المحور الثاني: توزيعات المعاينة

تمهيد:

إن الهدف من اخذ العينة هو معرفة خصائص مجتمعها، فاخذ العينات ليس القصد منه العينة لذاتها بل المجتمع الذي أخذت منه ، فالعينة وسيلة وليست الهدف.

وتقدم العينات تقديرات لخصائص مجتمعها، وهذه التقديرات تدور حول المتوسط الحقيقي لمجتمع الدراسة، أي أن متوسط العينة هو ليس متوسط مجتمعها، بل قيمة تمثل العينة ذاتها، وتعتمد في تقدير القيمة المحتملة لمتوسط المجتمع وفق حدود معينة للثقة .

إذا أخذت جميع العينات المحتملة من مجتمعها فيتوقع أن تكون متوسطات العينات موزعة بالتساوي حول متوسط مجتمع الدراسة، بعبارة أخرى ان متوسط متوسطات العينات يساوي متوسط مجتمعها.

نهدف من خلال هذا المحور الى:

- التعرف توزيع معاينة للوسط؛
-

أولا: تعريف المعاينة

هي عملية اختيار (*Process*) عدد كاف من عناصر المجتمع، بحيث يتمكن الباحث من خلال دراسة العينة المختارة وفهم خصائصها من تعميم هذه الخصائص على عناصر المجتمع الأصلي، ولابد أن نتذكر دوما بأن ناتج عملية المعاينة هو العينة (*Sample*) المرغوب بها.

ثانيا: توزيع المعاينة

وهو ذلك التوزيع التكراري لأحد التوابع الإحصائية المحسوب من بيانات العينات العشوائية ذات الحجم الواحد والتي يمكن سحبها من مجتمع إحصائي واحد.

نفرض أننا أخذنا عينه حجمها n من مجتمع ما، ثم حسبنا بعض المقاييس الإحصائية مثل المتوسط الحسابي، التباين، ... فإن كل مقياس من هذه المقاييس يعتبر متغير عشوائي في ذاته يختلف من عينه إلى أخرى – هذا المتغير العشوائي يخضع لتوزيع معين – هذا التوزيع يسمى بتوزيع العينة

فمثلاً نقول إن توزيع المعاينة للمتوسط الحسابي وهو عبارة عن توزيع جميع المتوسطات الحسابية للعينات المأخوذة من نفس هذا المجتمع ذو الحجم n ، وكذلك فإن توزيع المعاينة للتباين هو توزيع جميع التباينات المحسوبة من عينات لها نفس الحجم n ومأخوذة من نفس المجتمع ، وهكذا...

إذا أخذنا عينات متكررة من مجتمع ما وقمنا بحساب متوسط كل عينة، فإننا نجد أن معظم هذه المتوسطات تختلف عن بعضها البعض، ويسمى التوزيع الاحتمالي لمتوسطات العينات "توزيع المعاينة للوسط "

أنواع توزيعات المعاينة

أولا: توزيعات المعاينة للوسط

نظرية (1)

إذا كان X يخضع لتوزيع وسطه μ وتباينه σ^2، وكان $\bar{X}$ يمثل الوسط الحسابي للعينة ذات الحجم n والمسحوبة من هذا المجتمع فإن:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

$$Var(\bar{X}) = \sigma^2{}_{\bar{X}} = \frac{\sigma^2}{n}$$

نظرية (2)

إذا كانت $X_1, X_2, X_3, \dots \dots X_n$ عينة عشوائية من مجتمع طبيعي وسطه μ وتباينه σ^2، وكان $\bar{X}$ يمثل الوسط الحسابي للعينة ذات الحجم n والمسحوبة من هذا المجتمع فإن $\bar{X}$ يتبع التوزيع الطبيعي بوسط μ وانحراف معياري $\sigma/\sqrt{n}$ أي أن:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

يخضع لتوزيع طبيعي معياري

مثال: أخذت عينة عشوائية حجمها 9 من الأطفال حديثي الولادة في أحد المستشفيات، فإذا علم أن وزن الطفل حديث الولادة يخضع لتوزيع طبيعي 2900 غرام وانحرافه المعياري 600 غرام.

(أ) أوجد معدل وتباين والانحراف المعياري للوسط الحسابي لأوزان الأطفال في العينة.

$$X \sim N(2900, (600)^2)$$

$$\bar{X} \sim N\left(2900, \frac{(600)^2}{n}\right)$$

$$\mu_{\bar{X}} = \mu, \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

(ب) أوجد احتمال أن الوسط الحسابي لأوزان الأطفال في العينة يزيد عن 3100 غرام.

$$P(\bar{X} > 3100) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{3100 - \mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(Z > \frac{3100 - 2900}{600/\sqrt{9}}\right)$$

$$= P\left(Z > \frac{200}{200}\right)$$

$$= P(Z > 1)$$

$$= 1 - P(Z < 1) = 1 - 0.8413 = 0.1587$$

(ج) أوجد احتمال أن الوسط الحسابي لأوزان الأطفال في العينة يقع بين 2700 و 3200 غرام.

$$P(2700 < \bar{X} < 3200) = P\left(\frac{2700-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{3200-\mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{2700-2900}{600/\sqrt{9}} < Z < \frac{3200-2900}{600/\sqrt{9}}\right)$$

$$= P\left(\frac{-200}{200} < Z < \frac{300}{200}\right)$$

$$= P(-1 < Z < 1.5)$$

$$= P(Z < 1.5) + P(Z < 1) - 1$$

$$= 0.9332 + 0.8413 - 1 = 0.7745$$

نظرية (3): النهاية المركزية (تقارب التوزيعات)

إذا كانت $X_1, X_2, X_3, \dots \dots X_n$ عينة عشوائية من مجتمع وسطه μ وتباينه σ^2، وكان $\bar{X}$ يمثل الوسط الحسابي للعينة ذات الحجم n والمسحوبة من هذا المجتمع فإن $\bar{X}$ يقترب من التوزيع الطبيعي بوسط μ وانحراف معياري $\sigma/\sqrt{n}$ كلما كبرت n $(n \geq 30)$ أي أن:

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

يخضع لتوزيع طبيعي معياري

مثال: مجتمع حجمه 900 بمتوسط 20 µ =و . 12= σنستخرج كل العينات الممكنة. أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة:

- n = 36
- n = 64

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

(2) $n = 64 : N = 900 \Rightarrow \dfrac{n}{N} = \dfrac{64}{900} = 0.071 > 0.05$

$$\Rightarrow \sigma_m = \frac{12}{\sqrt{64}}\sqrt{\frac{900-64}{900-1}} = 1.92$$

$E(m) = \mu = 20$

نظرية 4: إذا كانت P نسبة صفة معينة في مجتمع ما وسحبنا عينات عشوائية حجم كل منها n، فإن مختلف النسب $\hat{P}_i$ $(i=1\ ;....\ ;n)$، تشكل توزيع المعاينة للنسبة، ويتميز التوزيع النظري لمعاينة النسبة $\hat{P}$ بالخصائص التالية:

توقع (متوسط توزيع المعاينة للنسبة) $E(\hat{P}) = \mu_{\hat{P}} = P$.

وخطأ معاينة (الانحراف المعياري لتوزيع المعاينة للنسبة) $\sigma_{\hat{P}} = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{p(1-p)}{n}}$

نكتب: $\hat{P} \approx N(p, \sigma_{\hat{P}}^2)$ (في حالة $n \geq 30$، أو في حالة مجتمع طبيعي)

مثال:

سحبنا عينة عشوائية حجمها 31 من نقاط الطلبة في امتحان الاحصاء الرياضي فكانت كالتالي:

{19.5، 11.5، 14، 11، 16.5، 18، 7.5، 3.5، 2.5، 2.5، 8.5، 10.5، 8.5، 2، 5، 3، 2.5، 1، 2.5، 10، 6، 6، 15، 6، 8.5، 5.5، 1.5، 12، 3، 18.5، 10.5}

1. أحسب $\hat{P}$ نسبة الطلبة الحاصلين على المعدل في العينة.
2. أوجد التوزيع العيني لنسبة الطلبة الحاصلين على المعدل في العينة من مجموع الطلبة البالغ 300.

الحل:

1. حساب $\hat{P}$ نسبة الطلبة الحاصلين على المعدل في العينة.

لدينا 12 علامة أكبر أو تساوي 10 $(12=n_a)$ ⇐ $\hat{p}=n_a/n=12/31=0.387$

2. التوزيع العيني لنسبة الطلبة الحاصلين على المعدل في العينة (علامة أكبر أو تساوي 10).

القيمة المتوقعة لنسبة الطلبة الحاصلين على المعدل في العينة: $E(\hat{P})=\mu_{\hat{P}}=P=0.38$

لحساب تباين توزيع المعاينة للنسبة (خطأ المعاينة) لابد من اختبار النسبة $n/N = 31/300 =$ $0.103> 0.05$ ⇐ نستخدم معامل التصحيح لحساب تباين توزيع المعاينة للنسبة.

$$\sigma_{\hat{P}}^2=\frac{pq}{n}.\frac{N-n}{N-1}=\frac{p(1-p)}{n}.\frac{N-n}{N-1}=\frac{0.387(1-.387)}{31}.\frac{300-31}{300-1}=\boxed{0.0069}$$

ونكتب: $\hat{P}\sim N(P,\sigma_{\hat{P}}^2) >\Rightarrow \hat{P}\sim N(0.387,0.0069)$

نظرية 5: في حالة $m,\ n\geq 30$، فإن توزيع المعاينة للفرق بين متوسطين يتوزع تقريبا وفق التوزيع الطبيعي، ونكتب:

$$(\bar{X}-\bar{Y})\sim N(\mu_X-\mu_Y,\sigma_{\bar{X}-\bar{Y}}^2)$$

وعليه فإن توزيع المتغير المعياري $(\bar{X}-\bar{Y})$ يقترب من التوزيع الطبيعي المعياري ونكتب:

$$Z=\frac{(\bar{X}-\bar{Y})-(\mu_{\bar{X}}-\mu_{\bar{Y}})}{\sqrt{\frac{\sigma_{\bar{X}}^2}{n}+\frac{\sigma_{\bar{Y}}^2}{m}}}\sim N(0,1)$$

نظرية 6: توزيع المعاينة للفرق بين نسبتي عينتين حجمهما كبير($m,\ n\geq 30$) ومسحوبتين من مجتمعين مستقلين، فإن توزيع المعاينة للفرق بين **نسبتي عينتين** يتبع التوزيع الطبيعي تقريبا، ونكتب:

$$(\hat{P}_X - \hat{P}_Y) \sim N(\mu_{PX-PY}, \sigma^2_{PX-PY})$$

وعليه فإن توزيع المتغير المعياري للفرق بين نسبتي عينين يقترب من التوزيع الطبيعي المعياري ونكتب:

$$Z = \frac{(\hat{P}_X - \hat{P}_Y) - (\mu_{\hat{P}_X} - \mu_{\hat{P}_Y})}{\sqrt{\sigma^2_{\hat{P}_X - \hat{P}_Y}}} \sim N(0,1)$$

ثانيا: توزيع المعاينة للتباين

نظرية 7: اذا كان X المتغير العشوائي من مجتمع ما و S^2 متغيرة عشوائية تمثل تباين عينة مسحوبة من ذات المجتمع فإن القيمة المتوقعة لتباين العينة يعطى كالتالي:

(1) السحب بدون ارجاع

(أ) إذا كان حجم المجتمع غير منته

(ب) إذا كانت نسبة حجم العينة إلى المجتمع $\left(\frac{n}{N}\right) < 0.05$

(2) السحب بالارجاع

إذا كان السحب بدون ارجاع أو إذا كان حجم المجتمع N صغير أو منته أي أن $\left(\frac{n}{N}\right) \geq 0.05$

$$E(S^2) \begin{cases} \mu_{S^2} = \sigma^2_X\left(\frac{n-1}{n}\right) \\ \mu_{S^2} = \sigma^2_X\left(\frac{n-1}{n}\right)\left(\frac{N-n}{N}\right) \end{cases}$$

نظرية 8: انطلاقا من عينة عشوائية حجمها n من المشاهدات المستقلة مسحوبة من مجتمع طبيعي بتباين σ_X^2 فإن:

$$\chi_v^2 = \frac{(n-1)S_X^2}{\sigma_X^2}$$

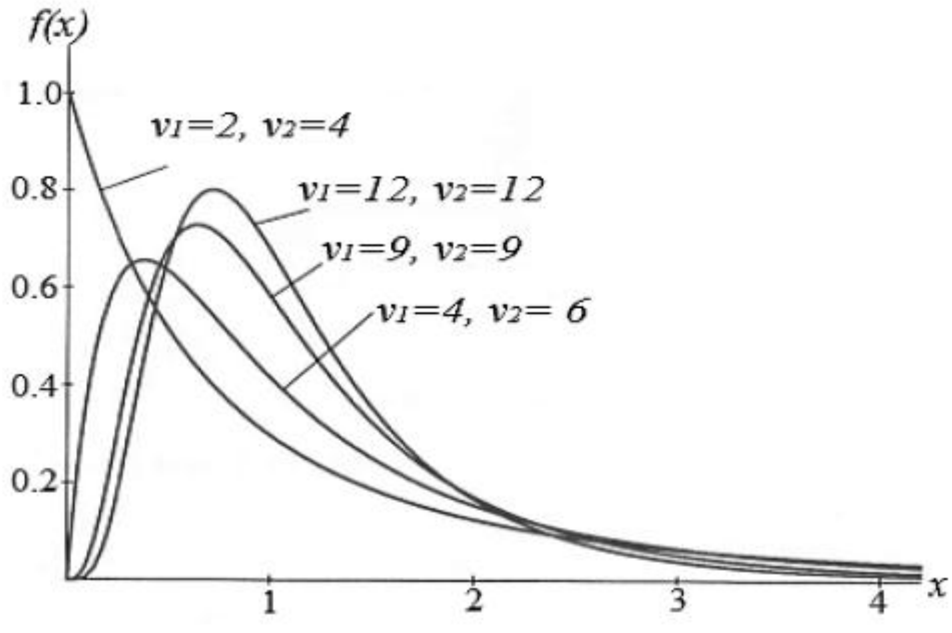
تتبع توزيع كاي مربع (*chi-square*) بدرجة حرية $v=n-1$.

توزيع المعاينة لنسبة تبايني عينيتين:

اذا كان لدينا $X \sim \chi^2_{v1}$ و $Y \sim \chi^2_{v2}$ و X و Y مستقلان عن بعضهما، $Z = \frac{(X/n)}{(Y/m)}$ يتبع توزيع F (*Fisher*) بدرجتي حرية v_1 و v_2 ($v_1= n-1$ و $v_2=m-1$). ونكتب: $Z \sim F_{v_1 v_2}$

ويشير v_1 إلى درجة حرية البسط، و v_2 يشير إلى درجة حرية المقام. وعليه توزيع فيشر هو توزيع نسبة متغيرين مستقلين يتبعان توزيع χ^2 بدرجتي حرية v_1 و v_2. وهذه الدرجات هي التي تحدد شكل التوزيع، حيث أن توزيع فيشر ملتوي نحو اليمين وتقل درجة الالتواء كلما زادت درجات أحدى درجات الحرية أو كلاهما حسب ما هو موضح في الشكل التالي:

التمثيل البياني لتوزيع فيشر F حسب درجة الحرية



مما سبق نستنتج أن توزيع المعاينة لنسبة تبايني عينتين مستقلتين يتبع توزيع فيشر، ويمكن صياغة النظرية كما يلي:

نظرية 9: ان نسبة تباين عينتين مستقلتين حجمهما على الترتيب n , m، مسحوبتين من مجتمعين طبيعيين تبايناهما σ^2_X , σ^2_Y على التوالي تكتب كما يلي:

$$F=\frac{\left[\frac{S_X^2 n}{n-1}\right]\frac{1}{\sigma_X^2}}{\left[\frac{S_Y^2 m}{m-1}\right]\frac{1}{\sigma_Y^2}}=\frac{\hat{S}_X^2/\sigma_X^2}{\hat{S}_Y^2/\sigma_Y^2}\rightarrow F_{v_1,v_2}$$