الفصل السادس

التوزيع الطبيعي

٦ـ١ مقدمة : رأينا في الفقرة (٤ـ٤) أن المتحولات العشوائية المستمرة تولد فراغ عينة مستمراً تكون نقاطه متراصة إلى بعضها كنقاط محور موجه وبالتالي فإنها بالإضافة إلى كونها لا نهائية في عددها غير قابلة للعد . وكأمثلة تقليدية لمتحولات عشوائية مستمرة نذكر أطوال وأوزان البشر ، أخطاء القياسات في تجربة مخبرية ، عمر مصباح كهربائي الخ . كما رأينا في تلك الفقرة أنه للحصول على نموذج احتمالي ، في مثل هذه الحالة ، نبدأ باختيار منحن مستمر يمثل ما سميناه بتابع الكثافة الاحتمالية . ولا بد أن يحقق مثل هذا التابع ، ولنرمز له بـ $f(x)$ شرطين :

١ ـ $f(x) \geqslant 0$ من أجل جميع قيم x .

٢ ـ المساحة تحت $f(x)$ تساوي الواحد تماماً .

وعندئذ يكون احتمال أي حادثة عددية عبارة عن مساحة تحت منحني الكثافة هذا . وكنتيجة نجد أن احتمال أن يفترض المتحول x قيمة معينة a ، مثلاً ، أي $p(x=a)$ هو المساحة تحت المنحني فوق النقطة a من محور الفواصل وهي صفر . وهكذا فإن مثل هذا الحل لمشكلة إيجاد نموذج احتمالي لفراغ عينة مستمر يحتم علينا القول بأن احتمال أن يكون لمتحول عشوائي مستمر قيمة معينة هو احتمال يساوي الصفر . وهذا تعبير واقعي عن استحالة توصل الإنسان إلى أجهزة قياس دقيقة بصورة مطلقة . ولذلك تبقى مثل هذه النتيجة مقبولة طالما بقي الإنسان غير قادر على الإدعاء بأن قياساته للأطوال أو الأوزان أو لعمر مصباح

كهربائي أو لنتيجة تجربة يقوم بها في المخبر الخ . هي قياسات لا تخضع لأي خطأ على الإطلاق إذ مهما أوتي جهاز القياس من الدقة ومهما بلغت مهارة الإنسان الذي يستخدم الجهاز فلا بد من ارتكاب خطأ مهما كان صغيراً .

وبينما تتخذ منحنيات الكثافة أشكالاً مختلفة نلاحظ أن عدداً كبيراً من المتحولات العشوائية التي نواجهها في الطبيعة لها منحن تكراري أو منحن كثافة له تقريباً شكل الجرس أو ، كما نعبر عن ذلك إحصائياً، له بصورة تقريبية شكل منحني التكرار الطبيعي أو شكل التوزيع الطبيعي .

ونعرف رياضياً تابع الكثافة الاحتمالية الطبيعي على الشكل :

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

وهي معادلة منحن له شكل الجرس المبين في الشكل (٦ـ١) . حيث :

π = عددثابت يساوي تقريباً 3.1416

e = عدد ثابت يساوي تقريباً 2.7183 .

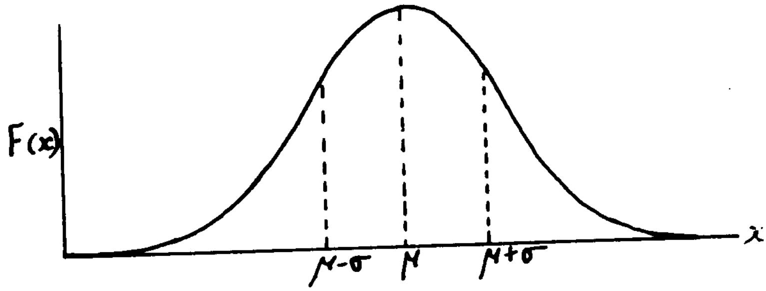
σ = وسيط يساوي الانحراف المعياري للتوزيع ، وللمنحنيات الطبيعية المختلفة انحرافات معيارية مختلفة إلا أنه من أجل منحن معين يبقى σ ثابتاً .

μ = وسيط يساوي متوسط التوزيع ، وللمنحنيات الطبيعية المختلفة متوسطات مختلفة ، إلا أن μ في منحن معين تبقى ثابتة .

ويُبرهن في الحساب التكاملي أن المساحة تحت المنحني $e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ تساوي $\sigma\sqrt{2\pi}$ تماماً وبالتالي فإن المساحة تحت المنحني الطبيعي $f(x)$ كما عرفناه أعلاه تساوي الواحد تماماً .

ونلاحظ أن المنحني متناظر حول النقطة μ وأن النقطة $x = \mu$ هي النقطة التي يتمركز عندها التوزيع وينتشر على جانبيها بصورة متناظرة . وأن التابع يبلغ نهايته العظمى عند النقطة $x = \mu$ ويعاني انعطافاً عند النقطتين $\mu - \sigma$ و $\mu + \sigma$ كما نلاحظ أنه من أجل قيم صغيرة لِ σ يكون انتشار المنحني على جانبي μ

محدوداً بينما ينتشر إلى مسافات أبعد ويأخذ شكلاً أكثر إنبساطاً (المساحة تحت المنحني في جميع الأحوال يجب أن تبقى مساوية للواحد) عندما يكون σ كبيراً .



شكل ٦ـ١ منحني الكثافة الاحتمالية الطبيعي

ونادراً ما نجد في الواقع العملي متحولات تمتد قيمها بين « اللانهاية السالبة » و« اللانهاية الموجبة » ، أي من أقصى المحور الحقيقي الموجه على اليسار إلى أقصى هذا المحور على اليمين . ومن المؤكد أن أطوال وأوزان البشر ، أو عمر مصباح كهربائي لا يحقق ذلك . ومع هذا فإن مضلعات التكرار النسبي المرسومة لأنواع كثيرة من القياسات ستنتج شكلاً يشبه شكل الجرس والذي يمكن اعتباره صورة تقريبية للمنحني المرسوم في الشكل (٦ـ١) . وتلقي نظرية النهاية المركزية وهي أهم نظرية في الإحصاء ضوءاً على هذه الظاهرة المفيدة جداً .

٦ـ٢ **نظرية النهاية المركزية** : تعرض نظرية النهاية المركزية ، وتحت شروط عامة جداً ، أن كلاً من مجموع ومتوسط عينة عشوائية مسحوبة من مجتمع ما يمتلك ، عند تكرار هذه العينات عدداً كبيراً من المرات ، توزيعاً له على وجه التقريب شكل الجرس . وربما كان من الأفضل إيضاح هذه العبارة بمثال .

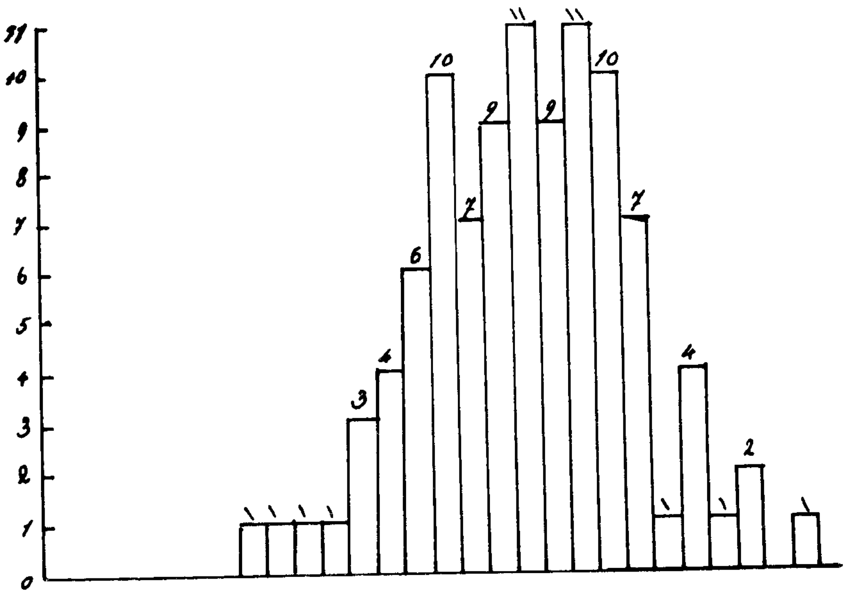
لنعتبر المجتمع المتولد عن قذف قطعة زهر عدداً كبيراً جداً من المرات وقد رأينا توزيعه الاحتمالي في الشكل (٤ـ٢) . لنسحب عينة من خمسة قياسات $n = 5$ من المجتمع وذلك بقذف قطعة الزهر خمس مرات وتسجيل الملاحظات الخمس الناتجة . ثم نحسب مجموع هذه الملاحظات الخمس Σx_i ومتوسطها $\bar{x}$

جدول ٦ـ١ عينات من مجتمع قذف قطعة الزهر

رقم العينة	قياسات العينة	$\sum x_i$	$\bar{x}$	رقم العينة	قياسات العينة	$\sum x_i$	$\bar{x}$
1	3, 5, 1, 3, 2	14	2.8	51	2, 3, 5, 3, 2	15	3.0
2	3, 1, 1, 4, 6	15	3.0	52	1, 1, 1, 2, 4	9	1.8
3	1, 3, 1, 6, 1	12	2.4	53	2, 6, 3, 4, 5	20	4.0
4	4, 5, 3, 3, 2	17	3.4	54	1, 2, 2, 1, 1	7	1.4
5	3, 1, 3, 5, 2	14	2.8	55	2, 4, 4, 6, 2	18	3.6
6	2, 4, 4, 2, 4	16	3.2	56	3, 2, 5, 4, 5	19	3.8
7	4, 2, 5, 5, 3	19	3.8	57	2, 4, 2, 4, 5	17	3.4
8	3, 5, 5, 5, 5	23	4.6	58	5, 5, 4, 3, 2	19	3.8
9	6, 5, 5, 1, 6	23	4.6	59	5, 4, 4, 6, 3	22	4.4
10	5, 1, 6, 1, 6	19	3.8	60	3, 2, 5, 3, 1	14	2.8
11	1, 1, 1, 5, 3	11	2.2	61	2, 1, 4, 1, 3	11	2.2
12	3, 4, 2, 4, 4	17	3.4	62	4, 1, 1, 5, 2	13	2.6
13	2, 6, 1, 5, 4	18	3.6	63	2, 3, 1, 2, 3	11	2.2
14	6, 3, 4, 2, 5	20	4.0	64	2, 3, 3, 2, 6	16	3.2
15	2, 6, 2, 1, 5	16	3.2	65	4, 3, 5, 2, 6	20	4.0
16	1, 5, 1, 2, 5	14	2.8	66	3, 1, 3, 3, 4	14	2.8
17	3, 5, 1, 1, 2	12	2.4	67	4, 6, 1, 3, 6	20	4.0
18	3, 2, 4, 3, 5	17	3.4	68	2, 4, 6, 6, 3	21	4.2
19	5, 1, 6, 3, 1	16	3.2	69	4, 1, 6, 5, 5	21	4.2
20	1, 6, 4, 4, 1	16	3.2	70	6, 6, 6, 4, 5	27	5.4
21	6, 4, 2, 3, 5	20	4.0	71	2, 2, 5, 6, 3	18	3.6
22	1, 3, 5, 4, 1	14	2.8	72	6, 6, 6, 1, 6	25	5.0
23	2, 6, 5, 2, 6	21	4.2	73	4, 4, 4, 3, 1	16	3.2
24	3, 5, 1, 3, 5	17	3.4	74	4, 4, 5, 4, 2	19	3.8
25	5, 2, 4, 4, 3	18	3.6	75	4, 5, 4, 1, 4	18	3.6
26	6, 1, 1, 1, 6	15	3.0	76	5, 3, 2, 3, 4	17	3.4
27	1, 4, 1, 2, 6	14	2.8	77	1, 3, 3, 1, 5	13	2.6
28	3, 1, 2, 1, 5	12	2.4	78	4, 1, 5, 5, 3	18	3.6
29	1, 5, 5, 4, 5	20	4.0	79	4, 5, 6, 5, 4	24	4.8
30	4, 5, 3, 5, 2	19	3.8	80	1, 5, 3, 4, 2	15	3.0
31	4, 1, 6, 1, 1	13	2.6	81	4, 3, 4, 6, 3	20	4.0
32	3, 6, 4, 1, 2	16	3.2	82	5, 4, 2, 1, 6	18	3.6
33	3, 5, 5, 2, 2	17	3.4	83	1, 3, 2, 2, 5	13	2.6
34	1, 1, 5, 6, 3	16	3.2	84	5, 4, 1, 4, 6	20	4.0
35	2, 6, 1, 6, 2	17	3.4	85	2, 4, 2, 5, 5	18	3.6
36	2, 4, 3, 1, 3	13	2.6	86	1, 6, 3, 1, 6	17	3.4
37	1, 5, 1, 5, 2	14	2.8	87	2, 2, 4, 3, 2	13	2.6
38	6, 6, 5, 3, 3	23	4.6	88	4, 4, 5, 4, 4	21	4.2
39	3, 3, 5, 2, 1	14	2.8	89	2, 5, 4, 3, 4	18	3.6
40	2, 6, 6, 6, 5	25	5.0	90	5, 1, 6, 4, 3	19	3.8
41	5, 5, 2, 3, 4	19	3.8	91	5, 2, 5, 6, 3	21	4.2
42	6, 4, 1, 6, 2	19	3.8	92	8, 2, 1, 2, 1	14	2.8
43	2, 5, 3, 1, 4	15	3.0	93	6, 3, 1, 5, 2	17	3.4
44	4, 2, 3, 2, 1	12	2.4	94	1, 3, 6, 4, 2	16	3.2
45	4, 4, 5, 4, 4	21	4.2	95	6, 1, 4, 2, 2	15	3.0
46	5, 4, 5, 5, 4	23	4.6	96	1, 1, 2, 3, 1	8	1.6
47	6, 6, 6, 2, 1	21	4.2	97	6, 2, 5, 1, 6	20	4.0
48	2, 1, 5, 5, 4	17	3.4	98	3, 1, 1, 4, 1	10	2.0
49	6, 4, 3, 1, 5	19	3.8	99	5, 2, 1, 6, 1	15	3.0
50	4, 4, 4, 4, 4	20	4.0	100	2, 4, 3, 4, 6	19	3.8

ويبين الجدول (٦ـ١) نتائج تكرار هذه العملية مائة مرة . كما يبين الشكل (٦ـ٢) المضلع التكراري للقيم المائة لِ $\bar{x}$ (أو $\sum x_i$) . وسيلاحظ القارىء النتيجة الهامة التالية :

وهي أنه بالرغم من أن التوزيع الاحتمالي لِ x له شكل أفقي تماماً إلا أن المضلع التكراري لمائة من قيم $\bar{x}$ (وهو يقدم صورة أولية عن شكل التوزيع الاحتمالي للمتحول $\bar{x}$ أو للمتحول $\sum x_i$) يتخذ شكلاً مقبباً قريباً من شكل الجرس وكلما زدنا حجم العينات المسحوبة عن خمسة كلما اعتدل شكل المضلع التكراري ليقترب أكثر فأكثر من شكل التوزيع الطبيعي . وبعبارة أخرى لو أخذنا في مثالنا أي لو قذفنا قطعة الزهر عشر مرات بدلاً من خمس ثم سجلنا نتائج مائة من هذا الحجم $(n = 10)$ ورسمنا المضلع التكراري للقيم المائة لِ $\bar{x}$ فإننا سنجد شكلاً أكثر قرباً من شكل الجرس . ولا بد من ملاحظة أنه للحصول على فكرة



شكل ٦ـ٢ المضلع التكراري لمتوسطات العينات المائة المسحوبة من مجتمع قذف قطعة الزهر

أدق عن شكل التوزيع الاحتمالي لِـ $\bar{x}$ نحتاج نظرياً إلى عدد لا نهائي من العينات أو لنقل بصورة عملية اننا نحتاج إلى عدد من العينات أكبر بكثير من المائة التي تضمنتها تجربتنا هنا . ومع هذا فإن الشكل الذي تقدمه العينات المائة كاف لتوضيح الفكرة الأساسية التي تتضمنها نظرية النهاية المركزية والتي نعرضها في العبارة التالية :

نظرية النهاية المركزية : إذا سحبنا عينات عشوائية حجم كل منها n من مجتمع متوسطه μ وانحرافه المعياري σ محدودان فإن توزيع متوسط العينة $\bar{x}$ يكون من أجل قيم كبيرة لِـ n مطابقاً تقريباً للتوزيع الطبيعي بمتوسط يساوي μ وانحراف معياري $\sigma/\sqrt{n}$ وستزداد دقة التقريب كلما ازداد n .

ويمكن إعادة صياغة النظرية لتتفق مع $\sum_{i=1}^{n} x_i$ بدلاً من $\bar{x}$. أي أن توزيع $\sum_{i=1}^{n} x_i$ ينزع أيضاً إلى أن يصبح طبيعياً بمتوسط يساوي $n\mu$ وانحراف معياري يساوي $\sqrt{n}\,\sigma$ وذلك عندما تصبح n كبيرة .

وتبدو أهمية نظرية النهاية المركزية من زاويتين فهي توضح أولاً نزوع العديد من المتحولات العشوائية لأن يكون توزيعها بصورة تقريبية هو التوزيع الطبيعي . إذ يمكن أن نتصور طول الانسان مركباً من عدد من العناصر العشوائية مثل طول الأم ، طول الأب ، نشاط الغدة أو الغدد التي لها علاقة بالطول ، البيئة أو المحيط بأنواعه ، التغذية الخ . وإذا كانت آثار هذه العوامل تنضاف إلى بعضها لتنتج واقعاً معيناً بالنسبة لطول الإنسان فعندئذ يمكن اعتبار الطول كمجموع لعدد من المتحولات العشوائية وهكذا تفعل نظرية النهاية المركزية فعلها ويكون توزيع الطول هو على وجه التقريب التوزيع الطبيعي وذلك بصرف النظر عن توزيعات المركبات التي تشكل المجموع . وهذا بالطبع محاولة للتعليل ليس أكثر إذ أن ما يجري في الحقيقة غير معروف لنا بصورة دقيقة ولكن ما يمكن قوله على أي حال هو أن نظرية النهاية المركزية توضح سبب وجود العديد من المتحولات العشوائية التي نصادفها في الطبيعة والتي نعتبر أن توزيعها

الاحتمالي هو التوزيع الطبيعي .

ومن زاوية أخرى نجد أن العطاء الأكثر أهمية لنظرية النهاية المركزية يتعلق بمسألة الاستقراء الإحصائي . فالعديد من الاحصاءات التي تلعب دور « التقدير » أو « صانع القرار » والتي تُستخدم للقيام باستقراءات حول وسطاء المجتمع (مثل p في التوزيع الثنائي ، μ في التوزيع الطبيعي الخ) هذه الاحصاءات تأخذ شكل مجموع لقياسات العينة أو متوسط هذه القياسات . وإذا كانت الحال كذلك وكانت n كبيرة بكفاية فيمكننا اعتبار التوزيع الطبيعي تقريباً جيداً للتوزيع الاحتمالي لهذا التقدير أو الإحصاء . ويمكن عندها استخدام القاعدة التجريبية التي ذكرناها في الفصل الثاني بالنسبة لإحصاء كهذا . وسنجد في الفقرات القادمة والفصول القادمة العديد من الاستخدامات المفيدة للغاية ، في هذا المضمار ، لنظرية النهاية المركزية .

والناحية التي تقلق هنا هو السؤال التالي : كم يجب أن يكون كبر حجم العينة n حتى يصبح التقريب الناشىء عن تطبيق نظرية النهاية المركزية تقريباً جيداً من وجهة النظر العملية . ولسوء الحظ فإنه لا يوجد جواب عام ومحدد تماماً لهذا السؤال ، فالأمر يتعلق بالتوزيع الاحتمالي الموافق للمجتمع الذي جاءت منه العينة والغاية التي سنستخدم التقريب من أجلها وهكذا . وغالباً ما يكون لكل حالة حكمها ، معتمدين بصورة رئيسية على الخبرة السابقة والتجربة . ونشعر بشيء من الراحة عندما ننظر إلى مثال قذف قطعة الزهر المذكور أعلاه فقد لاحظنا أن المضلع التكراري للقيم المائة لِ $\bar{x}$ قريب من شكل الجرس بالرغم من أن حجم العينة الذي استخدمناه n لم يتعد الخمسة وبالرغم من أن توزيع المجتمع الذي تأتي منه العينات هو خط أفقي وبعيد جداً عن شكل الجرس . وفي العديد من الحالات تفعل نظرية النهاية المركزية فعلها حتى من أجل عينات صغيرة الحجم .

٦ـ٣ ـ العينات العشوائية : أشرنا فيما سبق إلى تعابير مثل « عينة تمثيلية

للمجتمع » وأخذ العينة « بطريقة عشوائية » و« عينات عشوائية » ولم نحاول إعطاء تعريف لهذه التعابير . ويلاحظ القارىء أن تطبيق نظرية النهاية المركزية كما ذكرناها أعلاه رهن بكون العينة عشوائية . فما هي العينة العشوائية ومن أين تكتسب أهميتها بالنسبة للاستقراء الإحصائي ؟

ونعود بالقارىء ثانية إلى مثال قذف قطعة زهر المذكور في الفقرة (٣-١) حيث رغبنا في استقراء ما إذا كانت قطعة الزهر المقذوفة متوازنة أم لا ؟ ونعود أيضاً إلى أسلوب اتخاذ قرار بالنسبة لقبول أو رفض وسقة من البضاعة الواردة إلى مصنع ، وإلى اختبار فرضية تتعلق بفعالية لقاح ضد الزكام . ففي كل حالة سحبنا عينة من المجتمع للقيام باستقراء معين وهي في هذه الحالات اتخاذ قرار .

وإذا لاحظنا بعد الحصول على العينة أنها من النوع الغير محتمل (احتمال الحصول عليها تحت الفرضية الابتدائية هو احتمال بسيط) نستنتج أن الفرضية الابتدائية غير مقبولة ونرفضها . وإذا وجدنا أن العينة محتملة جداً نستنتج أن الفرضية الابتدائية مقبولة ولا نرفضها . وبعبارة أخرى فإنه لا بد لنا من معرفة احتمال الحصول على عينة كتلك التي لاحظناها لكي نصل إلى استقراء احصائي . ونكرر هنا عبارة ذكرناها في الفصل الثالث وهي أن سير الاحتمال ينطلق من المجتمع إلى العينة بينما يعكس الاحصاء ، على الوجه الآخر ، الاتجاه فيستخدم الاحتمال كأداة للقيام باستقراء حول المجتمع معتمداً على المعلومات التي تحويها عينة . ويجب أن يكون واضحاً أن لطريقة أخذ العينة أثرها في حساب احتمال الحصول على عينة محددة .

لنفرض الآن أننا سحبنا عينة من n من القياسات من مجتمع يحوي N قياساً . فكم هو عدد العينات التي يمكن الحصول عليها أو عدد الأوجه التي كان يمكن أن تكون عليها العينة الموجودة بين أيدينا ؟ إن هذا العدد هو عدد متوافقات N شيئاً مأخوذ n منها في وقت واحد أي : $\binom{N}{n} = \frac{N!}{n!(N-n)!}$

وإذا حصلنا على العينة بطريقة يكون معها لكل من هذه العينات نفس الاحتمال في أن تكون هي العينة الملحوظة ، فعندئذ نقول أن الطريقة عشوائية وأن العينة الناتجة هي عينة عشوائية . وفي العديد من الحالات يكون المجتمع ذهنياً كما في حالة القيام بقياس في تجربة مخبرية . وهنا ننظر إلى المجتمع على أنه العدد اللانهائي من القياسات التي كان سيمكن الحصول عليها لو قمنا بتكرار التجربة مرة بعد أخرى . وإذا رغبنا بعينة من 10 قياسات من هذا المجتمع فإننا ببساطة نكتفي بالقيام بعشر تكرارات للتجربة . وفي هذه الحالة نحرص على أن تكون هذه التكرارات مستقلة عن بعضها البعض . وفي الحقيقة تمثل كل ملاحظة أو قياس متحولاً عشوائياً بحد ذاتها إلا أن لكل منها نفس التوزيع الاحتمالي باعتبارها مسحوبة من نفس المجتمع وكون العينة عشوائية يعني أن القياسات التي تحويها العينة مستقلة عن بعضها البعض .

وليس من الضروري أن تكون العينة بكاملها عشوائية ففي تصميم التجارب حيث نحرص بصورة عامة على الحصول على أكبر قدر من المعلومات من حجم للعينة محدد سلفاً ، وفي الطرق المتبعة لأخذ عينة حيث نحرص على أن تمثل العينة المجتمع الذي يأتي منه أفضل تمثيل ، نضطر لتجاوز العشوائية بصورة جزئية فتكون العينة مأخوذة وفق خطة مرسومة سلفاً في أجزاء منها وفي أجزاء أخرى تكون عشوائية .

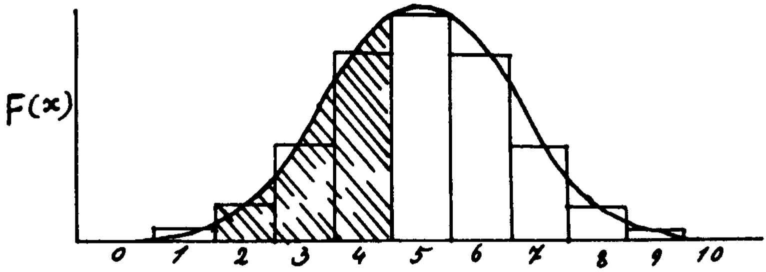
٦ـ٤ تقريب التوزيع الثنائي بالتوزيع الطبيعي : رأينا في الفصل السابق عدة تطبيقات للتوزيع الثنائي اقتضت جميعها حساب احتمال أن يأخذ x ، وهو عدد النجاحات من بين n تكراراً ، قيمة معينة أو يقع ضمن مجال معين . وقد اقتصرنا على أمثلة تكون فيها n صغيرة بسبب المشقة التي يتضمنها حساب $p(x)$ عندما تكون n كبيرة . ولنفرض الآن أننا نحتاج إلى حساب احتمال وقوع x ضمن مجال معين من أجل $n = 1000$ فمع أن مثل هذا العمل ليس مستحيلاً إلا أنه ممتنع إلى الحد الذي نريد تجنبه . وتقدم نظرية النهاية المركزية حلاً لهذه

المشكلة ذلك لأنه يمكن النظر إلى عدد النجاحات x من بين n تكراراً كمجموع يحقق شروط نظرية النهاية المركزية . فإذا اصطلحنا على أن يوافق النتيجة S (أو النجاح) العدد 1 ويوافق النتيجة F (أو الفشل) العدد صفر . فعندئذ تكون نتيجة التكرارات الـ n متوالية من الأرقام وكل واحد منها صفر أو واحد أي متوالية من النوع $x_1, x_2, \ldots, x_n$ حيث تأخذ كل x_i إما القيمة 1 أو القيمة صفر . وهكذا يمكن التعبير عن عدد النجاحات x كمجموع لِ n من المتحولات x_i أي $x = \sum_{i=1}^{n} x_i$ حيث يتوزع كل x_i وفق التوزيع الثنائي الموافق للحالة الخاصة التي يكون فيها $n = 1$ ، $P(1) = p$ و $P(0) = q$.

ويسمى مثل هذا التوزيع عادة بالتوزيع الثنائي النقطي . وهكذا تصبح التكرارات الـ n المستقلة عينة عشوائية من مجتمع التوزيع الثنائي النقطي و $x = \sum_{i=1}^{n} x_i$ هو مجموع قيم العينة . ووفقاً لنظرية النهاية المركزية يكون التوزيع التقريبي لـ x من أجل n كبيرة بكفاية هو التوزيع الطبيعي بمتوسط يساوي np وتشتت يساوي npq . وبالتالي يمكن استخدام المساحات تحت منحني الكثافة الطبيعي هذا لحساب احتمالات تتعلق بالمتحول x ولكن بصورة تقريبية .

وعلى سبيل المثال ، لنعتبر التوزيع الثنائي من أجل $n = 10$ ، $p = \frac{1}{2}$ فعندئذ $\mu = np = 5$ و $\sigma = \sqrt{npq} = 1.58$ ويبين الشكل (٦-٣) الاحتمال الموافق لحادثة معينة وذلك عند استخدام كل من التوزيع الثنائي والتوزيع الطبيعي كما تحدده نظرية النهاية المركزية . ونظرة أولية للشكل تبين أن التقريب جيد تماماً حتى في حالة $n = 10$ ويسهم تناظر التوزيع الثنائي في حالة $p = \frac{1}{2}$، في كون التقريب على هذه الدرجة من الجودة حتى من أجل قيمة صغيرة لِ n .

واحتمال أن يكون x مساوياً لِ 2 ، 3 ، أو 4 يساوي تماماً مساحة المستطيلات الثلاثة المقامة فوق $x = 2$ ، $x = 3$ ، و $x = 4$. ويمكن تقريب هذه المساحة تحت المنحني الطبيعي من $x = 1.5$ إلى $x = 4.5$ وهي المساحة المظللة في الشكل (٦-٣) . ويقدم جدول التوزيع الطبيعي في الملحق المساحات تحت



شكل ٦ـ٣ مقارنة التوزيع الثنائي بالتوزيع الطبيعي في حالة $n = 10$ ، $P = \frac{1}{2}$

المنحني الطبيعي . وسنناقش في الفقرة القادمة استخدام هذا الجدول .

وعندما تكون n صغيرة و p قريبة من الصفر أو الواحد فإن شكل المضلع الاحتمالي سيكون منحازاً (أي يتجمع معظمه) إلى جانب القيمة $x = 0$ أو $x = n$ على الترتيب . أي أنه سيكون بعيداً جداً عن وضع التناظر . وفي مثل هذه الحالات سيكون التقريب سيئاً . وبصورة عامة فإنه كلما ابتعدت p عن القيمة $\frac{1}{2}$ كلما ابتعد شكل المضلع الاحتمالي للتوزيع الثنائي عن التناظر .

وإذا تذكرنا القاعدة التجريبية في الفصل الثاني فإن 95% تقريباً من القياسات من مجتمع طبيعي ستقع ضمن إنحرافين معياريين على جانبي المتوسط (أي ضمن المجال $\mu \pm 2\sigma$) وجميع القياسات تقريباً ستقع ضمن المجال $\mu \pm 3\sigma$. ومن الصعب أن يكون التوزيع الثنائي متناظراً إذا كانت القياسات من المجتمع الثنائي منتشرة إلى مدى انحرافين معياريين على جانبي المتوسط . ولتحديد متى يكون التقريب باستخدام التوزيع الطبيعي مناسباً ، نحسب إذن $\mu = np$ و $\sigma = \sqrt{npq}$ فإذا وقع المجال $\mu \pm 2\sigma$ ضمن مدى التوزيع الثنائي أي بين 0 و n ، فسيكون التقريب جيداً .

٦ـ٥ **المساحات تحت منحني الكثافة الطبيعي** : نلاحظ أن معادلة منحني الكثافة الطبيعي ، كما عرفناه في الفقرة (٦ـ١) ، تحوي وسيطين μ و σ. وبإعطاء قيم مختلفة لكل من هذين الوسيطين ، نحصل على منحنيات مختلفة ، أي أن معادلة المنحني كما عرفناه في الفقرة (٦ـ١) لا تمثل منحنياً واحداً وإنما

أسرة من المنحنيات لا نهاية لعددها . ووضع جدول للمساحات تحت كل من هذه المنحنيات غير ممكن . وسنجد أنه يمكن وضع جدول واحد بحيث يكفي لحساب المساحات تحت أي منحني كثافة طبيعي . وأسهل طريقة لتحقيق ذلك هو أن نحسب المساحات الواقعة ضمن عدد محدد من الانحرافات المعيارية على جانبي المتوسط ، أي بنفس الطريقة التي عبرنا فيها عن القاعدة التجريبية في الفصل الثاني .

وبما أن المنحني متناظر فيمكن التبسيط بإقامة جدول للمساحات بين المتوسط μ والنقاط x الواقعة إلى يمين μ. فإذا فرضنا نقطة x أكبر من μ فإن المسافة بين x و μ هي $x-\mu$ وإذا عبرنا عنها بدلالة الانحراف المعياري σ ولنفرض أنها تساوي z مرة الانحراف المعياري σ فيمكننا أن نكتب $x-\mu = z\sigma$ وإذا قسنا المسافات على محور الفواصل بوحدة قياس تساوي σ (وعندها يكون $\sigma = 1$ حكماً) فإن القياس $x-\mu$ مقيساً بالوحدة الجديدة يصبح $\frac{x-\mu}{\sigma}$ أي يساوي z . وهكذا نكتب المتحول الجديد z بدلالة المتحول القديم x على الشكل :

$$Z = \frac{x-\mu}{\sigma} \qquad (2)$$

ونلاحظ أنه يوافق كل قيمة لِـ x قيمة واحدة لِـ z والعكس بالعكس . ومن أجل $z = 0$ يكون $x = \mu$. ويمكن البرهان على أن معادلة منحني الكثافة للمتحول الجديد z هي :

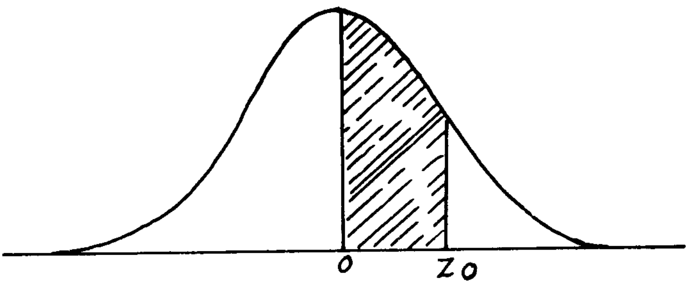
$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} , \quad -\infty < Z < +\infty \qquad (3)$$

وهذه المعادلة خالية من μ و σ . ويدعى التوزيع الطبيعي الموافق لتابع الكثافة هذا بالتوزيع الطبيعي المعياري . وللاختصار يدعى منحني الكثافة نفسه بمنحنى التوزيع الطبيعي المعياري . ويبين الشكل (٦ــ٤) التمثيل البياني لهذا المنحني . كما يقدم جدول التوزيع الطبيعي في الملحق المساحات تحت هذا المنحني بين $z = 0$ و $z = z_0$ ، ونقصد المساحة المظللة في الشكل (٦ــ٤) حيث z_0 أي نقطة إلى

يمين المبدأ $z = 0$. ومن تناظر المنحني وكون المساحة تحته مساوية للواحد فإن المساحة إلى يمين $z = 0$ هي 0.5 والمساحة إلى يسار $z = 0$ هي 0.5 . وبالإشارة إلى جدول التوزيع الطبيعي في الملحق نلاحظ أن العمود الأول يحوي قيم z بفاصل يساوي $\frac{1}{10}$ من قيمة إلى القيمة التي تليها . والمنزلة العشرية الثانية من قيمة z معطاة في الصف الأفقي الأول من الجدول . والرقم الموجود في صلب الجدول والموافق لقيمة معينة لـ z بمنزلتين عشريتين (أي بفاصل يساوي $\frac{1}{100}$ بين قيمة والقيمة التي تليها) هو المساحة تحت المنحني بين $z = 0$ وهذه القيمة . وهكذا فإن المساحة بين المبدأ $z = 0$ و $z = .70$ مثلاً موجودة في العمود الثاني من الجدول حذاء القيمة 0.7 وتساوي $.2580$. وبصورة مشابهة فإن المساحة بين $z = 0$ و $z = .84$ هي العدد من العمود السادس (أي العمود الموافق لِ $.04$) المحاذي لِ 0.8 من قيم z في العمود الأول . أما المساحة بين $z = 0$ و $z = 1.0$ فهي $.3413$. والمساحة الواقعة ضمن إنحراف معياري واحد على جانبي μ هي $2\,(.3413) = .6826$. والمساحة الواقعة ضمن انحرافين معياريين على جانبي μ هي ضعف المساحة التي يعطيها الجدول بين $z = 0$ و $z = 2.0$ أي $2\,(.4772) = .9544$. وهذه الأعداد تذكرنا بالنسب التي تعطيها القاعدة التجريبية المذكورة في الفصل الثاني . ونختم هذه الفقرة ببعض الأمثلة .

مثال ٦ـ١ أحسب قيمة z ولتكن z_0 بحيث يقع (وإلى أربعة أرقام عشرية) $.95$. تماماً من المساحة ضمن $\pm z_0$ من الانحرافات المعيارية عن المتوسط .

سيقع نصف المساحة 0.95 إلى يسار المتوسط ونصفها إلى يمين المتوسط لأن التوزيع الطبيعي متناظر . أي نريد القيمة z_0 الموافقة لمساحة $.475$. ويقع الرقم $.475$. في الصف الموافق لِ $z = 1.9$ والعمود $.06$. ومنــه يكون $z_0 = 1.96$. وهي قيمة قريبة جداً من القيمة 2 التي استخدمناها في القاعدة التجريبية .

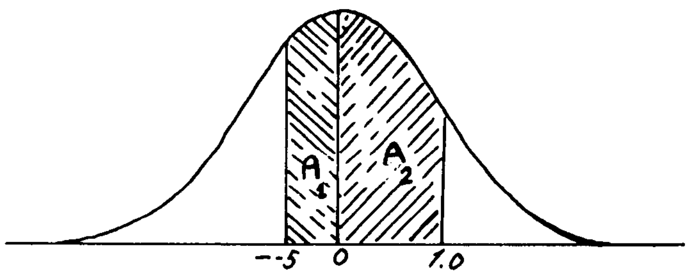


شكل ٦ـ٤ توزيع طبيعي معياري

مثال ٦ـ٢ أحسب المساحة بين5. - = zو 1.0 = z كما هو مبين في الشكل (٦ـ٥) .

المساحة المطلوبة تساوي مجموع المساحتين A_1 و A_2 المبينتين في الشكل . ومن جدول التوزيع الطبيعي نجد أن $A_2 = .3413$. والمساحة A_1 تساوي المساحة بين 0 = zو 5. = z أو $A_1 = .1915$. وهكذا يكون مجموع المساحتين :

$$A = A_1 + A_2 = .1915 + .3413 = .5328.$$



شكل ٦ـ٥ المساحة تحت المنحني الطبيعي في المثال ٦ـ٢

مثال ٦ـ٣ : ليكن x متحولاً عشوائياً يتوزع وفق التوزيع الطبيعي بمتوسط يساوي 2 وتشتت يساوي 16 والمطلوب حساب احتمالات الحوادث العددية التالية : $P(-1 < x < 3\ 5)$ $P(x > 1)$, $P(x < 3)$

$$P(x<3) = P\left(\frac{x-\mu}{\sigma} < \frac{3-\mu}{\sigma}\right) = P\left(z < \frac{3-2}{4}\right) = P(z < -\ .25)$$

$$= P(z > .25) = .5 - P(0 < Z < .25)$$

ومن جدول التوزيع الطبيعي نجد أن $P(0 < z < .25) = .0987$ وبالتالي :

$$P(x < 3) = .5 - .0987 = .4013$$

$$P(x > 1) = P\left(\frac{x-\mu}{\sigma} > \frac{1-\mu}{\sigma}\right) = P\left(z > \frac{1-2}{4}\right) = P(z > -.25)$$

$$= P(-.25 < z < 0) + P(z > 0) = P(0 < z < .25) + .5$$

$$= .0987 + .5 = .5987.$$

$$P(-1 < x < 3.5) = P\left(\frac{-1-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{3.5-\mu}{\sigma}\right) = P(-.75 < Z < .375)$$

$$= P(-.75 < z < 0) + P(0 < z < .375)$$

$$= P(0 < z < .75) + P(0 < z < .375)$$

ومن جدول التوزيع الطبيعي نجد أن $P(0 < z < .75) = .2734$ ولحساب $P(0 < z < .375)$ نأخذ القيمة الواقعة في منتصف الطريق بين $P(0 < z < .37)$ و $P(0 < z < .38)$ أي بين .1443 .. و.1480 وهي القيمة .1462

ومنه يكون :

$$P(-1 < x < 3.5) = .2734 + .1462 = .4196$$

مثال ٦ـ٤ : ليكن توزيع x هو التوزيع الطبيعي بمتوسط يساوي 10 وتشتت يساوي 4 إحسب احتمال وقوع x بين 11 و 13.6 أي $P(11 < x < 13.6)$.

$$P(11 < x < 13.6) = P\left(\frac{11-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{13.6-\mu}{\sigma}\right)$$

$$= P\left(\frac{11-10}{2} < z < \frac{13.6-10}{2}\right) = P(.5 < Z < 1.8)$$

$$= P(0 < z < 1.8) - p(0 < z < .5) = .4641 - 1915 = .2726$$

ونقدم فيما يلي عدداً من الأمثلة على استخدام التقريب الطبيعي للتوزيع الثنائي :

مثال ٦ـ٥ بالإشارة إلى التجربة الثنائية الموضحة في الشكل (٦ـ٣) حيث

$n = 10$ و $p = \frac{1}{2}$. احسب احتمال أن يكون x مساوياً لِـ 2 أو 3 أو 4 بدقة تصل الرقم العشري الرابع مستخدماً جدول التوزيع الثنائي . ثم أحسب الاحتمال الموافق مستخدماً التقريب الطبيعي للتوزيع الثنائي .

الاحتمال المطلوب ولنرمز له بِـ p_1 يساوي :

$$P_1 = \sum_{x=2}^{4} P(x) = \sum_{x=0}^{4} P(x) - \sum_{x=0}^{1} P(x) = .3770 - .0108 = .3662 .$$

ولاستخدام التقريب الطبيعي يجب أن نحسب المساحة بين $x_1 = 1.5$ و $x_2 = 4.5$ حيث $\mu = np = 5$ و $\sigma = \sqrt{npq} = 1.58$. وهكذا نكتب :

$$P_2 = P(1.5 < x < 4.5) = P\left(\frac{1.5-5}{1.58} < \frac{x-\mu}{\sigma} < \frac{4.5-5}{1.58}\right)$$

$$= P(-2.22 < z < -.32)$$

$$= P(-2.22 < z < 0) - P(-.32 < z < 0)$$

$$= P(0 < z < 2.22) - P(0 < z < .32) = .4868 - .1255 = .361$$

وهذه القيمة تتفق برقمين عشريين مع القيمة الحقيقية p_1 .

مثال ٦ـ٦ مدى الثقة بقطعة الكترونية هو احتمال أن نختار واحدة من كومة إنتاج فنجدها تؤدي المهمة التي صممت من أجلها . اخترنا عينة من ألف قطعة ووجدنا من بينها $x = 27$ قطعة لا تعمل . أحسب احتمال مثل هذه النتيجة بفرض أن مدى الثقة لمثل هذه القطع هو 98 . .

إن احتمال الحصول على قطعة عاطلة هو $p = .02$ ومنه :

$$\mu = np = 1000\,(.02) = 20$$

$$\sigma = \sqrt{npq} = \sqrt{1000\,(.02)(.98)} = 4.43$$

والمطلوب حساب :

$$P = p\,(x \geqslant 27) = P\,(27) + P\,(28) + \ldots + P\,(999) + P\,(1000)$$

والتقريب الطبيعي هو المساحة تحت المنحني الطبيعي إلى يمين $x = 26.5$ (لاحظ أنه يجب استخدام $x = 26.5$ بدلاً من $x = 27$ بحيث نشمل المستطيل

الاحتمالي المقام فوق النقطة $x = 27$) ومنه نكتب :

$$P(x \geqslant 26.5) = P\left(\frac{x-\mu}{\sigma} \geqslant \frac{26.5-\mu}{\sigma}\right) = P\left(Z \geqslant \frac{26.5-20}{4.43}\right) = P(Z \geqslant 1.47)$$

$$= .5 - P(0 < z < 1.47) = .5 - .4292 = .0708$$

مثال ٦ـ٧ اختبرنا لقاحاً جديداً ضد الزكام . وقد أعطي اللقاح لمائة شخص وروقبوا من حيث إصابتهم بالزكام لمدة سنة وقد نجا 68 منهم من الإصابة بالزكام . ولنفرض أننا نعلم من معلومات سابقة أن احتمال عدم الإصابة بالزكام هي بصورة طبيعية وبدون استخدام اللقاح 5 . أية نتائج يمكنك استخلاصها من هذه التجربة حول فعالية اللقاح ؟

بترجمة المسألة إلى فرضية تتعلق بوسيط المجتمع الثنائي نجد أن المطلوب هو اختبار الفرضية بأن احتمال النجاة من الزكام هي $p = .5$. وإذا فرضنا أن محتوى اللقاح لا يزيد قابلية الإصابة بالزكام فيصبح البديل للفرضية الابتدائية هو أن نرفضها عندما يكون عدد من نجوا من الإصابة بالزكام x كبيراً . لنحسب الآن وباستخدام التقريب الطبيعي للتوزيع الثنائي احتمال الحصول على 68 أو أكثر على أساس الفرض بأن النجاة من الزكام تمت بصورتها الطبيعة وبدون تأثير اللقاح فنجد :

$$\sigma = \sqrt{npq} = \sqrt{100(.5)(.5)} = 5$$

$$\mu = np = 100(.5) = 50$$

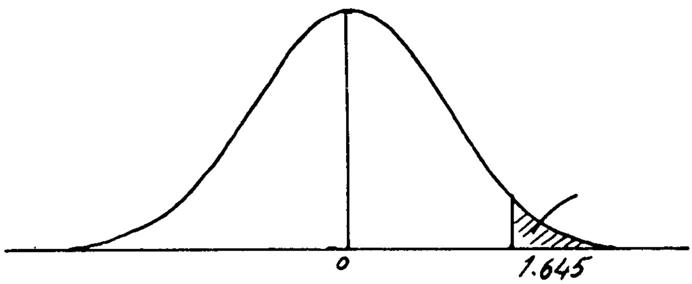
$$P(x \geqslant 68) = P\left(\frac{x-\mu}{\sigma} \geqslant \frac{68-50}{5}\right) = P(z \geqslant 3.6) = .0000$$

هذا الاحتمال هو من الصغر بحيث لا يذكره الجدول الطبيعي وهذا يدعونا إما الى الاعتقاد بأن حادثة نادرة جداً قد وقعت أو أن اللقاح مفيد بالفعل في منع الاصابة بالرشح .

ورفض الفرضية يطرح سؤالاً إضافياً . فما هو مدى فعالية اللقاح وهل هو فعال إلى المدى الذي يمكن معه إنتاجه على نطاق واسع من وجهة النظر

الاقتصادية ؟ وتقود الإجابة على الجزء الأول من السؤال إلى مسألة التقدير التي سنناقشها في الفصل القادم . أما الإجابة على الجزء الثاني من السؤال فهي عبارة عن قرار اقتصادي يتعلق بمدى الطلب على مثل هذا اللقاح وبأكلاف الانتاج والتسويق الخ .

مثال ٦ــ٨ يجري عادة تحديد α وهو احتمال الخطأ من النوع I وموقع وشكل منطقة الرفض في اختبارٍ إحصائي قبل إجراء التجربة . لنفرض أننا نرغب القيام بالتجربة المذكورة في المثال السابق واختبار الفرضية $p = .5$. فما هي منطقة الرفض المناسبة لهذا الاختبار إذا رغبنا في أن يكون α مساوياً لِ $.05$ (أنظر الشكل ٦ــ٦) .



شكل ٦ــ٦ موقع منطقة الرفض في المثال ٦ــ٨

عرضنا في المثال السابق أنه يمكن استخدام x عدد الناجين من الإصابة كإحصاء للاختبار وأن منطقة الرفض هي المنطقة الموافقة لقيم كبيرة لِ x أي الذيل الأعلى (أو الأيمن) للتوزيع الاحتمالي لِ x . وإذا اكتفينا بأن تكون α حوالي $.05$ فالمطلوب هو حساب قيمة للمتحول x ولنرمز لها بِ x_α بحيث أن :

$$P(x \geqslant x_\alpha) \doteq .05$$

حيث $\doteq$ تعني « يساوي تقريباً » . أي أننا نريد :

$$P(x < x_\alpha) = .95$$

أو :

$$P\left(\frac{x-np}{\sqrt{npq}} < \frac{x_\alpha - np}{\sqrt{npq}}\right) = P(Z < Z_\alpha) \doteq .95$$

او :

$$P(o < z < z_\alpha) = .45$$

ونلاحظ من جدول التوزيع الطبيعي أن المساحة .4495 توافق z = 1.64 و 0.4504 توافق z = 1.65 وهكذا تكون المساحة .45 موافقة من خلال قاعدة التناسب الطردي لِ z = 1.645 ، أي أن :

$$z_\alpha = 1.645$$

ومنه :

$$\frac{x_\alpha - 50}{5} = 1.645$$

أو :

$$x_\alpha = 58.225$$

ومن الواضح أن عدد من لم يصابوا بالزكام يجب أن يكون عدداً صحيحاً أي إما أن تكون نقطة البدء بالنسبة لمنطقة الرفض هي x = 58 أو x = 59 .

ولنفرض أننا تبنينا منطقة الرفض $x \geqslant 59$ فعندئذ يكون احتمال الخطأ من النوع I هو :

$$P(x \geqslant 59) = \alpha ,$$

ويمكن حساب قيمة α باستخدام التقريب الطبيعي أي المساحة الموافقة لِ $x \geqslant 58.5$ وهكذا نكتب :

$$\alpha = P(x \geqslant 58.5) = P\left(z \geqslant \frac{58.5 - 50}{5}\right) = P(z \geqslant 1.7)$$

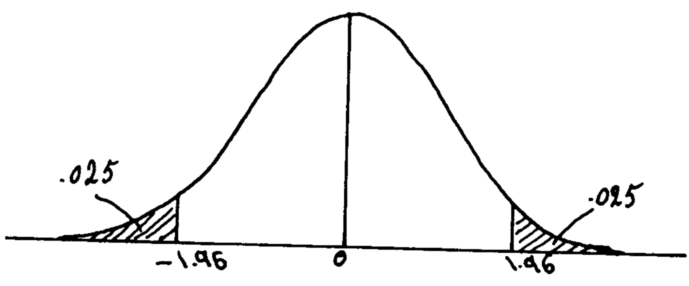
$$= .5 - P(0 < z < 1.7) = .5 - .4554 = .0446.$$

وبينما تقدم الطريقة الأولى قيمة أدق لِ α إلا أن الفروق العملية بين $\alpha = .0446$ و $\alpha = .05$ بسيطة جداً . وعندما تكون n كبيرة فإننا نوفر الجهد والوقت باستخدام z كإحصاء للاختبار بدلاً من x . وهكذا نرفض الفرضية في المثال (٦ـ٧) إذ كان z أكبر أو يساوي 1.645

مثال ٦ـ٩ تعتقد إدارة حصر التبغ أن 10% من جميع المدخنين يفضلون

تدخين اللفافة من النوع A . ولاختبار مدى صحة هذا الاعتقاد اختير 2500 من المدخنين بطريقة عشوائية وذلك من مجتمع كافة المدخنين وسئلوا عن اللفافة المفضلة لديهم . وكانت النتيجة أن $x = 218$ عبروا عن تفضيلهم للسيجارة A . فهل تقدم هذه المعلومات دليلاً كافياً للتراجع عن الاعتقاد بأن 10% من المدخنين يفضلون السيجارة A ؟ استخدم $\alpha = .05$.

الفرضية الابتدائية المطلوب اختبارها هي أن $p = .1$ حيث p هو احتمال أن يفضل مدخن السيجارة A . وذلك في مقابل الفرضية البديلة بأن p هي أكبر أو أقل من .1 ويمكن تحديد مواقع منطقة الرفض كما هو مبين في الشكل (٦ـ٧) . وسنرفض الفرضية الابتدائية إذا كان $z > 1.96$ أو $z < -1.96$ ونلاحظ هنا أننا وضعنا نصف القيمة α في كل من ذيلي المنحني الطبيعي ذلك لأننا نريد رفض الفرضية الابتدائية عندما تكون p أكبر أو أصغر من 0.1. ويدعى هذا الاختبار بالاختبار الإحصائي الثنائي الذيل ويقابلة في المثالين (٦ـ٧) و(٦ـ٨) اختبار وحيد الذيل . وذلك عندما كانت الفرضية البديلة هي أن قيمة p أكبر فقط من القيمة التي تحددها الفرضية الابتدائية .



شكل ٦ـ٧ موقع منطقة الرفض في المثال ٦ـ٩

ولدينا هنا :

$$\mu = np = (2500)(.1) = 250$$

$$\sigma = \sqrt{npq} = 15.$$

أي أن قيمة z الموافقة لِ x = 218 هي . .

$$z = \frac{218 - 250}{15} = -2.1$$

وبما أن قيمة z تقع ضمن منطقة الرفض (أصغر من 1.96 -) نرفض الفرضية الابتدائية ونستنتج أن أقل من 10% من المدخنين يفضلون السيجارة A .

ما هو احتمال أن نكون قد اتخذنا القرار الصحيح ؟ والجواب بالطبع هو 1 أو 0 في هذا المثال بالذات وذلك وفقاً لما إذا كان القرار صحيحاً أو مخطئاً . إلا أننا نعلم بأنه إذا كررنا استخدام هذا الاختبار مرة بعد أخرى في مسائل مشابهة فإن احتمال رفض الفرضية الابتدائية مع أنها صحيحة هو $\alpha = .05$ وهذا يولّد عندنا ثقة معقولة بأننا اتخذنا في هذا المثال القرار الصحيح .

٦-٦ بعض التوزيعات المستمرة المستخدمة في طرق الاحصاء : سنستعرض في هذه الفقرة وباختصار توزيعات مستخدمة على نطاق واسع في الاحصاء التطبيقي وهي (i) التوزيع χ^2 (كاي مربع) ، (ii) التوزيع t أو توزيع ستيودنت ، (iii) التوزيع F أو توزيع سنيديكور .

(i) يدعى تابع الكثافة : (4)

$$f(u) = \frac{u^{\frac{1}{2}\nu - 1}\, e^{-\frac{1}{2}u}}{(\frac{1}{2}\nu - 1)!\, 2^{\frac{1}{2}\nu}} \quad , \quad u > 0$$

بتابع كثافة التوزيع χ^2 بِـ ν درجة من الحرية أو اختصاراً التوزيعχ^2 بِـ ν درجة من الحرية . (χ و ν هما الحرفان اليونانيان « كاي » و« نو») ولنفهم مصطلح « درجة من الحرية » بأنه مجرد تعبير خاص عن الوسيط ν الوارد في عبارة التوزيع . وν بصورة عامة هي عدد صحيح موجب . ويعطي الشكل (٦-٨) تمثيلاً لهذا التوزيع من أجل عدة قيم لِ ν . ومن أجل قيم لِ ν أكبر من 6 يكون للمنحني نفس الشكل بصورة عامة كما في حالة $\nu = 6$. وعندما تزداد ν عن القيمة 6 تتناقص قيمة النهاية العظمى للمنحني وينتشر أبعد إلى اليمين بينما يقترب الذيل الأيمن من محور الفواصل بسرعة أقل .

ويقدم جدول كاي مربع في الملحق قيم u التي يقع على يمينها % α من المساحة الكلية تحت المنحني (وهي تساوي الواحد) ويقع على يسارها%$(1-\alpha)$ من المساحة . وذلك من أجل قيم متعددة لِ ν . ونذكر هنا بدون برهان أنه إذا كان توزيع المتحول z هو التوزيع الطبيعي المعياري فإن توزيع المتحول z^2 هو التوزيع χ^2 بدرجة واحدة من الحرية أي أن معادلة منحني الكثافة الاحتمالي هي المعادلة F(u) المذكورة أعلاه بعد تبديل ν بالواحد أي :

$$f(u) = \frac{1}{\sqrt{2\pi}} \, u^{-\frac{1}{2}} \, e^{-\frac{1}{2}u} \qquad u > 0 \qquad (5)$$

حيث $\sqrt{\pi} = (-\frac{1}{2})!$ كما يُبرهن في حساب التفاضل والتكامل . وإذا كانت $Z_1, Z_2, \dots, Z_n$ عينــة عشوائيــة (متحولات مستقلــة) من التوزيع الطبيعي المعياري أي بمتوسط يساوي الصفر وتشتت يساوي الواحد . فإن توزيع المتحول :

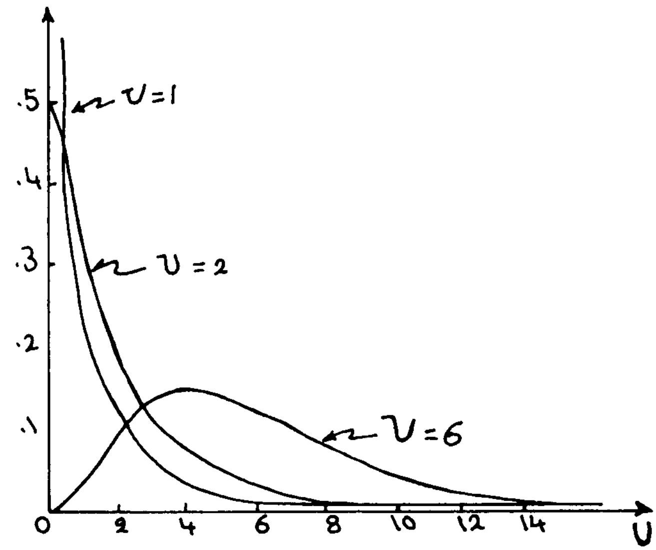
$$U = \sum_{i=1}^{n} Z_i^2 \qquad (6)$$

هو التوزيع χ^2 المذكور أعلاه بِـ n درجة من الحرية أي أننا نحصل على معادلة منحني الكثافة الاحتمالية للمتحول $\sum_{1}^{n} Z_i^2$ بوضع $n = \nu$ في معادلة F(u) المذكورة أعلاه .

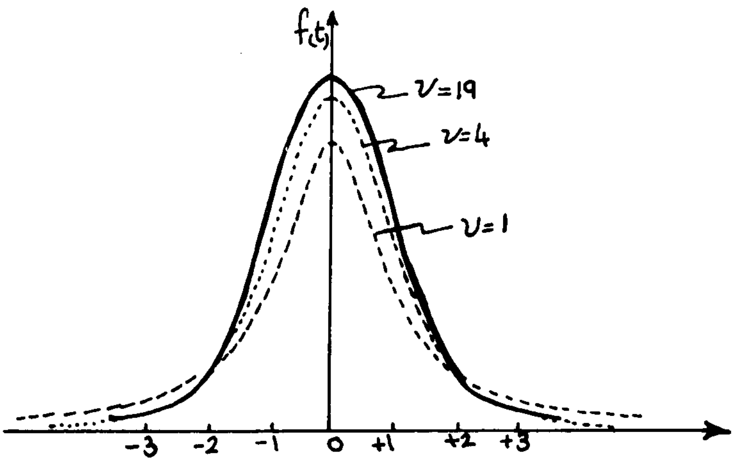
(ii) يدعى تابع الكثافة :

$$f(t) = \frac{[\frac{1}{2}(\nu-1)]!}{\sqrt{\nu\pi}\,[\frac{1}{2}(\nu-2)]!} \cdot \frac{1}{\left[1+\frac{t^2}{\nu}\right]^{\frac{1}{2}(\nu+1)}}, \quad -\infty < t < +\infty \qquad (7)$$

بتابع كثافة التوزيع t بِـ ν درجة من الحرية أو اختصاراً التوزيع t بِـ ν درجة من الحرية حيث ν عدد صحيح موجب عادة . ويقدم الشكل (٦–٩) عدة تمثيلات بيانية لهذا المنحني من أجل قيم متعددة لِ ν والمنحني كما نلاحظ متناظر حول الصفر . ويبلغ نهايته العظمى من أجل $t = 0$.



شكل ٦ـ٨ أمثلة من التوزيع χ^2 من أجل $\nu = 1,2,6$



شكل ٦ـ٩ أمثلة من التوزيع t من أجل $\nu = 1,4,19$

ويقدم جدول التوزيع t في الملحق القيمة المطلقة أو القيمة الموجبة لِـ t التي يقع على يمينها $\frac{\alpha}{2}$ % من المساحة الكلية تحت المنحني وذلك من أجل قيم مختلفة لِـ α و ν .

ويمكن البرهان على أنه إذا كان المتحول العشوائي u يتبع التوزيع χ^2 بـ ν درجة من الحرية وكان المتحول العشوائي z يتبع التوزيع الطبيعي المعياري أي بمتوسط يساوي الصفر وتشتت يساوي الواحد فإن المتحول العشوائي t وهو النسبة :

$$t = \frac{Z}{\sqrt{u/\nu}} \tag{8}$$

يتبع التوزيع t المذكور أعلاه بـ ν درجة من الحرية .

(iii) يدعى تابع الكثافة :

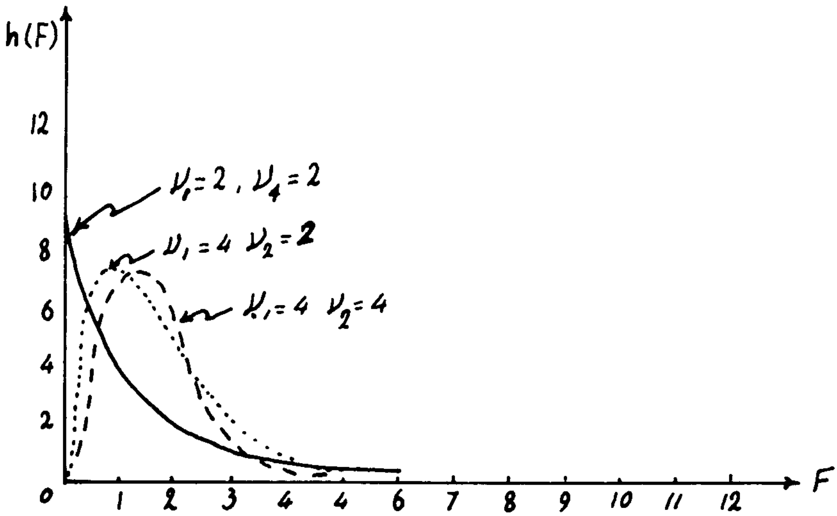
$$h(F) = \frac{\left[\frac{1}{2}(\nu_1+\nu_2)-2\right]!}{\left[\frac{1}{2}(\nu_1-2)\right]!\left[\frac{1}{2}(\nu_2-2)\right]!}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1} \frac{F^{\frac{1}{2}(\nu_1-2)}}{(1+\nu_1F/\nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}, \quad F>0 \tag{9}$$

بتابع كثافة التوزيع F أو توزيع سنيديكيور بـ ν_1 و ν_2 درجة من الحرية ونرمز له بـ $F(\nu_1, \nu_2)$ حيث ν_1, ν_2 هما عادة عددان صحيحان موجبان ولنلاحظ أن $F(\nu_1,\nu_2)$ لا يساوي $F(\nu_2, \nu_1)$ أي إذا أخذت كل من ν_1 و ν_2 موقع الأخرى في عبارة تابع الكثافة فإننا نحصل على تابع جديد يختلف عن التابع الأول . ويقدم الشكل (٦ ـ ١٠) عدة أمثلة من المنحني h (F) وذلك من أجل قيم مختلفة لِـ ν_1 و ν_2 . ونقدم في الملحق جدولين للتوزيع F يعطيان قيم F التي يقع 5% و 1% ، على الترتيب ، من المساحة الكلية تحت المنحني على يمينها . ونلفت النظر أيضاً وبدون برهان إلى الحقيقة النظرية التالية : إذا كان المتحول u_1 يتبع التوزيع χ^2 بـ ν_1 درجة من الحرية ويتبع المتحول u_2 التوزيع χ^2 بـ ν_2 درجة من الحرية وكان المتحولان u_1 و u_2 مستقلين عن بعضهما فعندئذ يتبع المتحول F المعرف بالنسبة :

$$F = \frac{u_1/\nu_1}{u_2/\nu_2} \tag{10}$$

توزيع سينديكور أو التوزيع F بـ ν_1 و ν_2 درجة من الحرية . ونلاحظ أن درجة الحرية المذكورة أولاً هي درجة حرية المتحول الموجود في الصورة والثانية هي درجة حرية المتحول الموجود في المخرج .

وإذا أخذنا عينتين عشوائيتين مستقلتين من مجتمع طبيعي معياري ولتكن العينة الأولى $Z_1, Z_2, \dots, Z_n$ ، والعينة الثانية $Z'_1, Z'_2, \dots, Z'_m$ فعندئذ يتبع مجموع



شكل (٦-١٠) أمثلة من التوزيع F

مربعات مقادير العينة الأولى $\sum_{i=1}^{n} Z_i^2$ التوزيع χ^2 بـِ n درجة من الحرية . ويتبع مجموع مربعات مقادير العينة الثانية $\sum_{i=1}^{m} Z_i'^2$ التوزيع χ^2 بـ m درجة من الحرية والتوزيعان مستقلان لأن العينتين مستقلتين بالفرض . وهكذا نجد أن المتحول المعرف بالنسبة :

$$F = \frac{\sum_{i=1}^{n} Z_i^2 / n}{\sum_{i=1}^{m} Z_i'^2 / m} \qquad (11)$$

يتبع التوزيع F بـِ n و m درجة من الحرية .

وتُستخدم هذه التوزيعات ، وعلى الأخص التوزيع F ، على نطاق واسع في طرق الاحصاء . كما سنجد في الفصول القادمة من هذا الكتاب وخاصة الجزء الثاني .

ونلاحظ أن $t^2 = \frac{z^2}{u/\nu}$ هو نسبة $\chi^2(1)$ إلى $\chi^2(\nu)/\nu$ أي أن توزيع المتحول $[t(\nu)]^2$ هو بالتعريف $F(1,\nu)$. وهذا يعني أن التوزيع $F(1,\nu)$ يكافىء مربع التوزيع $t(\nu)$.

تمارين

١ ـ باستخدام جدول التوزيع الطبيعي أحسب الاحتمالات التالية :

$P(-.2 < z < .2)$, $P(.3 < z < 1.56)$, $P(-.9 < z < 0)$, $P(0 < z < 1.2)$
$p(z < 1.35)$, $P(z < -.32)$, $P(z > -.75)$, $P(-1.3 < z < 1.74)$

٢ ـ أوجد النقطة z_0 بحيث أن :

$P(-Z_0 < z < z_0) = .90$, $P(z < z_0) = .05$, $P(z < z_0) = .8643$,
$P(z > z_0) = .5$, $(P(-z_0 < z < z_0) = .99$, $P(-z_0 < z < z_0) = .95$

٣ ـ تتوزع معدلات مجتمع كبير من طلبة الكليات تقريباً وفق التوزيع الطبيعي بمتوسط يساوي 2.4 وانحراف معياري يساوي .8 . فما هي نسبة الطلاب الذين تتجاوز معدلاتهم 3.0 ؟ (المعدل التام هو 4) .

٤ ـ بالإشارة إلى التمرين السابق إذا فرضنا أننا شطبنا أسماء الطلاب الذين تكون معدلاتهم 1.9 أو أقل فماذا ستكون نسبة الطلاب الذين تُشطب أسماؤهم ؟

٥ ـ يتوزع عمر نوع من الغسالات الكهربائية وفق التوزيع الطبيعي بمتوسط يساوي 3.1 سنة وإنحراف معياري هو 1.2 سنة . إذا كانت الغسالات مكفولة لمدة سنة فما هي نسبة الغسالات المباعة التي سيضطر المصنع إلى استبدالها بغسالة جديدة ؟

٦ ـ وجدنا أن الفترة الزمنية الضرورية لإنجاز اختبار للذكاء مخصص لطلبة الكليات تتوزع احتمالياً وفق التوزيع الطبيعي بمتوسط يساوي 70 دقيقة وإنحراف معياري يساوي 12 دقيقة . كم يجب أن نحدد زمن الاختبار إذا

أردنا إتاحة وقت كاف لِـ 90% من الطلاب لإتمام الاختبار ؟

٧ ـ لتكن التجربة الثنائية حيث $n = 25$ ، $p = .4$. احسب $P(8 \leqslant x \leqslant 11)$ مستخدماً :

أ ـ التوزيع الثنائي . (جدول التوزيع الثنائي في الملحق) .

ب ـ التقريب الطبيعي للتوزيع الثنائي .

٨ ـ أعد نفس المطلوب في التمرين السابق من أجل $P(x \leqslant 4)$ حيث نفرض الآن أن $n = 25$ ، $p = .2$ في التوزيع الثنائي .

٩ ـ وجد تاجر أن احتمال إتمام صفقة بيع عند قدوم زبون هي 0.3 فإذا استقبل التاجر 50 زبوناً في اليوم فما هو احتمال ألا يقل عدد صفقاته عن 10 ؟ (افترض أن x عدد الصفقات يتبع التوزيع الثنائي) .

١٠ ـ أخذنا عينة من الناخبين في مدينة معينة في انتخابات الادارة المحلية . ولنفرض أن المرشح A سيربح إذا استطاع الفوز بِـ 40% من أصوات الناخبين . إذا حاز هذا المرشح على 920 صوتاً من عينة تضم 2500 ناخباً ، فهل تتناقض هذه النتيجة مع الفرض بأن A سيفوز ؟

١١ ـ نُظمت ماكينة لتقديم شراب مرطب بحيث تلفظ على المتوسط μ أونزة للكأس الواحدة . إذا كان ملء الكأس من الأونزات يتوزع وفق التوزيع الطبيعي بإنحراف معياري يساوي 0.3 أونزة ، فما هي القيمة التي يجب تحديدها لِـ μ بحيث تفيض الكؤوس ذات السعة « 8 أونزة » بنسبة 1% فقط ؟

١٢ ـ نقوم باختبار إحصائي لاختبار الفرضية بأن وسيط توزيع ثنائي p يساوي 0.1 . فإذا كان حجم العينة $n = 400$ ورغبنا أن يكون α مساوياً لِـ .05 (اختبار ثنائي الذيل) . حدّد موقع منطقة الرفض إذا كان إحصاء الاختبار (أ) المتحول الطبيعي المعياري z ، (ب) المتحول الثنائي x .

١٣ ـ أحسب β في التمرين السابق إذا كانت القيمة الحقيقية لِـ p هي 0.15

١٤ ـ تستخدم شركة صناعية 3000 مصباح كهربائي . ويتوزع عمر المصباح وفق التوزيع الطبيعي بمتوسط يساوي 500 ساعة وإنحراف معياري يساوي 50 ساعة . ولكي يكون عدد المصابيح المحترقة خلال أوقات الإنتاج أقل ما يمكن تُستبدل كل المصابيح بعد فترة معينة . كيف يجب تحديد فترة إستبدال المصابيح بحيث لا يحترق بين موعدين للاستبدال أكثر من 1% من المصابيح .

١٥ ـ تقول مديرية الإذاعة والتلفزيون أن 20% من المشاهدين يتابعون برناجاً معيناً .. وقد وجدنا في عينة عشوائية من 1000 مشاهد أن $X = 184$ من بينهم يتابعون البرنامج . فهل تقدم هذه المعلومات دلالة كافية لنقض ما تقوله مديرية الإذاعة والتلفيزيون ؟

١٦ ـ استخدم الجدول χ^2 لحساب الاحتمالات التالية :

(أ) $P(\chi^2 \geqslant 26.119)$ من أجل $\nu = 14$.

(ب) $P(\chi^2 > 26.119)$ من أجل $\nu = 14$.

(جـ) $P(\chi^2 \geqslant 26.119)$ من أجل $\nu = 19$.

(د) $P(23.337 \leqslant \chi^2 \leqslant 42.980)$ من أجل $\nu = 24$

(هـ) $P(\chi^2 < 9.348)$ من أجل $\nu = 3$

١٨ ـ إستخدم جدول التوزيع t لحساب الاحتمالات التالية :

(أ) $P(|t| \geqslant 2.015)$ من أجل $\nu = 5$.

(ب) $P(|t| > 2.015)$ من أجل $\nu = 5$.

(جـ) $P(|t| \geqslant 2.015)$ من أجل $\nu = 8$.

(د) $P(t \leqslant -2.015)$ من أجل $\nu = 5$.

(هـ) $P(-2.015 \leqslant t \leqslant 2.015)$ من أجل $\nu = 5$.

(و) $P(.7 \leqslant t \leqslant 1.8)$ من أجل $\nu = 26$.

(ز) $P(|t| \geqslant 3)$ من أجل $\nu = 17$.

١٨ ـ استخدم جدول التوزيع F لحساب الاحتمالات التالية :

(أ) $P(F \geqslant 7.79)$ من أجل $\nu_1 = 11$ و $\nu_2 = 6$.

(ب) $P(F > 7.79)$ من أجل $\nu_1 = 11$ و $\nu_2 = 6$.

(جـ) $P(F \geqslant 4.03)$ من أجل $\nu_1 = 11$ و $\nu_2 = 6$.

(د) $P(F \geqslant 7.79)$ من أجل $\nu_1 = 2$ و $\nu_2 = 10$.

(هـ) $P(F \leqslant 0.20)$ من أجل $\nu_1 = 9$ و $\nu_2 = 14$.