

Teil II - Deskriptive Statistik

Tarek Carls

21. Oktober 2023

1.1 Ziel der deskriptiven Statistik

Hauptziel: Beschreibung der Stichprobe bezüglich erfasster Merkmale (Variablen).

- Daten in Tabellen zusammenfassen und ordnen.
- Erstellen von Diagrammen und Grafiken.
- Kenngrößen berechnen, z.B. Mittelwert, Standardabweichung.
- Unterscheidung: Univariate vs. multivariate Statistik.
- Fokus auf einzelnen Merkmalen.
- Hilft Daten zu verstehen.
- Wählt geeignete Verfahren für bi- und multivariate Analysen aus.
- In späteren Schritten werden dann komplexere Annahmen mit multivariaten, induktiven Verfahren getestet

1.2 Messniveaus

Merkmale in empirischen Untersuchungen haben unterschiedliche Messniveaus:

Nominalniveau:

- Keine geordneten Merkmalsausprägungen.
- Beispiele: Geschlecht, Haarfarbe.

Ordinalniveau:

- Geordnete Merkmalsausprägungen ohne gleiche Abstände.
- Beispiel: Schulnoten.

Intervallniveau:

- Geordnet, gleiche Abstände, kein natürlicher Nullpunkt.
- Beispiele: Temperatur in Celsius.

Verhältnissniveau:

- Geordnet, gleiche Abstände, natürlicher Nullpunkt.
- Beispiele: Gewicht, Länge.

1.3 Lage- und Streuungsmaße

Lagemaße: Zeigen, wo sich Stichprobenwerte konzentrieren.

Streuungsmaße: Beschreiben die Variabilität der Messwerte.

Wichtige Lagemaße:

- Modus (häufigster Wert)
- Median (mittlerer Wert)
- Mittelwert (Durchschnitt)
- Quartile

Wichtige Streuungsmaße:

- Spannweite
- Interquartilsabstand
- Standardabweichung

Häufigkeitstabellen

- Grundlage für erste Datenanalyse.
- Einfache und übersichtliche Darstellung.
- Informationen:
 - Absolute Häufigkeiten: Wie oft tritt eine Ausprägung auf?
 - Relative Häufigkeiten: Welcher Anteil der Gesamtstichprobe?
- Besonders sinnvoll bei kategorialen Daten.

1.4 Balkendiagramme

- Grafische Darstellung von kategorialen Daten.
- Jeder Balken repräsentiert eine Kategorie.
- Länge der Balken: Absolute oder relative Häufigkeiten.
- Gut geeignet für übersichtliche Vergleiche.

1.4 Kreisdiagramme

- V.a. für Nominalniveau.
- Darstellung der Anteile einzelner Kategorien an einem Ganzen.
- Jedes Segment repräsentiert eine Kategorie.
- Farbkodierung hilft bei der Unterscheidung.

1.4 Histogramme

- Häufig für kontinuierliche Daten genutzt.
- Darstellung der Verteilung von Daten.
- Bildung von Klassen aus den Daten.
- Keine Lücken zwischen Balken: Zeigt Kontinuität.
- Wichtiger Indikator für die Datenverteilung (z.B. Normalverteilung).

1.4 Boxplots

- Übersichtliche Darstellung von Streuungsmaßen.
- Zeigt Median, Quartile und Ausreißer.
- Besonders nützlich für den Vergleich mehrerer Gruppen.
- Interpretation:
 - Box: Interquartilsabstand (IQR).
 - Whisker: Datenbereich ohne Ausreißer.
 - Punkte: Mögliche Ausreißer.

Häufigkeitstabellen

- Grundlage für erste Datenanalyse.
- Einfache und übersichtliche Darstellung.
- Informationen:
 - Absolute Häufigkeiten: Wie oft tritt eine Ausprägung auf?
 - Relative Häufigkeiten: Welcher Anteil der Gesamtstichprobe?
- Besonders sinnvoll bei kategorialen Daten.

Wie machen wir das?

- Prozedur in SPSS: Analysieren → Deskriptive Statistiken → Häufigkeiten.
- Auswahl der Variable für die Häufigkeitstabelle.
- Ergebnis zeigt Anzahl gültiger Angaben und fehlende Werte.

2.2 Bedingte Häufigkeitsauszählungen

- Erstellung von Häufigkeitstabellen für spezifische Subgruppen.
- Beispiel: Häufigkeitstabelle nur für männliche Patienten.
- Auswahl in SPSS: Daten → Fälle auswählen.
- Danach verfahren wir genau so wie bei normalen Häufigkeitstabellen

2.3 Formate für Häufigkeitstabellen

- Auswahl des Formats über Mehrere Variablen auswählen.
- Optionen: Variablen vergleichen und Ausgabe nach Variablen ordnen.
- Möglichkeit, lange Häufigkeitstabellen zu unterdrücken.

3 Berechnung von statistischen Kennwerten

- Demonstration der Berechnung von Lage- und Streuungsmaßen in SPSS.
- Anwendung sowohl auf die gesamte Datendatei als auch auf spezifische Subgruppen.
- Wie machen wir das in SPSS?
 - Prozedur in SPSS: „Analysieren → Deskriptive Statistiken“.
 - Ergebnisse zeigen Kennwerte wie Mittelwert, Median, Modus, Standardabweichung, Varianz, Minimum, Maximum und Spannweite.
 - Anwendbar auf quantitative Variablen mit mindestens Intervallniveau.
- Möglichkeit, Kennwerte für spezifische Subgruppen zu berechnen.
 - Beispiel: Berechnung für die Variable SSportliche Betätigung nur für Patienten über 60 Jahre.
 - Temporäre Fallauswahl in SPSS zur Selektion spezifischer Gruppen.

4. Einleitung - Erstellen von Grafiken

- SPSS bietet viele Möglichkeiten, diverse Grafiken zu erstellen.
- Auswahl der Grafik basiert auf Messniveau und Merkmalsausprägungen.

4.1 Balkendiagramm - Schritte

- Anwendung:

- 1 Menüpunkt: *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten*
- 2 Variable auswählen, für die das Balkendiagramm angefertigt werden soll.
- 3 Schalter *Diagramme* betätigen.
- 4 Auswahl: *Balkendiagramme*.
- 5 Optionen: *Häufigkeiten* oder *Prozentwerte*.
- 6 Einstellungen bestätigen.

- Interpretation:

- Balkenhöhe zeigt absolute Häufigkeiten der Merkmalsausprägung.
- Häufigkeitstabellen ermöglichen genauere Interpretation.
- Häufigkeiten können direkt im Balkendiagramm angezeigt werden.

4.2 Kreisdiagramm - Schritte

- Anwendung:
 - Ähnliche Schritte wie beim Balkendiagramm.
 - Menüpunkt: *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten*
 - Schalter *Diagramme* betätigen.
 - Auswahl: *Kreisdiagramm*.
 - Option: *Prozentwerte*.
- Interpretation:
 - Jede Merkmalsausprägung stellt ein farbiges „Kuchenstück“ dar.
 - Legende erläutert die Farben der „Kuchenstücke“.
 - Diagramm kann bearbeitet werden, um genauere Informationen anzuzeigen.

4.3 Histogramm - Schritte

- Vorgehen:
 - Ähnliche Schritte wie bei den anderen Diagrammen.
 - Variable auswählen.
 - Menüpunkt: *Analysieren* → *Deskriptive Statistiken* → *Häufigkeiten*
 - Schalter *Diagramme* betätigen.
 - Auswahl: *Histogramm*.
 - Option: *Normalverteilungskurve im Histogramm anzeigen*.
- Interpretation:
 - Zeigt Verteilung der Daten.
 - Histogramme sind besonders nützlich für stetige Daten.
 - Interpretation basiert auf Form und Verteilung der Balken.

4.4 Boxplot: Eine Einführung

- Geeignet für die grafische Darstellung von stetigen Merkmalen.
- Boxplot gibt Aufschluss über Median, Quartile und Ausreißer.
- Einen Boxplot in SPSS erstellen:
 - Menüpunkt: Analysieren → Deskriptive Statistiken → Explorative Datenanalyse.
 - Wählen Sie die passende Variable aus.
 - Aktivieren Sie den Punkt "Diagramme".
 - Deaktivieren Sie die Checkbox *Stengel-Blatt* und bestätigen Sie.
 - Klicken Sie auf OK im Hauptdialog.

Wie man einen Boxplot interpretiert

- x-Achse: Bezieht sich auf die Variable.
- y-Achse: Skala für die Merkmalsausprägungen.
- Median, obere und untere Quartile sind sichtbar.
- Schnurrhaare repräsentieren den Wertebereich.
- Überprüfung auf Normalverteilung:
 - Median sollte mittig in der Box liegen.
 - Schnurrhaare sollten ungefähr gleich lang sein.
 - Ausreißer sollten sich gleichmäßig verteilen.
 - Eine Variable kann als normalverteilt betrachtet werden, wenn die oben genannten Kriterien erfüllt sind.
- Zusätzliche Statistiken: Aktiviere den Punkt *Beides* im Dialog.
- Dies gibt eine Tabelle mit deskriptiven Statistiken aus.
- Erlaubt eine genaue Interpretation des Boxplots.

5. Einleitung

- SPSS bietet die Möglichkeit, Tabellen und Grafiken nach ihrer Erstellung entsprechend den individuellen Wünschen des Benutzers anzupassen.
 - Tabelle im Viewer mit linker Maustaste aktivieren.
 - Roter Pfeil als Indikator zeigt erfolgreiche Aktivierung.
 - Rechte Maustaste öffnet Menüfenster.
 - Menüpunkt *Inhalt bearbeiten* auswählen.
- Bearbeitungsoptionen:
 - Wahl: Bearbeiten im Viewer oder separatem Fenster.
 - Anpassen von Tabellendimensionen und Rahmen.
 - Farbanpassung von einzelnen Zellen.
 - Wichtig: Änderungen immer bestätigen.

5.2 Zugriff auf den Diagramm-Editor

- Wie kann ich Grafiken editieren?
 - 1 Doppelklick auf Grafik oder Aktivieren.
 - 2 Roter Pfeil zeigt erfolgreiche Aktivierung.
 - 3 Kontextmenü über rechte Maustaste.
 - 4 *Inhalt bearbeiten* → *In separatem Fenster*.
- Optionen:
 - 1 Balken auswählen und Kontextmenü öffnen.
 - 2 *Elemente* für Anzeigen von Häufigkeiten pro Balken.
 - 3 Farbanpassung über *Auswählen* → *Alles Balken*.
 - 4 Fenster *Eigenschaften* für weitere Anpassungen.
- Weitere Anpassungen:
 - Farbänderungen über Registerkarte *Füllung und Rahmen*.
 - 3D-Effekte über Registerkarte *Tiefe und Winkel*.
 - Bestätigen mit *Zuweisen*.

Einführung

- Deskriptive Betrachtung als erster Schritt in statistischen Analysen.
- Kreuztabellen bieten einen Einblick in die deskriptiven Beziehungen zwischen zwei Variablen.
- Beispiel: Unterschiede zwischen Männern und Frauen in Bezug auf das Auftreten von chronischen Krankheiten.

Kreuztabellen in SPSS

- Menü: Analysieren → deskriptive Statistik → Kreuztabellen.
- Zeilen: Unabhängige Variable.
- Spalten: Abhängige Variable.
- Zeilenprozentage aktivieren für Interpretation.

Beispiel – Kreuztabelle: Geschlecht und Raucherstatus

		Raucher		Randsumme
		Ja	Nein	
Geschlecht	Männlich	30	70	100
	Weiblich	40	60	100
	Randsumme	70	130	200

Chi-Quadrat-Test

- Prüft Unterschiede zwischen Gruppen hinsichtlich eines Merkmals.
- Anwendbar auf 2x2-Kreuztabellen oder mehr.
- Angewendet auf das Beispiel von vorhin:
 - H_0 : Kein Unterschied bezüglich Raucherstatus zwischen Männern und Frauen.
 - H_1 : Es gibt einen Unterschied.
 - Formale Beschreibung:
 - $H_0 : p_{\text{raucherF}} = p_{\text{raucherM}}$
 - $H_1 : p_{\text{raucherF}} \neq p_{\text{raucherM}}$

Hypothesen des Chi-Quadrat-Tests

- H_0 : Kein Unterschied bezüglich chronischen Erkrankungen zwischen Männern und Frauen.
- H_1 : Es gibt einen Unterschied.
- Formale Beschreibung:
 - $H_0 : p_{\text{chronF}} = p_{\text{chronM}}$
 - $H_1 : p_{\text{chronF}} \neq p_{\text{chronM}}$
- Durchführung:
 - Menü: Analysieren → Deskriptive Statistiken → Kreuztabellen.
 - Statistiken → Chi-Quadrat aktivieren.

Einführung

- Vergleich von stetigen Variablen mittels t-Tests.
- Unterschiedliche t-Tests je nach Datensituation.
 - Einstichproben t-Test
 - Zweistichproben t-Test (unabhängige Stichproben)
 - Zweistichproben t-Test (abhängige Stichproben)
- Wichtigste Voraussetzung: Annahme der Normalverteilung der Variablen.

Der Einstichproben t-Test

- Testet, ob der Mittelwert einer Stichprobe von einem festen Wert abweicht.
- Beispiel: Haben Patienten während der Reha täglich 60 Minuten Sport getrieben?
- Hypothesen: $H_0 : \mu_{sport} = 60$ Minuten und $H_1 : \mu_{sport} \neq 60$ Minuten.

t-Test für zwei unabhängige Stichproben

- Vergleich der Mittelwerte von zwei unabhängigen Gruppen.
- Beispiel: Zufriedenheitsscore in zwei verschiedenen Reha-Zentren.
- Ergebnisse abhängig vom Levene-Test zur Überprüfung der Varianzgleichheit.

t-Test für zwei abhängige Stichproben

- Vergleich von zwei Messungen pro Subjekt (z.B. vorher-nachher-Vergleich).
- Beispiel: Unterschied im Körpergewicht vor und nach der Reha.
- Signifikanter Unterschied, wenn p-Wert kleiner als 0,05.

Einführung

- Analyse des Zusammenhangs zwischen zwei Variablen.
- Beispiel: Zusammenhang zwischen Körpergrösse und Gewicht.
- Ziel: Beschreibung des Zusammenhangs zwischen den Variablen.

8.2 Korrelationskoeffizient nach Pearson

- Anwendung bei vermutetem linearem Zusammenhang.
- Voraussetzung: Beide Variablen stetig oder intervallskaliert.
- Hypothesenbeispiel:
 - H_0 : Kein Zusammenhang zwischen Körpergröße und Gewicht.
 - H_1 : Zusammenhang zwischen Körpergröße und Gewicht existiert.
 - Korrelationskoeffizient von 1: Perfekter positiver linearer Zusammenhang.
 - Korrelationskoeffizient von -1: Perfekter negativer linearer Zusammenhang.
 - Korrelationskoeffizient von 0: Kein linearer Zusammenhang.
 - Beispiel: Korrelationskoeffizient von 0,965 zwischen Körpergröße und Gewicht zeigt starken positiven Zusammenhang.

Korrelationskoeffizient nach Spearman

- Geeignet für ordinale Variablen oder Mischung aus stetigen und ordinalen Variablen.
- Nicht zwingend linearer Zusammenhang erforderlich.
- Rangkorrelationskoeffizient basierend auf Rängen.