



# المعهد العربي للتدريب والبحوث الإحصائية

دورة متوسطة حول تصميم عينات المسوح الإحصائية

7

Cluster and Multi-Stage Sample العينة العنقودية ومتعددة المراحل

إعداد:

نايف عابد

nayif@pcbs.gov.ps nayifabed@yahoo.com

آذار - 2021

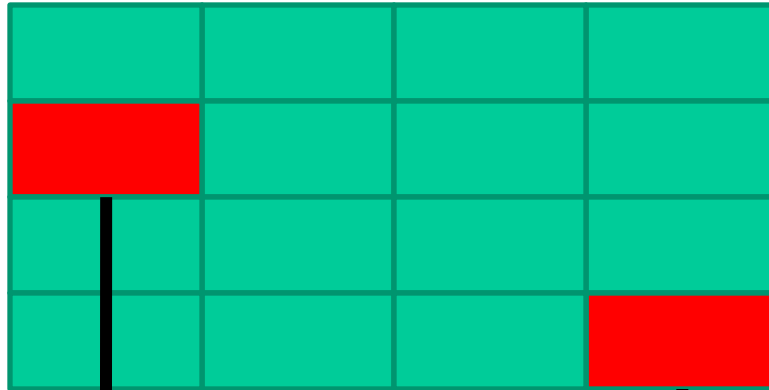
## مفهوم العينة العنقودية:

- يتم تقسيم مجتمع الهدف إلى مجموعات تسمى عناقيد clusters (مثل تقسيم التجمعات السكانية الكبيرة إلى مناطق عد، او بلوكات)
- في العينة العنقودية يتم اختيار عينة من العناقيد ويتم الوصول إلى كافة الوحدات داخل العناقيد التي تم اختيارها في العينة.
- في تصميم العينات العنقودية يتم الاستفادة من التقسيم العنقودي الاعتيادي الموجود في مجتمع الهدف. مثل : مناطق العد الجغرافية/التعدادية، المدارس/الصفوف الدراسية...الخ.
- من الممكن تشكيل عناقيد افتراضية من خلال تجميع الوحدات المتقاربة من بعضها البعض.

## مميزات العينة العنقودية والمتعددة المراحل:

1. يتم استخدامها عند عدم وجود إطار معينة تفصيلي يحتوي على جميع عناصر المجتمع بشكل مباشر.
2. تقليل التكاليف المرتفعة والوقت عند اختيار وحدات متجاورة
3. الاستفادة من التقسيمات الإدارية في السيطرة على الأخطاء غير الإحصائية بشكل أكثر فعالية
4. تستخدم العينة متعددة المراحل إذا كان من غير العملي أن يتم عمل مسح/حصر كل الوحدات في العنقود بسبب كبر حجمه.
5. سحب عينة متعددة المراحل يزيد في عدد العناقيد في المرحلة الأولى وهذا يؤدي إلى انتشار العينة وشمولها لمجتمع الدراسة
6. التصميم متعدد المراحل يتصف بالمرونة حيث انه من الممكن استخدام أسلوب سحب العينات في كل مرحلة مختلفة عن المراحل الأخرى.

إطار العناقيد

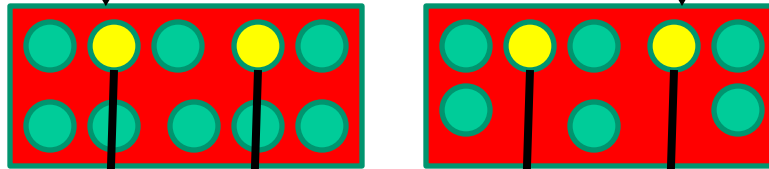


وحدات المعاينة الرئيسية

Primary sampling units (PSU)

المرحلة الأولى

First Stage

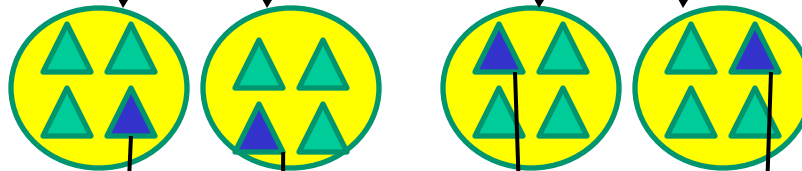


وحدات المعاينة الثانوية

Secondary sampling units (SSU)

المرحلة الثانية

Second Stage



وحدات المعاينة النهائية

Ultimate sampling units (USU)

المرحلة الثالثة

Third Stage



## مقارنة بين العينة العنقودية والعينة الطبقية

العينة الطبقية	العينة العنقودية
يتم سحب العينة من كل طبقة (شمول كل الطبقات)	يتم سحب عينة من العناقيد (شمول عينة من العناقيد)
تباين التقدير يعتمد على تباين القيم داخل الطبقة	تباين التقدير يعتمد على التباين بين العناقيد (كلما زاد عدد العناقيد كلما قل التباين)

# العينة العنقودية من مرحلة واحدة (العناقيد المتساوية في الحجم)

## Single-stage cluster sampling (equal)

مخطط المجتمع المكون من عناصر عددها  $NM$  ، حيث  $N$  عدد العناقيد في المجتمع و  $M$  عدد العناصر بكل عنقود، و  $y$  هو المتغير قيد الاهتمام

العناصر	العناقيد						
Elements	1	2	3	...	$i$	...	$N$
1	$y_{11}$	$y_{21}$	$y_{31}$	...	$y_{i1}$	...	$y_{N1}$
2	$y_{12}$	$y_{22}$	$y_{32}$	...	$y_{i2}$	...	$y_{N2}$
...	...	...	...	...	...	...	...
$j$	$y_{1j}$	$y_{2j}$	$y_{3j}$	...	$y_{ij}$	...	$y_{Nj}$
...	...	...	...	...	...	...	...
$M$	$y_{1M}$	$y_{2M}$	$y_{3M}$	...	$y_{iM}$	...	$y_{NM}$
Cluster total مجموع العنقود	$y_1$	$y_2$	$y_3$	...	$y_i$	...	$y_N$
Cluster mean متوسط العنقود	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	...	$\bar{y}_i$	...	$\bar{y}_N$

# العينة العنقودية من مرحلة واحدة (العناقيد المتساوية في الحجم)

## Single-stage cluster sampling (equal)

مخطط العينة المكون من عناصر عددها  $nM$  ، حيث  $n$  عدد العناقيد في العينة و  $M$  عدد العناصر بكل عنقود ، و  $y$  هو المتغير قيد الاهتمام

العناصر	العناقيد Clusters						
Elements	1	2	3	...	$i$	...	$n$
1	$y_{11}$	$y_{21}$	$y_{31}$	...	$y_{i1}$	...	$y_{n1}$
2	$y_{12}$	$y_{22}$	$y_{32}$	...	$y_{i2}$	...	$y_{n2}$
...	...	...	...	...	...	...	...
$j$	$y_{1j}$	$y_{2j}$	$y_{3j}$	...	$y_{ij}$	...	$y_{nj}$
...	...	...	...	...	...	...	...
$M$	$y_{1M}$	$y_{2M}$	$y_{3M}$	...	$y_{iM}$	...	$y_{nM}$
Cluster total مجموع العنقود	$y_1$	$y_2$	$y_3$	...	$y_i$	...	$y_n$
Cluster mean متوسط العنقود	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	...	$\bar{y}_i$	...	$\bar{y}_n$

تقدير المتوسط من العينة العنقودية من مرحلة واحدة (العناقيد المتساوية في الحجم)

متوسط العنقود

$$\bar{y}_i = \frac{1}{M} (y_{i1} + y_{i2} + \dots + y_{iM}) = \frac{y_i}{M} = \frac{1}{M} \sum_{j=1}^M y_{ij}$$

متوسط العينة (عدد n من العناقيد)

$$\bar{y} = \overline{y_n} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i$$

متوسط المجتمع (عدد N من العناقيد)

$$\bar{Y} = \overline{Y_N} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i$$



تقدير تباين المتوسط من العينة العنقودية من مرحلة واحدة (العناقد المتساوية في الحجم)

$$V(\bar{y}_n) = \frac{N - n}{N} \frac{1}{n} \frac{1}{M^2} \frac{\sum_{i=1}^N \left( y_i - \frac{\sum_{i=1}^N y_i}{N} \right)^2}{N - 1}$$

$$V(\bar{y}_n) = \frac{N - n}{N} \frac{1}{n} \frac{\sum_{i=1}^N (\bar{y}_i - \bar{Y})^2}{N - 1}$$

$$= \frac{1-f}{n} S_b^2$$

$$S_b^2 = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y}_n)^2}{n-1}$$

نستبدل تباين المجتمع  $S_b^2$  بتباين العينة  $S_b^2$

$$v(\bar{y}_n) = \frac{1-f}{n} S_b^2$$

يصح لدينا تقدير لتباين المتوسط

يمكن التعبير عن التباين للمتوسط في العينة العنقودية بدلالة معامل الارتباط الذاتي  $\rho$   
 intracluster (or intraclass) correlation coefficient ICC

$$S^2 = \frac{\sum_{i=1}^N \sum_{j=1}^M (y_{ij} - \bar{Y})^2}{NM - 1}$$

$$\rho = \frac{E(y_{ij} - \bar{Y})(y_{jk} - \bar{Y})}{E(y_{ij} - \bar{Y})^2} = \frac{2 \sum_{i=1}^N \sum_{j=1 < k}^M (y_{ij} - \bar{Y})(y_{jk} - \bar{Y})}{(M - 1)(NM - 1)S^2}$$

$$V(\bar{y}_n) = \frac{(1-f)(NM-1)}{nM^2(N-1)} S^2 [1 + (M-1)\rho]$$

إذا كانت  $N$  كبيرة فان :  $NM - 1 \approx NM$        $N - 1 \approx N$

$$V(\bar{y}_n) \approx \frac{1-f}{nM} S^2 [1 + (M-1)\rho]$$

# اثر تصميم العينة *Deff* Design Effect

Deff = التباين للتقدير حسب تصميم العينة / التباين للتقدير حسب العينة العشوائية البسيطة

$$Deff = \frac{v(\bar{y}_{clus})}{v(\bar{y}_{sr})} \quad \text{بدون إرجاع ولنفس حجم العينة}$$

التباين لتقدير المتوسط في حالة العينة العشوائية البسيطة:

$$V(\bar{y}_{nM}) = \frac{NM - nM}{NM} \frac{S^2}{nM} = \frac{1-f}{nM} S^2$$

$$V(\bar{y}_n) = \frac{1-f}{nM} S^2 [1 + (M - 1)\rho] \quad \text{التباين لتقدير المتوسط في حالة العينة العنقودية:}$$

$$Deff = \frac{V(\bar{y}_n)}{V(\bar{y}_{nM})} = [1 + (M - 1)\rho] \quad \text{اثر تصميم العينة}$$

# العينة العنقودية من مرحلة واحدة (العناقيد غير المتساوية في الحجم) Single-stage cluster sampling (unequal)

مخطط المجتمع المكون من  $N$  عناقيد و  $y$  هو المتغير قيد الاهتمام

العناصر	Clusters العناقيد						
Elements	1	2	3	...	$i$	...	$N$
1	$y_{11}$	$y_{21}$	$y_{31}$	...	$y_{i1}$	...	$y_{N1}$
2	$y_{12}$	$y_{22}$	$y_{32}$	...	$y_{i2}$	...	$y_{N2}$
...	...	...	...	...	...	...	...
$j$	$y_{1j}$	$y_{2j}$	$y_{3j}$	...	$y_{ij}$	...	$y_{Nj}$
...	...	...	...	...	...	...	...
$M_i$	$y_{1M_1}$	$y_{2M_2}$	$y_{3M_3}$	...	$y_{iM_i}$	...	$y_{NM_N}$
Cluster total مجموع العنقود	$y_1$	$y_2$	$y_3$	...	$y_i$	...	$y_N$
Cluster mean متوسط العنقود	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	...	$\bar{y}_i$	...	$\bar{y}_N$

# العينة العنقودية من مرحلة واحدة (العناقيد غير المتساوية في الحجم) Single-stage cluster sampling (unequal)

مخطط العينة المكون من  $n$  عناقيد و  $y$  هو المتغير قيد الاهتمام

العناصر	Clusters العناقيد						
Elements	1	2	3	...	$i$	...	$N$
1	$y_{11}$	$y_{21}$	$y_{31}$	...	$y_{i1}$	...	$y_{N1}$
2	$y_{12}$	$y_{22}$	$y_{32}$	...	$y_{i2}$	...	$y_{N2}$
...	...	...	...	...	...	...	...
$j$	$y_{1j}$	$y_{2j}$	$y_{3j}$	...	$y_{ij}$	...	$y_{Nj}$
...	...	...	...	...	...	...	...
$M$	$y_{1M_1}$	$y_{2M_2}$	$y_{3M_3}$	...	$y_{iM_i}$	...	$y_{NM_N}$
Cluster total مجموع العنقود	$y_1$	$y_2$	$y_3$	...	$y_i$	...	$y_N$
Cluster mean متوسط العنقود	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	...	$\bar{y}_i$	...	$\bar{y}_N$

تقدير المتوسط من العينة العنقودية من مرحلة واحدة (العناقد غير المتساوية في الحجم)

$$M_0 = \sum_{i=1}^N M_i$$

عدد العناصر في كل المجتمع

$$y_i = \sum_{j=1}^{M_i} y_{ij}$$

عدد العناصر في كل عنقود

$$\bar{M} = \frac{\sum_{i=1}^N M_i}{N} = \frac{M_0}{N}$$

متوسط عدد العناصر في كل عنقود

$$\bar{Y} = \frac{\sum_{i=1}^N \sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^N M_i} = \frac{\sum_{i=1}^N M_i \bar{y}_i}{\sum_{i=1}^N M_i} = \frac{\sum_{i=1}^N M_i \bar{y}_i}{M_0}$$

متوسط المجتمع

$$\bar{Y}_N = \frac{\sum_{i=1}^N \bar{y}_i}{N}$$

$$\bar{y}_n = \frac{N}{nM_0} \sum_{i=1}^n M_i \bar{y}_i = \frac{1}{n} \sum_{i=1}^n \left( \frac{M_i \bar{y}_i}{\bar{M}} \right)$$

متوسط العينة

## العينة العنقودية متعددة المراحل (ذات مرحلتان)

### Multi-stage cluster sampling (Two Stage)

مخطط المجتمع المكون من  $N$  عناقيد و  $y$  هو المتغير قيد الاهتمام

Cluster العناقيد	$M_i$	Population elements عناصر المجتمع	Total المجموع	Cluster mean متوسط العنقود
1	$M_1$	$y_{11}, y_{12}, \dots, y_{1j}, \dots, y_{1M_1}$	$Y_1 = \sum_{j=1}^{M_1} y_{1j}$	$\bar{Y}_1 = \frac{Y_1}{M_1}$
2	$M_2$	$y_{21}, y_{22}, \dots, y_{2j}, \dots, y_{2M_2}$	$Y_2 = \sum_{j=1}^{M_2} y_{2j}$	$\bar{Y}_2 = \frac{Y_2}{M_2}$
...	...	...	...	...
$i$	$M_i$	$y_{i1}, y_{i2}, \dots, y_{ij}, \dots, y_{iM_i}$	$Y_i = \sum_{j=1}^{M_i} y_{ij}$	$\bar{Y}_i = \frac{Y_i}{M_i}$
...	...	...	...	...
$N$	$M_N$	$y_{N1}, y_{N2}, \dots, y_{Nj}, \dots, y_{NM_N}$	$Y_N = \sum_{j=1}^{M_N} y_{Nj}$	$\bar{Y}_N = \frac{Y_N}{M_N}$

تقدير المتوسط من العينة العنقودية من مرحلتين

$$M_0 = \sum_{i=1}^N M_i$$

عدد العناصر في كل المجتمع

$$Y = \sum_{i=1}^N Y_i = \sum_{i=1}^N \sum_{j=1}^{M_i} y_{ij}$$

المجموع في المجتمع

$$\bar{M} = \frac{\sum_{i=1}^N M_i}{N} = \frac{M_0}{N}$$

متوسط عدد العناصر في كل عنقود في المجتمع

$$\bar{Y} = \frac{\sum_{i=1}^N \sum_{j=1}^{M_i} y_{ij}}{\sum_{i=1}^N M_i} = \frac{Y}{M_0} = \frac{\sum_{i=1}^N Y_i}{M_0} = \frac{\sum_{i=1}^N Y_i}{M_0} = \frac{\sum_{i=1}^N M_i \bar{Y}_i}{M_0}$$

متوسط المجتمع

$$\bar{Y}_i = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i} = \frac{Y_i}{M_i}$$

متوسط العنقود



# العينة العنقودية متعددة المراحل (ذات مرحلتان)

## Multi-stage cluster sampling (Two Stage)

مخطط العينة المكون من  $n$  عناقيد و  $y$  هو المتغير قيد الاهتمام

Unit العناقيد	$M_i$	$m_i$	Sample observation عناصر المجتمع	Total المجموع	Cluster mean متوسط العنقود
1	$M_1$	$m_1$	$y_{11}, y_{12}, \dots, y_{1j}, \dots, y_{1m_1}$	$y_1 = \sum_{j=1}^{m_1} y_{1j}$	$\bar{y}_1 = \frac{y_1}{m_1}$
2	$M_2$	$m_2$	$y_{21}, y_{22}, \dots, y_{2j}, \dots, y_{2m_2}$	$y_2 = \sum_{j=1}^{m_2} y_{2j}$	$\bar{y}_2 = \frac{y_2}{m_2}$
...	...	...	...	...	...
$i$	$M_i$	$m_i$	$y_{i1}, y_{i2}, \dots, y_{ij}, \dots, y_{im_i}$	$y_i = \sum_{j=1}^{m_i} y_{ij}$	$\bar{y}_i = \frac{y_i}{m_i}$
...	...	...	...	...	...
$n$	$M_n$	$m_n$	$y_{n1}, y_{n2}, \dots, y_{nj}, \dots, y_{nm_n}$	$y_N = \sum_{j=1}^{m_n} y_{nj}$	$\bar{y}_n = \frac{y_n}{m_n}$

$$m_0 = \sum_{i=1}^n m_i , \bar{m} = \frac{m_0}{n}$$

متوسط عدد العناصر في كل عنقود في العينة

$$y = \sum_{i=1}^n y_i = \sum_{i=1}^n \sum_{j=1}^{m_i} y_{ij}$$

تقدير المجموع في العينة

$$\bar{y}_i = \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i} = \frac{y_i}{m_i}$$

تقدير متوسط العينة في العنقود

$$\bar{y} = \frac{y}{m_0}$$

تقدير متوسط العينة

$$v(\bar{y}_{ts}) = (1 - f_1) \left( \frac{1}{\bar{M}^2} \right) \frac{S_b^2}{n} + \frac{1}{nN\bar{M}^2} \sum_{i=1}^N M_i^2 (1 - f_{2i}) \left( \frac{S_i^2}{m_i} \right)$$

تقدير تباين المتوسط

$$f_1 = \frac{n}{N} , f_{2i} = \frac{m_i}{M_i}$$

# تمارين تطبيقية على برنامج SPSS