وزارة التعليم العالي والبحث العلمي

جامعة – البليدة 02 – لونيسي علي

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

قسم العلوم الاقتصادية

دروس عبر الخط في مقياس الإحصاء 03

موجهة لطلبة السنة الثانية ليسانس علوم اقتصادية

من إعداد :

د . غزغازي محمد

السنة الجامعية

2023/2022

محاور المقياس

المحور الأول : المتغيرات العشوائية وتوزيعاتها

المحور الثاني : توزيعات المعاينة

المحور الثالث : التقدير

المحور الرابع : اختبار الفرضيات

أولا: مفهوم المتغيرة العشوائية المتقطعة وتوزيعها الاحتمالي

1 مفهوم المتغيرة العشوائية

هي قيمة متغيرة يلحق بقيمها احتمالات تحقق كل قيمة. يرمز للمتغيرة ع بحرف لاتيني كبير. ونميز بين م ع المتقطعة وم العشوائبة المتصلة أو المستمرة.

مثال : في تجربة إلقاء مكعب نرد يمكن أن نسمي الوجه الذي يستقر عليه الكعب متغيرة عشوائية X.

القيم الممكنة لX هي: 1، 2، 3، 4، 5، 6. بكل قيمة يمكن أن نلحق احتمال تحققها، وهو هنا 1/6. ونكتب مثلا :

$$P(X = 1) = f(1) = 1/6,\ \ P(X = 2) = f(2) = 1/6 ,\ldots$$

لاحظ أن القيم الممكنة ل X (1، 2، 3، 4، 5، 6) هي متنافية، ولذلك فإن مجموع احتمالاتها يساوي 1. $\sum_{x=1}^{6} P(X = x) = 1$

مثال 2. في تجربة إلقاء قطعة نقدية مرتين يمكن أن نعين المتغيرة العشوائية X التي تمثل عدد مرات الحصول على كتابة. في هذه الحالة القيم الممكنة ل X هي 0، 1، 2. لا حظ أنه يمكن تعيين متغيرات عشوائية أخرى انطلاقا منن نفس التجربة، مثلا Y عدد مرات الحصول على صورة، وهي متغيرة تأخذ القيم 0، 1، 2، ثم المتغيرة Z بحث Z = X - Y ...

القيم الممكنة ل X هي 0، 2، -2. الاحتمالات الملحقة بقيمها يمكن حسابها كما يلي:

P(Z = 0) = P(X – Y = 0) = P(X = 0 et Y = 0 ou X = 1 et Y = 1 ou X = 2 et Y = 2) =>

P(Z = 0) = 0 + P(X = 1 et Y = 1) + 0 = 2 * 0.5^2 = **0.5**

2 المتغيرة العشوائية المتقطعة

و تسمى أيضا م ع منفصلة، وهي التي تأخذ عددا منتهيا من القيم الممكنة في مجال مغلق.

مثال: داخل المجال المغلق [2, 5] المتغيرة X المعرفة في المثال الأول تأخذ 4 قيم ممكنة.

3 التوزيع الاحتمالي للمتغيرة المتقطعة

هي مجموعة القيم الممكنة مع الاحتمالات المرتبطة بقيم المتغيرة. نرمز للمتغيرة بحرف كبير وللقيم التي تأخذها المتغيرة بحرف صغير. نعبر عن احتمال قيمة معينة كما يلي P(X = x) ونكتب أيضا : f(x). وتسمى الدالة f(x) دالة الكثافة الاحتمالية.

مثال: التوزيع الاحتمالي لم ع للمثال الأول (إلقاء مكعب نرد) يكتب كما يلي:

X	1	2	3	4	5	6	
P(X = x)	1/6	1/6	1/6	1/6	1/6	1/6	1

مثال 2. التوزيع الاحتمالي ل X، عدد مرات الصورة في رميتين لقطعة نقدية:

X	0	1	2	
P(X = x)	1/4	2/4	1/4	1

4 شروط دالة الكثافة للمتغيرة المتقطعة

نعبر عن احتمال قيمة معينة كما يلي $P(X = x)$ ونكتب أيضا : $f(x)$ وتسمى الدالة $f(x)$ دالة الكثافة الاحتمالية. لكي يمكن اعتبار دالة ما، أيا كانت، دالة كثافة احتمالية يجب أن يتحقق شرطان اثنان:

$$1)\qquad f(x) \geq 0$$

$$2)\qquad \sum_x f(x) = 1$$

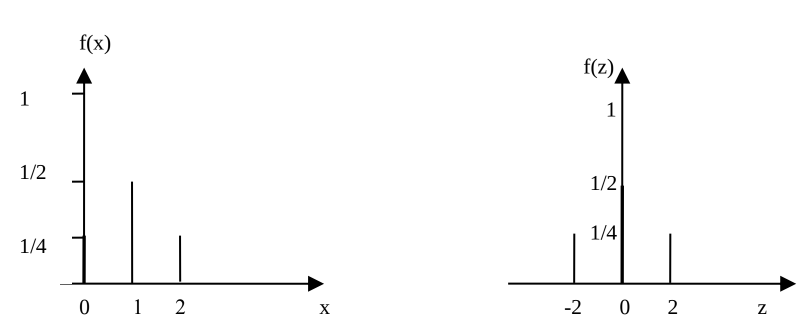
مثال: نأخذ دالة الكثافة ل X نتيجة لإلقاء حجر نرد: $f(1) = f(2) = f(3) = \ldots\ f(6) = 1/6 \geq 0$,

الشرط الأول محقق، والشرط الثاني أيضا لأن: $\Sigma f(x) = 1/6 + 1/6 + \ldots + 1/6 = 6(1/6) = 1$

5 التمثيل البياني لدالة الكثافة الاحتمالية ل م ع المتقطعة

تمثل المتغيرة العشوائية المتقطعة ليس من خلال منحنى ولكن من خلال أعمدة متوازية على محور X.

مثال: نمثل بيانيا منحنيات دوال الكثافة ل X و Z المعرفة على إلقاء قطعة نقدية مرتين.



رسم 1 التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المتقطعة

6 دالة التوزيع F(x) للمتغيرة العشوائية المتقطعة

تعرف دالة التوزيع – وتسمى أيضا "الدالة التجميعية" –كما يلي:

$$\mathbf{F(x) = P(X \leq x)}$$

ويمكن استنتاج دالة التوزيع من دالة الكثافة الاحتمالية f(x) كما يلي:

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u)$$

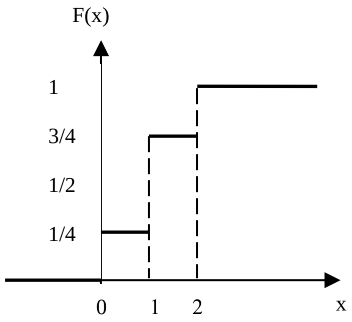
إذا كانت X تأخذ عددا منتهيا من القيم فإن F(x) يمكن تعريفها كما يلي:

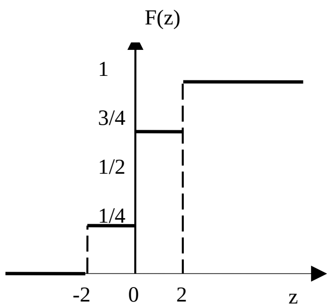
$$F(x) = \begin{cases} 0 & , \quad -\infty < x < x_1 \\ f(x_1) & , \quad x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & , \quad x_2 \leq x < x_3 \\ \cdots & \\ f(x_1) + f(x_2) + \ldots + f(x_n), & \quad x_n \leq x < +\infty \end{cases}$$

مثال: أوجد قيم F(x) وF(z) للأمثلة السابقة ومثلهما بيانيا.

X	0	1	2
f(x)	1/4	1/2	1/4
F(x)=P(X≤x)	1/4	3/4	1

Z	-2	0	2
f(x)	1/4	1/2	1/4
F(x)=P(X≤x)	1/4	3/4	1





رسم 2 التمثيل البياني لدالة التوزيع للمتغيرة العشوائية المتقطعة

ملاحظة. تأخذ دالة التوزيع للم ع المتقطعة شكلا سلميا، وهي لا تكون متناقصة في أي مجال، وأكبر قيمة ممكنة لها هي 1.

ثانيا: مفهوم المتغيرة العشوائية المستمرة وتوزيعها الاحتمالي

1 تعريف المتغيرة العشوائية المستمرة

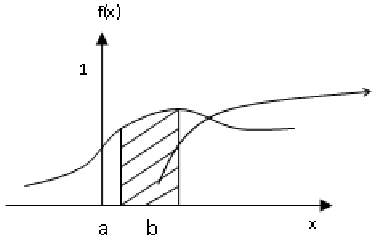
هي متغيرة ع تأخذ عددا لا متناهيا من القيم في مجال محدود، أو هي تأخذ أي قيمة داخل هذا المجال. من أجل هذا فأن وحدات قياس المتغيرة الميتمرة تكون مستمرة مثل الزمن، الوزن، المسافة، الحجم، ...

2 التوزيع الاحتمالي المستمر

هو مجموعة القيم التي يمكن أن تأخذها المتغيرة ع المستمرة والاحتمالات الملحقة بها. نسمي توزيعا كهذا دالة الكثافة الاحتمالية أو دالة الاحتمال، وهي ممثلة بمنحنى متصل.

لاحظ أنه بما أن X تأخذ عددا لا متناهيا من القيم داخل أي مجال، فإن احتمال قيمة بعينها هو احتمال يؤول إلى الصفر. $P(X=x) \rightarrow 0$ لذلك فإن دالة الكثافة تستعمل لحساب احتمال مجال. ويكون ذلك بحساب المساحة تحت منحنى f(x) بين حدود المجال:

$$\boxed{P(a<x<b)=\int_a^b f(x)dx}$$



رسم3 التمثيل البياني لدالة الكثافة للمتغيرة العشوائية المستمرة

لاحظ أن إشارة التكامل هنا تقابل إشارة المجموع في حالة المتغيرة ع المتقطعة.

3 خصائص دالة الكثافة الاحتمالية للمتغيرة العشوائية المستمرة

باستبدال إشارة التكامل بإشارة المجموع نجد أن شروط دالة الكثافة الاحتمالية للم ع المستمرة تكتب كما يلي :

$$1)\ f(x) \geq 0$$

$$2)\int_{-\infty}^{+\infty} f(x)dx = 1$$

من البديهي إذا أن منحنى دالة الكثافة لا يمكن أن ينزل أسفل محور الم ع، كما أن المساحة الإجمالية بين المنحنى والمحور الأفقي تساوي الواحد.

هذه الخصائص تفيدنا في حساب احتمالات بعض المجالات من خلال احتمالات مجالات أخرى.

$$\boxed{P(a<x\leq b)=P(X\leq b)-P(X\leq a)=F(b)-F(a)}$$

مثال: أوجد قيمة الثابت C التي تحقق الشرطين الأول والثاني لدالة الكثافة الاحتمالية في الدالة التالية:

- ✓ أحسب احتمال أن تكون X تنتمي للمجال من 1 إلى 2.
- ✓ أحسب احتمال أن تكون X لا تنتمي للمجال من 1 إلى 2.

$$f(x)=\begin{cases} cx^2 & 0<x<3 \\ 0 & \text{sinon} \end{cases}$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow \int_{-\infty}^{0} 0dx + \int_{0}^{3} Cx^2dx + \int_{3}^{+\infty} 0dx = 1 \Rightarrow C\left[\frac{x^3}{3}\right]_0^3 = 9C = 1 \Rightarrow C = 1/9$$

لكي تكون x دالة كثافة يجب أن يكون C = 1/9 .

$$P(1<x\le 2)=\int_1^2 f(x)dx=\int_1^2 (1/9)x^2dx=\frac{1}{9}\left[\frac{x^3}{3}\right]_1^2=\frac{1}{9}\left[\frac{8-1}{3}\right]=\frac{7}{27}$$

$$P(1>x>2)=1-P(1<x<2)=1-\frac{7}{27}=\frac{20}{27}$$

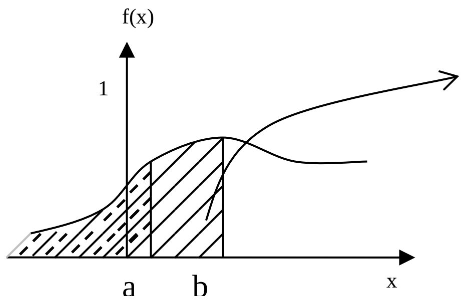
4 دالة التوزيع F(x) للمتغيرة العشوائية المستمرة

تعرف دالة التوزيع للمتغيرة المستمرة كما يلي:

$$\boxed{F(x)=P(X\le x)=\int_{-\infty}^{x} f(u)du}$$

لدالة التوزيع أهمية أكبر بالنسبة للمتغيرة المستمرة. السبب في ذلك أننا نهتم، في حالة المتغيرة المستمرة، باحتمال مجال وليس باحتمال نقطة، ولحساب احتمال مجال من الأيسر التعويض في دالة التوزيع بدلا من حساب التكامل في كل مرة. يتضح ذلك من القاعدة التالية: بفرض a, b نقطتان من مجال تعريف X، بحيث b > a . لحساب احتمال أن تكون X تنتمي إلى المجال]a, b] :

$$\boxed{P(a<x\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)}$$



مثال:

أوجد دالة التوزيع للمتغيرة المذكورة في المثال السابق.

استخدم دالة التوزيع لحساب الاحتمال: P(1< x <2).

رسم 4 حساب الاحتمال من خلال دالة التوزيع

$$*\, x<0: \qquad F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{0} 0du=0$$

$$*\, 0\le x<3: \qquad F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{0} 0du+\int_0^x \frac{1}{9}u^2du=\frac{1}{9}\left[\frac{u^3}{3}\right]_0^x=\frac{x^3}{27}$$

$$*\, x\ge 3: \qquad F(x)=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{0} 0du+\int_0^3 \frac{1}{9}u^2du+\int_3^x 0du=0+\frac{1}{9}\left[\frac{u^3}{3}\right]_0^3+0=\frac{27}{27}=1$$

$$F(x)=\begin{cases}0 & x<0\\ x^3/27 & 0\le x<3\\ 1 & x\ge 3\end{cases}\qquad P(1<x\le 2)=F(2)-F(1)=\frac{2^3}{27}-\frac{1^3}{27}=\frac{7}{27}$$

5 قاعدة لايبنيز Règle de LEIBNITZ

تفيد هذه القاعدة الرياضية العامة في استنتاج أن مشتقة دالة التوزيع هي دالة الكثافة:

$$\boxed{\frac{d\int_{-\infty}^{x} f(u)du}{dx}=f(x)\Rightarrow\frac{dF(x)}{dx}=f(x)}$$

مثال: أوجد دالة الكثافة للمتغيرة X إذا كانت دالة التوزيع كما يلي:

$$F(x)=\begin{cases}1-e^{-2x} & x\ge 0\\ 0 & \text{sinon}\end{cases}$$

$$*\,x<0: f(x)=F'(x)=(0)'=0$$

$$*\,x\ge 0: F(x)=1-e^{-2x}\Rightarrow f(x)=\left(1-e^{-2x}\right)'=2e^{-2x}$$

$$f(x)=\begin{cases}2e^{-2x} & x>0\\ 0 & \text{sinon}\end{cases}$$

6 خلاصة المبحث الأول و الثاني

يتم تعريف التوزيع الاحتمالي (أو القانون الاحتمالي) لمتغيرة عشوائية من خلال تحديد القيم الممكنة للمتغيرة و الاحتمالات المقابلة لها.

يتم هذا التحديد إما من خلال جدول (جدول التوزيع الاحتمالي) أو دالة، تسمى دالة كثافة الاحتمالية.

لكي نقول عن دالة ما أنها دالة كثافة احتمالية يجب أن تكون موجبة دوما و أن يكون مجموع الاحتمالات مساويا للواحد.

الدالة التجميعية (أو دالة التوزيع) تمثل احتمال مجال من أصغر قيمة للمتغيرة إلى نقطة ما:

$F(x)=P(X\le x)=\sum_{u\le x} f(u)$ في حالة م متقطعة و $F(x)=P(X\le x)=\int_{-\infty}^{x} f(u)du$ في حالة م مستمرة.

نظرا لتعريفها تأخذ دالة التوزيع مســــارا متزايدا (أو ثابتا على أجزاء من المجال). تبرز أهمية الدالة التجميعية أكثر عندما تكون المتغيرة مســـتمرة لأننا نهتم حينها باحتمالات مجالات. يمكن اســـتنتاج دالة الكثافة من خلال اشتقاق دالة التوزيع.

التوقع الرياضي والتباين

في العديد من الحالات لا يكفي حســـاب احتمال تحقق حدث أو أحداث معينة بل نحتاج للخروج بتوقع معين يلخص الوضــعية المطروحة أمامنا. من جهة أخرى قد يصــعب المفاضــلة بين خيارات متاحة مقيمة بمبالغ معينة بسبب ارتباط كل مبلغ بمخاطرة مختلفة؛ من المعروف أن الاستثمارات الأكثر مردودية هي تلك التي تتضـمن أكبر مخاطرة، فكيف يمكن أخذ في الحسبان المخاطرة والمبلغ المتوقع وبطريقة دقيقة وموضوعية؛ إن طريقة التوقع وبقية المفاهيم الأخرى الواردة أعلاه يمكن أن تساعدنا في ذلك.

ثالثا: التوقع الرياضي Espérance mathématique

1 تعريف التوقع

يعرف التوقع الرياضي لمتغيرة عشوائية متقطعة كما يلي:

$$\boxed{E(X)=\sum_{i=1}^{n}x_iP(X=x_i)=\sum_{i=1}^{n}x_i f(x_i)}$$

و يعرف التوقع الرياضي لمتغيرة عشوائية مستمرة كما يلي:

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx$$

نرمز للتوقع أحيانا ب μ أو μ_X.

مثال: نلقي قطعة نقدية 4 مرات. أحسب العدد المتوقع من المرات التي نحصل فيها على وجه.

X عدد مرات وجه	0	1	2	3	4	المجموع
P(X)	$(1/2)^4$	4 $(1/2)^4$	6 $(1/2)^4$	4 $(1/2)^4$	$(1/2)^4$	16/16 = 1
XP(X)	0	4 $(1/2)^4$	12 $(1/2)^4$	12 $(1/2)^4$	4 $(1/2)^4$	32/16 = 2

$$E(X)=\sum_{i=1}^{n}x_iP(X=x_i)=32/16=2$$

العدد المتوقع هو مرتين من بين 4 رميات.

مثال 2. نلقي قطعة نقدية مرة واحدة. يربح اللاعب 20 دج إذا حصل على الرقم 2، ويربح 40 دج إذا حصل على الرقم 4، و60 دج إذا حصل على الرقم 6، ويخسر 10 دج إذا حصل على الرقم 1، 30 دج إذا حصل على الرقم 3، و50 دج إذا حصل على الرقم 5. تحقق مما إذا كانت العبة متوازنة (هل توقع الربح يساوي توقع الخسارة).

الجواب هو أن اللعبة غير متوازنة لأن توقع الربح أكبر من توقع الخسارة $E(x) = 30/6 = 5 > 0$

نتيجة الرمي	1	2	3	4	5	6	المجموع
نتيجة المراهنة X	-10	20	-30	40	-50	60	-
P(X=x)	1/6	1/6	1/6	1/6	1/6	1/6	6/6
X*P(X=x)	-10/6	20/6	-30/6	40/6	-50/6	60/6	(120-90)/6>0

مثال3. أوجد التوقع الرياضي للمتغيرة ذات دالة الكثافة التالية:

$$f(x)=\begin{cases}\frac{1}{2}x & 0<x<2\\ 0 & \text{sinon}\end{cases}$$

$$E(x)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{0}x.0.dx+\int_{0}^{2}x(\frac{1}{2}x)dx+\int_{2}^{\infty}x.0.dx$$

$$E(x)=0+\frac{1}{2}\left[\frac{x^3}{3}\right]_0^2+0=\frac{4}{3}$$

2 توقع دالة

يستخدم توقع دالة عند حساب عدد من المقاييس مثل التباين، العزوم المركزية والعزوم المرتبطة بالأصل.

لتكن X م ع لها دالة كثافة f(x)، وy = g(x) م ع تابعة لها.

$$E(Y)=E(g(x))=\sum_x g(x)f(x)$$

في حالة X م متصلة:

$$E(Y)=E(g(x))=\int_{-\infty}^{+\infty}g(x)f(x)dx$$

مثال1. نلقي قطعة نقدية مرتين. نسمي X عدد مرات الحصول على صورة، و$Y=X^2$. أحسب E(X) وE(Y) .

X	0	1	2	المجموع
P(X = x)	1/4	1/2	1/4	1
X*P(X)	0	1/2	1/2	E(X) =1
X^2	0	1	4	
X^2*P(X)	0	1/2	1	$E(X^2)$ = 3/2

مثال2: لتكن X م ع ذات دالة الكثافة التالية، و$Y = g(x) = 3x^2 - 2x$. أحسب E(Y).

$$f(x)=\begin{cases}\frac{1}{2}x & 0<x<2\\ 0 & \text{sinon}\end{cases}$$

$$E(Y)=E(g(x))=\int_{-\infty}^{+\infty}g(x)f(x)dx=\int_{-\infty}^{+\infty}(3x^2-2x)(0)dx+\int_{-\infty}^{+\infty}(3x^2-2x)(x/2)dx+\int_{-\infty}^{+\infty}(3x^2-2x)(0)dx$$

$$E(Y)=0+\int_{0}^{2}(3x^2-2x)(x/2)dx+0=\frac{1}{2}\left[\frac{3x^4}{4}-\frac{2x^3}{3}\right]_0^2=\frac{1}{2}\left[12-\frac{16}{3}\right]=\frac{20}{6}=\frac{10}{3}$$

3 خصائص التوقع الرياضي

E (C) = C
E(CX) = CE(X)
E(X+Y) = E(X) + E(Y)
E(XY) = إذا كانت المتغيرتان مستقلتان.
E(X)E(Y)

مثال: تتوقع مؤسسة أن تتلقى كل شهر 3 طلبيات من العميل A و4 من B .

- أحسب العدد المتوقع من الطلبيات المتلقاة من A في السنة.
- أحسب العدد الإجمالي المتوقع من الطلبيات المتلقاة في شهر.

يعين كل عميل من طرفه مندوبا عن كل طلبية لمتابعة إتمامها. كم تتوقع أن يلزم من مقابلة لتعريف مندوبي العميل A بمندوبي B.

E(12A) = 12E(A) = 12(3) = 36
E(A*B) = E(A)*E(B) = 4*3 = 12.
E(A + B) = E(A) + E(B) = 3 + 4 = 7

رابعا: التباين والانحراف المعياري Variance et écart type

1 تعريف التباين

يعرف التباين لمتغيرة عشوائية كما يلي:

$$\boxed{\sigma_X{}^2=Var(X)=E\left[(X-\mu)^2\right]}$$

و الانحراف المعياري هو جذر التباين. $\sigma=\sqrt{V(X)}=\sqrt{\sigma^2}$.

في حالة المتغيرة العشوائية المتقطعة : $V(X)=\sum_x (x-\mu)^2 f(x)$

في حالة المتغيرة العشوائية مستمرة : $V(X)=\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx$

مثال. نلقي قطعة نقدية مرتين. نسمي X عدد مرات الحصول على صورة، أحسب V(X) .

X	0	1	2	المجموع
P(X = x)	1/4	1/2	1/4	1
X*P(X)	0	1/2	1/2	E(X) = μ =1

$(X-\mu)^2$	1	0	1	
$(X-\mu)^2$ * P(X)	1/4	0	1/4	V(X) = 1/2

مثال 2. لتكن X م ع ذات دالة الكثافة التالية؛ أحسب تباين X.

$$f(x)=\begin{cases}\frac{1}{2}x & 0<x<2\\ 0 & \text{sinon}\end{cases} \qquad \sigma^2=\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx. \qquad \mu=4/3$$

$$\sigma^2=\int_{-\infty}^{0}\left(x-\frac{4}{3}\right)^2 *0dx+\int_{0}^{2}\left(x-\frac{4}{3}\right)^2\frac{x}{2}dx+\int_{2}^{+\infty}\left(x-\frac{4}{3}\right)^2 0dx=0+\frac{1}{2}x\left(x^2-\frac{8x}{3}+\frac{16}{9}\right)_0^2+0$$

$$\sigma^2=\frac{1}{2}\left(x^3-\frac{8x^2}{3}+\frac{16x}{9}\right)_0^2=\frac{1}{2}\left(8-\frac{32}{3}+\frac{32}{9}\right)=\frac{4}{9}$$

2 خصائص التباين

$V(X) = E(X^2) - E(X)^2$

$V(CX) = C^2V(X)$, $V(C) = 0$

في حالة استقلال المتغيرتان عن بعضهما

$V(X - Y) = V(X) + V(Y)$, $V(X + Y) = V(X) + V(Y)$

مثال: نلقي قطعة نقدية مرتين. نسمي X عدد مرات الحصول على صورة، أحسب V(X) باستخدام الصيغة

. $V(X) = E(X^2) - E(X)^2$

لتكن المتغيرة $Y = 2X$. أحسب V(Y) ، نلقي حجر نرد ونسمي Z النتيجة المحصل عيها. أحسب تباين المتغيرة W حيث: $W = Z - Y$

X	0	1	2	المجموع
P(X = x)	1/4	1/2	1/4	1
X P(X)	0	1/2	1/2	$E(X) = \mu = 1$
$(X)^2$	0	1	4	
$(X)^2$ * P(X)	0	1/2	1	$E(X)^2 = 3/2$

$V(X) = E(X^2) - E(X)^2 = (3/2) - 1 = 1/2$

$V(Y) = V(2X) = 2^2V(X) = 4 (1/2) = 2.$

$V(W) = V(Z-2X) = V(Z) +V(2X) = V(Z) + 2^2V(X).$

$V(Z) = E(Z^2) - E(Z)^2 = (1/6)[1^2+2^2+3^2+4^2+5^2+6^2]-$
$(1/6)[1+2+3+4+5+6] = 70/6$
$V(W) = (70/6) + 4\ (1/2) = 82/6 = 13.67$

3 المتغيرة المعيارية Variable centrée réduite

يمكن أن نلحق بأي متغيرة عشوائية X متغيرة معيارية (تسمى أيضا المتغيرة المركزية) ويرمز لها X* . تلحق المتغيرة المعيارية بالمتغيرة الحقيقية من أجل المقارنة لأن المتغيرة المعيارية ليس لها وحدة كالمتر أو الساعة ... وإنما هي تعبر عن كل قيمة x ل X من خلال المسافة بين x والتوقع μ محسوبة لس بالوحدة الأصلية وإنما بالانحرافات المعيارية.

$$\boxed{X^* = \frac{X-\mu}{\sigma}}$$

من خلال خصائص التوقع والتباين نستخرج التوقع والتباين للمتغيرة المعيارية.

$$E(X^*) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}\left[E(X)-E(\mu)\right] = 0$$

$$V(X^*) = E\left[(X^*-E(X^*))^2\right] = E\left[(X^*-0)^2\right] = E(X^{*2}) = E\left(\frac{(X-\mu)^2}{\sigma^2}\right) = \frac{1}{\sigma^2}E(X-\mu)^2 = \frac{\sigma^2}{\sigma^2} = 1$$

مثال: أحسب X* من أجل متوسط 70، انحراف معياري 5، وX يساوي: 55، 60، 50، 75، 80، 70.

الجواب: القيم هي: -3، -2، -4، 1، 2، 0.

مثال 2. يتدرب عاملان أحمد وعلي من أجل المشاركة في ماراتون عيد العمال 1 ماي. يشترط يوم المسابقة أن يكون وزن المترشح لا يتجاوز المجال $\mu \pm 1.5\sigma$.

إذا كان الوزن المتوسط بالكغ هو $\mu = 70$ والانحراف المعياري هو 5 كغ. هل سيقبل العاملان أحمد وعلي إذا كان وزنهما: 77 كغ، و80 كغ؟

الجواب: مجال القبول هو من 62.5 كغ إلى 77.5 كغ، لذلك فسيرفض علي ويقبل أحمد.

4 خلاصة

التوقع الرياضي و التباين هي أهم المؤشرات المعبرة عن خصائص المتغيرة و يحسبان كما يلي، حسب كون المتغيرة متقطعة أو مستمرة:

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

لكل من التوقع و التباين 4 خصائص أساسية تتمثل فيما يلي:

$E(C) = C$	توقع عدد ثابت	$V(C) = 0$

$E(CX) = CE(X)$	$V(CX) = C^2V(X)$
$E(X+Y) = E(X) + E(Y)$	في حالة استقلال المتغيرتان عن بعضهما $V(X + Y) = V(X) + V(Y)$ $V(X - Y) = V(X) + V(Y)$,
$E(XY) = E(X)E(Y)$ في حالة استقلال المتغيرتان عن بعضهما.	$V(X) = E(X^2) - E(X)^2$

من أجل المقارنة بين المتغيرات تستخدم المتغيرة المعيارية التي تسمح بالتعبير عن قيمة X ليس من خلال وحداتها الأصلية (كغ، متر، زمن، ...) وإنما بعدد الانحرافات المعيارية التي تفصل بين القيمة X والتوقع الرياضي.

التوقع الرياضي والتباين للمتغيرة المعيارية $X^* = \frac{X-\mu}{}$ هما على التوالي 0 و 1.

لحساب التوقع الرياضي لدالة ما في X (مثلا التباين، أو X^2) نضرب قيم الدالة في الاحتمالات المقابلة لـX:

$$E(Y) = E(g(x)) = \sum_x g(x) f(x)$$

$$E(Y) = E(g(x)) = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

خامسا: التوزيعات الاحتمالية الشائعة المستمرة

1 التوزيع الطبيعي أو توزيع لابلاس قوس[1] - D. Normale ou D. de Laplace Gausse

يعد التوزيع الطبيعي من أهم التوزيعات الاحتمالية شائعة الاستخدام لما له من خصائص تنطبق على نسبة كبيرة من الظواهر الطبيعية والاجتماعية والاقتصادية. فلو اخترنا بالصدفة مئة أو ألفا من المارين في شار ع ما وقسنا أطوالهم لوجدنا نسبة كبيرة منها قريبة من متوسط ما، ونسبة قليلة من طوال القامة ونسبة مقاربة لها من قصار القامة. ومثل هذا بالنسبة للأوزان. ولو مثلنا هذه البيانات في معلم متعامد متجانس لكان

[1] باسم العالمان الرياضيان الفيزيائيان و الفلكيان الفرنسي Pière Simon de Laplace (1749-1827) والألماني Carl Freidrich Gauss – الصورة لهذا الأخير–، (1777-1855) الذين كانا من أوائل من اكتشف هذا القانون. أما من أعطاه تسمية التوزيع الطبيعي فهو Pearson في 1893. أنظر J. J. Droesbeke (1997)، ص329 .

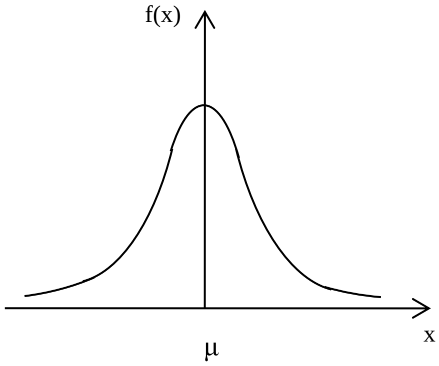
المنحنى الذي يمثل النسبة، أو ما يمكن أن نسميه الاحتمال، ذا شكل جرسي متماثل حول المتوسط وهي صفات التوزيع الطبيعي (الشكل9) :

(أ) صيغة القانون

تكتب دالة الكثافة لمنحنى للتوزيع الطبيعي كما يلي:

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

حيث μ و σ هما على التوالي التوقع والانحراف المعياري. ونكتب $X \sim N(\mu, \sigma)$



الشكل العام للتوزيع الطبيعي

دالة التوزيع (الدالة التجميعية) للتوزيع الطبيعي تكتب كما يلي:

$$F(x)=P(X\le x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-1/2\left(\frac{v-\mu}{\sigma}\right)^2}dv$$

المتغيرة المركزية أو المعيارية : تستخدم المتغيرة المعيارية $Z = (X-\mu)/\sigma$ لتكوين الجداول الإحصائية للاحتمالات:

$F(z) = P(Z \le z)$ أو $P(0 \le Z \le z)$،

حيث تسمح بكتابة الدالة f وF بدلالة مجهول واحد Z بدلا من 3 مجاهيل x وμ و σ وذلك كما يلي:

$$\boxed{f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\qquad -\infty<z<\infty}$$

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du$$

بالنظر إلى العلاقة الخطية بين المتغيرتين X و Z، فإن Z تتبع نفس توزيع X أي التوزيع الطبيعي. ونعلم أن:

$$V(Z) = 1 \quad E(Z) = 0$$

(ب) خصائص التوزيع الطبيعي

الدالة المتجددة للعزوم : $M_x(t) = \exp\left(t\mu + \frac{t^2\mu^2}{2}\right)$

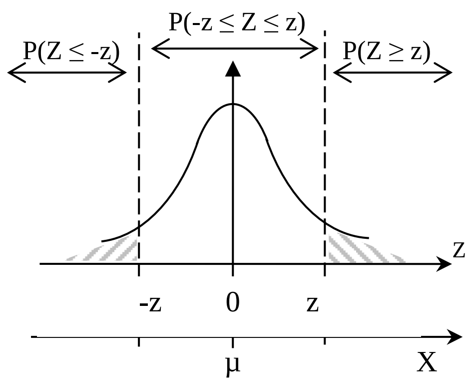
من خصائص التوزيع الطبيعي أنه يعتبر معتدلا لا مدببا ولا مفلطحا، حيث يعتبر معامل التفلطح $\alpha_4 = 3$ للتوزيع الطبيعي معيارا لاعتدال المنحنيات.

من خصائص التوزيع الطبيعي أيضا أنه متماثل $\alpha_3 = \frac{\mu_3}{\sigma^3} = 0$ حول القيمة المتوقعة

تماثل منحنى X حول المتوسط (أنظر الشكل 3) يعني تماثل لمنحنى Z حول 0، مما يعني أنه من أجل أي قيمة للمتغيرة المعيارية $z > 0$:

$$P(0 \le Z \le z) = P(-z \le Z \le 0) = P(-z \le Z \le z) / 2$$

$$P(Z \le -z) = 1 - P(Z \le z) = P(Z \ge z)$$



استخدام تماثل الوزيع الطبيعي في حساب الاحتمالات

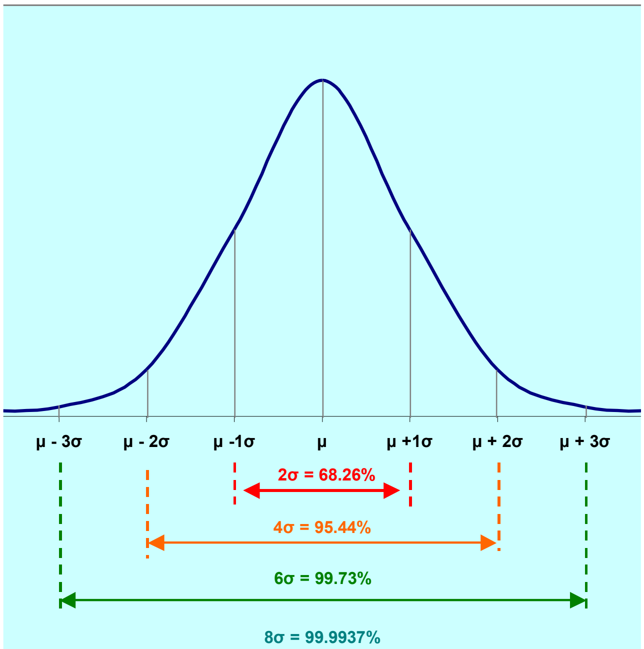
و لقد تم باستخدام المتغيرة المعيارية Z حساب الاحتمالات (المساحات) تحت المنحنى ومنها خاصة:

$$P(-\sigma \le X \le \sigma) = P(-\sigma \le Z \le \sigma) = 0.6837,$$

$$P(-2\sigma \le X \le 2\sigma) = P(-2 \le Z \le 2) = 0.9544,$$

$$P(-3\sigma \le X \le 3\sigma) = P(-3 \le Z \le 3) = 0.9973.$$

هذه القيم وغيرها متوفرة في الجداول الإحصـــائية التي نجدها في الكثير من المراجع، كما يمكن حســـابها باستخدام الحاسوب.



المساحات الأساسية تحت منحنى التوزيع الطبيعي

مثال: باستعمال الجداول الاحصائية 1) أحسب : $P(0 \leq Z \leq z)$ حيث $z = 1, 2, 3$

2) أحسب $P(-z \leq Z \leq z)$ من أجل نفس القيم ل z.

1) 0.3413 ، 0.47725 ، 0.49865

2) 0.6827، 0.9545، 0.9973

مفهوم العينات :تعرف العينة بأنها ذلك الجزء من المجتمع الذي يجري اختيارها على وفق طرائق وقواعد علمية بحيث تمثل المجتمع تمثيلا صحيحا

مميزات استخدام العينات في البحوث :

-العينات تكتفي بعدد محدود من المفردات وليس جميعها، وذلك اقتصادا في الجهد والنفقات.

-أنها سريعة في إعطاء نتائج البحوث مقارنة بأسلوب الحصر الشامل .

-تتيح للباحث التعميق في مصادر الأحكام واتخاذ القرارات.

-تستخدم لأنها أقل عرضة للأخطاء مقارنة بالأساليب الأخرى.

-يعد استخدامها (العينات) من الوسائل المعنية بإثراء البحوث العلمية الأصلية.

-أنها طريقة مناسبة، حيث إمكانية تحديد مدى الثقة في نتائجها، وكذا نسبة تمثيلها للمجتمع.

عيوب استخدام العينة (أخطاء المعاينة) :

-اخذ عينة من مصدر خاطئ، كأن تستخدم دليل الهاتف للحصول على عينة تمثل الرأي العام .

-التحيز الشخصي، ويحدث ذلك حينما يأخذ الباحث عينته المختارة من فئة معينة لها خصائص مميزة عن المجتمع الكلي.

-جمع بيانات ناقصة، فمثلا إهمال العامل الجغرافي عند دراسة المستوى الاقتصادي للسكان بتقسيم الأسر المبحوثة حسب دخولها.

-خطأ الصدفة، يزداد احتمال ورود هذا الخطأ كلما صغر حجم العينة.

أنواع العينات:

-**العينات غير الاحتمالية** :وهي تلك العينات التي يتم اختيارها بطريقة غير عشوائية، أي التي لا تعتمد على نظرية الاحتمالات ومن عيوبها أنها لا تمثل مجتمع البحث تمثيلا دقيقا، ومن ثم فإن نتائجها لا تصلح للتعميم على المجتمع كله، ومن أمثلة هذا النوع من العينات أن يختار الباحث عينة يرى أنها تمثل المجتمع الأصلي الذي يقوم بدراسته تمثيلا صادقا.

العينات الاحتمالية:

-العينة العشوائية البسيطة: هي العينة التي تختار وحدتها من الإطار الخاص بها، على أساس يهيئ فرص انتقاء متكافئة لجميع وحدات المجتمع المسحوبة منها.

-العينة العشوائية الطبقية: في هذه الحالة ينبغي تقسيم المجتمع إلى أقسام أو طبقات مختلفة ثم يأخذ من كل قسم أو طبقة عينة متجانسة بطريقة عشوائية، على أن يكون حجم كل طبقة في العينة متناسبة مع حجم الطبقة المناظرة لها في المجتمع الأصلي.

-العينة العشوائية المنتظمة: يتم اختيار وحداتها بحيث تكون المسافة أو المدة بين كل وحدة وأخرى ثابتة لجميع وحدات العينة.

-العينة العشوائية العنقودية: وهي عينة تختار عن طريق استخدام تجمعات (عناقيد) تختار من المجتمع الأصلي بدلا من انتقاء المفردات بصفة مباشرة من هذا المجتمع .

نظرية المعاينة Distribution d'échantillonnage :

تنتشر في مجتمعاتنا المعاصرة عمليات الاستقصاء، ففي عالم الأعمال تقوم المؤسسات عن طريق مصالح التسويق ومصالح البحوث والتطوير بإجراء استقصاءات للإطلاع على توجهات المستهلكين، وفي وسائل الإعلام لا يمر يوم دون أن يعلن عن نتائج استقصاء أجرته مجلة أو جامعة حول مواضيع سياسية أو اجتماعية متعددة، منها الاستقصاءات المثيرة للجدل حول الآراء السياسية للمواطنين أثناء الحملات الانتخابية. فما هي الأسس النظرية

الرياضية التي تستند عليها الاستقصاءات المختلفة ؟ أو كيف يمكن الاستدلال من خلال بيانات عينة على خصائص المجتمع الذي أخذت منه؟ الإجابة على هذه الأسئلة و غيرها تتطلب فهم العلاقات الرياضية بين الخصائص المختلفة للمجتمع مثل المتوسط، التباين وغيرها، والخصائص المناظرة لها في العينة وهو ما سنتناوله في هذا الفصل .

- **مفهوم المعاينة <u>L'échantillonnage</u>**:المعاينة هي عملية الحصول على عينة من المجتمع الإحصائي، فلو أردنا سحب عينة من المجتمع حجمها n وكان حجم المجتمع N وكان السحب مع الإرجاع فعدد الطرق الممكنة لسحب العينة هو N^n ، أما لو كان السحب بدون إرجاع فعدد الطرق الممكنة هو C_N^n حيث : $C_N^n = \frac{N!}{n!(N-n)!}$ ؛ فمثلا لو كان حجم المجتمع 10 وحجم العينة 03 وكان السحب مع الإرجاع فيمكن سحب 10^3 = 1000 عينة مختلفة ، أما إذا كان السحب بدون إرجاع فيمكن سحب $C_{10}^3 = \frac{10!}{3!(10-3)!} = 120$ ؛ عينة مختلفة.

ذكرنا في المحور الأول أهمية التوزيع الطبيعي في التطبيقات الإحصائية الواسعة والتوزيعات المشتقة منه (مربع كاي، ستيودنت، فيشر) وسنذكر الآن دون برهان نظرية هامة تسمى نظرية النهاية المركزية والتي تلعب دورا هاما في الإحصاء التطبيقي حيث تمكننا من تقريب أي توزيع احتمالي للتوزيع الطبيعي عند تحقق شروط معينة.

- **<u>العينة النفاذية والعينة غير النفاذية Echantillon exhaustif et non exhaustif</u>**

عندما يكون السحب بالإرجاع حيث يمكن أن تظهر المفردة أكثر من مرة في العينة، نسمي هذه المعاينة غير نفاذية لأن تكرار العملية لا يؤدي إلى تقليص عدد المفردات في المجتمع، والعكس نسمي المعاينة بدون إرجاع معاينة نفاذية.

- **<u>إحصائية المعاينة Statistique de l'échantillonnage</u>**

لتقدير معالم المجتمع (متوسط المجتمع μ، تباين المجتمع σ² النسبة p ...) ننطلق من بيانات العينة، حيث نحتاج إلى حساب معالم مثل متوسط العينة$\bar{X}$ ، تباين العينة S²، النسبة في العينة$\hat{P}$. بصفة عامة، نسمي كل قيمة تحسب انطلاقا من بيانات العينة من أجل تقدير قيمة معالم المجتمع إحصائية المعاينة. نظريا (رياضيا) إحصائية المعاينة هي كل دالة في المتغيرات العشوائية التي تمثل القيم المحصل عليها في العينة.

<u>I - توزيع المعاينة للمتوسطات</u>

1 <u>متوسط توزيع المعاينة للمتوسطات</u>

<u>مثال 1:</u> ليكن المجتمع 1، 3، 5، 6، 8. ما هي القيمة المتوقعة لمتوسط عينة مسحوبة بالإرجاع مكونة من مفردتين $E(\bar{X})$؟ أحسب متوسط المجتمع μ. قارن بين $E(\bar{X})$ وμ. من أجل تحديد ذلك أحسب جميع الحالات الممكنة للمتوسط $\bar{X}$ حسب كل عينة.

<u>الحل :</u>العينات الممكنة ذات الحجم n = 2 من مجتمع حجمه 5 عددها: 5*5 =25

العينات الممكنة				
(1, 1)	(3, 1)	(5, 1)	(6, 1)	(8, 1)
(1, 3)	(3, 3)	(5, 3)	(6, 3)	(8, 3)
(1, 5)	(3, 5)	(5, 5)	(6, 5)	(8, 5)
(1, 6)	(3, 6)	(5, 6)	(6, 6)	(8, 6)
(1, 8)	(3, 8)	(5, 8)	(6, 8)	(8, 8)

المتوسطات الممكنة للعينة (معاينة غ نفاذية) $\bar{X}_i$					$(\sum \bar{X}_i)$
1	2	3	3,5	4,5	
2	3	4	4,5	5,5	
3	4	5	5,5	6,5	115
3,5	4,5	5,5	6	7	
4,5	5,5	6,5	7	8	

القيمة المتوقعة لـ$\bar{X}$ هي متوسط قيمها وهي : $E(\bar{X}_i) = (\sum \bar{X}_i) / 25 = 4,6$

حساب متوسط المجتمع: $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

مثال2. أوجد نفس مطالب المثال 1. في حالة السحب بدون إرجاع.

<u>**الحل :**</u> العينات الممكنة عددها: $10 C_5^2 =$

العينات الممكنة بدون إرجاع			
(1, 3)			
(1, 5)	(3, 5)		
(1, 6)	(3, 6)	(5, 6)	
(1, 8)	(3, 5)	(5, 8)	(6, 8)

المتوسطات الممكنة للعينة أو توزيع المعاينة للمتوسطات (معاينة نفادية) $\bar{X}_i$				$(\sum \bar{X}_i)$
2				
3	4			46
3,5	4,5	5,5		
4,5	5,5	6,5	7	

القيمة المتوقعة لـ$\bar{X}_i$ هي متوسط قيمها وهي: $E(\bar{X}_i) = \mu_{\bar{X}} = (\sum \bar{X}_i) / 10 = 4,6$

متوسط المجتمع : $\mu = (1 + 3 + 5 + 6 + 8)/5 = 4.6$

نظرية 1. إذا كانت X م ع تمثل مجتمع ما و$\bar{X}_i$ متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع، فإن القيمة المتوقعة لمتوسط العينة $E(\bar{X})$ تكتب كما يلي: $E(\bar{X}) = \mu_{\bar{X}} = \mu$

<u>البرهان :</u> لنرمز بـ Xi لقيم المتغيرة الأصلية X .

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n}\sum_i E(Xi) = \frac{1}{n}\sum_i \mu = \frac{1}{n}n\mu = \mu.$$

2 تباين توزيع المعاينة للمتوسطات

أ - حالة المعاينة بالإرجاع

مثال. أحسب تباين المجتمع في المثال 1، أحسب التباين (والانحراف المعياري) لتوزيع المعاينة للمتوسطات $\delta_{\bar{X}}$ علما أن العينة مسحوبة بالإرجاع (غ نفادية)، قارن بين تباين المجتمع وتباين متوسطات العينات الممكنة (توزيع المعاينة للمتوسطات).

الحل :

$\bar{X}$				
1	2	3	3,5	4,5
2	3	4	4,5	5,5
3	4	5	5,5	6,5
3,5	4,5	5,5	6	7
4,5	5,5	6,5	7	8

$$\boldsymbol{\delta^2_{\bar{X}}} = [\sum(\bar{X}_i - E(\bar{X}))^2\ ]/25 = 2.92;$$

$$\boldsymbol{\delta^2} = [\sum i\ (xi - \mu)^2\ ]/5 = 5.84$$

$$2.92 = 5.84 / 2$$

هذا المثال يمهد للنظرية التالية:

نظرية 2. إذا كانت X م ع تمثل مجتمع ما و $\bar{X}_i$ متغيرة ع تمثل متوسط عينة مسحوبة من ذات المجتمع بالإرجاع، فإن تباين $\bar{X}_i$ (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي: $\boldsymbol{\delta^2_{\bar{X}} = \frac{\delta^2}{n}}$ حيث n حجم العينة.

البرهان: لنرمز ب Xi لقيم المتغيرة الأصلية X .

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_i Xi\right) = \frac{1}{n^2}\sum_i V(Xi) = \frac{1}{n^2}\sum_i \sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

ب - حالة المعاينة بدون إرجاع.

مثال: في المثال 1 أحسب تباين المتوسطات الممكنة للعينة $\boldsymbol{\delta^2_{\bar{x}}}$ في حالة المعاينة بدون إرجاع، قارن بين تباين المجتمع وتباين المتوسطات الممكنة للعينة.

المتوسطات الممكنة للعينة أو توزيع المعاينة للمتوسطات (معاينة نفادية) $\bar{X}_i$			
2			
3	4		
3,5	4,5	5,5	
4,5	5,5	6,5	7

الحل : تباين المتوسطات الممكنة للعينة:

$\boldsymbol{\delta^2_{\bar{X}}} = [\sum(\bar{X}_i - E(\bar{X}_i))^2\ ]/10 = 2.19$

تباين المجتمع: $\boldsymbol{\delta^2} = [\sum i\ (xi - \mu)^2\ ]/5 = 5.84$

أو بطريقة ثانية:

$\boldsymbol{\delta^2} = E(X^2) - E(X)^2$

$= (1 + 9 + 25 + 36 + 64) / 5 - 4.6^2 = 5.84)$

المقارنة بين تباين متوسط العينة و تباين المجتمع: $2.19 = \frac{5.84}{2}\left(\frac{5-2}{5-1}\right)$

هذا يمهد للنظرية التالية:

نظرية 3. إذا كانت X م ع تمثل مجتمع ما حجمه N و $\bar{X}_i$ متغيرة ع تمثل متوسط عينة حجمها n مسحوبة من ذات المجتمع بدون إرجاع، فإن تباين $\bar{X}_i$ (تباين توزيع المعاينة للمتوسطات) يكتب كما يلي:

$$\boldsymbol{\delta^2_{\bar{X}} = \frac{\delta^2}{n}\left(\frac{N-n}{N-1}\right)}$$

وتسمى النسبة $\boldsymbol{\frac{N-n}{N-1}}$: معامل الإرجاع.

3 طبيعة توزيع $\bar{X}$

ندرس طبيعة توزيع متوسط توزيع المعاينة للمتوسطات من خلال النظريات التالية:

نظرية 4. إذا كان المجتمع موزع طبيعيا بمتوسط μ وتباين δ^2 فإن متوسط العينة المسحوبة منه يتبع أيضا التوزيع الطبيعي بمتوسط μ وتباين δ^2/n، ونكتب $\bar{X} \longrightarrow N(\mu\ ;\ \delta^2/n)$

نظرية 5. (نظرية النهاية المركزية): إذا كان المجتمع الذي تسحب منه العينة ذو متوسط μ وتباين δ^2 لكن ليس بالضرورة طبيعيا فإن المتغيرة المعيارية ل $\bar{X}_i$ **أي** $\boldsymbol{Z = \frac{\bar{X}_i-\mu}{\delta/\sqrt{n}}}$ تؤول إلى التوزيع الطبيعي المعياري عندما يكون n كبيرا $(n \geq 30)$ ونكتب: $Z \longrightarrow N(0, 1)$.

في حالة المجتمع محدود والمعاينة نفاذية نستبدل العبارة $\delta/\sqrt{n}$ بـــ $\boldsymbol{\delta_{\bar{X}} = \frac{\delta}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}}$

عمليا يستخدم الإحصائيين هذه الصيغة المعدلة بمعامل الإرجاع للانحراف المعياري عندما $n/N \geq 0.05$

مثال 1: مجتمع حجمه 900 بمتوسط $\mu = 20$ و$\delta = 12$. نستخرج كل العينات الممكنة.

أحسب المتوسط والانحراف المعياري لتوزيع المعاينة للمتوسطات في حالة:

(1) حجم العينة $n = 36$،

(2) $n = 64$.

الحل : $\mathbf{E(\bar{X}_i) = \mu = 20}$

(1) $n = 36$: $n/N = 36/900 = 0.04 < 0.05 \Rightarrow \boldsymbol{\delta_{\bar{X}}} = \frac{\boldsymbol{\delta}}{\sqrt{\boldsymbol{n}}} = \frac{\mathbf{12}}{\sqrt{\mathbf{36}}} = 2$

(2) $n = 64$: $N = 900$; $\frac{n}{N} = \frac{64}{900} = 0{,}071 > 0{,}05$

$$\boldsymbol{\delta_{\bar{X}}} = \frac{\mathbf{12}}{\sqrt{\mathbf{64}}}\sqrt{\left(\frac{\mathbf{900-64}}{\mathbf{900-1}}\right)} = \mathbf{1,395}$$

مثال2. باستخدام معطيات المثال السابق ($n = 36$) أحسب احتمال أن يكون $\bar{X}$ محصورا بين 18 و22.

أحسب نفس الاحتمال في حالة $n = 64$.

الحل :

$$\text{I/ } \boldsymbol{z_1} = \frac{\bar{\boldsymbol{X}}_{\mathbf{1}}-\boldsymbol{\mu}}{\boldsymbol{\delta}/\sqrt{\boldsymbol{n}}} = \frac{\mathbf{18-20}}{\mathbf{12}/\sqrt{\mathbf{36}}} = \mathbf{-1} \ ; \ \boldsymbol{Z_2} = \mathbf{1} \ \rightarrow \boldsymbol{p}(\mathbf{-1} < Z < 1) = \mathbf{0,682}$$

$$\text{II / } \boldsymbol{z_1} = \frac{\bar{\boldsymbol{X}}_{\mathbf{1}}-\boldsymbol{\mu}}{\boldsymbol{\delta}/\sqrt{\boldsymbol{n}}}\frac{\mathbf{1}}{\sqrt{\left(\frac{\boldsymbol{N-n}}{\boldsymbol{N-1}}\right)}} = \frac{\mathbf{18-20}}{\mathbf{1,395}} = \mathbf{-1,45} \ ; \ \boldsymbol{Z_2} = \mathbf{1,45}$$

$$\rightarrow \boldsymbol{p}\left(\mathbf{18} < \bar{\boldsymbol{X}} < 22\right) \rightarrow \boldsymbol{p}(\mathbf{-1,45} < Z < 1{,}45) = \mathbf{0,85}$$

إذا كان تباين المجتمع الطبيعي مجهول :

في هذه الحالة نقدر تباين المجتمع بتباين العينة المسحوبة منه ونفرق هنا بين حالتين :

- **عندما يكون حجم العينة n ≥ 30 :**

يكون توزيع الوسط الحسابي للعينة هو : $\bar{X} \longrightarrow \left(\mu\, ; \frac{S^2}{n}\right)$

حيث S^2 هو تباين العينة ويحسب من العلاقة : $S^2 = \frac{\sum(xi - \bar{x})^2}{n-1}$

ويكون التوزيع المعياري المقابل هو : $Z = \frac{\bar{x} - \mu}{S/\sqrt{n}} \longrightarrow$ **N(0 ;1)**

مثال : إذا كان : X⟶N (7; σ²) ، فما التوزيع الاحتمالي لمتوسط عينة عشوائية بحجم 36 من مجتمع X و تباينها 9 ، ثم احسب $P(\bar{X} > 8)$.

الحل :

حيث أن تباين المجتمع مجهول فإن : $\bar{X} \longrightarrow N\left(\mu\, ; \frac{S^2}{n}\right) = N\left(7\, ; \frac{9}{36}\right)$

ولحساب الاحتمال المطلوب لاحظ أن : $Z = \frac{\bar{X} - 7}{3/\sqrt{36}} = \frac{\bar{X} - 7}{0{,}5} \longrightarrow$ **N(0 ;1)**

وبالتالي يكون الاحتمال المطلوب :

$$P(\bar{X} > 8) = P\left(Z > \frac{8-7}{0{,}5}\right) = 1 - P(Z < 2) = 1 - 0{,}97725 = 0{,}0225$$

- **عندما يكون حجم العينة n < 30 :**

يكون توزيع الوسط الحسابي للعينة هو :

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n-1}} \longrightarrow t\,(n\text{-}1)$$

مثال: إذا كان X⟶N (8; σ²) أوجد التوزيع الاحتمالي لوسط العينة {8,9,10,9,8,9,7,7,5,8}

ثم احسب $P(7 < \bar{X} < 10)$.

الحل :

بما أن تباين المجتمع X مجهول وحجم العينة n =10 وهو أقل من 30 يكون توزيع الوسط الحسابي للعينة هو $\bar{X} \longrightarrow t(9)$ ؛ ولحساب الاحتمال المطلوب نجد أولا تباين العينة :

$$\bar{X} = \frac{\sum X}{n} = \frac{80}{10} = 8$$

$$S^2 = \frac{\sum (X-8)^2}{9} = 2$$

وبالتالي يكون التوزيع : $T = \frac{\bar{X}-8}{1{,}414/\sqrt{9}} \longrightarrow t\,(9)$

ويكون الاحتمال المطلوب :

$$P(7 < \bar{X} < 10) = P\left(\frac{7-8}{1{,}414/\sqrt{9}} < \frac{\bar{x}-8}{1{,}414/\sqrt{9}} < \frac{10-8}{1{,}414/\sqrt{9}}\right)$$

$$= P(-2{,}12 < T_9 < +4{,}24)$$

$$= P(T_9 < 4{,}24) - [1 - P(T_9 < 2{,}12)]$$

$$= (0{,}995) - (0{,}025) = 0{,}97$$

ثالثا : إذا كان المجتمع طبيعي ذو حجم محدود وكان حجم العينة n < 30 وكان السحب بدون إرجاع فإنه :

- **عند معلومية σ^2 يكون توزيع الوسط الحسابي للعينة هو :**

$$\bar{X} \longrightarrow N\left(\mu\,;\frac{\sigma^2}{n}\,\frac{N-n}{N-1}\right)$$

مثال : من مجتمع بتوزيع طبيعي حجمه 200 مشاهدة وتباينه $\sigma^2 = 64$ تم سحب عينة عشوائية بدون إرجاع بحجم 50 مشاهدة، فما هو توزيع الوسط الحسابي للعينة ، ثم احسب

$P(\bar{X} > 32)$ لما $\mu = 30$.

الحل : حيث أن تباين المجتمع معلوم فإن توزيع الوسط الحسابي للعينة هو :

$$\bar{X} \longrightarrow \mathrm{N}\left(30\,;\frac{64}{50}\ \frac{200-50}{200-1}\right) = N(30;0{,}96)$$

وبالتالي : $Z = \frac{\bar{X}-30}{\sqrt{0{,}96}} \longrightarrow \mathrm{N}\ (0\ ;\ 1)$

ويكون الاحتمال المطلوب هو :

$$P(\bar{X} > 32) = P\left(Z > \frac{32-30}{\sqrt{0{,}96}}\right) = P\ (Z > 2{,}02)$$

$$= 1 - P(Z < 2{,}02) = 1 - 0{,}97831 = 0{,}02169$$

- **عند مجهولية σ^2 يكون توزيع الوسط الحسابي للعينة هو :**

$$\bar{X} \longrightarrow N\left(\ \mu\,;\frac{S^2}{n}\ \frac{N-n}{N-1}\right)$$

حيث تم استبدال تباين العينة **S^2** بدلا من تباين المجتمع **σ^2** في الحالة السابقة .

مثال : أعد حل المثال السابق إذا كان **σ^2** مجهول وكان تباين العينة **$S^2 = 49$** .

الحل : حيث أن تباين المجتمع مجهول فإن توزيع الوسط الحسابي للعينة هو :

$$\bar{X} \longrightarrow \mathrm{N}\left(30\,;\frac{49}{50}\ \frac{200-50}{200-1}\right) = N(30;0{,}74)$$

وبالتالي : $Z = \frac{\bar{X}-30}{\sqrt{0{,}74}} \longrightarrow \mathrm{N}\ (0\ ;\ 1)$

ويكون الاحتمال المطلوب هو :

$$P(\bar{X} > 32) = P\left(Z > \frac{32-30}{\sqrt{0{,}74}}\right) = P\ (Z > 2{,}32)$$

$$= 1 - P(Z < 2{,}32) = 1 - 0{,}98983 = 0{,}01017$$

ملاحظة : عند سحب عينة من مجتمع توزيعه مجهول أو يتبع توزيعا غير التوزيع الطبيعي، فإن توزيع الوسط الحسابي للعينة التي حجمها $n \geq 30$ يُقَرَّب إلى التوزيع الطبيعي وذلك حسب نظرية النهاية المركزية ، حيث يكون :

$$\bar{X} \longrightarrow \mathrm{N}\left(\mu\,;\frac{S^2}{n}\right)$$

مثال : تم سحب عينة عشوائية بحجم 36 مشاهدة من مجتمع مجهول التوزيع، فبلغ الوسط الحسابي للعينة 10 بتباين 12 ، فما هو التوزيع الاحتمالي لمتوسط قيم العينة .

الحل : مباشرة من الملاحظة السابقة يكون :

$$\bar{X} \longrightarrow N\left(10\,;\frac{12}{36}\right)$$

II - توزيع الفرق بين متوسطي عينتين :

نظرية 1 : إذا كان $X_1 \longrightarrow N(\mu_1; \sigma_1^2)$ وسحبت من مجتمعه عينة حجمها n_1 و $X_2 \longrightarrow N(\mu_2; \sigma_2^2)$ وسحبت من مجتمعه عينة حجمها n_2 وكان المجتمعين مستقلين وعرفنا المتغيرين العشوائيين $\bar{X}_1$ و $\bar{X}_2$ فإن $\bar{X}_1 - \bar{X}_2$ يمثل متغيرا عشوائيا للفرق بين متوسطي العينتين .ويمكن الحصول على الوسط الحسابي والتباين له كالتالي :

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2}$$

$$\sigma^2{}_{\bar{X}_1-\bar{X}_2} = \sigma^2{}_{\bar{X}_1} + \sigma^2{}_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

وسنتحدث الآن عن حالات مختلفة لإيجاد توزيع الفرق بين وسطي العينتين :

1 – عند معلومية σ_1^2 و σ_2^2 :

$$\bar{X}_1 - \bar{X}_2 \longrightarrow N\left(\mu_1 - \mu_2;\ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \longrightarrow N(0;1)$$

مثال : إذا كان $X_1 \longrightarrow N(30; 25)$ وسحبت من مجتمعه عينة حجمها 30 مشاهدة

و$X_2 \longrightarrow N(20; 16)$ وسحبت من مجتمعه عينة حجمها 35 مشاهدة أوجد توزيع الفرق بين متوسطي العينتين ؛

ثم احسب الاحتمال $P(\bar{X}_1 - \bar{X}_2 < 12)$.

الحل : بالتطبيق المباشر للعلاقة : $\bar{X}_1 - \bar{X}_2 \longrightarrow N\left(\mu_1 - \mu_2;\ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

نحصل على : $\bar{X}_1 - \bar{X}_2 \longrightarrow N\left(30-20;\ \frac{25}{30}+\frac{16}{35}\right) = N(10; 1{,}29)$

وبالتالي فإن :

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - 10}{\sqrt{1{,}29}}$$

فيكون الاحتمال المطلوب هو :

$$P(\bar{X}_1 - \bar{X}_2 < 12) = P\left(\frac{(\bar{X}_1 - \bar{X}_2) - 10}{\sqrt{1{,}29}} < \frac{12-10}{\sqrt{1{,}29}}\right)$$

$$= P\left(Z < \frac{12-10}{\sqrt{1{,}29}}\right)$$

$$= P(Z < 1{,}76) = 0{,}9608$$

2 – عند مجهولية σ_1^2 و σ_2^2 وحجم العينتين كبير:

يتم استبدال تبايني المجتمعين في الصيغة السابقة بتبايني العينتين.

$$\bar{X}_1 - \bar{X}_2 \longrightarrow N\left(\mu_1 - \mu_2;\ \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)$$

<u>مثال :</u> معمل ينتج 700 كلغ من الأرز كمعدل يومي ، سحبت منه عينة عشوائية تمثل إنتاج 40 يوما فكان الانحراف المعياري لها 40 كلغ. معمل آخر ينتج 500 كلغ كمعدل يومي ، سحبت منه عينة عشوائية تمثل إنتاج 35 يوما فكان الانحراف المعياري لها 20 كلغ. فما هو التوزيع الاحتمالي للفرق بين متوسطي العينتين، ثم احسب الاحتمال التالي :

$$P(180 < \bar{X}_1 - \bar{X}_2 < 210)$$

<u>الحل :</u> لدينا من المسألة $S_1^2 = 1600$ و $S_2^2 = 400$ و حيث أن $n_1=40$ و $n_2=35$ وكليهما أكبر من 30 إذن يكون توزيع الفرق بين العينتين هو :

$$\bar{X}_1 - \bar{X}_2 \longrightarrow N\left(700 - 500;\ \frac{1600}{40} + \frac{400}{35}\right)$$

$$\bar{X}_1 - \bar{X}_2 \longrightarrow N(200;\ 51{,}43)$$

وبالتالي :

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - 200}{\sqrt{51,43}} \longrightarrow N\ (0; 1)$$

لذلك يكون الاحتمال المطلوب هو :

$$P(180 < \bar{X}_1 - \bar{X}_2 < 210) = P(\frac{180 - 200}{\sqrt{51,43}} < \frac{(\bar{X}_1 - \bar{X}_2) - 200}{\sqrt{51,43}} < \frac{210 - 200}{\sqrt{51,43}})$$

$$= P\ (-2,79 < Z < 1,39) = P(Z < 1,39) + P(Z < 2,79) - 1$$

$$= 0,91774 + 0,99736 - 1 = 0,9151$$

ملاحظة 1 : التوزيع أعلاه يصلح أيضا عند السحب من مجتمعين غير طبيعيين أو مجهولي التوزيع بشرط كبر حجم العينة ، وذلك استنادا لنظرية النهاية المركزية .

ملاحظة 2 : عند الحديث عن توزيع المعاينة لمجموع إحصائيتين $(\bar{X}_1 + \bar{X}_2)$ فإن الوسط الحسابي والتباين لهذا النوع من التوزيع يكون وفق الصيغتين التاليتين :

$$\mu_{\bar{x}_1 + \bar{x}_2} = \mu_{\bar{x}_1} + \mu_{\bar{x}_2} = \mu_1 + \mu_2$$

$$\sigma^2{}_{\bar{X}_1 + \bar{X}_2} = \sigma^2{}_{\bar{X}_1} + \sigma^2{}_{\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

2 – عند مجهولية $\boldsymbol{\sigma_1^2}$ و $\boldsymbol{\sigma_2^2}$ وحجم العينتين صغير (أقل من 30) :

إن قيمة التباينات في المجتمعات غالبا ما تكون مجهولة، وعليه عند تقدير الانحراف المعياري للفرق $(\bar{X}_1 - \bar{X}_2)$علينا أن نستخدم تبايني العينتين ؛ بالإضافة لذلك إذا كانت العينتين مستقلتين وصغيرتا الحجم فإن توزيع المعاينة لــ $(\bar{X}_1 - \bar{X}_2)$ ليس له توزيع طبيعي ، بل له توزيع **ستيودنت *(t)*** بدرجات حرية ***(v)*.**

ونميز هنا حالتين :

أ - حالة تبايني المجتمعين مجهولين ومتساويين :

ليكن $\boldsymbol{\sigma_1^2}$ هو التباين الخاص بالمجتمع الأول وهو مجهول القيمة ، وكان $\boldsymbol{\sigma_2^2}$ التباين الخاص بالمجتمع الثاني وهو مجهول القيمة أيضا ، وكان $\boldsymbol{\sigma_2^2 = \sigma_1^2}$ ؛ فإن التباين المشترك لهما (وهو مجهول أيضا) هو $\boldsymbol{\sigma^2}$ حيث :

$\boldsymbol{\sigma^2 = \sigma_1^2 = \sigma_2^2}$ ومنه تكون صيغة الانحراف المعياري للفرق $(\bar{X}_1 - \bar{X}_2)$ على الشكل التالي :

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

وبما أن التباين المشترك σ^2 مجهول القيمة فإننا نستطيع تقديره باستخدام تبايني العينيتين S_1^2 و S_2^2 ؛ حيث يكون تقدير التباين هو متوسط مرجح للقيم S_1^2 و S_2^2 وتكون الترجيحات مبينية على أساس حجم العينات . ولكي يكون تقدير التباين تقديرا غير متحيزا لـ σ^2 فإننا يجب أن نستخدم درجات الحرية ($V_1= n_1 - 1$) و ($V_2= n_2 - 1$) كترجيحات بدلا من استخدام حجم العينيتن n_1 و n_2 بشكل مباشر.

وبناء على ذلك فإن مقدر التباين المشترك σ^2 هو S_p^2 ويعطى بالصيغة التالية :

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

ومنه يكون :

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

حيث يسمى S_p^2 **بتيباين العينة التجميعي "pooled simple variance"** وذلك لأنه مكون من تجميع المعلومات عن العينيتين معا .

نظرية 2 : إذا كان $\bar{X}_1$ هو الوسط الحسابي لعينة عشوائية حجمها n_1 سحبت من مجتمع وسطه μ_1 وتباينه σ_1^2 ، وكان $\bar{X}_2$ هو الوسط الحسابي لعينة عشوائية حجمها n_2 سحبت من مجتمع آخر مستقل عن الأول وسطه μ_2 وتباينه σ_2^2، وكان حجم العينتين صغير (أقل من 30) وتبايني المجتمعين مجهولين ومتساويين فإن المتغير :

$$T = \frac{[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\ ]}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \to t_v \ ; \quad v = n_1 + n_2 - 2$$

مثال : أخذت عينة عشوائية حجمها 18 من مجتمع إحصائي وسطه 35 ؛ وأخذت عينة عشوائية أخرى حجمها 22 من مجتمع إحصائي وسطه 33، فإذا كان تبايني العينتين هما على التوالي : 6 و 9 ، وكان تبايني المجتمعين مجهولين ومتساويين . فأوجد احتمال أن يكون الفرق بين متوسطي هاتين العينتين أقل من 3 .

الحل :

لدينا : $n_1 = 18\ ;\ n_2 = 22\ ;\ \mu_1 = 35;\quad \mu_2 = 33\ ;\ S_1^2 = 6\ ;\ S_2^2 = 9$

بما أن تبايني المجتمعين مجهولين ومتساويين نستخدم توزيع t بدرجة حرية $v = n_1 + n_2 - 2$

وذلك لإيجاد الاحتمال التالي :

$$P(\bar{X}_1-\bar{X}_2<3)=P\left(\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\ \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\ \frac{1}{n_2}\right)}}<\ \frac{3-(35-33)}{\sqrt{S_p^2\left(\frac{1}{18}+\ \frac{1}{22}\right)}}\right)$$

الآن نقوم بحساب S_p^2 حيث :

$$s_p^2\ =\frac{(n_1\ -\ 1)\ s_1^2\ +\ (n_2\ -\ 1)\ s_2^2}{n_1\ +\ n_2\ \ -\ 2}$$

$$S_p^2=\ \frac{(18-1)*6+(22-1)*9}{18+22-2}=7{,}66$$

$$P(\bar{X}_1-\bar{X}_2<3)=P\left(T<\frac{1}{\sqrt{(7{,}66)*\left(\frac{1}{18}+\frac{1}{22}\right)}}\right)$$

$$P(T_{38}<1{,}1376)=0{,}9$$

إذن :
$$P(\bar{X}_1-\bar{X}_2<3)=0{,}9$$

ب - **<u>حالة تبايني المجتمعين مجهولين وغير متساويين :</u>**

إذا كان $\bar{X}_1$ و $\bar{X}_2$ هما الوسطين الحسابيين لعينتين مستقلتين صغيرتي الحجم تم سحبهما من مجتمعين ذو متوسطين μ_1 و μ_2 وذو تباينين مجهولين وغير متساويين فإننا نستطيع تقدير الانحراف المعياري للفرق $(\bar{X}_1-\bar{X}_2)$ باستخدام تبايني العينيتين مباشرة S_1^2 و S_2^2 وذلك كما يلي :

$$\sigma_{\bar{X}_1-\bar{X}_2}=\ \sqrt{\frac{S_1^2}{n_1}+\ \frac{S_2^2}{n_2}}$$

وعليه فإن :
$$T=\ \frac{\left[(\bar{X}_1-\bar{X}_2)-\ (\mu_1-\ \mu_2)\ \right]}{\sqrt{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)}}\rightarrow t_v$$

ويتم حساب درجة الحرية بالصيغة المركبة التالية :
$$v = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n_1-1)}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2-1)}}$$

مثال : إذا كانت نقاط طلبة السنة الثانية علوم اقتصادية في مقياس الإحصاء لديها وسط حسابي قدره 15 ، وكانت نقاط طلبة السنة الثانية محاسبة ومالية في نفس المقياس لديها وسط حسابي قدره 10 ؛ وقمنا باختيار عينة عشوائية من المجموعة الأولى حجمها 25 طالب ، وعينة من المجموعة الثانية قدرها 20 طالب ؛ وكان تباين العينة الأولى 6 وتباين العينة الثانية 4 . **فأوجد احتمال أن يكون الفرق بين متوسطي العينيتين أكثر من 6** ، مع العلم أن تبايني المجتمعين مجهولين وغير متساويين .

الحل :

لدينا : $\boldsymbol{n_1 = 25\,;\ n_2 = 20\,;\ \mu_1 = 15;\quad \mu_2 = 10\,;\ S_1^2 = 6\,;\ S_2^2 = 4}$

بما أن تبايني المجتمعين مجهولين وغير متساويين نستخدم توزيع t بدرجة حرية

$$v = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{(n_1-1)}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{(n_2-1)}} = \frac{\left(\frac{6}{25}+\frac{4}{20}\right)^2}{\frac{\left(\frac{6}{25}\right)^2}{(25-1)}+\frac{\left(\frac{4}{20}\right)^2}{(20-1)}} = 43$$

وذلك لإيجاد الاحتمال التالي :

$$P(\bar{X}_1-\bar{X}_2 > 6) = P\left(\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)}} > \frac{6-(15-10)}{\sqrt{\left(\frac{6}{25}+\frac{4}{20}\right)}}\right)$$

$$P\,(T_{43} > 1{,}5076) = 1 - P\,(T_{43} \le 1{,}5076) = 1 - 0{,}95 = 0{,}05$$

إذن :
$$P(\bar{X}_1-\bar{X}_2 > 6) = 0{,}05$$

III - توزيع النسبة للعينة :

في المجتمعات الإحصائية التي تتبع توزيع ذي الحدين نسبة لخاصية معينة ندرسها في المجتمع الإحصائي ونرمز لها بالرمز P ، مثل نسبة الصالح من إنتاج مصنع معين، حيث يتم حساب النسبة بقسمة مجموع مفردات الخاصية X على حجم المجتمع N أي أن $\mathbf{P} = \frac{\mathbf{X}}{\mathbf{N}}$ ، فلو أخذنا جميع العينات التي حجمها n من مجتمع حجمه N قد تختلف النسب لهذه العينات، مما يعني أن سلوك هذه النسب هو سلوك متغير عشوائي وبالتالي هذا المتغير العشوائي له توزيع احتمالي مستمد من توزيع المجتمع الذي سحبت منه العينات، وسنرمز لنسبة الظاهرة في العينة بالرمز $\widehat{\mathrm{P}}$ وبعدد مفردات الخاصية فيها X فتكون نسبة الخاصية في العينة $\widehat{\mathbf{P}} = \frac{\mathbf{X}}{\mathbf{n}}$.

إن هذه النسبة تمثل متغير عشوائي له توزيع احتمالي يسمى توزيع النسبة حيث يكون وسطه الحسابي :

$$\mu_{\widehat{\mathrm{P}}} = E(\widehat{\mathrm{P}}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = P$$

ويكون تباينه هو :

$$\sigma^2{}_{\widehat{\mathrm{P}}} = V(\widehat{\mathrm{P}}) = V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} * np\,(1-p) = \frac{p(1-p)}{n}$$

وفي حالة العينات الكبيرة n ≥ 30 – وهي موضوع دراستنا – يقرب التوزيع للتوزيع الطبيعي وذلك استنادا لنظرية النهاية المركزية ونعبر عن ذلك بالصيغة

$$\mathbf{Z} = \frac{\widehat{\mathbf{P}} - \boldsymbol{\mu}_{\widehat{\mathbf{P}}}}{\boldsymbol{\sigma}_{\widehat{\mathbf{P}}}} \longrightarrow \mathbf{N(0\,,1)}$$

مثال : مصنع ينتج عادة 25% عبوات كبيرة الحجم ، سحبت من إنتاجه عينة حجمه 2200 عبوة تبين أن منها 500 عبوة كبيرة الحجم .

أوجد توزيع النسبة للعبوات الكبيرة ثم احسب احتمال أن المصنع ينتج أقل من 26% من العبوات الكبيرة في فترة إجراء البحث .

الحل : لاحظ أن المجتمع هنا هو مجتمع ذو حدين بنسبة نجاح P = 0,25 ، ولاحظ أن حجم العينة كبير وبحساب وسط التوزيع وتباينه نجد أن :

$\mu_{\hat{p}} = p = 0{,}25$

$.\ \frac{n}{N} = \frac{500}{2200} = 0{,}227 \gg 0{,}05 \rightarrow \sigma^2{}_{\hat{p}} = \frac{p(1-p)}{n} * \frac{N-n}{N-1} = \frac{0{,}25*0{,}75}{500} * \frac{2200-500}{2200-1} = 0{,}00031$

$\mathrm{n} = 500 \gg 30 \rightarrow \widehat{\mathrm{P}} \longrightarrow \mathrm{N}(0\,;1)$ ؛ (حسب نظرية النهاية المركزية)

$$\mathrm{Z} = \frac{\widehat{\mathrm{P}} - \mu_{\widehat{\mathrm{P}}}}{\sigma_{\widehat{\mathrm{P}}}} = \frac{\widehat{\mathrm{P}} - 0{,}25}{0{,}0176} \longrightarrow \mathrm{N}(0\,,1)$$

هذا التوزيع عبارة عن توزيع ثنائي خاص بالمتغيرات العشوائية المنقطعة ، ولكي نتمكن من استعمال التوزيع الطبيعي في حساب الاحتمال المطلوب يتم تقريب التوزيع الثنائي المنقطع إلى التوزيع الطبيعي المستمر وهو ما يتطلب القيام بتصحيح الاستمرارية حسب القاعدة التالية :

$$P\,(\bar{p} < p) = P\left(\bar{p} < p - \frac{1}{2n}\right)$$

$$P\ (\bar{p}\ > p) = P\ \left(\bar{p} > p + \frac{1}{2n}\right)$$

$$P\ (\widehat{\mathrm{P}} < 0{,}26) = P\ \left(p < 0{,}26 - \ \frac{1}{2\ (500)}\right) = P(p < 0{,}259)$$

$$\rightarrow P\left(Z < \frac{0{,}259-0{,}25}{0{,}0176}\right) = P\ (Z < 0{,}51) = \ 0{,}69497.$$

IV - توزيع الفرق بين نسبتين :

بفرض أن$X_1 \longrightarrow b\ (N_1\ ; P_1)$ وسحبت من مجتمعه عينة حجمها $n_1 \geq 30$ و$X_2 \longrightarrow b(N_2\ ; P_2)$ وسحبت عينة من مجتمعه عينة حجمها $n_2 \geq 30$ وكان المجتمعين مستقلين فإن : $\widehat{\mathrm{P}}_1 - \widehat{\mathrm{P}}_2$ يمثل متغيرا عشوائيا للفرق بين نسبتي العينتين. ويكون الوسط الحسابي والتباين له كالتالي :

$$\mu_{\hat{p}_1 - \hat{p}_2} = \mu_{\hat{p}_1} - \ \mu_{\hat{p}_2} = P_1 - \ P_2$$

$$\sigma^2{}_{\hat{p}_1 - \hat{p}_2} = \sigma^2{}_{\hat{p}_1} + \ \sigma^2{}_{\hat{p}_2} = \frac{P_1(1 - \ P_1)}{n_1} + \frac{P_2(1 - \ P_2)}{n_2}$$

ويكون توزيع فرق النسب :

$$Z = \frac{\left(\widehat{\mathrm{P}}_1 - \widehat{\mathrm{P}}_2\right) - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \longrightarrow \mathrm{N}(0;1)$$

<u>مثال :</u> إذا علم أن نسبة الذكور في مؤسسة (أ) تبلغ 0,3 وفي مؤسسة (ب) تبلغ 0,2 ، فإذا سحبت عينتين عشوائيا الأولى من المؤسسة (أ) بحجم 100 والثانية من المؤسسة (ب) بحجم 200 فما احتمال أن يكون الفرق بين نسبتي الذكور في العينتين أكبر من 06% .

<u>الحل :</u> لاحظ أن التوزيع هو توزيع فرق بين نسبتين بوسط حسابي وتباين :

$$\mu_{\hat{p}_1 - \hat{p}_2} = \ P_1 - \ P_2 = 0{,}3 - 0{,}2 = 0{,}1$$

$$\sigma^2{}_{\hat{p}_1 - \hat{p}_2} = \ \sigma^2{}_{\hat{p}_1} + \ \sigma^2{}_{\hat{p}_2} = \ \frac{0{,}3 * 0{,}7}{100} + \frac{0{,}2 * 0{,}8}{200} = 0{,}0029$$

ويكون توزيع فرق النسب :

$$Z = \frac{\left(\widehat{\mathrm{P}}_1 - \widehat{\mathrm{P}}_2\right) - 0{,}1}{0{,}054} \longrightarrow \mathrm{N}(0;1)$$

فيكون الاحتمال المطلوب هو :

$$P\big((\hat{p}_1-\hat{p}_2)>0{,}06\big)=P\left(\hat{p}_1-\hat{p}_2>0{,}06+\frac{1}{2(100+200)}\right)=P\,(\hat{p}_1-\hat{p}_2>0{,}0617)$$

$$\rightarrow P\left(Z>\frac{0{,}0617-\ 0{,}1}{0{,}054}\right)\ =P(Z>-0{,}71)=P(Z\leq 0{,}71)=0{,}76115$$

V- توزيع التباين للعينة :

يكون توزيع التباين كالتالي : $\boldsymbol{\chi_0^2}=\ \frac{(\mathrm{n}-1)\delta^2}{\sigma^2}\longrightarrow \boldsymbol{\chi^2}_{(\boldsymbol{n}-\mathbf{1})}$

وكتطبيق على هذا التوزيع نستعرض المثال التالي:

مثال : سحبت عينة عشوائية بحجم 20 مشاهدة من مجتمع توزيع طبيعي تباينه 9 ، احسب احتمال أن يزيد تباين العينة عن 15 .

الحل : حيث أنّ $\boldsymbol{\chi_0^2}=\ \frac{19\ \delta^2}{9}\longrightarrow \boldsymbol{\chi^2}_{(\mathbf{19})}$

و الاحتمال المطلوب : $P(\delta^2>15)=P\ \left(\frac{19\,S^2}{9}>\frac{19*15}{9}\right)=P\left(\chi_0^2>31{,}67\right)$

$$=1-P\left(\chi_0^2\leq 31{,}67\right)=0{,}025$$

لاحظ أن القيمة 31,67 غير موجودة في الجدول رقم 3 الخاص بتوزيع $\boldsymbol{\chi^2}$ فنأخذ القيمة الموالية لها وهي 32,85 ، ويكون الاحتمال الموافق لها هو 0,975 .

VI- توزيع النسبة بين تباين عينتين :

يعتبر هذا التوزيع من التوزيعات الهامة التي تبحث في تجانس المجتمعات ونلجأ لحساب النسب بين التباينات وليس الفرق بينها لسهولة دراسة النسب وتفسيرها .
فإذا سحبت عينة حجمها $\mathbf{n_1}$ وتباينها $\boldsymbol{\delta_1^2}$ من مجتمع $\mathbf{N}(\boldsymbol{\mu_1};\ \boldsymbol{\sigma_1^2})$ وعينة أخرى حجمها $\mathbf{n_2}$ وتباينها $\boldsymbol{\delta_2^2}$ من مجتمع $\mathbf{N}(\boldsymbol{\mu_2};\ \boldsymbol{\sigma_2^2})$ مستقل عن المجتمع الأول فإن :

$$\frac{(n_1-1)\delta_1^2}{\sigma_1^2}\rightarrow\ \boldsymbol{\chi}^2_{(\boldsymbol{n_1}-\mathbf{1})}\ \ ;\ \frac{(n_2-1)\delta_2^2}{\sigma_2^2}\rightarrow\ \boldsymbol{\chi}^2_{(\boldsymbol{n_2}-\mathbf{1})}$$

وبالتالي :

$$\frac{\dfrac{(n_1-1)\delta_1^2}{\sigma_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)\delta_2^2}{\sigma_2^2}\Big/(n_2-1)}\ \ \rightarrow F(n_1-1;n_2-1)$$

بعد تبسيط الطرف الأيسر نحصل على :

$$F = \frac{\delta_1^2 / \sigma_1^2}{\delta_2^2 / \sigma_2^2} \rightarrow F(n_1 - 1; n_2 - 1)$$

وإذا تساوى تبايني المجتمعين فإن :

$$\frac{\delta_1^2}{\delta_2^2} \rightarrow F(n_1 - 1; n_2 - 1)$$

مثال :سحبت عينة حجمها 15 من مجتمع طبيعي تباينه 9 ، وسحبت عينة أخرى حجمها 21 من مجتمع طبيعي تباينه 25 مستقل عن المجتمع الأول .

أوجد احتمال النسبة بين تبايني العينتين أقل من أو يساوي 0,8 .

الحل : حيث أنّ :

$$F = \frac{\delta_1^2 / 9}{\delta_2^2 / 25} \rightarrow F(14; 20)$$

فإنّ :

$$P\left(\frac{\delta_1^2}{\delta_2^2} \leq 0{,}8\right) = P\left(\frac{\delta_1^2 / 9}{\delta_2^2 / 25} \leq 0{,}8\left(\frac{25}{9}\right)\right)$$

$$= P(F \leq 2{,}22) = 0{,}95$$

في المحور السابق درسنا من خلال مجموعة من النظريات العلاقة الرياضية بين معالم العينة والمعالم المناظرة لها في المجتمع مثل المتوسط، التباين، النسبة...كما درسنا العلاقة بين شكل توزيع المجتمع وشكل التوزيع الاحتمالي لمعالم العينة. تظهر هذه العلاقات كتوصيف لخصائص العينة ومعالمها ولكنها تستخدم أكثر لتقدير خصائص ومعالم المجتمع محل الدراسة، وهذا ما سنتعرف عليه في هذا المحور.

أولا: مفاهيم أساسية

2 بعض خصائص المقدر

لتقدير معلمة من معالم مجتمع محل دراسة، نحتاج إلى اختيار الإحصائية المناسبة في العينة لتقدير هذه المعلمة. غالبا ما تكون المعلمة المناظرة في العينة هي أحسن مقدر، كأن نقدر متوسط المجتمع μ من خلال متوسط العينة μ_m. تسمى الإحصائية المستخدمة في التقدير المقدر.

(أ) المقدر غير المتحيز

نقول عن إحصـــائية ما بأنها مقدر غير متحيز sans biais لمعلمة المجتمع إذا كان متوســـطها أو توقعها الرياضي مساويا لمعلمة المجتمع.

مثال: نقول عن متوسط العينة m أنه مقدر غير متحيز لمتوسط المجتمع μ لأن $E(m) = \mu$. في المقابل نسمي الإحصائية S^2 في معاينة بالإرجاع أنها مقدر متحيز ل σ^2 لأن $E(S^2) = \sigma^2 (n-1)/n \neq \sigma^2$ ، بينما تعتبر الاحصائية $S'^2 = S^2n/(n-1)$ مقدرا غير متحيز في معاينة بالإرجاع.

(ب) الكفاءة

تتعلق كفاءة (efficacité) مقدر ما بمقدار التباين لتوزيع المعاينة للإحصائية، فإذا كان لمقدرين (إحصائيتين) نفس المتوسط نقول عن المقدر ذو توزيع المعاينة الأقل تباينا أنه الأكثر كفاءة.

مثال: لكل من توزيعي المعاينة للمتوسط والوسيط نفس المتوسط هو متوسط المجتمع μ ، لكن يعتبر المتوسط m مقدرا أكثر كفاءة لمتوسط المجتمع μ من الوسيط لأن تباين توزيع المعاينة للمتوسطات $V(m) = \sigma^2/n$ أقل من تباين توزيع المعاينة للوسيط :

$$.V(méd) = \sigma^2\pi/2n = (\sigma^2/n)\ (3.14159/2) \quad > \quad \sigma^2/n$$

من البديهي أن استخدام مقدرات فعالة وغير متحيزة هو الأفضل، إلا أنه قد يلجأ لمقدرات أخرى لسهولة الحصول عليها.

(ج) التقارب convergeance

نقول عن مقدر أنه متقارب إذا كان يؤول إلى قيمة المعلمة المقدرة عندما يؤول حجم العينة إلى ما لا نهاية.

مثال: يعتبر متوسط العينة مقدرا متقاربا لمتوسط المجتمع لأن:

$$E(m) = \mu \quad , \quad V(m) = \frac{\sigma^2}{n} \xrightarrow[n \to \infty]{} 0.$$

3 التقدير النقطي والتقدير بمجال.

قد نحتاج إلى تقدير لمعلمة مجتمع بقيمة واحدة ونقول عن هذا التقدير أنه **تقدير نقطي**، و أحيانا أخرى نحتاج إلى تقدير معلمة المجتمع بنقطتين يحددان مجال لقيمة المعلمة ونقول عن هذا النوع من التقدير أنه **تقدير بمجال**.

مثال : إذا قدرنا دخل الأسرة في منطقة ما ب 18000 دج، نكون قد قدرنا دخل الأسرة تقديرا نقطيا. يكون تقديرنا بمجال إذا قلنا مثلا أن الدخل يساوي 18000 ± 2000 أي أنه يتراوح بين 16000 و20000دج.

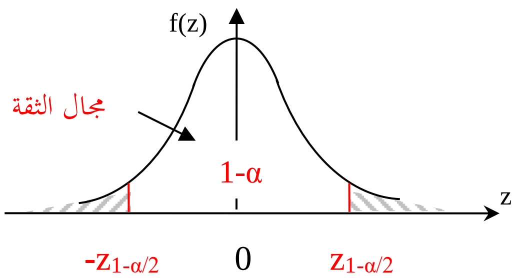
(أ) درجة التأكد

لكي يكون التقدير علميا ينبغي تقييم احتمال أن تكون المعلمة تنتمي فعلا إلى المجال المحدد، لذلك نلحق بالمجال ما يسمى بدرجة أو مستوى الثقة، ويرمز له ب p. الاحتمال المعاكس يسمى احتمال الخطأ ويرمز له ب α ، ويسمى أيضا ""مستوى المعنوية".

مثال: دخل الأسرة في المنطقة (أ) ينتمي إلى المجال [16000، 20000] بمستوى معنوية 5 % أي بمستوى ثقة 95 % . وتسمى الحدود 16000 و20000 **حدود الثقة**.

(ب) تعيين حدود مجال الثقة

تحدد حدود الثقة من خلال معاملات الثقة التي بدورها تحدد من خلال مستوى المعنوية (مستوى الثقة). ففي حالة استخدام التوزيع الطبيعي للتقدير تكون القيمتين ± 1.96 معاملات الثقة من أجل مستوى ثقة 95% بينما القيمتين ±2.58 تمثلان معاملات الثقة من أجل مستوى ثقة 99 % .



رسم 1 مجال الثقة للتوزيع الطبيعي

مثال: ليكن μ_s و σ_s متوســط وانحراف معياري توزيع المعاينة لإحصــائية ما s حيث $\mu_s = \mu$. إذا كان توزيع المعاينة ل s توزيعا طبيعيا (كما هو الحال بالنســبة لأغلب الإحصــائيات عندما ($n \geq 30$)) فإننا نقدر مثلا وبالنظر إلى توزيع s أن:

القيمتين $\mu_s \pm 1.96\sigma_s$ تمثلان **حدود الثقة** ب 95 %، و $\mu_s \pm 2.58\sigma_s$ حدود الثقة ب 99 %.

في حالة التوزيع الطبيعي يرمز لحدود الثقة ب Z_c أو $Z_{1-\alpha/2}$ (أنظر الرسم).

ثانيا: التقدير بمجال

1-مجال الثقة للمتوسط

يقدر متوسط المجتمع μ من خلال الإحصائية m.

(ج) أ – تقدير μ باستخدام التوزيع الطبيعي

نستخدم التوزيع الطبيعي لتحديد مجال الثقة إذا علمنا أن المجتمع الذي سحبت منه العينة يتبع التوزيع الطبيعي.

وفي حالة العينة الممتدة ($n \geq 30$) يمكن كذلك الاستفادة من نظرية النهاية المركزية[1] أن m تتبع التوزيع الطبيعي.

تكتب حدود مجال الثقة كما يلي:

$m \pm z \cdot \frac{S}{\ }$ أو $m \pm z_c \frac{S'}{\sqrt{\ }}$

وفي حالة σ مجهول: $m \pm z_c \cdot \frac{\sigma}{\sqrt{n}}$

و نستخدم هذه الصيغة إلا إذا كان المجتمع محدود (ذا حجم N) والمعاينة نفادية حيث تصبح الصيغة كالآتي:

$$m \pm z_c \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

إلا أنه غالبا ما يكون الانحراف المعياري للمجتمع σ مجهولا، ولذلك نعوض σ في الصيغ السابقة بالمقدر S أو S'.

الجدول الآتي يبين قيم z_c التي تمثل حدود مجال الثقة بحسب مستوى الثقة :

[1] التي تخص في الحقيقة توزيع مجموع قيم العينة –في حالة كون العينة كبيرة بما فيه الكفاية– وليس المتوسط.

0.5	0.8	0.90	0.95	0.98	0.99	مستوى الثقة $1-\alpha$
0.5	0.2	0.10	0.05	0.02	0.01	α مستوى المعنوية
0.75	0.9	0.95	0.975	0.99	0.995	$1-\alpha/2$
0.674	1.282	1.645	1.96	2.326	82.5	$Z_{1-\alpha/2}$

مثال : نقدر أن μ يوجد داخل المجال $m \pm 1.96\sigma_m$ بمستوى ثقة 95% (0.95) أي بمستوى معنوية 5 % (0.05)، وداخل المجال $m \pm 2.58\sigma_m$ بمستوى ثقة 99% أي بمستوى معنوية 0.01...

(د) تقدير μ باستخدام التوزيع t :

في حالة العينة الصغيرة ($n < 30$) وσ مجهول نستخدم توزيع ستيودنت لتحديد مجالات الثقة ل μ.

مثلا القيم $-t_{0.975}$ ؛ $t_{0.975}$ تحد 95 % من المساحة تحت المنحنى ونقول أن $-t_{0.975}$; $t_{0.975}$ تمثل القيم الحرجة أو معاملات الثقة عند مستوى ثقة 95 % ونكتب:

$$-t_{0.975} < \frac{m-\mu}{\hat{S}/\sqrt{n}} < t_{0.975}$$

ومنه نستخلص مجال الثقة ل μ كما يلي:

$$m - t_{0.975}\cdot\frac{\hat{S}}{\sqrt{n}} < \mu < m + t_{0.975}\cdot\frac{\hat{S}}{\sqrt{n}}$$

2 – مجال الثقة للنسبة

(أ) حالة المجتمع غير محدود أو المعاينة غير نفادية و العينة الممتدة ($n \geq 30$) :

لتكن s إحصائية تمثل نسبة "نجاحات" في عينة ذات حجم $n \geq 30$ مستخرجة من مجتمع ثنائي حيث p هي نسبة النجاحات. تستعمل التوزيع الطبيعي لتقدير p فنعين حدود الثقة ل p كما يلي: $p' \pm z_c\sigma_p$ أين p' نسبة النجاحات في العينة،

نعلم من الفصل السابق أن $\sigma_p = \sqrt{\frac{pq}{n}}$ ومنه يحدد مجال الثقة ل p كما يلي:

$$p' \pm z_c\cdot\sqrt{\frac{p(1-p)}{n}}$$

(ب) في حالة كون المجتمع محدود ذا حجم N والمعاينة نفادية:

$$p' \pm z_c\cdot\sqrt{\frac{p(1-p)}{n}}\cdot\sqrt{\frac{N-n}{N-1}}$$

3 مجال الثقة للتباين

لتقدير التباين والانحراف المعياري لمجتمع بمجال ثقة نستعمل الخاصية : $\chi^2_{n-1} \sim \frac{(n-1)\hat{S}^2}{\sigma^2} = \frac{nS^2}{\sigma^2}$.

مثال: مجال الثقة ب95 % يحدد كما يلي:

$$\chi^2_{0.025} \leq \frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi^2_{0.975}$$

ومنه نستنتج مجال الثقة ل σ كما يلي:

$$\frac{\sqrt{n}S}{\chi_{0.975}} \leq \sigma \leq \frac{\sqrt{n}S}{\chi_{0.025}} \qquad ou \qquad \frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}} \leq \sigma \leq \frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$$

نظرا لأن توزيع ك2 غير متماثل فإن المجال أعلاه ليس الأمثل، إذ توجد طريقة لتضييق مجال الثقة أكثر إذا لم نشأ أن تكون أطراف المنحنى متساوية، وهذا بخلاف التوزيعات المتماثلة كالطبيعي وستيودنت.

4 مجالات الثقة لنسبة تباينين

رأينا سابقا (نظرية 11 من الفصل 5) أنه إذا كان لدينا مجتمعان طبيعيان تبايناهما σ^2_1 , σ^2_2 وسحبنا منهما عينتين عشوائيتين حجمهما على التوالي n_2 , n_1 فإن :

$$F = \frac{\left[\frac{S_1^2 n_1}{n_1-1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2 n_2}{n_2-1}\right]\frac{1}{\sigma_2^2}} = \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \to F_{n_1-1;\, n_2-1}$$

إذا يمكن تكوين تقدير بمجال ل F عند مستوى ثقة 0.98 كما يلي:

$$F_{0.01} \leq \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \leq F_{0.99}$$

و من ثم يمكن تقدير النسبة بين تبايني المجتمعين كما يلي:

$$\frac{1}{F_{0.99}}\frac{\hat{S}_1^2}{\hat{S}_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{F_{0.01}}\frac{\hat{S}_1^2}{\hat{S}_2^2}$$

5 خلاصة

لتقدير إحصائية مجتمع نستخدم نظريات توزيع المعاينة. هذه النظريات تتناول خصائص إحصائيات العينة من متوسط العينة، النسبة في العينة، ... و علاقتها بالإحصائيات المناظرة لها في المجتمع.

جدول1 توزيع المعاينة للمتوسطات حسب طبيعة توزيع المجتمع، معلومية التباين و حجم العينة.

قانون المجتمع	تباين المجتمع (σ^2)	n	$\sigma_{\bar{x}}$	قانون $\bar{x}$
طبيعي	معلوم	$n < 30$ أو $n \geq 30$	$\sigma/\sqrt{n}$	$N(\mu ; \sigma/\sqrt{n})$
	غير معلوم	$n \geq 30$	$S'/\sqrt{n}$	$N(\mu ; S'/\sqrt{n})$
		$n < 30$	$S'/\sqrt{n}$	$t_{\alpha; n-1}$
غير معلوم	معلوم	$n \geq 30$	$\sigma/\sqrt{n}$	$N(\mu ; \sigma/\sqrt{n})$
	غير معلوم	$n \geq 100$	$S'/\sqrt{n}$	$N(\mu ; S'/\sqrt{n})$

جدول 2 تحديد مجال الثقة للنسبة، للتباين وللنسبة بين تباينين

المجتمع	التوزيع الاحتمالي للإحصائية	مجال الثقة
مجتمع غير محدود أو معاينة غير نفادية و عينة ممتدة ($n \geq 30$)	التوزيع الطبيعي	$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}}$
مجتمع محدود ذا حجم **N** والمعاينة نفادية	التوزيع الطبيعي	$p' \pm z_c \cdot \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

غير معلوم	$\frac{nS^2}{\sigma^2}=\frac{(n-1)\hat{S}^2}{\sigma^2}\sim\chi^2_{n-1}$	$\frac{\sqrt{n}S}{\chi_{0.975}}\le\sigma\le\frac{\sqrt{n}S}{\chi_{0.025}}$ أو $\frac{\sqrt{n-1}\hat{S}}{\chi_{0.975}}\le\sigma\le\frac{\sqrt{n-1}\hat{S}}{\chi_{0.025}}$
مجتمعين طبيعيين، أو عينتين مسحوبتين من مجتمع طبيعي واحد.	$F=\frac{\left[\frac{S_1^2n_1}{n_1-1}\right]\frac{1}{\sigma_1^2}}{\left[\frac{S_2^2n_2}{n_2-1}\right]\frac{1}{\sigma_2^2}}=\frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2}\to F_{n_1-1;\,n_2-1}$	مثلا عند مستوى ثقة 0.98: $\chi^2_{0.025}\le\frac{nS^2}{\sigma^2}=\frac{(n-1)\hat{S}^2}{\sigma^2}\le\chi^2_{0.975}$ $\frac{1}{F_{0.99}}\frac{\hat{S}_1^2}{\hat{S}_2^2}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{1}{F_{0.01}}\frac{\hat{S}_1^2}{\hat{S}_2^2}$

6 ملحق. مجالات الثقة للفروق والمجاميع

إذا كانت S_1 وS_2 إحصائيتا معاينة لها توزيع يقترب من التوزيع الطبيعي، والعينتان مستقلتان، تكتب حدود الثقة للفروق بين المعالم التي تمثلها الإحصائيتين كما يلي:

$$S_1-S_2\ \pm\ z_c\cdot\sigma_{S_1-S_2}=S_1-S_2\ \pm\ z_c\cdot\sqrt{\sigma^2{}_{S_1}+\sigma^2{}_{S_2}}$$

في حالة المجموع :

$$S_1+S_2\ \pm\ z_c\cdot\sigma_{S_1-S_2}=S_1+S_2\ \pm\ z_c\cdot\sqrt{\sigma^2{}_{S_1}+\sigma^2{}_{S_2}}$$

مثال: إذا كانت الإحصائيتان هما متوسطا عينتين مستقلتين، مسحوبتين من مجتمعين غير محدودين، نحدد مجال الثقة للفرق (و للمجموع) بين متوسطي المجتمعين $\mu_1-\mu_2$ كما يلي :

$$m_1-m_2\ \pm\ z_c\cdot\sigma_{m_1-m_2}=m_1-m_2\ \pm\ z_c\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

مثال 2: إذا كانت الإحصائيتان هما نسبتان في عينتين مستقلتين، مسحوبتان من مجتمعين غير محدودين :

$$p'_1-p'_2\ \pm\ z_c\cdot\sigma_{p'_1-p'_2}=p'_1-p'_2\ \pm\ z_c\cdot\sqrt{\frac{pq_1}{n_1}+\frac{pq_2}{n_2}}$$

ثالثا: طرق تأسيس المقدر

أحد الطرق لاختيار مقدر معلمة ما للمجتمع أن نأخذ مباشرة نظيرتها في العينة، وإذا كان هذا المقدر لا يتصف بالخصائص المطلوبة نجري عليه تعديلا (استخدام S'^2 بدلا من S^2 لتقدير σ^2). توجد طرق أخرى لتحديد المقدر الأنسب منها طريقة المعقولية العظمى والتي تدعى أيضا طريقة الاحتمال الأكبر والتي تنسب إلى العالم فيشر وكذا طريقة العزوم.

1 طريقة العزوم

ليكن المطلوب تقدير عدد K من معالم المجتمع : $\theta_1, \theta_2, \ldots, \theta_k$. نكون جملة معادلات عددها K. تتضمن كل معادلة مساواة العزم المرتبط بالأصل من الدرجة k لمتغيرة المجتمع X : $\mu'_k = E(X^k)$، بنظيره لمتغيرة المعاينة x :

$$m'_k = (1/n)\sum_i x_i^k \qquad k = 1, 2, \quad , K$$

مثال: ليكن $X \sim B(20; p)$. تقدير p بطريقة العزوم انطلاقا من عينة يتم كما يلي:

لدينا عدد المعالم المراد تقديرها $K = 1$ إذا نحتاج إلى معادلة واحدة : $\mu = 20p$.ومنه $p = 20/\mu$،

نأخذ إذا كمقدر لـp القيمة: p' ونحسبها كما يلي : $p' = m/20$.

في حالة تقدير معلمتين للمجتمع نحتاج أن نستعمل جملة المعادلتين: $\mu = m \quad , \quad \mu'_2 = m'_2$

مثال2 : لتكن $X \sim N(\mu; \sigma^2)$. نسحب عينة ذات متوسط m، وتباين S^2. لتقدير μ وσ^2 نحتاج إلى حل جملة المعادلتين:

$$\begin{cases}\mu'_2 = m'_2 \\ \mu = m\end{cases} \text{ or } \begin{cases}\mu'_2 = \mu^2 + \sigma^2 \\ m'_2 = m^2 + S^2\end{cases} \quad \text{la solution est} \quad \begin{cases}\hat{\mu} = m \\ \hat{\sigma}^2 = S^2\end{cases}$$

هذه الطريقة قد تعطي مقدرات متحيزة كما في هذه الحالة.

2 طريقة المعقولية العظمى (طريقة الاحتمال الأكبر)

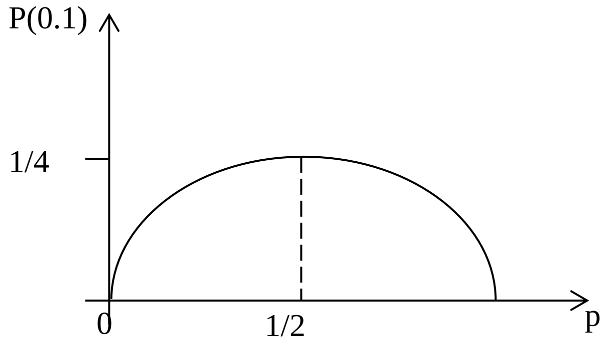
حالة كون متغيرة المجتمع متقطعة : نريد تقدير معلمة θ واحدة للمجتمع، ولدينا عينة غير نفادية (المتغيرات التي تمثل قيم المحصل عليها في العينة مستقلة) لها نفس التوزيع للمجتمع. من البديهي أن احتمال تحقق عينة بذاتها مرتبط ب قيمة المعلمة المجهولة : $P(x_1, x_2, \ldots, x_n) = L(\theta)$. هناك قيمة ل θ تعظم احتمال الحصول على العينة المحصل عليها، ونفترض أن تلك القيمة هي الصحيحة بما أن العينة حصلت بالفعل. تتمثل طريقة المعقولية العظمى في البحث عن هذه القيمة. أي البحث عن θ التي تعظم $L(\theta)$، حيث :

$$L(\theta) = f(x_1, \ldots, x_n ; \theta) = f(x_1) \cdot f(x_2) \ldots f(x_n).$$

تعتمد طريقة المعقولية العظمى على تعظيم دالة الاحتمال المشتركة $L(\theta)$.

مثال: ليكن $X \sim B(p)$، حيث النجاح هو وجود الخاصية " أ " لدى فرد مسحوب عشوائيا من المجتمع. نرد تقدير p من خلال عينة حجمها 2. ما هي القيمة p' ل p التي تجعل النتيجة 1، 0 هي الأكثر احتمالا؟ أي ما هي p' التي تجعل $p(0.1) = pq$ أكبر ما يمكن؟

من الواضح أن أكبر قيمة ل$p(0.1)$ هي ¼ والقيمة التي تحققها هي $p' = 1/2$ ، وبهذا نجيب على التساؤل.



رسم 2 أقصى قيمة ل P(0,1)

في المحور السابق تناولنا كيفية تقدير معالم المجتمع من خلال بيانات العينة وبعض خصائص المقدر الجيد. في هذا المحور سنتناول كيفية اختبار فرضيات موضوعة حول معالم مجتمع أو أكثر. يحتاج الدارس أحيانا في مرحلة ما من بحثه إلى اختبار فرضية أو أكثر بخصوص المجتمع المدروس. من أمثلة ذلك: اختبار فرضية بخصوص معدل الدخل في منطقة معينة، اختبار فرضية نسبة شفاء لدواء معين، ... يتم ذلك بصياغة فرضية عن المجتمع المدروس (أو المجتمعات المدروسة) ومن ثم محاولة الحصول على دليل إحصائي ينفي أو يثبت هذه الفرضية وذلك من خلال بيانات عينة (أو أكثر) عشوائية بسيطة. تخص الفرضية أحد معالم المجتمع كالمتوسط، النسبة أو التباين، ونعتمد في إثباتها أو رفضها على خصائص إحصائية معاينة مختارة. من أجل ذلك يعتمد هذا الدرس، كما هو الحال بالنسبة لدرس التقدير، على درس المعاينة.

أولا: اختبار المتوسط

يتناول هذا الاختبار متوسط المجتمع (μ)، مثل متوسط الدخل، متوسط وزن منتج معين، .. ويؤكد اختبار المتوسط فرضية مساواته لقيمة ما μ_0. و للقيام بالاختبار نستخرج عينة عشوائية نحسب فيها المتوسط m ثم نستخدم التوزيع الاحتمالي ل m لقياس قرب أو بعد هذه القيمة من μ_0.

1 اختبار ثنائي الاتجاه للمتوسط.

لنتناول هذا المثال: نريد اختبار فرضية حول متوسط دخل الطالب في السنة الأولى من تخرجه، ولتكن القيمة الافتراضية هي 15000دج كمتوسط للدخل الشهري. نحتاج إلى الخطوات التالية: تحديد الفرضيات، تحديد قاعدة القرار، حساب القيمة الجدولية للمتغيرة، حساب القيمة الفعلية للمتغيرة، اتخاذ القرار.

(أ) **تحديد الفرضيات (الصفرية والبديلة):**

$$H_0 : \mu = \mu_0 \ \boxed{\leftrightarrow} \ H_1 : \mu \neq \mu_0$$

تسمى الفرضية H_0 **الفرضية الصفرية** أو فرضية العدم، ويؤدي الاختبار إما إلى رفضها ونكتب RH_o وفي هذه الحالة نقبل **الفرضية البديلة** أو المعاكسة أو عدم رفضها ونكتب $R'H_0$. μ_0 هي القيمة الافتراضية ل μ وهي في هذه الحالة 15000 لذلك نكتب الفرضيات كما يلي:

$$H_0 : \mu = 15000 \ \boxed{\leftrightarrow} \ H_1 : \mu \neq 15000$$

عادة ما تكون μ_0 محددة بناءا على بيانات عينة عشوائية بسيطة ($\mu_0 = m$)، وفي هذه الحالة يمكن استخدام الخاصية

$m \sim N(\mu, \sigma^2/n)$ لاجراء الاختبار، حيث أنه تحت H_0 فإن : $m \sim N(\mu_0 , \sigma^2/n)$

مما يعني معلومية احتمال أن يكون m قريب إلى درجة ما من μ_0 فمثلا :

$$P(\mu_0 - 1.64(\sigma_m) \leq m \leq \mu_0 + 1.64(\sigma_m)) = 0.90$$

$$P(\mu_0 - 1.96(\sigma_m) \leq m \leq \mu_0 + 1.96(\sigma_m)) = 0.95$$

$$P(\mu_0 - 2.58(\sigma_m) \leq m \leq \mu_0 + 2.58(\sigma_m)) = 0.99$$

وبصفة عامة نكتب:

$$P[\mu_0 - z_{1-\alpha/2}(\sigma_m) \leq m \leq \mu_0 + z_{1-\alpha/2}(\sigma_m)] = 1-\alpha$$

أو حسب الكتابة الأكثر شيوعا:

$$P(-z_{1-\alpha/2} \leq \frac{m-\mu_0}{\sigma_m} \leq z_{1-\alpha/2}) = 1-\alpha$$

حيث:

- $(m - \mu_0)/\sigma_m$: (**متغيرة القرار**) هي المتغيرة المعيارية ل m ونرمز لها ب z_c، حيث $z \sim N(0, 1)$.
- σ_m تحدد كما يلي: $\sigma_m = \sigma/\sqrt{n}$ في حالة المعاينة بالإرجاع (أو $n \leq 0.05N$) و $\sigma_m = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ في الحالة المعاكسة.
- $1 - \alpha/2$: المساحة على يسار z .
- n : حجم العينة.

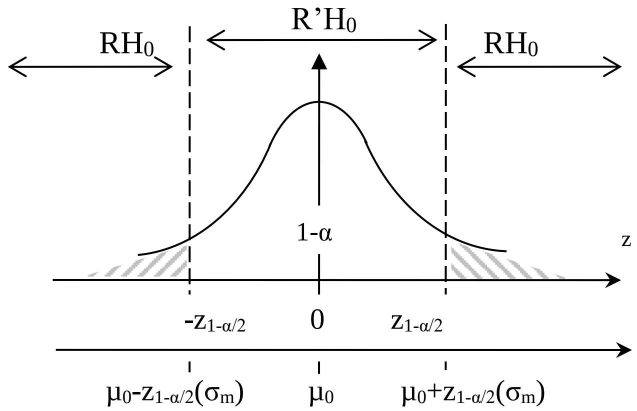
يمكن إذا كان m خارج المجال $1-\alpha$، أن نرفض الفرضية الصفرية التي حدد على أساسها هذا المجال ونقبل بالتالي الفرضية البديلة.

تسمى هذه (الخطة) قاعدة القرار.

(ب) تحديد قاعدة القرار

تكتب قاعدة القرار في المثال الذي بين أيدينا، وهي قاعدة **اختبار ثنائي الاتجاه** (أنظر الشكل 1) ، كما يلي:

$$\begin{cases} RH_0 \ \ si \ \left|\dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}}\right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{ sinon.} \end{cases} \qquad \text{أو} \qquad \begin{cases} RH_0 \ \ si \ z_c = \dfrac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} \notin \left[-z_{1-\alpha/2};\ z_{1-\alpha/2}\right] . \\ \overline{R}H_0 \ \text{ sinon.} \end{cases}$$



رسم 3 منطقتي القبول و الرفض في حالة قاعد القرار الثنائية

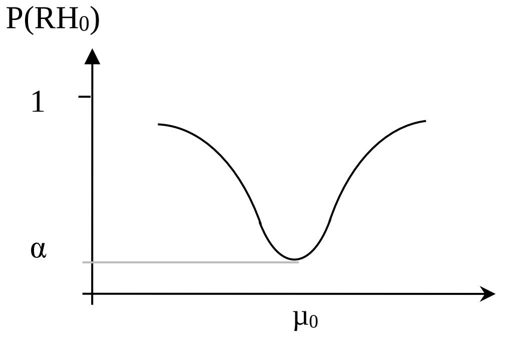
تتضـمن هذه الخطة مخاطرة تتمثل في الوصـول إلى قرار خاطيء: فقد تكون الفرضـية H_0 صـحيحة بينما تقودنا قيمة m المحصـــلة إلى رفضــها، ويســـمى هذا **الخطأ من النوع الأول**، واحتماله α ، ويكتب : $P(RH_0 / H_0) = \alpha$،

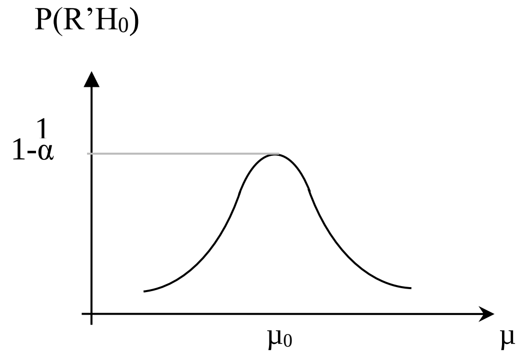
و قد تقودنا قيمة m إلى قبول H_0 فيما هي خاطئة، ويسمى هذا **الخطأ من النوع الثاني** واحتماله $1-\alpha$ ويكتب :

$$P(R'H_0 / H_1) = 1-\alpha$$

و يمكن تقليص احتمال أحد الخطأين على حساب الثاني، ولكن لا يمكن تقليص احتمال كلا الخطأين معا إلا بزيادة حجم العينة.

و يقيس احتمال رفض الفرضية الصفرية $P(RH_0)$ قوة الاختبار (أنظر الشكل 2) فيما يقيس احتمال قبولها $P(R'H_0)$ فعالية الاختبار (أنظر الشكل 2). ويتوقف كلا الاحتمالين على القيمة الحقيقية ل μ .





رسم5 (1) منحنى الفعالية — منحنى القوة (2)

(ج) حساب z الجدولية:

ويرمز لها ب z_t حيث، وهي المشار إليها في قاعدة القرار (الشكل الثاني)، وفي حالتنا (اختبار ثنائي بمستوى معنوية 5%) :

$$z_t = z_{1-\alpha/2} = z_{1-0.05/2} = z_{1-0\cdot025} = z_{0.975}$$

ومن الجدول نجد أن $z_{0.975} = 1.96$.

(د) حساب z الفعلية:

ويرمز لها ب z_c ,وهي المتغير المعيارية ل m (أنظر قاعدة القرار الشكل الأول) :

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{15800 - 1500}{1500 / \sqrt{100}} = 5.33$$

(هـ) القرار:

نقرر قبول أو رفض H_0 حسب قاعدة القرار. وفي حالتنا نرفض H_0 لأن $z_c > z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج.

2 الاختبار أحادي الاتجاه للمتوسط.

يتميز الاختبار الثنائي عن الأحادي في الفرضية البديلة التي هي عدم مساواة في الاختبار الثنائي وأكبر تماما أو أصغر تماما (حسب الحالة) في الاختبار الأحادي، وهذا يترتب عليه تغيير في قاعدة القرار.

(أ) الاختبار أحادي الاتجاه من اليمين.

لنرجع إلى المثال السابق مع تغيير محدد هو أننا نريد اختبار ما إذاكان متوسط الدخل للخريج 15000دج أم أكثر (اختبار من اليمين).

أ- **الفرضيات** : $H_0 : \mu = \mu_0 \boxed{\leftrightarrow} H_1 : \mu > \mu_0$

في هذه الحالة $\mu_0 = 15000$ لذلك نكتب : $H_0 : \boxed{\leftrightarrow} H_1 : \mu > 15000$

$\mu = 15000$

ب- **قاعدة القرار:**

$$\begin{cases} RH_0 \; si \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} > z_{1-\alpha} \; . \\ \overline{R}H_0 \; \text{sinon.} \end{cases}$$

ج- **حساب z الجدولية:** (اختبار على اليمين بمستوى معنوية 5%) : $z_t = z_{1-\alpha} = z_{1-0.05} = z_{0.95}$

ومن الجدول نجد أن $z_{0.95} = 1.645$

د- **حساب z الفعلية:**

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{15800 - 1500}{1500 / \sqrt{100}} = 5.33$$

هـ- **القرار:** نرفض H_0 لأن $z_c > z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج وإنما هو أكبر.

(ب) الاختبار أحادي الاتجاه من اليسار

نعود إلى مثالنا ونفترض أن متوسط العينة كان 14200دج ونريد أن نختبر ما إذا كان متوسط الدخل مساوي أم أقل من 15000دج.

أ- **الفرضيات** : $H_0 : \mu = 15000 \boxed{\leftrightarrow} H_1 : \mu < 15000$

ب- **قاعدة القرار:**

$$\begin{cases} RH_0 \; si \dfrac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} < -z_{1-\alpha} \; . \\ \overline{R}H_0 \; \text{sinon.} \end{cases}$$

ج- **حساب z الجدولية:** (اختبار على اليسار بمستوى معنوية 5 %) :

$z_t = -z_{1-\alpha} = -z_{1-0.05} = -z_{0.95} = -1.645$

د- **حساب z الفعلية:**

$$z_c = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{14200 - 1500}{1500 / \sqrt{100}} = -5.33$$

هـ- **القرار:** نرفض H_0 لأن $z_c < z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف أقل من 15000دج .

3 استخدام S كمقدر ل σ في اختبار المتوسط.

في الأمثلة السابقة افترضنا أن σ معلوم، في الواقع غالبا ما يكون الانحراف المعياري مجهولا ونحتاج بالتالي إلى استخدام الانحراف المعياري للعينة (S) عند حساب σ_m (أنظر درس التقدير)، حيث نعوض العبارة

$$\sigma m = \sigma/\sqrt{n} \quad \text{ب} \quad \sigma_m = \frac{S}{\sqrt{n-1}} \quad \text{أو} \quad \sigma_m = \frac{S'}{\sqrt{n}}$$

مثال: في المثال السابق نفترض أن الانحراف المعياري للدخل الشهري للطالب مجهول، لكن الانحراف المعياري للعينة $S = 1600$. كيف يمكن اختبار ما إذا كان الدخل الشهري أقل من 15000دج؟

الخطوات أ، ب ، ج تبقى بدون تغيير.

د- **حساب z الفعلية**: $z_c = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} = \frac{14200-1500}{1600/\sqrt{99}} = -4.97$

هـ- **القرار**: نرفض H_0 لأن $z_c < z_t$ ونقبل H_1 أي أن متوسط دخل الخريج حديث التوظيف ليس 15000دج وإنما هو أقل.

4 استخدام التوزيع t في اختبار المتوسط.

في حالة $n < 30$ و σ (الانحراف المعياري للمجتمع) مجهولا، لا يمكن استخدام التوزيع الطبيعي، ولكن لدينا :

$$\frac{\overline{X}-\mu}{S/\sqrt{n-1}} \sim t_{n-1} \quad \text{و تحت } H_0 \ (\mu = \mu_0) : \quad \frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} \sim t_{n-1}$$

يمكن إذا استخدام التوزيع ستيودنت (بشرط أن يكون توزيع المجتمع طبيعيا أو على الأقل جرسي الشكل) .

و تتغير قاعدة القرار تبعا لهذا التغيير فتكتب في حالة الاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left|\frac{\overline{X}-\mu_0}{S/\sqrt{n-1}}\right| > t_{n-1;\,1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:

$$\begin{cases} RH_0 \ si \frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} > t_{n-1,\,1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:

$$\begin{cases} RH_0 \ si \frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} < -t_{n-1;\,1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

5 خلاصة

يتم اختبار الفرضيات من خلال 5 خطوات متتالية وهي:

- تحديد الفرضيات (الصفرية والبديلة)
- تحديد قاعدة القرار
- حساب القيمة الجدولية للمتغيرة
- حساب القيمة الفعلية للمتغيرة
- اتخاذ القرار.

تتحدد كيفية إتمام كل خطوة حسب طبيعة الاختبار (ثنائي أو أحادي الاتجاه)، حسب طبيعة المجتمع و طبيعة و حجم العينة، ... و تسخدم في ذلك نظريات توزيع المعاينة.

ثانيا: اختبار النسبة واختبار التباين

1 اختبار النسبة

يتعلق هذا الاختبار بنسبة مفردات المجتمع التي تتصف بخاصية ما (p)، حيث يؤكد الاختبار أو ينفي صحة فرضية معينة بخصوص قيمة p . يرمز للقيمة الافتراضية ب p_0 وتكتب الفرضية كما يلي:

$$H_0 : p = p_0$$

للقيام بالاختبار نستخدم خصائص p' النسبة في العينة (أنظر توزيع المعاينة للنسبة : نظرية 6).

$$E(p') = \mu_{p'} = p \qquad ; \qquad \sigma^2{}_{p'} = \frac{pq}{n}$$

عند $n \geq 30$: $p' \approx N\ (p, \sigma_{p'})$ (نظرية موافر- لابلاس)

استنادا إلى هذه الخصائص وتحت H0 : $p' \approx N\left(p_0 ; \sqrt{\frac{p_0 q_0}{n}}\right)$

و من ثم يمكن تحديد قاعدة القرار بحسب طبيعة الاختباركما يلي:

في حالة الاختبار الثنائي:
$$\begin{cases} RH_0 \ si \left| \dfrac{p'-p_0}{\sqrt{p_0 q_0 / n}} \right| > z_{1-\alpha/2} \ . \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليمين:
$$\begin{cases} RH_0 \ si \ \dfrac{p'-p_0}{\sqrt{p_0 q_0 / n}} > z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

في حالة اختبار من اليسار:
$$\begin{cases} RH_0 \ si \ \dfrac{p'-p_0}{\sqrt{p_0 q_0 / n}} < -z_{1-\alpha} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

مثال: تقدر الدوائر الرسمية نسبة المتخرجين الجامعيين الذين يحصلون على عمل في السنة الأولى التي تلي تخرجهم ب 70 % . وجدت دراسة أجريت على عينة من 900 طالب أن نسبة الحصول على عمل 67 % . كيف يمكن اختبار ما إذا كانت النسبة الرسمية صحيحة أم مبالغ فيها، بمستوى معنوية 5 %.

$$H_0 : p = 0.70 \ \boxed{\leftrightarrow} \ H_1 : p < 0.70$$

$$\frac{p'-p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.67-0.7}{\sqrt{0.7(0.3)/900}} \cong -196.34 < -z_{1-0.05} = -1.64$$

ومنه نرفض الفرضية H_0 .

2 اختبار التباين

لاختبار صدقية فرضية بخصوص قيمة تباين مجتمع ما،

$$H_0 : \sigma^2 = \sigma_0^2 \ \boxed{\leftrightarrow} \ H_1 : \sigma^2 = \sigma_0^2$$

نستعمل المقدر غير المنحاز $\hat{S}^2 = \dfrac{\sum_i (X_i - \overline{X})^2}{n-1}$. حيث في حالة العينة الكبيرة ($n \geq 50$ في أحسن الأحوال) ، وتحت H_0 فإن $\mu_4 = E(X-\mu)^4$
$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \approx N(0.1).$$

حيث μ_4 هو العزم المركزي من الدرجة الرابعة. وبهذا الشكل تكتب قاعدة القرار للاختبار الثنائي كما يلي:

$$\begin{cases} RH_0 \ si \left| \dfrac{\hat{S}^2 - \sigma_0^2}{\sqrt{(\mu_4 - \sigma_0^4)/n}} \right| > z_{1-\alpha/2} \\ \overline{R}H_0 \ \text{sinon.} \end{cases}$$

وفي حالة μ_4 مجهول يمكن استخدام كمقدر : $m_4 = E(xi - m)^4$.

وإذا كان المجتمع طبيعيا، حيث $\mu_4 = 3\sigma^4$ ، فإن **متغيرة القرار** يمكن أن تكتب كما يلي:

$$T = \frac{\hat{S}^2 - \sigma_0^2}{\sigma_0^2\sqrt{2/n}} \approx N(0.1).$$

ثالثا: اختبار المقارنة بين مجتمعين

يتناول هذا الاختبار مقارنة بين مجتمعين من خلال المتوسط أو التباين لكل منهما ... وسنركز هنا على متغيرة القرار، إذ من السهل على الطالب استنتاج كيفية إتمام الخطوات الأخرى على ضوء ما سبق.

1 اختبار تساوي متوسطي مجتمعين

الغرض من الاختبار هو تأكيد أو نفي تساوي متوسطي مجتمعين من خلال عينتين عشوائيتين بسيطتين مستقلتين. تكتب الفرضيات (في حالة الاختبار الثنائي)كما يلي: $H_0 : \mu_1 = \mu_2 \ \boxed{\leftrightarrow} \ H_1 : \mu_1 \neq \mu_2$

لتحديد متغيرة القرار نعتمد في الاختبار على متغيرة القرار T أو T' بحسب الحالة (نترك للطالب استنتاج قاعدة القرار)، حيث نميز بين حالة كون تباينا المجتمعين معلومين وحالة كون تباينا المجتمعين مجهولين.

(أ) تباينا المجتمعين معلومين

1- المجتمعين طبيعيين:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0.1)$$

2- مجتمعين ما ($n_1 , n_2 \geq 30$):

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx N(0,1)$$

(ب) تباينا المجتمعين مجهولين

1- المجتمعان طبيعيان:

إذا كان تباينا المجتمعين متساويين

$$T'=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim t_{n_1+n_2-2}$$

2- مجتمعين ما $(n_1\ ,\ n_2\geq 30)$:

$$T=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{S_1^2}{n_1-1}+\frac{S_2^2}{n_2-1}}}\approx N(0,1)$$

مثال: نسحب من مجتمعين طبيعيين متساويي التباين عينتين حجم الأولى 18 وحجم الثانية 21. وجدنا النتائج التالية:

$m_1 = 81,\ m_2 = 76,\ S^2_1 = 9,\ S^2_2 = 8.$ كيف يمكن اجراء اختبار تساوي متوسطي المجتمعين بمستوى معنوية 5 % .

$$H_0 : \mu_1 = \mu_2 \ \boxed{\leftrightarrow}\ H_1 : \mu_1 \neq \mu_2$$

$$T'=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{81-76}{\sqrt{\frac{18(9)+21(8)}{18+21-2}\left(\frac{1}{18}+\frac{1}{21}\right)}}\cong 5.43$$

$$t_{0.975;37}\cong 2.336<5.43 \qquad \Rightarrow RH_0$$

2 اختبار تساوي تبايني مجتمعين

الغرض من الاختبار هو تأكيد أو نفي تساوي تباينا مجتمعين من خلال عينتين عشوائيتين بسيطتين مستقلتين.

تكتب الفرضيات (في حالة الاختبار الثنائي)كما يلي: $H_0 : \sigma^2_1 = \sigma^2_2 \ \boxed{\leftrightarrow}\ H_1 : \sigma^2_1 \neq \sigma^2_2$

نعتمد في الاختبار على متغيرة القرار T أو T' بحسب الحالة، حيث نميز بين حالة كون المجتمعين طبيعيين أم غير ذلك.

(أ) مجتمعين طبيعيين

1- الحالة العامة:

$$T'=\frac{\hat{S}_1^2}{\hat{S}_2^2}\sim F_{n_1-1;\ n_2-1}$$

2- في حالة $n_1\ ,\ n_2\geq 30$

$$T=\frac{1}{2}\ln\left(\frac{\hat{S}_1^2}{\hat{S}_2^2}\right)/\sqrt{\frac{1}{2}\left(\frac{1}{n_1-1}+\frac{1}{n_2-1}\right)}\approx N(0;1)$$

(ب) مجتمعين ما $(n_1, n_2 \geq 50)$

1– $\mu_4^{(1)}$; $\mu_4^{(2)}$ معروفين :

$$T = (\hat{S}_1^2 - \hat{S}_2^2) \,/\, \sqrt{\frac{\mu_4^{(1)} - \hat{S}_1^4}{n_1} - \frac{\mu_4^{(2)} - \hat{S}_2^4}{n_2}} \approx N(0;1)$$

2– في حالة $\mu_4^{(1)}$; $\mu_4^{(2)}$ غير معروفين : نعوض μ_4 ب m_4 .

مثال : نسحب من مجتمعين طبيعيين عينتين حجم الأولى 18 وحجم الثانية 21. وجدنا النتائج التالية: $m_1 = 81, m_2 = 76, S^2_1 = 9, S^2_2 = 8.$ كيف يمكن اجراء اختبار تساوي متوسطي المجتمعين بمستوى معنوية 5 % ؟

$$H_0 : \sigma^2_1 = \sigma^2_2 \ \boxed{\leftrightarrow}\ H_1 : \sigma^2_1 \neq \sigma^2_2$$

$S'^2_1 = S^2_1 * n_1 / (n_1-1) = 9\,(18)/17 \approx 9.53$; $S'^2_2 = S^2_2 * n_2 / (n_2-1) = 8\,(21)/20 = 8.4$

$S'^2_1 / S'^2_2 \approx 1.135$; $F_{0.05\,;17\,;20} \approx 2.17$

$$T < F_{\alpha\ ;\, n1-1\ ;n2-1} \Rightarrow R'H_0$$

رابعا: اختبار الاستقلال والتجانس

1 اختبار التجانس

لنعد إلى اختبار النسبة، ونفترض أن لدينا عددا k من الخصائص المتنافية، نسبة تحقق كل منها في المجتمع p_i حيث $\sum p_i = 1$. نريد اختبار فرضية تساوي هذه النسب:

$$H_0 : p_i = p_{i0} \ \ (i = 1, 2, \ldots k)\ \boxed{\leftrightarrow}\ H_1 : p_i \neq p_{i0}$$

(الفرضية البديلة هي أن إحدى النسب النظرية p_{i0} على الأقل غير مساوية للقيمة الحقيقية.)

متغيرة القرار : لإنجاز الاختبار نستخرج عينة نحسب فيها عدد مرات تحقق الخصائص (n_i) . إذا تحققت الشروط

$n \geq 30$ ، $p_{i0} \geq 1$ وعلى الأقل في 80 % من الحالات $np_{i0} \geq 5$ نبرهن أن :

$$T = \sum_i \frac{(n_i - np_{i0})^2}{np_{i0}} \approx \chi^2_{k-1}$$

2 اختبار التعديل

تستخدم هذه الطريقة أيضا لاختبار تعديل توزيع معين بتوزيع آخر، وفي هذه الحالة نقارن بين تكرارات العينة (التكرارات الحقيقية) n_i وتكرارات افتراضية n_{i0} ، حيث تصاغ الفرضيات كما يلي :

$H_0: n_i = n_{i0}$ $(i = 1, 2, \dots k)$ $\boxed{\leftrightarrow}$ H_1: على الأقل إحدى التكرارات النظرية n_{i0} غير مساوية للتكرار الحقيقي n_i

$$T = \sum_i \frac{(n_i - n_{i0})^2}{n_{i0}} \approx \chi^2_{k-m-1}$$

m عدد من معالم من المجتمع المقدرة انطلاقا من بيانات العينة لتحديد التكرارات النظرية.

مثال: يتقدم إلى انتخابات معينة 3 مرشحين: أ، ب وج. نريد اختبار فرضية بمستوى معنوية 5 % حول شعبيتهم كما يلي:

$$H_0 : p_1 = 0.4,\ p_2 = 0.35,\ p_3 = 25$$

أجري استجواب ل 400 ناخب فكان توزع فئات المساندين على التوالي : 170، 135، 95 .

لدينا $n = 400 \geq 30$، الأعداد الافتراضية $np_{i0} = 160, 140, 100 \geq 1$ ، وأكثر من 80 % من $np_{i0} \geq 5$.

$$T = \sum_i \frac{(n_i - np_{i0})^2}{np_{i0}} = \frac{(170-160)^2}{160} + \frac{(135-140)^2}{140} + \frac{(95-100)^2}{100} = 1.05$$

$X^2_{2\,;\,0.95} = 5.99 > 1.05 \Rightarrow R'H_0$.

المصادر والمراجع :

#	عنوان المرجع	المؤلف	دار النشر و السنة
1	مقدمة في نظرية الاحتمالات	جبار عبد مضجي	دار المسيرة للنشر والتوزيع - 2011
	الإحصاء للإداريين والاقتصاديين	دلال القاضي وآخرون	دار الحامد – 2005
3	مقدمة في الإحصاء الرياضي	جبار عبد مضجي	دار المسيرة للنشر والتوزيع - 2015
	اساليب الاحصاء التطبيقي	حسن ياسين طعمة	دار صفاء للنشر والتوزيع - 2009
5	الاحصاء الاستدلالي من الجانب النظري و التطبيقي منهج كمي	بوعظم كمال	دار زهران للنشر والتوزيع - 2012
	الاحتمالات : النظرية و التطبيق	مجدي الطويل	دار النشر للجامعات - 2000
7	نظرية الاحتمالات	عبد الحميد ربيع غيطان	دار الكتب الأكاديمية - 2004
	الإحصاء والاحتمالات النظرية والتطبيق	**علي عبد السلام العماري**	**دار الحكمة ليبيا – 2013**
9	مبادئ في الاحتمالات والإحصاء	**مبارك اسبر ديب**	**جامعة تشرين -سوريا - 2009**
	Probabilités : Exercices corrigés avec rappels de cours	**Adil ELMARHOUM Mohamed DIOURI**	**la collection Sciences et Techniques – Maroc 2014**