

# Uitleg over alle onderdelen van statistiek

## Scatterplot:

### **Wat is het?**

Een scatterplot toont de relatie tussen twee variabelen door punten op een grafiek te plaatsen, waarbij elk punt een waarde voor beide variabelen vertegenwoordigt.

### **Hoe werkt het?**

Het helpt bij het identificeren van patronen, trends en outliers in de gegevens, en kan laten zien of er een correlatie is tussen de twee variabelen.

## Histogram:

### **Wat is het?**

Een histogram toont de verdeling van een enkele variabele door de frequentie van de waarden in verschillende intervallen weer te geven.

### **Hoe werkt het?**

Het laat zien hoe de gegevens over verschillende waarden zijn verdeeld en kan helpen bij het identificeren van de centrale tendens, spreiding en asymmetrie van de gegevens.

## Boxplot:

### **Wat is het?**

Een boxplot toont de verdeling van gegevens op basis van een vijf-getallen samenvatting: minimum, eerste kwartiel, mediaan, derde kwartiel en maximum.

### **Hoe werkt het?**

Het helpt bij het identificeren van outliers, spreiding en vergelijken van verdelingen tussen verschillende datasets.

## Z-scores:

### **Wat is het?**

Z-scores normaliseren gegevens door ze te transformeren naar een schaal met een gemiddelde van 0 en een standaardafwijking van 1.

### **Hoe werkt het?**

Het wordt berekend door de waarde van een gegevenspunt af te trekken van het gemiddelde en te delen door de standaardafwijking. Dit helpt bij het vergelijken van gegevenspunten op verschillende schalen.

## T-Tests:

### **Wat is het?**

T-tests worden gebruikt om te testen of er een statistisch significant verschil is tussen de gemiddelden van twee groepen.

### **Hoe werkt het?**

Het berekent een t-statistiek en p-waarde op basis van de steekproefgegevens, waarbij een lage p-waarde aangeeft dat de gemiddelden statistisch significant verschillen.

## ANOVA:

### **Wat is het?**

ANOVA (Analysis of Variance) wordt gebruikt om statistisch significante verschillen tussen de gemiddelden van drie of meer groepen te testen.

### **Hoe werkt het?**

Het berekent een p-waarde op basis van de variantie tussen de groepen en binnen de groepen, waarbij een lage p-waarde wijst op significante verschillen tussen de groepen.

## Correlation Matrix:

### **Wat is het?**

Een correlatiematrix toont de correlatiecoëfficiënten tussen meerdere variabelen.

### **Hoe werkt het?**

Het wordt gebruikt om relaties tussen variabelen te identificeren, waarbij een hoge correlatie kan wijzen op een voorspellende relatie tussen variabelen.

## PCA (Principal Component Analysis):

### **Wat is het?**

PCA is een dimensionaleitsreductietechniek die wordt gebruikt om een groot aantal variabelen te transformeren naar een kleiner aantal ongecorreleerde variabelen (hoofdcomponenten).

### **Hoe werkt het?**

Het identificeert de richtingen (hoofdcomponenten) die de meeste variantie in de gegevens vastleggen en projecteert de gegevens op deze nieuwe ruimte.

## Logistic Regression:

### **Wat is het?**

Logistische regressie is een statistisch model dat wordt gebruikt voor binomiale classificatieproblemen.

### **Hoe werkt het?**

Het past de logistische functie toe op de gegevens om de kans te voorspellen dat een bepaalde gebeurtenis plaatsvindt, waarbij de uitvoerwaarden tussen 0 en 1 liggen.

## Bias en Betrouwbaarheid:

### **Wat is het?**

Bias verwijst naar systematische fouten die leiden tot onnauwkeurige resultaten, terwijl betrouwbaarheid verwijst naar de consistentie van een maat.

### **Hoe werkt het?**

Bias wordt berekend als het verschil tussen de gemiddelde voorspelde waarden en de gemiddelde werkelijke waarden, terwijl betrouwbaarheid de verhouding van de variantie van de voorspelde waarden tot de totale variantie meet.