اختلافات كبيرة.

الفرضية الرابعة: توزيع العدد (الشرطي على النموذج) يتبع توزيع بواسون Poisson. أحد عواقب ذلك هو أن العدد المشاهد والمتوقع يجب أن يكون متساوياً (في الواقع، متشابه جداً). بشكل أساسي، هذا ما يقوله النموذج يتنبأ بالعدد الملحوظ بشكل جيد. يمكن اختبار ذلك بعدد من الطرق، ولكن هناك طريقة واحدة لحساب المتوقع ورسمها مع العدد المشاهد لمعرفة ما إذا كانت متشابهة.

الفرضية الخامسة: متوسط وتباين توزيع النموذج متطابقان. هذا هو نتيجة للافتراض رقم 4؛ حيث أن تباين توزيع Poisson له نفس قيمة المتوسط.

الفرضية الساسة: يتطلب الامر عدم وجود قيم متطرفة أو قيم مفقودة

ويمكن تلخيص هذه الافتراضات في النقاط التالية:

1. قيم المتغير التابع يجب ان تكون أعداد معدودة واذا لم يتحقق هذا الفرض فان نموذج انحدار بواسون يكون غير ملائم استخدامه.
2. قيم المتغير التابع يحب ان تكون اعداد موجبة غير سالبة ولا كسرية لأن توزيع بواسون توزيع منفصل قيم المتغير لهذا التوزيع يجب أن تكون اعداد صحيحة موجبة
3. قيم المتغير التابع يجب ان تتبع توزيع بواسون
4. قيم المتغير المستقل قد تكون مستمرة أو ترتيبية أو اسمية
5. المشاهدات يجب ان تكون مستقلة

7.2 المقارنة بين انحدار بواسون والانحدار الخطي الكلاسيكي:

هناك عدة اختلافات بين الانحدارين منها :

1. في انحدار بواسون نفترض أن الأخطاء تتبع توزيع بواسون بينما في الانحدار يشترط تبعية الأخطاء للتوزيع الطبيعي
2. في انحدار بواسون يشترط ان تكون البيانات دائما اعداد صحيحة موجبة بينما في الانحدار الكلاسيكي قد تكون موجبة ام سالبة او كسرية.
3. في انحدار بواسون نجد أن قيم المتغير التابع أعداد صحيحة موجية وغير سالبة هذا يضمن أن البيانات دائما موجبة بينما في الانحدار العادي قد تكون القيم موجبة ام سالبة أو كسرية.

8.2 الحالات التي لا يستخدم فيها انحدار بواسون

1. إذا احتوت بيانات Y على أصفار، فلن يستخدم عادةً انحدار بواسون.
2. إذا كان بيانات Y غير صحيحة، فلن تستخدم انحدار بواسون.
3. لا توجد فجوات بين قيم متغير الاستجابة، على سبيل المثال، 0 و 5: يتم تمثيل كل الأعداد الصحيحة بينهما من خلال احتمالات كبيرة. هذا يعني، على سبيل المثال، إذا كانت بيانات Y الخاصة بك يمكن أن تكون مضاعفات 5 فقط، مثل 0 ، 5 ، 10 ، 15 ، وما إلى ذلك ، فلن تستخدم الانحدار بواسون.
4. لا يوجد حد علوي مميز على Y. وهذا يعني أنه إذا كانت بيانات Y الخاصة بك مقيدة أعلاه، كما هو الحال في استجابة الاستبيان التي يمكن أن تكون 0 أو 1 أو 2 أو 3 أو 4 ، مع عدم وجود شيء أعلى ، فلن تستخدم انحدار بواسون.
5. إذا كان توزيع البيانات لمتغير الاستجابة ملتوي نحو اليسار فلن تستخدم انحدار بواسون لأن توزيع بواسون ملتوي نحو اليمين.

9.2 تقدير معلمات النموذج: هناك عدة طرق يمكن استخدامها في تقدير معالم النموذج المجهولة. في هذا الجزء سيتم التطرق الي طريقة الامكان الأعظم بشكل موجز للحصول علي تقديرات معلمة نموذج انحدار بواسون المجهولة وتوضيح سبب عدم امكانية استخدام طريقة المربعات الصغري.

1.9.2 طريقة المربعات الصغرى الاعتيادية (OLS)

تعتبر طريقة المربعات الصغرى العادية (OLS) ، إحدى التقنيات الإحصائية الأكثر استخداما في البحث التطبيقي لتقدير معالم نماذج الانحدار حيث ترتكز هذه الطريقة OLS على افتراضات معينة منها أن قيم المتغير التابع هو قيم مستمرة وموزعة بشكل طبيعي ومرتبطة خطيًا بالمتغيرات المستقلة على وجه الخصوص وهذه الافتراضات في كثير من الأحيان قد لا تتحقق مع بعض البيانات مثل انحدار بواسون الذي يفترض أن بيانات المتغير التابع أعداد صحيحة غير سالبة (بيانات غير مستمرة) وأن الأخطاء تتوزع وفقا لتوزيع بواسون (غير طبيعي) وبالتالي لا تعمل طريقة المربعات الصغرى الاعتيادية بصورة جيدة وأن المقدرات التي يتم ايجادها غير كفؤة وبالتالي لا يمكن الاعتماد عليها في عمليات التنبؤ.

2.9.2 طريقة الامكان الأعظم MLE

Maximum Likelihood Estimation (MLE) of a Poisson regression model

تعد طريقة الامكان الأعظم احدي أهم طرق التقدير والأكثر استخداما لما تمتاز به تقديراتها من خصائص جيدة خاصة في حالة العينات الكبيرة حيث تم اقتراح هذه الطريقة من قبل العالم R.A.Fisher . مبدأ هذه الطريقة هو إيجاد تقدير لمعلمة المجتمع المجهولة التي تجعل دالة الإمكان الأعظم في نهايتها العظمي وأنها تعطي مقدرات كافية ان وجدت وتكون غير متحيزة وتمتاز بأقل تباين. وهناك بعض الخصائص لهذه الطريقة سنذكرها دون اللجوء الي تقديم البراهين اللازمة لها حتي لا نخرج عن الهدف الرئيسي لهذه الدراسة ومن هذه الخصائص هي:

1. أن تقديرات الإمكان الأعظم هي تقديرات متسقة.
2. أن تقديرات الإمكان الأعظم هي تقديرات أكثر كفاءة من بين جملة من تقديرات أخري متاحة.
3. أن التوزيع الاحتمالي لتقدير الإمكان الأعظم تؤول الي التوزيع الطبيعي عندما يكون حجم العينة كبير.

ولتطبيق هذه الطريقة في تقدير معلمة نموذج انحدار بواسون نفرض أن $Y_1,...,Y_n$ متغيرات عشوائية مستقلة لها توزيع بواسون بالمعدل λ لها دالة كتلة احتمالية تأخذ الشكل التالي:

$$P(Y-y)=\frac{\lambda^{y}e^{-\lambda}}{y!},y=0,1...$$

وبفرض أن المعدل λ لها علاقة غير خطية بالمتغيرات المستقلة تأخذ الشكل التالي $\lambda = e^{x\beta}$ وبأخذ اللوغارتم الطبيعي نحصل علي الشكل التالي :

$$\ln(\lambda)=x^{t}\beta=\beta_0+\beta_1 x_1+...+\beta_p x_p=\beta_0+\sum_{j=1}^{p}\beta_j x_j$$

ولتقدير معالم نموذج انحدار بواسون نستخدم طريقة الامكان الاعظم لقياسات العينة حيث تعرف بأنها التوزيع المشترك لتلك القياسات فاذا رمزنا لدالة الامكان بالرمز L وبالتالي يتم تعريفها علي النحو التالي: !

$$L(Y,\lambda)=\Pi P(Y=y_i)=\prod_{i=1}^{n}\frac{\lambda^{y}e^{-\lambda}}{y}=\frac{\prod_{i=1}^{n}\lambda_i^{\ y_i}e^{-\lambda}}{\prod_{i=1}^{n}y_i!}$$

وبأخذ اللوغارتم لدالة الإمكان الأعظم نحصل علي الشكل التالي:

$$\ln L(Y,\lambda)=\ln\frac{\prod_{i=1}^{n}\lambda_i^{\ y_i}e^{-\lambda}}{\prod_{i=1}^{n}y_i!}=\sum_{i=1}^{n}y_i\ln\lambda-n\lambda-\ln y!_i,\ where\,\lambda=e^{x'\beta}$$

وبإيجاد المشتقة الأولي ومساواته بالصفر نحصل علي الشكل التالي:

$$\frac{\partial L}{\partial\beta}=\sum_{i=1}^{n}x_i(y_i-e^{x'\beta})=0$$

وبحل هذه المعادلة باستخدام التحليل بطريقة نيوتن رأفسون نحصل علي تقدير المعلمة والتي تأخذ الشكل التالي:

$$\hat{\beta}=\arg solve=\sum_{i=1}^{n}x_i(y_i-e^{x'\beta})=0$$

الفصل الثالث: التطبيق العملي

1.3 تمهيد: هذا الفصل يتضمن الجانب العملي (التطبيقي) حيث سيتم توفيق نموذج انحدار بواسون لبيانات الدراسة باستخدام برنامج R ويشمل ذلك إيجاد معالم نموذج انحدار بواسون واختبار معنويتها الاحصائية وتشخيص النموذج المقدر.

2.3 البرنامج التطبيقي R

تعتبر برمجية R واحدة من أهم البرمجيات في الإحصاء وفي عدة مجالات أخرى حيث تدّرس اليوم في مواد مختلفة في الجامعات وتستخدم في الحسابات الإحصائية بصورة عملية. وتعتبر لغة R هي برمجية إحصائية اخترعها Ross Ihaka و Robert Gentlema ولهذا سميت بلغة R وبيئة تطوير (Language and development environment) متخصصة في تحليل وتمثيل البيانات الإحصائية . وتتكونR من حزمة رئيسية "Core" يمكن توسيعها بواسطة حزم أخرى موجودة بمستودعات المشروع الرئيسي ."CRAN" يبلغ عدد هذه الحزم لحد الآن 2456 .وهي تقدم مكتبات للطرق الإحصائية الأساسية والمتقدمة كالإحصاء الوصفي، والاختبارات الإحصائية، وتخطيط التجارب، وتحليل الارتباطات الخطية، واللاخطية، وتحليل المتتاليات الزمنية، والتحليل متعدد المتغيرات ...الخ.

وعلى الرغم من أن هذه البرمجية مجانية ومصدر مفتوح للجميع، إلا أن كثير من الطلاب والباحثين ليس لديهم في أغلب الأحيان معرفة كافية باستخدامها بصورة مهنية. وتعد لغة R من اللغات التي برزت حديثا وتزايدت بشكل سريع بمجال البرمجة العلمية في قطاعي الإحصاء والمعلوماتية الحيوية (Bioinformatics) حيث باتت معتمدة على نطاق واسع في كثير من الجامعات ومراكز البحوث العلمية، حيث نرى استخدامها والإشارة إليها في العديد من المقالات المنشورة بالمجلات العلمية المحكّمة.

وسيتم في هذا البحث تطبيق برنامج R **لتقدير نموذج انحدار بواسون حيث أن هناك** عدة برامج إحصائية يمكن تطبيقها لا جراء انحدار بواسون والتي يمكن تلخيصها وعرضها قي الجدول التالي:

جدول (1.3): يوضح البرامج التي يمكن استخدامها لتقدير معالم نموذج انجدار بواسون

البرنامج	الوصف
MATLAB	MATLAB Statistics Toolbox: Poisson regression can be performed using the "glmfit" and "glmval" functions.
Microsoft Excel:	Excel is not capable of doing Poisson regression by default. One of the Excel Add-ins for Poisson regression is Xpost
R	The function for fitting a generalized linear model in R is glm(), and can be used for Poisson Regression
SAS:	Poisson regression in SAS is done by using GENMOD
Stata:	Stata has a procedure for Poisson regression named "poisson"
mPlus:	mPlus allows for Poisson regression using the command COUNT IS when specifying the data
SPSS:	In SPSS, Poisson regression is done by using the GENLIN command

3.3 وصف البيانات Data Descriptive :

ارتكزت الدراسة الحالية علي البيانات التي استخدمت من قبل كل من Deb and Trivedi (1997) بتحليل البيانات المتعلقة بالطلب على الرعاية الطبية من قبل كبار السن والبالغ عددهم 4406 أفراد تتراوح أعمارهم بين 66 عامًا فأكثر لسنة 1987/88 19 وهذه البيانات موجودة في حزمة R . تم أخذ هذه البيانات من أحد مكتبات برنامج R المسماه 'SenSrivastava' الموجود علي الموقع http://www.econ.queensu.ca/jae/1997-v12.3/deb-trivedi/

حيث كان الهدف هو نمذجة الطلب على الرعاية الطبية | عدد الأطباء / غير الأطباء وزيارات العيادات الخارجية بالمستشفيات | من قبل المتعاونين المتاح للمرضى. هنا ، عدد زيارات الطبيب هو المتغير التابع (ofp) بينما المتغيرات المستقلة الأولي تتعلق باستخدام متغيرات الحالة الصحية (hosp) عدد الإقامات في المستشفى، الحالة الصحية المتصورة ذاتيا (health) عدد الحالات المزمنة (numchron) ، وكذلك المتغيرات الاجتماعية والاقتصادية مثل متغيرات الجنس (gender) ، والمدرسة (school) (عدد سنوات التعليم) ، والقطاع الخاص مؤشر التأمين (privins). ويمكن عرض المعلومات الخاصة بالمتغير التابع (المعتمد) والمتغيرات المستقلة (التوضيحية) كما هو موضح في الجدول التالي :

جدول (2.3) يوضح المعلومات الخاصة بمتغيرات الدراسة

المعني الإنجليزي	الترجمة	الرمز المستخدم	نوع المتغير
the number of physician office visits	عدد زيارات الطبيب	Ofp	التابع
number of hospital stays	عدد الاقامة في المستشفى	Hosp	مستقل
self-perceived health status	الحالة الصحية المتصورة ذاتيا	Health	مستقل
number of chronic conditions	عدد الحالات المزمنة	Numchron	مستقل
Gender	الجنس	Gender	مستقل
number of years of education	عدد سنوات التعليم	School	مستقل
Private insurance indicator	مؤشر التأمين	Privins	مستقل

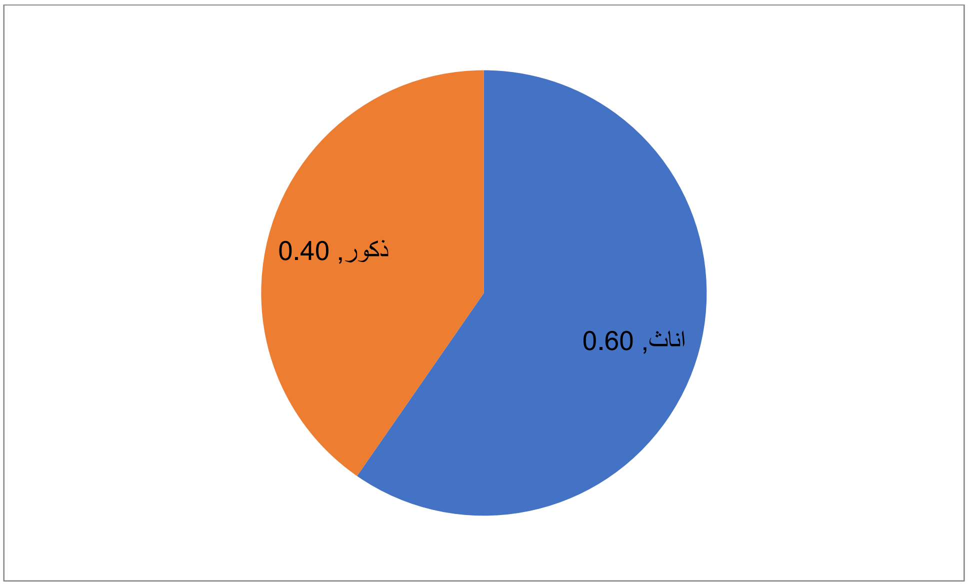
4.3 الإحصاء الوصفي

1- الجنس: يبين الجدول (3.3) توزيع أفراد عينة الدراسة حسب الجنس حيث نلاحظ أن أعلي نسبة كانت 60% وهي للإناث أما أقل نسبة كانت للذكور وقد بلغت 40%.

جدول (3.3) يوضح توزيع أفراد عينة الدراسة حسب الجنس

الجنس	العدد	النسبة
اناث	2628	0.60
ذكور	1778	0.40
المجموع	4406	100

كما تم استخدام القطاع الدائري لتوضيح الجنس كما هو مبين في الشكل (1.3)



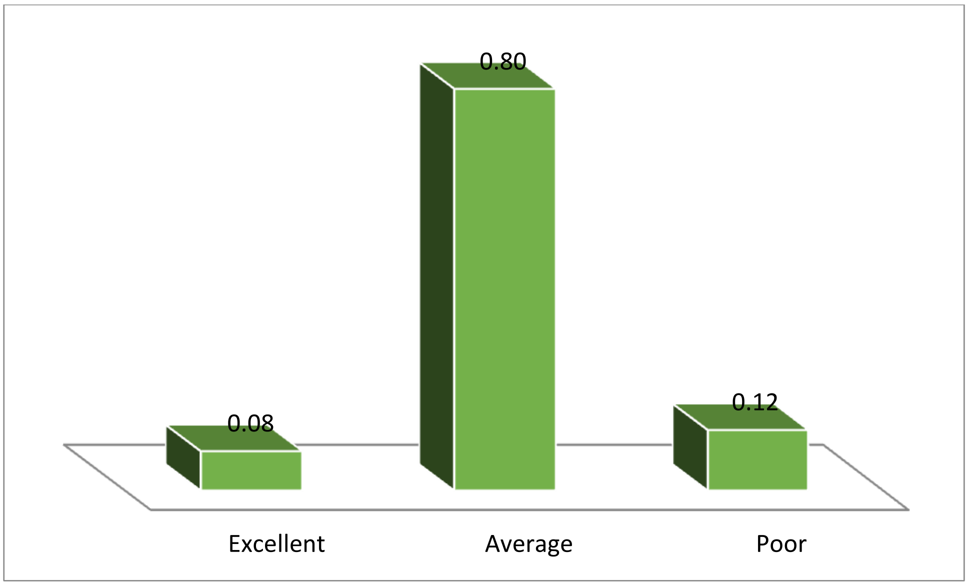
الشكل (1.3) قطاع دائري توضح النسب المئوية لعينة الدراسة حسب الجنس

2- **الحالة الصحية:** يبين الجدول (4.3) توزيع أفراد عينة الدراسة حسب **الحالة الصحية** حيث نلاحظ أن أعلي نسبة كانت 80% هي لفئة المستوي الثاني (المتوسط) بينما كانت أقل نسبة بلغت (8%) وهي للفئة الأخيرة وهي لفئة المستوي الممتاز(Excellent).

جدول (4.3) يوضح توزيع أفراد عينة الدراسة حسب الحالة الصحية.

Health الحالة الصحية	العدد	النسبة
ضعيفة	554	0.12
متوسط	3509	0.80
ممتازة	343	0.08
المجموع	4406	100

كما تم استخدام الأعمدة البيانية لتوضيح عينة الدراسة حسب الحالة الصحية كما هو مبين في الشكل (2.3)



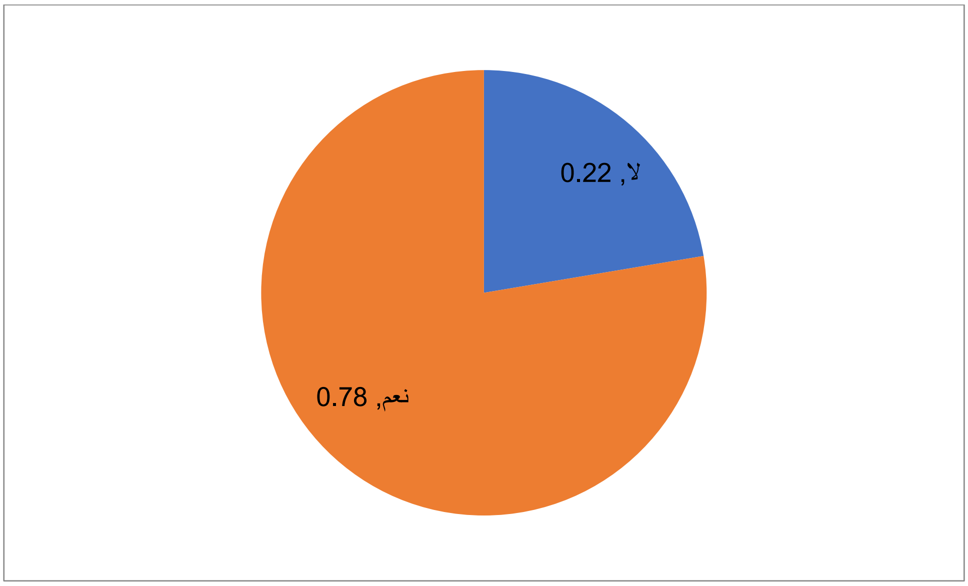
الشكل (2.3) أعمدة بيانية توضح النسب المئوية لعينة الدراسة حسب الحالة الصحية

3- **مؤشر التأمين:** يبين الجدول (5.3) توزيع أفراد عينة الدراسة حسب **مؤشر التأمين** حيث نلاحظ أن أقل نسبة هي للفئة الثانية (نعم لديها تأمين) وقد بلغت 9.1% وان اعلي نسبة هي 43.9% وهي للفئة الأولي (لا تستخدم التامين).

جدول (5.3) يوضح توزيع أفراد عينة الدراسة حسب مؤشر التأمين

Privins مؤشر التأمين	العدد	النسبة
لا	985	0.22
نعم	3421	0.78
المجموع	4406	100

كما تم استخدام قطاع دائري لتوضيح عينة الدراسة حسب مؤشر التأمين **كما هو مبين في الشكل (3.3)**



الشكل (3.3) قطاع دائري يوضح النسب المئوية لعينة الدراسة حسب مؤشر التأمين

جدول (6.3) يوضح المقاييس الإحصائية لبعض المتغيرات

School عدد سنوات التعليم	Numchron عدد الحالات المزمنة	Hosp عدد الاقامة في المستشفى	Ofp عدد الزيارات	الاحصاءة
0	0	0	0	القيمة الصغرى
18	8	8	89	القيمة العظمي
10.19	1.54	0.29	5.77	المتوسط
11	1	0	4	الوسيط

نلاحظ من خلال الجدول أعلاه لمتغير عدد الزيارات تراوحت بين (0،89) حيث بلغ متوسط عدد الزيارات خمسة زيارات. بينما تساوت عدد الإقامة في المستشفى وعدد الحالات المزمنة حيث بلغت 8 حالات. أما متغير عدد سنوات التعليم فقد تراوحت بين (0،18) حين كان متوسط سنوات التعليم لعينة الدراسة عشرة سنوات.

5.3 تطبيق نموذج انحدار بواسون باستخدام برنامج R

يمكن استخدام برنامج R في تقدير **نموذج انحدار بواسون** وغيرها من التوزيعات الأخرى, ولكن قبل أن نبدأ بهذه التطبيقات علينا أولاً أن نوضح الشكل العام المستخدم في تقدير معالم النموذج في برنامج R ثم نقوم بالتطبيق علي البيانات السابقة للدراسة وتلخيص النتائج المتحصل عليها من خلال هذه التطبيقات, والجدول التالي يوضح الشكل العام للدالة المستخدمة في تقدير معالم نموذج انحدار بواسون باستخدام برنامج R.

جدول (7.3) **يوضح الشكل العام** للدالة المستخدمة في تقدير معالم نموذج انحدار بواسون **باستخدام برنامج R**

الشكل العام	fm_pois <- glm(ofp ~ ., data = dt, family = poisson)
الرموز	**الوظيفة التي تقوم بها**
Glm	اسم الدالة المستخدمة لتقدير معالم النموذج
Ofp	اسم المتغير التابع
~ .	العلامة لوضع كل المتغيرات المستقلة الموجودة في البيانات أو كتابة كل متغير مستقل والفصل بينهما بعلامة +
Data	كتابة اسم ملف البيانات
Family	كتابة اسم التوزيع المراد تقدير معالم وهنا توزيع بواسون أما اذا أردنا تقدير معالم توزيع اخر مثل توزيع الهندسي السالب بدلا من توزيع بواسون

والان سنقدر معالم نموذج انحدار بواسون باستخدام برنامج R عن طريق تطبيق الصيغة التالية:

fm_pois <- glm(ofp ~ ., data = dt, family = poisson)

ثم كتابة الأمر summary(fm_pois) فنحصل علي النتائج التالية:

جدول (8.3) يوضح المعالم المقرة والأخطاء المعيارية واختبار معنوية المتغيرات المستقلة

المتغيرات	المعامل	الخطأ المعياري	احصاءة الاختبار	القيمة الاحتمالية	القرار
الثابت	1.02	0.02	43.25	2e-16	معنوي
Hosp (H)	0.16	0.01	27.47	2e-16	معنوي
Healthpoor(HP)	0.24	0.02	13.91	2e-16	معنوي
Healthexcellent(HE)	0.36-	0.03	11.94-	2e-16	معنوي
Numchron(N)	0.14	0.04	32.02	2e-16	معنوي
Gendermale(G)	0.11-	0.01	8.67-	2e-16	معنوي
School(S)	0.02	0.01	14.18	2e-16	معنوي
Privinsyes(P)	0.20	0.02	11.96	2e-16	معنوي

وبالتالي يمكن كتابة النموذج المقدر علي الصورة التالية:

$$ofp = 1.02 + 0.16H + 0.24HP - 0.36HE + 0.14N - 0.11G + 0.02S + 0.2P$$

ولاختبار معنوية المتغيرات المستقلة للنموذج المقدر نلاحظ أن قيم احصائي الاختبار للمتغيرات الصحية كانت كبيرة بينما قيم احصائي الاختبار للمتغيرات الاجتماعية والاقتصادية صغيرة وهذا قد يكون بسبب وجود قيم متطرفة كثيرة في المتغيرات الاجتماعية والاقتصادية التي أثرت علي

قيم اختبار وولد (Wald). ونتيجة لوجود التشتت في مجموعة البيانات هذه نقترح إعادة حساب اختبارات Wald باستخدام الأخطاء القياسية (sandwich standard errors).

وذلك بتطبيق الامر التالي:

```
coeftest(fm_pois, vcov = sandwich)
```

حيث تم الحصول علي النتائج التالية المدرجة في جدول (9.3)

جدول (9.3) يوضح المعالم المقرة والأخطاء المعيارية واختبار معنوية المتغيرات المستقلة

المتغيرات	المعامل	الخطأ المعياري	احصاءة الاختبار	القيمة الاحتمالية	القرار
الثابت	1.02	0.06	15.94	2e-16	معنوي
Hosp (H)	0.16	0.02	7.50	5.93e-14	معنوي
Healthpoor(HP)	0.24	0.05	4.59	4.9e-06	معنوي
Healthexcellent(HE)	0.36-	0.07	4.67-	2.95e-06	معنوي
Numchron(N)	0.14	0.01	11.36	2.2e-16	معنوي
Gendermale(G)	0.11-	0.03	3.17-	0.001	معنوي
School(S)	0.02	0.01	5.14	2.71e-7	معنوي
Privinsyes(P)	0.20	0.04	4.67	2.91e-06	معنوي

ومن خلال النتائج المبينة في الجدول أعلاه تبين أن المتغيرات المستقلة مازالت معنوية ولكن يبدو أن الأخطاء القياسية أكثر ملاءمة وهذا يدل أن هذا الاختبار يفضل استخدامه بدلا من وولد عندما يكون هناك تشتت في البيانات تحت الدراسة.

الفصل الرابع: الاستنتاجات والتوصيات

4.1 المقدمة

هدفت هذه الدراسة إلي التركيز علي نموذج انحدار بواسون حيث تم توضيح المفاهيم الأساسية لتوزيع بواسون وخصائصه الرياضية بالإضافة الي التعريف بنموذج انحدار بواسون من حيث شكله الرياضي والافتراضات الأساسية له في الفصل الثاني. أما الجانب التطبيق تم شرحه في الفصل الثالث مع تطبيق عملي باستخدام برنامج SPSS وتوضيح الخطوات اللازمة لتشغيل البرنامج في إيجاد نموذج انحدار بواسون وكيفية إعطاء الأوامر المختلفة والتعامل مع الأزرار المناسبة لمربعات الحوار في الفصل الثالث. في هذا الفصل سنتناول الاستنتاجات والتوصيات:

4.2 الاستنتاجات: Conclusion هناك عدة استنتاجات يمكن تلخيصها في النقاط التالية:

1. لاحظنا أن توزيع أفراد عينة الدراسة حسب الجنس كانت للإناث حيث بلغت النسبة 60% وهي بينما نسبة للذكور هي الأقل وقد بلغت 40%.
2. نلاحظ أن أعلي نسبة لتوزيع أفراد عينة الدراسة حسب الحالة الصحية كانت 80% هي لفئة المستوي الثاني (المتوسط) بينما أقل نسبة قد بلغت (8%) وهي للفئة الأخيرة وهي لفئة المستوي الممتاز.
3. نلاحظ أن أقل نسبة لتوزيع أفراد عينة الدراسة حسب التأمين قد بلغت 9.1% للذين يستخدمون التأمين وان اعلي نسبة هي 43.9% وهي للفئة الأولي التي لا تستخدم التأمين.
4. نلاحظ أن عدد الزيارات تراوحت بين (0،89) حيث بلغ متوسط عدد الزيارات خمسة زيارات. بينما تساوت عدد الإقامة في المستشفى وعدد الحالات المزمنة حيث بلغت 8 حالات. أما عدد سنوات التعليم فقد تراوحت بين (0،18) حين كان متوسط سنوات التعليم لعينة الدراسة عشرة سنوات.
5. نلاحظ أن قيم احصائي الاختبار للمتغيرات الصحية كانت كبيرة بينما قيم احصائي الاختبار للمتغيرات الاجتماعية والاقتصادية صغيرة وهذا قد يكون بسبب وجود قيم متطرفة كثيرة في المتغيرات الاجتماعية والاقتصادية التي أثرت علي قيم اختبار وولد (Wald).
6. نتيجة لوجود التشتت في مجموعة البيانات تم إعادة حساب اختبارات Wald باستخدام الأخطاء القياسية (sandwich standard errors).

4.3 التوصيات Recommendation: اعتمادا علي ما توصل إليه الباحث من استنتاجات نعرض بعض التوصيات علي النحو التالي:

1. استخدام نموذج انحدار بواسون لتحليل البيانات التي تكون بهيئة بيانات معدودة (Count Data) لمتغير الاستجابة وأن هذه البيانات هي أعداد صحيحة غير سالبة
2. عند استخدام انحدار بواسون لتحليل بيانات ظاهرة معينة يجب التحقق من الافتراضات الخاصة بهذا النموذج من أجل الحصول علي نتائج صحيحة يمكن الاعتماد عليها.
3. يجب الانتباه الي توزيع الأخطاء لأنه في انحدار بواسون نفترض أن الأخطاء تتبع توزيع بواسون بينما في الانحدار يشترط تبعية الأخطاء للتوزيع الطبيعي
4. في انحدار بواسون يشترط ان تكون البيانات دائما اعداد صحيحة موجبة بينما في الانحدار الكلاسيكي قد تكون موجبة ام سالبة او كسرية.
5. عدم استخدام طريقة المربعات الصغري الاعتيادية (OLS) في تقدير معالم نموذج انحدار بواسون حيث ترتكز هذه الطريقة OLS على افتراضات معينة منها أن قيم المتغير التابع هو قيم مستمرة وموزعة بشكل طبيعي ومرتبطة خطيًا بالمتغيرات المستقلة على وجه الخصوص وهذه الافتراضات في كثير من الأحيان قد لا تتحقق مع بعض البيانات مثل انحدار بواسون الذي يفترض أن بيانات المتغير التابع أعداد صحيحة غير سالبة (بيانات غير مستمرة) وأن الأخطاء تتوزع وفقا لتوزيع بواسون (غير طبيعي) وبالتالي لا تعمل طريقة المربعات الصغرى الاعتيادية بصورة جيدة وأن المقدرات التي يتم ايجادها غير كفؤة وبالتالي لا يمكن الاعتماد عليها في عمليات التنبؤ.
6. عند وجود قيم متطرفة في مجموعة البيانات يفضل استخدام اختبار Wald باستخدام الأخطاء القياسية (sandwich standard errors).

4.4 الدراسات المستقبلية Futures work

1. دراسة مقارنة طرائق تقدير معلمات نموذج انحدار بواسون
2. دراسة نموذج انحدار بواسون المتعدد
3. دراسة نموذج انحدار بواسون عندما تعاني البيانات من مشكلة التعدد الخطي
4. دراسة نموذج انحدار بواسون المتعدد عندما تعاني البيانات من القيم المتطرفة
5. دراسة نموذج انحدار بواسون اللامعلمي والشبه اللامعلمي
6. دراسة انحدار هيرميت Hermite regression .

References :المراجع

1. علي عبدالسلام العماري + علي حسين العجيلي (2010) الإحصاء والاحتمالات النظرية والتطبيق ، دار الحكمة.
2. أمير حنا هرمز (1990) الإحصاء الرياضي دار الكتاب للطباعة والنشر.
3. أحمد عودة (1991) مقدمة في النظرية الإحصائية مطابع جامعة الملك سعود.
4. لمياء محمد علي + ايثار حسين (2017) مقارنة بين طريقة بيز وطريقة الإمكان الأعظم الكاملة لتقدير نموذج انحدار بواسون الهرمي وتطبيقها على وفيات الأمهات في بغداد.
5. تحليل الانحدار الخطي د. محمد عبدالرحمن إسماعيل مركز البحوث.
6. جمال إبراهيم + سمير سليم (2004) تحليل الارتباط ونماذج الانحدار البسيط منشورات جامعة السابع من ابريل.
7. A Comparison of Regression Models for Count Data in Third Party Automobile Insurance
8. , (2014)Annelie Johansson.
9. Esin Avcı1*, Sibel Alturk2, & Elif Neyran Soylu3 (2015), Comparison Count Regression Models for Over dispersed Alga Data.