

Syllabus

February 15, 2021

1. Einführung. Definition von (Bio)Statistik. Beschreibende und schliessende Statistik. Heterogene Daten. Statistische Einheit (Merkmalsträger oder Untersuchungseinheiten). Merkmale und deren Auswahl. Grundgesamtheit und Stichprobe. Skalenniveaus.
2. Beschreibende Statistik. Tabellen (Urliste, Primärtafel, Häufigkeitstabelle, ~~Stamm-Blatt-Diagramm~~). Charakteristische Maßzahlen. Lageparameter: Variationsbreite (Range), (arithmetischer) Mittelwert, Median, Modus oder Modaler Wert, geometrischer Mittelwert, harmonischer Mittelwert. Streuungsmaße: Varianz und Standardabweichung, *mittlere Fehler des Mittelwertes*, Kumulative Verteilung, Quantile, Variationskoeffizient. Graphiken: Balkendiagramm, Komponentenbalkendiagramm, Histogramm, relative und absolute Häufigkeiten, Klassenbreite nach Sturges, Boxplot. Grammar of graphics (*ggplot*). Univariate und bivariate Daten. *face_wrap* und *face_grid*.
3. Schliessende Statistik. Schätzen und Test-Theorie. Statistische Hypothesen: Nullhypothese und alternative Hypothese. Fehler 1. (α) und 2. Art (β). Signifikanzniveau. False negatives und false positives. Punkt- und Intervallschätzer. Konfidenzintervall. Freiheitsgrade. t -Verteilung und Test.
4. Regression. Antwort- und Prädiktor-Variablen. Intercept. Slope. Fehlerterm. Einfache und mehrfache Regression. Ziele der Regression. Kausalität. ANOVA-Tabelle. F -Test und t -Test. Standardfehler (error of the mean). Standardfehler einer neuen Beobachtung. Konfidenzintervall. Modellvergleiche. Bestimmtheitsmaß (coefficient of determination), R^2 , Angepasstes R^2 . Akaike Information Criterion. Multiple Lineare Regression. Haupteffekte und Interaktionen. Modelldiagnostik. Unabhängigkeit, Homoskedastizität, Normalverteilung von Residuen, Linearität. Q-Q plot. Shapiro-Wilk-Test. Hebelwert (leverage). Component plus residual plot. Non-constant Variance Score Test (*ncvTest*). Variance inflation factor. Einflüßreicher Messwerte: Ausreißer, Cooks D-Wert, Added-variable plots. Generalisierte Lineare Modelle. Logistische Regression. Poisson-Regression.
5. Zusammenhänge in den Daten. Wahrscheinlichkeitsverteilungen. Normalverteilung oder Gauß-Verteilung. *Binomiale Verteilung*. Poisson-Verteilung. Zusammenhänge in kategoriellen Daten. Kontingenztabelle. Pearsonsche Chi-Quadrat-Test (χ^2 -Test). Exakter Test nach Fisher. Yates-Korrektur. G-Test. Relative Chance (Odds-Ratio). Kontingenzkoeffizient. Zusammenhänge in kontinuierlichen Daten. Pearson-Korrelationskoeffizient, Spearman-Rangkorrelationskoeffizient. Kendall-tau.
6. ANOVA. Einfache und mehrfache ANOVA. Unabhängige ANOVA. Levene-Test. Bartlett-Test. Welch's F. Robuste ANOVA. Kontraste. *Post hoc*-Tests. Bonferroni- und Benjamini-Hochberg-Test. Tukey HSD paarweise Vergleiche. ANOVA mit Messwiederholungen (repeated measures).

Spherizität. Mauchly-Test. Greenhouse-Geisser-Korrektur und Huynh-Feldt-Korrektur von F -Verhältnis. ANOVA als hierarchisches GLM-Modell. Faktorielle ANOVA. ANOVA mit gemischtem Design.

7. Nichparametrische Tests. U-Test von Mann und Whitney (Wilcoxon-Rangsummen-Test). Wilcoxon-Test für Paardierenz (Wilcoxon's signed-ranks test). Kruskal-Wallis-Test. Friedman-Test.
8. Diskriminanzanalyse. Lineare Diskriminanzanalyse (LDA). Versuchplanung.

Literatur

1. A. Field. Discovering Statistics using R.
2. Peter Bruce, Andrew Bruce & Peter Gedeck. Practical Statistics for Data Scientists
3. R. Kabacoff. R in Action.
4. M. Crawley. The R Book.
5. W. Köhler et al. Biostatistik. Eine Einführung für Biologen und Agrarwissenschaftler