**Testatfragen zu Vorlesung IV**

31. Warum vermeidet man beim Mittelwertvergleich für drei oder mehr unabhängige Stichproben das mehrfache Testen?

32. Beschreibe den Ablauf einer einfaktoriellen ANOVA.

33. Einfaktorielle ANOVA: Wie lauten Null- und Alternativhypothese beim Globaltest?

34. Einfaktorielle ANOVA: Wie ist die Prüfgröße beim Globaltest verteilt?

35. Erläutere: Bonferroni-Korrektur beim multiplen Testen.

36. Beim Vergleich von zwei Stichproben $\mathcal{S}_1$ und $\mathcal{S}_2$ von Lebensaltern ergab sich eine "stochastic superiority" von $\widehat{A} = 0.400$. Was bedeutet das?

37. Wie lauten die Shift-Hypothesen beim U-Test, und welche Folgerungen kann man daraus ziehen?

38. Wie werden die Rangsummen $R'$, $R''$ beim U-Test gebildet?

39. Erkläre, wie die Korrektur der Ränge für Bindungen erfolgt.

40. Wie werden beim Wilcoxon-Rangsummentest die Wertepaare mit Differenz Null eingeordnet?

**Anmerkungen**

∗ *NYHA-Klasse* (p. 4-02): Siehe [ DKG (2013) ], p. 7. "NYHA" bezieht sich auf die New York Heart Association, die diese Klassifizierung bereits 1928 eingeführt hat.

∗ *6-Minuten-Gehtest* (p. 4-02): Die Wegstrecke, die eine gesunde Person in sechs Minuten zurücklegen würde, wurde in Abhängigkeit von Alter, Größe und Gewicht in [ Enright/Sherrill (1998) ] abgeschätzt.

∗ *Mehrfaches Testen bei R unabhängigen Stichproben* (p. 4-03): Es entstehen insgesamt $\binom{R}{2} = R \cdot (R-1)/2$ ungeordnete Paare von Stichproben, für die jeweils ein Test durchgeführt werden müßte.

∗ *ANOVA* (pp. 4-04 ff.): Die Grundgedanken der Methode stammen aus [ Fisher (1925) ]. "Post hoc" (lateinisch): "nachher".

∗ *Zerlegung der Gesamt-Varianz und Berechnungsschema* (pp. 4-05 ff.): Ich folge [ Brown (2020) ], habe aber die dort verwendeten Bezeichnungen geändert. Die Abkürzung "SS" bedeutet "sum of squares". Es lohnt sich, die Zerlegung noch einmal nachzurechnen:

$$SS_{\mathsf{Gesamt}} = \sum_n \sum_r \big( Z(n,r) - \widehat{\mu} \big)^2 = \sum_n \sum_r Z(n,r)^2 - 2 \cdot \widehat{\mu} \cdot \underbrace{\sum_n \sum_r Z(n,r)}_{N \cdot \hat{\mu}} + N \cdot \widehat{\mu}^2 = \sum_n \sum_r Z(n,r)^2 - N \cdot \widehat{\mu}^2$$

$$SS_{\mathsf{RestZwSubj}} = \sum_r \big( N_{(r)} - 1 \big) \cdot \widehat{\sigma}^2_{(r)} = \sum_r \Big( \sum_n \big( Z(n,r) - \widehat{\mu}_{(r)} \big)^2 \Big) = \sum_r \Big( \sum_n Z(n,r)^2 - N_{(r)} \cdot \widehat{\mu}^2_{(r)} \Big)$$

(Rechnung in jeder einzelnen Gruppe analog zur ersten Zeile)

$$\begin{aligned} SS_{\mathsf{Gruppe}} = \sum_r N_{(r)} \cdot \big( \widehat{\mu}_{(r)} - \widehat{\mu} \big)^2 &= \sum_r N_{(r)} \cdot \widehat{\mu}^2_{(r)} - 2 \cdot \widehat{\mu} \cdot \sum_r N_{(r)} \cdot \widehat{\mu}_{(r)} + \sum_r N_{(r)} \cdot \widehat{\mu}^2 \\ &= \sum_r N_{(r)} \cdot \widehat{\mu}^2_{(r)} - 2 \cdot \widehat{\mu} \cdot \underbrace{\sum_n \sum_r Z(n,r)}_{N \cdot \hat{\mu}} + N \cdot \widehat{\mu}^2 = \sum_r N_{(r)} \cdot \widehat{\mu}^2_{(r)} - N \cdot \widehat{\mu}^2 \end{aligned}$$

Tatsächlich erhalten wir: $SS_{\mathsf{Gesamt}} = SS_{\mathsf{Gruppe}} + SS_{\mathsf{RestZwSubj}}$.

∗ *Bonferroni-Korrektur* (p. 4-10): Sie wurde 1936 von dem italienischen Mathematiker C. E. Bonferroni eingeführt und wird wegen ihrer Einprägsamkeit bis heute benutzt. Andere Verfahren zur Korrektur des Signifikanzniveaus verringern die Power nicht so stark, sind aber komplizierter.

∗ *Robuste ANOVA nach Welch* (p. 4-11): Auf die formelmäßige Darstellung wird in unserem Kurs verzichtet. Wo immer möglich, sollte der Vorteil der Unempfindlichkeit gegenüber inhomogenen Varianzen genutzt werden.

∗ *Hundertjährige Wetterereignisse* (p. 4-12 f.): Der Begriff "hundertjähriges Ereignis" stammt aus der Geowissenschaft. Dabei ist zu beachten, daß sich die Wahrscheinlichkeit des Auftretens "in the long run" ergibt: der konkrete Abstand zwischen zwei aufeinanderfolgenden hundertjährigen Ereignissen derselben Kategorie kann viel kürzer oder länger als 100 Jahre ausfallen.

∗ *Anzahl möglicher Extremereignisse* (p. 4-13): Durch unsere Überschlagsrechnung wird die Anzahl der möglichen Extremereignisse grob unterschätzt. Der maximale Vorhersagehorizont für das beim Deutschen Wetterdienst für Deutschland und Mitteleuropa verwendete Regionalmodell ICON EU beträgt 120 Stunden (5 Kalendertage), siehe `www.dwd.de/DE/leistungen/modellvorhersagedaten/modellvorhersagedaten.html` (15.10.2026). Wetterereignisse an zwei Kalendertagen mit einer Woche Abstand werden daher als unabhängig angesehen, ebenso Ereignisse in aufeinanderfolgenden Monaten oder Jahreszeiten. Ein "Rekordereignis" bezieht sich jedoch nicht nur auf den Zeitpunkt, sondern auch auf den Ort bzw. die Region des Eintretens, wobei die Zahl von 20 (über das Bundesgebiet oder Mitteleuropa verteilten) Lokalitäten natürlich viel zu niedrig ist.

∗ *Mindestens ein Wetterextrem pro Monat* (p. 4-13): Tritt ein einzelnes extremes Wetterereignis mit der Wahrscheinlichkeit $p$ einmal im Kalenderjahr auf, so kann die Wahrscheinlichkeit $q$ dafür, daß im laufenden Jahr mindestens 12 verschiedene der 4400 möglichen Ereignisse vorkommen, wie folgt berechnet werden:

$$q = 1 - P_0 - P_1 - P_2 - \ ... \ - P_{11} \quad \text{mit} \quad P_j = \binom{4400}{j} \cdot (1-p)^{4400-j} \cdot p^j \ , \ 0 \leqslant j \leqslant 11 \, .$$

$P_j$ ist gerade die Wahrscheinlichkeit dafür, daß im laufenden Jahr genau $j$ der 4400 Ereignisse eintreten. Für unser ursprüngliches $p = 0.01$ beträgt $q$ immer noch 99.99 %. Angesichts des Umfangs der Informationen, die uns heute zur Verfügung stehen, ist es also völlig normal, daß man jeden Monat mindestens ein hundertjähriges Wetterextrem beobachten kann.

∗ *Erste Beobachtung zur stochastic superiority* (p. 4-16): Nach dem Prinzip "Fläche gleich Wahrscheinlichkeit" gilt sowohl

$$\begin{aligned} Prob\{\, x' < c \,\} \cdot Prob\{\, x'' > c \,\} &= \int_{-\infty}^{c} f'(x)\, dx \cdot \int_{c}^{+\infty} f''(x)\, dx \quad \text{als auch} \\ Prob\{\, x' > c \,\} \cdot Prob\{\, x'' < c \,\} &= \int_{c}^{+\infty} f'(x)\, dx \cdot \int_{-\infty}^{c} f''(x)\, dx \, , \end{aligned}$$

aber wegen der Identität $f'(x) = f''(x)$ für alle $x \in \mathbb{R}$ stimmen beide Ausdrücke überein, und damit auch ihre Integrale bezüglich $c$.

∗ *Zweite Beobachtung zur stochastic superiority* (p. 4-17): Die Werte für $\widehat{A}$ wurden mit der weiter unten auf p. 4-25 eingeführten Formel berechnet.

∗ *Mann-Whitney-U-Test* (pp. 4-18 ff.): Der Test wurde zuerst von [ Wilcoxon (1945) ] vorgeschlagen und später von [ Mann/Whitney (1947) ] vollständig ausgearbeitet. Wir bezeichnen in unserem Kurs zur besseren Unterscheidung den Test für unabhängige Stichproben als "Mann-Whitney-U-Test" und den Test für verbundene Stichproben als "Wilcoxon-Rangsummentest". In der Literatur werden die Bezeichnungen der beiden Tests meist nicht unterschieden (z.B. "Wilcoxon-Mann-Whitney-Test"), so daß man stets nachprüfen muß, welches Verfahren konkret gemeint ist.

∗ *Hypothesen beim Mann-Whitney-U-Test* (p. 4-18): Sie werden oft dahingehend zusammengefaßt, daß der Test die "zentralen Tendenzen" zweier Gesamtpopulationen vergleicht. [ MANN/WHITNEY (1947) ] haben ursprünglich die Basis-Hypothesen benutzt. Die Shift-Hypothesen und ihre Folgerungen für Mediane und Mittelwerte werden natürlich auch dann verwendet, wenn die Gestalt der Verteilungen nicht exakt übereinstimmt, sondern nur annähernd gleich ist.

∗ *Existenz von Median und Mittelwert* für eine metrische Variable mit stetiger Verteilungsdichte (p. 4-19): Im Gegensatz zum Median, der immer gebildet werden kann, existiert der Mittelwert bzw. Erwartungswert nicht in allen Fällen, weil es Dichtefunktionen $f(x)$ gibt, bei denen man Probleme mit der Berechnung des uneigentlichen Integrals $\int_{-\infty}^{\infty} x \cdot f(x)\, dx$ bekommt. Zum Beispiel gibt es für eine $F$-Verteilung mit beliebigem $df_1 \geqslant 1$ und $df_2 = 2$ keinen Mittelwert.

∗ *Korrektur der Ränge für Bindungen* (p. 4-21): Stimmen mehr als zwei aufeinanderfolgende Messungen überein, ist die Mittelwertbildung entsprechend über mehr als zwei Ränge zu erstrecken.

∗ *Rechnen mit Rangsummen* (p. 4-21 f.):

$$R' + R'' \;=\; \sum_{k=1}^{N'+N''} k \;=\; \frac{(N' + N'') \cdot (N' + N'' + 1)}{2} \;=\; \frac{N' \cdot (N' + 1)}{2} + \frac{N'' \cdot (N'' + 1)}{2} + N' \cdot N''\,.$$

∗ *Auswertung des Mann-Whitney-U-Tests* (p. 4-23 f.): In der Literatur wird oft $U = \mathrm{Min}\,(U', U'')$ bzw. $\widehat{T} = \mathrm{Min}\,(\widehat{T}', \widehat{T}'')$ gebildet und die Auswertung anschließend auf $U$ oder $\widehat{T}$ bezogen. Unsere Formulierung der Auswertebedingungen ist dazu völlig äquivalent.

∗ *Kritische Werte beim Mann-Whitney-U-Test für* $N' + N'' \leqslant 30$ (p. 4-23):
Siehe z.B. `https://statologie.de/mann-whitney-u-tabelle/` (15.10.2026).

∗ *Korrektur von* $\sigma_U$ *für viele Bindungen* (p. 4-22): Liegen nicht nur einzelne Bindungen vor oder treten die Bindungen in Bändern auf, so muß bei der Berechnung von $\widehat{T}'$, $\widehat{T}''$ anstelle von $\sigma_U$ die korrigierte Größe $\sigma_{U,\mathsf{korr}}$ verwendet werden:

$$\sigma_{U,\mathsf{korr}} = \sqrt{\frac{N' \cdot N'' \cdot (N' + N'' + 1)}{12} - \frac{N' \cdot N'' \cdot S}{12 \cdot (N' + N'' - 1) \cdot (N' + N'')}} \quad \text{mit} \quad S = \sum_{m=1}^{M} (b_m^3 - b_m)\,,$$

darin sind $b_m$ die Anzahlen der verbundenen Ränge in den einzelnen Ranggruppen und $M$ die Gesamtzahl der Ranggruppen mit Bindungen.

∗ *Wilcoxon-Rangsummentest* (pp. 4-26 ff.): Siehe oben zur Bezeichnung des Tests.

∗ *Auswertung des Wilcoxon-Rangsummentests* (p. 4-29): In der Literatur wird oft $R = \mathrm{Min}\,(R^-, R^+)$ bzw. $\widehat{T} = \mathrm{Min}\,(\widehat{T}^-, \widehat{T}^+)$ gebildet und die Auswertung anschließend auf $R$ oder $\widehat{T}$ bezogen. Unsere Formulierung der Auswertebedingungen ist dazu völlig äquivalent.

∗ *Kritische Werte beim Wilcoxon-Rangsummentest für* $N^* \leqslant 30$ (p. 4-29):
Siehe z.B. `https://statologie.de/ wilcoxon-vorzeichen-rang-tabelle` (15.10.2026).

∗ *Korrektur von* $\sigma_R$ *für viele Bindungen* (p. 4-29): Liegen nicht nur einzelne Bindungen vor oder treten die Bindungen in Bändern auf, so muß bei der Berechnung von $\widehat{T}^-$, $\widehat{T}^+$ anstelle von $\sigma_R$ die korrigierte Größe $\sigma_{R,\mathsf{korr}}$ verwendet werden:

$$\sigma_{R,\mathsf{korr}} = \sqrt{\frac{N^* \cdot (N^* + 1) \cdot (2\,N^* + 1)}{24} - \frac{S}{48}} \quad \text{mit} \quad S = \sum_{m=1}^{M} (b_m^3 - b_m)\,,$$

darin sind $b_m$ die Anzahlen der verbundenen Ränge in den einzelnen Ranggruppen und $M$ die Gesamtzahl der Ranggruppen mit Bindungen.

∗ *Kruskal-Wallis-Test* (p. 4-32): Die Prüfgröße $\widehat{T}$ "fällt" hier sozusagen "vom Himmel". Wir verzichten auf ihre exakte Herleitung, die weitgehend analog zur Zerlegung der Gesamt-Varianz bei der einfaktoriellen ANOVA verläuft. Zu allen Größen, die wir bei der Zerlegung der Gesamt-Varianz auf p. 4-06 eingeführt haben, gibt es ein nichtparametrisches Gegenstück. In der Literatur wird die Prüfgröße meist nicht mit $\widehat{T}$, sondern mit $H$ bezeichnet.

∗ *Kritische Werte beim Kruskal-Wallis-Test für kleine Stichproben* (p. 4-32): Siehe z.B. `https://dataanalytics.org.uk/critical-values-for-the-kruskal-wallis-test/` (15.10.2026).

∗ *Korrektur der Prüfgröße für viele Bindungen* (p. 4-32): Liegen nicht nur einzelne Bindungen vor oder treten die Bindungen in Bändern auf, so muß beim Kruskal-Wallis-Test anstelle der Prüfgröße $\widehat{T}$ die korrigierte Prüfgröße $\widehat{T}_{korr}$ verwendet werden:

$$\widehat{T}_{korr} = \widehat{T} \,/\, \Big( 1 - \frac{S}{M^3 - M} \Big) \quad \text{mit} \quad S = \textstyle\sum_{m=1}^{M} (b_m^3 - b_m)\,,$$

darin sind $b_m$ die Anzahlen der verbundenen Ränge in den einzelnen Ranggruppen und $M$ die Gesamtzahl der Ranggruppen mit Bindungen.

### Quellenverzeichnis


[ Brown (2020) ] Brown, S.: *Comparing more than two means: One-way ANOVA.* Elektronisch veröffentlicht unter `brownmath.com/stat/anova1.htm` (03.11.2025)

[ DKG (2013) ] Deutsche Gesellschaft für Kardiologie — Herz- und Kreislaufforschung e.V. (Hrsg.): *Herzinsuffizienz. Leitlinien für die Diagnose und Behandlung der akuten und chronischen Herzinsuffizienz.* Düsseldorf 2013

[ Enright/Sherrill (1998) ] Enright, P. L.; Sherrill, D. L.: *Reference equations for the six-minute walk in healthy adults.* Am. J. Respir. Crit. Care Med. **158** (1998) : 1384 – 1387

[ Fisher (1925) ] Fisher, R. A.: *Statistical Methods for Research Workers.* Oliver and Boyd; Edinburgh - London 1925. Online verfügbar z.B. unter `psychclassics.yorku.ca/Fisher/Methods/index.htm` (15.10.2026)

[ Mann/Whitney (1947) ] Mann, H. B.; Whitney, D. R.: *On a test of whether one of two random variables is stochastically larger than the other.* Ann. Math. Statist. **18** (1947) : 50 – 60

[ Wilcoxon (1945) ] Wilcoxon, F.: *Individual comparisons by ranking methods.* Biometrics Bull. **1** (1945) : 80 – 83


### Bildnachweis

∗ p. 4-02: Grafische Darstellung entnommen aus [ DKG (2013) ], p. 7.

∗ p. 4-12: Foto: Helga Kattinger, veröffentlicht unter Pixabay-Lizenz (gebührenfreie, nichtkommerzielle Nutzung)