المحاضرة الأولى: توزيعات المعاينة

1- مقدمة:

العينة هي عبارة عن فئة أو مجموعة جزئية من المجتمع يتم اختيار وتحليل بياناتها وذلك بهدف تقدير معالم المجتمع غير المعلومة، أو اختبار فروض بشأنها، وبشكل عام لاستنباط معلومات عن معالم المجتمع المسحوب منه العينة، نقوم بحساب الوسط الحسابي للعينة على سبيل المثال ويكون هذا الوسط تقديرا لمتوسط المجتمع المجهول، ويسمى هذا التقدير: **إحصاءة** (statistic).

والقاعدة العامة تقول أن أي دالة في المتغيرات العشوائية المكونة لعينة المشاهدات تسمى **إحصاءة**.

وعليه إذا كانت لدينا عينة عشوائية $(X_1,X_2,\ldots\ldots\ldots\ldots, Xn)$ حجمها n فإن الوسط الحسابي لهذه العينة هو الإحصاءة $\overline{X}$ ، حيث:

$$\overline{X} = \Sigma Xi / n \quad \ldots\ldots\ldots\ldots(1)$$

بينما تباين العينة Sample variance هو الإحصاءة S^2 :

$$S^2 = \Sigma(x_i - \overline{X})^2 / n-1 \quad \ldots\ldots.(2)$$

و ما يجب الإشارة إليه، هو أنه في حالة عدم معرفة تباين المجتمع فإنه يتم استخدام تباين العينة S^2 كتقدير له، أما الانحراف المعياري للعينة S فهو عبارة عن الجذر التربيعي لتباين العينة S^2، ويستخدم أيضا كتقدير للانحراف المعياري للمجتمع σ .

كذلك استخدمنا (n -1) في مقام المعادلة (2) بدلا من n وذلك لكي يكون التقدير الناتج تقديرا غير متحيزا (Unbiased Estimate) للمعلمة σ^2، أي يكون:

$$E(S^2) = \sigma^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

إن العلاقة (3) تكون صحيحة إذا كان المجتمع المسحوب منه العينة غير محدود (غير منته)، أو عندما يكون حجم المجتمع (N) كبير جدا. أما إذا كان المجتمع محدود ومكون من القيم التالية: $(X_1, X_2, \ldots\ldots\ldots\ldots X_N)$

يكون:

$$E(S^2) = \frac{N}{N-1}\sigma^2 \quad .... (4)$$

مع العلم أن:

$$\sigma^2 = \Sigma(X_i - \mu)^2 / N \quad .... (5)$$

$$\mu = \Sigma X_i / N \quad .... (6)$$

2- المعاينة بإرجاع والمعاينة بدون إرجاع:

إذا كان لدينا صندوق يحتوي على سبعة بطاقات مرقمة من 1 إلى 7 وقمنا بسحب بطاقة من هذا الصندوق، فانه لدينا الخيار في إرجاع هذه البطاقة أو عدم إرجاعها قبل إجراء عملية السحب الثانية. ففي الحالة الأولى فان البطاقة يمكن أن تظهر عدة مرات، بينما في الحالة الثانية يمكن أن تظهر البطاقة مرة واحدة فقط، وبالتالي ففي العينات التي يمكن أن نختار فيها مفردات المجتمع أكثر من مرة تسمى **بالمعاينة بإرجاع**. بينما إذا كانت المفردة في المجتمع لا يمكن اختيارها أكثر من مرة فتسمى **المعاينة بدون إرجاع**.

وفي الحالة العامة إذا كان لدينا مجتمعا من المفردات يتبع توزيعا معينا، ونريد اختيار عينة حجمها n من هذا المجتمع، فانه يمكن اختيار هذه العينة وفق الطريقتين التاليتين (السحب مع الإرجاع والسحب بدون إرجاع)، حيث نجد أن عدد العينات التي سوف يتم سحبها من هذا المجتمع وفق الطريقتين السابقتين يتم تحديده كما يلي:

N^n في حالة السحب مع الإرجاع.

C_N^n في حالة السحب بدون إرجاع.

3- توزيع المعاينة:

بفرض أننا قمنا باختيار عينة عشوائية حجمها n من مجتمع معين، ثم بعد ذلك قمنا بحساب مقياسا معينا لهذه العينة وليكن على سبيل المثال الوسط الحسابي $\overline{X}$. ثم اخترنا عينة ثانية لها نفس الحجم n وقمنا بحساب نفس المقياس السابقة، واخترنا عينة ثالثة وحسبنا منها المقياس نفسه، وهكذا بالنسبة لجميع العينات التي يمكن سحبها من هذا المجتمع، فانه سيتوفر لدينا عدد كبير من القيم للوسط الحسابي $\overline{X}$، هذه القيم تكون مجتمعا آخر عدد مفرداته اكبر بكثير من عدد مفردات المجتمع الأصلي.

ومن هذا المنطلق يمكن النظر إلى هذا المقياس على انه متغير عشوائي يأخذ قيما مختلفة (طبعا هي القيم التي حصلنا عليها من هذه العينات)، ويتبع توزيعا معينا قد يختلف عن توزيع المجتمع الأصلي، ويسمى هذا التوزيع **بتوزيع المعاينة** لهذا المقياس سواء كان هذا المقياس هو الوسط الحسابي للعينة $\overline{X}$ أو الانحراف المعياري لها S.

وبالتالي يمكن القول بأن توزيع المعاينة لأية إحصاءة من إحصاءات العينة، $\overline{X}$)، S^2) هو التوزيع الاحتمالي لجميع القيم الممكنة لهذه الإحصاءة، والتي نحصل عليها عند سحب كل العينات بنفس الحجم والطريقة ومن نفس المجتمع.

مثال1: إذا كان لدينا مجتمع مكون من خمس وحدات (N=5)، وكانت قيم ظاهرة معينة لهذه الوحدات هي: 1.5 ، 3 ، 6 ، 4.5 ، 7.5

المطلوب:

1- أحسب الوسط الحسابي (μ) والتباين (σ^2).
2- أحسب الوسط الحسابي $\overline{X}$ لجميع العينات البسيطة الممكنة والتي حجم كل منها ثلاث وحدات، وكذلك حدد جدول التوزيع الاحتمالي للوسط الحسابي ($\overline{X}$)، ومنه أحسب القيمة المتوقعة والتباين للمتغير($\overline{X}$).
3- أحسب التباين S^2 لجميع العينات العشوائية الممكنة التي حجم كل منها ثلاث وحدات واكتب التوزيع الاحتمالي للمتغير S^2 وتحقق من أن:

$$E(S^2)=\frac{N}{N-1}\sigma^2$$

الحل :

1- الوسط الحسابي وتباين المجتمع:

$$\mu=\Sigma X_i / N = (1.5 + 3 + 6 + 4.5 + 7.5) / 5 = 4.5$$

$$\sigma^2 =\Sigma(X_i - \mu)^2 / N$$

$$\sigma^2 = [(1.5-4.5)^2 +(3-4.5)^2 + (6-4.5)^2 +(4.5-4.5)^2 + (7.5-4.5)^2] / 5$$

$$\sigma^2 = 4.5$$

2- بما أن السحب تم بدون إرجاع فإن العينات الممكنة عددها عشرة وذلك حسب:

$$C_N^n = C_5^3 = \frac{5!}{3!2!} = 10$$

والجدول الموالي يوضح مختلف هذه العينات وأوساطها الحسابية.

الجدول (1) العينات العشرة الممكنة والوسط الحسابي المقابل لكل منهما

رقم العينة	القيم المختلفة للعينة x_i (x_1, x_2, x_3)	$\sum x_i$	$\bar{X}_i$
01	(1.5, 3, 6)	10.5	3.5
02	(1.5, 3, 4.5)	9	3
03	(1.5, 3, 7.5)	12	4
04	(1.5, 6, 4.5)	12	4
05	(1.5, 6, 7.5)	15	5
06	(1.5, 4.5, 7.5)	13.5	4.5
07	(3, 6, 4.5)	13.5	4.5
08	(3, 6, 7.5)	16.5	5.5
09	(3, 4.5 ,7.5)	15	5
10	(6, 4.5, 7.5)	18	6

من خلال هذا الجدول فإننا نستطيع تحديد جدول التوزيع الاحتمالي للوسط الحسابي ($\bar{X}$) لعينة عشوائية حجمها ثلاث وحدات، وذلك كما يلي:

الجدول (2): التوزيع الاحتمالي للوسط الحسابي ($\bar{X}$) لعينة عشوائية حجمها ثلاث وحدات.

$\bar{X}_i$	3	3.5	4	4.5	5	5.5	6
$P(\bar{X}_i)$	1/10	1/10	2/10	2/10	2/10	1/10	1/10

انطلاقا من هذا الجدول فإن القيمة المتوقعة للوسط الحسابي ($\bar{X}$) هي:

$$E(\bar{X}) = \Sigma \bar{X}_i P(\bar{X}_i) \text{.........(7)}$$

$$E(\overline{X})=3(1/10)+3.5(1/10)+4(2/10)+4.5(2/10)+5(2/10)+5.5(1/10)+6(1/10) = 4.5$$

مما سبق نستنتج أن: $\mathbf{E(\overline{X})=\mu = 4.5}$

أما تباين الوسط الحسابي للعينة فيحسب بالعلاقة التالية:

$$Var(\overline{X}) = E(\overline{X})^2-[E(\overline{X})]^2 ....(8)$$

حيث:

$$E(\overline{X})^2 = \Sigma \overline{X}_i^2 . P(\overline{X}_i)$$

$$E(\overline{X})^2=3^2(1/10)+3.5^2(1/10)+4^2(2/10)+4.5^2(2/10)+5^2(2/10)+(5.5)^2(1/10)+6^2(1/10) = 21$$

$$\Rightarrow Var(\overline{X}) = 21-(4.5)^2 = \mathbf{0.75}$$

3- حساب التباين S^2 لجميع العينات العشوائية الممكنة: ويتم ذلك كما يلي:

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

$$= \Sigma(x_i-\overline{X})^2 / 3-1$$

يمكن تلخيص قيم S^2 المناظرة لكل عينة من العينات العشرة الممكنة في الجدول (3).

الجدول (3): قيم S^2 المناظرة للعينات العشرة

S^2	$\Sigma(x_i-\overline{X})^2$	$\overline{X}_i$	القيم المختلفة للعينة (x_1, x_2, x_3)	رقم العينة
5.25	10.5	3.5	(1.5, 3, 6)	1
2.25	4.5	3	(1.5, 4.5, 3)	2
9.75	19.5	4	(1.5, 3, 7.5)	3
5.25	10.5	4	(1.5, 6, 4.5)	4
9.75	19.5	5	(1.5, 6, 7.5)	5
9	18	4.5	(1.5, 4.5 ,7.5)	6

7	(3, 6, 4.5)	4.5	4.5	2.25
8	(3, 6, 7.5)	5.5	10.5	5.25
9	(3, 4.5, 7.5)	5	10.5	5.25
10	(6, 4.5, 7.5)	6	4.5	2.25

وانطلاقا من هذا الجدول فإنه يمكننا الحصول على التوزيع الاحتمالي للمتغير S^2 كما هو موضح في الجدول (4) التالي:

الجدول (4): التوزيع الاحتمالي للمتغير S^2

S^2	2.25	5.25	9	9.75
$P(S^2)$	3/10	4/10	1/10	2/10

إذن من خلال الجدول (4) نجد:

$$E(S^2) = \sum S^2.P(S^2)......................................(9)$$

$$=2.25(\frac{3}{10}) + 5.25(4/10) + 9(1/10) + 9.75(2/10)$$

$$= 5.625$$

التحقق من أن: $E(S^2) = \frac{N}{N-1}\sigma^2$

من المعادلة (4) نجد: $\frac{N}{N-1}\sigma^2 = [\ 5/(5\text{-}1)\ ]4.5 = 5.625$

إذن نستنتج أن: $\mathbf{E(S^2) = \frac{N}{N-1}\sigma^2}$