**Interview Questions About Statistics**

**What is statistics, and what are branches of statistics?**

Statistics is mathematical science pertaining to the collection, presentation, analysis, and interpretation of data. It is used for:

* Analyze Primary Data
* Build a Statistical Model
* Predict the outcomes

The two main branches of statistics are [descriptive statistics](https://explorable.com/descriptive-statistics) and [inferential statistics](https://explorable.com/inferential-statistics).

**Please explain descriptive statistics.**

Helps organize data and focuses on the main characteristics of the data. It provides a summary of data numerically or graphically. It describes the basic features of the data in a study, such as mean, median, mode, SD, correlation.

Descriptive statistics are either measures of central tendency or measures of spread.

* Measures of central tendency: mean, median, and mode
* Measures of spread: standard deviation, variance, the minimum and maximum variables, range and the kurtosis and skewness.

**What are the data types? Please explain.**

1. Numerical data represents some quantifiable information that is measurable and is further divided into two subcategories:

* Discrete data, which is integer based (e.g. number of people)
* Continuous data, which is decimal based (e.g. price, distance, temperature).

2. Categorical data is qualitative data that is used to classify data into categories (think of an enumeration in programming). For example, gender, car brands, country of residence etc.

**Interviewfragen zu Statistiken**

**Was ist Statistik und was sind Zweige der Statistik?**

Statistik ist eine mathematische Wissenschaft, die sich mit der Erhebung, Darstellung, Analyse und Interpretation von Daten befasst. Es wird genutzt für:

• Primärdaten analysieren

• Erstellen Sie ein statistisches Modell

• Sagen Sie die Ergebnisse voraus

Die beiden Hauptzweige der Statistik sind die deskriptive Statistik und die Inferenzstatistik.

**Bitte erläutern Sie die deskriptive Statistik.**

Hilft beim Organisieren von Daten und konzentriert sich auf die Hauptmerkmale der Daten. Es bietet eine numerische oder grafische Zusammenfassung der Daten. Es beschreibt die grundlegenden Merkmale der Daten in einer Studie, wie z. B. Mittelwert, Median, Modus, Standardabweichung, Korrelation.

Die deskriptive Statistik ist entweder ein Maß für die zentrale Tendenz oder ein Maß für die Streuung.

• Maße der zentralen Tendenz: Mittelwert, Median und Modus

• Streuungsmaße: Standardabweichung, Varianz, minimale und maximale Variablen, Bereich sowie Kurtosis und Schiefe.

**Welche Datentypen gibt es? Bitte erkläre.**

1. Numerische Daten stellen einige quantifizierbare Informationen dar, die messbar sind und weiter in zwei Unterkategorien unterteilt sind:

• Diskrete Daten, die ganzzahlbasiert sind (z. B. Anzahl der Personen)

• Kontinuierliche Daten, die auf Dezimalbasis basieren (z. B. Preis, Entfernung, Temperatur).

2. Kategoriale Daten sind qualitative Daten, die verwendet werden, um Daten in Kategorien zu klassifizieren (denken Sie an eine Aufzählung in der Programmierung). Zum Beispiel Geschlecht, Automarken, Wohnsitzland etc.

**İstatistiklerle İlgili Mülakat Soruları**

**İstatistik nedir ve istatistiğin dalları nelerdir?**

İstatistik, verilerin toplanması, sunumu, analizi ve yorumlanması ile ilgili matematik bilimidir. İçin kullanılır:

• Birincil Verileri Analiz Edin

• İstatistiksel Bir Model Oluşturun

• Sonuçları tahmin edin

İstatistiğin iki ana dalı, tanımlayıcı istatistikler ve çıkarımsal istatistiklerdir.

**Lütfen tanımlayıcı istatistikleri açıklayın.**

Verileri düzenlemeye yardımcı olur ve verilerin ana özelliklerine odaklanır. Sayısal veya grafiksel olarak verilerin bir özetini sağlar. Ortalama, medyan, mod, SD, korelasyon gibi bir çalışmadaki verilerin temel özelliklerini tanımlar.

Tanımlayıcı istatistikler ya merkezi eğilim ölçüleri ya da yayılma ölçüleridir.

• Merkezi eğilim ölçüleri: ortalama, medyan ve mod

• Yayılma ölçüleri: standart sapma, varyans, minimum ve maksimum değişkenler, aralık ve basıklık ve çarpıklık.

**Veri türleri nelerdir? Lütfen açıkla.**

1. Sayısal veriler, ölçülebilir ve ayrıca iki alt kategoriye ayrılan bazı niceliksel bilgileri temsil eder:

• Tamsayıya dayalı ayrık veriler (örn. kişi sayısı)

• Ondalık tabanlı sürekli veriler (örneğin fiyat, mesafe, sıcaklık).

2. Kategorik veriler, verileri kategorilere ayırmak için kullanılan nitel verilerdir (programlamada bir numaralandırma düşünün). Örneğin cinsiyet, otomobil markaları, ikamet ettiğiniz ülke vb.

3. Ordinal data represents discrete and ordered units, e.g. champions league rank (1st, 2nd, 3rd), bug priority (low, critical or showstopper), or hotel rating (1–5\*).

**What is variable and what are categorical, discrete, and continuous variables?**

A variable is a characteristic that describes a member of the sample.

* Categorical variables contain a finite number of categories or distinct groups. Gender, material type, and payment method.
* Discrete variables are numeric variables that have a countable number of values between any two values.
* Continuous variables are numeric variables that have an infinite number of values between any two values.

**What is the Central Limit Theorem and why is it important?**

The CLT states that the mean values from a group of samples will be normally distributed about the population mean, even if the population itself is not normally distributed.

The central limit theorem states that, when samples from a data set with a known variance are aggregated, their mean roughly equals the population mean.

Thus, even though we might not know the shape of the distribution where our data comes from, the central limit theorem says that we can treat the sampling distribution as if it were normal.

Suppose that we are interested in estimating the average height among all people. Collecting data for every person in the world is impossible. While we can’t obtain a height measurement from everyone in the population, we can still sample some people. The question now becomes, what can we say about the average height of the entire population given a single sample. The Central Limit Theorem addresses this question exactly.

3. Ordinale Daten stellen diskrete und geordnete Einheiten dar, z. Champions-League-Rang (1., 2., 3.), Fehlerpriorität (niedrig, kritisch oder Showstopper) oder Hotelbewertung (1–5\*).

**Was ist variabel und was sind kategoriale, diskrete und kontinuierliche Variablen?**

Eine Variable ist ein Merkmal, das ein Mitglied der Stichprobe beschreibt.

• Kategoriale Variablen enthalten eine endliche Anzahl von Kategorien oder unterschiedlichen Gruppen. Geschlecht, Materialtyp und Zahlungsmethode.

• Diskrete Variablen sind numerische Variablen, die eine zählbare Anzahl von Werten zwischen zwei beliebigen Werten haben.

• Kontinuierliche Variablen sind numerische Variablen, die eine unendliche Anzahl von Werten zwischen zwei beliebigen Werten haben.

**Was ist der zentrale Grenzwertsatz und warum ist er wichtig?**

Die CLT gibt an, dass die Mittelwerte aus einer Gruppe von Stichproben um den Mittelwert der Grundgesamtheit normalverteilt sind, selbst wenn die Grundgesamtheit selbst nicht normalverteilt ist.

Der zentrale Grenzwertsatz besagt, dass, wenn Stichproben aus einem Datensatz mit bekannter Varianz aggregiert werden, ihr Mittelwert ungefähr gleich dem Mittelwert der Grundgesamtheit ist.

Obwohl wir die Form der Verteilung, aus der unsere Daten stammen, möglicherweise nicht kennen, besagt der zentrale Grenzwertsatz, dass wir die Stichprobenverteilung so behandeln können, als ob sie normal wäre.

Angenommen, wir sind daran interessiert, die durchschnittliche Größe aller Menschen zu schätzen. Es ist unmöglich, Daten für jeden Menschen auf der Welt zu sammeln. Wir können zwar nicht von jedem in der Bevölkerung eine Körpergröße messen, aber wir können dennoch einige Personen beproben. Nun stellt sich die Frage, was wir über die durchschnittliche Größe der gesamten Bevölkerung bei einer einzigen Stichprobe sagen können. Der Zentrale Grenzwertsatz befasst sich genau mit dieser Frage.

3. Sıralı veriler, ayrık ve sıralı birimleri temsil eder, örn. şampiyonlar ligi sıralaması (1., 2., 3.), hata önceliği (düşük, kritik veya şaşırtıcı) veya otel derecelendirmesi (1-5\*).

**Değişken nedir ve kategorik, ayrık ve sürekli değişkenler nelerdir?**

Değişken, örneğin bir üyesini tanımlayan bir özelliktir.

• Kategorik değişkenler, sınırlı sayıda kategori veya farklı grup içerir. Cinsiyet, malzeme türü ve ödeme yöntemi.

• Ayrık değişkenler, herhangi iki değer arasında sayılabilir sayıda değere sahip sayısal değişkenlerdir.

• Sürekli değişkenler, herhangi iki değer arasında sonsuz sayıda değere sahip sayısal değişkenlerdir.

**Merkezi Limit Teoremi nedir ve neden önemlidir?**

CLT, bir örnek grubundan elde edilen ortalama değerlerin, popülasyonun kendisi normal olarak dağılmasa bile popülasyon ortalaması hakkında normal olarak dağılacağını belirtir.

Merkezi limit teoremi, bilinen bir varyansa sahip bir veri setinden örnekler toplandığında, ortalamalarının kabaca popülasyon ortalamasına eşit olduğunu belirtir.

Bu nedenle, verilerimizin geldiği dağılımın şeklini bilmesek bile, merkezi limit teoremi, örnekleme dağılımına normalmiş gibi davranabileceğimizi söylüyor.

Tüm insanların ortalama boyunu tahmin etmekle ilgilendiğimizi varsayalım. Dünyadaki her insan için veri toplamak imkânsızdır. Popülasyondaki herkesten boy ölçümü elde edemesek de yine de bazı insanları örnekleyebiliriz. Şimdi soru, tek bir örnek verilen tüm popülasyonun ortalama yüksekliği hakkında ne söyleyebiliriz. Merkezi Limit Teoremi tam olarak bu soruyu ele alır.

**Bitte erläutern Sie Populations- und Beispielbegriffe. Was ist der Unterschied zwischen einer Grundgesamtheit und einer Stichprobe?**

***Population***: Eine Bevölkerung ist die Gruppe, aus der Daten erhoben werden sollen.

**Stichprobe**: Eine Teilmenge der Bevölkerung

Eine Population beschreibt alle Mitglieder einer Gruppe, während eine Stichprobe eine Teilmenge von Mitgliedern ist, die Sie mit Zeit und Ressourcen messen können.

**Was sind die Maße der Variabilität (Streuung)? Bitte erkläre.**

Range: Range ist die Differenz zwischen der niedrigsten und der höchsten Zahl eines Datensatzes. Um die Reichweite zu berechnen, subtrahieren wir den Mindestwert vom Höchstwert.

Es zeigt uns, wie vielfältig der Datensatz ist, d. h. wie gestreut er ist, aber auch hier ist er, wie der Mittelwert, sehr empfindlich gegenüber Ausreißern.

Varianz: Die Varianz misst, wie verteilt die Daten sind. Sie beschreibt, wie weit Werte vom Mittelwert entfernt liegen. Berechnet als Summe der quadratischen Abstände von jedem Punkt zum Mittelwert.

* Mittelwert berechnen.
* Nehmen Sie die Differenz zwischen jedem Wert und dem Mittelwert
* Quadrieren Sie diese Differenz.
* Summieren Sie alle Unterschiede
* Teilen Sie schließlich durch die Gesamtzahl der Beobachtungen.

Es gibt einen Unterschied zwischen der SAMPLE-Varianz und der POPULATION-Varianz, die der Bessel-Korrektur unterliegt (𝒏 − 𝟏)

* Für die vollständige Population dividieren wir durch die Anzahl der Datenpunkte (n)
* Proben, wir dividieren durch die Anzahl der Datenpunkte minus 1(n - 1)

**Standardabweichung:** Die Standardabweichung (dargestellt durch den griechischen Buchstaben Sigma — σ) ist einfach die Quadratwurzel der Varianz. Es ist ein Maß für die Streuung in Bezug darauf, wie viele Standardabweichungen es vom Mittelwert entfernt ist, und wird verwendet, um zu beurteilen, welcher Datenpunkt ein Ausreißer ist.

Vorteil: Gleiche Einheiten wie die Probe und aussagekräftig.

**Lütfen ana kütleyi ve örnek terimleri açıklayınız. Bir popülasyon ve bir örnek arasındaki fark nedir?**

***Popülasyon***: Nüfus, verilerin toplanacağı gruptur.

**Örnek**: Popülasyonun bir alt kümesi

Bir popülasyon, bir grubun tüm üyelerini tanımlarken, bir örneklem, zaman ve kaynakların ölçmenize izin verdiği bir üye alt kümesidir.

**Değişkenlik (dağılım) ölçüleri nelerdir? Lütfen açıklayiniz.**

Aralık: Aralık, bir veri kümesinin en düşük ve en yüksek sayısı arasındaki farktır. Aralığı hesaplamak için minimum değeri maksimum değerden çıkarırız.

Bize veri kümesinin ne kadar çeşitli olduğunu, yani ne kadar yayılmış olduğunu gösterir, ancak yine, ortalama gibi, aykırı değerlere karşı gerçekten hassastır.

Varyans: Varyans, verilerin ne kadar yayıldığını ölçer. Değerlerin ortalamadan ne kadar uzakta olduğunu açıklar. Her noktadan ortalamaya olan uzaklıkların karelerinin toplamı olarak hesaplanır.

• Ortalama hesaplayın.

• Her değer ile ortalama arasındaki farkı alın

• Bu farkın karesini alın.

• Tüm farklılıkları toplayın

• Son olarak, toplam gözlem sayısına bölün.

NUMUNE varyansı ile Bessel'in düzeltmesine tabi POPÜLASYON varyansı arasında bir fark var (𝒏 − 𝟏)

• Tam popülasyon için veri noktası sayısına böleriz (n)

• örnekler, eksi 1(n - 1) veri noktası sayısına böleriz

**Standart Sapma:** Standart sapma (Yunanca sigma — σ harfiyle gösterilir) varyansın sadece kareköküdür. Ortalamadan kaç standart sapma uzakta olduğuna göre bir dağılım ölçüsüdür ve hangi veri noktasının aykırı olduğunu yargılamak için kullanılır.

**Fayda**: Örnekle aynı birimler ve üzerinde konuşulması anlamlı.

**Please explain population and sample terms. What is the difference between a population and a sample?**

***Population***: A population is the group from which data is to be collected.

***Sample***: A subset of population

A population describes all of the members of a group, while a sample is a subset of members that time and resources allow you to measure.

**What are the measures of variability (dispersion)? Please explain.**

Range: Range is the difference between the lowest and the highest number of a dataset. To calculate the range, we subtract the minimum from the maximum value.

It shows us how varied the dataset is, i.e. how spread it is, but again, like mean, it is really sensitive to outliers.

***Variance***: Variance measures how spread out the data is. It describes how far values lie from the mean. Calculated as the sum of square distances from each point to the mean.

* Calculate mean.
* Take difference between each value and the mean
* Square this difference.
* Sum all differences
* Finally, divide by the total number of observations.

There’s a difference between the SAMPLE variance and the POPULATION variance subject to Bessel's correction (𝒏 − 𝟏)

* For full population, we divide by the number of data points (n)
* samples, we divide by the number of data points minus 1(n - 1)

***Standard Deviation:*** Standard deviation (represented by the Greek letter sigma — σ) is just the square root of the variance. It is a measure of dispersion in terms of how many standard deviations it is away from the mean, and used to judge which data point is an outlier.

***Benefit***: Same units as the sample and meaningful to talk about.

**What is an interquartile range?**

The IQR is a number that indicates how spread the middle half (i.e. the middle 50%) of the dataset is and can help determine outliers. It is the difference between the Q3 and Q1.

Generally speaking, outliers are those data points that fall outside from the Q1 – 1.5 x IQR and Q3 + 1.5 x IQR range.

**What are levels of measurement?**

Based on the nature of variables, it is classified into four types:

**Nominal Variables**: They are ones which have two or more categories, and it is impossible to order the values. Predetermined categories and can’t be sorted.

Ex. gender and blood group

**Ordinal Variables:** They have values in a logical order. However, the relative distance between two data values is not clear. Can be sorted and lacks scale.

Ex. the size of a coffee cup – large, medium, and small

The ratings of a product – bad, good, and best

**Interval Variable:** With an interval scale, equal differences between scale values do have equal quantitative meaning. An interval scale provides more quantitative information than the ordinal scale. The interval scale does not have a true zero point. Provides scale and lacks a ‘zero’ point.

Ex. The Fahrenheit degree scale used to measure temperature. The distance between two compartments in a train.

**Ratio Variable:** They are similar to interval scales in that equal differences between scale values have equal quantitative meaning. It has a true zero point.

Ex. the system of inches used with a common ruler.

**Was ist ein Quartilsabstand?**

Der IQR ist eine Zahl, die angibt, wie gestreut die mittlere Hälfte (d. h. die mittleren 50 %) des Datensatzes ist, und dabei helfen kann, Ausreißer zu ermitteln. Es ist der Unterschied zwischen Q3 und Q1.

Im Allgemeinen sind Ausreißer diejenigen Datenpunkte, die außerhalb des Bereichs Q1 – 1,5 x IQR und Q3 + 1,5 x IQR liegen.

**Was sind Messniveaus?**

Basierend auf der Art der Variablen wird sie in vier Typen eingeteilt:

**Nominale Variablen:** Sie haben zwei oder mehr Kategorien, und es ist unmöglich, die Werte zu ordnen. Vorgegebene Kategorien und können nicht sortiert werden.

Ex. Geschlecht und Blutgruppe

**Ordnungsvariablen:** Sie haben Werte in einer logischen Reihenfolge. Der relative Abstand zwischen zwei Datenwerten ist jedoch nicht klar. Kann sortiert werden und hat keinen Maßstab.

Ex. die Größe einer Kaffeetasse – groß, mittel und klein

Die Bewertungen eines Produkts – schlecht, gut und am besten

**Intervallvariable:** Bei einer Intervallskala haben gleiche Unterschiede zwischen Skalenwerten die gleiche quantitative Bedeutung. Eine Intervallskala liefert mehr quantitative Informationen als die Ordinalskala. Die Intervallskala hat keinen echten Nullpunkt. Bietet Skalierung und keinen „Null“-Punkt.

Ex. Die Fahrenheit-Grad-Skala zur Messung der Temperatur. Der Abstand zwischen zwei Abteilen in einem Zug.

**Verhältnisvariable:** Sie ähneln Intervallskalen darin, dass gleiche Unterschiede zwischen Skalenwerten die gleiche quantitative Bedeutung haben. Es hat einen wahren Nullpunkt.

Ex. das Zollsystem, das mit einem gewöhnlichen Lineal verwendet wird.

**Çeyrekler arası aralık nedir?**

IQR, veri kümesinin orta yarısının (yani orta %50'sinin) ne kadar yayılmış olduğunu gösteren ve aykırı değerlerin belirlenmesine yardımcı olabilecek bir sayıdır. Q3 ve Q1 arasındaki fark budur.

Genel olarak, aykırı değerler, Q1 – 1.5 x IQR ve Q3 + 1.5 x IQR aralığının dışında kalan veri noktalarıdır.

**Ölçüm seviyeleri nelerdir?**

Değişkenlerin doğasına göre dört türe ayrılır:

**Nominal Değişkenler:** İki veya daha fazla kategoriye sahip olan ve değerleri sıralamak mümkün olmayan değişkenlerdir. Önceden belirlenmiş kategoriler ve sıralanamaz.

Örn. cinsiyet ve kan grubu

**Ordinal Değişkenler:** Mantıksal bir sırada değerlere sahiptirler. Ancak, iki veri değeri arasındaki göreli mesafe net değildir. Sıralanabilir ve ölçeği yoktur.

Örn. bir kahve fincanı boyutu - büyük, orta ve küçük

Bir ürünün derecelendirmeleri – kötü, iyi ve en iyi

**Aralık Değişkeni:** Bir aralık ölçeğinde, ölçek değerleri arasındaki eşit farklar eşit nicel anlama sahiptir. Bir aralık ölçeği, sıralı ölçekten daha fazla nicel bilgi sağlar. Aralık ölçeğinin gerçek bir sıfır noktası yoktur. Ölçek sağlar ve 'sıfır' noktasından yoksundur.

Örn. Sıcaklığı ölçmek için kullanılan Fahrenheit derece ölçeği. Bir trende iki kompartıman arasındaki mesafe.

**Oran Değişkeni:** Ölçek değerleri arasındaki eşit farkların eşit nicel anlama sahip olması bakımından aralıklı ölçeklere benzerler. Gerçek bir sıfır noktası vardır.

Örn. ortak bir cetvelle kullanılan inç sistemi.

**What is the difference between covariance and correlation?**

Correlation and Co-variance both are used as a measure to check how two variables change with respect to each other.

**Covariance**:

* Measure of change of how two variables change with respect to each other.
* It is unit dependent.
* Difference in scale can output different co-variance.
* It varies from - infinity to infinity
* Example: Height vs weight (kg) and Height vs weight (lbs) will have different covariance values

**Correlation**:

* Unit measure of change between two variables change with respect to each other. (Basically it’s normalized covariance value)
* It is unit independent.
* Difference in scale doesn’t affect the correlation value
* It varies from -1 to 1
* Example: Height vs weight(kg) and Height vs weight(lbs) will have same correlation values.

**What are depend or independ events?**

**Independent Events**

An independent series of events occur when the outcome of one event has no effect on the outcome of another.

Independent are those events whose outcome does not influence the probability of the outcome of the other event. Due to this reason:

P(A|B) = P(A)

P(B|A) = P(B)

**Dependent Events**

A dependent event occurs when the outcome of a first event does affect the probability of a second event.

**Was ist der Unterschied zwischen Kovarianz und Korrelation?**

Korrelation und Kovarianz werden beide als Maß verwendet, um zu überprüfen, wie sich zwei Variablen in Bezug zueinander ändern.

**Kovarianz:**

* Änderungsmaß dafür, wie sich zwei Variablen relativ zueinander ändern.
* Es ist einheitenabhängig.
* Skalenunterschiede können unterschiedliche Kovarianzen ausgeben.
* Es variiert von - unendlich bis unendlich
* Beispiel: Größe vs. Gewicht (kg) und Größe vs. Gewicht (lbs) haben unterschiedliche Kovarianzwerte

**Korrelation**:

- Einheitsmaß der Änderung zwischen zwei Variablen, die sich relativ zueinander ändern. (Im Grunde ist es ein normalisierter Kovarianzwert)

* Es ist geräteunabhängig.
* Skalenunterschiede wirken sich nicht auf den Korrelationswert aus
* Es variiert von -1 bis 1
* Beispiel: Größe vs. Gewicht (kg) und Größe vs. Gewicht (lbs) haben dieselben Korrelationswerte.

**Was sind abhängige oder unabhängige Ereignisse?**

**Unabhängige Veranstaltungen**

Eine unabhängige Reihe von Ereignissen tritt auf, wenn das Ergebnis eines Ereignisses keine Auswirkung auf das Ergebnis eines anderen hat.

Unabhängig sind solche Ereignisse, deren Ausgang die Wahrscheinlichkeit des Ausgangs des anderen Ereignisses nicht beeinflusst. Aus diesem Grund:

P(A|B) = P(A)

P(B|A) = P(B)

**Abhängige Ereignisse**

Ein abhängiges Ereignis tritt auf, wenn das Ergebnis eines ersten Ereignisses die Wahrscheinlichkeit eines zweiten Ereignisses beeinflusst.

**Kovaryans ve korelasyon arasındaki fark nedir?**

Korelasyon ve Co-varyans, iki değişkenin birbirine göre nasıl değiştiğini kontrol etmek için bir ölçü olarak kullanılır.

**kovaryans:**

* İki değişkenin birbirine göre nasıl değiştiğinin değişiminin ölçüsü.
* Birim bağımlıdır.
* Ölçekteki fark, farklı ortak varyans üretebilir.
* Sonsuzdan sonsuza kadar değişir
* Örnek: Boy - ağırlık (kg) ve Boy - ağırlık (lbs) farklı kovaryans değerlerine sahip olacaktır

**Korelasyon**:

- İki değişken arasındaki değişimin birim ölçüsü birbirine göre değişir. (Temel olarak normalleştirilmiş kovaryans değeridir)

* Birim bağımsızdır.
* Ölçekteki farklılık korelasyon değerini etkilemez
* -1 ile 1 arasında değişir
* Örnek: Boy - kilo(kg) ve Boy - kilo(lbs) aynı korelasyon değerlerine sahip olacaktır.

**Bağımlı veya bağımsız olaylar nelerdir?**

**Bağımsız Etkinlikler**

Bağımsız bir olaylar dizisi, bir olayın sonucunun diğerinin sonucu üzerinde hiçbir etkisi olmadığında meydana gelir.

Bağımsız, sonucu diğer olayın sonucunun olasılığını etkilemeyen olaylardır. Bu sebepten dolayı:

P(A|B) = P(A)

P(B|A) = P(B)

**Bağımlı Olaylar**

Bir ilk olayın sonucu ikinci bir olayın olasılığını etkilediğinde bağımlı bir olay meydana gelir.

**What is conditional probability? Please explain.**

Conditional probability is the likelihood of an event occurring, based on the occurrence of a previous event.

The notation for conditional probability is P(A|B), read as ‘the probability of A given B’. The formula for conditional probability is:

The probability of A given B is equal to the probability of A and B occurring over the probability of B alone occurring.

Conditional probability does not indicate that there is necessarily a causal relationship between the two events, nor that both events occur simultaneously.

**What is Bayes' Teorem? Please explain.**

Bayes' theorem describes the [probability](https://en.wikipedia.org/wiki/Probability) of an [event](https://en.wikipedia.org/wiki/Event_(probability_theory)), based on prior knowledge of conditions that might be related to the event. The probability of A given B is equal to the probability of B given A times the probability of A, over the probability of B.

For example, if cancer is related to age, then, using Bayes' theorem, a person's age can be used to more accurately assess the probability that they have cancer.

A is called the proposition and B is called the evidence.

P(A) is called the prior probability of proposition and P(B) is called the prior probability of evidence.

P(A|B) is called the posterior.     P(B|A) is the likelihood.

Posterior= (Likelihood)\*(Proposition prior probability) / Evidence prior probability

The more variables that are in play, and the more certain we become about those variables, the more certain an accurate conclusion can be drawn, using conditional probabilities!

**Was ist bedingte Wahrscheinlichkeit? Bitte erkläre.**

Die bedingte Wahrscheinlichkeit ist die Wahrscheinlichkeit, dass ein Ereignis eintritt, basierend auf dem Eintreten eines früheren Ereignisses.

Die Notation für bedingte Wahrscheinlichkeit ist P(A|B), gelesen als „die Wahrscheinlichkeit von A gegeben B“. Die Formel für die bedingte Wahrscheinlichkeit lautet:

Die Wahrscheinlichkeit, dass A bei B gegeben ist, ist gleich der Wahrscheinlichkeit, dass A und B auftreten, gegenüber der Wahrscheinlichkeit, dass B allein auftritt.

Die bedingte Wahrscheinlichkeit gibt nicht an, dass notwendigerweise ein Kausalzusammenhang zwischen den beiden Ereignissen besteht oder dass beide Ereignisse gleichzeitig auftreten.

**Was ist Bayes' Teorem? Bitte erkläre.**

Das Theorem von Bayes beschreibt die Wahrscheinlichkeit eines Ereignisses, basierend auf vorherigem Wissen über Bedingungen, die mit dem Ereignis zusammenhängen könnten. Die Wahrscheinlichkeit von A bei B ist gleich der Wahrscheinlichkeit von B bei A multipliziert mit der Wahrscheinlichkeit von A über der Wahrscheinlichkeit von B.

Wenn beispielsweise Krebs mit dem Alter zusammenhängt, kann das Alter einer Person unter Verwendung des Bayes-Theorems verwendet werden, um die Wahrscheinlichkeit, dass sie an Krebs erkrankt ist, genauer einzuschätzen.

A heißt Aussage und B heißt Beweis.

P(A) heißt A-Propositionswahrscheinlichkeit und P(B) A-Beweiswahrscheinlichkeit.

P(A|B) wird als Posterior bezeichnet. P(B|A) ist die Wahrscheinlichkeit.

Posterior = (Wahrscheinlichkeit)\*(Vorherige Propositionswahrscheinlichkeit) / Evidenz-vorherige Wahrscheinlichkeit

Je mehr Variablen im Spiel sind und je sicherer wir uns über diese Variablen werden, desto sicherer kann eine genaue Schlussfolgerung gezogen werden, indem bedingte Wahrscheinlichkeiten verwendet werden!

**Koşullu olasılık nedir? Lütfen açıkla.**

Koşullu olasılık, önceki bir olayın meydana gelmesine bağlı olarak bir olayın meydana gelme olasılığıdır.

Koşullu olasılığın gösterimi P(A|B)'dir ve "A'nın verilen B olasılığı" olarak okunur. Koşullu olasılık formülü:

Verilen bir B'nin olasılığı, A ve B'nin meydana gelme olasılığına, B'nin tek başına meydana gelme olasılığına eşittir.

Koşullu olasılık, iki olay arasında mutlaka bir nedensel ilişki olduğunu veya her iki olayın aynı anda meydana geldiğini göstermez.

**Bayes Teoremi nedir? Lütfen açıkla.**

Bayes teoremi, olayla ilgili olabilecek koşulların ön bilgisine dayalı olarak bir olayın olasılığını tanımlar. Verilen A'nın olasılığı, verilen A'nın olasılığı çarpı A'nın olasılığı bölü B'nin olasılığına eşittir.

Örneğin, kanser yaşla ilgiliyse, o zaman Bayes teoremi kullanılarak bir kişinin yaşı, kanser olma olasılığını daha doğru bir şekilde değerlendirmek için kullanılabilir.

A'ya önerme, B'ye kanıt denir.

P(A) önermenin önceki olasılığı, P(B) ise kanıtın önceki olasılığı olarak adlandırılır.

P(A|B) arka denir. P(B|A) olasılıktır.

Arka= (Olasılık)\*(Önerme önceliği olasılığı) / Kanıt önceliği olasılığı

Oyunda ne kadar çok değişken varsa ve bu değişkenler hakkında ne kadar emin olursak, koşullu olasılıklar kullanılarak o kadar kesin sonuç çıkarılabilir!

**What is poisson distributions? Specify discrete probability distributions categories.**

The Poisson distribution gives the probability of a number of events occurring in a fixed interval of time, if these events happen:

* with a known average rate and
* Independently of the time since the last event.

A Poisson distribution considers the number of successes per unit of time\* over the course of many units

\* Or any other continuous unit, e.g. distance

The average number of homes sold by the Acme Realty Company is 2 homes per day. What is the probability that exactly 3 homes will be sold tomorrow?   μ=2, x=3, e=2.71...

**P(x; μ) = (e-μ) (μx) / x!**

A classic example here is the number of phone calls received by a call centre.

 — Or —

If we know the average number of things that happen in a given time period, another use-case can be to predict the odds of getting another value instead, on a given future time. E.g. my Medium posts get an average of 1,000 views per day; I can use the Poisson probability mass function to estimate the probability of having 1,500 visits.

**What is probability?**

* Probability is a value between 0 and 1 that a certain event will occur
* the act of flipping a coin is called a trial.
* Each trial of flipping a coin can be called an experiment.
* Each mutually exclusive outcome is called a simple event.
* The sample space is the sum of every possible simple event.

**Was sind Poisson-Verteilungen? Spezifizieren Sie diskrete Wahrscheinlichkeitsverteilungskategorien.**

Die Poisson-Verteilung gibt die Wahrscheinlichkeit an, dass eine Reihe von Ereignissen in einem festen Zeitintervall auftreten, wenn diese Ereignisse eintreten:

• mit einem bekannten Durchschnittskurs und

• Unabhängig von der Zeit seit dem letzten Ereignis.

Eine Poisson-Verteilung berücksichtigt die Anzahl der Erfolge pro Zeiteinheit\* über viele Einheiten hinweg

\* Oder jede andere kontinuierliche Einheit, z. Distanz

Die durchschnittliche Anzahl der von der Acme Realty Company verkauften Häuser beträgt 2 Häuser pro Tag. Wie groß ist die Wahrscheinlichkeit, dass morgen genau 3 Häuser verkauft werden? μ=2, x=3, e=2,71...

P(x; μ) = (e-μ) (μx) / x!

Ein klassisches Beispiel hierfür ist die Anzahl der Anrufe, die ein Call Center erhält.

- Oder -

Wenn wir die durchschnittliche Anzahl von Dingen kennen, die in einem bestimmten Zeitraum passieren, kann ein weiterer Anwendungsfall darin bestehen, die Chancen vorherzusagen, stattdessen zu einem bestimmten zukünftigen Zeitpunkt einen anderen Wert zu erhalten. Z.B. meine Medium-Beiträge werden durchschnittlich 1.000 Mal pro Tag angesehen; Ich kann die Poisson-Wahrscheinlichkeits-Massenfunktion verwenden, um die Wahrscheinlichkeit von 1.500 Besuchen zu schätzen.

**Was ist Wahrscheinlichkeit?**

* Wahrscheinlichkeit ist ein Wert zwischen 0 und 1, dass ein bestimmtes Ereignis eintritt
* Der Vorgang des Werfens einer Münze wird als Versuch bezeichnet.
* Jeder Versuch, eine Münze zu werfen, kann als Experiment bezeichnet werden.
* Jedes sich gegenseitig ausschließende Ergebnis wird als einfaches Ereignis bezeichnet.
* Der Abtastraum ist die Summe aller möglichen einfachen Ereignisse.

**Poisson dağılımları nedir? Ayrık olasılık dağılımları kategorilerini belirtin.**

Poisson dağılımı, eğer bu olaylar gerçekleşirse, sabit bir zaman aralığında bir dizi olayın meydana gelme olasılığını verir:

• bilinen bir ortalama oran ile ve

• Son olaydan itibaren geçen zamandan bağımsız olarak.

Bir Poisson dağılımı, birçok birim boyunca birim zamandaki\* başarı sayısını dikkate alır.

\* Veya başka bir sürekli birim, örn. mesafe

Acme Realty Company tarafından satılan ortalama ev sayısı günde 2 evdir. Yarın tam olarak 3 evin satılmış olma olasılığı nedir? μ=2, x=3, e=2.71...

**P(x; μ) = (e-μ) (μx) / x!**

Burada klasik bir örnek, bir çağrı merkezi tarafından alınan telefon aramalarının sayısıdır.

- Veya -

Belirli bir zaman diliminde gerçekleşen şeylerin ortalama sayısını biliyorsak, başka bir kullanım durumu, belirli bir gelecek zamanda bunun yerine başka bir değer elde etme olasılığını tahmin etmek olabilir. Örneğin. Medium gönderilerim günde ortalama 1.000 görüntüleniyor; 1.500 ziyaret olma olasılığını tahmin etmek için Poisson olasılık kütle fonksiyonunu kullanabilirim.

**Olasılık nedir?**

• Olasılık, belirli bir olayın gerçekleşeceğine dair 0 ile 1 arasında bir değerdir.

• yazı tura atma eylemine deneme denir.

• Yazı tura atmanın her denemesi bir deney olarak adlandırılabilir.

• Her birbirini dışlayan sonuca basit olay denir.

• Örnek uzay, olası her basit olayın toplamıdır.

**Binom dağılımı nedir?**

"Binom", bir denemenin iki ayrık, birbirini dışlayan sonucu olduğu anlamına gelir. İki olası sonucu olan bir deney düşünelim: ya başarı ya da başarısızlık. Deneyin birkaç kez tekrarlandığını ve tekrarların birbirinden bağımsız olduğunu varsayalım. Sonucun başarılı olduğu toplam deney sayısı, dağılımına binom dağılımı adı verilen rastgele bir değişkendir.

n denemede x başarıyı gözlemleme olasılığını verir.

Diyelim ki bir zar 5 kez atılıyor. Tam olarak 2 dörtlü gelme olasılığı nedir?

b(x; n, P) = { n! / [ x! (n - x)! ] } \* Px \* (1 - P)n - x

x=2, n=5, p=1/6

**Bernoulli dağılımı nedir?**

Bernoulli Denemesi, yalnızca iki olası sonucun olduğu rastgele bir deneydir: başarı veya başarısızlık

• bir dizi deneme n, şu sürece ikili bir dağılımı izleyecektir:

1. Başarı olasılığı p sabittir
2. Denemeler birbirinden bağımsızdır

n denemede x başarıyı gözlemleme olasılığını verir

• tek bir denemede başarı olasılığı p ile gösterilir

• p'nin tüm denemeler için sabit olduğunu varsayar

**İstatistikte z ve t testleri ne amaçla kullanılır?**

• “İki örneğin aynı popülasyondan gelme olasılığı nedir?

• Her iki test de iki grubun ortalamasının birbirinden istatistiksel olarak farklı olup olmadığını değerlendirir.

• z-testi, örneklem büyüklüğü büyük olduğunda ve popülasyon standart sapması biliniyorsa uygun olan normal dağılımı izler.

• T-testi, örneklem büyüklüğü küçük olduğunda ve popülasyon standart sapması bilinmediğinde uygun olan t-dağılımını takip eder.

**Was ist Binomialverteilung?**

„Binomial“ bedeutet, dass es zwei diskrete, sich gegenseitig ausschließende Ergebnisse einer Studie gibt. Betrachten wir ein Experiment mit zwei möglichen Ergebnissen: entweder Erfolg oder Misserfolg. Angenommen, das Experiment wird mehrmals wiederholt und die Wiederholungen sind voneinander unabhängig. Die Gesamtzahl der Experimente, bei denen sich das Ergebnis als erfolgreich herausstellt, ist eine Zufallsvariable, deren Verteilung Binomialverteilung genannt wird.

Gibt die Wahrscheinlichkeit an, x Erfolge in n Versuchen zu beobachten.

Angenommen, ein Würfel wird fünfmal geworfen. Wie groß ist die Wahrscheinlichkeit, genau 2 Vieren zu bekommen?

b(x; n, P) = { n! / [ x! (n - x)! ] } \* Px \* (1 - P)n - x

x=2, n=5, p=1/6

**Was ist die Bernoulli-Verteilung?**

Ein Bernoulli-Versuch ist ein Zufallsexperiment, bei dem es nur zwei mögliche Ergebnisse gibt: Erfolg oder Misserfolg

• eine Reihe von Versuchen n folgt einer binären Verteilung, solange

1. Die Erfolgswahrscheinlichkeit p ist konstant
2. Prüfungen sind voneinander unabhängig

Gibt die Wahrscheinlichkeit an, x Erfolge in n Versuchen zu beobachten

* die Erfolgswahrscheinlichkeit bei einem einzelnen Versuch wird mit p bezeichnet
* nimmt an, dass p für alle Versuche fest ist

**Zu welchem ​​Zweck werden z- und t-Tests in der Statistik verwendet?**

* Wir verwenden Z- und t-Verteilungen zur Beantwortung der Frage „Wie groß ist die Wahrscheinlichkeit, dass zwei Stichproben aus derselben Grundgesamtheit stammen?
* Beide Tests beurteilen, ob sich die Mittelwerte zweier Gruppen statistisch voneinander unterscheiden oder nicht.
* Der z-Test folgt der Normalverteilung, was angemessen ist, wenn der Stichprobenumfang groß und die Standardabweichung der Grundgesamtheit bekannt ist.
* Der T-Test folgt der t-Verteilung, was geeignet ist, wenn die Stichprobengröße klein und die Standardabweichung der Grundgesamtheit nicht bekannt ist.

**What is binomial distribution?**

“Binomial” means there are two discrete, mutually exclusive outcomes of a trial. Let’s consider an experiment having two possible outcomes: either success or failure. Suppose the experiment is repeated several times and the repetitions are independent of each other. The total number of experiments where the outcome turns out to be a success is a random variable whose distribution is called binomial distribution.

Gives the probability of observing x successes in n trials.

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

b(x; n, P) = { n! / [ x! (n - x)! ] } \* Px \* (1 - P)n - x

x=2, n=5, p=1/6

**What is bernoulli distribution?**

A Bernoulli Trial is a random experiment in which there are only two possible outcomes: **success** or **failure**

* a series of trials n will follow a binary distribution so long as

1. The probability of success p is constant
2. Trials are independent of one another

Gives the probability of observing x successes in n trials

* the probability of success on a single trial is denoted by p
* assumes that p is fixed for all trials

**For what purpose are z and t tests used in statistics?**

* We use Z- and t-Distributions to answer the question “ What is the probability that two samples come from the same population?
* Either test assess whether mean of two groups are statistically different from each other or not.
* z-test follows normal-distribution, which is appropriate when the sample size is large, and the population standard deviation is known.
* T-test follows t-distribution, which is appropriate when the sample size is small, and the population standard deviation is not known.

**What do you understand by the term normal distribution?**

It is a set of continuous variables spread across a normal curve or in the shape of a bell curve. It can be considered as a continuous probability distribution and is useful in statistics. It is the most common distribution curve and it becomes very useful to analyze the variables and their relationships when we have the normal distribution curve. The normal distribution curve is symmetrical. The non-normal distribution approaches the normal distribution as the size of the samples increases. It is also very easy to deploy the Central Limit Theorem. This method helps to make sense of data that is random by creating an order and interpreting the results using a bell-shaped graph.

Data is usually distributed in different ways with a bias to the left or to the right or it can all be jumbled up. However, there are chances that data is distributed around a central value without any bias to the left or right and reaches normal distribution in the form of a bell shaped curve. The random variables are distributed in the form of an symmetrical bell shaped curve.

**Properties of Normal Distribution:**

Unimodal -one mode Symmetrical -left and right halves are mirror images Bell-shaped -maximum height (mode) at the mean Mean, Mode, and Median are all located in the center Asymptotic.

**What do you mean by confidence interval in statistical analysis?**

It is an interval estimate for a parameter value. It is constructed in a way so that, in the long run, a given proportion of these intervals will include the unknown true parameter value. The proportion is given by the "level of confidence". For instance, you can expect that at least 90% of (a large series of) 90% confidence intervals will include the unknown true values of the parameters.

**Was versteht man unter dem Begriff Normalverteilung?**

Es ist ein Satz kontinuierlicher Variablen, die über eine normale Kurve oder in Form einer Glockenkurve verteilt sind. Sie kann als kontinuierliche Wahrscheinlichkeitsverteilung betrachtet werden und ist in der Statistik nützlich. Es ist die häufigste Verteilungskurve und es wird sehr nützlich, die Variablen und ihre Beziehungen zu analysieren, wenn wir die Normalverteilungskurve haben. Die Normalverteilungskurve ist symmetrisch. Die Nicht-Normalverteilung nähert sich der Normalverteilung an, wenn die Größe der Stichproben zunimmt. Es ist auch sehr einfach, den zentralen Grenzwertsatz anzuwenden. Diese Methode hilft dabei, zufällige Daten zu verstehen, indem eine Reihenfolge erstellt und die Ergebnisse mithilfe eines glockenförmigen Diagramms interpretiert werden.

Daten werden normalerweise auf unterschiedliche Weise verteilt, mit einer Neigung nach links oder rechts, oder sie können alle durcheinander gebracht werden. Es besteht jedoch die Möglichkeit, dass Daten ohne Verzerrung nach links oder rechts um einen zentralen Wert herum verteilt sind und eine Normalverteilung in Form einer glockenförmigen Kurve erreichen. Die Zufallsvariablen werden in Form einer symmetrischen glockenförmigen Kurve verteilt.

**Eigenschaften der Normalverteilung:**

Unimodal -ein Modus Symmetrisch -linke und rechte Hälfte sind Spiegelbilder glockenförmig -maximale Höhe (Modus) am Mittelwert Mittelwert, Modus und Median befinden sich alle in der Mitte Asymptotisch.

**Was meinst du mit Konfidenzintervall in der statistischen Analyse?**

Es ist eine Intervallschätzung für einen Parameterwert. Es ist so konstruiert, dass langfristig ein bestimmter Anteil dieser Intervalle den unbekannten wahren Parameterwert enthält. Der Anteil wird durch das „Level of Confidence“ angegeben. Sie können beispielsweise davon ausgehen, dass mindestens 90 % (einer großen Reihe von) 90 %-Konfidenzintervallen die unbekannten wahren Werte der Parameter enthalten.

**Normal dağılım teriminden ne anlıyorsunuz?**

Normal bir eğri boyunca veya bir çan eğrisi şeklinde yayılan bir dizi sürekli değişkendir. Sürekli bir olasılık dağılımı olarak düşünülebilir ve istatistikte faydalıdır. En yaygın dağılım eğrisidir ve normal dağılım eğrisine sahip olduğumuzda değişkenleri ve aralarındaki ilişkileri analiz etmek çok faydalı olur. Normal dağılım eğrisi simetriktir. Normal olmayan dağılım, örneklerin boyutu arttıkça normal dağılıma yaklaşır. Merkezi Limit Teoremi'ni uygulamak da çok kolaydır. Bu yöntem, bir sıra oluşturarak ve sonuçları çan şeklindeki bir grafik kullanarak yorumlayarak rastgele olan verileri anlamlandırmaya yardımcı olur.

Veriler genellikle sola veya sağa doğru farklı şekillerde dağıtılır veya hepsi karıştırılabilir. Bununla birlikte, verilerin sola veya sağa herhangi bir sapma olmaksızın merkezi bir değer etrafında dağılma ve çan şeklindeki bir eğri şeklinde normal dağılıma ulaşma şansı vardır. Rastgele değişkenler simetrik bir çan şeklinde eğri şeklinde dağıtılır.

**Normal Dağılımın Özellikleri:**

Tek modlu -bir mod Simetrik -sol ve sağ yarımlar ayna görüntüleridir Çan şeklindeki -ortalamada maksimum yükseklik (mod) Ortalama, Mod ve Medyan tümü merkezde Asimptotik yer alır.

**İstatistiksel analizde güven aralığı ile ne demek istiyorsunuz?**

Bir parametre değeri için bir aralık tahminidir. Uzun vadede bu aralıkların belirli bir oranı bilinmeyen gerçek parametre değerini içerecek şekilde yapılandırılmıştır. Oran, "güven düzeyi" ile verilir. Örneğin, %90'lık güven aralıklarının en az %90'ının (geniş bir seri) parametrelerin bilinmeyen gerçek değerlerini içermesini bekleyebilirsiniz.

**What is a p-value? Why is it important?**

When you perform a hypothesis test in statistics, a p-value can help you determine the strength of your results. P-value is a number between 0 and 1. Based on the value it will denote the strength of the results. The claim which is on trial is called Null Hypothesis.

**What are measures of central tendency?**

Mean: The mean (represented by the greek letter mu— μ) is the average of a dataset. To calculate the mean, we sum up all the values and divide it by the number of values.

Median: The median is the middle of a dataset. To calculate the median, we sort all the values (in ascending or descending order) and take the one that is in the middle.

If there is an even number of data points, then we calculate the mean of the two that fall in the middle.

The median is less susceptible to outliers than the mean, and hence we need to take into consideration how the data distribution looks like, to choose which one to use. The median is much closer to most of the values in the series!

Mode: The mode is the most common/occurring value in the dataset. To calculate the mode, we locate the number that occurs more frequently.

Mode is usually only relevant to discrete numerical data — not to continuous data.

**What do you understand by Type I vs Type II error?**

Type I error is committed when the null hypothesis is true and we reject it, also known as a ‘False Positive’.

Type II error is committed when the null hypothesis is false and we fail to reject it, also known as ‘False Negative'.

**Was ist ein p-Wert? Warum ist es wichtig?**

Wenn Sie einen Hypothesentest in der Statistik durchführen, kann Ihnen ein p-Wert dabei helfen, die Stärke Ihrer Ergebnisse zu bestimmen. Der P-Wert ist eine Zahl zwischen 0 und 1. Basierend auf dem Wert gibt er die Stärke der Ergebnisse an. Die Behauptung, die vor Gericht steht, heißt Nullhypothese.

**Was sind Maße der zentralen Tendenz?**

Mittelwert: Der Mittelwert (dargestellt durch den griechischen Buchstaben mu— μ) ist der Durchschnitt eines Datensatzes. Um den Mittelwert zu berechnen, summieren wir alle Werte und dividieren ihn durch die Anzahl der Werte.

Median: Der Median ist die Mitte eines Datensatzes. Um den Median zu berechnen, sortieren wir alle Werte (in aufsteigender oder absteigender Reihenfolge) und nehmen den mittleren.

Wenn es eine gerade Anzahl von Datenpunkten gibt, berechnen wir den Mittelwert der beiden, die in die Mitte fallen.

Der Median ist weniger anfällig für Ausreißer als der Mittelwert, und daher müssen wir berücksichtigen, wie die Datenverteilung aussieht, um auszuwählen, welche verwendet werden soll. Der Median liegt viel näher an den meisten Werten in der Reihe!

Modus: Der Modus ist der häufigste/vorkommende Wert im Datensatz. Um den Modus zu berechnen, suchen wir die Zahl, die häufiger vorkommt.

Der Modus ist normalerweise nur für diskrete numerische Daten relevant  –  nicht für kontinuierliche Daten.

**Was versteht man unter Fehlern 1. und 2. Art?**

Fehler 1. Art wird begangen, wenn die Nullhypothese wahr ist und wir sie ablehnen, auch bekannt als „falsch positiv“.

Ein Typ-II-Fehler wird begangen, wenn die Nullhypothese falsch ist und wir sie nicht zurückweisen, auch bekannt als „Falsch-Negativ“.

**p değeri nedir? Neden önemli?**

İstatistikte bir hipotez testi yaptığınızda, bir p değeri, sonuçlarınızın gücünü belirlemenize yardımcı olabilir. P değeri 0 ile 1 arasında bir sayıdır. Değere bağlı olarak sonuçların gücünü gösterir. Yargılanmakta olan iddiaya Boş Hipotezi denir.

**Merkezi eğilim ölçüleri nelerdir?**

Ortalama: Ortalama (Yunanca mu— μ harfiyle gösterilir) bir veri kümesinin ortalamasıdır. Ortalamayı hesaplamak için tüm değerleri toplar ve değer sayısına böleriz.

Medyan: Medyan, bir veri kümesinin ortasıdır. Medyanı hesaplamak için tüm değerleri (artan veya azalan sırada) sıralar ve ortadakini alırız.

Çift sayıda veri noktası varsa, ortada kalan ikisinin ortalamasını hesaplarız.

Medyan, aykırı değerlere ortalamadan daha az duyarlıdır ve bu nedenle, hangisini kullanacağımızı seçmek için veri dağılımının nasıl göründüğünü dikkate almamız gerekir. Medyan, serideki değerlerin çoğuna çok daha yakındır!

Mod: Mod, veri kümesinde en yaygın/ortaya çıkan değerdir. Modu hesaplamak için daha sık meydana gelen sayıyı buluruz.

Mod genellikle yalnızca ayrık sayısal verilerle ilgilidir—— sürekli verilerle ilgili değildir.

**Tip I ve Tip II hatasından ne anlıyorsunuz?**

Tip I hata, sıfır hipotezi doğru olduğunda ve biz onu "Yanlış Pozitif" olarak da bilinen reddettiğimizde yapılır.

Tip II hata, sıfır hipotezi yanlış olduğunda ve "Yanlış Negatif" olarak da bilinen onu reddetmediğimizde yapılır.

**What is hypothesis testing? What are hypothesis testing types? Please explain with examples.**

A hypothesis is some kind of assumption and hypothesis testing is an inferential statistical technique to determine whether there is enough evidence in a data sample to infer that a certain condition holds true for the entire population. It is used to test the validity of a claim that is made about a population.

* Take a random sample
* Analyze the properties of the sample
* Test whether or not the identified conclusions correctly represent the population
* A hypothesis is generated about a population parameter.

**Null Hypothesis:** The null hypothesis is assumed to be true unless there is a strong evidence to the contrary. This claim that’s on trial is called the [null hypothesis](https://www.dummies.com/education/math/statistics/how-to-determine-a-p-value-when-testing-a-null-hypothesis/).

**Alternative Hypothesis:** The alternative hypothesis is assumed to be true when the null hypothesis is proved false. The alternative hypothesis is the one you would believe if the null hypothesis is concluded to be untrue.

**Framing Hypothesis:**

* At the start of the experiment, the null hypothesis is assumed to be true.
* If the data fails to support the null hypothesis, only then can we look to an alternative hypothesis

\* If testing something assumed to be true, the null hypothesis can reflect the assumption:

**Claim**:  “Our product has an average shipping weight of 3.5kg.”

**Null hypothesis:** average weight = 3.5kg

**Alternate hypothesis:** average weight ≠ 3.5kg

\* If testing a claim we want to be true, but can’t assume, we test its opposite:

**Claim**:  “This prep course improves test scores.”

**Null hypothesis:** old scores ≥ new scores

**Alternate hypothesis:** old scores < new scores

The null hypothesis should contain an equality (=, ≤ ,≥)

The alternate hypothesis should not have an equality (≠,<,>)

**What is hypothesis testing? What are hypothesis testing types? Please explain with examples.**

A hypothesis is some kind of assumption and hypothesis testing is an inferential statistical technique to determine whether there is enough evidence in a data sample to infer that a certain condition holds true for the entire population. It is used to test the validity of a claim that is made about a population.

* Take a random sample
* Analyze the properties of the sample
* Test whether or not the identified conclusions correctly represent the population
* A hypothesis is generated about a population parameter.

Null Hypothesis: The null hypothesis is assumed to be true unless there is a strong evidence to the contrary. This claim that's on trial is called the null hypothesis.

Alternative Hypothesis: The alternative hypothesis is assumed to be true when the null hypothesis is proved false. The alternative hypothesis is the one you would believe if the null hypothesis is concluded to be untrue.

Framing Hypothesis:

* At the start of the experiment, the null hypothesis is assumed to be true.
* If the data fails to support the null hypothesis, only then can we look to an alternative hypothesis

\* If testing something assumed to be true, the null hypothesis can reflect the assumption:

Claim: “Our product has an average shipping weight of 3.5kg.”

Null hypothesis: average weight = 3.5kg

Alternate hypothesis: average weight ≠ 3.5kg

\* If testing a claim we want to be true, but can't assume, we test its opposite:

Claim: “This prep course improves test scores.”

Null hypothesis: old scores ≥ new scores

Alternate hypothesis: old scores < new scores

The null hypothesis should contain an equality (=, ≤ ,≥)

The alternate hypothesis should not have an equality (≠,<,>)

**Hipotez testi nedir? Hipotez test türleri nelerdir? Lütfen örneklerle açıklayınız.**

Bir hipotez, bir tür varsayımdır ve hipotez testi, bir veri örneğinde belirli bir koşulun tüm popülasyon için geçerli olduğunu çıkarsamak için yeterli kanıt olup olmadığını belirlemek için çıkarımsal bir istatistiksel tekniktir. Bir popülasyon hakkında ileri sürülen bir iddianın geçerliliğini test etmek için kullanılır.

* Rastgele bir örnek alın
* Numunenin özelliklerini analiz edin
* Belirlenen sonuçların popülasyonu doğru bir şekilde temsil edip etmediğini test edin
* Bir popülasyon parametresi hakkında bir hipotez oluşturulur.

Boş Hipotezi: Aksine güçlü bir kanıt olmadıkça boş hipotezin doğru olduğu varsayılır. Deneme aşamasında olan bu iddiaya boş hipotez denir.

Alternatif Hipotez: Sıfır hipotezinin yanlış olduğu kanıtlandığında alternatif hipotezin doğru olduğu varsayılır. Alternatif hipotez, boş hipotezin doğru olmadığı sonucuna varılırsa inanacağınız hipotezdir.

Çerçeveleme Hipotezi:

* Deneyin başlangıcında, sıfır hipotezinin doğru olduğu varsayılır.
* Veriler boş hipotezi desteklemiyorsa, ancak o zaman alternatif bir hipoteze bakabiliriz.

\* Doğru olduğu varsayılan bir şeyi test edersek, boş hipotez şu varsayımı yansıtabilir:

İddia: "Ürünümüzün ortalama kargo ağırlığı 3,5 kg'dır."

Sıfır hipotezi: ortalama ağırlık = 3.5kg

Alternatif hipotez: ortalama ağırlık ≠ 3.5kg

\* Doğru olmasını istediğimiz ancak varsayamadığımız bir iddiayı test ediyorsak, tersini test ederiz:

İddia: "Bu hazırlık kursu, test puanlarını artırır."

Boş hipotez: eski puanlar ≥ yeni puanlar

Alternatif hipotez: eski puanlar < yeni puanlar

Boş hipotez bir eşitlik içermelidir (=, ≤ ,≥)

Alternatif hipotezin bir eşitliği olmamalıdır (≠,<,>)

**What Is the Goal Of A/B Testing?**

It is a statistical hypothesis testing for randomized experiment with two variables A and B. The objective of A/B Testing is to detect any changes to the web page to maximize or increase the outcome of an interest. Example: Identify the click-through rate for a banner ad.

A/B testing is a fantastic method for figuring out the best online promotional and marketing strategies for business. It can be used to test everything from website copy to sales emails to search ads An example of this could be identifying the click-through rate for a banner ad.

**Why statistical knowladge are important in data science?**

As Josh Wills once said,

“Data Scientist is a person who is better at statistics than any programmer and better at programming than any statistician.”

**Was ist das Ziel von A/B-Tests?**

Es handelt sich um einen statistischen Hypothesentest für ein randomisiertes Experiment mit zwei Variablen A und B. Das Ziel des A/B-Tests besteht darin, Änderungen an der Webseite zu erkennen, um das Ergebnis eines Interesses zu maximieren oder zu steigern. Beispiel: Identifizieren Sie die Klickrate für eine Bannerwerbung.

A/B-Tests sind eine fantastische Methode, um die besten Online-Werbe- und Marketingstrategien für Unternehmen herauszufinden. Es kann verwendet werden, um alles zu testen, von Website-Kopien über Verkaufs-E-Mails bis hin zu Suchanzeigen. Ein Beispiel dafür könnte die Ermittlung der Klickrate für eine Bannerwerbung sein.

**Warum ist statistisches Wissen in der Datenwissenschaft wichtig?**

Wie Josh Wills einmal sagte:

„Data Scientist ist eine Person, die besser in Statistik ist als jeder Programmierer und besser im Programmieren als jeder Statistiker.“

**A/B Testinin Amacı Nedir?**

İki değişken A ve B ile rastgele deney için istatistiksel bir hipotez testidir. A/B Testinin amacı, bir ilginin sonucunu en üst düzeye çıkarmak veya artırmak için web sayfasındaki herhangi bir değişikliği tespit etmektir. Örnek: Bir banner reklam için tıklama oranını tanımlayın.

A/B testi, işletmeler için en iyi çevrimiçi promosyon ve pazarlama stratejilerini bulmak için harika bir yöntemdir. Web sitesi kopyasından satış e-postalarına ve arama reklamlarına kadar her şeyi test etmek için kullanılabilir. Buna bir örnek, bir banner reklam için tıklama oranını belirlemek olabilir.

**Veri biliminde istatistiksel bilgi neden önemlidir?**

Josh Wills'in bir zamanlar dediği gibi,

"Veri Bilimcisi, istatistikte herhangi bir programcıdan daha iyi ve programlamada herhangi bir istatistikçiden daha iyi olan bir kişidir."

**What Is the Goal Of A/B Testing?**

It is a statistical hypothesis testing for randomized experiment with two variables A and B. The objective of A/B Testing is to detect any changes to the web page to maximize or increase the outcome of an interest. Example: Identify the click-through rate for a banner ad.

A/B testing is a fantastic method for figuring out the best online promotional and marketing strategies for business. It can be used to test everything from website copy to sales emails to search ads An example of this could be identifying the click-through rate for a banner ad.

**Why statistical knowladge are important in data science?**

As Josh Wills once said,

“Data Scientist is a person who is better at statistics than any programmer and better at programming than any statistician.”

**Was ist das Ziel von A/B-Tests?**

Es handelt sich um einen statistischen Hypothesentest für ein randomisiertes Experiment mit zwei Variablen A und B. Das Ziel des A/B-Tests besteht darin, Änderungen an der Webseite zu erkennen, um das Ergebnis eines Interesses zu maximieren oder zu steigern. Beispiel: Identifizieren Sie die Klickrate für eine Bannerwerbung.

A/B-Tests sind eine fantastische Methode, um die besten Online-Werbe- und Marketingstrategien für Unternehmen herauszufinden. Es kann verwendet werden, um alles zu testen, von Website-Kopien über Verkaufs-E-Mails bis hin zu Suchanzeigen. Ein Beispiel dafür könnte die Ermittlung der Klickrate für eine Bannerwerbung sein.

**Warum ist statistisches Wissen in der Datenwissenschaft wichtig?**

Wie Josh Wills einmal sagte:

„Data Scientist ist eine Person, die besser in Statistik ist als jeder Programmierer und besser im Programmieren als jeder Statistiker.“

**A/B Testinin Amacı Nedir?**

İki değişken A ve B ile rastgele deney için istatistiksel bir hipotez testidir. A/B Testinin amacı, bir ilginin sonucunu en üst düzeye çıkarmak veya artırmak için web sayfasındaki herhangi bir değişikliği tespit etmektir. Örnek: Bir banner reklam için tıklama oranını tanımlayın.

A/B testi, işletmeler için en iyi çevrimiçi promosyon ve pazarlama stratejilerini bulmak için harika bir yöntemdir. Web sitesi kopyasından satış e-postalarına ve arama reklamlarına kadar her şeyi test etmek için kullanılabilir. Buna bir örnek, bir banner reklam için tıklama oranını belirlemek olabilir.

**Veri biliminde istatistiksel bilgi neden önemlidir?**

Josh Wills'in bir zamanlar dediği gibi,

"Veri Bilimcisi, istatistikte herhangi bir programcıdan daha iyi ve programlamada herhangi bir istatistikçiden daha iyi olan bir kişidir."