الإحصاء - المحاضرة الخامسة

Statistics and probabilities- Lecture 5

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

2021

في هذه المحاضرة سيتم التطرق لـ

بعض التوزيعات الإحتمالية المستمرة Some continuous probability distributions

- التوزيع المنتظم uniform distribution،

- التوزيع الأسي Exponential distribution،

- التوزيع الطبيعي Normal distribution.

- مفهوم الكثافة الإحتمالي Probability density.

- توزيع ستيودنت Student distribution.

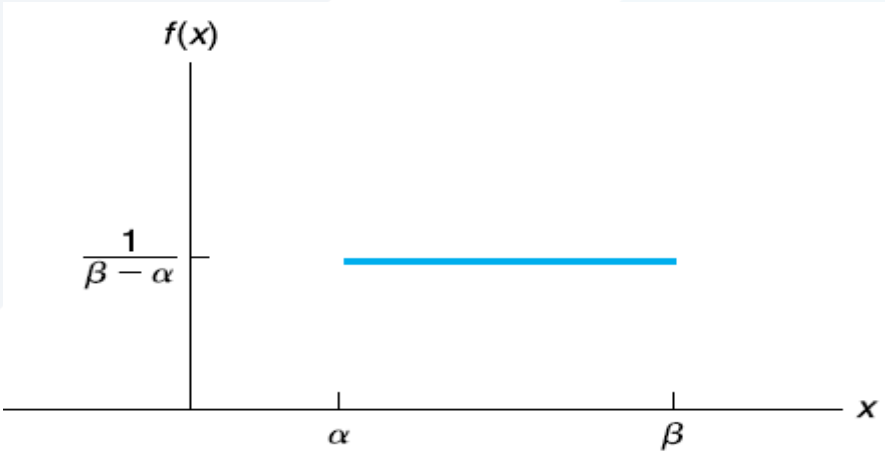
9-2-2- التوزيعات الاحتمالية المستمرة Continuous probability distribution:

9-2-2-1- التوزيع المنتظم *uniform distribution*:

يخضع متغير عشوائي X_i للتوزيع المنتظم، إذا كان تابع الاحتمال المترابط به يُعطى كما يأتي:

$$p(X=i)=\begin{cases}\dfrac{1}{\beta-\alpha} & \alpha\le X\le\beta\\ 0 & otherwise\end{cases}$$

بعبارة أخرى إن احتمال أن يأخذ X أي قيمة ضمن المجال $[\alpha,\beta]$ هو احتمال متساوٍ ويُعطى بمقلوب طول هذا المجال $\frac{1}{\beta-\alpha}$، و أنّ احتمال أن يأخذ المتغير أي قيمة خارج هذا المجال يساوي الصفر، والشكل الآتي يعبر عن دالة الاحتمال للتوزيع المنتظم:



ومن الأمثلة الواضحة عن التوزيع المنتظم (لمتغير منقطع) هو المتغير العشوائي الذي يعبر عن الرقم العلوي عند رمي حجر النرد، حيث يُلاحظ أنّ احتمال أن يأخذ المتغير أي قيمة ضمن المجال $[1,6]$ هو احتمال متساوٍ ويقدر بـ$\frac{1}{6}$.

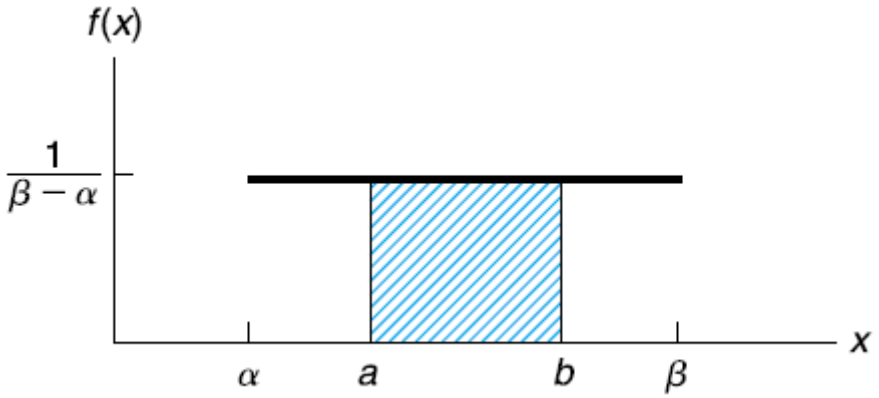
ملاحظة: قد يتم التساؤل هنا لماذا لم يتم القسم على 5 باعتبار أنّه وفقاً للقانون $\frac{1}{6-1}=\frac{1}{5}$ والجواب هنا إنّ قانون الحتمال للتوزيع المنتظم للمتغير المنقطع *disrcrete uniform value* يُعطى كما يأتي:

$$(X=i)=\begin{cases}\dfrac{1}{n} & \alpha\le X\le\beta\\ 0 & otherwise\end{cases}$$

حيث n، هي عدد القيم في المجال $[\alpha,\beta]$.

كذلك من مزايا التوزيع المنتظم أنّه إذا كان X متغير عشوائي يخضع للتوزيع المنتظم ضمن المجال $[\alpha, \beta]$، فإنّ احتمال أن يأخذ هذا المتغير أي قيمة ضمن مجال جزئي (يقع ضمن المجال الأصلي) وليكن $[a, b]$يساوي طول المجال الجزئي $[a, b]$ على طوال المجال الأصلي $[\alpha, \beta]$ أي:

$$p(a \leq X \leq b) = \frac{b-a}{\beta-\alpha}$$



يُعطى التوقع الرياضي للتوزيع المنتظم:

$$E(X) = \frac{\alpha+\beta}{2}$$

بينما التباين $Var(X)$ يُعطى كما يأتي:

$$Var(X) = \frac{(\beta-\alpha)^2}{12}$$

مثال 30:

إذا كان X_i متغير يخضع للتوزيع المنتظم ضمن المجال $[0, 10]$، فما هو احتمال:

a) أن يقع $2 < X < 9$

b) أن يقع $1 < X < 4$

c) $X < 5$

d) $X > 6$

e) ما هو التوقع $E(X)$، والتباين $Var(X)$

الحل :

بالنسبة للاحتمال الأول، يلاحظ أنّ المجال $]2,9[$ هو مجال جزئي ضمن المجال $[0,10]$ بالتالي وفقاً للخاصية السابقة وباعتبار أنّ المجال الأصلي $[0,10] = [\alpha,\beta]$ فإنّ احتمال

$2 < X < 9$ يساوي

$$p(a < X < b) = \frac{b-a}{\beta-\alpha}$$

$$p(2 < X < 9) = \frac{9-2}{10-0} = \frac{7}{10}$$

أمّا احتمال $1 < X < 4$

$$p(1 < X < 4) = \frac{4-1}{10-0} = \frac{3}{10}$$

أمّا احتمال $X < 5$ فهو احتمال أن $0 < X < 5$ أي يقع ضمن المجال $]0,5[$ يساوي :

$$p(0 < X < 5) = \frac{5-0}{10-0} = \frac{5}{10}$$

أمّا احتمال $X > 6$ فهو احتمال أن $6 < X < 10$ أي يقع ضمن المجال $]6,10[$ يساوي:

$$p(6 < X < 10) = \frac{10-6}{10-0} = \frac{4}{10}$$

التوقع،

$$E(X) = \frac{\alpha+\beta}{2} = \frac{0+10}{2} = 5$$

بينما التباين $Var(X)$ يُعطى كما يأتي:

$$Var(X) = \frac{(\beta-\alpha+1)^2-1}{12}$$

$$= \frac{(10-0+1)^2-1}{12} = \frac{120}{12} = 10$$

9-2-2-2- التوزيع الأسي Exponential random variable:

في توزيع بواسون Poisson distribution يمكن حساب احتمال حدوث عيب في مصباح كهربائي أو سلك نحاسي ما، ولكن يعتبر من المهم أيضاً معرفة الوقت المنقضي لكي يحدث أول عطل في المصباح الكهربائي أو معرفة طول سلك النحاس حتى نقطة اكتشاف العيب، ويعتبر المتغير العشوائي الذي يعبر عن هذا المقدار متغير مستمر و يخضع للتوزيع الإحتمالي الأسي والذي يُعطى تابع كثافته الإحتمالي Probability density function كما يأتي:

$$f(X) = \begin{cases} \lambda e^{-\lambda x} & X \geq 0 \\ 0 & X < 0 \end{cases}$$

حيث λ يعبر عن معامل التوزيع أو ما يعرف بالمعدل.

يعتبر التوزيع الأسي ملائماً من الناحية العملية في حساب احتمال أن يبقى نظام ما قيد العمل لفترة زمنية معينة أو بعبارة أخرى يتعلق بالتوزيع الاحتمالي لمقدار الزمن حتى وقوع حدث ما. مثل الوقت المستغرق لحدوث عطل في مصباح كهربائي أو في آلة ما، الزمن اللازم لاستهلاك بطارية جهاز ما، الزمن المستغرق لاستقبال مكالمة خاطئة.

ولفهم صيغة تابع الكثافة الإحتمالي للتوزيع الأسي، سيتم تعريف متغير Y يعبر عن عدد النقاط المعيبة في جزء من سلك نحاسي يبلغ طوله x سم فإذا كان متوسط عدد العيوب في السنتمتر الواحد λ، وبالتالي لحساب الاحتمال باستخدام توزيع بواسون:

$$p(X > x) = p(Y = 0) = e^{-\lambda x}\frac{(\lambda x)^0}{0\,!} = e^{-\lambda x}$$

حيث يعبر المقدار λx متوسط عدد العيوب في السلك ذو الطول x.

وبالتالي،

$$F(x) = p(X < x) = 1 - e^{-\lambda x}$$

ويعرف تابع الكثافة الإحتمالي بأنّه مشتق تابع التوزيع، أي أنّ تابع الكثافة الإحتمالي للتوزيع الأسي:

$$f(X) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

يُعطى تابع التوزيع الاحتمالي الأسي $F(x)$ والتوقع الرياضي والتباين كما يأتي:

$$F(x) = p(X \leq x) = 1 - e^{-\lambda x}$$

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

مثال 31:

في شركة ضخمة للشبكات الحاسوبية إذا كان عدد مرات الدخول (log-on) للنظام يعبر عن متغير عشوائي خاضع لتوزيع بواسون بمتوسط 25 مرة دخول في الساعة، وفي حال أرادت الشركة تقييم مواردها ومستوى الطاقم لمقابلة احتياجات خدمات الزبائن، كان لا بدّ من حساب الاحتمالات الآتية:

ما هو احتمال عدم وجود دخول للنظام خلال فترة ستة دقائق؟

الحل:

بفرض X متغير عشوائي يعبر عن الوقت بالساعات من البداية وحتى حالة أول دخول للنظام، بالتالي إنّ المتغير X يخضع للتوزيع الاسي بمعامل $\lambda = 25$، وبما إنّ متوسط عدد مرات الدخول مُعطى خلال الساعة سيتم التعبير عن الوقت بالساعات أيضاً أي أنّ 6 دقائق تساوي 0.1 ساعة، وبالتالي:

$$p(X > 0.1) = e^{-25*0.1} = 0.082$$

وما هو احتمال أن يكون الدخول الثاني للنظام بين 2 و 3دقائق، أي بين 0.033 و 0.05 ساعة؟

$$p(0.033 < X < 0.05) = e^{-25*0.05} - e^{-25*0.033} = 0.125$$

ما هو الوقت المقابل لاحتمال يعادل 90% لعدم وجود حالة دخول للنظام ؟

أي $p(X > x) = e^{-25*x} = 0.9$

$$-25 * x = \ln(0.9)$$

$$x = 0.00421 \text{ ساعة}$$

أي 0.25 دقيقة.

يمكن أيضاً حساب مدة الوقت المتوسط أو المتوقع للدخول التالي (أي بين دخول وخول آخر) للنظام باستخدام:

$$E(X) = \frac{1}{\lambda} = \frac{1}{25} = 0.04 \text{ ساعة} = 2.4 \text{ دقيقة}$$

أما تباين مدة الوقت للدخول التالي للنظام :

$$Var(X) = \frac{1}{\lambda^2} = \frac{1}{625} = 0.0016 \text{ ساعة} = 0.096 \text{ دقيقة}$$

خواص التوزيع الأسي:

الخاصية الأولى : اللاذاكرة Memoryless:

التوزيع الأسي يتمتع بخاصية وهي أنّه بلا ذاكرة (Memoryless) بمعنى آخر سلوك المغير في الماضي ليس له تأثير على سلوكه المستقبلي. بحيث أنّ كل لحظة هي بداية لفترة عشوائية جديدة ، والتي لها نفس التوزيع بغض النظر عن مقدار الوقت المنقضي بالفعل. وعبارة "بلا ذاكرة" تعني أنّه "ينسى" ما سبقه. أي إذا واصلنا الانتظار ، فإن طول الوقت الذي يتم انتظاره لا يزيد أو يقلل من احتمالية وقوع الحدث. يمكن تحديد أي وقت على أنه وقت صفر. وبفرض أنّ زلزال يضرب منطقة ما فإنّ احتمال وقوع زلزال آخر في غضون أسبوع أو شهر أو عشر سنوات من تلك النقطة كلها متساوية.

يمكن صياغة ما سبق بلغة الاحتمال كما يأتي:

$$p(X > s + t \,|X > t) = p(X > s) = 1 - p(x \leq s)$$

$$p(X > s) = 1 - \left(1 - \boldsymbol{e}^{-\lambda s}\right) = \boldsymbol{e}^{-\lambda s}$$

لتوضيح ذلك:

$$p(X > s + t \,|X > t) = \frac{p(X>s+t \text{ and } X>t)}{p(X>t)}$$

$$p(X > s + t \,|X > t) = \frac{p(X>s+t\,)}{p(X>t)}$$

$$p(X > s + t \,|X > t) = \frac{\boldsymbol{e}^{-\lambda(s+t)}}{\boldsymbol{e}^{-\lambda t}}$$

$$p(X > s + t \,|X > t) = \boldsymbol{e}^{-\lambda s} = \boldsymbol{p(X > s)}$$

مثال 32 :

بفرض أنّ عدد الاميال X التي يمكن أن تقطعها سيارة ما قبل أن تنفذ بطاريتها يخضع للتوزيع الأسي مع قيمة متوسطة تبلغ 10000 ميل، فإذا أراد شخص قطع رحلة 5000 ميل، فما هو احتمال أن تقطع السيارة هذه المسافة من دون ان تنفذ بطاريتها.

يمكن الاستفادة من خاصية الذاكرة الضعيفة للمتغير الخاضع للتوزيع الأسي :

$$p(X > 5) = 1 - \left(1 - \boldsymbol{e}^{-\lambda 5}\right) = \boldsymbol{e}^{-\frac{1}{10}*5} = \boldsymbol{e}^{-\frac{1}{2}} = 0.604$$

أمّا إذا كان X لا يخضع للتوزيع الأسي فإنّ الاحتمال المطلوب يساوي:

$$p(X > s + t \,|X > t) = \frac{p(X>t+5\,)}{p(X>t)} = \frac{1-F(t+5)}{1-F(t)}$$

حيث يعبر t عن الوقت المنقضي على استخدام بطارية السيارة قبل البدء بالرحلة، بمعنى آخر إنّ حساب هذا الاحتمال عندما لا يكون X خاضع للتوزيع الأسي يحتاج لمعلومة إضافية عن t، بينما يعتبر هذا المقدار غير مهم في الحالة التي يخضع فيها X للتوزيع الأسي.

الخاصية الثانية:

ليكن $X_1, X_2, \dots \dots . X_n$ سلسلة من n متغير عشوائي مستقل يخضع كل منها للتوزيع الأسي وفاً للمعاملات الآتية :

$$\lambda_1, \lambda_2, \dots \dots . \lambda_n$$

فإنّ المقدار $\min(X_1, X_2, \dots \dots . X_n)$ يخضع للتوزيع الأسي بمعامل $\sum_{i=1}^{n} \lambda_i$،

يمكن برهان ذلك بسهولة انطلاقاً من الاحتمال الآتي:

$$p[min(X_1, X_2, \dots \dots . X_n) > x] = p(X_1 > x, X_2 > x, \dots \dots . X_n > x)$$
$$= \prod_{i=1}^{n} p(X_i > x) = \prod_{i=1}^{n} e^{-\lambda_i x} = e^{-\sum_{i=1}^{n} \lambda_i x}$$

مثال 33:

بفرض أن نظام ما مكوّن من عدة مكونات n مستقلة عن بعضها البعض، وبفرض أنّه عند حدوث أي عطل في أي مكوّن فإنّ هذا الأمر ينعكس على كامل النظام، فما هو احتمال أن يعمل النظام للفترة t؟

بالاستفادة من الخاصية السابقة يمكن صياغة الاحتمال الآتي وهو احتمال أن يعمل النظام للفترة t:

$$p[min(X_1, X_2, \dots \dots . X_n) > t] = e^{-\sum_{i=1}^{n} \lambda_i t}$$

9-2-2-3- التوزيع الطبيعي *Normal distribution*:

يعتبر التوزيع الطبيعي من أشهر التوزيعات الاحتمالية للمتغيرات المستمرة[1]، ويستمد اسمه من كونه يُستخدم لحساب احتمال وقوع الأحداث الطبيعية أو المرتبطة بالحياة الطبيعية.

إذا كان لدينا ثابتين $\mu \in \mathbb{R}$ و $\sigma > 0$ ، فنقول أن المتغير العشوائي X_i يخضع للتوزيع الطبيعي إذا كان قانون احتماله يعطى بالعلاقة الآتية:

$$f_x(X_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{X_i-\mu}{\sigma})^2}, \quad X \in \mathbb{R}: \ -\infty < X < +\infty$$

[1] قبل البدء بقراءة هذا المبحث يُفضل مراجعة ما تمّ عرضه عن تابع الكثافة الإحتمالية density function وتابع التوزيع الإحتمالي التراكمي cumulative function في المحاضرات السابقة.

بحيث يمثل μ المتوسط الحسابي للمجتمع *arithmetic mean of distribution* و σ الانحراف المعياري *standard deviation* للمجتمع والذي يكون غالباً غير معلوم.

نلاحظ أن شكل تابع الكثافة الاحتمالي يملك الخصائص الآتية:

1- للتوزيع الاحتمالي الطبيعي يشبه شكل الجرس bell-shaped، بحيث يملك قمة واحد (one peak) في مركز التوزيع. كذلك في التوزيع الطبيعي تكون قيم كل من الوسط الحسابي، الوسيط والمنوال متساوية وتتمركز في وسط التوزيع. ويعود شكل الجرس إلى طبيعة الحوادث في الحياة الطبيعية أو الواقعية. حيث من الملاحظ، أنّ الحوادث ذات القيم المتوسطة (المتكررة أو العادية) يكون احتمال حدوثها عالياً مقارنة بالحوادث الشاذة أو المتطرفة. فمثلاً، قد يكون من المحتمل جداً أن يكون طول شخص بالغ 1.7م (كقيمة متوسطة)، ولكن بالمقابل قد يكون من النادر جداً في هذه الحياة أن نرى شخص طوله 2.5 م. من أجل ذلك نلاحظ في الشكل السابق أنّه في المنتصف يكون تابع الكثافة ذو قيم عالية بينما على الجوانب ينخفض حتى يصبح قريباً جداً من المحور الأفقي.

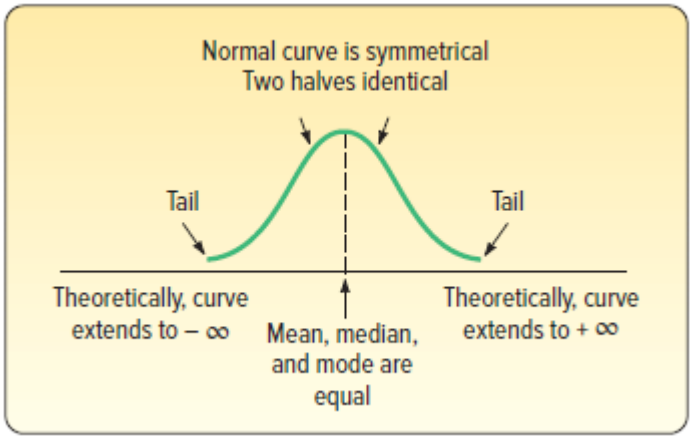
تبلغ إجمالي المساحة تحت منحنى التوزيع الاحتمالي الواحد، وتقع نصف المساحة تحت المنحنى على يمين قيمة الوسط الحسابي (أو مركز التوزيع)، بينما يقع النصف الآخر على يساره.

2- التوزيع الإحتمالي الطبيعي هو توزيع متناظر symmetrical distribution حول المتوسط، بمعنى إذا تمّ قطع منحنى التوزيع الطبيعي بشكل عمودي في قيمة المركز إلى قسمين، فإنّ أحد القسمين سيكون مرآة للقسم الآخر. كذلك تكون إجمالي مساحة كل قسم مساوية لـ 0.5.

3- يلاحظ أنّ منحنى التوزيع الطبيعي ينخفض على جانبي قيمة المركز (المتوسط) ليصبح قريباً من المحور X دون أن يلامسه، ويطلق على هذا الجزء من المنحنى من الجانبين بالذيل tail.

4- يتحدد موقع التوزيع (location) عن طريق قيمة المتوسط μ، بينما يتحدد تشتت التوزيع عن طريق الانحراف المعياري standard deviation، σ.

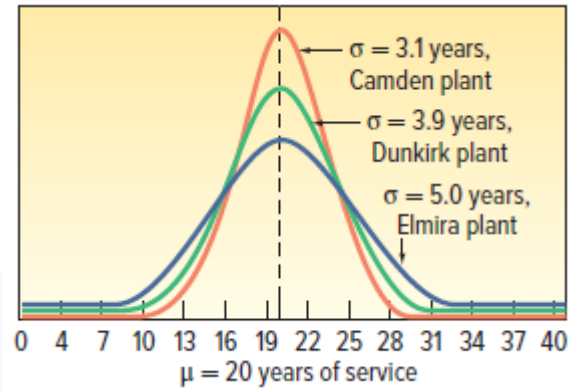
الخصائص السابقة تظهر واضحة في الشكل الآتي:



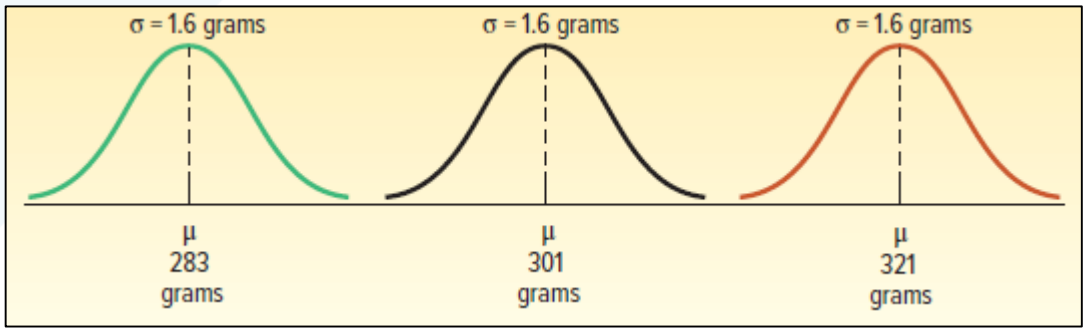
إضافة للخصائص السابقة نورد الملاحظات الآتية:

ملاحظة 1: تجدر الإشارة إلى أن شكل وموقع التوزيع الطبيعي يختلف حسب قيمة المتوسط μ والانحراف المعياري.

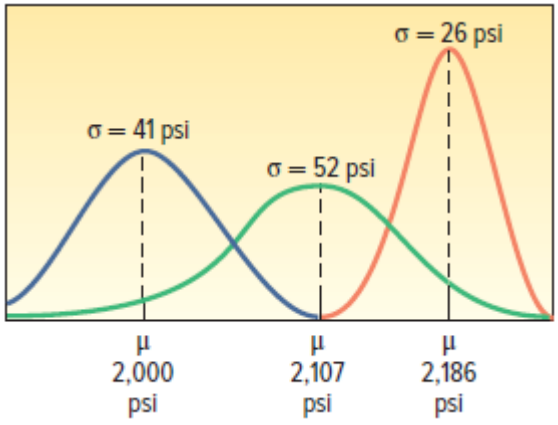
فمثلاً، قد نجد عدة توزيعات طبيعية لها نفس الموقع (*location*) أي لها نفس المتوسط μ ولكنّها تختلف في الشكل، كما في المثال الآتي الذي يعرض التوزيع الطبيعي لمدة خدمة العمال في ثلاث مصانع (*Camden, Dunkirk, Elmira*):



كذلك قد يكون هناك عدة توزيعات احتمالية طبيعية تملك نفس الشكل ولكن تختلف فيا الموقع كما المثال الآتي الذي يعرض منحنى التوزيع الإحتمالي لأوزان ثلاث صناديق من الحبوب:



وأخيراً يوجد عدة توزيعات احتمالية طبيعية تختلف فيما بينها من ناحية الشكل والموقع معاً، كما في المثال الآتي الذي يوضح التوزيع الطبيعي لقوى الشد مقاسة بالرطل (*Psi*) لثلاث أنواع من الأكبال:



ملاحظة 2: في التوزيعات الإحتمالية المنقطعة (مثل توزيع بيرنوللي، الثنائي، بواسون) قد يكون للإحتمالات النقطية (الحصول على قيمة محددة) أهمية أو معنى مثل حساب احتمال نجاح عدد من الطلاب، أو احتمال تأخر عدد من رحلات الطيران وغالياً ما يطلق على التوزيع الإحتمالي للمتغير المنقطع بتابع الكتلة الإحتمالي *Probability Mass function (PMD)*. ولكن في التوزيعات الإحتمالية المستمرة قد تكون الإحتمالات النقطية غير مهمة مثلاً قد لا يهتم مزارع ما باحتمال ان يكون محصوله بالضبط 3 طن من دون زيادة او نقصان، بل يهتم أكثر باحتمال أن يكون محصوله بين 2.5 و 3 طن مثلاً، أي يهتم باحتمالات مجالية وليس نقطية. من أجل ذلك يطلق على التوزيع الإحتمالي بتابع الكثافة الإحتمالية *probability density function (PDF)* أي مقدار كثافة المساحة تحت منحنى التوزيع الإحتمالي. وللتوضيح أكثر إنّ احتمال ان يأخذ متغير منقطع (عدد الأولاد مثلاً) قيمة ما مثل 2 هو احتمال وارد جداً، ولكن احتمال أن يساوي متغير مستمر (كمعدل هطول الأمطار) قيمة تساوي 3 مم بالضبط من دون زيادة أو نقصان ذرة مياه صغيرة هو احتمال قليل جدا بل قريب جداً من الصفر(مثلاً إذا قمت برسم خط عمودي معبر عن احتمال لقيمة ما، والسؤال ما هي مساحة هذا الخط ؟؟؟). لذلك في التوزيعات الاحتمالية للمتغيرات المستمرة لا يتم استخدام نفس المنهجية، بمعنى آخر لا يتم حساب احتمال ان يساوي المتغير قيم معينة بل حساب احتمال أن يدور المتغير في مجال حول نقطة معينة $[x, x+dx]$. فإذا كان الإحتمال الأول يساوي الصفر فإنّ الثاني قد يكون قريب من الصفر ولكن لا يساوي الصفر (0.00001 مثلاً) يتم الحديث عن تابع الكثافة الإحتمالي[2] *probability density function* أي يعبر عن كثافة الاحتمال القريب من نقطة ما ($P(|x-3|>\varepsilon)$ حيث ε عدد متناه في الصغر. أو بعبارة اخرى إنّ احتمال أن يقع المتغير X في المجال $[x, x+dx]$ هو[3] $f(x)dx$ بمعنى آخر $P(x<X<x+dx)=f(x)dx$ وهو المساحة المحصورة تحت منحى $f(x)$ بين x و $x+dx$، حيث dx هو الفرق الضئيل جدا بين قيمتين لـ X، و $f(x)$ هو

[2] هنا يتم الحديث عن كثافة الإحتمال وليس الاحتمال، ولكن السؤال هل يمكن ل pdf أن يكون أكبر من الواحد؟ في بعض الاحيان نعم فمثلاً لو في التوزيع المنتظم $f(x)=\frac{1}{b-a}$، فما لو كان المجال هو $[0, 0.5]$، هنا نلاحظ انّ $f(x)$ يساوي 2 !!! ولكن من ناحية أخرى غنّ تابع التوزيع الإحتمالي التراكمي *CDF* أو المساحة تحت المنحى يجب ان تساوي الواحد !!!

[3] قارن مع القانون الفيزيائي أنّ الكتلة mass تساوي الكثافة * الحجم !!!!.

تابع الكثافة الإحتمالي *PDF* الذي يعبر عن مقدار تركز الإحتمال في مقدار dx من المتغير المستمر أي مقدار كثافة الاحتمال حول x.

في التوزيعات المستمرة أيضاً يكون تابع التوزيع التراكمي (*CDF*) *cumulative probability function* والذي يساوي تكامل تابع الكثافة الإحتمالي *pdf* أي جبرياً إن تابع الكثافة الإحتمالي *PDF* هو مشتق التوزيع الإحتمالي التراكمي *CDF* والذي يأخذ شكل حرف (*S*). وهنا يجدر القول أنّ شكل حرف (*S*) يدل على أنّ قيمة المشتق لتابع *CDF* والمقابلة لقيم المتغير X الواقعة في المنتصف هي اكبر (ميل أكبر) من غيرها مقارنة بالقيم المتطرفة (الواقعة على يمين أو يسار القيم الواقعة في المنتصف):

أي انّ الكثافة الإحتمالية تُعطى رياضياً كما يأتي:

$$f(x) = \lim_{\Delta \to 0} \frac{p(x < X \leq x + \Delta)}{\Delta}$$

$$f(x) = \lim_{\Delta \to 0} \frac{F(x + \Delta) - F(x)}{\Delta} = \frac{\partial F(x)}{dx}$$

كذلك في التوزيعات الإحتمالية المستمرة فإنّ المساحة أسفل منحنى التوزيع الإحتمالي تمثل احتمالات متعددة. ويكون إجمالي المساحة تحت المنحنى تساوي الواحد. ولأن التوزيع الإحتمالي الطبيعي هو توزيع متناظر *symmetric* حول المتوسط. فإنّ المساحة على يمين المتوسط تساوي 0.5 وتساوي المساحة على يسار المتوسط. فمثلاً في مثال قوى الشد في الكابلات السابق، إنّ احتمال أن يكون قوة الشد لإحدى الكابلات أقل من 2000 رطل (*Psi*) يساوي 0.5، ويساوي احتمال أن يكون قوة الشد لإحدى الكابلات أعلى من 2000 رطل ويساوي أيضاً 0.5.

- **التوزيع الطبيعي المعياري *standard normal probability distribution*:**

نظراً لتعدد أشكال ومواقع التوزيع الطبيعي قد يلجأ المختصون إلى تحويل القيم العادية للمتغير إلى قيم معيارية عن طريق طرح الوسط الحسابي *arithmetic mean* (μ) من كل قيمة من القيم الأصلية وقسمتها على الانحراف المعياري *standard deviation* (σ).

$$Z_i = \frac{X_i - \mu}{\sigma}$$

وعند التعامل مع العينة بدلاً من المجتمع يتم التعويض بقيم المتوسط الحسابي $\bar{x}$، والانحراف المعياري s للعينة:

$$z_i = \frac{X_i - \bar{x}}{s}$$

ويتم التحويل لقيم معيارية لعدة أسباب، نذكر منها:

1. التمكن من مقارنة خصائص العينات مع بعضها حيث تكون القيم المعيارية قابلة للمقارنة لأنها تأخذ كل من مقياس التشتت (الوسط الحسابي) ومقايس النزعة المركزية (بعين الاعتبار).

2. تحويل القيم الكبيرة إلى قيم صغيرة مع المحافظة على خصائصها والمعلومات المتوفرة بها، الأمر الذي يسهل عمليات الحساب والتحليل

3. ملاءمة طبيعية تابع التوزيع الذي يعطي قيم احتمالية للقيم السالبة الأمر الذي لا يتماشى مع طبيعة البيانات في الحياة العادية (مثل أطوال البشر، إنتاج المحصولات، هطولات الأمطار) وبالتالي عند تحويل القيم العادية إلى قيم معيارية يكون من الممكن قبول القيم السالبة.

4. كما رأينا سابقاً أنّه يمكن التعبير عن التوزيع الإحتمالي بثلاث طرق (جدول احتمالي، أو منحنى بياني، أو قانون احتمال) ولكن نظراً لوجود عدد كبير من التوزيعات الطبيعية (حسب اختلاف قيم المتوسط، والانحراف المعياري) يكون من الصعب توفير جدول للقيم الإحتمالية. لذلك اقتضت الحاجة إلى توحيد قيمة المتوسط والانحراف المعياري لبناء جدول التوزيع الإحتمالي، وهذا ما يوفره التوزيع الطبيعي المعياري.

المثال الآتي يوضح الأسباب السابقة:

لنفترض أنّ لدينا ثلاث قيم:

10 20 30

يُلاحظ من القيم السابقة أنّ القيمة الواقعة في المنتصف 20 تمثل قيمة الوسط الحسابي *arithmetic mean* والوسيط *Median*. وأنّ القيمة الثالثة 30 تساوي مثلي القيمة الأولى 10.

أمّا الانحراف المعياري يساوي:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{3}(x_i - \bar{x})^2}{3}}$$

$$\sigma = \sqrt{\frac{(10-20)^2 + (20-20)^2 + (30-20)^2}{3}}$$

$$\sigma = \sqrt{\frac{100+0+100}{3}} = \sqrt{\frac{200}{3}} \approx 8.165$$

لتحويل القيم الأصلية السابقة لقيم معيارية يتم طرح قيمة الوسط الحسابي (20) من كل قيم وقسمة الناتج على الانحراف المعياري *8.16*:

المعيارية القيم *Standard values*

القيم الأصلية

$$z_i = \frac{X_i - \bar{x}}{s}$$

10 $$z_1 = \frac{X_1 - \bar{x}}{s} = \frac{10 - 20}{8.165} = -1.225$$

20 $$z_2 = \frac{X_2 - \bar{x}}{s} = \frac{30 - 20}{8.165} = 0$$

30 $$z_i = \frac{X_3 - \bar{x}}{s} = \frac{30 - 20}{8.165} = 1.225$$

يلاحظ من الجدول السابق أنّه تمّ تحويل القيم الأصلي (*30 20 10*) إلى قيم معيارية

(*-1.225 0 1.225*)

حيث تتمتع هذه القيم بصغر قيمتها مع الحفاظ على نفس المعلمات التي كانت متضمنة في القيم الأصلية (القيمة الثانية 0 هي قيمة الوسط الحسابي والقيمة الثالثة 1.225 تساوي مثلي القيمة الأولى *-1.225*). هذا بالإضافة إلى ظهر قيم سالبة الأمر الذي يساعد في تمثيلها بيانياً في منحنى التوزيع الطبيعي المعياري.

ملاحظة: إذا تمّ حساب الوسط الحسابي للقيم المعيارية نجد أنّه يساوي الصفر، كذلك إذا تمّ حساب الانحراف المعياري فهو يساوي الواحد. بالتالي إنّ قيمة الوسط الحسابي والانحراف المعياري للقيم المعيارية دائما تساوي **الصفر والواحد** على الترتيب.

ملاحظة: إن القيمة المعيارية Z تقيس مقدار البعد الاختلاف أو الفرق بين أي قيمة محدد والوسط الحسابي مقاساً بالانحراف المعياري. والمثال الآتي يبين أيضاً أحد فوائد تحويل البيانات إلى الشكل المعياري:

مثال 34:

لنفرض أن علامات الطلاب في أحد المقررات كانت كالتالي:

20، 15، 26، 32، 18، 28، 35، 14، 26، 22، 17

وكانت علامة النجاح هي 30 من 60، ويتضح من البيانات السابقة أن معظم علامات الطلاب أقل من المعدل المطلوب للنجاح، لذلك قرر أستاذ المقرر بأن يعتبر الطالب راسب في حال كان أقل من متوسط العلامات المحققة بأكثر من انحراف معياري واحد.

المتوسط الحسابي للعلامات المحققة هو 23، والانحراف المعياري 6,6

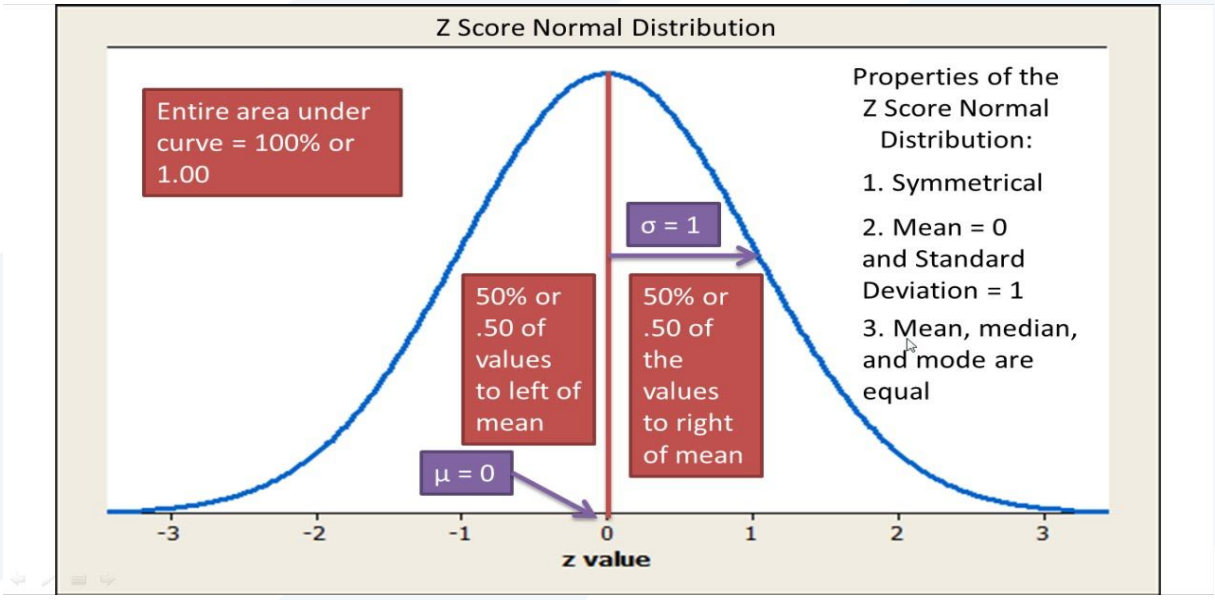
وبالتالي تصبح العلامات بعد طرح المتوسط الحسابي من كل منها وتقسيم الناتج على الانحراف المعياري كالتالي:

-0,45، -1,21، 0,45، 1,36، -0,76، -0,76، 1,82، -1,36، 0,45، -0,45، -0,15، -0,91. وبالتالي نرى أن هناك طالبين فقط يستحقان الرسوب.

يُطبّق قانون التوزيع الطبيعي في مجالات واسعة، في الاقتصاد، في الفيزياء، وعلم الأحياء. يعتبر هذا التوزيع هو التوزيع الذي تقترب منه عدة توزيعات منقطعة. إضافة إلى ذلك، يعد قانون الاحتمال الطبيعي التوزيع التقريبي لمجموع عدة متغيرات عشوائية مستقلة.

يُمكن تمثيل تابع الكثافة للتوزيع الطبيعي المعياري *Standard normality density function* بالشكل الآتي:

إن التوزيع الطبيعي المعياري يمتلك نفس الخصائص التي يمتلكها أي توزيع طبيعي ويّعطى قانون احتماله أو (تابع



كثافته الإحتمالية):

$$f_x(z) = \Phi(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z_i)^2}, \quad x \in \mathbb{R}: \ -\infty < x < +\infty$$

الجدول الآتي يعرض جزء من قيمة الاحتمال أو قيمة المساحة المحصورة بين القيمة المعياري $z = 0$ وأي قيمة أخرى:

z	0.00	0.01	0.02	0.03	0.04	0.05	...
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	
⋮							

مثلاً: إنّ إجمالي المساحة الواقعة بين القيمة المعيارية صفر، والقيمة 1.51 تساوي 0.4345. والسؤال كيف تم إيجاد هذه القيمة؟

إنّ القيمة 1.51 يمكن تجزئتها إلى جزئين الأول 1.5 والثاني 0.01 أي $1.51 = 1.5 + 0.01$، بالتالي يتم النظر إلى سطر القيمة 1.5 وعمود القيمة 0.01 ومن ثمّ تؤخذ القيمة الواقعة في تقاطع السطر والعمود وهي القيمة 0.4345.

مثال آخر، إنّ إجمالي المساحة المحصورة بين القيمة المعيارية صفر والقيمة المعيارية 1.73 هي 0.4582، لماذا؟

لأنّ القيمة 1.71 يمكن تجزئتها إلى 1.7 و 0.03 أي يتم النظر إلى سطر القيمة 1.7 وعمود القيمة 0.03 ومن ثم تؤخذ القيمة الإحتمالية الواقعة في تقاطع السطر والعمود وهي 0.4582.

- **تطبيقات التوزيع الطبيعي المعياري:**

يعتبر التوزيع الطبيعي المعياري مهم جداً لتحديد الاحتمالات لأي متغير عشوائي يخضع للتوزيع الطبيعي. والإجراء الأساسي يتمثل أولاً بإيجاد قيمة Z المعيارية كما راينا سابقاً، ومن ثمّ استخاجمها لإيجاد عدة احتمالات لقيم المتغير العشوائي، التطبيقات الآتية توضح هذا الأمر:

مثال 35:

على افتراض أنّ الدخل الأسبوعي لشركة تأجير السيارات يتبع التوزيع الطبيعي بمتوسط $\mu = 1000$ دولار، وانحراف معياري $\sigma = 100$ دولار، ما هي القيمة المعيارية Z لسائق يحقق 1100 دولار أسبوعياً، ولسائق يحقق 900 دولار أسبوعياً.

إنّ القيمة المعيارية تساوي:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

من أجل دخل 1100 دولار تصبح القيمة المعيارية

$$Z_i = \frac{1100 - 1000}{100} = 1$$

أي أنّ دخل 1100 هو أعلى من المتوسط بمقدار انحراف معياري واحد.

ومن أجل دخل 1100 دولار تصبح القيمة المعيارية

$$Z_i = \frac{900 - 1000}{100} = -1$$

أي أنّ دخل 900 هو أقل من المتوسط بمقدار انحراف معياري واحد.

- **إيجاد المساحة أسفل منحنى التوزيع الطبيعي Finding area under the normal curve:**

بالعودة للمثال السابق رقم 23، إذا طلب حساب أن يقع الدخل الأسبوعي للسائق في شركة تأجير السيارات بين 1000 دولار، و 1100 دولار؟

يلاحظ أنّ متغير الدخل هو متغير مستمر ويتبع التوزيع الطبيعي، وبالتالي إنّ حساب الاحتمال المطلوب يعادل حساب المساحة الواقعة تحن المنحني والمحصورة بين 1000 و 1100 دولار. ولإيجاد المساحة المذكورة يجب استخدام التوزيع الطبيعي المعياري لأنّ هذا التوزيع يملك متوسط معوم وثابت (يساوي الصفر) وانحراف معياري معلوم وثابت (يساوي الواحد). بالتالي يجب تحول القيمة 1000 و القيمة 1100 إلى قيم معيارية كالآتي:

$$Z_i = \frac{X_i - \mu}{\sigma} = \frac{1000 - 1000}{100} = 0$$

$$Z_i = \frac{X_i - \mu}{\sigma} = \frac{1100 - 1000}{100} = 1$$

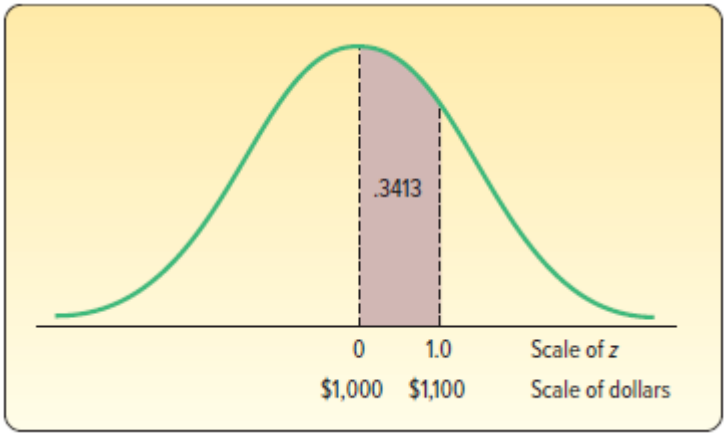
أي إنّ حساب احتمال أن يقع الدخل الأسبوعي بين 1000 و1100 يعادل احتمال أن تقع القيم المعياري بين 0 و1.

الجدول الآتي يبين المساحة تحت المنحنى لعدة قيم معيارية:

z	0.00	0.01	0.02
⋮	⋮	⋮	⋮
0.7	.2580	.2611	.2642
0.8	.2881	.2910	.2939
0.9	.3159	.3186	.3212
1.0	.3413	.3438	.3461
1.1	.3643	.3665	.3686
⋮	⋮	⋮	⋮

وفقاً لما تمّ عرضه سابقاً عن كيفية استخدام هذا الجدول، وبما إنّ المطلوب هو حساب المساحة الواقعة بين القيمة المعيارية صفر والقيمة المعيارية 1، نجد أنّ قيمة هذه المساحة هي 0.3413 (هي القيمة المظللة باللون الأحمر

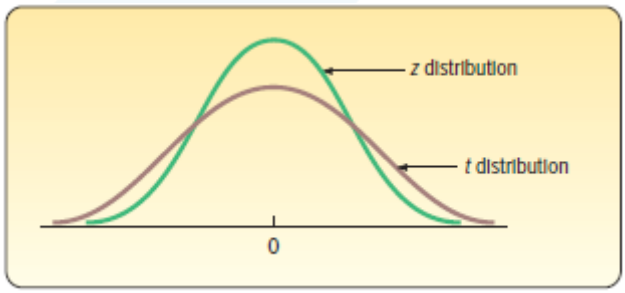
الغامق). وذلك لأنّ القيمة 1 يمكن تقسيمها إلى 1.0 و 0.00 بالتالي يتم النظر إلى سطر القيمة 1.0 وعمود القيمة 0.00 وتؤخذ القيمة الواقعة عند تقاطه هذا السطر مع العمود وهي القيمة 0.3413. وهو ما يتفق مع الشكل الآتي:



9-2-2-3- توزيع ستودنت *student's distribution*:

توزيع ستيودنت هو توزيع احتمالي مستمر ويملك نفس خصائص التوزيع الطبيعي (شكل الجرس، التناظر، وجود قيم متطرفة من الجانبين). ويشتق توزيع ستيودنت من التوزيع الطبيعي ولكن يتم استخدامه لحساب قيمة الاحتمال في العينات الصغيرة، لأنّ قيمة المتوسط الحسابي والانحراف المعياري يتم حسابهما من العينة عندما يكون الانحراف المعياري للمجتمع غير معلوم بسبب عدم القدرة على جمع عدد كبير من أفراد المجتمع.

والشكل الآتي يعرض مقارنة بين التوزيع الطبيعي المعياري Z normal distribution وتوزيع ستيودنت t-distribution:



حيث يلاحظ أنّ توزيع ستيودنت ذو قمة أوسع من التوزيع الطبيعي (more flatter)، ويمتد بشكل أكبر على الجانبين، وذلك لأنّ الانحراف المعياري لتوزيع ستيودنت أكبر من الانحراف المعياري في التوزيع الطبيعي.