تقويم أساليب المعاينة العشوائية الطبقية في تقدير المساحة المزروعة لمحصول الشلب في النجف.

إقبال محمود علوان*، رشا عادل سعيد** و سعاد مهدي رشيد***
* قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.
** قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.
*** كلية العلوم، جامعة النهرين.

الخلاصة

يهدف البحث إلى تقييم الأساليب الإحصائية المستخدمة في تقديرات إنتاج المحاصيل الحقلية الرئيسية وهو الرز والوصول إلى السبل الفضلى التي تزيد من دقة التقديرات. وباستخدام أنواع مختلفة من مقدرات المعاينة ,تمت المقارنة بين تباين متوسط المعاينة العشوائية البسيطة ($\bar{y}$)(var) وتباين متوسط المعاينة العشوائية الطبقية ($\bar{y}_{st}$)(var) وتباين متوسط مقدر الانحدار المنفصل ($\bar{y}_{lrs}$)(var) وتباين متوسط مقدر النسبة التجميعي ($\hat{\bar{y}}_{RC}$)(var) بينت نتائج المقارنة بعد التطبيق إن أساليب التقدير لمقدر النسبة التجميعي ومقدر الانحدار المنفصل تعطي أفضل التقديرات,ولغرض الحصول على أفضل تقسيم لطبقات المعاينة العشوائية الطبقية فقد استخدمت طريقة $cum.f^{4/5}$ التقريبية في تعيين الحدود المثلى للطبقات وتم وضع تقسيمات طبقية جديدة بموجبها,تم حساب تباين متوسط المعاينة العشوائية الطبقية وفق التقسيمات الطبقية الجديدة وبالمقارنة مع التقسيم المتبع حالياً في الجهاز المركزي للإحصاء فأن الطبقات الجديدة أدت إلى تقليل التباين بشكل كبير وواضح وبناءاً عليه فقد أوصى البحث بإتباع التقسيمات الطبقية الجديدة التي اقترحتها الدراسة.

المقدمة والهدف

تعتمد مديرية الإحصاء الزراعي في الجهاز المركزي للإحصاء في تقديرات إنتاج المحاصيل الحقلية على أسلوب المعاينة العشوائية الطبقية وتستخدم في ذلك أسلوبين مختلفين من أساليب المسح الأسلوب الأول هو المسح بالعينة الخاضع للطرائق الموضوعية (Objective Methods) والأسلوب الثاني هو أسلوب العد الشامل (Complete enumerate) [3]. تم في هذا البحث إيجاد أفضل الطرائق الممكنة لتقدير المساحة المزروعة لمحصول الشلب واستخدمت في ذلك البيانات الخاصة بعام 2008 من خلال تقويم الأساليب المتبعة في المعاينة العشوائية الطبقية و دراسة الأسس المتبعة وفق المعاينة العشوائية الطبقية وباستخدام أنواع مختلفة من مقدرات المعاينة تمت المقارنة بين تباين متوسط المعاينة العشوائية البسيطة وتباين متوسط المعاينة العشوائية الطبقية وتباين متوسط مقدر الانحدار المنفصل وتباين متوسط مقدر النسبة التجميعي وبينت نتائج المقارنة بعد التطبيق أن أساليب التقدير لمقدر النسبة التجميعي تعطي أفضل التقديرات لأنها تحقق أقل تباين ولغرض الحصول على أفضل تقسيم لطبقات المعاينة العشوائية الطبقية فقد استخدمت طريقة $cum.f^{4/5}$ التقريبية في تعيين الحدود المثلى للطبقات وتم وضع تقسيمات طبقية جديدة بموجبها تم حساب تباين متوسط المعاينة الطبقية وفق التقسيمات الطبقية الجديدة وبالمقارنة مع التقسيم المتبع حالياً في الجهاز المركزي للإحصاء فأن الطبقات الجديدة أدت إلى تخفيض ملحوظ في التباين.

1-الجانب النظري.

المعاينة العشوائية البسيطة [2]

Simple Random Sampling

في المعاينة العشوائية البسيطة يكون تقدير التباين للمتوسط $v(\bar{y})$ هو:

$$\text{var}(\bar{y}) = \frac{S^2}{n}(1-f) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1.1)$$

حيث إن $f = n/N$ ويسمى كسر المعاينة Sampling Fraction

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

المعاينة العشوائية الطبقية [1]

Stratified Random Sampling

في المعاينة العشوائية الطبقية يكون تقدير التباين للمتوسط $v(\bar{y}_{st})$ هو:

$$\text{var}(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h}(1-f_h) \quad \ldots\ldots\ldots\ldots (1.2)$$

حيث ان:

$$S_h^2 = \frac{\sum_{i=1}^{nh} (y_{hi} - \bar{y}_h)^2}{n_h - 1}$$

و

$$f_h = n_h / N_h$$

و

$$W_h = N_h / N$$

وزن (ترجيحية) الطبقة Stratum weight وان:

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$

تقدير الانحدار المنفصل [4]

Separate Regression Estimate

في مقدر الانحدار المنفصل يكون تباين المتوسط $var(\bar{y}_{lrs})$ هو:

$$\text{var}(\bar{y}_{lrs}) = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)}{n_h}\left[\frac{1}{n_h - 2}\right]\left[\sum_{i}^{nh}(y_{hi} - \bar{y}_h)^2 - b_h^2 \sum_{i}^{h}(x_{hi} - \bar{x}_h)^2\right] \quad \ldots\ldots\ldots\ldots (1.3)$$

تقدير النسبة التجميعي [6]

Estimate Combined Ratio

يأخذ نسبة تجميعية واحدة لكافة الطبقات ومن خلال اخذ تقدير لمجموعي المجتمعين y,x .

إذا رمزنا لتقدير النسبة التجميعي بالرمز $\hat{Y}_{RC}$ (C ترمز لكون التقدير تجميعيا) فان:

$$\hat{Y}_{RC} = \frac{\hat{Y}_{ST}}{\hat{X}_{ST}} X = \frac{\bar{y}_{st}}{\bar{x}_{st}} X = \hat{R}_C X$$

حيث أن:

$$\hat{Y}_{st} = \sum N_h \bar{y}_h$$

$$\hat{X}_{st} = \sum N_h \bar{x}_h$$

فإذا كان حجم العينة الكلي كبيرا جدا يكون تباين المتوسط لمقدر النسبة التجميعي كالأتي:

$$var(\hat{\bar{Y}}_{RC}) = \sum \frac{N_h^2(1-f_h)}{n_h}(S_{yh}^2 + R^2 S_{xh}^2 - 2R\, r_h S_{yh} S_{xh}) \quad \ldots\ldots\ldots\ldots (1.4)$$

بناء الطبقات والحدود التقريبية المثلى,[4] ،[5]

لغرض تقليل التباين في تقديرات العينة يتم تقسيم المجتمع إلى طبقات بافتراض إن الطبقات تكونت باستخدام متغير الدراسة (Y) وبافتراض إن n, L ثوابت فأن طريقة تشكيل الحدود المثلى للطبقات على ضوء تقليل تباين المتوسط الطبقي تكون بالاعتماد على الطريقة الآتية:

الطريقة التقريبية المستخدمة

The using Approximated method

طريقة Cum.f $^{4/5}$:

بفرض أن:

$$z(y) = \int_a^b f^{4/5}(y)\,dy$$

وان هذه القيم لها دالة كثافة احتمالية Probability density function f(y) وتباين محدد σ في الفترة [a,b] و بتجزئة $Cum.f^{4/5}$ إلى مديات (Ranges) متساوية فان يكون:

$$Z_h(y) = \int_{y_{h-1}}^{y_h} f^{4/5}(y)\ dy, \quad h = 1,2,\ldots,L-1$$

وبافتراض إن μ_h هو المتوسط التقريبي لـf(y) داخل الطبقة h فان :

$$Z_h(y) = \int_{y_{h-1}}^{y_h} m_h^{4/5}\, dy = m_h^{4/5}(y_h - y_{h-1})$$

$$m_h^{4/5} = \frac{Z_h(y)}{(y_h - y_{h-1})}$$

$$m_h = \left[\frac{Z_h(y)}{(y_h - y_{h-1})}\right]^{5/4}$$

أن تباين المتوسط الطبقي باستخدام التوزيع الأمثل عند إهمال كسر المعاينة هو:

$$V_{opt}(\bar{y}_{st}) = \frac{1}{n}\left(\sum_{h=1}^{L} W_h \sigma_h\right)^2$$

حيث ان:

$$W_h = m_h(y_h - y_{h-1})$$

$$s_h^2 = \frac{(y_h - y_{h-1})^2}{12}$$

وبالتعويض عن قيمة σ_h , μ_h,W_h يكون:

$$V_{opt}(\bar{y}_{st}) = \frac{1}{n}\left[\sum_{h=1}^{L} M_h(y_h - y_{h-1})\frac{(y_h - y_{h-1})}{(12)^{1/2}}\right]^2$$

$$V_{opt}(\bar{y}_{st}) = \frac{1}{12n}\left[\sum_{h=1}^{L}\frac{Z_h^{5/4}(y)}{(y_h - y_{h-1})^{5/4}}(y_h - y_{h-1})^2\right]^2$$

$$= \frac{1}{12n}\left[\sum_{h=1}^{L} Z_h^{5/2}(y)\right] \quad \text{.....(1.5)}$$

إن المعادلة السابقة تصبح أقل ما يمكن عندما تكون $Z_h(y)$ ثابتا لجميع قيم h وذلك باستخدام مضاعف لاكرانج تحت القيد $\sum_{h=1}^{L} Z_h(y) = Z(y)$ هو مستقل

عند اختيار حدود الطبقات أي إن:

$$Z_h(y) = \frac{Z(y)}{L}$$

وان تباين المتوسط الطبقي في هذه الحالة يصبح :

$$V_{opt}(\bar{y}_{st}) = \frac{Z^{5/2}(y)}{12nL^2}$$

إذ إن تباين متوسط المعاينة الطبقية للطريقة التقريبية $Cum.f^{m/n}$ يكون باستخدام القاعدة العامة التالية :

$$V_{opt}(\bar{y}_{st}) = \frac{Z^{2n/m}(y)}{12nL^2}$$

حيث إن :

$$Z(y) = \int_{y_{h-1}}^{y_h} f^{m/n}\, dy$$

2- الجانب التطبيقي.

تم في هذا البحث استخدام البيانات الخاصة بمحصول الشلب لعام 2008 لمحافظة النجف واقضينها حيث تحتل محافظة النجف المرتبة الأولى من بين المحافظات في القطر من حيث المساحة المزروعة والإنتاج لمحصول الشلب حيث أن نسبة إنتاجها 51.34% لسنة 2008 على مستوى المحافظات وذلك لما تتميز به هذه المحافظة من تربة مناسبة لزراعته.

البيانات وكيفية الحصول عليها .

تم الحصول على جميع البيانات اللازمة من قبل مديرية الإحصاء الزراعي في الجهاز المركزي للإحصاء الخاصة بمحصول الشلب لعام 2008 لأقضية محافظة النجف بما فيها بيانات العينة المقدرة والفعلية* للمساحة المزروعة لكل حائز بالدونم على مستوى الاقضية حيث إن محافظة النجف تتألف من ثلاث أقضية (النجف والمناذرة والكوفة).

تقدير التباين للمتوسط للعينة العشوائية الطبقية والعينة العشوائية البسيطة ولمقدر الانحدار المنفصل ولمقدر النسبة التجميعي.

تم تقدير التباين المتوسط للعينة العشوائية البسيطة والعينة العشوائية الطبقية و لمقدر الانحدار المنفصل ولمقدر النسبة التجميعي. حسب المعادلات (1.1) و(1.2) و(1.3) و(1.4) على التوالي وبمقارنة التباينات نلاحظ إن أعلى تباين متحقق من جراء تطبيق أسلوب المعاينة العشوائية البسيطة وأن أتباع أسلوب المعاينة العشوائية الطبقية يخفض التباين بشكل كبير جدا وعند المقارنة بين نتائج متوسطي مقدر الانحدار المنفصل ومقدر النسبة ألتجميعي نجد إن إتباع أي من الأسلوبين علا وفق المعاينة العشوائية الطبقية يعطيان نتائج أفضل من إتباع أسلوب المعاينة الطبقية لوحدها وان إتباع أسلوب تقدير النسبة التجميعي هو الأفضل لأنه يحقق اقل تباين في جميع الاقضية و النتائج موضحة في الجدول رقم (1) وأن جميع التطبيقات أجريت باستخدام ألبرنامج الإحصائي Minitab 11 for Windows.

الجدول رقم (1)

تباين المتوسطات.

القضاء / تباين المتوسط	النجف	الكوفة	المناذرة
$Var(\bar{y})$	10.4192	367.2138	1717.7833
$Var(\bar{y}_{st})$	0.0073	0.2056	0.6279
$Var(\bar{y}_{lrs})$	0.0072	0.0861	0.0815
$Var(\hat{\bar{y}}_{RC})$	0.0071	0.0088	0.0021

بناء الطبقات والحدود المثلى للطبقات

* **يقصد ببيانات العينة الفعلية هي العينة التي تخضع للقياس والحصاد.**

إن طريقة $Cum.f^{1/2}$ المقترحة من قبل Dalenius & Hodgges تتطلب إن يكون f(y) ثابتا تقريبا أي يتبع التوزيع المنتظم وحيث إن فترات الفئات في التوزيع الأصلي غير متساوية الطول لذا فقد تم إعادة توزيع بيانات الإطار توزيعا منتظما على فترات متساوية واستخدم في هذا التطبيق إطار قضاء المناذرة لمحصول الشلب لكونه يمثل مجتمعا كبيرا نسبيا ويحتوي على مفردات تضم كافة الفئات وبتطبيق عدد من الطرائق التقريبية لمعرفة أي منها تحقق اقل تباين باستخدام المعادلة (1.5) وبحساب التباين الخاص بكل طريقة بافتراض إن حجم العينة الكلية نفسه المستخدم من قبل مديرية الإحصاء الزراعي لغرض أن تكون المقارنة متكافئة بين نتائج التقسيمات الطبقية الجديدة والتقسيمات الطبقية السابقة وبتطبيق عدد من الطرائق التقريبية على البيانات الخاصة بهذا البحث وهذه الطرائق التقريبية المختلفة مع التباينات الخاصة بها موضحة في الجدول رقم (2) والذي يبين إن الطريقة التي تحقق اقل تباين هي الطريقة $Cum.f^{4/5}$.

جدول رقم (2)

الطرائق التقريبية $Cum.f^{m/n}$وتبايناتها.

الطريقة $Cum.f^{m/n}$	$V_{opt}(\bar{y}_{st})$
$Cum.f^{2/6}$	12599811.2700
$Cum.f^{1/2}$	56991.4753
$Cum.f^{4/7}$	15162.1213
$Cum.f^{2/3}$	4469.1719
$Cum.f^{3/4}$	1751.7610
$Cum.f^{4/5}$	1140.4153

وبعد أجراء التطبيق على كل طريقة من هذه الطرائق وذلك بتقسيم $Cum.f^{4/5}$/ L حيث إن (L) تمثل عدد الطبقات المناسبة لتقسيم المجتمع المبحوث وتحديد نقاط التقسيم ومن خلال القيم

المماثلة عندها لـ Y_h كما موضح في الجدول رقم (3).

الجدول رقم (3)

نتائج تقسيم $Cum.f^{4/5}$ على قيم مختلفة لـ (L) لإطار قضاء المناذرة لمحصول الشلب.

L	$Cum.f^{4/5}/L$
3	0.5071
4	0.8803
5	0.3042
6	0.2535

ومن ملاحظة الجدول رقم (4) والذي يمثل النسب المئوية للتكرارات N_h لكل فئة والذي يبين إن نسبة التكرارات في الفئات الأولى تكون كبيرة جدا ثم تنخفض تدريجيا وفي هذه الحالة لا يمكن اعتماد الطرائق التي تسبب زيادة التكرارات في الفئات الأولى وكذلك لا يمكن اعتماد الطرائق التي تجعل الفئات الأخيرة كبيرة.

جدول رقم (4)

النسب المئوية لحجم كل فئة لمحصول الشلب 2008.

الفئة (الطبقة)	حدود الفئة (الطبقة)	النسبة المئوية
الفئة الأولى	اقل من 6	% 22.8000
الفئة الثانية	6 واقل من 12	% 42.8700
الفئة الثالثة	12 واقل من26	% 23.6500
الفئة الرابعة	26 واقل من51	% 6.2400
الفئة الخامسة	51واقل من 101	% 2.4200
الفئة السادسة	101 واقل من 201	% 1.3600
الفئة السابعة	201 واقل من 401	% 0.5100
الفئة الثامنة	401 واقل من 1001	% 0.1000
الفئة التاسعة	1001 واقل من 2000	% 0.0300

التقسيمات وحدود الطبقات المقترحة .

ولغرض تطبيق طريقة $Cum.f^{4/5}$ فقد تم تقسيم الإطار إلى ثلاثة مجاميع ثم تطبيق الطريقة على كل مجموعة بشكل منفصل ثم دمج الفئات للمجموعة الأولى مع الفئات في المجموعة الثانية ثم الفئات في المجموعة الثالثة ومنه تم الحصول على فئات التقسيم والحدود الطبقية المثلى الجديدة. لذلك فقد تم إعادة تشكيل الطبقات التسع بحدود تقريبية مثلى مع مراعاة أن يكون حجم العينة الكلية نفسه المستخدم من

قبل مديرية الإحصاء الزراعي لغرض أن تكون المقارنة متكافئة. وأن أي زيادة في عدد الطبقات بصحبه زيادة في حجم العينة وبالتالي زيادة في الكلفة الكلية لذلك لم يتم زيادة عدد الطبقات. كما موضح في الجدول رقم (5) التالي:

جدول رقم (5)

التقسيم الجديد لحدود الفئات.

الفئة (الطبقة)	حدود الفئة (الطبقة)
الفئة الأولى	اقل من8
الفئة الثانية	8 واقل من 14
الفئة الثالثة	14 واقل من32
الفئة الرابعة	32 واقل من52
الفئة الخامسة	52 واقل من 92
الفئة السادسة	92 واقل من 202
الفئة السابعة	202 واقل من302
الفئة الثامنة	302واقل من 402
الفئة التاسعة	402واقل من 5000

واستنادا إلى هذه الطبقات والحدود الطبقية المثلى الجديدة للفئات فقد تم إعادة تقسيم بيانات الأطر الزراعية بموجبه والخاصة بمحصول الشلب للاقضية بنفس حجم العينة الكلية المستخدم من قبل مديرية الإحصاء الزراعي ولغرض المقارنة بين نتائج التقسيمات الجديدة والتقسيمات الطبقية السابقة فقد تم حساب التباين الطبقي $(\mathrm{Var})\,\bar{y}_{st}$ وفق التقسيمات الجديدة نلاحظ إن الحدود الطبقية المثلى الجديدة أدت إلى خفض التباين للاقضية كافة بشكل واضح وكبير,كما موضح قي الجدول رقم (6) التالي:

جدول رقم (6)

التباين الطبقي $\mathrm{Var}\ \bar{y}_{st}$ وفق الحدود الطبقية المثلى الجديدة لفئات أقضية محافظة النجف.

القضاء	$\mathrm{Var}\ \bar{y}_{st}$
المناذرة	0.1150
الكوفة	0.1205
النجف	0.0072

الاستنتاجات.

1-أثبتت الدراسة كفاءة استخدام التقدير بطريقة مقدر الانحدار المنفصل ومقدر النسبة التجميعي في تقدير مساحة الشلب نجد إن إتباع أي من الأسلوبين على وفق المعاينة العشوائية الطبقية يعطيان نتائج أفضل من إتباع أسلوب المعاينة الطبقية لوحدها وان إتباع أسلوب تقدير النسبة التجميعي هو الأفضل لأنه يحقق اقل تباين.

2-أثبتت طريقة $Cum.f^{4/5}$ التقريبية المستخدمة كفاءتها في تعيين حدود مثلى للطبقات والتي أدت إلى تقليل التباين بشكل كبير وواضح.

التوصيات.

بناءا على الاستنتاجات الواردة سابقاً فإن البحث أعطى التوصيات الآتية:

1-يوصي البحث باستخدام التقسيمات الطبقية الجديدة المقترحة عند أتباع أسلوب المعاينة العشوائية الطبقية في تقديرا لمساحة و الإنتاج لمحصول الشلب لأنه يحقق أقل تباين.

2-استخدام التقدير بطريقة مقدر الانحدار المنفصل ومقدر النسبة التجميعي في تقدير المساحة لمحصول الشلب بدلا من أسلوب المعاينة العشوائية الطبقية الاعتيادية لان هذا الأسلوب يخفض التباين بشكل كبيرو واضح.

المصادر

أولا : المصادر العربية.

[1] الناصر، عبد المجيد حمزة وصفاء يونس الصفاوي (2001) "العينات نظري وتطبيقي" دار الكتب للطباعة والنشر جامعة الموصل.

[2] حمد، عدنان شهاب ومهدي محسن العلاق (2001) "أساليب المعاينة في ميدان التطبيق" المعهد العربي للتدر يب والبحوث، بغداد.

[3] هيئة التخطيط – الجهاز المركزي للإحصاء ,دائرة الإحصاء الزراعي ,2009 "أنتاج الشلب وزهرة الشمس لسنة 2008".

ثانيا:المصادر الأجنبية.

[4] Cochran, W, G., "Sampling Techniques", 3rd Edition, New York: John Wiley and Sons, (1977).
[5] Jones, H. L., "The Analysis of Variance of Data from Stratified Subsample ", JASA, vol.63, (1968), pp.64-86.
[6] Konijn, H. S., "Statistical Theory of Sample Survey Design and Analysis", North–Holland Publishing Company-Amsterdam., (1973).

Abstract

The aim of this work is to evaluate the statistical procedures used in estimating the main field crop productions (rice), and to obtain the best mean that could increase the precision of the estimating. By using different sorts of sampling estimators, a comparison was made among the variances of the mean for simple random sampling (var($\bar{y}$)), stratified random sampling (var($\bar{y}_{st}$)), separate regression estimator (var($\bar{y}_{lrs}$)), and combined ratio estimator (var($\hat{\bar{y}}_{RC}$)), . The results indicate that the methods of estimation for separate regression estimator and combined ratio estimator give best estimations. The approximate cum.$f^{4/5}$ method was used to determine the optimum stratum boundaries, new strata was putting according to the approximate cum.$f^{4/5}$ method and then var ($\bar{y}_{st}$) was calculated. In comparison with strata used nowadays in central statistical organization, the new strata lead to wide reduce the variance, the study recommendation of applying the objective methods with the new stratification which has been suggested by the study.