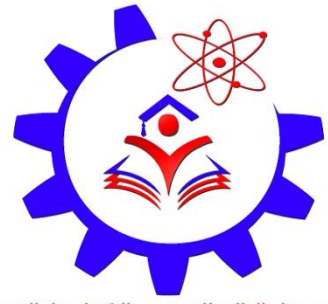


Ministry of Higher Education
Higher Future Institute of Engineering and
Technology in Mansoura
Department of Mathematics and
Basic Sciences



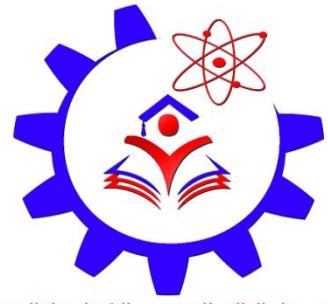
معهد المستقبل العالي للهندسة و التكنولوجيا بالمنصورة
Higher Future Institute of Engineering and Technology at Mansoura

احصاء ونظرية الاحتمالات



د. حاتم عبدالعزيز العجمي

Mathematics and statistics



شُقَّ مصطلح الإحصاء باللغة الإنجليزية (Statistics) من الكلمة الإيطالية (Statista)، والكلمة الألمانية (Statistik)، والكلمة اللاتينية (Status)، والتي هي عبارة عن مصطلحات تُعنى بمعلومات الدولة بالإنجليزية (Political state)، حيث كانت بداية استخدام هذا المصطلح لجمع البيانات التي تخص أفراد الدولة، لغاية إنشاء قاعدة بيانات يتم من خلالها فرض الضرائب لتحسين الوضع المادي للدولة، كما تم تعريف الإحصاء على أنه العلم الذي يهتم بجمع البيانات الرقمية، ومن ثم تنظيمها، وترتيبها، وتحليلها، بهدف الوصول إلى نتائج معينة لتوضيح ظاهرة أو حالة ما، أو بأنه العلم الذي يهتم بالطريقة التي يتم من خلالها جمع البيانات والمعلومات وتحويلها إلى صورة عددية، حيث تُجمع البيانات من خلاله بشكل منظم، وفيما يخص استخدامات علم الإحصاء فهي كثيرة؛ كاستخدامه في العلوم الهندسية، و العلوم الطبية، وعلم الاجتماع، والاقتصاد، والصناعة، والكيمياء، والرياضة، والإدارة، وغيرها العديد من المجالات.

د. حاتم عبدالعزيز العجمي

Mathematics and statistics

المحتويات

7	أساس الاحتمال	: الفصل الأول
21	الإحصاء الوصفي	: الفصل الثاني
31	المتغيرات العشوائية المنفصلة	: الفصل الثالث
43	المتغيرات العشوائية المستمرة	: الفصل الرابع
51	أمثلة على المتغيرات العشوائية	: الفصل الخامس
67	نظرية المعاينة	: الفصل السادس
85	نظرية التقدير	: الفصل السابع
95	اختبارات الفروض والعنوية	: الفصل الثامن
111	تمهيد المنحنيات والانحدار والارتباط	: الفصل التاسع
129	توزيعات احتمالية أخرى	: الفصل العاشر
144	موضوعات رياضية	: ملحق (A)
		المساحات تحت المنحنى الطبيعي المعياري من 0	: ملحق (B)
148	حتى Z	
150	توزيع t ستودنت	: ملحق (C)
152	توزيع كاي تربيع	: ملحق (D)
154	قيم النسب 95%، 99% لتوزيع F	: ملحق (E)
158	قيم $e^{-\lambda}$: ملحق (F)
160	أرقام عشوائية	: ملحق (G)
161	قائمة المصطلحات العلمية (انجليزي/عربي)	

الفصل الأول

أساس الاحتمال

Basic Probability

في هذا الفصل:

- ✓ التجارب العشوائية
- ✓ فراغات العينة
- ✓ الأحداث
- ✓ مفهـوم الاحتمال
- ✓ بديهيات الاحتمالات
- ✓ بعض النظريات الهامة على الاحتمال
- ✓ تعيين الاحتمالات
- ✓ الاحتمال الشرطي
- ✓ نظريات حول الاحتمال الشرطي
- ✓ الأحداث المستقلة
- ✓ نظرية أو قاعدة بيز
- ✓ تحليل التوافق
- ✓ المبادئ الأساسية للعد

✓ التباديل

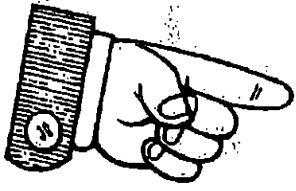
✓ التوافيق

✓ معاملات ذات الحدين

✓ تقريب ستيرلنج للمضروب $n!$

Random Experiments

التجارب العشوائية



نحن متعودون على أهمية التجارب في مجالات العلوم والهندسة. فالتجريب مفيد في الاستخدام لافتراض أن إجراء التجارب تحت شروط متقاربة سوف يعطى نتائج متساوية. وفي هذه الظروف سوف نكون قادرين على تحديد قيم المتغيرات التي تؤثر على نتائج التجربة. وعلى أى حال، في بعض التجارب لا نتمكن من تحديد قيم بعض المتغيرات وبالتالي سوف تتغير النتائج من إجراء تجربة إلى أخرى مع أن معظم الشروط تظل كما هي. وتوصف هذه التجارب بالتجارب العشوائية:

مثال 1.1: عند رمي زهرة طاولة، فالنتيجة التي سوف تحدث تكون أحد الأرقام التالية $\{1, 2, 3, 4, 5, 6\}$.

Example 1.1. If we toss a die, the result of the experiment is that it will come up with one of the numbers in the set $\{1, 2, 3, 4, 5, 6\}$.

Sample Spaces

فراغات العينة

تتكون المجموعة S من كل النتائج الممكنة للتجربة العشوائية ويطلق عليها اسم فراغ العينة Sample Space، ويطلق على كل نتيجة نقطة من العينة Sample Point. وغالباً ما يوجد أكثر من فراغ للعينة يمكن

أن يصف نتائج التجربة، ولكن يوجد في العادة واحد فقط يعطى المعلومات الأكثر.

مثال 1.2: إذا تم رمي زهرة طاولة، فإن فراغ العينة يكون {1, 2, 3, 4, 5, 6} بينما يوجد آخر يعطى {زوجي، فردي}. ومن الواضح أن الأخير غير كافي ليحدد، مثلاً ما إذا كان الناتج يقبل القسمة على 3.

Example 1.2. If we toss a die, then one sample space is given by {1, 2, 3, 4, 5, 6} while another is {even, odd}. It is clear, however, that the latter would not be adequate to determine, for example, whether an outcome is divisible by 3.

وإذا كان من الممكن تمثيل فراغ العينة بيانياً فإنه من الأفضل استخدام الأرقام بدلاً من الحروف ما أمكن ذلك:

ويسمى فراغ العينة المحدد Finite Sample إذا كان به عدد محدود من النقط. أما إذا كان به عدد غير محدود من النقط والتي يمكن عدّها فيسمى بفراغ العينة الغير محدود أو اللانهائي Countably Infinite Sample Space. أما إذا وجد به عدد لا نهائي من النقط لا يمكن عدّه داخل فترة على محور x مثل $0 \leq x \leq 1$ فإنه يسمى بفراغ العينة اللانهائي Noncountably Infinite Sample Space. وفراغ العينة المحدود Finite أو المحدود الممكن عدّه Countably Finite يطلق عليه عادة فراغ العينة المنفصل Discrete Sample Space، بينما اللانهائي الذي لا يمكن عدّه يسمى بفراغ العينة الغير منفصل Nondiscrete Sample Space.

مثال 1.3: فراغ العينة الناتج من رمي زهرة نرد يعطى فراغ عينة منفصل. بينما اختيار أي رقم، ليس من الضروري رقم صحيح بين الواحد والعشرة يعطى فراغ عينة غير منفصل.

Example 1.3. The sample space resulting from tossing a die yields a discrete sample space. However, picking any number, not just integers, from 1 to 10, yields a nondiscrete sample space.

الحدث هو مجموعة جزئية A Subset من مجموعة فراغ العينة S ؛ أي أنه مجموعة من النتائج الممكنة. فإذا كان ناتج التجربة يمثل عنصراً في الحدث A فإننا نقول أن الحدث A يتحقق. والحدث الذي يتكون من نقطة واحدة من فراغ العينة S يسمى بالحدث البسيط Simple أو الحدث الأولي Elementary Event والأحداث الخاصة حينما يكون هو S نفسها فإنه يمثل الحدث المؤكد Sure or Certain Event حيث أنه من الضروري أن يحدث عنصر من S ، بينما المجموعة الفارغة \emptyset Empty set والتي تسمى بالحدث المستحيل Impossible Event لأن أي عنصر في \emptyset لا يمكن حدوثه.

ويستخدم عمليات المجموعات على الأحداث في S فإننا نحصل على أحداث أخرى في S . وكمثال إذا كانت A ، B تمثل أحداثاً، فإن:

1. $A \cup B$ تمثل الحدث « حدوث A أو B أو الاثنين معاً ». وتسمى $A \cup B$ اتحاد الأحداث A و B "Union".

2. $A \cap B$ تمثل الحدث « حدوث الحدين A ، B في نفس الوقت ». وتسمى $A \cap B$ تقاطع الأحداث A و B "Intersection".

3. A' تمثل الحدث "not A " أي مكمل الحدث A "Complement of A ".

4. $A - B = A \cap B'$ وتمثل الحدث " A وليس B " وخاصة $A' = S - A$.

وإذا كانت الأحداث A ، B تمثل أحداثاً منفصلة، أي أن $A \cap B = \emptyset$ فإننا نقول في الغالب أن الأحداث متباعدة بالتبادل Mutually Exclusive. وهذا يعني أن الأحداث لا يمكن أن تحدث مع بعضها في نفس الوقت. ونقول أن مجموعة الأحداث A_1, A_2, \dots, A_n تكون أحداثاً منفصلة إذا كان كل اثنين منهما يمثلان أحداثاً منفصلة.

يوجد في الغالب في أى تجربة عشوائية عدم تأكد من أن حدث معين سوف يحدث أم لا. وكمقياس للفرصة Chance أو للاحتمال Probability يمكن التوقع بحدوث الحدث بتحديد رقم يقع بين الصفر والواحد الصحيح. فإذا كنا متأكدين من حدوث الحدث فإننا نقول إن الاحتمال يكون 100% أو واحد صحيح. وإذا كنا متأكدين أن الحدث لن يحدث فإننا سنقول إن احتمال حدوثه صفر. وإذا كان الاحتمال مساوياً $\frac{1}{4}$ فإننا نقول إنه يوجد فرصة تمثل 25% لحدوث الحدث، وإن احتمال عدم حدوثه يساوى 75%؛ أى أن مكمل عكس الحدث هو 75% إلى 25% أو 3 إلى 1.

ويوجد أسلوبان لتقدير احتمال حدوث حدث ما.

1. **الأسلوب الكلاسيكي Classical Approach**: إذا كان يمكن حدوث الحدث بطرق عددها h من الطرق الكلية والتي عددها n . وكلها متساوية، فإن احتمال حدوث الحدث يساوى h/n .

2. **الأسلوب التكراري Frequency Approach**: إذا كان بعد تكرار التجربة n مرة، وكانت n كبيرة جداً وتكرر حدوث الحدث h مرة منها، كان احتمال حدوث الحدث يساوى h/n . ويسمى هذا الأسلوب بالأسلوب التجريبي أيضاً Empirical Probability.

وكلا الأسلوبين التقليدي والتكراري له خلفية، فالأقل يتمثل في احتمالات الحدوث المتساوية "Equally Likely" وهي كلمة غامضة، والثاني يتمثل في أن تكون n كبيرة وهي أيضاً تخضع للغموض، ومن هنا قدم الرياضيون لنظرية الاحتمالات بعدد من البديهيات Axiomatic Approach.

بديهيات الاحتمالات The Axioms of Probability

افتراض أننا لدينا فراغ العينة S . إذا كانت S منفصلة Discrete فكل المجموعات الجزئية تتبع الأحداث والعكس صحيح. وإذا كانت S غير منفصلة Nondiscrete فإن مجموعات جزئية خاصة (تسمى مقيسة) تتبع الأحداث. لكل حدث A في الفئة C من الأحداث فإننا نلحق به رقم حقيقي $P(A)$ ، وتسمى P دالة الاحتمال Probability Function وتسمى $P(A)$ باحتمال حدوث الحدث The Probability of the Event وذلك بتحقق البديهيات الآتية:

البديهية الأولى: لكل حدث A في الفئة C يكون

$$P(A) \geq 0$$

البديهية الثانية: للحدث المؤكد S في الفئة C يكون

$$P(S) = 1$$

البديهية الثالثة: لأي عدد من الأحداث المتنافية بالتبادل Mutually

Exclusive Event (A_1, A_2, \dots في الفئة) يكون

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

وفي حالة خاصة في حالة الأحداث المتنافية A_1, A_2 يكون

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

بعض النظريات الهامة على الاحتمال

Some Important Theorems on Probability

نظرية 1-1: إذا كانت $A_1 \subset A_2$ فإن

$$P(A_1) \leq P(A_2) \text{ and } P(A_2 - A_1) = P(A_2) - P(A_1) \quad (1)$$

نظرية 1-2: لكل حدث A يكون،

$$0 \leq P(A) \leq 1 \quad (2)$$

أى أن الاحتمال يكون بين الصفر والواحد الصحيح.

نظرية 1-3: للمجموعة الفارغة Empty Set \emptyset

$$P(\emptyset) = 0 \quad (3)$$

أى أن الحدث المستحيل يكون احتمال حدوثه مساوياً للصفر.

نظرية 1-4: إذا كانت A' تمثل مكمل الحدث A فإن

$$P(A') = 1 - P(A) \quad (4)$$

نظرية 1-5: إذا كانت $A = A_1 \cup A_2 \cup \dots \cup A_n$

وكانت الأحداث A_1, A_2, \dots, A_n أحداثاً متنافية فإن

$$P(A) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (5)$$

نظرية 1-6: إذا كانت الأحداث B و A تمثل أى حدثين. فإن

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (6)$$

وبصفة عامة إذا كان لدينا الأحداث A_1, A_2, A_3 فإن

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_2 \cap A_3) -$$

$$P(A_3 \cap A_1) + P(A_1 \cap A_2 \cap A_3)$$

ويمكن التعميم على n من الأحداث.

نظرية 1-7: لأى حدثين A, B فإن

$$P(A) = P(A \cap B) + P(A \cap B') \quad (7)$$

Assignment of Probabilities

تعيين الاحتمالات

إذا تكون فراغ العينة S من عدد محدود من النتائج a_1, a_2, \dots, a_n فإن

النظرية 1-5 تعطى

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1 \quad (8)$$

حيث أن A_1, A_2, \dots, A_n تمثل أحداثاً أولية بحيث $A_i = \{a_i\}$

وهذا يعنى أنه يمكن اختيار أى أرقام غير سالبة للاحتتمالات من هذه الأحداث البسيطة حتى تتحقق المعادلة السابقة. وفى حالة خاصة إذا افترضنا تساوى الاحتمالات لكل الأحداث البسيطة. فإن

$$P(A_k) = \frac{1}{n} \quad k = 1, 2, \dots, n \quad (9)$$

وإذا كان الحدث A يحدث إذا تحقق h من الأحداث البسيطة فإن:

$$P(A) = \frac{h}{n} \quad (10)$$

وهى تعادل الأسلوب الكلاسيكى للاحتتمالات. ونحن نتبع بالطبع أسلوباً آخر فى تحديد الاحتمالات مثل الأسلوب التكرارى. وتحديد الاحتمالات يتطلب نموذج رياضى Mathematical Model يكون قد تم اختباره بنفس الطريقة فى المجالات الطبيعية والعلوم الأخرى حتى يتحدد نجاحه.



تذكروا!

إن احتمال حدوث حدث معين لا بد وأن يكون بين الصفر والواحد الصحيح.

الاحتمال الشرطى Conditional Probability

إذا كان لدينا الحدثان A, B . حيث أن $P(A) > 0$ وكان $P(B|A)$ يرمز لاحتمال حدوث B بشرط حدوث A . وحيث أن الحدث A معروف حدوثه، فإنه يعطى فراغ عينة جديد يحل محل فراغ العينة الأصلية S .

وهذا يقودنا إلى التعريف

$$P(B|A) \equiv \frac{P(A \cap B)}{P(A)} \quad (11)$$

أو

$$P(A \cap B) \equiv P(A)P(B|A) \quad (12)$$

وهذا يعنى أن احتمال حدوث الحدثين A و B معاً يكون مساوياً لاحتمال حدوث الحدث A مضروباً فى احتمال حدوث B بشرط حدوث A . وتسمى $P(B|A)$ باحتمال حدوث B بشرط A ، Conditional Probability of B Given A ؛ أى أن B سوف يحدث بشرط حدوث A . ومن السهل إثبات أن الاحتمال الشرطى يحقق بديهيات الاحتمال السابقة.

نظريات حول الاحتمال الشرطى

Theorem on Conditional Probability

نظرية 1-8: لأى ثلاثة أحداث A_1, A_2, A_3 يكون لدينا

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \quad (13)$$

أى أن احتمال الأحداث A_1, A_2, A_3 سوياً يكون مساوياً لاحتمال حدوث الحدث A_1 مضروباً فى احتمال حدوث الحدث A_2 بشرط حدوث A_1 مضروباً حتى احتمال حدوث الحدث A_3 بشرط حدوث الحدثين A_1, A_2 سوياً. ويمكن تعميم هذه النتيجة بسهولة على n حدث.

نظرية 1-9: إذا كان الحدث A سوف ينتج فى أحد الأحداث المتنافية

فيما بينها الآتية A_1, A_2, \dots, A_n ، فإن

$$\begin{aligned} &= P(A_1)P(A | A_1) + P(A_2)P(A | A_2) + \dots \\ &+ P(A_n)P(A | A_n) \end{aligned} \quad (14)$$

Independent Events

الأحداث المستقلة

إذا كان $P(B|A) = P(B)$ ، وهذا يعنى أن حدوث الحدث B لا يتأثر بحدوث أو عدم حدوث الحدث A . وبالتالي فإننا نقول أن الأحداث A, B تمثل أحداثاً مستقلة Independent Events. أى أن:

$$P(A \cap B) = P(A)P(B) \quad (15)$$

ونلاحظ أنه إذا تحققت هذه المعادلة فإننا نقول أن الحدثين A, B هما حدثان مستقلان.

ونقول أن الثلاثة أحداث A_1, A_2, A_3 هي أحداث مستقلة إذا كان كل حدثين مستقلين فيما بينهما. أى أن

$$P(A_j \cap A_k) = P(A_j)P(A_k) \quad j \neq k \quad \text{حيث} \quad j, k = 1, 2, 3 \quad (16)$$

وأيضاً

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \quad (17)$$

لا بد أن تحدث هاتان الخاصيتان لكي يمكننا القول أن الأحداث A_1, A_2, A_3 هي أحداث مستقلة. واستقلال أكثر من ثلاثة أحداث يمكن تعريفه بتعميم هذه النتيجة.



Bayes' Theorem or Rule

نظرية أو قاعدة بيز

نفترض أننا لدينا الأحداث المتنافية فيما بينها الآتية A_1, A_2, \dots, A_n ويكون اتحادها يمثل فراغ العينة وأن أحدها لا بد وأن يحدث. فإنه إذا كان لدينا أى حدث A فسوف تكون لدينا النظرية الهامة الآتية:

نظرية 1-10 قاعدة بيز:

$$P(A_k | A) = \frac{P(A_k)P(A|A_k)}{\sum_{j=1}^n P(A_j)P(A|A_j)} \quad (18)$$

وهذا يمكننا من إيجاد احتمالات حدوث الأحداث A_1, A_2, \dots, A_n الممكن. ولهذا السبب فإن نظرية بيز غالباً ما يشار إلى أنها نظرية على احتمال الأسباب .A Theorem on the Probability of Causes

تحليل التوافيق Combinatorial Analysis



يكون في كثير من الحالات عدد النقاط في فراغ العينة ليس كبيراً، ونكون في حاجة إلى عدد نقط العينة حتى يمكننا إيجاد الاحتمال. وتنشأ المشاكل عندما نحتاج للعد وتوجد صعوبة في التطبيق. ومن هنا نقوم بتحليل التوافيق Combinatorial Analysis والتي يمكن أن تسمى بطرق حساب متقدمة .A Sophisticated Way of Counting

المبادئ الأساسية للعد

Fundamental Principle of Counting

إذا كان من الممكن تحقيق شيء ما بطرق مختلفة عددها n_1 ، وتحقيق شيء آخر بطرق مختلفة عددها n_2, \dots ، وأخيراً يمكن تحقيق شيء k بطرق عددها n_k . فإن هذه الأشياء التي عددها k يمكن تحقيقها معاً طبقاً لأي ترتيب بطرق عددها $n_1 n_2 \dots n_k$.

التباديل Permutations

افترض أننا لدينا n شيء مختلف ونريد ترتيب r منها في خط. وحيث أنه يمكن ترتيب الأول منها بطرق عددها n وترتيب الثاني بطرق عددها

$(n-1), \dots$ ، وترتيب الأخير منها بطرق عددها $(n-r+1)$. وبمبدأ العد الأساسي السابق فإن عدد الترتيبات المختلفة لها جميعاً أو التباديل كما يشار إليها غالباً يكون في الصورة:

$${}_n P_r = n(n-1)\dots(n-r+1) \quad (19)$$

وبسمى الترتيب Arrangements أو التباديل Permutations والذي يمثل ضرب r من العوامل. ونسمى المقدار ${}_n P_r$ بعدد تباديل n من الأشياء تأخذ r مرة.

مثال 1.4: نحتاج إلى أن يجلس 5 رجال و 4 نساء في صف بحيث تمثل المرأة المقاعد الزوجية. ما عدد طرق الترتيب الممكنة؟

Example 1.4. It is required to seat 5 men and 4 women in a row so that the women occupy the even places. How many such arrangements are possible?

يمكن إجلاس الرجال بطرق عددها ${}_5 P_5$ والنساء بطرق عددها ${}_4 P_4$ وحيث أنهم سوف يجلسون سوياً فإن عملية الترتيب الكلية لهم معاً

$$= {}_5 P_5 \cdot {}_4 P_4 = 5! 4! = (120)(24) = 2880$$

وفي الحالة الخاصة عند $r = n$ فإن

$${}_n P_n = n(n-1)(n-2)\dots 1 = n! \quad (20)$$

والتي تسمى مضروب n . n Factorial ويمكن أن تكتب التباديل ${}_n P_r$ على الشكل

$${}_n P_r = \frac{n!}{(n-r)!} \quad (21)$$

بدلالة المضروب. وعندما $n = r$ فإنه يمكن أن نستنتج أن $0! = 1$ وهذا هو تعريف مضروب 0 أي $0!$.

افتراض أن مجموعة تتكون من n شيء بحيث تكون مقسمة إلى النوع الأول n_1 (أي أنها لا تختلف فيما بينها)، n_2 تمثل النوع الثاني، \dots ، n_k

النوع الأخير. بحيث أن $n_1 + n_2 + \dots + n_k = n$. وبالتالي فإن عدد التباديل لهذه الأشياء هو:

$${}^n P_{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!} \quad (22)$$

Combinations

التوافيق

في حالة التباديل قد ركزنا اهتمامنا على ترتيب للأشياء. وكمثال فإن abc . وفي كثير من المشاكل نهتم فقط بالاختيار بدون الأخذ في الاعتبار الترتيب. ويسمى هذا الاختيار بالتوافيق Combination. وكمثال لذلك فإن abc تكون هي bca أي نفس الاختيار.

والعدد الكلي لاختيار r شيء من n شيء (تسمى أيضاً توافيق n

شيء تأخذ r في المرة) ويرمز له بالرمز ${}_n C_r$ أو $\binom{n}{r}$ حيث

$$\binom{n}{r} = {}_n C_r = \frac{n!}{r!(n-r)!} \quad (23)$$

ويكتب أيضاً

$$\binom{n}{r} = \frac{n(n-1)\dots(n-r+1)}{r!} = \frac{{}_n P_r}{r!} \quad (24)$$

ومن السهل إثبات أن

$$\binom{n}{r} = \binom{n}{n-r} \quad \text{or} \quad {}_n C_r = {}_n C_{n-r} \quad (25)$$

مثال 1.5: من 7 حروف ساكنة، 5 حروف متحركة كم عدد الكلمات التي يمكن تكوينها من 4 حروف ساكنة و 3 حروف متحركة؟ ليس من الضروري أن يكون للكلمات معنى.

Example 1.5. From 7 consonants and 5 vowels, how many words can be formed consisting of 4 different consonants and 3 different vowels? The words need not have meaning.

اختيار الأربعة حروف الساكنة من 7 حروف يكون بطرق عددها 7C_4 وعدد طرق اختيار ثلاثة حروف متحركة من خمسة حروف عددها 5C_3 ويمكن ترتيبها مع بعضها بطرق عددها ${}^7P_7 = 7!$ بالتالي فإن عدد الطرق هو:

$${}^7C_4 \cdot {}^5C_3 \cdot 7! = 35 \cdot 10 \cdot 5040 = 1,764,000$$

Binomial Coefficients

معاملات ذات الحدين

تسمى الأرقام الناتجة من صيغة التوافيق بمعاملات ذات الحدين لأنها تتبع من مفكوك ذات الحدين

$$(x + y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \dots + \binom{n}{n}y^n \quad (26)$$

تقريب ستيرلنج للمضروب $n!$

Stirling's Approximation to $n!$

حيثما تكون n كبيرة، فإن إيجاد $n!$ يكون غير عملياً وفي هذه الحالة نلجأ إلى التقريب ذي الشكل التالي:

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \quad (27)$$

حيث أن $e = 2.71828 \dots$ والذي يعتمد على اللوغاريتم الطبيعي. والرمز \sim يعنى أن النسبة في الطرف الأيسر إلى الطرف الأيمن تقترب من الواحد عندما تقترب n من ما لانهاية.

وللحاسبات التكنولوجية أثر على قيمة صيغة تقريب ستيرلنج للحسابات الرقمية ولكنه يظل صالحاً لتقدير التقريبات النظرية (انظر الملحق A).

الفصل الثاني

الإحصاء الوصفي

Descriptive Statistics

في هذا الفصل:

- ✓ الإحصاء الوصفي
- ✓ مقاييس النزعة المركزية
- ✓ الوسط
- ✓ الوسيط
- ✓ المنوال
- ✓ مقاييس التشتت
- ✓ التباين والانحراف المعياري
- ✓ المئين
- ✓ المدى الربيعي
- ✓ الالتواء

Descriptive Statistics

الإحصاء الوصفي

عندما نعطي تقريراً عن مجموعة من البيانات، فإنه من المفيد أن نصفها بلغة شائعة لأكثر عدد من الناس. ولذلك فإنه تم إيجاد مصطلحات تساعدنا على ذلك. وسوف نناقش طرقاً تصف المركز،

التشتت، والشكل الناتج عند إعطاء فئة من البيانات.

مقاييس النزعة المركزية Measures of Central Tendency

مقياس النزعة المركزية يعطى قيمة مفردة تمثل كل القيم المنتجة أو تتوسط كل القيم الناتجة عن تجربة ما. ويمثل الوسط الحسابي Arithmetic Mean المقياس الرئيسى للنزعة المركزية. وبينما يكون الوسط هو الأكثر استخداماً فإنه يوجد مقاييس أخرى يمكن استخدامها، وهى الوسيط Median والمووال Mode.

★ ملاحظة!

يوجد طرق متعددة لمقاييس النزعة المركزية لمجموعة من البيانات ومنها الوسط الحسابي والوسيط والمووال، ولكل منها مزاياه وعيوبه اعتماداً على البيانات والهدف منها.

إذا كان لدينا مجموعة من n من الأرقام، x_1, x_2, \dots, x_n فإن الوسط والذي يرمز له دائماً بالرمز \bar{x} أو μ يكون

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

مثال 2.1: اعتبر مجموعة الأرقام التالية:

Example 2.1. Consider the following set of integers:

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

الوسط الحسابي \bar{x} لهذه المجموعة هو

$$\bar{x} = \frac{1+2+3+4+5+6+7+8+9}{9} = 5$$

Median

الوسيط

الوسيط هو تلك القيمة x بحيث أن $P(X \leq x) = \frac{1}{2}$ و $P(X > x) = \frac{1}{2}$ وبلغت
أخرى يكون الوسيط القيمة التي يكون نصف القيم x_1, x_2, \dots, x_n أكبر
منها ونصفها أصغر منها.

مثال 2.2: اعتبر مجموعة الأرقام

Example 2.2. Consider the following set of integers:

$$S = \{1, 6, 3, 8, 2, 4, 9\}$$

فإذا أردنا إيجاد الوسيط، فإننا نريد إيجاد القيمة x بحيث يكون
نصف عدد القيم قبل x يكون مساوياً لنصف عدد القيم الآخر بعد x .
فنبداً أولاً بترتيب القيم

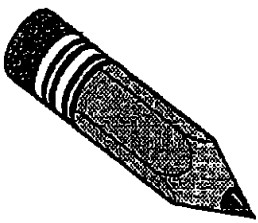
$$S = \{1, 2, 3, 4, 6, 8, 9\}$$

ونلاحظ أن الرقم 4 هو الرقم الذي قبله ثلاثة أرقام وبعده ثلاثة أرقام.
وبالتالي فإن 4 هو الوسيط. وفي بعض الأحيان من الممكن ألا يكون
الوسيط أحد الأرقام المشاهدة.

مثال 2.3: اعتبر مجموعة الأرقام الآتية:

Example 2.3. Consider the following set of integers:

$$S = \{1, 2, 3, 4, 6, 8, 9, 12\}$$



وحيث أن الأرقام مرتبة أصلاً فإنه من الملاحظ أنه
ليس لدينا قيمة واحدة متوسط للمجموعة وبدلاً من
ذلك يوجد لدينا رقمان متوسطان وهما 4، 6

وبالتالى فإن الوسيط يكون رقماً بين 4 و 6. فى أغلب الأحيان هو متوسط الرقمين ويكون الوسيط مساوياً

$$\frac{4+6}{2}=5$$

وبصفة عامة، إذا كان لدينا n قيمة مرتبة وكانت n عدداً فردياً، فإن وسيط البيانات يقع بالضبط فى المنتصف. ويمكن إيجاد موقعه فى المجموعة باستخدام العلاقة $\frac{n+1}{2}$. أما إذا كانت n عدداً زوجياً فإن الوسيط يكون متوسط القيمتين اللتين تقعان فى متوسط المجموعة المرتبة. ويكون موقعهما $\frac{n}{2}$ ، $\frac{n}{2}+1$.

Mode

الموال

الموال لمجموعة البيانات يمثل القيمة التى تحدث غالباً؛ وبلغه أخرى التى يكون لها أكبر احتمال حدوث. وفى بعض الأحيان يكون لدينا أكثر من قيمة لها أكبر احتمال حدوث، وفى هذه الحالة يكون للتوزيع منوالان Bimodal إذا كانت قيمتين أو ثلاثة Trimodal أو أكثر Multimodal على الترتيب.

مثال 2.4: اعتبر التالى الناتج من رمى زهرة بها عشرة جوانب:

Example 2.4. Consider the following rolls of a ten-sided die:

$$R = \{2, 8, 1, 9, 5, 2, 7, 2, 7, 9, 4, 7, 1, 5, 2\}$$

ف نجد أن الرقم الأكثر تكراراً هو الرقم 2. فقد ظهر 4 مرات. ولذلك فإن الموال للمجموعة R هو الرقم 2. لاحظ أنه إذا ظهر الرقم 7 مرة أخرى فيكون عدد مرات ظهوره هو 4 مرات. وفى هذه الحالة يكون لدينا منوالان هما 7 و 2.

اعتبر المجموعتين التاليتين من الأرقام الصحيحة:

$$S = \{5, 5, 5, 5, 5, 5\} \text{ and } R = \{0, 0, 0, 10, 10, 10\}$$



فإذا حسبنا الوسط الحسابي لكل منهما نجده 5 مع
أنهما مجموعتان من البيانات مختلفتان اختلافاً تاماً.
وبالتالي نحتاج إلى إحصاء وصفى بجانب مقياس النزعة
المركزية والذي سوف نسميه مقياس التشتت. وسوف
نقيس التشتت Dispersion أو الانتشار Scatter للقيم حول

الوسط الحسابي للبيانات. فإذا كانت البيانات مركزة حول الوسط
الحسابي فإن المقياس يكون صغيراً؛ بينما إذا كانت البيانات مبعثرة
بعيداً عن الوسط فإن المقياس سوف يكون كبيراً. والمقياسان اللذان
نستخدمهما عادة هما التباين والانحراف المعياري.

التباين والانحراف المعياري

Variance and Standard Deviation

كمية ذات أهمية كبيرة في الاحتمالات والإحصاء تسمى التباين. فالتباين
والذي نرمز له بالرمز σ^2 لمجموعة من الأرقام x_1, x_2, \dots, x_n يمكن
إيجاده كالتالي:

$$\sigma^2 = \frac{[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2]}{n} \quad (2)$$

والتباين يمثل رقم غير سالب. والجذر التربيعي الموجب للتباين
يسمى بالانحراف المعياري.

مثال 2.5: أوجد التباين والانحراف المعياري للمجموعة التالية من
درجات اختبار.

Example 2.5. Find the variance and standard deviation for the following set of test scores:

$$T = \{75, 80, 82, 87, 96\}$$

وحيث أننا نوجد التشتت حول الوسط الحسابي فإننا نحتاج لإيجاد الوسط.

$$\mu = \frac{75 + 80 + 82 + 87 + 96}{5} = 84$$

وباستخدام الوسط يمكن إيجاد التباين.

$$\sigma^2 = \frac{[(75-84)^2 + (80-84)^2 + (82-84)^2 + (87-84)^2 + (96-84)^2]}{5}$$

وذلك يؤدي إلى التالي:

$$\sigma^2 = \frac{[(81) + (16) + (4) + (9) + (144)]}{5} = 50.8$$

وبالتالي فإن التباين لهذه المجموعة من الدرجات هو 50.8 ولإيجاد الانحراف المعياري والذي نرسم له بالرمز σ ، فإننا نوجد الجذر التربيعي للتباين

$$\sigma = \sqrt{\sigma^2} = \sqrt{50.8} = 7.1274118$$

والتباين والانحراف المعياري بصفة عامة أكثر الكميات استخداماً لقياس التشتت ويوجد على أي حال مقاييس أخرى يمكن الإشارة إليها.

✓ يجب أن تعلم

إنه من الممكن القسمة على $(n-1)$ عند إيجاد التباين بدلاً من القسمة على n . ورغم أن النتيجة سوف تختلف إلا أنه عندما تكبر n فإن الفرق سوف يكون صغيراً جداً.

يكون من الأوفق في الغالب تقسيم مجموعة البيانات الترتيبية باستخدام الأحداث الرأسية لكي نحدد النسبة الأقل من نقطة معينة بالنسبة إلى إجمالي البيانات الكلية. والقيم التي تقابل كل مساحة تسمى قيم المئين Percentile Values، واختصاراً المئين Percentiles. وكمثال نسبة الدرجات التي تقع أقل من x_α هي α ؛ أي كمية الدرجات أقل من $x_{0.10}$ أو 10% ونسمى $x_{0.10}$ على أنها 10th Percentile أو العشير. والمثال الآخر هو الوسيط. حيث أن نصف نقاط البيانات تقع أقل منه فيكون هو المئين الخمسين 50th ويرمز له بالرمز $x_{0.50}$ (أو العشير الخامس 5th Decile).

والمئين الخامس والعشرون 25th Percentile يمثل وسيط القيم قبل الوسيط. كما أن المئين الخامس والسبعين 75th Percentile يمثل وسيط القيم بعد الوسيط. ويسمى المئين 25 25th Percentile بالربيع الأول 1st Quartile. كما أن المئين 75 75th Percentile يسمى بالربيع الثالث 3rd Quartile. وبالتالي يعتبر الوسيط هو الربيع الثاني 2nd Quartile.

Interquartile Range

المدى الربيعي

ومقياس آخر من مقاييس التشتت هو المدى الربيعي Interquartile Range. وهو عبارة عن الفرق بين الربيع الثالث والربيع الأول. وبعبارة أخرى

$$x_{0.75} - x_{0.25}$$

مثال 2.6: أوجد المدى الربيعي لمجموعة البيانات التالية لدرجات الجولف.

Example 2.6. Find the interquartile range from the following set of golf scores:

$$S = \{67, 69, 70, 71, 74, 77, 78, 82, 89\}$$

وحيث أننا لدينا 9 نقط وأن البيانات مرتبة تصاعدياً فإن الوسيط يقع في $\frac{9+1}{2}$ أى عند الموقع 5. أى أن الوسيط = القيمة رقم 5 أى يساوى 74.

والربيع الأول $x_{0.25}$ يقع قبل الوسيط بين الأربعة بيانات الأولى. ويكون متوسط القيمة الثانية والقيمة الثالثة؛ أى أن $x_{0.25} = 69.5$.

كما أن الربيع الثالث $x_{0.75}$ يمثل متوسط القيمة السابعة والثامنة. حيث أنه يمثل وسيط الأربعة قيم بعد الوسيط فيكون $x_{0.75} = 80$.

وبالتالى يكون فى النهاية المدى الربيعى مساوياً

$$x_{0.75} - x_{0.25} = 80 - 69.5 = 11.5$$

ومقياس أخير للتشتت يستحق الذكر هو نصف المدى الربيعى Semi-interquartile Range، وكما يدل الاسم فإنه بكل بساطة يساوى نصف المدى الربيعى.

مثال 2.7: أوجد نصف المدى الربيعى للبيانات السابقة

Example 2.7. Find the semiinterquartile range for the previous data set.

$$\frac{1}{2}(x_{0.75} - x_{0.25}) = \frac{1}{2}(80 - 69.5) = 5.75$$

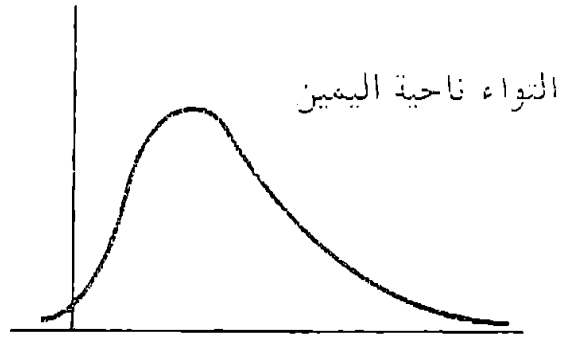
Skewness

الالتواء

وأخر شىء فى الإحصاء الوصفى يمكن أن نشير إليه فى هذا الباب هو شكل توزيع البيانات. فمنها ما يكون متماثلاً Symmetrical Data أو أن تكون البيانات موزعة بانتظام أو أن يكون عدد البيانات فى قيمها العليا أكبر من عددها فى قيمها الصغرى.

وغالباً ما تكون البيانات غير متماثلة حول أى قيمة ولكن بدلاً من ذلك قد تكون القيم الكبيرة أكثر قليلاً أو أقل قليلاً. وفى الحالة التى يكون

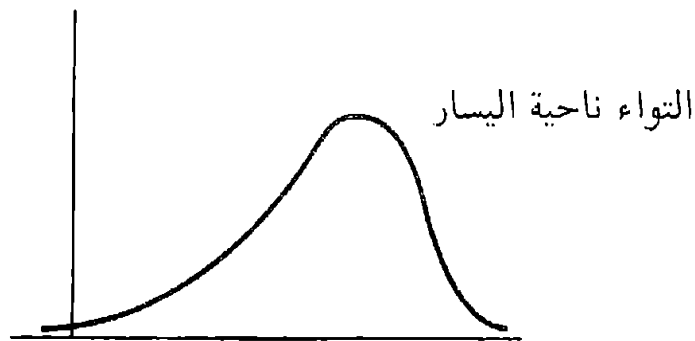
فيها عدد البيانات الكبيرة أكبر يكون الالتواء لليمين Skewed to the Right.



شكل 2-1

التواء ناحية اليمين Skewed to the Right

وفي الحالة التي يكون فيها عدد البيانات الأقل أكثر فإننا نقول أن البيانات ملتوية ناحية اليسار Skewed to the Left.



شكل 2-2

التواء ناحية اليسار Skewed to the Left

★ هام
إذا وجد التواء في مجموعة البيانات ناحية اليمين أو اليسار فإنه يوجد فرصة كبيرة أن تكون هناك قيمة شاذة Outlier في مجموعة بياناتك. والقيم الشاذة تؤثر بشدة على قيمة الوسط الحسابي والانحراف المعياري لمجموعة البيانات وبالتالي فإذا كان توزيع البيانات ملتويًا فيجب أن نفكر في مقاييس أخرى لقياس النزعة المركزية والتشتت بدلاً من الوسط الحسابي والانحراف المعياري.

الفصل الثالث

المتغيرات العشوائية المنفصلة

Discrete Random Variables

في هذا الفصل.

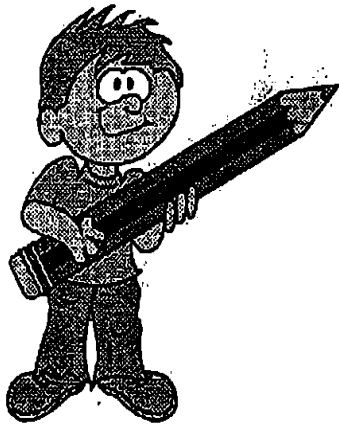
- ✓ المتغيرات العشوائية
- ✓ التوزيع الاحتمالي المنفصل
- ✓ دوال التوزيع للمتغيرات العشوائية
- ✓ دوال التوزيع للمتغيرات العشوائية المنفصلة
- ✓ القيم المتوقعة
- ✓ التباين والانحراف المعياري
- ✓ بعض النظريات على التوقع
- ✓ بعض النظريات على التباين

Random Variables

المتغيرات العشوائية

افتراض أننا لكل نقطة في فراغ العينة نخصص عدد ما. فإننا نحصل على دالة Function معرفة على فراغ العينة. وتسمى هذه الدالة بالمتغير العشوائي Random Variable (أو المتغير التصادفي Stochastic Variable) أو بالتحديد دالة المتغير Random Function (أو الدالة التصادفية Stochastic Function). ويرمز لها عادة بالحروف الكبيرة مثل X أو Y .

وفى العموم يمثل المتغير العشوائى ظاهرة طبيعية أو هندسية أو أى معنى آخر.



ويأخذ المتغير العشوائى عدداً محدوداً Finite Number أو عدد غير محدود يمكن عدّه Countably Infinite Number من القيم ويسمى فى هذه الحالة متغيراً عشوائياً منفصلاً Discrete Random Variable وحينما يأخذ المتغير عدداً لانهائياً من القيم لا يمكن عدّه يسمى بالمتغير العشوائى المتصل أو الغير منفصل Nondiscrete Random Variable.

التوزيع الاحتمالى المنفصل

Discrete Probability Distribution

اعتبر X متغيراً عشوائياً منفصلاً وافترض أنه يأخذ القيم x_1, x_2, \dots, x_n مرتبة طبقاً لترتيب معين. وتأخذ القيم بالاحتمالات الآتية:

$$P(X = x_k) = f(x_k) \quad k = 1, 2, \dots \quad (1)$$

ومن الأفضل وضع دالة الاحتمال والتى نشير إليها بالتوزيع الاحتمالى probability distribution فى الصورة:

$$P(X = x) = f(x) \quad (2)$$

لكل $x = x_k$ وتكون المعادلة السابقة مساوية للصفر لأى فئة أخرى x . حيث $f(x) = 0$.

وبصفة عامة $f(x)$ تكون دالة احتمال إذا كان:

1. $f(x) \geq 0$
2. $\sum_{x} f(x) = 1$

ويؤخذ هذا المجموع على كل قيم x الممكنة.

مثال 3.1: عند رمي قطعة عملة مرتين. وكانت X تمثل عدد الصور الممكن الحصول عليها. فإنه لكل نقطة في فراغ العينة يرتفق رقم لـ X . كالتالي

Example 3.1. Suppose that a coin is tossed twice. Let X represent the number of heads that can come up. With each sample point we can associate a number for X as follows:

نقطة العينة	HH	HT	TH	TT
X	2	1	1	0

الآن يمكننا إيجاد دالة الاحتمال التي تتبع المتغير العشوائي X وبافتراض أن قطعة العملة متوازنة. فإن

$$P(HH) = \frac{1}{4}, \quad P(HT) = \frac{1}{4}, \quad P(TH) = \frac{1}{4}, \quad P(TT) = \frac{1}{4}$$

وبالتالي

$$P(X = 0) = P(TT) = \frac{1}{4}$$

$$P(X = 1) = P(HT \cup TH) = P(HT) + P(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(X = 2) = P(HH) = \frac{1}{4}$$

وبالتالي فإن دالة الاحتمال يمكن الحصول عليها كالتالي

x	0	1	2
$f(x)$	1/4	1/2	1/4

دوال التوزيع للمتغيرات العشوائية

Distribution Functions for Random Variables

تعرف دالة التوزيع التراكمية The Cumulative Distribution أو اختصاراً دالة التوزيع The Distribution Function للمتغير العشوائي X كالتالي

$$F(x) = P(X \leq x) \quad (3)$$

حينما تكون x أى رقم حقيقى. أى أن $-\infty \leq x \leq \infty$
أى أن دالة التوزيع التراكمية تحدد احتمال أن التغير العشوائى
يأخذ أى قيمة أقل من أو تساوى x .
ولدالة التوزيع الخصائص التالية:

1. تكون $F(x)$ دالة غير تناقصية بمعنى أن $F(x) \leq F(y)$ إذا كانت $x \leq y$.

$$2. \lim_{x \rightarrow \infty} F(x) = 1, \quad \lim_{x \rightarrow -\infty} F(x) = 0$$

3. $F(x)$ تكون مستمرة من اليمين؛ بمعنى أن $\lim_{x \rightarrow 0^+} F(x+h) = F(x)$

لكل قيم x .

دوال التوزيع للمتغيرات العشوائية المنفصلة

Distribution Functions for Discrete Random Variables

دالة التوزيع للمتغير العشوائى X يمكن الحصول عليها من دال الاحتمال
باستخدام التالى، لكل x فى المدى $(-\infty, \infty)$

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ f(x_1) + \dots + f(x_n) & x_n \leq x < \infty \end{cases} \quad (4)$$

من الواضح أن دالة الاحتمال للمتغير العشوائى المنفصل يمكن الحصول عليها من دالة التوزيع إذا قمنا بعمل التالى

$$f(x) = P(x) = \lim_{u \rightarrow x} F(u) \quad (5)$$

القيم المتوقعة Expected Values

يوجد مبدأ هام فى الإحصاء والاحتمالات وهو التوقع الرياضى Mathematical Expectation والقيمة المتوقعة Expected Value أو باختصار التوقع Expectation للمتغير العشوائى. بالنسبة للمتغيرات العشوائية المنفصلة X والتى تأخذ القيم x_1, x_2, \dots, x_n فإن توقع X يعرف كالتالى

$$E(X) = x_1 P(X = x_1) + \dots + x_n P(X = x_n) = \sum_{j=1}^n x_j P(X = x_j) \quad (6)$$

أو إذا كان $P(x = x_j) = f(x_j)$

$$E(X) = x_1 f(x_1) + \dots + x_n f(x_n) = \sum_{j=1}^n x_j f(x_j) = \sum_x x f(x) \quad (7)$$

حيث أن الجمع الأخير يؤخذ على كل قيم x . لاحظ أنه عندما تكون كل الاحتمالات متساوية فإن

$$E(X) = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (8)$$

والتي هى ببساطة الوسط الحسابى للقيم x_1, x_2, \dots, x_n

مثال 3.2: افترض وجود لعبة تستخدم فيها زهرة طاولة واحدة من المفترض أنها متوازنة. ويكسب اللاعب \$20 إذا ظهر الوجه 2، و\$40 إذا ظهر الوجه 4، ويخسر \$30 إذا ظهر الوجه 6، بينما لا يخسر ولا يكسب إذا ظهر أى وجه آخر. فأوجد توقع المبلغ المتوقع الذى يكسبه.

Example 3.2. Suppose that a game is to be played with a single die assumed fair. In this game a player wins \$20 if a 2 turns up; \$40 if a 4 turns up; loses \$30 if a 6 turns up; while the player neither wins nor loses if any other face turns up. Find the expected sum of money to be won.

نفترض أن X يمثل المتغير العشوائى الذى يعطى المبلغ الذى يكسبه أو يخسره. والمبالغ التى يكسبها عندما تظهر الزهرة الأرقام 1, 2, 3, 4, 5, 6 هي x_1, x_2, \dots, x_6 على الترتيب. بينما الاحتمالات التى يفترض أن يتحقق بها ذلك هي $f(x_1), f(x_2), \dots, f(x_6)$ وتكون دالة الاحتمال للمتغير X هي

x	0	+20	0	+40	0	-30
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

ولذلك فإن القيمة المتوقعة أو التوقع هو

$$E(X) = (0)\left(\frac{1}{6}\right) + (20)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (40)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (-30)\left(\frac{1}{6}\right) = 5$$

وهذا يعنى أن اللاعب سوف يتوقع أن يكسب \$5. فى هذه اللعبة المتوازنة أو العادلة فإن اللاعب سوف يتوقع أن يدفع \$5 لكى يلعب هذه اللعبة.



تذكرا!
تمثل القيمة المتوقعة للمتغير العشوائى المنفصل بقياس التوزعة المركزية!

التباين والانحراف المعياري

Variance and Standard Deviation

سبق وأن ذكرنا أن توقع المتغير العشوائى المنفصل غالبًا يمثل متوسطه ونرمز له بالرمز μ . وكما لاحظنا فى الباب الثانى وجود كمية أخرى هامة فى الاحتمال والإحصاء هى التباين. فإذا كانت X تمثل متغيراً عشوائياً يأخذ القيم x_1, x_2, \dots, x_n بدالة الاحتمال $f(x)$. فإن علاقة التباين هى

$$\sigma_X^2 = E[(X - \mu)^2] = \sum_{j=1}^n (x_j - \mu)^2 f(x_j) = \sum_x (x - \mu)^2 f(x) \quad (9)$$

وفى حالة خاصة عندما تكون كل الاحتمالات متساوية. فإن

$$\sigma_X^2 = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n} \quad (10)$$

وهو التباين لمجموعة n من الأرقام تأخذ القيم x_1, x_2, \dots, x_n

مثال 3.3: أوجد التباين للعبة فى المثال السابق (3.2).

Example 3.3. Find the variance for the game played in Example 3.2.

قد كانت دالة الاحتمال كالتالى

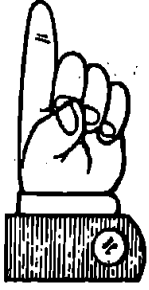
x_j	0	+20	0	+40	0	-30
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

كما نتذكر أن التوقع كان $\mu = 5$ وبالتالي فإن التباين نحصل عليه كالتالى:

$$\begin{aligned} \sigma_X^2 &= (0-5)^2 \left(\frac{1}{6}\right) + (20-5)^2 \left(\frac{1}{6}\right) + (0-5)^2 \left(\frac{1}{6}\right) + (40-5)^2 \left(\frac{1}{6}\right) \\ &+ (0-5)^2 \left(\frac{1}{6}\right) + (-30-5)^2 \left(\frac{1}{6}\right) = \frac{2750}{6} = 458.33\bar{3} \end{aligned}$$

وبالتالى يمكن إيجاد الانحراف المعياري بإيجاد الجذر التربيعي للتباين

$$\sigma_x = \sqrt{458.333} = 21.40872096$$



لاحظ أنه إذا كانت X لها اتجاهات معينة أو وحدات قياس مثل السنتيمتر (cm)، فإن تباين X يأخذ الوحدات cm^2 بينما يأخذ الانحراف المعياري نفس وحدات المتغير X ؛ أى cm. ولهذا السبب يستخدم الانحراف المعياري غالبًا.

بعض النظريات على التوقع

Some Theorems on Expectation

نظرية 3-1: إذا كانت c تمثل أى ثابت. فإن

$$E(cX) = cE(X) \quad (11)$$

نظرية 3-2: إذا كانت X, Y أى متغيرين عشوائيين. فإن

$$E(X+Y) = E(X) + E(Y) \quad (12)$$

نظرية 3-3: إذا كانت X, Y أى متغيرين عشوائيين مستقلين. فإن

$$E(XY) = E(X)E(Y) \quad (13)$$

ملاحظة! ★

تتحقق هذه الخصائص لأي نوع من المتغيرات العشوائية، وليس فقط للمتغيرات العشوائية المنفصلة. سوف نبحث نوع آخر من المتغيرات العشوائية في الباب القادم.

بعض النظريات على التباين

Some Theorems on Variance

نظرية 3-4:

$$\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2 \quad (14)$$

حيث أن $\mu = E(x)$.

نظرية 3-5: إذا كانت c تمثل ثابت. فإن

$$\text{Var}(cX) = c^2 \text{Var}(X) \quad (15)$$

نظرية 3-6: الكمية $E[(X - a)^2]$ تكون أقل ما يمكن حينما

$$a = \mu = E(X) \quad (16)$$

نظرية 3-7: إذا كانت X, Y أي متغيرين عشوائيين مستقلين. فإن

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \quad \text{or} \quad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \quad (17)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) \quad \text{or} \quad \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

لا تنس

أن هذه النظريات تنطبق على التباين وليس على الانحراف المعياري. فتأكد أنك لا بد وأن تحول الانحرافات المعيارية إلى التباينات قبل تطبيق هذه النظريات.

ويمكن بسهولة تعميم النظرية (3-7) على أكثر من متغيرين عشوائيين مستقلين. فالتباين لمجموع عدد من المتغيرات العشوائية المستقلة يكون مساوياً لمجموع تبايناتها.

وتنطبق هذه النظريات أيضاً على المتغيرات العشوائية الغير منفصلة كما تنطبق على المتغيرات المنفصلة.

مثال 3.4: اعتبر X, Y تمثل الأحداث العشوائية المستقلة عند رمي زهرة طاولة متوازنة. فاحسب القيمة المتوقعة للمتغير $X + Y$ ، وتباين $X + Y$.

Example 3.4. Let X and Y be the random independent events of rolling a fair die. Compute the expected value of $X + Y$, and the variance of $X + Y$.

دالة الاحتمالات لكل من X و Y على حدة هي

x_j	1	2	3	4	5	6
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

ومنها يمكن إيجاد

$$\mu_X = \mu_Y = 3.5 \quad \text{و} \quad \sigma_X^2 = \sigma_Y^2 = 2.9166\bar{6}$$

وتوجد طريقتين لحساب $E(X + Y)$ وتباين $Var(X + Y)$

ففي الطريقة الأولى نقوم بحساب التوزيع الاحتمالي للمتغير $X + Y$ ثم نوجد التوقع والتباين منها. لاحظ أن القيم الممكنة للمتغير $X + Y$ هي

2, 3, ..., 11, 12

$x + y$	2	3	4	5	6	
$f(x + y)$	1/36	2/36	3/36	4/36	5/36	
$x + y$	7	8	9	10	11	12
$f(x + y)$	6/36	5/36	4/36	3/36	2/36	1/36

ثم نوجد التوقع كما يلي:

$$E(X + Y) = (2)\left(\frac{1}{36}\right) + (3)\left(\frac{2}{36}\right) + \dots + (11)\left(\frac{2}{36}\right) + (12)\left(\frac{1}{36}\right) = \frac{252}{36} = 7$$

ويتبع ذلك أن التباين يكون:

$$Var(X + Y) = \left[(2 - 7)^2 \left(\frac{1}{36}\right) + \dots + (12 - 7)^2 \left(\frac{1}{36}\right) \right] = \frac{210}{36} = 5.833\bar{3}$$

أما الطريقة الثانية فاستخدام النظرية (3-2) والنظرية (3-7) يجعل ذلك أكثر سهولة.

فباستخدام نظرية (3-2).

$$E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7.$$

وباستخدام نظرية (3-7)

$$Var(X + Y) = Var(X) + Var(Y) = 2.9166\bar{6} + 2.9166\bar{6} = 5.833\bar{3}$$

وحيث أن $X = Y$ فإنه يمكننا أيضاً إيجاد التوقع باستخدام نظرية (3-1):

$$E(X + Y) = E(X + X) = E(2X) = 2[E(X)] = 2(3.5) = 7$$

وعلى أي حال لا يمكننا استخدام نظرية (3-5) لإيجاد التباين حيث أننا استخدمنا أصلاً نفس التوزيع للمتغير X ، كما أن X ليست مستقلة عن نفسها. لاحظ أننا سوف نوجد التباين الخطأ إذا طبقنا هذه النظرية:

$$Var(X + X) = Var(2X) = (2^2)Var(X) = 4Var(X) = 11.66\bar{6}$$

الفصل الرابع

المتغيرات العشوائية المستمرة

Continuous Random Variables

في هذا الفصل:

- ✓ المتغيرات العشوائية المستمرة
- ✓ التوزيع الاحتمالي المستمر
- ✓ دوال التوزيع للمتغيرات العشوائية المستمرة
- ✓ القيم المتوقعة
- ✓ التباين
- ✓ خصائص القيم المتوقعة والتباينات
- ✓ التفسيرات البيانية

المتغيرات العشوائية المستمرة

Continuous Random Variables

يقال على المتغير العشوائي غير المنفصل X مستمر مطلقاً Absolutely Continuous أو مستمر Continuous فقط إذا كانت له دالة التوزيع على الشكل:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad (1)$$

حيث الدالة $f(x)$ لها الخصائص التالية:

$$1. f(x) \geq 0$$

$$2. \int_{-\infty}^{\infty} f(x) dx = 1$$

التوزيع الاحتمالي المستمر

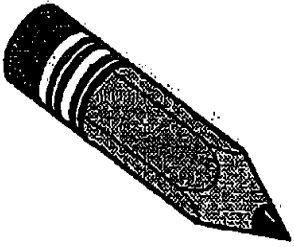
Continuous Probability Distribution

ومن التعريف السابق والخصائص نجد أن المتغير العشوائي المتصل X يكون احتمال أن X تأخذ قيمة معينة للصفري، بينما فترة الاحتمال للمتغير X أن يقع بين قيمتين مختلفتين، نفترض أنها a, b ويكون

$$P(a < X < b) = \int_b^a f(x) dx \quad (2)$$

مثال 4.1: إذا تم اختيار شخص ما من بين مجموعة كبيرة من البالغين فإن احتمال أن يكون طوله X 68 بوصة (أي 68.000 بوصة) بالضبط سيكون صفراً. بينما يكون الاحتمال أكبر من الصفر إذا كانت X بين 67.000 بوصة و68.000 بوصة.

Example 4.1. If an individual were selected at random from a large group of adult males, the probability that his height X is precisely 68 inches (i.e., 68.000... inches) would be zero. However, there is a probability greater than zero that X is between 67.000... inches and 68.000... inches.



والدالة التي تحقق هذه المتطلبات $f(x)$ تسمى بدالة الاحتمال أو دالة التوزيع للمتغير العشوائي المتصل. وتسمى غالباً بدالة كثافة الاحتمال Probability Density Function. وأي دالة $f(x)$ تحقق الشرطين السابقين تسمى دالة الكثافة ويمكن أن تحصل على الاحتمالات من العلاقة (2).

مثال 4.2: أوجد الثابت c لكي تكون الدالة $f(x)$ دالة كثافة احتمال ،
ثم أوجد $P(1 < X < 2)$.

Example 4.2. Find the constant c such that the function

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 0 & \text{غير ذلك} \end{cases}$$

is a density function, and then find $P(1 < X < 2)$.

لاحظ أنه إذا كان $c \geq 0$ فإن الخاصية الأولى سوف تتحقق. وبالتالي فإن $f(x)$ لا بد وأن تحقق الخاصية الثانية لكي تصبح دالة كثافة احتمال،
والآن

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^3 cx^2 dx = \frac{cx^3}{3} \Big|_0^3 = 9c$$

وحيث أن هذا لا بد وأن يساوي 1. فإن $c = \frac{1}{9}$. وتكون دالة كثافة
الاحتمال هي

$$f(x) = \begin{cases} \frac{1}{9}x^2 & 0 < x < 3 \\ 0 & \text{غير ذلك} \end{cases}$$

ثم لإيجاد الاحتمال. نجد أن،

$$P(1 < X < 2) = \int_1^2 \frac{1}{9}x^2 dx = \frac{x^3}{27} \Big|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

دوال التوزيع للمتغيرات العشوائية المستمرة

Distribution Functions for Continuous Random Variables

نتذكر تعريف دالة التوزيع التراكمية أو دالة التوزيع للمتغير العشوائي

$$F(x) = P(X \leq x) \quad (3)$$

حيث أن x أى رقم حقيقى. أى أن $-\infty \leq x \leq \infty$ وبالتالى،

$$F(x) = \int_{-\infty}^x f(x) dx \quad (4)$$

مثال 4.3: أوجد دالة التوزيع فى المثال (4.2).

Example 4.3. Find the distribution function for example 4.2.

$$F(x) = \int_{-\infty}^x f(x) dx = \int_0^x \frac{1}{9} x^2 dx = \frac{x^3}{27}$$

حيث $x \leq 3$

ويوجد علاقة بين دالة التوزيع The Distribution Function ودالة كثافة الاحتمال Density Function ولكى نوجد هذه العلاقة، نفترض احتمال أن المتغير العشوائى X يأخذ القيم على x وأن قيمه تكون فى المدى $x, x + \Delta x$.

فاحتمال أن X تقع بين $x, x + \Delta x$ هى

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(u) du \quad (5)$$

وكانت Δx صغيرة جداً، فإننا نحصل على التقريب

$$P(x \leq X \leq x + \Delta x) \approx f(x)\Delta x \quad (6)$$

ومن (1) بمفاضلة الطرفين - نحصل على

$$\frac{dF(x)}{dx} = f(x) \quad (7)$$

لكل القيم حينما تكون $f(x)$ دالة مستمرة، أى أن تفاضل دالة التوزيع يعطى دالة كثافة الاحتمال.

Expected Values

القيمة المتوقعة

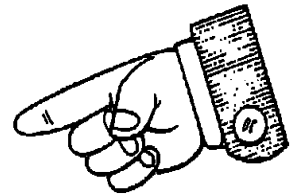
إذا كانت X تمثل متغيراً عشوائياً مستمراً له كثافة الاحتمال $f(x)$. فإن

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (3)$$

مثال 4.4: دالة كثافة الاحتمال للمتغير X هي

Example 4.4. The density function of a random X is given by

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{غير ذلك} \end{cases}$$



والقيمة المتوقعة للمتغير X هي

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^2 x \left(\frac{1}{2}x \right) dx = \int_0^2 \frac{x^2}{2} dx = \frac{x^3}{6} \Big|_0^2 = \frac{4}{3}$$

Variance

التباين

إذا كانت X متغيراً عشوائياً متصلاً له دالة كثافة الاحتمال $f(x)$. فإن التباين يكون

$$\sigma_X^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad (9)$$

بشرط تقارب التباين.

مثال 4.5: أوجد التباين والانحراف المعياري للمتغير العشوائي في المثال السابق (4.4)، باستخدام أن المتوسط تم إيجاده وهو

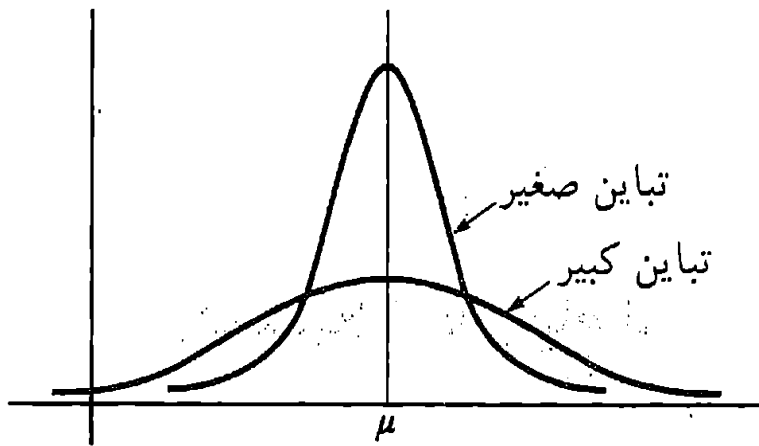
$$\mu = E(x) = \frac{4}{3}$$

Example 4.5. Find the variance and standard deviation of the random variable from Example 4.4, using the fact that the mean was found to be $\mu = E(X) = \frac{4}{3}$.

$$\sigma^2 = E\left[\left(X - \frac{4}{3}\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \frac{4}{3}\right)^2 f(x) dx = \int_{-\infty}^{\infty} \left(x - \frac{4}{3}\right)^2 \left(\frac{1}{2}x\right) dx = \frac{2}{9}$$

وبالتالي يكون الانحراف المعياري هو $\sigma = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$

تذكر أن التباين (أو الانحراف المعياري) يمثل مقياس للتشتت أو الانتشار لقيم المتغير العشوائي حول الوسط μ . وإذا اتجهت البيانات إلى التمرکز قريبا من الوسط فإن التباين يكون صفرًا. بينما إذا انتشرت البيانات بعيدًا عن الوسط فإن التباين يكون كبيرًا. والرسم الموجود في الشكل (4.1) يوضح ذلك لمتغيرين لهما نفس المتوسط μ .



شكل 4-1

خصائص القيم المتوقعة والتباينات

Properties of Expected Values and Variances

لقد تم مناقشة نظريات متعددة عن التوقع والتباين للمتغير العشوائي في الباب الثالث. وحيث أن هذه النظريات تطبق على أي نوع

من المتغيرات العشوائية، فإننا يمكن تطبيقها على المتغيرات العشوائية المتصلة كما تم في المتغيرات العشوائية المنفصلة.

مثال 4.6: إذا كان لدينا دالة كثافة الاحتمال في المثال (4.4) فأوجد $E(3X)$ ، $Var(3X)$.

Example 4.6. Given the probability density function in Example 4.4, find $E(3X)$ and $Var(3X)$.

فإذا استخدمنا طريقة الحساب المباشرة.

$$E(3X) = \int_{-\infty}^{\infty} 3x f(x) dx = \int_0^2 3x \left(\frac{1}{2}x\right) dx = \int_0^2 \frac{3}{2}x^2 dx = \frac{x^3}{2} \Big|_0^2 = 4$$

أما باستخدام نظرية (3.1)، (3.2) على الترتيب، فإننا نحصل على نفس النتيجة كالتالي:

$$E(3X) = 3E(X) = 3\left(\frac{4}{3}\right) = 4$$

أو

$$E(3X) = E(X + X + X) = E(X) + E(X) + E(X) = \frac{4}{3} + \frac{4}{3} + \frac{4}{3} = 4$$

باستخدام نظرية (3.5) لإيجاد التباين

$$Var(3X) = 3^2 Var(X) = 9\left(\frac{2}{9}\right) = 2$$

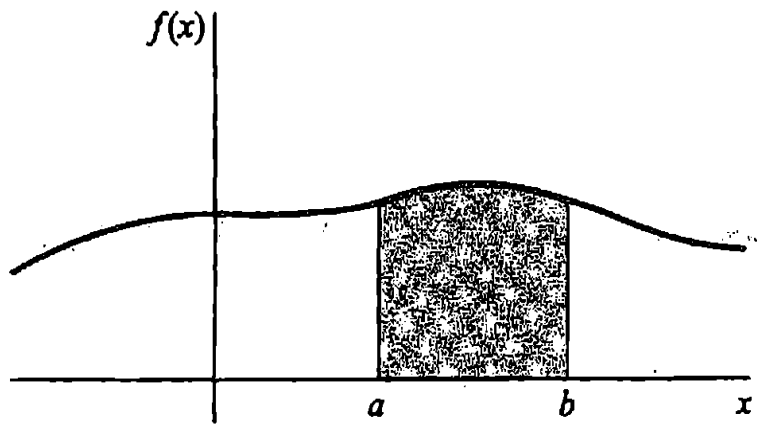
ملاحظة! 

هذه النظريات صحيحة كما رأينا ونجعل عملية الحساب سهلة فتعلمها واستفد من استخدامها فهي ليست للعرض فقط.

Graphical Interpretations

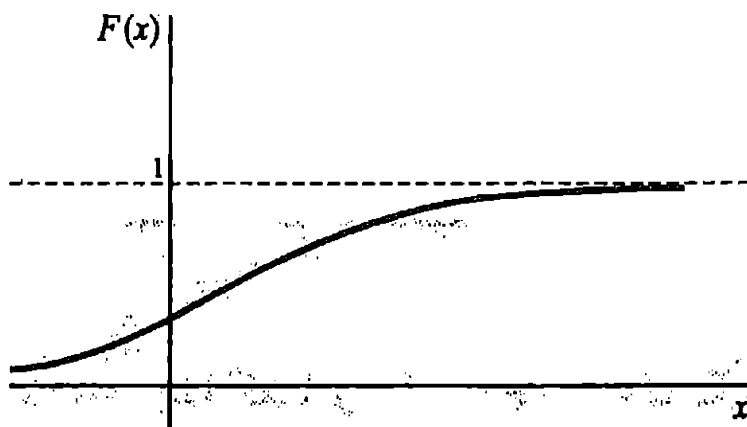
التفسيرات البيانية

إذا كانت $f(x)$ تمثل دالة كثافة الاحتمال للمتغير العشوائي X ، فإنه يمكننا تمثيلها بيانياً بالمنحنى $y=f(x)$ ، كما في الشكل (4-2) وحيث أن $f(x) \geq 0$ فإن المنحنى لن يكون تحت المحور الأفقى x -axis والمساحة الكلية المحدودة بالمنحنى والمحور الأفقى x -axis تكون مساوية للواحد الصحيح طبقاً للخاصية 2. وهندسياً يكون احتمال أن X تقع بين a ، b بمعنى $P(a < x < b)$ يمثل المساحة المظللة في الشكل (4-2).



شكل 4-2

كما أن دالة التوزيع $F(x) = P(X \leq x)$ تكون دالة تزايدية Monotonically Increasing Function وذلك بين صفر وواحد صحيح ويمثلها المنحنى في الشكل (4-3) التالي:



شكل 4-3

الفصل الخامس

أمثلة على المتغيرات العشوائية

Examples of Random Variables

في هذا الفصل:

- ✓ التوزيع ذو الحدين
- ✓ خصائص توزيع ذو الحدين
- ✓ التوزيع الطبيعي
- ✓ أمثلة على التوزيع الطبيعي
- ✓ توزيع بواسون
- ✓ العلاقة بين توزيع ذي الحدين والتوزيع الطبيعي
- ✓ العلاقة بين توزيع ذي الحدين وتوزيع بواسون
- ✓ العلاقة بين توزيع بواسون والتوزيع الطبيعي
- ✓ نظرية الحد المركزية
- ✓ قانون الأعداد الكبيرة

The Binomial Distribution

التوزيع ذو الحدين

نفترض أننا لدينا تجربة عشوائية مثل رمي قطعة عملة أو زهرة طاولة على التوالي أو سحب كرة من صندوق بالتوالي. كل عملية رمي أو



سحب تسمى محاولة Trial. وكل محاولة من المحاولات المتتالية يرتبط بها احتمال حدوث حدث ما مثل حدث الحصول على صورة من رمي قطعة العملة، أربع نقط عند رمي الزهرة أو اختيار كرة بلون معين عند السحب من الصندوق. وفي كل حالة من الحالات السابقة يظل الاحتمال - احتمال حدوث الحدث - ثابتاً لا يتغير من محاولة إلى أخرى (كما في رمي قطعة العملة أو رمي الزهرة). وهذه المحاولات تكون مستقلة Independent وتسمى في الغالب محاولات برنولي Bernoulli Trials وذلك بعد أن درسها جيمس برنولي James Bernoulli في نهاية القرن السابع عشر.

افترض أن p يمثل احتمال حدوث حدث ما في محاولات برنولي (يسمى باحتمال النجاح Probability of Success). وبالتالي يكون $q = 1 - p$ يمثل احتمال عدم حدوث الحدث في محاولة واحدة (يسمى باحتمال الفشل Probability of Failure). واحتمال حدوث الحدث بالضبط x مرة في n محاولة (أي x مرة نجاح، $n - x$ مرة فشل) يتحقق بالمعادلة

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \quad (1)$$

حيث أن التغير العشوائي X يشير إلى عدد مرات النجاح في n محاولة كما أن $x = 0, 1, \dots, n$

مثال 5.1: احتمال الحصول على 2 صورة بالضبط في 6 محاولات من رمي قطعة عملة متكامل التوازن هو

Example 5.1. The probability of getting exactly 2 heads in 6 tosses of a fair coin is

$$P(X = 2) = \binom{6}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{6-2} = \frac{6!}{2!4!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = \frac{15}{64}$$

وتسمى دالة الاحتمال المنفصلة $f(x)$ غالباً توزيع ذا الحدين Binomial Distribution. حيث $x = 1, 2, \dots, n$ والتي تتبع الحدود المتتالية فى مفكوك ذى الحدين.

$$(q + p)^n = q^n + \binom{n}{1} q^{n-1} p + \binom{n}{2} q^{n-2} p^2 + \dots + p^n = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} \quad (2)$$

والحالة الخاصة فى توزيع ذى الحدين عندما $n = 1$ تسمى توزيع برنولى Bernoulli Distribution.

خصائص توزيع ذو الحدين

Properties of Binomial Distributions

كما فى أى توزيع آخر نريد أن نعرف الاحصاءات الوصفية للتوزيع ذو الحدين.

الوسط	$\mu = np$
التباين	$\sigma^2 = np(1-p)$
الانحراف المعياري	$\sigma = \sqrt{np(1-p)}$

مثال 5.2: ارمى قطعة عملة 100 مرة. واحسب عدد مرات ظهور الصورة. أوجد الوسط الحسابى والتباين والانحراف المعياري لهذه التجربة.

Example 5.2. Toss a fair coin 100 times, and count the number of heads that appear. Find the mean, variance, and standard deviation of this experiment.

فى 100 رمية لقطعة عملة متكاملة التوازن، فإن التوقع أو الوسط لعدد

$$\mu = np = 100 \times 0.5 = 50 \text{ مرات ظهور الصورة هو}$$

$$\text{والتباين: } \sigma^2 = (100)(0.5)(0.5) = 25$$

وهذا يعنى أن الانحراف المعياري σ هو $\sigma = \sqrt{(100)(0.5)(0.5)} = \sqrt{25} = 5$

The Normal Distribution

التوزيع الطبيعي

أحد أهم التوزيعات الاحتمالية المتصلة هو التوزيع الطبيعي Normal Distribution. والذي يسمى أحياناً بتوزيع جاوس Gaussian Distribution وكثافة احتمال هذا التوزيع هي

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty \quad (3)$$

حيث أن μ ، σ تمثلان الوسط والانحراف المعياري على الترتيب ودالة التوزيع له هي:

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(v-\mu)^2/2\sigma^2} dv \quad (4)$$

فإذا كانت X لها دالة التوزيع السابقة فإننا نقول أن X موزعة توزيعاً طبيعياً Normally Distributed بالتوقع μ والتباين σ^2 .
وإذا اعتبرنا Z متغيراً عشوائياً يمثل بالعلاقة

$$Z = \frac{X - \mu}{\sigma} \quad (5)$$

فإنها تسمى بالمتغير القياسي التابع للمتغير العشوائي X . والقيمة المتوقعة للمتغير Z هي الصفر بانحراف معياري الوحدة. وفي هذه الحالة تكون دالة كثافة الاحتمال للمتغير Z عندما يكون $\mu = 0$ ، $\sigma^2 = 1$ كالتالي:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (6)$$

وتمثل دالة كثافة الاحتمال المعيارية Standard Normal Density Function. ودالة التوزيع التي تتبعها هي

$$F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z e^{-u^2/2} du \quad (7)$$

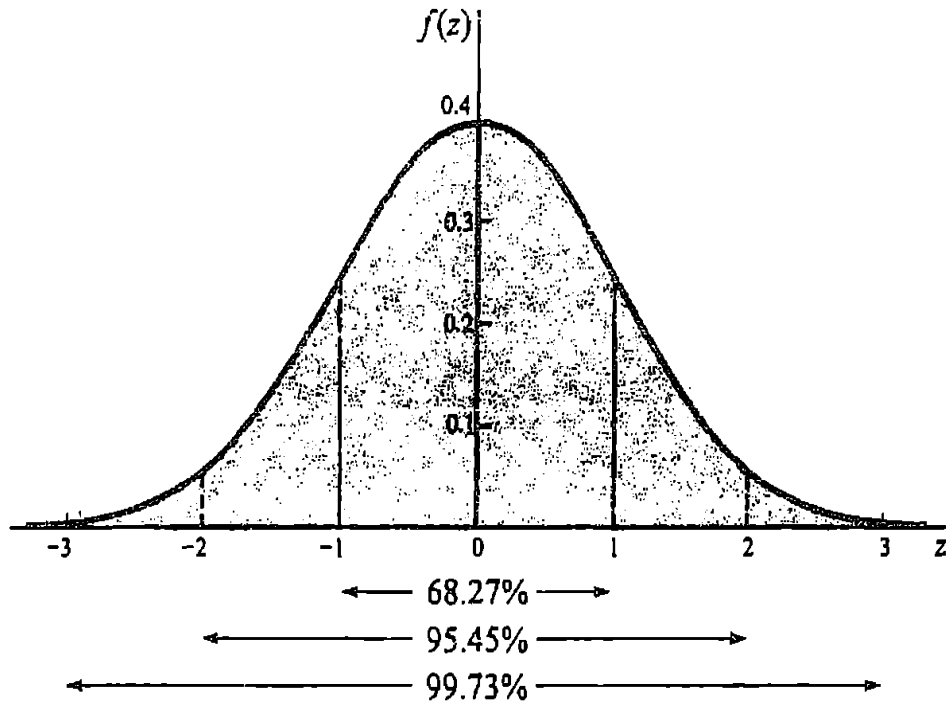
وتسمى أحياناً القيمة Z للمتغير المعياري Z بالدرجة المعيارية
Standard Score.

والرسم البياني لدالة كثافة الاحتمال للتوزيع الطبيعي المعياري يسمى بالمنحنى المعتاد المعياري Standard Normal Curve موضح في الشكل (5-1). حيث نرى على هذا الرسم تظهر المساحات بين 1، 2، 3 انحراف معياري من الوسط (أى بين +1، -1، +2، -2، +3، -3) تساوى 68.27%، 95.45%، 99.73% على الترتيب وذلك من المساحة الكلية التى تساوى الواحد الصحيح. وهذا يعنى أن

$$P(-1 \leq Z \leq 1) = 0.6827$$

$$P(-2 \leq Z \leq 2) = 0.9545$$

$$P(-3 \leq Z \leq 3) = 0.9973$$



شكل 5-1

ملاحظة! ★

للتوزيع الطبيعي أهمية كبيرة، حيث يستخدم كثيراً في الحياة العملية،
فأكد من فهم طريقة استخدامه.

ويوجد في الملحق B جدول يعطى المساحة تحت المنحنى الطبيعي المعياري بين النقط $z = 0$ ، وأي قيمة موجبة للمتغير المعياري z . ومن ذلك الجدول يمكننا إيجاد المساحات بين أي نقطتين متتاليتين على محور z باستخدام خاصية التماثل لمنحنى التوزيع الطبيعي $z = 0$

أمثلة على التوزيع الطبيعي

Examples of the Normal Distribution

وحيث أن هذا التوزيع له أهمية كبيرة فإننا نقدم عدداً من الأمثلة لكيفية استخدامه.

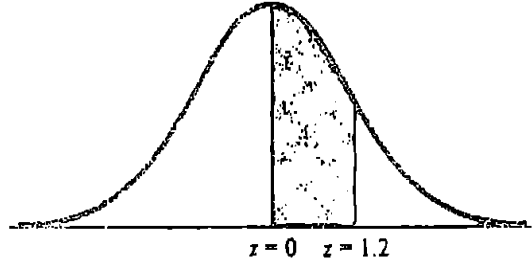
مثال 5.3: أوجد المساحة تحت منحنى التوزيع الطبيعي المعياري بين $z = 0$ ، $z = 1.2$

Example 5.3. Find the area under the standard normal curve between $z = 0$ and $z = 1.2$.

باستخدام جدول الملحق B. واستخدم العمود z حتى نصل إلى 1.2 ونبحث تحت العمود 0. نجد أن النتيجة هي 0.3849 وهي تمثل احتمال أن z تقع بين 0، 1.2. وبالتالي $P(0 \leq Z \leq 1.2) = 0.3849$

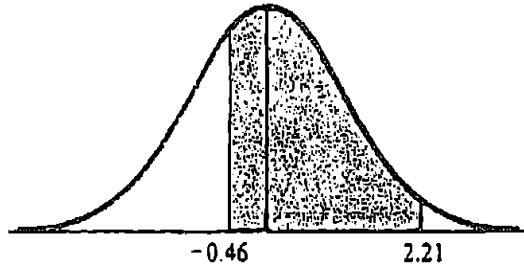
مثال 5.4: أوجد المساحة تحت المنحنى الطبيعي المعياري بين $z = 2.21$ و $z = -0.46$

Example 5.4. Find the area under the standard normal curve between $z = -0.46$ and $z = 2.21$.



شكل 5-2

اعتبر أن الصورة التالية لكثافة احتمال التوزيع الطبيعي.



شكل 5-3

لو نظرنا للشكل (5-3) نجد أن المساحة المطلوبة تنقسم إلى جزئين. الأولى المساحة بين $z = -0.46$ ، $z = 0$ والثانية المساحة بين $z = 0$ ، $z = 2.21$

وحيث أن التوزيع الطبيعي متماثل، فإن المساحة بين $z = -0.46$ ، $z = 0$ هي نفس المساحة بين $z = 0$ ، $z = 0.46$ وباستخدام جدول الملحق B نجدها 0.1772 وبلغة أخرى

$$P(-0.46 \leq Z \leq 0) = P(0 \leq Z \leq 0.46) = 0.1772$$

ونستخدم جدول الملحق B لإيجاد المساحة بين $z = 0$ ، $z = 2.21$ وهي

$$P(0 \leq Z \leq 2.21) = 0.4864$$

وبالتالي تكون المساحة المطلوبة هي المساحة بين $z = -0.46$ ، $z = 2.21$ أي

$$\begin{aligned}
& \text{المساحة الكلية} = (\text{المساحة بين } z = -0.46 \text{ و } z = 0) \\
& + (\text{المساحة بين } z = 0 \text{ و } z = 2.21) \\
& = 0.4864 + 0.1722 = \\
& = 0.6586 \\
& \text{أى أن: } P(0.46 \leq Z \leq 2.21) = 0.6636
\end{aligned}$$

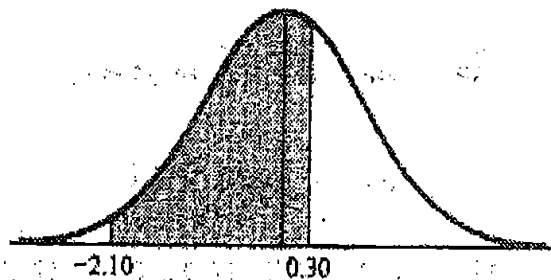
مثال 5.5: إذا كان متوسط وزن 500 طالب في إحدى الكليات هو 151 رطلاً بانحراف معياري 15 رطلاً. افترض أن الأوزان موزعة توزيعاً طبيعياً. أوجد عدد الطلبة الذين يكون وزنهم (a) بين 120 رطلاً، 155 رطلاً. (b) أكثر من 185 رطلاً.

Example 5.5. The mean weight of 500 male students at a certain college is 151 lb and the standard deviation is 15 lb. Assuming the weights are normally distributed, find how many students weigh (a) between 120 and 155 lb, (b) more than 185 lb.

(a) حيث أن الأوزان مسجلة لأقرب رطل وبالتالي فإن الأوزان بين 120 و 155 رطلاً تعنى أن نحسب بين 119.5 و 155.5 رطل. وبالتالي فإن

$$-2.10 = \frac{119.5 - 151}{15} = \text{رطل بالوحدات المعيارية}$$

$$0.30 = \frac{155.5 - 151}{15} = \text{رطل بالوحدات المعيارية}$$



شكل 5-4

وتكون نسبة الطلاب المطلوبة هي = (المساحة بين $z = -2.10$ ، $z = 0.30$)

(المساحة بين $z = -2.10$ ، $z = 0$)

+ (المساحة بين $z = 0$ ، $z = 0.30$)

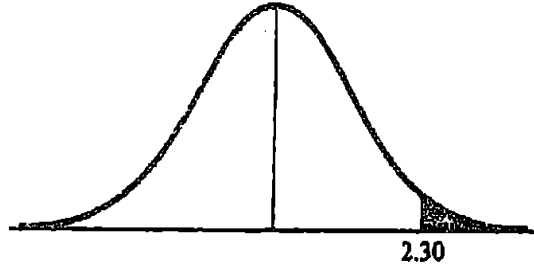
$$0.6000 = 0.4821 + 0.1179$$

وهذا يعني أنه من بين 500 طالب نجد أن 60% يقع وزنهم بين 120، 155 رطلاً. ويكون عددهم

$$\text{طالب } 500 \times 0.6000 = 300$$

(b) لاحظ أن عدد الطلبة الذين يزيد وزنهم عن 185 رطلاً يكون وزنهم على الأقل 185.5 رطل. وبالتالي فإن

$$\text{الدرجة المعيارية للوزن } 185.5 \text{ رطل} = \frac{185.5 - 151}{15} = 2.30$$



شكل 5-5

وتكون النسبة المطلوبة هي

$$P(z \geq 2.30) = \text{(المساحة في الطرف الأيمن بعد } 2.30) =$$

$$\text{(المساحة في الطرف الأيمن } z = 0) =$$

$$- \text{(المساحة بين } z = 0 \text{ ، } z = 2.30) =$$

$$0.0107 = 0.5 - 0.4893 =$$

وبالتالي يكون عدد الطلبة الذين يكون وزنهم أكبر من 185 رطلاً.

$$\text{طلاب } 500 \times 0.0107 = 5$$

وإذا كان W يمثل وزن طالب اختيار بطريقة عشوائية. فإننا نلخص الطريقة السابقة كالتالي

$$P(119.5 \leq W \leq 155.5) = 0.6000 \quad P(W \geq 185.5) = 0.0107$$

Poisson Distribution توزيع بواسون

افترض أن X تمثل متغيراً عشوائياً منفصلاً يأخذ القيم $0, 1, 2, \dots$ بحيث أن دالة الاحتمال للمتغير X هي

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots \quad (8)$$

حيث أن λ تمثل ثابتاً موجباً. يسمى هذا التوزيع بتوزيع بواسون Poisson Distribution (نسبة إلى مكتشفه S.D. Poisson في بداية القرن التاسع عشر). والمتغير الذي يكون له هذا التوزيع يسمى بمتغير له توزيع بواسون.

ويمكن الحصول على قيم توزيع بواسون باستخدام الملحق F، والذي يعطى قيم $e^{-\lambda}$ لقيم λ المختلفة.

مثال 5.6: إذا كان احتمال أن يعاني شخص معين كرد فعل سيئ لإعطائه حقنة من دواء معين هو 0.001. احسب احتمال أنه من بين 2000 شخص (a) يوجد بالضبط 3 يعانون من حقنة هذا الدواء. (b) أكثر من 2 يعانون من الحقن بهذا الدواء.

Example 5.6. If the probability that an individual will suffer a bad reaction from injection of a given serum is 0.001, determine the probability that out of 2000 individuals, (a) exactly 3, (b) more than 2, individuals will suffer a bad reaction.

افترض أن X تمثل عدد الذين يعانون من الأثر لهذا الدواء وموزع

توزيعاً برنولياً. وحيث أن الأثر السيئ نادر الوقوع. فإننا نعتبر أنه موزع توزيعاً بواسونياً. أي أن

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \lambda = np = (2000)(0.001) = 2$$

(a)

$$P(X = 3) = \frac{2^3 e^{-2}}{3!} = 0.180$$

(b)

$$\begin{aligned} P(X > 2) &= 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - \left[\frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} \right] \\ &= 1 - 5e^{-2} \\ &= 0.323 \end{aligned}$$

ويمكن الحصول على تقدير أكثر دقة للاحتمالات باستخدام توزيع ذا الحدين ولكن باستخدام عمليات حسابية أكثر.

العلاقة بين توزيع ذي الحدين والتوزيع الطبيعي

Relationships between Binomial and Normal Distributions

عندما تكون n كبيرة ولم تكن أي من p ، q قريبة جداً من الصفر. فإنه يمكن تقريب التوزيع ذا الحدين إلى توزيع طبيعي بمتغير عشوائى معيارى بحيث أن:

$$Z = \frac{X - np}{\sqrt{npq}} \quad (9)$$

وحيثما X يمثل متغيراً عشوائياً يمثل عدد مرات النجاح فى n

محاولة برنولية وفيها احتمال النجاح يكون مساوياً p . ويكون التقريب جيداً عندما تزيد n . ومن الناحية العملية يكون التقريب جيداً عندما يكون كلا من $np, nq \geq 5$. وتقريب توزيع ذا الحدين إلى توزيع طبيعي يوضح كالتالي

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{X - np}{\sqrt{npq}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du \quad (10)$$

وبلغة أخرى نقول أن المتغير العشوائي المعياري $\frac{(X - np)}{\sqrt{npq}}$ يكون معتاداً تقاربياً *Asymptotically Normal*.

مثال 5.7: احسب احتمال الحصول على عدد صور شاملة بين 3، 6 من رمى قطعة عملة متكاملة التوازن 10 مرات (a) باستخدام توزيع ذا الحدين (b) باستخدام تقريب ذا الحدين إلى توزيع طبيعي.

Example 5.7: Find the probability of getting between 3 and 6 heads inclusive in 10 tosses of a fair coin by using (a) the binomial distribution and (b) the normal approximation to the binomial distribution.

(a) نفترض أن X متغير عشوائي يمثل عدد الصور في تجربة، في قطعة عملة عشرة مرات. وبالتالي

$$P(X=3) = \binom{10}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^7 = \frac{15}{28}, \quad P(X=4) = \binom{10}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6 = \frac{105}{512}$$

$$P(X=5) = \binom{10}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{63}{256}, \quad P(X=6) = \binom{10}{6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4 = \frac{105}{512}$$

ويكون الاحتمال المطلوب

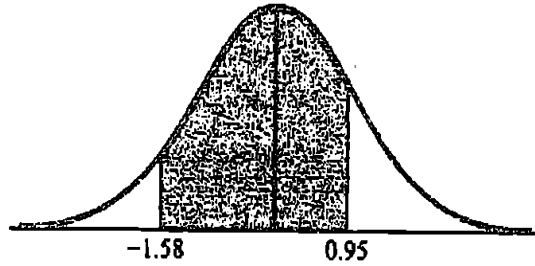
$$P(3 \leq X \leq 6) = \frac{15}{28} + \frac{105}{512} + \frac{63}{256} + \frac{105}{512} = 0.7734$$

(b) وبمعاملة البيانات على أنها مستمرة، وهذا يعني أن عدد الصور 3 إلى 6 يمكن أن نعتبره من 2.5 إلى 6.5 صور. كما أن التوقع والتباين هما: $\mu = np = 10 \times 0.5 = 5$

$$\sigma = \sqrt{npq} = \sqrt{(10)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 1.58.$$

وبالتالي فإن الدرجة المعيارية المناظرة للرقم 2.5 هي: $\frac{2.5-5}{1.58} = -1.58$

والمناظرة للقيمة 6.5 هي: $\frac{6.5-5}{1.58} = 0.95$



شكل 5-6

ويكون الاحتمال المطلوب = (المساحة بين $z = -1.58$ ، $z = 0.95$)

= (المساحة بين $z = -1.58$ ، $z = 0$)

+ (المساحة بين $z = 0$ ، $z = 0.95$)

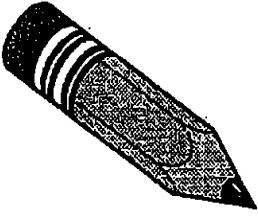
$$= 0.7718 = 0.4429 + 0.3289$$

وتعتبر هذه النتيجة جيدة مقارنة بالنتيجة المضبوطة في (a) ويتلاشى الاختلاف كلما زاد حجم العينة.

العلاقة بين توزيع ذي الحدين وتوزيع بواسون

Relationships between Binomial and Poisson Distributions

في حالة توزيع ذو الحدين عندما تكون n كبيرة وبقترب احتمال النجاح



p من الصفر؛ أى أن $q = 1 - p$ تقترب من الواحد صحيح، فإن الحدث يكون نادر الوقوع Rare Event. وفى الحالات العملية عندما نعتبر الحدث نادر الوقوع ويكون عدد المحاولات لا يقل عن 50.

بينما أن np تكون أقل من 5 ($np < 5$). فإن توزيع ذا الحدين يقترب من توزيع بواسون بالمتوسط $\lambda = np$. وهذا متوقع بمقارنة توقع وتباين كل منهما. وذلك بإحلال $\lambda = np, q \approx 1, p \approx 0$ فى معادلة الوسط والتباين لدى الحدين فإننا نحصل على توقع وتباين التوزيع البواسونى.

العلاقة بين توزيع بواسون والتوزيع الطبيعي

Relationships between Poisson and Normal Distributions

حيث أنه توجد علاقة بين توزيع ذى الحدين والتوزيع الطبيعي وأيضاً بين توزيع ذى الحدين والتوزيع البواسونى، فإننا نتوقع علاقة بين توزيع بواسون والتوزيع الطبيعي. وهذه هى الحالة فى الحقيقة. فإذا كان X متغيراً عشوائياً له توزيع بواسونى بدالة الاحتمال التالية

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (11)$$

وأن

$$\frac{X - \lambda}{\sqrt{\lambda}} \quad (12)$$

يكون هو المتغير العشوائى المعيارى المناظر، وحينئذ فإن

$$\lim_{\lambda \rightarrow \infty} P\left(a \leq \frac{X - \lambda}{\sqrt{\lambda}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du \quad (13)$$

وهذا يعنى أن توزيع بواسون يقترب من التوزيع الطبيعي عندما $\lambda \rightarrow \infty$ أو $\frac{\lambda - \mu}{\sqrt{\lambda}}$ يكون له توزيع طبيعي تقاربى.

نظرية الحد المركزية Central Limit Theorem

يقودنا التشابه بين التوزيعات الثلاثة ذو الحدين وبواسون والطبيعى إلى التساؤل عن وجود توزيعات أخرى بالإضافة إلى ذى الحدين وبواسون تقترب من التوزيع الطبيعي كحالة نهائية. وهذه الملاحظة تكشف عن نظرية الحد المركزية على أن مجموعة من التوزيعات تخضع لهذا التقارب من التوزيع الطبيعي.

نظرية 5-1: نظرية الحد المركزية (Central Limit Theorem) افترض أننا لدينا عدداً من المتغيرات العشوائية المستقلة X_1, X_2, \dots, X_n لها نفس التوزيعات (بمعنى أن كل منها له نفس دالة الاحتمال فى حالة التوزيعات المنفصلة ولها نفس دالة كثافة الاحتمال فى حالة التوزيعات المتصلة) ولها التوقع المحدود μ والتباين σ^2 . فإن

$$S_n = X_1 + X_2 + \dots + X_n \quad (n = 1, 2, \dots),$$

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du \quad (14)$$

وبالتالى فإن المتغير العشوائى $\frac{(S_n - n\mu)}{\sigma\sqrt{n}}$ والذى يمثل متغير معتاد قياسى للمتغير S_n يكون له توزيع طبيعى معيارى. وهذه النظرية صحيحة أيضاً تحت شروط أكثر عمومية، فمثلاً تكون صحيحة عندما تكون المتغيرات X_1, X_2, \dots, X_n عشوائية ومستقلة بنفس المتوسط وبنفس التباين ولكن ليس من الضروري أن يكون لكل منها نفس التوزيع.

نظرية 5-2: قانون الأعداد الكبيرة (Law of Large Numbers) افترض المتغيرات العشوائية المستقلة بالتبادل x_1, x_2, \dots, x_n (منفصلة أو متصلة) لكل منها التوقع المحدود μ والتباين σ^2 . فإنه إذا كان

$$S_n = X_1 + X_2 + \dots + X_n \quad (n = 1, 2, \dots)$$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0 \quad (15)$$

وحيث أن S_n/n تمثل الوسط الحسابي للمجموع $X_1 + X_2 + \dots + X_n$ فإن هذه النظرية تقرر احتمال أن الوسط الحسابي S_n/n يختلف عن العينة المتوقعة μ بأكثر من ε يقترب من الصفر عندما $n \rightarrow \infty$ وتبعاً لذلك فقد نتوقع صحة النتيجة أن $\lim_{n \rightarrow \infty} S_n/n = \mu$ ولكنها تكون بالفعل خطأ. وعلى أى حال يمكن إثبات أن $\lim_{n \rightarrow \infty} S_n/n = \mu$ باحتمال واحد صحيح. وهذه النتيجة تسمى بالقانون القوي للأعداد الكبيرة The Strong Law of Large Numbers. وتسمى نظرية (5-2) بالقانون الضعيف للأعداد الكبيرة The Weak Law of Large Numbers.

الفصل السادس

نظرية المعاينة

Sampling Theory

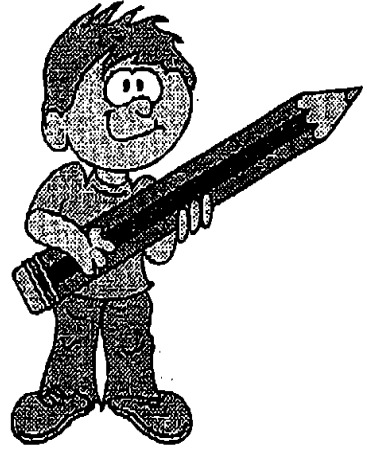
في هذا الفصل:

- ✓ المجتمع والعينة
- ✓ المعاينة
- ✓ العينات العشوائية، الأرقام العشوائية
- ✓ معالم المجتمع
- ✓ إحصاءات العينة
- ✓ توزيعات المعاينة
- ✓ الوسط الحسابي للعينة
- ✓ توزيع المعاينة للمتوسطات
- ✓ توزيع المعاينة للنسب
- ✓ توزيع المعاينة للمجاميع والفروق
- ✓ تباين العينة
- ✓ التوزيعات التكرارية
- ✓ توزيعات التكرار النسبي

نهتم في الحياة العملية بتقرير النتائج على المجموعات الكبيرة من الأفراد أو الأشياء. وبدلاً من أن نختبر كل المجموعة التي تسمى بالمجتمع Population والتي غالباً ما يكون صعب أو مستحيل فحصها، فإننا نقوم بعمل فحص لجزء صغير من هذا المجتمع وهذا الجزء يسمى بالعينة Sample. ونقوم بعمل ذلك بهدف الاستدلال على عدد من الحقائق حول المجتمع محل الدراسة وذلك اعتماداً على نتائج العينة، وهذه العملية تسمى بالاستدلال الإحصائي Statistical Inference. كما أن عملية اختيار مفردات العينة تسمى بالمعاينة Sampling.

مثال 6.1: عندما نرغب في استخلاص نتائج حول نسبة المعيب من العينة من إنتاج المسامير لمصنع معين خلال فترة 6 أيام الأسبوع وذلك باختيار 20 مسماراً يومياً. في هذه الحالة نعتبر كل المسامير المنتجة طول أيام الأسبوع الستة هي المجتمع بينما تمثل العينة 120 مسماراً والتي تم سحبها على أساس 20 مسماراً يومياً.

Example 6.1. We may wish to draw conclusions about the percentage of defective bolts produced in a factory during a given 6-day week by examining 20 bolts each day produced at various times during the day. In this case all bolts produced during the week comprise the population, while the 120 selected bolts constitute a sample.



ولابد أن تسجل عدداً من الملاحظات. أولاً: أن كلمة مجتمع Population ليس لها نفس المعنى بالضرورة في لغة الحياة اليومية مثلما تقول أن «مجتمع القاهرة الآن 16 مليون نسمة». ثانياً: إنه غالباً ما تعنى كلمة مجتمع Population المشاهدات Observations أو القياسات Measurements أكثر من أنها تعنى الأفراد Individuals أو الأشياء

Objects. ثالثاً: إن المجتمع يمكن أن يكون محدوداً Finite أو غير محدود Infinite وذلك اعتماداً على حجم المجتمع Population Size والذي نرسم له بالرمز N . وبالمثل سنسمى حجم العينة للعدد الذي اختير في العينة وسنرمز له بالرمز n وهو محدود عامة.

المعاينة

إذا سحبنا شيئاً من صندوق فإن لنا الاختيار أمام إعادة الشيء إلى الصندوق قبل سحب المرة التالية With Replacing أو عدم إعادته قبل السحب التالي Without Replacing. ففي الحالة الأولى يمكن لشيء معين أن يظهر في الاختيار في مرة تالية ومرة تالية بينما في الحالة الثانية فإنه يمكن أن يظهر في الاختبار مرة واحدة فقط. والمعاينة من مجتمع بحيث أنه يمكن اختيار العضو منه أكثر من مرة يسمى بالمعاينة مع الإعادة Sampling With Replacement بينما المعاينة من المجتمع بحيث لا يختار العضو منه أكثر من مرة يسمى بالمعاينة بدون إعادة Sampling Without Replacement.

✓ يجب أن تعرف

المجتمع المحدود الذي تتم فيه المعاينة مع الإعادة يعتبر نظرياً غير محدود حيث أن العينات لأي حجم يتم سحبها دون أن يتفص حجم المجتمع. ولأهداف عملية فإن المعاينة من مجتمع محدود كبير جداً تعتبر معاينة من مجتمع لانهائي.

العينات العشوائية، الأرقام العشوائية

Random Samples, Random Numbers

من الواضح أن صلاحية النتائج الخاصة بمجتمع تعتمد على صحة

اختيار العينة التي تمثل المجتمع تمثيلاً صادقاً بما فيه الكفاية. وأحد مشاكل الاستدلال الإحصائي الهامة هو عملية اختيار العينة.

وأحد الطرق لاختيار العينة من مجتمع محدود التي يمكن فيها اعتبار أن كل مفردة لها نفس فرصة الاختيار في العينة وهي ما نسميها بالعينة العشوائية Random Sample. والمعينة العشوائية من مجتمع محدود نسبياً تتم باستخدام جدول أعد خصيصاً لذلك الغرض.

وحيث أن الاستدلال الإحصائي من العينة على المجتمع لا يمكن أن يكون مؤكداً فإننا يجب أن نستخدم لغة الاحتمالات في أي بيان يتم استنتاجه.

معالم المجتمع Population Parameters

يعتبر المجتمع معروفاً إذا تم معرفة التوزيع الاحتمالي $f(x)$ (دالة الاحتمال أو دالة الكثافة) للمتغير العشوائي X . ففي المثال (6.1) إذا كانت X تمثل متغيراً عشوائياً والتي تكون قيمة عدد المسامير المعيبة المنتجة كل يوم من أيام الأسبوع فإن X لها التوزيع الاحتمالي $f(x)$.

وإذا كانت X لها توزيع طبيعي. فإننا نقول إن المجتمع موزع توزيعاً طبيعياً Normally Distributed أو إننا لدينا مجتمعاً طبيعياً Normal Population. وبنفس الأسلوب إذا كانت X موزعة طبقاً لذي الحدين فإننا نقول أن المجتمع موزع طبقاً لذي الحدين، أو إننا لدينا مجتمع ذي الحدين Binomial Population.

وبالطبع فإنه من المؤكد إنه إذا كان المجتمع موزعاً توزيعاً طبيعياً فإن كميات مؤكدة لا بد وأن تظهر في $f(x)$ مثل μ ، σ . والمنوال، والالتواء فيمكن إيجادها بدلالة هذه القيمة وتسمى هذه القيم بمعالم المجتمع Population Parameters.



تذكراً!

إذا تم إعطاء المجتمع فهذا يعنى معرفة $f(x)$ وبالتالي
نعرف معالم المجتمع أيضاً.

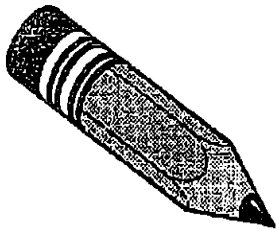
وتظهر لنا مشكلة رئيسية عندما لا نعرف على وجه الدقة التوزيع الاحتمالى للمجتمع $f(x)$. ولكن ربما قد نعرف فكرة عنه أو على الأقل يمكن عمل بعض الفروض حوله، أو نحو سلوك $f(x)$. وكمثال على ذلك عندما نفترض لأسباب ما أن المجتمع له توزيع طبيعى. وفى هذه الحالة ربما لا نعرف قيمة معلمة أو معلمتين من معالمه الاثنين μ ، σ . وبالتالي فإننا يمكن أن نرغب فى الاستدلال الاحصائى حول هذه المعالم.

Sample Statistics

إحصاءات العينة

يمكننا اختيار عينات عشوائية من المجتمع ونستخدمها للحصول على قيم تستخدم للتقدير والاختبار حول معالم المجتمع.

ولكى نتصور ذلك نفترض أننا نرغب فى الحصول على النتائج حول أطوال 12000 طالب وذلك من خلال فحص عينة حجمها 100 طالب فقط، اختيرت عشوائياً من هذا المجتمع. وفى هذه الحالة تكون X متغيراً



عشوائياً يمثل قيمة أطوال الطلاب. وللحصول على العينة من 100 طالب. نختار أولاً الطالب الأول بطريقة عشوائية من المجتمع. وهذه المفردة يمكن أن تأخذ أى قيمة x_1 مثلاً من الأطوال

الموجودة. ونقول إن x_1 هى قيمة المتغير العشوائى X_1 ، حيث يستخدم الدليل 1 دلالة على المفردة الأولى المختارة. ثم نختار المفردة الثانية للعينة والتي يمكن أن تأخذ القيمة x_2 من الأطوال الموجودة. وهى

القيمة التي يأخذها المتغير X_2 . ونستمر في هذه العملية حتى نصل إلى القيمة x_{100} للمتغير X_{100} . حيث أن العينة حجمها 100. وللسهولة نفترض أن عملية السحب هذه مع الإعادة؛ أي أن المفردة يمكن أن تختار أكثر من مرة في العينة. وفي هذه الحالة حيث أن حجم العينة صغير جداً بالنسبة إلى حجم المجتمع فإن المعاينة بدون إعادة تعطى نفس النتائج فعلاً في حالة المعاينة مع الإعادة.

وفي الحالة العامة يكون حجم العينة n والقيم هي x_1, x_2, \dots, x_n للمتغيرات X_1, X_2, \dots, X_n . وفي هذه الحالة تكون في المعاينة مع الإعادة X_1, X_2, \dots, X_n ، متغيرات مستقلة وذات توزيعات متماثلة ولكل منها دالة الاحتمال $f(x)$. ويكون التوزيع المشترك هو

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = f(x_1) f(x_2) \dots f(x_n)$$

والكمية التي نحصل عليها من العينة بهدف تقدير معلمة المجتمع تسمى إحصاء العينة Sample Statistics. وإحصاء العينة ذات الحجم n يمكن أن يعرف رياضياً كدالة في التغيرات العشوائية X_1, X_2, \dots, X_n . أي أن $g(X_1, X_2, \dots, X_n)$ والدالة $g(X_1, X_2, \dots, X_n)$ تمثل هي الأخرى متغيراً عشوائياً آخر تمثل قيمها بالدالة $g(x_1, x_2, \dots, x_n)$.

☆ ملاحظة!

كلمة إحصاء Statistic تطلق غالباً على المتغيرات العشوائية أو قيمها، ويكون واضحاً المفهوم الخاص بها من سياق الموضوع.

وبصفة عامة نجد أنه لكل معلمة من معالم المجتمع يوجد إحصاء يحسب من العينة. وعادة ما تكون طريقة الحصول على هذا الإحصاء

من العينة مماثلة لطريقة الحصول على المعلمة المناظرة له من المجتمع المحدود. حيث أن العينة تتكون من مجموعة محدودة. وكما سوف نرى أن ذلك لن يعطى دائماً أحسن تقدير Best Estimate للمعلمة وأحد المشاكل الهامة في نظرية المعاينة هو تقرير إحصاء العينة الذي يعطى تقديراً للمعلمة ما في المجتمع. سنعتبر تلك المشاكل فيما بعد.

وسوف نستخدم الحروف الإغريقية Greek Letters مثل μ , σ للتعبير عن معالم المجتمع والحروف الرومانية s, m... إلخ للقيمة المقابلة لإحصاءات العينة.

توزيعات المعاينة Sampling Distributions

كما رأينا إحصاء العينة Sample Statistics والذي يحسب من X_1, X_2, \dots, X_n يكون دالة في هذه المتغيرات العشوائية ولذلك يكون هو الآخر متغيراً عشوائياً. والتوزيع الاحتمالي لإحصاء العينة يسمى غالباً بتوزيع المعاينة Sampling Statistic لهذا الإحصاء.

وبصورة بديلة يمكن أن نعتبر كل العينات الممكنة ذات الحجم n والتي يمكن أن تسحب من المجتمع، ولكل عينة تقوم بحساب هذا الإحصاء. بهذه الطريقة يمكن أن نحصل على التوزيع للإحصاء والذي يمثل توزيع المعاينة.

ولتوزيع المعاينة يمكن حساب الوسط والتباين والانحراف المعياري.. إلخ. وأحياناً يسمى الانحراف المعياري أيضاً بالخطأ المعياري Standard Error.

الوسط الحسابي للعينة The Sample Mean

نفترض أن X_1, X_2, \dots, X_n تشير إلى المتغيرات العشوائية المستقلة ذات التوزيعات المتشابهة للعينة العشوائية ذات الحجم n . وبالتالي فإن

متوسط العينة Sample Mean يكون متغيراً عشوائياً يعرف كالتالى

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (1)$$

وإذا كانت x_1, x_2, \dots, x_n ترمز للعينة التى تحصل عليها من عينة حجمها n ، فإن متوسط هذه العينة يرمز له بالرمز

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2)$$

توزيع المعاينة للمتوسطات

Sampling Distribution of Means

اعتبر $f(x)$ التوزيع الاحتمالى لمجتمع معين يتم سحب عينة منه حجمها n . وبالتالى فإنه من الطبيعى أن ندرس التوزيع الاحتمالى لإحصاءات المعاينة \bar{X} ، والتى تسمى بتوزيع المعاينة للوسط الحسابى للعينة أو توزيع المعاينة للوسط وتلخص النظريات التالية ذلك.

نظرية 6-1: الوسط الحسابى لتوزيع المعاينة لمتوسطات والذى يرمز له بالرمز $\mu_{\bar{X}}$ يمكن الحصول عليه كالتالى

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \quad (3)$$

حيث μ هى متوسط المجتمع.

وهذه النظرية تقرر أن القيمة المتوقعة لمتوسط العينة هى متوسط المجتمع.

نظرية 6-2: إذا كان المجتمع لانهاى Infinite والمعاينة عشوائية أو أن المجتمع محدود والمعاينة كانت مع الإعادة، فإن تباين توزيع المعاينة للمتوسطات والذى يرمز له بالرمز $\sigma_{\bar{X}}^2$ ، يكون فى الصورة

$$E[(\bar{X} - \mu)^2] = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \quad (4)$$

نظرية 6-3: إذا كان المجتمع ذا الحجم N ، وكانت المعاينة بدون إعادة وكان حجم العينة $n \leq N$ فإن العلاقة السابقة (4) تصبح

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right) \quad (5)$$

حيث تكون $\mu_{\bar{X}}$ من النظرية 6-1. نلاحظ أن النظرية (6-3) تؤول إلى (6-2) عندما تؤول N إلى ما لانهاية؛ أي أن $N \rightarrow \infty$.

نظرية 6-4: إذا كانت المعاينة من مجتمع طبيعي بالتوقع μ والتباين σ^2 ، فإن متوسط المعاينة يكون موزعاً توزيعاً طبيعياً بالتوقع μ والتباين σ^2/n .

نظرية 6-5: افترض أن المجتمع الذي سحبت منه العينات له توزيع احتمالي بالمتوسط μ والتباين σ^2 وليس من الضروري أن يكون طبيعياً. فإن المتغير المعياري الذي يعبر عن \bar{X} هو

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad (6)$$

يكون موزعاً توزيعاً طبيعياً تقاربياً Asymptotically Normal؛ أي أن

$$\lim_{n \rightarrow \infty} P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du \quad (7)$$

وهذه النظرية هي نتيجة من نظرية الحد المركزية. ويفترض هنا أن المجتمع لانهاية أو أن المعاينة تكون مع الإعادة فإذا كان غير ذلك فإن ما سبق يكون صحيحاً إذا تم استبدال σ/\sqrt{n} في نظرية 6-5 بالتباين $\sigma_{\bar{X}}^2$ في نظرية (6-3).

مثال 6.2: 500 كرة رمان بلى لها متوسط وزن 5.02 أوقية وانحراف معياري 0.30 أوقية. فاحسب احتمال أن 100 كرة رمان بلى اختيرت من هذه المجموعة لها أوزان مشتركة أكثر من 510 أوقية.

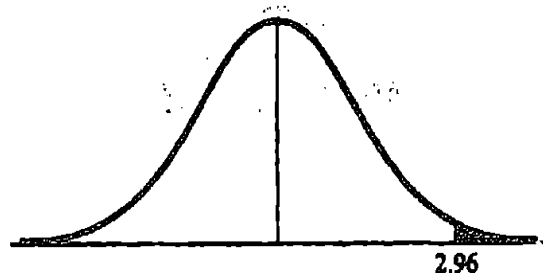
Example 6.2. Five hundred ball bearings have a mean weight of 5.02 oz and a standard deviation of 0.30 oz. Find the probability that a random sample of 100 ball bearings chosen from this group will have a combined weight of more than 510 oz.

بالنسبة لتوزيع المعاينة للمتوسطات فإن $\mu_{\bar{X}} = \mu = 5.02$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{0.30}{\sqrt{100}} \sqrt{\frac{500-100}{500-1}} = 0.027$$

وبالتالي فإن الأوزان المشتركة تزيد عن 510 أوقية إذا كان متوسط الوزن للمائة كرة رمان البلى يزيد عن 5.10 أوقية.

$$Z = \frac{5.10 - 5.02}{0.027} = 2.96$$



شكل 6-1

ويكون الاحتمال المطلوب هو = (المساحة يمين $z = 2.96$)

= (المساحة يمين $z = 0$)

- (المساحة يمين $z = 0$ و $z = 2.96$)

$$= 0.0015 = 0.5 - 0.4985$$

ولذلك فإنه يوجد ثلاث فرص من 2000 عند سحب عينة حجمها 100 كرة رمان بلى بوزن كلى يزيد عن 510 درهماً.

توزيع المعاينة للنسب

Sampling Distribution of Proportions

افترض أن لدينا مجتمعاً غير محدود موزعاً توزيع ذي الحدين باحتمال وجود صفة معينة p واحتمال عدم وجود هذه الصفة $1-p$ ، وكمثال على ذلك يكون المجتمع هو كل رميات قطعة عملة متكاملة التوازن ويكون احتمال الحصول على الصورة $p = 1/2$.

اعتبر أن كل العينات الممكنة ذات الحجم n المسحوبة من هذا المجتمع. وفي كل عينة نحدد الإحصاء لنسبة النجاح p . وفي هذا المثال تكون النسبة تمثل نسبة عدد الصور التي نحصل عليها إلى حجم العينة n ، أو عدد المحاولات n . وبالتالي نحصل على توزيع المعاينة الذي متوسطه μ_p ، وانحرافه المعياري σ_p في الصورة

$$\mu_p = p \quad \sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}} \quad (8)$$

والتي نحصل عليها من النظرية (6-1)، (6-2) على الترتيب وذلك بوضع $\mu = p$ ، $\sigma = \sqrt{pq}$.

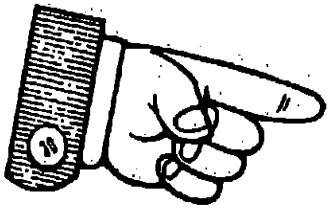
وللقيم الكبيرة n ($n \geq 30$) فإن توزيع المعاينة يكون توزيعاً طبيعياً تقاربياً، كما في نظرية (6-5).

وفي حالة المجتمعات المحدودة والتي تكون فيها المعاينة بدون الإعادة فإن المعادلة σ_p تستبدل بالمعادلة $\sigma_{\bar{x}}$ كما في نظرية (6-3) مع $\sigma = \sqrt{pq}$.

لاحظ أن المعادلات μ_p ، σ_p يمكن إيجادها بسهولة بقسمة الوسط على n للأولى وقسمة الانحراف المعياري على n بالنسبة إلى الثانية (np و \sqrt{npq}) وذلك بالنسبة لتوزيع ذي الحدين.

توزيع المعاينة للفروق وللمجموع

Sampling Distribution of Differences and Sums



نفترض وجود مجتمعين. وتم سحب عينة عشوائية حجمها n_1 من المجتمع الأول وحساب الإحصاء S_1 . وهذا يصل بنا إلى توزيع المعاينة لهذا الإحصاء بحيث يكون متوسطه μ_{S_1} وانحرافه المعياري σ_{S_1} وبالمثل لكل عينة عشوائية من المجتمع الثاني حجمها n_2 تم حساب الإحصاء S_2 وله توزيع بمتوسط μ_{S_2} وانحرافه المعياري σ_{S_2} .
ويأخذ كل التوافق الممكنة للعينات من المجتمعين فإنه يمكن أن نحصل على التوزيع الاحتمالي للفروق بين الإحصائين $(S_1 - S_2)$ Sampling Distribution of Difference. ويكون للمتوسط والانحراف المعياري لهذا التوزيع والذي نرسم له $\mu_{S_1 - S_2}$ ، $\sigma_{S_1 - S_2}$ على الترتيب في الصورة:

$$\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2} \quad \sigma_{S_1 - S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \quad (9)$$

وذلك بشرط استقلال العينات عن بعضها البعض وبلغة أخرى يشترط استقلال المتغيرات S_1 ، S_2 .

فإذا كانت S_1 ، S_2 يمثلان متوسطات العينات من المجتمعين والتي نرسم لها بالرموز \bar{X}_1 ، \bar{X}_2 على الترتيب. فإن توزيع المعاينة للفرق بين المتوسطات في المجتمعات اللانهائية ذات التوقعات μ_1 ، μ_2 والانحرافات المعيارية σ_1 ، σ_2 على الترتيب هو

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2 \quad (10)$$

و

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (11)$$

باستخدام النظريات (6-1)، (6-2). وهذه النتيجة تكون صحيحة أيضاً في حالة المجتمعات المحدودة عندما تكون المعاينة مع الإعادة. ويكون المتغير المعتاد المعياري هو

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (12)$$

وهذه النتيجة توزع توزيعاً قريباً جداً من التوزيع الطبيعي عندما تكون كل من n_1 و n_2 كبيرة ($n_1, n_2 \geq 30$). وبالطبع نحصل على نفس النتيجة في حالة المجتمعات الغير محدودة وتتم المعاينة بدون إعادة باستخدام النظريات (6-1)، (6-3).

وبخصوص الفرق بين نسبتين في مجتمعين من مجتمعات ذي الحدين لهما البارامترات p_1, q_1 و p_2, q_2 على الترتيب. وتكون في هذه الحالة S_1, S_2 يمثلان نسب النجاح p_1, p_2 بحيث يكون متوسطها وانحرافها المعياري للفرق كالتالي

$$\mu_{p_1-p_2} = \mu_{p_1} - \mu_{p_2} = p_1 - p_2 \quad (13)$$

و

$$\sigma_{p_1-p_2} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} \quad (14)$$

وبدلاً من أن نأخذ الفرق بين الإحصاءات نأخذ أحياناً المجموع لهذه الإحصاءات. وفي حالة توزيع المعاينة لمجموع الإحصاءات S_1 و S_2 ، فيكون الوسط والانحراف المعياري على الترتيب

$$\mu_{S_1+S_2} = \mu_{S_1} + \mu_{S_2} \quad \sigma_{S_1+S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \quad (15)$$

وذلك بافتراض أن العينات مستقلة: ويمكن الحصول حينئذ على نتائج مماثلة لـ $\mu_{\bar{X}_1 - \bar{X}_2}$ و $\sigma_{\bar{X}_1 - \bar{X}_2}$.

مثال 6.3: وجد أن نسبة الإنتاج المعيب المنتج بواسطة آلة معينة هي أنه 2%. ما احتمال أن عند شحن 400 وحدة من هذا الإنتاج يكون به 3% أو أكثر معيب؟

Example 6.3. It has been found that 2% of the tools produced by a certain machine are defective. What is the probability that in a shipment of 400 such tools, 3% or more will prove defective?

$$\mu_p = p = 0.02 \text{ and } \sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.02(0.98)}{400}} = \frac{0.14}{20} = 0.007$$

ويستخدم معامل تصحيح المتغيرات المنفصلة، وهو $\frac{1}{2n} = \frac{1}{800} = 0.00125$

فنحصل على

$$z = \frac{0.03 - (0.00125 - 0.02)}{0.007} = 1.25$$

ويكون الاحتمال المطلوب =

المساحة تحت المنحنى الطبيعي يمين القيمة $z = 1.25 = 0.1056$ وإذا لم نستخدم معامل التصحيح فإننا نحصل على الاحتمال مساوياً 0.0764.

The Sample Variance

تباين العينة

إذا كانت X_1, X_2, \dots, X_n ترمز إلى المتغيرات العشوائية لعينة حجمها n ، فإن المتغير العشوائي الذي يعطى تباين العينة Sample Variance هو

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n} \quad (16)$$

لقد رأينا في النظرية (6-1) أن $E(\bar{X}) = \mu$. وإذا وجدنا أن $E(S^2) = \sigma^2$ أي أن القيمة المتوقعة لهذا الإحصاء تكون مساوية لمعلمة المجتمع، فإننا نقول أن الإحصاء يمثل مقدارا غير متحيز Unbiased Estimator وأن العينة تمثل تقديراً غير متحيز Unbiased Estimate لهذه المعلمة. أي أن:

$$E(S^2) = \mu_{S^2} = \frac{n-1}{n} \sigma^2 \quad (17)$$

والذي يكون قريباً جداً من المعلمة σ^2 فقط عند أحجام العينات الكبيرة ($n \geq 30$) . وبالتالي فإن التقدير غير المتحيز يعرف على أساس أن

$$\hat{S}^2 = \frac{n}{n-1} S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1} \quad (18)$$

ومن ثم

$$E(\hat{S}^2) = \sigma^2 \quad (19)$$

ولذلك نرى أن بعض الإحصائيين يختارون \hat{S}^2 عند إيجاد تباين العينة بدلاً من S^2 . أي بالقسمة على $(n-1)$ بدلاً من القسمة على n في المقام الخاص بتعريف S^2 وذلك لأنه بفعل هذا نجد أن كثيراً من النتائج التالية تكون أكثر بساطة.

التوزيعات التكرارية Frequency Distributions

إذا كان لدينا عينة كبيرة (أو مجتمع)، فإنه يكون من الصعب ملاحظة الخصائص المختلفة أو حساب الإحصاءات المختلفة من وسط حسابي أو انحراف معياري. ولهذا السبب فإنه من المفيد تجميع أو تنظيم البيانات الخام Raw Data. ولكي نتصور ذلك نفترض أننا لدينا عينة

تتكون من أطوال 100 طالب في جامعة XYZ، فإننا نقوم بترتيب البيانات Data في فئات Classes أو مجموعات Categories ونقوم بتحديد عدد الطلاب في كل فئة وتسمى بالتكرارات Class Frequency. ويبين جدول (6-1) هذا التنظيم ويسمى بالتوزيع التكراري Frequency Distribution أو الجدول التكراري Frequency Table.

جدول (6-1) يوضح الجدول أطوال 100 طالب في جامعة XYZ

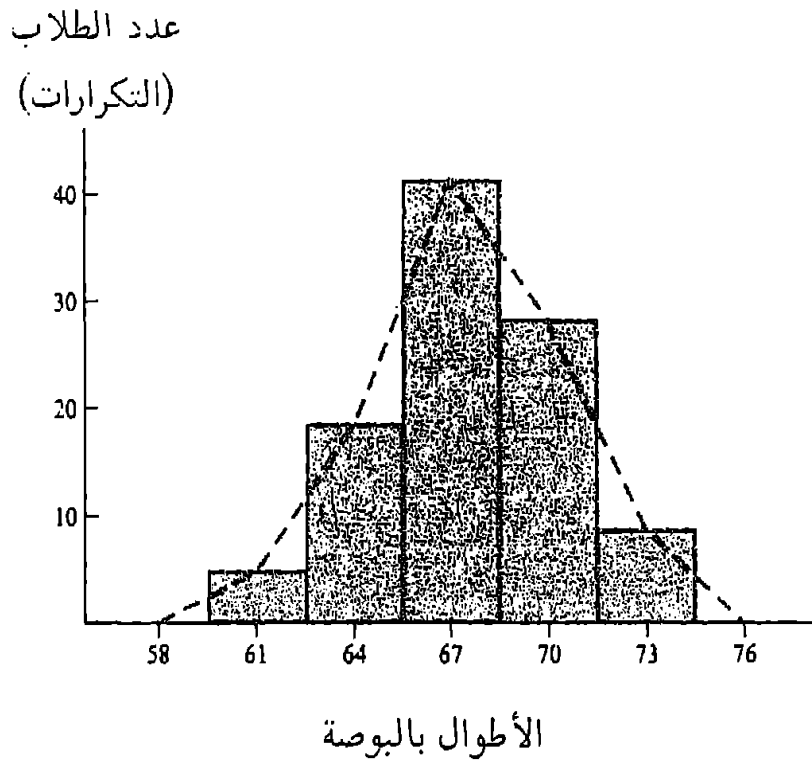
الأطوال بالبوصة	عدد الطلاب
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
المجموع	100



فأول فئة أو صنف يتكون من الأطوال من 60 إلى 62 بوصة، وتكتب 60-62 وتسمى فترة الفئة Class Interval. وحيث أنه يوجد 5 طلبة لهم هذه الأطوال أي داخل هذه الفئة، فإن تكرار هذه الفئة Class Frequency هو 5. وحيث أن الطول المسجل 60 بوصة يكون فعلاً بين 59.5، 60.5 بوصة، بينما الطول المسجل 62 بوصة يكون بين 61.5، 62.5 بوصة، فإننا نسجل فترات الفئات من 59.5-62.5، والفئة التالية تكون 62.5-65.5.. وهكذا في الفئة 59.5-62.5 تسمى الأرقام 62.5، 59.5 بحدود الفئات Class Boundaries. وطول الفئة نرسم له بالرمز c . حيث أنه متساوي لكل الفئات في هذه الحالة وهو الفرق بين الحد الأعلى والحد الأدنى للفئة. وفي هذه الحالة يكون مساوياً $c = 62.5 - 59.5 = 3$

ومركز الفئة The Midpoint of the Class الذي يمثل الفئة يسمى بعلامة
الفئة Class Mark. وفي الجدول (6-1) نجد أن علامة الفئة الذي يمثل
الفئة 60-62 هو 61.

والرسم البياني للتوزيع التكراري يسمى بالهستوجرام Histogram أو
المدرج التكراري. وهو كما هو واضح في الشكل (6-2) أو المضلع
التكراري Frequency Polygon ويتم بربط مراكز الفئات في أعلى المدرج
التكراري. وأنه من المثير أن نجد أن شكل الرسم يشير إلى أن العينة
قد اختيرت من مجتمع للأطوال يتوزع طبيعياً.



شكل 6-2

توزيعات التكرار النسبي

Relative Frequency Distributions

إذا كان في الجدول (6-1) تم تسجيل التكرارات النسبية أو النسب المئوية
بدلاً من عدد الطلبة في كل فئة، فإن النتيجة تسمى النسب أو توزيع

التكرارات النسبية Percentage Frequency Distribution. فمثلاً النسبة أو التكرار النسبي الذي يوجد في الفئة 63-65 هو $\frac{18}{100}$ أو 18%. والمدرج التكراري التابع للتوزيع التكراري النسبي لا يختلف عما هو موجود في الشكل (6-1) ما عدا أن المحور الرأسي يكون مثلاً لتكرارات نسبية بدلاً من تكرارات مطلقة. ويكون مجموع لمساحات الأعمدة واحد صحيح أو 100%.

ويمكن اعتبار التكرارات النسبية توزيع احتمالي. حيث تم استبدال الاحتمالات بالتكرارات النسبية. حيث أن التكرارات النسبية يمكن أن تعتبر احتمالات تجريبية Empirical Probabilities، وبالتالي فإن توزيعات التكرار النسبي تعرف على أنها توزيعات احتمالية تجريبية Empirical Probability Distributions.

الفصل السابع

نظرية التقدير

Estimation Theory

في هذا الفصل:

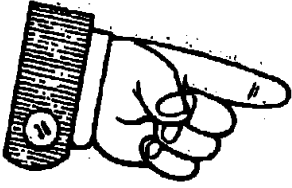
- ✓ التقديرات غير المتحيزة والتقديرات الكفاء
- ✓ التقديرات بنقطة والتقديرات بفترة
- ✓ تقديرات فترة الثقة لعالم المجتمع
- ✓ فترات الثقة للمتوسطات
- ✓ فترات الثقة للنسب
- ✓ فترات الثقة للفرق والمجاميع

التقديرات غير المتحيزة والتقديرات الكفاء

Unbiased Estimates and Efficient Estimates

لقد ذكرنا في الباب السادس أنه يقال أن الإحصاء غير متحيز Unbiased Estimator لمعلمة المجتمع إذا كان متوسط أو توقع الإحصاء مساوياً لقيمة المعلمة. وقيمة الإحصاء في هذه الحالة يقال إنها تقدير غير متحيز Unbiased Estimate للمعلمة.

إذا كان توزيع المعاينة لإحصائين له نفس الوسط، فإن الإحصاء الأقل تباين يسمى المقدر الأكثر كفاءة More Efficient Estimator للوسط.



والقيمة التابعة لهذا الإحصاء الكفاء تسمى بالتقدير الكفاء Efficient Estimate. ومن الواضح أن الإنسان يحاول أن يحصل في نفس الوقت على التقديرات الغير متحيزة والكفاء في نفس الوقت ولكن ذلك يكون غير ممكن دائماً.

التقديرات بنقطة والتقديرات بفترة

Point Estimates and Interval Estimates

يسمى التقدير بواسطة قيمة واحدة لمعلمة المجتمع بالتقدير بواسطة نقطة A Point Estimate، أما تقدير معلمة المجتمع بواسطة رقمين من الممكن أن تقع بينهما معلمة المجتمع فيسمى بالتقدير بفترة Interval Estimate.

مثال 7.1: إذا قلنا أن المسافة هي 5.28 قدم فإننا نكون قد أوجدنا تقدير بنقطة. أما إذا قلنا من ناحية أخرى أن المسافة هي بين 5.28 ± 0.03 قدم؛ أي أن المسافة بين 5.25 قدم و5.31 قدم فإننا نكون قد استخدمنا التقدير بواسطة فترة.

Example 7.1. If we say that a distance is 5.28 feet, we are giving a point estimate. If, on the other hand, we say that the distance is 5.28 ± 0.03 feet, i.e., the distance lies between 5.25 and 5.31 feet, we are giving an interval estimate.

وغالباً ما تسمى العبارة التي تحدد الخطأ أو الدقة لتقدير بالصلاحيه Reliability.

تقديرات فترة الثقة لعالم المجتمع

Confidence Interval Estimates of Population Parameters

اعتبر μ ، σ يمثلان الوسط والانحراف المعياري (الخطأ العياري) لتوزيع المعاينة للإحصاء S . وبالتالي إذا كان توزيع المعاينة للإحصاء

S له توزيع طبيعي تقاربي (والذي وجدنا أنه سوف يكون له توزيع طبيعي عندما تكون $n \geq 30$) فإننا نتوقع وجود S تقع في الفترة $\mu - \sigma$ إلى $\mu + \sigma$ ، $\mu - 2\sigma$ إلى $\mu + 2\sigma$ أو $\mu - 3\sigma$ إلى $\mu + 3\sigma$ بالنسب المئوية 68.27%، 95.45%، 99.73% على الترتيب. وأن نقول نتوقع وجود أو نحن على ثقة في وجود μ في الفترات.

$S - \sigma$ إلى $S + \sigma$ ، $S - 2\sigma$ إلى $S + 2\sigma$ أو $S - 3\sigma$ إلى $S + 3\sigma$

بالاحتمالات 68.27%، 95.45%، 99.73% على الترتيب.

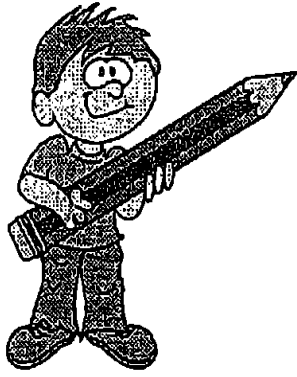
ولهذا سنسمى هذه الفترات المتتالية بأنها 68.27%،

95.45%، 99.73% فترات الثقة Confidence Intervals

لتقدير μ . (أى لتقدير معلمة المجتمع في هذه

الحالة تكون S غير متحيزة). وتسمى حدود هذه

الفترات Confidence Limits $(S \pm \sigma, S \pm 2\sigma, S \pm 3\sigma)$



تسمى 68.27%، 95.45% و 99.73% بحدود الثقة Confidence Intervals

وبالمثل تكون $S \pm 1.96\sigma$ ، $S \pm 2.58\sigma$ هي 95%، 99% (أو 0.95، 0.99)

هي حدود الثقة للمعلمة μ . وتسمى النسبة المئوية للثقة غالباً بمستوى

الثقة Confidence Level. وتسمى الأرقام 1.96، 2.58.. إلخ في حدود

الثقة بالقيم الحرجة Critical Values ويرمز لها بالرمز z_c والتي يمكن

إيجادها مقابل معامل الثقة. ويوجد في جدول (7-1) عدداً من قيم z_c

تتبع عدد من مستويات الثقة التي تستخدم في الحياة العملية. وفي

حالة مستويات الثقة التي لا توجد بالجدول، فإننا نقوم بإيجادها من

جدول مساحات التوزيع الطبيعي في الملحق B.

جدول 7-1

مستوى الثقة	99.73%	99%	98%	96%	95.49%
z_c	3.00	2.58	2.33	2.05	2.00

مستوى الثقة	95%	90%	80%	68.27%	50%
z_c	1.96	1.645	1.28	1.00	0.6745

وفى حالة ما يكون الإحصاء له توزيع معاينة يختلف عن التوزيع الطبيعي فإنه يتم عمل التعديل المناسب للحصول على فترات ثقة مناسبة.

فترات الثقة للمتوسطات

Confidence Intervals for Means

سوف نرى كيف نشأ فترة ثقة للوسط الحسابى للمجتمع باستخدام حالات مختلفة. ففي الحالة الأولى عندما يكون حجم العينة كبيراً ($n \geq 30$) وفي الحالة الثانية عندما يكون حجم العينة صغيراً ($n < 30$) ويكون المجتمع موزعاً توزيعاً طبيعياً.

في حالة العينات الكبيرة ($n \geq 30$) Large Samples ($n \geq 30$)

إذا كان الإحصاء S هو متوسط العينة \bar{X} ، فإن حدود الثقة 95%، 99% لتقدير متوسط المجتمع μ هي $\bar{X} \pm 1.96\sigma_{\bar{X}}$ ، $\bar{X} \pm 2.58\sigma_{\bar{X}}$ على الترتيب وبصفة عامة تكون حدود الثقة هي $\bar{X} \pm z_c \sigma_{\bar{X}}$ حيث أن z_c تعتمد على مستوى الثقة المطلوب ونحصل عليه من جدول (7-1). وباستخدام قيم $\sigma_{\bar{X}}$ والتي سبق تقديمها فى الباب السادس، نجد أن حدود الثقة لمتوسط المجتمع هي:

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}} \quad (1)$$

فى حالة المعاينة من مجتمع غير محدود أو أن تكون المعاينة مع الإعادة

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (2)$$

إذا كان المعاينة من مجتمع محدود N أو أن تكون المعاينة بدون إعادة.

ويكون الانحراف المعياري للمجتمع σ غير معروف فى العادة وبالتالي

فإننا لكي نوجد حدود الثقة هذه لا بد وأن نقوم بتقديره من العينة بالتقديرات S أو $\hat{\sigma}$.

مثال 7.2: أوجد فترة الثقة 95% لتقدير متوسط أطوال 1546 طالباً في جامعة XYZ. بأخذ عينه من 100 طالب (افتراض أن متوسط العينة \bar{x} هو 67.45 بوصة وانحراف المعياري في العينة $\hat{\sigma}$ هو 2.93 بوصة).

Example 7.2. Find a 95% confidence interval estimating the mean height of the 1546 male students at XYZ University by taking a sample of size 100. (Assume the mean of the sample, \bar{x} , is 67.45 and that the standard deviation of the sample, $\hat{\sigma}$, is 2.93 inches.)

حدود الثقة 95% هي $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$

وباستخدام، بوصة $\bar{x} = 67.45$ و بوصة $\hat{\sigma} = 2.93$
كتقدير للانحراف العياري σ فإن حدود الثقة تكون:

$$67.45 \pm 1.96 \left(\frac{2.93}{\sqrt{100}} \right) \text{ بوصة}$$

أو أن

$$67.45 \pm 0.57 \text{ بوصة}$$

وبالتالي فإن فترة الثقة 95% للوسط الحسابي للمجتمع μ هي بين 66.88، 68.02 بوصة؛ أي أن: $66.88 < \mu < 68.02$ وذلك بدرجة ثقة 95%. ويكافئ ذلك القول بأننا على ثقة 95% أن متوسط المجتمع الحقيقي يقع بين 66.88، 68.02 بوصة.

في حالة العينات الصغيرة ($n < 30$) والمجتمع طبيعي

Small Samples ($n < 30$) and Population Normal

سوف نستخدم توزيع t في هذه الحالة (انظر الباب العاشر) للحصول

على مستويات الثقة. وكمثال إذا كانت $t_{0.975}$ ، $-t_{0.975}$ تمثل قيم T بحيث 2.5% من المساحة تقع في نهاية كل طرف من توزيع t ، وبالتالي فإن حدود الثقة للتوزيع T تكون

$$-t_{0.975} < \frac{(\bar{X} - \mu)\sqrt{n}}{\hat{S}} < t_{0.975} \quad (3)$$

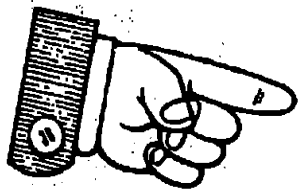
ومن هذه العلاقة يمكن تقدير فترة الثقة للمعلمة μ لتقع في الفترة.

$$\bar{X} - t_{0.975} \frac{\hat{S}}{\sqrt{n}} < \mu < \bar{X} + t_{0.975} \frac{\hat{S}}{\sqrt{n}} \quad (4)$$

وذلك بدرجة ثقة 95%. وبصفة عامة تكون حدود الثقة لمتوسطات المجتمع تكون:

$$\bar{X} \pm t_c \frac{\hat{S}}{\sqrt{n}} \quad (5)$$

حيث أن t_c يمكن إيجادها من الجدول بالملحق C . وبمقارنة العلاقة (5) مع العلاقة (1) تبين أنه لا بد وأن تقوم بإحلال t_c بدلاً من z_c في حالة العينات الصغيرة. وعندما تكون $n > 30$ تكون z_c ، t_c تقريباً متساوية. كما يلاحظ أيضاً أنه في العينات الصغيرة تظهر \hat{S} في العلاقة (5) على أساس استخدام الانحراف المعياري للعينة بدلاً من الانحراف المعياري للمجتمع الذي عادة ما يكون مجهولاً كما في (1).



حجم العينة له أهمية كبيرة! فنحن نقوم بإشياء
فترات الثقة معتمد على حجم العينة ولهذا
بأنك تعرف الأنماط التي سوف تستخدمه.

فترات الثقة للنسب

Confidence Intervals for Proportions

افترض أن الإحصاء S يمثل نسبة النجاح في عينة حجمها $n \geq 30$ سميت عشوائياً من مجتمع ذي الحدين به نسبة النجاح p (احتمال النجاح). فإن حدود الثقة للنسبة p هي $P \pm z_c \sigma_p$. حيث أن P تمثل نسبة النجاح في العينة والتي حجمها n . ونستخدم القيمة σ_p والتي تم حسابها في الباب السادس، وبالتالي فإن حدود الثقة للنسبة في المجتمع هي

$$P \pm z_c \sqrt{\frac{pq}{n}} = P \pm z_c \sqrt{\frac{p(1-p)}{n}} \quad (6)$$

في حالة المعاينة من مجتمع غير محدود أو حالة المعاينة مع الإعادة من مجتمع محدود. وبالمثل تكون الحدود هي

$$P \pm z_c \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}} \quad (7)$$

إذا كانت المعاينة بدون إعادة من مجتمع محدود حجمه N . ويلاحظ أن هذه النتائج قد تم إيجادها من العلاقات (1)، (2) وذلك بإحلال \bar{x} مكان المقدار P ومكان σ المقدار \sqrt{pq} .

ولكى نوجد حدود الثقة نستخدم تقدير العينة P محل المعلمة p .

مثال 7.3: من المعاينة من 100 صوت اختيروا بطريقة عشوائية من أصوات إحدى الدوائر دلت على أن النسبة 55% يصوتون لصالح مرشح معين. فأوجد حدود الثقة 99% لنسبة الأصوات الكلية التي ستكون في صالح هذا المرشح.

Example 7.3. A sample poll of 100 voters chosen at random from all voters in a given district indicate that 55% of them were in favor of a particular candidate. Find the 99% confidence limits for the proportion of all voters in favor of this candidate.

حدود الثقة 99% لصالح المرشح p هي $p \pm 2.58 \sigma_p$.

$$\begin{aligned} P \pm 2.58 \sigma_p &= P \pm 2.58 \sqrt{\frac{p(1-p)}{n}} \\ &= 0.55 \pm 2.58 \sqrt{\frac{(0.55)(0.45)}{100}} \\ &= 0.55 \pm 0.13 \end{aligned}$$

أى أن حدود النسبة هي $0.42 \leq p \leq 0.68$.
أى أن حدود الثقة لنسبة المصوتين لهذا المرشح سوف تكون بين 42%،
68% بدرجة ثقة 99%.

فترات الثقة للفروق والمجاميع

Confidence Intervals for Differences and Sums

إذا كانت S_1 ، S_2 يمثلان إحصاءان للعينة بتوزيعات معتادة تقاربية، فإن
حدود الثقة للفرق بين المعالم التابعة لهذه الإحصاءات S_1 ، S_2 هي

$$S_1 - S_2 \pm z_c \sigma_{S_1 - S_2} = S_1 - S_2 \pm z_c \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \quad (8)$$

بينما حدود الثقة لمجموع المعالم هو

$$S_1 + S_2 \pm z_c \sigma_{S_1 + S_2} = S_1 + S_2 \pm z_c \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2} \quad (9)$$

بشرط استقلال العينات.

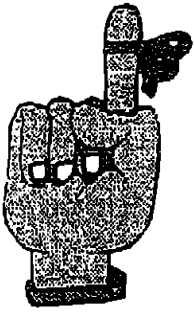
فمثلاً حدود الثقة للفرق بين متوسطى مجتمعين فى حالة ما تكون
المجتمعات غير محدودة وتكون انحرافاتهما المعيارية معروفة σ_1 ، σ_2 هي

$$\bar{X}_1 - \bar{X}_2 \pm z_c \sigma_{\bar{X}_1 - \bar{X}_2} = \bar{X}_1 - \bar{X}_2 \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10)$$

حيث أن \bar{x}_1, n_1 و \bar{x}_2, n_2 يمثلان الأوساط الحسابية وأحجام العينات المسحوبة من المجتمعات على الترتيب. وبالمثل تكون حدود الثقة للفرق بين نسبتى مجتمعين حينما تكون المجتمعات غير محدودة هي

$$P_1 - P_2 \pm z_{\alpha/2} \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}} \quad (11)$$

حيث أن P_1, P_2 يمثلان النسبة في كل عينة على الترتيب. كما أن n_1 و n_2 يمثلان أحجام العينات المسحوبة من المجتمعات.



تذكر!
إن تباين الفرق بين متوسطين هو نفسه تباين مجموع المتوسطين؛ أى أن:
$$\sigma_{X+Y}^2 = \sigma_{X-Y}^2$$

مثال 7.4: في عينة عشوائية من 400 شخص أعمارهم تزيد عن 21 سنة و 600 مراهق يشاهدون برنامج معين فى التلفزيون، أجاب 100 من الذين تزيد أعمارهم عن 21 سنة و 300 من المراهقين أنهم يحبون هذا البرنامج. أوجد حدود الثقة بمستوى 99.73% للفرق بين النسبتين للبالغين والمراهقين الذين يشاهدون البرنامج ومعجبون به.

Example 7.4. In a random sample of 400 adults and 600 teenagers who watched a certain television program, 100 adults and 300 teenagers indicated that they liked it. Construct the 99.73% confidence limits for the difference in proportions of all adults and all teenagers who watched the program and liked it.

حدود الثقة للفرق بين النسبتين هي بتطبيق العلاقة (11) حيث أن الدليل 1، 2 يرمزان للمراهقين والبالغين على الترتيب. حيث أن

$$Q_1 = 1 - P_1 \quad , \quad Q_2 = 1 - P_2$$

كما أن:

$$P_1 = \frac{300}{600} = 0.50 \quad , \quad P_2 = \frac{100}{400} = 0.25$$

وهي نسب المراهقين P_1 والبالغين P_2 الذين يشاهدون البرنامج ويعجبون به. وبالتالي فإن حدود الثقة بدرجة 99.73% هي

$$0.50 - 0.25 \pm 3 \sqrt{\frac{(0.50)(0.50)}{600} + \frac{(0.25)(0.75)}{400}} = 0.25 \pm 0.09 \quad (12)$$

أي أن فترة الثقة بدرجة 99.73% للفرق بين النسبة الحقيقية تقع بين 0.16، 0.34.

الفصل الثامن

اختبارات الفروض والمعنوية

Test of Hypothesis and Significance

في هذا الفصل:

- ✓ القرارات الإحصائية
- ✓ الفروض الإحصائية
- ✓ اختبارات الفروض والمعنوية
- ✓ أخطاء النوع الأول والنوع الثاني
- ✓ مستوى المعنوية
- ✓ الاختبار باستخدام التوزيع الطبيعي
- ✓ اختبار الجانب الواحد واختبار الجانبين
- ✓ القيمة P
- ✓ اختبارات خاصة
- ✓ العلاقة بين نظرية التقدير واختبارات الفروض

Statistical Decisions

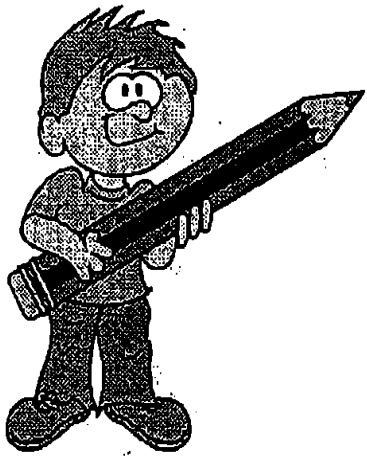
القرارات الإحصائية

غالبًا ما يتم في الحياة العملية اتخاذ قرارات حول المجتمعات بالاعتماد على معلومات من العينة. وتسمى بالقرارات الإحصائية. ومن

أمثلة ذلك إذا ما رغبتنا في تقرير فعالية دواء معين في علاج مرض ما بناء على بيانات العينة، أو ميزة أسلوب تدريس معين عن أسلوب آخر، أو أن تكون عملة معينة غير متوازنة.

الفروض الإحصائية Statistical Hypothesis

في محاولة للوصول إلى قرار فإنه من المفيد وضع افتراضات Assumptions أو تخمينات Guesses حول المجتمعات. ومن هذه الافتراضات التي قد تكون صحيحة أو غير صحيحة ما نسميه بالفروض الإحصائية Statistical Hypothesis وهي بصفة عامة عبارة عن تقريرات Statements حول التوزيعات الاحتمالية للمجتمعات.



وكمثال على ذلك إذا ما أردنا أن نعرف ما إذا كانت قطعة عملة معينة غير متوازنة Loaded، فإننا نقوم بوضع فرض Hypothesis أن قطعة العملة متوازنة Fair بمعنى أن احتمال الحصول على الصورة $p=0.5$. وبالمثل إذا ما أردنا أن نعرف أن أسلوب ما أفضل من أسلوب آخر، فإننا نقوم

بوضع الفرض أنه لا يوجد اختلاف N بين الأسلوبين (أي أن الاختلاف يرجع إلى تقلبات المعاينة من نفس المجتمع) ومثل هذا الفرض يطلق عليه عادة الفرض العدمي Null Hypothesis ويرمز له بالرمز H_0 .

وأي فرض يختلف عن فرض العدم المعطى يسمى بالفرض البديل Alternative Hypothesis. ومن أمثلة ذلك إذا كان الفرض العدمي هو $p=0.5$ فإن الفروض البديلة الممكنة هي $p \neq 0.5$ ، $p=0.7$ أو $p > 0.5$. ويرمز للفرض البديل بالرمز H_1 .

اختبارات الفروض والمعنوية

Tests of Hypothesis and Significance

في حالة افتراض أن فرض معين صحيح True ونجد أن النتيجة في العينة العشوائية تعطي نتيجة تؤكد غير ذلك طبقاً للصدفة المطلقة التي تخضع لها نظرية المعاينة فإننا نقول إننا أمام اختلاف معنوي Significant Differences ونرفض الفرض (أو على الأقل لا نقبله معتمدين على الشواهد التي حصلنا عليها). وكمثال على ذلك إذا حصلنا على 16 صورة عند رمي قطعة عملة عشرين مرة، فإننا نرفض الفرض الذي يرى أن قطعة العملة متوازنة مع أنه يمكن تصور أنه ربما نحن على خطأ.

✓ يجب أن تعرف

إن الطرق التي تمكنتنا من تحديد إما قبول أو رفض الفروض أو تحديد أن العينات المشاهدة تختلف اختلافاً معنوياً عن النتائج المتوقعة تسمى باختبارات الفروض أو اختبارات المعنوية أو قواعد اتخاذ القرارات.

أخطاء النوع الأول والنوع الثاني

Type I and Type II Errors

إذا تم رفض فرض وكان هذا الفرض صحيحاً فإننا نكون قد وقعنا في خطأ من النوع الأول Type I Error وإذا قمنا من ناحية أخرى بقبول فرض كان يجب أن يرفض فإننا نكون قد وقعنا في خطأ من النوع الثاني a Type II Error وفي كل حالة فإن هناك قرار خطأ أو تقرير خطأ قد حدث.

وحتى تكون أية اختبارات فروض أو قواعد قرارات جيدة فلا بد أن يكون الهدف هو تقليل هذه الأخطاء وهي ليست بالعملية السهلة، فإنه لحجم عينة معين نحاول تخفيض نوع معين من الخطأ رغم أنه يتبع ذلك بتزايد النوع الآخر من ناحية أخرى. وفي الحياة العملية يكون أحد هذه الأخطاء أكثر خطورة من الآخر وبالتالي فإننا نحاول أن نوفق أو نصل إلى تحجيم أكثرهما خطورة مع التوفيق بينهما. والطريقة الوحيدة التي تساعدنا على تخفيض الاثنين معاً هي زيادة حجم العينة والتي ربما قد تكون أو لا تكون ممكنة.

Level of Significance

مستوى المعنوية

في اختبار فرض معين أكبر احتمال يمكن الوصول معه إلى خطأ من النوع الأول يسمى بمستوى المعنوية Level of Significance. ويحدد هذا الاحتمال مقدماً قبل البدء في إجراء الاختبار أي قبل إجراء عملية المعاينة حتى لا تؤثر النتائج التي نحصل عليها في قرارنا.

وفي العادة نستخدم في الحياة العملية مستويات المعنوية 0.05 أو 0.01 وهذا لا يمنع من استخدام مستويات أخرى للمعنوية. فإذا تم كمثل اختيار 0.05 أو 5% كمستوى معنوية وذلك لتصميم اختبار فرض ما وهذا يعني أنه يوجد 5 فرص من 100 أنه يتم رفض الفرض الذي كان يجب قبوله؛ أي أنه بينما أن الفرض العدمي هو الصحيح فإننا أمام 95% ثقة أننا سوف نتخذ القرار الصحيح. وبالتالي فإننا نرفض الفرض

★ ملاحظة!

اختيار مستوى المعنوية قبل بداية اختبارات الفروض يساعد بدرجة كبيرة في الاختيار بين رفض أو قبول الفرض العدمي.

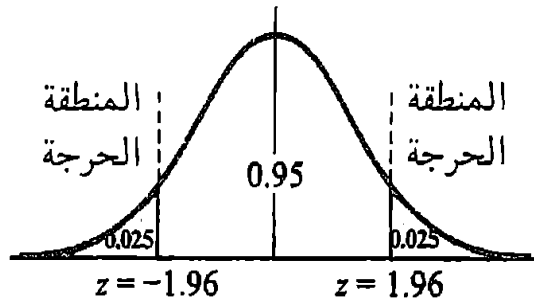
بمستوى معنوية 0.05. Rejected at a 0.05 Level of Significance والتي تعنى
أنا نقع فى خطأ باحتمال 0.05.

الاختبار باستخدام التوزيع الطبيعي

Test Involving the Normal Distribution

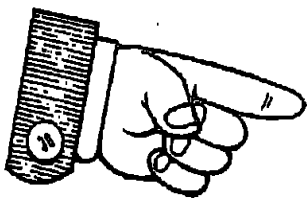
ولتوضيح الفكرة المقدمة سابقاً، نفترض أنه تحت فرض معين فإن توزيع
المعينة للإحصاء S له توزيع طبيعى بالتوقع μ والانحراف المعياري σ_s .
وتوزيع هذا المتغير المعياري $Z = \frac{S - \mu_s}{\sigma_s}$ يكون توزيعاً طبيعياً معيارياً

توقعه صفر وتباينة الوحدة كما هو موجود فى الشكل (8-1). والقيم
المتطرفة للمتغير Z تقودنا إلى مناطق الرفض للفرض العدمى.



شكل 8-1

وكما هو مبين بالرسم نجد أننا نستطيع أن نكون على ثقة بأنه إذا
كان 95% كان الفرض صحيحاً، فإن الدرجة المعيارية Z لإحصاء العينة
الفعلى S يقع بين -1.96، +1.96 (حيث أن المساحة تحت المنحنى
الطبيعى المعياري بين هاتين القيمتين تمثل 95% من المساحة الكلية).



فإذا تم اختيار عينة عشوائية ووجدنا أن الدرجة
المعيارية للإحصاء الخاص بها يقع خارج المدى
-1.96، +1.96 فإننا نستنتج أن هذا الحدث ينتج باحتمال
5% هو الجزء المظلل فى الرسم - الشكل (8-1)

وذلك إذا كان الفرض العدمي صحيح. ونقول أن الدرجة Z تختلف معنوياً عما هو متوقع تحت هذا الفرض وبالتالي فإننا نرفض هذا الفرض العدمي.

والجزء المظلل ويمثل 0.05 هو مستوى المعنوية لهذا الاختبار. وهو يمثل احتمال الخطأ الذي تقع فيه عند رفض الفرض العدمي؛ أي احتمال حدوث خطأ من النوع الأول ولذلك نقول أن الفرض رفض عند مستوى معنوية 0.05، أو أن الدرجة Z لإحصاء معاينة معين يكون معنوياً Significant عند مستوى معنوية 0.05.

والمجموعة من الدرجات المعيارية Z خارج المدى -1.96 ، $+1.96$ تمثل المنطقة الحرجة Critical Region أو منطقة رفض الفرض العدمي. Region of Rejection of the Hypothesis. ومجموعة الدرجات المعيارية داخل المدى $+1.96$ ، -1.96 يمكن أن تسمى إذن منطقة القبول Region of Acceptance of the Hypothesis أو المنطقة غير المعنوية The Region of Nonsignificance.

وعلى أساس الملاحظات السابقة يمكننا صياغة قاعدة الاختبار التالية:

(a) رفض الفرض عند مستوى معنوية 0.05 إذا كانت الدرجة العيارية z للإحصاء S تقع خارج المدى 1.96 ، -1.96 (أي إنه إذا كانت $z > 1.96$ أو $z < -1.96$). وهذا مرادف لأن نقول أن إحصاء العينة المشاهدة يكون معنوياً عند مستوى معنوية 0.05.

(b) قبول الفرض (أو يمكن إذا أردنا عدم اتخاذ قرار مطلقاً) بخلاف ذلك.

ويجب أن نلاحظ أنه يمكن استخدام مستويات معنوية أخرى. كمثال مستوى المعنوية 0.01 وبالتالي نقوم بإحلال 2.58 بدلاً من 1.96 (انظر

جدول (8-1) ويمكن استخدام جدول (7-1) على اعتبار أن مستوى المعنوية مضافاً إلى مستوى الثقة يساوى 100%.

اختبار الجانب الواحد واختبار الجانبين

One-Tailed and Two-Tailed Tests

فى الاختبار السابق قد ركزنا اهتمامنا على القيم المتطرفة Extreme Values للإحصاء S أو القيم z التابعة له على جانبي الوسط؛ أى على ذيلى التوزيع. ولهذا السبب يقال على مثل هذا الاختبار الاختبار ذى الذيلين Two-tailed Tests أو الاختبار ذى الجانبين Two-sided Tests.

وغالباً ما نهتم فقط بالقيم المتطرفة فى طرف واحد أو فى جانب واحد فقط. أى فى ذيل واحد من التوزيع. وكمثال على ذلك حينما نقوم باختبار الفرض الذى يرى أن عملية أفضل من العملية الأخرى. فمثل هذا النوع من الاختبارات يسمى بالاختبارات ذات الذيل الواحد One-tailed Tests أو الاختبارات ذات الاتجاه الواحد One-sided Tests. وفى مثل هذا الاختبار تكون المنطقة الحرجة فى اتجاه واحد فقط من التوزيع بمساحة تكون مساوية لمستوى المعنوية.

والجدول (8-1) الذى يعطينا قيم z للاختبارات فى اتجاه واحد واتجاهين لمستويات المعنوية المختلفة وسوف يفيد كثيراً الرجوع إليه. ويمكن إيجاد القيم الحرجة للدرجة العيارية z لمستويات المعنوية الأخرى من جداول المساحة تحت التوزيع الطبيعى.

جدول 8-1

مستوى المعنوية α	0.10	0.05	0.01	0.005
القيم الحرجة z للاختبارات ذات الذيل الواحد	-1.28 or 1.28	-1.645 or 1.645	-2.33 or 2.33	-2.58 or 2.58
القيم الحرجة z للاختبارات ذات الذيلين	-1.645 and 1.645	-1.96 and 1.96	-2.58 and 2.58	-2.81 and 2.81

القيمة P

P Value

فى معظم الاختبارات نعتبر الفرض العدمى H_0 هو تأكيد على أن معلمة المجتمع تكون مساوية لقيمة معينة وأن الفرض البديل H_1 يكون هو أحد ثلاث قيم أخرى:

- (i) أن المعلمة أكبر من القيمة المحددة (اختبار الذيل الأيمن).
- (ii) أن المعلمة أقل من القيمة المحددة (اختبار الذيل الأيسر).
- (iii) أن المعلمة إما أكبر من أو أقل من القيمة المحددة (اختبار الذيلين).

وفى حالة (i)، (ii) يكون H_1 له اتجاه مفرد بالنسبة إلى المعلمة. أما فى الحالة الثالثة تكون H_1 ثنائية الاتجاه. وبعد عمل الاختبار وحساب الإحصاء S فإن القيمة P (the P Value) للاختبار تكون هى احتمال أن قيمة S فى الاتجاه (أو الاتجاهات) H_1 فى طرف التوزيع مثل ما تحقق فعلاً عندما يكون H_0 صحيح.

وكمثال على ذلك، نفترض أن الانحراف المعياري σ لمجتمع طبيعى معروف ويساوى 3، وأن H_0 ترى أن الوسط μ يكون مساوياً 12. وتم سحب عينة حجمها 36 من هذا المجتمع أعطت متوسط 12.95، فإن إحصاء الاختبار المحسوب من العينة يكون

$$Z = \frac{\bar{X} - 12}{\sigma / \sqrt{n}} = \frac{\bar{X} - 12}{0.5},$$

والذى يمثل المتغير المعتاد العياري عندما يكون H_0 صحيحاً. وتكون قيمة إحصاء الاختبار Z هى

$$Z = \frac{12.95 - 12}{0.5} = 1.9.$$

وتكون القيمة P لهذا الاختبار تعتمد على الفرض البديل كالتالي:

(i) عندما $H_1: \mu > 12$ (الحالة الأولى). فإن القيمة p تمثل احتمال أن العينة العشوائية ذات الحجم 36 تعطى الوسط الحسابي 12.95 أو أكثر إذا كان الوسط الحسابي الحقيقي (متوسط المجتمع) 12؛ أي أن $P(Z \geq 1.9) = 0.029$ وبلغة أخرى أنه يوجد ثلاث فرص في كل 100 فرصة يكون $\bar{x} \geq 12.95$ إذا كان المتوسط μ مساوياً 12.

(ii) عندما تكون $H_1: \mu < 12$ (الحالة الثانية) فإن القيمة P تمثل احتمال أن متوسط العينة العشوائية ذات الحجم 36 تعطى متوسطاً مقداره 12.95 أو أقل إذا كان المتوسط الحقيقي للمجتمع مساوياً 12 أي أن $P(Z \leq -1.9) = 0.971$ وبلغة أخرى أنه يوجد 97 فرصة من 100 يكون فيها متوسط العينة $\bar{x} \leq 12.95$ إذا كان المتوسط الحقيقي μ مساوياً 12.

(iii) عندما تكون $H_1: \mu \neq 12$ (الحالة الثالثة) فإن القيمة P تمثل احتمال أن متوسط العينة العشوائية 0.95 يختلف عن متوسط المجتمع $\mu = 12$ بمعنى أن $\bar{x} \geq 12.95$ أو $\bar{x} \leq 11.05$ إذا كان المتوسط الحقيقي يكون مساوياً للقيمة 12. وتكون هنا القيمة P هي $P(Z \geq 1.9) + P(Z \leq -1.9) = 0.057$ والتي تقول أنه يوجد 6 حالات في كل مائة حالة أن $|\bar{x} - 12| \geq 0.095$ إذا كانت القيمة الحقيقية مساوية للقيمة 12؛ (أي $\mu = 12$).

وقيم P الصغيرة تعطينا دليلاً لرفض الفرض العدمي H_0 لصالح الفرض البديل. كما أن قيم P الكبيرة تعطينا دليلاً قوياً عادلاً لعدم رفض الفرض العدمي في صالح الفرض البديل. ففي الحالة (i) في المثال السابق فإن القيمة الصغيرة $P = 0.029$ تمثل دليلاً قوياً عادلاً على أن متوسط المجتمع أكبر من 12، بينما في الحالة (ii) فإن القيمة الكبيرة $P = 0.971$ تمثل دليل

قوى على أن $H_0: \mu = 12$ لا يجب أن يرفض في صالح الفرض البديل $H_1: \mu < 12$ أما في الحالة (iii) فإن القيمة $P = 0.057$ تمدنا بدليل لرفض H_0 في صالح الفرض البديل $H_1: \mu \neq 12$ ولكن ليس بدليل أكثر من رفض الفرض H_0 في صالح الفرض البديل $H_1: \mu > 12$.

ويجب أن نذكر أن القيمة P ومستوى المعنوية لا تقدم لنا معياراً لرفض أو عدم رفض الفرض العدمي لنفسه ولكن الرفض أو عدم الرفض لصالح الفرض البديل. وكما سبق في المثال التصويرى السابق أن نتائج اختبار متماثلة ومستويات معنوية مختلفة يمكن أن توصلنا إلى استنتاجات مختلفة بالنظر إلى نفس الفرض العدمي بالعلاقة مع فروض مختلفة بديلة.

حيث يكون إحصاء الاختبار S Test Statistic هو متغير عشوائى معتاد معيارى فإن جدول الملحق B سوف يكون كافياً لحساب القيمة P ولكن حينما تكون S هي إحدى المتغيرات العشوائية t ، F ، أو كاي تربيع Chi-square فإن كل منهما له توزيع مختلف يعتمد على درجات الحرية فإنه يوجد برامج جاهزة لحساب القيمة P بالإضافة إلى الجداول بالملاحق (C)، (D)، (E) والعديد من الجداول الأخرى.

مثال 8.1: متوسط فترة حياة 100 لمبة فلورسنت منتجة من مصنع معين هو 1570 ساعة بانحراف معيارى 120 ساعة. إذا كانت μ تمثل متوسط حياة لكل اللمبات الناتجة بواسطة المصنع، فاختر الفرض الذى يرى أن، ساعة $\mu = 1600$ ، ضد الفرض البديل، ساعة $\mu \neq 1600$ ، استخدام مستوى معنوية 0.05 وأوجد القيمة P للاختبار.

Example 8.1. The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. If μ is the mean lifetime of all the

bulbs produced by the company, test the hypothesis $\mu = 1600$ hours against the alternative hypothesis $\mu \neq 1600$ hours. Use a significance level of 0.05 and find the P value of the test.

لا بد وأن نقرر أحد هذين الفرضين

$$H_0: \mu = 1600 \text{ ساعة} , H_1: \mu \neq 1600 \text{ ساعة}$$

سوف نستخدم الاختبار ذا الاتجاهين حيث أن $\mu \neq 1600$ يشمل كل القيم التي تكون أكبر من وأقل من 1600 ساعة في طرفي التوزيع. وللاختبار ذي الاتجاهين بمستوى معنوية 0.05، فإننا يكون لدينا قاعدة القرار التالي:

1. رفض H_0 إذا كانت الدرجة المعيارية \bar{z} لمتوسط العينة خارج المدى -1.96 إلى 1.96.
2. قبول H_0 غير ذلك.

والإحصاء محل البحث هو الوسط الحسابي \bar{X} . والتوزيع الاحتمالي للوسط الحسابي X يكون له المتوسط $\mu_{\bar{X}} = \mu$ والانحراف المعياري $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. حيث أن μ و σ يمثلان الوسط الحسابي والانحراف المعياري للمجتمع لكل اللببات المنتجة بواسطة الشركة.

وتحت ظل الفرض العدمي H_0 حيث، ساعة $\mu = 1600$ ،

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{100}} = 12$$

باستخدام الانحراف المعياري للعينة كتقدير σ . وحيث أن Z هي

$$Z = \frac{\bar{X} - 1600}{12} = \frac{1570 - 1600}{12} = -2.50$$

فتقع خارج المدى -1.96 إلى 1.96. وبالتالي فإننا نرفض H_0 عند مستوى معنوية 0.05.

أما القيمة P في حالة اتجاهين هي

$$P(Z \leq -2.50) + P(Z \geq 2.50) = 0.0124$$

والتي تمثل احتمال أن متوسط فترة حياة اللمبات أقل من 1570 ساعة أو أكثر من 1630 ساعة تحدث بالصدفة حينما تكون H_0 صحيحة.

اختبارات خاصة Special Tests

في حالة العينات الكبيرة، تقترب كثيراً من الإحصاءات من التوزيع الطبيعي بالمتوسط μ_s والانحراف المعياري σ_s . وفي هذه الأحوال نستخدم هذه النتائج لكي نصيغ قواعد الاختبار Decision Rules أو اختبارات الفروض والمعنوية. والحالات الآتية تمثل حالات خاصة للإحصاءات ذات الأهمية العملية. وفي كل حالة تعتبر النتائج على المجتمعات الغير محدودة Infinite أو في حالة المعاينة مع الإعادة نطبق التوزيع الطبيعي. وفي حالة المعاينة بدون إعادة من مجتمعات محدودة فلا بد وأن يوجد تعديل بالنتائج وسوف نعتبر فقط الحالات ذات العينات الكبيرة $n \geq 30$.

1. المتوسطات Means: هنا تكون $S = \bar{X}$ متوسط العينة $\mu_s = \mu_{\bar{X}} = \mu$ يكون مساوياً متوسط المجتمع، $\sigma_s = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ حيث أن σ تمثل

الانحراف المعياري للمجتمع و n هي حجم العينة والمتغير المعياري هو

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad (1)$$

يمكن للضرورة أن نستخدم الانحراف المعياري s من العينة (أو \hat{s})

كتقدير للانحراف المعياري للمجتمع σ .

ولاختبار الفرض العدمي H_0 أن متوسط المجتمع $\mu = a$. فإننا

نستخدم الإحصاء (1). وبالتالي، فإذا كان الفرض البديل $H_1: \mu \neq a$ نستخدم الاختبار في اتجاهين. ونقبل H_0 (أو على الأقل لا نرفض) بمستوى معنوية 0.05 إذا ما عينة ذات الحجم n لها متوسط \bar{x} تعطى

$$-1.96 \leq \frac{\bar{x} - a}{\sigma / \sqrt{n}} \leq 1.96 \quad (2)$$

ونرفض الفرض العدمي في الحالات غير ذلك. ولمستويات المعنوية الأخرى فإننا نقوم بتغيير العلاقة (2) تبعاً لذلك. ولكي نختبر الفرض H_0 ضد الفرض البديل H_1 أن الوسط الحسابي للمجتمع أكبر من a ، فإننا نستخدم الاختبار في اتجاه واحد. ونقبل H_0 (أو على الأقل لا نرفضه) عند مستوى معنوية 0.05 إذا

$$\frac{\bar{x} - a}{\sigma / \sqrt{n}} < 1.645 \quad (3)$$

(انظر الجدول 8-1) ونرفضه في الحالات غير ذلك. ولاختبار H_0 ضد الفرض البديل H_1 أن الوسط الحسابي للمجتمع أقل من a ، فإننا نقبل H_0 عند مستوى معنوية 0.05 إذا

$$\frac{\bar{x} - a}{\sigma / \sqrt{n}} > 1.645 \quad (4)$$

2. **النسب Proportions**: هنا تكون $S = P$ هي نسبة النجاح في العينة $\mu_S = \mu_P = p$ حيث أن p هي النسبة في المجتمع. n تمثل حجم العينة، $\sigma_S = \sigma_P = \sqrt{\frac{pq}{n}}$ حيث $q = 1 - p$ والمتغير المعياري

$$Z = \frac{P - p}{\sqrt{pq/n}} \quad (5)$$

وفي حالة $P = X/n$ تمثل النسبة في العينة، X عدد مرات النجاح الفعلي في العينة. فإن العلاقة (5) تصبح

$$Z = \frac{X - np}{\sqrt{npq}} \quad (6)$$

وبالمثل سوف نقوم بعمل الاختبار في اتجاه واحد أو اتجاهين كما تم في حالة الوسط الحسابي.

3. **الفرق بين متوسطين Differences of Means:** افترض أن \bar{X}_1 ، \bar{X}_2 يمثلان الأوساط الحسابية في العينات الكبيرة n_1 ، n_2 المسحوبة من مجتمعين على الترتيب لهما المتوسطات μ_1 ، μ_2 بالانحرافات المعيارية σ_1 ، σ_2 . واعتبر الفرض العدمي H_0 أنه لا يوجد اختلاف بين متوسطي المجتمعين؛ بمعنى أن $H_0: \mu_1 = \mu_2$ وكما سبق أن أثبتنا في الباب السادس أن توزيع المعاينة للفرق بين متوسط مجتمعين له توزيع طبيعي تقاربي وأن الانحراف المعياري هو

$$\mu_{\bar{X}_1 - \bar{X}_2} = 0 \quad \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (7)$$

حيث يمكن استخدام الانحرافات المعيارية المشاهدة إذا احتجنا إلى ذلك s_1 ، s_2 (أو \hat{s}_1 و \hat{s}_2) كتقديرات للانحرافات المعيارية σ_1 ، σ_2 . ويكون المتغير المعتاد المعياري كالتالي

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}} \quad (8)$$

وكما سبق في الجزء (1) السابق، فإنه يمكن أن نختبر الفرض العدمي ضد الفرض البديل (أو معنوية الفرق المشاهد) عند مستوى المعنوية المناسب.

4. **الفرق بين النسب Differences of Proportions:** افترض أن P_1 ، P_2 تمثل النسب المشاهدة في العينات الكبيرة ذات الأحجام n_1 ، n_2 على الترتيب والمسحوبة من المجتمعات والتي بها النسب هي

P_1, P_2 . اعتبر الفرض العدمى هو أنه لا يوجد اختلاف بين نسب المجتمعات؛ أى أن $p_1 = p_2$. وبالتالي فإن العينات تعتبر مسحوبة من مجتمع واحد.

ومن دراستنا للفرق بين نسبتين فى الباب السادس سوف نضع $p_1 = p_2 = p$ وقد وجدنا أن الفرق بين نسبتين له توزيع طبيعى تقاربى بالمتوسط والانحراف المعيارى

$$\mu_{P_1-P_2} = 0 \quad \sigma_{P_1-P_2} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \quad (9)$$

حيث أن: $\bar{p} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$ تستخدم لتقدير النسبة المشتركة p .

وباستخدام المتغير المعيارى

$$Z = \frac{P_1 - P_2 - 0}{\sigma_{P_1-P_2}} = \frac{P_1 - P_2}{\sigma_{P_1-P_2}} \quad (10)$$

ونستطيع ملاحظة الفروق عند أى مستوى معنوية مناسب وبالتالى نطبق اختبار الفرض العدمى.

ويمكن بالمثل تصميم اختبارات مماثلة للإحصاءات الأخرى.

العلاقة بين نظرية التقدير واختبارات الفروض

Relationship between Estimation Theory and Hypothesis Testing

من الملاحظات السابقة نجد وجود علاقة بين نظرية التقدير بواسطة فترات الثقة ونظرية اختبارات الفروض. وكمثال على ذلك نلاحظ أن النتيجة (2) بقبول الفرض H_0 بمستوى معنوية 0.05 يكون معادلاً للنتيجة (1) فى الباب السابع الذى يقودنا إلى فترة الثقة 95%.

$$\bar{x} - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{1.96\sigma}{\sqrt{n}} \quad (11)$$

وبالتالى فإنه على الأقل فى الاختبارات فى اتجاهين فإننا نقوم فعلاً بتطبيق فترات الثقة فى الباب السابع لعمل اختبارات الفروض. وبالمثل نصل إلى نفس النتيجة لاختبارات الجانب الواحد، حيث تتطلب إعداد فترات ثقة فى اتجاه واحد.

مثال 8.2: اعتبر المثال (8.1). فإن فترة الثقة 95% للمثال (8.1) هى

Example 8.2. Consider Example 8.1. A 95% confidence interval for Example 8.1 is the following

$$1570 - \frac{(1.96)(120)}{\sqrt{100}} \leq \mu \leq 1570 + \frac{(1.96)(120)}{\sqrt{100}}$$

حيث أن

$$1570 - 23.52 \leq \mu \leq 1570 + 23.52$$

وهذا يقودنا إلى فترة ثقة بين 1546.48 و 1593.52 بدرجة ثقة 95%. وواضح أنه فى حدود هذه الفترة لا يشمل المتوسط الذى نختبره وهو 1600 وهذا يقودنا إلى رفض الفرض العدمى $H_0: \mu = 1600$.

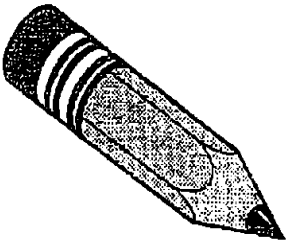
الفصل التاسع

تمهيد المنحنيات والانحدار والارتباط Curve Fitting, Regression, and Correlation

في هذا الفصل:

- ✓ تمهيد المنحنيات
- ✓ الانحدار
- ✓ طريقة المربعات الصغرى
- ✓ خط المربعات الصغرى
- ✓ خط انحدار المربعات الصغرى بدلالة تباينات وتغاير العينة
- ✓ الخطأ المعياري للتقدير
- ✓ معامل الارتباط الخطى
- ✓ معامل الارتباط العام
- ✓ الارتباط وعدم الاستقلال

Curve Fitting



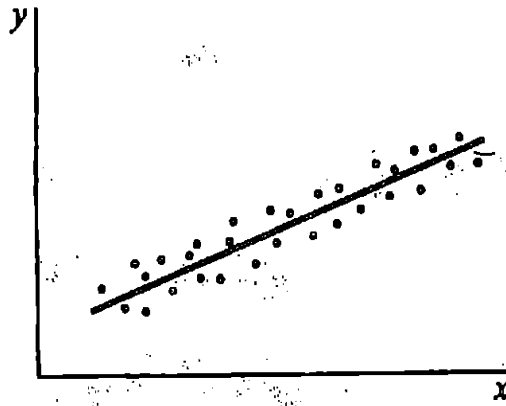
تمهيد المنحنيات

كثيراً جداً ما توجد في التطبيقات العملية علاقة بين متغيرين أو أكثر ونريد التعبير عنها في صيغة رياضية بتحديد معادلة تربط هذه المتغيرات.

وأول خطوة هي جمع بيانات للقيم المتقابلة لهذين المتغيرين. وكمثال نفترض أن x ، y ترمزان إلى طول ووزن الإنسان البالغ على الترتيب. وبالتالي فإن عينة من n فرد تبين لنا الأطوال x_1, x_2, \dots, x_n والأوزان المقابلة y_1, y_2, \dots, y_n

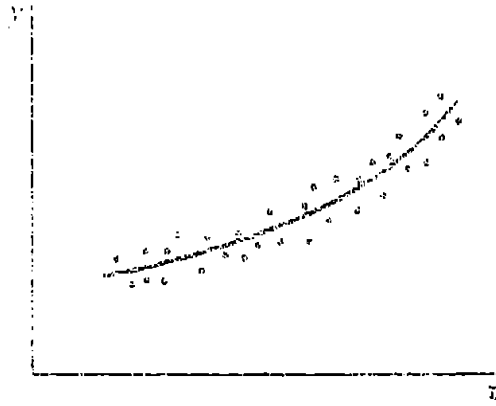
والخطوة التالية هي رسم هذه النقط $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ على رسم بياني. ويسمى أحياناً شكل هذه النقط على الرسم بشكل الانتشار .Scatter Diagram

ولكننا في الغالب نصور شكل منحنى البيانات تقريباً من شكل الانتشار. ويسمى هذا الخط بالمنحنى التقريبي Approximating Curve. وفي شكل (9-1) كمثال تظهر البيانات علاقة بين المتغيرات. تمثل تقريباً بصورة جيدة في خط مستقيم. وبالتالي نقول أنه توجد علاقة خطية بين المتغيرات Linear Relationship.

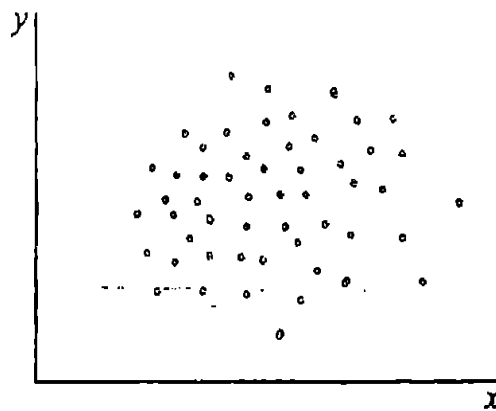


شكل 9-1

أما في شكل (9-2) يتضح أنه توجد علاقة أيضاً بين المتغيرات ولكنها ليست علاقة خطية ونسميها علاقة غير خطية Nonlinear Relationship بين المتغيرات. أما في شكل (9-3) يظهر شكل الانتشار عدم وجود علاقة على الإطلاق بين المتغيرات.



الشكل 9-2



الشكل 9-3

والمشكلة العامة لإيجاد المعادلات التي تعطي المنحنيات التقريبية لمجموعة من البيانات تسمى بتمهيد المنحنيات Curve Fitting. وفي التطبيق يظهر نوع المعادلة من شكل الانتشار. ففي الشكل (9-1) يمكننا استخدام معادلة الخط المستقيم:

$$y = a + bx$$

بينما في الشكل (9-2) يمكننا محاولة استخدام منحنى الدرجة الثانية أو القطع المكافئ بالمعادلة:

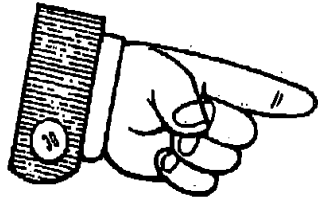
$$y = a + bx + cx^2$$

وتحقيقاً للغرض من هذا الكتاب سوف نقتصر على البيانات التي تعطي علاقة خطية فقط.

وأحياناً يساعدنا رسم شكل الانتشار بعد تحويل المتغيرات Transformed Variables. وكمثال على ذلك المتغيرات $\log x$, $\log y$ تقودنا إلى عمل خط مستقيم فإنه يمكن محاولة اعتبار أن $\log y = a + bx$ تمثل معادلة للمنحنى التقاربي.

Regression

الانحدار



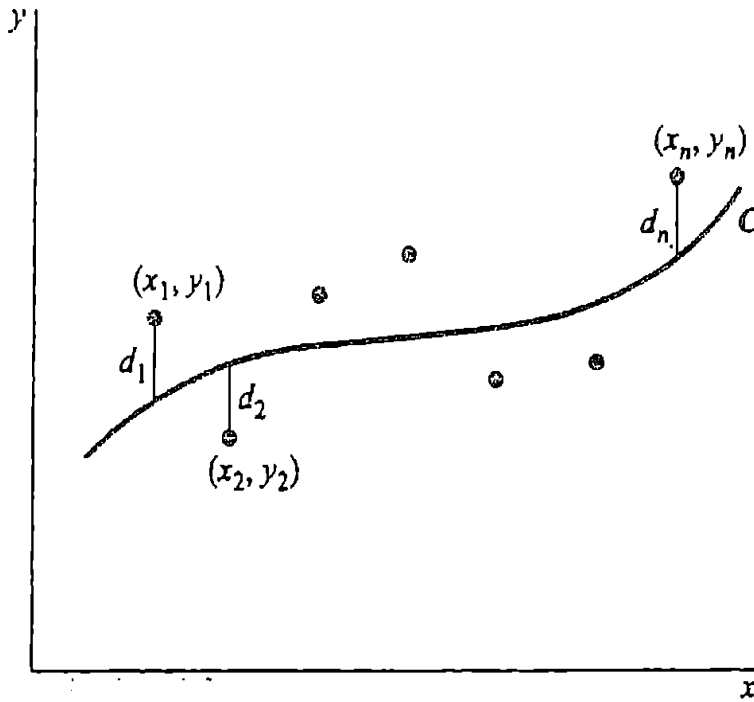
أحد أهم أهداف تمهيد المنحنيات هو تقدير أحد المتغيرات (المتغير التابع) من المتغير الآخر (المتغير المستقل). ونسعى عملية التقدير هذه بالانحدار Regression. فإذا أردنا تقدير y باستخدام x في معادلة، فإننا نسمى هذه المعادلة بمعادلة انحدار y على x ، (A Regression Equation of y on x) والمنحنى التابع يسمى بمنحنى انحدار y على x . (A Regression Curve of y on x). وحيث أننا نستخدم فقط الحالة الخطية فإننا نقول أنه خط انحدار y على x ، (A Regression Line of y on x).

طريقة المربعات الصغرى The Method of Least Squares

عموماً يوجد أكثر من منحنى يمكن تمهيدته لمجموعة من البيانات. ولكي نتحاشى الحكم الشخصي في وجود أكثر من خط أو أكثر من منحنى تقريبي لتمثيل هذه البيانات فإنه من الضروري لتعريف «أفضل خط ممهد» "Best-Fitting Line" أو «أفضل منحنى ممهد» "Best-Fitting Parabola" وهكذا.

ولكي نصل للتعريف المطلوب، نفترض الشكل (4-9) والذي تم وضع النقط $(x_1, y_1), \dots, (x_n, y_n)$ عليه. وبالنسبة للقيمة (x_1, y_1) فإنه يوجد اختلافات بين y_1 والقيمة التابعة على الخط C . وسوف نرمز لهذا الاختلاف بالرمز d_1 والذي يسمى عادة بخطأ الانحراف Deviation Error أو

الباقى Residual ويكون موجباً أو سالباً أو مساوياً للصفر. وبالمثل من القيم التابعة والمقابلة لـ x_1, x_2, \dots, x_n نحصل على الانحرافات d_1, d_2, \dots, d_n



شكل 9-4

ومقياس تمهيد الخط C لمجموعة من البيانات يرتبط بالكمية $d_1^2 + d_2^2 + \dots + d_n^2$ ويكون التمهيد جيداً إذا كان هذا المقدار صغيراً. أما إذا كان هذا المقدار كبيراً فإن التمهيد يكون رديئاً. وبالتالي نقوم بعمل التعريف التالى:

تعريف: من بين كل المنحنيات التى تمثل مجموعة من النقط تقاربياً، يوجد منحنى واحد يمثل أحسن تمهيد Best Fitting للمجموعة وهو عندما تكون:

$$d_1^2 + d_2^2 + \dots + d_n^2 = \text{نهاية صغرى}$$

والمنحنى الذى له هذه الخاصية يقال أنه يمهد البيانات بإحساس ومفهوم المربعات الصغرى ويسمى بمنحنى انحدار المربعات الصغرى Least-Squares Regression Curve أو ببساطة منحنى المربعات الصغرى.

والخط الذي له هذه الخاصية يسمى بخط المربعات الصغرى Least Squares Line. والقطع المكافئ الذي له هذه الخاصية يسمى بقطع المربعات الصغرى وهكذا Least-Squares Parabola .. إلخ.

وعادة يتم توظيف المفهوم الجديد عندما x تمثل المتغير المستقل Independent Variable وتمثل y المتغير التابع Dependent Variable. أما إذا اعتبرنا x هي المتغير التابع، فإن التعريف سوف يعدل بأخذ الانحرافات الأفقية بدلاً من الانحرافات الرأسية. وهذان التعريفان يقودانا إلى منحنيين مختلفين للمربعات الصغرى. فالأول به y هي المتغير التابع والثاني به x هي المتغير التابع. ولكننا سوف نعتبر دائماً y هي المتغير التابع، x هي المتغير المستقل. إلا إذا نص على غير ذلك.

✓ يجب أن تعرف

أنه من الممكن تعريف منحنى مربعات صغرى آخر باعتبار المسافات المتعامدة من نقط البيانات إلى المنحنى بدلاً من المسافات الرأسية أو الأفقية. وعلى أي حال لا تستخدم هذه الطريقة غالباً.

خط المربعات الصغرى The Least-Squares Line

باستخدام التعريف السابق يمكننا أن نبين أن خط المربعات الصغرى التقريبي لمجموعة من النقط $(x_1, y_1), \dots, (x_n, y_n)$ يكون له المعادلة

$$y = a + bx \quad (1)$$

حيث أنه يمكن إيجاد قيمة الثوابت a, b على المعادلات الآتية حلاً
آنياً

$$\sum y = an + b \sum x$$

$$\sum xy = a \sum x + b \sum x^2 \quad (2)$$

والتي تسمى بالمعادلات الطبيعية Normal Equations لخط المربعات الصغرى. مع ملاحظة أنه تم اختصار $\sum y$ ، $\sum xy$ بدلاً من $\sum_{j=1}^n y_j$ ، $\sum_{j=1}^n x_j y_j$.

من السهل تذكر المعادلة الطبيعية (2) بملاحظة أن المعادلة الأولى تم إيجادها بجمع المعادلة (1) على جميع النقط، أما المعادلة الثانية فتم إيجادها بضرب طرفي المعادلة (1) في x ثم إجراء عملية الجمع بعد ذلك، وبالطبع ليس هذا باشتقاق المعادلات الطبيعية ولكن فقط حتى يمكن تذكرها.

وتكون قيمة كل من الثوابت a ، b كالتالي

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n \sum x^2 - (\sum x)^2} \quad , \quad b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} \quad (3)$$

كما أنه يمكن كتابة قيمة ما على الشكل

$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (4)$$

حيث أن \bar{x} ، \bar{y} هي الأوساط الحسابية لكل من x ، y . كما أن $\bar{x} = (\sum x) / n$ وبقسمة طرفي العلاقة الطبيعية الأولى في (2) على n نحصل على

$$\bar{y} = a + b\bar{x} \quad (5)$$

وبالتالي فإننا نوجد قيمة b من العلاقة (3) أو (4) أولاً ثم استخدام العلاقة (5) لإيجاد قيمة a . حيث أن $a = \bar{y} - b\bar{x}$.

وذلك يكون مساوياً لكتابة خط المربعات الصغرى كالتالى

$$y - \bar{y} = b(x - \bar{x}) \quad \text{or} \quad y - \bar{y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} (x - \bar{x}) \quad (6)$$

والعلاقة (6) تبين أن الثابت b يمثل ميل الخط (1) The Slope of the Line. وهو الثابت الأساسى لتحديد الخط. ومن (6) يتضح لنا أيضاً أن خط المربعات الصغرى يمر من خلال النقطة (\bar{x}, \bar{y}) والتي تسمى المركز The Centroid أو مركز ثقل البيانات The Center of Gravity of the Data. والميل لخط الانحدار يكون مستقلاً عن نقطة الأصل للمحاور. وهذا يعنى أنه إذا تم عمل التحويل (فى الغالب يسمى بعملية نقل المحاور) كالتالى

$$x = x' + h \quad y = y' + k \quad (7)$$

حيث أن h, k تمثل ثوابت. فإن قيمة b تكون

$$b = \frac{n \sum x'y' - (\sum x')(\sum y')}{n \sum x'^2 - (\sum x')^2} = \frac{\sum (x - \bar{x}')(y - \bar{y}')}{\sum (x' - \bar{x}')^2} \quad (8)$$

حيث أن x, y تم تغييرهما بـ x', y' ولهذا فإن قيمة b لا تتغير بالتحويلة (7). ونلاحظ أن a والتي تحدد الجزء المقطوع من المحور على y والتي تعتمد على نقطة الأصل وبالتالي فإنها سوف تتغير. وفى الحالة الخاصة $h = \bar{x}, k = \bar{y}$ نجد أن (8) تصبح

$$b = \frac{\sum x'y'}{\sum x'^2} \quad (9)$$

وغالباً ما نستخدم العلاقات (8)، (9) لتبسيط العمليات الحسابية فى حالة الحصول على خط المربعات الصغرى.

والملاحظة السابقة تنطبق أيضاً على خط انحدار x على y . والنتائج نحصل عليها في التعبير في x ، y . وعلى ذلك فإن خط انحدار المربعات الصغرى x على y هو

$$x - \bar{x} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (y - \bar{y})^2} (y - \bar{y}) \quad (10)$$

وتلاحظ أن العلاقة (10) تختلف عن العلاقة (6).



تذكر!

أنك يجب أن تحاول إيجاد معادلة خط الانحدار إذا فقط إذا كانت مجموعة البيانات لها علاقة خطية.

مثال 9.1: الجدول 9-1 يعطى الأطوال x ، y لعينة من 12 أباً وأكبر الأبناء. فأوجد خط انحدار المربعات الصغرى.

Example 9.1. Table 9-1 shows the respective heights x and y of a sample of 12 fathers and their oldest sons. Find the least-squares regression line of y on x .

جدول 9-1

أطوال الأب x بالبوصة	65	63	67	64	68	62
أطوال الابن y بالبوصة	68	66	68	65	69	66

أطوال الأب x بالبوصة	70	66	68	67	69	71
أطوال الابن y بالبوصة	68	65	71	67	68	70

خط الانحدار y على x هو $y = ax + b$ ويمكن إيجاده بحل المعادلات الطبيعية

$$\sum y = an + b \sum x \quad \text{and} \quad \sum xy = a \sum x + b \sum x^2$$

ونقوم بإيجاد المجاميع كالتالى:

جدول 9-2

x	y	x ²	xy	y ²
65	68	4225	4420	4624
63	66	3969	4158	4356
67	68	4489	4556	4624
64	65	4096	4160	4225
68	69	4624	4692	4761
62	66	3844	4092	4356
70	68	4900	4760	4624
66	65	4356	4290	4225
68	71	4624	4828	5041
67	67	4489	4489	4489
69	68	4761	4692	4624
71	70	5041	4970	4900
$\sum x = 800$	$\sum y = 811$	$\sum x^2 = 53,418$	$\sum xy = 54,107$	$\sum y^2 = 54,849$

وباستخدام هذه المجاميع نحصل على المعادلات الطبيعية

$12a + 800b$	$=$	811
$800a + 53,418b$	$=$	$54,107$

ويحل هاتين المعادلتين نحصل على $a = 35.82$ ، $b = 0.476$ وبالتالى تكون معادلة انحدار المربعات الصغرى y على x هي $y = 35.82 + 0.476x$

خط انحدار المربعات الصغرى بدلالة تباينات وتغاير العينة

The Least-Squares Regression Line in Terms of Sample Variances and Covariance

تباينات وتغاير العينة للمتغيرات x ، y هي

$$s_x^2 = \frac{\sum (x - \bar{x})^2}{n}, \quad s_y^2 = \frac{\sum (y - \bar{y})^2}{n}, \quad s_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} \quad (II)$$

ويمكن كتابة خطوط انحدار المربعات الصغرى y على x وأيضاً x على y بدلالة التباينات والتغاير كما يلي على الترتيب

$$y - \bar{y} = \frac{s_{xy}}{s_x^2} (x - \bar{x}) \quad \text{and} \quad x - \bar{x} = \frac{s_{xy}}{s_y^2} (y - \bar{y}) \quad (12)$$

كما أن معامل ارتباط العينة The Sample Correlation Coefficient هو

$$r = \frac{s_{xy}}{s_x s_y} \quad (13)$$

وبالتالى فإن (12) يمكن أن تكتب كالتالى

$$\frac{y - \bar{y}}{s_y} = r \frac{(x - \bar{x})}{s_x} \quad \text{and} \quad \frac{x - \bar{x}}{s_x} = r \frac{(y - \bar{y})}{s_y} \quad (14)$$

وبالنظر إلى حقيقة أن $\frac{x - \bar{x}}{s_x}$ ، $\frac{y - \bar{y}}{s_y}$ تمثل القيمة المعيارية للعينة أو

الدرجات المعيارية. فإن النتائج (14) تمثل أبسط طريقة لإيجاد خطوط الانحدار. ومن الواضح أن خطى الانحدار فى (14) مختلفين ماعدا عندما $r = \pm 1$ حيث فى هذه الحالة نجد أن فقط العينة تقع كلها على الخط والتي تمثل الارتباط والانحدار المستقيم التام Perfect Linear Correlation and Regression.

ومن المهم أن نذكر أنه إذا كان خطى الانحدار (14) $x = c + dy$ $y = ax + b$ على الترتيب. فإن

$$bd = r^2 \quad (15)$$

وحتى الآن لم نحدد أو ندرس معنوية معامل الارتباط ولكننا قد قمنا فقط بتعريفه بدلالة التباينات والتغاير فقط.

الخطأ المعياري للتقدير Standard Error of Estimate

إذا اعتبرنا أن y_{est} ترمز إلى القيمة المقدرة للمتغير y بالنسبة إلى قيمة معينة x والتي نحصل عليها من منحنى الانحدار y على x . وبالتالي فإن مقياس التشتت حول خط الانحدار نحصل عليه من الكمية

$$s_{y,x} = \sqrt{\frac{\sum (y - y_{est})^2}{n}} \quad (16)$$

والتي تسمى بالخطأ المعياري المقدر y على x The Standard Error of x Estimate y on x . وكما ذكرنا من قبل أن $\sum (y - y_{est})^2 = \sum d^2$ إنه من بين كل خطوط الانحدار الممكنة فإن خط انحدار المربعات الصغرى تعطي أقل تقدير للخطأ المعياري.

وفي حالة خط الانحدار $y_{est} = a + bx$ بحيث أن a, b يمكن إيجادهما من العلاقات (2). ونحصل على

$$s_{y,x}^2 = \frac{\sum y^2 - a \sum y - b \sum xy}{n} \quad (17)$$

أو

$$s_{y,x}^2 = \frac{\sum (y - \bar{y})^2 - b \sum (x - \bar{x})(y - \bar{y})}{n} \quad (18)$$

ويمكننا أيضاً التعبير عن $s_{y,x}^2$ لخط انحدار المربعات الصغرى بدلالة التباين ومعامل الارتباط

$$s_{y,x}^2 = s_y^2 (1 - r^2) \quad (19)$$

والذي نستنتج منه أن $r^2 \leq 1$ أي أن $-1 \leq r \leq 1$. والخطأ المعياري للتقدير له خصائص مرادفة لخصائص الانحراف المعياري. وكمثال لذلك إذا تم إيجاد زوج من الخطوط موازية لخط

انحدار y على x على مسافات عمودية $s_{y,x}$ ، $2s_{y,x}$ ، $3s_{y,x}$ على الترتيب فإننا نجد إذا كانت n كبيرة بما فيه الكفاية فإننا نجد أنه بين هذه الأزواج من الخطوط حوالي 68% ، 95% ، 99.7% من نقط العينة على الترتيب.

وحيث أنه يوجد تقدير غير متحيز لتباين المجتمع بالعلاقة $s^2 = ns^2/(n-1)$ وبالتالي فإنه يوجد تقدير غير متحيز لمربع الخطأ المعياري المقدر. ونحصل عليه باستخدام $s_{y,x}^2 = ns_{y,x}^2/(n-2)$ ولهذا السبب يفضل بعض الإحصائيين استخدام العلاقة (16) بالقسمة على $(n-2)$ بدلاً من n .

ويمكن تعديل الملاحظات السابقة بسهولة لتنطبق على خط انحدار x على y أو على الانحدار غير الخطي أو الانحدار المتعدد (ودائماً نستخدم $s_{y,x}$ ليرمز إلى تقدير الخطأ المعياري).

معامل الارتباط الخطي

The Linear Correlation Coefficient

لقد قمنا حتى الآن بتعريف معامل الارتباط (13) بدون أن تختبر معنويته. ومحاولة لإيجاد ذلك تذكر الشكل (19) والتعريفات s_y ، $s_{y,x}$ نجد أن

$$r^2 = 1 - \frac{\sum (y - y_{est})^2}{\sum (y - \bar{y})^2} \quad (20)$$

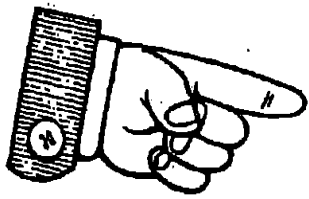
ويمكننا أن نبين أن

$$\sum (y - \bar{y})^2 = \sum (y - y_{est})^2 + \sum (y_{est} - \bar{y})^2 \quad (21)$$

والمقدار الموجود في الجانب الأيسر للمعادلة (21) تسمى بالتغيرات الكلية Total Variation. أما المقدار الأول في الجانب الأيمن يسمى بالمتغيرات غير المفسرة Unexplained Variation. ويسمى المقدار الثاني

بالتغيرات المشروحة أو المفسرة Explained Variation. وهذه التسمية لهذه الانحرافات $(y - y_{est})$ نشأت لأنها تحدث بطريقة عشوائية أو لا يمكن التنبؤ بها بينما تسمى الانحرافات $(y_{est} - \bar{y})$ بالمشروحة أو المفسرة بواسطة خط انحدار المربعات الصغرى والتي تؤول إلى نمط محدد. ومن المعادلة (20)، (21) نجد أن

$$r^2 = \frac{\sum (y_{est} - \bar{y})^2}{\sum (y - \bar{y})^2} = \frac{\text{الانحرافات المشروحة}}{\text{الانحرافات الكلية}} \quad (22)$$



ولذلك فإن r^2 يمكن أن يفسر على أنه نسبة الانحرافات الكلية المفسرة بواسطة خط انحدار المربعات الصغرى. وبلغه أخرى r تقيس كيفية تمثيل خط انحدار المربعات الصغرى لبيانات العينة،

فإذا كان التباين الكلي تم تفسيره كلياً بواسطة خط الانحدار؛ أى أن $r^2 = 1$ أو $r = \pm 1$ فإننا نقول أنه يوجد ارتباط تام كلى Perfect Linear Correlation (أى أنه فى هذه الحالة يوجد انحدار خطى تام Perfect Linear Regression). ومن ناحية أخرى إذا كان كل الاختلاف لم يفسر فإن الاختلاف المفسر يكون مساوياً للصفر وبالتالي فإن $r = 0$. وفى التطبيق فإن الكمية r^2 تسمى معامل التحديد Coefficient of Determination ودائماً تقع بين الصفر والواحد الصحيح.

وبالتالى يمكننا حساب معامل الارتباط من أحد هذه النتائج.

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}} \quad (23)$$

أو

$$r^2 = \frac{\sum (y_{est} - \bar{y})^2}{\sum (y - \bar{y})^2} = \frac{\text{الانحرافات المفسرة}}{\text{الانحرافات الكلية}} \quad (24)$$

والتي تكون دائماً متساوية في حالة الانحدار الخطي. والصيغة (23) تسمى غالباً بشكل العزوم للانحدار الخطي Product-Moment Formula وتوجد أشكال أخرى متكافئة بالاستخدام العملي مثل

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (25)$$

وأيضاً

$$r = \frac{\bar{xy} - \bar{x}\bar{y}}{\sqrt{(x^2 - \bar{x}^2)(y^2 - \bar{y}^2)}} \quad (26)$$

وإذا تم عمل تحويلة (7) فإننا نحصل على

$$r = \frac{n \sum x'y' - (\sum x')(\sum y')}{\sqrt{[n \sum x'^2 - (\sum x')^2][n \sum y'^2 - (\sum y')^2]}} \quad (27)$$

والتي تبين أن r لا تتأثر بأي تغيير في المحاور. وفي حالة خاصة، عندما تضع $h = \bar{x}$ ، $k = \bar{y}$ فإن (27) تصبح

$$r = \frac{\sum x'y'}{\sqrt{(\sum x'^2)(\sum y'^2)}} \quad (28)$$

والتي تكون غالباً مفيدة في الحساب.

معامل الارتباط الخطي قد يكون إما سالباً أو موجباً. فإذا كانت r موجبة، فإن y تزيد بزيادة x (يكون ميل خط انحدار المربعات الصغرى موجباً) بينما إذا كانت r سالبة فإن y تنقص بزيادة x (يكون الميل سالباً). ودائماً نأخذ إشارة معامل الارتباط في الاعتبار إذا استخدمنا المعادلات (23)، (25)، (26)، (27) أو (28) بينما إذا تم استخدام (24) لحساب r فإننا يجب أن نستخدم الإشارة الصحيحة.

معامل الارتباط العام

Generalized Correlation Coefficient

التعريف (23) (أو التعريفات المتماثلة له (25) حتى (28)) لمعامل الارتباط يتضمن قيم المعاينة فقط x, y . وبالتالي فإنه يؤدي إلى نفس الرقم لكل أشكال منحنيات الانحدار ولكنه عديم الفائدة لقياس درجة التوفيق ما عدا حالة الانحدار المستقيم حيث ينطبق مع العلاقة (24). وهذا التعريف يعطى:

$$r^2 = \frac{\sum (y_{est} - \bar{y})^2}{\sum (y - \bar{y})^2} = \frac{\text{التباين المفسر}}{\text{التباين الكلي}} \quad (29)$$

يعكس شكل منحنى الانحدار (من خلال y_{est}) وبالتالي يكون مناسباً كتعريف لمعامل الارتباط العام r . Generalized Correlation Coefficient ونستخدم (29) للحصول على معاملات الارتباط غير الخطية (حيث يقيس جودة توفيق منحنى الانحدار غير الخطى Nonlinear Regression Curve للبيانات). العلاقة (19) بين معامل الارتباط وتقدير الخطأ المعياري قائمة أيضاً بالمثل في حالة الارتباط غير الخطى.

مثال 9.2: أوجد معامل التحديد ومعامل الارتباط في المثال (8.2).

Example 9.2. Find the coefficient of determination and the coefficient of correlation from Example 8.2.

معامل التحديد r^2 حيث

$$r^2 = \frac{\text{التباين المشروح}}{\text{التباين الكلي}} = \frac{19.22}{38.92} = 0.4938$$

ويكون معامل الارتباط r هو

$$r = \pm \sqrt{0.4938} = \pm 0.7027$$

وحيث أن r_{xy} تزيد بزيادة x (أي أن ميل خط الانحدار موجب) فإن الارتباط يكون موجباً وبالتالي فإن $r = 0.7027$ أو $r = 0.70$ إلى رقمين معنويين.

وحيث أن معامل الارتباط يقيس بصفة عامة كيف أن منحنى الانحدار (أو السطح) يوائم بيانات العينة، وبالتالي فإنه يكون من غير المنطقي استخدام معامل الارتباط الخطي حينما تكون البيانات غير خطية. افترض أنه تم تطبيق العلاقة (23) على بيانات غير خطية وحصل على قيمة رقمية تعتبر أقل من واحد بكثير. وبالتالي فإنه لا يستنتج عدم وجود ارتباط ولكن يستنتج بوجود ارتباط خطي ضعيف Little Linear Correlation. (ويصل إلى هذا الاستنتاج أحياناً هؤلاء غير العارفين بالأسس الخاصة بنظرية الارتباط). ولكنه في الحقيقة ربما يكون هناك ارتباط غير خطي كبير Large Nonlinear Correlation.

الارتباط وعدم الاستقلال Correlation and Dependence

حينما يوجد معامل ارتباط غير مساوي للصفر بين متغيرين X ، Y فإنه يوجد عدم استقلال Dependence بالمعنى الاحتمالي بالإضافة إلى ذلك فإننا نستخدم العلاقة (6) للتنبؤ بقيمة المتغير التابع Y من قيمة المتغير المستقل X .

✓ يجب أن تعرف

إنه من الضروري أن تدرك أن "الارتباط Correlation" و "عدم الاستقلال Dependence" بالمعنى السابق ليس من الضروري أن يدل ضمناً على علاقة سببية مباشرة للاعتماد المتبادل بين X ، Y .

مثال 9.3: إذا كانت X تمثل مرتبات المدرسين في سنة بينما تمثل Y حجم الجريمة، فإن معامل الارتباط قد يختلف عن الصفر وقد يمكننا إيجاد خط انحدار للتنبؤ بقيمة متغير من الآخر. ولكنه من الصعب أن نقول أنه توجد علاقة تبادلية مباشرة بين X ، Y .

Example 9.3. If X represents teachers' salaries over the years while Y represents the amount of crime, the correlation coefficient may be different from zero and we may be able to find a regression line predicting one variable from the other. But we would hardly be willing to say that there is a direct interdependence between X and Y .

الفصل العاشر

توزيعات احتمالية أخرى

Other Probability Distributions

في هذا الفصل:

- ✓ التوزيع المتعدد
- ✓ التوزيع الهايبرجيومتري
- ✓ التوزيع المنتظم
- ✓ توزيع كوشي
- ✓ توزيع جاما
- ✓ توزيع بيتا
- ✓ توزيع كاي تربيع χ^2
- ✓ توزيع t ستيودنت
- ✓ توزيع F
- ✓ العلاقات بين توزيعات كاي تربيع، t و F

التوزيع المتعدد The Multinomial Distribution

نفترض أننا لدينا الأحداث المتنافية بالتبادل A_1, A_2, \dots, A_k بالاحتمالات المقابلة p_1, p_2, \dots, p_k بحيث أن $p_1 + p_2 + \dots + p_k = 1$ فإذا

كانت X_1, X_2, \dots, X_k تمثل متغيرات عشوائية تعطى على الترتيب عدد مرات حدوث الأحداث A_1, A_2, \dots, A_k من المحاولات الكلية التي عددها n . كما أن $X_1 + X_2 + \dots + X_k = n$ فإن

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \quad (1)$$

حيث أن $n_1 + n_2 + \dots + n_k = n$ تمثل دالة الاحتمال المشتركة للمتغيرات العشوائية X_1, X_2, \dots, X_k ويسمى هذا التوزيع والذي يمثل تعميم للتوزيع ذي الحدين بالتوزيع المتعدد Multinomial Distribution حيث أن العلاقة السابقة تمثل الحد العام في المفكوك المتعدد Multinomial Expansion $(p_1 + p_2 + \dots + p_k)^n$.

التوزيع الهايبرجيومتري

The Hypergeometric Distribution

نفترض أننا لدينا صندوق به b كرة زرقاء، r كرة حمراء. ودعنا نقوم بعمل n محاولة من تجربة سحب كرة بطريقة عشوائية من الصندوق ومعرفة لونها ثم إعادتها إلى الصندوق. وهذا النوع من التجارب يشير غالباً إلى المعاينة مع الإعادة Sampling With Replacement. فإذا كانت X تمثل في هذه الحالة متغيراً عشوائياً يرمز إلى عدد الكرات الزرقاء المختارة (حالات النجاح) في n محاولة، فباستخدام توزيع ذا الحدين، فإن احتمال الحصول على x حالة نجاح هي

$$P(X = x) = \binom{n}{x} \frac{b^x r^{n-x}}{(b+r)^n}, \quad x = 0, 1, \dots, n \quad (2)$$

حيث أن: $q = 1 - p = \frac{r}{(b+r)}$ ، $p = \frac{b}{(b+r)}$

وإذا تم عمل تعديل على ما سبق بجعل السحب بدون إعادة Sampling

Without Replacement أي أنه لا يتم إعادة الكرة المسحوبة قبل سحب الكرة التالية، فإن

$$P(X = x) = \frac{\binom{b}{x} \binom{r}{n-x}}{\binom{b+r}{n}}, \quad x = \max(0, n-r), \dots, \min(n, b) \quad (3)$$

ونسى هذه العلاقة بالتوزيع الهايبرجيومترى. ويكون متوسطة وتبايناً على الشكل التالى:

$$\mu = \frac{nb}{b+r}, \quad \sigma^2 = \frac{nbr(b+r-n)}{(b+r)^2(b+r-1)} \quad (4)$$

وإذا اعتبرنا مجموع عدد الكرات الزرقاء والحمراء هو N ، بينما نسبة الكرات الزرقاء هي p والحمراء هي $q = 1 - p$. فإن

$$p = \frac{b}{b+r} = \frac{b}{N}, \quad q = \frac{r}{b+r} = \frac{r}{N} \quad \text{or} \quad b = Np, \quad r = Nq$$

وهذا يقودنا إلى

$$P(X = x) = \frac{\binom{Np}{x} \binom{Nq}{n-x}}{\binom{N}{n}} \quad (5)$$

كما أن التوقع والتباين هو

$$\mu = np, \quad \sigma^2 = \frac{npq(N-n)}{N-1} \quad (6)$$

ونذكر أنه عندما $N \rightarrow \infty$ (إذا N تكون كبيرة بالمقارنة مع n). فإننا

نحصل على

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad (7)$$

$$\mu = np, \quad \sigma^2 = npq \quad (8)$$

وهو نفس متوسط وتباين توزيع ذي الحدين. وهذه النتائج المتوقعة. حيث أنه عندما تكبر N فإن المعاينة مع الإعادة تكون هي نفسها المعاينة بدون الإعادة.

مثال 10.1: يحتوى صندوق على 6 مرات زرقاء، 4 كرات حمراء. ثم إجراء تجربة على أساس سحب كرة ومشاهدة لونها ثم عدم إعادتها للصندوق. فأوجد بعد خمسة محاولات للتجربة، احتمال سحب ثلاث كرات زرقاء

Example 10.1 A box contains 6 blue marbles and 4 red marbles. An experiment is performed in which a marble is chosen at random and its color is observed, but the marble is not replaced. Find the probability that after 5 trials of the experiment, 3 blue marbles will have been chosen.

عدد الطرق المختلفة لاختيار 3 كرات زرقاء من 6 كرات هو $\binom{6}{3}$.

كما أن عدد الطرق المختلفة لاختيار العدد الباقي، كرتين من اللون الأحمر هو $\binom{4}{2}$. ولذلك فإن عدد العينات والتي يكون فيها 3 كرات

زرقاء و 2 كرة حمراء هو $\binom{6}{3}\binom{4}{2}$. كما أن عدد الطرق المختلفة

لاختيار 5 كرات من 10 كرات (6+4) هو $\binom{10}{5}$. ولذلك فإن الاحتمال

المطلوب هو

$$\frac{\binom{6}{3}\binom{4}{2}}{\binom{10}{5}} = \frac{10}{21}$$

The Uniform Distribution

التوزيع المنتظم

يكون للمتغير العشوائي X توزيعاً منتظماً في الفترة $a \leq x \leq b$ Uniformly Distributed إذا كانت دالته للكثافة هي

$$f(x) = \begin{cases} 1/(b-a) & a \leq x \leq b \\ 0 & \text{غير ذلك} \end{cases} \quad (9)$$

ويسمى التوزيع بالتوزيع المنتظم وتكون دالة التوزيع هي

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x < b \\ 1 & x \geq b \end{cases} \quad (10)$$

ويكون المتوسط والتباين كالتالي

$$\mu = \frac{1}{2}(a+b), \quad \sigma^2 = \frac{1}{12}(b-a)^2 \quad (11)$$

The Cauchy Distribution

توزيع كوشي

يكون للمتغير العشوائي X توزيع كوشي أو موزع كوشي إذا كانت دالة كثافة احتمال X كالتالي

$$f(x) = \frac{a}{\pi(x^2 + a^2)} \quad a > 0, -\infty < x < \infty \quad (12)$$

وتكون دالة الكثافة متماثلة حول $x=0$ وبالتالي يكون الوسيط مساوياً للصفر. كما أنه لا يوجد له متوسط حسابي وتباين.

The Gamma Distribution

توزيع جاما

يكون للمتغير العشوائي X توزيع جاما أو موزع جاما إذا كان له دالة الكثافة التالية

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (\alpha, \beta > 0) \quad (13)$$

حيث أن $\Gamma(\alpha)$ تمثل دالة جاما (انظر مرفق A) ويكون المتوسط والتباين كالتالي

$$\mu = \alpha\beta \quad \sigma^2 = \alpha\beta^2 \quad (14)$$

توزيع بيتا The Beta Distribution

المتغير العشوائي X له توزيع بيتا أو موزع بيتا إذا كان له كثافة الاحتمال الآتية

$$f(x) = \begin{cases} \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} & 0 < x < 1 \\ 0 & \text{غير ذلك} \end{cases} \quad (\alpha, \beta > 0) \quad (15)$$

حيث أن $B(\alpha, \beta)$ هي دالة بيتا (انظر مرفق A). وبالنظر إلى علاقة دوال بيتا بدوال جاما، فإن دالة بيتا يمكن أن تكتب على الصورة.

$$f(x) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{غير ذلك} \end{cases} \quad (16)$$

ولكل قيم α, β الموجبة. ويكون المتوسط والتباين هو

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \quad (17)$$

ويوجد منوال وحيد عندما تكون $\alpha > 1, \beta > 1$ كالتالي

$$X_{\text{mode}} = \frac{\alpha - 1}{\alpha + \beta - 2} \quad (18)$$

توزيع كاي تربيع χ^2 The Chi-Square Distribution

إذا كانت X_1, X_2, \dots, X_n تمثل متغيرات عشوائية معتادة مستقلة لكل منها التوقع صفر وتباين الوحدة. اعتبر المتغير.

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2 \quad (19)$$

وتسمى χ^2 Chi-Square بكاي تربيع. كما أنه لكل قيم $x \geq 0$,

$$P(\chi^2 \leq x) = \frac{1}{2^{v/2} \Gamma(v/2)} \int_0^x u^{(v/2)-1} e^{-u/2} du \quad (20)$$

وتكون $P(\chi^2 \leq x) = 0$ ، $x < 0$.

ويسمى هذا التوزيع توزيع كاي تربيع وتسمى v بعدد درجات الحرية Number of Degrees of Freedom. ويكون التوزيع السابق له دالة كثافة الاحتمال

$$f(x) = \begin{cases} \frac{1}{2^{v/2} \Gamma(v/2)} x^{(v/2)-1} e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (21)$$

وتوزيع كاي تربيع يمثل حالة خاصة من توزيع جاما عندما تكون

$$\alpha = \frac{v}{2} \text{ و } \beta = 2 \text{ . ولذلك فإن}$$

$$\mu = v, \quad \sigma^2 = 2v \quad (22)$$

وعندما تكون $v (v \geq 30)$ كبيرة فإنه يمكننا إثبات أن المقدار $\sqrt{2\chi^2} - \sqrt{2v-1}$ يوزع توزيعاً طبيعياً توقعه صفر وتباينه الوحدة؛ أي

توزيع طبيعي معياري، ونذكر ثلاث نظريات سوف تكون مفيدة وهي كما يلي:

نظرية 10-1: إذا كان لدينا المتغيرات العشوائية المستقلة والموزعة توزيعاً طبيعياً X_1, X_2, \dots, X_v توقعة صفر وتباينه الوحدة، فإن $\chi^2 = \chi_1^2 + \chi_2^2 + \dots + \chi_v^2$ يمثل كاي تربيع بدرجات حرية v .

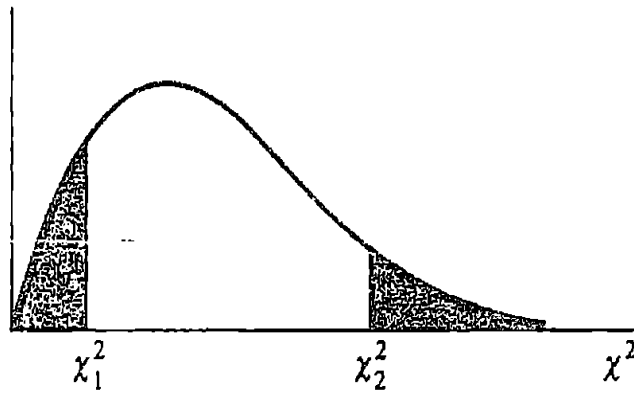
نظرية 10-2: إذا كان المتغيرات العشوائية المستقلة U_1, U_2, \dots, U_k لكل منها توزيع كاي تربيع بدرجات حرية v_1, v_2, \dots, v_k على الترتيب. فإن المجموع W حيث أن $W = U_1 + U_2 + \dots + U_k$ يكون له توزيع كاي تربيع بدرجات حرية $v_1 + v_2 + \dots + v_k$.

نظرية 10-3: إذا كان لدينا المتغيرات العشوائية المستقلة V_1, V_2 . فإذا افترضنا أن V_1 موزعة توزيع كاي تربيع بدرجات حرية v_1 . بينما المجموع $V = V_1 + V_2$ يكون موزعاً توزيع كاي تربيع بدرجات حرية v ، فإن V_2 يكون موزعاً توزيع كاي تربيع بدرجات حرية $v_2 = v - v_1$. (يكون $v > v_1$).

ويربط التوزيعات كاي تربيع وتوزيع t وتوزيع F وأي توزيعات أخرى فإنه من الشائع استخدام نفس الرمز Same Symbol لكل من المتغير والقيم التي يأخذها المتغير. ولذلك فإن قيم النسبة المئوية Percentile Values لتوزيع كاي تربيع لدرجات الحرية v يرمز لها بالرمز $\chi_{p,v}^2$ أو اختصاراً χ_p^2 إذا كانت v مفهومة وليس بالرمز $\chi_{p,v}^2$ أو χ_p (انظر الملحق D). وعلى القارئ أن يكون حريصاً خاصة عند تبديل المتغيرات في دوال كثافة الاحتمال.

مثال 10.2: الرسم البياني لتوزيع كاي تربيع بدرجات حرية 5 في الشكل (10-1). أوجد القيم χ_1^2 و χ_2^2 التي عندها يكون الجزء المظلل في الطرف الأيمن = 0.05، وأيضاً حتى يكون الجزء المظلل في الطرفين مساوياً 0.05.

Example 10.2. The graph of the chi-square distribution with 5 degrees of freedom is shown in Figure 10-1. Find the values for χ_1^2, χ_2^2 for which the shaded area on the right = 0.05 and the total shaded area = 0.05.



شكل 10-1

إذا كان الجزء المظلل في الطرف الأيمن هو 0.05. فإن المنطقة شمال χ_2^2 هي $(1 - 0.05) = 0.95$. كما أن χ_2^2 تمثل 95% أي $\chi_{0.95}^2$.

وبالرجوع إلى الجدول المرفق D وقراءة أسفل العمود المعنون ν حتى تصل إلى الرقم 5. وبالتالي للعمود المعنون $\chi_{0.95}^2$ لنفس السطر 5، نجد أن القيمة هي 11.1 والتي تمثل قيمة كاي تربيع χ^2 .

وثانياً، حيث أن التوزيع غير متماثل Not Symmetric فإنه يوجد أكثر من قيمة تكون المساحة بالنسبة لها 0.05. وكمثال على ذلك تكون المساحة المظللة في الطرف الأيمن هي 0.04 وتكون في الطرف الأيسر هي 0.01. ولكنه يكون في العادة اعتبار المساحتين المظلتين في كلا الطرفين متساويتين إلا إذا ذكر غير ذلك؛ أي أن كل منهما يكون مساوياً 0.025.

فإذا كانت المساحة المظللة في الطرف الأيمن هي 0.025، فإن المساحة على شمال χ_2^2 هي $(1 - 0.025) = 0.975$ وتمثل χ_2^2 هي نسبة 97.5% و $\chi_{0.975}^2$ وبالتالي نحصل على 12.8 من جدول (D). وبالمثل إذا كانت المساحة المظللة في الطرف الأيسر هي 0.025، فإن المساحة على شمال χ_1^2 هي 0.025 وتمثل χ_1^2 2.5%، $\chi_{0.025}^2$ تكون مساوية 0.831. ولذلك فإن القيم المطلوبة هي 12.8، 0.831.

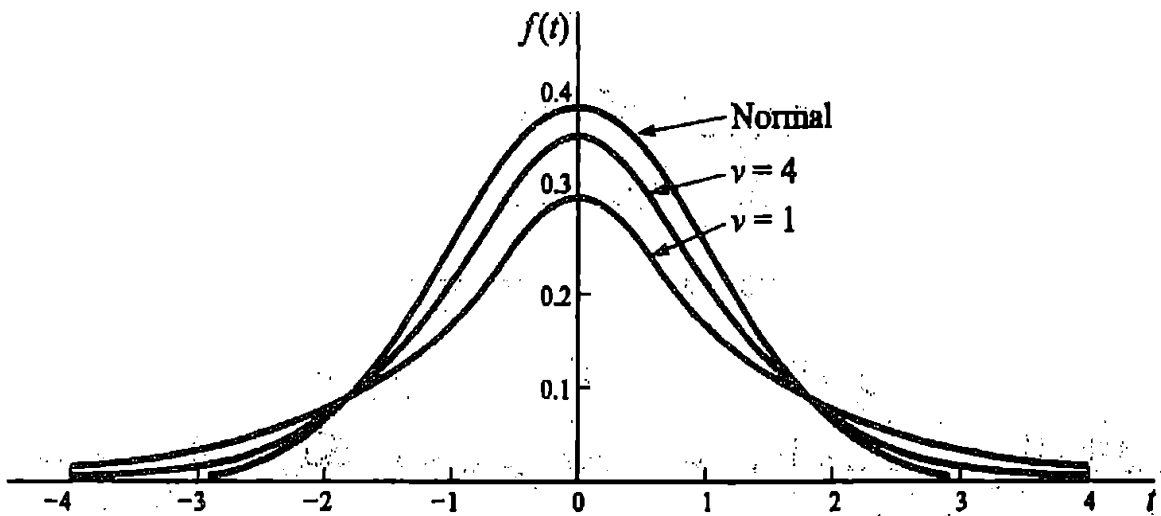
Student's t Distribution

توزيع t ستيودنت

إذا كان المتغير العشوائي له كثافة الاحتمال التالية

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \quad -\infty < t < \infty \quad (23)$$

فإننا نقول أن له توزيع t ستيودنت وباختصار توزيع t بدرجات حرية ν . فإذا كانت ν كبيرة ($\nu \geq 30$) فإن الرسم البياني للدالة $f(t)$ تقترب من المنحنى الطبيعي كما هو مبين في الشكل 10-2.



شكل 10-2

والقيم المئوية للتوزيع t بدرجات حرية ν يرمز لها بالرمز $t_{p,\nu}$ أو باختصار t_p تكون مفهومة. وللحصول عليها يكون من الجدول بالملحق (C). وحيث أن توزيع t متماثل فإن $t_{1-p} = -t_p$ أى أنه $t_{0.05} = -t_{0.95}$ والتوقع والتباين لتوزيع t هو

$$\mu = 0 \quad \text{and} \quad \sigma^2 = \frac{\nu}{\nu-2} \quad (\nu > 2) \quad (24)$$

والنظرية التالية مهمة فى العمل التالى.

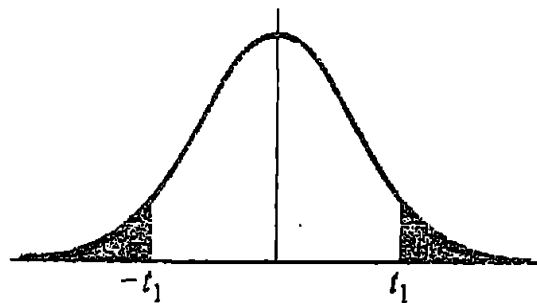
نظرية 10-4: إذا كان لدينا المتغيران العشوائيان المستقلان Y, Z ، حيث أن Y تكون موزعة توزيعاً طبيعياً توقعه صفر وتباينه الوحدة، بينما Z تكون موزعة طبقاً لتوزيع كاي تربيع بدرجات حرية ν . فإن المتغير العشوائى.

$$T = \frac{Y}{\sqrt{Z/\nu}} \quad (25)$$

يكون له توزيع t بدرجات حرية ν .

مثال 10.3: إذا كان منحنى توزيع t بدرجات حرية 9 الموضح بالشكل 10-3. فأوجد القيم t_1 التى تمثل المساحة المظلل بالنسبة لها 0.05 فى الطرف الأيمن وتكون المساحة الغير مظلمة هى 0.99.

Example 10.3. The graph of Student's t distribution with 9 degrees of freedom is shown in Figure 10-3. Find the value of t_1 for which the shaded area on the right = 0.05 and the total unshaded area = 0.99.



شكل 10-3

إذا كانت المساحة المظللة في الطرف الأيمن هي 0.05 فإن المساحة على شمال t_1 هي $(1 - 0.05) = 0.95$ وأن t_1 تمثل نسبة 0.95، أي $t_{0.95}$. وبالرجوع إلى جدول الملحق (C) فننظر أسفل العمود المعنون v حتى نصل إلى الرقم 9، ثم ننظر يميناً حتى العمود المعنون $t_{0.95}$ فتكون النتيجة هي القيمة المطلوبة t .

ثم إذا كانت المساحة الغير مظلة هي 0.99، فإن المساحة المظللة هي $(1 - 0.99) = 0.01$ وتكون المساحة المظللة ناحية اليمين هي $\frac{0.01}{2} = 0.005$.

ومن الجدول نجد أن $t_{0.995} = 3.25$

The F Distribution

F توزيع

المتغير العشوائي الذي يكون له توزيع F اختصاراً لاسم فيشر (Named after R.A. Fisher) بدرجات حرية v_1, v_2 Degrees of Freedom يكون بكثافة الاحتمال

$$f(u) = \begin{cases} \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} v_1^{v_1/2} v_2^{v_2/2} u^{(v_1/2)-1} (v_2 + v_1 u)^{-(v_1 + v_2)/2} & u > 0 \\ 0 & u \leq 0 \end{cases} \quad (26)$$

وقيمة المئين لتوزيع F بدرجات الحرية v_1, v_2 يرمز لها بالرمز F_{p, v_1, v_2} . والجدول الذي يعطينا هذه القيم هو بالملحق (E) ويكون القيم $F_{0.99}, F_{0.95}$ والمتوسط والتباين لهذا التوزيع هو

$$\mu = \frac{v_2}{v_2 - 2} \quad (v_2 > 2) \quad \text{and} \quad \sigma^2 = \frac{2v_2^2(v_1 + v_2 + 2)}{v_1(v_2 - 4)(v_2 - 2)^2} \quad (27)$$

وللتوزيع منوال وحيد في القيمة

$$u_{\text{mode}} = \left(\frac{v_1 - 2}{v_1} \right) \left(\frac{v_2}{v_2 + 2} \right) \quad (v_1 > 2) \quad (28)$$

والنظريات التالية مهمة في العمل التالي.

نظرية 10-5: إذا كان لدينا المتغيران العشوائيان المستقلان V_1 ، V_2 بحيث أنه يكون لكل منهما كاي تربيع بدرجات حرية v_1 ، v_2 على الترتيب، فإن المتغير

$$V = \frac{V_1 / v_1}{V_2 / v_2} \quad (29)$$

يكون له توزيع F بدرجات حرية v_1 ، v_2 .

نظرية 10-6:

$$F_{1-p, v_2, v_1} = \frac{1}{F_{p, v_1, v_2}} \quad (30)$$



تذكرا!

بينما نستخدم بصفة خاصة مع العينات الصغيرة كل من توزيع t وتوزيع χ^2 وتوزيع F فإنهم جميعاً بالمثل صالحون في حالة العينات الكبيرة.

العلاقات بين توزيعات كاي تربيع، t و F

Relationships Among Chi-Squares, t , and F

نظرية 10-7:

$$F_{1-p, 1, v} = t_{1-(p/2), v}^2 \quad (31)$$

نظرية 10-8:

$$F_{p, v, \infty} = \frac{\chi_{p, v}^2}{v} \quad (32)$$

مثال 10.4: تأكد من نظرية 10-7 ببيان أن $F_{0.95} = t_{0.975}^2$.

Example 10.4. Verify Theorem 10-7 by showing that $F_{0.95} = t_{0.975}^2$.

وذلك بمقارنة النواتج في العمود الأول للجدول $F_{0.95}$ في الملحق (E) مع نظيره في توزيع t تحت $t_{0.975}$. فنجد أن

$$161 = (12.71)^2, 18.5 = (4.30)^2, 10.1 = (3.18)^2, 7.71 = (2.78)^2, \text{ etc.,}$$

وهكذا..

والتي تمثل التأكد المطلوب.

مثال 10.5: تحقق من النظرية 10-8 عندما تكون $p = 0.99$.

Example 10.5. Verify Theorem 10-8 for $p = 0.99$.

فنقارن نواتج الصف الأخير من $F_{0.99}$ في جدول الملحق (E) (المناظر لدرجة الحرية $v_2 = \infty$) مع النواتج في الملحق (D) تحت $\chi_{0.99}^2$. فنجد أن:

$$6.63 = \frac{6.63}{1}, 4.61 = \frac{9.21}{2}, 3.78 = \frac{11.3}{3}, 3.32 = \frac{13.3}{4}, \text{ etc.,}$$

والتي تمثل التحقق المطلوب.

ملحق (A)

Mathematical Topics رياضية

Special Sums

مجاميع خاصة

يمثل التالي بعض مجاميع متسلسلات. ومن المعروف أن $0! = 1$ وحينما تكون المتسلسلات لانهاية فإن مدى التقارب يكون مذكوراً.

$$1. \sum_{j=1}^m j = 1 + 2 + 3 + \dots + m = \frac{m(m+1)}{2}$$

$$2. \sum_{j=1}^m j^2 = 1^2 + 2^2 + 3^2 + \dots + m^2 = \frac{m(m+1)(2m+1)}{6}$$

$$3. e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{j=0}^{\infty} \frac{x^j}{j!} \quad \text{all } x$$

$$4. \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots = \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{(2j+1)!} \quad \text{all } x$$

$$5. \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots = \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j}}{(2j)!} \quad \text{all } x$$

$$6. \frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots = \sum_{j=0}^{\infty} x^j \quad |x| < 1$$

$$7. \ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots = -\sum_{j=1}^{\infty} \frac{x^j}{j} \quad -1 \leq x < 1$$

Eulers' Formulas

صيغ أويلر

$$8. \quad e^{i\theta} = \cos \theta + i \sin \theta, \quad e^{-i\theta} = \cos \theta - i \sin \theta$$

$$9. \quad \cos \theta = \frac{e^{i\theta} + e^{-i\theta}}{2}, \quad \sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

The Gamma Function

دالة جاما

نرمز لدالة جاما بالرمز $\Gamma(n)$ وهي

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt \quad n > 0$$

ونلاحظ أن الصيغة المتكررة هي

$$\Gamma(n+1) = n\Gamma(n)$$

وحيث أن $\Gamma(1)$. فإن فك دالة جاما، $n > 0$ باستخدام الدالة المتكررة السابقة. فإذا كانت n رقمًا صحيحًا موجبًا، فإن

$$\Gamma(n+1) = n!$$

ولهذا السبب تسمى أحيانًا دالة المضروب Factorial Function. ومن أهم خصائص دالة جاما

$$\Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin p\pi}$$

وعندما $p = \frac{1}{2}$ ، فإن

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

وللقيم الكبيرة n فإننا نحصل على صيغة تقريب ستيرلنج Stirling's
Asymptotic Formula التالية

$$\Gamma(n+1) \sim \sqrt{2\pi n} n^n e^{-n}$$

The Beta Function

دالة بيتا

نرمز لدالة بيتا بالرمز $B(m, n)$ وتعرف كالتالي

$$B(m, n) = \int_0^1 u^{m-1} (1-u)^{n-1} du \quad m > 0, n > 0$$

وترتبط بدالة جاما كالتالي

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$$

Special Integrals

تكاملات خاصة

$$10. \int_0^{\infty} e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} \quad a > 0$$

$$11. \int_0^{\infty} x^m e^{-ax^2} dx = \frac{\Gamma\left(\frac{m+1}{2}\right)}{2a^{(m+1)/2}} \quad a > 0, m > -1$$

$$12. \int_0^{\infty} e^{-ax^2} \cos bx dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} e^{-b^2/4a} \quad a > 0$$

$$13. \int_0^{\infty} e^{-ax} \cos bx dx = \frac{a}{a^2 + b^2} \quad a > 0$$

$$14. \int_0^{\infty} e^{-ax} \sin bx dx = \frac{b}{a^2 + b^2} \quad a > 0$$

$$15. \int_0^{\infty} x^{p-1} e^{-ax} dx = \frac{\Gamma(p)}{a^p} \quad a > 0, p > 0$$

$$16. \int_{-\infty}^{\infty} e^{-(ax^2+bx+c)} dx = \sqrt{\frac{\pi}{a}} e^{(b^2-4ac)/4a} \quad a > 0$$

$$17. \int_0^{\infty} e^{-(ax^2+bx+c)} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}} e^{(b^2-4ac)/4a} \operatorname{erfc}\left(\frac{b}{2\sqrt{a}}\right) \quad a > 0$$

حيث

$$\operatorname{erfc}(u) = 1 - \operatorname{erf}(u) = 1 - \frac{2}{\sqrt{\pi}} \int_0^u e^{-x^2} dx = \frac{2}{\sqrt{\pi}} \int_u^{\infty} e^{-x^2} dx$$

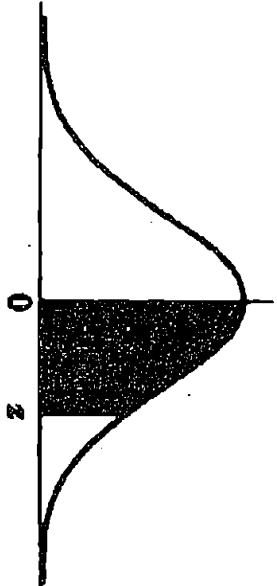
ويسمى دالة مكمل الخطأ Complementary Error Function.

$$18. \int_0^{\infty} \frac{\cos \omega x}{x^2 + a^2} dx = \frac{\pi}{2a} e^{-a\omega} \quad a > 0, \omega > 0$$

$$19. \int_0^{\pi/2} \sin^{2m-1} \theta \cos^{2n-1} \theta d\theta = \frac{\Gamma(m)\Gamma(n)}{2\Gamma(m+n)} \quad m > 0, n > 0$$

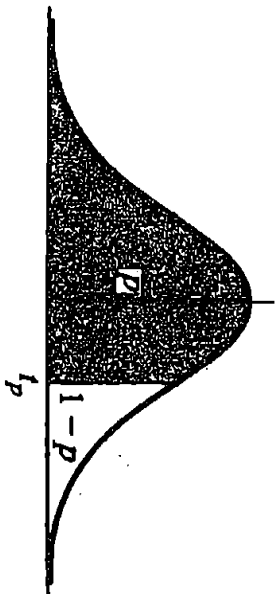
(B) ملاحق

المساحات تحت المنحنى الطبيعي المعياري من 0 حتى Z



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319

1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000



(C) ملحق

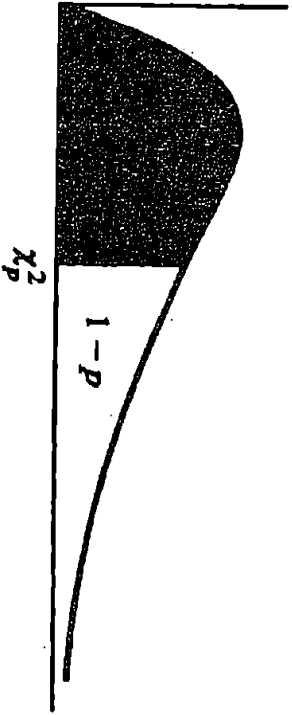
توزیع t سٹیوڈنٹ

ν	$t_{.55}$	$t_{.60}$	$t_{.70}$	$t_{.75}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.158	.325	.727	1.000	1.376	3.08	6.31	12.71	31.82	63.66
2	.142	.289	.617	.816	1.061	1.89	2.92	4.30	6.96	9.92
3	.137	.277	.584	.765	.978	1.64	2.35	3.18	4.54	5.84
4	.134	.271	.569	.741	.941	1.53	2.13	2.78	3.75	4.60
5	.132	.267	.559	.727	.920	1.48	2.02	2.57	3.36	4.03
6	.131	.265	.553	.718	.906	1.44	1.94	2.45	3.14	3.71
7	.130	.263	.549	.711	.896	1.42	1.90	2.36	3.00	3.50
8	.130	.262	.546	.706	.889	1.40	1.86	2.31	2.90	3.36
9	.129	.261	.543	.703	.883	1.38	1.83	2.26	2.82	3.25
10	.129	.260	.542	.700	.879	1.37	1.81	2.23	2.76	3.17
11	.129	.260	.540	.697	.876	1.36	1.80	2.20	2.72	3.11
12	.128	.259	.539	.695	.873	1.36	1.78	2.18	2.68	3.06

13	.128	.259	.538	.694	.870	1.35	1.77	2.16	2.65	3.01
14	.128	.258	.537	.692	.868	1.34	1.76	2.14	2.62	2.98
15	.128	.258	.536	.691	.866	1.34	1.75	2.13	2.60	2.95
16	.128	.258	.535	.690	.865	1.34	1.75	2.12	2.58	2.92
17	.128	.257	.534	.689	.863	1.33	1.74	2.11	2.57	2.90
18	.127	.257	.534	.688	.862	1.33	1.73	2.10	2.55	2.88
19	.127	.257	.533	.688	.861	1.33	1.73	2.09	2.54	2.86
20	.127	.257	.533	.687	.860	1.32	1.72	2.09	2.53	2.84
21	.127	.257	.532	.686	.859	1.32	1.72	2.08	2.52	2.83
22	.127	.256	.532	.686	.858	1.32	1.72	2.07	2.51	2.82
23	.127	.256	.532	.685	.858	1.32	1.71	2.07	2.50	2.81
24	.127	.256	.531	.685	.857	1.32	1.71	2.06	2.49	2.80
25	.127	.256	.531	.684	.856	1.32	1.71	2.06	2.48	2.79
26	.127	.256	.531	.684	.856	1.32	1.71	2.06	2.48	2.78
27	.127	.256	.531	.684	.855	1.31	1.70	2.05	2.47	2.77
28	.127	.256	.530	.683	.855	1.31	1.70	2.05	2.47	2.76
29	.127	.256	.530	.683	.854	1.31	1.70	2.04	2.46	2.76
30	.127	.256	.530	.683	.854	1.31	1.70	2.04	2.46	2.75
40	.126	.255	.529	.681	.851	1.30	1.68	2.02	2.42	2.70
60	.126	.254	.527	.679	.848	1.30	1.67	2.00	2.39	2.66
120	.126	.254	.526	.677	.845	1.29	1.66	1.98	2.36	2.62
∞	.126	.253	.524	.674	.842	1.28	1.645	1.96	2.33	2.58

(D) ملحق

توزیع کای تربیع χ^2

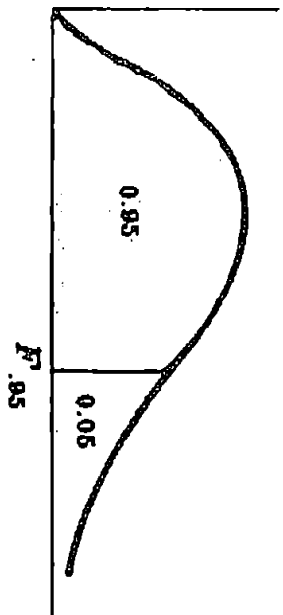


ν	$\chi_{.005}^2$	$\chi_{.01}^2$	$\chi_{.025}^2$	$\chi_{.05}^2$	$\chi_{.10}^2$	$\chi_{.25}^2$	$\chi_{.50}^2$	$\chi_{.75}^2$	$\chi_{.90}^2$	$\chi_{.95}^2$	$\chi_{.975}^2$	$\chi_{.99}^2$	$\chi_{.995}^2$	$\chi_{.999}^2$
1	.0000	.0002	.0010	.0039	.0158	.102	.455	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	.0100	.0201	.0506	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	.0717	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3	32.9

13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6	43.8
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6	51.2
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100	104	112
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	102	107	112	116	125
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	108	113	118	124	128	137
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109	118	124	130	136	140	149

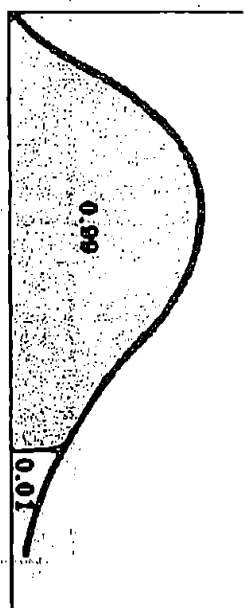
ملحق (F)

قيم النسب 95% ، 99% لتوزيع F



$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13

15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00



$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5981	6023	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17

14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

ملحق (F)

$e^{-\lambda}$
قيم

$(0 < \lambda < 1)$

λ	0	1	2	3	4	5	6	7	8	9
0.0	1.0000	.9900	.9802	.9704	.9608	.9512	.9418	.9324	.9231	.9139
0.1	.9048	.8958	.8869	.8781	.8694	.8607	.8521	.8437	.8353	.8270
0.2	.8187	.8106	.8025	.7945	.7866	.7788	.7711	.7634	.7558	.7483
0.3	.7408	.7334	.7261	.7189	.7118	.7047	.6977	.6907	.6839	.6771
0.4	.6703	.6636	.6570	.6505	.6440	.6376	.6313	.6250	.6188	.6126
0.5	.6065	.6005	.5945	.5886	.5827	.5770	.5712	.5655	.5599	.5543
0.6	.5488	.5434	.5379	.5326	.5273	.5220	.5169	.5117	.5066	.5016
0.7	.4966	.4916	.4868	.4819	.4771	.4724	.4677	.4630	.4584	.4538
0.8	.4493	.4449	.4404	.4360	.4317	.4274	.4232	.4190	.4148	.4107
0.9	.4066	.4025	.3985	.3946	.3906	.3867	.3829	.3791	.3753	.3716

($\lambda = 1, 2, 3, \dots, 10$)

λ	1	2	3	4	5	6	7	8	9	10
$e^{-\lambda}$.36788	.13534	.04979	.01832	.006738	.002479	.000912	.000335	.000123	.000045

ملاحظة : للحصول على قيم $e^{-\lambda}$ لأي قيم أخرى ل λ استخدم قوانين الأسس

Example: $e^{-3.48} = (e^{-3.00})(e^{-0.48}) = (.04979)(.6188) = .03081.$

ملحق (G)

أرقام عشوائية

51772	74640	42331	29044	46621	62898	93582	04186	19640	87056
24033	23491	83587	06568	21960	21387	76105	10863	97453	90581
45939	60173	52078	25424	11645	55870	56974	37428	93507	94271
30586	02133	75797	45406	31041	86707	12973	17169	88116	42187
03585	79353	81938	82322	96799	85659	36081	50884	14070	74950
64937	03355	95863	20790	65304	55189	00745	65253	11822	15804
15630	64759	51135	98527	62586	41889	25439	88036	24034	67283
09448	56301	57683	30277	94623	85418	68829	06652	41982	49159
21631	91157	77331	60710	52290	16835	48653	71590	16159	14676
91097	17480	29414	06829	87843	28195	27279	47152	35683	47280

قائمة المصطلحات العلمية (إنجليزي/عربي)

(A)			
Alternative hypothesis	الفرض البديل	Coefficient of determination	معامل التحديد
Areas under the standard normal curve		Combinations	توافيق
المساحات تحت المنحنى الطبيعي المعياري		Combinatorial analysis	تحليل التوافيق
Arithmetic mean	الوسط الحسابي	Complementary error function	
Asymptotically normal		دالة مكمل الخطأ	
	توزيعاً طبيعياً تقاربياً	Conditional probability	الاحتمال الشرطي
(B)		Confidence intervals	فترات الثقة
Basic probability	أساس الاحتمال	differences and sums	
Bayes' theorem	نظرية بايز	للفرق وللمجموع	
Bernoulli trials and distribution		means	للمتوسطات
	محاولات وتوزيع برنولي	population parameters	لمعالم المجتمع
Best fitting curve	توفيق أفضل منحنى	proportions	لالنسب
Beta distribution	توزيع بيتا	Confidence level	مستوى الثقة
Beta function	دالة بيتا	Confidence limits	حدود الثقة
Binomial coefficients		Continuous probability distribution	
	معاملات ذي الحدين	توزيع احتمالي متصل	
Binomial distribution		Continuous random variables	
	توزيع ذي الحدين	متغيرات عشوائية متصلة	
Binomial expansion	مفكوك ذات الحدين	Correlation and dependence	
(C)		الارتباط وعدم الاستقلال	
Cauchy distribution	توزيع كوشي	Covariance	تغاير
Central limit theorem		Critical values	القيم الحرجة
	نظرية الحد المركزية	Curve fitting, regression and correlation	
Centroid	مركز متوسط		
Chi-square distribution	توزيع كاي تربيع	تمهيد المنحنيات والانحدار والارتباط	

	(D)		Generalized correlation coefficient	
Degrees of freedom	درجات الحرية		معامل الارتباط العام	
Dependent variable	متغير تابع		(H)	
Descriptive statistics	إحصاء وصفي	Histogram	المدرج التكراري	
Deviation error	خطأ الانحراف	Hypergeometric distribution		
Discrete probability distribution			توزيع هايبرجيومتري	
	توزيع احتمالي منفصل	Hypothesis and significance		
Discrete random variables			الفروض والمعنوية	
	متغير عشوائي منفصل	(I)		
Dispersion	تشتت	Independent events	أحداث مستقلة	
	(E)	Independent variables	متغيرات مستقلة	
Elementary events	الأحداث الأولية	Interquartile range	المدى الربيعي	
Empirical probability	الاحتمال التجريبي	Interval probability	فترة الاحتمال	
Estimates	تقديرات	(L)		
confidence interval	بفترة ثقة	Law of large numbers		
point and interval	بنقطة وفترة		قانون الأعداد الكبيرة	
standard error	الخطأ المعياري	Least squares line	خط المربعات الصغرى	
unbiased and efficient		Least squares method		
	غير متحيز وكفاء		طريقة المربعات الصغرى	
Estimation theory	نظرية التقدير	Level of significance	مستوى المعنوية	
Eulers' formulas	صيغ أويلر	Linear correlation coefficient		
Expected values	القيم المتوقعة		معامل الارتباط الخطي	
	(F)	Linear regression	انحدار خطي	
Factorial function	دالة المضروب	Linear relationship	علاقة خطية	
F distribution	توزيع F	(M)		
Frequency distributions	توزيعات تكرارية	Mathematical topics	موضوعات رياضية	
	(G)	Mean	متوسط	
Gamma distribution	توزيع جاما	Masures of central tendency		
Gamma function	دالة جاما		مقاييس النزعة المركزية	
Gaussian distribution	توزيع جاوس	Measures of dispersion	مقاييس التشتت	

Median	وسيط	Random samples	عينات عشوائية
Method of least squares	طريقة المربعات الصغرى	Random variables	متغيرات عشوائية
Mode	منوال	Region of acceptance	منطقة القبول
Multinomial distribution	توزيع متعدد	Region of nonsignificance	منطقة عدم المعنوية
(N)		Region of rejection	منطقة الرفض
n factorial	مضروب n	Region of significance	منطقة المعنوية
Nonlinear relationship	علاقة غير خطية	Regression	انحدار
Normal distribution	توزيع طبيعي	Reliability	صلاحية
Null hypothesis	فرض العدم	(S)	
(O)		Sample mean	متوسط العينة
One-tailed tests	اختبار الجانب الواحد	Sample spaces and points	فراغات ونقاط العينة
(P)		Sample statistics	إحصاءات العينة
Parabola	قطع مكافئ	Sample variance	تباين العينة
Percentiles	المئين	Sampling distributions	توزيعات المعاينة
Permutations	تباديل	Sampling theory	نظرية المعاينة
Poisson distribution	توزيع بواسون	Scatter	انتشار
Polygon graph	رسم المضلع	Scatter diagram	شكل الانتشار
Population and sample	المجتمع والعينة	Skewness	التواء
Principle of counting	مبادئ العد	Slope	ميل
Probability	احتمال	Special integrals	تكاملات خاصة
Probability distributions	توزيعات احتمالية	Special sums	مجاميع خاصة
Product-moment formula	صيغة العزوم	Special tests	اختبارات خاصة
P Value	قيمة P	Standard deviation	انحراف معياري
(Q)		Standard error	خطأ معياري
Quadratic curve	منحنى من الدرجة الثانية	Standard normal curve	المنحنى المعتاد المعياري
(R)			
Random experiments	تجارب عشوائية		
Random numbers	أرقام عشوائية		

Standard normal density function	دالة الكثافة للتوزيع المعتاد المعياري	Chi-square	نظرية كاي تربيع
Standard score	الدرجة المعيارية	expectation	نظرية التوقع
Standard variable	متغير معياري	F distribution	نظرية توزيع F
Statistical decisions	قرارات إحصائية	law of large numbers	قانون الأعداد الكبيرة
Statistical hypothesis	فروض إحصائية	probability	نظرية الاحتمال
Stirling's approximation to n!	تقريب ستيرلنج لمضروب n	sampling distribution of means	توزيع المعاينة للمتوسطات
Stirling's asymptotic formula	صيغة ستيرلنج التقريبية	Student's t	نظرية توزيع t
Stochastic variable	متغير تصادفي	variance	نظرية التباين
Student's t distribution	توزيع t ستيودنت	Transformed variables	متغيرات محورة
Sums of series	مجاميع المتسلسلات	Translation of axes	نقل المحاور
(T)		Two-tailed tests	اختبارات الجانبيين
t distribution (see Student's t distribution)	توزيع t (انظر توزيع t ستيودنت)	Type I and Type II errors	أخطاء من النوع الأول والنوع الثاني
Test involving normal distribution	اختبار باستخدام التوزيع الطبيعي	(U)	
Test of hypothesis and significance	اختبار الفروض والمعنوية	Unbiased estimate	تقدير غير متحيز
Theorems	نظريات	Uniform distribution	التوزيع المنتظم
Bayes'	نظرية بايز	(V)	
central limit	نظرية الحد المركزية	Values of $e^{-\lambda}$	قيمة $e^{-\lambda}$
		Variance	تباين
		Variation	تغير - اختلاف