المحور 01: نظرية المعاينة وتوزيعاتها

إنّ الهدف من نظرية المعاينة هو تحديد طرق المعاينة التي تحقق الخصائص الرئيسية التالية:

- أن تكون العينة ممثلة تمثيلا جيدا لمجتمع الدراسة، أي تحمل نفس خصائص المجتمع؛
- أن تكون التقديرات التي نحصل عليها لخصائص المجتمع من بيانات العينة دقيقة ويمكن قياس مصداقيتها؛
- أن تكون تكلفة اختيار العينة صغيرة جدا.

المحاضرة 01 : نظرية المعاينة، بعض المفاهيم الإحصائية والرياضية

من الضروري قبل الدخول في تفاصيل هذا المحور أن نتطرق إلى التعريف ببعض المصطلحات الإحصائية والتمييز بين المفاهيم مثل المجتمع الإحصائي والعينة، المعاينة بالإرجاع وبدون إرجاع، معالم المجتمع، إحصاءات العينة وخطأ المعاينة، نظرية النهاية المركزية، قانون الأعداد الكبيرة ....

- **المجتمع الإحصائي والعينة الإحصائية:**

✓ **يقصد بالمجتمع الإحصائي أو مجتمع الدراسة (Population)** جميع العناصر التي تشملها دراسة ما (يطلق عليها مفردات أو وحدات). فعلى سبيل المثال إذا كان الهدف من الدراسة هو التعرف على أعمار مجموعة من الطلبة بالمركز الجامعي ميلة في فترة ما، فإن المجتمع الإحصائي هو جميع الطلبة المسجلون بالمركز الجامعي ميلة في تلك الفترة.

✓ **تعرّف العينة على أنها جزء من المجتمع (Sample)**، فدراسة العينة هي دراسة لمجموعة فقط من مفردات المجتمع بدلا من إجراء حصر شامل للمجتمع ككل. فعلى سبيل المثال، إذا كان عدد الطلبة الذين تقدموا لامتحان مقياس الإحصاء 100 طالب، فالمجتمع في هذه الحالة هو علامات 100 طالب في مقياس الإحصاء. وإذا أخذنا علامات 20 طالب فقط من بين الذين تقدموا لهذا الامتحان نقول أنه تم اختيار عينة من علامات 20 طالب من مجتمع علامات مقياس الإحصاء.

- **المعاينة بالإرجاع والمعاينة بدون إرجاع:**

عندما يكون السحب بالإرجاع يعني المفردة التي تم سحبها بمكن أن تظهر أكثر من مرة في العينة، في هذه الحالة تسمى المعاينة بالمعاينة بالإرجاع. والعكس نسمي المعاينة بدون إرجاع إذا لم يكن السحب بدون إرجاع. فلو أردنا سحب عينة من المجتمع حجمها $\boldsymbol{n}$ وكان حجم المجتمع $\boldsymbol{N}$ وكان السحب بالارجاع فإن عدد الطرق المكنة لسحب عينة هو $\boldsymbol{N^n}$. أما إذا كان السحب بدون ارجاع فعدد الطرق المكنة هو $\mathbf{C_N^n} = \frac{N!}{(N-n)!\,n!}$.

- **معالم المجتمع واحصائيات العينة:**

✓ يسمى أي مقياس إحصائي في المجتمع بالمعلمة (Parameter) وعادة ما يرمز لمعالم المجتمع بالحروف الإغريقية، فالوسط الحسابي في المجتمع يرمز له بالرمز $\boldsymbol{\mu}$ (ميو)، والتباين في المجتمع بالرمز $\boldsymbol{\sigma^2}$ (سيجما مربع)، والنسبة في المجتمع بالرمز $\mathbf{P}$.

✓ يسمى أي مقياس إحصائي في العينة بـ "إحصائية" statistic وعادة ما يرمز للوسط الحسابي في العينة بالرمز $\mathbf{\bar{X}}$، وللتباين في العينة بـ $\boldsymbol{S^2}$ أو $\boldsymbol{\sigma_{\bar{X}}^2}$. وتجدر الإشارة إلى أن إحصائيات العينة تكون دائما عبارة عن متغير عشوائي (Random variable) عكس معالم المجتمع التي تكون دائما ثابتة.

فإذا كان حجم المجتمع هو N من المفردات فإن المتوسط الحسابي للمجتمع والذي يرمز له بالرمز μ هو:

$$\mu = \frac{\sum X_i}{N}$$

وهي قيمة ثابتة لهذا المجتمع.

أما التباين لهذا المجتمع فهو متوسط مجموع مربعات الانحرافات، أي:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$$

والاحصائية قيمة متغيرة، لأنها تختلف من عينة لأخرى داخل المجتمع الواحد. فمثلا في عينة حجمها n فإنّ المتوسط الحسابي لها هو:

$$\bar{X} = \frac{\sum X_i}{n}$$

والتباين هو:

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

- **خطأ المعاينة:**

قد تكون إحصائية العينة مساوية لمعلمة المجتمع، أو قد تكون أصغر أو أكبر منها. ويسمى الفرق بين إحصائية العينة ومعلمة المجتمع بخطأ المعاينة.

- **عزوم العينة والمجتمع (sample moments and population):**

لدينا متوسط العينة مساو لـ

$$\bar{X} = \frac{1}{n}\sum X_i$$

وهو يمثل المتوسط الحسابي لقياسات العينة وفي نفس الوقت عبارة عن تقدير لمتوسط المجتمع μ. عندئذ فإن التوقع (الأمل الرياضي) للمتغير العشوائي $\bar{X}$ هو:

$$\boldsymbol{E(\bar{X})} = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\,E\left(\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}\sum \mu = \boldsymbol{\mu}$$

أما التباين فيكون:

$$\boldsymbol{V(\bar{X})} = V\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\,V\left(\sum X_i\right) = \frac{1}{n^2}\sum V(X_i) = \frac{1}{n^2}\sum \sigma^2 = \boldsymbol{\frac{\sigma^2}{n}}$$

وهذا يعني أنّ متوسط العينة $\boldsymbol{\bar{X}}$ هو تقدير لمتوسط المجتمع $\boldsymbol{\mu}$، وله عزما من الرتبة الأولى حول نقطة الأصل هو $\boldsymbol{\mu}$ وعزما مركزيا من الرتبة الثانية هو $\boldsymbol{\frac{\sigma^2}{n}}$. وهذا يعني أنّ $\bar{X}$ هو متغير عشوائي يتبع وفق دالة احتمالية $f(\bar{X})$ أو $P(\bar{X})$ بمتوسط قدره $\boldsymbol{\mu}$ وتباين $\boldsymbol{\frac{\sigma^2}{n}}$. وبذلك الدالة الاحتمالية لـ $\bar{X}$ تعتمد على دالة التوزيع الاحتمالي للمتغير العشوائي X الذي أختيرت منه تلك العينة.

كذلك نفس الشيء بالنسبة لتباين العينة S^2:

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

$$\boldsymbol{E(S^2) = E\left[\frac{\sum (X_i - \bar{X})^2}{n-1}\right]}$$

$$= \frac{1}{n-1}\left[\sum E(X_i-\mu)^2 - nE(\bar{X}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] = \frac{(n-1)\,\sigma^2}{n-1} = \boldsymbol{\sigma^2}$$

أما بالنسبة لتباين S^2 فيكون:

$$\boldsymbol{V(S^2)} = V\left[\frac{\sum(X_i - \bar{X})^2}{n-1}\right] = \frac{1}{n-1}V\left[\sum(X_i - \bar{X})^2\right]$$
$$= \boldsymbol{\frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^2\right)}$$

علما أنّ: $\mu_4 = E(X-\mu)^4$

وبذلك، تباين العينة S^2 له عزما من الرتبة الأولى حول نقطة الأصل σ^2 وعزما مركزيا من الرتبة الثانية هو $\boldsymbol{\frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^2\right)}$. وهذا يعني أن S^2 متغير عشوائي يتبع دالة احتمالية مثل $f(S^2)$ بمتوسط قدره σ^2 وتباين قدره $V(S^2)$.

- **قانون الأعداد الكبيرة Law of Large Numbers (L.L.N) ومتباينة تشبيشيف Chybeshev `s Inequality**

✓ **قانون الأعداد الكبيرة:**

دائما عند المعاينة، يكون هدفنا الأساسي هو جعل الفرق المطلق بين $\bar{X}$، μ أي $|\bar{X} - \mu|$ قريبا من الصفر. وبذلك حجم العينة يلعب دورا كبيرا جدا في تحقيق هذا الهدف، وهو ما يطلق عليه بـ "قانون الأعداد الكبيرة".

يفترض قانون الأعداد الكبيرة أنه إذا كان $E(X_i) = \mu$ و $V(X_i) = \sigma^2$. وليكن $\bar{X}$ يمثل المتوسط الحسابي لهذه المتغيرات وأن: $E(\bar{X}) = \mu$، $V(\bar{X}) = \frac{\sigma^2}{n}$، عندئذ سوف يكون:

$$\boldsymbol{P\{|\bar{X} - \mu| \geq c\} \leq \frac{\sigma^2}{nc^2} \quad , c > 0}$$

وعندما تكون $\boldsymbol{n \to \infty}$ عند ذلك:

$$\lim_{n\to\infty} P\{|\bar{X} - \mu| \geq c\} \leq \lim_{n\to\infty} \frac{\sigma^2}{nc^2} = 0$$

إذن:

$$\boldsymbol{\lim_{n\to\infty} P\{|\bar{X} - \mu| \geq c\} = 0}$$

بعبارة أخرى:

$$\boldsymbol{\lim_{n\to\infty} P\{|\bar{X} - \mu| < c\} = 1}$$

وهذا يعني أنّه باحتمال قدره 1، $\bar{X}$ يقترب من μ عندما $n \to \infty$.

✓ **متباينة تشيبيشيف Chebyshevs Inequality:**

تنص هذه المتباينة على أنه إذا كان X متغير عشوائي بدالة احتمالية $f(x)$ وبمتوسط وتباين محدودين هما على التوالي μ ، σ^2، ولنفرض أنّ k عدد موجب، عندئذ سوف يكون:

$$\boldsymbol{P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2} \quad , k > 0}$$

إذا وضعنا $\boldsymbol{c = k\sigma}$، تصبح علاقة تشيبيشيف هي:

$$\boldsymbol{P\{|X-\mu| \geq c\} \leq \frac{\sigma^2}{c^2}}$$

- **نظرية النهاية المركزية The Central Limit Theorem(C.L.T)**

إذا كانت $X_1, X_2, \dots \dots . X_n$ متغيرات عشوائية مستقلة من مجتمع لا نهائي (تتوزع وفق نفس دالة الكثافة الاحتمالية $f(x)$، ولنفرض أن: $E(X_i) = \mu$ ، $V(X_i) = \sigma^2$. فإذا سحبنا عينة عشوائية كبيرة حجمها n من هذا المجتمع فإن توزيع المتوسط الحسابي للعينة $\bar{X}$ سوف يتبع بالتقريب التوزيع الطبيعي بمتوسط حسابي μ وتباين $\frac{\sigma^2}{n}$.