

Statistik I

1. Motivation: Wozu braucht man Statistik?

- Organisation und Ablauf
- Welche Geschichte erzählen die Daten: Deskriptive Statistik
- Ein Signal nachweisen: Induktive Statistik
- Definition: Univariate und bivariate Deskription (Definition)
- Rohdaten: Merkmale, Merkmalsträger, Merkmalsausprägungen
- Transparenz ist sehr wichtig!

2. Qualitative Merkmale

- Nominales und ordinales Messniveau (qualitative Merkmale)
- Beschreibung von qualitativen Merkmalen durch Grafiken
 - Einfache und gruppierte Säulendiagramme interpretieren
 - Das gruppierte Säulendiagramm aus zwei Perspektiven (Rolle der abhängigen bzw. unabhängigen Variable)
- Definition: Absolute und relative Häufigkeit
- Definition: 2×2 Kreuztabellen
- Definition: Univariate und bivariate Deskription in 2×2 Kreuztabellen

3. Wie analysiere ich qualitative Merkmale?

- Wiederholung: 2×2 Kreuztabellen
- Der Φ -Koeffizient in 2×2 Kreuztabellen
- Univariate Deskription und Exploration
 - Lagemaße: Modus und Median
 - Quantile und kumulative Häufigkeit
- Bivariate Deskription und Exploration
 - Kreuztabellen ($k \times m$), bedingte Häufigkeit
 - χ^2 -Koeffizient
 - Erklärung: Zufallsschwankung (Signal und Rauschen)

4. Signal und Rauschen trennen: Was ist Zufall

- Die Bedingte Verteilung eines nominalen Merkmals geg. eines dichotomen Merkmals in der Theorie
- Beispiel: Merkmal X ist dichotom (0 oder 1), Y hat vier Ausprägungen, $f_{Y|X}$ durch Festlegung von W.keiten
- Theoretische Verteilung $f_{Y|X}$ und beobachtete Daten: Man beobachtet in verschiedenen Stichproben unterschiedliche Verteilungen
- Unter der Bedingung $f_{Y|X=0} = f_{Y|X=1}$ welche Werte nimmt der χ^2 -Koeffizient an?

- Ab wann gehen wir davon aus, dass der beobachtete Werte des χ^2 -Koeffizient auf einen Unterschied zwischen $f_{Y|X=0}$ und $f_{Y|X=1}$ hinweist?
- χ^2 -Verteilung, ihre 95%-Quantil grafisch Verstehen
- Cramer's V und der (korr.) Kontingenzkoeffizient

5. Univariate Deskription von quantitativen Merkmalen

- Metrische Merkmale: Intervall- und Verhältnisskala
- Lagemaße und Streuungsmaße: Modus, Median, Quantile und Mittelwert
- Spannweite, Interquartilsabstand, Varianz und Standardabweichung
- Graphische Darstellung: Boxplot und Histogramm
- Kumulative Häufigkeit und emp. Verteilungsfunktion
- Lageregel: symmetrische und unsymmetrische Verteilungen

6. Die Normalverteilung als univariate Modellverteilung

- Historische Einstieg in die Normalverteilung
- Definition: Modell bzw. Modellverteilung mit der Intention zu Beschreiben
- Erwartungswert beschreibt die Lage des Symmetrieachse
- Standardabweichung beschreibt Breite der Kurve
- Die 68 – 95 – 99.7-Regel

7. Multivariate Deskription und Exploration von quantitativen Merkmalen

- Standardisierung
- Kovarianz und Korrelation nach Pearson bei standardisierten Merkmalen
- Korrelation nach Pearson allgemein.
- Was ist ein gleichsinniger bzw. gegensinniger linearer Zusammenhang?
- Korrelation und zweidimensionale lineare Regression (deskriptiv)
- Graphische Darstellung: Streudiagramm und Regressionsgerade

8. Bivariate Analyse von quantitativen und qualitativen Merkmalen

- Monotone Zusammenhänge und Korrelation nach Spearman
- Kandells τ_α
- Vergleich von Boxplots
- Die durch die Regressionsgerade erklärte Streuung: Was ist ein Modellfehler? (rein deskriptiv, einfach Beschreiben, wie gut das Modell ist)

9. Vergleich von Mittelwerten und das Konzept des Konfidenzintervalls

- Vergleich von Mittelwerten in zwei Gruppen
- Wie groß muss der Unterschied sein, damit man ihn nicht durch Zufallsschwankungen erklären kann.
- Was ist ein Konfidenzintervall (KI) am Beispiel der Mittelwertdifferenz
 - bei verbundener Messung
 - bei unverbundener Messung
- Interpretation des KI als Menge aller Nullhypothesen, die durch die Daten nicht abgelehnt werden