



جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية علوم الحاسوب وتكنولوجيا المعلومات
قسم الإحصاء والمعلوماتية

المقارنة بين الطرق الكلاسيكية وطرق بايز

في تقدير معلومات التوزيع الطبيعي

بمحة مقدم

الى

كلية علوم الحاسوب وتكنولوجيا المعلومات / جامعة القادسية
وهو جزء من متطلبات نيل درجة البكالوريوس علوم في الإحصاء والمعلوماتية

من قبل الطالب

محمد عباس فاضل

إشراف

أ.م. جبران عبد الامير خطار

1438هـ 2017 م

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

لِيَرْفَعَ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا
الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ

صدق الله العلي العظيم

سورة المجادلة-11

الاهداء

- امتنان وشكر الى الله سبحانه وتعالى والى النبي محمد (ص) والى
بيته (صلوات الله عليهم اجمعين)
- الى منيع الاحترام والتقدير والدي العزيز.
- الى منيع الحب والحنان والدي العزيزة.
- الى الذين مهدوا لنا طريق العلم والمعرفة اساتذتي الاعزاء
- الى منيع الاخاء اخوتي وأخواتي
- الى جميع أصدقائي الاوفياء

الباحث

محمد عباس فاضل

شكر وتقدير

- أنه لمن دواعي الفخر والاعتزاز والتقدير أن اتقدم بحميم الشكر الجزيل والتقدير الفائق لأستاذي المشرف على البحث الاستاذ جبران عبد الأمير خطار لما قدمه لي من توجيهات وإرشادات كانت لها الأثر في تكامل البحث.

- كما اتقدم بالشكر والتقدير للأساتذة الأفاضل في قسم الإحصاء والمعلوماتية ولا استثنى أحداً منهم والشكر موصول إلى أعضاء لجنة المناقشة.
- والله ولي التوفيق

الباحث

محمد عباس فاضل

الفصل الاول

1- مقدمة البحث

نظرية التخمين

لنظرية التخمين اهمية كبيرة في تطبيقات النظرية الاحصائية في الجوانب العلمية . وبشكل عام يمكن النظر الى نظرية التخمين على انها جزئين متكاملين الاول يهتم بالبحث على افضل مخمن لمعلمة مجهولة تخضع لمجتمع معين وهنا غالبا ما يسمى التضمين بنقطة (Point estimation) في حين الجزء الثاني يهتم بالبحث عن افضل فترة يمكن حساب قيمة المعلمة المجهولة خلالها وهذا غالبا ما يسمى بالتخمين بالفترة (Interval estimation) المؤثر الاحصائي الذي تستخدم قيمته للتخمين $T(\theta)$ حيث ان $\theta(t)$ يسمى بالمخمن او المقدّر (estimator)

هناك عدة طرق لإيجاد المخمنات منها :

- 1- طريقة العزوم
- 2- طريقة الترجيح الاعظم
- 3- طريقة التباين الاقل
- 4- طريقة المربعات الصغرى
- 5- طرق بايز.

في بحثنا هذا سنتناول طريقتين فقط وهي طريقتين العزوم وطريقة المربعات الصغرى فقط كي نقارنهما مع طرق بايز في

البحث عن افضل مخمن في تخمين معلمات التوزيع الطبيعي
 $M(6^2)$.

2- الجانب النظري

يتضمن الفصل الاول التوزيع الطبيعي الذي سوف نقوم
بتخمين معلماته (M, σ^2) .

ان التوزيع الطبيعي هو من الادوات الكثيرة الاستخدام في
التحليل الاحصائية وفي المجالات الصناعية والطبية ويستخدم
في دراسة البواقي في تحليل الانحدار ويمكن تعريف منحني
التوزيع الطبيعي بقيمتين .

المتوسط M الانحراف المعياري σ^2

ويعتبر التوزيع الطبيعي من التوزيعات المستمرة المتصلة .

نفرض لدينا X يتوزع توزيعا طبيعيا $(X \sim N(M, \sigma^2))$

$$f_x = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - M)^2}{2\sigma^2}\right)$$

حيث ان :

M هي القيمة المتوقعة

σ^2 هو التباين

ملاحظة عندما $M = 0$ يسمى توزيع طبيعي معياري

$$\sigma^2 = 1$$

خصائص التوزيع الطبيعي

هذا التوزيع من أكثر التوزيعات الاحتمالية استخداماً ، بل يشق منه كل التوزيعات الاحتمالية الأخرى المستخدمة في الاستدلال الإحصائي .

ومن خصائص هذا التوزيع ما يلي :-

1- الوسط الحسابي

2- والتباين σ^2

3- منحني هذا التوزيع متماثل على جانبي الوسط M

4- كيفية حساب الاحتمالات $P(x_1 < x < x_2)$

5- التوقع التوزيعي الطبيعي $E(x) = M$

6- التباين التوزيعي الطبيعي $Var(x) = \sigma^2$

لحساب احتمالات التوزيع الطبيعي يحدد بالمساحة التالية أسفل المنحني

وتحسب هذه المساحة (الاحتمال) بإيجاد التكامل التالي :

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x) dx$$

$$= \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

وهذا التكامل يصعب حسابه حيث لجأ الإحصائيون إلى عمل تحويله Transform ، يسهل استخدام توزيعها الاحتمالي في حساب مثل هذه الاحتمالات وهذه التحويلة هي :

$$z = \left(\frac{x - \mu}{\sigma} \right)$$

ويعرف المتغير الجديد z بالمتغير الطبيعي المقياسي

Standard Normal Variable وهذا المتغير له دالة كثافة احتمال بالصورة التالية :

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty, \pi = 22/7$$

ومن خصائص هذا التوزيع مايلي :

1- متوسطه هو : $E(z) = 0$

2- تباينه هو : $\text{Var}(z)=1$

ومن ثم يعبر عن توزيع المتغير z بالرموز $Z \sim N(0,1)$

طرق تخمين معلمات التوزيع الطبيعي

1- طريقة العزوم *Method of moment*

تعتبر هذه الطريقة من اقدم الطرق لايجاد المخرنات اوجدها العالم بيرسون .

نفرض ان $[Xi], i = 1, 2, \dots, n$ يكون عشوائي لديه P.d.f

$f(x, \theta)$ عندما المعلمة $\theta = [\theta_1, \theta_2, \theta_k]$ معلومة .

هدفنا هو تقدير المعلمة θ مفهوم طريقة العزوم يعتمد على فكرة التقدير عن طريق عينات بسيطة .

للاعتماء على عءء من المعلمات هناك نظام من المعاءلات

$$\hat{m}_j = m_j(\theta_1, \theta_2, \theta_B, \dots \dots \theta_k), j = 1, 2, \dots k$$

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^n X_i^j \quad \text{عندما}$$

وهذا يسمى بالعزم الرائي حول الصفر

$$\bar{X}_i^j = \frac{1}{n} \sum_{i=1}^n X_i^j \quad \text{بما ان}$$

$$m_j = E \bar{X}_i^j \quad \text{اذن}$$

2- طريقة المربعات الصغرى *OLS*

حيث ان قيمة y الناتجة من معادلة خط الانحدار $y_i = ax_j + b$ هي قيمة تقديرية ، وبفرض ان القيمة الفعلية المقابلة لـ $\hat{y}_j$ هي y_j فيكون هناك فرق بين القيمة الفعلية والقيمة المقدرة وهذا ما

يسمى بالخطأ أو الانحراف (e) ، وقد يكون هذا الانحراف سالب
او موجب اي ان

$$e = y_i - \hat{y}_i$$

وتتلخص طريقة المربعات الصغرى في تحديد
الثوابت a, b والتي تجعل مجموع مربعات الانحراف اقل ما
يمكن اي ايجاد a, b التي تجعل الكمية :

$$\begin{aligned} d = \sum e^2 &= \sum_{i=1}^n (y_i - (ax_i + b))^2 \\ &= \sum (ax_i + b - y_i)^2 \end{aligned}$$

يتفاضل d جزئياً بالنسبة الى a ومساواة الناتج بالصفر نحصل على :-

$$\frac{\partial}{\partial a} \sum^n (ax_i + b - y_i)^2 = 0$$

$$\therefore 2a(\sum^n x_i)((\sum^n x_i) + 2bn((\sum^n x_i) - z(\sum^n x_i)_{i=1}(\sum^n x_i)_{i=1} \\ = 0$$

$$\therefore b_u + a \sum_{i=1}^n X_i = \sum_{i=1}^n y_i \quad (1)$$

بالمثل يتفاضل d جزئياً بالنسبة الى b و مساواة الناتج بالصفر ونحصل على

$$\therefore b_u + a \sum_{i=1}^n X_i + a \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad (2)$$

بحل المعادلتين (1) ، (2) نجد ان

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - [\sum_{i=1}^n x_i]^2} \quad (3)$$

$$\left[\frac{i = 1 \sum^n}{n} \right] - a \left[\frac{\sum_{i=1}^n x_i}{m} \right] \quad (4)$$

بالمثل يمكن تحديد معاملات $(\hat{d}, \hat{b})$ خط انحدار x على y كما يلي :

$$n \sum x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

$$a^1 \frac{i = 1}{n_{i=1} \sum y_i^2 - [\sum_{i=1}^n y_i]^2} \quad (5)$$

$$b^1 = \left[\frac{\sum_{i=1}^n x_i}{n} \right] - \left[\frac{\sum_{i=1}^n y_i}{n} \right] \quad (6)$$

وتهدف طريقة المربعات الصغرى الى ايجاد معادلة خطية تقديرية

- المقدار e_i يخضع لقانون التوزيع الطبيعي بوسط $m=0$ وبتباين .. ويمكن ان ياخذ قيمة سالبة او معدومة .

- الوسط الحسابي لقيم e_i توقعه يكون معدوماً $E(e_i) = 0$ وهكذا يعني ان مجموع القيم العشوائية التي هي اكبر من القيم الحقيقية تتساوى مع تلك التي هي اقل منها او انها متنوعة بشكل متناظر .

- ان حد الخطأ يكون نفس التباين

$$var(e_i) = E(e_i) = \sigma^2 i$$

$$e_i \rightarrow N(0, \sigma^2 i)$$

$$E(e_i, e_j) = 0$$

حيث ان $i \neq j$

حيث ان حد الخطأ e_i يكون مستقلاً عن المتغير المستقل X_i عند كل مشاهدة

$$(0 \ v(e_i x_i) = E(e_i x_i) = 0$$

حجم العينة مهما اختلفت تكون القيمة كالاتي :

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

حيث ان التوقع الرياضي للمتغير يكون كالآتي

$$E(y_i) = E(a + \beta x_i + e_i) \quad \therefore \text{توقع حد الخطأ يكون معدوم}$$

$$\therefore a\beta x_i \text{ هي معلمه المتغير } x$$

$$\text{Var}(y_i) = E(y_i - (y_i))^2$$

$$= E[a + \beta x_i + e_i - a\beta x_i]^2 = E(e_i)^2 = \sigma^2$$

$$\text{Var}(y_i) = E(e_i)^2 = \sigma^2$$

3- مفهوم نظرية بايز

يركز اسلوب بايز في التقدير بمفهومه بشكل عام على استخدام معلومات مسابقة حول المعالم المجهولة $\theta = (\theta_1, \theta_2, \dots, \theta_{\frac{p}{k}})$ المطلوب تقديرها على اعتبار ان هذه المعالم متغيرات عشوائية و ليست كميات ثابتة ويضاف على تلك المعلومات المسابقة معلومات العينة المشاهدة ، حيث تمثل تلك المعلومات على شكل دالة احتمالية (*Prior p.d.f*) ويمكن ان نعرفها بأنها الدالة التي تمثل كل المعلومات والخيرات حول المعالم المراد تقديرها و التي تم التوصل اليها سابقا من خلال التحليل او المراقبة لتلك المعالم .

اما دالة التوزيع الحالي لمشاهدات العينة الحالية (y) تعتمد على المعالم (θ) ويرمز لها بالرمز $(f(y)(\theta))$.

ان مقدار بيز لأي معلمه يعتمد على دالتين الاولى تعرف بدالة الكثافة الاحتمالية اللاحقة (*p.d.f*) والثانية دالة الخسارة (*lossfunction*) فدالة الكثافة الاحتمالية اللاحقة يمكن تعريفها

بأنها دالة تمثل كل المعلومات حول المعالم المراد تقديرها بعد مشاهدتنا لمعلومات العينة الحالية وبعبارة أخرى أنها تركيب بين المعلومات الأولية وبيانات العينة الحالية حيث يمكن التعبير رياضياً عما تقدم كالآتي :

$$f(y, \theta) = f(y|\theta) \cdot g(\theta)$$

$$f(\theta|y) = \frac{f(y|\theta) \cdot g(\theta)}{f(y)}$$

$$f(\sigma|y) \propto f(\theta) \cdot f(y|\theta) \dots \dots \dots (1)$$

حيث أن :

$\propto$ تمثل الكمية التناسبية

أما دالة الخسارة يمكن تعريفها بأنها الدالة التي تحقق الشرطين الآتيين معاً :

$$1. L(\hat{\theta}, \theta) \geq 0 \quad \forall \hat{\theta}, \forall \theta$$

$$2. L(\hat{\theta}, \theta) = 0 \quad \forall \hat{\theta} = \theta$$

لذلك مقدر بيز يعتمد على إيجاد قيم $(\hat{\theta})$ التي تقلل توقع الخسارة

أي أن :

$$\text{Min } \hat{\theta} E[L(\hat{\theta}, \theta) = |y] = \text{Min } \hat{\theta} \int L(\hat{\theta}, \theta) f(\theta|y) d\theta$$

وتسمى دالة الخسارة المستخدمة في عملية التقدير النقطي بدالة خسارة مربع الخطأ والتي تكتب بالشكل الآتي :

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 \dots \dots \dots (2)$$

في حين تسمى بدالة خسارة مربع الخطأ الموزون في حالة وجود متجه من المعالم ويمكن كتابتها بالشكل الآتي :

$$L(\hat{\theta}, \theta) = (\hat{\theta}, \theta) Q L(\hat{\theta}, \theta) \dots \dots \dots (3)$$

حيث أن :

Q: مصفوفة موجبة محددة غير عشوائية و متماثلة .

عندما تكون Q مصفوفة قطرية فيطلق على دالة الخسارة تسمية دالة الخسارة التربيعية والتي تكتب كالآتي :

$$L(\hat{\theta}, \theta) = \sum_{i=1}^e q_i (\hat{\theta}_i, \theta_i)^2 \dots \dots \dots (4)$$

أن مقدار بيز النقطة يعتمد دالة خسارة مربع الخطأ بمثل الوسط الحسابي لدالة الكثافة الاحتمالية اللاحقة $f(\hat{\theta}, y)$ كما موضح أدناه

$$\begin{aligned} E(L(\hat{\theta}, \theta)|y) &= E(L(\hat{\theta}, \theta)|y)^2 \\ &= \int (\hat{\theta}, \theta)^2 f(\theta|y) e|\theta \\ &= \hat{\theta}^2 \int f(\theta|y) d\theta \\ &\quad - 2 \hat{\theta}^2 \int \theta f(\theta|y) d\theta + \int \hat{\theta}^2 f(\theta|y) d\theta \dots \dots \dots (5) \end{aligned}$$

وبأخذ المشتقة الأولى للمعادلة (5) بالنسبة إلى $(\hat{\theta})$ ومساواتها للصفر

لغرض الحصول على اقل توقع خسارة يكون لدينا كالاتي :

$$\frac{dE[L(\hat{\theta}, \theta)|y]}{d\hat{\theta}} = 0$$

$$2\hat{\theta} \int f((\theta) | y) d\theta - 2 \int \theta f(\theta | y) d\theta = 0$$

$$(\hat{\theta}(1) - E(\theta|y)) = 0$$

$$\hat{\theta} = E((\theta|y))$$

و بأخذ المشتقة الثانية لمعادلة (5) بالنسبة الى $\hat{\theta}$ نلاحظ انها قيمة موجبة مما يدل على ان المعادلة (5) في نهايتها الصغرى .

في حين مقدر بايز لمتجه من المعالم فيمثل الوسط الحسابي للدالة الاحتمالية اللاحقة .

ان استخدام اسلوب بايز في التقدير يقتضي وجود دالة كثافة احتمالية اولية لذلك من المهم التشخيص لدالة اولاً قبل اجراء عملية التقدير باسلوب بايز .

الفصل الثاني

تخمين معلمات التوزيع الطبيعي

1- التخمين بطريقة العزوم MLE

$$\therefore \bar{X} = \sum \frac{x_i}{n}, \bar{M}_l = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{M}_l = E(x) = M$$

$$\therefore \bar{M}_l = \bar{m}_l$$

$$\therefore \hat{M} = \bar{X}$$

$$E(X^2) = Var(x) + [E(x)]^2$$

$$= \sigma^2 + M^2$$

$$\frac{1}{n} \sum_{i=1}^n x_i^2 = \sigma^2 + M^2$$

$$\therefore \sigma^2 = \frac{1}{n} \sum x_i^2 - \hat{M}$$

$$\frac{1}{n} \sum x_i^2 - \bar{X}$$

2- التخمين بطريقة المربعات الصغرى OLS

اولا : توضيح الهدف من الطريقة

تهدف طريقة المربعات الصغرى كما سبقت الإشارة الى ايجاد معادلة خطية تقديرية ، يكون خطها اقرب ما يمكن الى جميع نقاط شكل الانتشار .

$$y_i = a + \beta x_i + \varepsilon_i \dots \dots \dots (1)$$

حيث ان الشق الاول من المعادلة وهو : $a + \beta x_i$ ، يعرف بالمتغير التوضيحي او المفسر ، بينما الجزء الثاني وهو ε_i فيعبر عن المتغيرات غير الموضحة او غير المفسرة ، والذي هو عبارة عن انحراف القيم التقديرية عن القيم الحقيقية للمتغير التابع ، ويمكن للشكل ان يوضح ذلك :

حيث عبرنا عن معادلة المستقيم التي يمكن ان تمثل المعادلة الحقيقية تقريبا والتي نسميها بالمعادلة المقدرة على الشكل :

$$\hat{y}_i = \hat{a} + \hat{\beta} x_i$$

كما هو واضح في الرسم اعلاه ، فإن معادلة المستقيم المفترض ان يكون اقرب ما يمكن الى جميع نقاط شكل الانتشار هي :

$$\hat{y}_i = \hat{\alpha} + \beta x_i \dots \dots \dots (2)$$

بينما المعادلة الحقيقية والتي يستحيل ايجادها مادامت نقاط شكل الانتشار ليست على استقامة تامة هي :

$$y_i = a + \beta x_i + \varepsilon_i$$

حيث ε_i هي قيمة عشوائية تعادل قيمة الانحرافات عن المنحني الخطي للدالة ، وهي قيمة غير قابلة للقياس وتسمى بالمتغير العشوائي او حد الخطأ .

ثانيا : الفرضيات الاساسية

لتقدير معالم النموذج بطريقة المربعات الصغرى يفترض ما يلي :

فرضيات متعلقة بحد الخطأ

1- المقدار ε_i يخضع لقانون التوزيع الطبيعي ، اي ان قيمها تكون خاضعة للصدفة في اي فترة زمنية ، ويمكنها ان تاخذ ايه قيمة موجبة او سالبة او معدومة .

2- الوسط الحسابي لقيم ε_i اي توقعه يكون معدوما : $E(\varepsilon_i) = 0$ ، وهذا يعني ان مجموع القيم العشوائية التي هي اكبر من القيمة الحقيقية تتساوى مع تلك التي اقل منها ، اي انها متوزعة بشكل متناظر .

3- فرض تجانس تباين حد الخطأ ، اي ان تباينه حول وسطه يكون ثابتا بالنسبة لجميع مشاهدات قيم المتغير المستقل ، وهذا يتضمن ان احد الخطأ يكون نفس التباين ، ويعبر عن ذلك كما يلي :

$$Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma_i^2$$

ويعبر عن الفرضيات الثلاث على النحو :

$$\varepsilon_i \rightarrow N(0, \sigma^2_i)$$

4- انعدام الارتباط الذاتي بين حدود الاخطاء ، اي ان حد الخطأ لمشاهدة ما خلال فترة ما لا يرتبط بحد الخطأ لمشاهدة اخرى في فترة اخرى ، ويعبر عن ذلك على النحو :

$$E(\varepsilon_i \varepsilon_j) = 0$$

حيث : $j \neq i$

5- قيم حد الخطأ ε_i تكون مستقلة عن المتغير المستقل x_i عند كل مشاهدة ، ويعني ذلك ان تباينهما المشترك يكون معدوما .
اي :

$$Cov(\varepsilon_i x_i) = 0$$

فرضيات متعلقة بالمتغير التابع والمستقل :

6- انتظام قيم المتغير المستقل ، وعدم تغيرها من عينة لاخرى
وينجر عن ذلك انه مهما اختلف حجم العينة تكون القيمة

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}$$

عبارة عن رقم نهائي غير معدوم (وهي تعبر عن تباين المتغير المستقل).

7- ان المتغير التابع يكون دالة خطية في المتغير العشوائي ε_i ،
و بالتالي فان المتغير التابع y_i يكتسب الصفة العشوائية من المتغير ε_i كما يكتسب منه ايضا صفة التوزيع الطبيعي ،
وهذا يجعلنا نكتب التوقع الرياضي للمتغير التابع على الشكل :

$$E y_i = E(\alpha \beta x_i + \varepsilon_i)$$

وبما ان توقع حد الخطأ معدوم يكون :

$$E y_i = \alpha \beta x_i$$

ذلك لان : x_i ، α β عبارة عن ثوابت

ويكون تباين y_i على النحو التالي :

$$\begin{aligned} Var(y_i) &= E[y_i - (y_i)]^2 = E[\alpha + \alpha \beta x_i + \varepsilon_i - \alpha - \beta x_i]^2 \\ &= E(\varepsilon_i)^2 = \sigma^2 \end{aligned}$$

$$Var(y_i) = E(\varepsilon_i)^2 = \sigma^2$$

والنتيجة هي ان y_i تخضع ايضا للقانون الطبيعي بتوقع رياضي هو :

$$E(y_i) = \alpha + \beta x_i$$

وبأنحراف معياري :

$$E(\varepsilon_i)^2 = \sigma^2$$

$$y_i \rightarrow N(\alpha + \beta x_i, \sigma^2_i)$$

ثالثا : اشتقاق معالم خط الانحدار بطريقة المربعات الصغرى

الاثبات :

نفترض ان المعادلة $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$ هي معادلة تقديرية للمعادلة الحقيقة :

$$y_i = a + \beta x_i + \varepsilon_i$$

حيث ان قيمة $\hat{y}_i$ تكافئ بما هو عليه في التوزيع الطبيعي القياسي
قيمة الـ F_i

وان القيمة المقدرة تنحرف عن القيمة الحقيقية بالمقدار e_i حيث :

E_i هي قيمة تقديرية للمقدار ε_i هي الفرق بين القيم الحقيقية y_i
(للمجتمع) والقيم المقدرة $\hat{y}_i$

(للعينة) اي :

$$e_i(y_i - \hat{y}_i) \dots \dots \dots (3)$$

يسمى هذا المقدار بالبواقي .

وعليه يمكن ان نكتب:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

بالنسبة للمجتمع :

اما بالنسبة للعينة :

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i + \varepsilon_i$$

ويكون :

$\hat{\alpha}$: القيمة التقديرية لـ α

$\hat{\beta}$: القيمة التقديرية لـ β

E_i : القيمة التقديرية لحد الخطأ ε_i

وحتى تكون المعادلة التقديرية اقرب ما يمكن الى جميع نقاط شكل
الانتشار ، يجب ان يكون :

$$\sum_{i=1}^n e_i \sum_{i=1}^n (y_i - \hat{y}_i) \Rightarrow 0 \dots \dots \dots (4)$$

اي مجموع البواقي يؤول الى الصفر ، وهكذا يكافئ كذلك مربعات البواقي تؤول الى ادنى قيمة ممكنة ، اي :

$$\sum_{i=1}^n e_i^2 \sum_{i=1}^n (y_i - \hat{y}_i)^2 \Rightarrow \text{ادنى قيمة} \dots \dots \dots (4)$$

اي مجموع البواقي يؤول الى الصفر ، ويكافئ كذلك مربعات البواقي تؤول الى ادنى قيمة ممكنة ، اي :

$$\sum_{i=1}^n e_i^2 \sum_{i=1}^n (y_i - \hat{y}_i)^2 \Rightarrow \text{ادنى قيمة} \dots \dots \dots (5)$$

بتعويض $\hat{y}_i$ المقدرة بقيمتها حسب المعادلة (2) نجد :

$$\sum_{i=1}^n e_i^2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \rightarrow \text{ادنى قيمة} \dots \dots \dots (6)$$

وحتى تاخذ المعادلة رقم (6) قيمتها الدنيا ، يجب ان يتحقق شرطان :

الشرط الاول (اللازم) : المشتقات الجزئية الاولى بالنسبة للمعلمتين $\hat{\alpha}$ ، $\hat{\beta}$ ، تساوي الصفر .

الشرط الثاني (الكافي) : المشتقات الثانية بالنسبة للمعلمتين ، يجب ان تكون اكبر من الصفر .

اذن لايجاد قيم المعالم المقدرة التي تجعل مربعات البواقي في ادنى قيمة لها ، يجب ان نساوي المشتقة الاولى بالنسبة لميل الدالة الى الصفر ، ثم نبحث عن قيمة الميل بدلالة بقية المتغيرات ، ونقوم بنفس الشيء بالنسبة للمعلمة الثانية .

نضع :

$$k = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \dots\dots\dots (6)$$

بالاشتقاق جزئيا بالنسبة لـ: K

$$\frac{\partial k}{\partial \hat{\alpha}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \dots\dots\dots (8)$$

ومنه يكون :

$$\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \dots\dots\dots (9)$$

بفك القوس نجد :

$$\frac{\partial k}{\partial \hat{\alpha}} \sum_{i=1}^n y_i - N\hat{\alpha} - \hat{\beta} \sum_{i=1}^n x_i = 0 \dots\dots\dots (10)$$

بقسمة طرفي المعادلة (10) على المقدار N نجد :

$$\bar{y} - \hat{\alpha} - \hat{\beta}\bar{X} = 0 \dots\dots\dots (11)$$

ومنه يكون :

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \dots \dots \dots (12)$$

حيث ان قيمة الـ $\hat{\alpha}$ تكافئ قيمة $\hat{\sigma}^2$

بأشتقاق K جزئيا بالنسبة للمعلمة الثانية نجد :

$$\frac{\partial k}{\partial \hat{\alpha}} - 2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0 \dots \dots \dots (13)$$

اي :

$$\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0 \dots \dots \dots (14)$$

تسمى المعادلة (10) و (14) بالمعادلتين الطبيعيين او معادلتى المربعات الصغرى الطبيعتين بفك المعادلة (14) نجد :

$$\sum_{i=1}^n y_i x_i - \hat{\alpha} \sum_{i=1}^n x_i - \hat{\beta} \sum_{i=1}^n x_i^2 = 0 \dots \dots \dots (15)$$

وهذا يكافىء :

$$\sum_{i=1}^n y_i x_i - \hat{\beta} \sum_{i=1}^n x_i^2 - \hat{\beta} \sum_{i=1}^n x_i = 0 \dots \dots \dots (16)$$

بتعويض المعادلة (12) في المعادلة (16) نجد :

$$\sum_{i=1}^n y_i x_i - \hat{\beta} \sum_{i=1}^n x_i^2 + (\bar{y} - \hat{\beta} \bar{x}) \sum_{i=1}^n x_i \dots \dots \dots (17)$$

ومنه نجد :

$$\sum_{i=1}^n y_i x_i - \hat{\beta} \sum_{i=1}^n x_i^2 + \bar{y} \sum_{i=1}^n x_i - \hat{\beta} \bar{x} \sum_{i=1}^n x_i \dots \dots \dots (18)$$

بإخراج $\hat{\beta}$ عامل مشترك وتحويل $\bar{y} \sum_{i=1}^n x_i$ الى الطرف الايسر نجد :

$$\sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i = \hat{\beta} \left(\sum_{i=1}^n x_i^2 + -\bar{x} \sum_{i=1}^n x_i \right) \dots \dots \dots (19)$$

واخيرا تكون قيمة $\hat{\beta}$ كما يلي :

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} \dots \dots \dots (20)$$

يمكن استخدام هذه المعادلة في تقدير ميل الدالة ، غير انه من الصعب استخدامها عندما تكون ارقام المتغيرات كبيرة ، لذلك يتم استخدام معادلة اخرى ، تعتمد على الانحرافات عن الوسط الحسابي ، والتي يتم اشتقاقها باتباع المنهجية التالية :

في بسط المعادلة 20 نطرح ونضيف المقدار : $\bar{x} \sum_{i=1}^n y_i$

في مقام المعادلة 20 نطرح ونضيف المقدار : $\bar{y} \sum_{i=1}^n x_i$

فنحصل على العبارة التالية :

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - \bar{x} \sum_{i=1}^n y_i + \bar{x} \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n x_i + \bar{x} \sum_{i=1}^n x_i} \dots \dots \dots (21)$$

الطرف الاخير في البسط هو : $N\bar{x}\bar{y}$ الطرف الاخير في المقام هو $N\bar{x}^2$ بالتعويض في 21 واخراج المجموع نجد :

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x} \bar{y})}{\sum_{i=1}^n (x_i^2 - 2\bar{x} x_i - \bar{x}^2)} \dots \dots \dots (22)$$

لاحظ ان البسط هو جداء المقدارين :

$$\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})$$

والمقام هو :

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

ومنه يمكن كتابة المعادلة 22 على النحو التالي :

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \dots \dots \dots (23)$$

حيث ان $\hat{\beta}$ تكافئ الـ $\hat{M}$

واضح ان المعادلة 23 سهلة الاستخدام ، كما انها تختصر الارقام ،
لكونها تعتمد على الانحرافات عن الوسط الحسابي .

تسمى طريقة ايجاد المعالم المقدرة لدالة الانحدار بهذه الطريقة ،
بطريقة المربعات الصغرى .

يمكن التأكيد بان القيمتين : $\hat{\beta}, \hat{\alpha}$ ، المقدرتين بهذه الطريقة تؤديان
الى نهاية صغرى لمربعات البواقي ، بمراجعة المشتقات الجزئية
الثانية للمعادلة (7) وذلك للمعلمتين ، اذ نجد هذه المشتقات موجبة .

3- التخمين بطريقة بايز

تخمن معلمات التوزيع الطبيعي $N(\mu, \sigma^2)$ حيث تتم طريقة التخمين كالآتي :

1- عندما يكون المتوسط (μ) غير معروف $(-\infty < \mu < \infty)$ وتكون قيمة $(\varepsilon > 0)$ معروفة .

وتفترض ايضا اعطاء معلمتين (μ_p, ε_p) .

علماً ان قيمة $\varepsilon_0 = \frac{1}{\sigma_0^2}$ حيث ان ε_0 هي للتوزيع الاساسي للعينة

$$\varepsilon_1 = \frac{1}{\sigma_1^2}$$

حيث ان ε_1 هي لتوزيع الـ *Postorio*

$$\varepsilon_p = \frac{1}{\sigma_p^2}$$

حيث ان ε_p هي لتوزيع الـ *Prior*

$$\mu_1 = \frac{n\varepsilon_0\bar{x} + \mu_p\varepsilon_p}{n\varepsilon_0 + \varepsilon_p}$$

حيث ان ε_p ، ε_0 هي قيم يتم فرضها باعتبار ان μ غير معلومة و الـ μ_p تمثل متوسط توزيع *prior* للتوزيع الطبيعي (الاساسي)

μ_1 تمثل متوسط توزيع الـ *(posterior)* للتوزيع الطبيعي (الاساسي)

حيث يمكن كتابة *(posterior)* لتوزيع (μ) كالآتي :

$$f(M|x) \propto f_n(M|x) x P(\mu)$$

$$= \left(\frac{\varepsilon_0}{2\pi}\right)^{n/2} \prod_{i=1}^n \exp\left(\frac{\varepsilon_0}{2}(x_i - M)^2\right) \times \left(\frac{\varepsilon_p}{2\pi}\right) \exp\left(\frac{\varepsilon_p}{2}(\mu - \mu_p)^2\right)$$

$$\propto \exp\left(-\frac{1}{2}\left[\varepsilon_0 \sum_{i=1}^n (x_i - \mu)^2 + \varepsilon_p(\mu - \mu_p)^2\right]\right)$$

$$(x_i - \mu)^2 = (x_i - \bar{x})^2 + (\mu - \bar{x})^2 + 2((x_i - \bar{x})(\bar{x} - \mu))$$

حيث نبسط ما ورده اعلاه لنحصل على :

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n\varepsilon_0(\mu - \bar{x})^2$$

ونجد :

$$f(\mu|x) \propto \exp\left(-\frac{1}{2}\left[\varepsilon_0 \sum_{i=1}^n (x_i - \bar{x})^2 + n(\mu - \bar{x})^2 + \varepsilon_p(\mu - \mu_p)^2\right]\right)$$

$$\propto \exp\left(-\frac{1}{2}\left[n\varepsilon_0(\mu - \bar{x})^2 + \varepsilon_p(\mu - \mu_p)^2\right]\right)$$

لاحظ ان هذا النموذج (2) $-Q(\mu)$ هو من الدرجة الثانية متعددة حيث تستطيع كتابة $Q(\mu)$ كالآتي :

$$[Q(\mu) = n\varepsilon_0(\mu - \bar{x})^2 + \varepsilon_p(\mu - \mu_p)^2]$$

$$n\varepsilon_0\mu^2 - 2n\varepsilon_0\bar{x}\mu + n\varepsilon_0\bar{x}^2 + \varepsilon_p\mu^2 - 2\varepsilon_p\mu_p\mu + \varepsilon_p\mu_p^2$$

$$= (n(\varepsilon_0 + \varepsilon_p)) \left[\mu^2 - \frac{2n\varepsilon_0\bar{x} + \mu_p\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} + C_1 \right]$$

$$= (n(\varepsilon_0 + \varepsilon_p)) \left[\mu - \frac{n\varepsilon_0\bar{x} + \mu_p\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} \right]^2 + C_2$$

حيث ان C_2, C_1 هي ثوابت حيث لا تعتمد على μ

$$f(\mu|x) \propto \exp(-Q(\mu)) \propto \exp\left(-\frac{n\varepsilon_0 + \varepsilon_p}{2} \left[\mu - \frac{n\varepsilon_0\bar{x} + \mu_p\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} \right]^2\right)$$

$$\therefore \mu_1 = \frac{n\varepsilon_0}{n\varepsilon_0 + \varepsilon_p} \bar{X} + \frac{\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} \mu_p$$

2- عندما يكون المتوسط μ معروف وقيمة $\varepsilon (\varepsilon > 0)$ غير معروفة

$$\mu_1 = \frac{\sigma_0^2 \mu_p + n\sigma_p^2 \bar{X}}{\sigma_0^2 + n\sigma_p^2} \text{ and } \sigma_1^2 = \frac{\sigma_0^2 \sigma_p^2}{\sigma_0^2 + n\sigma_p^2}$$

حيث يمكننا كتابة (Poserior) لتوزيع (μ) كالآتي :

$$f(\sigma|x) \propto f_n(x|\mu) \times p(\mu)$$

$$= \left(\frac{\varepsilon_0}{2\pi}\right)^{n/2} \prod_{i=1}^n \exp\left(-\frac{\varepsilon_0}{2}(x_i - \sigma)^2\right) \times \left(\frac{\varepsilon_p}{2\pi}\right) \exp\left(-\frac{\varepsilon_p}{2}(\sigma - \mu_p)^2\right)$$

$$\propto \exp\left(-\frac{1}{2} \left[\varepsilon_0 \sum_{i=1}^n (x_i - \sigma)^2 + \varepsilon_p (\sigma - \mu_p)^2 \right] \right)$$

$$(x_i - \sigma)^2 = (x_i - \bar{x})^2 + (\sigma - \bar{x})^2 + 2((x_i - \bar{x})(\bar{x} - \sigma))$$

حيث نبسط ما ورد اعلاه لنحصل على :

$$\sum_{i=1}^n (x_i - \sigma)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\sigma - \bar{x})^2$$

ونجد :

$$\begin{aligned} f(\sigma|x) &\propto \exp \left(-\frac{1}{2} \left[\varepsilon_0 \sum_{i=1}^n (x_i - \bar{x})^2 + n\varepsilon_0(\sigma - \bar{x})^2 \right. \right. \\ &\quad \left. \left. + \varepsilon_p(\sigma - \sigma_p)^2 \right] \right) \\ &\propto \exp \left(-\frac{1}{2} [n\varepsilon_0(\sigma - \bar{x})^2 + \varepsilon_p(\sigma - \sigma_p)^2] \right) \end{aligned}$$

لاحظ ان هذا النموذج $(-Q(\sigma|2))$ هو من الدرجة الثانية متعددة الحدود حيث تستطيع كتابة $Q(\sigma)$ كالآتي :

$$\begin{aligned} (Q(\sigma) &= n\varepsilon_0(\sigma - \bar{x})^2 + \varepsilon_p(\sigma - \mu\sigma_p)^2] \\ n\varepsilon_0\sigma^2n\bar{x} - 2n\varepsilon_0\bar{x}\sigma + \varepsilon_p\sigma^2 - \varepsilon_p\sigma_p^2M - 2\varepsilon_p\sigma_p\sigma \\ &= (n(\varepsilon_0 + \varepsilon_p))[\sigma^2 - 2\frac{n\varepsilon_0\bar{x} + \sigma_p\varepsilon_p}{n\varepsilon_0 + \varepsilon_p}\sigma + C_1 \\ &= (n(\varepsilon_0 + \varepsilon_p))[\sigma - \frac{n\varepsilon_0\bar{x} + \sigma_p\varepsilon_p}{n\varepsilon_0 + \varepsilon_p}] + C_2 \end{aligned}$$

حيث ان C_2, C_1 هي ثوابت حيث هذه الثوابت لا تعتمد على σ

$$f(\mu|x) \propto \exp \left(-Q \left(\frac{(\sigma)}{2} \right) \right) \propto \exp \left(-\frac{n\varepsilon_0 + \varepsilon_p}{2} [\sigma \right.$$

$$-\frac{\sigma_p \varepsilon_p}{n\varepsilon_0 + \varepsilon_p}]^2)$$

$$\therefore \sigma_1 = \frac{n\varepsilon_0}{n\varepsilon_0 + \varepsilon_p} \bar{X} + \frac{\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} \sigma_p$$

الفصل الثالث

الجانب العملي

تم الحصول على البيانات للمرضى من مستشفى الديوانية التعليمي / شعبة الاحصاء للذين تعرضوا لعلاjin لمرض ضغط الدم بحيث وضحت التجربة استجابات المرض للأدوية التي معموله مع نظام الرنين ويرمز لها للاختصار بـ RAAS وهو بروتين يوجد في الكلية يزيد ضغط الدم ويتأثر بفعالية البلازما بصورة ضعيفة .

وتشير الدراسة الى اخذ عينة من الاشخاص المصابين بمرض ضغط الدم والذي يبلغ عددهم (٥٠ شخص) ، وتم اعطاء لكل مريض نوعين من الادوية .

النوع الاول (ARB) والنوع الثاني (ACE-1) والذي تم اخذه بعد مرور الشهر الاول وسجلت استجابات كل شخص مريض لكل نوع من انواع الادوية وكانت النتائج كما مبينة في الجدول ادناه .

يرمز لعامل الاستجابة بـ (1) وعدم الاستجابة بـ (0). ويمثل الـ X_i مجاميع الاستجابة للدوائيين .

i	ARB	ACE-1	Xi
1	1	1	2
2	0	1	1
3	0	1	1
4	1	1	2
5	1	1	2
6	1	1	2
7	0	0	0
8	1	0	1
9	1	0	1
10	1	1	2
11	1	1	2
12	0	1	1
13	0	0	0
14	1	0	1
15	0	0	0
16	1	0	1
17	0	1	1
18	1	1	2
19	1	1	2
20	1	1	2
21	0	1	1
22	1	0	1
23	0	1	1
24	0	0	0
25	0	0	0
26	1	1	2
27	0	0	0
28	1	1	2
29	1	1	2
30	1	0	1
31	1	1	2
32	0	1	1
33	1	1	2
34	0	0	0
35	1	0	1
36	1	1	2
37	1	1	2
38	0	1	1
39	0	0	0
40	0	1	1
41	1	1	2
42	1	1	2
43	1	0	1
44	0	0	0
45	0	1	1
46	1	1	2
47	1	1	2
48	0	1	1
49	1	1	2
50	0	0	0
			$\sum Xi = 63$

وقد تم اختبار البيانات باستخدام برنامج Easyfit باستخدام مربع كاي Chi-square وكانت النتائج كالآتي :

Chi-Squared					
Deg. of freedom	2				
Statistic	2.2343				
P-Value	0.32721				
Rank	10				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Reject?	No	No	No	No	No

١- ايجاد نسبة الخطأ MSE بطريقة العزوم MLE

$$MSE = \frac{\sum_{i=1}^n (F - \hat{F})^2}{n}$$

حيث ان :

$$F = \frac{i+3/8}{n+0.25}$$

هي القيمة المحسوبة بالقانون

$\hat{F}$: تمثل قيمة F الجدولية (F table) .

نقوم بايجاد قيمة المعلمتين μ , σ المقدرتين بطريقة العزوم MLE

$$\hat{\mu} = \bar{X} \rightarrow \bar{X} = \sum_{i=1}^{50} \frac{Xi}{n} \rightarrow \bar{X} = \frac{65}{50} = 1.26$$

$$\hat{\sigma}^2 = \sum_{i=1}^{50} Xi^2 - \bar{X}^2 = -456.251 \rightarrow \hat{\sigma} = 21.360$$

ثم بعد ذلك نقوم بتعويض قيمة تلك المعلمتين في قانون التوزيع الطبيعي القياسي والذي هو :

$$Z = \frac{X - \mu}{\sigma}$$

حيث Z هي القيمة القياسية المحسوبة والتي من خلالها نستطيع استخراج قيمة F الجدولية .

والجدول التالي يبين نتائج هذه الطريقة .

i	ARB	ACE-1	Xi	(Xi)^2	Z	F table	F
1	1	1	2	4	1.942	0.97	0.2775
2	0	1	1	1	0.942	0.827	0.2975
3	0	1	1	1	0.942	0.827	0.3175
4	1	1	2	4	1.942	0.97	0.3375
5	1	1	2	4	1.942	0.97	0.3575
6	1	1	2	4	1.942	0.97	0.3775
7	0	0	0	0	-0.057	0.477	0.3975
8	1	0	1	1	0.942	0.827	0.4175
9	1	0	1	1	0.942	0.827	0.4375
10	1	1	2	4	1.942	0.97	0.4575
11	1	1	2	4	1.942	0.97	0.4775
12	0	1	1	1	0.942	0.827	0.4975
13	0	0	0	0	-0.057	0.477	0.5175
14	1	0	1	1	0.942	0.827	0.5375
15	0	0	0	0	-0.057	0.477	0.5575
16	1	0	1	1	0.942	0.827	0.5775
17	0	1	1	1	0.942	0.827	0.5975
18	1	1	2	4	1.942	0.97	0.6175
19	1	1	2	4	1.942	0.97	0.6375
20	1	1	2	4	1.942	0.97	0.6575
21	0	1	1	1	0.942	0.827	0.6775
22	1	0	1	1	0.942	0.827	0.6975
23	0	1	1	1	0.942	0.827	0.7175
24	0	0	0	0	-0.057	0.477	0.7375
25	0	0	0	0	-0.057	0.477	0.7575
26	1	1	2	4	1.942	0.97	0.7775
27	0	0	0	0	-0.057	0.477	0.7975
28	1	1	2	4	1.942	0.97	0.8175
29	1	1	2	4	1.942	0.97	0.8375
30	1	0	1	1	0.942	0.827	0.8575
31	1	1	2	4	1.942	0.97	0.8775
32	0	1	1	1	0.942	0.827	0.8975
33	1	1	2	4	1.942	0.97	0.9175
34	0	0	0	0	-0.057	0.477	0.9375
35	1	0	1	1	0.942	0.827	0.9575
36	1	1	2	4	1.942	0.97	0.9775
37	1	1	2	4	1.942	0.97	0.9975
38	0	1	1	1	0.942	0.827	1.0175
39	0	0	0	0	-0.057	0.477	1.0375
40	0	1	1	1	0.942	0.827	1.0575
41	1	1	2	4	1.942	0.97	1.0775
42	1	1	2	4	1.942	0.97	1.0975
43	1	0	1	1	0.942	0.827	1.1175
44	0	0	0	0	-0.057	0.477	1.1375
45	0	1	1	1	0.942	0.827	1.1575
46	1	1	2	4	1.942	0.97	1.1775
47	1	1	2	4	1.942	0.97	1.1975
48	0	1	1	1	0.942	0.827	1.2175
49	1	1	2	4	1.942	0.97	1.2375
50	0	0	0	0	-0.057	0.477	1.2575
			$\Sigma Xi = 63$			$\Sigma Ft = 40.853$	$\Sigma F = 38.375$

$$MSE_{MLE} = \frac{\sum(F - \hat{F})^2}{n} = \frac{(38.375 - 40.853)^2}{50} = 0.122$$

٢- ايجاد نسبة الخطأ بطريقة المربعات الصغرى OLS

نقوم بايجاد قيمة المعلمتين μ , σ المقدرتين بطريقة المربعات الصغرى OLS ، ثم بعد ذلك نقوم بتعويض تلك المعلمتين بقانون التوزيع الطبيعي القياسي .

$$Z = \frac{X - \mu}{\sigma}$$

في طريقة المربعات الصغرى OLS لنموذج الانحدار الخطي للتوزيع الطبيعي

$$\hat{Y}_i = \hat{\sigma} + \hat{\beta} X_i$$

وعليه فإن : $\hat{\beta}$ تكافئ $\hat{\mu}$ ، والـ $\hat{\sigma}$ تكافئ $\hat{\alpha}$ ، وقيمة $\hat{\beta}$ تكافئ $\hat{Y}_i$ تكافئ F الجدولية .
اذن :

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{(-826.05)(0.009)}{(-456.251)} = \frac{-7.43445}{-456.251} = 0.016$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} = 0.81706 - (0.016 * 63) = 0.190$$

$$\hat{\alpha} = \hat{\sigma}^2 = 0.190 \rightarrow \hat{\sigma} = 0.435$$

والجدول التالي يوضح نتائج هذه الطريقة .

i	ARB	ACE-1	Xi	(Xi)^2	Z	F table	F
1	1	1	2	4	1.963	1.18	0.2775
2	0	1	1	1	0.963	0.136	0.2975
3	0	1	1	1	0.963	0.136	0.3175
4	1	1	2	4	1.963	1.18	0.3375
5	1	1	2	4	1.963	1.18	0.3575
6	1	1	2	4	1.963	1.18	0.3775
7	0	0	0	0	-0.036	0.98	0.3975
8	1	0	1	1	0.963	0.136	0.4175
9	1	0	1	1	0.963	0.136	0.4375
10	1	1	2	4	1.963	1.18	0.4575
11	1	1	2	4	1.963	1.18	0.4775
12	0	1	1	1	0.963	0.136	0.4975
13	0	0	0	0	-0.036	0.98	0.5175
14	1	0	1	1	0.963	0.136	0.5375
15	0	0	0	0	-0.036	0.98	0.5575
16	1	0	1	1	0.963	0.136	0.5775
17	0	1	1	1	0.963	0.136	0.5975
18	1	1	2	4	1.963	1.18	0.6175
19	1	1	2	4	1.963	1.18	0.6375
20	1	1	2	4	1.963	1.18	0.6575
21	0	1	1	1	0.963	0.136	0.6775
22	1	0	1	1	0.963	0.136	0.6975
23	0	1	1	1	0.963	0.136	0.7175
24	0	0	0	0	-0.036	0.98	0.7375
25	0	0	0	0	-0.036	0.98	0.7575
26	1	1	2	4	1.963	1.18	0.7775
27	0	0	0	0	-0.036	0.98	0.7975
28	1	1	2	4	1.963	1.18	0.8175
29	1	1	2	4	1.963	1.18	0.8375
30	1	0	1	1	0.963	0.136	0.8575
31	1	1	2	4	1.963	1.18	0.8775
32	0	1	1	1	0.963	0.136	0.8975
33	1	1	2	4	1.963	1.18	0.9175
34	0	0	0	0	-0.036	0.98	0.9375
35	1	0	1	1	0.963	0.136	0.9575
36	1	1	2	4	1.963	1.18	0.9775
37	1	1	2	4	1.963	1.18	0.9975
38	0	1	1	1	0.963	0.136	1.0175
39	0	0	0	0	-0.036	0.98	1.0375
40	0	1	1	1	0.963	0.136	1.0575
41	1	1	2	4	1.963	1.18	1.0775
42	1	1	2	4	1.963	1.18	1.0975
43	1	0	1	1	0.963	0.136	1.1175
44	0	0	0	0	-0.036	0.98	1.1375
45	0	1	1	1	0.963	0.136	1.1575
46	1	1	2	4	1.963	1.18	1.1775
47	1	1	2	4	1.963	1.18	1.1975
48	0	1	1	1	0.963	0.136	1.2175
49	1	1	2	4	1.963	1.18	1.2375
50	0	0	0	0	-0.036	0.98	1.2575
			$\Sigma Xi = 63$			$\Sigma Ft = 37.164$	$\Sigma F = 38.375$

$$MSE_{ols} = \frac{\sum (F - \hat{F})^2}{n} = \frac{(38.375 - 37.164)^2}{50} = 0.029$$

٣- ايجاد نسبة الخطأ بطريقة بايز Bayes

نقوم بايجاد قيمة المعلمتين μ , σ المقدرتين بطريقة بايز

$$\mu = \frac{n\varepsilon_0}{n\varepsilon_0 + \varepsilon_p} \bar{X} + \frac{\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} \mu_p = \frac{(50)(2)}{(50)(2) + 2} (63) + \frac{2}{(50)(2) + 2} (3) = 61.822$$

$$\sigma = \frac{n\varepsilon_0}{n\varepsilon_0 + \varepsilon_p} \bar{X} + \frac{\varepsilon_p}{n\varepsilon_0 + \varepsilon_p} \sigma_p = \frac{(50)(2)}{(50)(2) + 2} (63) + \frac{2}{(50)(2) + 2} (0.707) = 61.777$$

ثم بعد ذلك نقوم بتعويض تلك المعلمتين بقانون التوزيع الطبيعي القياسي .

$$Z = \frac{X - \mu}{\sigma}$$

والجدول التالي يوضح نتائج هذه الطريقة .

i	ARB	ACE-1	X_i	$(X_i)^2$	Z	F table	F
1	1	1	2	4	-0.96	0.168	0.2775
2	0	1	1	1	-0.98	0.163	0.2975
3	0	1	1	1	-0.98	0.163	0.3175
4	1	1	2	4	-0.96	0.168	0.3375
5	1	1	2	4	-0.96	0.168	0.3575
6	1	1	2	4	-0.96	0.168	0.3775
7	0	0	0	0	-1	0.158	0.3975
8	1	0	1	1	-0.98	0.163	0.4175
9	1	0	1	1	-0.98	0.163	0.4375
10	1	1	2	4	-0.96	0.168	0.4575
11	1	1	2	4	-0.96	0.168	0.4775
12	0	1	1	1	-0.98	0.163	0.4975
13	0	0	0	0	-1	0.158	0.5175
14	1	0	1	1	-0.98	0.163	0.5375
15	0	0	0	0	-1	0.158	0.5575
16	1	0	1	1	-0.98	0.163	0.5775
17	0	1	1	1	-0.98	0.163	0.5975
18	1	1	2	4	-0.96	0.168	0.6175
19	1	1	2	4	-0.96	0.168	0.6375
20	1	1	2	4	-0.96	0.168	0.6575
21	0	1	1	1	-0.98	0.163	0.6775
22	1	0	1	1	-0.98	0.163	0.6975
23	0	1	1	1	-0.98	0.163	0.7175
24	0	0	0	0	-1	0.158	0.7375
25	0	0	0	0	-1	0.158	0.7575
26	1	1	2	4	-0.98	0.163	0.7775
27	0	0	0	0	-1	0.158	0.7975
28	1	1	2	4	-0.96	0.168	0.8175
29	1	1	2	4	-0.96	0.168	0.8375
30	1	0	1	1	-0.98	0.163	0.8575
31	1	1	2	4	-0.96	0.168	0.8775
32	0	1	1	1	-0.98	0.163	0.8975
33	1	1	2	4	-0.96	0.168	0.9175
34	0	0	0	0	-1	0.158	0.9375
35	1	0	1	1	-0.98	0.163	0.9575
36	1	1	2	4	-0.96	0.168	0.9775
37	1	1	2	4	-0.96	0.168	0.9975
38	0	1	1	1	-0.98	0.163	1.0175
39	0	0	0	0	-1	0.158	1.0375
40	0	1	1	1	-0.98	0.163	1.0575
41	1	1	2	4	-0.96	0.168	1.0775
42	1	1	2	4	-0.96	0.168	1.0975
43	1	0	1	1	-0.98	0.163	1.1175
44	0	0	0	0	-1	0.158	1.1375
45	0	1	1	1	-0.98	0.163	1.1575
46	1	1	2	4	-0.96	0.168	1.1775
47	1	1	2	4	-0.96	0.168	1.1975
48	0	1	1	1	-0.98	0.163	1.2175
49	1	1	2	4	-0.96	0.168	1.2375
50	0	0	0	0	-1	0.158	1.2575
			$\Sigma X_i = 63$			$\Sigma F_t = 8.2$	$\Sigma F = 38.375$

$$MSE_{Bayes} = \frac{\sum (F - \hat{F})^2}{n} = \frac{(38.375 - 8.2)^2}{50} = 18.21$$

الاستنتاجات

نلاحظ من خلال دراستنا للبحث عن افضل طريقة لتخمين معلمات التوزيع الطبيعي حيث وجد ان :

- 1- طريقة المربعات الصغرى OLS هي افضل طريقة لتخمين معلمات التوزيع الطبيعي لانها تحتوي ع اقل نسبة خطأ MSE.
- 2- ثم تاتي طريقة العزوم MLE بالمرتبة الثانية ثم بعد ذلك تاتي طريقة بايز Bayes.

المصادر

1- احصاء رياضي (2) اميرحنا هرمز (1981)

2-Kevin P.(2007)

Conjugate Bayesian analysis of the Gaussain distribution .

3-Songfeng. Z

Bayesian Approoch to progmeter estimation .

4-الاحصاء الحياتي الدكتور عبد الخالق عبد الجبار النقيب 2011م

5-الاحصاء الاستدلالي الدكتور حسن ياسين طعمة الدكتور ايمان حسين حنوش (2012)

فائنة المختبرات

الموضوع	رقم الصفحة
الاية	أ
الاهداء	ب
شكر وتقدير	ت
الفصل الاول	1
الفصل الثاني	13
الفصل الثالث	29
استنتاجات	38
المصادر	39
فهرست المحتويات	40