نموذج مبنى تقديريا في مجتمع محدود المعاينة

أ. صباح هادي أ. سليم الغرابي م.د. إيمان محمد

كلية الادارة والاقتصاد- جامعة بغداد – قسم الاحصاء

المستخلص

المجتمع هو مجموعات من المفردات تشترك في صفة او صفات وتكون موضوع دراسة او بحث ويطلق على هذه المجموعات احصائيا مجتمع الدراسة (او اختصارا المجتمع **population**)
وقد يكون المجتمع مجموعة ما من البشر او اشجار انواع معينة من الفاكهة او الحيوانات او انتاج دولة ما لسلع معينة خلال فترة زمنية محددة ... الخ . والمجتمع قد يكون محدودا اذا كان يمكن حصر عدد افراده مثل سكان مدينة ما او طلاب مرحلة دراسية معينة وقد يكون المجتمع غير محدود (لانهائي) اذا كان لا يمكن حصر عدد افراده مثل النجوم او الكواكب او الكائنات الحية بمياه البحار . ولغرض دراسة صفات معينة لمجتمع ما فان البيانات الاحصائية تجمع بأحد اسلوبين هما اسلوب الحصر الشامل وفيه تجمع البيانات عن كل مفردة من مفردات المجتمع واسلوب المعاينة وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يطلق عليه عينة تحمل صفات المجتمع المسحوبة منه ثم يتم تعميم النتائج على المجتمع بأكمله .

يعرض هذا البحث تقدير بعض مقدرات المجتمع محدود المعاينة حيث تم استخدام تقدير معدل النموذج $u_{ns} = E_m(\bar{y}_{ns})$ والحصول على أفضل نموذج غير متحيز لمقدرات معدل مجتمع محدود باستخدام النتيجة التي استخدمها **Fuller**. كذلك تم عرض بعض المقدرات الحصينة للمتوسطات في مجتمع محدود المعاينة والتي تكون ملائمة في حالة وجود بعض المشاهدات الشاذة وهذه المقدرات الحصينة تشتق على أساس الدوال الفعالة التنبوئية الحقيقية.

المصطلحات الرئيسية للبحث/ مجتمع محدود المعاينة، معدل المجتمع، التعدد الخطي، انحدار الحرف، التقديرات الحصينة.

مجلة العلوم
الاقتصادية والإدارية
المجلد ٢٢ العدد ٨٧
الصفحات ٣٨٤-٣٩٠

١. المقدمة

توصل Fuller (1970) إلى تقدير مجتمع محدود المعاينة لنموذج مبنى إستلالياً. ثم علق Rao (1975) على هذه النتيجة المهمة بأن هناك كثيراً من التقديرات المعلمية التقليدية المعرفة بمقدر حصين هو مناسب لمجتمع محدود المعاينة. ومن هنا نجد أنه يمكن الحصول على هذا المقدر بواسطة تقدير تقليدي معلمي أو حصين لطرائق تقدير المقدر لمجتمع محدود كلياً. أي أن معدل مربع الخطأ للمقارنة على المقدر الأصلي تؤدي بالنتيجة إلى المقارنة باستخدام الخطأ للمقدر المشترك لمجتمع محدود كلياً. في بحث Fuller المذكور آنفاً ثم افتراض N من القياسات y_j حيث أن j=1,2,…,N على مجتمع محدود الوحدات. نفرض أن هذه القياسات مستقلة ولها معدلات متساوية.

إن الهدف من هذا البحث هو الشمولية أي بعبارة أخرى توسيع النموذج لمعدلات غير متساوية وسوف نبرهن أن هذه النتيجة التي حصلنا عليها يمكن استخدامها في الحصول على أفضل نموذج غير متحيز لمقدرات معدل مجتمع محدود.

ويهدف أيضاً إلى الحصول على المقدرات الحصينة للمتوسطات في مجتمع محدود المعاينة والتي تكون ملائمة في حالة وجود بعض المشاهدات الشاذة وهو من المواضيع التي اهتم بها Malay Ghosh (2008).

٢. الجانب النظري

أفرض أن القياسات y_j (j=1,2,…,N) هي مستقلة وأن $E_m(y_j)=\mu_j$ التنبؤ الاتي يعبر عن مجتمع محدود كلياً كالآتي:

$$Y = n\bar{y}_s + (N-n)\bar{y}_{ns} \qquad \dots(1)$$

حيث أن $\bar{y}_s$ هو معدل لعينة الوحدات، وأن $\bar{y}_{ns} = \sum_{j\in s} y_j \big/ (N-n)$ هو ليس معدل لعينة الوحدات.

لتقدير Y (معدل النموذج) فإن:

$$E_m(\bar{y}_{ns}) = \mu_{ns} = \sum_{j\in s}\mu_j \big| (N-n)$$

أفرض أن لدينا تقديرين لمعدل النموذج μ_{ns} هما $\hat{\mu}_s^{(2)}, \hat{\mu}_s^{(1)}$ وعليه فإن التقديرين لـ Y هما:

$$\hat{Y}_t = n\bar{y}_s + (N-n)\hat{\mu}_s^{(t)} \quad , \quad t=1,2 \qquad \dots(2)$$

ويستخدم الدليل s على $\hat{\mu}^{(t)}$ للتعبير عن الاعتماد على وحدات العينة. هذه النتيجة هي نفس الشكل كما في Fuller (1970). ولأجل أي معاينة مصممة لحجم عينة ثابت n، فإن لكل عينة s ،

$$E_m(\hat{\mu}_s^{(1)} - \mu_{ns})^2 \le E_m(\hat{\mu}_s^{(2)} - \mu_{ns})^2$$

وأن

$$E_m E_p(\hat{Y}_1 - Y)^2 \le E_m E_p(\hat{Y}_2 - Y)^2$$

نلاحظ أنه لا افتراض حول التباين لـ y_j جعل غير الذي هو:

$$Var(y_j) = \sigma_j^2 < \infty \ \text{ for } j=1,2,\dots,N$$

المجتمع المحدود لـ MSE ولـ $\hat{y}_T$ هو معدل فوق النموذج

$$E_m E_p(\hat{Y}_t - Y)^2 = (N-n)^2 E_m E_p(\hat{\mu}_s^{(t)} - \bar{y}_{ns})^2 \qquad \dots(3)$$

عند تعويض (1) و(2) في الجانب الأيسر لـ (3) وعند إضافة وطرح μ_{ns} تحت التوقعات وتربيع الحدود الخارجية الجانب الأيمن لـ (3) يمكن كتابته كما يأتي:

$$(N-n)^2\left\{E_mE_p(\hat{\mu}_s^{(t)}-\mu_{ns})^2-2E_mE_p(\hat{\mu}_s^{(t)}-\mu_{ns})(\bar{y}_{ns}-\mu_{ns})+E_mE_p(\bar{y}_{ns}-\mu_{ns})^2\right\} \quad ...(4)$$

بما أن عوامل التوقع E_m , E_p مستقلة وبما أن المشاهدات مستقلة، يمكن التعبير عن المعادلة (4) كما يأتي:

$$(N-n)^2\left\{E_pE_m(\hat{\mu}_s^{(t)}-\mu_{ns})^2-2E_pE_m(\hat{\mu}_s^{(t)}-\mu_{ns})E_m(\bar{y}_{ns}-\mu_{ns})+E_mE_p(\bar{y}_{ns}-\mu_{ns})^2\right\} \quad ...(5)$$

لاحظ أن $\hat{\mu}_s^{(t)}$ تعتمد على مشاهدات المعاينة، حيث أن $\bar{y}_{ns}$ تعتمد على واحدة غير معاينة، وعليه يوجد نموذج مستقل عند استعمال النموذج المفروض $E_m(y_j)=\mu_j$ فإن $E(\bar{y}_{ns}-\mu_{ns})=0$ في (5) وكنتيجة فإن (5) تؤدي إلى:

$$(N-n)^2\left\{E_pE_m(\hat{\mu}_s^{(t)}-\mu_{ns})^2+E_mE_p(\bar{y}_{ns}-\mu_{ns})^2\right\}$$

ولأي نموذج معطى فإن $E_mE_p(\bar{y}_{ns}-\mu_{ns})$ هو ثابت. لنفس المجتمع محدود المعاينة، نفرض النموذج البيزي الهرمي (وهو شرطي على μ):

$$y_j \overset{ind}{\sim} N(\mu_{aj},\sigma_j^2)$$

حيث أن:

a_j , σ_j^2 ثوابت معلومة

μ غير معلومة ولها توزيع **Uniform**

وكذلك فإن:

$$\bar{y}_{ns}=\frac{\sum_{j=1}^{n}a_j\,\sigma_j^{-2}\,y_j}{\sum_{i=1}^{n}a_i^2\,\sigma_i^{-2}}$$

حيث أن

$$a_s=(a_1,\dots,a_n)^T,$$
$$a_u=(a_{n+1},a_{n+2},\dots,a_N)$$
$$y_s=(y_1,\dots,y_n)^T$$
$$y_u=(y_{n+1},y_{n+2},\dots,y_N)^T$$

وأن:

$$\sum u=\operatorname{diag}(\sigma_{n+1}^2,\dots,\sigma_N^2)$$

إن التوزيع اللاحق لـ θ/y_s هو $N\left(\bar{y}_{ns}\,,\,1\Big/\sum_{j=1}^{n}a_j^2\sigma_j^{-2}\right)$ وأن توزيع التنبؤ اللاحق أو التوزيع اللاحق المتنبأ به لـ y_s هو:

$$N(\bar{y}_{ns}a_u, \sum u + a_u a_u^T / \sum a_j^2 \sigma_j^{-2})$$

ومن النموذج المعطى فإن المقدر لمتوسط المجتمع المحدود $(\bar{y}_s = N^{-1}\sum_{j=1}^{N} y_j)$ هو:

$$\tilde{\bar{y}}_p = N^{-1}\left\{\sum_{j=1}^{n} y_j + \bar{y}_{ns}\sum_{i=n+1}^{N} a_j\right\}$$

وفي الجانب العملي، إذا كان **i=1,…,n** ، $\sigma_i^2 = \sigma_{xi}^2$ ، $a_i = x_i$ فإن المقدر الناتج لمتوسط المجتمع المحدود هو مقدر نسبي $(\bar{y}_s / \bar{x}_s)\bar{x}_{ns}$ حيث أن:

$$\bar{y}_s = n^{-1}\sum_{i=1}^{n} y_i$$

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

$$\bar{x}_{ns} = N^{-1}\sum_{j=1}^{N} x_j$$

وعند اختيار $a_i = x_i$ و $\sigma_i^2 = x_i^2$ سنحصل على مقدر متوسط المجتمع:

$$N^{-1}\left[\sum_{i=1}^{n} y_i + n^{-1}\sum_{i=1}^{n}(y_i / x_i)\right]$$

وهذا المقدر قدّم من قبل **Royall (1970)**. على أية حال فإن مقدرات (y_i / a_i) ليس بالضرورة أن تكون بياناتها شاذة، فقد تكون بعض مركبات y_i / a_i مختلفة تماماً عن متوسط المجتمع $\bar{y}_{ns}$ وقد يحدث هذا عندما تكون نسبة y_i / a_i أقل جزئياً في المقدار من النسب الأخرى y_j / a_j لكن تباينها σ_i^2 / a_i^2 أكثر بكثير من التباينات الأخرى $(i \neq j)$, σ_j^2 / a_j^2.

وبالاعتماد على نظرية (1) في المصدر (3) فإن المقدر الحصين لمتوسط المجتمع محدود المعاينة هو:

$$\bar{y}_{ns}^{(R)} = N^{-1}\left[\sum_{i=1}^{n} y_i + \hat{\mu}_R \sum_{j=n+1}^{N} a_j\right]$$

٣. الجانب التطبيقي

يمكن ايجاد النموذج الأمثل غير المتحيز لمقدرات **Y** أو $\bar{y}$. أفرض أن $\hat{\mu}_s$ هي أصغر مربعات موزونة لتقدير μ_{ns} . فإن لكل عينة معطاة وبواسطة نظرية كاوس ماركوف فإن التباين لـ $E_m(\hat{\mu}_s - \mu_{ns})^2$ هو الأصغر بين التباينات الخطية غير المتحيزة s لمقدرات μ_{ns} وهنا عدم التحيز هو بالنسبة إلى النموذج. عند تطبيق البند (2) فإن:

$$\hat{Y} = n\bar{y}_s + (N - n)\hat{\mu}_s$$

هو مقدر أمثل لـ **Y** من بين كل المقدرات الخطية لنموذج غير متحيز لـ **Y** ذلك أن $E_m E_p(\hat{Y}-Y)^2$ هو الأصغر بين التباينات للمقدرات في تلك الفئة. هذا التطبيق متوازي. مثال ذلك النتيجة لـ **Royall (1970)** ، **Royall and Cumber Land (1981)** وكمثال على ذلك، نفرض النموذج:

$$y_j = \alpha + \beta X_j + e_j$$

$$E_m(e_j)=0 \quad , \quad E_m(e_j^2)=\sigma^2 \quad , \quad E(e_j, e_k)=0 \quad \dots(6)$$

والتي تظهر في **Cumberland, Royall (1981)** في نموذج (6)، $\mu_j = \alpha + \beta X_j$ حيث أن **X** هي **covariate** ، $E_m(e_j, e_k)=0$ تنتج من استقلالية المشاهدات، والتقديرات لأصغر مربعات لـ β, α هما:

$$\hat{\beta} = \sum_{j\in s} X_j(y_j - \bar{y}_s) \Big/ \sum_{j\in s}(X_j - \bar{X}_s)^2$$

$$\hat{\alpha} = \bar{y}_s - \hat{\beta}\bar{X}_s$$

المقدر الناتج لـ **Y** هو:

$$\hat{Y} = n\bar{y}_s + \sum_{j\in s}\hat{\mu}_j$$

$$= n\bar{y} + \sum_{j\in s}(\hat{\alpha} - \hat{\beta}X_j)$$

$$= N\left\{\bar{y}_s + \hat{\beta}(\bar{X} - \bar{X}_s)\right\}$$

الطريقة في البند (2) يمكن أن تطبق فقط للنماذج التي يوجد فيها خطأ يكون غير مترابط، نظرية **Royall (1976) (2.1)** تحتوي أحسن نموذج خطي غير متحيز للمقدر **Y** عندما تكون المشاهدات مترابطة والنتائج غير المتشابهة في النموذج الأمثل غير متحيزة التقدير نظرياً. النتيجة في البند (2) ربما تطبق خارج الفئة لنموذج غير متحيز التقديرات.

أفرض أن النموذج (6) يطبق ولكن عدة **Covariates** تظهر في (6) بدل $y_j = \alpha + \beta X_j + e_j$ بواسطة $y_j = X_j^T\beta + e_j$ حيث أن $\beta^T = (\alpha, \beta_1, \beta_2, \dots, \beta_p)$ هو متجه لمعاملات الانحدار وإن $X_j^T = (1, X_{1j}, X_{2j}, \dots, X_{pj})$ هو متجه لـ **Covariates** على j^{th} وحده أكثر من ذلك أفرض أن y_s هي **n x 1** متجه لمعاينة قياسات الأشياء المرغوبة، وأن X_s هي **n x (p+1)** مصفوفة لمعاينة **Covariates**. الصفوف لـ X_s تتكون من X_j^T لكل $j \in s$ المقدر لأصغر مربعات لـ β هو:

$$\hat{\beta} = (X_s^T X_s)^{-1} X_s^T y_s$$

والنتيجة هي أحسن نموذج خطي غير متحيز للمقدر **Y** هو:

$$\hat{Y}_{blue} = N\left\{\bar{y}_s + (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)(\bar{X} - \bar{X}_s)\right\}$$

$$= n\bar{y}_s + X_{ns}^T\hat{\beta}$$

حيث أن $\bar{X}, \bar{X}_s$ هما **p x 1** متجهات لمعدلات الـ **Covariates** المجتمع والعينة وعندما

$$X_{ns}^T = (N-n, N\bar{X}_1 - n\bar{X}_1, N\bar{X}_2 - n\bar{X}_2, \dots, N\bar{X}_p - n\bar{X}_p)$$

تكون مشكلة Multicol-Linearity موجودة في الـCovariates فإن استخدام الانحدار الـridge ربما يكون مناسب (أنظر (1)) Draper and Smith (1981) الـridge العادي المقدر لـ β هو:

$$\hat{\beta}_k = (X_s^T X_s + KI)^{-1} X_s^T y_s$$

حيث أن K هي أي ثابت وأن I هي (p+1) x (p+1) مصفوفة أحادية. المقدر الناتج لـ Y هو:

$$\hat{Y}_{ridge} = n\bar{y}_s + X_{ns}^T \hat{\beta}_k$$

عند تطبيق النتيجة في البند (2) مقارنة مع الكفاية النسبية المطبقة تحت النموذج لانحدار ridge تحمل فوق المقدرات المكونة لتقدير Y، [8] Vinod and Ullah (1981)، كمثال، الشروط التي تجعل تباين النموذج لـ $X_{ns}^T\hat{\beta}$ هو أكبر من نموذج MSE لـ $X_{ns}^T\hat{\beta}_k$. في التطبيق الشرط الضروري والكافي لأجل تقدير ridge يكون أكثر كفاية من المقدر لأصغر مربعات هو:

$$0 < K < 2/[-\min(0, \psi)] \quad \dots (7)$$

حيث أن ψ هي أصغر قيمة ذاتية لـ $(X_s^T X_s)^{-1} - (\beta\beta^T/\sigma^2)$ وكنتيجة المتباينة في (7) تتحقق لكل عينة s شرط ضروري وكافي لأجل:

$$E_m E_p (\hat{Y}_{ridge} - Y)^2 < E_m E_p (\hat{Y}_{blue} - Y)^2 \quad \dots (8)$$

إذا كانت العينة مستقلة، الشرط في المعادلة (7) يكون:

$$0 < K < 2\sigma^2/\beta^T\beta$$

٤. المصادر

1. Draper N.R. and Smith, H. (1981), "*Applied Regression Analysis*". 2nd ed., New York, John Wiley.
2. Fuller, W.A. (1970), "*Simple Estimators for the Mean of Skewed Population: Technical Report* ". Iowa State University, Dept. of Statistics.
3. Malay Ghosh, (2008), "*IMS Collections Beyond Parametric in Interdisciplinary Research*". Vol. (1), 116-122. Institute of Mathematical Statistics.
4. Rao, N.J.K. (1975), "*On the Foundations of Survey Sampling*". In a survey of Statistical Design and Linear Models, and J.N. Srivastava Amsterdam North-Holland, pp. 489-505.
5. Royall, R.M. (1970), "*On Finite Population Sampling Theory Under Certain Linear Regression Models*". Biometrika, 75; 377-387.
6. Royall, R.M. (1976), "*The Linear Least-Squares Prediction Approach to Two-State Sampling*". Journal of the American Statistical Association, 71, 657-664.
7. Royall, R.M. and Cumberland, W.G. (1984), "*The Finite Population Linear Regression Estimator and Estimators of it's Variance-An Empirical Study*". Journal of the American Statistical Association, 76, 924-930.
8. Seber, G.A.F. (1977), "*Linear Regression Analysis*". New York, John Wiley.
9. Vinod, H.D. and Ulaah, A. (1981), "*Recent Advances in Regression Methods*". New York, Marcel Dekker.

10- www.minshawi.com/vb/attachment.php?attachment

Model Estimated Building in Finite Population Sampling

Abstract

The population is sets of vocabulary common in character or characters and it's study subject or research . statistically , this sets is called study population (or abridgement population) such as set of person or trees of special kind of fruits or animals or product any country for any commodity through infinite temporal period term ... etc.

The population maybe finite if we can enclose the number of its members such as the students of finite school grade . and maybe infinite if we can not enclose the number of it is members such as stars or aquatic creatures in the sea . when we study any character for population the statistical data is concentrate by two method ، the first method is census which we concentrate the data for each singular of population ، and the second method is sampling method which we concentrate the data for part of population such as this part (sample) have the sane characters of population which we taken .

This research proposes estimation for some of parameters in finite population sampling, such we use the estimation of average of the model $u_{ns} = E_m(\bar{y}_{ns})$ and obtaining of the Best Unbiased Estimator of average of finite population by the Fuller use. This research also proposes some robust estimators of the finite population mean which suitable in the presence of some outlying observations. The robust estimators are derived on the basis of certain predictive influence functions.

Keywords/ Finite Population Sampling, Average of Population, Multi collinearity, Ridge Regression, Robust Estimation.