Inhaltdis

[Einführung in die Experimentplanung und die Bedeutung der Statistik 2](#_Toc522370666)

[Einführung und Problembeschreibung 2](#_Toc522370667)

[Warum Versuche/Experimente 2](#_Toc522370668)

[Wissenserwerb nach shewhart 2](#_Toc522370669)

[Aspekte der industriellen Versuchsplanung 2](#_Toc522370670)

[Das Prozessmodell des Experiments 2](#_Toc522370671)

[Begriffe 2](#_Toc522370672)

[DoE: Design of Experiment 2](#_Toc522370673)

[Versuchsplanung 2](#_Toc522370674)

[Begriffe 2](#_Toc522370675)

[Vorgehensweise 2](#_Toc522370676)

[Fehlerfortpflanzung 2](#_Toc522370677)

[Fehlerrechnung 2](#_Toc522370678)

[Ableitungsregeln 3](#_Toc522370679)

[Statistische Grundbegriffe 3](#_Toc522370680)

[SkalenE 3](#_Toc522370681)

[Nominalskala 3](#_Toc522370682)

[Ordinalskala 3](#_Toc522370683)

[Metrische Skala(Kardinalskala) 3](#_Toc522370684)

[Informationsniveau 3](#_Toc522370685)

[Ablauf der Statistischen UNtersuchung 3](#_Toc522370686)

[Datamining & datafarming 3](#_Toc522370687)

[Datamining 3](#_Toc522370688)

[datafarming 3](#_Toc522370689)

[Häufigkeitsverteilungen 3](#_Toc522370690)

[Einfache häufikeit 3](#_Toc522370691)

[Kumulierte Häufigkeit(SUmmenhäufigkeit) 3](#_Toc522370692)

[Klassifizierte Häufigkeit 3](#_Toc522370693)

[Grundlagen der Beschreibenden Statistik 4](#_Toc522370694)

[Einführung und Problembeschreibung 4](#_Toc522370695)

[Mittelwerte 4](#_Toc522370696)

[Modus 4](#_Toc522370697)

[Median 4](#_Toc522370698)

[Quantil 4](#_Toc522370699)

[Arithmetisches Mittel 4](#_Toc522370700)

[Harmonisches Mittel 4](#_Toc522370701)

[Geometrisches Mittel 4](#_Toc522370702)

[Streunugsmasse 5](#_Toc522370703)

[Spannweite R 5](#_Toc522370704)

[Zentraler Quartilsabstand(ZQA) 5](#_Toc522370705)

[Mittlere Absolute Abweichung 5](#_Toc522370706)

[Varianz 5](#_Toc522370707)

[Standardabweichung 5](#_Toc522370708)

[Variationskoeffizient 5](#_Toc522370709)

[Zeitreihen, Regression, korrelation 5](#_Toc522370710)

[Motivation 5](#_Toc522370711)

[Regressionsanalyse 5](#_Toc522370712)

[Zeitreihe 6](#_Toc522370713)

[Wahrscheinlichkeit 6](#_Toc522370714)

[Kombinatorik und Wahrscheinlichkeit 8](#_Toc522370715)

[KOmbinatorik 8](#_Toc522370716)

[Bedingte Wahrscheinlichkeit 9](#_Toc522370717)

[Unabhängige Wahrscheinlichkeit 10](#_Toc522370718)

[VOLLSTÄNDIGE WAHRSCHEINLICHKEIT 10](#_Toc522370719)

[Zufallsvariable und die wichtigsten theoretischen Verteilungen 10](#_Toc522370720)

[Zufallsvariabeln 10](#_Toc522370721)

[Diskrete Wahrscheinlichkeitsfunktion 10](#_Toc522370722)

[Kontinuierliche Verteilungsdichtefunktion 10](#_Toc522370723)

[Wichtige diskrete Verteilungen 11](#_Toc522370724)

[Wichtigsten stetigen theoretischen Verteilungen 11](#_Toc522370725)

[Schliessende Statistik 14](#_Toc522370726)

[Zufallsstichproben 14](#_Toc522370727)

[Modellbildung 15](#_Toc522370728)

[Testverteilungen 15](#_Toc522370729)

[Normalverteilung 15](#_Toc522370730)

[t-Verteilung 16](#_Toc522370731)

[Schätzverfahren & **Abschätzung** der unbekanntem Parameter **μ** und **σ** 17](#_Toc522370732)

[Schätzverfahren 17](#_Toc522370733)

[Schätzfunktion 17](#_Toc522370734)

# Einführung in die Experimentplanung und die Bedeutung der Statistik

## Einführung und Problembeschreibung

### Warum Versuche/Experimente

* ***Funktionsumfang*** der Produkte ***erhöhen***(Kundenanforderungen gerecht werden)
* ***Kosten*** müssen ***gesenkt*** werden
* **Entwicklungszeit**/Fertigung, etc. **verkürzen**
* Dazu Zusammenhänge/Parameter untersuchen(analysieren + verbessern)
* Verbesserungen anwenden – überprüfen – bewerten
* Gezielte Versuche/Expertimente sind erforderlich

### Wissenserwerb nach shewhart

Ein Bild, das Text, Karte enthält.

Mit hoher Zuverlässigkeit generierte BeschreibungWissenserwerb durch:

* Wahrnehmung
* Beobachten
* Lernen
* Denken und Schliessen
* Erklärungen(Modelle) finden und beschreiben
* Hypothesen formulieren & validieren

🡪dafür braucht es Experimente(am realen System oder Modell)

### Aspekte der industriellen Versuchsplanung

Die Hindernisse beim Experimentieren/Erkenntnisgewinn:

* ***Komplexität*** – Vielgestaltigkeit(viele Faktoren)
* ***Kompliziertheit*** – Schwierigkeit
  + Schwierige/nicht verstandenen/unbekannte/unerklärbare Zusammenhänge oder Faktoren
* ***Rauschen, Dynamik*** – Fehlereinflüsse 🡪 Messungenauigkeit
  + Unterschiedliche Ergebnisse bei gleiche Faktoren

### Ein Bild, das Screenshot enthält. Mit hoher Zuverlässigkeit generierte BeschreibungDas Prozessmodell des Experiments

|  |
| --- |
| Experimente werden immer nach einem bestimmen Schema durchgeführt:  1. Ausgangssituation beschreiben  2. Untersuchungsziele festlegen / Zielgrössen definieren  3. Faktoren auswählen und gewichten  4. Versuchsplanung erstellen  5. Versuche durchführen  6. Ergebnisse auswerten und Vertrauensintervalle bestimmen  7. Ergebnisse interpretieren und Massnahmen ableiten  8. Überprüfen der «Verbesserungen» |

### Begriffe

|  |
| --- |
| ***Systematischer Fehler*** – Fehler, welcher «mit System» auftritt. Kann deshalb rausgerechnet werden   * Messapparatur falsch geeicht * verunreinigte Salzlösung |

|  |
| --- |
| ***Zufälliger Fehler*** –zufällig entstandener Fehler, kann nicht rausgerechnet(vorhergesehen) werden.   * Thermisches Rauschen eines Sensors |

## DoE: Design of Experiment

### Versuchsplanung

***So wenig Versuche wie möglich***.

Zwei verschiedene Ansätze:

* Immer nur eine der Grösse(Faktoren) verändern
* Alle möglichen Kombinationen der Faktoren verändern
  + Kein Probieren – strukturelles Vorgehen

### Begriffe

|  |
| --- |
| ***Zielgrössen*** – beschreiben das Ergebnis eines Versuchs(Messwerte oder errechnete Grössen[Achtung! Fehlerfortplanzung] |
| **Einflussgrössen** – Grössen die die Zielgrössen beinflussen können   * Steuergrössen – einstellbar evtl. schwankend * Störgrössen – nicht vorgegeben/vorhersehbar |
| ***Faktoren*** – für den Versuch wesentlichen Einflussgrössen |
| ***Faktorstufen*** – Werte die die Faktoren im Versuch annehmen sollen:   * Quantitative Faktoren – Stufen mit Zahlenwerten * Qualitative Faktoren – Namen, Beschreib./Bezeic. |

### Ein Bild, das Screenshot enthält. Mit sehr hoher Zuverlässigkeit generierte BeschreibungVorgehensweise

**Ausgangssituation spezifizieren / Problem beschreiben / Ziel definieren**

* Kunde und dessen Bedürfnisse definieren
* Liegen bereits Daten vor?
* Welche Probleme müssen gelöst werden
* Welche Ressourcen (Zeit und Geld) stehen zur Verfügung (Kosten/Nutzen Analyse)
* Betroffene Gruppen und deren Beziehung untereinander listen (Wiederstände, Supporter, Wissensträger)
* ***Ist das Problem verstanden?***
* ***Was ist das Problem genau?***

**Untersuchungsziel festlegen**

* Wertdefinition von Prozessergebnissen oder -parameter
* Versuch Streuung/Robustheit zu reduzieren
* Wichtigsten Störgrössen erkennen/definieren
* Prozesse optimieren durch fertigen + lernen
* Funktion/Zuverlässigkeit nachweisen

**-Zielgrössen/Faktoren festlegen**

* Zielgrössen beschreiben(Kunden- & Zielorientiert)
* Quantifizierung 🡪 Zielgrössen müssen messbar sein
* Vollständige Ergebnisse & Eigenschaften berücksichtigen
* Jede Zielgrösse möglichst grundlegenden Zusammenhang
* ***Erst möglich viele(alle) Einflussgrössen sammeln und dann auf die (wesentlichen) Faktoren reduzieren***

Unterstützung durch Diagramme, Tabellen, etc.

**Lösen des Problems**

* ***Analytisch*** – schnellstes, billigstes
* ***Experimentell***
  + Planung durchführen
  + Realen Objekt oder Simulation?
  + Versuch durchführen/dokumentieren/auswerten

***🡪Experiment auswerten 🡪 Massnahmen/Weiteres ableiten***

|  |  |
| --- | --- |
| Fehlerfortpflanzung | |
| Weiterziehen eines gemessenen Fehlers in einer Rechnung. Auch das Resultat wird den Messfehler enthalten. | |
| Fehlerrechnung | |
| * Der Begriff ist irreführend denn man kann den Fehler nicht berechnen sondern nur schätzen. * Benötigt um ***Bereich abzuschätzen***, in denen der tatsächliche Wert mit einer gewisser Wahrscheinlichkeit liegt   Daher wird immer ein Angabe über die Genauigkeit des Wertes angegeben! | |
| ***Absoluter Fehler ∆x*** | Mit dem absoluten Fehler wird das Intervall berechnet oder geschätzt, in welchem sich das Ergebnis befinden sollte. Der absolute Fehler wird immer mit einer Masseinheit angegeben.  Formel: ***relativer Fehler \* Messung(Wert)***  ***Bei Summen und Differenzen addieren sich die absoluten Fehler***.  Gut für Ergebnisdarstellungen(z.B 5.8cm ± 0.1cm) |
| ***Implizite Fehlerannahme*** | Oft gibt man den Fehlereingabe nicht für jeden Wert an. Es hat sich daher eine Regel eingebürgert.  Fehler bei Nennwerteingabe:   * ***Mindestens einer halben Einheit der letzten(rechten) Stelle*** * Ein Bild, das Screenshot enthält.    Mit sehr hoher Zuverlässigkeit generierte Beschreibung***Höchstens 4 Einheiten der letzten bedeutsamen Stelle*** |
| ***Relativer Fehler*** | Mit dem Relativen Fehler wird prozentuale Abweichung des Ergebnis berechnet, daher ist er auch ohne einheitenlos(in Prozent):  Relativer Fehler:  Mit Hilfe des relativen Fehler lässt sich gut Abschätzen, welcher Faktor verbessert werden sollte. (der mit dem grösseren Fehleranteil)  ***Bei Produkten und Quotienten addieren sich die relativen Fehler***.  Gut um Gefühl der Messgenauigkeit zu erhalten(Bei Messgeräten relativer Fehler bezogen auf eingestellten Messbereich)  Beispiel: Voltmeter hat bei 20-100V eine Abweichung(relativer Fehler) von 0.001=0.1% |
| ***Beispiel Absoluter Fehler***  ***Lineal*** | Auf beiden Seiten wird abgelesen und die Fehlertoleranzen addiert ergeben den absoluten Fehler.  **Ein Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungLänge wird abgelesen:** L = 28.15cm – 22.35cm = 5.8cm  **Abgelesene Unsicherheit der Werte ist** ± 0.05cm  //Differenz🡪Addieren des absoluten Fehlers:  **Addieren des absoluten Fehlers:** ΔL = 0.05 cm **+** 0.05 cm = 0.1 cm  **Resultat notieren**: 5.8cm ***± 0.1cm***  Der Fehler befindet sich also im Intervall [5.7 ; 5.9] |
| ***Beispiel Relativer Fehler*** | **Länge:** L = 5.8cm  **absoluter Fehler:** ΔL = 0.1cm  **Einfügen in die Formel**: = |
| ***Beispiel Relativer Fehler***  ***Fläche*** | C:\Users\sever\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\8117010C.tmpEs soll die Fläche bestimmt werden.  **Länge:** 5.8cm ± 0.1cm  **Breite**: 0.9 ± 0.1 cm  **Fläche:** 5.8 \* 0.9 = 5.22 cm2  **Relativer Fehler Breite:** ΔB = 0.1/0.9 = 11.1%  **Relativer Fehler Länge:** : ΔL = 0.1/5.8cm = 1.7%  //Multiplikation: Addieren des relativen Fehlers  **Relativer Fehler Fläche:** ΔB + ΔL = 12.8%  **Absoluter Fehler Fläche:** ΔA = relativer Fehler \* Messung(Wert) = 0.128 \* 5.22 = 0.668cm2  **Ergebnis:** 5.2cm2 ± 0.7cm2. |
| ***Methode mehr als zwei Messgrössen*** | Sobald Exponenten vorkommen 🡪 diese im Produkt nach vorne nehmen |
| ***Methode min/max*** | Wenn mann nur Fehlergrenzen berechnen muss(nur absoluter Fehler) |

## Ableitungsregeln

# Statistische Grundbegriffe

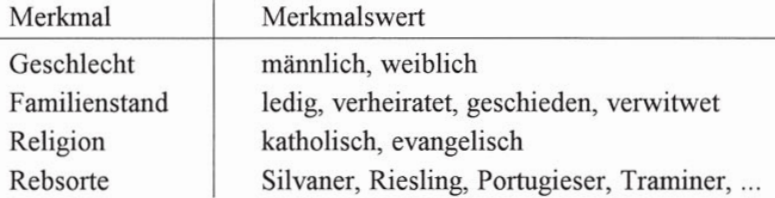
|  |
| --- |
| **Merkmalsträger**: Gegenstand der statistischen Untersuchung.  «Träger der Merkmale» z.B Werkzeugmaschine |
| **Grundgesamtheit**: Menge aller Merkmalsträger mit übereinstimmenden Abgrenzungsmerkmalen(zeitlich, räumlich,sachlich).  Z.B. Menge aller Werkzeugmaschinen Standort Zürich  Alle Maschinen Typ B-299  Alle Ausfälle in den letzten 30 Tagen |
| **Merkmal(Information)**: Eigenschaft des Merkmalträgers.  Z.B. Ausfallraten, Servicedauer, Leistung, etc. |
| **Merkmalswert**: Wert des eines Merkmals des Merkmalträgers.  Z.B. Maschine XY hat Leistung 4000W oder Ausfallrate 0.01% |

## SkalenE

In Skalen sind mögliche Merkmalswerte nach einer Ordnung abgelegt.

### Nominalskala

Auf der Nominalskala sind als Skalenwerte ***Namen*** abgetragen, die ***gleichberechtigt*** bzw. ***gleichbedeutend*** nebeneinander angeordnet sind.



### Ordinalskala

Ein Bild, das Screenshot enthält.

Mit sehr hoher Zuverlässigkeit generierte BeschreibungAuf der Ordinalskala sind als Skalenwerte ***Klassenbezeichnungen*** abgetragen. Die Skalenwerte stehen jetzt ***nicht mehr gleichberechtigt*** nebeneinander, sondern sind entsprechend ihrer Klasse in ***auf- oder absteigender Folge*** auf der Skala angeordnet.

### Metrische Skala(Kardinalskala)

Die Metrische Skala beinhaltet reelle Zahlen entsprechend auf- oder absteigender Folge.

#### Intervallskala

Ein Bild, das Screenshot enthält.

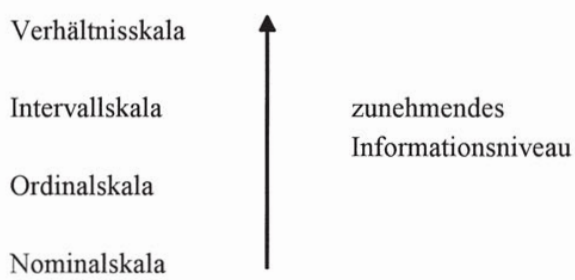
Mit hoher Zuverlässigkeit generierte BeschreibungSkalenwert 0 ist ein willkürlich gewählter Punkt. Somit kann man zwischen zwei Merkmalswerten den Abstand(Intervall) messen, jedoch kein Verhältnis bilden werden(8:00 nicht doppelt 4:00):

#### Verhältnisskala

Ein Bild, das Screenshot enthält.

Mit hoher Zuverlässigkeit generierte BeschreibungAuf der Verhältnisskala entspricht der Skalenwert Null dem natürliche, absoluten Null Punkt. Negative Werte somit nicht möglich, dafür kann man das Verhältnis zwischen zwei Werten messen.(20Jahre ist doppel so viel wie 10Jahre)

### Informationsniveau



## Ablauf der Statistischen UNtersuchung

1. ***Planung***
   1. **Festlegung der folgenden Punkte**: Merkmal, Merkmalträger, Erhebungstechnik, Aufbereitungsverfahren, Darstellungsform, statistische Analyseverfahren
2. ***Datenerhebung***
   1. **Konkretisierung Untersuchungsziel**: Abgrenzungsmerkmale, Untersuchungsmerkmale, Datenumfang, erwartetes Ergebnis (Ziel/Problem festlegen)
   2. **Erhebungstechniken**: Messungen, Zählung, Befragung, Beobachtung
   3. **Herkunft der Daten**:
   * Primärstatistik
     + Daten erstmalig erhoben
     + Exakt fürs Ziel gerichtet/erzeugt
   * Sekundärstatistik
     + Bereits vorliegendes Material
     + Kostengünstig
     + Nicht aktuell & nicht exakt
   1. **Erhebungsumfang**: Teil- oder Vollerhebung
      * Bei experiment immer nur Teilerhebung(Stichprobe)🡪schliessende Statis.
3. ***Datenaufbereitung*** ***und -darstellung***
   1. **Daten von Urliste überführen in**: Strichliste, Häufigkeitstabelle
4. ***Analyse und Interpretation***

## Datamining & datafarming

### Datamining

Vereinfacht gesagt ist Data-Miniing das Auswerten von grossen Daten und daraus Rückschlüsse wie Regelmäßigkeiten, Gesetzmäßigkeiten und verborgener Zusammenhänge zu finden.

### 

### datafarming

Vereinfacht gesagt will man mittels Data-Farming Daten erzeugen, welche dann wieder für Experimente gebraucht werden können.

|  |  |  |  |
| --- | --- | --- | --- |
| Häufigkeitsverteilungen | | | |
| Einfache häufikeit | | | |
| Gibt an, ***wie häufig ein Merkmalswert*** xi aufgetreten ist. Dieser kann ***absolut oder relativ*** ausgedrückt werden. | | | |
| ***absolute einfache Häufigkeit hi*** | **Anzahl** der Merkmalsträger mit dem Merkmals-wert xi (i = 0, .. v)  Wie viel Mal kommt der Wert xi gesamthaft vor? | | |
| ***relative einfache Häufigkeit fi*** | **Anteil** der Merkmalsträger mit dem Merkmals-wert i (i: 1., .. , v)  Wie ist das Verhältnis/Anteil dieses Merkmalswert gegenüber allen Werten? | | |
| ***Gesamtzahl aller Merkmalsträger***  ***n*** | Summe aller Merkmalsträger, welche vorkommen.  Wie viele Merkmalsträger gibt es insgesamt? | | |
| ***Anzahl verschiedener Merkmalswerte***  ***v*** | Summe aller verschiedener Merkmalsträger, welche vorkommen.  Wie viele verschiedene Merkmalsträger gibt es? | | |
| ***Formeln*** |  |  | Ein Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung | | |
| ***Beispiel*** | **Ausfälle Maschinen B200**  **Absolute** **Häufigkeit**(ablesbar)  h1 = 30, h2 = 20, h3 = 10  **Gesamtanzahl**  n = 30+20+10 = 60  **Relative** **Häufigkeit**  f1 = 30/60 = 0.5  f2 = 20/60 = 0.33  f3 = 10/60 = 0.16  **Anzahl verschiedene Merkmalsträger** = 3 | | |
| Kumulierte Häufigkeit(SUmmenhäufigkeit) | | | |
| ***Absolute kumulierte Häufigkeit***  ***Hi*** | **Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungAnzahl** der Merkmalsträger mit einem Merkmalswert der **kleiner oder gleich** xi ist | | |
| ***Relative kumulierte Häufigkeit*** | **Summe der Anteile** der Merkmalsträger mit dem Merkmalswert kleiner gleich i. | | |
| ***Formeln*** | Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung |  | | |
| Klassifizierte Häufigkeit | | | |
| Bei mehr als 10 bis 15 verschiedenen Merkmalswerten ist die Darstellung nicht mehr überschaubar, man wechselt darum auf die Klassifizierte Häufigkeit. Dabei werden ***Klassen von Merksmalsweren(Intervalle)*** gebildet.  Ein Bild, das drinnen enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung | | | |

|  |  |
| --- | --- |
| Grundlagen der Beschreibenden Statistik | |
| Einführung und Problembeschreibung Mit ***Mittelwerte*** können typische Eigenschaften der Häufigkeitsverteilung beschrieben werden. Sie sind zudem sind kompakter als Tabellierte Häufigkeiten. Mithilfe von Kenngrössen, den Parametern, können aussagekräftige Grössen definiert werden, welche die grosse Menge von Messdaten verdichten. Diese Menge nennt man ***Streuungsmasse***.  Anhand von empirischen Daten versucht man Kenngrössen Daten zu beschrieben/charakterisieren. Mittelwerte | |
| Modus | |
| ***Grundlagen*** | Der Modus ist der Wert, welcher ***am häufigsten beobachtet*** wird.   * Für jede Verteilung bestimmbar. * Von Ausreissern unbeeinflussbar * Schnelle und einfache Ermittlung * Als Mittelwert geeignet wenn seine Häufigkeit die anderen Häufigkeiten dominiert |
| ***Formeln bei gleicher Klassenbreite*** | |  |  | | --- | --- | | * : Unterer Klassengrenze | * : oberere Klassengrenze | | * :Modusklasse | * : Klasse links von Modus | | * : Klasse rechts von Modus |  | |
| ***Formel bei ungleicher Klassenbreite*** | |  |  | | --- | --- | | * : Unterer Klassengrenze | * : oberere Klassengrenze | | * : Dichte Modusklasse | * : Dichte Klasse links von Modus | | * : Dichteklasse rechts von Modus | | |
| ***Dichte Formel*** |  |
| Median | |
| ***Grundlagen*** | Derjenige Merkmalswert, dessen Merkmalsträger in der Rangordnung aller Merkmalsträger genau ***in der Mitte*** ist.   * Mindestens Ordinalskala(weil die Merkmalswerte eine Rangordnung aufzeigen müssen) * Unbeeinflusst von Ausreissern * Gut geeignet für schiefe Verteilungen * Vermittelt einen guten Eindruck für die «Mitte» |
| ***Formeln*** | |  |  | | --- | --- | | * : Kumulierte Klassenhäufigkeit | Gesamtzahl aller Merkmalsträger | |
| Quantil | |
| ***Grundlagen*** | Ein Quantil ist ein Merkmalswert, durch den die Gesamtheit in **zwei Teile** zerlegt. Zusätzlich gibt es noch   |  |  |  | | --- | --- | --- | | ***Quartil*** = 4 Teile | ***Dezile*** = zehn Teile | ***Perzentile*** = 100 | |
| ***Formeln*** | Berechnung erfolgt analog des Medians(anstatt n/2🡪n/4 oder n/10, etc.):  Beispiel: 3, Quantil(3n/4 anstatt n/2): |
| Arithmetisches Mittel | |
| ***Grundlagen*** | Das arithmetische Mittel ist die ***Summe*** aller Merkmalswerte ***geteilt*** durch die ***Anzahl*** der Merkmalswerte. Das ist der Durchschnitt, wie man in «im Volksmund» kennt.   * , da Addition notwendig von Merkmalswerten * In der Praxis sehr oft benutzt - Geeignet für symmetrische, nicht für schiefe Verteilungen * Ausreisser können zu starken Verzerrungen füren(weil «Durchschnitt verfälscht») |
| ***Formel*** | ***Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung***  ***Bei der Klassifizierten Häufigkeit ergibt sich xj' aus (Klassenmittelwert)*** |
| Harmonisches Mittel | |
| ***Grundlagen*** | Man bildet den Durchschnitt zu Merkmalswerten, welche in Summe gesehen relativ gleich weit entfernt sind.   * Merkmal muss verhältnisskaliert sein(Verhältnissskala , wegen Bildung Quotienten) * Merkmalswerte alle positiv oder alle negativ * Wird oft verwendet für ***Durschnittgeschwindigkeit***(hm/h oder z.B Flasche/h) |
| ***Formeln*** | ***Ein Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung*** |
| ***Beispiel*** | ***Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungEin Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung*** |
| Geometrisches Mittel | |
| ***Grundlagen*** | Ist die n-te Wurzel aus dem Produkten aller Merkmalswerte.   * Wird verwendet um durcschnittliche ***Entwicklungen*** oder Wachstum zu berechnen z.B Veränderungen von Geldbeträgen über Jahre oder Kappenlänge nach Anzahl Waschgängen * Merkmalswerte > 0(wegen Division) |
| ***Formel/Beispiel*** | Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungEin Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung |
| ***Beispiel gleicher Klassen*** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Klassenbreite*** | ***absolute Häufigkeit(hi)*** | ***relative Häufigkeit(fi)*** | ***Häufigkeitsdichte*** | ***Kumulierte absolute Häufigkeit*** | | 0 - 100 | 100 | 0.083 | 100/100 =1 | 100 | | 100 – 200 | 300 | 0.25 | 300/100 =3 | 400 | | 300 - 400 | 500 | 0.417 | 500/100 = 5 | 900 | | 500 - 600 | 300 | 0.25 | 300/1000 = 3 | 1200 |   **//600 in Klasse 3** |
| ***Beispiel ungleicher Klassenbreite*** | |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***Klassenbreite*** | ***absolute Häufigkeit*** | ***relative Häufigkeit*** | ***Häufigkeitsdichte*** | ***Kumulierte absolute Häufigkeitsdichte*** | | 0 - 100 | 100 | 0.01 | 100/100 =1 | 100 | | 100 - 500 | 2400 | 0.24 | 2400/400 =6 | 2500 | | 500 - 1000 | 4500 | 0.45 | 4500/500 = 9 | 7000 | | 1000 - 2000 | 3000 | 0.30 | 3000/1000 = 3 | 10000 | |
| ***Beispiel 3. Quantil*** | * + 1. ***Bestimmen der Klasse, in welchem das 3. Quantil liegt(n = 245)***  1. 183.75 > als 145 🡪 4. Klasse muss genommen werden    * 1. ***Berechnen:*** |

|  |  |
| --- | --- |
| Streunugsmasse | |
| Spannweite R | |
| ***Grundlage*** | Die Spannweite R ist die Differenz aus dem grössten und dem kleinsten beobachteten Merkmalswert.   * Braucht mindestens Intervallskala * Extrem empfindlich auf Ausreisser |
| ***Formel*** | ***Ein Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungEin Bild, das Objekt, Uhr enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung*** |
| Zentraler Quartilsabstand(ZQA) | |
| ***Grundlagen*** | Ein Bild, das Himmel, Karte, Text enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungDer zentrale Quartilsabstand ist die Entfernung zwischen den zwei Merkmalswerten, welche die zentral gelegenen 50% der Merkmalsträger einordnen.   * Mindestens Intervallskala * Ausreisser-Problem tritt nicht auf * Geeignet wenn Kernbereich(50%) intressiert |
| ***Formel*** |  |
| ***Beispiel*** | ***Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungEin Bild, das Himmel enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung*** |
| Mittlere Absolute Abweichung | |
| ***Grundlagen*** | Die mittlere absolute Abweichung ist die ***durchschnittliche Entfernung*** aller beobachteten Merkmalswerte ***vom arithmetischen Mittel(oder Median).*** Man rechnet die Differenz von jedem Punkt und dem Mittelwert und teilt diese nachher durch die Anzahl Werte.   * Mindestens Intervallskala * Ausreisser können auch hier das Bild verzerren |
| ***Formel*** | Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibungx-Mittel hier arith. Mittel oder Median. |
| ***Beispiel*** | Ein Bild, das Wand, Himmel, Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibungn = 3 + 10 + 4 = 17  δ = 1/17 \* (6.12 + 10.40 + 0.16) = 0.98 |
| Varianz | |
| ***Grundlagen*** | Varianz ist die ***Summe der quadierten Abweichungen***, ***dividiert durch die Anzahl der Merkmalsträger***. Der Unterschied zur mittleren absoluten Abweichung ist das Quadrat. Somit ***keine negativen Werte.*** |
| ***Formel*** | ***Ein Bild, das Objekt enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung*** |
| Standardabweichung | |
| ***Grundlagen*** | Die Standardabweichung ist die ***Quadratwurzel aus der Varianz***. Die Berchnungen werden meistens mit der Standardabweichungen gerechnet. |
| ***Formel*** |  |
| ***Beispiel*** | **Ein Bild, das Screenshot enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung** |
| Variationskoeffizient | |
| ***Grundlagen*** | Der Variationskoeffizient misst **die relative Streuung**. Er setzt die Streung in Relation zu Lage der Häufigkeitsverteilung.  Er ist der Quotient aus Standardabweichung und arithmetischem Mittel multipliziert mit 100. |
| ***Formel*** |  |
| ***Beispiel*** | ***Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung*** |
| ***Überblick - Boxplot*** | ***Ein Bild, das Text enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungBoxplot = Darstellung von Mittelwerten und Streuungsmassen um einen schnellen Einblick zu bekommen, schnelle Übersicht wo die Daten liegen und wie diese verteilt sind.*** |
| Zeitreihen, Regression, korrelation | |
| Motivation | |
| 1. Bei Experimenten entstehen Tupel, ein Faktorwert(control) und ein Ergebniswert(response). 2. Ein Bild, das Objekt enthält.     Mit sehr hoher Zuverlässigkeit generierte BeschreibungMit diesen Tupel(Punkten) kann man in ein Streungsdiagramm (Punktwolke) erstellen 3. Nun wollen wir eine Funktion finden, welche diese Punkte bestmöglich annähert. 4. Achtung! Eine Extrpolation kann gefährlich sein, wenn man zu wenig Werte hat, kann es falsche Funktion liefern. | |
| Regressionsanalyse | |
| ***Grundlagen*** | * Man will einen Zusammehnag zwischen einer unabhängigen Grösse(x-Werte) und abhängigen Werten(y-Werte) anhand von Einzelmessungen modellieren. * Die Form bzw. Tendenz des Zusammenhangs durch eine mathematische Funktion(Regressionsfunktion) beschreiben |
| ***Lineare Regression*** | In der linearen Regressionsanalyse wird die Tendent ***durch die Funktion(Regressionsgerade)*** bestimmt***:***   * ***Geradengleichung: y =a + b·x*** //y wird duch Parameter bestimmt(Standard) * ***Proportionalität: y = c·x*** //bei unbekannter Abhängigkeit zusätzlich |
| ***Ansatz der kleinsten Quadrate*** | 1. partielles Ableiten des Ausdrucks nach a und nach b 2. nullsetzen der beiden partiellen Ableitungen 3. auflösen der beiden Gleichungen nach a und b   Ergibt für: :  **und**  Analog folgt für**:**  **und** |
| ***Beispiel*** | Student muss 6h arbeiten. Wie viel Zeit hat er für das Studium?  und  Werte einsetzen:  🡪  🡪 **38.43**  ***y = 38.43 - 0.49 ∙ x mit y(6) = 35.5 h*** 🡪 ***der Student kann 35.5h für das Studium aufwenden, wenn er 6 Stunden(y=6) arbeitet.*** |
| ***Linearisierung*** | ***Allgemeine Exponentialfunktion: y = a ex***  ***Allgemeine Logarithmusfunktion: y = a + b·lnx***  ***Wenn der Verlauf Exponentiell*** und nicht linear ist müssen die ***y-Werte logarithmiert*** werden.  Dann wie oben weiterfahren und am Schluss Logarithmieren und Kehrwert bilden: |
| ***Newton-Algorithmus*** | ***Polynome: y = an xn + an-1 xn-1 + … + a0***  ***Ansatz: f(x) = a0 + a1(x−x1) + a2(x−x1)(x−x2) + ... +an(x−x1)(x−x2)···(x− xn)***  Koeffizienten können iterativ bestimmt werden: |
| ***Schema Newton-Algorithmus*** |  |
| ***Beispiel*** | ***Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungWarum nicht bechtet?*** |
| Zeitreihe | |
| ***Definition*** | Zusammenhang zwischen Merkmalsträger x und Zeitpunkten ti.  Zeitreihe ist ***zeitlich geordnete Folge von Merkmalswerten***  • Kosten/Gewinnentwicklung  • Verbrauch von Ressourcen  • Auftragseingang  • Bearbeitungs- oder Durchlaufzeiten |
| ***Aufgaben/Ziel*** | ***erkennen der Struktur*** und der Gesetzmäßigkeiten einer Zeitreihe  Mittel zu Prognose - ***erkennen eines Trends*** |
| ***Trend*** | Blick in die Zukunft 🡪 beschreibt **langfristige Grundrichtung** einer Zeitreihe  ***Um ihn streuen die Zeitreihenwerte*** im Zeitablauf |
| ***Probleme*** | * periodische Schwankungen (Schweinezyklus) * Besondere Ereignisse 🡪 Ausreisser(unvorhersehbar)   🡪 ***Lösung*** durch sogenanntes ***Glätten*** des Verlaufs mittels linearer Regression |
| ***Zusammenhang*** | Das ***Erkennen des Zusammenhangs*** zwischen zwei oder mehr Merkmalen nennt man ***Korrelation***:  - Kraftstoffart – Leistung - Autounfälle – Alter -Lernaufwand – Lernerfolg  Dabei Untersucht man die Merkmale x & . Dabei intressiert man sich für folgende Fragen:   * 1. besteht ein Zusammenhang zwischen X und Y?   2. von welcher Form ist der Zusammenhang?   3. von welcher Stärke (Intensität) ist der Zusammenhang? |
| ***Abhängigkeit*** | * Zwei Merkmale sind statistisch ***unabhängig***, ***wenn* der Wert des einen Merkmals nicht davon abhängt, welchen Wert das andere Merkmal besitzt**, sonst sind sie abhängig * **Formale Abhängigkeit - *zahlenmäßig begründete Abhängigkeit*** * **Sachliche Abhängigkeit –** ist der Wert des einen Merkmales ***ursächlich***(***kausal***) für den Wert des anderen?(Ursache – Wirkung kann nicht bestimmt werden) |
| ***Korrelationsanalyse*** | ***Stärke (Intensität) des Zusammenhangs*** feststellen 🡪 wie stark ist der Einfluss von x auf y |
| ***Korrelationskoeffizient von  Bravais Pearson*** | 1): ***Kovarianz*** ***🡪 Streuung der Merkmalsträger***(kombinationen) um den Mittelpunkt/Durschnitt  Berechnung analog Varianz:    2.) Normierung der Kovarianz es entsteht der **Korrelationskoeffizient r**  zu:      Daraus folgt:   * ***Positivem r*** 🡪der lineare Zusammenhang der Merkmale X und Y ist ***gleichläufig*** * ***Negativem*** ***r*** 🡪 der lineare Zusammenhang der Merkmale X und Y ist ***gegenläufig*** * 0: kein **linearer** Zusammenhang * ±1: sehr starker **linearer** Zusammenhang |
|  |  |
| Wahrscheinlichkeit | |
| ***Zufallsexperimente*** | Ein Zufallsexperiment ist ein Vorgang, der als ***beliebig oft wiederholbar*** angesehen werden kann und dessen Endzustand oder ***Ergebnis vom Zufall abhängt***. |
| ***Wahrscheinlichkeitsräume*** | * ***Ergebnismenge*** Ω ***🡪*** alle mögliche Ergebnisse des Experiments * ***System A der Ereignisse 🡪*** Elemente Teilmenge von Ω * ***Wahrscheinlichkeitsmass P 🡪*** jedem Ereignis A wird Wahrscheinlichkeit P(a) zugeordnet |
| ***System der Ereignisse*** | weisst gewisse ***Eigenschaften*** auf, die wir aus der ***Mengenlehre*** kennen:   * dass beide Ereignisse zugleich **A ∩ B** * zumindest eines von beiden Ereignissen  **A ∪ B** * das Gegenereignis (komplementäre) zu einem Ereignis  **AC**   Es seien A, B ⊂ Ω zwei Ereignisse die ein System A der Ereignisse bilden.     |  |  |  | | --- | --- | --- | | ***A oder B*** | ***A und B*** | ***A komplementär*** | |  |  |  |   Speziell gilt:  Die ***leere*** ***Menge*** ∅ bildet das ***unmögliche Ereignis***.  Die ***Ergebnismenge*** Ω bildet das ***sichere*** ***Ereignis***  ***Ein-Elementige Teilmengen*** der Ergebnismenge Ω werden als ***Elementarereignisse*** bezeichnet.  Zwei Ereignisse A und B heißen ***unvereinbar***, wenn A ∩ B = ∅  Die leere Menge ist ein Ereignis.  Endliche Vereinigungen, endliche Durchschnitte und Differenzen sind Ereignisse.  ***Auf der Menge der Ergebnisse eines Experiments werden Teilmengen definiert, jede Teilmenge stellt ein Ereignis dar und ihr kann eine Wahrscheinlichkeit p zugeordnet werden.*** |
| ***Wahrscheinlichkeitsmass*** | Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungEin Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung |
| ***Laplace-Experiment*** | Ist die ***Ergebnismenge*** Ω = {ω1, ω2,…,ωn} ***endlich*** und wird jedem Elementarereignis aus Ω ***die*** ***gleiche Wahrscheinlichkeit,*** zugeordnet, so gilt für die Wahrscheinlichkeit eines beliebigen Ereignisses: |
| ***Beispiele*** | Werfen eines Würfels mit Ω = {1, 2,…,6} und P( {ωi} ) = 1 / 6.  Werfen einer Münze mit Ω = { {Wappen}, {Zahl} } und P( {Wappen} ) = P( {Zahl} ) = 1 / 2. |
| ***Omega Algebra*** | Eine Omega Algebra lieht vor, wenn zum System der Ereignisse noch die leere Menge und der Ergebnisraum zugeordnet werden und ferner die Regeln von Kolmogorov gelten. |
| Kombinatorik und Wahrscheinlichkeit | |
| KOmbinatorik – Regeln für das zählprinzip | |
| ***Grundsatz*** | * ***Mengen 🡪 Reihenfolge unwichtig*** * ***Tupel 🡪 Reihenfolge wichtig*** |
| ***Zählprinzip*** | Gibt es in einem n-Tupel für die Besetzung   * + der ersten Stelle **k**1 Möglichkeiten   + der zweiten Stelle **k2** Möglichkeiten   + der dritten Stelle **k3** Möglichkeiten   + der n-ten Stelle **kn** Möglichkeiten   ***dann gibt es insgesamt k1 • k2 • k3 • … • kn verschiedene n-Tupel*** |
| ***Beispiel*** | Beispiele sind Pferde oder Autorennen, da es bei jeder Stelle ein Auto weniger hat:  Die Reihenfolge der ersten 5 Plätze eines Pferderennen mit 18 gestarteten Pferden:  18 • 17 • 16 • 15 • 14 = 1.028.160 🡪entspricht Geordnete Probe ohne zurücklegen |
| ***Urnenmodell der Kombinatorik*** | In der Kombinatorik wird viel mit einem Urnenmodell gearbeitet. Es werden daraus beliebig viele Kugeln gezogen. Dabei gibt es folgende Unterscheidungen:   * Die Kugeln werden zurückgelegt(mit Wiederholungen) oder nicht(ohne Wiederholungen) * Die Reihenfolge wird beachtet(geordnet) oder nicht(ungeordnet)   Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungBei der ungeordneten Kugel Ziegung entspricht (1;2) 🡪 (2;1), etc.  Ziehung von k Kugeln aus einer Urne mit n = 3 Kugeln |

|  |  |
| --- | --- |
| ***Geordnete Proben***  ***Für Tupel***  ***Reihenfolge wichtig*** | Bei geordneten Proben wird auf die Reihenfolge geachtet.  Es gibt immer eine Menge m mit ***n Elementen***.  Bilden wir nun ***k-Tupel, deren Einträge aus der Menge M*** stammen, so zwei Möglichkeiten:   1. ***Mit Wiederholung***   Ein Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungEs kann jedes Mal wieder die gleiche Zahl(oder gleiches Element) vorkommen   1. ***Ohne Wiederholen(ohne zurücklegen)***   Ein Bild, das Objekt enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungBei jedem entnehmen gibt es eine Kugel weniger(n!-Möglichkeiten), da die Reihenfolge eine Rolle spielt noch geteilt durch(n-k)!  Mehr Ergebnisse da Reihenfolge wichtig:  Ein Bild, das Shoji, Sport, Sportwettkampf enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung |
| ***Permutationen*** | Die Zahl der Permutationen einer n-Menge ist ***n!***  Das ist ein Spezialfall von einer geordneten Probe ohne Wiederholung(n = k)  Beispiel: Festlegung der Sitzordnung von 10 Gästen ist: 10! = 3628800 |
| ***Permutation vs Kombinatorik*** | Permutation Mass für alle Kombinationsmöglichkeiten. Kombinatorik Regel für Berechnung der Möglichkeiten, wenn Bedingungen die Möglichkeiten einschränken |
| ***Ungeordnete Proben***  ***für Mengen***  ***Reihenfolge unwichtig*** | Bei der ungeordneten Proben wird nicht auf die Reihenfolge geachtet.  Es gibt ebenfalls eine Menge m mit n Elementen.  Es gibt wieder die 2 Fälle:   1. Ein Bild, das Objekt enthält.     Mit hoher Zuverlässigkeit generierte Beschreibung ***Mit Wiederholung***   Sind aus ***n verschiedenen Elementen*** ***k Elemente mit Wiederholung*** auszuwählen und ist die ***Anordnung ohne Bedeutung***.   1. Ein Bild, das Himmel enthält.     Mit sehr hoher Zuverlässigkeit generierte Beschreibung***Ohne Wiederholung(ohne zurücklegen)***   ***Die Anzahl der k-elementigen Teilmengen aus einer n-elementigen Menge.***  ***Gleich wie bei der geordneten Probe, jedoch da die Ordnung keine Rolle spielt muss man noch durch k! teilen.***  ***Anzahl aller möglichen Teilmengen k aus n.***  k: Anzahl der zu entnehmenden Elemente  ***Anzahl aller möglicher Element-Kombinationen mit Wiederholung***  Da die Reigenfolge unwichtig ist gibt es weniger Ergebnisse:  Ein Bild, das Shoji enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung |
| ***Übersicht*** |  |
| ***Beispiel Geordnete Probe ohne Wiederholung*** | Man 10 verschiedene Fruchtsorten und will jeweils eine Anordnung mit 4 machen. Dabei ist jede Sorte nur einmal vorhanden. Wie viele Kombinationen gibt es? |
| ***Beispiel Geordnete Probe mit Wiederholungen*** | Man hat wieder 10 verschiedene Früchte. Man will genau 7 Früchte aufstellen(also kann auch die gleiche Frucht mehrmals vorkommen):  nk = 107 = 10'000'000 Kombinationen |
| ***Beispiel Geordnete Probe mit Wiederholungen Schwierig*** | Man hat nun verschiedene Früchte einmal und eine Frucht zwei Mal. Nun will man alle Elemente anordnen:  //geteilt durch zwei, weil ein Element zwei Mal vorkommt🡪 Es können nur halb soviel Kombinationen auftreten.(Wäre das Element 3x vorgekommen 🡪 :3) |
| ***Beispiel Ungeordnete Proben ohne Wiederholung*** | Aus einer 30 köpfigen Schulklasse gewinnen 4 Schüler einen Preis. Wie viele Auswahlmöglichkeiten gibt es?  = |
| ***Beispiel Schachturnier*** | Klasse mit 30 Schüler möchte Spiel spielen, wobei alle gegen alle einmal spielen: |
| ***Beispiel Ungeordnete Probe mit Wiederholung*** | Eine Firma stellt Gummibärli mit fünf verschiedenen Geschmackssorten her. Die Tüten, welche Sie verkaufen beinhalten jeweils 12 Gummibärli. Wie viele verschiedene Gummibärlimischungen gibt es?  //beachte n ist die Anzahl verschiedener Gummibärli und k die Anzahl, welche in einen Pack hineingefüllt wird. |
| ***Beispiel Ungeordnete Probe mit Wiederholung*** | Die 6 grossen Parteien in der Schweiz kämpfen um 3 Sitze im Nationalrat. Wie viele Kombis gibt’s |
| ***Beispiel Stichproben*** | Eine Lieferung besteht aus 30 Glühbirnen, es werden jeweils 3 ohne zurücklegen entnommen.  Wie viele Strichproben sind möglich?  (30!)/(3! \*(30-3)!) =4060 Proben  Wie gross ist die Wahrscheinlichkeit, dass in einer Strichprobe genau 2 defekte Glühbirnen enthalten sind wenn von 30 genau 6 defekt sind=  P(2 defekte) = |
| ***Beispiel Lotto*** | Wie gross ist es im Lotto(5 aus 45) genau zwei richtige zu erhalten?  P(2 richtige) = |
| Bedingte Wahrscheinlichkeit | |
| ***Grundlagen*** | Es sei (Ω, A, P) ein Wahrscheinlichkeitsraum.  P(A∩B)  Sind A, B ∈ A Ereignisse und gilt P(B) > 0, so heißt    Die bediente Wahrscheinlichkeit von A unter der Bedingung B.  Man vergleicht also zwei Wahrscheinlichkeiten miteinander. |
| ***Beispiel Impfungen*** | Versucht, ob eine Impfung gegen eine Erkrankung helfen kann:  **Wahrscheinlichkeit einer Erkrankung**     * mit Impfung      * Die Erkrankung hängt also davon ab von der Bedingung «geimpft» oder «nicht geimpft» |
| ***Bemerkung*** | Die bedingte Wahrscheinlichkeit tritt bei mehrstufigen Experimenten auf, wenn man in den einzelnen Stufen Ereignisse protokolliert.   * Die Pfadregel besagt für den rot gezeichneten Pfad |
| ***Vorgehen*** | Ein Bild, das Himmel enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungUnter den 20 Schülern einer Klasse sind 4 Raucher. Von den 12 männlichen Schülern sind 3 Raucher. Gesucht wird ein neuer Klassensprecher.  Wie groß ist der Anteil der Männer unter der Bedingung, dass es sich um einen Raucher handelt?   1. Bezeichnungen auswählen   R: „Der ausgewählte Schüler ist Raucher.“ M: „Der ausgewählte Schüler ist männlich.“ ¯R: „Der ausgewählte Schüler ist Nichtraucher.“ ¯M: „Der ausgewählte Schüler ist weiblich.“   1. Gesucht definieren: Gesucht ist 2. Tabelle ausfüllen: 3. Wahrscheinlichkeit rauchender Mann berechnen |
| Unabhängige Wahrscheinlichkeit | |
| ***Grundlagen*** | Zwei Ereignisse A und B heißen stochastisch unabhängig, falls P(A\B) = P(A) gilt:  P(A) ist nicht von P(B) abhängig.   * Wenn also keine bedingte Wahrscheinlichkeit vorliegt, dann eine unabhängige. |
| VOLLSTÄNDIGE WAHRSCHEINLICHKEIT | |
| ***Grundlagen*** | Seien B1,…,Bn paarweise disjunkte Ereignisse, dann gilt:    Die Wahrscheinlichkeit A ist die Wahrscheinlichkeit A in Abhängigkeit Bi multipliziert mit der Wahrscheinlichkeit Bi  Wobei i der Index für die Anzahl disjunkten Ereignisse ist. |
| ***Rechenregeln allgemein*** |  |
| ***Beispiel Karten*** | Ein Bild, das Objekt, Uhr enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungAus einem Skatblatt (32 Karten, davon 4 Asse) werden nacheinander zwei Karten gezogen, ohne dass die erste Karte wieder zurückgelegt wird.  Wie groß ist die Wahrscheinlichkeit beim zweiten Zug ein Ass zu ziehen? |
| ***Beispiel*** | Ein Bild, das Text, Karte enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungEin Koch- Lehrling versalzt seine Suppe mit einer Wahrscheinlichkeit von 0.5. Wenn er verliebt ist –ein Zustand, in dem er sich mit Wahrscheinlichkeit 0.4 befindet – ist es ganz schlimm: Dann versalzt er nämlich 80 % seiner Suppen. §  P(S) allgemein = 0.5 P(L) = 0.4 🡪 P(¬L) = 0.6  P(S|L) = 0.8 🡪 P(S|¬L) = 0.2 🡪 P(S∩L) = 0.8 \* 0.4 = 0.32 🡪 P(S∩¬L) = 0.2 \* 0.4 = 0.08  P(S) = P(S∩L) + P(S∩¬L) 🡪 0.5 = 0.32 + P(S∩¬L) 🡪 P(S∩¬L) = 0.5 – 0.32 = 0.18  P(¬S∩¬L) = 1 - P(S∩L) - P(¬S∩L) - P(S∩¬L) = 1 – 0.32 – 0.08 – 0.18 = 0.42 |

|  |  |
| --- | --- |
| Zufallsvariable und die wichtigsten theoretischen Verteilungen | |
| Zufallsvariabeln | |
| ***Grundlagen*** | Zufallsvariable ist:   * Keine Variable * Ist nicht zufällig   Der Name ist völlig irreführend:  ***Die Zufallsvariable X ist eine Abbildung/Funktion die jedem Ergebnis* ω *eines Zufallsexperimentes einen Wert(aus dem Ereignisraum Ω) zuordnet.***  ***Für das gilt:***    ***Ein Bild, das Text, Karte enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung***Jede Menge Ai muss eine gewisse Wahrscheinlichkeit beinhalten: |
| ***Beispiele*** | Ein Roulett Spieler setzt einen Chip «Ereignis 1. Dutzend»:  Ein Bild, das Screenshot enthält.  Mit hoher Zuverlässigkeit generierte BeschreibungWenn die Zahlen 1-12 kommen, gibt es 3 Geldstücke, ansonsten wird das gesetzte weggenommen: |
| Diskrete Wahrscheinlichkeitsfunktion | |
| ***Grundlagen*** | Wie der Name bereits verrät, weist diese Funktion beliebigen Werten die Wahrscheinlichkeit zu.   * http://members.chello.at/gut.jutta.gerhard/kurs/wahrschv3.gifDie Summe aller Einzelwahrscheinlichkeit ist 1, entspricht der Fläche * Die Wahrscheinlichkeit P, dass ein Ereignis xi eintritt kann unmittelbar abgelesen werden * Die Wahrscheinlichkeit eines Intervalles xi, xi+1,.. xn ergibt sich: |
| Kontinuierliche Verteilungsdichtefunktion | |
| ***Grundlagen*** | http://upload.wikimedia.org/wikipedia/commons/thumb/5/5e/Breit-Wigner.png/220px-Breit-Wigner.pngZeigt die Verteilung der Wahrscheinlichkeiten   * Die Fläche unter der Kurve entsprich 1, Summe aller Wahrscheinlichkeiten(Integral) * Es kann nur die Wahrscheinlichkeit eines Bereiches berechnet werden(Weil nur Fläche(Integral) bestimmbar: |
| ***Einfache Mittelwert*** | Der Mittelwert einer Verteilungsfunktion kann berechnet werden: |
| ***Beispiel*** | Augensumme eines Würfels mit zwei Würfen:  Gesamtanzahl n an Möglichkeiten: 6\*6 = 36  Ein Bild, das Screenshot enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung |
| ***Varianz(Streuwert)*** | Auch der Streuwert – die Varianz – kann berechnet werden:  oder |
| ***Beispiel*** | Augensumme eines Würfels mit zwei Würfen:  Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungGesamtanzahl n an Möglichkeiten: 6\*6 = 36  n = 36 Beispiel für i = 2 (2-7)2\*1/36 = 0.7 |
| Wichtige diskrete Verteilungen | |
| ***Bernoulli Prozess*** | Als Bernoulli-Prozess wird eine ***wiederholte Durchführung eines Zufallsvorgangs*** bezeichnet, wobei gilt:   * + Bei jeder Wiederholung interessiert nur, ob ein bestimmtes Ereignis eintritt oder nicht   + Die Wiederholungen sind unabhängig.   + Die ***Erfolgswahrscheinlichkeit p*** bleibt gleich   + Die ***Misserfolgswahrscheinlichkeit q*** ist dann: ***q = 1-p***   + ***Erwartungswert: E(X) = p***   + ***Varianz: σ2 = pq*** |
| ***Binominal-Verteilung*** | Ist die ***bekannteste und wichtigste*** diskrete Verteilung:  **Wahrscheinlichkeitsfunktion**:  **Erwartungswert**:  **Varianz**  **Anwendung:**   * + - Das Zufallsexperiment unterscheidet nur zwei Ergebnisse     - Das Experiment wird ***n-mal wiederholt*** (Zufallsstichprobe vom Umfang n)     - ***Gesucht***: Die ***Wahrscheinlichkeit*, dass bei n-maliger Durchführung des Experimentes das Ereignis**        * **genau**       * **mindestens**       * **höchstens**   **x-mal eintrifft** |
| ***Beispiel*** | Gegen eine Krankheit wurde ein neues Medikament entwickelt. Die Heilungschance liegt bei 90%. Wie gross ist die Wahrscheinlichkeit, dass bei 5 zufällig gewählten Patienten mindestens 4 geheilt werden?(Addieren wegen midestens!!) |
| ***Poissonverteilung*** | Man interessiert sich dafür, ***wie hoch die Wahrscheinlichkeit*** ist, dass das ***Ereignis E in einem Intervall genau oder höchstens x-mal eintritt***, wenn bekannt ist, dass in diesem Intervall das Ereignis im Mittel μ - mal auftritt.  ***μ*** gibt also eine Rate pro Zeitintervall an(im ***Durchschnitt***)  **Wahrscheinlichkeitsfunktion**:  **Erwartungswert = Varianz:**  Unabhängige Ereignisse 🡪 Poisson Verteilt(Errinerungsfrei) |
| ***Beispiel*** | Die Anzahl X der Telefonanrufe, die in einer Telefonvermittlung im Mittel pro Minute ankommen sei Ps(1) – verteilt.   * + Wie gross ist die Wahrscheinlichkeit, dass in einer Minute ***μ = 1***     - Genau ein Anruf ankommt 🡪 P(1) = 0.368     - Höchstens ein Anruf 🡪 p(0) + p(1) = 0736     - Mindestens ein Anruf 1 – p(1) = 1 - 0.368 = 0.632     - Zwei oder drei Anrufe ankommen P(2) + p(3) = 0.245   + Wie gross ist die Wahrscheinlichkeit, dass in 5 Minuten genau 6 Anrufe ankommen: ***μ = 5***   + P(6) = 0.146 |
| Wichtigsten stetigen theoretischen Verteilungen | |
| ***Rechteckverteilung*** | * Rechteckverteilungen eignen sich zur Beschreibung von Vorgängen, bei denen die ***Ergebnisse nur Zahlen eines bestimmten Intervalls [a, b]*** sein können. * Die ***Wahrscheinlichkeit***, dass ein Ergebnis in ein ***bestimmtes*** ***Teilintervall*** fällt, wird nur ***durch*** ***dessen*** ***Länge*** bestimmt. * ***Alle Ergebnisse*** eines bestimmten Intervalls [a, b] sind ***gleich wahrscheinlich*** 🡪Gleichverteilung im Intervall [a, b] |
| ***Beispiel*** | Zwei Linien  Aufgabe 1 — Rechteck  Sie besuchen Freunde an der Uni Erfurt und wollen von dort mit der nächsten S-Bahn  weiterfahren und zwar zwischen 17 und 18 Uhr. Zwischen 17 und 18 Uhr fahren die  Straßenbahnen der Linien 3 und 6 ab Haltestelle Universität Richtung „Urbicher Kreuz" bzw.  „Rieth" jeweils im 10-Minuten Takt. Der erste Zug der Linie 3 fährt um 17:05 Uhr, der erste  Zug der Linie 6 um 17:00 Uhr. Wenn ein Sie zufällig zu einer gleichveneilten Zeit zwischen  17:00 Uhr und 18:00 Uhr die Haltestelle erreichen und einfach in die nächste Straßenbahn  einsteigt, wie groß ist die dann die Wahrscheinlichkeit, dass Sie eine Bahn in Richtung  „Rieth" nehmen?  Freihandzeichnungen Freihandzeichnungen ￼￼ ￼￼￼￼￼ ￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼￼  P(blau) = Fläche blau = 50% |
| ***Dreieckverteilung*** | Plot of the Triangular PMF f(x)=\begin{cases}   \frac{2(x-a)}{(b-a)(c-a)}, & \text{wenn } a \le x < c\\   \frac{2}{b-a},             & \text{wenn } x = c\\   \frac{2(b-x)}{(b-a)(b-c)}, & \text{wenn } c < x \le b. \end{cases}Erwartungswert oder Median:  Varianz: |
| ***Beispiele*** | 1. Beschreibung von Bedienprozessen Ziel ist die Ermittlung oder Beschreibung der Bearbeitungszeit, einer Dienstleistung.   z.B. Servicezeit an einer Kasse, Zeit eines Produktionsprozesses |
| ***Exponentialverteilung*** | **Exponentialverteilung mit Parameter λ > 0:**  **λ** = Irgendeine Rate z.B Anrufsrate |
| ***Beispiele*** | * Zeitspanne zwischen zwei Anrufen in einer Telefonzentrale. * Dauer eines Telefongesprächs. * **Lebensdauer eines Geräts**, wenn Defekte durch äußere Einflüsse  und **nicht durch Verschleiß** verursacht werden. |
| ***Beispiel Anruf*** | Ein Software-Hersteller hat für seine Kunden in Norddeutschland die Hotline ND eingerichtet. An Werktagen rufen zwischen 20.00 und 21 .00 Uhr durchschnittlich 5 Kunden an. ***Wie groß ist die Wahrscheinlichkeit, dass zwischen zwei Anrufen höchstens 5 bzw. 12 Minuten vergehen***?  **λ** = 5 Anrufe pro Stunde 🡪 5/60 🡪 1/12 Anruf/Minute  -> höchstens 5 Minuten: 1 – e-1/12 \* 5= 0.34 und -> höchstens 12 Minuten: 0.63  **🡪 Beachte: Immer die gleiche Masseinheit verwenden!** |
| ***Weilbull-Verteilung***  ***W(α, β)-Verteilung*** | **Weibull-Verteilung mit Parametern α (scale) > 0 und β (shape) > 0:** |
| ***Anwendung*** | Beschreibung der **Lebensdauer von Geräten oder Materialien mit Abnutzungserscheinungen**   * Ein Bild, das Screenshot enthält.    Mit sehr hoher Zuverlässigkeit generierte Beschreibungα ist ein Skalierungsparameter, β ist ein Formparameter. |
| ***Zentrale Grenzwertsatz*** | Der Zentrale Grenzwertsatz besagt, dass irgendwann, wenn man genug Zufallsvariablen(Werte) hat, alles näherungsweise normalverteilt ist. |
| ***Normalverteilung*** | Die wichtigste stetige Verteilung. Spielt in der schliessenden Statistik eine wichtige Rolle.  **Wahrscheinlichkeitsdichteverteilung**  **Verteilungsfunktion**(Integral lässt sich nicht lösen)  **Erwartungswert**  **Varianz** |
| ***Wichtige Anmerkungen*** | * Die Normalverteilung ist eine ***stetige symmetrische Verteilung*** * μ enspricht dem Erwartungswert(x-Wert des Scheitelpunktes) 🡪xmax = μ * Wird ***μ verändert, so hat es eine Verschiebung nach links oder rechts*** zu Folge * σ entspricht der Standardabweichung * Wird ***σ verändert, so hat es eine Veränderung in der Breite der Kurve*** zu Folge * Die Fläche zwischen zwei Werten entspricht der Wahrscheinlichkeit, dass x in diesem Bereich liegt.   Bildergebnis für normalverteilung |
| ***Wichtige Beziehungen*** | Es gibt folgende wichtige Beziehungen ***zwischen der Verteilfunktion F(X) und der Normalverteilung:*** |
| ***Beispiele*** | σ= 1 und m = 0, in dem Fall ist z = x   * + z = -∞ und z = 0 🡪 0.5   + z = 0 und z = 1.2 🡪0.385   + z = -1.2 und 0 🡪0.385   + z = 0.81 und z = 1.94 = 0.183 |

|  |  |
| --- | --- |
| ***Positive z Werte*** |  |
| ***Negative z Werte*** |  |
| ***Beispiel Zucker Abfüllanlage*** | Messungen ergeben, dass das Gewicht der Zucker- Pakete in Gramm um den Mittelwert μ= 1000 g um die Standardabweichung σ = 5g schwankt. Es kann von einer  Normalverteilung ausgegangen werden.   * Wie gross ist die Wahrscheinlichkeit, dass der tatsächliche Inhalt der Packung zwischen 990 Gramm und 1010 Gramm liegt   z1 = (x1- μ) / σ = (990-1000)/5 = -2  z1 = (x2- μ) / σ = (1010-1000)/5 = 2  p(2) = 0.9772  p(-2) = 0.0228  P(2) – P(-2) = 0.9544   * Wie gross ist die Wahrscheinlichkeit das in einer Packung mehr als 1015 Gramm enthalten sind   Z = (x- μ) / σ = (1015-1000)/5 =  P(3) = 0.9987  1 – p(3) = 0.0013   * Wie viel wiegt zu 90% ein zufällig entnommenes Paket höchstens?   0.9 aus Tabelle gelesen 🡪 z=1.285  1.285 = (x- 1000) / 5 🡪 x = 1006.43 g   * In welchem Bereich schwankt das Gewicht eines Paketes in 98% Prozent aller Fälle symmetrisch zum Mittelwert?   P(2.3) = 0.99  P(-2.3) = 0.01  2.3 = (x- 1000) / 5 = = 1011.5  -2.3 = (x- 1000) / 5 = 988.5  Es schwankt zwischen 988.5g und 1011.5 g  Dabei ist die Schwankung vom Mittelwert aus: |
| ***Beispiel Flugzeug*** | Bei einem bestimmten Grossraumflugzeug ist die Auslastung pro Flug näherungsweise normalverteilt, mit den Parametern μ = 150 Passagiere und σ = 25 Passagiere.  Mit welcher Anzahl von Passagieren ist mit einer Wahrscheinlichkeit von 90% **mindestens** zu rechnen?  P(1.285) = 0.9 //aus der Tabelle gelesen  1.285 = (x – 150) / 25 = 117.875  Es ist mindestens mit 118 Personen zu rechnen. |

|  |  |
| --- | --- |
| Schliessende Statistik | |
| ***Grundlagen*** | Folgende Fakten sind wichtig:   * Die schließende Statistik befasst sich mit ***Messreihen***,  deren ***konkrete Ausprägungen vom Zufall*** beeinflusst werden. * Diesen Messreihen können ***ebenfalls*** ***charakteristische*** ***Größen*** etwa in Form des ***arithmetischen*** ***Mittels*** und der ***Varianz*** zugeordnet werden. * Hier stellt sich die ***Frage*** nach der ***Gestalt*** ***des*** ***unbekannten*** ***Verteilungsgesetzes***  unter dem die vorliegende Realisierung zustande kam. * ***Anhand der charakteristischen Größen*** kann man die eingesetzte ***Verteilungsfunkton*** ***annehmen*** oder ***ablehnen***.   + Man hat eine angenommene Verteilungsfunktion und eine theoretische Verteilung, diese werden verglichen um eine passende Funktion zu finden. |
| ***Beispiel*** | ***Ziel der Schliessenden Statistik ist es aus einer Teilmenge einer Menge auf die ganze Menge zu schliessen***:  Beispielsweise nimmt man 1000 Studenten und untersucht, ob und wie sich der Alkoholkonsum von Studenten auf ihre Noten auswirkt.  Dann schliesst man aus den erhobenen Daten auf alle Studenten. |
| Zufallsstichproben | |
| ***Vorgehen / Problembereiche*** | 1. Die Stichproben- oder Datenanalyse 2. Die Auswertung oder Analyse der erhobenen Daten |
| ***Einfache Zufallsstichprobe*** | * Bekannteste Form der Stichprobe * Es werden Elemente(Proben) so ausgewählt, dass ***die Wahrscheinlichkeit für alle Elemente gleich gross ist.***   **Beispiel**: Ich befrage einzelne Studenten aus allen Hochschulen in der Schweiz über ihre Motivation zum Lernen. |
| ***Geschichtete Stichprobe*** | * Man ***bildet*** ***Gruppen***(Schichten) und ***entnimmt*** jeweils verschiedene ***Elemente*** ***aus*** den jeweiligen ***Gruppen***(Schichten) * Dabei haben ***Elemente in den*** gleichen ***Gruppen*** ähnliche oder noch besser ***gleiche*** ***Merkmale***(Homogenität der Elemente)   **Beispiel**: Ich gruppiere die Studenten der Schweiz anhand der Studienrichtigung und befrage sie über die Motivation zum Lernen. Dabei werden aus jeder Schicht(Studiengang) einige Personen befragt. |
| ***Klumpen Stichprobe*** | Es ***existieren*** ***Teilmengen***(Klumpen), ***welche*** die ***Grundgesamtheit*** relativ ***gut*** ***abbilden***. Dann werden Proben aus den Klumpen entnommen, die dann ausgewertet werden.  **Beispiel:** Die HSR und die HTW haben in den letzten Jahren in der Umfrage «Motivation beim Lernen» ähnlich abgestummen wie die Grundgesamtheit(alle Studenten Schweiz). Deshalb werden jetzt nur noch Proben aus dem Klumpen HTW und HSR genommen. |
| ***Systematische Stichprobe*** | Es werden ***Regeln***(eine ***Systematik***) ***definiert***, die auf wiederholbare Weise eine Zufallsähnliche Stichprobenauswahl ermöglicht.  **Beispiel:** Bei der Umfrage werden jeweils 10 Studenten aus jedem Kanton befragt. |
| ***Mehrstufige Stichprobe*** | Ist letztlich eine ***Kombination*** ***der*** oben benannten ***Verfahren*** |
| Modellbildung | |
| ***Vorgehen*** | 1. **Isolierende Abstraktion** Ermitteln der Merkmalsträger und Merkmale, die zur Modellbildung erforderlich sind. ***Suche den Kern des Experiments*** und blende Unwichtiges aus. 2. Verteilungsfunktion bestimmen. Daraus lassen sich beliebig viele Daten für die Experimentdurchführung generieren. 3. Daten für Modellverifikation und Modellvalidierung festlegen 4. Festlegen der control/respose-Daten |
| ***Anforderungen der Experimentauswertung*** | * + - Wie sieht die Stichprobenfunktion und ihre Wahrscheinlichkeitsverteilung aus?     - Wie kann ein Vertrauensintervall für den Mittelwert bzw. die Varianz einer Messreihe ermittelt werden?     - Wie kann der Stichprobenumfang, d.h. die Anzahl der Experimente, festgelegt werden?     - Welche etablierten Testverfahren gibt es? |
| Testverteilungen | |
| ***Grundlagen*** | Sind Verteilungen, die bei vielen statistischen Tests Verwendung finden.   * ***Normalverteilung***, wenn die Stichprobe ***n ≥ 30*** ist * ***t-Verteilung*** (auch Student t-Verteilung), wenn die ***Stichprobe < 30*** ist * Chi Quadrat Verteilung * Viele Stichproben sind annähernd Normalverteilt(Zentraler Grenzwertsatz) |
| Normalverteilung | |
| ***Konfidenzintervalle für den Wert m der*** | Liegt eine konkrete Stichprobe  {x1, x2 , x3, …, xn} vor,  so gilt für das **Stichprobenmittel**(der Erwartungswert = der Mittelwert)  **Stichprobenfunktion(= Stichprobenmittelwert)** |
| ***Satz 1*** | Es sei X ein Zufallsvariable mit dem Erwartungswert μ und der Varianz σ2.  **Stichprobenmittelwert**  Ein Bild, das Screenshot enthält.  Mit hoher Zuverlässigkeit generierte Beschreibung**Erwartungswert**  **Varianz** |
|  |  |
| ***Grafische Darstellung*** | Je mehr n man hat, desto näher kommt man an die Normalverteilung  (Zentraler Grenzwertsatz):  Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung |
| ***Satz 2*** | **Es gelten die Voraussetzungen von Satz 1. Ist *X* darüber hinaus normalverteilt, ist auch der Stichprobenmittelwert Normalverteilt.**  ***σ die Standardabweichung***  ***μ der Mittelwert***  ***n Anzahl der Stichproben***  **Stichprobenwert** |
| ***Beispiel*** | Sie haben das Simulationsmodell einer Produktionseinrichtung erstellt, vom realen System wissen sie, dass diese Maschine im ***normalverteilten Mittel 10 Stück pro Sekunde*** produziert und die ***Standardabweichung von 1 Stück pro Sekunde*** besitzt. Sie führen in der Simulationsumgebung ***25 Experimente*** (Replications) durch.  Gegeben: μ = 10 Stk./s und σ = 1 Stk./s und n=25   * 1. Mit welcher Wahrscheinlichkeit wird der mittlere Ausstoss zwischen 9.8 und 10.2 Stück pro Sekunde liegen?   Gesucht:  Z1 = 🡪 P(-1) = 0.1587(aus Tabelle)  Z2 = 🡪 P(1) = 0.8413(aus Tabelle)  P(-1 < x < 1) = 1 – (0.8413 – 0.1587) = 0.3174 = 31.74%   * 1. Mit welcher Wahrscheinlichkeit wird der Ausstoss über 10.2 liegen   Z =🡪 P(1) = 0.8413(aus Tabelle)  1 – 0.8413 = 0.1587 = 15.87%   * 1. Wie verändert sich das Ergebniss, wenn sie statt dessen 100 Experimente machen?   Z = *🡪* P(2) 0.9772(aus Tabelle)  1 - 0.9772= 0.0228 🡪 weniger Varianz - annähernd normalverteilt(Zentraler Grenzwertsatz) |
| ***Konfidenzintervall*** | **Konfidenzintervall**    **Beispiel:** |
| Nach unten begrenzt: Die Grenze ist unten. Mindestens: Nach oben begrenzt z,B (1-p(untere Grenze)):  Ein Bild, das Objekt, Himmel enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung | |
| t-Verteilung | |
| ***Grundlagen*** | T-Verteilungen werden eingesetzt, wenn n< 30. T-Verteilungen haben, wie Normalverteilungen ein glockenförmigen Verlauf, sind aber flacher und breiter. |
| ***Formeln*** |  |
| ***Freiheitsgrade r*** | Der **Freiheitsgrad r** bestimmt bei einer Gleichung, wie viele Parameter «**frei**» **wählbar** sind. |
|  | |

|  |  |
| --- | --- |
| Schätzverfahren & **Abschätzung** der unbekanntem Parameter **μ** und **σ** | |
| Schätzverfahren | |
| ***Grundlagen*** | Haben die Aufgabe den oder die ***unbekannten*** ***Parameter*** der Verteilung eines Merkmals in der Grundgesamtheit ***anhand einer Stichprobe zu schätzen***.  Die Schätzung kann   * + - durch die Angabe eines einzigen Wertes erfolgen → **Punktschätzung**     - durch die Angabe eines Intervalls → **Intervallschätzung** |
| Schätzfunktion | |
| ***Grundlagen*** | * mathematisches Instrument * Bindeglied zwischen Grundgesamtheit und Stichproben * Verteilung der Schätzfunktion approximativ bekannt * ***Ziel*** der Schätzfunktionen ist,   + ***von einer Stichprobe auf die Grundgesamtheit zu schliessen***,   + Den Fehler einer falschen Schätzung zu minimieren oder zu bestimmen   + **Punktschätzung**,     - ***Parameterschätzung*** für den ***Mittelwert*** μ, die ***Varianz*** σ2 und für ***unbekannte*** ***Wahrscheinlichkeiten*** (auch Anteilswerte genannt) p   + **Intervallschätzung**     - ***Bestimmung*** ***von*** ***Vertrauensintervallen*** für die oben aufgeführten Parameter und das damit verbundene ***Risiko*** einer ***Fehlentscheidung*** oder ***Fehlinterpretation*** ,     - Die ***Angabe*** des ***Fehlers*** oder der ***Genauigkeit*** ***einer*** ***Schätzung*** wird auch als ihre Zuverlässigkeit bezeichnet |
| ***Gütekriterien für Schätzfunktionen*** | Ist  Ein kleiner Restfehler ist ein Gütekriterium für eine Schätzfunktion. |
| ***Konstruktion der Schätzfunktion für μ*** | * Minimiere * Also minimiere * Differenzieren nach und 0 setzen ergibt die Schätzfunktion zu |
| ***Punktschätzung*** | Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte BeschreibungEs kann immer die t-Verteilung genommen werden, weil diese bei n > 30 ungefähr normalverteilt ist & bei n>30 besser als die Normalverteilung ist. |
| ***Intervallsschätzung für das Stichprobenmittel*** | Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung |
| ***Erstellung eines Konfidenzintervalls***  ***Bei konkreten Werten*** | 1. Feststellung der Verteilungsform von   entscheiden ob, kleiner oder grössergleich 0.05   1. Feststellung der Varianz von ggf. schätzen mit s2 2. Ermittlung des Quantilwertes ***z aus Tabelle*** oder Rechner 3. Berechnung des maximalen Schätzfehlers *Der* ***maximale******Schätzfehler*** *ist das* ***Produkt*** *aus* ***Quantilswert*** *und* ***Standardabweichung*** *von X .* 4. Ermittlung der Konfidenzgrenzen *Die* ***untere*** *und die* ***obere******Konfidenzgrenze*** *ergeben sich durch* ***Subtraktion*** *bzw.* ***Addition*** *des* ***maximalen******Schätzfehlers*** *vom bzw.* ***zum******Stichprobenmittel*** |
| ***Beispiel Wurstfabrik*** | In einer Wurstfabrik werden u.a. Leberwürste hergestellt. Aus langjährigen Messreihen ist bekannt, dass das Füllgewicht der Leberwürste normalverteilt ist. Das ***Soll***-***Mindestgewicht*** der Würste ***beträgt*** ***125 g.*** Aus der ***Tagesproduktion*** ***von 600*** ***Würsten*** ***wurden 26 Würste*** zufällig ohne Zurücklegen ***entnommen*** und gewogen. Die Messergebnisse für das Füllgewicht (in g) betrugen dabei:   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 128.4 | 123.8 | 123.5 | 126.9 | 125.5 | 123.1 | 124.9 | | 123.1 | 126.6 | 121.9 | 125.3 | 123.4 | 122.1 | 124 | | 123.3 | 123.2 | 123.2 | 124 | 122.8 | 127.1 | 125.7 | | 127.1 | 125.8 | 123.7 | 125.9 | 124.9 |  |  |  * Erstellung des zentralen 95%-Konfidenzintervalls für μ   1. Feststellung der Verteilungsform von :  =  2. Feststellung der Varianz von ggf. schätzen mit s2    3. Ermittlung des Quantilwertes z aus Tabelle oder Rechner(97.5%)  Z = 1.96  4. Berechnung des maximalen Schätzfehlers  *Der maximale Schätzfehler ist das Produkt aus Quantilswert und Standardabweichung von X .*  E = 1.96 \* 1.31 = 2.57  Ermittlung der Konfidenzgrenzen:  124.5g ± 2.57 = [121.93, 127.07] |
| ***Varianzen für die  Schätzfunktion*** | Ein Bild, das Screenshot enthält.  Mit sehr hoher Zuverlässigkeit generierte Beschreibung |
| ***Erstellung Konfidenzintervall***  ***Bei Prozentsätzen*** | 1. Feststellung der Verteilungsform von P die Schätzfunktion ist aprox. normalverteilt, wenn n\*P\*(1-P) > 9 normalverteilt 2. Feststellung der Varianz von P  siehe vorherige Folie 3. Ermittlung des Quantilswertes z 4. Berechnung des maximalen Schätzfehlers *Der maximale Schätzfehler ist das Produkt aus Quantilswert und Standardabweichung von P* 5. Ermittlung der Konfidenzgrenzen *Die untere und die obere Konfidenzgrenze ergeben sich durch Substraktion bzw. Addition des maximalen Schätzfehlers vom bzw. zum Stichprobenmittel P* |
| ***Beispiel*** | Ein Chemieunternehmen möchte den Bekanntheitsgrad eines von ihm hergestellten Waschmittels in Erfahrung bringen. Dazu werden ***400 Personen*** zufällig ausgewählt und befragt. Das Waschmittel war bei ***30 % der Befragten bekannt***.   * Erstellen des zentralen 95%-Konfidenzintervalls   1) Feststellung der Verteilungsform von P  N\*P\*(1-P) = 400 \* 0.3 \* 0.7 = 84 🡪 84>9 🡪normalverteilt  2) Feststellung der Varianz von P  3) Ermittlung des Quantilswerts(0.975)  Z = 1.96  4) Berechnung des maximalen Schätzfehlers  e = 1.96 \* 0.023 = 0.045 = 4.5 %  5) Ermittlung der Konfidenzgrenzen  30% ± 4.5 = [25.5, 34.5] |
| ***Anzahl Proben mit zurücklegen*** | Bestimmung des Z-Wertes***(Varianz nicht doppelt quadrieren)*** |
| ***Beispiel*** | Es ist bei einer ***Konfidenz von 95 %*** und einer bekannten ***Varianz von1.44 g2*** die ***erforderliche Anzahl von Proben zu ermitteln***, wenn eine ***Genauigkeit von 0.2 g*** gefordert wird. Annahme Normalverteilung, dann folgt für z = 1.96!    ***Achtung! Die Varianz ist = muss also nicht nochmals quadriert werden!*** |
| ***Anzahl Proben ohne Zurücklegen*** |  |
| ***Beispiel*** | Einer Lieferung von ***1000*** ***Paketen*** ***Zucker*** ist mit einer ***95 % Konfidenz*** bei einem Fehler von ***e = 0.2*** g zu untersuchen, ob bei einer bekannten Standardabweichung von 1.2 g der garantierte Mittelwert eingehalten wird. Wie viele Proben sind aus der Lieferung mindestens zu entnehmen?  Z Aus Tabelle = 1.96 |
| ***Konfidenzintervall für die Varianz I*** | *P*  *P* |
| ***Beispiel*** | Gegeben sind die folgenden Werte aus dem vorangegangenen Beispiel der Wurstfabrik:   * 1. *Erstellung des (zentralen) 95% Konfidenzintervalls für*   2. *Erstellung des noch oben begrenzten*   *95 % Konfidenzintervalls*    *= 0.95*  *14.6114,* |
| ***Länge des Konfidenzintervalls*** | Die Länge des Konfidenzintervalls variiert mit der Anzahl der Stichproben:  ***Wenige Stichproben 🡪 Weniger Vertrauen 🡪 Grosses Konfidenzintervall***  ***Viele Stichproben 🡪 Mehr Vertrauen 🡪 Kleiners Konfidenzintervall***  ***Mehr Vetrauen/Sicherheit durch:***   * + - * 1. ***Mehr Stichproben***         2. ***Grösseres Konfidenzintervall*** |
| ***Vorgehen wenn Sigma nicht vorgegeben*** | **bestimmen:**    ***Standardabweichung bestimmen***    ***Z bestimmen*** |
| ***Hypothesentest*** | **Co = //Maximalwert des Hypothese**  **Z in diesem Fall 1-0.01 = 0.99** |
|  | |
| Bildergebnis fÃ¼r chi quadrat tabelle | |