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.

أولا: 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 عشوائيا، 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 بالحصر الشامل.

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

- قد تكون إحصاءة العينة مساوية لمعلمة المجتمع وقد تكون أصغر منها وأكبر منها، ويسمى الفرق بين إحصاءة العينة ومعلمة المجتمع بخطأ المعاينة أو الخطأ العشوائي، وهذا بافتراض أن العينة عشوائية وعدم وجود أخطاء غير عشوائية (الأخطاء في تجميع البيانات وتدوينها وتبويبه).

- **عزوم العينة والمجتمع:** متوسط العينة مساو لـ:

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

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

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

أما التباين فيكون:

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

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

كذلك نفس الشيئ بالنسبة لتباين العينة S^2:

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

$$E(S^2) = \sigma^2$$

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 σ^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 σ^2 وتباين قدره $V(S^2)$.

- **التوزيعات التقاربية:** يمكن توضيحها من خلال ما يلي:

- **التقارب بالاحتمال:** نقول أن متتالية متغيرات عشوائية عشوائية $X_1; X_2; \ldots; X_n$ $\{X_n, n \geq 1\}$ متقاربة بالاحتمال إلى متغير عشوائي x إذا كان:

$$\lim_{n\to+\infty} P(|X_n - X| \geq \xi) = 0$$

ويرمز له:

$$X_n \xrightarrow{P} X$$

مثال: لتكن $X_1; X_2; \dots; X_n$ متتالية متغيرات عشوائية حيث:

$$p(X_n = 1) = \frac{1}{n}$$

$$p(X_n = 0) = 1 - \frac{1}{n}$$

أثبت أن: $X_n \xrightarrow{p} 0$

نتيجة: $X_1; X_2; \dots; X_n$ و $Y_1; Y_2; \dots; Y_n$ متتاليتين للمتغيرات العشوائية ولتكن $f: R \longrightarrow R$ دالة مستمرة لنفرض أن:

$$X_n \xrightarrow{p} X$$

$$Y_n \xrightarrow{p} Y$$

فإن:

$$\forall\, \alpha, \beta \in R \qquad \alpha X_n + \beta\gamma_n \xrightarrow{p} \alpha X + \beta Y$$

$$X_n Y_n \xrightarrow{p} XY$$

$$f(X_n) \xrightarrow{p} f(X)$$

- **التقارب بالمتوسط:** نقول عن متتالية من المتغيرات العشوائية القابلة للمكاملة أنها متقاربة بالمتوسط نحو متغير عشوائي x إذا كان :

$$\lim_{n\to+\infty} E(|X_n - X|) = 0$$

ويرمز له:

$$X_n \xrightarrow{m} X$$

و نقول أن متتالية من المتغيرات العشوائية $\{X_n; n \geq 1\}$ متقاربة بالمتوسط التربيعي نحو المتغير نحو المتغير العشوائي X إذا كان:

$$\lim_{n\to+\infty} E(|X_n - X|^2) = 0$$

مثال: لتكن $\{X_n; n \geq 1\}$ متتالية من المتغيرات العشوائية

$$\forall n \geq 1 \qquad P(X_n = 0) = 1 - e^{-n}$$

$$P(X_n = 1) = e^{-n}$$

بين أن $\{X_n; n \geq 1\}$ متقاربة نحو 0 بالمتوسط

- **التقارب شبه المؤكد:** نقول عن متتالية من المتغيرات العشوائية أنها متقاربة بشكل شبه مؤكد نحو متغير عشوائي x إذا كان :

$$\sum_{n} p(|X_n - X| \geq \xi) < \infty \qquad \forall \xi > 0$$

ويرمز له:

$$X_n \overset{ps}{\to} X$$

مثال: $X_1; X_2; \ldots\ldots; X_n$ متتالية متغيرات عشوائية حيث:

$$P(X_n = 0) = 1 - \frac{1}{n^2}$$

$$P(X_n = 1) = \frac{1}{n^2}$$

أثبت أن: $X_n \overset{ps}{\to} X$

- **التقارب بالقانون:** إذا كانت $X_1; X_2; \ldots\ldots; X_n$ متغيرات عشوائية ولتكن $\{F_n; n \geq 1\}$ متتالية توابع التوزيع لها:

$$\forall n \geq 1\ F_n(x) = F_{X_n}(x)$$

أو:

$$\lim_{n \to +\infty} E(f(X_n)) = E[f(X)]$$

ويرمز له:

$$X \overset{loi}{\to} X_n$$

بشرط دالة f مستمرة ومحدودة.

مثال: $X_1; X_2; \ldots\ldots; X_n$ متتالية متغيرات عشوائية مستقلة لها نفس دالة الكثافة الاحتمالية حيث:

$$f(x) = \alpha x^{-(\alpha+1)} \qquad x \in [1, +\infty[$$

نضع من أجل $n \geq 1$

$$y_n = n^{-\frac{1}{\propto}}\ \ \mathrm{Max}\ X_k \quad 0 \leq k \leq 1$$

بين أن: $X \overset{loi}{\to} X_n$

- **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 $X_1; X_2; \ldots\ldots; X_n$ لكي يؤول مجموعها Y_n إلى التوزيع الطبيعي عندما $n \to +\infty$، وبشكل عام يمكن التعبير عن المحتوى العام لمبرهنات النهاية المركزية كالآتي:

- **مبرهنة النهاية المركزية بصيغة ليابونوف:**

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.

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

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

- **مبرهنة النهاية المركزية بصيغة موافر ولابلاس:** 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 $N(np; npq)$ بازدياد n أي أن:

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

- **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 تشبيشيف:** هي متراجحة مشهورة ترجع إلى عالم الرياضيات الروسي بافنوتي تشبيشيف 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 تشبيشيف، وتنص نظرية تشبيشيف على مايلي:

- احتمال المتغير العشوائي X يأخذ القيمة x من القيمة الممكنة لـ X تنحرف عن متوسطه μ بالمقدار σk 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 σ و μ 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 تنحرف عن متوسطه μ بالمقدار $k\sigma_{\bar{X}}$ على الأكثر هو على الأقل $1 - \frac{1}{k^2}$ ويعبر عنه بالعبارة الاحتمالية التالية:

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

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

المثال الأول: ليكن $X \backsim N(64.4)$ أوجد الحد الأعلى لاحتمال حدوث الحادثة $\{|X - \mu| \geq 6\}$ باستخدام متباينة تشبيشيف.

المثال الثاني: إذا كان لدينا مجتمعا من طلاب جامعة ما له متوسط $\mu = 70$ كلغ وانحراف معياري 6 كلغ أخذت منه عينة عشوائية حجمها 36 طالبا، أوجد احتمال انحراف متوسطها $\bar{X}$ عن μ بطريقتين حسب العبارة الاحتمالية التالية:

$$P(|\bar{X} - \mu| \leq 2)$$

- **القانون الضعيف للأعداد الكبيرة:** يعتبر هذا القانون من نتائج نظرية تشبيشيف ويستفاد منه بشكل خاص في نظرية المعاينة، وتصاغ هذه النظرية بالشكل التالي:

لتكن المتغيرات $X_1; X_2; \dots; X_n$ متغيرات عشوائية مستقلة لها نفس التوزيع الاحتمالي ولكل منها نفس المتوسط والتباين إذا كان:

$$X = X_1 + X_2 + \cdots + X_n$$

$$\lim_{n \to +\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} E(X)\right| < \xi\right) = 1 \qquad \xi > 0$$

- **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:

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

ولنرمز ب 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:

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

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

- **التوزيع الاحتمالي الهامشي لإحصاءة مرتبة:** إذا كانت $\mathrm{X_1;X_2;\ldots\ldots;X_n}$ عينة عشوائية من توزيع مستمر دالة كثافته الاحتمالية $f(x)$ ودالة توزيعه $F(x)$ وكانت $Y_1;Y_2;\ldots;Y_n$ الاحصاءات المرتبة للعينة $\mathrm{X_1;X_2;\ldots\ldots;X_n}$ ، فإن التوزيع الاحتمالي للإحصاءة المرتبة Y_i يعرف بدالة الكثافة التالية:

$$g(y_i)=\frac{n!}{(i-1)!\,(n-i)!}f(y_i)[F(y_i)]^{i-1}[1-F(y_i)]^{n-i}$$

بوضع $i=1$ و $n=i$ نتحصل على دالة الكثافة الاحتمالية لتوزيع Y_1 و Y_n على التوالي:

$$\mathrm{g(y_1)=n[1-F(y_1)]^{n-1}f(y_1)}$$
$$\mathrm{g(y_n)=n[F(y_n)]^{n-1}f(y_n)}$$

المثال الأول:

بفرض $\mathrm{X_1;X_2;\ldots\ldots;X_n}$ عينة عشوائية من التوزيع المنتظم $R[0,1]$ أوجد توزيع كل من Y_n و Y_1

المثال الثاني:

إذا كانت $X_1;X_2;\ldots\ldots;X_n$ عينة عشوائية من توزيع أسي أوجد توزيع كل من $y_{n-1};y_2$

- **التوزيع المشترك لإحصائيتين مرتبتين:** إذا كانت $\mathrm{X_1;X_2;\ldots\ldots;X_n}$ عينة عشوائية من مجتمع دالة كثافته الاحتمالية $f(x)$ ودالة توزيعه $F(x)$ فإن دالة الكثافة الاحتمالية المشتركة للاحصائيتين المرتبتين Y_J و Y_i هي:

$$g(y_i;y_j)=\frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}[F(y_i)]^{i-1}\left[1-F(y_j)\right]^{n-j}\left[F(y_j)-F(y_i)\right]^{j-i-1}f(y_i)f(y_j)$$

المثال الأول: بفرض $\mathrm{X_1;X_2;\ldots\ldots;X_n}$ عينة من التوزيع المنتظم، أوجد التوزيع المشترك للاحصائيتين y_n و y_1 .

المثال الثاني: بفرض $\mathrm{X_1;X_2;\ldots\ldots;X_n}$ عينة عشوائية من مجتمع كثافة توزيعه

$$f(x)=\frac{1}{2}e^{-\frac{x}{2}}\qquad\qquad x>0$$

أوجد التوزيع المشترك للاحصائيتين y_n و y_1