نظرية المعاينة Sampling Theory فص 11

تهتم نظرية العينات بدراسة العلاقة بين المجتمع والعينات المسحوبة منه فيما يسمى بالاستدلال الاحصائي. (Statistical Inference). هناك عدة طرق لاخذ العينات من المجتمع لاستخدامها في الاستدلال الاحصائي ومن اشهر الطرق هي الطريقة العشوائية وبها تتساوى العينات في فرص الاختيار.

والمجتمع قد يكون محدودًا إذا كان يمكن حصر عدد أفراده مثل سكان مدينة ما أو طلاب مرحلة تعليمية معينة ، وقد يكون المجتمع غير محدود (لا نهائي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار وعند دراسة صفة ما أو صفات معينه لمجتمع ما ، فإن البيانات الإحصائية عن تلك الصفة أو الصفات تجمع بأحد أسلوبين.

اولا: اسلوب الحصر (Census) :وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب يتطلب وفرة في الوقت والمال والمجهود الفني وتزداد هذه المتطلبات وتتضاعف كلما ازداد حجم المجتمع (عدد أفراد المجتمع). وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول وتدعمها بإمكانيات ضخمه مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

الثاني: اسلوب المعاينة (Sampling method): وفيه يتم جمع البيانات عن جزء من مفردات المجتمع ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله. أن إسلوب المعاينة يقصد به دراسة خصائص المجتمع من خلال دراسة عينه مسحوبة منه ، ونجاح هذا الإسلوب يعتمد على أن تحمل العينة أقصى درجة من دقة التمثيل للمجتمع المسحوبة منه.

اقسام العينات

1. **العينات العشوائية**

A. العينة العشوائية البسيطة
B. العينة المنتظمة
C. العينة العشوائية الطبقية
D. العينة متعددة المراحل أو العنقودية

2. **العينات غير العشوائية**

A. العينة العمدية أو المقصودة
B. العينة الحصصيه

مراحل تصميم خطة المعاينة: خطة المعاينة هي وصف للإجراءات المنهجية الخاصة بكيفية تنفيذ كل مرحلة من مراحل المعاينة، والمتمثلة فيما يلي:

- تحديد مشكلة وهدف الدراسة: تعد هذه المرحلة من أهم مراحل تصميم خطة المعاينة، فهي تسمح بتحديد الإطار العام للدراسة المتضمن طبيعة وحجم البيانات المطلوبة، متغيرات الدراسة، المجتمع الإحصائي المدروس وغيرها من المتغيرات.

- تحديد المجتمع الإحصائي عن طريق تحديد المتغيرات والمفردات المراد دراستها والحدود الزمنية والمكانية.

- تحديد البيانات المطلوبة، مصادرها وطريقة جمعها وفقا لموضوع الدراسة، هدفها، فرضياتها وتكاليف الحصول عليها وغيرها.

- تحديد حجم العينة الذي يحقق التوافق بين دقة المعلومات وتكاليف الدراسة، والذي يعتمد على تطبيق القواعد الإحصائية الخاصة بنظرية المعاينة وخبرة الباحث.

- جمع البيانات عن مفردات العينة المختارة، وهي مرحلة هامة لأنها عرضة لكثير من أخطاء القياس.

- مراجعة وترميز البيانات باستخدام طرق إحصائية معينة للوصول إلى نتائج ودراستها.

- تحليل البيانات وتقدير معالم المجتمع: بعد الحصول على إحصاءات العينة يتم دراستها وتحليلها للاستدلال على خصائص المجتمع.

المعالم والإحصاءات:

المقاييس الإحصائية التي تحسب من بيانات مجتمع الدراسة بأكمله يطلق عليها معالم المجتمع (**Parameters of population**). أما المقاييس الإحصائية التي تحسب من بيانات عينه مسحوبة من مجتمع الدراسة فيطلق عليها إحصاءات (**Statistics**). ويعتبر كل إحصاء منها بمثابة تقدير أو قيمه تقديرية لمعلمة المجتمع المناظر، فيكون المتوسط الحسابي المحسوب من بيانات العينة تقدير لمعلمة المجتمع المناظرة وهي المتوسط الحسابي المسحوب منه هذه العينة وهكذا. ويجب ألا يغيب عن الأذهان بأن حساب قيمة المتوسط الحسابي من بيانات العينة ليس هدفًا في حد ذاته ولكن وسيلة للتعرف على المتوسط الحسابي للمجتمع موضوع الدراسة. وهكذا بالحال بالنسبة لباقي المقاييس الإحصائية التي تحسب من العينة. للتفرقة بين المعالم والإحصاءات يجب أن نرمز لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال يرمز للمتوسط الحسابي للمجتمع بالرمز μ بينما يرمز للمتوسط الحسابي للعينة بالرمز m

بينما يرمز للانحراف المعياري للعينة بالرمز S للانحراف المعياري للمجتمع بالرمز σ وهكذا.

توزيعات المعاينة (Sampling Distributions):

إذا سحبنا عينة عشوائية وحسبنا متوسطها الحسابي سوف نجده قيمة ثابتة، أما إذا سحبنا عدة عينات (بعض العينات) بنفس الحجم من نفس المجتمع فمن المتوقع أن يأخذ المتوسط الحسابي قيما مختلفة في هذه العينات، أما إذا قمنا بسحب كل العينات الممكنة والتي لها نفس الحجم n وحسبنا المتوسطات الحسابية لهذه العينات، فإن المتوسط الحسابي العام $\bar{X}$ لمتوسطات هذه العينات يجب أن يساوي تماما الوسط الحسابي للمجتمع μ، ويسمى توزيع المتوسطات الحسابية بتوزيع المعاينة للمتوسط الحسابي.

توزيع المعاينة للمتوسط الحسابي $\bar{X}$ بالنسبة لعينات عشوائية ذات الحجم n مأخوذة من مجتمع يتوزع توزيعا طبيعيا متوسطه $\boldsymbol{\mu}$ وتباينه $\boldsymbol{\sigma}$ يكون مساويا لمتوسط المجتمع (سواء كان السحب بالإرجاع أو بدون إرجاع)، أما تباينه σ_m^2 فيكون

في حال الارجاع

$$\sigma_m^2 = \frac{\boldsymbol{\sigma}^2}{n}$$

في حال عدم الارجاع

$$\sigma_m^2 = \frac{\boldsymbol{\sigma}^2}{n}\left(\frac{N-n}{N-1}\right)$$

ملاحظات:

- N يمثل حجم المجتمع وقد يكون محدود الحجم (منته) أو غير محدود (غير منته)، فإذا كان حجم المجتمع محدود وكان السحب بالإرجاع، فيعتبر المجتمع هنا غير منته أو غير محدود الحجم.
- القيمة $\left(\frac{N-n}{N-1}\right)$ تسمى بمعامل التصحيح أو الانتهاء.
- لمعرفة نوعية السحب نستعين بالقاعدة التالية:

سحب بدون إرجاع $\quad \frac{\mathbf{n}}{\mathbf{N}} \geq 0.05$

سحب بالارجاع $\quad \frac{\mathbf{n}}{\mathbf{N}} < 0.05$

5 متوسط توزيع المعاينة للمتوسطات

مسألة 1: ليكن المجتمع 1، 3، 5، 6، 8. ما هي القيمة المتوقعة لمتوسط عينة (m) مسحوبة **بالإرجاع** مكونة من مفردتين ؟ أحسب متوسط المجتمع μ. قارن بين m وμ. من أجل تحديد ذلك أحسب جميع الحالات الممكنة للمتوسط m_i حسب كل عينة.
العينات الممكنة العينات الممكنة ذات الحجم n = 2 من مجتمع حجمه 5 عددها: 5*5 =25

العينات الممكنة j→ ، i↓				
(1, 1)	(3, 1)	(5, 1)	(6, 1)	(8, 1)
(1, 3)	(3, 3)	(5, 3)	(6, 3)	(8, 3)
(1, 5)	(3, 5)	(5, 5)	(6, 5)	(8, 5)
(1, 6)	(3, 6)	(5, 6)	(6, 6)	(8, 6)
(1, 8)	(3, 8)	(5, 8)	(6, 8)	(8, 8)

المتوسطات الممكنة للعينة (معاينة غ نفادية) $m_i = (i+j)/2$				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

القيمة المتوقعة m لـ m_i هي متوسط قيمها وهي $m = (\sum_i m_i) / 25 = 4,6$.

حساب متوسط المجتمع: $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

مثال2. أوجد نفس مطالب المثال 1. في حالة السحب **<u>بدون إرجاع</u>**. العينات الممكنة عددها:

$$\binom{5}{2} = 10$$

المتوسطات الممكنة للعينة أو **توزيع المعاينة للمتوسطات (معاينة بدون ارجاع)** m_i			
2			
3	4		
3,5	4,5	5,5	
4,5	5,5	6,5	7

العينات الممكنة **<u>بدون إرجاع</u>**			
(1, 3)			
(1, 5)	(3, 5)		
(1, 6)	(3, 6)	(5, 6)	
(1, 8)	(3, 5)	(5, 8)	(6, 8)

القيمة المتوقعة m لـ m_i هي متوسط قيمها وهي:

$$E(m) = \mu_m = (\Sigma_i\, m_i) / 10 = 4{,}6$$

متوسط المجتمع : $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

نظرية 1. إذا كانت م ع تمثل مجتمع ما وm متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لمتوسط العينة E(M) تكتب كما يلي: $E(M) = \mu_m = \mu$

البرهان : لنرمز ب X_i لقيم المتغيرة الأصلية X .

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n}\sum_i E(Xi) = \frac{1}{n}\sum_i \mu = \frac{1}{n}n\mu = \mu.$$

6 تباين توزيع المعاينة للمتوسطات

مثال. أحسب تباين المجتمع في المسألة 1 اعلاه، أحسب التباين (والانحراف المعياري) لتوزيع المعاينة للمتوسطات σ^2_m علما أن العينة مسحوبة بالإرجاع (**غ نفادية**)، قارن بين تباين المجتمع وتباين متوسطات العينات الممكنة (توزيع المعاينة للمتوسطات).

m_i				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

$$\sigma^2_m = [\Sigma_i (m_i - m)^2]/25 = 2.92;$$
$$\sigma^2 = [\Sigma_i (x_i - \mu)^2]/5 = 5.84$$
$$\mathbf{2.92 = 5.84 / 2}$$

هذا المثال يمهد للنظرية التالية:

نظرية 2. إذا كانت م ع تمثل مجتمع ما و m_i متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع

بالإرجاع، فإن تباين m_i (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي: $\boxed{\sigma_m^2 = \frac{\sigma^2}{n}}$ حيث n حجم العينة.

البرهان: لنرمز ب X_i لقيم المتغيرة الأصلية X .

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n^2}\sum_i V(Xi) = \frac{1}{n^2}\sum_i \sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

أ- طبيعة توزيع m

ندرس طبيعة توزيع متوسط توزيع المعاينة للمتوسطات من خلال النظريات التالية:

سنتكلم عن الحالات المختلفة لأيجاد توزيع المعاينة للمتوسط الحسابي

اولاً **(نظرية** 4): عندما يكون مجتمع موزع توزيعاً طبيعياً بمتوسط μ وتباين σ^2 معلوم اي : $m \sim N(\mu,\sigma^2)$ فانه وبغض النظر عن حجم العينة n التي تسحب من هذا المجتمع سوف يكون توزيع المتوسط الحسابي للعينة هو $m \sim N(\mu,\frac{\sigma^2}{n})$ ولحساب الاحتمالات للمتغير العشوائي يتم ايجاد القيم المعيارية له (اي تحويل قيمته الى المعيارية) حيث:

$$Z = \frac{m-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

مثال: اذا كان تباين مجتمع هو $\sigma^2 = 9$ و متوسطه هو $\mu = 12$ فاوجد التوزيع الاحتمالي لمتوسط عينة عشوائية حجمها 40 سحبت من المجتمع ثم اوجد قيمة الاحتمال $(P(m \leq 13)$

الحل:

لحساب قيمة الاحتمال نحول الى المتغير المعياري

$$Z = \frac{m-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{13-12}{\sqrt{\frac{9}{40}}} = 2.1 \quad from\ the\ table\ P(Z \leq 2.1) = 0.982136$$

مثال: مجتمع حجمه 900 بمتوسط $\mu = 20$ و$\sigma = 12$. نستخرج كل العينات الممكنة. أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة: (1) حجم العينة $n = 36$.

(1) $n = 36: \quad n/N = 36/900 = 0.04 < 0.05 \Rightarrow \sigma_m = \sigma/\sqrt{n} = 12/\sqrt{36} = 2$

$$E(m) = \mu = 20$$

مثال2. باستخدام معطيات المثال السابق $(n = 36)$ أحسب احتمال أن يكون m محصورا بين 18 و22.

$$Z_1 = \frac{m_1 - \mu}{\sigma/\sqrt{n}} = \frac{18-20}{12/\sqrt{36}} = -1\,, \quad Z_2 = 1 \quad \Rightarrow P(18 < m < 22) = P(Z_1 < Z < Z_2) = 0.6827$$

ثانياً : عندما يكون تباين المجتمع الطبيعي مجهول (σ^2)مجهول في هذه الحالة نستبدل تبيان المجتمع (σ^2) بتباين العينة (S^2)المسحوبة منه ولكي نميز بين حالتين:

في حال العينات الكبيرة $(n \geq)$ يكون توزيع المتوسط الحسابي للعينة هو:

$$m \sim N\left(\mu, \frac{S^2}{n}\right)$$

ويكون التوزيع الطبيعي المعياري هو

$$Z = \frac{m-\mu}{\sqrt{\frac{S^2}{n}}}$$

علما ان تباين العينة يحسب من العلاقة الاتية

$$S^2 = \frac{\sum(x_i - \bar{x}^2)}{n-1}$$

مثال: اذا كان تباين مجتمع هو σ^2 مجهولاً و متوسطه هو $\mu = 12$ فاوجد التوزيع الاحتمالي لمتوسط عينة عشوائية حجمها 40 سحبت من المجتمع حيث ان تباينها يساوي 16 ثم اوجد قيمة الاحتمال $(\mathrm{P}(m \leq 13)$

الحل:

لحساب قيمة الاحتمال نحول الى المتغير المعياري

$$Z = \frac{m-\mu}{\sqrt{\frac{S^2}{n}}} = \frac{13-12}{\sqrt{\frac{16}{40}}} = 1.58 \quad from\ the\ table\ P(Z \leq 1.58) = 0.942947$$

في حالة العينات الصغيرة $(n < 30)$

يتم تطبيق قانون يختلف كون التوزيع سيكون مختلف ويتبع الى توزيع (T) او ما يسمى بتوزيع ستودنت والذي يتطلب قوانين وجداول خاصة كونه يختلف قليلا عن التوزيع الطبيعي كون تباينه يكون اكثر من واحد بقليل اي انه اكثر تشتتا من التوزيع الطبيعي المعياري.

7 المتوسط والتباين

ليكن لدينا مجتمعين نسحب من كل منهما عينة عشوائية، نحسب في كل عينة محسوبة من المجتمع الأول الإحصائية S_1 ونحسب نفس الاحصائية (المتوسط مثلا أو التباين ...) في كل عينة من المجتمع الثاني ونسميها S_2. إن الفرق $S_1 - S_2$ يشكل بدوره متغيرة عشوائية لها المتوسط والتباين التاليين:

$$\mu_{(S1-S2)} = \mu_{S1} - \mu_{S2} \qquad ; \qquad \sigma^2_{(S1-S2)} = \sigma^2_{S1} + \sigma^2_{S2}$$

مثال 1. إذا كانت الاحصائية هي **المتوسط** فإن:

$$\mu_{(m1-m2)} = \mu_{m1} - \mu_{m2} = \mu_1 - \mu_2 \qquad ; \qquad \sigma^2_{(m1-m2)} = \sigma^2_{m1} + \sigma^2_{m2} = \sigma^2/n_1 + \sigma^2/n_2$$

مثال 2. إذا كانت الاحصائية هي **النسبة** فإن:

$$\mu_{(p1-p2)} = \mu_{p1} - \mu_{p2} = p_1 - p_2 \qquad ; \qquad \sigma^2_{(p1-p2)} = \sigma^2_{p1} + \sigma^2_{p2} = p_1q_1/n_1 + p_2q_2 / n_2$$

إذا كان الاهتمام هو على مجموع الاحصائيتين بدلا من الفرق بينهما فإن:

$$\mu_{(S1+S2)} = \mu_{S1} + \mu_{S2} \qquad ; \qquad \sigma^2_{(S1+S2)} = \sigma^2_{S1} + \sigma^2_{S2}$$

8 طبيعة توزيع المعاينة للفرق بين متوسطين

نظرية 7 : في حالة $n_1 \geq 30$ و n_2 ، يقترب توزيع المتغيرة المعيارية للفرق بين متوسطين من التوزيع الطبيعي المعياري. ونكتب:

$$\mu_{m1-m2} \approx N(0, 1)$$

مثال: ليكن المجتمع U_1 : 3، 7، 8. والمجتمع U_2 : 2، 4. تحقق من أن :

$$\mu_{(U1-U2)} = \mu_{U1} - \mu_{U2} \; ; \qquad \sigma^2_{(U1-U2)} = \sigma^2_{U1} + \sigma^2_{U2} \; .$$

<table>
<tr><td colspan="3">U_1</td><td colspan="2" rowspan="2">**$U_1 - U_2$**</td></tr>
<tr><td>8</td><td>7</td><td>3</td></tr>
<tr><td>**6**</td><td>**5**</td><td>**1**</td><td>2</td><td rowspan="2">U_2</td></tr>
<tr><td>**4**</td><td>**3**</td><td>**-1**</td><td>4</td></tr>
</table>

$$\mu_{U1} = (3+7+8)/3 = 6 \; ; \qquad \mu_{U2} = (2+4)/2 = 3$$

$$\Rightarrow \mu_{U1} - \mu_{U2} = 6 - 3 = 3$$

$$\mu_{U1-U2} = (1+5+6-1+3+4)/6 = 3$$

$$\sigma^2_{U1} = (3^2+7^2+8^2)/3 - 6^2 = 14/3 \; ;$$

$$\sigma^2_{U2} = (2^2+4^2)/2 - 3^2 = 1 \qquad \Rightarrow \sigma^2_{U1} + \sigma^2_{U2} = 17/3$$

$$\sigma^2_{U1-U2} = (1^2+5^2+6^2+1^2+3^2+4^2)\,/\,6 - 3^2$$

$$= (1+25+36+1+9+16)\,/\,6 - 9 = 17/3$$

مثال: شركتان متنافستان بمنتج معين، كل من منتجاتهما يتصف بالصفات المبينة بالجدول ادناه، **ما احتمال** ان العينة العشوائية لإنتاج الشركة **A** لها متوسط عمر يزيد عن متوسط عمر عينة الشركة **B** بما لا يقل عن سنة واحدة.

الشركة – صفات المنتج	B	A
متوسط العمر (سنة) μ	6	6.5
الانحراف المعياري(سنة) σ	0.81	0.64
حجم العينة المسحوبة n	49	36

الحل: المطلوب $P(\mu_1 - \mu_2) \geq 1$

$\mu = \mu_1 - \mu_2 = 6.5 - 6.0 = 0.5$; $\sigma^2_p = \sigma^2_{(U1-U2)} = \sigma^2_{U1} + \sigma^2_{U2}$

$\sigma^2_p = \frac{0.81}{36} + \frac{0.64}{49} = 0.0356$; $\rightarrow \sigma_p = 0.189$

$Z = \frac{m-\mu}{\sigma_p} = \frac{1-0.5}{0.189} = 2.65$

$P(\mu_1 - \mu_2) \geq 1 = P(Z > 2.65)$

$= 1 - P(Z \leq 2.65)$ using St. N. dist. Table

$= 1 - 0.9960 = 0.0040$ which was wanted.