المحاضرة الأولى:

1– مفاهيم أساسية:

المجتمع الإحصائي: يشير إلى كافة العناصر المكونة لمجموعة من الأشياء التي يراد إخضاعها لتجربة أو دراسة ما، وعليه المجتمع يمثل كافة القياسات، القيم، أو المفردات وليس الأفراد، أو الأشياء التي تم قياسها (الأوزان، آراء الناخبين، ....)، كما أن المجتمع قد يكون محدود أي يمكن حصر عدد مفرداته كعدد طلاب السنة الثانية قسم علوم التسيير، أو غير محدود أي من الصعب حصر عدد مفرداته كعدد النجوم في السماء مثلا ...، ونرمز عادة لحجم المجتمع بـ N.

العينة: هي مجموعة جزئية من المجتمع، أي أنها نموذج يشمل جزء من مفردات المجتمع الأصلي، ويرمز لحجم العينة بـ n، وعادة ما تكون العينة محدودة الحجم، وبالتالي تغنينها العينة عن دراسة كل مفردات المجتمع خاصة في حالة المجتمعات الغير محدودة والتي يصعب حصر مفرداتها.

المعلمة: يقصد بمعالم التوزيع مجموعة من الخصائص مثل: المتوسط μ ، الانحراف المعياري σ، ... الخ ومن خصائص المجتمع أيضا طبيعة توزيعه الاحتمالي كأن يكون طبيعيا أو غيره.

الإحصاءة: تشير إلى كل الخصائص المتعلقة بالعينة، مثل: متوسط ، تباين ، ... الخ.

المعاينة: هي عملية اختيار عدد كاف من عناصر المجتمع، بحيث يتمكن الباحث من خلال دراسة العينة المختارة وفهم خصائصها من تعميم هذه الخصائص على عناصر المجتمع الأصلي، ولابد أن نتذكر دوما بأن ناتج عملية المعاينة هو العينة المرغوب بها.

وتختلف أنواع العينات باختلاف طريقة المعاينة (طريقة سحبها)، ووفقا لدرجة دقتها وتمثيلها للمجتمع الأصلي، وعموما نميز بين العينات الاحتمالية وغير الاحتمالية.

وتعتبر المعاينة الاحتمالية من أكثر العينات استعمالا وتمتاز بكونها ممثلة للمجتمع الإحصائي بشكل جيد، وهذا النوع ينقسم بدوره إلى معاينة مقيدة ومعاينة غير مقيدة، والتي لا تستند إلى أي قيد أو شرط وتسمح المعاينة الغير مقيدة بالحصول على عينة عشوائية بسيطة.

العينة العشوائية البسيطة: نقول عن عينة أنها عشوائية إذا كان لكل مفردة في المجتمع نفس الاحتمال لأن تكون في العينة، وهذه العينة من أكثر العينات استعمالا وتخص المجتمعات المتجانسة (قد يصعب تحقيق ذلك في الواقع)، أي تلك التي لها خصائص متشابهة أو متقاربة، مثلا: مجتمع الطلبة متجانس من حيث السن، ويمكن الحصول

على عينة عشوائية عن طريق القرعة مثلا، ونميز بين نوعين من المعاينة العشوائية باختلاف طريقة السحب: معاينة نفاذية، ومعاينة غير نفاذية.

العينة الغير نفاذية: عندما يكون السحب بالإرجاع حيث يمكن أن تظهر المفردة أكثر من مرة في العينة، وتسمى هذه العينة غير نفاذية لأن تكرار العملية لا يؤدي إلى تقليص عدد المفردات في المجتمع، وفي هذه الحالة نستخدم القانون التالي لإيجاد عدد العينات الممكنة N^n.

العينة النفاذية: تسمى المعاينة بدون إرجاع معاينة نفاذية حيث يمكن أن تظهر المفردة في العينة أكثر من مرة واحدة، وفي هذه الحالة نستخدم القانون التالي لاستخراج العينات الممكنة:

$$C_N^n = \frac{N!}{(N-n)!\,n!}$$

مثال: ليكن لدينا المجتمع الإحصائي التالي: {1، 3، 5}، أحسب عدد العينات ذات الحجم n=2 التي يمكن استخراجها ثم حددها: 1- في حالة السحب بالإرجاع، 2- حالة السحب بدون إرجاع.

الحل:

1- استخراج كل العينات في حالة السحب بالإرجاع:

$N^n=3^2=9$، العينات الممكنة هي 9 عينات.

(1،1) -(3،3) -(5،5) – (1، 3) – (1، 5) – (3، 1) – (3، 5) – (5، 1) -(5، 3).

2- استخراج كل العينات في حالة السحب بدون بالإرجاع:

$$C_N^n = \frac{N!}{(N-n)!\,n!} \rightarrow C_3^2 = \frac{3!}{(3-2)!\,2!} = \frac{3\times 2\times 1}{1\times 2} = 3$$

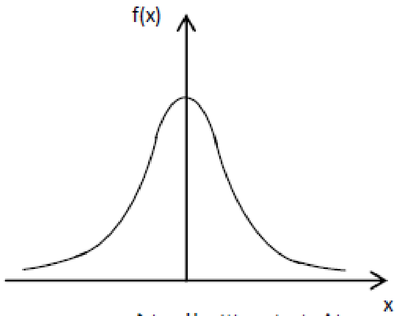
(1، 3) – (1، 5) – (3، 5)

كما وتنقسم المعاينة الغير مقيدة إلى: العينة العنقودية، العينة الطبقية، العينة المنتظمة العشوائية، وتنقسم المعاينة غير الاحتمالية إلى: العينة المسيرة، العينة القصدية، العينة الحصصية، كرة الثلج.

أهم التوزيعات الاحتمالية المستمرة:

هناك عدة توزيعات احتمالية إلا أننا سنعرض أهمها وهي: التوزيع الطبيعي، توزيع ستيودنت، توزيع مربع كاي وتوزيع فيشر.

1– التوزيع الطبيعي: يعتبر التوزيع الطبيعي من أهم التوزيعات المتصلة حيث يلعب دورا أساسيا في عملية المعاينة، كما يدرس سلوك المتغيرات العشوائية المتصلة مثل درجة الحرارة والطول والوزن والدخل والأخطاء العشوائية الناتجة عن تحليل الانحدار، ويمكن توضيح ذلك بالشكل العام للتوزيع الطبيعي.



الشكل العام للتوزيع الطبيعي

تكتب دالة كثافة لمنحنى التوزيع الطبيعي كما يلي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (-\infty < x < \infty)$$

حيث x متغير عشوائي حقيقي، e هو الأساس اللوغارتمي الطبيعي،

σ هو الانحراف المعياري وهو قيمة موجبة

ونقول أن x يتبع توزيع طبيعي على أساس القيمتين (μ, σ^2)، ونكتب: $x \to N(\mu, \sigma^2)$

التوزيع الطبيعي المعياري (القياسي):

هو توزيع طبيعي متصل وسطه الحسابي الصفر (0) وتباينه (1).

حيث أن إيجاد الدالة الأصلية للدالة $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2}$ صعب جدا ولذلك غالبا ما يتم الاعتماد على النشر المحدود لسلسلة تايلور في جوار نقط متتالية، ولأجل ذلك تم وضع جداول خاصة بقيم التوزيع الطبيعي وذلك بعد معرفة قيم المعلمتين μ و σ اللذان يختلفان ويتغيران من متغير عشوائي لآخر وهذا شبه مستحيل لأنه يجب علينا تكوين ما لا نهاية من الجداول حسب قيمهما، ولهذا اختيرت قيمتين ثابتتين هما: $\mu = 0$ و $\sigma = 1$

دالة كثافة الاحتمالية للتوزيع الطبيعي المعياري/

لدينا: $x \to N(\mu, \sigma^2)$ وبعد تحويله للتوزيع الطبيعي المعياري نرمز له بالرمز z، حيث $Z = \frac{x-\mu}{\sigma}$

ومنه دالة الكثافة الاحتمالية لمتغير عشوائي يتبع التوزيع الطبيعي المعياري تعطى كما يلي:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

وهناك جداول خاصة تحسب هذا التكامل وهو جدول التوزيع الطبيعي.

الصيغة المعتمدة في إيجاد قيمة الاحتمال:

- $P(z < a) = Q(a)$
- $P(z \geq a) = 1 - P(z < a)$
- $P(z < -a) = 1 - P(z < a)$
- $P(z > -a) = 1 - P(z < -a) = 1 - [1 - P(z < a)] = P(z < a)$
- $P(a < z < b) = P(z < b) - P(z < a)$.

مثال: أوجد $p(z \leq 2.5)$

نقوم باستخراج القيمة من جدول التوزيع الطبيعي المعياري كتالي:

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

$\Phi(2.5)$

إذن القيمة هي: $p(z \leq 2.5) = 0.9938$

مثال: إذا كان العمر الافتراضي لمشغل الأقراص المرنة والمنتج بمعرفة إحدى شركات الكمبيوتر يتبع التوزيع الطبيعي بمتوسط 1000 ساعة وانحراف معياري 100 ساعة، فإذا تم اختيار وحدة مشغل من إنتاج هذه الشركة بطريقة عشوائية، فأوجد احتمال أن يكون عمرها الافتراضي:

- بين 1000 و1150 ساعة.
- أقل من 930 ساعة.
- أكبر من 780 ساعة.

الحل:

لدينا: $X \to N(1000, 10000)$

احتمال أن يكون العمر الافتراضي لمشغل الأقراص المرنة بين 1000 و1150 ساعة.

$$P\,(1000 \le X \le 1150) = P\left(\frac{1000-1000}{100} \le \frac{X-\mu}{\sigma} \le \frac{1150-1000}{100}\right)$$
$$= P(0 \le Z \le 1.5) = Q(1.5) - Q(0) = 0.9332 - 0.5 = 0.4332$$

احتمال أن يكون العمر الافتراضي لمشغل الأقراص المرنة أقل من 930 ساعة.

$$P\,(X < 930) = P\left(\frac{X-\mu}{\sigma} < \frac{930-1000}{100}\right) = P\left(Z < \frac{-70}{100}\right)$$
$$= P(Z < -0.7) = 1 - Q(0.7) = 1 - 0.7580 = 0.2420$$

احتمال أن يكون العمر الافتراضي لمشغل الأقراص المرنة أكبر من 780 ساعة.

$$P\,(X > 780) = P\left(\frac{X-\mu}{\sigma} > \frac{780-1000}{100}\right) = P\left(Z > \frac{-220}{100}\right)$$
$$= P(Z > -2.2) = 1 - (1 - Q(2.2)) = 0.9861$$

توزيع مربع كاي:

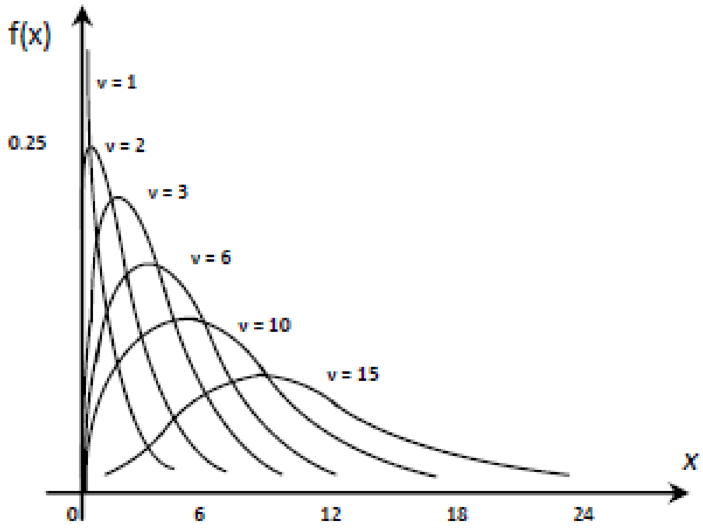
توزيع مربع كاي هو من أكثر التوزيعات استخداما في مجال اختبار الفروض بأنواعها، ويمكن تعريفه كما يلي:

لتكن $x_1, x_2, \ldots\ldots, x_v$ متغيرات عشوائية مستقلة كل منهما تتبع التوزيع الطبيعي المعياري $(\mu = 0, \sigma = 1)$

المتغيرة $x = x_1^2 + x_2^2 + \ldots + x_v^2$، ولها دالة الكثافة التالية:

$$\boxed{f(x) = \begin{cases} \dfrac{x^{(\nu/2)-1} e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} & x > 0 \\ 0 & x \le 0 \end{cases}}$$

ويمثل الرسم البياني التالي شكل توزيع مربع كاي



حيث يتضح من منحنى التوزيع أنه يقترب من التماثل كلما ازدادت درجة الحرية ونعبر عن المتغير العشوائي الذي يتبع توزيع مربع كاي بـ $X \sim \chi^2_v$ أو $X^2(\alpha,v)$

يمكن الاعتماد على جداول خاصة بهذا التوزيع في إيجاد قيم مربع كاي النظرية وعند مستوى معنوية ودرجة حرية معينة.

ولإيجاد قيمة الاحتمال المقابل لقيمة معينة وحسب جدول مربع كاي يجب أن تكون الصيغة كالآتي:

$P(X^2 > X^2_{n,\alpha}) = \alpha$، حيث أن α تمثل مستوى المعنوية (قيمة الاحتمال).

خصائص التوزيع:

- قيمة المتغير العشوائي هنا هي كمية موجبة دائما.
- الوسط الحسابي لقيم المتغير في هذا التوزيع هي (v).
- إن تباين قيم المتغير العشوائي في هذا التوزيع هو $(2v)$.

مثال1: أوجد القيم:

$X^2(0.95,7) = 2.167$

$X^2(0.1,17) = 24.769$

مثال 2: إذا كان المتغير العشوائي x^2 يتبع توزيع مربع كاي بدرجة حرية 10

- جد الوسط الحسابي والتباين، ثم اوجد $P(x^2 > 12.55)$، $P(x^2 < 20.48)$،

.$P(15.99 < x^2 < 29.59)$

الحل: من المعلومات المتوفرة: أي أن n=10، وحيث أن الوسط الحسابي يمثل حجم العينة أي أنه

μ= n=10 وبالتالي فإن التباين يساوي 2n وعليه فإن لتباين يساوي 20.

أما لقيم الاحتمالية فيمكن حسابها من خلال جداول توزيع مربع كاي عندما n=10.

$P(x^2 > 12.55) = 0.25$

$P(x^2 < 20.48) = 1- P(x^2 > 20.48) = 1-0.025 = 0.975$

$P(15.99 < x^2 < 29.59) = p(x^2 <29.59) – P(x^2 < 15.99)$

$= (1-P(x^2>29.59) – (1-P(x^2 > 15.99))$

$=p(x^2 > 15.99) – P (x^2 > 29.59)$

$= 1.10- 0.001 = 0.099$

توزيع ستيودنت:

ويعتبر من احد توزيعات المعاينة المهمة جدا والذي يستخدم عادة في الاختبارات الخاصة بحجوم العينات الصغيرة وأن هذا التوزيع هو بالأساس مشتق من حاصل قسمة متغيرين مستقلين، المتغير الأول الموجود في البسط وهو المتغير ذو توزيع طبيعي معياري، والمتغير الثاني الموجود في المقام ما هو إلا الجذر التربيعي الموجب لمتغير ذو توزيع مربع كاي مقسوما على درجة حريته، ويعرف المتغير العشوائي (t) بالشكل التالي:

$$t = \frac{z}{\sqrt{\frac{x^2}{n}}}$$

ودالة كثافته الاحتمالية معطاة بالصيغة التالية:

$$f(t) = c\left(1 + \frac{t^2}{v}\right)^{-v+\frac{1}{2}} \qquad (-\infty < t < \infty)$$

وهذا التوزيع يسمى توزيع t حيث تمثل درجات الحرية (v=n-1) وc ثابت يعتمد على قيمة v ليجعل المساحة تحت المنحنى تساوي 1.

ونقول: $x \rightarrow t_{(\alpha,n)}$ أو $x \rightarrow t_{\alpha(n)}$

يتم استخراج القيم من جدول توزيع ستيودنت بمستوى معنوية α ودرجة حرية n مع الأخذ بعين الاعتبار أن: -

$t_{(\alpha ,n)} = t_{(1-\alpha, n)}$

شروط نوزيع t:

يمكن تحديد الشروط الثلاثة لاستخدام توزيع t كما يلي:

1- أن يكون المجتمع المسحوبة منه العينة له توزيع طبيعي.

2- الانحراف المعياري (σ) (أو التباين) للمجتمع غير معروف (مجهول)

3- العينة تكون صغيرة الحجم (حجمها أقل تماما من 30 مشاهدة ($n<30$))

ويمكننا القراءة في جدول ستيودنت حيث: α عدد موجب.

$P(T<a) =1- p(T>a)$

$P(T<-a) = p(T>a)$

$P(T>-a) =1- p(T>a)$

$P(a<T<b) = P(T<b)- P(T<a)$

مثال:

1- إذا كان X متغير عشوائي يتبع توزيع t بدرجات حرية 15، أوجد كل من:

- $T_{(0.005, 15)}$
- $t_{(0.1, 15)}$
- α، حيث $t_{(\alpha, 15)}=1.615$

2- بالاعتماد عاة جدول ستيودنت أوجد ما يلي:

- $t_{(0.05,36)}$
- $t_{(0.975 ,36)}$

الحل:

1-

- بالبحث في جدول t عند تقاطع الصف 15 والعمود 0.005 فإننا نحصل على القيمة 2.947.

- بالبحث في جدول t عند تقاطع الصف 15 والعمود 0.1 فإننا نحصل على القيمة 1.341.
- بالبحث في جدول t عند تقاطع الصف 15 والعمود 0.00 نبحث على القيمة 1.615 أو أقرب قيمة لها لنجد 1.753 وبالاسقاط عموديا نجد α=0.05 .

2- بالرجوع إلى جدول ستيودنت نجد أن درجة الحرية 35 لا توجد في الجدول ولذا نأخذ القيمة التي تليها مباشرة (أكبر منها ومتوفرة في الجدول) أي v=40:

- بالبحث في جدول t عند تقاطع الصف 40 والعمود 0.05 فإننا نحصل على القيمة 1.684.
- نلاحظ أن القيمة 0.975 غير موجودة في الجدول وبالتالي نستخدم القاعدة:

$$-t_{(\alpha, n)} = t_{(1-\alpha, n)} \longrightarrow t_{(\alpha, n)} = -t_{(1-\alpha, n)}$$

$$\longrightarrow t_{(0.975, 40)} = -t_{(1-0.975, 40)} = -t_{(0.025, 40)} = -2.021$$

توزيع فيشر: F:

إن توزيع فيشر يعتبر أحد توزيعات المعاينة ويستند بالأساس على التوزيع الطبيعي وله استخدامات كثيرة في المجالات الإحصائية التطبيقية وبالذات في مجال تصميم وتحليل التجارب، فعلى افتراض أن هناك متغيرين الأول هو: $x_1^2 \sim x_{n1}^2$

والثاني هو: $x_2^2 \sim x_{n2}^2$

فإن F يعبر عنها بالشكل التالي: $f = \frac{x_1^2/n1}{x_2^2/n2} \sim f_{(n1,n2)}$

أي أن المتغير العشوائي (f) يسلك وفق توزيع F بدرجة حرية هي n1 و n2 للبسط والمقام على التوالي.

مثال: إذا علمت أن المتغير العشوائي (f) يسلك وفق توزيع F بدرجة حرية هي 4 و 6 للبسط والمقام على التوالي، أوجد قيمة $P(F > 4.53)$

الحل: إن قيمة الاحتمال تستخرج من الجدول الخاص بالتوزيع: $P(F > 4.53) = 0.05$