المحاضرة 1: أهم قوانين التوزيعات الاحتمالية المستمرة

تمهيد:

تلعب التوزيعات الاحتمالية دورا هاما في التطبيقات الإحصائية، وتنقسم التوزيعات الاحتمالية إلى قسمين:

- توزيعات احتمالية مستمرة.
- توزيعات احتمالية متقطعة.

سوف نكتفي ببعض التوزيعات الهامة والتي لها أهمية خاصة في التطبيقات الإحصائية وعلى الخصوص نظرية العينات.

1- التوزيعات الاحتمالية المستمرة:

• التوزيع الطبيعي (Normal Distribution):

يعتبر من أهم التوزيعات الاحتمالية وأكثرها استخداما، ذلك أن التوزيع الطبيعي يستخدم بكثرة في وصف الكثير من الظواهر الطبيعية والاقتصادية والاجتماعية (من أوزان وأطوال...) التي نصادفها في حياتنا اليومية. ولو مثلنا بيانات هذه الظواهر على معلم متعامد ومتجانس لحصلنا على منحنى كثافة، أو منحنى تكرار له تقريبا شكل الجرس، أو كما نعبر عنه إحصائيا شكل منحنى التكرار الطبيعي أو شكل التوزيع الطبيعي.

فإذا كان X متغير عشوائي يتبع التوزيع الطبيعي بمتوسط حسابي μ وتباين ^{2} فإن دالة الكثافة الاحتمالية له تكون بالشكل التالي:

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

ونعبر عن هذه الدالة بالصيغة التالية:

$$\boldsymbol{X\sim N(\mu,\sigma^2)}$$

وتقرأ: X تتبع التوزيع الطبيعي بمتوسط حسابي μ وتباين σ^2.

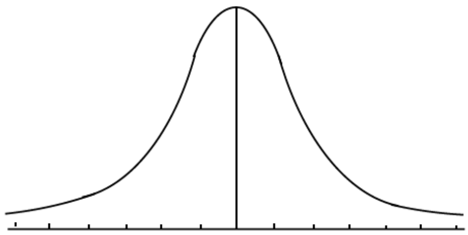
كما أن تابع التوزيع الاحتمالي لهذا المتغير العشوائي (تابع التوزيع الطبيعي) يعرّف بما يلي:

$$\boldsymbol{P(X<x)=F(x)=\int_{-\infty}^{x}f(x)dx}$$

وفي الحساب التكاملي:

$$\int_{-\infty}^{+\infty}\boldsymbol{e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx=\ \sigma\sqrt{2\pi}}$$

الرسم التالي يوضح منحنى التوزيع الطبيعي:



ومن خلال هذا المنحنى يتضح أن التوزيع متماثل (Symétrique) حول الوسط الحسابي **μ** وأن أكثر المشاهدات تقع حول الوسط الحسابي وأقلها في الطرفين.

ونظرا لصعوبة حساب التكامل السابق أو عمل جداول لحساب هذا التكامل عند قيم مختلفة لـ **μ** و σ^2 لاستخدامها في حساب الاحتمال (وهي مسألة بدون فائدة)، فإنه بإمكاننا تحويل أي توزيع طبيعي إلى توزيع طبيعي معياري.

نظرية (1-1): إذا كان المتغير العشوائي **X** يتبع التوزيع الطبيعي بمتوسط **μ** و تباين σ^2، أي: $X \sim N(\mu, \sigma^2)$ فإن المتغير العشوائي **Z** يتبع التوزيع الطبيعي المعياري بمتوسط $\mu = 0$ وتباين $\sigma^2 = 1$، أي $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

حسب النظرية **(1-1)**، تكون دالة الكثافة الاحتمالية لـ **Z** كما يلي:

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

أما دالة التوزيع تكون وفق الصيغة التالية:

$$P(Z < z) = F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

لتسهيل القراءة سنرمز لدالة الكثافة الاحتمالية في هذه الحالة بـ $\varphi(z)$ بدلا من **f(z)** ولدالة التوزيع بـ $\emptyset(z)$ بدلا من **F(z)**. وبالتالي إذا كان $Z \sim N(0, 1)$ فإن كثافته $\varphi(z)$ ودالة توزيعه $\emptyset(z)$.

إن حساب قيم دالة التوزيع للمتغير الطبيعي المعياري يعتمد على جدول يعطي قيم دالة التوزيع $\emptyset(z)$ ابتداء من الصفر وبفاصل 0.01 بين كل قيمة والقيمة التي تليها (أنظر الملحق 2). وقد وضعت قيم z ابتداء من 3.4- وبفاصل 0.1 في العمود الأيسر، ووضعت القيم العشرية الثانية من قيم z في السطر الرأسي. أما قيم المساحات أو القيم الاحتمالية (أي قيم الدالة $\emptyset(z)$) فقد وضعت في صلب الجدول.

ملاحظة:

بسبب التماثل، قيم: $\emptyset(-z) = 1 - \emptyset(z)$

لتوضيح استخدام هذا الجدول نأخذ التمرين التالي:

مثال (1-3):

إذا كانت المتغيرة العشوائية X تتبع التوزيع الطبيعي بمتوسط $\mu = 5$ وتباين $^2 = 9$ أي: $X \sim N(5,9)$

والمتغيرة Z تتبع التوزيع الطبيعي المعياري، أي $Z \sim N(0,1)$ أوجد القيم الاحتمالية التالية: $P(X < 7)$ ،

- بالنسبة لقيم الاحتمال الخاصة بالمتغير العشوائي X، يجب أن نحول المتغير العشوائي X إلى Z المعيارية كما يلي:

$$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sqrt{\sigma^2}} = \frac{X-5}{\sqrt{9}} \sim N(0,1)$$

وبالتالي:

- $P(X < 7) = P\left(Z < \frac{7-5}{\sqrt{9}}\right) = P\left(Z < \frac{2}{3}\right) = \emptyset(0.66) = \mathbf{0.745373}$

وهذه القيمة مستخرجة من جدول التوزيع الطبيعي المعياري (الملحق 2)، وتمثل تقاطع السطر **0.6** مع العمود **0.06**.

- $P(4.2 < X < 5.5) = P\left(\frac{4.2-5}{\sqrt{9}} < Z < \frac{5.5-5}{\sqrt{9}}\right) = P(-0.26 < Z < 0.16) = \emptyset(0.16) - \emptyset(-0.26) =$
$0.563559 - 0.397432 = \mathbf{0.166127}$

وتستخرج القيمة $\emptyset(0,16) = \mathbf{0.563559}$ من جدول التوزيع الطبيعي المعياري بالنسبة للقيم الموجبة وهي تمثل تقطع السطر **0,10** مع العمود **0,06**. أما القيمة $\emptyset(-0,26) = \mathbf{0.397432}$ فتم استخراجها أيضا من جدول التوزيع الطبيعي المعياري بالنسبة للقيم السالبة وهي تمثل تقاطع السطر **0,2-** مع العمود **0,06-**.

- بالنسبة لقيم الاحتمال الخاصة بالمتغير العشوائي Z، فتستخرج بشكل مباشر باستخدام جدول التوزيع الطبيعي المعياري (الملحق 2) نجد:

- $P(Z < -1.47) = \emptyset(-1.47) = 0.070781$

وهذه القيمة هي تقاطع السطر 1.4- مع العمود 0.07-.

وبنفس الطريقة نجد:

- $P(1.14 < Z < 2.89) = P(Z < 2.89) - P(Z \leq 1.14) = \emptyset(2.89) - \emptyset(1.14) =$
$0.9980 - 0.8728 = 0.1252$

- $P(Z < 1.25) = \emptyset(1.25) = 0.894350$
- $P(Z > 1.30) = 1 - P(Z \leq 1.30) = 1 - \emptyset(1.30) = 1 - 0.903199 = 0.096801$

يمكن أيضا استخدام جدول دالة التوزيع الطبيعي المعياري بشكل عكسي، كما هو موضح في التمرين التالي:

مثال (1-4):

إذا كان $Z \sim N(0,1)$ وكان لديك $P(Z < a) = 0.978822$ و $P(Z > b) = 0.252346$، أوجد قيمة كل من a و b.

- من جدول التوزيع الطبيعي المعياري نبحث عن قيمة الاحتمال 0.978822، سنجد هذه القيمة تقع عند العمود 4 والسطر 19 وبالتالي قيمة a = 2.03.
- قبل البحث عن قيمة b نحوّل أولا العلاقة من أكبر إلى أصغر ،

$P(Z > b) = 0.252346 \Rightarrow P(Z \leq b) = 1 - 0.252346 = 0.747654$

من جدول التوزيع الطبيعي المعياري نبحث عن قيمة الاحتمال 0.747654، سنجد هذه القيمة تقع بين العمودين 7 و 8 والسطر 7 وبالتالي قيمة b بالتقريب تساوي 0.66 (b = 0.66).

• توزيع مربع كاي (Chi–Square):

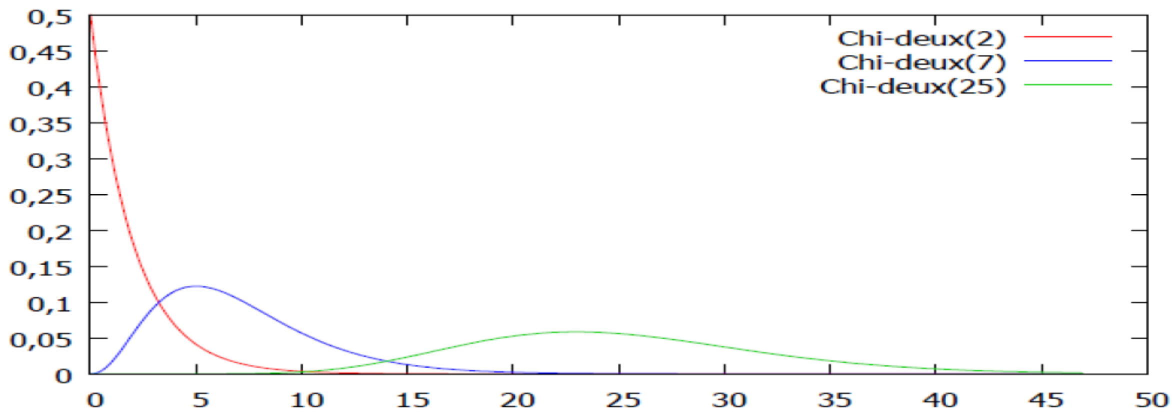
نقول عن المتغير العشوائي X أنه يتبع توزيع مربع كاي χ^2 بدرجة حرية ϑ (vartheta) إذا كانت دالة كثافته الاحتمالية على الشكل التالي:

$$f(x) = \frac{1}{\Gamma\left(\frac{\vartheta}{2}\right)(2)^{\frac{\vartheta}{2}}} . x^{\frac{\vartheta}{2}-1} . e^{-\frac{x}{2}} \qquad , x \geq 0$$

توزيع مربع كاي يعتبر كحالة خاصة من توزيع جاما $(Gamma)$ بالمعلمتين: $\alpha = \frac{\vartheta}{2}$ و $\beta = 2$.

ودرجة الحرية ϑ تدل على معلمة توزيع مربع كاي، ونكتب: $X \sim \chi^2_{(\vartheta)}$

الشكل البياني التالي يوضح دالة التوزيع لمربع كاي عند قيم مختلفة لـ ϑ.



يتضح من التمثيل البياني الخاص بمربع كاي أنّ الالتواء موجب ويقترب من التماثل كلما كبرت درجة الحرية. بعبارة أخرى، كلما كان حجم العينة كبيرا $n \geq 30$ يقترب التوزيع من التوزيع الطبيعي.

سنذكر هنا بعض النظريات الهامة الخاصة بهذا التوزيع التي سنستخدمها في تطبيقاتنا ولكن بدون برهان.

نظرية (1-2): إذا كان المتغير العشوائي $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ فإن $Z^2 = (\frac{X-\mu}{\sigma})^2 \sim \chi^2_{(1)}$.

نظرية (1-3): إذا كانت المتغيرات العشوائية $X_1, X_2, \dots, X_n$ مستقلة عن بعضها وكل متغيرة تتبع التوزيع الطبيعي المعياري $N(0,1)$، فإن: $\sum_{i=1}^{n} X_i^2 \sim \chi^2_n$.

نظرية (1-4): إذا كانت $U_1, U_2, \dots, U_n$ متغيرات عشوائية مستقلة تتبع توزيع مربع كاي بدرجات حرية $\vartheta_1, \vartheta_2, \dots, \vartheta_n$ على التوالي، فإن: $\sum_{i=1}^{n} U_i \sim \chi^2_{(\vartheta)}$.

نظرية (1-5): إذا أخذت عينات عشوائية كل بحجم n وتباين ^{2} من مجتمع طبيعي التوزيع بمتوسط حسابي μ و تباين σ^2، فإن: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$. حيث: $s^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$.

ونعبّر عن قيمة المتغير العشوائي الذي يتبع توزيع مربع كاي بالرمز $\chi^2_{(\alpha,\vartheta)}$ وهي القيمة التي يقع على يمينها مساحة α عن منحنى مربع كاي بدرجة حرية ϑ وتحسب قيم هذا المتغير العشوائي والاحتمالات المرافقة له باستخدام جدول مربع كاي الموضح في الملحق 3، والتمرين التالي يوضح كيفية استخدام ذلك الجدول الخاص بالتوزيع.

مثال (1-5): أوجد القيم $\chi^2_{(0.025,20)}$، $\chi^2_{(0.01,5)}$، $\chi^2_{(0.975,7)}$

الحل:

لإيجاد قيمة $^2_{(0.025,20)}$ نختار من العمود الأيسر الخاص بدرجات الحرية ϑ القيمة **20** ومن السطر العلوي الخاص بالمساحات (أي الاحتمالات) نختار القيمة **0.025**. وبتقاطع العمود مع السطر نحصل على القيمة المطلوبة، وبالتالي يكون $\chi^2_{(0.025,20)} = 34.17$.

بنفس الطريقة نحصل على $^2_{(0.01,5)} = 15.09$ وأيضا $\chi^2_{(0.975,7)} = 1.69$.

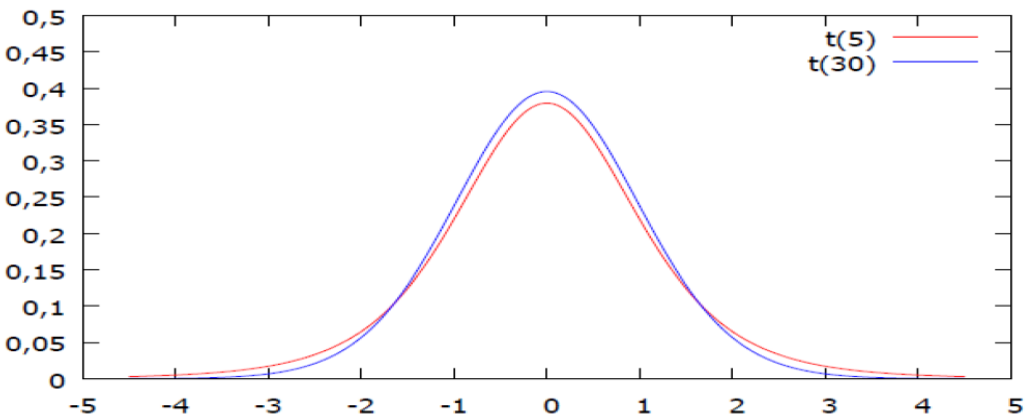
• توزيع ستودنت "t" Student:

إذا كانت دالة الكثافة الاحتمالية للمتغير العشوائي t على الصورة:

$$f(t) = \frac{\Gamma\left(\frac{\vartheta+1}{2}\right)}{\Gamma\left(\frac{\vartheta}{2}\right)\sqrt{\vartheta\pi}}\left(1+\frac{t^2}{\vartheta}\right)^{-\frac{n+1}{2}} \qquad , -\infty < t < +\infty$$

فإن هذا التوزيع يسمى توزيع **t**، حيث ϑ تمثل درجات الحرية وهي في نفس الوقت معلمة هذا التوزيع، ويلعب توزيع **t** دورا مهما عند سحب العينات الصغيرة الحجم $(n < 30)$ من مجتمع مجهول التباين.

ويتشابه توزيع **t** مع التوزيع الطبيعي القياسي من حيث الشكل الجرسي إلاّ أنه أكثر انخفاضا منه، وعندما تزداد درجات الحرية فإن توزيع **t** يقترب من التوزيع الطبيعي القياسي. والشكل التالي يوضح ذلك.



سنذكر هنا بعض النظريات الهامة الخاصة بهذا التوزيع ولكن بدون برهان.

نظرية (1-6): إذا كان المتغيران العشوائيان المستقلان Y و Z، حيث أن Y يتوزع توزيعا طبيعيا قياسيا $Y \sim N(0,1)$ و Z يتوزع توزيع مربع كاي $Z \sim \chi^2_\vartheta$ فإن المتغير العشوائي $T = \frac{Y}{\sqrt{Z/\vartheta}} \sim t_\vartheta$.

نظرية (1-7): إذا أخذت عينات عشوائية صغيرة كل بحجم n وتباين ^{2} من مجتمع طبيعي التوزيع بمتوسط حسابي **μ** وتباين σ^2، فإن: $T = \frac{\bar{X}-\mu}{\sqrt{s^2/n}} \sim t(n-1)$

ونعبّر عن قيمة المتغير العشوائي الذي يتبع توزيع t بالرمز $_{(\alpha,\vartheta)}$ وهي القيمة التي يقع على يمينها مساحة α على منحنى توزيع t بدرجة حرية ϑ. وتحسب قيم هذا المتغير العشوائي والاحتمالات المرافقة له باستخدام جدول توزيع ستودنت

ملاحظة:

بسبب التوزيع المتماثل حول الصفر، يكون: $t_{(\alpha,\vartheta)} = -\mathbf{t}_{(1-\alpha,\vartheta)}$

مثال (1-6): أوجد قيم t في الحالات التالية: $t_{(0.05,15)}$، $t_{(0.95,7)}$، $t_{(0.995,10)}$

الحل:

لإيجاد قيمة $_{(0.05,15)}$ نختار من العمود الأيسر الخاص بدرجات الحرية القيمة **15** ومن السطر العلوي الخاص بالاحتمالات القيمة **0.05**. وبتقاطع السطر مع العمود نحصل على القيمة المطلوبة، وبالتالي يكون $t_{(0.05,15)} = 1.753$.

لإيجاد قيمة $_{(0.95,7)}$ نلاحظ أن قيمة $\alpha = 0.95$ غير متوفرة في جدول توزيع ستودنت، وبذلك سوف نستعمل تماثل المنحنى حول الصفر فنحصل على:

$$t_{(0.95,7)} = -\mathbf{t}_{(1-0.95,7)} = -\mathbf{t}_{(0.05,7)} = -1.895$$

وبالمثل نستخرج قيمة $t_{(0.995,10)}$ فتكون قيمته:

$$t_{(0.995,10)} = -\mathbf{t}_{(1-0.995,10)} = -\mathbf{t}_{(0.005,10)} = -3.169$$

• توزيع فيشر "F" Fisher:

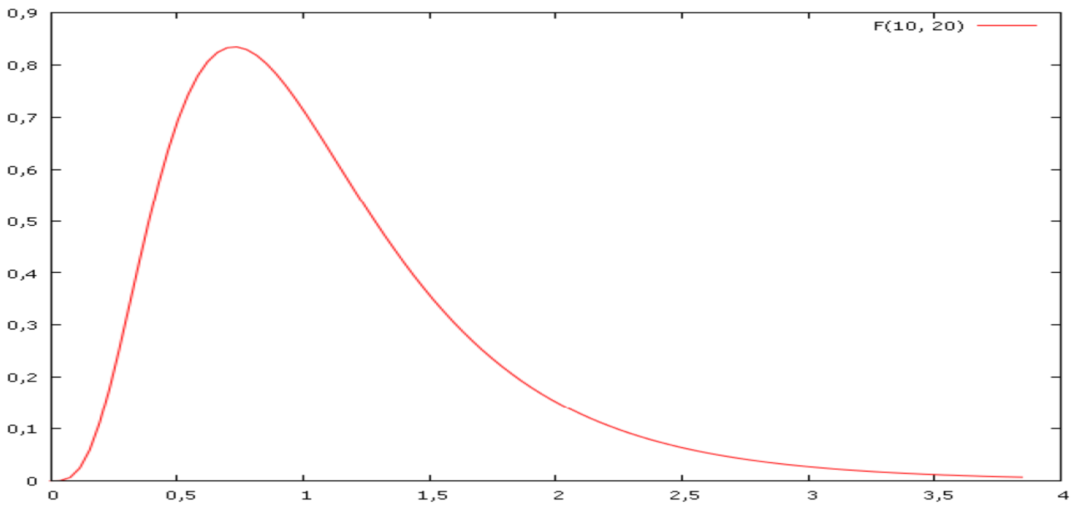
إذا كان المتغيران العشوائيان المستقلان X_1 و X_2 يتوزعان توزيع مربع كاي بدرجتي حرية ϑ_1 و ϑ_2 على التوالي ($X_1 \sim \chi^2_{\vartheta_1}$ و $X_2 \sim \chi^2_{\vartheta_2}$) فإن المتغير X المعرف بالصيغة الآتية: $X = \frac{X_1/\vartheta_1}{X_2/\vartheta_2}$ له دالة الكثافة الاحتمالية التالية:

$$g(X) = \frac{\Gamma\left(\frac{\vartheta_1+\vartheta_2}{2}\right)}{\Gamma\left(\frac{\vartheta_1}{2}\right)\Gamma\left(\frac{\vartheta_2}{2}\right)} (\vartheta_1)^{\frac{\vartheta_1}{2}} (\vartheta_2)^{\frac{\vartheta_2}{2}} X^{\left(\frac{\vartheta_1}{2}\right)-1} (\vartheta_1 + \vartheta_2 X)^{\frac{\vartheta_1+\vartheta_2}{2}} \quad , X > 0$$

وعليه نقول بأن المتغير X يتوزع توزيع فيشر بدرجة حرية $_1$ و ϑ_2 ويعبّر عنه بالشكل التالي:

$$X \sim F_{(\vartheta_1,\vartheta_2)}$$

ويقترب شكل توزيع فيشر من الشكل التالي:



ونعبّر عن قيمة المتغير العشوائي الذي يتبع توزيع فيشر بالرمز $_{(\alpha,\vartheta_1,\vartheta_2)}$ وهي القيمة التي يقع على يمينها مساحة α على منحنى توزيع F بدرجة حرية ϑ_1 في البسط و ϑ_2 في المقام.

والجدول (5) الموضح في الملاحق خاص بقيم المتغير العشوائي الذي يتبع توزيع فيشر $_{(\alpha,\vartheta_1,\vartheta_2)}$ حسب قيم α التالية: $\{0.01$ ، 0.025 ، 0.05 ، 0.10 ، $0.25\}$.

مثال (1- 7): أوجد القيم التالية: $F_{(0.01,10,12)}$، $F_{(0.025,8,6)}$، $F_{(0.05,12,7)}$، $F_{(0.10,5,4)}$ ، $F_{(0.25,7,9)}$

الحل:

لإيجاد قيمة $_{(0.01,10,12)}$ نستخدم جدول توزيع فيشر الخاص بقيمة $\alpha = 0.01$، ومن السطر في أعلى الجدول الخاص بدرجات حرية البسط نختار القيمة **10** ومن العمود الخاص بدرجات حرية المقام نختار القيمة **12**. وبتقاطع السطر مع العمود نجد أن قيمة $F_{(0.01,10,12)} = 4.85$.

وبتغيير قيم α واستخدام نفس الطريقة نحصل على $F_{(0.025,8,6)} = 5.60$ و $F_{(0.05,12,7)} = 3.57$ و $F_{(0.10,5,4)} = 4.05$ وأيضا $F_{(0.25,7,9)} = 1.60$.

قد يحدث أحيانا أن تكون قيم α هي إحدى القيم التالية غير المتوفرة في جداول فيشر:

$$\{0.01 \text{ ، } 0.025 \text{ ، } 0.05 \text{ ، } 0.10 \text{ ، } 0.25\}$$

وحتى نتمكن من استخراج قيم المتغير العشوائي الذي يتبع توزيع فيشر في هذه الحالة، نستعين بالنظرية التالية:

نظرية (1-8): بفرض أن $0 < n_2$ ، $_1$ فإن:

$$F(1-\alpha, n_1-1, n_2-1) = \frac{1}{F(\alpha, n_2-1, n_1-1)}$$

مثال (1- 8): أوجد القيمة $F_{(0.90,5,4)}$

الحل:

لاحظ أنّ $\alpha = 0.90$ وفي مثالنا هذا غير متوفرة ضمن قيم α في جداول فيشر، لذلك سنستعين بتطبيق النظرية (1−8):

$$F_{(0.90,5,4)} = \frac{1}{F(0.10,4,5)} = \frac{1}{3.52} = 0.284$$