تقدير متوسط المجتمع في المعاينة العشوائية الطبقية مع وجود بيانات مفقودة

د. محمد مزيد دريباتي*

د. منذر بوبو**

هديل درويش***

(تاريخ الإيداع 29 / 4 / 2021. قُبِل للنشر في 9 / 6 /2021)

□ ملخّص □

يهدف البحث الحالي إلى اقتراح مقدّر جديد لمتوسط المجتمع في المعاينة العشوائية الطبقية مع وجود بيانات مفقودة. القيم المفقودة التي يدرسها البحث تتمثل بأخطاء القياس وأخطاء عدم الاستجابة معا وفي وقت واحد. من أجل دراسة فعالية المقدّر المقترح تمت مقارنة النتائج التي حصلنا عليه من هذا المقدّر مع عدد من المقدّرات السابقة التي سبق ذكرها في الدراسات العلمية. تمت مقارنة النتائج بين المقدرات المختلفة على بيانات مولدة وبيانات حقيقية وتمت مقارنة خطأ الانحياز ومتوسط مربعات الخطأ لكل منها.

الكلمات المفتاحية: التقدير، متوسط المجتمع، المعاينة العشوائية الطبقية، البيانات المفقودة.

*أستاذ مساعد – قسم الإحصاء الرياضي– كلية العلوم– جامعة تشرين– اللاذقية– سورية.

**أستاذ مساعد – قسم القياس والتقويم– كلية التربية– جامعة تشرين– اللاذقية– سورية.

***طالبة دكتوراه – قسم الإحصاء الرياضي– كلية العلوم– جامعة تشرين– اللاذقية– سورية.

Estimation of Population Mean in the Presence of Missing Data Under Stratified Random Sampling

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

(Received 29 / 4 / 2021. Accepted 9 / 6 /2021)

☐ ABSTRACT ☐

This paper aims to propose a new estimator of the population mean in stratified random sampling with missing data. The missing values studied by the paper are measurement errors and non-response errors together and at the same time. To study the effectiveness of the proposed estimator, the results obtained from this estimator were compared with several previous estimates that were previously mentioned in scientific studies. The results were compared between the different estimates on both generated and real data, and the bias error and mean squares of error were compared for each of them.

Keywords: Estimation, population mean, stratified random sampling, missing data.

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

مقدمة:

يحاول الباحثون عادةً إجراء البحوث على المجتمعات بشكل كامل، ولكن صعوبة ذلك –نتيجة كبر حجم المجتمع والتكلفة والوقت الذي يحتاجه إنجاز مثل هذه البحوث– أجبرت الباحثين على الاعتماد على أساليب المعاينة الإحصائية في تحليل البيانات الإحصائية التي يتم جمعها إما عن طريق إجراء دراسة تجريبية أو من الاستبيانات التي يقوم الباحث بتطبيقها، ويفترض عادة أن جميع البيانات الممثلة لحالات المتغيرات المدروسة تقاس بشكل كامل وصحيح وهو أمر غير محقق دوماً، حيث أنه وفي أثناء تبويب البيانات في الجداول الإحصائية من أجل إجراء الدراسة الإحصائية قد يصادف الباحث بيانات مفقودة لأسباب عدة منها عدم الاستجابة أو أخطاء في القياس وهي ناتجة عن سبب مقصود أو غير مقصود. ونظراً لصعوبة إعادة التطبيق أو إمكانية الحصول على القيم المفقودة مرة أخرى يكون الباحث مجبر على قبول البيانات التي قام بجمعها ومعالجتها على الرغم من وجود بعض القيم المفقودة في بعض المتغيرات المدروسة لأن إمكانية حذف هذه المتغيرات أو الأفراد التي تحوي القيم المفقودة في بعض الأحيان قد تكون صعبة نتيجة أهمية وجودها رغم وجود القيم المفقودة فيها. ونحاول في بحثنا هذا تقليل الأخطاء التي قد تنتج عن تقدير المتوسط الحسابي في حال وجود هذه القيم من خلال المقدّر المقترح.

أهمية البحث وأهدافه:

تكمن أهمية هذا البحث في التقليل من الأخطاء الناتجة عن تقدير المتوسط الحسابي في حالة المعاينة العشوائية الطبقية دون الحاجة إلى إعادة جمع معطيات جديدة مما يقلل من الوقت والتكلفة والجهد في إجراء البحوث مع المحافظة على الدقة والكفاءة والفاعلية للمقدّر.

طرائق البحث ومواده:

تم في هذا البحث اقتراح مقدّر جديد لمتوسط المجتمع ومقارنة النتائج التي نحصل عليها مع مقدرات أخرى تمت دراستها سابقاً في أبحاث أخرى. وقبل البدء بعرض المقدّر الجديد المقترح سوف نقدم بعض التعريفات والمصطلحات التي توضح بعض المفاهيم المستخدمة حتى لا نقع في التعميم لأن مفهوم القيم المفقودة قد يأخذ معانٍ مختلفة حسب طريقة التعامل معه وحسب طريقة حدوثه.

ثم نقدم عرضاً لأهم الفرضيات التي استندنا عليها في اقتراحنا للمقدّر، وبالاعتماد على بعض المقدّرات القياسية المبرهنة سابقا نثبت كفاءة مقدّرنا المقترح نظريا ثم تجريبا عن طريق المقارنة مع بيانات قياسية وباستخدام برنامج R.

تعريفات واصطلاحات:

–ليكن لدينا مجتمع منته بحجم N، قسم إلى L طبقة (طبقة متجانسة داخل كل منها ومتباينة فيما بينها)، وفق أسس المعاينة العشوائية الطبقية، بحيث يكون N_h حجم الطبقة h و $h=1,2,3,\ldots L$ و $N=\sum_{h=1}^{L} N_h$ [7]

سنعتبر أن المعطيات المجموعة من المجتمع قد قُسمت الى مجموعتين رئيسيتين حسب المعاينة المضاعفة وذلك حسب الاستجابة:

تمثل المجموعة الأولى الحالات المستجيبة وحجمها N_{ah} في الطبقة h وتمثل المجموعة الثانية الحالات غير المستجيبة وحجمها N_{bh} في الطبقة h .

–نسحب بطريقة المعاينة العشوائية البسيطة من كل طبقة h عينة بحجم n_h وتقسم هذه العينة إلى عينة من الحالات المستجيبة بحجم n_{ah} وعينة من الحالات غير المستجيبة أو القياسات الخاطئة بحجم n_{bh} وبحيث يكون حجم العينة الطبقية الكلي $n=\sum_{h=1}^{L} n_h$ ، ثم نقوم باختيار عينة جزئية (sub–sample) من كل عينة غير مستجيبة n_{bh} بحجم

n_{2h} حيث $n_{2h}=\dfrac{n_{bh}}{k_h};k_h>1$. [5]

–نصطلح أن(Y^*_{hi},X^*_{hi}) و (y^*_{hi},x^*_{hi}) هي القيم الحقيقية والقيم المسجلة للمتغيرات (Y,X) في العينة الطبقية n_h من كل طبقة h حيث $i=1,2,..,n_h$ ونعرف أخطاء القياس كما يلي:

$U^*_{hi}=y^*_{hi}-Y^*_{hi}$: خطأ القياس بالنسبة لمتغير الدراسة Y في الطبقة h .

$V^*_{hi}=x^*_{hi}-X^*_{hi}$: خطأ القياس بالنسبة للمتغير المساعد X في الطبقة h .

ويكون S^2_{hY} و S^2_{hX} تباينات الطبقة h من المجتمع للمتغيرين Y وX على الترتيب وذلك بالنسبة للحالات **المستجيبة** من المجتمع.

ويكون $S^2_{hY(2)}$ و $S^2_{hX(2)}$ تباينات الطبقة h من المجتمع للمتغيرين Y و X على الترتيب وذلك بالنسبة للحالات **غير المستجيبة** من المجتمع

ويكون S^2_{hU} و S^2_{hV} تباينات الطبقة h من المجتمع للمتغيرين YوX على الترتيب وذلك بالنسبة للحالات **المستجيبة** من المجتمع وبوجود أخطاء قياس،

ويكون $S^2_{hU(2)}$ و $S^2_{hV(2)}$ تباينات الطبقة h من المجتمع للمتغيرين YوX على الترتيب وذلك بالنسبة للحالات **غير المستجيبة** من المجتمع وبوجود أخطاء قياس.

و C_{hY} و C_{hX} معاملا التغاير للحالات **المستجيبة** في الطبقة h بالنسبة للمتغيرين YوX على الترتيب.

و $C_{hY(2)}$ و $C_{hX(2)}$ معاملا التغاير للحالات **غير المستجيبة** في الطبقة h بالنسبة للمتغيرين YوX على الترتيب:

$$C_{hY}=\frac{S_{hY}}{\bar{Y}_h} \text{ و } C_{hX}=\frac{S_{hX}}{\bar{X}_h} \text{ و } C_{hY(2)}=\frac{S_{hY(2)}}{\bar{Y}_h} \text{ و } C_{hX(2)}=\frac{S_{hX(2)}}{\bar{X}_h}.\ [2]$$

ρ_{hYX} معامل الارتباط بين المتغيرين YوX بالنسبة للحالات **المستجيبة** في الطبقة h , $\rho_{hYX(2)}$ معامل الارتباط بين المتغيرين YوX بالنسبة للحالات **غير المستجيبة** في الطبقة h .

$$\rho_{hYX}=\frac{\operatorname{cov}(Y_h,X_h)}{S_{hY}S_{hX}},\ \rho_{hYX(2)}=\frac{\operatorname{cov}(Y_{h(2)},X_{h(2)})}{S_{hY(2)}S_{hX(2)}} \quad [3]$$

فرضيات البحث:

1–إن محاولة إعادة الحصول على البيانات بطريقة جديدة –ويفضل أن تكون إحدى وسائل التواصل المباشر– يزيد كثيرا من معدل الاستجابة وبالتالي حجم العينة الجزئية أصغر بكثير من حجم العينة الأولي [5] ومنه نفترض وجود استجابة كاملة في العينة الجزئية وذلك في كل طبقة h

2–القيم الحقيقية للمتغيرات مستقلة عن أخطاء القياس لأن القيم الحقيقية ليس لها علاقة بالتوزع الطبيعي لأخطاء القياس ومنه من أجل كل طبقة من المجتمع يكون $E(\sum_{i=1}^{n_h} V^*_{hi})=0$ و $E(\sum_{i=1}^{n_h} U^*_{hi})=0$ [12-13-14]

3–إن أخطاء القياس أو عدم الاستجابة أو كلاهما معا تكون مرتبطة بمتغير الدراسة والمتغير المساعد أيضا ولكن بصورة مستقلة أي أن حدوث خطأ في متغير الدراسة لا يعني بالضرورة وجوده في المتغير المساعد أيضا والعكس صحيح [6]

الدراسة المرجعية:

1–كان Hansen and Hurwitz [5] أول من درس أثر البيانات المفقودة في التقدير وذلك عام (1946) حيث وضعا مقدّر لمتوسط المجتمع في ظل وجود بيانات مفقودة وذلك في المعاينة العشوائية الطبقية بالشكل:

$T_{HH}=\sum_{h=1}^{L} P_h \overline{y^*_h}$ حيث عرفا $\overline{y^*_h}$ بالشكل:

$$P_h=\frac{N_h}{N} \text{ و } \bar{y}_{2h}=\frac{1}{n_{2h}}\sum_{i=1}^{n_{2h}} y_{bhi} \text{ و } \bar{y}_{ah}=\frac{1}{n_{ah}}\sum_{i=1}^{n_{ah}} y_{ahi} \text{ و } \bar{y}^*_h=\frac{n_{ah}}{n_h}.\bar{y}_{1h}+\frac{n_{bh}}{n_h}.\bar{y}_{2h}$$

واصطلحا بعد دراسة نظرية وتجريبية أن تباين تقدير المتوسط:

$$\text{Var}(T_{HH})=\sum_{h=1}^{L} p_h^2[\lambda_h(S^2_{hY}+S^2_{hU})+\theta_h(S^2_{hY(2)}+S^2_{hU(2)})]$$

حيث أن:
$$\theta_h=\frac{W_{b_h}(k_h-1)}{n_h};W_{b_h}=\frac{N_{b_h}}{N_h};\lambda_h=n_h^{-1}-N_h^{-1}$$

2–ثم بدأ الباحثون بالتركيز على دراسة تقدير المتوسط في ظل وجود أخطاء القياس وعدم الاستجابة معا وفي وقت واحد وبوجود متغير مساعد في البحث حيث وضع Rao's [8–9–10] عام 1990 مقدّر جديد بالشكل:

$$T_R=\sum_{h=1}^{L} P_h \frac{\bar{y}^*_h}{\bar{x}_h}\bar{X}_h$$

وقام بحساب متوسط مربعات الخطأ لهذا المقدّر بالشكل:

$$MSE(T_R)=\sum_{h=1}^{L} P_h^2[\lambda_h\bar{Y}_h^2(C^2_{hY}+C^2_{hX}-2\rho_{hYX}C_{hY}C_{hX})+\theta_h S^2_{hY(2)}+\lambda_h\bar{Y}_h^2(\frac{S^2_{hU}}{\bar{Y}_h^2}+\frac{S^2_{hV}}{\bar{X}_h^2})+\theta_h S^2_{hU(2)}]$$

3–ثم طور Srivastava في عام 2010 [15] مقدّر النسبة الذي وضعه كوكران في عام 1977 [2] بالشكل:

$$T_S=\sum_{h=1}^{L} P_h(\frac{\bar{y}^*_h}{\bar{x}^*_h})\bar{X}_h$$

وقام بحساب متوسط مربعات الخطأ لهذا المقدّر بالشكل:

$$\begin{aligned}MSE(T_S)=&\sum_{h=1}^{L} P_h^2[\lambda_h\bar{Y}_h^2(C^2_{hY}+C^2_{hX}-2\rho_{hYX}C_{hY}C_{hX})+\\&\theta_h\bar{Y}_h^2(C^2_{hY(2)}+C^2_{hX(2)}-2\rho_{hYX(2)}C_{hY(2)}C_{hX(2)})+\\&+\lambda_h\bar{Y}_h^2(\frac{S^2_{hU}}{\bar{Y}_h^2}+\frac{S^2_{hV}}{\bar{X}_h^2})+\theta_h\bar{Y}_h^2(\frac{S^2_{hU(2)}}{\bar{Y}_h^2}+\frac{S^2_{hV(2)}}{\bar{X}_h^2})]\end{aligned}$$

4–ثم وضع Sing and Kumar [12–13–14] عام 2011 مقدّر آخر:

$$T_{SK}=\sum_{h=1}^{L}P_h\bar{y}_h^{*}(\frac{\bar{X}_h}{\bar{x}_h^{*}})(\frac{\bar{X}_h}{\bar{x}_h})$$

وقام بحساب متوسط مربعات الخطأ لهذا المقدّر بالشكل:

$$MSE(T_{SK})=\sum_{h=1}^{L}P_h^2[\lambda_h\bar{Y}_h^2(C_{hY}^2+4C_{hX}^2-4\rho_{hYX}C_{hY}C_{hX})+$$
$$\theta_h\bar{Y}_h^2(C_{hY(2)}^2+C_{hX(2)}^2-2\rho_{hYX(2)}C_{hY(2)}C_{hX(2)})+$$
$$+\lambda_h\bar{Y}_h^2(\frac{S_{hU}^2}{\bar{Y}_h^2}+4\frac{S_{hV}^2}{\bar{X}_h^2})+\theta_h\bar{Y}_h^2(\frac{S_{hU(2)}^2}{\bar{Y}_h^2}+\frac{S_{hV(2)}^2}{\bar{X}_h^2})]$$

5-ثم قدم Azeem and Haneef عام 2017 [1] مقدّر جديد للنسبة بالشكل :

$$T_A=\sum_{h=1}^{L}P_h\bar{y}_h^{*}(\frac{\bar{x}_h'^{*}}{\bar{x}_h})(\frac{\bar{x}_h'^{*}}{\bar{x}_h^{*}})$$

وقام بحساب متوسط مربعات الخطأ لهذا المقدّر بالشكل:

$$MSE(T_A)=\sum_{h=1}^{L}P_h^2\{\lambda_h\bar{Y}_h^2[C_{hY}^2+(\frac{N_h+n_h}{N_h-n_h})C_{hX}^2-2(\frac{N_h+n_h}{N_h-n_h})\rho_{hYX}C_{hY}C_{hX}]+$$
$$\theta_h\bar{Y}_h^2[C_{hY(2)}^2+(\frac{N_h+n_h}{N_h-n_h})C_{hX(2)}^2-2(\frac{N_h+n_h}{N_h-n_h})\rho_{hYX(2)}C_{hY(2)}C_{hX(2)}]+$$
$$\lambda_h\bar{Y}_h^2[\frac{S_{hU}^2}{\bar{Y}_h^2}+(\frac{N_h+n_h}{N_h-n_h})^2\frac{S_{hV}^2}{\bar{X}_h^2}]+\theta_h\bar{Y}_h^2[\frac{S_{hU(2)}^2}{\bar{Y}_h^2}+(\frac{N_h+n_h}{N_h-n_h})^2\frac{S_{hV(2)}^2}{\bar{X}_h^2}]\}$$

بعد استعراض المقدّرات السابقة ودراستها بشكل جيد قمنا باقتراح المقدّر التالي:

المقدّر المقترح:

$$M=\sum_{h=1}^{L}P_h\bar{y}_h^{*}\left[q_{1h}\exp\left\{\frac{\bar{X}_h-\bar{x}_h^{*}}{\bar{X}_h+\bar{x}_h^{*}}\right\}+q_{2h}\exp\left\{\frac{N_h(\bar{X}_h-\bar{x}_h^{*})}{N_h(\bar{X}_h+\bar{x}_h^{*})-2n_h\bar{x}_h^{*}}\right\}\right]$$

حيث: $P_h=\frac{N_h}{N}$

q_1 و q_2 :ثوابت عددية يتم اختيارها بحيث يتم تحقق أصغر متوسط مربعات خطأ للمقدّر المقترح M حيث

$$q_1+q_2=1$$

إيجاد خطأ الانحياز للمتوسط المقترح: $Bias(M)=E(M-\bar{Y})$

لنفترض أن:

$$\delta_{hy}^{*}=\sum_{i=1}^{n_h}(Y_{hi}^{*}-\bar{Y}_h)\text{...............}(1)$$

$$\delta_{hU}^{*}=\sum_{i=1}^{n_h}U_{hi}^{*}=\sum_{i=1}^{n_h}(y_{hi}^{*}-Y_{hi}^{*})\text{...............}(2)$$

$$\delta_{hX}^{*}=\sum_{i=1}^{n_h}(X_{hi}^{*}-\bar{X}_h)\text{...............}(3)$$

$$\delta^*_{hV} = \sum_{i=1}^{n_h} V^*_{hi} = \sum_{i=1}^{n_h} (x^*_{hi} - X^*_{hi})...............(4)$$

بجمع (1) و (2) نجد:

$$\delta^*_{hy} + \delta^*_{hU} = \sum_{i=1}^{n_h}(Y^*_{hi} - \overline{Y}_h) + \sum_{i=1}^{n_h}(y^*_{hi} - Y^*_{hi}) \Rightarrow \delta^*_{hy} + \delta^*_{hU} = \sum_{i=1}^{n_h} Y^*_{hi} - \sum_{i=1}^{n_h}\overline{Y}_h + \sum_{i=1}^{n_h} y^*_{hi} - \sum_{i=1}^{n_h} Y^*_{hi}$$

$$\Rightarrow \delta^*_{hy} + \delta^*_{hU} = -n_h\overline{Y}_h + n_h\overline{y}^*_h \Rightarrow \overline{y}^*_h = \overline{Y}_h + \frac{\delta^*_{hY} + \delta^*_{hU}}{n_h}$$

ومنه:

$$\overline{y}^*_h = \overline{Y}_h + \xi_{0h}$$

حيث ξ_{0h} متحول عشوائي

$$\xi_{0h} = \frac{\delta^*_{hY} + \delta^*_{hU}}{n_h}$$

وبنفس الطريقة نحصل على:

$$\overline{x}^*_h = \overline{X}_h + \xi_{1h}$$

حيث:

$$\xi_{1h} = \frac{\delta^*_{hX} + \delta^*_{hV}}{n_h}$$

ولدينا حسب الفرض(2) وحسب [5]:

$$E(\xi_{0h}) = E(\xi_{1h}) = 0$$

وبالتالي نجد:

$$E(\xi_{0h})^2 = E(\overline{y}^*_h - \overline{Y}_h)^2 = \operatorname{var}(\overline{y}^*_h) = \lambda_h(S^2_{hY} + S^2_{hU}) + \theta_h(S^2_{hY(2)} + S^2_{hU(2)})$$

$$E(\xi_{1h})^2 = E(\overline{x}^*_h - \overline{X}_h)^2 = \operatorname{var}(\overline{x}^*_h) = \lambda_h(S^2_{hX} + S^2_{hV}) + \theta_h(S^2_{hX(2)} + S^2_{hV(2)})$$

$$E(\xi_{0h}\xi_{1h}) = E\left[(\overline{y}^*_h - \overline{Y}_h)(\overline{x}^*_h - \overline{X}_h)\right] = \operatorname{cov}(\overline{y}^*_h, \overline{x}^*_h)$$

$$= \rho_{h\overline{y}^*,\overline{x}^*}.\sqrt{\operatorname{var}(\overline{y}_h)}\sqrt{\operatorname{var}(\overline{x}_h)} = \lambda_h\rho_{hYX}S_{hY}S_{hX} + \theta_h\rho_{hYX(2)}S_{hY(2)}S_{hX(2)}$$

ولدينا:

$$\exp\left(\frac{\overline{X}_h - \overline{x}^*_h}{\overline{X}_h + \overline{x}^*_h}\right) = \exp\left(\frac{\overline{X}_h - \overline{X}_h - \xi_{1h}}{\overline{X}_h + \overline{X}_h + \xi_{1h}}\right) = \exp\left(\frac{-\xi_{1h}}{2\overline{X}_h + \xi_{1h}}\right) = \exp\left(-\frac{\xi_{1h}}{2}.\frac{1}{\overline{X}_h + \frac{\xi_{1h}}{2}}\right) = \exp\left(-\frac{d}{\overline{X}_h + d}\right); d = \frac{\xi_{1h}}{2}$$

$$f(d) = e^{-\frac{d}{\overline{X}_h + d}} \Rightarrow f(0) = e^{-\frac{0}{\overline{X}_h + 0}} = e^0 = 1$$

$$f'(d) = \frac{(-1)(\overline{X}_h + d) - (+1)(-d)}{(\overline{X}_h + d)^2} e^{-\frac{d}{\overline{X}_h + d}} == \frac{-\overline{X}_h + d - d}{(\overline{X}_h + d)^2} e^{-\frac{d}{\overline{X}_h + d}} = \frac{-\overline{X}_h}{(\overline{X}_h + d)^2} e^{-\frac{d}{\overline{X}_h + d}}$$

$$f'(0) = \frac{-\overline{X}_h}{(\overline{X}_h + 0)^2} e^{-\frac{d}{\overline{X}_h + d}} = -\frac{1}{\overline{X}_h}$$

$$f''(d) = \frac{-2(\overline{X}_h + d)(-\overline{X}_h)}{(\overline{X}_h + d)^4} e^{-\frac{d}{\overline{X}_h + d}} + \frac{\overline{X}^2_h}{(\overline{X}_h + d)^4} e^{-\frac{d}{\overline{X}_h + d}} \Rightarrow$$

$$f''(0) = \frac{2\overline{X}^2_h}{\overline{X}^4_h} + \frac{\overline{X}^2_h}{\overline{X}^4_h} = \frac{2}{\overline{X}^2_h} + \frac{1}{\overline{X}^2_h} = \frac{3}{\overline{X}^2_h}$$

وبالنشر حسب تايلور حيث نكتفي بالتقريب من الدرجة الثانية:

$$q_1 \exp\left(\frac{\overline{X}_h - \overline{x}^*_h}{\overline{X}_h + \overline{x}^*_h}\right) = q_{1h}\left[f(0) + f'(0)\frac{d}{1!} + f''(0)\frac{d^2}{2!}\right] = q_{1h}\left[f(0) + f'(0)\frac{\xi_{1h}}{2} + f''(0)\frac{\xi_{1h}^2}{8}\right]$$

$$= q_{1h}\left[1 + \left(-\frac{1}{\overline{X}_h}\frac{\xi_{1h}}{2}\right) + \left(\frac{3}{\overline{X}_h^2}\right)\frac{\xi_{1h}^2}{8}\right] = q_{1h} - \frac{1}{2\overline{X}_h}\xi_{1h}q_{1h} + \frac{3}{8}.\frac{1}{\overline{X}_h^2}\xi_{1h}^2 q_{1h}$$

$$\exp\left\{\frac{N_h(\overline{X}_h - \overline{x}^*_h)}{N_h(\overline{X}_h + \overline{x}^*_h) - 2n_h\overline{x}^*_h}\right\} =$$

$$= \exp\left(\frac{N_h\overline{X}_h - N_h(\overline{X}_h - \xi_{1h})}{N_h\overline{X}_h - 2n_h\left(\overline{X}_h + \xi_{1h}\right) + N_h\left(\overline{X}_h + \xi_{1h}\right)}\right) = \exp\left(\frac{-N_h\xi_{1h}}{2N_h\overline{X}_h - 2n_h\overline{X}_h - 2n_h\xi_{1h} + N_h\xi_{1h}}\right)$$

$$f(\xi_{1h}) = \exp\left(\frac{-N_h\xi_{1h}}{\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}}\right) \Rightarrow f(0) = e^0 = 1$$

$$f'(\xi_{1h}) = \left\{\frac{\{-N_h[\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}] - \left(N_h - 2n_h\right)(N_h\xi_{1h})\}}{\left[\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}\right]^2}\right\} . e^{\frac{-N_h\xi_{1h}}{\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}}}$$

$$f'(0) = \frac{-N_h\left(2N_h - 2n_h\right)\overline{X}_h}{\left(2N_h - 2n_h\right)^2\overline{X}_h^2} . e^0 = \frac{-N_h}{\left(2N_h - 2n_h\right)\overline{X}_h}$$

$$f''(\xi_{1h}) = \{\frac{-2(N_h - 2n_h)[(2N_h - 2n_h)\overline{X}_h + (N_h - 2n_h)\xi_{1h}].[-N_h(2N_h - 2n_h)\overline{X}_h]}{[(2N_h - 2n_h)\overline{X}_h + (N_h - 2n_h)\xi_{1h}]^4} .$$

$$e^{\frac{-N_h\xi_{1h}}{\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}}} + \frac{(-N_h)^2(2N_h - 2n_h)^2\overline{X}_h^2}{[(2N_h - 2n_h)\overline{X}_h + (N_h - 2n_h)\xi_{1h}]^4} . e^{\frac{-N_h\xi_{1h}}{\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}}}$$

$$= \{\frac{-2(N_h - 2n_h)(2N_h - 2n_h)\overline{X}_h - 2(N_h - 2n_h)^2\xi_{1h}].[-N_h(2N_h - 2n_h)\overline{X}_h]}{[(2N_h - 2n_h)\overline{X}_h + (N_h - 2n_h)\xi_{1h}]^4}$$

$$. e^{\frac{-N_h\xi_{1h}}{\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}}} + \frac{N_h^2(2N_h - 2n_h)^2\overline{X}_h}{[(2N_h - 2n_h)\overline{X}_h + (N_h - 2n_h)\xi_{1h}]^4} . e^{\frac{-N_h\xi_{1h}}{\left(2N_h - 2n_h\right)\overline{X}_h + \left(N_h - 2n_h\right)\xi_{1h}}}$$

$$f''(0) = \frac{-2N_h(N_h - 2n_h)(2N_h - 2n_h)^2\overline{X}_h + N_h^2(2N_h - 2n_h)^2\overline{X}_h^2]}{(2N_h - 2n_h)^4\overline{X}_h^4} = \frac{2N_h(N_h - 2n_h) + N_h^2}{(2N_h - 2n_h)^2\overline{X}_h^2}$$

$$= \frac{N_h(2N_h - 4n_h + N_h)}{(2N_h - 2n_h)^2\overline{X}_h^2} = \frac{N_h(3N_h - 4n_h)}{(2N_h - 2n_h)^2\overline{X}_h^2}$$

ومنه حسب تايلور نجد:

$$\exp\left\{\frac{N_h(\bar{X}_h-\bar{x}^*_h)}{N_h(\bar{X}_h+\bar{x}^*_h)-2n_h\bar{x}^*_h}\right\}=f(0)+f'(0)\xi_{1h}+f''(0)\frac{\xi_{1h}^2}{2!}=1-\frac{N_h}{(2N_h-2n_h)\bar{X}_h}\xi_{1h}+\frac{N_h(3N_h-4n_h)}{(2N_h-2n_h)^2\bar{X}_h}.\frac{\xi_{1h}^2}{2}$$

وبالتعويض:

$$M-\bar{Y}=\sum_{h=1}^{L}\{[\bar{Y}_h+\xi_{0h}][q_{1h}-\frac{1}{2}.\frac{1}{\bar{X}_h}\xi_{1h}.q_{1h}+\frac{3}{8}.\frac{1}{\bar{X}_h^2}.\xi_{1h}^2.q_{1h}+1-\frac{N_h}{(2N_h-2n_h)}\frac{1}{\bar{X}_h}\xi_{1h}+$$
$$\frac{N_h(3N_h-4n_h)}{2(2N_h-2n_h)^2}.\frac{1}{\bar{X}_h^2}.\xi_{1h}^2-q_{1h}+\frac{N_h}{(2N_h-2n_h)}.\frac{1}{\bar{X}_h}.q_{1h}.\xi_{1h}-\frac{N_h(3N_h-4n_h)}{2(2N_h-2n_h)^2}.\frac{1}{\bar{X}_h^2}.q_{1h}.\xi_{1h}^2]\}$$
$$=\sum_{h=1}^{L}[q_{1h}\bar{Y}_h+q_{1h}\xi_{0h}-\frac{1}{2}\frac{\bar{Y}_h}{\bar{X}_h}\xi_{1h}q_{1h}-\frac{1}{2}\frac{1}{\bar{X}_h}\xi_{0h}\xi_{1h}q_{1h}+\frac{3}{8}.\frac{\bar{Y}_h}{\bar{X}_h^2}.\xi_{1h}^2.q_{1h}+\frac{3}{8}.\frac{1}{\bar{X}_h^2}.\xi_{0h}\xi_{1h}^2.q_{1h}$$
$$\bar{Y}_h+\xi_{0h}-\frac{N_h}{2N_h-2n_h}\frac{\bar{Y}_h}{\bar{X}_h}\xi_{1h}-\frac{N_h}{2N_h-2n_h}\frac{1}{\bar{X}_h}\xi_{0h}\xi_{1h}+\frac{N_h(3N_h-4n_h)}{2(2N_h-2n_h)^2}.\frac{\bar{Y}_h}{\bar{X}_h^2}.\xi_{1h}^2+$$
$$\frac{N_h(3N_h-4n_h)}{2(2N_h-2n_h)^2}.\frac{1}{\bar{X}_h^2}.\xi_{0h}.\xi_{1h}^2-q_{1h}\bar{Y}_h-q_{1h}\xi_{0h}+\frac{N_h}{2N_h-2n_h}\frac{\bar{Y}_h}{\bar{X}_h}\xi_{1h}q_{1h}+\frac{N_h}{2N_h-2n_h}\frac{1}{\bar{X}_h}\xi_{0h}\xi_{1h}q_{1h}$$
$$-\frac{N_h(3N_h-4n_h)}{2(2N_h-2n_h)^2}.\frac{\bar{Y}_h}{\bar{X}_h^2}.\xi_{1h}^2q_{1h}-\frac{N_h(3N_h-4n_h)}{2(2N_h-2n_h)^2}.\frac{1}{\bar{X}_h^2}.\xi_{0h}.\xi_{1h}^2q_{1h}-\bar{Y}_h]$$

و بعد الإصلاح و الاختزال و بإهمال الحدود فوق التربيعية ل ξ_{0h},ξ_{1h} نجد :

$$M-\bar{Y}=\sum_{h=1}^{L}\{\xi_{01}-\frac{1}{2}[\frac{N_h}{N_h-n_h}-q_{1h}(\frac{n_h}{N_h-n_h})]\frac{\bar{Y}_h}{\bar{X}_h}\xi_{1h}+\frac{1}{8}.\frac{\bar{Y}_h}{\bar{X}_h^2}[(\frac{N_h}{N_h-n_h}.\frac{3N_h-4n_h}{N_h-n_h})$$
$$-q_{1h}[(\frac{N_h}{N_h-n_h}.\frac{3N_h-4n_h}{N_h-n_h})-3]]\xi_{1h}^2-\frac{1}{2\bar{X}_h}[\frac{N_h}{N_h-n_h}-q_{1h}(\frac{n_h}{N_h-n_h})]\xi_{01}\xi_{1h}\}$$

ولحساب الانحياز:

$$Bias(M)=E(M-\bar{Y})=E(\sum_{h=1}^{L}[\frac{1}{8}\frac{\bar{Y}_h}{\bar{X}_h}\alpha_h\xi_{1h}^2-\frac{1}{2\bar{X}_h}\beta_h\xi_{0h}\xi_{1h}])$$

حيث:

$$\alpha_h=[(\frac{N_h}{N_h-n_h}.\frac{3N_h-4n_h}{N_h-n_h})-q_{1h}[(\frac{N_h}{N_h-n_h}.\frac{3N_h-4n_h}{N_h-n_h})-3]$$
$$\beta_h=\frac{N_h}{N_h-n_h}-q_{1h}(\frac{n_h}{N_h-n_h})$$
$$Bias(M)=\lambda_h[\frac{\bar{Y}_h}{8\bar{X}_h^2}\alpha_h(S_{hX}^2+S_{VX}^2)-\frac{1}{2\bar{X}_h}\beta_h\rho_{hYX}S_{hY}S_{hX}]+\theta_h[\frac{\bar{Y}_h}{8\bar{X}_h^2}\alpha_h(S_{hX(2)}^2+S_{VX(2)}^2)-\frac{1}{2\bar{X}_h}\beta_h\rho_{hYX(2)}S_{hY(2)}S_{hX(2)}]$$

ويحسب متوسط مربعات الخطأ بالعلاقة:

$$MSE(M)=E(M-\bar{Y})^2$$

وبتربيع طرفي العلاقة $M-\bar{Y}$ وحساب التوقع مع أخذ القوى حتى المرتبة التربيعية فقط :

$$E(M-\bar{Y})^2 = E(\sum_{h=1}^{L}[\xi_{0h}-\frac{1}{2}\beta_h\frac{\bar{Y}_h}{\bar{X}_h}\xi_{1h}]^2) = \sum_{h=1}^{L}E(\xi_{0h}^2-\beta_h\frac{\bar{Y}_h}{\bar{X}_h}\xi_{0h}\xi_{1h}+\frac{1}{4}\beta_h{}^2\frac{\bar{Y}_h^2}{\bar{X}_h^2}\xi_{1h}^2)$$

$$MSE(M) = \sum_{h=1}^{L}P_h^2\{\lambda_h\bar{Y}_h^2[C_{hY}^2+\frac{1}{4}\beta_h^2.C_{hX}^2-\beta_h\rho_{hYX}C_{hY}C_{hX}]+\theta_h\bar{Y}_h^2[C_{hY(2)}^2+\frac{1}{4}\beta_h^2C_{hX(2)}^2-$$

$$\beta_h\rho_{hYX(2)}C_{hY(2)}C_{hX(2)}]+\lambda_h\bar{Y}_h^2[\frac{S_{hU}^2}{\bar{Y}_h^2}+\frac{1}{4}\beta_h^2\frac{S_{hV}^2}{\bar{X}_h^2}]+\theta_h\bar{Y}_h^2[\frac{S_{hU(2)}^2}{\bar{Y}_h^2}+\frac{1}{4}\beta_h^2\frac{S_{hV(2)}^2}{\bar{X}_h^2}]\}$$

ولإيجاد القيمة المثلى لمتوسط مربعات الخطأ نحن بحاجة لإيجاد قيمة β_h التي تجعله أصغر ما يمكن وبالتالي نقوم بإيجاد المشتق الأول ل $MSE(M)$ بالنسبة ل ونساويه بالصفر و نقوم بحساب β_h :

$$\frac{\partial MSE(M)}{\partial\beta_h}=0\Rightarrow 2[\frac{1}{4}\lambda_h\bar{Y}_h^2C_{hX}^2+\frac{1}{4}\theta_h\bar{Y}_h^2C_{hX(2)}^2+\frac{1}{4}\lambda_h\frac{S_{hV}^2}{\bar{X}_h^2}\bar{Y}_h^2+\frac{1}{4}\theta_h\frac{S_{hV(2)}^2}{\bar{X}_h^2}\bar{Y}_h^2]\beta_h-$$

$$[\lambda_h\bar{Y}_h^2\rho_{hYX}C_{hY}C_{hX}+\theta_h\bar{Y}_h^2\rho_{hYX(2)}C_{hY(2)}C_{hX(2)}]=0$$

$$\Rightarrow$$

$$\beta_h(0pt)=\frac{2[\lambda_h\rho_{hYX}C_{hY}C_{hX}-\theta_h\rho_{hYX(2)}C_{hY(2)}C_{hX(2)}]}{\lambda_h[\frac{S_{hX}^2+S_{hV}^2}{\bar{X}_h^2}]+\theta_h[\frac{S_{hX(2)}^2+S_{hV(2)}^2}{\bar{X}_h^2}]}$$

المقارنة مع المقدّرات السابقة :

للمقارنة بين المقدّرات من حيث الكفاءة نقارن متوسط مربعات الخطأ مع الأخذ بعين الاعتبار أن وجود خطأ في القياس أو عدم الاستجابة أو الاثنان معا يؤدي إلى زيادة متوسط مربعات الخطأ للمقدّرات، وبالتالي كلما حاولنا التقليل من المصطلحات S_U^2 وS_V^2 و$S_{U(2)}^2$ و$S_{V(2)}^2$ و$S_{X(2)}^2$ و$S_{Y(2)}^2$ كلما قللنا من متوسط مربعات الخطأ MSE،أي كلما حصلنا على أمثال أقل لهذه المصطلحات كلما حصلنا على مقدّر أكثر كفاءة،ولكي يكون المقدّر المقترح في بحثنا M هو الأفضل أو الأكثر كفاءة يجب أن يتحقق :

MSE(M)=Min

بعد الطرح والاختزال نلاحظ:

- إن:

$$MSE(T_R)-MSE(M)>0$$
$$MSE(T_S)-MSE(M)>0$$

إذا كان $1-\frac{1}{4}\beta_h^2>0$ وهو محقق عندما :

$$\frac{\lambda_h\rho_{hYX}C_{hY}C_{hX}-\theta_h\rho_{hYX(2)}C_{hY(2)}C_{hX(2)}}{\lambda_h[\frac{S_{hX}^2+S_{hV}^2}{\bar{X}_h^2}]+\theta_h[\frac{S_{hX(2)}^2+S_{hV(2)}^2}{\bar{X}_h^2}]}<1$$

وهذا الشرط محقق دوما لأن: $0<\rho_{hXY}<1$

- كما أنه:

$$MSE(T_{SK}) - MSE(M) > 0$$

إذا كان:

$$4-\frac{1}{4}\beta_h^2 > 0 \Rightarrow \{\frac{\lambda_h \rho_{hYX} C_{hY} C_{hX} - \theta_h \rho_{hYX(2)} C_{hY(2)} C_{hX(2)}}{\lambda_h[\frac{S_{hX}^2+S_{hV}^2}{\bar{X}_h^2}]+\theta_h[\frac{S_{hX(2)}^2+S_{hV(2)}^2}{\bar{X}_h^2}]}\} > 4$$

$$\Rightarrow$$

$$\frac{\lambda_h \rho_{hYX} C_{hY} C_{hX} - \theta_h \rho_{hYX(2)} C_{hY(2)} C_{hX(2)}}{\lambda_h[\frac{S_{hX}^2+S_{hV}^2}{\bar{X}_h^2}]+\theta_h[\frac{S_{hX(2)}^2+S_{hV(2)}^2}{\bar{X}_h^2}]} > 2$$

وهو محقق دوما محقق حيث : $\rho_{hXY} < 1 \Rightarrow \rho_{hXY}^2 << 1$

- وأيضا:

$$MSE(T_A) - MSE(M) > 0$$

إذا كان:

$$\frac{1}{4}[\frac{N+2n}{N-n}]^2 - \frac{1}{4}\beta_h^2 > 0 \Rightarrow [\frac{N+2n}{N-n}] > \beta_h$$

وهو محقق دوما ذلك لأن: $N+2n > N-2 \Rightarrow \frac{N+2n}{N-n} > 1$

حيث: $\beta_h^2 < 1; \rho_{hXY}^2 << 1$

أي أن مقدرنا أفضل عند تحقق الشروط الأربعة السابقة.

دراسة محاكاة:

بالنسبة للدراسة التجريبية,قمنا بتوليد مجتمعين خاضعين للتوزيع الطبيعي,لكل منهما معالم مختلفة,واستخدمنا بيانات قياسية لأربع مجتمعات مستخدمة سابقا وقمنا بحساب معالمها وذلك باستخدام لغة R,فكانت نتائج المقارنات كما يلي:

المجتمع الأول:

$$X_1 = rnorm(1500,3,7), Y_1 = X_1 + rnorm(1500,0,1), y_1 = Y_1 + rnorm(1500,1,3)$$
$$x_1 = X_1 + rnorm(1500,1,3)$$
$$X_2 = rnorm(1300,4,9), Y_2 = X_2 + rnorm(1300,0,1), y_2 = Y_2 + rnorm(1300,1,3)$$
$$x_2 = X_2 + rnorm(1300,0,3)$$
$$X_3 = rnorm(1200,4,8), Y_3 = X_3 + rnorm(1200,0,1), y_3 = Y_3 + rnorm(1200,1,3)$$
$$x_3 = X_3 + rnorm(1200,0,1)$$
$$X_4 = rnorm(1000,5,10), Y_4 = X_4 + rnorm(1000,0,1), y_4 = Y_4 + rnorm(1000,1,3)$$
$$x_4 = X_4 + rnorm(1000,0,1)$$

$N_1 = 1000, N_2 = 1200, N_3 = 1300, N_4 = 1500, n_1 = 200, n_2 = 210, n_3 = 220, n_4 = 215$

$\bar{Y}_1 = 3.20716, \bar{Y}_2 = 4.582486, \bar{Y}_3 = 3.79172, \bar{Y}_4 = 5.670808$

$\bar{X}_1 = 3.241139, \bar{X}_2 = 4.627208, \bar{X}_3 = 4.627208, \bar{X}_4 = 5.666893$

$S^2_{1Y} = 53.52266, S^2_{2Y} = 84.23239, S^2_{3Y} = 66.3915, S^2_{4Y} = 106.3107$

$S^2_{1X} = 52.84269, S^2_{2X} = 82.98812, S^2_{3X} = 65.46337, S^2_{4X} = 104.2774$

$\rho_{1YX} = 0.9903624, \rho_{2YX} = 0.9939457, \rho_{3YX} = 0.9924618, \rho_{4YX} = 0.9950779$

Estimator	10%من عدم الاستجابة			20%من عدم الاستجابة		
	k_h			k_h		
	2	4	8	2	4	8
T_{HH}	0.094074	0.114968	0.156755	0.102687	0.140807	0.217047
T_R	0.012512	0.015776	0.025523	0.015675	0.031452	0.14012
T_S	0.012504	0.034279	0.038896	0.034520	0.030145	0.031452
T_{SK}	0.027283	0.015799	0.100546	0.042132	0.024125	0.074562
T_A	0.012567	0.015775	0.028564	0.017412	0.015114	0.041422
M	0.012463	0.014008	0.022157	0.014145	0.012245	0.030232

المجتمع الثاني:

$X_1 = rnorm(1000, 4, 9), Y_1 = X_1 + rnorm(1000, 0, 1), y_1 = Y_1 + rnorm(1000, 1, 3)$

$x_1 = X_1 + rnorm(1000, 1, 3)$

$X_2 = rnorm(1000, 5, 10), Y_2 = X_2 + rnorm(1000, 0, 1), y_2 = Y_2 + rnorm(1000, 1, 3)$

$x_2 = X_2 + rnorm(1000, 1, 3)$

$X_3 = rnorm(1000, 3, 7), Y_3 = X_3 + rnorm(1000, 0, 1), y_3 = Y_3 + rnorm(1000, 1, 3)$

$x_3 = X_3 + rnorm(1000, 1, 3)$

$X_4 = rnorm(1000, 4, 8), Y_4 = X_4 + rnorm(1000, 0, 1), y_4 = Y_4 + rnorm(1000, 1, 3)$

$x_4 = X_4 + rnorm(1000, 1, 3)$

عدد الطبقات =4

$N_1 = 1000, N_2 = 1000, N_3 = 1000, N_4 = 1000, n_1 = 200, n_2 = 200, n_3 = 200, n_4 = 200$

$\bar{Y}_1 = 4.039068, \bar{Y}_2 = 5.670898, \bar{Y}_3 = 2.902825, \bar{Y}_4 = 3.616159$

$\bar{X}_1 = 3.968049, \bar{X}_2 = 5.666863, \bar{X}_3 = 2.918596, \bar{X}_4 = 3.643237$

$S^2_{1Y} = 81.46952, S^2_{2Y} = 106.310, S^2_{3Y} = 45.99937, S^2_{4Y} = 67.69053$

$S^2_{1X} = 81.06883, S^2_{2X} = 104.2774, S^2_{3X} = 45.99937, S^2_{4X} = 66.19725$

$\rho_{1YX} = 0.9939164, \rho_{2YX} = 0.9950779, \rho_{3YX} = 0.9896319, \rho_{4YX} = 0.9926346$

Estimator	10%من عدم الاستجابة	20%من عدم الاستجابة

	k_h			k_h		
	2	4	8	2	4	8
T_{HH}	0.095391	0.116525	0.158794	0.105952	0.148191	0.232694
T_R	0.044213	0.021475	0.035124	0.023412	0.028415	0.04578
T_S	0.023214	0.034279	0.037221	0.031123	0.031454	0.05748
T_{SK}	0.031214	0.021741	0.051145	0.034154	0.064545	0.21456
T_A	0.024012	0.032001	0.040021	0.024412	0.027415	0.04956
M	0.021021	0.021120	0.031257	0.017845	0.024515	0.03845

المجتمع الثالث:

بيانات قياسية [4] حيث: عدد الطلاب: X ، عدد المدرسين: Y، عدد الطبقات= 6

$N_1 = 106, N_2 = 106, N_3 = 94, N_4 = 171, N_5 = 24, N_6 = 173$

$n_1 = 15, n_2 = 15, n_3 = 12, n_4 = 20, n_5 = 23, n_6 = 15$

$\bar{Y}_1 = 1563.774, \bar{Y}_2 = 2212.594, \bar{Y}_3 = 9384.309, \bar{Y}_4 = 5588.012,$

$\bar{Y}_5 = 966.955, \bar{Y}_6 = 404.3988, \bar{X}_1 = 24375.59, \bar{X}_2 = 27421.70,$

$\bar{X}_3 = 72409.95, \bar{X}_4 = 74364.68, \bar{X}_5 = 26441.72, \bar{X}_6 = 9842.15$

$S^2_{1Y} = 41281746, S^2_{2Y} = 133437791, S^2_{3Y} = 894457433, S^2_{4Y} = 820445636,$

$S^2_{5Y} = 5710999, S^2_{6Y} = 894440.3, S^2_{1X} = 2419565835, S^2_{2X} = 3301722268,$

$S^2_{3X} = 25842911895, S^2_{4X} = 81569146488, S^2_{4X} = 81569146488, S^2_{5X} = 2061412416$

$S^2_{6X} = 353245374, \rho_{1YX} = 0.8156414, \rho_{2YX} = 0.815414, \rho_{3YX} = 0.9011201$

$\rho_{4YX} = 0.9858761, \rho_{5YX} = 0.7130988, \rho_{6YX} = 0.893595.$

Estimator	10%من عدم الاستجابة			20%من عدم الاستجابة		
	k_h			k_h		
	2	4	8	2	4	8
T_{HH}	271948.3	3340426	4582312	3025218	4287631	6722457
T_R	489521.3	478565.6	475632.4	412354.2	385675.4	542573
T_S	245697.8	312456.5	345212.1	310245.7	365784.7	612354.5
T_{SK}	325461.5	298564.5	241578.3	287456.2	412578.2	510254.8
T_A	251453.4	321456.7	247845	252451.2	368741.8	624532.2
M	214526.5	224124.4	237845.4	241023.5	343345.7	461250.5

المجتمع الرابع:

بيانات قياسية [11] حيث الناتج القومي الإجمالي عام 1982: X، حجم المجتمع (عدد السكان) بالمليون: Y، عدد الطبقات= 5

$$N_1 = 38, N_2 = 14, N_3 = 11, N_4 = 33, N_5 = 24, n_1 = 17, n_2 = 6, n_3 = 4, n_4 = 12, n_5 = 11$$
$$\bar{Y}_1 = 13.03684, \bar{Y}_2 = 27.35, \bar{Y}_3 = 23.13636, \bar{Y}_4 = 79.65455, \bar{Y}_5 = 20.2833$$
$$\bar{X}_1 = 1029.158, \bar{X}_2 = 25671.57, \bar{X}_3 = 5028.818, \bar{X}_4 = 7533.939, \bar{X}_5 = 16315.25$$
$$S_{1Y}^2 = 270.9083, S_{2Y}^2 = 3906.929, S_{3Y}^2 = 1339.405, S_{4Y}^2 = 45082.17, S_{5Y}^2 = 368.9423$$
$$S_{1X}^2 = 3667896, S_{2X}^2 = 6568461403, S_{3X}^2 = 63348743, S_{4X}^2 = 440717912,$$
$$S_{5X}^2 = 40844121, \rho_{1YX} = 0.7439544, \rho_{2YX} = 0.969956, \rho_{3YX} = 0.9768227,$$
$$\rho_{4YX} = 0.2948897, \rho_{5YX} = 0.9011072$$

Estimator	10%من عدم الاستجابة			20%من عدم الاستجابة		
	k_h			k_h		
	2	4	8	2	4	8
T_{HH}	219.5673	280.1731	401.3808	284.9066	386.1869	648.7416
T_R	179.2575	321.7635	420.7414	187.5365	452.3623	685.4642
T_S	1014.5359	956.1247	178.7683	325.4594	274.3145	545.2432
T_{SK}	1152.3358	1452.6753	1243.4359	1243.4156	1025.7453	1145.6424
T_A	148.1247	741.2596	420.5086	177.5756	189.4668	325.4505
M	128.8993	423.4513	156.4462	146.5203	187.6756	213.5535

المجتمع الخامس:

بيانات قياسية [4] حيث: مساحة الأرض بالهكتار : X، إنتاج القمح بالطن: Y، عدد الطبقات= 4

$$N_1 = 47, N_2 = 30, N_3 = 29, N_4 = 13, n_1 = 15, n_2 = 10, n_3 = 10, n_4 = 5$$
$$\bar{Y}_1 = 443.5447, \bar{Y}_2 = 68.68276, \bar{Y}_3 = 17.06667, \bar{Y}_4 = 52.52308$$
$$\bar{X}_1 = 160.262, \bar{X}_2 = 29.70345, \bar{X}_3 = 11.54667, \bar{X}_4 = 23.62308$$
$$S_{1Y}^2 = 74026.75, S_{2Y}^2 = 2871.781, S_{3Y}^2 = 244.1282, S_{4Y}^2 = 4451.124$$
$$S_{1X}^2 = 8377.401, S_{2X}^2 = 316.4532, S_{3X}^2 = 91.45775, S_{4X}^2 = 682.9703$$
$$\rho_{1YX} = 0.9583838, \rho_{2YX} = 0.779071, \rho_{3YX} = 0.8719665, \rho_{4YX} = 0.9922591$$

Estimator	10%من عدم الاستجابة			20%من عدم الاستجابة		
	k_h			k_h		
	2	4	8	2	4	8
T_{HH}	249.1903	369.0362	608.7280	354.24	386.1843	648.7410
T_R	0.4532	0.7454	0.9801	1.4531	6.1453	10.4523
T_S	0.4231	0.6124	0.7453	1.6231	5.9862	11.2453
T_{SK}	322.1465	721.4321	965.2	243.2530	652.4521	886.1236
T_A	0.4123	0.4254	0.7856	1.6231	5.7856	10.2546
M	0.1452	0.3412	0.7231	1.3321	4.8962	10.0035

المجتمع السادس:

بيانات قياسية [11] حيث: حجم المجتمع (عدد السكان) عام 1982: X، حجم المجتمع (عدد السكان) عام 1983: Y، عدد الطبقات= 5

$N_1 = 38, N_2 = 14, N_3 = 11, N_4 = 33, N_5 = 24, n_1 = 17, n_2 = 6, n_3 = 4, n_4 = 12, n_5 = 11$

$\bar{Y_1} = 13.03684, \bar{Y_2} = 27.35, \bar{Y_3} = 23.13636, \bar{Y_4} = 79.65455, \bar{Y_5} = 20.2833$

$\bar{X_1} = 11.88421, \bar{X_2} = 26.18571, \bar{X_3} = 21.88182, \bar{X_4} = 75.24242, \bar{X_5} = 20.09582$

$S_{1Y}^2 =, S_{2Y}^2 = 3906.929, S_{3Y}^2 = 1339.405, S_{4Y}^2 = 45082.17, S_{5Y}^2 = 368.9423$

$S_{1X}^2 = 222.4889, S_{2X}^2 = 3683.071, S_{3X}^2 = 1174.032, S_{4X}^2 = 41280.19,$

$S_{5X}^2 = 364.7839, \rho_{1YX} = 0.9996193, \rho_{2YX} = 0.9998693, \rho_{3YX} = 0.9998858,$

$\rho_{4YX} = 0.9993071, \rho_{5YX} = 0.9998059$

Estimator	10%من عدم الاستجابة			20%من عدم الاستجابة		
	k_h			k_h		
	2	4	8	2	4	8
T_{HH}	630.3935	835.3869	1155.3740	713.2246	1053.8800	1735.1910
T_R	62.1478	67.4523	98.9986	187.5	92.5238	210.4566
T_S	56.4853	64.1548	98.2136	325.4	89.1436	188.1253
T_{SK}	624.5862	795.2598	1052.4569	943.4	875.1258	1658.1258
T_A	62.3542	64.1258	98.2368	177.5	87.2546	179.1456
M	55.4586	63.4569	97.4538	146.5	85.2658	149.3256

References:

1- Azeem; Hanif M." *Joint influence of measurement error and non response on estimation of population mean"*. Communications in Statistics-Theory and Methods. (2017).46(4):1679±16

2- Cochran, W.G. *"Sampling Techniques, 3rd Edition*". New York: John Wiley & Sons.(1977). Inc5

3- Cochran, W.G. *"Errors of measurement in statistics*". Technometrics.(1990). 10(4), 637-666.

4-"FBS." *Crops area production by districts"*, Islamabad; 2011.

5- Hansen, M.H; Hurwitz, W.N.*"The problem of non-response in sample surveys*". J. Amer. Statist. Assoc.(1946).41, 517-529 Statistical Planning and Inference.(2009). 139(8):2552±2558

6- Okafor, F.C;Lee, H." *Double sampling for ratio and regression estimation with sub sampling the non-respondent*". Survey Methodology . (2000). 26, 183-188.

7- Rabie, Abdel Hamid, Samra, Adel; Al-Sayyad, Jalal "*An Introduction to Statistics for Students of Economic and Administrative Studies*." First Edition, King Abdulaziz University, Saudi Arabia, (2015).

8- Rao, P.S.R.S."*Ratio estimation with sub sampling the non-respondents*". Survey Methodology. (1986). 12(2), 217-230.
9- Rao, P.S.R.S. *"Ratio and Regression estimates with sub sampling of non-respondents. Statistical Association Meetings".* Sept".(1987)., 2-16, Tokyo, Japan.
10- Rao, P.S.R.S.*"Regression estimators with sub sampling of non-respondents".* In Data Quality Control. Theory and Pragmatics, (Gunar E. Liepins and V.R.R. Uppuluri,
eds) Marcel Dekker, New York, . (1990). 191-2008.
11- SaÈrndal C, Swensson B, Wretman J. *"Model Assisted Survey Sampling*". New York: Springer. (1992).
12-Singh, H.P; Kumar, S."*Estimation of mean in presence of non-response using two-phase sampling scheme*". Statistical Papers. (2008). 51, 559-582.
13-Singh, H.P; Kumar, S. "*A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information*". SORT (2009).33, 71-84.
14- Singh, H.P; Kumar, S. "*Combination of regression and ratio estimate in presence of non-response*". Braz. J. Statist. Assoc. (2011). 25(2), 205-217.
15-Srivastava, S. K. "*Estimation of population mean in repeat surveys in the presence of measurement errors*". J. Ind. Soc. Agri. Statist.(2010). 53:125–33.