الإحصاء والاحتمالات- المحاضرة الخامسة

Statistics and probabilities- Lecture 5

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

في هذه المحاضرة سيتم التطرق لـ

أشهر التوزيعات الإحتمالية المستمرة continuous probability distribution

- التوزيع الطبيعي Normal distribution.
- توزيع كاي مربع Chi squares distribution.
- - توزيع فيشر Fisher distribution.
- - مفهوم مجال الثقة Interval of confidence.

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

9-2-2-1- 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:

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

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 واحدة (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.

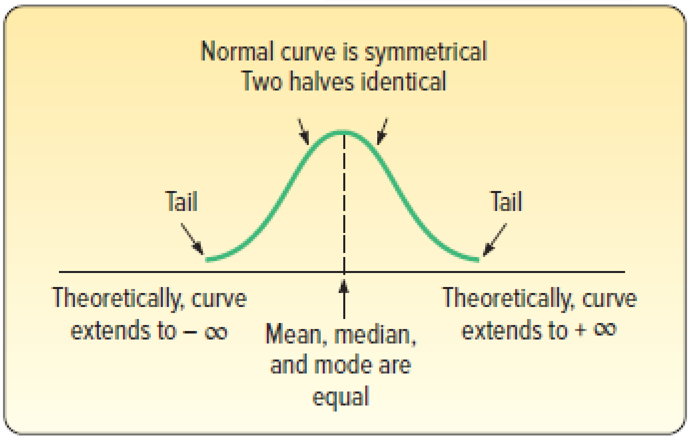
3- يلاحظ أنّ منحنى التوزيع الطبيعي ينخفض على جانبي قيمة المركز (المتوسط) ليصبح قريباً من المحور X دون أن يلامسه، ويطلق على هذا الجزء من المنحنى من الجانبين بالذيل tail.

---

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

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

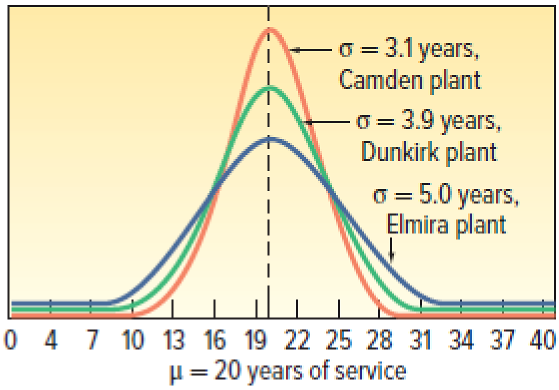
الخصائص السابقة تظهر واضحة في الشكل الآتي:



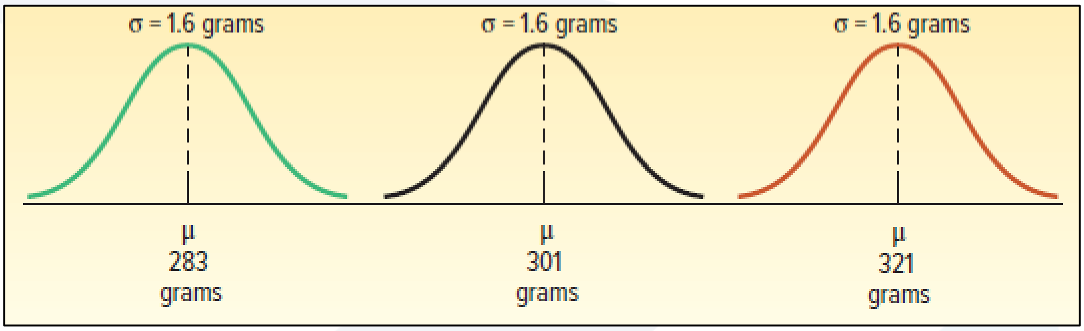
إضافة للخصائص السابقة نورد الملاحظات الآتية:

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

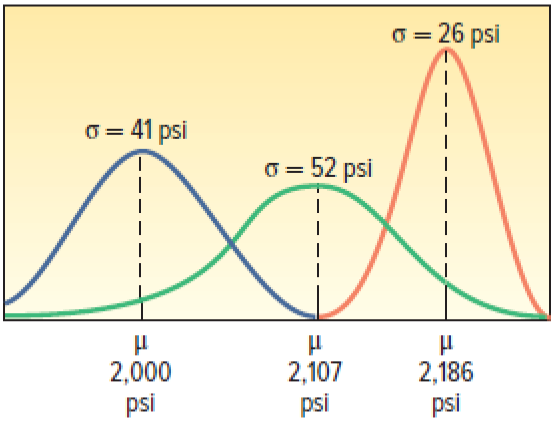
فمثلاً، قد نجد عدة توزيعات طبيعية لها نفس الموقع (*location*) أي لها نفس المتوسط μ ولكنّها تختلف في الشكل، كما في المثال الآتي الذي يعرض التوزيع الطبيعي لمدة خدمة العمال في ثلاث مصانع (*Camden, Dunkirk, Elmira*):



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 وغالباً 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).

هنا 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 أخرى 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 منحنى $f(x)$ بين x و $x+dx$، حيث dx هو الفرق الضئيل جدا بين قيمتين لـ X، و $f(x)$ هو 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:

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

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

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.

[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 !!!!.

- **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*:

القيم المعيارية *Standard values* $z_i = \frac{X_i - \bar{x}}{s}$	القيم الأصلية
$z_1 = \frac{X_1 - \bar{x}}{s} = \frac{10-20}{8.165} = -1.225$	10
$z_2 = \frac{X_2 - \bar{x}}{s} = \frac{20-20}{8.165} = 0$	20
$z_i = \frac{X_3 - \bar{x}}{s} = \frac{30-20}{8.165} = 1.225$	30

يلاحظ من الجدول السابق أنّه تمّ تحويل القيم الأصلي (*10 20 30*) 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 19:

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

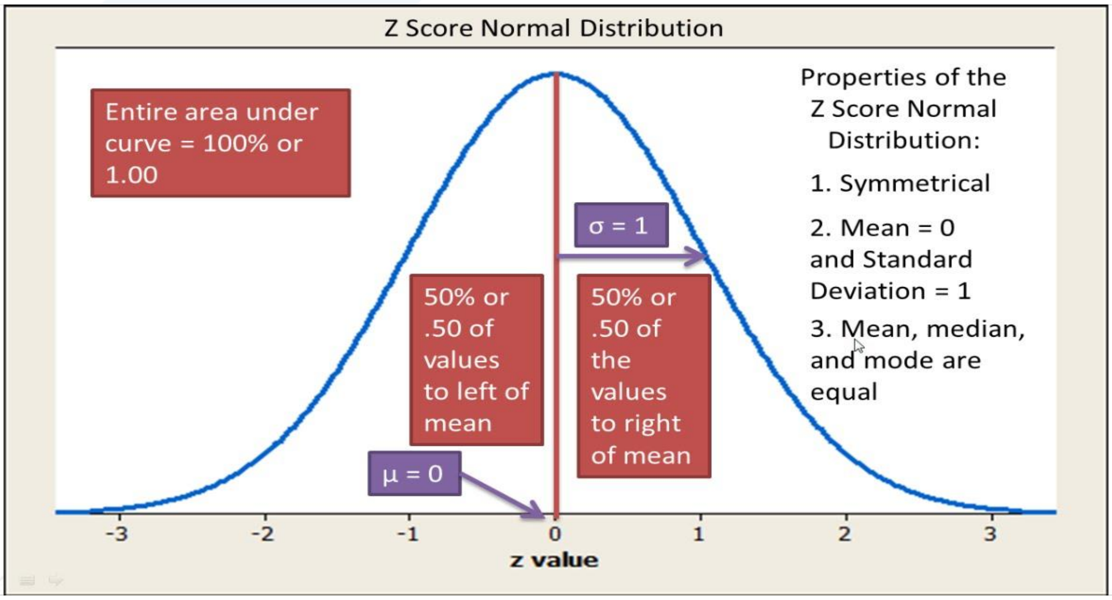
وبالتالي تصبح العلامات بعد طرح المتوسط الحسابي من كل منها وتقسيم الناتج على الانحراف المعياري كالتالي:

-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 بمكن 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 استخجامها لإيجاد عدة احتمالات لقيم المتغير العشوائي، التطبيقات الآتية توضح هذا الأمر:

مثال 20:

على افتراض أنّ الإنتاج السنوي لمصنع السيارات يتبع التوزيع الطبيعي بمتوسط $\mu = 1000$ سيارة، وانحراف معياري $\sigma = 100$ دولار، ما هي القيمة المعيارية Z لقيمة إنتاج تبلغ 1100 سيارة سنوياً، ولقيمة إنتاج يحقق 900 سيارة سنوياً.

إنّ القيمة المعيارية تساوي:

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

من أجل قيمة إنتاج 1100 سيارة تصبح القيمة المعيارية

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

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

ومن أجل قيمة إنتاج 900 سيارة تصبح القيمة المعيارية

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

أي أنّ قيمة إنتاج 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 احتمال أن تقع قيمة الإنتاج بين 1000 سيارة، و 1100 سيارة؟

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

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

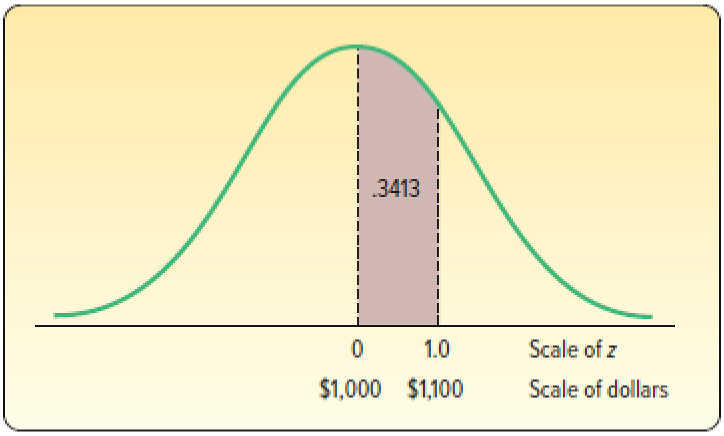
$$Z_i = \frac{X_i - \mu}{\sigma} = \frac{1100 - 1000}{100} = 1$$

أي إنّ حساب احتمال أن تقع قيمة الإنتاج بين 1000 و1100 يعادل احتمال أن تقع القيم المعيارية 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 تقاطع هذا السطر مع العمود وهي القيمة 0.3413. وهو ما يتفق مع الشكل الآتي:



مثال:

يعتقد المهندسون المدنيون أن مقدار الوزن W (بوحدات 1000 رطل) يعبر عن مقدار الوزن الذي يمكن أن تتحمله مسافة معينة من الجسر دون حدوث أضرار هيكلية، يتم توزيعه عادةً بمتوسط 400 وانحراف معياري 40، وعلى فرض أنّ وزن إجمالي المركبات العابرة على هذا الجسر يخضع للتوزيع الطبيعي بمتوسط 360 وانحراف معياري 30 فإنّ المطلوب: حساب احتمال حدوث ضرر في هذا الجسر.

الحل:

بفرض أنّ X متغير عشوائي يعبر عن إجمالي وزن المركبات العابرة على الجسر، بالتالي احتمال حدوث ضرر في الجسر هو احتمال أن يكون X أكبر من 400،

$$P(X > 400)$$

وهو يساوي احتمال أن تكون القيمة المعيارية Z أكبر من

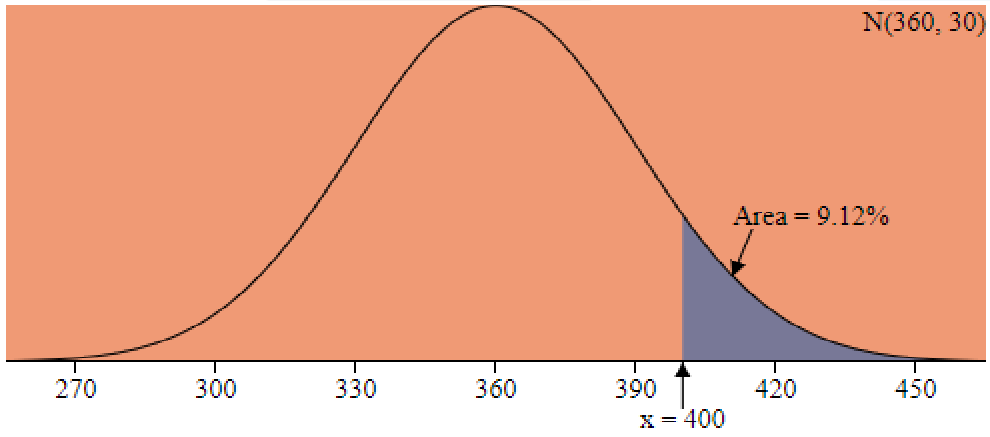
$$\frac{400-360}{30} = 1.33$$

$$P(X > 400) = P(Z > 1.33)$$

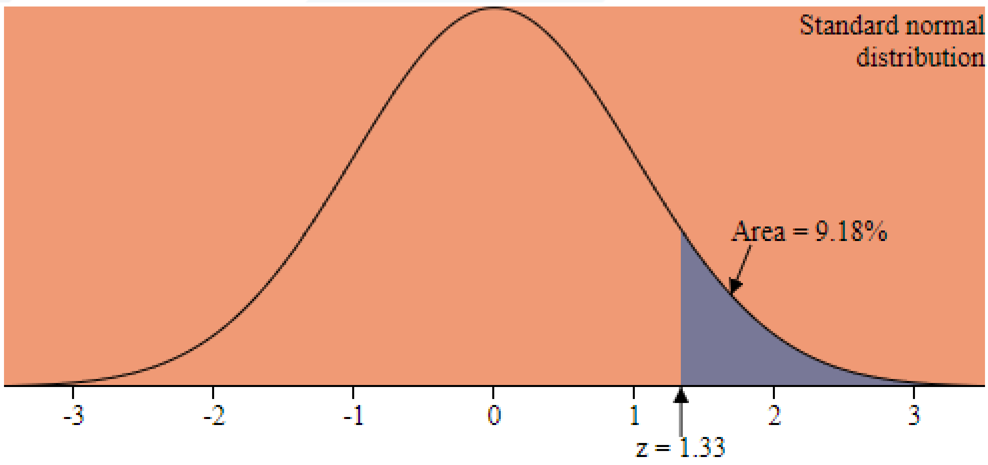
وللتبسيط فإنّ هذا الاحتمال يساوي :

$$P(Z > 1.33) = 1 - P(Z \leq 1.33)$$

من الشكل الآتي:



أو بالشكل المعياري:



إنّ الاحتمال $P(Z \leq 1.33)$ كما رأينا سابقاً يساوي، المساحة بين القيمة 1.33 والمتوسط 0 مضافاً إليها قيمة المساحة المحصورة تحت المنحنى والتي تقع على يسار المتوسط وتبلغ 0.5.

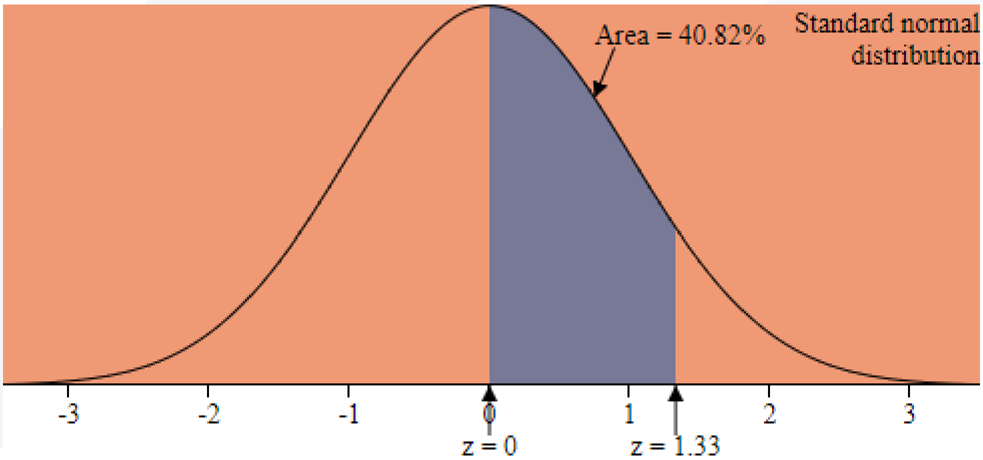
$$P(Z \leq 1.33) = P(Z \leq 0) + P(0 < Z < 1.33)$$

ولحساب المساحة المحصورة بين المتوسط و القيمة 1.33 سيتم النظر للجدول:

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	
⋮							

بالنظر للجدول ، يلاحظ أنّ المساحة المحصورة بين المنحنى و المتوسط 0، وهو يساوي احتمال:

$$P(0 < Z < 1.33) = 0.4082$$



أي أنّ :

$$P(Z \leq 1.33) = P(Z \leq 0) + P(0 < Z < 1.33)$$

$$P(Z \leq 1.33) = 0.5 + 0.4082 = 0.9082$$

بالتالي:

$$P(Z > 1.33) = 1 - P(Z \leq 1.33)$$

$$P(Z > 1.33) = 1 - 0.9082 = 0.0918$$

أي ضمن المعطيات القائمة احتمال حدوث خلل هيكلي في الجسر هو 9.18% وهو احتمال كبير يتطلب من المعنيين إعادة تقييم معايير تشييد هذا الجسر.

مثال:

بالاعتماد على معطيات المثال السابق، المطلوب ما هو الوزن الذي من الممكن أن يتحمله الجسر إذا كان احتمال حدوث خلل هيكلي هو 2.56%:

الحل:

بفرض أنّ الوزن المطلوب هو X بالتالي إنّ قيمته المعيارية :

$$Z = \frac{X - 360}{30} = z$$

ومن المعطيات يمكن الاستنتاج أنّ:

$$P(Z > z) = 0.0256$$

وكما تمّ عرضه سابقاً:

$$P(Z > z) = 1 - P(Z \leq z) = 0.0256$$

أي :

$$P(Z \leq z) = 1 - 0.0256 = 0.9744$$

ولكن:

$$P(Z \leq z) = P(Z \leq 0) + P(0 < Z < z)$$

$$P(Z \leq z) = 0.5 + P(0 < Z < z) = 0.9744$$

أي:

$$P(0 < Z < z) = 0.9744 - 0.5 = 0.4744$$

بالتالي سيتم النظر للجدول والبحث عن القيمة المقابلة للاحتمال 0.744:

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	$\boxed{0.4744}$	
⋮							

وهي القيمة $1.95 = 0.05 + 1.9$ أي أنّ $z = 1.95$

بالعودة للقيمة المعيارية:

$$\frac{X - 360}{30} = z = 1.95$$

وبحل هذه المعادلة:

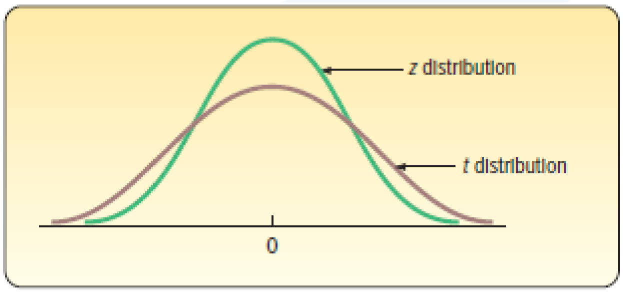
$$X - 360 = 58.5 \Longrightarrow X = 58.5 + 360 = 418.5$$

أي عندما يكون إجمالي الوزن للمركبات العابرة على الجسر هو 418.5 (1000 رطل) فإنّ احتمال حدوث خلل هيكلي على هذا الجسر هو 2.56%.

9-2-2-3- توزيع ستيودنت *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.

مفاهيم إحصائية مهمة: مفهوم درجات الحرية *Degrees of freedom*:

يمكن تبسيط مفهوم درجات الحرية بأنّه مرتبط بحجم العينة و باستقلالية مفردات هذه العينة. أي أنّه يعبر عن عدد المفردات أو المشاهدات المستقلة التي تم استخدامها لحساب المؤشر الإحصائي (متوسط مشاهدات العينة مثلاً).

ماذا يعني استقلالية مفردات العينة؟

فرضاً إذا كان لدينا 5 مفردات في عينة ما وهي : 1 2 3 4 ؟ وكان متوسط هذه المفردات هو 3، من الواضح أنّ المفردة الخامسة المفقودة لا يمكن أن تبقى حرّة بل يمكن القول أن المفردة الخامسة هي حتماً 5 لكي يبقى المتوسط هو 3، بمعنى آخر أنّنا لا نملك الحرّية باختيار المفردة الخامسة. أي أنّ درجة الحرية صفر.

بينما إذا كان لدينا مفردتان مفقودتان 1 2 3 ؟ ؟ وكان المتوسط هو 5 هنا نلاحظ اننا نملك كامل الحرية لاختيار المفردة الرابعة المفقودة، ولكن بعد تحديدها سنكون مجبرين لقيمة واحدة فقط نعطيها للمفردة الخامسة لكي يبقى المتوسط هو 5، أي أنّ درجة الحرية هي درجة حرّة واحدة.

ونستمر إذا كان لدينا ثلاث قيم مفقود نلاحظ أنّه لدينا كامل الحرية لاختيار مفردتين مستقلتين، أي أنّه ليدنا 2 درجة حرية، وهكذا إلى أن نصل إلى حالة وجود 5 قيم مفقودة عندها يكون لدينا أربع قيم مفقودة. وبتعميم هذا المثال نلاحظ ان درجات الحرية يساوي عدد القيم ناقص واحد. هنا قمنا بطرح واحد لأنّه لدينا مؤشر إحصائي واحد نقوم بحسابه.

ولكن إذا كان لدينا أكثر من مؤشر إحصائي مثل حالة نموذج الانحدار، نقوم بطرح عدد معاملات الانحدار (أو بالبارمترات) في هذا النموذج. مثلاً إذا كان لدينا نموذج انحدار بثلاث معاملات تكون درجة الحرية هي عدد المشاهدات ناقص 3.

في نفس السياق عند حساب المتوسط الحسابي للعينة وفقاً للصيغة:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

فإنّ هناك n مشاهدة حرّة لحساب هذه الصيغة.

أمّا لحساب صيغة التباين للعينة:

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

فإنّ عدد المشاهدات (أو القيم) الحرّة المستخدمة لحساب هذه الصيغة هو $n-1$ وليس n ، إذ تمّ استخدام n القيمة لحساب المتوسط وبعد ذلك تمّ استخدام هذا المتوسط لحساب التباين. أي بمعنى آخر للحفاظ على قيمة المتوسط $\bar{x}$ هناك $n-1$ قيمة حرّة بينما تبقى لمشاهدة أو القيمة الأخيرة مقيّدة:

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

لماذا مفهوم درجة الحرية مهم؟

إنّ مفهوم درجات الحرية هو مفهوم مهم جداً يؤثر على موثوقية التحليل الإحصائي وعلى قوة الاختبارات المستخدمة لدراسة دلالة التقديرات الإحصائية. كذلك يلعب مفهوم درجة الحريّة في تحديد عدد المقدرات التي يمكن تقديرها، حيث لا يمكن تقدير عدد كبيرة من المجاهيل باستخدام عدد قليل من درجات الحرية.

مما سبق يتضح أهمية وجود حجم كبير للعينة المدروسة، لأنّ كبر حجم العينة يساعد في زيادة درجات الحرية وبالتالي نملك الموثوقية لعميم النتائج المستخلصة من العينة على المجتمع الكامل، وبالمقابل لا يمكن تعميم تقدير مستمد من عدد قليل من درجات الحريّة على مجتمع كامل.

9-2-2-4- توزيع كاي مربع *Chi squares* :

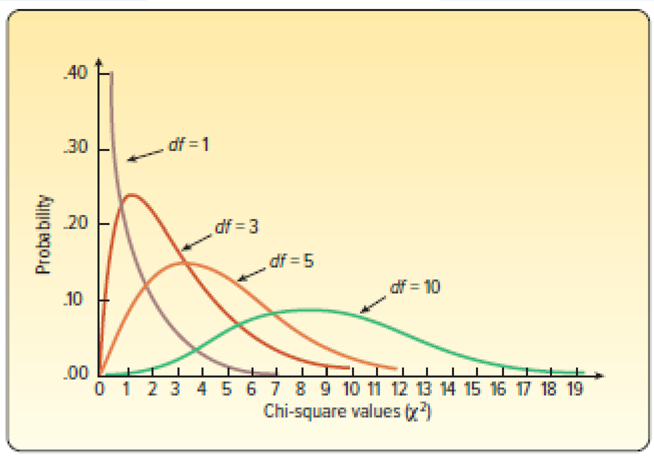
ليكن Z_1 متغير خاضع $N(0,1)$ فإنّ $X = Z_1^2$ يخضع توزيع كاي مربع بـ 1 درجة حرية.

ليكن Z_n ،..... Z_2، Z_1 ، n متغير عشوائي خاضع للتوزيع الطبيعي المعياري $N(0,1)$ ، بالتالي :

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

يخضع لتوزيع كاي مربع $X \sim \chi_n^2$ مع n درجة حرية degrees of freedom.

والذي يُعطى في الشكل الآتي:



من خواص توزيع كاي مربع:

1- إجمالي المساحة تحت المنحنى الإحتمالي تساوي 1.
2- يبدأ المنحنى من القيمة لصفر ويمتد باتجاه اللانهاية ويقترب من المحور الأفقي دون أن يلامسه.
3- منحنى توزيع كاي مربع ملتو نحو اليمين.
4- عندما تزداد درجات الحرية يصبح توزيع كاي مربع مشابهاً للتوزيع الطبيعي.

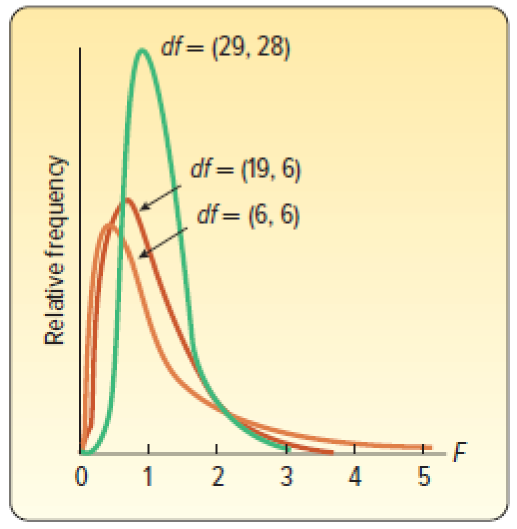
9-2-2-5- توزيع فيشر *Fisher distribution* :

إذا كان χ_n^2 و χ_m^2 متغيران خاضعان للتوزيع كاي مربع مع n و m دجة حرية على الترتيب. فإنّ المقدار :

$$\boldsymbol{F_{n/m}} = \frac{\chi_n^2/n}{\chi_m^2/m}$$

يخضع لتوزيع فيشر بدرجتي حرية، درجة حرية للبسط n و درجة حرية للمقام m.

والشكل الآتي يوضح المنحنى الإحتمالي لتوزيع فيشر.



من خواص توزيع فيشر :

1- يأخذ القيم الموجبة فقط من 0 إلى اللانهاية.
2- توزيع فيشر ملتو نحو اليمين *right skewed*،
3- كلّما ازدادت قيمة توزيع فيشر نحو قيم كبيرة فإنّ التوزيع الإحتمالي يقترب نحو المحور الأفقي دون أن يلامسه.