

المحور الأول

نظرية المعاينة وتوزيعاتها

## تمهيد:

تهتم نظرية المعاينة بدراسة العلاقة بين المجتمع والعينات المسحوبة منه ضمن ما يسمى الاحصاء الاستدلالي، وذلك من خلال دراسة بيانات العينة ثم تعميم النتائج المتوصل إليها على المجتمع بأكمله.

## أولاً: مفاهيم احصائية ورياضية

من بين أهم المفاهيم المرتبطة بالمعاينة الاحصائية ما يلي:

- **المجتمع الاحصائي:** هو مجموعة المفردات أو العناصر التي تشترك في خصائص أو خاصية محددة تميزها عن غيرها من المجتمعات.

- **العينة:** هي جزء من المجتمع الاحصائي يتم اختيارها منه بطرق مختلفة بهدف تعميم النتائج المتوصل إليها من خلال دراستها على المجتمع، والذي يتحدد مدى تمثيل العينة له بعدة عوامل منها: حجم العينة وتباين خصائص المجتمع.

- **العينة العشوائية البسيطة:** هي عينة يتم اختيارها من المجتمع عشوائياً، بحيث يكون لكل مفردة من مفردات المجتمع نفس الفرصة في تكوين العينة (نفس احتمالية الظهور).

- **الاحصاءة:** هي دالة في مفردات العينة كالوسط الحسابي، الانحراف المعياري للعينة، معامل الارتباط.... الخ.

- **المعلمة:** هي دالة في مفردات المجتمع الاحصائي المدروس بطريقة علمية للاستدلال على خواصه.

- **المعاينة:** هي عملية اختيار جزء من مفردات المجتمع الاحصائي المدروس بطريقة علمية للاستدلال على خواصه.

- **توزيع المعاينة (التوزيع العيني):** هو التوزيع الاحتمالي للاحصاءات العينة الذي يمكن الحصول عليه من خلال دراسة العلاقة بين احصاءة العينة ومعالم المجتمع، عن طريق سحب عينة عشوائية حجمها  $n$  وتكرار العملية للحصول على جميع العينات الممكنة.

## ملاحظات:

- تسمى دراسة جميع مفردات المجتمع بالحصص الشامل.

- معالم المجتمع تكون دائماً ثابتة في حين أن احصاءات العينة متغيرات عشوائية.

- قد تكون احصاءة العينة مساوية لمعلمة المجتمع وقد تكون أصغر منها وأكبر منها، ويسمى الفرق بين احصاءة

العينة ومعلمة المجتمع بخطأ المعاينة أو الخطأ العشوائي، وهذا بافتراض أن العينة عشوائية وعدم وجود أخطاء

غير عشوائية (الأخطاء في تجميع البيانات وتدوينها وتبويبها).

- عزوم العينة والمجتمع: متوسط العينة مساو لـ:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

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

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

أما التباين فيكون:

$$V(X) = \frac{\sigma^2}{n}$$

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

كذلك نفس الشيء بالنسبة لتباين العينة  $S^2$ :

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

أما بالنسبة لتباين  $S^2$  فيكون:

$$V(S^2) = \frac{1}{n} \left( \mu - \frac{n-3}{n-1} \sigma^2 \right)$$

علما أن:  $\mu_4 = E(X - \mu)^4$

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

- نظرية النهاية المركزية: تعد مبرهنات النهاية المركزية مجموعة نتائج لنظرية الاحتمالات تنص على أن مجموعة عدة متغيرات عشوائية مستقلة ومتشابهة التوزع تميل إلى التوزع حسب توزيع احتمالي معين، وهي تبين الشروط العامة المفروضة على المتغيرات العشوائية  $X_1; X_2; \dots; X_n$  لكي يؤول مجموعها  $Y_n$  إلى التوزيع الطبيعي عندما  $n \rightarrow +\infty$ ، وبشكل عام يمكن التعبير عن المحتوى العام لمبرهنات النهاية المركزية كالاتي:

• مبرهنة النهاية المركزية بصيغة لياونوف:

إذا كانت  $\{X_n\}$  متتالية من المتغيرات العشوائية تحقق الشروط التالية:

▪ القيم المتوقعة  $E(X_i) = \mu$  موجودة.

▪ التباينات  $Var(X_i) = \sigma^2$  موجودة.

▪ العزوم المركزية المطلقة من الرتبة الثالثة  $E(|X_i - m|^3)$  موجودة.

$$\lim_{n \rightarrow +\infty} \frac{\sum [(|X_1 - m|^3)]}{(\sum_{i=1}^n \sigma_i^2)^{\frac{3}{2}}} = 0$$

فإن توزيع المجموع  $Y_n = \sum_{i=1}^n X_i$  ينتهي إلى التوزيع الطبيعي  $N(n\mu; n\sigma^2)$  عندما تكون  $n$  كبيرة، وإذا

رمزنا  $Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$  فإن هذا المتغير يخضع تقريبا للتوزيع الطبيعي المعياري الطبيعي المعياري.

• مبرهنة النهاية المركزية بصيغة موافر ولا بلاس: إذا كانت متتالية المتغيرات العشوائية مستقلة ولكل منها توزيع

بيرنولي  $B(1, P)$  فإن توزيع المجموع  $Y_i = \sum_{i=1}^n X_i$  يتقارب من التوزيع الطبيعي  $N(np; npq)$  بازياد

$n$  أي أن:

$$\lim_{n \rightarrow +\infty} B(n, P) = N(np; npq)$$

- قانون الأعداد الكبيرة: عند المعاينة يكون الهدف الأساسي جعل الفرق المطلق قريبا من الصفر، ولتحقيق هذا

الهدف يجب أن يكون حجم العينة كبيرا وهو ما يطلق عليه قانون الأعداد الكبيرة.

- متراجحة تشببشيف: هي متراجحة مشهورة ترجع إلى عالم الرياضيات الروسي بافنوتي تشببشيف تمكن من الفهم

الدقيق لكيفية أن التباين يقيس التغير حول المتوسط للمتغير العشوائي.

فإذا علم التوزيع الاحتمالي للمتغير العشوائي  $X$  فإنه يمكن تحديد تباينه  $V(X)$  ومتوسطه  $E(X)$ ، ولكن

من غير الممكن إيجاد التوزيع الاحتمالي للمتغير العشوائي بمجرد معرفة تباينه ومتوسطه، وبالتالي فإنه من غير

الممكن حساب الاحتمالات لهذا التوزيع في هذه الحالة، إلا أنه توجد إمكانية لحساب حد أعلى أو حد أدنى لهذه

الاحتمالات وذلك باستخدام متباينة تشببشيف، وتنص نظرية تشببشيف على مايلي:

• احتمال المتغير العشوائي  $X$  يأخذ القيمة  $x$  من القيمة الممكنة  $L$   $X$  تحرف عن متوسطه  $\mu$  بالمقدار  $\sigma k$

على الأكثر يكون على الأقل هو  $1 - \frac{1}{k^2}$ ، ويعبر عن ذلك بالعبارة الاحتمالية التالية:

$$P(|X - \mu| \leq \sigma k) \geq 1 - \frac{1}{k^2}$$

أو:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

حيث:  $k$  ثابت أكبر من الواحد ( $k > 1$ )

أي أن هذه القيمة داخل الفترة  $[\mu - k\sigma, \mu + k\sigma]$  وإذا كانت  $\sigma$  و  $\mu$  مجهول للمجتمع وقدرت من العينة بالمقدارين  $S^2$  و  $\bar{X}$  فإن الفترة السابقة تصبح  $[\bar{X} - \sigma k, \bar{X} + \sigma k]$

• إذا كان المتغير العشوائي  $\bar{X}$  لجميع متوسطات العينات التي حجمها  $n$  من المجتمع الأصلي ذو المتوسط  $\mu_{\bar{X}}$  وتباين  $\sigma_{\bar{X}}^2$  فإن احتمال انحراف القيمة  $\bar{X}$  لأي من متوسطات العينات الممكنة تتحرف عن متوسطه  $\mu$  بالمقدار  $k\sigma_{\bar{X}}$  على الأكثر هو على الأقل  $1 - \frac{1}{k^2}$  ويعبر عنه بالعلاقة الاحتمالية التالية:

$$P(|\bar{X} - \mu| \leq k\sigma_{\bar{X}}) \geq 1 - \frac{1}{k^2}$$

أي أن هذه القيمة تقع داخل الفترة  $(\mu - k\sigma_{\bar{X}}, \mu + k\sigma_{\bar{X}})$

- الاحصاءات المرتبة: إذا كانت  $X_1; X_2; \dots; X_n$  عينة عشوائية من توزيع مستمر و  $x = x_1 + x_2 + \dots + x_n$  قيم لـ  $X$ ، فالقيم الملاحظة  $x_1; x_2; \dots; x_n$  يمكن ترتيبها تصاعديا حسب قيمها (من الأصغر إلى الأكبر) ونرمز لها  $y_1; y_2; \dots; y_n$  حيث:

$$y_1 \leq y_2 \leq y_3 \leq \dots \leq y_n \dots \dots \dots (1)$$

ولنرمز بـ  $X_k$  للمتغير العشوائي الذي يأخذ القيمة  $x_k; k = 1, \dots, n$  عند أي قيمة ملاحظة  $x$ ، وبذلك نعرف متتالية جديدة من المتغيرات العشوائية  $Y_1; Y_2; \dots; Y_n$  حيث:

$$Y_1 \leq Y_2 \leq Y_3 \leq \dots \leq Y_n \dots \dots \dots (2)$$

تدعى المتغيرات الجديدة  $Y_k$  بالاحصاءات المرتبة للعينة ويدعى المتغير  $Y_k$  بالاحصاءة ذو الترتيب  $k$ ، ومن ثم تدعى  $Y_1; Y_2; \dots; Y_n$  بالعينة المرتبة، كما تدعى السلسلة المعرفة بالعلاقة (2) بالمتسلسلة المتغيرة للعينة  $X_1; X_2; \dots; X_n$ ، وهذا ما يعني أن السلسلة المتغيرة لـ  $X$  ما هي إلا عينة عناصرها مرتبة تصاعديا حسب قيمها.