جامعة السودان للعلوم والتكنولوجيا
كلية الدراسات العليا
(كلية العلوم)

الطرائق التقريبية لإيجاد حدود الطبقات في المعاينة العشوائية الطبقية

رسالة مقدمة إلى قسم الإحصاء التطبيقي-كلية العلوم/جامعة السودان للعلوم والتكنولوجيا لنيل درجة الماجستير في علوم الإحصاء

من قبل
صفاء دفع الله محمد

بإشراف
الدكتور بسام يونس إبراهيم
أستاذ العمليات التصادفية المشارك

رجب 1424هـ سبتمبر 2003م

﴿ وَأَناَّ مِنَّا الصَّلَحُونَ وَ مِنَّا دونَ ذَلِكَ كُنَّا
طَرآئِقَ قِدَدًا﴾

الآية (11) من سورة الجن

الإهداء

إلى التي علّمتني حب الأهل والعشيرة والوطن وكان دعاؤها سر نجاحي
والدتي الحبيبة التي خاضت روحها إلى بارئها
إلى الذي غرس في نفسيّ حبّ العلم والعلماء
والدي العزيز المفضال
إلى من يعمل لأجل رفع راية العلم والمعرفة
أستاذي الجليل الدكتور بسام يونس إبراهيم
إلى من أمضى معي مسيرة هذا الدرب الوعر فكانوا خير معين وأنا أتسلق جبال هموم العلم والمعرفة
أخواني وأخواتي الأعزاء
إلى الشموع التي تحترق لتضيء لنا الطريق
أساتذتي الإجلاء
إلى من كن خير الأحبة والأصحاب
صديقاتي وزميلاتي

إليكم جميعاً أهدي هذا الجهد المتواضع

الباحثة

الشكر والعرفان

الحمد والشكر لله الأكرم الذي عّلمَ بالقلم عّلمَ الإنسان ما لم يعلم.

والشكر أجزله والحمد نسوقه بعد المولى عز وجل إلى إنسان جاد بعلمه وما بخل، وعمل بهمة، ليل نهار في تعميق رسالة العلم بلا كلل ولا تواني بالرغم من زخم الحياة.

الشكر أسوقه إلى أخ العروبة والإسلام أبن الرافدين أرض الحضارة ومهد العلوم.

الشكر آيات أزفها إلى عالم تفانى وجاد لنا بعلمه في غزارة وتواصل يتحدى النازلات، ولا ينقطع لدى المعضلات حتى صار هذا البحث واقعاً ملموساً.

ذلك العالم هو أستاذنا الجليل الدكتور بسام يونس إبراهيم أستاذ العمليات التصادفية المشارك بقسم الإحصاء التطبيقي-كلية العلوم/جامعة السودان للعلوم والتكنولوجيا.

كما أخص بالشكر كل الذين همسوا في أذني لرفع شأني وشأن علمي ووقفوا بجانبي في إخراج هذا البحث.

وأخيراً أرجو أن أكون قد أضفت شيئاً ولو قليلاً من وراء هذا الجهد المتواضع، فإن أصبت فمن الله وإن أخطأت فمن نفسي، ولا أدعي بكمال هذا العمل فالكمالُ لله وحده وهو ولي التوفيق .

الباحثة

الملخص

يعد موضوع العينات أحد فروع الإحصاء التطبيقي من المواضيع المهمة في البحث العلمي، حيث تنقسم البحوث الإحصائية من حيث درجة الشمول إلى بحوث بطريقة الحصر الشامل وأخرى بطريقة العينات. تستخدم طريقة الحصر الشامل عندما يكون الباحث غير ملم بطبيعة مفردات المجتمع تحت الدراسة إذ أنه لا يستطيع اختيار عينة تصلح لتمثيل هذا المجتمع. أما البحث بطريقة العينات فهو المستخدم في حالة كون الباحث يملك كمية من المعلومات عن المجتمع تساعده على اختيار العينة المناسبة التي تمثل ذلك المجتمع تمثيلاً جيداً، أو في حالة تعذر أو استحالة الحصر الشامل لأسباب مختلفة.

إن طريقة أسلوب المعاينة معروفة منذ زمن بعيد يستخدمها الشخص العادي في حياته اليومية من دون أن ينتبه إلى كونها طريقة عملية تؤدي إلى استنتاجات سليمة إذا بنيت على أسس صحيحة.

ولكي يكون حكمنا على الكل دقيقاً وصحيحاً وجب علينا أن نهتم بالطريقة التي نختار بها هذا الجزء حتى نحصل على النتيجة الدقيقة، وعليه يجب أن نختار طريقة للمعاينة بحيث تكون خواص المجتمع بما فيها الاختلاف بين المفردات منعكسة في العينة بأحسن ما يسمح به حجم العينة. وهكذا نرى أن اختيار العينات ليس العينات مجرد استخدام جزء من المجتمع بدلاً منه فحسب، وإنما أسلوب يستند على قواعد مستمدة من النظرية الإحصائية والتي تعتمد على نظرية الاحتمالات وعلى قواعد رياضية كثيرة، لذلك أصبحت نظريات العينات أساس كثير من الدراسات النظرية والعملية.

إن إحدى طرائق سحب العينة من المجتمع هي طريقة المعاينة العشوائية الطبقية والتي تستخدم في حالة المجتمعات غير المتجانسة. فإذا قسم المجتمع الذي له توزيع بدالة كثافة احتمالية $f(y)$ معرفة ضمن الفترة $[a,b]$ ، إلى L من الطبقات بالحدود $a = y_0, y_1, y_2, \dots, y_L = b$ فإنه يتوجب إيجاد $L-1$ من هذه الحدود بحيث $y_1 < y_2 < \cdots < y_{L-1}$.

لقد وجد Dalenius & Hodges (1959) أن تباين الوسط الحسابي الطبقي $V_{Ney}(\bar{y}_{st})$ هو أقل ما يمكن عند تحقق المعادلة:

$$\left[\frac{(y_h-\mu_h)^2+\sigma_h^2}{\sigma_h}\right]=\left[\frac{(y_h-\mu_{h+1})^2+\sigma_{h+1}^2}{\sigma_{h+1}}\right], h=1,2,\dots,L-1$$

وذلك في حالة كون متغير الدراسة هو نفسه متغير الطبقية.

أما في حالة استخدام المتغير المساعد فإن:

$$\frac{\beta^2(x_h-\mu_h)^2+(\beta^2\sigma_{x_h}^2+2\sigma^2)}{(\beta^2\sigma_{x_h}^2+\sigma^2)^{\frac{1}{2}}}=\frac{\beta^2(x_h-\mu_{h+1})^2+(\beta^2\sigma_{x_{h+1}}^2+2\sigma^2)}{(\beta^2\sigma_{x_{h+1}}^2+\sigma^2)^{\frac{1}{2}}}, h=1,2,\dots,L-1$$

حيث أن:

$$\mu_h = \frac{1}{W_h} \int_{y_{h-1}}^{y_h} y f(y) dy$$

يمثل متوسط الطبقة h ، وأن:

$$W_h = \int_{y_{h-1}}^{y_h} f(y) dy$$

يمثل وزن الطبقة h ، بينما تباين الطبقة h هو:

$$\sigma_h^2 = \frac{1}{W_h} \int_{y_{h-1}}^{y_h} y^2 f(y) dy - \mu_h^2$$

يلاحظ من المعادلات في أعلاه أن الحل المباشر لهما غير ممكن وذلك لأن y_h وكلاً من μ_h و σ_h^2 يعتمد حساب كلاً منهما على الآخر، فإن الحل التام لهذه المعادلات يتطلب طرائق تكرارية غالباً ما تكون معقدة، هذا ما تم تناوله في الفصل الثاني.

أما الفصل الثالث فقد تناول الطرائق التقريبية المقترحة سابقاً والطريقة التي تم اقتراحها في هذا البحث بالاعتماد على توزيع نيمان في حالة كون متغير الدراسة هو متغير الطبقية وفي حالة استخدام المتغير المساعد. إن هذه الطرائق التقريبية تعطي صيغ تقريبية لـ $V_{Ney}(\bar{y}_{st})$ ، حيث أن أول طريقة تقريبية لحالة متغير الدراسة هو متغير الطبقية هي طريقة $Cum\, f^{\frac{1}{2}}$ المقترحة من Dalenius & Hodges (1959) حيث وجدا فيها أن تباين الوسط الحسابي الطبقي هو:

$$V_{Ney}^{**}(\bar{y}_{st}) = \frac{K^4(y)}{12nL^2}$$

كذلك اقترح كل من القصاب والطائي (1994) طريقة $Cum\, f^{\frac{2}{3}}$ ووجدا أن تباين الوسط الحسابي الطبقي هو:

$$V_{Ney}^{***}(\bar{y}_{st}) = \frac{M^3(y)}{12nL^2}$$

أما في هذه الدراسة فقد اقترحت الباحثة الطريقة التقريبية $Cum\, f^{\frac{1}{3}}$ ووجدت أن تباين الوسط الحسابي الطبقي هو:

$$V_{Ney}^{*}(\bar{y}_{st}) = \frac{H^6(y)}{12nL^2}$$

أما في حالة المتغير المساعد، فقد اقترح القصاب والطائي (1994) طريقة $Cum\, f^{\frac{2}{3}}$ على فرض أن:

$$y = \alpha + \beta x + e$$

ووجد أن تباين الوسط الحسابي الطبقي هو:

$$V^{***}_{Ney}(\bar{y}_{st})=\frac{\sigma^2_y}{n}\left\{\frac{\rho^2 M^3(x)}{12L^2\sigma^2_x}+(1-\rho^2)M^3(x)\right\}$$

في حين توصل Setling (1968) إلى أن تباين الوسط الحسابي الطبقي في طريقة $Cum\,f^{\frac{1}{2}}$ هو:

$$V^{**}_{Ney}(\bar{y}_{st})==\frac{\sigma^2_y}{n}\left[\frac{\rho^2 K^4(x)}{12L^2\sigma^2_x}+(1-\rho^2)K^4(x)\right]$$

بينما توصلت الباحثة في هذه الدراسة إلى أن تباين الوسط الحسابي الطبقي في الطريقة المقترحة $Cum\,f^{\frac{1}{3}}$ هو:

$$V^{*}_{Ney}(\bar{y}_{st})=\frac{\sigma^2_y}{n}\left[\frac{\rho^2 H^6(x)}{12L^2\sigma^2_x}+(1-\rho^2)H^6(x)\right]$$

كما تناول الفصل الثالث المقارنات بين الطرائق التقريبية الثلاث اعتماداً على تباين الوسط الحسابي الطبقي عن طريق حساب نسب التباين بينها.

أما الفصل الرابع فقد تناول التطبيقات النظرية في حالة متغير الدراسة هو متغير الطبقية وفي حالة استخدام المتغير المساعد. تناول هذه التطبيقات حساب تباين الوسط الحسابي الطبقي للطرائق التقريبية الثلاث بالاعتماد على التوزيعات النظرية (المنتظم، الأسي، والطبيعي)، إذ تبين أن أقل تباين كان في طريقة $Cum\,f^{\frac{2}{3}}$ ثم يليه في طريقة $Cum\,f^{\frac{1}{2}}$ ، أما أكبر تباين فكان في طريقة $Cum\,f^{\frac{1}{3}}$ ، ومن هذا نجد أن طريقة $Cum\,f^{\frac{2}{3}}$ يمكن وصفها بأنها أكثر كفاءةً من الطرق الأخرى.

أما الفصل الخامس فقد تناول التطبيقات العملية للطرائق التقريبية الثلاث في دراسة قيم الصادرات والواردات المنقولة جواً عبر مطار الخرطوم الدولي للفترة (1970-1998) حيث تم الحصول على البيانات من الجهاز المركزي للإحصاء في وزارة المالية. وقد تم إثبات النتائج النظرية التي تم التوصل إليها في الفصل الرابع، حيث تم حساب تباين الوسط الحسابي الطبقي $V_{Ney}(\bar{y}_{st})$ للطرائق التقريبية الثلاث، وبعد إجراء المقارنة بين التباينات لوحظ أن التباين في طريقة $Cum\,f^{\frac{2}{3}}$ كان أقل تباين ثم يليه في طريقة $Cum\,f^{\frac{1}{2}}$ وأكبر تباين كان في طريقة $Cum\,f^{\frac{1}{3}}$ وكما موضح في الجدولين (5-3) و (5-4).

ABSTRACT

Sampling is a branch of Applied Statistics. It is the most important topic in the scientific research. Statistical researches are broadly divided into two: researches based on sampling and those which do not consider sampling. A research, which is not based on sampling, is used when a researcher is not familiar with the observations of population under study, he cannot choose a sample to represent this population. On the other hand a research which is based on samples is used when the researcher knows a lot of information about the population which is turn helps him to choose the right samples that can well represent the population.

The sampling method is known since the older days. A common man in his daily life uses it without knowing that it is scientific method and it may lead to correct conclusion if it is based on correct grounds.

In order for our reasoning to be precise and correct we should consider the method by which to choose this part. The sampling system used should reflect the different strata of the population. The choice of samples is not merely using a part of population rather than a sample. It is based on rules drawn from statistical theories that are based on probabilities theory and many mathematical rules. Therefore, the samples' theories have formed the basis of many practical and theoretical studies. One of the methods by which a sample can be drawn from a population is simple stratified sampling method which used in case of heterogeneous populations.

If the population which have the distribution with probability density function $f(y)$ *defined on* $[a,b]$ *, is divided into* L *stratums with boundaries* $a = y_0, y_1, y_2, \dots, y_L = b$ *, it must found* $L-1$ *from this*

boundaries such that $y_1 < y_2 < \cdots < y_{L-1}$.

In (1959) Dalenius & Hodges have found that $V_{Ney}(\bar{y}_{st})$ *is minimum when the following equation exists in case of stratum variable:*

$$\left[\frac{(y_h-\mu_h)^2+\sigma_h^2}{\sigma_h}\right]=\left[\frac{(y_h-\mu_{h+1})^2+\sigma_{h+1}^2}{\sigma_{h+1}}\right],\ h=1,2,\ldots,L-1$$

whereas in case of instrumental variables:

$$\frac{\beta^2(x_h-\mu_h)^2+(\beta^2\sigma_{x_h}^2+2\sigma^2)}{(\beta^2\sigma_{x_h}^2+\sigma^2)^{\frac{1}{2}}}=\frac{\beta^2(x_h-\mu_{h+1})^2+(\beta^2\sigma_{x_{h+1}}^2+2\sigma^2)}{(\beta^2\sigma_{x_{h+1}}^2+\sigma^2)^{\frac{1}{2}}},h=1,2,\ldots,L-1$$

such that:

$$\mu_h=\frac{1}{W_h}\int_{y_{h-1}}^{y_h} yf(y)dy$$

is the strata mean h, and

$$W_h=\int_{y_{h-1}}^{y_h} f(y)dy$$

is the weight of strata h, and the variance of strata h is:

$$\sigma_h^2=\frac{1}{W_h}\int_{y_{h-1}}^{y_h} y^2 f(y)dy-\mu_h^2$$

It is very difficult to solve above equations because y_h , μ_h *and* σ_h^2 *each of them depend on others, so we need the iteration methods as treated in chapter two.*

Chapter three consists the suggested approximation methods and the suggested method in this study depending on Neyman allocation in both cases: strata and instrumental variables. These methods are gives approximate formulas to $V_{Ney}(\bar{y}_{st})$ *, such that the first method in case of* *strata variable is* $Cum\,f^{\frac{1}{2}}$ *which suggested in (1959) by Dalenius & Hodges, they found that the variance of strata mean is:*

$$V^{**}_{Ney}(\bar{y}_{st})=\frac{K^4(y)}{12nL^2}$$

also Al-Qassab & Al-Taye in (1994) suggest that $Cum\,f^{\frac{2}{3}}$ *method, and they found that the variance of strata mean is:*

$$V^{***}_{Ney}(\bar{y}_{st})=\frac{M^3(y)}{12nL^2}$$

Whereas in this study the $Cum\,f^{\frac{1}{3}}$ *method is suggested and it is found that that the variance of strata mean in case of strata variable is:*

$$V^{*}_{Ney}(\bar{y}_{st})=\frac{H^6(y)}{12nL^2}$$

In case of instrumental variable, Al-Qassab & Al-Taye in (1994) suggest that $Cum\,f^{\frac{2}{3}}$ *by assuming that* $y=\alpha+\beta x+e$ *they found that the variance of strata mean is:*

$$V^{***}_{Ney}(\bar{y}_{st})=\frac{\sigma_y^2}{n}\left\{\frac{\rho^2 M^3(x)}{12L^2\sigma_x^2}+(1-\rho^2)M^3(x)\right\}$$

Selting in (1968) arrived to that the variance of strata mean in $Cum\,f^{\frac{1}{2}}$ *method is:*

$$V^{**}_{Ney}(\bar{y}_{st})==\frac{\sigma_y^2}{n}\left[\frac{\rho^2 K^4(x)}{12L^2\sigma_x^2}+(1-\rho^2)K^4(x)\right]$$

Whereas the researcher in this study arrived to that the variance of strata mean in $Cum\,f^{\frac{1}{3}}$ *method is:*

$$V^{*}_{Ney}(\bar{y}_{st}) = \frac{\sigma_y^2}{n}\left[\frac{\rho^2 H^6(x)}{12L^2\sigma_x^2} + (1-\rho^2)H^6(x)\right]$$

Chapter three also contains comparisons among the three approximate methods depending on the variance of strata mean by computing the variance ratios among them.

Chapter four consist the theoretical applications of the approximated methods in both cases strata and instrumental variables. These applications depend on the uniform, exponential and normal distributions it is show that the least variance is found in $Cum\, f^{\frac{2}{3}}$ *method and the maximum variance is found in* $Cum\, f^{\frac{1}{3}}$ *, so* $Cum\, f^{\frac{2}{3}}$ *method can be described as more efficient method than others.*

Chapter five consist the practical applications of the approximated methods according to the imports and exports of Sudan during (1970-1998). It is proved that the theoretical results which introduced in chapter four.

فهرس الجداول