الفصل الرابع

التوزيعات الاحتمالية

Probability Distributions

في بعض الأحيان لا يقتصر الاهتمام على معرفة الاحتمالات المصاحبة لحدث معين ولكن يتعدى ذلك إلى معرفة توزيع تلك الاحتمالات على كامل المدى لجميع الأحداث الممكنة. على سبيل المثال، لو افترضنا زهرة نرد مرقمه من ١ - ٦ وتم تعريف المتغير العشوائي X بأنه الرقم الذي يظهر عند رمي الزهرة يمكن كتابة الاحتمالات المرتبطة بكل قيمه من قيم المتغير العشوائي X كما في المعادلة رقم (4.1) وجميع الاحتمالات متساوية لافتراضنا أن الزهرة سليمة من العيوب :

$$P(X=1) = 1/6, \quad P(X=2) = 1/6, \ldots, P(X=6) = 1/6 \tag{4.1}$$

وتسمى قيم المتغير العشوائي X والاحتمالات المصاحبة لها بالتوزيع الاحتمالي ويمكن استبدال كتابة كل احتمال مصاحب في حالة تساوي الاحتمالات بالصيغة الرياضية التالية:

$$P(X) = 1/6 \text{ for } (X = 1, 2, \ldots, 6) \tag{4.2}$$

ويكون مجموع قيم الاحتمالات المصاحبة للمتغير X مساوي للواحد.

وهـذه الخاصية تـشمل جميع التوزيعـات الاحتماليـة، وكمـا هـو معلـوم فـإن الاحتمالات يجب أن تكون موجبه أو غير سالبة.

مـن جهـة أخـري فـإن هنـاك علاقـة بـين التوزيعـات الاحتماليـة والتوزيعـات التكرارية ، حيث إن التوزيعات التكرارية عبارة عن قائمة تشمل جميع النتائج للمتغير التي تم تقسيمها إلى فئات مع التكرارات لكل فئة.

في حين أن التوزيع الاحتمالي يعبر عن جميع النتائج الممكنة للمتغير العشوائي مع الاحتمال المصاحب لكل قيمة بدلاً من التكرار.

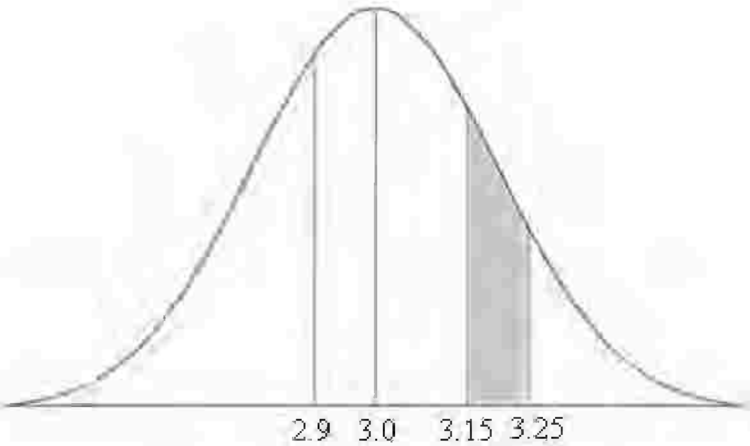
وتهدف العديد من المشكلات الإحصائية إلى اختيار التوزيـع الأفضل المناسـب للمتغير العشوائي كما تمت الإشـارة لـذلك سـابقاً. وقبـل اختيـار التوزيعـات الاحتماليـة المستخدمة غالباً في الإحصاء، سيتم النظر إلى دوال الكثافة الاحتمالية.

دوال الكثافة الاحتمالية Probability Density Functions

دالـة الكثافـة الاحتماليـة، ويرمـز لهـا بـ pdf ، قـد تكـون منفـصلة أو متـصلة. وتُعرّف دالة الكثافة الاحتماليـة المنفـصلة بأنهـا عبـارة عـن دالـة نقطيه علـى فـراغ عينـة منتهي ، وبذلك فإنها تأخذ قيم عددية منتهية فقط.

وتشير الـ pdf للاحتمال المصاحب لكل ناتج ممكن للمتغير العشوائي المنفـصل. أمـا دالـة الكثافـة الاحتماليـة المتـصلة فهـي مجموعـة دوال تعـرض التوزيـع الاحتمـالي للنواتج الممكنة لمدى من القيم للمتغير العشوائي المتصل.

ويوضح الشكل رقم (٤،١) قيمة الاحتمال بيانياً للمتغير العشوائي المتصل X ، الـذي يعبّـر عـن قطـر عمـود المحـرك المنتج مـن شـركة الآلات الزراعيـة بقطـر يـتراوح بين٣،١٥ إلى ٣،٢٥ والذي هو عبارة عن المساحة المظللة.



الشكل رقم (٤.٩). التوزيع الاحتمالي المتصل.

وتعبر دالة الكثافة الاحتمالية pdf عن مجموعة مستقلة من التوزيع الاحتمالي بحيث تكون قاعدة لتحديد الاحتمالات المصاحبة لأحداث التجربة، بينما التوزيع الاحتمالي عبارة عن عرض منظم أو ترتيب للاحتمالات عند تحديدها. من هذا المنطلق يمكن دراسة بعض التوزيعات الاحتمالية الشائعة الاستخدام.

التوزيع الاحتمالي ثنائي الحدين Binominal Probability Distribution

مجموعة المشكلات الإحصائية التي نصادفها كثيراً تشمل أحداث عشوائية بحيث يكون لها نتيجتين فقط. هذه النتائج تحدث بدون نمط ثابت ولكن احتمال حدوث هذه النتائج يبقى ثابتاً لأي محاولة.

هذا النوع من المسائل يسمى محاولات برنولي. يتكون نظام برنولي من ناتجين أو حدثين فقط، حيث إن النتائج أو الحدث الذي نهتم بدراسته أو نسعى لمعرفته يسمى

نجاح والناتج الآخر يسمى فشل. في بعض الأحيان يكون لدينا معرفه بمستوى من الثقة بمقدار النجاح المتحصل عليه من عدد كبير من المحاولات والذي يدعى المعلمة أو العامل المتغير.

دالة الكثافة الاحتمالية لمحاولات برنولي تسمى دالة الكثافة الاحتمالية ثنائيه الحدين والتي تعطي الاحتمال P لعدد r نجاحات في عدد n محاولة من التجربة. وبالطبع فإن معرفتنا لعدد حالات النجاح r يُمكّننا من معرفه عدد حالات الفشل والتي تساوي $n - r$ ؛ نظراً لأن لدينا ناتجين فقط للتجربة وبذلك يمكن حساب احتمال الفشل والذي يساوي $1- P$ والذي يعبر عنه أحياناً بـq. ويمكن صياغة دالة الكثافة الاحتمالية ثنائية الحدين رياضياً في المعادلة رقم (4.3):

$$P\ (r \setminus np) = {}_{n}C_{r}\ p^{r}(1-p)^{n-r} or\ {}_{n}C_{r}\ P^{r}\ q^{n-r} \qquad (4.3)$$

ويمكن إيضاح ذلك بالمثال التالي: نفرض رمي حجر نرد سليم أربع مرات ونرغب في معرفة احتمال الحصول على رقم ١ ثلاث مرات في هذه الأربع رميات.

ولحل هذه المسألة نحسب أولاً عدد الطرق الممكن الحصول بها على رقم ١ ثلاث مرات في الأربع رميات باستخدام التوافيق حيث

$${}_{4}C_{3}\ =\ 4!/(4-3)!\ 3! = 4$$

وعليه يمكن حساب احتمال وقوع أي من هذه الطرق باستخدام قاعدة ضرب الاحتمالات. فمثلاً احتمال الحصول على رقم واحد ثلاث مرات بشكل متتالٍ (١) وقيمة أخرى غير الواحد مثلاً (١َ) يمكن حسابها كالتالي:

$$P(1111') = (1/6)(1/6)(1/6)(5/6) =\ 5/1296$$

وعليه فإن احتمال الحصول على رقم ١ في المحاولات الثلاث الأولى من أربع محاولات يمكن حسابها كالتالي :

$$(4)(5/1296\ ) = 20/1296 \quad \text{أو} \quad 5/324$$

ويمكن استخدام دالة الكثافة الاحتمالية للتوزيع ثنائي الحدين لحساب ذلك الاحتمال كالتالي :

$$(4)(1/6)^3\,(5/6)^1 = 5/324 \quad \text{أو} \quad {}_4C_3\,p^3\,q^1 \quad \text{أو} \quad {}_4C_3 p\,p\,p\,q$$

ويمكن إيضاح ذلك بمثال آخر: وباستخدام المعادلة رقم (4.3) بافتراض أن لدينا عشرة مواليد من العجول ونرغب في معرفة احتمال أنها تحتوي على ذكر واحد. يمكن استخدام دالة الكثافة الاحتمالية لتوزيع ثنائي الحدين حيث $r=1$, $n=10$, $P=0.5$ والتعويض في المعادلة (4.3).

$$P(r=1\,|\,n=10, p=0.5) = {}_{10}C_1(0.5)^1(0.5)^9 = 10!/1!9!(0.5)(0.00195) = 0.0098$$

يمكن تفسير هذه النتيجة بأنها غير متماثلة.

وتجدر الإشارة إلى أن المصطلح ثنائي الحدين أتى من طريقه صياغة دالة الكثافة الاحتمالية. حيث إن قيمتها لأي قيم معطاة لـ r ، n ، p مكافئة للحد العام في مفكوك ثنائي الحدين والذي يعبر عنه بـ $(p+q)^n$.

وتعطي صياغة دالة الكثافة الاحتمالية ثنائية الحدين طريقة حساب الاحتمال لأي عدد محتمل للنجاح في أي عدد معطى من المحاولات، وترتيب تلك الاحتمالات بطريقة معينة يعطي دالة التوزيع الاحتمالي ثنائي الحدين. ولذلك فإن دالة الكثافة الاحتمالية ثنائية الحدين تعرّف كامل التوزيعات، توزيع واحد لكل زوج من قيم n ، P ، معالم تلك التوزيعات.

وحيث إن التوزيعات التي تكون فيها قيمه $P=0.5$ متماثلة، فإن التوزيعات التي تكون فيها قيم P صغيره تكون متماثلة بصوره معقولة خاصة عندما تكون قيمه n كبيرة.

ومن هذا المنطلق يمكن استخدام المنحنى الطبيعي المتماثل كتقريب لثنائي الحدين والعكس صحيح.

هناك حالات عندما نرغب في إيجاد الاحتمال لعدد (r أو أكثر) من حالات النجاح أو (r أو أقل) في عدد n محاولة. وفي الحقيقة فإننا نهدف إلى إيجاد الاحتمال المصاحب للمساحة تحت أحد أطراف التوزيع ثنائي الحدين. ويمكن الحصول على هذه الاحتمالات بجمع احتمالات ثنائي الحدين لأحداث معينة. وللتجارب المشتملة على عدد كبير من المحاولات، فإن جمع هذه الاحتمالات يحتاج لوقت؛ كما أنه مزعج خاصة إذا كانت قيمة r ليست قريبة من الصفر أو قريبه من n . ولهذا السبب تم إيجاد جداول خاصة بتوزيع ثنائي الحدين والتي تعطي مساحات الأطراف للمدى الكبير من قيم n ، P ، r (انظر الملحق جدول رقم (٢) للمثال الجدولي).

الجدول رقم (٤,١). نتائج سبع محاولات لاختيار ذكور وإناث الحيوانات عشوائياً.

المحاولة	الناتج	قيمة X	تعريف المتغير
١	F	١	X_1
٢	M	٠	X_2
٣	F	١	X_3
٤	F	١	X_4
٥	M	٠	X_5
٦	F	١	X_6
٧	M	٠	X_7

متوسط وتباين محاولات برنولي
The Mean and Variance of A Bernoulli Process

من السهولة اشتقاق متوسط وتباين توزيع برنولي بطريقه واضحة ؛ نظراً لإمكانية وصف محاولات برنولي بالعديد من الطرق ونفترض أن لدينا تجربة سحب حيوانين مع الإعادة من حظيرة كبيره تحتوي على مجموعة كبيره من الحيوانات عددها n ونصف تلك الحيوانات ذكور M والنصف الآخر إناث F ، ولذلك فإن نسبة الإناث تساوي $P = \frac{F}{n}$. وعليه فإن احتمال النجاح لاختيار حيوان أنثى F تساوي $P = \frac{1}{2}$ ونسبة الفشل (اختيار ذكر M) تساوي $(1 - P) = \frac{1}{2}$. وبدلاً من استخدام دالة التوزيع الاحتمالي ثنائي الحدين في هذه الحالة يمكن استخدام طريقه أخرى. بفرض أننا سحبنا عينة عشوائية من مجموعة حيوانات عددها n في حظيرة وتم تعريف متغير عشوائي منفصل X بحيث تكون قيمته ١ في حاله (سحب أنثى) وقيمته تساوي صفر في حالة الفشل . بهذه الطريقة تم تعريف مجموعة n من المتغيرات العشوائية المستقلة ، $X_1, X_2, \ldots, X_n$ بحيث يعبر كل متغير عن محاولة مستقلة ولغرض إيضاح المثال بحيث يكون أكثر وضوح نفترض أننا سحبنا عدد سبعة حيوانات بطريقة عشوائية من الحظيرة وكانت النتيجة كما يلي : $\{F, M, F, F, M, F, M\}$.

وتم عرض هذه البيانات جدولياً في الجدول رقم (٤،١). وبالتالي يمكن حساب القيمة المتوقعة للمتغير X باستخدام المعادلة التالية.

$$\mu = E(X) = \Sigma X \cdot P(X) = (1)(p) + (0)(1-p) = p \qquad (4.4)$$

ويمكن إيجاد التباين للمتغير العشوائي X باستخدام الصيغة الموضحة في المعادلة رقم (4.5) والذي يعبر عنه بـ $Var(X)$ أو σ^2 ويمكن إيجاد التباين للمثال آنفاً كالتالي :

$$\sigma^2 = E(X-\mu)^2 \quad \text{أو} \quad E(X-p)^2 = \sum (X-p)^2 \cdot P(X) \qquad (4.5)$$

ويمكن حساب التباين لعدد حالات النجاح r في مجموعة المحاولات n باستخدام المعادلة (4.6)) التي يمكن التعبير عنها كالتالي :

$$\sigma^2 = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p(1-p) \qquad (4.6)$$

وتعبر r عن مجموع المتغيرات العشوائية حيث يمكن إيضاحها في المعادلة التالية:

$$r = X_1 + X_2 + \ldots + X_n \qquad (4.7)$$

ويمكن حساب القيمة المتوقعة لـ r باستخدام المعادلة التالية :

$$E(r) = E(X_1) + E(X_2) + \ldots + E(X_n) = p + p + \ldots + p = np \qquad (4.8)$$

وبطريقة مشابهة يمكن التعبير عن التباين كما في المعادلة رقم (4.9) :

$$\text{Var}(r) = p(1-p) + \ldots + p(1-p) = np(1-p) = npq \qquad (4.9)$$

ويمكن حساب الانحراف المعياري بإيجاد الجذر التربيعي للتباين.

ويمكن إيضاح ما سبق من خلال المثال التالي :

بائع معروضات زراعية لديه مخزن به قطع غيار منها ١٠٪ معيبة فإذا افترضنا أن اختيار القطعة المعيبة هو حالة نجاح وعبّرنا عنه بـ S واختيار قطعة صالحة تعبر عن الفشل وعبّرنا عنه بـ F عليه يمكن حساب القيمة المتوقعة لمرات النجاح لعينة من وحدتين باستخدام المعادلة رقم (4.8) كالتالي :

$$E(r) = np = (2)(0.1) = 0.2$$

والتباين يمكن حسابه باستخدام المعادلة رقم (4.9).

$$Var(r) = npq = (2)\ (0.1)\ (0.9) = 0.18$$

وعليه فإن الانحراف المعياري :

$$S = \sqrt{0.18} = 0.42$$

توزيع بواسون الاحتمالي Poisson Probability Distribution

المجموعة الثانية من المسائل الإحصائية يمكن وصفها باحتمالات نجاح صغيرة لأي مجموعة من محاولات عديدة لتجربة معينه. على سبيل المثال هذه المسائل تشمل تطبيقات تأمين المحاصيل، وعدد الكوارث المتعلقة بحالات الطقس في فترة زمنية معينة، وعدد الحراثات القادمة لمخزن الحبوب خلال عملية الحصاد، ومن ثم نوع الدورة أو الصف أو خط الانتظار، والطلب على المواد من مخزن شركة المعدات الزراعية ، عدد العيوب في منتجات مصانع الآلات الزراعية.

ويتميز توزيع بواسون الاحتمالي بملاءمته لعرض هذا النوع من المسائل الإحصائية علاوة على إمكانية استخدام صيغ محدّدة للتوزيع ثنائي الحدين كتوزيع احتمالي لعرض هذه الحالات. وتكون صيغ ثنائي الحدين أكثر ملاءمة عندما يكون عدد محاولات برنولي n كبير واحتمال النجاح p صغير. وفي هذه الحالة فإن الحسابات باستخدام دالة التوزيع الاحتمالي ثنائي الحدين طويلة ومملة ولكنها بسيطة باستخدام دالة التوزيع الاحتمالي لبواسون.

وفي الحقيقة فإن توزيع بواسون لا يمكن وصفه باستخدام أحداث لها ناتجين فقط واحتمال ثابت للنجاح كما في حالة محاولات برنولي. ولكن من خلال شروط معينة يمكن استخدام توزيع بواسون لحلّ مسائل برنولي. على سبيل المثال، هياكل المحراث تصنّع باستمرار بعملية طرق ويكون سمك الجدار للأنبوب أقل من المواصفات المقبولة.

ففي عينة لـ ١٠٠٠قدم من هذا الأنبوب وجد أن ٢٠ مكاناً منها معيبة (أقل من المواصفات) بصورة عشوائية بمتوسط ٢ لكل ١٠٠ قدم. ففي حالة ثبات احتمال وجود عدد معين من العيوب لكل ١٠٠ قدم يمكن اعتبار مسألة بواسون كحالة برنولي بعرض طول قصير جداً كمحاولة مستقلة بناتج صفر أو أكثر من النجاح.

وكلما كان بالإمكان المحافظة على متوسط np مساوي لـ ٢ معيب لكل ١٠٠ قدم يمكن تجزئة ال١٠٠ قدم إلى ١٠ أجزاء بطول ١٠ قدم أو بطول ١ قدم أو بأي تقسيم مناسب.

وكلما كان عدد القطع كبير فإن احتمال وجود أكثر من عيب في القطعة يكون صغير ويكون الاختلاف بين دالة التوزيع الاحتمالي لثنائي الحدين وبواسون ضئيل.

ولذلك يمكن القول بأنه في حالة مسائل برنولي الاحتمالات مرتبطة بقيمة الاحتمال p للنجاح في عدد n محاولات مستقلة وعدد r من النجاح.

في توزيع بواسون، يمكن التعبير عن الاحتمالات لعدد معطى من النجاحات l ، لوحدة من المسافة (مثل طول قصير للأنبوب) وعدد حالات النجاح m في قيمة معطاة من المسافة . وتجدر الإشارة بأن الحرف l مماثل للحرف p وكذلك الحرف m مماثل للحرف r . وبذلك فإن العدد المتوقع من النجاحات np لمحاولات برنولي مماثل للـ lm (معادلة رقم 4.10) للعدد المتوقع للنجاحات في قيمة معطاة من المسافة لتوزيع بواسون.

$$\mu = lm \qquad (4.10)$$

ويمكن إيضاح ذلك من خلال المثال التالي:

في دراسة لوصول شاحنات الحبوب للرافعات كان احتمال وصول الشاحنات في أي دقيقة مختارة $p = 0.0333$ في حين أن الرقم المتوقع للوصول في نصف ساعة

يمكــن حــسابه كالتــالي $np = (30)(0.0333) = 1$ ويمكــن اختيــار الدقيقــة كوحدة للمسافة l ونصف الساعة كقيمة معطاة للمسافة m وبذلك يمكن حساب الرقم المتوقع u.

$$\mu = (0.0333)(30) = 1$$

ولذلك يمكن حساب احتمالات ثنائي الحدين لعدد r من حالات النجاح في n محاولة باستخدام دالة الكثافة الاحتمالية لثنائي الحدين الموضحة في المعادلة رقم (4.11) التالية :

$$P(r \backslash \mathrm{np}) = {}_nC_r p^r q^{n-r} \qquad (4.11)$$

وكما هو معلوم فإنه عندما تكون قيمه p صغيره فإن التوزيع الاحتمالي سيكون ملتو باتجاه اليمين. وعليه فإنه للقيم الصغيرة للاحتمال p لعدد كبير من المحاولات n مع بقاء ناتج np ثابت فإن نهاية التوزيع ثنائي الحدين هي توزيع بواسون والذي يمكن عرضه رياضياً في المعادل رقم (4.11) التالية :

$$\lim P(r \backslash n, p) = \frac{(np)^r}{r!} e^{-np} \qquad (4.12)$$

حيث e هي الأساس للوغاريثم الطبيعي والتي تساوي ٢,٧١٨٢٨ وبالتعويض عن قيمة np بـ μ يمكن كتابة الصيغة الرياضية لتوزيع بواسون في الشكل التالي :

$$P(r \backslash \mu) = \frac{\mu^r}{r!} e^{-\mu} \qquad (4.13)$$

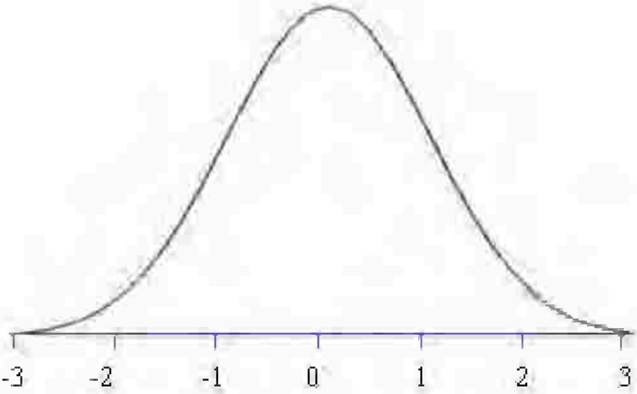
ولذلك فإن هناك معلمة واحدة لتوزيع بواسون والتي يمكن تحديد كامل التوزيــع عنــد معرفتهــا. حيــث إن متوســط التوزيــع $E(r) = \mu$ وكــذلك

التباين $Var(r) = \mu$. ويمكن إيجاد الانحراف المعياري بحساب الجذر التربيعي للتباين μ . ويمكن الحصول على هذه النتائج مباشرة من التوزيع ثنائي الحدين ؛ حيث إن القيمة المتوقعة $E(r) = np$ والتباين $Var(r) = n\,p\,q$ وتوزيع بواسون الذي يعبّر عن النهاية للتوزيع ثنائي الحدين عندما تقترب n من ما لانهاية و p تقترب من الصفر حيث $np = \mu$ تبقى ثابتة. ولذلك فإن npq تقترب من $np = \mu$ كلما اقتربت قيمة q من الواحد الصحيح.

ويمكن عرض المثال التالي لتوضيح ذلك حيث قام قسم الفحص بمصنع الآلات الزراعية بفحص خرطوم الهيدوليك البالغ طوله ٣٠٠ قدم لمعرفة العيوب وتسجيلها لكل طول من الخرطوم. ومن خبرات سابقه فإن عدد العيوب للطول يتبع توزيع بواسون ، فلو افترضنا اختبار ٢٠ وحدة طول ووجد منها ١٠ تالفة أو معيبة ولذلك فإن تقدير التباين والقيمة المتوقعة $0.5 = 10/20 = \mu$ ويمكن إيجاد الانحراف المعياري بأخذ الجذر التربيعي للتباين والذي يساوي ٠,٧.

التوزيع الاحتمالي الطبيعي The Normal probability Distribution

في حالة المتغيرات العشوائية المنفصلة فإن التوزيع الاحتمالي الأكثر استخداماً هو توزيع ثنائي الحدين في حين أن التوزيع الاحتمالي الطبيعي هو الأكثر استخداماً في حالة المتغيرات العشوائية المتصلة. ويتميز منحنى التوزيع الطبيعي بأنه متماثل ويشبه شكل الجرس، كما في الشكل رقم (٤,٢). ويقع متوسط التوزيع μ في وسط التوزيع، كما أنها تتحدد معالمه بالمتوسط μ والانحراف المعياري σ .



الشكل رقم (٤,٢)- التوزيع الاحتمالي الطبيعي المعياري.

ونظراً لأن هذا التوزيع نظري أو توزيع حقيقي مرتبط بجميع قيم المجتمع، فإننا نستخدم الحروف الإغريقية الخاصة بالمجتمع μ ، σ لعرض قيم المتوسط والانحراف المعياري بدلاً من استخدام $\overline{X}$ ، S والتي سوف نستخدمها لاحقاً عندما نتطرق للمتوسط و الانحراف المعياري للعينات المسحوبة من التوزيع.ويتميز التوزيع الطبيعي ببعض الخصائص يمكن إيجازها فيما يلي:

١- المساحة تحت التوزيع الواقعة بين $\mu-\sigma$ و $\mu+\sigma$ تمثل تقريباً ٦٨ ٪ من المساحة الإجمالية الواقعة تحت المنحنى.

٢- المساحة تحت التوزيع الواقعة بين $\mu-2\sigma$ و $\mu-2\sigma$ تمثل تقريباً ٩٥٪ من المساحة الإجمالية الواقعة تحت المنحنى.

٣- المساحة تحت التوزيع الواقعة بين $\mu-3\sigma$ و $\mu+3\sigma$ تمثل تقريباً ٩٩,٧ ٪ من المساحة الإجمالية الواقعة تحت المنحنى.

وعليه فإن مدى المتغير العشوائي X نظرياً يقع بين سالب مالا نهاية وموجب ما لا نهاية تحت المنحنى الطبيعي إلا أنه تقريباً معظم الاحتمالات تتواجد في المدى بين

المتوسط مضاف له ثلاثة انحرافات معيارية والمتوسط مطروحاً منه ثلاثة انحرافات معيارية.

ويتحدد موقع المنحنى وشكله الطبيعي بقيم المتوسط والانحراف المعياري، حيث تحدد قيمة المتوسط مركز التوزيع على خط الأعداد الحقيقية بينما تحدد قيم الانحراف المعياري انتشار المنحنى حول المتوسط. وحيث إن جميع المنحنيات الطبيعية الخاصة بالتوزيع الاحتمالي النظري تتميز بأن إجمالي المساحة تحت المنحنى تساوي الواحد ولذلك فإن زيادة الانحراف المعياري تؤدي لانخفاض ارتفاع المنحنى بحيث يكون مفرطحاً وممتداً أكثر على خط الأعداد. ونظراً لأن شكل المنحنى يتحدد كلياً بواسطة الانحراف المعياري فإنه يمكن معايرة كل المنحنيات الطبيعية باستخدام انحراف معياري واحد بإجراء تعديل بسيط في المتغيرات.وعليه فإن أبسط منحنى يمكن استخدامه للمعايرة بين المنحنيات الطبيعية هو المنحنى الذي متوسطه صفر وانحرافه المعياري يساوي الواحد الصحيح (الشكل رقم ٤,٢) وعليه فعند الحاجة لحساب الاحتمالات لأي منحنى طبيعي يتم أولاً تحويل ذلك المنحى إلى المنحنى الطبيعي القياسي. ويمكن التعبير عن دالة الكثافة الاحتمالية للمنحنى الطبيعي باستخدام المعادلة التالية:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad (4.14)$$

وتتسم هذه المعادلة بصعوبة تقييمها ولذلك يستخدم الجدول رقم ٧ بالملحق للحصول على الاحتمالات للتوزيع الطبيعي القياسي. واستخدام هذا الجدول ممكن نظراً لأن أي نقطة على المحور السيني X للتوزيع الطبيعي القياسي يقابلها نقطة على المحور السيني X لأي توزيع طبيعي ويمكن تحديد قيمتها بمعرفة كم انحراف معياري تبعد عن المتوسط.

وعلى سبيل المثال يمكن إيضاح ذلك بافتراض أن لدينا منحنيين تختلف فقط في الانحراف المعياري بحيث إن متوسط المنحنى الأول يساوي صفر وانحراف معياري يساوي الواحد والمنحنى الثاني لمتوسط صفر وانحراف معياري يساوي ثلاثة. وبذلك يمكن تحويل المنحنى الثاني إلى الأول بتغير منحنى X بقسمة كل قيمه على ٣ (قيمه الانحراف المعياري) وعليه فإن قيمة $X = 6$ على المنحنى الطبيعي يعادلها $X = 2$ على المنحنى الطبيعي القياسي وهكذا بالنسبة لباقي القيم.

وبصفة عامه، لأي نقطه x على المنحنى الطبيعي ذي المتوسط μ وانحراف معياري σ هناك قيمه مقابله Z على المنحنى الطبيعي القياسي، وعليه فإن النقطة x تقع على بعد Z انحراف معياري على يمين المتوسط أو $X = \mu + z\sigma$ والتي يمكن صياغتها كمعادلة رياضيه كالتالي:

$$z = (X - \mu) / \sigma \qquad (4.15)$$

هذه الصيغة لـ Z تعبّر عن الطريقة المستخدمة لمعايرة القيم للمتغير X والتي تم استخدامها سابقاً. وباستخدام هذه الصيغة الرياضية يمكن تحويل جميع القيم لأي منحنى طبيعي إلى قيم مناظره لها في المنحنى الطبيعي القياسي ومن ثم حساب الاحتمالات تحت أي منحنى طبيعي باستخدام الجدول ٧ بالملحق. فمثلاً بافتراض أن x تنتمي لمنحنى طبيعي بمتوسط ٢٣٠ وانحراف معياري ٢٠ ($x \sim N(230,20)$ ونرغب في حساب احتمال أن قيمه x أكبر من أو تساوي ٢٨٠ ؟

في هذه الحالة يتم إيجاد القيمة المناظرة لقيمة x على المنحنى الطبيعي القياسي Z باستخدام المعادلة رقم (4.15) .

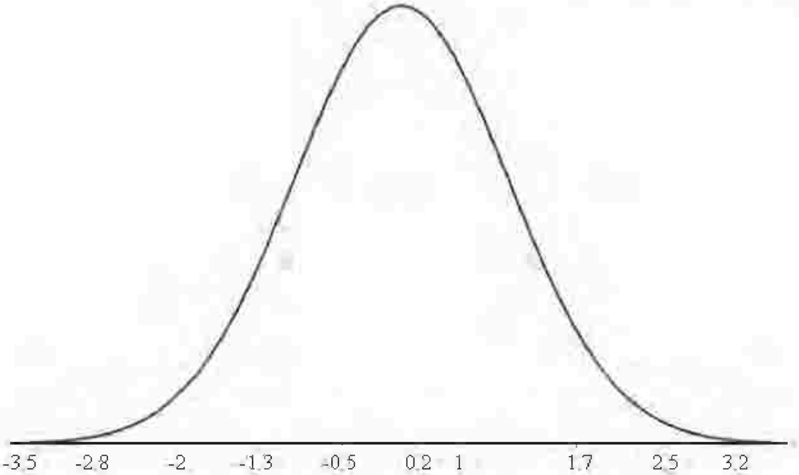
$$z = \frac{280-230}{20} = \frac{50}{20} = 2.5$$

وباستخدام جدول رقم ٧ في الملحق لإيجاد الاحتمال بين قيمة المتوسط وقيمة Z نجد أن الاحتمال يساوي ٠,٤٩٣٨ وعليه فإن إيجاد احتمال أن قيمة X أكبر أو تساوي ٢٨٠ يمكن حسابها بطرح قيمة الاحتمال السابق من ٠,٥ والتي تساوي ٠,٠٠٦٢.

والآن يمكن النظر لمثال آخر ما احتمال أن تكون قيمه X أقل من ٢٧٠ ؟

بنفس الطريقة السابقة يمكن حساب قيمة Z $z = \frac{270-230}{20} = \frac{40}{20} = 2$

من الجدول رقم ٧ نجد أن الاحتمال بين المتوسط وقيمة Z تساوي ٠,٤٧٧٢ والاحتمال يسار المتوسط يساوي ٠,٥ وعليه فإن احتمال أن قيمة أقل من ٢٧٠ يساوي $0.50 + 0.4772 = 0.9772$.



الشكل رقم (٤,٣). التوزيع الاحتمالي الطبيعي المعياري

ويجب ملاحظة أن هذه الاحتمالات عبارة عن المساحات تحت المنحنى وجمع أو طرح تلك الاحتمالات مكافئ لجمع أو طرح المساحات (الشكل رقم ٤,٣) حيث

إن إجمالي المساحة تحت المنحنى تساوي الواحد الصحيح إلا أن إيجاد الاحتمال لقيمة محددة لـ X غير معرّف ؛ لأنه ليس هناك مساحة يمكن إيجادها لنقطه محددة بل يجب أن تكون بين نقطتين.

ودائماً تصاغ مسائل الاحتمالات لمدى من القيم للمتغير العشوائي X عند استخدام التوزيعات المستمرة مثل التوزيع الطبيعي. وفي المثال السابق ليس هناك فرق عند القول أوجد احتمال أن قيمة X أقل من ٢٧٠ ؛ أو القول أوجد احتمال أن قيمة X أقل من أو تساوي ٢٧٠ لأن الإجابة واحده.

التقريب الطبيعي لاحتمالات ثنائي الحدين

Normal Approximation to Binomial Probabilities

عندما يكون لدينا تجربة مشابهة لمحاولات برنولي ونرغب في حساب الاحتمال لـ r من حالات النجاح في عدد n من المحاولات بحيث إن n كبيرة، يمكن في بعض الأحيان استخدام التوزيع الطبيعي كتقريب لتوزيع ثنائي الحدين. فمثلاً إذا تم اختيار عينة n=20 عشوائياً من عمليات تصنيعية باحتمال يساوي p = 0.4 ونرغب في معرفة احتمال الحصول على خمسة عيوب، أي r = 5 وباستخدام دالة الكثافة الاحتمالية للتوزيع ثنائي الحدين يمكن الحصول على التالي :

$$P(r=5,\ n=20,\ \ p=0.4)={}_{20}C_5(0.4)^5(0.6)^{15}=0.0746$$

ويمكن حساب التقريب الطبيعي لاحتمال ثنائي الحدين بحساب المتوسط E(r) والانحراف المعياري ثم إيجاد المساحة التقريبية للمنحنى الطبيعي كالتالي :

$$\mathrm{E}(r)=np=20(0.4)=8$$

$$\sigma_r=\sqrt{npq}=\sqrt{(20)(0.4)(0.6)}=2.19.$$

ولإيجاد الاحتمال لقيمة r نستخدم طريقه تقريبية، كما ناقشنا سابقاً أن المنحنى الطبيعي عبارة عن منحنى مستمر، ويصعب إيجاد الاحتمال لنقطة محددّه مثل $r = 5$. في هذه الحالة يمكن التغلّب على هذه المشكلة برسم الشكل البياني للتوزيع ثنائي الحدين حيث نجد أن $r = 5$ هي عبارة عن مركز الفئة التي حدّها الأعلى ٥,٥ وحدّها الأدنى ٤,٥. وباستخدام هذه النقاط كقيم لـ r يمكن إيجاد المساحة للمنحنى الطبيعي والتي تعطي تقريب للمساحة المناظرة وبالتالي الاحتمال الذي نرغب في إيجاده.

وعليه يمكن معايرة القيم بحساب قيمة Z ثم إيجاد الاحتمالات المناظرة باستخدام الجدول رقم ٧ بالملحق. وأخيراً يتم طرح الاحتمالات المتحصل عليها لإيجاد الاحتمال المناظر للمساحة كالتالي.

$$z_1 = \frac{4.5-8}{2.19} = \frac{-3.5}{2.19} = -1.60 \qquad z_2 = \frac{5.5-8}{2.19} = \frac{-2.5}{2.19} = -1.14$$

باستخدام الجدول في الملحق يمكن إيجاد الاحتمالات:

$$z_1 = -1.6 \qquad p = 0.4452$$
$$z_2 = -1.14 \qquad p = 0.3729$$

وعليه فإن الاحتمال:

$$p(4.5 \le r \le 5.5) = 0.4452 - 0.3729 = 0.0723$$

وهذا الاحتمال قريب جداً من الاحتمال الحقيقي لثنائي الحدين. وعليه فإننا حصلنا على إجابة مقاربه باستخدام التقريب الطبيعي. والجدير بالذكر أن طريقة الحساب بصفة عامة أسهل في حالة استخدام التقريب الطبيعي بدلاً من استخدام دالة الكثافة الاحتمالية لثنائي الحدين وخاصة عند الحاجة للاحتمالات التجميعية.

ودالة الكثافة الاحتمالية للتوزيع الطبيعي هي عبارة عن نهاية التوزيع ثنائي الحدين عندما تصبح n كبيرة جداً وعليه فإن التقريب الطبيعي يكون مكافئاً لثنائي

الحدين عندما تكون n كبيرة. وكقاعدة عامة يمكن استخدام التقريب الطبيعي في أي وقت ما دام المتوسط أكبر من ٥ عندما يكون الاحتمال أقل من ٠,٥ و/أو nq >5 والاحتمال أكبرمن٠,٥ وتكون نتيجة التقريب الطبيعي ضعيفة في أطراف التوزيع ثنائي الحدين في حين أنها تعطي نتائج جيدّه كلما كانت حول المتوسط.

تمارين Exercises

١- أعطت شركة العروض التجارية المحلية للزراعة عرض للعملاء بخصم ١ ٪ للعميل الذي يدفع كامل القيمة المستحقة خلال عشرة أيام. في الماضي ٢٠٪ من العملاء دفعوا خلال عشرة أيام وفي الشهر الحالي أصدرت سبع فواتير المطلوب حساب الاحتمالات التالية.

أ) احتمال حصول عميلين على الخصم.

ب) حصول عملين على الأكثر على الخصم.

ج) عدم حصول أي عميل على الخصم.

د) جميع العملاء السبعة حصلوا على الخصم.

هـ) حصول خمسة عملاء أو أكثر على الخصم.

٢- أوجد المتوسط والانحراف المعياري للتوزيع الاحتمالي ثنائي الحدين في السؤال رقم (١).

٣- مسئول الشراء في أحد الأسواق الكبيرة يعرف من خبرته في الماضي أن ٢ ٪ من الطماطم الوارد من المكسيك غير صالح ويجب أن يعدم، قام المسئول باختبار شحنة بعد وصولها بأخذ عينة عشوائية لعدد أربع من الطماطم. المطلوب إيجاد الاحتمالات التالية :

أ) لا يوجد طماطم غير صالح.

ب) عدد ٢ من الطماطم غير صالح.

ج) عدد ٢ من الطماطم غير صالح على الأقل.

د) على الأكثر واحد من الطماطم غير صالح.

٤- من خبرة الطبيب البيطري السابقة فإن ٣٠ ٪ من العجول المصابة بالأمراض يمكن شفاؤها. تم اختبار عدد ستة عجول في مشروع محلي مصابة بالمرض. المطلوب حساب الاحتمالات التالية :

أ) شفاء جميع العجول الستة.

ب) شفاء عجلين على الأكثر.

ج) عدم شفاء أي من العجول.

د) ارسم التوزيع الاحتمالي باستخدام الشكل التخطيطي وحدّد هل هو متماثل أم ملتو ؟ احسب المتوسط والانحراف المعياري.

٥- يتوزع عدد الحوادث الشهرية في شركات تعبئة اللحوم تبعاً لتوزيع بواسون إذا كان متوسط التوزيع ٠,٥ خلال الشهر الأخير. المطلوب حساب الاحتمالات التالية :

أ) عدم وجود حوادث.

ب) حصول حادث واحد.

ج) حصول أكثر من حادث - (استخدام قاعدة المتممة).

٦- تتوزع المكالمات الواردة لشركة زراعية تبعاً لتوزيع بواسون بمتوسط ثلاث مكالمات في الساعة. ما احتمال أن يصل للشركة خلال الساعة القادمة المكالمات التالية :

أ) على الأقل مكالمتين ؟

ب) مكالمة واحدة ؟

ج) أن لا يصل للشركة أي مكالمات؟

٧- يجرى قسم المبيعات بشركة لبيع بذور القطن ٥٠٠ مكالمة يومياً خلال موسم العمل . فإذا كان احتمال عقد صفقة بيع خلال المكالمة يساوي ٠,٠٢ استخدم تقريب توزيع بواسون للتوزيع ثنائي الحدين لإيجاد احتمال الحصول خلال ال ٥٠٠ مكالمة على :

أ) عشر صفقات بيع فقط.

ب) أكثر من خمسة عشر صفقه بيع.

٨- للتوزيع الطبيعي القياسي أوجد الاحتمالات التالية عندما تكون قيمة z :

أ) أقل من ١ .

ب) بين ١ ، ١,٥ .

ج) أكبر من ٢,١٧.

د) بين صفر و - ٠,٨٣ .

هـ) أقل من - ١,٦٦ .

ز) بين - ١,٢ و ٠,٣٤ .

٩- باستخدام جدول التوزيع الطبيعي القياسي أوجد قيمة الاحتمالات التالية :

أ) $p(z < -z_0) = 0.05$

ب) $p(z > z_0) = 0.1628$

ج) $p(-z_0 < z < z_0) = 0.7540$

د) $p(z > z_0) = 0.0485$

١٠- يتوزع وزن الفواكه المعبأة من قبل شركة الفواكه طبيعياً بمتوسط $u=10$ أوقية وانحراف معياري $\sigma=0.4$ أوقيه. إذا تم اختيار أحد العبوات عشوائياً ما احتمال أن يكون وزنها :

أ) أقل من عشر أوقيات.

ب) أقل من ١٠,٧ أوقيه.

ج) بين ٩,٦ و ١٠,٧ أوقيه.

١١- يتوزع وزن العجول حديثة الولادة طبيعياً بمتوسط ٤ أرطال وانحراف معياري ٠,٥ رطل. إذا تم اختيار أحد العجول من مجموعة كبيره من العجول حديثة الولادة فما احتمال أن يكون وزنه :

أ) أكثر من ٣,٥ رطل.

ب) بين ٤,٥ و ٥ رطل.

ج) بين ٣,٥ و ٤,٥ رطل.

١٢- إذا كان عدد أزهار الأقحوان المزروعة في أصيص مقاس ٦ بوصه تتوزع طبيعياً بمتوسط ١٦ وانحراف معياري ٢ ما احتمال أن أحد الأصص المختارة من بيت محمي كبير مملؤ بالأصص المزروعة يحتوي على :

أ) أقل من ١٤ زهرة.

ب) أقل من ٢٠ زهرة.

ج) من ١٢ إلى ٢٠ زهرة.

د) ٩٠ ٪ من الأصيص سوف تحتوي على أكثر من .... زهرة.