**Klausurvorbereitung**

**Theorie**

**Statistik**

* **Lehre von Verfahren und Methoden zur Gewinnung, Erfassung, Analyse, Charakterisierung, Abbildung, Nachbildung und Beurteilung von beobachtbaren Daten über die Wirklichkeit** (Empirie
* **systematische Zusammenstellung von Zahlen und Daten zur Beschreibung von Zuständen, Entwicklungen und Phänomen**

**Beschreibende Statistik**

* **sammelt Daten bei allen Untersuchungseinheiten**, über die man Informationen erhalten will
* empirische Daten **durch Tabellen, Kennzahlen, (Maßzahlen oder Parameter) und Grafiken**
* **Beurteilung von Daten basierend auf Stichproben**
* Dient der Daten-Beobachtung / Datenbeurteilung
* **Konzentration auf das Wesentliche**

**Schließende Statistik**

* **Schluss von einer Teilmenge (Stichprobenergebnissen) auf das Eigenschaften der Grundgesamtheit (GG)**
* **Grund für Stichproben: enorme Größe der GG**
* **Stichprobe: Auswahl einer Teilmenge aus der GG** zu der Informationen gesammelt werden sollen
* **Datenbeurteilung (Analyse) durch Schlüsse auf Basis unvollständiger Daten**
* Dient dem **Beweis oder der Widerlegung von Hypothesen, die sich auf GG beziehen**
* Basiert auf Wahrscheinlichkeitsrechnung

**Auswahlverfahren zur Stichprobe oder Umfang zur Stichprobe?**

* Das **Auswahlverfahren zur Erhebung der Stichproben und Bestimmung des Stichprobenumfangs ist wichtiger als der Stichprobenumfang.**
* **Auswahlverfahren**: **Art und Weise, wie Elemente der Stichprobe zweckmäßig ausgewählt werden**
* Auswahlverfahren bestimmt Stichprobenumfang.
* **schlechte Auswahl zur Stichprobe liefert unabhängig vom Stichprobenumfang keine brauchbaren Erkenntnisse über die GG** (Stichprobe ist „verzerrt“ bzw. „nicht repräsentativ

**willkürliche Auswahl**

* **unkontrollierte Aufnahme eines Elements aus der GG in die Stichprobe**
* **kein Auswahlplan**
* **Interviewer sind frei in der Auswahl ihrer Interviewpartner**.

**zufällige Auswahl (Random-Auswahl)**

* **Zufallsgesteuert**
* **jedes Element aus GG gelangt mit gleicher Wahrscheinlichkeit** **in die Stichprobe**
* **Voraussetzung: Liste/Datei zu allen Elementen der GG**
* **ermöglicht Berechnung von Stichprobenfehler** (durch Wahrscheinlichkeitsrechnung)
* **Einfache Zufallsstichprobe**:
  + **Jede Stichprobe hat dieselbe Chance ausgewählt zu werden**
* **Geschichtete Zufallsstichprobe**
  + **Elemente der GG werden in Gruppen** (Schichten) **eingeteilt** (ein Element gehört genau zu einer Gruppe (Schicht))
  + **nach Gruppierung werden einfache Zufallsstichproben aus jeder Gruppe genommen**
* **Klumpenstichprobe**
  + **Einfache Zufallsauswahl aus zusammengefassten Elementen (Klumpen, Cluster)**
  + Stichprobe wird aus allen Elementen bestimmter Cluster genommen

**Quota-Auswahl**

* **bewusstes Auswahlverfahren**
* ausgewählte **Quotierungsmerkmale sollen in der Stichprobe dieselbe Verteilung wie in der Grundgesamtheit** erreichen.
* **Quotenpläne** **für ausgewählte Merkmale (Geschlecht, Alter, Beruf)** **ermöglichen dieselbe Verteilung in der Stichprobe wie in der GG**
* Verteilung der Quotierungsmerkmale in der GG muss bekannt sein

**TED-Umfrage im Fernsehen**

* **willkürliche Auswahl**
* **kein Auswahlplan**   
  (jeder Zuschauer kann an Umfrage einmal, mehrmals oder nicht teilnehmen

**Warum „bis unter“ zum Wertebereich bei klassierten Daten?**

„bis unter“

* **ermöglicht eindeutige Klassierung bei diskreten metrischen Merkmalen**
* **vermeidet Überschneidungen bei den Merkmalsausprägungen**   
  (Klassen 100 bis 200 € und 200 bis 300 € beinhalten beide einen Umsatz von 200,00 €)
* **vermeidet Lücken bei den Merkmalsausprägungen**   
  (Klassen 100 bis 199 € und 200 bis 300 beinhalten keinen Umsatz zwischen 199,01 und 199,99 €)

**Probleme bei Klassierung von Daten**

* Zielkonflikt: Übersichtlichkeit versus Informationsverlust
* Berechnung von exakten statistischen Kennzahlen (z.B. Mittelwerte) ist bei klassierten Daten nicht möglich   
  Mittelwerte, Quantile sind bei klassierten Daten nur Näherungswerte
* Näherungswerte können nur unter bestimmten Annahmen (z.B. Gleichverteilung) berechnet werden

**offenen Randklassen**

* besitzen keine Klassenuntergrenze (z. B. „bis unter 50 kg“) oder keine Klassenobergrenze (z. B. „120 kg und schwerer“)
* Berechnung von Klassenbreite und Klassenmitte nicht möglich

**Berechnung der Klassenbreite bi für die Klasse „150 b.u. 180 cm“**

* **Klassenbreite bi** = **xk – xk-1** = **180 – 150 = 30 cm**

**Berechnung der Klassenmitte mi für die Klasse „150 b.u. 180 cm“**

* Klassenmitte **mi = (xk + xk-1) ½** = (180 + 150) / 2 = 330 / 2 = 165 cm

**Histogramm**

* **flächenproportionale Darstellung der Häufigkeiten von klassierten Daten**
* **x-Achse muss Skala mit geordneten Werten und gleichen Abständen zwischen den Werten** (z. B. 200, 400, 600, 800) sein
* **Rechtecke liegen auf x-Achse nebeneinander** (ohne Abstände)
* **Rechteckbreite = Klassenbreite bi**   
  (bei Klasse „150 b.u. 180“ wird die Rechteckbreite zwischen den Werten 150- 180 auf der x-Achse gezeichnet)
* **Rechteckhöhe** = absolute Klassenhäufigkeit hi / Klassenbreite bi 🡆 **hi / bi**

**Qualitative Merkmale = Rangmerkmale**

* **Nicht metrische Skalen**
* **lassen sich nicht mit Zahlen messen**
* **sind immer diskret** (**haben nur eine abzählbare Menge möglicher Merkmalswerte**)
* **Codierung und Rangordnung möglich**
* **Keine Rechenoperationen** mit den Merkmalsausprägungen zulässig
* **Skalenniveau: Nominalskala und Ordinalskala (Rangskala)**  
  **Beispiele:**  
  Geschlecht (Nominalskala)  
  Güteklasse (Rangskala)  
  Umsatzklasse (Rangskala)  
  Wohnort (Nominalskala)  
  Kundenzufriedenheit, gemessen auf einer Skala von „1 = sehr zufrieden“ bis „5 = sehr unzufrieden“ (Rangskala)  
  Beruf (Nominalskala)  
  Steuerklasse (Rangskala)  
  Einkommensklasse (Rangskala)

**Skalenniveau: nicht metrische Skalen**

* **bei qualitativen Merkmalen**
* Keine Rechenoperationen zulässig
* nur Vergleichsoperationen zulässig = ≠ < >
* **Nominalskala**
* **Verhältnisskala (Ratioskala)**
* **Absolutskala**

**Quantitative Merkmale = metrische Merkmale**

* **alle** **quantitativen Merkmale sind metrisch**!
* **Metrische Skalen sind Kardinalskalen (endliche (abzählbare) Anzahl (Menge) der Elemente)**
* Unterscheidung in
  + **stetige Merkmale**
    - Menge der Merkmalsausprägungen überabzählbar
    - **Intervall der reellen Zahlen** (weitere Zwischenwerte zwischen 2 Ausprägungen)   
      z.B. **Gewicht, Alter (wenn nicht Alter als ganzzahliger Wert in Jahren 🡆 dann ist es diskret), Fahrzeit**
  + **diskrete Merkmale**
    - **Menge der Merkmalsausprägungen ist endlich bzw. abzählbar**   
      (i.d.R. **ganze Zahlen**)   
      z.B. Kinderzahl, Sitzplätze, das monatliche Gehalt

**Skalenniveau: metrische Skalen**

* **bei quantitativen Merkmalen**
* **Rechenoperationen sind zulässig (+ - \* /)**
* Skala **hat Maßeinheit und Nullpunkt**(muss kein natürlicher Nullpunkt sein   
  (z. B. Temperatur in °C ist Intervallskala 🡆 hat Nullpunkt, aber keinen natürlichen Nullpunkt)
* Metrische Skalen (Kardinalskalen) **können auf alle statistische Verfahren angewandt** werden
* **Intervallskala**
* **Verhältnisskala (Ratioskala)**
* **Absolutskala**

z. B.

**Häufigkeiten,**

**arithmetisches Mittel**,

**Streuungsmaße** wie:

Standardabweichung,

Varianz,

Spannweite

* Metrische Skalen (Kardinalskalen) **sind abwärtskompatibel**
  + jedes Verfahren, dass auf niedrigem Skalenniveau angewandt werden kann, darf auch auf höherem Skalenniveau angewandt werden

z. B. darf Median-Berechnung zu einer Ordinalskala auch auf einer  
 Kardinalskala (Intervallskala) angewandt werden

Intervallskala

bei quantitativen Merkmalen / metrischen Merkmalen

Differenzbildung

Es werden nur Differenzen zwischen den Merkmalsausprägungen verglichen

z B. Längendifferenz: 100 cm – 50 cm

hat keinen natürlichen Nullpunkt,

natürlicher Nullpunkt ist von Natur aus gegeben

Der natürliche Nullpunkt ist nicht veränderbar.

Zur Einordnung der Skala die Frage stellen

Wie wurde der natürliche Nullpunkt bestimmt?

(natürlich oder willkürlich vom Menschen bestimmt)

nicht natürlicher (willkürlicher) Nullpunkt ist z. B. °C

Herr Celsius hätte jeden beliebigen anderen Punkt als 0°C festlegen

können

natürlicher Nullpunkt durch physikalische Größen (z. B. Gramm)

oder durch logische Annahmen (0 € Gewinn)

es lassen sich Reihenfolgen und quantifizierbare Abstände bilden

gleichgroße Intervalle zwischen den Ausprägungen (z. B. Temperaturskala °C enthält gleich große Intervalle von 1°C)

gleichgroße Intervallabstände sind interpretierbar

z. B. 6°C – 2°C = 12°C – 8°C

keine Verhältnisbildung möglich

Ausprägungen können auch Wert < 0 haben

Daten können alle (unendlich viele) Ausprägungen innerhalb eines Intervalls annehmen

z. B. Längendifferenzen, IQ, Temperatur in Celsius, Geburtsjahr

=, ≠, , +, -

**Grundgesamtheit**

* die Menge aller möglichen Erhebungseinheiten

**Stichprobe**

* eine n-elementige Teilmenge der Grundgesamtheit mit N Elementen (Merkmalsträgern)

**Merkmal**

* Im Rahmen der statistischen Erhebung relevante Eigenschaften der Merkmalsträger
* Statistische Variable  
  **Beispiele**:  
  Familienstand  
  Alter  
  Größe  
  Gewicht  
  Umsatz  
  Einkommen

**Merkmalsausprägung**

* mögliche Ausformungen eines Merkmals
* Merkmalswerte (Variablenwerte) / Beobachtungswerte in einer Werteliste  
  **Beispiele:**  
  {„ledig, verheiratet, geschieden“}  
  {„zufrieden“, „eher unzufrieden“, „unzufrieden“}  
  {„1 Std.“, 2 Std.“, „3 Std.“}  
  {10.000, 20.000, 30.000 }

**Merkmalsträger**

* Einzelnes Objekt einer statistischen Untersuchung, Träger der Informationen  
  Untersuchungseinheit, Erhebungseinheit   
  Beispiele:  
  Menschen  
  Kunden

**Merkmalswerte**

* einzelner Messwert / Beobachtungswert (Ergebnis) aus Merkmalsausprägungen (Werteliste)  
  **Beispiele:**  
  „ledig“ aus den Merkmalsausprägungen {„ledig, verheiratet, geschieden“}  
  „zufrieden“ aus den Merkmalsausprägungen {„zufrieden“, „eher unzufrieden“, „unzufrieden“}