المحور الأول: نظرية المعاينة وتوزيعاتها

تمهيد

تعرفنا في المحور التمهيدي على مفهوم مهم جدا وهو توزيع المعاينة، وقد عرفناه حينها بأنه توزيع القيم المأخوذة بواسطة إحصاءة ما (المتوسط الحسابي، التباين، ...) في جميع العينات الممكنة من نفس الحجم من نفس المجتمع. وقبل الخوض في مسألة المعاينة، دعونا نستعرض مبدأ عام سيسمح لنا بفهم ما سنقوم به خلال هذا المحور والمحور الذي سيليه (نظرية التقدير) وإظهار المغزى من دراسة العلاقة بين المجتمع والعينة

المطلب الرئيسي في نهاية المطاف سيكون حساب احتمال معين

$$P(X \geq 100) = ?$$
$$P(\bar{X} \leq 85) = ?$$
$$P(120 \leq \bar{X} \leq 150) = ?$$
$$P(0.45 \leq \bar{p} \leq 0.6) = ?$$
$$P(180 \leq \bar{X}_1 - \bar{X}_2 \leq 210) = ?$$

وفي كل تلك الحالات، سيتوجب علينا تحويل المتغير العشوائي X أو المتوسط الحسابي إلى القيمة المعيارية Z (التوزيع الطبيعي) أو إلى القيمة T (توزيع ستودنت) كل حسب الوضعيات التي سنصادفها فيما سيأتي لاحقا

فلو أخذنا على سبيل المثال لا الحصر الحالة التالية:

$$P(\bar{X} \leq 85) = ?$$

فإن هذا سيجرنا حتما (بعد التأكد من أن التوزيع المدروس هو التوزيع الطبيعي) لتحويل قيمة الاحتمال إلى:

$$P(\bar{X} \leq 85) = P\left(z \leq \frac{85 - \boldsymbol{\mu_{\bar{X}}}}{\boldsymbol{\sigma_{\bar{X}}}}\right) = ?$$

أين سيتحتم علينا تحديد قيمة كل من:

$\boldsymbol{\mu_{\bar{X}}}$: وهو ما يعرف بمتوسط توزيع المعاينة للمتوسطات (سيتم شرحه لاحقا)

$\boldsymbol{\sigma_{\bar{X}}}$: وهو ما يعرف بالانحراف المعياري لتوزيع المعاينة للمتوسطات (سيتم شرحه كذلك لاحقا)

وعليه، قبل حساب أي احتمال، فإنه يصبح من الواجب علينا تحديد قيمة كل من متوسط متوسطات العينات والانحراف المعياري لمتوسطات العينات. وهنا سوف يطلب منا في الغالب (بشكل مباشر أو غير مباشر) تكوين **توزيع المعاينة** لمتوسطات العينات

$$\boldsymbol{\mu_{\bar{X}}} = ?$$
$$\boldsymbol{\sigma_{\bar{X}}} = ?$$

هاتان القيمتان تخصان العينات (إحصاءات)، وسيتم حسابهما بالاعتماد على معلمات المجتمع، وهذا هو بالضبط – عموما – الهدف الأساسي من دراسة موضوع **المعاينة وتوزيعاتها**

مصطلحات أساسية وقواعد

☜ **المعاينة:** هي مجموعة من العينات الممكن سحبها من المجتمع الإحصائي. وفي المعاينة يتم حساب:

✓ **متوسط متوسطات العينة (ويسمى كذلك متوسط توزيع المعاينة للمتوسطات)** ويرمز له بالرمز $\mu_{\bar{X}}$ وهو يساوي:

$$\boldsymbol{\mu_{\bar{X}}} = \frac{\sum \bar{X}}{\text{عدد العينات الممكن سحبها}}$$

$$\mu_{\bar{X}} = \frac{\sum F_i \bar{X}}{\sum F_i}$$

✓ **تباين متوسطات العينة (ويسمى كذلك تباين توزيع المعاينة للمتوسطات)** ويرمز له بالرمز $\sigma_{\bar{X}}^2$ وهو يساوي:

$$\sigma_{\bar{X}}^2 = \frac{\sum(\bar{X}_i - \mu_{\bar{X}})^2}{\text{عدد العينات الممكن سحبها}}$$

$$\sigma_{\bar{X}}^2 = \frac{\sum F_i(\bar{X}_i - \mu_{\bar{X}})^2}{\sum F_i}$$

طرق سحب العينات

وبما أننا نتحدث عن اختيار أو سحب عينات من مجتمع ما، فإنه جدير بالذكر أو التذكير بأن هناك طريقتين لسحب العينات:

الطريقة الأولى: السحب يكون مع الإرجاع أو بالإرجاع

الطريقة الثانية: السحب يكون بدون إرجاع

وحسب كل طريقة، فإنه من الضروري جدا التمييز بين كل حالة من الحالات المذكورة في الجدول التالي

	سحب بإرجاع	سحب بدون إرجاع
متوسط المتوسطات	$\mu_{\bar{X}} = \mu$	$\mu_{\bar{X}} = \mu$
تباين المتوسطات	$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$	$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$
عدد العينات الممكن سحبها	N^n	$C_N^n = \frac{N!}{n!\,(N-n)!}$

يسمى المقدار $\left(\frac{N-n}{N-1}\right)$ بمعامل التصحيح

قواعد أساسية

القاعدة الأولى

عندما يكون السحب بإرجاع، أو عندما تكون النسبة بين حجم العينة n وحجم المجتمع N:

$$\frac{n}{N} < 0.05$$

فإن:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \blacksquare \Rightarrow \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

القاعدة الثانية

عندما يكون السحب بدون إرجاع، أو عندما تكون النسبة بين حجم العينة n وحجم المجتمع N:

$$\frac{n}{N} \geq 0.05$$

فإن:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \blacksquare \Rightarrow \sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

القاعدة الثالثة

إذا كان لدينا جدول من هذا الشكل:

X	X_1	X_2	...	X_N	Σ
$P(X)$	$P(X_1)$	$P(X_2)$	...	$P(X_N)$	1

فإن هذا النوع من الجداول يسمى **جدول التوزيع الاحتمالي**. وانطلاقا من هذا الجدول يتم حساب **معلمات المجتمع** بالشكل:

$$\mu = \sum XP(X)$$

$$\sigma^2 = \sum X^2P(X) - \mu^2 \Rightarrow \sigma = \sqrt{\sigma^2}$$

قاعدة الرابعة

إذا كان لدينا جدول من هذا الشكل:

$\bar{X}$	$\bar{X}_1$	$\bar{X}_2$	...	$\bar{X}_n$	Σ
$P(\bar{X})$	$P(\bar{X}_1)$	$P(\bar{X}_2)$	...	$P(\bar{X}_n)$	1

فإن هذا النوع من الجداول يسمى **جدول توزيع المعاينة**. وانطلاقا من هذا الجدول يتم حساب **إحصاءات العينة** بالشكل:

$$\mu_{\bar{X}} = \sum \bar{X}P(\bar{X})$$

$$\sigma_{\bar{X}}^2 = \sum \bar{X}^2P(\bar{X}) - \mu_{\bar{X}}^2 \Rightarrow \sigma_{\bar{X}} = \sqrt{\sigma_{\bar{X}}^2}$$

توزيع المعاينة للمتوسط الحسابي

يعتمد شكل ونوع توزيع المعاينة للمتوسط الحسابي للعينات على توزيع المجتمع الأصلي الذي اختيرت منه هذه العينات العشوائية، وسندرس طبيعة توزيع المعاينة للمتوسطات من خلال الحالات التالية:

الحالة الأولى: المجتمع يتبع التوزيع الطبيعي

الحالة الثانية: المجتمع لا يتبع التوزيع الطبيعي

الحالة الأولى: المجتمع يتبع التوزيع الطبيعي

وهنا كذلك يمكن التمييز بين حالتين:

أولا – عندما يكون تباين المجتمع الطبيعي معلوم

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

ملاحظة هامة:

عندما يكون **السحب بدون إرجاع**، فيتم إدخال **معامل التصحيح** على تباين العينة. ويصبح عندئذ:

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

ثانيا – عندما يكون تباين المجتمع الطبيعي مجهولا

أ- عندما يكون حجم العينة كبيرا $(n \geq 30)$

يصبح حينها التوزيع الطبيعي المعياري بالشكل:

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

إذا كان السحب مع الإرجاع، فنقوم باستبدال العلاقات السابقة بإضافة معامل التصحيح، وتصبح العلاقات كالتالي:

$$\sigma_{\bar{X}}^2 \cong \frac{S^2}{n-1} \times \frac{N-n}{N-1}$$

وتصبح:

$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N-1}}}$$

ب- عندما يكون حجم العينة صغيرا $(n < 30)$

حينها يصبح التوزيع بالشكل:

$$T = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n-1}}} \sim t_{n-1}$$

ملاحظة:

إذا كان السحب مع الإرجاع، فنقوم باستبدال العلاقات السابقة بإضافة معامل التصحيح، وتصبح العلاقات كالتالي:

$$\sigma_{\bar{X}}^2 \cong \frac{S^2}{n-1} \times \frac{N-n}{N-1}$$

وتصبح:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n-1}}\sqrt{\frac{N-n}{N-1}}}$$

الحالة الثانية: المجتمع لا يتبع التوزيع الطبيعي

قد يحدث في كثير من الأحيان أن يكون المجتمع الذي تسحب منه العينات **لا يتبع التوزيع الطبيعي**، فقد يكون ملتويا نحو اليمين أو نحو اليسار. وفي هذه الحالة نطبق **نظرية النهاية المركزية**

نظرية النهاية المركزية (نظرية تقارب التوزيعات)

إذا كانت: $x_1,\ x_2,\ \dots,x_n$ مفردات عينة عشوائية مكونة من n مفردة، ومسحوبة من مجتمع **لا يتبع التوزيع الطبيعي**، متوسطه الحسابي μ وتباينه σ^2، فإن توزيع المعاينة **يقترب من التوزيع الطبيعي** بمتوسط μ وتباين $\frac{\sigma^2}{n}$ كلما ازداد حجم العينة n $(n \geq 30)$

توزيع المعاينة للفرق بين متوسطي عينتين

قد يرغب أي باحث، في كثير من الأوقات، في إجراء مقارنة بين مجتمعين قصد معرفة الاختلاف بينهما من خلال مقارنة متوسطيهما، أي من خلال حساب الفرق بين متوسطيهما الحسابيين $(\mu_1 - \mu_2)$ حيث:

μ_1 يرمز للمتوسط لحسابي للمجتمع الأول

μ_2 يرمز للمتوسط لحسابي للمجتمع الثاني

الحالة الأولى: الفرق بين عينتين مستقلتين

الحالة الثانية: الفرق بين عينتين مرتبطتين

الحالة الأولى: الفرق بين عينتين مستقلتين

أولا- توزيع المعاينة للفرق بين متوسطي عينتين مسحوبتين من مجتمعين يتبعان التوزيع الطبيعي بتباينين معلومين

بما أن، في هذه الحالة، المجتمعين يتبعان التوزيع الطبيعي، فيكون:

$$\mu_{\bar{X}_1-\bar{X}_2}=\mu_{\bar{X}_1}-\mu_{\bar{X}_2}=\mu_1-\mu_2$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2}=\sigma^2_{\bar{X}_1}+\sigma^2_{\bar{X}_2}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$$

وبالتالي، تصبح الصيغة الرياضية للمتغير المعياري بالشكل:

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\backsim N(0,\ 1)$$

ملاحظة:

إذا كان **أحد المجتمعين أو كليهما لا يتبع التوزيع الطبيعي**، فيكون توزيع المعاينة للفروق **قريبا من التوزيع الطبيعي** بشرط أن تكون العينتين كبيرتين حسب نظرية النهاية المركزية

ثانيا- توزيع المعاينة للفرق بين متوسطي عينتين كبيرتين مسحوبتين من مجتمعين يتبعان التوزيع الطبيعي بتباينين مجهولين (n_1 و n_2 كبيرين)

$$\mu_{\bar{X}_1-\bar{X}_2}=\mu_{\bar{X}_1}-\mu_{\bar{X}_2}=\mu_1-\mu_2$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2}=\sigma^2_{\bar{X}_1}+\sigma^2_{\bar{X}_2}\simeq\frac{S_1^2}{n_1-1}+\frac{S_2^2}{n_2-1}$$

وبالتالي، تصبح الصيغة الرياضية للمتغير المعياري بالشكل:

$$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1-1}+\frac{S_2^2}{n_2-1}}}\backsim N(0,\ 1)$$

ثالثا- توزيع المعاينة للفرق بين متوسطي عينتين مسحوبتين من مجتمعين يتبعان التوزيع الطبيعي بتباينين مجهولين وعلى الأقل حجم إحدى العينتين صغير (n_1 و n_2 صغيرين أو أحدهما صغير)

وهنا يمكن التمييز بين حالتين:

الحالة الأولى: عندما يكون التباينين مجهولين ومتساويين

الحالة الثانية: عندما يكون التباينين مجهولين وغير متساويين

أ. عندما يكون التباينين مجهولين ومتساويين

$$S_P^2=\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}$$

حيث يكون:

متوسط الفرق بين متوسطي العينتين

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

تباين الفرق بين متوسطي العينتين

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = S_P^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

وتعطى إحصاءة ستودنت بالشكل:

$$T = \frac{(\bar{X}_1-\bar{X}_2)-\mu_1-\mu_2}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

ب. عندما يكون التباينين مجهولين وغير متساويين

في مثل هكذا حالات، يكون المتغير العشوائي معقد جدا. وقد درس هذه المشكلة كثير من الاحصائيين، ومن أهمهم فيشر وبهرين (Fisher – Behren) لذلك سميت هذه المشكلة باسميهما. وقد اقترحت عدة حلول لهذه المشكلة، أكثرها استخداما هو اعتبار أن توزيع المتغير هو توزيع قريب من توزيع ستودنت بدرجة حرية V تحسب بالصيغة:

$$V = \frac{\left[\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1-1}+\frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2-1}}$$

ويكون في هذه الحالة:

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}$$

ومنه:

$$T = \frac{(\bar{X}_1-\bar{X}_2)-\mu_1-\mu_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \backsim T_V$$

الحالة الثانية: الفرق بين عينتين مرتبطتين

وهنا نميز بين حالتين:

الحالة الأولى: حجم العينة كبير وتباين المجتمع مجهول

الحالة الثانية: حجم العينة صغير وتباين المجتمع مجهول

أ. عندما يكون حجم العينة كبير وتباين المجتمع مجهول

في هذه الحالة سوف يتبع متوسط الفرق للعينتين التوزيع الطبيعي، ويتم حساب معلمتي التوزيع كما يلي:

$$\mu_{\bar{d}} = \frac{\sum d_i}{n} = \frac{\sum(x_i-y_i)}{n}$$

$$\sigma^2_{\bar{d}} = \frac{\sigma^2_d}{n} \approx \frac{S^2_d}{n-1}$$

حيث:

$$S_d^2 = \frac{\sum(d_i - \mu_{\bar{d}})^2}{n}$$

ونكتب:

$$\bar{d} \sim N\left(\mu_{\bar{d}}, \frac{S_d^2}{n-1}\right)$$

ويكون عندئذ:

$$Z = \frac{\bar{d} - \mu_{\bar{d}}}{\frac{S_d}{\sqrt{n-1}}} \sim N(0,\ 1)$$

ب. عندما يكون حجم العينة صغير وتباين المجتمع مجهول

في هذه الحالة سوف يتبع متوسط الفرق للعينتين توزيع ستودنت بدرجة حرية V، ويتم حساب معلمتي التوزيع كما يلي:

$$\mu_{\bar{d}} = \frac{\sum|d_i|}{n} = \frac{\sum|x_i - y_i|}{n}$$

$$\sigma_{\bar{d}}^2 = \frac{\sigma_d^2}{n} \approx \frac{S_d^2}{n-1}$$

ونكتب:

$$\bar{d} \sim T_{(n-1)}\left(\mu_{\bar{d}}, \frac{S_d^2}{n-1}\right)$$

ويكون عندئذ:

$$T = \frac{\bar{d} - \mu_{\bar{d}}}{\frac{S_d}{\sqrt{n-1}}}$$

ملاحظة

قد يصادف الباحث، في بعض الأحيان، حالة توزيع المعاينة لمجموع متوسطين، فيعطى عندئذ المتوسط الحسابي والتباين لهذا النوع من التوزيع بالشكل:

$$\mu_{\bar{X}_1+\bar{X}_2} = \mu_1 + \mu_2$$

$$\sigma_{\bar{X}_1+\bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

توزيع المعاينة للنسب

سنوضح فيما يأتي كيفية تحديد المتوسط، التباين، وطبيعة التوزيع الاحتمالي لنسبة خاصية ما في العينة. بفرض أن لدينا مجتمعا معينا وأن بعض مفرداته تتوفر فيها صفة معينة (صفة محل الدراسة) وأن نسبة هذه المفردات في المجتمع هي p، فإذا سحبنا من هذا المجتمع عينة عشوائية من الحجم n وكان عدد من يتصفون بهذه الصفة في تلك العينة هو x فإن:

$$\bar{p} = \frac{x}{n}$$

حيث:

p تمثل نسبة المجتمع

$\bar{p}$ تمثل نسبة العينة

نحصل على توزيع للنسبة $\bar{p}$ ومعالمها:

$$\mu_{\bar{p}} = p$$

وكذلك:

حالة السحب بإرجاع	حالة السحب بدون إرجاع
$\sigma^2_{\bar{p}} = \frac{pq}{n} \blacksquare \Rightarrow \sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$	$\sigma^2_{\bar{p}} = \frac{pq}{n}\left(\frac{N-n}{N-1}\right) \blacksquare \Rightarrow \sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$

وكل ذلك عندما يكون: $n \geq 30$ (حجم العينة كبير)

وعندها تكون الإحصاءة Z للمتغير العشوائي $\bar{p}$ فتعطى بالعلاقة التالية:

$$Z = \frac{\bar{p} - \mu_{\bar{p}}}{\sigma_{\bar{p}}} = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}} \backsim N(0,\ 1)$$

توزيع المعاينة للفرق بين نسبتي عينتين

نميز هنا بين حالتين:

الحالة الأولى: السحب مع الإرجاع

الحالة الثانية: السحب بدون إرجاع

الحالة الأولى: السحب مع الإرجاع

يكون متوسط الفرق بين نسبتي العينتين المستقلتين بالشكل:

$$\mu_{\bar{p}_1 - \bar{p}_2} = \mu_{\bar{p}_1} - \mu_{\bar{p}_2} = p_1 - p_2$$

وتباين الفرق بين نسبتي العينتين المستقلتين يكون بالشكل:

$$\sigma^2_{\bar{p}^2 - \bar{p}^2} = \sigma^2_{\bar{p}_1} + \sigma^2_{\bar{p}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

ويعطى عندئذ المتغير العشوائي الطبيعي القياسي بالشكل:

$$z = \frac{(\mu_{\bar{p}_1} - \mu_{\bar{p}_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

الحالة الثانية: السحب بإرجاع

يكون متوسط الفرق بين نسبتي العينتين المستقلتين بالشكل:

$$\mu_{\bar{p}_1 - \bar{p}_2} = \mu_{\bar{p}_1} - \mu_{\bar{p}_2} = p_1 - p_2$$

وأما تباين الفرق بين نسبتي العينتين المستقلتين في هذه الحالة فيكون بالشكل:

$$\sigma^2_{\bar{p}^2 - \bar{p}^2} = \sigma^2_{\bar{p}_1} + \sigma^2_{\bar{p}_2} = \left(\frac{p_1 q_1}{n_1} \times \frac{N_1 - n_1}{N_1 - 1}\right) + \left(\frac{p_2 q_2}{n_2} \times \frac{N_2 - n_2}{N_2 - 1}\right)$$

ويعطى عندئذ المتغير العشوائي الطبيعي القياسي بالشكل:

$$z = \frac{(\mu_{\bar{p}_1} - \mu_{\bar{p}_2}) - (p_1 - p_2)}{\sqrt{\left(\frac{p_1 q_1}{n_1} \times \frac{N_1 - n_1}{N_1 - 1}\right) + \left(\frac{p_2 q_2}{n_2} \times \frac{N_2 - n_2}{N_2 - 1}\right)}}$$