

Statistics and Probability Olasılık ve İstatistik



Eng. Abdulrahman Hamdi

Bu derste ne öğreneceğiz ?

01 İstatistiğine giriş

Merkezi Eğilim ve Dağılım Ölçüler

02 Popülasyon ve Örnek

Veri Türleri ve Toplanması

03 Olasılığın giriş

Temel kombinasyon ve
permütasyon ve olasılık
Toplam Olasılık Kuralı

04 Olasılık

Koşullu Olasılık
Bağımsızlık
Bayes teoremi

Bu derste ne öğreneceğiz ?

05 Rastgele Değişkenler ve Ayırık Dağılımlar

Ayrık Dağılımlar
Beklenen Değer ve Varyans
Bernoulli, Binom ve Poisson

07 Veri Analizi

Merkezi Limit Teoremi
Sıklık Dağılımı
Korelasyon

06 Rastgele Değişkenler ve Sürekli Dağılımlar

Sürekli Dağılımlar
Normal Dağılım
Standard Dağılım

08 İstatistik

Tahmin
Karar
Hipotez

01



İstatistiğine giriş

Eng: Abdulrahman Hamdi

Giriş:

Neden İstatistik?

Bir şeyi anlamak/öğrenmek için öncelikle onunla ilgili verileri toplamak gerekir. Veri toplama ve analiz etme, kalıpları anlamaya, bilinçli kararlar almaya ve bilim, iş dünyası ve tıp gibi birçok alanda gelecekteki sonuçları tahmin etmeye yardımcı olur.

لماذا الإحصاء؟

لفهم/تعلم شيء ما، يجب أولاً جمع البيانات المتعلقة به وتحليلها لفهم الأنماط، واتخاذ قرارات مستنيرة، وتوقع النتائج المستقبلية في مجالات متعددة مثل العلوم، الأعمال، والطب.

Merkezi Eğilim ve Dağılım Ölçüler

Merkezi Eğilim Ölçüleri:

Aritmetik Ortalama: Veri setindeki tüm değerlerin toplamının, değer sayısına bölünmesiyle elde edilir.

Medyan: Veri setindeki orta değer; veri sıralandığında ortadaki değer.

Mod: En sık tekrar eden veri değeri.

Dağılım Ölçüleri:

Standart Sapma: Veri değerlerinin ortalamadan ne kadar uzaklaştığını ölçer.

Varyans: Standart sapmanın karesi olup, veri dağılımının ne kadar yaygın olduğunu gösterir.

مقاييس الاتجاه المركزي:

المتوسط الحسابي: حاصل قسمة مجموع القيم على عدد القيم في مجموعة البيانات.

الوسيط: القيمة المتوسطة في مجموعة البيانات عند ترتيبها ترتيباً تصاعدياً.

المodal: القيمة التي تتكرر بشكل أكبر في مجموعة البيانات.

مقاييس التوزيع:

الانحراف المعياري: يقيس مدى تباعد القيم عن المتوسط.

التباين: مربع الانحراف المعياري، ويعكس مدى انتشار البيانات.

Merkezi Eğilim ve Dağılım Ölçüler

Merkezi Eğilim Ölçüleri:

Ortalama, bir veri setindeki değerlerin merkezi eğilimini veya genel eğilimini temsil eden bir ölçüdür. En yaygın kullanılan ortalama türleri şunlardır:

Aritmetik Ortalama: Tüm değerlerin toplamının, değer sayısına bölünmesiyle bulunur.

Geometrik Ortalama: Tüm değerlerin çarpımının, değer sayısının köküne alınmasıyla hesaplanır.

Harmonik Ortalama: Değerlerin terslerinin ortalamasının tersidir.

Ağırlıklı Ortalama: Her bir değer belirlir bir ağırlıkla çarpılıp, bu çarpımların toplamının, ağırlıkların toplamına bölünmesiyle bulunur.

مقاييس الاتجاه المركزي:

المتوسط: هو مقياس يمثل الاتجاه المركزي أو الاتجاه العام لمجموعة بيانات. الأنواع الأكثر شيوعاً من المتوسطات هي: **المتوسط الحسابي:** يُحسب بجمع جميع القيم وقسمة الناتج على عدد القيم.

المتوسط الهندسي: يُحسب بأخذ الجذر النوني لمنتج جميع القيم، حيث تشير n إلى عدد القيم.

المتوسط التوافقي: هو معكوس متوسط القيم المعكوسة.

المتوسط الوزني: يُحسب بضرب كل قيمة بوزن معين، ثم جمع هذه المضاعفات وقسمة الناتج على مجموع الأوزان.

02



Popölasyon ve Örnek

Eng: Abdulrahman Hamdi

Popölasyon ve Örnek

POPÖLASYON

Ölçülebilir kaliteye parametre denir.
Popölasyon tam bir kümedir.
Raporlar, görüşlerin doğru bir temsilidir.
Belirli bir grubun tüm üyelerini içerir.

ÖRNEKLEM

Ölçülebilir kaliteye istatistik denir.
Örneklem, popölasyonun bir alt kümesidir.
Raporlarda hata payı ve güven aralığı bulunur.
Tüm popölasyonu temsil eden bir alt kümedir.

المجتمع

تسمى الجودة القابلة للقياس "معلمة".
المجتمع هو مجموعة كاملة.
التقارير هي تمثيل دقيق للآراء.
يشمل جميع أعضاء مجموعة معينة.

العينة

تسمى الجودة القابلة للقياس "إحصائية".
العينة هي مجموعة فرعية من المجتمع.
تحتوي التقارير على هامش خطأ وفترة ثقة.
هي مجموعة فرعية تمثل المجتمع بأكمله.

Popülasyon ve Örnek

ÖRNEKLEM

ortalama, medyan, mod;
varyasyon, standart sapma.

POPÜLASYON

hipotez testi,
güven aralığı,
regresyon.

Popölasyon ve Örnek

POPÖLASYON

Genellikle **N** ile gösterilir.

Araştırma konusundaki tüm bireylerin veya öğelerin tamamını içerir.

Popölasyon, parametrelerle (örneğin, ortalama (μ) ve standart sapma (σ) tanımlanır.

ÖRNEKLEM

Genellikle (**n**) ile gösterilir.

Popölasyonun bir alt kümesidir.

Örneklem, istatistiklerle (örneğin, örneklem ortalaması ($\bar{x}$) ve örneklem standart sapması (**s**) tanımlanır.

المجتمع

يُعرَف بالرمز N

هو مجموعة كاملة وشاملة لجميع الأفراد أو العناصر قيد الدراسة.

يتم استخدام معلمات مثل المتوسط (μ) والانحراف المعياري (σ) لوصف المجتمع

العينة

يُعرَف بالرمز n

هي مجموعة فرعية تُختار من المجتمع.

تُستخدم الإحصائيات مثل المتوسط ($\bar{x}$) والانحراف المعياري للعينة (s) لوصف العينة

تحتوي على هامش خطأ وفترة ثقة لتقدير القيم في المجتمع.

Merkezi Eğilim ve Dağılım Ölçüler

Örnekler:

10 8 11 15 15 56 85 75 8

Veri setini kullanarak aşağıdaki soruları cevaplayın:

- 1- Aritmetik ortalama kaçtır?
- 2- Geometrik ortalama kaçtır?
- 3- Harmonik ortalama kaçtır?
- 4- Mod kaçtır?

10 8 11 15 12 56 85 75 19 20

Veri setini kullanarak aşağıdaki soruları cevaplayın:

- 1- Varyans kaçtır?
- 2- Standart Sapma kaçtır?
- 3- Ortanca kaçtır?
- 4- Varyans ve Standart Sapma kaçtır? (**Popülasyon**)

Merkezi Eğilim ve Dağılım Ölçüler

Örnekler:

1- Aşağıdakilerden hangisi popülasyonun bir özelliği değildir?

- A) Parametrelerle tanımlanır
- B) Araştırma konusundaki tüm bireyleri içerir
- C) Örneklem standart sapması ile tanımlanır
- D) Genellikle N harfi ile gösterilir

2- Popülasyon hangi harf ile gösterilir?

- A) n
- B) s
- C) N
- D) $\bar{x}$

3- Örneklem için kullanılan sembol ve istatistiklerden hangisi doğrudur?

- A) N ve ortalama (μ)
- B) n ve örneklem ortalaması ($\bar{x}$)
- C) σ ve örneklem standart sapması (s)
- D) n ve popülasyon ortalaması (μ)

4- Bir öğrencinin ders notları ve bu derslerin ağırlıkları aşağıdaki gibidir:

Matematik: 80 (Ağırlık: 3)

Fizik: 70 (Ağırlık: 4)

Kimya: 90 (Ağırlık: 2)

Bu derslerin ağırlıklı ortalaması nedir?

03

Olasılığına giriş



Eng: Abdulrahman Hamdi

SAYMA

Toplama Yoluyla Sayma: A ve B sonlu ve ayırık iki küme olmak üzere, bu iki kümenin birleşim kümesinin eleman sayısı $s(A \cup B) = s(A) + s(B)$ olarak bulunur. Sonlu ve ayırık iki kümenin birleşiminin eleman sayısını bu şekilde bulmaya Toplama Yoluyla Sayma denir.

Çarpma Yoluyla Sayma: İki ayırık olaydan birincisi a farklı yolla, ikincisi b farklı yolla gerçekleşiyorsa, bu iki olay birlikte $a \cdot b$ farklı yolla gerçekleşir. Yani, $s(A) \cdot s(B) = a \cdot b$ formülüyle ifade edilir.

العد عن طريق الجمع: إذا كان لدينا مجموعتان A و B نهائيتان ومنفصلتان، فإن عدد العناصر في اتحاد هاتين المجموعتين يُحسب باستخدام العلاقة التالية:
 $s(A \cup B) = s(A) + s(B)$ يُطلق على هذه الطريقة في حساب عدد عناصر اتحاد مجموعتين منفصلتين "العد عن طريق الجمع".

العد عن طريق الضرب: إذا كان لدينا حدثين منفصلين، يمكن أن يحدث الأول بعدة طرق مختلفة قدرها a، ويمكن أن يحدث الثاني بعدة طرق مختلفة قدرها b، فإن عدد الطرق التي يمكن أن يحدث بها الحدثان معاً هو $a \cdot b$ ويتم التعبير عن ذلك بالعلاقة
 $s(A) \cdot s(B) = a \cdot b$.

SAYMA

Saymanın Temel İlkesi: Bir işlem n_1 yolla bunu takip eden, işlemler $n_2, n_3, n_4, \dots, n_r$ farklı biçimde yapılmış olsun. Bu işlemin tamamı $n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_r$ çarpımı kadar farklı yolla yapılabilir buna Saymanın Temel İlkesi denir.

Faktöriyel: $n \in \mathbb{N}^+$ olmak üzere 1 den n ye kadar olan doğal sayıların çarpımına Faktöriyeldir ve $n!$ ile gösterilir.

مبدأ العد الأساسي: إذا كان بالإمكان إجراء عملية ما بعدة طرق قدرها n_1 ، وكانت العمليات التي تليها يمكن أن تُجرى بعدة طرق قدرها $n_2, n_3, n_4, \dots, n_r$ ، فإن العدد الإجمالي للطرق الممكنة لإتمام هذه العمليات هو حاصل ضرب $n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_r$. يُعرف هذا بـ "مبدأ العد الأساسي".

العامل (فكتوريال): إذا كان n عدد طبيعي موجب فإن حاصل ضرب الأعداد الطبيعية من 1 إلى n يُسمى "العامل" ويرمز له بالرمز $n!$.

Permütasyon

Permütasyon (Sıralama): n ve $r \in N^+$ ve $n \geq r$ olmak üzere, n elemanlı bir kümenin birbirinden farklı r tane elemanının bir sıra üzerinde farklı her dizilişinin n r li permütasyonudur ve $P(n, r)$ ile gösterilir.

$$P(n, r) = \frac{n!}{(n - r)!}$$

Veya $P(n, r) = n \cdot (n - 1) \cdot (n - 2) \cdots (n - r + 1)$

Tekrarlı Permütasyon: n tane nesnenin $n_1, n_2, n_3, \dots, n_k$ tanesi kendi aralarında özdeş ve $n_1 + n_2 + n_3 + \dots + n_k = n$ olmak üzere, bu n tane nesnenin kendi aralarındaki birbirinden farklı sıralamalarının sayısı

التبديل (الترتيب):

إذا كان n و $r \in N^+$ و $r \leq n$ ، فإن كل ترتيب مختلف لـ r من العناصر المميزة لمجموعة تحتوي على n عناصر يُسمى التبديل (التوافق المرتبة) لـ n من r ويُرمز له بالرمز $P(n, r)$.

$$P(n, r) = \frac{n!}{(n - r)!}$$

$$P(n, r) = n \cdot (n - 1) \cdot (n - 2) \cdots (n - r + 1)$$

التبديل المكرر: إذا كان لدينا n من العناصر، منها

$n_1, n_2, n_3, \dots, n_k$ متماثلة فيما بينها و $n_1 + n_2 + n_3 + \dots + n_k = n$ ، فإن عدد الترتيبات المختلفة لهذه العناصر فيما بينها يُحسب بهذه الطريقة.

Kombinasyon

Kombinasyon (Gruplama): n ve $r \in N^+$ ve $n \geq r$ olmak üzere, n elemanlı bir kümenin r elemanlı her alt kümesine bu kümenin r li kombinasyonudur. $C(n, r)$ ile gösterilir.

$$C(n, r) = \frac{n!}{(n-r)! r!}$$

التوافيق:

إذا كان n و $N^+ \ni r$ و $r \leq n$ ، إذا كان لدينا مجموعة تحتوي على n عنصرًا، فإن كل مجموعة فرعية تحتوي على r عنصرًا منها تُسمى التوافيق (التوليف) لـ r من n . ويُرمز لها بالرمز $C(n, r)$.

$$C(n, r) = \frac{n!}{(n-r)! r!}$$

OLASILIK

OLASILIK: Bir deneyde elde edilen sonuçların her birine çıktıdır. Bir deneyde bütün çıktılarının kümesine "örnek uzay" denir. Örnek uzayın her bir çıktısına da "olay" denir.

Aynı örnek uzaydaki bir olaya ait olası durumların sayısı başka bir olaya ait olası durumların sayısına eşit ise bu olaylara "Eş Olası Olaylar" denir.

Bir örnek uzayda tanımlanan iki olayın ortak elemanı yoksa bu olaya "Ayrık Olay" denir.

Bir E Örnek Uzayında Bir A Olayının Olasılığı

$$P(A) = \frac{s(A)}{s(E)} = \frac{\text{A Olayının Eleman Sayısı}}{\text{Örnek Uzayın Eleman Sayısı}}$$

İhtimal: يُطلق على كل نتيجة من نتائج تجربة ما اسم "ناتج". ويُطلق على مجموعة جميع النواتج في تجربة معينة اسم "فضاء العينة". وكل ناتج في فضاء العينة يُسمى "حدثًا".

إذا كان عدد الاحتمالات الخاصة بحدث ما في نفس فضاء العينة مساويًا لعدد الاحتمالات الخاصة بحدث آخر، فإن هذه الأحداث تُسمى "أحداث متكافئة الاحتمال".

أما إذا لم يكن هناك عناصر مشتركة بين حدثين معرفين في نفس فضاء العينة، فإن هذه الأحداث تُسمى "أحداث متفرقة" أو "أحداث منفصلة".

OLASILIK

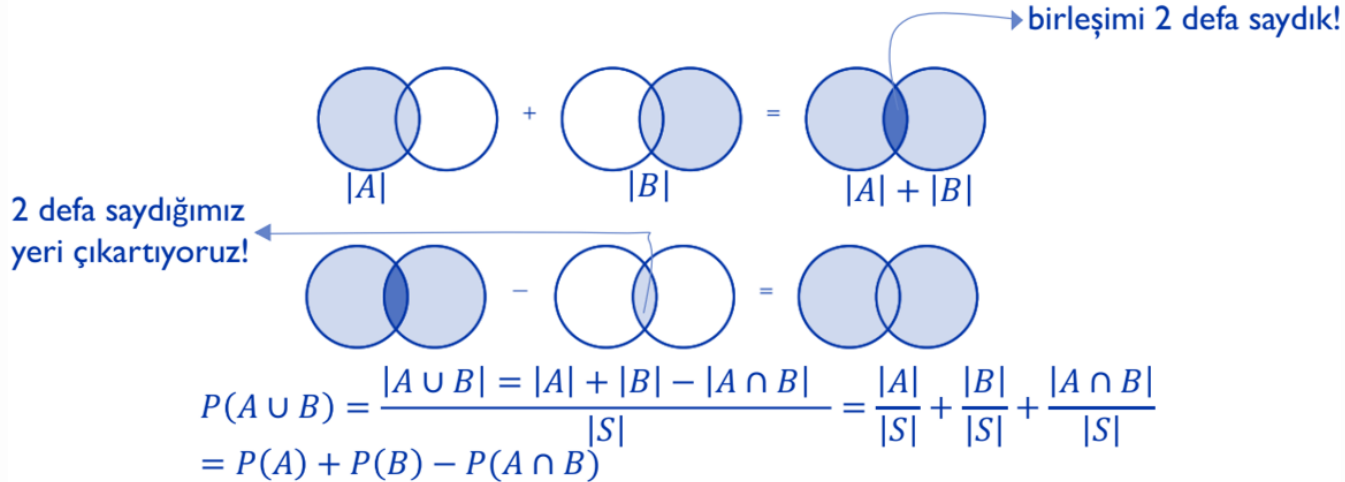
OLASILIK Özellikleri:

- 1) $0 \leq P(A) \leq 1$ Olasılık her zaman 0 ile 1 arasında bir sayıdır
- 2) $P(A) = 0$ ise (imkansız olay)
- 3) $P(1) = 1$ ise (kesin olay) Örnek uzayın olasılığı 1'dir
- 4) $P(A) + P(A') = 1$
- 5) $P(A) = 1 - P(A')$
- 6) A ve B ayrık iki olay ise A veya B olayının olasılığı
$$P(A \cup B) = P(A) + P(B) \text{dir}$$
- 7) A ve B ayrık olmayan iki olay ise $(A \cap B) \neq \emptyset$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

OLASILIK

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

İspat: :A ve B herhangi iki küme iken $|A \cup B| = |A| + |B| - |A \cap B|$



OLASILIK

ör.5 defa atılan bir bozuk paranın en az bir defa yazı gelme olasılığı nedir?

Çözüm.

Bir paranın 5 defa atılma deneyinde toplam $2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$ farklı sonuç elde edilebilir. Sonuçta $|S| = 32$ olur.

E olayı 5 defa atılan bir bozuk paranın en az bir defa yazı gelmesi olayı olsun. Bu durumda E^c olayı 5 atışta hiç yazı gelmeme durumu olur ve yalnızca bir sonuç içerir: $E^c = \{TTTTT\}$. E 'nin tamamlayıcısı olan E^c olayının olasılığını hesaplamak çok daha kolaydır: $P(E^c) = \frac{1}{32}$

$P(E) + P(E^c) = 1$ idi. O halde $P(E) = 1 - P(E^c) = 1 - \frac{1}{32} = \frac{31}{32} \approx 0.96$

OLASILIK

ör.Diyelimki bir mağazanın müşterilerinin %22'i Visa, %58 'i Mastercard, %14'ü her iki kartı birden kullanıyor. Bir müşterinin visa yada mastercard'ların en az birini kullanma olasılığı kaçtır?

Çözüm.

A olayı Visa kullanma olayı; B olayı Mastercard kullanma olayı olsun. Bu durumda $A \cup B$, Visa yada Mastercard'dan herhangi birini kullanma olayı olur.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.22 + 0.58 - 0.14 = 0.66$$

ör.Diyelimki bir kişinin bir mağazadan takım elbise alma olasılığı 0.3; kırıavat alma olasılığı 0.2; hem takım elbise hemde kırıavat alma olasılığı 0.1 olsun. Bu kişinin ne takım elbise nede kırıavat almama olasılığı kaçtır?

Çözüm.

A takım elbise alma olayı olsun. Bu durumda $P(A) = 0.3$.

B kırıavat alma olayı olsun. Bu durumda $P(B) = 0.2$.

Hem takım elbise hem kırıavat alma olasılığı: $P(A \cap B) = 0.1$

Takım elebise veya kırıavat alma olasılığı: $P(A \cup B) = 0.3 + 0.2 - 0.1 = 0.4$

Ne takım elbise nede kırıavat alma olasılığı: $1 - P(A \cup B) = 0.6$

04

Olasılık



Eng: Abdulrahman Hamdi

OLASILIK

Çarpım Kuralı :

Birinci elemanı A kümesinden, ikinci elemanı B kümesinden alınarak

oluşturulan ikililerin toplam sayısı: $|A| \times |B|$ dir.

Genelleştirirsek; $A_1, A_2, A_3, \dots, A_n$ n farklı küme olsun. Birinci elemanı A_1 'den,

ikinci elemanı A_2 'den,.... alınarak

oluşturulabilecek n elemanlı dizilerin sayısı:

$$|A_1| \times |A_2| \times |A_3| \dots \times |A_n|$$

قاعدة الضرب:

وتنص أن عدد الأزواج التي يمكن تشكيلها حيث يؤخذ
العنصر الأول من المجموعة A ، والعنصر الثاني من
المجموعة B هو: $|B| \times |A|$.

بالتعميم؛ دع $A_1, A_2, A_3, \dots, A_n$ تكون n
مجموعات مختلفة. عدد السلاسل ذات العناصر n
والتي يمكن تشكيلها حيث يؤخذ العنصر الأول من
 A_1 ، والثاني من A_2 ،... وهكذا هو:

$$|A_1| \times |A_2| \times |A_3| \dots \times |A_n|$$

OLASILIK

Şartlı(Koşullu) Olasılık Formülü:

A verilmişken B 'nin şartlı olasılığı: $P(B|A) = A$ 'nın içinde B 'nin de olma olasılığı oluşturulan ikililerin Deneyin sonucu A 'nın içinde olsun. Deneyin sonucunun B 'nin de içinde olması için sonucun hem A hem de B 'nin içinde yani $A \cap B$ 'nin içinde olması gerekir. Deneyin sonucu A 'nın içinde olduğundan A bizim yeni örnek uzayımız olur. Şu halde A olmuşken B 'nin olmasının olasılığı yani: $P(B|A)$

$$\square \quad P(B|A) = \frac{|A \cap B|}{|A|} = \frac{\frac{|A \cap B|}{|E|}}{\frac{|A|}{|E|}} \Rightarrow \mathbf{P(B|A) = \frac{P(A \cap B)}{P(A)}}$$

صيغة الاحتمال الشرطي:

إعطاء A ، احتمال B المشروط: $P(B|A) =$ احتمال وجود B داخل A . يجب أن يكون ناتج التجربة داخل A . لكي يكون ناتج التجربة داخل B أيضًا، يجب أن يكون النتيجة داخل A و B معًا، أي داخل $A \cap B$. نظرًا لأن ناتج التجربة داخل A ، فإن A يصبح فضاء العينة الجديد لدينا. وبالتالي، فإن احتمال حدوث B بشرط حدوث A هو: $P(B|A)$

OLASILIK

Şartlı(Koşullu) Olasılık Formül:

Ayrıca içler dışlar çarpımıyla

$$P(A \cap B) = P(B|A) \times P(A)$$

Okunuşu: A ve B olaylarının birlikte olma olasılığı, A 'nın olma olasılığı ile A varken B 'nin olma olasılığının çarpımlarına eşittir. Buna çarpım kuralı denir.

Sonuç olarak olasılık teorisinin iki çok önemli kuralı vardır:

1. Toplam Kuralı: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
2. Çarpım Kuralı: $P(A \cap B) = P(B|A) \times P(A)$

صيغة الاحتمال الشرطي:

بالتبادل الداخلي والخارجي.

$$P(A \cap B) = P(B|A) \times P(A)$$

احتمال حدوث الحدثين A و B معًا يساوي حاصل ضرب احتمال حدوث A في احتمال حدوث B بشرط حدوث A . يُطلق على هذا اسم قاعدة الضرب. وبالتالي، هناك قاعدتان مهمتان للغاية في نظرية الاحتمالات: الضرب والجمع

OLASILIK

Bağımsız olaylar:

Bir olayın gerçekleşmesinin ya da gerçekleşmemesinin, diğer bir olayın olma ihtimalini etkilemediği durumları ifade eder. Örneğin, bir paranın tura veya yazı gelmesi, başka bir olay olan zar atmanın sonucunu etkilemez. Aynı şekilde, bir zarı iki kez atma deneyinde, ilk zarın sonucu, ikinci zarın sonucunu etkilemez. Bu tür olaylar, bağımsız olaylar olarak adlandırılır.

الأحداث المستقلة :

هي تلك التي لا يؤثر حدوث حدث ما أو عدم حدوثه على احتمالية حدوث حدث آخر. على سبيل المثال، ظهور صورة أو كتابة عند رمي عملة لا يؤثر على نتيجة رمي نرد. أو عند رمي نرد مرتين، نتيجة الرمية الأولى لا تؤثر على نتيجة الرمية الثانية. تُسمى هذه الأحداث بالأحداث المستقلة.

OLASILIK

Bağımsız olaylar:

Eğer A ve B olayları bağımsızsa, A olayının gerçekleşmiş olması, B olayının gerçekleşme olasılığını etkilemez. Yani, $P(B|A) = P(B)$ olur. Bu durumda, A ve B olaylarının birlikte olma olasılığı, A'nın olma olasılığı ile B'nin olma olasılığının çarpımına eşit olur. Bu ilişki, çarpım kuralı olarak bilinir.

Matematiksel olarak: $P(A \cap B) = P(B) \times P(A)$

Burada, $P(A \cap B)$, A ve B olaylarının her ikisinin de gerçekleşme olasılığını, $P(A)$, A olayının olasılığını ve $P(B)$, B olayının olasılığını ifade eder. Eğer iki olay bağımsızsa, bu formül her zaman geçerlidir.

الأحداث المستقلة :

إذا كانت الأحداث A و B مستقلة، فإن حدوث A لا يؤثر على احتمالية حدوث B. بمعنى آخر، $P(B|A) = P(B)$ في هذه الحالة، احتمال حدوث كل من A و B معًا يساوي حاصل ضرب احتمالات حدوث كل منهما بشكل منفصل. وهذا يعبر عنه بقاعدة الضرب.

رياضياً: $P(A \cap B) = P(B) \times P(A)$

حيث $P(A \cap B)$ هو احتمال حدوث كل من A و B معًا، و $P(A)$ هو احتمال حدوث A، $P(B)$ هو احتمال حدوث B إذا كانت الأحداث مستقلة، فإن هذه الصيغة دائماً صحيحة.

OLASILIK

Bağımsız olaylar:

Örnek: Bir zar ve bir para atıyorsunuz. Zarın ve paranın sonuçları bağımsız olaylardır. Zarın sonucu, paranın sonucunu etkilemez ve paranın sonucu zarın sonucunu etkilemez.

OLASILIK

Bayes Teorimi:

Bayes Teoremi, olasılık kuramında, olayların gerçekleşme olasılıklarını güncellemek için kullanılan bir yöntemdir. Teorem, öncül bir olasılık bilgisi (öncül olasılık) ile yeni elde edilen bilgi veya veri (gözlemsel bilgi) arasındaki ilişkiyi tanımlar. Bayes Teoremi şu formülle ifade edilir:

$$\square \quad P(A|B) = \frac{P(A).P(B|A)}{P(B)}$$

$\square \quad P(A|B)$ B olayı gerçekleştiğinde A olayının gerçekleşme olasılığıdır (koşullu olasılık).

$\square \quad P(B|A)$ -----

$\square \quad P(A)$ A olayının gerçekleşme olasılığıdır (önsel olasılık).

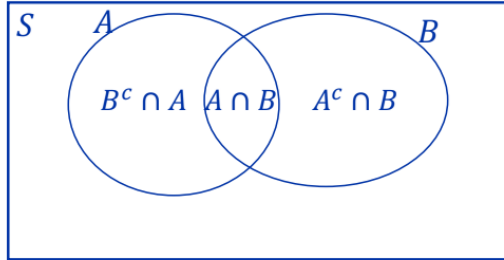
$\square \quad P(B)$ -----

نظرية بايس:

هي طريقة تُستخدم في نظرية الاحتمالات لتحديث احتمالات حدوث الأحداث بناءً على معلومات جديدة. تصف النظرية العلاقة بين الاحتمال المسبق (الاحتمال الأولي) والمعلومات الجديدة التي تم الحصول عليها (المعلومات الملاحظة). يتم التعبير عن نظرية بايس بالصيغة التالية:

OLASILIK

BAYES Teoremi



B kümesi ayrık olan (kesişimleri olmayan) $A \cap B$ ve $A^c \cap B$ 'nin birleşimi şeklinde yazılabilir:

$$B = (A \cap B) \cup (A^c \cap B).$$

O halde B 'nin olasılığı:

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

olur.

Bayes teoremi yeniden düzenlenirse:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

OLASILIK

Bayes Teorimi:

Örnek: Bir klinikte yapılan kanser testi gerçekte kanser olan hastaların %98'inde pozitif olarak sonuç veriyor, yani kanser diyor. Bu test, gerçekte kanser olmayan hastaların %97'sinde negatif olarak sonuç veriyor, yani kanser değilsin diyor. Ayrıca toplumdaki kişilerin 0.008'nin kanser olduğu biliniyor. Buna göre kanser testi pozitif çıkmış bir kişinin gerçekte kanser olma olasılığı nedir?

نظرية بايس:

مثال: في عيادة ما، يظهر اختبار السرطان نتيجة إيجابية لدى %98 من المرضى الذين يعانون بالفعل من السرطان، أي أن الاختبار يقول إنهم مصابون بالسرطان. هذا الاختبار يعطي نتيجة سلبية لدى %97 من الأشخاص الذين لا يعانون بالفعل من السرطان، أي أن الاختبار يقول إنهم غير مصابين بالسرطان. بالإضافة إلى ذلك، يُعرف أن 0.008 من الأشخاص في المجتمع يعانون من السرطان. بناءً على هذه المعطيات، ما هو احتمال أن يكون الشخص الذي حصل على نتيجة إيجابية في اختبار السرطان مصابًا بالسرطان بالفعل؟

OLASILIK

$$P(neden|sonuç) = \frac{P(sonuç|neden)P(neden)}{P(sonuç)}$$

Bayes Teorimi:

Genelde elimizde sonuç (yani gözlem, yani veri) vardır; ve biz tüm nedenler içersinde bu sonuca neden olabilecek en olası nedeni bulmak isteriz. Öte yandan çoğu kez $P(\text{Sonuç}|\text{Neden})$ 'i bulmak, $P(\text{Neden} | \text{Sonuç})$ 'i bulmaktan daha kolaydır. Diyelimki gözlemimiz (yani sonuç) öksürük. Buna neden olabilecek iki neden olsun: grip ve zatürre. Zatürre varken kişinin öksürmesi, grip varken kişinin öksürmesinden daha olasıdır. $P(\text{öksürük} | \text{Zatürre}) > P(\text{Zatürre} | \text{öksürük})$

نظرية بايس:

بشكل عام، لدينا النتيجة (أي الملاحظة أو البيانات)؛ ونريد أن نجد من بين جميع الأسباب السبب الأكثر احتمالاً الذي قد يؤدي إلى هذه النتيجة. من ناحية أخرى، في معظم الأحيان يكون من الأسهل العثور على (السبب | النتيجة) P بدلاً من العثور على (النتيجة | السبب) P . لنفترض أن ملاحظتنا (أي النتيجة) هي السعال. هناك سببان يمكن أن يكونا وراء هذا السعال: الإنفلونزا والالتهاب الرئوي. من المرجح أن يسعل الشخص في حالة الإصابة بالالتهاب الرئوي أكثر من سعال الشخص في حالة الإصابة بالإنفلونزا. $P(\text{السعال} | \text{الالتهاب الرئوي}) > P(\text{السعال} | \text{الإنفلونزا})$

OLASILIK

Bayes Teorimi:

Fakat genel olarak grip görölme olasılığı, zatürre görölme olasılığından çok daha yüksektir. $P(grip) \gg P(Zatürre)$

Bu yüzden teoremdeki $P(neden)$ katsayısı $P(Sonuç|Neden)$ 'i dengeler.

Diyelimki aynı sonuca yol açabilcek iki farklı neden olsun: $neden_1$ ve $neden_2$.

Ve biz bunlardan en olası nedeni ($P(Neden | Sonuç)$ değeri en yüksek nedeni)

bulmak istiyoruz. Bu durumda elimizde

$$P(neden_1|sonuç) = \frac{P(sonuç|neden_1)P(neden_1)}{P(sonuç)}$$

$$\text{ve } P(neden_2|sonuç) = \frac{P(sonuç|neden_2)P(neden_2)}{P(sonuç)}$$

نظرية بايس:

ولكن بشكل عام، احتمال حدوث الإنفلونزا أعلى بكثير من احتمال حدوث الالتهاب الرئوي.

(الإنفلونزا) $P \ll$ (الالتهاب الرئوي) P . لهذا السبب، معامل P (السبب) في النظرية يوازن P (النتيجة |

السبب). لنفترض أن هناك سببان مختلفان يمكن أن يؤديا إلى نفس النتيجة: $neden_1$ و $neden_2$.

ونحن نريد العثور على السبب الأكثر احتمالاً (السبب الذي تكون قيمته

(النتيجة | السبب) P هي الأعلى). في هذه الحالة لدينا

OLASILIK

Bayes Teorimi:

Karşılaştırmak istediğimiz iki değerinde paydası aynıdır: $P(\text{sonuç})$. O halde bu paydayı nedenleri karşılaştırıp, en olası nedeni ararken yok sayabiliriz. Nedenleri karşılaştırmada $P(\text{Sonuç}|\text{Neden})$ $P(\text{neden})$ çarpımına bakmak yeterli olur. Sonuç olarak $P(\text{Neden} | \text{Sonuç})$, $P(\text{Sonuç}|\text{Neden}) \times P(\text{neden})$ ile orantılıdır:

$$P(\text{Neden} | \text{Sonuç}) \propto P(\text{Sonuç}|\text{Neden}) \times P(\text{neden})$$

نظرية بايس:

القيمتان اللتان نريد مقارنتهما لهما نفس المقام:
 $P(\text{النتيجة})$. لذلك، يمكننا تجاهل هذا المقام أثناء مقارنة الأسباب والبحث عن السبب الأكثر احتمالاً. في مقارنة الأسباب، يكفي النظر إلى حاصل ضرب
 $P(\text{السبب} | \text{النتيجة}) \times P(\text{السبب})$. وبالتالي، فإن
 $P(\text{النتيجة}|\text{السبب})$ يتناسب مع
 $P(\text{السبب} | \text{النتيجة}) \times P(\text{السبب})$.

OLASILIK

Bayes Teorimi:

Diyelimki gözlemimiz (yani sonuç) öksürük. Buna neden olabilecek iki neden olsun: grip ve zatürre. Zatürre varken kişinin öksürmesi, grip varken kişinin öksürmesinden daha olasıdır.

$$P(\text{öksürük} \mid \text{Zatürre}) > P(\text{öksürük} \mid \text{grip})$$

Fakat genel olarak grip görülme olasılığı, zatürre görülme olasılığından çok daha yüksektir. $P(\text{grip}) \gg P(\text{Zatürre})$.

Bu yüzden $P(\text{neden})$ katsayısı, $P(\text{Sonuç} \mid \text{Neden})$ 'i dengeler.

$$P(\text{öksürük} \mid \text{Zatürre}) P(\text{zatürre}) \text{ Vs } P(\text{öksürük} \mid \text{grip}) P(\text{grip})$$

نظرية بايس:

نفترض أن ملاحظتنا (أي النتيجة) هي السعال. هناك سببان يمكن أن يكونا وراء هذا السعال: الإنفلونزا والالتهاب الرئوي. من المرجح أن يسعل الشخص في حالة الإصابة بالالتهاب الرئوي أكثر من سعال الشخص في حالة الإصابة بالإنفلونزا.

$$P(\text{الالتهاب الرئوي} \mid \text{السعال}) < P(\text{الإنفلونزا} \mid \text{السعال})$$

ولكن بشكل عام، احتمال حدوث الإنفلونزا أعلى بكثير من احتمال حدوث الالتهاب الرئوي. $P(\text{الإنفلونزا}) \gg P(\text{الالتهاب الرئوي})$.

لهذا السبب، معامل (السبب) P يوازن (السبب | النتيجة) P .

$$P(\text{الالتهاب الرئوي} \mid \text{السعال}) \times P(\text{الالتهاب الرئوي}) \text{ مقابل } P(\text{الإنفلونزا} \mid \text{السعال}) \times P(\text{الإنفلونزا})$$

OLASILIK

Bayes Teorimi:

Örnek: İki futbol takımı düşünelim: T1 ve T2. Varsayalım ki T1, T2 ile oynadığı maçların %65'ini kazanmış olsun. T1'in kazandığı maçların %70'i kendi sahasında gerçekleşmiş. Diğer yandan, T2'nin T1'i yendiği maçların %75'i kendi sahasında oynanmış. Eğer gelecek maç T1'in sahasında olacaksa, bu maçı hangi takımın kazanma olasılığı daha yüksektir?

نظرية بايس:

مثال: تخيل وجود فريقين لكرة القدم: T1 و T2. لنفترض أن T1 قد فاز في 65% من المباريات التي لعبها ضد T2. وأن 70% من المباريات التي فاز فيها T1 على T2 قد تمت على ملعبه. من ناحية أخرى، 75% من المباريات التي فاز فيها T2 على T1 قد تمت على ملعب T2 إذا كانت المباراة القادمة ستقام على ملعب T1 فمن المرجح أن يفوز أي فريق؟

OLASILIK

Bayes Teorimi:

Örnek: Kalem üreten bir fabrikada toplam kalem üretiminin %30'u A makinası, %50'si B makinası ve geriye kalan C makinası gerçekleştirmektedir. Bu makinaların üretimleri sırasında sırasıyla %4'ü, %2'si ve %5'i kusurlu kalemlerdir.

- a) Bir gün üretilen tüm kalemlerin seçilen bir kalemin kusurlu olma olasılığı kaçtır?
- b) Bir günün sonunda seçilen bir kalemin kusurlu olma bilgisi verildiğine göre, bu kalem'in A makinasından üretilmiş olma olasılığı kaçtır?

نظرية بايس:

مثال: في مصنع لإنتاج الأقلام، يتم إنتاج 30% من إجمالي الأقلام بواسطة ماكينة A ، و 50% بواسطة ماكينة B ، والجزء المتبقي بواسطة ماكينة C. أثناء عمليات الإنتاج لهذه الماكينات، تكون نسبة الأقلام المعيبة على التوالي 4%، و 2%، و 5%.

(أ) ما هي احتمالية أن يكون قلم مختار من بين جميع الأقلام المنتجة في يوم واحد معيباً؟

(ب) إذا تم إعطاء معلومات بأن القلم المختار معيب، ما هي احتمالية أن يكون قد تم إنتاجه بواسطة ماكينة A ؟

OLASILIK

Bayes Teorimi:

Örnek: Kalem üreten bir fabrikada toplam kalem üretiminin %30'u A makinası, %50'si B makinası ve geriye kalan C makinası gerçekleştirmektedir. Bu makinaların üretimleri sırasında sırasıyla %4'ü, %2'si ve %5'i kusurlu kalemlerdir.

- a) Bir gün üretilen tüm kalemlerin seçilen bir kalemin kusurlu olma olasılığı kaçtır?
- b) Bir günün sonunda seçilen bir kalemin kusurlu olma bilgisi verildiğine göre, bu kalem'in A makinasından üretilmiş olma olasılığı kaçtır?

نظرية بايس:

مثال: في مصنع لإنتاج الأقلام، يتم إنتاج 30% من إجمالي الأقلام بواسطة ماكينة A ، و 50% بواسطة ماكينة B ، والجزء المتبقي بواسطة ماكينة C. أثناء عمليات الإنتاج لهذه الماكينات، تكون نسبة الأقلام المعيبة على التوالي 4%، و 2%، و 5%.

(أ) ما هي احتمالية أن يكون قلم مختار من بين جميع الأقلام المنتجة في يوم واحد معيباً؟

(ب) إذا تم إعطاء معلومات بأن القلم المختار معيب، ما هي احتمالية أن يكون قد تم إنتاجه بواسطة ماكينة A ؟

05



Rastgele Değişkenler ve Ayrik Dağılımlar

Eng: Abdulrahman Hamdi

Rastgele Değişkenler ve Ayırık Dağılımlar

Rasgele Değişken (Random Variable):

Rasgele değişken, bir deneyin sonuçlarını sayısal değerlere atayan bir fonksiyondur. Olasılık teorisinde, rastgele bir olayın sayısal karşılığını temsil eder ve iki ana türü vardır: sürekli (continuous) ve ayırık (discrete).

Bir olasılık dağılımı değildir.

X, Y, Z gibi büyük harfler ile tanımlanır.

Türleri:

Ayrık Rasgele Değişken: Belirli ve sayılabilir değerler alır (örneğin, zar atışının sonucu).

Sürekli Rasgele Değişken: Belirli bir aralıktaki sonsuz sayıda değer alabilir (örneğin, sıcaklık ölçümü).

المتغير العشوائي:

المتغير العشوائي هو دالة تُسند نتائج تجربة ما إلى قيم عددية. في نظرية الاحتمالات، يمثل المتغير العشوائي المقابل العددي لحدث عشوائي، ويوجد نوعان رئيسيان: المتغير المتقطع والمتغير المستمر.

ليس توزيع احتمالي.

يتم تعريفه بواسطة أحرف كبيرة مثل X و Y و Z

الأنواع:

المتغير العشوائي المتقطع: يأخذ قيمًا محددة وقابلة للعد (مثل نتيجة رمي النرد).

المتغير العشوائي المستمر: يمكن أن يأخذ عددًا لا نهائيًا من القيم ضمن نطاق معين (مثل قياس درجة الحرارة).

Rastgele Değişkenler ve Ayırık Dağılımlar

Rasgele Değişken (Random Variable):

Kullanım Alanları:

Olasılık dağılımları ile olayların analizinde.

İstatistiksel modelleme ve veri analizi.

Finans, mühendislik ve sosyal bilimler gibi birçok alanda belirsiz olayların modellenmesinde.

Önemli Kavramlar:

Beklenen Değer (Expectation): Rasgele değişkenin uzun vadeli ortalaması.

Varyans: Değerlerin beklenen değerden sapmasının ölçüsü.

Olasılık Dağılımı: Değişkenin değerlerinin olasılıklarını tanımlar.

Rasgele değişkenler, belirsiz olayları sayısal olarak modellemeyi ve analiz etmeyi sağlar, bu da karar alma ve tahmin süreçlerinde önemli bir rol oynar.

المتغير العشوائي:

مجالات الاستخدام:

تحليل الأحداث باستخدام التوزيعات الاحتمالية.

النمذجة الإحصائية وتحليل البيانات.

نمذجة الأحداث غير المؤكدة في مجالات مثل المالية والهندسة والعلوم الاجتماعية.

المفاهيم المهمة:

القيمة المتوقعة: متوسط طويل الأجل للمتغير العشوائي.

التباين: قياس انحراف القيم عن القيمة المتوقعة.

التوزيع الاحتمالي: يحدد احتمالات قيم المتغير.

تساعد المتغيرات العشوائية على نمذجة الأحداث غير المؤكدة بشكل عددي وتحليلها، مما يلعب دورًا مهمًا في عمليات اتخاذ القرار والتنبؤ.

Rastgele Değişkenler ve Ayırık Dağılımlar

Ayrık Rastgele Değişkenler (Discrete Random Variables):

Belirli ve sayılabilir değerler alır (örneğin, zar atışının sonucu).

Eğer bir rastgele değişkenin alabileceği değerler sonlu sayıda elemanı olan bir kümeden seçiliyorsa (yani bu rastgele değişken sonlu tane değer alabiliyorsa) bu rastgele değişkene ayırık (discrete) rastgele değişken diyeceğiz.

المتغير العشوائي المتقطع:

يأخذ قيمًا محددة وقابلة للعد (على سبيل المثال، نتيجة رمي الزهر).

إذا كانت القيم التي يمكن أن يأخذها المتغير العشوائي مختارة من مجموعة تحتوي على عدد محدود من العناصر (أي أن هذا المتغير العشوائي يمكن أن يأخذ عددًا محدودًا من القيم)، فإننا نسمي هذا المتغير العشوائي متغيرًا عشوائيًا منفصلاً (discrete).

Rastgele Değişkenler ve Ayırık Dağılımlar

Olasılık Dağılımı (Probability Distribution):

Değişkenin değerlerinin olasılıklarını tanımlar.

Bir deneyde rastgele değişkenin alabileceği her değere karşılık bir olasılık vardır.

Diyelimki bir deneyde X rastgele değişkeni $x_1, x_2, \dots, x_n$ değerlerinden birini alsın.

Bu değerlerden birini alma olasılığını $P(X = x_i)$ şeklinde göstereceğiz.

التوزيع الاحتمالي:

يعرف احتمالات قيم المتغير.

في تجربة ما، يوجد احتمال لكل قيمة يمكن أن يأخذها المتغير

العشوائي. لنفترض أن المتغير العشوائي X يمكن أن يأخذ

واحدة من القيم $x_1, x_2, \dots, x_n$ في تجربة ما. سنعرض

احتمال أخذ أحد هذه القيم بالصيغة $P(X = x_i)$

Rastgele Değişkenler ve Ayırık Dağılımlar

Olasılık Dağılımı (Probability Distribution):

Örneğin bir önceki örnekte, X rastgele değişkeni (e, e, e) sonucu için 0 değerini alır. O halde $P(X = 0) = \frac{1}{8}$

X rastgele değişkeni $(k, e, e), (e, e, k), (e, k, e)$ sonuçları için 1 değerini alır. O halde $P(X = 1) = \frac{3}{8}$

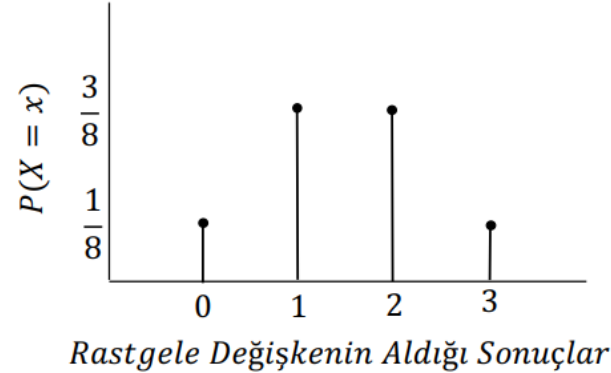
X rastgele değişkeni $(k, k, e), (k, e, k), (e, k, k)$ sonuçları için 2 değerini alır. O halde $P(X = 2) = \frac{3}{8}$

X rastgele değişkeni (k, k, k) , sonucu için 3 değerini alır. O halde $P(X = 3) = \frac{1}{8}$

Rastgele Değişkenler ve Ayırık Dağılımlar

Olasılık Dağılımı (Probability Distribution):

x	$P(X = x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
$\frac{+}{1}$	



$P(X = x_1), P(X = x_2), \dots, P(X = x_n)$ olasılıklarının hepsine birden X ayırık rastgele değişkeninin **olasılık dağılımı** diyeceğiz.

Doğal olarak $\sum_{i=1}^n x_i = 1$.

Rastgele Değişkenler ve Ayırık Dağılımlar

Olasılık Kütle Fonksiyonu (Probability Mass Function (PMF))

Ayrık olasılık dağılımı = olasılık kütle fonksiyonu (pmf)

Olasılık kütle fonksiyonu aldığı ayırık rastgele değişkeni, onun olasılığına götüren

fonksiyonun adıdır. p_X ile gösterilir.

$$p_X : X \rightarrow [0,1]$$

Bu fonksiyon aslında daha önce gösterdiğimiz $P(X = x)$ in kendisidir, yani X rastgele değişkenin x 'değerini alma olasılığını gösterir:

$$p_X(x) = P(X = x)$$

دالة الكتلة الاحتمالية (Probability Mass Function – PMF)

توزيع الاحتمال المنفصل = دالة الكتلة الاحتمالية (PMF)

دالة الكتلة الاحتمالية هي اسم الدالة التي تأخذ متغيراً عشوائياً منفصلاً وتربطه باحتماله. يتم تمثيلها بالرمز p_X

$$p_X : X \rightarrow [0,1]$$

هذه الدالة هي في الواقع نفسها التي تمثل $P(X = x)$ أي أنها تُظهر احتمال أن يأخذ المتغير العشوائي X القيمة x

$$p_X(x) = P(X = x)$$

Rastgele Değişkenler ve Ayırık Dağılımlar

Beklenen Değer (Expected Value)

Olasılık teorisindeki bir başka anahtar konu rastgele değişkenin beklenen değeridir.

X 'in beklenen değeri X 'in aldığı değerlerin ağırlıklı ortalamasıdır. Burada ağırlıklar değerlerin olasılıklarıdır (yani olasılık kütle fonksiyonudur).

X rastgele değişkenin beklenen değeri $E[X]$ ile gösterilir ve şöyle hesaplanır:

$$E[X] = \sum_{i=1}^n x_i P(X = x_i)$$

$x_1, x_2, \dots, x_n$ değerleri X rastgele değişkenin alabileceği değerler ve $P(X = x_i)$, x_i rastgele değişkeninin görülme olasılığıdır.

القيمة المتوقعة:

موضوع رئيسي آخر في نظرية الاحتمالات هو القيمة المتوقعة للمتغير العشوائي.

القيمة المتوقعة للمتغير العشوائي X هي المتوسط المرجح للقيم التي يأخذها X ، حيث تكون الأوزان هي احتمالات هذه القيم (أي دالة الكتلة الاحتمالية).

يُرمز للقيمة المتوقعة للمتغير العشوائي X بالرمز $E[X]$ ويُحسب كالتالي:

القيم $x_1, x_2, \dots, x_n$ هي القيم التي يمكن أن يأخذها المتغير العشوائي x_i ، $P(X = x_i)$ هي احتمالية رؤية القيمة x_i للمتغير العشوائي.

Rastgele Değişkenler ve Ayırık Dağılımlar

Beklenen Değer ~ Ortalama

Beklenen değer ile ortalama aslında birbirine çok yakın iki kavramdır. Eğer bir kümede her bir elemanın seçilme şansı aynı ise, bu kümenin ortalaması ile kümede seçilen elemanı gösteren rastgele değişkenin beklenen değeri aynı olur. $\bar{x}$ ile sembolize tanımlanır

ör. Diyelimki 5 kişilik bir arkadaş grubundaki kişilerin boyları cm cinsinden 178, 184, 169, 190, 164 olsun.

Bu kişilerden biri rastgele seçilsin. X rastgele değişkeni seçilen kişinin boyunu gösterebilir. X 'in alacağı değerler 178, 184, 169, 190, 164 olur ve her bir değişkeni alma olasılığı aynıdır: $P(X = x_i) = \frac{1}{5} (\forall i \in \{1,2,3,4,5\})$.

$$O \text{ halde } E[X] = 178 \times \frac{1}{5} + 184 \times \frac{1}{5} + 169 \times \frac{1}{5} + 190 \times \frac{1}{5} + 164 \times \frac{1}{5} = 177$$

Bu değer aynı zamanda gruptaki kişilerin boy ortalamasıdır:

$$\bar{x} = \frac{178 + 184 + 169 + 190 + 164}{5} = 177$$

القيمة المتوقعة ~ المتوسط

القيمة المتوقعة والمتوسط هما في الواقع مفهومان متشابهان للغاية. إذا كانت احتمالية اختيار كل عنصر في مجموعة ما متساوية، فإن المتوسط لتلك المجموعة سيكون هو نفسه القيمة المتوقعة للمتغير العشوائي الذي يمثل العنصر المختار من المجموعة. يرمز لها بالرمز $\bar{x}$

Rastgele Değişkenler ve Ayırık Dağılımlar

Rastgele Değişkenin Varyansı (Variance of Random Variables)

Değişkenleri özetlerken ortalama ve varyanstan faydalanıyorduk. Yani bir değişkenin ortalaması ve varyansı o değişkeni tanımlamaya, özetlemeye yetiyordu. Rastgele değişkenleri de özetlerken, beklenen değerini ve varyansını göstereceğiz. Beklenen değerin, onun ortalaması gibi düşünebiliriz. Fakat rastgele değişkenin aldığı değerlerin ortalamadan yani beklenen değerden ne kadar saptığını da bilmemiz gerekir. Örnek olarak aşağıdaki rastgele değişkenlerin beklenen değerleri aynıdır.

X rastgele değişkeni beklenen değerinden ne kadar sapması beklenir:

$(X - E[X])^2$ 'nin beklenen değeri nedir, yani $E[(X - E[X])^2]$?

$$\begin{aligned} E[(X - E[X])^2] &= E[(X^2 + (E[X])^2 - 2E[X]X)] \\ &= E[X^2] + E[E[X]^2] - E[2E[X]X] \\ &= E[X^2] + E[X]^2 - 2E[X]E[X] \\ &= E[X^2] + E[X]^2 - 2E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

O halde bir X rastgele değişkeninin varyansı:

$$\boxed{Var(X) = E[X^2] - (E[X])^2}$$

تباين المتغيرات العشوائية

عند تلخيص المتغيرات، كنا نستفيد من المتوسط والتباين. بمعنى أن المتوسط والتباين للمتغير كانا كافيين لتوصيف المتغير وتلخيصه. عند تلخيص المتغيرات العشوائية، سنوضح القيمة المتوقعة والتباين. يمكننا التفكير في القيمة المتوقعة كالم متوسط، ولكن من الضروري أيضًا معرفة مدى انحراف القيم التي يأخذها المتغير العشوائي عن المتوسط، أي القيمة المتوقعة. كمثال، القيم المتوقعة للمتغيرات العشوائية التالية هي نفسها.

Rastgele Değişkenler ve Ayırık Dağılımlar

Ortak Olasılık Dağılımı (Joint Probability Distribution)

Bazen bir deneyde tek bir rastgele değişkenin alabileceği değerle ilgilenmeyiz.

Birden fazla rastgele değişkenin birbirleriyle olan ilişkisi ile ilgileniriz.

Örneğin kişinin kanser teşhisi konulduğu yaşı ile kanserin ciddiyeti arasında bir ilişki arayabiliriz, yada üniversitedeki yıl ve alınan ders sayısı arasında bir ilişki.

X ve Y iki rastgele değişken olsun. $X: x_1, x_2, \dots, x_n$

değişkenlerini, $Y: y_1, y_2, \dots, y_m$

değişkenlerini alsın. Bu durumda, X 'in x_i değişkenini ve Y 'nin y_j değişkenini alma olasılığı:

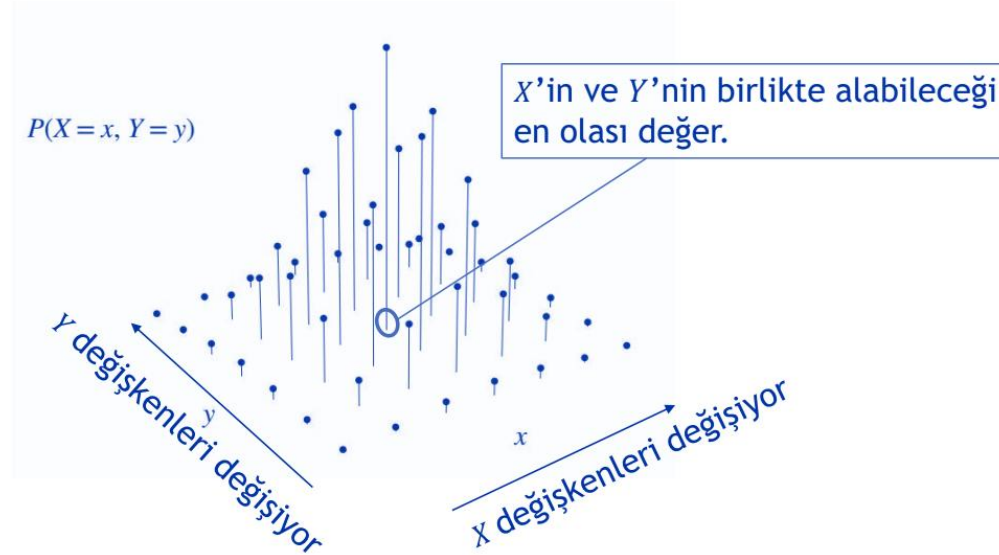
$$P(X = x_i, Y = y_j)$$

التوزيع الاحتمالي المشترك (Joint Probability Distribution)

في بعض الأحيان، لا نهتم فقط بالقيم التي يمكن أن يأخذها متغير عشوائي واحد في تجربة معينة. بل نكون مهتمين بالعلاقة بين أكثر من متغير عشوائي. على سبيل المثال، قد نبحث عن العلاقة بين العمر الذي تم فيه تشخيص الشخص بالسرطان وشدة المرض، أو العلاقة بين السنة الجامعية وعدد المواد الدراسية المأخوذة لنفترض أن لدينا المتغيرين العشوائيين X و Y . حيث $X: x_1, x_2, \dots, x_n$ يمثل قيم المتغير X ، و $Y: y_1, y_2, \dots, y_m$ يمثل قيم المتغير Y . في هذه الحالة، الاحتمالية أن يأخذ X القيمة x_i وأن يأخذ Y القيمة y_j تكون هي التوزيع الاحتمالي المشترك لهذين المتغيرين.

Rastgele Değişkenler ve Ayırık Dağılımlar

$P(X = x_i, Y = y_j)$ olasılıklarına X 'in ve Y 'nin ortak olasılık dağılımı diyeceğiz.



$$P(X = x_i, Y = y_j) = P(X = x_i) \times P(Y = y_j | X = x_i)$$

(çarpım kuralından)

Rastgele Değişkenler ve Ayırık Dağılımlar

ör. Bir zar atılma deneyinde hem zarın tek gelmesiyle hemde zarın asal gelmesi ile ilgilenelim. X rastgele değişkeni zar tekse 1, değilse 0 değerini alsın; Y rastgele değişkeni zar asalsa 1 değilse 0 değerini alsın.

$$X = \begin{cases} 1, \text{zar tekse} \\ 0, \text{zar tek değilse} \end{cases} \quad Y = \begin{cases} 1, \text{zar asalsa} \\ 0, \text{zar asal değilse} \end{cases}$$

Bu iki rastgele değişken birleşerek aşağıdaki gibi bir ortak olasılık dağılımı oluştururlar:

	$Y = 1$	$Y = 0$
$X = 1$	$P(X = 1, Y = 1)$	$P(X = 1, Y = 0)$
$X = 0$	$P(X = 0, Y = 1)$	$P(X = 0, Y = 0)$

X 'in 1 ve Y 'nin 0 almasının olasılığı. Kendine sor: ne zaman hem X , 1 değerini alır, hem de Y , 0 değerini alır. Bu zarın 1 gelmesiyle mümkündür.

O halde olasılık $\frac{1}{6}$ 'dır.

İkinci bir yöntem olarak çarpım kuralı ile bulabiliriz:

$$P(X = 1, Y = 0) = P(X = 1)P(Y = 0|X = 1) = \frac{3}{6} \cdot \frac{1}{3} = \frac{1}{6}$$

Rastgele Değişkenler ve Ayırık Dağılımlar

$$X = \begin{cases} 1, \text{zar tekse} \\ 0, \text{zar tek değilse} \end{cases} \quad Y = \begin{cases} 1, \text{zar asalsa} \\ 0, \text{zar asal değilse} \end{cases}$$

	$Y = 1$	$Y = 0$	Toplam
$X = 1$	$P(X = 1, Y = 1) = \frac{2}{6}$	$P(X = 1, Y = 0) = \frac{1}{6}$	$P(X = 1) = \frac{3}{6}$
$X = 0$	$P(X = 0, Y = 1) = \frac{1}{6}$	$P(X = 0, Y = 0) = \frac{2}{6}$	$P(X = 0) = \frac{3}{6}$
Toplam	$P(Y = 1) = \frac{3}{6}$	$P(Y = 0) = \frac{3}{6}$	1

Burası tek başına (X 'e bağımlı olmaksızın)
 Y 'nin 1 olma olasılığı

Rastgele Değişkenler ve Ayrik Dağılımlar

ör. Bir fakülte'deki öğrencilerin %75'i yerel, %25'i başka şehirlerden gelsin. Yerel öğrencilerin %60'ı bilgisayar mühendisliği, %30'u makine mühendisliği ve %10'u gıda mühendisliği bölümünde olsun. Başka şehirlerden gelen öğrencilerin %30'u bilgisayar mühendisliği, %40'ı makine ve %30'u inşaat okusun. X rastgele değişkeni öğrencinin yerel mi dışarıdan mı olduğunu; Y rastgele değişkeni öğrencinin bölümünü göstere. Buna göre X ve Y rastgele değişkenlerinin ortak olasılık dağılımı nasıl olur?

مثال: لنفترض أن 75% من الطلاب في كلية ما محليون، و25% يأتون من مدن أخرى. من بين الطلاب المحليين، 60% يدرسون هندسة الحاسوب، 30% يدرسون الهندسة الميكانيكية، و10% يدرسون الهندسة الغذائية. أما الطلاب القادمين من مدن أخرى، 30% منهم يدرسون هندسة الحاسوب، 40% يدرسون الهندسة الميكانيكية، و30% يدرسون الهندسة المدنية ليكن X متغير عشوائي يُمثل ما إذا كان الطالب محليًا أو من خارج المدينة، و Y متغير عشوائي يُمثل القسم الذي يدرسه الطالب. إذن، كيف سيكون التوزيع الاحتمالي المشترك للمتغيرين العشوائيين X و Y ؟

Rastgele Değişkenler ve Ayırık Dağılımlar

Çözüm.

		Y'nin alabileceği değerler		
X'in alabileceği değerler		Bilgisayar	Makine	İnşaat
	Yerel	$P(X = Yer, Y = Bil) =$ $P(X = Yer)P(Y = Bil X = Yer)$ $= 0.75 \times 0.6 = 0.45$	$P(X = Yer, Y = Mak) =$ $P(X = Yer)P(Y = Mak X = Yer)$ $= 0.75 \times 0.3 = 0.225$	$P(X = Yer, Y = İnş) =$ $P(X = Yer)P(Y = İnş X = Yer)$ $= 0.75 \times 0.1 = 0.075$
	Dış.	$P(X = Dış, Y = Bil) =$ $P(X = Dış)P(Y = Bil X = Dış)$ $= 0.25 \times 0.3 = 0.075$	$P(X = Dış, Y = Mak) =$ $P(X = Dış)P(Y = Mak X = Dış)$ $= 0.25 \times 0.4 = 0.1$	$P(X = Dış, Y = İnş) =$ $P(X = Dış)P(Y = İnş X = Dış)$ $= 0.25 \times 0.3 = 0.075$

Rastgele seçilen birinin yerel olması ve bilgisayar mühendisliği bölümünde olması olasılığı

	Bilgisayar	Makine	İnşaat	Toplam
Yerel	0.45	0.225	0.075	0.75
Dışarıdan	0.075	0.1	0.075	0.25
Toplam	0.525	0.325	0.15	1

Rastgele seçilen birinin makine mühendisliği bölümünde olması olasılığı. Dikkat edilirse bu, soruda açıkça verilmemiş, ortak olasılıklardan kendimiz hesapladık.

Rastgele Değişkenler ve Ayırık Dağılımlar

Marjinal Olasılık

Marjinal olasılık, ortak olasılıklar kullanılarak bulunan tek bir rastgele değişkenin belirli bir değeri alma olasılığıdır.

$$P(X = x_i) = \sum_{j=1}^m P(X = x_i, Y = y_j) \quad P(Y = y_j) = \sum_{i=1}^n P(X = x_i, Y = y_j)$$

marjinal olasılığı hesaplanırken X rastgele değişkeninin x_i değerini alması sabit iken, Y 'nin diğer değerleri alma olasılıklarının toplamını buluyoruz.

Bir önceki örnekten Y 'nin makine mühendiliği olması marjinal olasılığını hesaplayalım. Bir öğrenci yerel iken yada dışarıdan gelmiş iken makine mühendisi olabilir. Yani burada değişen şey öğrencinin nereden geldiğidir, makine mühendisliği sabittir.

الاحتمال الهامشي

هو احتمال أن يأخذ متغير عشوائي واحد قيمة معينة باستخدام التوزيعات الاحتمالية المشتركة عند حساب الاحتمال الهامشي، نبقى المتغير العشوائي X ثابتاً عند القيمة x_i ، ثم نجمع احتمالات القيم المختلفة التي يمكن أن يأخذها المتغير Y .
لنحسب الاحتمال الهامشي لكون Y هو الهندسة الميكانيكية من المثال السابق. يمكن للطالب أن يكون في الهندسة الميكانيكية سواء كان محلياً أو قادماً من مدينة أخرى. أي أن العامل المتغير هنا هو من أين أتى الطالب، بينما الهندسة الميكانيكية تظل ثابتة.

Rastgele Değişkenler ve Ayırık Dağılımlar

$$\begin{aligned} P(Y = makine) &= P(X = yerel, Y = makine) + P(X = dışarıdan, Y = makine) = \\ &= 0.225 + 0.1 = 0.325 \end{aligned}$$

	Bilgisayar	Makine	İnşaat	Toplam
Yerel	0.45	0.225	0.075	0.75
Dışarıdan	0.075	0.1	0.075	0.25
Toplam	0.525	0.325	0.15	1

→ X 'in marjinal dağılımı

↓
 Y 'nin marjinal dağılımı

Rastgele Değişkenler ve Ayırık Dağılımlar

Bazı Özel Ayırık Rastgele Değişkenler

1) Bernoulli Rastgele Değişkeni

Bir deneyde yalnızca 2 sonuç varsa (başarı, başarısızlık) bu bir Bernoulli deneyidir.

Bernoulli deneyleri örnekleri:

- Seçilen kişi başarılı, başarısız.
- Seçilen kişi kanser, kanser değil.
- Atılan bozuk para yazı, tura.
- Hava yağmurlu, kuru.
- Bir ilacın yan etkisi var, yan etkisi yok ...

Bernoulli deneyinde 2 sonuç olduğu için, bu sonuçların olasılıkları p ve $1 - p$ dir.

Bernoulli deneyinde 'başarı' sonucunu 1'e; 'başarısızlık' sonucunu 0'a götüren rastgele değişkene özel olarak Bernoulli Rastgele Değişkeni diyeceğiz.

بعض المتغيرات العشوائية المنفصلة الخاصة

1) المتغير العشوائي بيرنولي

إذا كان هناك نتيجتان فقط في تجربة (نجاح، فشل)، فإنها تُعتبر تجربة بيرنولي.

أمثلة على تجارب بيرنولي: الشخص المختار ناجح أو فاشل. الشخص المختار مصاب بالسرطان أو غير مصاب بالسرطان.

العملة الملقاة تظهر وجه أو ظهر.

الطقس ممطر أو جاف.

الدواء له آثار جانبية أو ليس له آثار جانبية ...

في تجربة بيرنولي، نظرًا لوجود نتيجتين، فإن احتمالات هذه النتائج هي p و $1 - p$. سنطلق على المتغير العشوائي الذي يأخذ القيمة 1 في حالة "النجاح" والقيمة 0 في حالة "الفشل" اسم المتغير العشوائي بيرنولي

Rastgele Değişkenler ve Ayrik Dağılımlar

Bazı Özel Ayrik Rastgele Değişkenler

1) Bernoulli Rastgele Değişkeni

X , Bernoulli rastgele değişkeni 1 yada 0 değerini alır.

Bu değişkenin $X = 1$ değerini alması (yani başarı olma) olasılığı:

$$P(X = 1) = p$$

$X = 0$ değerini alması (yani başarısızlık olma) olasılığı:

$$P(X = 0) = 1 - p$$

Bernoulli rastgele değişkeninin dağılımına yani $P(X = 1)$ ve $P(X = 0)$ değerlerine

Bernoulli olasılık dağılımı denir.

بعض المتغيرات العشوائية المنفصلة الخاصة

1) المتغير العشوائي لبيرنولي

المتغير العشوائي X لبيرنولي يأخذ القيمتين 1 أو 0. احتمالية

أن يأخذ هذا المتغير القيمة $X = 1$ (أي نجاح) هي:

$$P(X = 1) = p$$

واحتمالية أن يأخذ القيمة $X = 0$ (أي فشل) هي:

$$P(X = 0) = 1 - p$$

العشوائي لبيرنولي، أي القيمتين $P(X = 1)$ و $P(X = 0)$ اسم توزيع الاحتمال لبيرنولي

Rastgele Değişkenler ve Ayrik Dağılımlar

Bazı Özel Ayrik Rastgele Değişkenler

1) Bernoulli Rastgele Değişkeni

Bernoulli Rastgele Değişkeninin Beklenen Değeri

X , bir Bernoulli rastgele değişkeni olsun. Şu halde X 'in alabileceği değerler 1 ve 0 dir.

Ayrıca $P(X = 1) = p$ ve $P(X = 0) = 1 - p$ dir.

$$E[X] = 1 \times p + 0 \times (1 - p) = p$$
$$E[X] = p$$

Bir Bernoulli rastgele değişkenin beklenen değeri olma olasılığının ta kendisidir.

بعض المتغيرات العشوائية المنفصلة الخاصة

1) المتغير العشوائي لبيرنولي

القيمة المتوقعة للمتغير العشوائي لبيرنولي لنفترض أن X هو متغير عشوائي لبيرنولي. في هذه الحالة، القيم التي يمكن أن يأخذها X هي 1 و 0.

القيمة المتوقعة لمتغير عشوائي لبيرنولي هي بالضبط احتمالية النجاح نفسها.

Rastgele Değişkenler ve Ayırık Dağılımlar

Bazı Özel Ayırık Rastgele Değişkenler

2) Binomal Rastgele Değişkeni

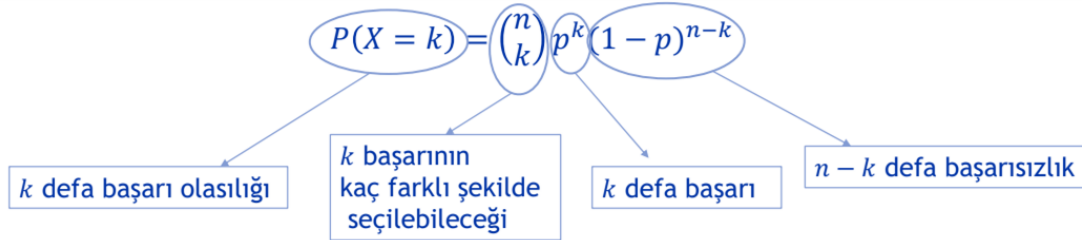
n tane birbirinden bağımsız Bernoulli deneyinde (yani iki sonuçlu deneyde) toplam

başarı sayısı binomial rastgele değişkeni verir.

Burada X rastgele değişkeni $0, 1, \dots, n$ değerlerini alabilir. Bu değerlerden birini

alması olasılığı

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$



بعض المتغيرات العشوائية المنفصلة الخاصة

(2) المتغير العشوائي ذو التوزيع ذو الحدين

في n تجربة بيرنولي مستقلة (أي تجارب ذات نتيجتين)، يكون إجمالي عدد النجاحات هو ما يُعرف بالمتغير العشوائي

ذو التوزيع ذو الحدين

هنا، يمكن للمتغير العشوائي X أن يأخذ القيم $0, 1, \dots, n$.

واحتمالية أن يأخذ إحدى هذه القيم هي ...

Rastgele Değişkenler ve Ayrik Dağılımlar

Bazı Özel Ayrik Rastgele Değişkenler

2) Binomal Rastgele Değişkeni

Binomial Rastgele Değişkenin Beklenen Değeri

$$E[X] = \sum_{i=0}^n i \times P(X = i)$$
$$E[X] = \sum_{i=0}^n i \times \binom{n}{i} p^i (1-p)^{n-i}$$
$$E[X] = np$$

Burada $E[X]$, n deneyde beklenen toplam başarı sayısı olarak yorumlanır.

بعض المتغيرات العشوائية المنفصلة الخاصة

2) المتغير العشوائي ذو التوزيع ذو الحدين

القيمة المتوقعة للمتغير العشوائي ذو التوزيع ذو الحدين

يمكن تفسير $E[X]$ على أنه العدد المتوقع الإجمالي للنجاحات في n تجربة

إذا كانت p هي احتمالية النجاح في كل تجربة، فإن القيمة المتوقعة لمتغير ذو توزيع ذو حدين X (أي العدد المتوقع للنجاحات) فإن: $E[X] = n.p$

حيث n هو عدد التجارب و p هو احتمال النجاح في كل تجربة

Rastgele Değişkenler ve Ayırık Dağılımlar

Bazı Özel Ayırık Rastgele Değişkenler Örnekler:

- 1) Bir A takımının, bir B takımını yenme olasılığı 0.3 olsun. Bu iki takım arasında yapılan 5 maçın 3'ünü A 'nın yenmesi olasılığı nedir?
- 2) 20 kişilik bir toplulukta 8 kişi araba kullanmayı biliyor olsun. Rastgele seçilen 5 kişiden ikisinin araba kullanma olasılığı nedir?
- 3) 100 defa atılan düzgün bir bozuk paranın toplam kaç defa tura gelmesi beklenir?
- 4) Yeni geliştirilen bir korona virüs aşısının 0.95 olasılıkla fayda sağladığı biliniyor. 8 kişiye bu aşı yapılıyor. Bu kişilerden en az ikisinin bu aşıdan fayda görme olasılığı nedir?
- 5) Bir önceki soru için en fazla 2 kişinin aşıdan fayda görmeme olasılığı nedir?

بعض المتغيرات العشوائية المنفصلة الخاصة أمثلة:

- 1) إذا كانت احتمالية فوز فريق A على فريق B هي 0.3، ما هي احتمالية أن يفوز A في 3 من أصل 5 مباريات ضد B ؟
- 2) في مجتمع مكون من 20 شخصًا، هناك 8 أشخاص يعرفون قيادة السيارة. ما هي احتمالية أن يكون من بين 5 أشخاص تم اختيارهم عشوائيًا شخصان يعرفان قيادة السيارة؟
- 3) عند رمي عملة معدنية منتظمة 100 مرة، كم مرة يُتوقع أن تظهر صورة؟
- 4) من المعروف أن اللقاح الجديد ضد فيروس كورونا لديه احتمالية 0.95 في تحقيق الفائدة. يتم إعطاء هذا اللقاح لـ 8 أشخاص. ما هي احتمالية أن يستفيد على الأقل شخصان من هؤلاء الأشخاص من اللقاح؟
- 5) بالنسبة للسؤال السابق، ما هي احتمالية أن يستفيد ما لا يزيد عن شخصين من اللقاح؟

06



Rastgele Değişkenler ve Sürekli Dağılımlar

Eng: Abdulrahman Hamdi

Rastgele Değişkenler ve Sürekli Dağılımlar

Şimdiye kadar ayrık rastgele değişken ile ilgilendik. Yani rastgele değişkenin alabileceği değerler sınırlı idi. Örneğin bir bozuk paranın üç kere atılma deneyinde, gelen tura sayılarını gösteren X rastgele değişkeninin alabileceği değerler yalnızca 0,1,2, yada 3'tü. Bunlardan başkası olamazdı.

Şimdi ise belirli bir aralıkta her değeri alabilen (yani sonsuz değer) alabilen rastgele değişken türü ile ilgileneceğiz. Bu rastgele değişken türüne sürekli rastgele değişken diyeceğiz.

Sürekli rastgele değişkenler genelde bir aletle ölçülen büyüklüklerdir:

- arabanın hızı
- kişinin kilosu, boyu
- ortamın sıcaklığı...

حتى الآن، تعاملنا مع المتغير العشوائي المنفصل. أي أن القيم التي يمكن للمتغير العشوائي أن يأخذها كانت محدودة. على سبيل المثال، في تجربة رمي قطعة نقود ثلاث مرات، كانت القيم التي يمكن للمتغير العشوائي X ، الذي يمثل عدد مرات ظهور الصورة، أن يأخذها هي فقط 0، 1، 2، أو 3. لم يكن بالإمكان أن يكون هناك أي قيم أخرى. الآن سنتعامل مع نوع من المتغير العشوائي الذي يمكنه أخذ أي قيمة ضمن نطاق معين (أي عدد لا نهائي من القيم). هذا النوع من المتغيرات العشوائية يسمى المتغير العشوائي المستمر. عادةً ما تكون المتغيرات العشوائية المستمرة مرتبطة بالكميات التي يتم قياسها بواسطة أداة، مثل:

- سرعة السيارة
- وزن الشخص أو طوله
- درجة حرارة البيئة...

Rastgele Değişkenler ve Sürekli Dağılımlar

Olasılık Yoğunluk Fonksiyonu (Probability Density Function - PDF)

sürekli bir rastgele değişkenin belirli bir değeri alma olasılığını modellemek için kullanılan bir fonksiyondur. Sürekli değişkenler sonsuz sayıda olası değer alabileceği için, herhangi bir tek değer olasılığı sıfırdır. Bu nedenle PDF, belirli bir aralık içinde bir değer olasılığını bulmak için kullanılır.

$$p_X : X \rightarrow [0,1]$$

Bu fonksiyon aslında daha önce gösterdiğimiz $P(X = x)$ in kendisidir, yani X rastgele değişkenin x 'değerini alma olasılığını gösterir:

$$p_X(x) = P(X = x)$$

دالة الكثافة الاحتمالية (Probability Density Function - PDF)

هي دالة تُستخدم لنمذجة احتمال أن يأخذ المتغير العشوائي المستمر قيمة معينة. وبما أن المتغيرات المستمرة يمكن أن تأخذ عددًا لا نهائيًا من القيم المحتملة، فإن احتمال أخذ أي قيمة مفردة يكون صفرًا. لذلك، يُستخدم PDF (دالة الكثافة الاحتمالية) لحساب احتمال وقوع قيمة المتغير في نطاق محدد.

Rastgele Değişkenler ve Sürekli Dağılımlar

Olasılık Yoğunluk Fonksiyonunun Özellikleri

1- Pozitiflik: PDF, herhangi bir x değeri için negatif olamaz.

Yani $f(x) \geq 0$ olmalıdır.

2- Bütün Olasılıkların Toplamı 1'dir: Bir rastgele değişkenin tüm olası değerleri için olasılık yoğunluk fonksiyonunun toplamı (integrali) 1 olmalıdır. Bu, tüm olası olayların gerçekleşme olasılığının 1 olduğunu ifade eder.

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

3- Belirli Bir Aralık İçindeki Olasılık: Bir sürekli rastgele değişkenin a ve b aralığı içinde bir değer alma olasılığı, bu aralıktaki PDF'nin integrali alınarak hesaplanır.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

دالة الكثافة الاحتمالية (Probability Density Function - PDF)

1- الإيجابية: لا يمكن أن تكون PDF سالبة لأي قيمة x . بمعنى أنه يجب أن يكون $f(x) \geq 0$

2- مجموع جميع الاحتمالات هو 1: يجب أن يكون مجموع (تكامل) دالة الكثافة الاحتمالية لجميع القيم الممكنة للمتغير العشوائي يساوي 1. وهذا يعبر عن أن احتمال وقوع جميع الأحداث الممكنة هو 1.

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

3- الاحتمال ضمن نطاق معين: يتم حساب احتمال أن يأخذ المتغير العشوائي المستمر قيمة ضمن النطاق a و b عن طريق أخذ تكامل PDF في هذا النطاق.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Rastgele Değişkenler ve Sürekli Dağılımlar

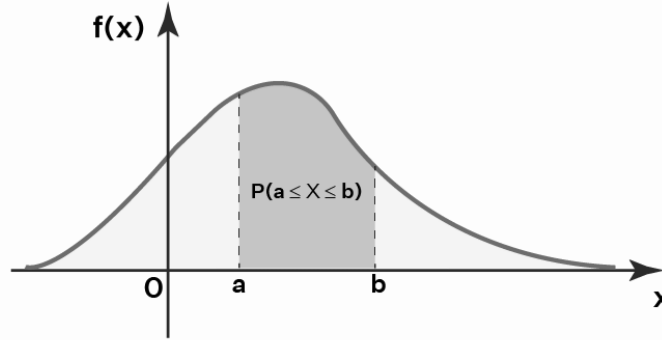
Olasılık Yoğunluk Fonksiyonunun Özellikleri

4- Bir Noktadaki Olasılık: Sürekli rastgele değişkenler için, bir noktadaki olasılık sıfırdır. Yani, $P(X=a)=0$ Bunun yerine, belirli bir aralıkta değer alma olasılığı hesaplanır.

دالة الكثافة الاحتمالية (Probability Density Function - PDF)

4- الاحتمال عند نقطة: بالنسبة للمتغيرات العشوائية المستمرة، فإن الاحتمال عند نقطة معينة يساوي صفرًا. أي أن $P(X=a)=0$ وبدلاً من ذلك، يتم حساب احتمال وقوع المتغير ضمن نطاق محدد.

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

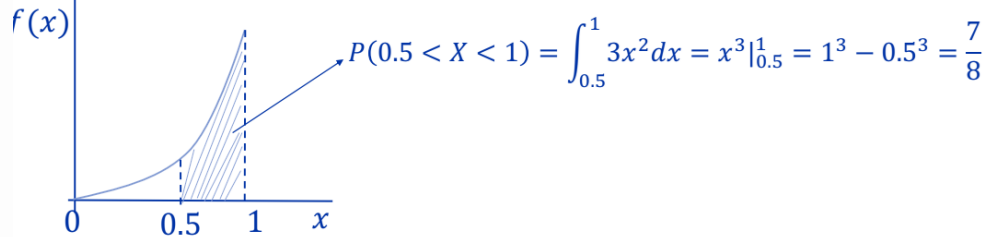


Rastgele Değişkenler ve Sürekli Dağılımlar

Örnek:

ör. X rastgele değişkenin olasılık yoğunluk fonksiyonu $f(x) = 3x^2$; ve X 'in alabilceği değerler $[0,1]$ aralığında olsun. Bu durumda $P(0.5 < X < 1)$ olasılığı ne olur:

Çözüm:



Bir önceki örnek için $P(X = 0.5) = ?$

$$P(X = 0.5) = \int_{0.5}^{0.5} 3x^2 dx = x^3 \Big|_{0.5}^{0.5} = 0.5^3 - 0.5^3 = 0$$

Sonuç X bir sürekli rastgele değişken iken X 'in belirli tek bir değeri alma olasılığı her zaman 0'dır.

Rastgele Değişkenler ve Sürekli Dağılımlar

Normal Dağılım (Çan Eğrisi):

Normal dağılım, en yaygın kullanılan sürekli dağılımlardan biridir. PDF'si şu şekildedir:

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

- μ ortalama (beklenen değer)
- σ standart sapma Bu fonksiyon

التوزيع الطبيعي (منحنى الجرس):

التوزيع الطبيعي هو أحد أكثر التوزيعات المستمرة استخدامًا. دالة الكثافة الاحتمالية PDF له تكون على الشكل التالي:

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

- μ المتوسط (القيمة المتوقعة)

- σ الانحراف المعياري

ويمثل هذا التوزيع البيانات التي تميل للتجمع حول المتوسط مع انتشار يعتمد على الانحراف المعياري.

Rastgele Değişkenler ve Sürekli Dağılımlar

Normal Dağılım (Çan Eğrisi):

X ' rastgele değişkeni ortalaması (μ) ve standart sapması (σ) ile karakterize olur.

Standart sapma (σ) , X 'in merkezi μ 'den ne kadar uzağa yayıldığını gösterir.

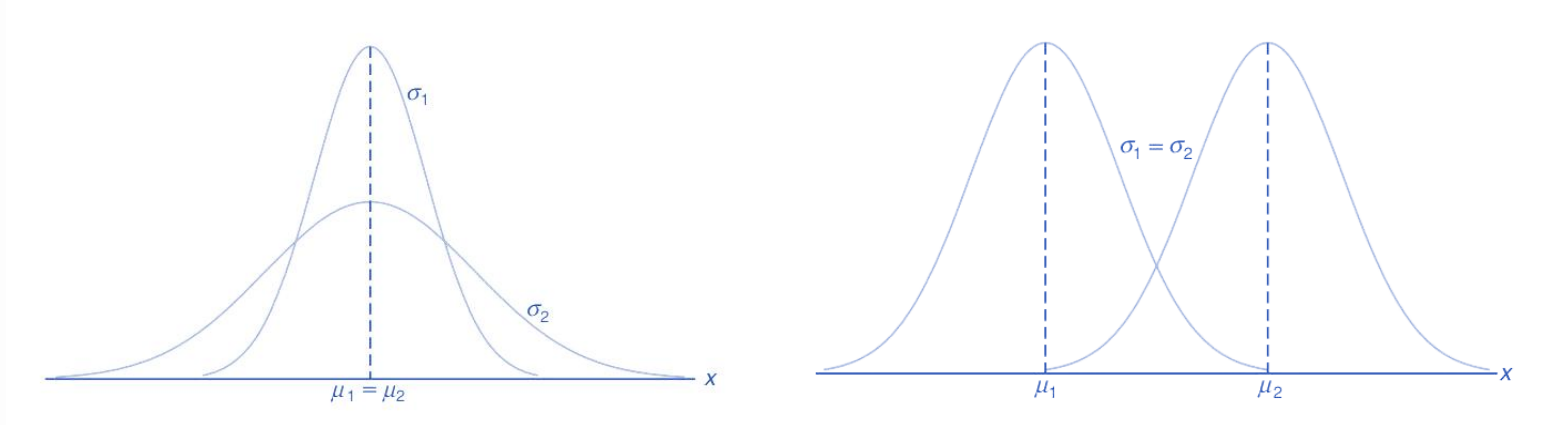
X ve Y ortalamaları aynı, fakat standart sapmaları farklı farklı olsun. Y 'nin standart sapması σ_2 , X 'nin standart sapması σ_1 'den büyük olsun: $\sigma_1 < \sigma_2$. Bu durumda aşağıda soldaki grafik elde edilir. Eğer standart sapmaları aynı, ortalamaları $\mu_1 < \mu_2$ olacak şekilde farklı ise sağdaki grafik elde edilir.

التوزيع الطبيعي (منحنى الجرس):

يتميز المتغير العشوائي X بمتوسط (μ) وانحراف معياري (σ). يمثل الانحراف المعياري (σ) مدى انتشار قيم X بعيداً عن المركز μ . إذا كان لكل من X و Y نفس المتوسط ولكن انحرافاتهما المعيارية مختلفة، وكان الانحراف المعياري لـ Y (σ_2) أكبر من الانحراف المعياري لـ X (σ_1) بحيث يكون: $\sigma_1 < \sigma_2$ ، فسيتم الحصول على الرسم البياني الموضح على اليسار. أما إذا كانت الانحرافات المعيارية متماثلة ولكن المتوسطات مختلفة بحيث يكون $\mu_1 < \mu_2$ فسيتم الحصول على الرسم البياني الموضح على اليمين.

Rastgele Değişkenler ve Sürekli Dağılımlar

Normal Dağılım (Çan Eğrisi):



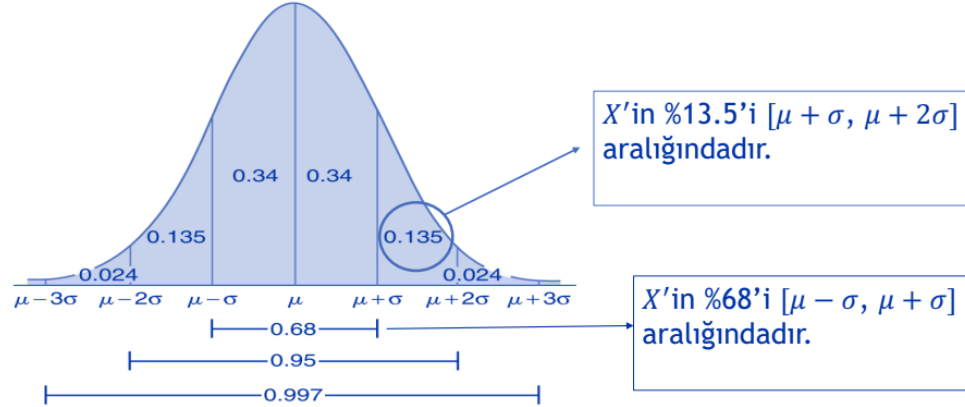
Rastgele Değişkenler ve Sürekli Dağılımlar

Normal Eğrinin Altında Kalan Alan

X normal sürekli rastgele değişkeninin olasılık yoğunluk fonksiyonun grafiğine normal eğri diyeceğiz. Bu eğrinin altında kalan alanı kullanarak, normal dağılıma sahip bir rastgele değişkenin belirli bir aralıkta değer almasının olasılığını

مساحة تحت منحنى التوزيع الطبيعي

سنطلق على الرسم البياني لدالة الكثافة الاحتمالية للمتغير العشوائي المستمر X اسم المنحنى الطبيعي. من خلال استخدام المساحة تحت هذا المنحنى، يمكننا حساب احتمال أن يأخذ المتغير العشوائي ذي التوزيع الطبيعي قيمة ضمن نطاق محدد.



Rastgele Değişkenler ve Sürekli Dağılımlar

Standard Normal Dağılım:

Bir X normal rastgele değişkeni, standart normal rastgele dönüştürülürken, X' 'den önce μ çıkartılır ardından bölüm satandart sapmaya σ bölünür.

$$Z = \frac{X - \mu}{\sigma}$$

التوزيع الطبيعي القياسي:

عند تحويل المتغير العشوائي X إلى متغير عشوائي طبيعي قياسي، يتم اتباع الخطوات التالية:

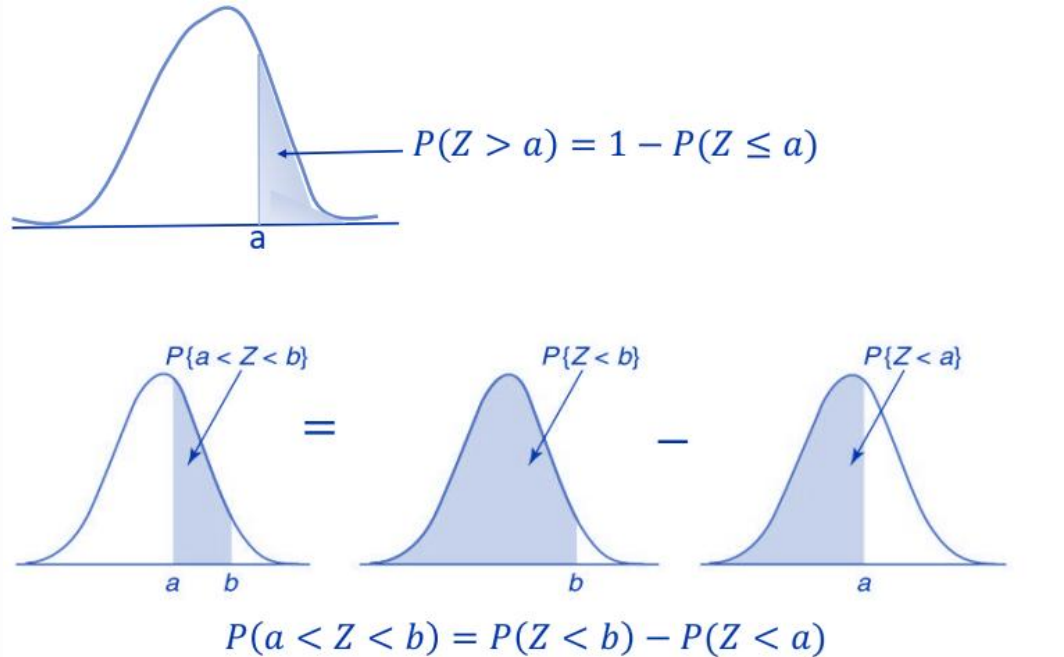
- 1- إزالة المتوسط: يُطرح المتوسط μ من المتغير X
 - 2- القسمة على الانحراف المعياري: يُقسم الناتج على الانحراف المعياري σ
- يمكن التعبير عن هذه العملية رياضياً بالشكل التالي:

$$Z = \frac{X - \mu}{\sigma}$$

حيث Z هو المتغير العشوائي الطبيعي القياسي. بفضل هذه العملية، يتم تحويل المتغير X إلى متغير له متوسط 0 وانحراف معياري 1، مما يسمح باستخدام جداول التوزيع الطبيعي القياسي لتحديد الاحتمالات.

Rastgele Değişkenler ve Sürekli Dağılımlar

Standard Normal Dağılım:



07



Veri Analizi

Eng: Abdulrahman Hamdi

Veri Analizi

İstatistik

İstatistiği iki başlık altında toplamıştık:
Tanımlayıcı İstatistik ve Çıkarımsal İstatistik.

Tanımlayıcı İstatistik (Descriptive Statistics):

Elimizdeki veriyi/numuneyi (sample'ı) tanımlar. Merkezi eğilim (mod, ortalama) ve dağınıklığı ölçer.

Çıkarımsal İstatistik (Inferential Statistics):

Elimizdeki verinin toplandığı popülasyon ile ilgili (büyük resim ile ilgili) çıkarım yapmamızı sağlar.

Bu haftaya kadar tanımlayıcı istatistik ile ilgilendik, elimizdeki veriyi anlamaya çalıştık. Şimdi ise çıkarımsal istatistik ile ilgileneceğiz; verinin toplandığı popülasyon hakkında bir çıkarım yapmaya çalışacağız.

الإحصاء

لقد قسمنا الإحصاء إلى قسمين: الإحصاء الوصفي والإحصاء الاستدلالي.

الإحصاء الوصفي Descriptive Statistics

يصف البيانات/العينة التي لدينا. يقيس الاتجاه المركزي (الوضع، المتوسط) والتشتت.

الإحصاء الاستدلالي Inferential Statistics

يساعدنا على إجراء استنتاجات حول المجتمع الذي تم جمع البيانات منه (الصورة الكبيرة).

حتى هذا الأسبوع، كنا نركز على الإحصاء الوصفي وحاولنا فهم البيانات التي لدينا. أما الآن، سننتقل إلى الإحصاء الاستدلالي وسنحاول التوصل إلى استنتاجات حول المجتمع الذي تم جمع البيانات منه.

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Popölasyonun her bir bireyine bir sayı eşleyelim. Bu sayı, örneğin bir kişinin boyu, kilosu, vücut sıcaklığı ya da bir fabrikada üretilen bir malın raf ömrü olabilir. Bu sayıları X_i rastgele değişkeni ile gösterelim (i . bireyin sayısal değeri). Bu rastgele değişkenlerin ortalaması μ ve standart sapması σ olsun (bunlar popölasyonun parametreleridir). Bu popölasyondan aynı büyüklükte k tane rastgele örnekler (sample'lar) alalım. Örneğin, bu büyüklük n olsun. Bu örneklerin ortalamalarını $\bar{x}_i$ ile gösterelim ($\bar{x}_i$ i .örneğin ortalamasıdır).

نظرية النهاية المركزية (Central Limit Theorem)

لنقم بتخصيص رقم لكل فرد من أفراد المجتمع. قد يكون هذا الرقم، على سبيل المثال، طول الشخص، وزنه، درجة حرارة جسمه، أو مدة صلاحية منتج ما تم تصنيعه في مصنع. دعنا نرمز لهذه الأرقام بالمتغير العشوائي X_i (القيمة العددية للفرد i). ليكن متوسط هذه المتغيرات العشوائية μ وانحرافها المعياري σ (هذه هي معايير المجتمع). لنأخذ من هذا المجتمع k عينات عشوائية بنفس الحجم. لنفترض أن حجم هذه العينات هو n . لنرمز بمتوسطات هذه العينات بـ $\bar{x}_i$ حيث $\bar{x}_i$ هو متوسط العينة i .

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneklerin ortalamaları $X_1, X_2, \dots, X_k$ 'da bir rastgele değişkendir ve normal dağılıma sahip olur.

Ortalamaların beklenen değeri populasyonun ortalaması μ 'ye eşit olur: $E(\bar{X}) = \mu$.

Ortalamaların standart sapması populasyonun standart sapmasının $\sqrt{n}$ 'e bölünmesiyle elde edilir: $\frac{\sigma}{\sqrt{n}}$

نظرية النهاية المركزية (Central Limit Theorem)

لنقم بمتوسطات العينات $X_1, X_2, \dots, X_k$ هي أيضاً متغيرات عشوائية وتتبع التوزيع الطبيعي.

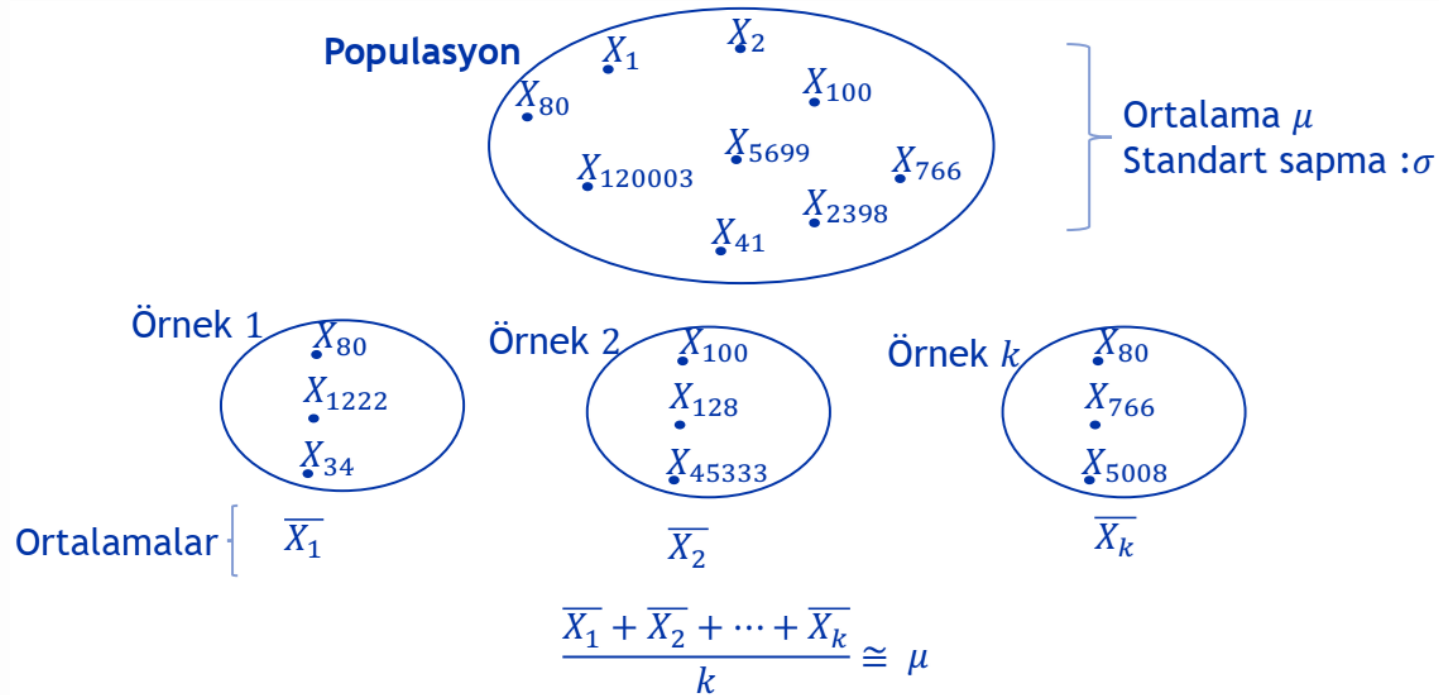
يكون التوقع الرياضي (القيمة المتوقعة) لمتوسطات العينات مساوياً لمتوسط المجتمع $E(\bar{X}) = \mu$

يتم حساب الانحراف المعياري لمتوسطات العينات بقسمة الانحراف المعياري

للمجتمع على الجذر التربيعي لـ n فتصبح المعادلة: $\frac{\sigma}{\sqrt{n}}$

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)



Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, 30 yaşındaki erkeklerin büyük tansiyonlarının ortalaması 120, standart sapması 10 olsun. Bu durumda, rastgele seçilen 9 kişilik bir grubun büyük tansiyonlarının ortalamasının 130'dan fazla olma olasılığı nedir?

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، لنفترض أن متوسط ضغط الدم الانقباضي للرجال البالغين من العمر 30 عامًا هو 120، وانحرافه المعياري هو 10. في هذه الحالة، ما هي احتمالية أن يكون متوسط ضغط الدم الانقباضي لمجموعة عشوائية مكونة من 9 أشخاص أكبر من 130؟

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, 30 yaşındaki erkeklerin büyük tansiyonlarının ortalaması 120, standart sapması 10 olsun. Bu durumda, rastgele seçilen 9 kişilik bir grubun büyük tansiyonlarının ortalamasının 130'dan fazla olma olasılığı nedir?

Çözüm.

30 yaşındaki tüm erkeklerin kümesi populasyonumuz olur. Bu populasyondaki kişilerin büyük tansiyonlarının ortalaması $\mu = 120$ ve standart sapması $\sigma = 10$ olarak verilmiş.

$\bar{X}$ rastgele değişkeni, 30 yaşındaki erkeklerden oluşan 9 kişilik herhangi bir grubun büyük tansiyonlarının ortalaması olsun. Biz $P(\bar{X} > 130)$ olasılığı ile ilgileniyoruz.

Merkezi limit teoremi gereği $\bar{X}$ rastgele değişkenleri normal dağılıma sahiptir.

$$P(\bar{X} > 130) = P\left(Z > \frac{130 - 120}{\frac{10}{\sqrt{9}}}\right) = P(Z > 3) = 1 - P(\leq 3) = 1 - 0.9987 = 0.0013$$

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، لنفترض أن متوسط ضغط الدم الانقباضي للرجال البالغين من العمر 30 عامًا هو 120، وانحرافه المعياري هو 10. في هذه الحالة، ما هي احتمالية أن يكون متوسط ضغط الدم الانقباضي لمجموعة عشوائية مكونة من 9 أشخاص أكبر من 130؟

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, bir fabrikada çalışan işçilerin kolesterol seviyelerinin ortalaması 202, standart sapması 14 olsun.

- i) İşçilerden 36 kişilik bir grup oluşturulduğunda, bu grubun kolesterol seviyelerinin 198 ile 206 arasında olma olasılığı nedir?
- ii) İşçilerden 64 kişilik bir grup oluşturulduğunda, bu grubun kolesterol seviyelerinin yine 198 ile 206 arasında olma olasılığı nedir?

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، على سبيل المثال، لنفترض أن متوسط مستويات

الكوليسترول للعمال في مصنع ما هو 202، وانحرافه المعياري هو 14.

أ- عندما يتم تشكيل مجموعة من 36 عاملاً، ما هي احتمالية أن تكون مستويات الكوليسترول في هذه المجموعة بين 198 و206؟

ب- عندما يتم تشكيل مجموعة من 64 عاملاً، ما هي احتمالية أن تكون مستويات الكوليسترول في هذه المجموعة أيضاً بين 198 و206؟

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, bir fabrikada çalışan işçilerin kolesterol seviyelerinin ortalaması 202, standart sapması 14 olsun.

- i) İşçilerden 36 kişilik bir grup oluşturulduğunda, bu grubun kolesterol seviyelerinin 198 ile 206 arasında olma olasılığı nedir?
- ii) ii) İşçilerden 64 kişilik bir grup oluşturulduğunda, bu grubun kolesterol seviyelerinin yine 198 ile 206 arasında olma olasılığı nedir?

Çözüm.

- i) 36 kişilik grupların kolestrol seviyelerinin ortalamasını gösteren $\bar{X}$ rastgele değişkeni merkezi limit teoremi gereği normal dağılıma sahip olur. O halde

$$P(198 \leq \bar{X} \leq 206) = P\left(\frac{198 - 202}{14/\sqrt{36}} \leq Z \leq \frac{206 - 202}{14/\sqrt{36}}\right) = P(-1.714 \leq Z \leq 1.714) \\ = P(Z \leq 1.714) - P(Z \leq -1.714) = 0.9564 - 0.0436 = 0.9128$$

$$ii) P(198 \leq \bar{X} \leq 206) = P\left(\frac{198-202}{14/\sqrt{64}} \leq Z \leq \frac{206-202}{14/\sqrt{64}}\right) = P(-2.286 \leq Z \leq 2.286) = 0.995$$

Sonuç olarak grubun büyüklüğü artırıldığında grubun ortalaması popülasyon ortalamasına yaklaşır.

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، على سبيل المثال، لنفترض أن متوسط مستويات الكوليسترول للعمال في مصنع ما هو 202، وانحرافه المعياري هو 14. أ- عندما يتم تشكيل مجموعة من 36 عاملاً، ما هي احتمالية أن تكون مستويات الكوليسترول في هذه المجموعة بين 198 و206؟
ب- عندما يتم تشكيل مجموعة من 64 عاملاً، ما هي احتمالية أن تكون مستويات الكوليسترول في هذه المجموعة أيضًا بين 198 و206؟

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, pil üreten bir fabrikanın ürettiği pillerin ömrünün ortalaması μ , standart sapması 100 olsun. Rastgele alınan 10 pilin ömürlerinin ortalamasının μ 'den 20 fazla olma olasılığı ne olur?

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، لنفترض أن متوسط عمر البطاريات التي ينتجها مصنع البطاريات هو μ ، وانحرافه المعياري هو 100. ما هي احتمالية أن يكون متوسط عمر 10 بطاريات تم اختيارها عشوائياً أكبر من μ بمقدار 20؟

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, pil üreten bir fabrikanın ürettiği pillerin ömrünün ortalaması μ , standart sapması 100 olsun. Rastgele alınan 10 pilin ömürlerinin ortalamasının μ 'den 20 fazla olma olasılığı ne olur?

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، لنفترض أن متوسط عمر البطاريات التي ينتجها مصنع البطاريات هو μ ، وانحرافه المعياري هو 100. ما هي احتمالية أن يكون متوسط عمر 10 بطاريات تم اختيارها عشوائياً أكبر من μ بمقدار 20؟

Çözüm.

$\bar{X}$ rastgele değişkeni 10 büyüklüğündeki pil gruplarının ortalama ömrü olsun. Merkezi limit teoremi gereği $\bar{X}$ normal dağılıma sahiptir.

$$\begin{aligned} P(\bar{X} \geq \mu + 20) &= P\left(Z \geq \frac{\mu + 20 - \mu}{\frac{100}{\sqrt{10}}}\right) = P(Z \geq 0.63) = 1 - P(Z < 0.63) \\ &= 1 - 0.7357 \\ &= 0.2643 \end{aligned}$$

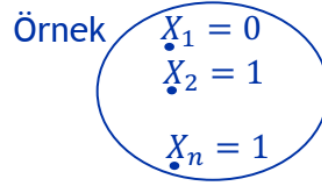
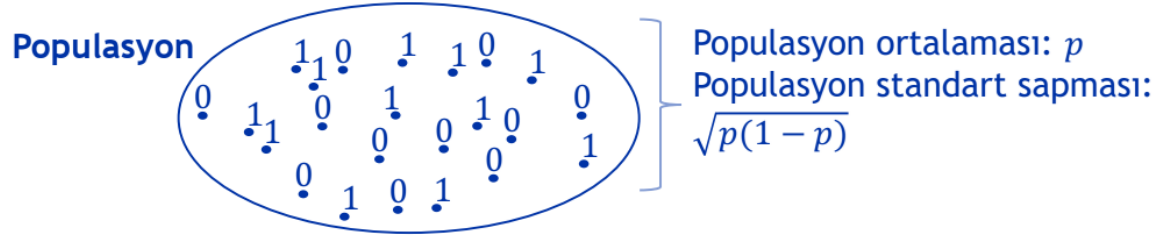
Veri Analizi

Merkezi Limit Teoremi Yardımıyla Normal Dağılım Kullanarak (Binomial Dağılıma Yaklaşmak)

Bernoulli rastgele değişkenlerinden oluşan bir popülasyonumuz olsun. (Yani popülasyonda yalnızca 0 (başarısızlık) ve 1 (başarı) değerleri var. Buradan n elemana sahip bir örnek alalım.

باستخدام نظرية النهاية المركزية لتقريب التوزيع الطبيعي (تقريب التوزيع الثنائي):

لنفترض أن لدينا مجتمعًا يتكون من متغيرات عشوائية برنولية. (أي أن المجتمع يحتوي فقط على القيم 0 (فشل) و 1 (نجاح)). دعنا نأخذ عينة تحتوي على n عنصرًا.



Bu örneğin ortalaması: $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{X}{n}$

Toplam başarı sayısı!

Veri Analizi

**Merkezi Limit Teoremi Yardımıyla Normal Dağılım
Kullanarak (Binomial Dağılıma Yaklaşmak)**

باستخدام نظرية النهاية المركزية لتقريب التوزيع الطبيعي
(تقريب التوزيع الثنائي):

Merkezi limit teoremi gereği $\bar{X} = \frac{X}{n}$ normal dağılıma sahip olur. Bunu standart

normal dağılımdaki karşılığı:
$$Z = \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\frac{X - np}{n}}{\sqrt{\frac{p(1-p)}{n}}} = \frac{X - np}{\sqrt{np(1-p)}}$$

$\frac{X - np}{\sqrt{np(1-p)}}$ yaklaşık olarak normal dağılıma sahip olur. Burada X , toplam başarı sayısını gösterdiğinden binomial rastgele değişken olur.

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, varsayalım ki bir popülasyonun %46'sı A adayını desteklesin. Bu popülasyondan rastgele seçeceğimiz 200 kişinin en az 100'ünün A adayını destekleme olasılığı nedir?

نظرية النهاية المركزية (Central Limit Theorem)

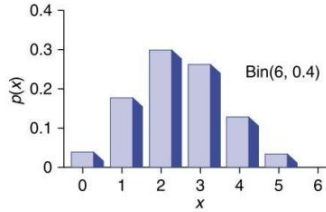
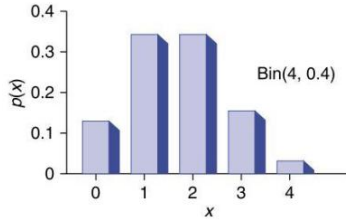
على سبيل المثال، لنفترض أن 46% من السكان يدعمون المرشح A. ما هي احتمالية أن يدعم على الأقل 100 من 200 شخص يتم اختيارهم عشوائيًا المرشح A؟

Veri Analizi

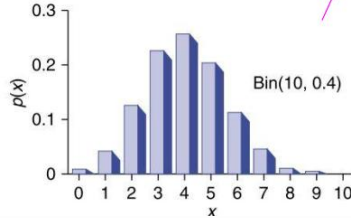
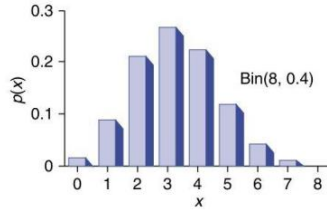
Çözüm.

X rastgele değişkeni seçilen 200 kişiden A'yı destekleyen toplam kişi sayısını gösterebilir. Biz $P(X \geq 100)$ ile ilgileniyoruz.

$$P(X \geq 100) = P\left(Z \geq \frac{100 - 200 \times 0.46}{\sqrt{200 \times 0.46(1 - 0.46)}}\right) = P(Z \geq 1.135) = 1 - P(Z < 1.135) = 0.0885$$



n (yani deney tekrar sayısı) arttıkça binomial dağılım, normal dağılıma benzer.



Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, bir fabrikanın ürettiği malların %10'u bozuk olsun. Bu fabrikanın ürettiği mallardan rastgele seçilen 8 malın:

- i) Hiçbirinin bozuk olmama olasılığı nedir?
- ii) ii) %15'inden fazlasının bozuk olma olasılığı nedir?

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، لنفترض أن 10% من المنتجات التي ينتجها مصنع ما معيبة. ما هي احتمالية أن تكون المنتجات الثمانية التي يتم اختيارها عشوائيًا:

أ- لا توجد أي منها معيبة؟
ب- أكثر من 15% منها معيبة؟

Veri Analizi

Merkezi Limit Teoremi (Central Limit Theorem)

Örneğin, bir fabrikanın ürettiği malların %10'u bozuk olsun. Bu fabrikanın ürettiği mallardan rastgele seçilen 8 malın:

- i) Hiçbirinin bozuk olmama olasılığı nedir?
- ii) ii) %15'inden fazlasının bozuk olma olasılığı nedir?

نظرية النهاية المركزية (Central Limit Theorem)

على سبيل المثال، لنفترض أن 10% من المنتجات التي ينتجها مصنع ما معيبة. ما هي احتمالية أن تكون المنتجات الثمانية التي يتم اختيارها عشوائيًا:
أ- لا توجد أي منها معيبة؟
ب- أكثر من 15% منها معيبة؟

Çözüm.

X rastgele değişkeni bu fabrikanın ürettiği mallardan rastgele seçilen 8 maldaki toplam bozuk sayısı olsun.

$p = 0.1$ başarı oranı

$$i) P(X = 0) = (1 - 0.1)^8 = 0.43$$

$$ii) P\left(X > 8 \times \frac{15}{100}\right) = P(X > 1.2) = P\left(Z > \frac{1.2 - 8 \times 0.1}{\sqrt{8 \times 0.1 \times (0.9)}}\right) = P(Z > 0.471) = 1 - P(Z \leq 0.471) = 1 - 0.68 = 0.32$$

Veri Analizi

Sıklık Dağılımı / Histogram

İstatistikte frekans dağılımı, verilerin belli kategorilere (veya aralıklara) gruplandırılması ve her bir kategori içinde yer alan değerlerin sayılmasını sağlayan bir yöntemdir. Histogram, verilerin dağılımını görsel olarak göstermek için kullanılan bir araçtır.

Histogramın Unsurları:

Kategoriler (veya aralıklar): Veri aralığı belirli kategorilere bölünür.

Bu kategoriler eşit boyutlu veya farklı boyutlarda olabilir.

Frekans: Her bir kategori içinde yer alan değerlerin sayısını temsil eder. Frekanslar, histogramdaki sütunların yüksekliği olarak gösterilir.

Eksenler: Yatay eksen (X), kategorileri temsil ederken, dikey eksen (Y) frekansı temsil eder.

توزيع التكرار / الهستوغرام

توزيع التكرار هو طريقة لعرض البيانات من خلال تجميعها في فئات (أو فواصل) وتعداد عدد القيم التي تقع ضمن كل فئة.

يستخدم الهستوغرام كأداة تصويرية لتوضيح توزيع البيانات بشكل مرئي. عناصر الهستوغرام:

الفئات (أو الفواصل): يتم تقسيم نطاق البيانات إلى فئات محددة. يمكن أن تكون هذه الفئات متساوية الحجم أو متفاوتة.

التردد: يُمثل عدد القيم التي تقع ضمن كل فئة. يتم عرض الترددات كارتفاعات الأعمدة في الهستوغرام.

المحاور: المحور الأفقي X يمثل الفئات، بينما المحور العمودي Y يمثل التردد.

Veri Analizi

Sıklık Dağılımı / Histogram

Histogram Oluşturma Adımları:

Veri Toplama: Analiz etmek istediğiniz veri kümesini toplayın.

Kategorileri Belirleme: Değer aralığını uygun kategorilere ayırın.

Frekansları Hesaplama: Her bir kategori içindeki değerlerin sayısını hesaplayın.

Histogramı Çizme: Her kategori için frekansları temsil etmek üzere sütunları kullanın.

Histogramın Kullanımları:

Veri dağılımını görselleştirmek.

Veri kalıplarını belirlemek, örneğin simetrik veya dağılmış olup olmadığını anlamak.

Aykırı değerleri veya sapmaları tespit etmek.

Histogram, verileri görsel bir şekilde anlamaya yardımcı olabilir, böylece analiz yapmayı ve karar verme sürecini kolaylaştırır.

توزيع التكرار / الهستوغرام

خطوات إنشاء الهستوغرام:

جمع البيانات: اجمع مجموعة البيانات التي ترغب في تحليلها.

تحديد الفئات: قم بتقسيم نطاق القيم إلى فئات مناسبة.

حساب الترددات: احسب عدد القيم التي تقع ضمن كل فئة.

رسم الهستوغرام: استخدم الأعمدة لتمثيل الترددات لكل فئة.

استخدامات الهستوغرام:

تصور توزيع البيانات.

تحديد أنماط البيانات، مثل التماثل أو التشتت.

اكتشاف القيم الشاذة أو الانحرافات.

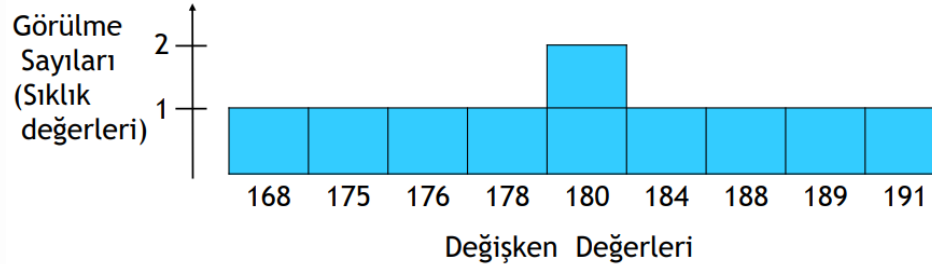
يمكن أن يساعد الهستوغرام في فهم البيانات بطريقة بصرية، مما يسهل تحليلها واتخاذ القرارات.

Veri Analizi

Sıklık Dağılımı / Histogram

Peki ya verideki değişkenimiz sayısal ise? Bu durumda histogram nasıl olur?

Yine diyelim ki 10 kişilik bu arkadaş grubundaki kişilerin boyları cm cinsinden şöyle olsun: 191, 168, 176, 175, 188, 180, 184, 180, 189, 178. Bu durumda 180 cm iki kez, diğer uzunluklar bir kez görülmüş olur. O halde histogram şöyle olacaktır:



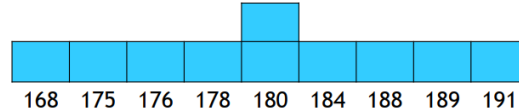
توزيع التكرار / الهستوغرام

إذا كانت المتغيرات في البيانات عددية، فإن الهستوغرام سيكون مشابهًا، ولكن سيتطلب معالجة القيم العددية بشكل خاص. إليك كيفية إنشاء الهستوغرام باستخدام بيانات الطول (بالسنتيمتر) لأفراد مجموعة مكونة من 10 أشخاص: بيانات الطول: 191، 168، 176، 175، 188، 180، 184، 180، 189، 178

Veri Analizi

Sıklık Dağılımı / Histogram

Bu histograma bakarak gruptaki kişilerin boyları hakkında hemen bir kanıya varmak ve bir sonuç çıkarmak güçtür. Değişken sayısal iken, bu değişkenin daha çok hangi değerleri aldığını göstermek için, yani bu değerlerin histogramı için genel yaklaşım, değerleri gruplara bölmektir. Buna İngilizcede "binning" (kutulamak) denir.



Değişken sayısal iken, bu değişkenin daha çok hangi değerler aldığını göstermek için, yani bu değerlerin histogramı için, genel yaklaşım değeri gruplara bölmektir: buna İngilizcede (binning, (kutulamak)) denir.

Bu veriden 3 grup oluşturalım:

165 - 175 arası değerler: 168, 175 (2 adet)

175 - 185 arası değerler: 176, 178, 180, 184 (5 adet)

185 - 195 arası değerler: 188, 189, 191 (3 adet)

Not : Yukarıdaki gruplama tamamen kişiseldir. İsteyen değişkenleri başka aralıklarla da gruplayabilir, ama grup aralıklarının aynı olmasına dikkat edilmelidir!

توزيع التكرار / الهستوغرام

عند النظر إلى هذا الهستوغرام، من الصعب التوصل إلى استنتاجات حول أطوال الأشخاص في المجموعة على الفور. عندما يكون المتغير عدديًا، يتم استخدام تقسيم هذا المتغير إلى مجموعات لإظهار القيم التي يأخذها بشكل أكبر، وهذا ما يعرف في الإنجليزية بـ "binning" أو التقسيم إلى صناديق.

Bu histograma bakarak gruptaki kişilerin boyları hakkında hemen bir kanıya varmak, bir sonuç çıkarmak güçtür.

Veri Analizi

Görelî Sıklık Dağılımı (Relative Frequency Distribution)

Veride her bir değerin görölme sayısını dikkate aldığımızda, ortaya çıkan sıklık dağılımı o veriye özgü (elimizdeki örneğe) gibi görülebilir. Verinin alındığı popölasyona yönelik daha genel bir sonuca varmak için, görölme sayıları normalleştirilerek her bir sayının [0-1] arası bir değeri alması sağlanabilir. Bu şekilde oluşturulan sıklık dağılımına "görelî sıklık dağılımı" denir.

Bir görölme sayısını [0-1] aralığına getirmek için, bu sayıyı toplam görölme sayısına (verideki toplam eleman sayısı) böleriz.

Bir önceki örnekte toplam 10 kişi vardı. Bu 10 kişinin 2'sinin boyu 165-175 arasında idi. Bu durumda, 165-175 cm aralığındaki görelî sıklık şöyle hesaplanır:

$$\text{Görelî Sıklık} = \frac{\text{Bu aralıktaki görölme sayısı}}{\text{Toplam görölme sayısı}} = \frac{2}{10} = 0.2$$

Yani, boyu 165-175 cm arasında olan kişilerin görelî sıklığı %20'dir.

توزيع التكرار النسبي

عند أخذ عدد مرات ظهور كل قيمة في البيانات بعين الاعتبار، يبدو أن توزيع التكرار الناتج خاص بتلك البيانات (بالنسبة للعينة الموجودة لدينا). للوصول إلى استنتاج أكثر عمومية بشأن المجتمع الذي تم أخذ البيانات منه، يمكن تطبيع أعداد مرات الظهور، بحيث تأخذ كل قيمة بين [0-1]. يُعرف توزيع التكرار الذي يتم إنشاؤه بهذه الطريقة بـ "توزيع التكرار النسبي".

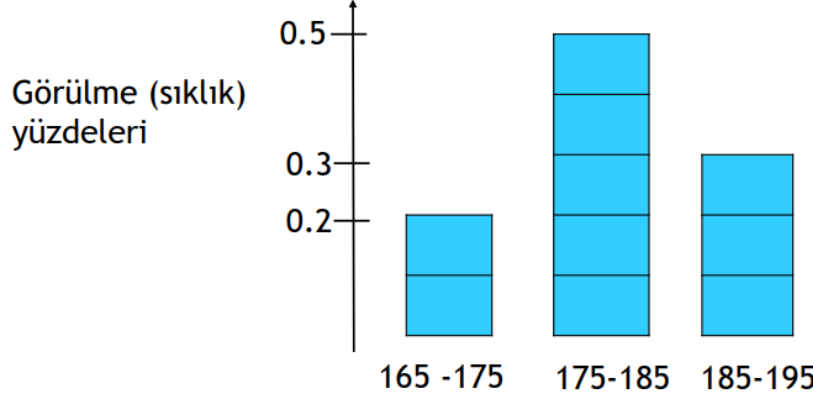
لتحويل عدد مرات الظهور إلى نطاق [0-1]، نقوم بقسمة هذا العدد على إجمالي عدد مرات الظهور (إجمالي عدد العناصر في البيانات). في المثال السابق، كان هناك إجمالي 10 أشخاص. وكان طول شخصين من هؤلاء الأشخاص بين 165-175 سم. في هذه الحالة، يتم حساب توزيع التكرار النسبي للمدى 165-175 سم كما يلي:

Veri Analizi

Görelî Sıklık Dağılımı (Relative Frequency Distribution)

توزيع التكرار النسبي

Histogramı görölme yüzdelerini kullanarak yaparsak:



Soru 2: Aşağıdaki veri, bir şirkette çalışan 50 kişinin son 6 haftadaki izinli gün sayılsını göstermektedir. Bu veriye ait histogramı görölme yüzdelerini gösterecek şekilde çiziniz.

2, 2, 0, 0, 5, 8, 3, 4, 1, 0, 0, 7, 1, 7, 1, 5, 4, 0, 4, 0, 1, 8, 9, 7, 0,
1, 7, 2, 5, 5, 4, 3, 3, 0, 0, 2, 5, 1, 3, 0, 1, 0, 2, 4, 5, 0, 5, 7, 5, 1

Veri Analizi

Sıklık Dağılımı Ne Zaman Normaldir?

İdeal dünyada veri, merkezi etrafında toplanmıştır. Bu, "normal" olandır. Verinin bu şekilde merkezi etrafında toplanmasına normal dağılım denir. Normal dağılıma sahip bir verinin histogramı çan eğrisi şeklinde olur.

Normal dağılım, veri setinin ortalaması etrafında simetrik bir dağılım sergiler; bu da, çoğu verinin ortalama değerine yakın olduğunu ve daha uç değerlerin daha az sıklıkta bulunduğunu gösterir. Histogramda bu durum, çan biçiminde bir eğri ile temsil edilir.

متى يكون توزيع التردد طبيعيًا؟

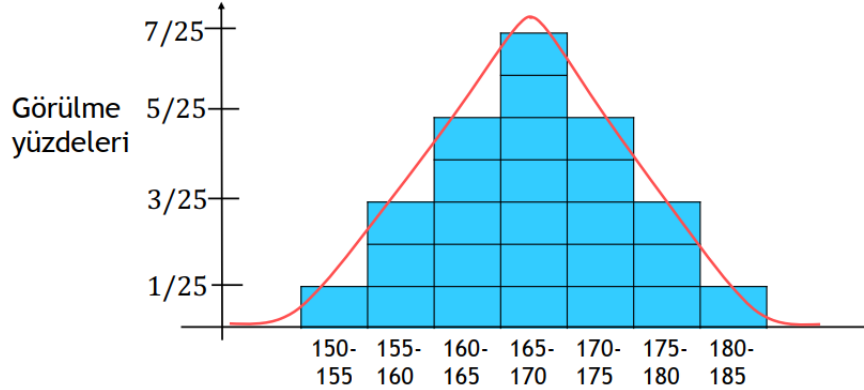
في عالم مثالي، تتجمع البيانات حول المركز. وهذا هو ما يُعتبر "طبيعيًا". يُعرف تجمع البيانات حول المركز بهذا الشكل بالتوزيع الطبيعي. يكون لهستوغرام البيانات التي تتمتع بتوزيع طبيعي شكل منحنى الجرس. يعرض التوزيع الطبيعي توزيعًا متماثلًا حول متوسط مجموعة البيانات؛ مما يعني أن معظم البيانات قريبة من القيمة المتوسطة وأن القيم الشاذة أقل تكرارًا. في الهستوغرام، يتم تمثيل هذه الحالة بمنحنى على شكل جرس.

Veri Analizi

Sıklık Dağılımı Ne Zaman Normaldir?

İdeal dünyada veri, merkezi etrafında toplanmıştır. Bu, "normal" olandır. Verinin bu şekilde merkezi etrafında toplanmasına normal dağılım diyeceğiz. Normal dağılıma sahip bir verinin histogramı çan eğrisi şeklinde olur.

ör.



متى يكون توزيع التردد طبيعياً؟

في عالم مثالي، تتجمع البيانات حول المركز. وهذا هو ما يُعتبر "طبيعياً". يُعرف تجمع البيانات حول المركز بهذا الشكل بالتوزيع الطبيعي. يكون لهستوغرام البيانات التي تتمتع بتوزيع طبيعي شكل منحنى الجرس. يعرض التوزيع الطبيعي توزيعاً متماثلاً حول متوسط مجموعة البيانات؛ مما يعني أن معظم البيانات قريبة من القيمة المتوسطة وأن القيم الشاذة أقل تكراراً. في الهستوغرام، يتم تمثيل هذه الحالة بمنحنى على شكل جرس.

Veri Analizi

Sıklık Dağılımı Ne Zaman Normaldir?

Ortalama boy olan 165–170 cm aralığındaki kişilerin sayısı en fazladır. Uç değerlerdeki boylara, yani 150–155 cm ve 180–185 cm sahip kişilerin sayıları en azdır. Grafikte ortalama boyun sağında ve solunda simetri vardır.

Sonuç olarak, eğer sıklık dağılımı normalse, veri merkezi etrafında toplanmıştır.

Bu şu demektir:

En fazla ortalama değer görülür.

Uç değerler (en küçük değerler ve en büyük değerler) en az görülür.

Ortalama değer altındaki ve üstündeki değerlerin görülme sayısı hemen hemen aynıdır.

متى يكون توزيع التردد طبيعيًا؟

إن عدد الأشخاص في نطاق الطول المتوسطة، الذي يتراوح بين 165-170 سم، هو الأكبر. بينما يكون عدد الأشخاص الذين لديهم أطوال في القيم الشاذة، مثل 150-155 سم و 180-185 سم، هو الأقل. في الرسم البياني، توجد تناظرية على يمين ويسار متوسط الطول.

نتيجة لذلك، إذا كان توزيع التكرار طبيعيًا، فهذا يعني أن البيانات تتجمع حول المركز.

وهذا يعني:

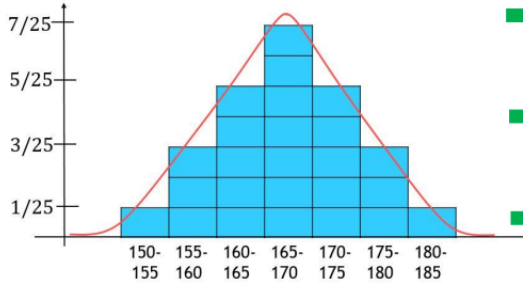
تُرى القيم المتوسطة في أغلب الأحيان.

تُرى القيم الشاذة (أقل القيم وأعلى القيم) في أقل الأحيان.

عدد مرات ظهور القيم الموجودة تحت ومتوسط القيمة وفوقها تقريبًا متساوي.

Veri Analizi

Sıklık Dağılımı Ne Zaman Normaldir?



→ Ortalama boy olan 165-170 boya sahip kişilerin sayısı en fazladır.

→ Uç değerdeki boylara: 150-155 ve 180-185 sahip kişilerin sayıları en azdır.

→ Grafikte ortalama boyun sağ ve solu simetriktir.

Sonuç olarak; eğer sıklık dağılımı normalse veri merkezi etrafına toplanmıştır.

Bu şu demektir:

- En fazla ortalama değer görülür.
- Uç değerler (en küçük değerler, en büyük değerler) en az görülür.
- Ortalama değer altındaki ve üstündeki değerlerin görülme sayısı hemen hemen aynıdır.

متى يكون توزيع التردد طبيعياً؟

إن عدد الأشخاص في نطاق الطول المتوسطة، الذي يتراوح بين 165-170 سم، هو الأكبر. بينما يكون عدد الأشخاص الذين لديهم أطوال في القيم الشاذة، مثل 150-155 سم و 180-185 سم، هو الأقل. في الرسم البياني، توجد تناظرية على يمين ويسار متوسط الطول. نتيجة لذلك، إذا كان توزيع التكرار طبيعياً، فهذا يعني أن البيانات تتجمع حول المركز.

وهذا يعني:

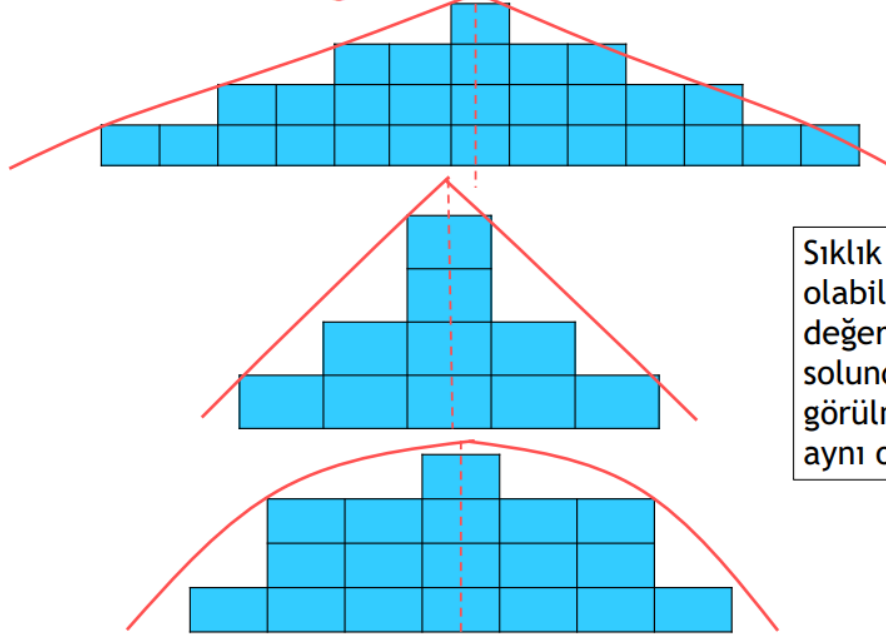
تُرى القيم المتوسطة في أغلب الأحيان.

تُرى القيم الشاذة (أقل القيم وأعلى القيم) في أقل الأحيان.

عدد مرات ظهور القيم الموجودة تحت ومتوسط القيمة وفوقها تقريباً متساوي.

Veri Analizi

Farklı Normal Sıklık Dağılımları

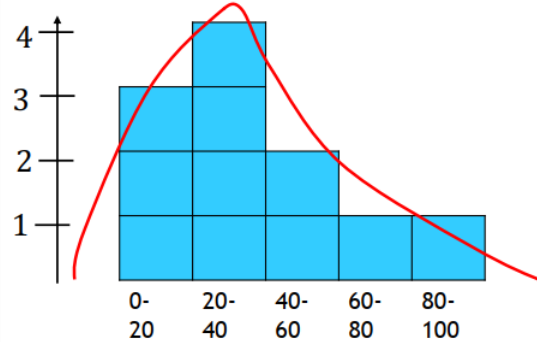


Sıklık dağılımının normal olabilmesi için ortalama değerin sağındaki ve solundaki değerlerin görülme sayılarının aynı olması gerekir.

Veri Analizi

Sıklık Dağılımının Normalden Sapması

Bazen en sık görülen değer ortalama değer olmayabilir. Örneğin, küçük değerler çok daha sık görülebilir. Örneğin, 11 kişilik bir arkadaş topluluğunda istatistik dersinden alınan vize notları şöyle olsun: 70, 50, 5, 45, 35, 17, 25, 27, 90, 5, 38. Bu değerleri 20'lik aralıklara böldüğümüzde: 0-20 arası: 3 değer 20-40 arası: 4 değer 40-60 arası: 2 değer 60-80 arası: 1 değer 80-100 arası: 1 değergörülür.



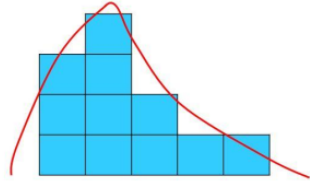
انحراف توزيع التردد عن الطبيعي

أحياناً قد لا يكون أكثر القيم تكراراً هو المتوسط. على سبيل المثال، قد تظهر القيم الصغيرة بشكل متكرر أكثر. على سبيل المثال، في مجموعة من 11 شخصاً، تكون الدرجات المستلمة في امتحان الإحصاء كالتالي: 70، 50، 5، 45، 35، 17، 25، 27، 90، 5، 38. عندما نقسم هذه القيم إلى نطاقات تبلغ 20: بين 0-20: 3 قيمين 20-40: 4 قيمين 40-60: 2 قيمين 60-80: 1 قيمة 80-100: 1 قيمة هذا ما يمكن رؤيته.

Veri Analizi

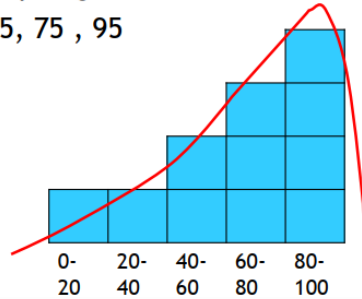
Sıklık Dağılımının Normalden Sapması

انحراف توزيع التردد عن الطبيعي



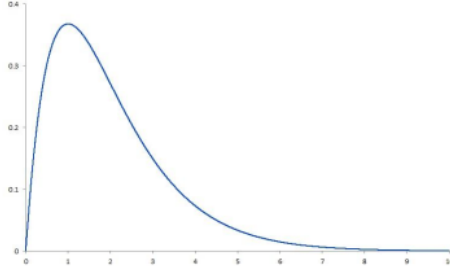
Bu şekilde küçük değerlerin daha çok görüldüğü dağılıma pozitif eğrilğe (positively skewed) sahip dağılım diyeceğiz.

Karşıt şekilde veride eğer büyük değerler daha çoksa, bu sıklık dağılımına negatif eğrilğe (negatively skewed) sahip dağılım denir. Bir önceki örnekte notlar eğer :
25, 50, 5, 52, 82, 100, 95, 78, 65, 75 , 95

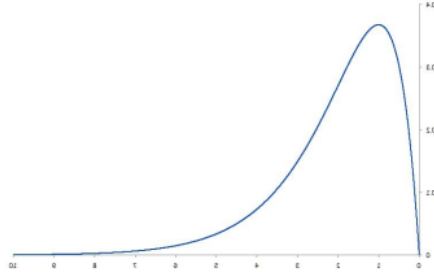


Veri Analizi

Pozitif Eğriliğe Sahip Sıklık Dağılımı



Negatif Eğriliğe Sahip Sıklık Dağılımı



Bu eğriye pozitif denmesinin nedeni kuyruğunun sayı doğrusunda pozitif yöne doğru uzaması; benzer şekilde aşağıdaki eğriye negatif denmesinin nedeni kuyruğunun negatif yöne doğru uzamasıdır.

يُطلق على هذا المنحنى "إيجابي" لأن ذيله يمتد في اتجاه الأعداد الموجبة على خط الأعداد. وبالمثل، يُطلق على المنحنى أدناه "سلبي" لأن ذيله يمتد في الاتجاه السالب.

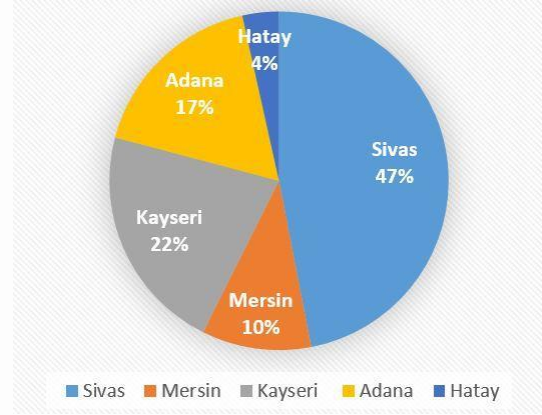
Veri Analizi

Daire Grafiđi (Pie Chart)

İlgilendiđimiz deđiřkenim sembolik türde bir kategorik deđiřken ise (yani bu deđiřkenler sözel ve sıralanamıyorsa) bu deđiřkenin farklı deđerlerinin veride ne sıklıkta görüldüğünü görselleřtirmek için daire grafiđi (pie chart) çizeriz.

ör. Diyelimki bir bölümünde 54 kiři Sivas'lı, 12 kiři Mersin'li, 25 kiři Kayseri'li, 20 kiři Adana'lı ve 4 kiři Hatay'lı olsun.

Bu veriden elde edilen daire grafiđi bölümdeki kiřilerin memleketlerini ve bu memleketlerin görölme yüzdelerini verir:



مخطط دائري

إذا كان المتغير الذي نهتم به متغيرًا فئويًا رمزيًا (أي أن هذه المتغيرات لفظية ولا يمكن ترتيبها)، فإننا نستخدم الرسم البياني الدائري لتصوير عدد المرات التي تظهر فيها القيم المختلفة لهذا المتغير في البيانات. على سبيل المثال، لنفترض أن هناك 54 شخصًا من سيفاس، 12 شخصًا من مرسين، 25 شخصًا من قيصري، 20 شخصًا من أضنة، و4 أشخاص من هاتاي في قسم معين. سيظهر الرسم البياني الدائري الناتج نسبة توزيع هذه الأماكن بين الأشخاص في القسم.

Veri Analizi

Korelasyon (Correlation)

iki ya da daha fazla değişken arasındaki ilişkiyi tanımlayan istatistiksel bir kavramdır. Korelasyon, bir değişkenin diğer bir değişkenle nasıl bir bağlantı içinde olduğunu gösterir. Yani, bir değişkendeki değişikliklerin diğer değişkende nasıl bir etki yarattığını anlamamızı sağlar.

Korelasyonun Temel Unsurları:

Korelasyon Katsayısı (r): Korelasyonun yönünü ve gücünü belirleyen sayısal bir değerdir. Genellikle -1 ile $+1$ arasında bir değer alır.

Pozitif korelasyon: Bir değişken artarken, diğeri de artar. ($r > 0$)

Negatif korelasyon: Bir değişken artarken, diğeri azalır. ($r < 0$)

Sıfır korelasyon: Değişkenler arasında herhangi bir ilişki yoktur ($r = 0$)

الارتباط

هو مفهوم إحصائي يصف العلاقة بين متغيرين أو أكثر. يوضح الارتباط كيف يرتبط أحد المتغيرات بمتغير آخر، أي كيف تؤثر التغيرات في متغير على الآخر. يساعدنا هذا في فهم تأثير التغيرات في متغير معين على المتغير الآخر.

العناصر الأساسية للارتباط:

معامل الارتباط: r هو قيمة عددية تحدد اتجاه وقوة الارتباط. يتراوح عادةً بين -1 و $+1$.

الارتباط الإيجابي: عندما يزيد أحد المتغيرات، يزداد المتغير الآخر أيضاً. $r > 0$

الارتباط السلبي: عندما يزيد أحد المتغيرات، يقل المتغير الآخر. $r < 0$

الارتباط الصفري: لا توجد علاقة بين المتغيرات. $r = 0$

Veri Analizi

Korelasyon (Correlation)

Korelasyon Türleri:

Doğrusal Korelasyon: İki değişken arasındaki ilişki bir doğru boyunca ilerler. Örneğin, eğitim seviyesi arttıkça gelir de artıyorsa, bu pozitif doğrusal korelasyondur.

Doğrusal Olmayan Korelasyon: İlişki doğrusal değildir, ancak değişkenler arasında hala bir bağ vardır. Örneğin, yaşla birlikte refah düzeyi önce artıp sonra düşebilir.

Korelasyon Katsayısının Yorumu:

$r = 1$: Mükemmel pozitif korelasyon (değişkenler arasında tam bir doğrusal ilişki vardır).

$r = -1$: Mükemmel negatif korelasyon (değişkenler tam ters doğrultuda hareket eder).

$r = 0$: Değişkenler arasında ilişki yoktur.

$0 < r < 1$: Zayıf veya güçlü pozitif ilişki.

$-1 < r < 0$: Zayıf veya güçlü negatif ilişki.

الارتباط

أنواع الارتباط:

الارتباط الخطي: العلاقة بين المتغيرين تتبع خطاً مستقيماً. على سبيل المثال، إذا زاد مستوى التعليم وزاد معه الدخل، فهذا ارتباط خطي إيجابي.

الارتباط غير الخطي: العلاقة ليست خطية، ولكن لا يزال هناك ارتباط بين المتغيرات. على سبيل المثال، قد يرتفع مستوى الرفاهية مع التقدم في العمر ثم ينخفض بعد ذلك.

تفسير معامل الارتباط:

$r = 1$ ارتباط إيجابي مثالي (يوجد علاقة خطية تامة بين المتغيرات).

$r = -1$ ارتباط سلبي مثالي (تتحرك المتغيرات في اتجاهين متعاكسين تماماً).

$r = 0$ لا توجد علاقة بين المتغيرات.

$0 < r < 1$ علاقة إيجابية ضعيفة أو قوية.

$-1 < r < 0$ علاقة سلبية ضعيفة أو قوية.

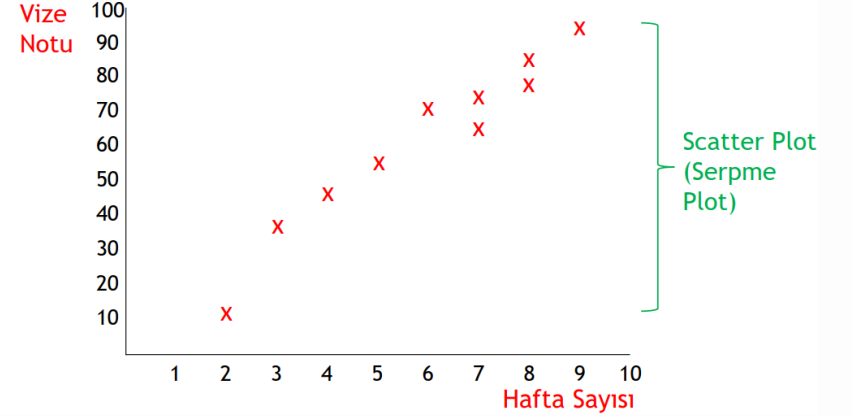
Veri Analizi

Korelasyon (Correlation)

Örneğin derse katılan hafta sayısı ve vize notları arasında bir ilişki var mı diye merak ediyoruz. Bunun için 10 kişiden aşağıdaki veriyi topluyoruz.

1	Hafta Sayısı	Vize Notu
2	5	55
3	3	37
4	2	11
5	8	84
6	9	94
7	7	65
8	7	74
9	4	46
10	6	71
11	8	78

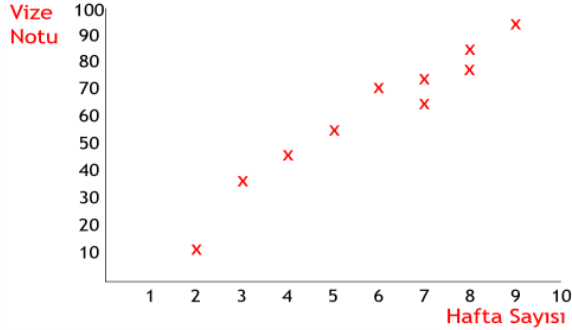
Bu veriyi yataydaki değişken hafta sayısı, dikeydeki değişken vize notu olacak şekilde görselleştirirsek:



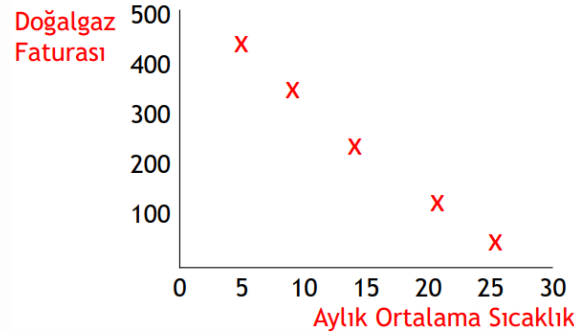
Veri Analizi

Korelasyon (Correlation)

الارتباط



Bu plotta, derse girilen hafta sayısı ile alınan vize notu arasında pozitif bir ilişki görülmektedir. Hafta sayısı arttıkça vize notu artar; hafta sayısı azaldıkça vize notu azalır (iki değişken birlikte hareket ederler). Bu durumda bu iki değişken arasında **pozitif korelasyon** vardır diyeceğiz.



Bu scatter plotta, aylık ortalama sıcaklık doğalgaz faturasının düştüğü görülür. Burada bir değişken arttıkça diğer değişken azalır (yada bir değişken azalırken diğeri artar). Bu durumda iki değişken arasında **negatif korelasyon** vardır diyeceğiz.

Veri Analizi

Korelasyonun Hesaplanması (Korelasyonun Sayısal Olarak Ölçülmesi)

Elimizdeki veri $(x_1, y_1), \dots, (x_N, y_N)$ çiftlerinden oluşsun.

$x_1, \dots, x_N$ değişkenlerinin ortalaması $\bar{x}$; $y_1, \dots, y_N$ değişkenlerinin ortalaması $\bar{y}$ olsun.

(x_i, y_i) çiftini ele alalım.

$(x_i - \bar{x})$, x_i değişkeninin kendi merkezi olan $\bar{x}$ 'den sapmasını verir.

$(y_i - \bar{y})$, y_i değişkeninin kendi merkezi olan $\bar{y}$ 'den sapmasını verir.

$$(x_i - \bar{x})(y_i - \bar{y})$$

çarpımında eğer:

1. $x_i > \bar{x}$ ve $y_i > \bar{y}$ ise bu çarpım pozitif olur (bu durumda her iki değer de merkezinden daha büyük)

Yada

2. $x_i < \bar{x}$ ve $y_i < \bar{y}$ ise bu çarpım yine pozitif olur (bu durumda her iki değer de merkezinden daha küçük).

Veri Analizi

Korelasyonun Hesaplanması (Korelasyonun Sayısal Olarak Ölçülmesi)

Sonuç olarak iki durumda da; iki değişken benzer hareket ediyorlar (ya ikisi birden merkezlerinden daha büyük, ya ikisi birden merkezlerinden daha küçük). O halde (x_i, y_i) çifti benzerdir; ve x ve y değişkenlerinin korelasyonuna pozitif katkıda bulunur.

Tersi olarak

$$(x_i - \bar{x})(y_i - \bar{y})$$

çarpımında eğer:

1. $x_i > \bar{x}$ ve $y_i < \bar{y}$ ise bu çarpım negatif olur olur (bu durumda birinci değişken merkezinden büyük iken ikinci değişken merkezinden küçüktür; bu iki değişken arasında uyumsuzluk vardır)
2. $x_i < \bar{x}$ ve $y_i > \bar{y}$ ise bu çarpım yine negatif olur olur (bu durumda birinci değişken merkezinden küçük iken ikinci değişken merkezinden büyüktür; bu iki değişken arasında yine uyumsuzluk vardır)

Veri Analizi

Korelasyonun Hesaplanması (Korelasyonun Sayısal Olarak Ölçülmesi)

O halde

$$(x_i - \bar{x})(y_i - \bar{y})$$

çarpımı negatif iken (x_i, y_i) çifti benzer değildir; ve x ve y değişkenlerinin korelasyonuna negatif katkıda bulunur.

$(x_1, y_1), \dots, (x_N, y_N)$ çiftlerinin x ve y değişkenlerinin korelasyonuna olan katkılarının ortalamasını alırsak:

$$\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Bu değere **kovaryans** denir σ_{xy} ile gösterilir. Fakat kovaryans değeri $(-\infty, +\infty)$ arasında bir değerdir. Yani herhangi bir sınırlanması yoktur. Veri setinden veri setine çok farklılık gösterebilir.

Elde ettiğimiz benzerlik değerinin herkesçe anlaşılır olması için, bir anlam ifade etmesi için standart bir aralık olan $[-1, 1]$ aralığında yer alması gerekir. Bunun için kovaryansı x 'in ve y 'nin standart sapmalarına böleceğiz.

Veri Analizi

Korelasyonun Hesaplanması (Korelasyonun Sayısal Olarak Ölçülmesi)

x 'in standard sapması σ_x :

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

y 'nin standard sapması σ_y :

$$\sigma_y = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2}$$

Kovaryans σ_{xy} , yukarıdaki standart sapmaların çarpımına bölünürse:

$$\frac{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2}}$$

Veri Analizi

Korelasyonun Hesaplanması (Korelasyonun Sayısal Olarak Ölçülmesi)

Bu ifade sadeleştirilerek korelasyon hesaplanır:

$$\rho = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

Bu değere Pearson Korelasyon Katsayısı diyeceğiz. Bu değer her zaman $[-1, 1]$ arasındadır.

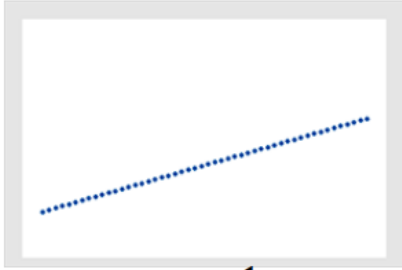
Eğer x ve y değişkenleri arasında negatif korelasyon varsa korelasyon negatiftir $[-1, 0)$ arası değer alır;

eğer x ve y değişkenleri arasında pozitif korelasyon varsa korelasyon pozitifdir $(0, 1]$ arası bir değer alır;

eğer x ve y değişkenleri arasında herhangi bir korelasyon yoksa, korelasyon 0'dır.

Veri Analizi

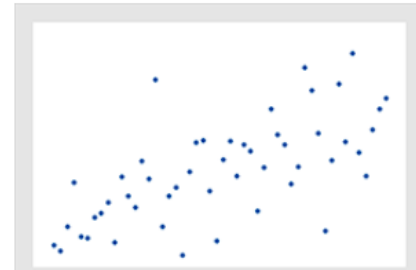
Farklı Büyüklükte Korelasyonlar



$$\rho = 1$$



$$\rho = 0.8$$



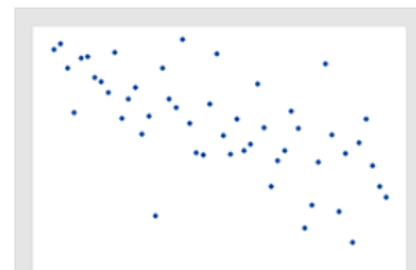
$$\rho = 0.6$$



$$\rho = -1$$



$$\rho = -0.8$$



$$\rho = -0.6$$

08



İstatistik

Eng: Abdulrahman Hamdi

İstatistik

Çıkarımsal istatistikte örneğin (sample'ın) alındığı popülasyonla ilgili çıkarım yaptığımızı söylemiştik. Bu çıkarımlar popülasyon parametreleri (ortalama, varyans, popülasyon oranı) ile ilgili tahminler (estimation) dir.

Popülasyon parametreleri ile ilgili tahminlerimizi iki başlık altında inceleyeceğiz:Nokta Tahminler ve Aralık Tahminler

Nokta Tahminler (Point Estimators)

1) Popülasyon Ortalamasının Nokta Tahmini

$X_1, X_2, \dots, X_n$ değerleri, popülasyondan alınan n gözlem olsun. Bu gözlemlerin oluşturduğu örneklem (sample) ortalaması olan $\bar{X}$ popülasyon ortalaması μ 'nun bir tahmini olarak kullanılabilir. Zaten merkezi limit teoreminden hatırlarsak popülasyondan rastgele eleman seçerek oluşturulan örneklerin ortalamalarının beklenen değeri μ idi. Bu yüzden herhangi bir örneğin ortalaması μ 'için bir tahmin olabilir.

في الإحصاء الاستنتاجي، ذكرنا أننا نقوم بإجراء استنتاج حول المجتمع (population) الذي أخذنا منه العينة (sample) هذه الاستنتاجات هي عبارة عن تقديرات (estimation) تتعلق بمعلومات المجتمع (متوسط، تباين، نسبة المجتمع). سندرس تقدير اتنا المتعلقة بمعلومات المجتمع تحت عنوانين رئيسيين: التقديرات النقطية والتقديرات الفترية.

التقديرات النقطية (Point Estimators)

1) التقدير النقطي لمتوسط المجتمع

لنفترض أن القيم $X_1, X_2, \dots, X_n$ هي مشاهدات مأخوذة من المجتمع بعدد n مشاهدات. يمكن استخدام متوسط العينة $\bar{X}$ (sample mean) الناتج عن هذه المشاهدات كتقدير لمتوسط المجتمع μ من نظرية الحد المركزي، نعلم أن القيمة المتوقعة لمتوسطات العينات العشوائية المأخوذة من المجتمع هي μ ولذلك، يمكن اعتبار متوسط أي عينة تقديراً لـ μ

İstatistik

ör. Bir fabrikada üretilen 8 otomobilin ilk kez arızalanmaları (bin km cinsinden) 175, 180, 185, 198, 148, 162, 191 ve 205 km sonra gerçekleşmiş olsun. Buna göre bu fabrikada üretilen otomobiller ortalama kaç bin km'den sonra ilk kez arızalanırlar?

Çözüm.

Seçilen 8 otomobilin ortalaması olan $\bar{X} = \frac{175+180+185+198+148+162+191+205}{8} = 180.5$

fabrikanın ürettiği tüm otomobillerden oluşan populasyonun ortalamasının bir tahmini olabilir.

Bu tahmin sonuçta bir tahmin olduğundan bir hata payı vardır. Bu hatanın miktarı örneğin standart sapması olan $\frac{\sigma}{\sqrt{n}}$ kadardır. (σ populasyon standart sapması, n örnek büyüklüğü).

Örnek üzerinden populasyon parametresi nokta tahmini olarak yapılırken, tahminin hatasına standart hata (standard error) denir.

Standart hata populasyon ortalaması için $\frac{\sigma}{\sqrt{n}}$ kadardır.

Bir önceki soru için varsayalım ki fabrikanın ürettiği tüm otomobillerin ilk kez arıza yapmalarının standart sapması 45 bin olsun. Bu durumda nokta tahminimiz olan 180.5 değerinin standart hatası $= \frac{45}{\sqrt{8}} = 15.95$ olur.

مثال: لنفترض أن أول عطل حدث لـ 8 سيارات تم إنتاجها في مصنع معين (بالكيلومترات) بعد 175، 180، 185، 198، 148، 162، 191 و 205 كيلومتراً. بناءً على ذلك، كم يبلغ متوسط المسافة التي تقطعها السيارات المنتجة في هذا المصنع (بالكيلومترات) قبل حدوث أول عطل؟

İstatistik

ör. Örnek büyüklüğü 1000 iken örneğin standart hatasını 4'te 1'e düşürmek için örneğe kaç eleman daha eklemek gerekir?

مثال: عندما يكون حجم العينة 1000، كم عدد العناصر الإضافية التي يجب إضافتها إلى العينة لتقليل الخطأ المعياري للعينة إلى ربع قيمته؟

Çözüm.

Örnek büyüklüğü 1000 iken standart hata: $\frac{\sigma}{\sqrt{1000}}$ dir. Bunun 4'te 1'i $\frac{\sigma}{4\sqrt{1000}} = \frac{\sigma}{\sqrt{16000}}$ dir. O halde örnek büyüklüğü 16000 olmalıdır, yani 15000 eleman daha eklenmelidir.

İstatistik

2) Populasyon Oranının Nokta Tahmini

Varsayalım ki bir populasyon elemanlarının belirli bir oranını merak ediyoruz. Örneğin, Türkiye'deki seçmenlerin bir X partisine oy verenlerin oranı, bir fabrikanın ürettiği bütün mallardan bozuk olanlarının oranı veya 18-25 yaş arası gençlerde sosyal medya bağımlılığı oranı gibi.

Populasyon oranını tahmin etmek için, populasyondan rastgele elemanlar seçerek bir örnek oluşturabiliriz. Daha sonra bu örneğin oranını, populasyon oranının bir nokta tahmini olarak kullanabiliriz.

Populasyon oranının nokta tahmini: $\hat{p} = \frac{X}{n}$

Burada X örnekte, ilgilendiğimiz olaya katılan eleman sayısıdır. Örneğin örnekteki bozuk ürün sayısı, belirli bir partiye oy verenlerin sayısı, sosyal medya bağımlılarının sayısı...

Bu tahminin standart hatası $\sqrt{\frac{p(1-p)}{n}}$ olur

2) التقدير النقطي لنسبة المجتمع

افترض أننا مهتمون بنسبة معينة من عناصر المجتمع. على سبيل المثال، نسبة الناخبين في تركيا الذين يصوتون لحزب معين، أو نسبة السلع المعيبة التي تم إنتاجها في مصنع معين، أو نسبة إيمان وسائل التواصل الاجتماعي بين الشباب الذين تتراوح أعمارهم بين 18 و 25 سنة. لتقدير نسبة المجتمع، يمكننا اختيار عناصر عشوائية من المجتمع لتكوين عينة. ثم، يمكننا استخدام نسبة هذه العينة كتقدير نقطي لنسبة المجتمع.

التقدير النقطي لنسبة المجتمع: $\hat{p} = \frac{X}{n}$

هنا، X في العينة هو عدد العناصر المشاركة في الحدث الذي نهتم به: على سبيل المثال، عدد المنتجات المعيبة في العينة، عدد الأشخاص الذين يصوتون لحزب معين، عدد الأشخاص المدمنين على وسائل التواصل الاجتماعي...

خطأ التقدير المعياري لهذا التقدير هو $\sqrt{\frac{p(1-p)}{n}}$

İstatistik

Bu tahminin standart hatası $\sqrt{\frac{p(1-p)}{n}}$ olur

Burada p , populasyonun oranıdır. Fakat p 'yı bilmediğimizden p 'nin yerine örnekten elde ettiğimiz tahmin olan $\hat{p}$ 'yi kullanırız.

ör. Bir bölümde son sınıflar için mezuniyet yapıp yapılmaması düşünüyor. Bunun için son sınıf öğrencilerinden rastgele 50 kişi seçilip, fikirleri soruluyor. 30 öğrenci mezuniyetin olmasını istiyor. Bu durumda tüm son sınıf öğrencileri arasında mezuniyet olmasını isteyen öğrencilerin oranı için bir tahmin yapınız. Bu yaptığınız tahminin standart hatası ne olur?

Çözüm.

Populasyon oranı için nokta tahmini $\hat{p} = \frac{30}{50} = 0.6$

Bu tahminin standart hatası: $\sqrt{\frac{0.6 \cdot 0.4}{50}} = 0.0693$

خطأ التقدير المعياري لهذا التقدير هو $\sqrt{\frac{p(1-p)}{n}}$

هنا، p هو نسبة المجتمع. ولكن بما أننا لا نعرف p ، فإننا نستخدم التقدير الذي نحصل عليه من العينة والذي يُرمز له $\hat{p}$ بدلاً من p

مثال، يتم التفكير في إجراء حفل تخرج لطلاب السنة النهائية في قسم معين. وللقيام بذلك، تم اختيار 50 طالبًا عشوائيًا من طلاب السنة النهائية وسؤالهم عن آرائهم. وأفاد 30 طالبًا بأنهم يرغبون في إقامة حفل التخرج. في هذه الحالة، قم بتقدير النسبة المئوية للطلاب الذين يرغبون في إقامة حفل التخرج بين جميع طلاب السنة النهائية. ما هو الخطأ المعياري لهذا التقدير الذي قمت به؟

İstatistik

ör. Önümüzdeki seçimlerde seçmenlerin kime oy vereceğini tahmin etmek için bir anket yapılıyor. 300 kişilik bu ankete 210 kişi X partisine oy vereceğini söylüyor. Buna göre tüm seçmenler içinde X partisine oy vereceklerin oranını tahmin ediniz. Bu tahminizin standart hatasını hesaplayınız.

Çözüm.

X partisine oy vereceklerin oranının tahmini $\hat{p} = \frac{210}{300} = 0.7$

Bu tahminin standart hatası: $\sqrt{\frac{0.7 \cdot 0.3}{300}} = 0.02$

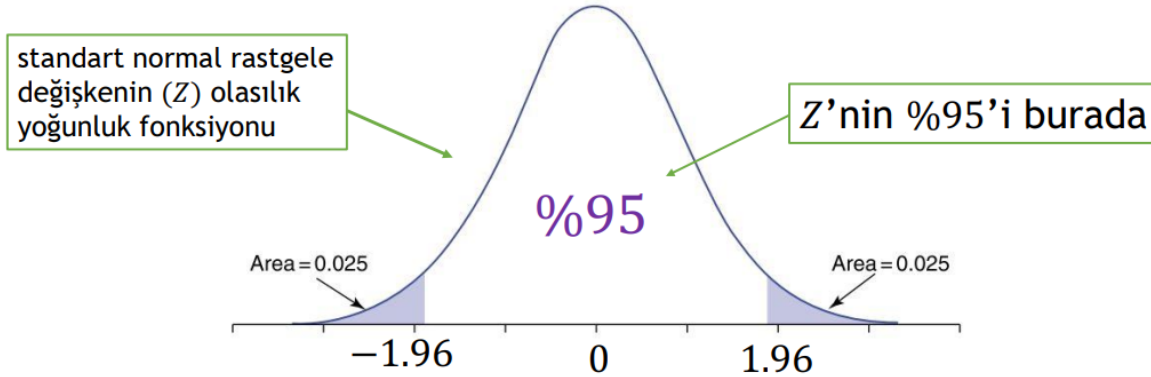
İstatistik

Aralık Tahminler (Interval Estimators)

Bir populasyon parametresinin aralık tahmini, bu parametreyi içerdği düşünülen aralıktır. Bu aralığın güven değeri, aralığın parametreyi içermesi olasılığıdır.

التقديرات الفترية

التقدير الفتري لمعلمة المجتمع هو النطاق الذي يُعتقد أن هذه المعلمة تقع ضمنه. القيمة الموثوقة لهذا النطاق هي احتمال احتواء النطاق على المعلمة.



Standart normal rastgele değişken olan Z 'nin 1.96'dan büyük olma olasılığı: $P(Z \geq 1.96) = 0.025$

Z 'nin 1.96'dan küçük olma olasılığı: $P(Z \leq -1.96) = 0.025$

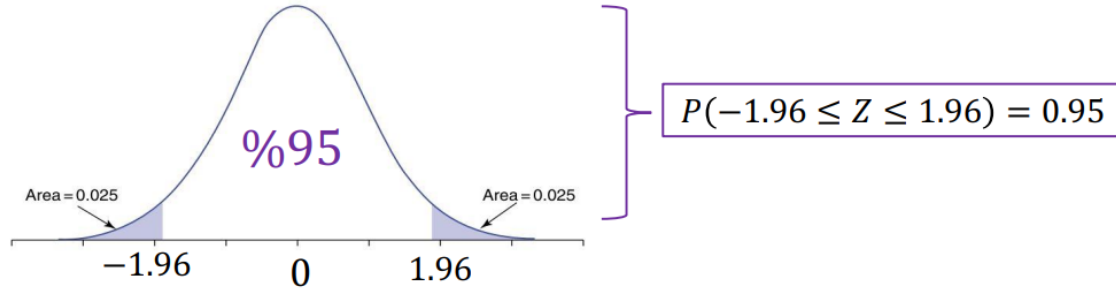
İstatistik

$\bar{X}$ rastgele değişkeni popülasyondan alınan bir n büyüklükteki örneğin ortalaması olsun. Merkezi limit teoremi gereği $\bar{X}$ normal dağılıma sahiptir, ortalaması (beklenen değeri) μ 'dür ve standart sapması $\frac{\sigma}{\sqrt{n}}$

$\bar{X}$ normal değişkenini standart hale getirmek için (ortalamasını 0 yapmak ve standart sapmasını 1 yapmak için):

$$Z \leftarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

لتكن $\bar{X}$ متغير عشوائي هو متوسط عينة من حجم n مأخوذة من المجتمع. وفقًا لنظرية الحد المركزي، فإن $\bar{X}$ يتبع توزيعًا طبيعيًا، حيث يكون متوسطه (القيمة المتوقعة) μ وانحرافه المعياري هو $\frac{\sigma}{\sqrt{n}}$ لتوحيد المتغير X العادي (أي لجعل متوسطه 0 وانحرافه المعياري 1)، نقوم بـ:



İstatistik

$-1.96 \leq Z \leq 1.96$ aralığını X ve μ 'yü içerecek şekilde düzenlersek:

$$\begin{aligned} -1.96 &\leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq 1.96 \\ -1.96 \times \frac{\sigma}{\sqrt{n}} &\leq \bar{X} - \mu \leq 1.96 \times \frac{\sigma}{\sqrt{n}} \\ \bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} &\leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}} \end{aligned}$$

olur. O halde populasyon ortalaması olan μ 'nün $\left[\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}} \right]$

aralığında olma olasılığı 0.95'tir. Bir başka deyişle biz %95 eminizki μ bu aralıktadır.

Bu aralığa μ 'nün %95 güven aralığı (confidence interval) denir.

احتمال أن تكون المعلمة ضمن هذا النطاق هو 0.95. بمعنى آخر، نحن واثقون بنسبة 95% أن μ يقع ضمن هذا النطاق. يُسمى هذا النطاق نطاق الثقة بنسبة 95% μ .

$$P\left(\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

İstatistik

ör. Daha önceki soruda bir fabrikada üretilen 8 otomobilin ilk kez arızalanmaların kaç bin km sonra gerçekleştiğini vermistik: 175, 180, 185, 198, 148, 162, 191 ve 205. Diyelimki bu fabrikada üretilen tüm otomobillerin (populasyonun) kaç bin km sonra arızalanmalarının standart sapması 45 olsun. Bu durumda üretilen otomobillerin ortalama kaç bin km den sonra arızalancağının %95 güven aralığını bulunuz.

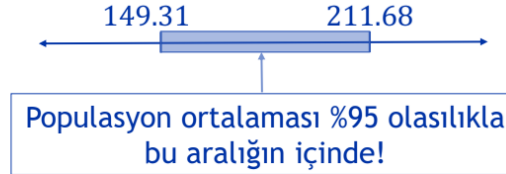
مثال: في السؤال السابق قدمنا بيانات حول المسافة التي قطعها 8 سيارات تم إنتاجها في مصنع معين قبل حدوث أول عطل: 175، 180، 185، 198، 148، 162، 191 و 205 كيلومتر. لنفترض أن انحراف المسافة المعياري لجميع السيارات المنتجة في المصنع (المجتمع) هو 45 كيلومتر. في هذه الحالة، حدد نطاق الثقة بنسبة 95% لعدد الكيلومترات التي ستقطعها السيارات قبل حدوث أول عطل.

Çözüm.

$$\bar{X} = \frac{175+180+185+198+148+162+191+205}{8} = 180.5$$

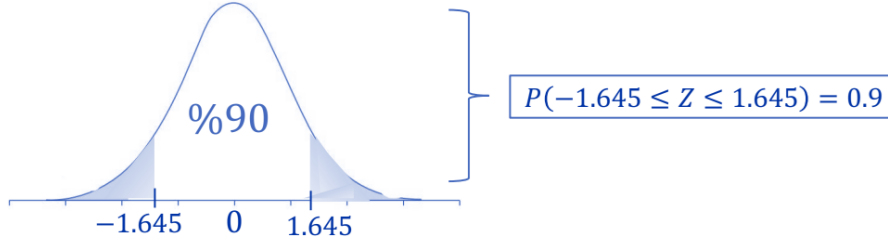
$$\%95 \text{ güven aralığı } \left[180.5 - 1.96 \times \frac{45}{\sqrt{8}}, 180.5 + 1.96 \times \frac{45}{\sqrt{8}} \right] = [149.31, 211.68].$$

Yani biz %95 eminizki populasyonun ortalaması [149.31, 211.68] aralığındadır.



İstatistik

Bir başka önemli güven aralığı %90 güven aralığıdır. Z standart normal rastgele değişkeninin $[-1.645, 1.645]$ aralığında olma olasılığı 0.9 dur:



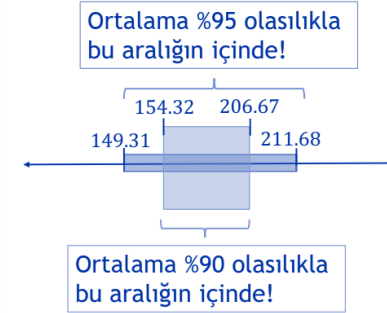
Populasyon ortalaması olan μ 'nün $\left[\bar{X} - 1.645 \times \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}}\right]$ aralığında olma olasılığı 0.9'dur.

Dikkat edilirse bu aralık, %95 aralığına göre daha dar bir aralıktır. O halde aralık darlaştıkça (genişledikçe), μ 'nün bu aralığa düşme şansı azalır (artar).

Bir önceki soru için %90 güven aralığını hesaplayalım.

$\bar{X} = 180.5$ ve $\sigma = 45$ idi. O halde μ 'nün %90 güven aralığı:

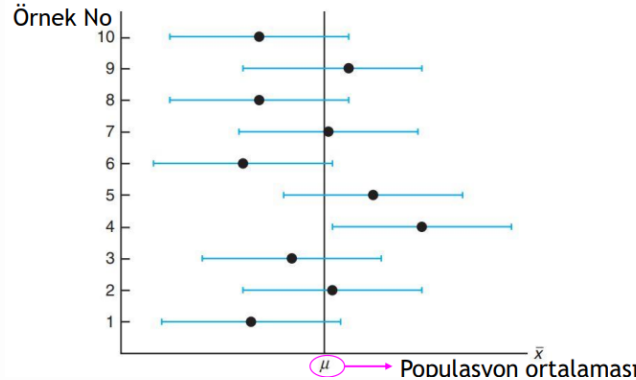
$$\left[180.5 - 1.645 \times \frac{45}{\sqrt{8}}, 180.5 + 1.645 \times \frac{45}{\sqrt{8}}\right] = [154.32, 206.67]$$



İstatistik

Unutulmaması gerekirkı güven aralıkları eldeki örnekten elde edilir. O yüzden aynı populayondan alınan farklı örneklerden elde edilen güven aralıkları birbirinden farklı olabilir. Örneğin aşağıda aynı populasıyondan alınan 10 farklı örnekten elde edilen güven aralıkları verilmiştir (mavi aralıklar). Bu örneklerin kendi ortalamaları ise nokta (•) ile verilmiştir. Görölüyorki 10 farklı örnekten elde edilen 10 farklı güven aralığının 9'u populasıyon ortalaması olan μ 'yü içeriıyor.

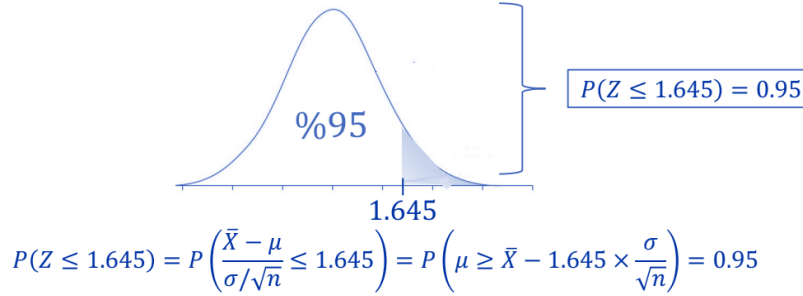
يجب أن نأخذ في الاعتبار أن نطاقات الثقة تُستخرج من العينة المتاحة. لذلك، قد تكون نطاقات الثقة المستخلصة من عينات مختلفة مأخوذة من نفس المجتمع مختلفة عن بعضها البعض. على سبيل المثال، في الأسفل، يتم عرض نطاقات الثقة المستخلصة من 10 عينات مختلفة مأخوذة من نفس المجتمع (النطاقات باللون الأزرق). أما المتوسطات الخاصة بهذه العينات فتُعرض بالنقطة (•). كما نرى، 9 من أصل 10 نطاقات ثقة مستخلصة من العينات المختلفة تحتوي على المتوسط الفعلي للمجتمع μ .



İstatistik

Alt ve Üst Güven Aralıkları

Bazen populasyon ortalamasının yalnızca bir üst sınırdan daha küçük olması, yada yalnızca bir alt sınırdan daha büyük olması ile ilgileniriz. Yani popülasyon ortalamasının alabileceği en büyük değer yada bu ortalamanın alabilceği en küçük değer ile ilgileniriz.



$$P\left(\mu \leq \bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

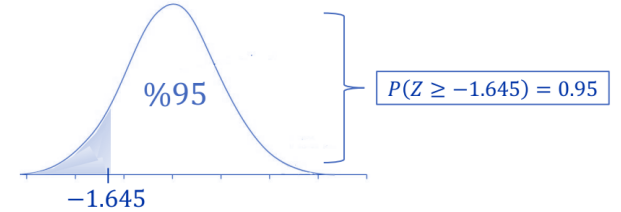
Populasyon ortalaması μ , %95 kesinlikle en fazla $\bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}}$ 'dir. Başka bir deyişle %95 eminizki populasyon ortalaması $\bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}}$ değerinden küçüktür.

النطاقات الثقة العليا والدنيا

في بعض الأحيان، نكون مهتمين فقط بأن يكون المتوسط السكاني أصغر من حد أعلى معين، أو أكبر من حد أدنى معين. بمعنى آخر، نحن مهتمون بالقيمة القصوى التي يمكن أن يأخذها المتوسط السكاني أو بالقيمة الدنيا التي يمكن أن يأخذها هذا المتوسط.

$$P\left(\mu \geq \bar{X} - 1.645 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Populasyon ortalaması μ , %95 kesinlikle en az $\bar{X} - 1.645 \times \frac{\sigma}{\sqrt{n}}$ 'dir. Başka bir deyişle %95 eminizki populasyon ortalaması $\bar{X} - 1.645 \times \frac{\sigma}{\sqrt{n}}$ değerinden büyüktür.



$$P(Z \geq -1.645) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq -1.645\right) = P\left(\mu \leq \bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

İstatistik

ör. 16 kişilik bir sınıfta öğrencilerin boyları cm cinsinden 188, 181, 165, 172, 181, 168, 165, 190, 177, 176, 170, 161, 182, 194, 157, 173 olsun. Bu sınıfın dahil olduğu okuldaki tüm kişilerin boy ortalamalarının standart sapmasının bilindiği ve bu standart sapmanın 14 olduğu varsayalım. Buna göre

- okuldaki kişilerin boy ortalamasını %90 ve %95 güven aralığı içerisinde veriniz.
- okuldaki kişilerin boy ortalaması %95 kesinlikle en fazla kaç olur?
- okuldaki kişilerin boy ortalaması %95 kesinlikle en az kaç olur?

مثال: في صف مكون من 16 طالبًا، تكون أطوال الطلاب بالسنتيمتر كما يلي: 188، 181، 165، 172، 181، 168، 165، 190، 177، 176، 170، 161، 182، 194، 157، 173. لنفترض أن انحراف المتوسط لجميع الأشخاص في المدرسة معروف ويبلغ 14. بناءً على هذه البيانات، قم بالإجابة على الأسئلة التالية:

- قدم نطاق الثقة لمتوسط أطوال الأشخاص في المدرسة عند مستوى ثقة 90% و 95%.
- ما هو أقصى قيمة قد يكون عليها متوسط أطوال الأشخاص في المدرسة بنسبة ثقة 95%؟
- ما هو أدنى قيمة قد يكون عليها متوسط أطوال الأشخاص في المدرسة بنسبة ثقة 95%؟

İstatistik

Çözüm. Örneğin ortalaması:

$$\bar{X} = \frac{188 + 181 + 165 + 172 + 181 + 168 + 165 + 190 + 177 + 176 + 170 + 161 + 1 + 194 + 157 + 173}{16}$$

$$\bar{X} = 175$$

Örnek büyüklüğü: $n = 16$

Populasyon standart sapması: $\sigma = 14$

Populasyon ortalaması μ 'nün %90 güven aralığı: $\left[\bar{X} - 1.645 \times \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}} \right]$

(yani bir başka deyişle biz %90 eminizki μ bu aralığın içindedir)

O halde %90 güven aralığı: $\left[175 - 1.645 \times \frac{14}{\sqrt{16}}, 175 + 1.645 \times \frac{14}{\sqrt{16}} \right] = [169.24, 180.75]$

Populasyon ortalaması μ 'nün %95 güven aralığı: $\left[\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}} \right]$

O halde %90 güven aralığı: $\left[175 - 1.96 \times \frac{14}{\sqrt{16}}, 175 + 1.96 \times \frac{14}{\sqrt{16}} \right] = [168.14, 181.86]$

Çözüm.

ii) $P\left(\mu \geq \bar{X} - 1.645 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$ idi. Şu halde $\bar{X}$, σ ve $\sqrt{n}$ yerlerine konulursa:

$$P\left(\mu \geq 175 - 1.645 \times \frac{14}{\sqrt{16}}\right) = P(\mu \geq 169.24) = 0.95$$

olur. Yani %95 eminizki populasyon ortalaması 169.24'ten büyüktür.

iii) Üst sınır için $P\left(\mu \leq \bar{X} + 1.645 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$ idi.

$$P\left(\mu \leq 175 + 1.645 \times \frac{14}{\sqrt{16}}\right) = 0.95$$

$$P(\mu \leq 180.75) = 0.95$$

%95 eminizki populasyon ortalaması 180.75'ten küçüktür.

İstatistik

Populasyon Standart Sapması Bilinmezken Ortalama İçin Güven Aralığı

Şu ana kadar populasyon ortalaması (μ) için bir güven aralığı oluştururken populasyon standart sapması olan σ 'yı bildiğimizi varsayıyorduk. Şimdi ise σ 'yı bilmediğimizi varsayalım; ki bu durumla karşılaşmak çok daha olasıdır.

Merkezi limit teoreminden gördükki

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

standart normal dağılıma sahipti (yani ortalaması 0, standart sapması 1 idi). Şimdi ise populasyon standart sapması σ 'yı bilmediğimizden, bunun yerine örneğin standart sapması olan s 'yi yazarsak:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

t-dağılımına (t-distribution); tam olarak t_{n-1} dağılımına sahip olur. Burada $n - 1$ serbestlik derecesidir (degree of freedom).

نطاق الثقة للمتوسط عندما لا يكون انحراف المجتمع المعياري معروفاً

حتى الآن، كنا نفترض أننا نعرف انحراف المجتمع المعياري σ عند إنشاء نطاق الثقة للمتوسط السكاني μ ولكن الآن، دعونا نفترض أننا لا نعرف σ وهو الوضع الأكثر احتمالاً.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{من نظرية الحد المركزي، رأينا أن}$$

كان لديه توزيع طبيعي قياسي (أي أن المتوسط كان 0 والانحراف المعياري كان 1). الآن، بما أننا لا نعرف انحراف المجتمع المعياري σ ، بدلاً من ذلك

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad \text{نستخدم الانحراف المعياري للعينه } s$$

سيكون التوزيع هو توزيع t-dağılım وبالتحديد توزيع t_{n-1} هنا $n - 1$ هو عدد درجات الحرية

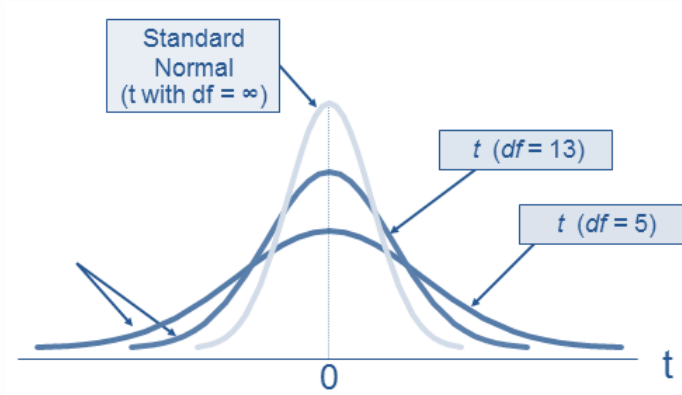
İstatistik

t-dağılımları

t-dağılımları genel olarak standart normal dağılıma benzer: çan eğrileri şeklindedir fakat biraz daha düzdür, pek sivri değildir. t-dağılımları tek bir dağılım değildir. Serbestlik derecesi $n - 1$ değiştikçe farklılık gösterirler. Genel olarak n (örnek büyüklüğü) arttıkça standart normal dağılıma yaklaşırlar.

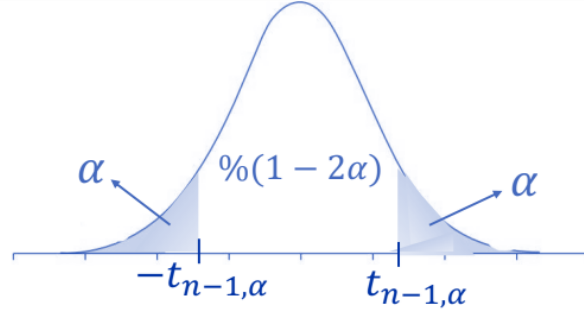
توزيع t -

توزيعات t بشكل عام تشبه التوزيع الطبيعي القياسي: هي منحنيات على شكل جرس، لكنها أكثر تسطحًا وأقل حدة. توزيعات t ليست توزيعًا واحدًا. تتغير حسب درجات الحرية $n-1$ بشكل عام، كلما زاد حجم العينة n ، اقتربت هذه التوزيعات من التوزيع الطبيعي القياسي.

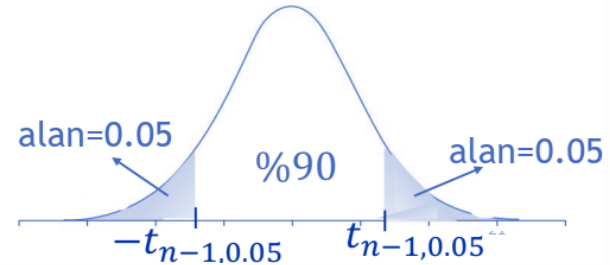
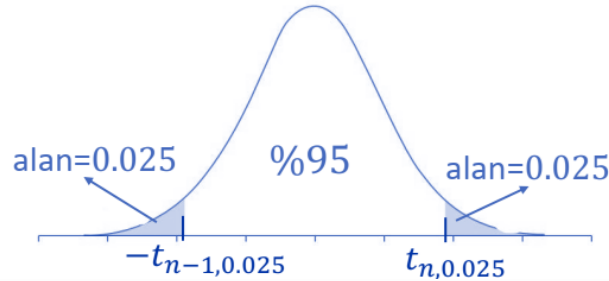


İstatistik

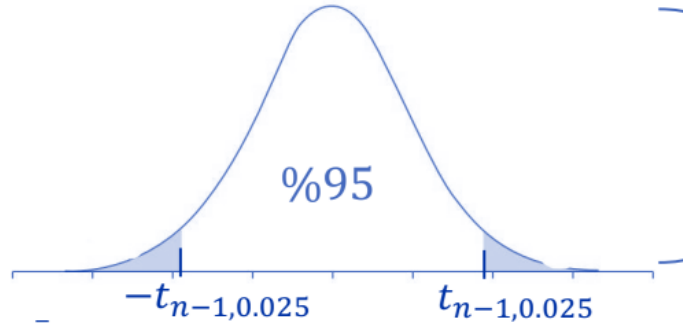
t-dağılımları



$t_{n,\alpha}$ 'nın sağındaki alanın büyüklüğü α kadardır. $-t_{n,\alpha}$ 'nın solundaki alanın büyüklüğü α kadardır. Biz genelde %95 ve %90 güven aralıkları ile ilgileniriz. %95 güven aralığını veren t değerleri $-t_{n,0.025}$ ve $t_{n,0.025}$; %90 güven aralığını veren t değerleri $-t_{n,0.05}$ ve $t_{n,0.05}$ dir.



İstatistik



$$P(-t_{n-1,0.025} \leq T \leq t_{n-1,0.025}) = 0.95$$

$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ idi. $-t_{n-1,0.025} \leq T \leq t_{n-1,0.025}$ eşitsizliğinde yerine yazılırsa, populasyon ortalaması için %95 güven aralığı:

$$\left[\bar{X} - t_{n-1,0.025} \times \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1,0.025} \times \frac{s}{\sqrt{n}} \right]$$

olur. Burada $t_{n,0.025}$ değerleri tablodan bulunur.

İstatistik

ör. Bir önceki soru için, okuldaki kişilerin boylarının standart sapmalarını bilmediğimiz varsayarak yalnızca elimizdeki veriye dayanarak okul boy ortalaması için %95 ve %90 güven aralıklarını bulalım.

Çözüm. Örneğin standart sapması:

$$s = \sqrt{\frac{1}{15} \times ((188 - 175)^2 + (181 - 175)^2 + \dots + (173 - 175)^2)} = 10.6$$

Table D.2 Percentiles $t_{n, \alpha}$ of t Distributions

n	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.601	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.660	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922

$$t_{15,0.025} = 2.131$$

مثال، في السؤال السابق، إذا افترضنا أننا لا نعرف انحرافات معايير أطوال الأشخاص في المدرسة، سنستخدم البيانات المتاحة لدينا فقط لتحديد نطاقات الثقة بنسبة 95% و 90% لمتوسط طول الطلاب في المدرسة.

Populasyon ortalaması için %95 güven aralığı:

$$\left[\bar{X} - t_{n-1,0.025} \times \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1,0.025} \times \frac{s}{\sqrt{n}} \right]$$
$$\left[175 - 2.131 \times \frac{10.6}{\sqrt{16}}, 175 + 2.131 \times \frac{10.6}{\sqrt{16}} \right] = [169.35, 180.64]$$

İkinci olarak %90 güve aralığını bulalım. Bunun için $t_{15,0.05}$ değerini tablodan bulmamız gerekir. Bir önceki slayttaki tabloya bakılırsa, bu değer 1.753'tür. O halde %90 güven aralığı:

$$\left[175 - 1.753 \times \frac{10.6}{\sqrt{16}}, 175 + 1.753 \times \frac{10.6}{\sqrt{16}} \right] = [170.35, 179.64]$$

Dikkat edilirse t-dağılımı yardımıyla bulunan %90 güven aralığı, yine t-dağılımı yardımıyla bulunan %95 güven aralığından daha dardır; tıpkı populasyon standart sapmasını bildiğimizi varsaydığımız durumda standart normal dağılımı yardımıyla bulduğumuz %90 ve %95 güven aralıkları gibi.

İstatistik

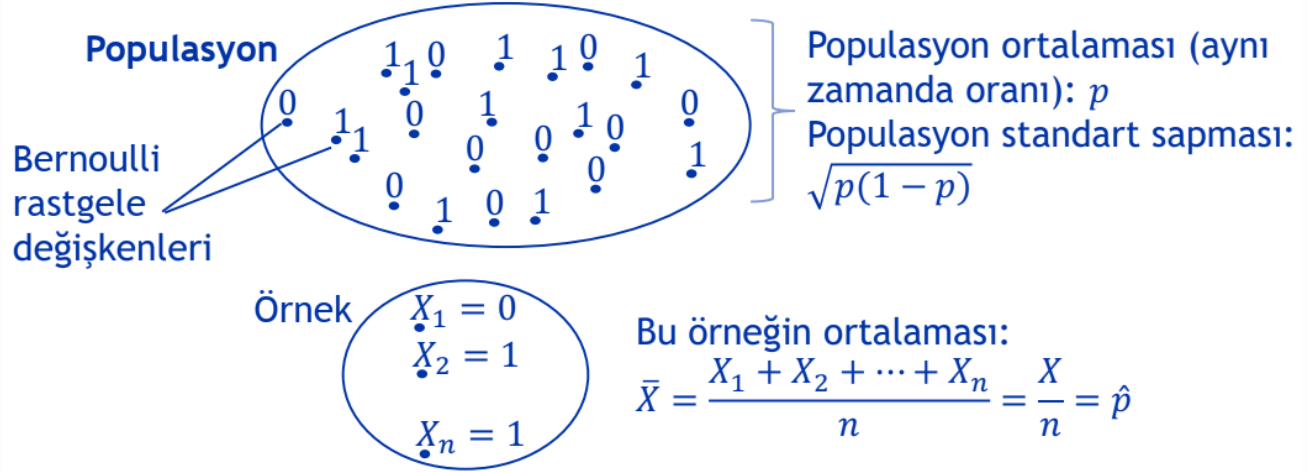
Populasyon Oranı İçin Güven Aralığı

Tıpkı populasyon ortalamasında olduğu gibi, populasyon oranı içinde güven aralığı oluşturabiliriz.

Daha önce normal dağılım yardımıyla binomial dağılıma yaklaştıktık:

نطاق الثقة للمتوسط السكاني

تمامًا كما هو الحال في المتوسط السكاني، يمكننا أيضًا إنشاء نطاق ثقة للمتوسط السكاني. في وقت سابق، استخدمنا التوزيع الطبيعي للاقترب من التوزيع الثنائي.



İstatistik

Örneğin ortalaması aynı zamanda oranı olan $\hat{p} = \frac{X_1+X_2+\dots+X_n}{n}$, merkezi limit teorem gereği normal dağılıma sahiptir. Örneklerden alınan bu oranların ortalaması p

(populasyon oranı), standart sapması $\frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$ dir.

Normal dağılıma sahip $\hat{p}$ 'yi standart hale getirirsek:

$$Z \leftarrow \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

olur. Populasyon oranı olan p yerine onun bir tahmini olan örnekten elde ettiğimiz $\hat{p}$ 'yi yazabiliriz. Bu durumda:

$$Z \leftarrow \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

yazabiliriz. Bu değişken standart normal dağılıma sahip olur!

على سبيل المثال، المتوسط الذي هو أيضًا النسبة $\hat{p} = \frac{X_1+X_2+\dots+X_n}{n}$ وفقًا لنظرية الحد المركزي، يتبع التوزيع الطبيعي. متوسط هذه النسب المستخلصة من العينات هو p النسبة السكانية، وانحرافها المعياري هو

$$\frac{\sqrt{p(1-p)}}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

إذا جعلنا $\hat{p}$ الذي يتبع التوزيع الطبيعي معياريًا، نحصل على:

$$Z \leftarrow \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

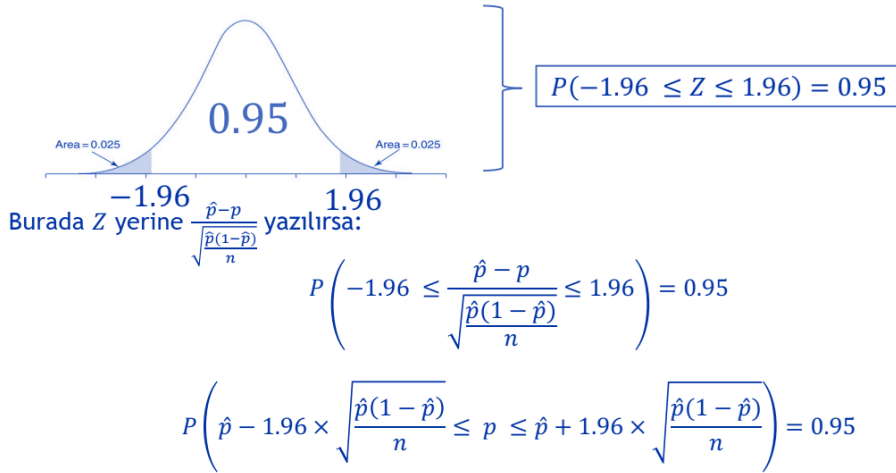
وبما أن النسبة السكانية غير معروفة، يمكننا استخدام تقديرها من العينة $\hat{p}$ بدلاً

$$Z \leftarrow \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

وهذا المتغير يتبع التوزيع الطبيعي القياسي!

İstatistik

Daha önceden de görmüştük ki:



Sonuç olarak populasyon oranı p için %95 güven aralığı:

$$\left[\hat{p} - 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

olur. Eğer 1.96 değeri yerine 1.645 yazılırsa p için %90 güven aralığı elde edilir:

$$\left[\hat{p} - 1.645 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.645 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

İstatistik

ör. Daha önceki örnekte önümüzdeki seçimlerde seçmenlerin kime oy vereceğini tahmin etmek için bir anket yapıldığını, 300 kişilik bu anketde 210 kişi X partisine oy vereceğini söylediğini belirtmiştik. Bu tahminin %95 ve %90 güven aralığı ne olur?

Çözüm.

$$\hat{p} = \frac{210}{300} = 0.7$$

مثال، في المثال السابق، ذكرنا أنه تم إجراء استطلاع رأي للتنبؤ بمن سيصوت الناخبون في الانتخابات القادمة، حيث أشار 210 شخص من أصل 300 في هذا الاستطلاع إلى أنهم سيصوتون لحزب X. فما هو نطاق الثقة بنسبة 95% و 90% لهذا التقدير؟

%95 güven aralığı:

$$\left[0.7 - 1.96 \times \sqrt{\frac{0.7 \times 0.3}{300}}, 0.7 + 1.96 \times \sqrt{\frac{0.7 \times 0.3}{300}} \right] = [0.64, 0.75]$$

%90 güven aralığı:

$$\left[0.7 - 1.645 \times \sqrt{\frac{0.7 \times 0.3}{300}}, 0.7 + 1.645 \times \sqrt{\frac{0.7 \times 0.3}{300}} \right] = [0.65, 0.74]$$

olur. Buna göre biz %95 eminizki tüm seçmeler içerisinde X partisine oy verme oranı [0.64, 0.75] arasındadır; %90 eminizki tüm seçmeler içerisinde X partisine oy verme oranı [0.65, 0.74] arasındadır.

İstatistik

Karar Vermek

Şimdiye kadar populasyon paramaterlerini (ortalama, populasyon oranı) tahmin etmekle ilgilendik. Fakat bir mühendisin eldeki veriyle yaptığı tek iş bu değildir. Genellikle mühendisten eldeki veri yardımıyla bir karar vermesi de beklenir. Örneğin iki farklı ölçüm cihazı ile yapılan ölçümler arasında bulunan farklar anlamlı mıdır, bir fabrikanın ürettiği mallardan bazıları alınarak oluşturulan örnek, bu fabrikanın ürettiği tüm malları yansıtır mı?

Hipotez Oluşturmak

Bir duruma karar verirken öncelikle bir hipotez oluştururuz; daha sonra elimizdeki verinin hipotezimizi destekleyip desteklemediğine bakarız. Eğer veri hipotezi desteklemezse, hipotezi redederiz, desteklerse hipotezi reddetmeyiz.

اتخاذ القرار

تمامًا كما هو الحال في المتوسط السكاني، يمكننا أيضًا إنشاء نطاق ثقة للمتوسط السكاني. في وقت سابق، استخدمنا التوزيع الطبيعي للاقتراب من التوزيع الثنائي.

حتى الآن كنا مهتمين بتقدير المعلومات السكانية (المتوسط، نسبة السكان). ولكن عمل المهندس ليس مقتصرًا على هذا فقط. عادةً ما يُتوقع من المهندس اتخاذ قرار بناءً على البيانات المتاحة. على سبيل المثال، هل الفروق بين القياسات التي أجرتها جهازين قياس مختلفين ذات دلالة؟ وهل العينة المأخوذة من بعض المنتجات التي تم إنتاجها في مصنع ما تمثل جميع المنتجات التي يتم إنتاجها في هذا المصنع؟

إنشاء الفرضية

عند اتخاذ القرار، نبدأ أولاً بإنشاء فرضية، ثم نتحقق من البيانات المتاحة لدعم هذه الفرضية أو عدم دعمها. إذا لم تدعم البيانات الفرضية، نقوم برفضها، وإذا دعمت الفرضية، لا نقوم برفضها.

İstatistik

Karar Vermek



Karar Verme Süreci

ör. Varsayalımki hipotezimiz bir fabrikanın ürettiđi malların %10'nun bozuk olduđudur. Bu hipotezi test etmek için fabrikanın ürettiđi mallardan rastgele 100 tane seçtiđimizi ve bunların 20 tanesinin bozuk olduđunu varsayalım. Eğer hipotezimiz doğru olaysaydı, böyle bir veriyi elde etme olasılıđımız çok düşük olurdu. Gerçekten: $p = 0.1$ iken 100 denemede toplam 20 başarının olması olasılıđı:

$$P(X = 20) = \binom{100}{20} 0.1^{20} 0.9^{80} = 0.002$$

Bu durumda elimizdeki veri hipotezi desteklemediđi için, hipotezimiz reddederiz.

İstatistik

Neyin Hipotez Olacağı

Gördükki elimizdeki veri hipotezi desteklemediğinde (yani hipotezin varlığında elimizdeki gibi bir verinin elde edilme olasılığı çok düşük olduğunda) hipotezi reddediyoruz, veri hipotezi desteklediğinde hipotezi reddetmiyoruz.

Yani karar verme sürecinde aksiyonlarımız yalnızca iki tanedir:

ya: hipotezi reddetmek,

yada : hipotezi reddetmemek.

Başka bir aksiyon mümkün değildir.

اتخاذ القرار

ما هو الفرض الذي سيتم اختبارهما رأينا، عندما لا تدعم البيانات الفرضية (أي عندما تكون احتمالية الحصول على مثل هذه البيانات في حال كانت الفرضية صحيحة منخفضة جدًا)، نقوم برفض الفرضية. وعندما تدعم البيانات الفرضية، لا نقوم برفض الفرضية.

بالتالي، في عملية اتخاذ القرار لدينا فقط خياران:

إما: رفض الفرضية،

أو: عدم رفض الفرضية.

لا يوجد خيار آخر.

İstatistik

Not!!!: Hipotezi reddetmemek hipotezi kabul etmek anlamına gelmez; yalnızca eldeki verinin hipotezi desteklemediği anlamına gelir. Eğer kanıtlamaya çalıştığımız şeyin tersini hipotez olarak düşünürsek, hipotezi reddetme durumunda kanıtlamaya çalıştığımız şeyin tersini reddetmiş oluruz. Yani tersi yanlış olur, böylece kendisi doğru olur. Bu, aynı Olmayan Ergi Yöntemi gibidir. Kanıtlamak istediğimiz şeyin tersinin yanlış olduğunu, böylece kendinin doğru olduğunu göstermeye çalışıyoruz.

ملاحظة!!

عدم رفض الفرضية لا يعني قبول الفرضية؛ بل يعني فقط أن البيانات المتاحة لا تدعم الفرضية. إذا فكرنا في العكس مما نحاول إثباته كفرضية، فإن رفض الفرضية يعني أننا قد رفضنا العكس مما نحاول إثباته. أي أن العكس يكون خاطئاً، وبالتالي يكون هو الصحيح. هذا يشبه طريقة الاستدلال بالنفي. نحن نحاول إثبات أن العكس مما نحاول إثباته خاطئ، وبالتالي نثبت صحة ما نحاول إثباته.

İstatistik

Hipotez Türleri:

1) Sıfır Hipotezi (Null Hypothesis)

Sıfır hipotezi göstermek (kanıtlamak) istediğimiz tezin mantıksal tersidir. Sıfır hipotezi H_0 ile gösterilir.

2) Alternatif Hipotezi (Alternative Hypothesis)

Alternatif hipotez göstermek (kanıtlamak) istediğimiz tezin kendisidir. Sıfır hipotezinin tersidir; H_1 ile gösterilir.

Sonuç olarak sıfır hipotezine, hipotezimiz gibi davranırız. Eğer elimizdeki veri sıfır hipotezini desteklemezse sıfır hipotezi reddedilir, alternatif hipotez kabul edilir. Eldeki veri sıfır hipotezini desteklerse, sıfır hipotezi rededilmez (ama kabul de edilmez).

اتخاذ القرار

1) الفرضية الصفرية

الفرضية الصفرية هي عكس ما نحاول إثباته (إثباته). يتم تمثيل الفرضية الصفرية بالرمز H_0 .

2) الفرضية البديلة

الفرضية البديلة هي ما نحاول إثباته (إثباته). هي العكس للفرضية الصفرية؛ ويتم تمثيلها بالرمز H_1 .

بناءً على ذلك، نتعامل مع الفرضية الصفرية كما لو كانت فرضيتنا. إذا كانت البيانات المتاحة لا تدعم الفرضية الصفرية، يتم رفض الفرضية الصفرية ويتم قبول الفرضية البديلة. إذا كانت البيانات تدعم الفرضية الصفرية، فلا يتم رفض الفرضية الصفرية (ولكن لا يتم قبولها أيضًا).

İstatistik

ör. Diyelimki aşırı kahve tüketiminin kansere yol açtığını göstermek istiyoruz. Bu durumda:

H_0 : Aşırı kahve tüketmek kansere neden olmaz.

H_1 : Aşırı kahve tüketmek kansere neden olur.

ör. Yeni bir ilaç çıkaran bir ilaç firması, bu ilacın mevcut ilaçdan daha etkili olduğunu göstermek istiyor. Bunun için iki grup insan topluyor: çalışma grubu ve kontrol grubu.

Çalışma grubundakilere yeni ilaç veriliyor; kontrol grubundakilere mevcut olan ilaç veriliyor.

H_0 : Kontrol ve çalışma grupları arasında fark yoktur.

H_1 : Kontrol ve çalışma grupları arasında fark vardır.

ör. Bir savcı sanığın suçlu olduğuna inanıyor.

H_0 : Sanık suçsuzdur.

H_1 : Sanık suçludur.

Eldeki veri (bulgu) sanığın suçsuz olduğu ile çelişirse, sanığın suçsuz olduğunu reddeder; sanığın suçlu olduğunu düşürüz; eğer bu veri sanığın suçsuz olduğu ile çelişmezse; sanığın suçsuzluğunu reddedemeyiz, sanığın suçsuz olduğunu çürütecek elimizde yeterince veri yoktur deriz.

مثال: لنفترض أننا نريد إثبات أن استهلاك القهوة المفرط يسبب السرطان. في هذه الحالة:

H_0 : استهلاك القهوة المفرط لا يسبب السرطان.

H_1 : استهلاك القهوة المفرط يسبب السرطان.

مثال آخر: شركة أدوية أصدرت دواءً جديدًا وترغب في إثبات أنه أكثر فعالية من الدواء الحالي. لهذا الغرض، تجمع الشركة مجموعتين من الأشخاص: مجموعة الدراسة ومجموعة التحكم. يتم إعطاء الدواء الجديد للمجموعة الدراسية؛ بينما يُعطى الدواء الحالي لمجموعة التحكم.

H_0 : لا يوجد فرق بين مجموعة التحكم ومجموعة الدراسة.

H_1 : يوجد فرق بين مجموعة التحكم ومجموعة الدراسة.

مثال آخر: مدعٍ عام يعتقد أن المتهم مذنب.

H_0 : المتهم بريء.

H_1 : المتهم مذنب. إذا كانت البيانات (الأدلة) تتناقض مع براءة المتهم، نرفض براءته ونبنى فكرة أنه مذنب. إذا لم تتناقض البيانات مع براءته، فلا يمكننا رفض براءته، ونقول أنه لا يوجد دليل كافٍ لدحض براءته.

KAYNAKLAR

➡ Olasılık ve İstatistik Notları, Doç.Dr. FIRAT İSMAİLOĞLU,
Sivas Cumhuriyet Üniversitesi, Bilgisayar Mühendisliği Bölümü, 2024

➡ Morris H DeGroot_ Mark J Schervish-Probability and statistics-Pearson
Education (2012)

➡ Olasılık ve İstatistik Notları, Dr.Mehmet Kazım Yetik, Karabük Üniversitesi,
2024



Eng: Abdulrahman Hamdi