مجلة جامعة تشرين للبحوث والدراسات العلمية – سلسلة العلوم الأساسية المجلد (40) العدد (1) 2018
Tishreen University Journal for Research and Scientific Studies - Basic Sciences Series Vol. (40) No. (1) 2018

تأثير كل من حجم ونوع المعاينة الإحصائية على تقديرات معاملات معادلة الانحدار الخطي البسيط

د. محمد مزيد دريباتي*
د. منذر بوبو**
هديل درويش***

(تاريخ الإيداع 5 / 11 / 2017. قُبِل للنشر في 8 / 2 /2018)

□ ملخّص □

يهدف البحث الحالي إلى دراسة تقديرات معاملات معادلة خط الانحدار الخطي البسيط باستخدام طريقة المربعات الصغرى وذلك عند حجوم عينات مختلفة وطرق معاينة مختلفة. وبذلك يكون هدف البحث هو محاولة لتحديد الحجم الأمثل والمعاينة الأفضل لتقدير هذه المعاملات. تم استخدام بيانات تجريبية لمجتمع مؤلف من 2000 فرداً من طلاب مدارس مناطق مختلفة من القطر. وقد تم في كل مرة تغيير حجم العينة وحساب المعاملات ثم مقارنة هذه المعاملات لحجوم عينات مختلفة مع معاملات المجتمع الحقيقي؛ وقد بينت النتائج أن تقديرات معاملات معادلة خط الانحدار تقترب من القيم الحقيقية لمعاملات معادلة خط الانحدار للمجتمع عندما يقترب حجم العينة من القيمة (325). كما تبين أن المعاينة بالطريقة العشوائية الطبقية ذات التوزيع المتناسب مع حجوم الفئات يعطي النتائج الأفضل والاكثر دقة لتقدير معادلة الانحدار الخطي بطريقة المربعات الصغرى.

الكلمات المفتاحية: الانحدار الخطي البسيط، طريقة المربعات الصغرى، المعاينة الإحصائية، العينة الإحصائية.

* أستاذ مساعد ، قسم الإحصاء الرياضي، كلية العلوم، جامعة تشرين. اللاذقية سورية.
** أستاذ مساعد ، قسم القياس والتقويم، كلية التربية، جامعة تشرين. اللاذقية سورية.
*** طالبة ماجستير، قسم الإحصاء، كلية العلوم، جامعة تشرين. اللاذقية، سورية.

مجلة جامعة تشرين للبحوث والدراسات العلمية – سلسلة العلوم الأساسية المجلد (40) العدد (1) 2018
Tishreen University Journal for Research and Scientific Studies - Basic Sciences Series Vol. (40) No. (1) 2018

The Effect of Both the Size and The Type of Statistical Sampling on The Estimates of Simple Linear Regression Equation Coefficients

Dr. Muhammad Mazyad Drybati *
Dr. Mounzer Boubou **
Hadil Darwish ***

(Received 5 / 11 / 2017. Accepted 8 / 2 /2018)

□ ABSTRACT □

These papers aim to study the estimation of the simple linear regression equation coefficients using the least square method at different sample sizes and different sampling methods. And so on, the main goal of this research is to try to determine the optimum size and the best sampling method for these coefficients. We used experimental data for a population consist of 2000 students from different schools all over the country. We had changed the sample size each time and calculate the coefficients and then compare these coefficients for different sample sizes with their coefficients of the real population; and the results have been shown that the estimation of the linear regression equation coefficients are close from the real values of the coefficients of the regression line equation for the population when the sample size closes the value (325). As it turns out that the Stratified random sampling with proportional distribution with class sizes gives the best and most accurate results to estimate linear regression equation with least square method.

Key Words: Linear Regression، Least Squares Method، Statistical Sampling، Statistical Sample.

* **Associate Professor, Department of Statistics, Tishreen University, Latakia, Syria.**
** **Associate Professor, Department of Measurement and Evaluation, Tishreen University, Latakia, Syria.**
*** **Master Student, Department of Statistics, Tishreen University, Latakia, Syria.**

مقدمة

يعتبر الانحدار الخطي البسيط واحداً من أهم وأبسط الأساليب الإحصائية المستخدمة في التنبؤ الاحصائي، ويقوم مبدأ الانحدار على دراسة العلاقة بين متغيرين يدعى الأول المتغير التابع ويرمز له عادة بالرمز (Y) والمتغير أو المتغيرات المستقلة والتي يرمز لها عادة بالرمز (X). ويتم تشكيل معادلة الانحدار من خلال تشكيل معادلة خط مستقيم وإن تقدير معاملات خط الانحدار يعتمد بشكل أساسي على حجم العينة ونوع المعاينة. إن هذه المعاملات تتأثر بشكل كبير (كما سنرى) بقيم المتوسطات الحسابية ولانحرافات المعيارية للمتغيرات الداخلة في حساب هذه المعادلة. ومن المعروف إن حجم المعاينة ونوعها تلعبان دورا مهما في تقدير كل من المتوسط والانحراف المعياري.

وبما أن دراسة المجتمع ككل امرأ ليس بالسهل؛ يلجأ الباحثون إلى استخدام نظرية المعاينة الإحصائية؛ والتي تقوم على دراسة بيانات العينة الإحصائية التي يتم سحبها من المجتمع ومن ثم باستخدام القوانين الإحصائية المناسبة لكل نوع من أنوع المعاينة يتم تقدير معالم المجتمع التي سحبت منه العينة. وقد أصبحت نظرية المعاينة الإحصائية جزءا لا يتجزأ من العلوم الإحصائية المختلفة بعد أن أصبح لهذه النظرية أسس علمية رياضية، وذلك لأن تطبيقها في البحوث العلمية يختصر كثيرا من الوقت والجهد اللازمين لعمليات البحث الشامل، وبالتالي نقتصد كمية لا بأس بها من التكاليف اللازمة لهذه العملية، وتظهر أهمية بحوث العينات في تصحيح معلومات البحث الشامل عندما تكون نسبة الأخطاء فيها كبيرة، أو عندما لا نتمكن من دراسة جزء ما من وحدات الموضوع المدروس، إضافة إلى القدرة على معرفة الدقة المتوفرة في معلومات البحث.

وفي هذا البحث استخدمنا نظرية العينات وخصائص سحب العينة العشوائية لتحسين نوع وحجم العينة المستخدمة في تقدير معاملات الانحدار الخطي البسيط لإعطاء أفضل نتائج ممكنة وبأقل تكلفة.

أهمية البحث وأهدافه:

تأتي أهمية البحث من أهمية الموضوع الذي يتم معالجته؛ فالانحدار الخطي البسيط هو من أهم الأساليب الإحصائية المستخدمة في بحوث التنبؤ والعلاقة الارتباطية الإحصائية في المجالات كافة. ومن خلال عمل الباحثين في مجال البحث العلمي مع غير المختصين يطرح علينا دائما السؤال التالي:" ما هو أقل حجم عينة يمكننا استخدامه من أجل تقدير معاملات الانحدار؟" من هنا كانت فكرة محاولة دراسة هذه التقديرات من أجل مساعدة الباحثين في الإجابة على هذا السؤال. كما تأتي الأهمية من طريقة الأسلوب المتبع في تحديد هذه التقديرات من خلال المقارنة بين حجوم وأنواع المعاينات المختلفة. أما هدف البحث الرئيس فيتحدد من خلال تحديد أفضل حجم ونوع معاينة لتقدير معاملات معادلة خط الانحدار الخطي البسيط باقل تكلفة ممكنة وذلك لأن عملية السحب للوحدات الإحصائية قد يترتب عليها تكلفة معينة نحاول التقليل منها مع المحافظة على الحصول على أفضل التقديرات الإحصائية.

طرائق البحث ومواده:

يستخدم البحث المنهج التجريبي في الحصول على النتائج من خلال التجريب ولعدد كبير من المرات بسحب عينات مختلفة وحجوم مختلفة. ويعتبر المنهج التجريبي واحداً من أهم المناهج المستخدمة في البحث العلمي.

الدراسات السابقة:

1- [8] قامت صفاء كريم كاظم عام 2009م من خلال بحثها "المقارنة بين تقديرات معالم نموذج الانحدار الخطي المتعدد باستخدام أسلوب OLS وأسلوب برمجة الأهداف الخطية" بدراسة طريقة المربعات الصغرىOLS مع طريقة أخرى هي أسلوب برمجة الأهداف الخطية Goal Programming، وذلك لتقدير معاملات الانحدار الخطي المتعدد، وبعد الحصول على التقديرات بالأسلوبين تبين أن الأسلوبين ملائمان في تقدير معلمات نموذج الانحدار الخطي المتعدد وذلك بعد المقارنة بين الطريقتين باستخدام اختبار Wilcoxon Test.

2- [12] في عام 2013م درس Pranap مقدرات المربعات الصغرى وقارنها بطرق أخرى لتحليل الانحدار الخطي، من خلال بحثه "Estimate Of The Regression Coefficient Based On kendall"'s Tau، ووجد أن مقدرات المربعات الصغرى عرضة لأخطاء فادحة وهي حساسة للتوزيعات غير الطبيعية حيث اقترح مقدر (نقطي ومجالي) قوي، فعال، بسيط وغير متحيز لمعاملات الانحدار الخطي يعتمد على معامل الارتباط الرتبي (Kendall Tau) للعينة ولكن هذا المقدر غير كاف ذلك لأنه حساس للتوزيعات ثقيلة الذيل.

3- [10] بحث Herve Abdi عام 2007م في بحثه "The Method Of Least Squares" استخدام طريقة المربعات الصغرى في تقدير المعالم الأكثر شيوعا في المجتمع كالمتوسط الحسابي، حيث وجد أن المربعات الصغرى هي طريقة مرنة في التقدير وتعتمد على أسس رياضية كالاشتقاق وتعطي تقديرات مثلى حيث تحاول تقليل مربعات الخطأ العشوائي، ولكنها تعاني حساسية عالية للقيم المتطرفة، ذلك لأن التربيع يؤدي إلى زيادة القيمة العددية المتطرفة وبالتالي زيادة تأثيرها سلبا في دقة التقديرات، وقد اقترح لحل هذه المشكلة استخدام تقنيات أقل حساسية لتأثير القيم المتطرفة، وهذه التقنيات هي حاليا قيد الدراسة والتطوير.

4- [11] قام Mengue Park عام 2002م من خلال بحثه "Regression Estimation Of The Mean In Survey Sampling"، بدراسة تأثير أسلوب المعاينة وتصميم العينة على تقدير معاملات معادلة خط الانحدار مستخدما معادلة الانحدار السكاني كمثال، حيث توصل إلى إيجاد مقدر انحدار ذو خصائص مرغوبة إحصائيا مثل عدم الانحياز والتباين الأدنى والاتساق، وقام تحت افتراض التوزيع الطبيعي بتصميم عينة عشوائية طبقية معينة من خلال دراسة محاكاة تجنب الباحث الوقوع في مشاكل عدم كفاية المعلومات التي تقدمها البيانات المتوفرة في العينة.

5- [14] في عام 1997م قام Fan & Wang بدراسة تحت عنوان "The Effects Of Sample Size, Estimation Methods, and Model, Specification On SEM Indxes" والتي ركزت على جانب من جوانب المعاينة، وهو الجانب العنقودي بهدف استعراض ووصف وإظهار كيف يؤثر الرسم العنقودي على الخطأ المعياري، وعلى النتائج اللاحقة للتحليل. وهذه الدراسة تطبيقية على عينة حجمها 150 حالة من مجتمع قوامه 1000 شخص.

ومن نتائج الدراسة: أنه باستعمال الرسم العنقودي كان هناك فروق حقيقية في تقدير المعالم في عدد من الحالات مقارنة باستخدام الطرق العادية في الإحصاء (أخذ عينات عشوائية)، وأن عدم الدقة في تقدير الخطأ، سيؤدي إلى نتائج مضللة في تقدير المعالم وقلة في الثقة بتلك التقديرات.

6- [6] قدم الغامدي عام 2000م في دراسته "أثر أسلوب اختيار العينة وحجمها على دقة تقدير معالم المجتمع الإحصائي" أساسا لتحسين تصميم العينة واتخاذ القرارات الدقيقة حول أهم خطوة من خطوات تصميم أبحاث العينات وبالتحديد ركز على حساب حجم انحراف إحصاءة العينة عن معلمة المجتمع الكلي ومدى تأثير أسلوب المعاينة وحجم

العينة في ضوء طبيعة البيانات المختلفة على حجم هذه الانحرافات، وذلك من خلال دراسة تطبيقية على عينة من علامات طلاب ثانوية جدة (9113 طالب)، واعتماد تكرار استخراج العينة، حيث توصل إلى أن حجم انحراف التقديرات الناتجة عن العينة للمتوسطات الحسابية يتوقف على أخطاء الباحث حيث تشمل أخطاء الانحياز التي يصعب التحكم بها وتقليلها وكذلك أخطاء المعاينة العشوائية، ووجد أن استخدام أسلوب المعاينة المناسب يقلل من كمية هذه الأخطاء.

7- [3] في عام 2012م درس محمد ابراهيم الشادري في بحثه "تأثير حجم العينة على قوة الاختبار الإحصائي"، تأثير حجم العينة على قوة اختبار T وقوة اختبار فيشر F.

حيث شملت عينة الدراسة عينات عشوائية من المفردات الإحصائية التي تم توليدها تراوحت بين 10 إلى 330 مفردة، حسب طبيعة الاختبار الإحصائي المستخدم حيث تحقق هذه البيانات مجموعة الافتراضات اللازمة لاختبار T و F وتوصل إلى اقتراح الحجم الأدنى الأمثل الذي يعطي قوة اختبار في حالة عينة واحدة وفي حالة عينتين (مترابطتين أو مستقلتين) سواء تساوت العينات بالحجم أو اختلفتا وذلك لكل من الاختبارين.

الجانب النظري:

فيما يلي التعاريف والمفاهيم الأساسية التي تتعلق بنظرية العينات وتحليل الانحدار وخواص طريقة المربعات الصغرى.

تعريف تحليل الانحدار(Regression): [2]

هوَ أسلوبٌ إحصائيٌّ يقومُ بصياغةِ دالةٍ رياضيَّةٍ لعمليَّةٍ ذات عوامل مؤثِّرةٍ عدة $x_1, x_2, \ldots, x_m$ لوصف متغيِّر ناتج Y من هذهِ العمليَّة والتَّحكُّمُ بهِ وتوقُّعُ قيمٍ غير معروفةٍ له.

ويكونُ شكلُ الدَّالة الرِّياضيَّةِ على الصورة :

$$Y = f\left(x_1, x_2, \ldots, x_m\right)$$

تسمَّى المتغيراتُ $x_1, x_2, \ldots, x_m$ بالمتغيراتِ المستقلةِ ، والمتغيرُ النَّاتجُ Y بالمتغيِّرِ التَّابعِ ، تعرفُ الدَّالةُ السَّابقةُ بدالةِ الانحدارِ الخطيِّ المتعددِ حيثُ يعبَّرُ عنها بالشكل :

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \varepsilon_i$$

طريقة المربعات الصغرى في تقدير خط الانحدار(LEAST SQUARES METHOD): [4]

تنصُّ نظريَّةُ غاوس – ماركوف أنَّ طريقةَ المربَّعاتِ الصُّغرى العاديّة (OLS :Ordinary Least Square) تُعطي أكفأ تقديرات خطيَّة غير متحيِّزة للمعالم بِشرط تَوافُر فروض معيَّنة .

فروض طريقة المربعات الصغرى OLS :

1. إنَّ الخطأَ العشوائيّ ε_i متغير عشوائي مستقل عن المتغيرات المستقلة x_i يرجع إلى الصدفة البحتة ويأخذ قيماً موجبة وسالبة وصفرية بحيث يكون متوسطه يساوي الصفر: $E(\varepsilon_i) = 0, \forall i = 1, \ldots, n$ أي أن مجموع الانحرافات الموجبة للخطأ العشوائي يساوي تماماً مجموع الانحرافات السالبة. ومتوسط قيم الانحرافات يساوي الصفر.

2. تجانس (ثبات) تباين الأخطاء العشوائية: وهو ما يعني أن تشتتها حول المتوسط ثابت وهو ما يعبر عنه بالعلاقة $Var(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$ ويكفل هذا الفرض أن كل مشاهدة يمكن الاعتماد عليها بنفس القدر، بمعنى أن البيانات التي تم جمعها لتقدير العلاقة يمكن أن الاعتماد عليها بنفس الدرجة، فكل مشاهدة توثر بنفس القوة في العلاقة

التي تقدرها، بحيث تكون تقديرات معاملات الانحدار كفؤ، وتكون اختبارات الفروض الخاصة بها غير متحيزة. كما أن المتغير العشوائي ε_i يتوزع توزيعاً طبيعياً بمتوسط صفر وتباين ثابت يعبر عنه بالعلاقة: $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$

3. إن المتغير التابع Y_i يعتمد على الحد العشوائي ε_i ويتبع أيضاً التوزيع الطبيعي حيث :

$$E(Y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 + \beta_1 x_i + E(\varepsilon_i) = \beta_0 + \beta_1 x_i$$

$$\sigma^2(Y_i) = \sigma^2(\beta_0 + \beta_1 x_i + \varepsilon_i) = \sigma^2(\varepsilon_i) = \sigma^2$$

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

4. لا يوجد ارتباط بينَ قيمِ المتغيِّرِ العشوائيِّ ε_i وأيٍّ منَ المتغيراتِ المستقلّة x_i: $\text{cov}(\varepsilon_i, x_i) = 0$

5. لا يوجد ارتباط بين قيم المتغير العشوائي ε_t في الفترة الزمنية t وقيمه في فترة سابقة أو لاحقة مثل $\varepsilon_t + 1$ و $\varepsilon_t - 1$: **أي ان:** $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$

6. في حالةِ الانحدارِ الخطيِّ المتعدِّد يُفترضُ عدمُ وجودِ علاقةِ ارتباطٍ قويَّةٍ بين المتغيراتِ المستقلّة: $\mathbf{cov}(x_i، x_j)=0$

7. افترض عدم وجود أخطاء في تحديدِ النّموذج.

8. افترض عدم وجود أخطاء في قياسِ المتغيراتِ المستقلّة.

9. في حالةِ جمعِ البياناتِ يُفترضُ عدمُ وجودِ أخطاء تجميع(aggregation).

مقدرات المربعات الصغرى: [13]

الهدفُ من طريقةِ المربَّعاتِ الصُّغرى هو إيجادُ تقديرات b_0 و b_1 ل β_0 وβ_1 على التَّرتيب ، يكون Q من أجلها أصغرَ ما يمكن وبمعنى أدق تُزوِّدنا هذهِ التقديرات بتوفيقٍ "جيِّد " لدالّة الانحدار .

يمكنُ الحصولُ على مقدَّراتِ β_0 وβ_1 التي تحقّقُ قاعدةَ المربعاتِ الدُّنيا ، بطريقتينِ أساسيّتينِ . في الأولى يمكنُ استخدامُ طرق البحثِ العدديَّةِ التي تحسبُ بصورةٍ متناسقةٍ معيارَ المربَّعاتِ الدُّنيا Q لتقديراتِ β_0 و β_1 مختلفة حتَّى نعثرَ على تلك التي تجعلُ Q أصغرَ مايمكن . وفي الطَّريقةِ الثانيةِ نجدُ بأسلوبٍ تحليليٍّ القيم b_0 و b_1 التي تجعلُ Q أصغرَ ما يمكن . وتكونُ الطريقةُ التحليليةُ ممكنةً عندما لا يكونُ النّموذجُ معقّداً من النَّاحيةِ الرِّياضيّةِ.

$$Q = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

ويمكنُ الحصولُ على b_0 و b_1 الّتي تجعل Q أصغرَ ما يمكن باشتقاقِ Q جزئيّاً بالنِّسبة لـ β_0 وβ_1:

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

وبفكِّ القوسين :

$$\sum_{i=1}^{n} y_i - n\, b_0 - b_1 \sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} x_i\, y_i - b_0 \sum_{i=1}^{n} x_i - b_1 \sum_{i=1}^{n} x_i^2 = 0$$

فنحصلُ على المعادلتين النَّاظميتين و بإعادةِ ترتيبِ الحدود :

$$\sum_{i=1}^{n} y_i = n\, b_0 + b_1 \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i\, y_i = b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} {x_i}^2$$

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} {x_i}^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$b_0 = \frac{1}{n}(\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i) = \bar{y} - b_1 \bar{x}$$

حيثُ $\bar{x}$، $\bar{y}$ متوسِّطات المشاهدات x_i y_i على التَّرتيب S_x^2 هو تباينُ المشاهدات :x

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

تعاريف أساسية في نظرية العينات: [9]

المجتمع (Population):

هو جميعُ الوحداتِ الإحصائية التي يرادُ إجراءُ البحثِ الإحصائي عليها ، ومن الضَّروريِّ تعريفُ هذهِ الوحداتِ بشكلٍ واضحٍ بحيثُ تَجمعُها صفةٌ واحدةٌ أو صفاتٌ مشتركةٌ ، ومعظمُ المجتمعاتِ الإحصائية مؤلفةٌ من وحداتٍ إحصائية تتغيَّرُ حسبَ الزَّمنِ (مجتمعات متجددة) ، وبعضُها الآخرُ مجتمعاتٌ ثابتةٌ لا تتغيَّرُ حسبَ الزَّمن. ويرمز لعدد واحدات المجتمع بالرمز N (حجم المجتمع).

العينة(Sample):

جزءٌ من المجتمعِ الإحصائيِّ يَتِم اختيارُهُ وفقَ أساليبِ المعاينةِ الإحصائيَّةِ ويشترط أن تكونَ ممثلةً للمجتمعِ الَّذي نقومُ بدراستهِ . ولكي تكونَ العينةُ ممثلةً للمجتمعِ يجبُ أن تتضمَّنَ خصائصَ المجتمعِ بشكلٍ يمكننا تعميمُ نتائجها لتقديرِ أهم معالمِ المجتمعِ الإحصائيِّ ويرمز لعدد وحدات العينة بالرمز n (حجم العينة) .

أسلوب المعاينة(Sampling Techniques):

هي عمليةُ اختيارِ عددٍ من واحداتِ المجتمعِ و إخضاعها للعملِ الإحصائيِّ، بحيثُ تكونُ النتائجُ التي تمَّ التوصُّلُ إليها بناءً على معطياتِ العينةِ تمثِّل مؤشِّراتِ المجتمعِ المرادُ تقديرُها .

تصميم العينة(Sample Design):

هي عمليةُ اختيارِ التَّركيبِ المناسبِ من عدَّةِ أنواعٍ من العيناتِ للوصولِ إلى العينةِ التي تُحقِّقُ النتائجَ المرجوَّةَ منها.

حجم المجتمع(Population Size):

يقصدُ بحجمِ المجتمعِ عددُ جميعِ وحداتِ المعاينةِ الَّتي يتكوَّنُ منها المجتمعُ ويُرمَزُ له عادةً بالرَّمز N.

حجم العينة(Sample Size):

هو عددُ وحداتِ المعاينةِ التي تمَّ اختيارها ويرمزُ لهُ عادةً بالرَّمز n ، ويعتبَرُ حجمُ العينةِ صغيراً إذا كانَ أقلَّ من 30 .

التباين(Variance):

يعدُّ التَّباينُ والانحرافُ المعياريُّ من أهمِّ مقاييسِ الانتشارِ أو التَّشتُّتِ ، التي تقيسُ مدى انتشارِ القيمِ عن بعضهَا أو عنْ قيمةٍ معينةٍ. ويعدُّ التَّباينُ أحدَ المقاييسِ التي تستخدمُ لقياسِ مدى تشتت القيمِ عن الوسطِ الحسابي ، إذ كلما كانتْ القيمُ بعيدةً عنهُ كانَ التباينُ أكبر ، والتباينُ هو عبارةٌ عن مجموعِ مربَّعاتِ انحرافاتِ القيمِ عن وسطِها الحسابي مقسوماً على عددِهَا ، ويمكننا التَّمييزُ بينَ تباينِ المجتمعِ (σ^2) وتباينِ العينةِ (S^2).

معامل ارتباط بيرسون: Person Correlation Coefficient:

هو مؤشر إحصائي يقيس قوة وجهة الارتباط بين متغيرين ، حيث تتراوح قيمته بين $(+1)$ و (-1) ، ويرمز له ب (r) ، وتدل إشارة r الموجبة على العلاقة الطردية ، بينما تدل إشارة r السالبة على العلاقة العكسية بين المتغيرات، وكلما اقتربت |r| من الواحد كان الارتباط بين المتغيرين قويا.

معامل التحديد: Multiple Coefficient of Determination:

ويعد مؤشر أساس في تقييم مدى معنوية العلاقة بين المتغير التابع (Y) والمتغيرات المستقلة (X_K) إذ (n، … ، 1 = k) ، بعبارة أخرى هو مقياس يوضح نسبة مساهمة المتغيرات المستقلة في تفسير التغير الحاصل في المتغير التابع.

النتائج والمناقشة:

جمع وتحليل بيانات المجتمع الكلي:

حصلنا على بيانات تجريبية لمجتمع مؤلف من 2000عنصر، يمثل نتائج الاختبارات الشفهية والكتابية لمادة الرياضيات لطلاب عدة مدارس مختارة عشوائيا. وقبل البدء بتحليل بيانات المجتمع، قمنا بإجراء اختبارات جودة توفيق نموذج الانحدار بواسطة برنامج SPSS (كما يظهر في الشكل 1) وجدنا أن قيمة معامل ارتباط بيرسون هي $r = 0.933$ وهذا يدل على وجود ارتباط قوي بين المتغيرين. وبلغت قيمة معامل التحديد: $R^2 = 0.871$ أي أن 87% من البيانات والانحرافات الكلية في قيم المتغير Y تفسرها العلاقة الخطية. ومنه يمكننا القول أن النموذج جيد التوفيق **يمثل علاقة خطية يمكن تعميم نتائجها.**

وبإجراء تحليل الانحدار على المجتمع نحصل على معادلة الانحدار التالية:

$$\hat{Y} = 0.115 + 0.987\,x \qquad (1)$$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.933[a]	.871	.871	1.191

a. Predictors: (Constant), x
b. Dependent Variable: y

الشكل (1)

تحديد الحجم المناسب للعينة:

قمنا بسحب **عينات عشوائية بسيطة** بحجوم **مختلفة** من المجتمع وتقدير معاملات معادلة الانحدار الخطي بطريقة المربعات الصغرى مستخدمين برنامج SPSS لكل عينة من العينات فحصلنا على النتائج المبينة في الجدول التالي:

جدول رقم (1) حساب تقديرات معاملات معادلة الانحدار عند حجوم مختلفة لعينات عشوائية بسيطة

حجم العينة	β_0	β_1	Std. β_0	Std. β_1	Std. x	Std. y	R^2	$\bar{\delta}$
50	2.17	0.901	1.321	0.058	2.852	2.815	0.833	1.0705
75	1.671	1.054	0.877	0.039	3.368	3.723	0.91	0.8116
100	1.28	0.933	0.901	0.04	3.107	3.152	0.846	0.6095
125	1.084	0.943	0.725	0.032	3.14	3.176	0.873	0.5065
150	0.653	0.966	0.643	0.028	3.514	3.602	0.887	0.2795
175	0.559	0.973	0.764	0.034	2.916	3.114	0.829	0.229
200	0.481	1.018	0.616	0.027	3.259	3.548	0.872	0.1985
225	0.413	0.975	0.531	0.023	3.435	3.558	0.886	0.155
250	0.371	0.974	0.526	0.023	3.214	3.338	0.937	0.1345
275	0.214	1.005	0.522	0.023	3.099	3.336	0.872	0.0585

300	0.21	1	0.515	0.023	2.997	3.22	0.865	0.054
325	0.191	0.98	0.475	0.021	3.262	3.429	0.877	0.0415
350	0.196	1.003	0.464	0.021	3.064	3.296	0.87	0.0485
375	0.199	0.984	0.444	0.02	3.155	3.324	0.872	0.0435
400	0.195	1.002	0.41	0.018	3.264	3.481	0.883	0.0475
600	0.199	0.98	0.355	0.016	3.055	3.217	0.866	0.0455

حيث:

$Std.\beta_0$: الخطأ المعياري للمعامل β_0

$Std.\beta_1$: الخطأ المعياري للمعامل β_1

$Std.x$: الخطأ المعياري للمتغير x والممثل لنتيجة الاختبار الشفهي.

$Std.y$: الخطأ المعياري للمتغير y والممثل لنتيجة الاختبار الكتابي.

R^2: معامل التحديد.

$\overline{\delta}$: متوسط الخطأ ويحسب كما يلي:

$$\delta\beta_0 = |\beta_0 - 0.115|$$
$$\delta\beta_1 = |\beta_1 - 0.987|$$
$$\overline{\delta} = \frac{\delta\beta_0 + \delta\beta_1}{2}$$

النتيجة:

بقراءة النتائج المدونة في الجدول السابق رقم (1) والتي تمثل القيم الوسطية لتكرار سحب 4 عينات عشوائية بسيطة بحجوم تجريبية مختلفة من (15 حجما) حيث بدأنا بالحجم $n = 50$ حتى الحجم $n = 400$ بزيادة 25 في كل مرحلة وتكرار سحب أربع عينات من كل حجم وتحليل نتائجها ثم سحب أربع عينات من الحجم $n = 600$ لتأكيد النتائج وأخذ نتائج العينة التي تعطي أكبر خطأ في كل حجم نستنتج مايلي:

1- إن قيمة كل من معاملي الانحدار الخطي البسيط $\beta_0 \& \beta_1$ المأخوذة من العينة العشوائية البسيطة لتقدير معاملي الانحدار في مجتمع حجمه $N = 2000$ ، تقترب من معالم المجتمع **بزيادة حجم العينة حتى الحجم $\boldsymbol{n = 325}$** ، حيث أن أي زيادة في الحجم بعد هذه القيمة لا يعطي أي تحسن ملحوظ في دقة التقديرات ، وبالتالي فإن الحجم الملائم للعينة هو $n = 325$

وبخطأ مقداره: $\delta\beta_0 = 0.076$

$\delta\beta_1 = 0.007$

ومتوسط خطأ مقداره: $\overline{\delta} = 0.0415$

2- إن الانحراف المعياري لكل من المتغيرين $x \& y$ لا يبدي أي تغير ملحوظ بتغير حجم العينة ، وهذا يرد إلى تجانس بيانات المجتمع وجودة توفيق النموذج بعيدا عن وجود القيم المتطرفة.

اختيار طريقة المعاينة الإحصائية:

حاول البحث أيضاً تحديد نوع العينة الأمثل والتي تعطينا الخطأ الأقل في تقديرات المربعات الصغرى لمعادلة الانحدار، قمنا بإجراء التحليل الإحصائي لعينات مسحوبة بطرق المعاينة العشوائية الثلاث : البسيطة، الطبقية ثم العنقودية مع ثبات العوامل الأخرى: الحجم $n = 325$ وتوزيع المجتمع الذي سحبت منه العينات (توزيع طبيعي). ثم قارنا بين معادلات الانحدار الناتجة.

المعاينة العشوائية البسيطة:

تعريف المعاينة العشوائية البسيطة: [5]،[7]

هي التَّصميمُ الَّذي يتساوى فيه احتمالُ انتقاءِ أيٍّ من العيناتِ ذاتِ الحجم n الممكنةِ التَّشكيلِ من مجتمعٍ مؤلفٍ من N عنصر بشرط أن يكونَ المجتمعُ متجانساً (أي أنَّ الفروقَ المتعلقةَ بالخاصةِ المدروسةِ في عناصرِ هذا المجتمعِ طفيفةٌ)، وتعتمدُ المعاينةُ العشوائيّةُ البسيطةُ على نظريَّةِ الاحتمالاتِ في اختيارِ وحداتِها وتقديرِ ثوابتها.

■ وبسحب عينة عشوائية بسيطة بحجم $n = 325$ وإجراء تحليل الانحدار عليها حصلنا على معادلة الانحدار التالية:

$$\hat{Y} = 0.191 + 0.98x$$

وبأخطاء معيارية: $\delta\beta_0 = 0.076$

$$\delta\beta_1 = 0.007$$

ومتوسط خطأ مقداره: $\overline{\delta} = 0.0415$

المعاينة العشوائية الطبقية:

تعريف المعاينة العشوائية الطبقية: [5]،[1]

عملية اختيار عدد من الوحدات من مجتمع مقسم إلى طبقات (بحيث تكون الطبقات غير متداخلة، وتكون المفردات ضمن الطبقة الواحدة متجانسة، بينما هناك فروق كبيرة بين الطبقات)، ويتم اختيار عينة عشوائية من كل طبقة، بحيث يكون السحب من الطبقات المختلفة مستقلا، ومجموع العينات المختارة من الطبقات تشكل العينة الطبقية العشوائية، وذلك للوصول إلى خصائص المجتمع من بيانات هذه العينة، إننا نعد كل طبقة مجتمعا صغيرا، تسحب منه عشوائيا عينة ذات حجم محدد، ونقوم بتقدير معالم المجتمع لكل طبقة على حدة، ثم تستخدم هذه التقديرات لتقدير معالم المجتمع كله.

رموز وتعاريف : [5]، [1]

N :عدد وحدات المجتمع.

L: عدد الطّبقات التي ينقسمُ إليها المجتمع.

N_h : عدد وحدات الطَّبقة ذاتِ الرُّتبة(h).

n: حجمُ العينة الطبقية.

n_h : حجمُ العينةِ المسحوبةِ من الطَّبقة ذاتِ الرُّتبة (h).

توزيع العينة على مختلف الطبقات :

تتوقَّفُ المؤشّراتُ الإحصائيّةُ المدروسةُ في المجتمعِ في المعاينةِ الطبقيَّةِ على توزيعِ العينةِ الإحصائيّةِ على الطَّبقات.

ولتوزيعِ العينةِ الكُلّيَّة n إلى عيناتٍ طبقيَّةٍ حجمُها n_h تُستَخدمُ عِدَّةُ طرقٍ لهذا التَّوزيع منها.

- **التوزيع المُتساوي:** يتمّ توزيعُ حجمِ العيّنة الكلية على مختلفِ الطّبقات بشكلٍ متساوٍ ، أي أنَّ أحجامَ جميعِ الطّبقات متساوية ، أي : $n_1 = n_2 = n_3 = ... = n_L$

ويساوي حجم كل طبقة : $n_h = \frac{n}{L}$

- **التوزيع المتناسب** : وهو يَشترِطُ أن نُوزِّع حجمَ العينة الكلية n على الطبقاتِ المختلفةِ بحيثُ يكونُ حجمُ العيِّناتِ الطبقيةِ التي نَسحَبُها من هذهِ الطبقاتِ متناسبةً مع حجمِ تلك الطبقات أي أنَّ التَّوزيعَ المتناسبَ مع حجومِ العيناتِ n_h يجبُ أن يُحقِّقَ الشَّرطَ التالي :

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \ldots = \frac{\mathrm{n}}{\mathrm{N}}$$

- **آلية توزيع العينة على الطبقات ومناقشة النتائج:**

1- قمنا في بحثنا كخطوة أولى بتقسيم المجتمع إلى 5 طبقات حسب المناطق الجغرافية (حسب المحافظات التي اخترنا منها المدارس) حيث حجم كل طبقة:

$$N_h = 400 \qquad ; h = 1،2، ....،5$$

وقمنا بتوزيع العينة الطبقية على الطبقات حسب التوزيع المتساوي حيث:

$$n_h = 65 \qquad ; h = 1،2، ....،5$$

وعن طريق برنامج SPSS سحبنا العينات من كل طبقة بطريقة المعاينة العشوائية البسيطة (لأن كل طبقة أصبحت مجتمع متجانس)، ثم شكلنا العينة الطبقية بحجم $n = 325$ وأجرينا تحليل الانحدار عليها فحصلنا على معادلة الانحدار التالية:

$$\hat{Y} = 0.12 + 1.003\,x$$

وبأخطاء معيارية: $\delta\beta_0 = 0.005$

$\delta\beta_1 = 0.016$

ومتوسط خطأ مقداره: $\overline{\delta} = 0.0105$

وبما أن الطبقات متساوية الحجم فإن التوزيع المتساوي هو نفسه التوزيع المتناسب مع حجوم الفئات.

2- كخطوة ثانية في دراسة العينة الطبقية، قمنا بتقسيم المجتمع إلى ثلاث طبقات، حسب تقدير درجات الطلاب (وسط، جيد، ممتاز)، حيث تكون بيانات كل طبقة متقاربة وبالتالي يقل تشتت البيانات.

فكانت الطبقة الأولى والتي تمثل تقدير "وسط" تضم الطلاب الذين حصلوا على درجات أقل أو تساوي 20 في الاختبار الكتابي والممثل للمتغير الأول وحجم هذه الطبقة $N_1 = 624$

والطبقة الثانية التي تمثل تقدير "جيد" تضم الطلاب الذين حصلو على درجات بين 20 و 25 في الاختبار الكتابي وحجم هذه الطبقة $N_2 = 1064$

والطبقة الثالثة التي تمثل تقدير "ممتاز"، تضم الطلاب الذين حصلو على درجات أكبر من 25 في الاختبار الكتابي وحجم هذه الطبقة $N_3 = 312$

باستخدام التوزيع المتساوي قمنا بتوزيع العينة الطبقية $n = 325$ على الطبقات الثلاث كما يلي:

$$n_1 = n_2 = n_3 = \frac{n}{3} \cong 108$$

وبسحب العينات من كل طبقة بالحجم السابق بطريقة المعاينة العشوائية البسيطة، وتشكيل العينة الطبقية ذات الحجم $n \cong 325$ ، وإجراء تحليل الانحدار عليها نحصل على معادلة الانحدار التالية:

$$\hat{Y} = 0.11 + 1.002\,x$$

وبأخطاء معيارية: $\delta\beta_0 = 0.005$

$$\delta\beta_1 = 0.015$$

ومتوسط خطأ مقداره: $\bar{\delta} = 0.01$

وزعنا العينة الطبقية على الطبقات الثلاث وفق التوزيع المتناسب مع حجوم الفئات كما يلي:

$$\frac{n}{N} = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3}$$

أي:

$$\frac{325}{2000} = \frac{n_1}{624} = \frac{n_2}{1064} = \frac{n_3}{312}$$

ومنه: $n_1 = 101$ ، $n_2 = 173$ ، $n_3 = 51$

قمنا بسحب العينات من الطبقات بطريقة المعاينة العشوائية البسيطة، وشكلنا العينة الطبقية بحجم $n = 325$، وأجرينا تحليل الانحدار على البيانات فحصلنا على معادلة الانحدار التالية:

$$\hat{Y} = 0.116 + 1.002\,x$$

وبأخطاء معيارية: $\delta\beta_0 = 0.001$ $\delta\beta_1 = 0.015$

ومتوسط خطأ مقداره: $\bar{\delta} = 0.008$

المعاينة العشوائية العنقودية: [5]،[7]

يقومُ أسلوبُ المعاينةِ العنقوديَّةِ على مبدأِ تقسيمِ المجتمعِ إلى مجموعاتٍ بشكلٍ مناسبٍ بحيثُ تكونُ هذهِ المجموعاتُ متقاربةٌ بالحجمِ ومتجانسةٌ بالنسبةِ للصفةِ المدروسةِ حيثُ كلُّ مجموعةٍ منَ المجموعات تُسمَّى عُنقود، وتشكِّلُ العناقيدُ المجتمعَ دونَ حذفٍ أو تكرارٍ، وتعد تكلفة المعاينة العنقودية أقل من تكلفة المعاينة العشوائية البسيطة، أو المعاينة الطبقية، واستخدام هذا النوع من المعاينات، يؤدي إلى توفيرُ التكاليف بسببِ عدم وجودِ مسافاتٍ كبيرة بين وحداتِ العيّنة، لأنها تقعُ بجانبِ بعضها ضمن العنقودِ الواحد، الذي يتمّ حصرُ جميعِ وحداتِه حصراً شاملاً.

وهكذا يمكننا القولُ أنّه يمكن استخدامُ المعاينةِ العنقوديةِ البسيطة بشكلٍ مناسبٍ للحصولِ على البيانات بأقلّ تكلفة في الحالتين التاليتين :

- عندما يكونُ إطار الوحداتِ الإحصائيةِ الّذي يتضمّن أسماءَها وعناوينَها غيرَ متوافر أو أنّ إعدادَه يتطلّب نفقاتٍ ضخمة.

- ضخامة نفقات الحصول على البيانات من الوحدات نتيجة انتشار الوحدات ووجود مسافات كبيرة بينها.

وتتمُّ هذه المعاينة كما يلي:

ليَكُن لدينا مجتمع U و مؤلَّف من N عنصر ولإجراءِ معاينةٍ عشوائيةٍ عنقوديةٍ على هذا المجتمع نتَّبعُ الخطوات التالية :

1. نقسّم هذا المجتمع إلى فئاتٍ مختلفةٍ وغيرَ متقاطعة وليكُن عدَدُها M وسنرمزُ لهذه الفئات بـ G_k (k:1،2،3،.....،M)

2. نسحبُ بطريقةِ المعاينةِ العشوائيَّةِ البسيطة من الفئاتِ G_k عيِّنةً من الفئاتِ بحجم وليكن m وبذلك نحصلُ على عينةٍ من الفئاتِ نُسمِّيهَا العينةَ الفِئَوِيَّة ونرمُز له بـ g ونسمي واحداتِها بواحِداتِ العينةِ الفئويةِ و نرمُز لها بـg_j حيث أن j:1،2 ، ....،(m).

3. بما أنَّ كلًّا من واحداتِ العينة الفئوية (أي g_j) يحتوي على جملةٍ من واحداتِ المجتمعِ الأوليَّةِ (العناصر) . لذلكَ نبحثُ عن عددِ عناصرِ المجتمعِ الموجودةِ في كلٍّ من هذه الفئات ، ولنرمُز لعددِ عناصرِ المجتمع الموجودة في الفئة g_j بـ N_j حيث j:1،2،....،(m) وبذلك نرتبُ الجدول التالي:

اسم الفئة في العينة الفئوية	g_1	g_2	g_3	....	g_{m-1}	g_m
عدد عناصر المجتمع في الفئة	N_1	N_2	N_3	....	N_{m-1}	N_m

4. نسحبُ بطريقةِ المعاينةِ العشوائيةِ البسيطةِ أيضاً من كلِّ فئاتِ العينةِ الفئويَّة (أي من كلٍ من g_j) عينة من الواحداتِ الأوليَّة بحجمٍ وليكن n_j ولنرمز لها بالرمز J فنحصل على جملةٍ من العيِّنات نسميها بالعينات الأوليَّة عددُها يساوي m (عدد الفئات) وأنَّ جملةَ هذه العينات تشكِّلُ عينةً من الواحداتِ الأوليةِ نُسمِّيها بالعينةِ الكليةِ ويكونُ حجمُهَا مساوياً لمجموعِ حجومِ العيناتِ الأوليةِ أي أن :

$$n = \sum_{j=1}^{m} n_j$$

وهنا نلاحظ أنَّنا نحتاجُ إلى ترقيمِ عناصرِ المجتمعِ الموجودةِ في الفئاتِ المسحوبةِ في العينةِ الفئويّة فقط .

نقومُ بتسجيلِ المعلوماتِ اللّازمة عن كلّ واحدةٍ من واحداتِ العيناتِ الأوليَّةِ (واحدات العينة الكلية) ثم نُجري الحساباتِ اللّازمة لتقديرِ معالمِ المجتمع.

و عند استخدام هذه المعاينة يجب مراعاة ما يلي:

1. أن يكونَ حجمُ العنقودِ صغيراً وعددُ العناقيدِ كبيراً .

2. عندَ تكوينِ العناقيد تُؤخَذ مفرداتُ المجتمعِ المتجاورة أو ضِمنَ منطقةٍ معيَّنة حيثُ تكونُ غالباً متشابهة للصفة المدروسة

3. أن تكونَ أحجامُ العناقيد متقاربة قَدرَ الإمكان .

4. يجب أن يكون كل عنقود موضحا ومعرفا لجامع البيانات.

■ قمنا بتقسيم المجتمع الممثل ل 2000 عنصر إلى عناقيد حسب المدارس، حيث يمثل كل عنقود مدرسة مشمولة بالبحث، فكان لدينا 20 عنقود، كل منها يحوي عدد من واحدات المجتمع، وعن طريق المعاينة العشوائية البسيطة قمنا بسحب 5 عناقيد، أي أن العينة الفئوية المسحوبة بحجم 5 $m = 5$ أي أن

$$N_h = 100 \quad ; \quad h = 1،2، \ldots ،5$$

ثم سحبنا من كل فئة من الفئات الخمس بطريقة المعاينة العشوائية البسيطة عينة أولية بحجم :

$$n_j = 65 \quad ; \quad j = 1،2، \ldots. ،5$$

ثم شكلنا العينة الكلية من جملة العينات الأولية بحجم: $n = 325$ وقمنا بتحليلها فحصلنا على معادلة الانحدار التالية:

$$\hat{Y} = 0.3 + 0.974\, x$$

وبأخطاء معيارية: $\delta\beta_0 = 0.185$, $\delta\beta_1 = 0.013$ ومتوسط خطأ مقداره: $\overline{\delta} = 0.099$

قمنا لتأكيد النتيجة، بإعادة سحب العينة الفئوية بحجم $m = 5$ ثم سحبنا عينات أولية من الفئات الجديدة ، وشكلنا عينة كلية جديدة ، وبعد تحليلها حصلنا على معادلة الانحدار التالية:

$$\hat{Y} = 0.597 + 0.959\, x$$

وبأخطاء معيارية: $\delta\beta_1 = 0.028$ ، $\delta\beta_0 = 0.482$

ومتوسط خطأ مقداره: $\overline{\delta} = 0.255$

ونلخص نتائج أساليب المعاينة العشوائية الثلاث بالجدول التالي:

جدول رقم (2) نتائج تقديرات معاملات معادلة خط الانحدار في حال المعاينات المختلفة

نوع المعاينة	β_0	β_1	Std. β_0	Std. β_1	Std. x	Std. y	R^2	$\overline{\delta}$
بسيطة	0.191	0.98	0.475	0.021	3.262	3.429	0.87	0.0415
طبقية موزعة جغرافيا بالتوزيع المتساوي	0.12	1.003	0.388	0.017	3.796	3.977	0.918	0.0105
طبقية موزعة حسب تقدير الدرجات بالتوزيع المتساوي	0.11	1.002	0.425	0.019	3.238	3.426	0.896	0.01
طبقية موزعة حسب تقدير الدرجات بالتوزيع المتناسب مع حجوم الفئات	0.116	1.002	0.425	0.019	3.242	3.432	0.896	0.008
العنقودية الأولى	0.3	0.974	0.465	0.021	3.116	3.254	0.87	0.099
العنقودية الثانية	0.597	0.959	0.471	0.021	3.105	3.202	0.865	0.255

النتيجة:

1) إن العينة العشوائية البسيطة تعطي نتيجة لا تتمتع بدقة كافية للتقدير، ذلك أن عمليات السحب فيها تتمتع بالعشوائية التامة.

2) إن العينة العشوائية الطبقية تعطي النتائج الأفضل والأدق لتقدير معاملات الانحدار الخطي البسيط بطريقة المربعات الصغرى، ذلك لأن الخطأ المعياري للعينة يتأثر بشكل عام بتشتت مفردات المجتمع الذي سحبت منه، لذا نجد أن الخطأ المعياري للعينة الطبقية أقل من الخطأ المعياري للعينة العشوائية البسيطة، نتيجة لإزالة قسم من تشتت المجتمع الإحصائي بإلغاء الاختلافات الكبيرة الموجودة ضمن الطبقة الواحدة.

وعندما نقوم بتصميم العينة وفق التوزيع المتناسب مع حجوم الفئات، نحصل على أفضل نتيجة معاينة وأقل خطأ معياري وأكبر عدد ممكن من البيانات على خط الانحدار.

3) إن نتائج المعاينة العشوائية العنقودية أقل دقة من المعاينة العشوائية البسيطة والطبقية، ويرد ذلك إلى أن العينة الكلية الممثلة للمجتمع العنقودي تسحب من فئات معينة من المجتمع، ورغم أن السحب يكون عشوائي بسيط، إلا أن عدد من الفئات (العناقيد) لا تدخل في التحليل، وبالتالي فإن بياناتها لا تكون مأخوذة في الحسبان عند تقدير معاملات معادلة الانحدار.

الاستنتاجات والتوصيات:

إن أهم النتائج التي توصلنا إليها:

- إن الحجم الأمثل لاختيار عينة لتقدير معاملات معادلة الانحدار الخطي البسيط بطريقة المربعات الصغرى هو

$n = 325$ وذلك لمجتمع ذو توفيق جيد وملائم للتحليل ومؤلف من 2000 عنصر، أي بنسبة 16.25% من حجم المجتمع الكلي.

- إن تصميم العينة بطريقة المعاينة العشوائية الطبقية ذات التوزيع المتناسب مع حجوم الفئات يعطي النتائج الأفضل والاكثر دقة لتقدير معادلة الانحدار الخطي البسيط بطريقة المربعات الصغرى.

ونوصي بالاستفادة من هذه النتائج في محاولة تقليل حجم العينة المطلوب عند توزيع العينة الطبقية توزيعا متناسبا مع حجوم الفئات.

المراجع

[1] أبو شعر، عبد الرزاق. العينات وتطبيقاتها في البحوث الاجتماعية. الطبعة الأولى، الإدارة العامة للبحوث، السعودية. 1997.

[2] ربيع، عبد الحميد؛ سمرة، عادل؛ الصياد، جلال. (2008): مقدمة في الإحصاء لطلاب الدراسات الاقتصادية والإدارية. الطبعة الأولى، جامعة الملك عبد العزيز، السعودية. 2015.

[3] شادري، محمد ابراهيم أحمد، 2012م، "تأثير حجم العينة على قوة الاختبار الإحصائي"، السعودية، جامعة أم القرى، كلية التربية، قسم علم النفس.

[4] عباسي، عبد الحميد. (2012): تشخيص الانحدار (مشاكله وعلاجها) ، تطبيقات في العلوم الاجتماعية. . الطبعة الثالثة، مركز الدراسات والاستثمارات الإحصائية والقياسية جامعة القاهرة، مصر. SPSS باستخدام 2012.

[5] علي، إبراهيم محمد. مدخل في نظرية العينات. الطبعة الأولى، منشورات جامعة حلب، سوريا. 1980.

[6] غامدي، عبد اللطيف محمد، 2000م ، "أثر أسلوب اختيار العينة وحجمها على دقة تقدير معالم المجتمع الإحصائي"، السعودية، جامعة أم القرى، قسم الإحصاء والبحوث.

[7] كنجو، أنيس. (1995): تقنية المعاينة الإحصائية (تأليف ويليام كوكران). الطبعة الأولى، قسم الإحصاء وبحوث العمليات كلية العلوم جامعة الملك سعود، السعودية. 1995.

[8] كاظم، صفاء كريم، 2009م ، "المقارنة بين تقديرات معالم نموذج الانحدار الخطي المتعدد باستخدام أسلوب OLS وأسلوب برمجة الأهداف الخطية"، العراق، جامعة المثنى، كلية العلوم، قسم الرياضيات وتطبيقات الحاسوب.

[9] معجم المصطلحات الإحصائية. (2015): المعهد العربي للتدريب والبحوث الإحصائية عام 2015 .

[10] Abdi, H, “The Method of Least Squares”, University of Texas, Dallas, USA, Neil Salkind, 2007.

[11] Park, M, “Regression Estimation of The Mean In Survey Sampling”, USA, IOWA State University, 2012.

[12] Sen, P.K, “Estimates of the Regression Coefficient Based on Kendall's Tau”, University of North Carolina, Chapil Hill, USA, 2013.

[13] WASSERMAN, W.D; KUTNER, M.S. Applied Liner Statistical Models: Regression ،Analysis of Variance and Experimental Designs. 3rd. ed, McGraw Hill, Inc U.S.A, 1990.

[14] Xitao, F. Thompson, B. Lin, W, “The Effects of Sample Size, Estimation Methods, and Model, Specification On Structural Equation Modeling Fit Indxes”, USA, Routledge, 1999.