

18. Li Y, Hou X-J, Liu T-S, Xu S-J, Huang Z-H, Yan P-Y, Xu X-Y, Dong R (2021) Risk factors for acute kidney injury following coronary artery bypass graft surgery in a Chinese population and development of a prediction model. *J Geriatr Cardiol* JGC 18:711–719.
19. Rasmussen SR, Kandler K, Nielsen RV, Cornelius Jakobsen P, Knudsen NN, Ranucci M, Christian Nilsson J, Ravn HB (2019) Duration of critically low oxygen delivery is associated with acute kidney injury after cardiac surgery. *Acta Anaesthesiol Scand* 63:1290–1297.
20. Merchant AM, Neyra JA, Minhajuddin A, Wehrmann LE, Mills RA, Gualano SK, Kumbhani DJ, Huffman LC, Jessen ME, Fox AA (2019) Packed red blood cell transfusion associates with acute kidney injury after transcatheter aortic valve replacement. *BMC Anesthesiol* 19:99.
21. Kindzelski BA, Corcoran P, Siegenthaler MP, Horvath KA (2018) Postoperative acute kidney injury following intraoperative blood product transfusions during cardiac surgery. *Perfusion* 33:62–70.
22. Karowni W, Vora AN, Dai D, Wojtyla D, Dakik H, Rao SV (2016) Blood Transfusion and the Risk of Acute Kidney Injury Among Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention. *Circ Cardiovasc Interv* 9:e003279.
23. Leahy MF, Hofmann A, Towler S, et al (2017) Improved outcomes and reduced costs associated with a health-system-wide patient blood management program: a retrospective observational study in four major adult tertiary-care hospitals: PATIENT BLOOD MANAGEMENT OUTCOMES. *Transfusion (Paris)* 57:1347–1358.
24. Scolletta S, Simioni P, Campagnolo V, et al (2019) Patient blood management in cardiac surgery: The "Granducato algorithm." *Int J Cardiol* 289:37–42.
25. Hajjar LA, Vincent J-L, Galas FRBG, et al (2010) Transfusion Requirements After Cardiac Surgery: The TRACS Randomized Controlled Trial. *JAMA* 304:1559.
26. Reeves BC, Pike K, Rogers CA, et al (2016) A multicentre randomised controlled trial of Transfusion Indication Threshold Reduction on transfusion rates, morbidity and health-care resource use following cardiac surgery (TITRe2). *Health Technol Assess* 20:1–260.
27. Bracey AW, Radovancevic R, Riggs SA, Houston S, Cozart H, Vaughn WK, Radovancevic B, McAllister HA, Cooley DA (1999) Lowering the hemoglobin threshold for transfusion in coronary artery bypass procedures: effect on patient outcome. *Transfusion (Paris)* 39:1070–1077.
28. Patel NN, Murphy GJ (2017) Evidence-Based Red Blood Cell Transfusion Practices in Cardiac Surgery. *Transfus Med Rev* 31:230–235.
29. Magruder JT, Dungan SP, Grimm JC, Harness HL, Wierschke C, Castillejo S, Barodka V, Katz N, Shah AS, Whitman GJ (2015) Nadir Oxygen Delivery on Bypass and Hypotension Increase Acute Kidney Injury Risk After Cardiac Operations. *Ann Thorac Surg* 100:1697–1703.
30. Newland PF, Baker RA (2017) Low Oxygen Delivery as a Predictor of Acute Kidney Injury during Cardiopulmonary Bypass. *J Extra Corpor Technol* 49:224–230.
31. Ranucci M, Johnson I, Willcox T, et al (2018) Goal-directed perfusion to reduce acute kidney injury: A randomized trial. *J Thorac Cardiovasc Surg* 156:1918–1927.e2.
32. Mukaida H, Matsushita S, Yamamoto T, Minami Y, Sato G, Asai T, Amano A (2021) Oxygen delivery-guided perfusion for the prevention of acute kidney injury: A randomized controlled trial. *J Thorac Cardiovasc Surg* S0022522321005110.

Die Perfusiologie 2024(1): 15–20 | <https://doi.org/10.47624/dp.033.OPNF5043>

Die Perfusiologie stellt in der Rubrik **Tutorials** in Folge relevante Methoden für wissenschaftliche Arbeiten zur klinischen Perfusion und technischen Medizin vor.

Statistik Teil 9: Die 1-Weg ANOVA

Einführung

In den vorangegangenen beiden Tutorials wurde detailliert der Vergleich von Mittelwerten im 1- und 2-Stichprobenfall besprochen [1,2]. In diesem Tutorial wird nun die Situation von drei oder mehr unabhängigen oder abhängigen Gruppen behandelt. Man spricht in diesem Fall auch von einer 1-Weg ANOVA (ANalysis Of VAriance) oder 1-Weg–Varianzanalyse. Im Fall von abhängigen Gruppen verwendet man zur Unterscheidung zum Teil auch die Abkürzung rANOVA, was für eine ANOVA mit Messwiederholungen (repeated measures) steht. Im Folgenden werden verschiedene Varianten der 1-Weg ANOVA vorgestellt, darunter auch Varianten, die auf Permutationen oder Bootstrap basieren, und Kriterien angeben, die bei der Testauswahl helfen sollen. Alle Tests werden anhand von realen Daten genauer demonstriert.

k-Stichproben Tests

Im k-Stichprobenfall ($k \geq 3$) werden die Mittelwerte (arithmetisches Mittel) von k-Gruppen miteinander verglichen. Dies führt auf die folgende Nullhypothese:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

Die Alternative lautet entsprechend, dass mindestens zwei Gruppen unterschiedliche Mittelwerte besitzen und lässt sich etwas schwieriger in einer expliziten Form angeben. Die Gruppen können hierbei abhängig oder unabhängig sein. Im Fall von abhängigen Gruppen liegen üblicherweise Messwiederholungen von k verschiedenen Zeitpunkten vor.

Im Unterschied zum 2-Stichprobenfall ist bei einem signifikanten Ergebnis in der 1-Weg ANOVA nicht automatisch klar, welche Gruppen sich hinsichtlich ihrer Mittelwerte unterscheiden. Daher werden in diesem Fall üblicherweise sogenannte **Post-hoc-Tests** durchgeführt, bei denen die Gruppen paarweise

Prof. Dr. Matthias Kohl

Department of Medical and Life Sciences
Institute of Precision Medicine
Hochschule Furtwangen
Jakob-Kienzle-Str. 17,
78054 Villingen-Schwenningen (Germany)
Phone: +49 (0) 7720 307-4635 · E-Mail: kohl@hs-furtwangen.de
www.hs-furtwangen.de · www.life-data-science.org

M. Kohl, F. Münch



TUTORIAL

Fazitbox

Pro und Contra 1-Weg ANOVA:

Pro

- Kann man von normalverteilten Werten ausgehen und sind keine Ausreißer zu erwarten, so stellen die Welch 1-Weg ANOVA (unabhängige Gruppen) und die 1-Weg ANOVA mit gemischten Effekten und unstrukturierter Kovarianzmatrix (abhängige Gruppen) die besten Tests dar.
- Permutations- und Bootstrap-Varianten der 1-Weg ANOVA sollten verwendet werden, falls die genaue Verteilung der Werte unbekannt ist und keine Ausreißer zu erwarten sind. Hierbei sollten Verfahren gewählt werden, die auch bei Heteroskedastizität (unabhängige Gruppen) bzw. Nicht-Sphärität/Nicht-Zirkularität (abhängige Gruppen) zuverlässige Ergebnisse liefern.

Contra

- Rechnet man mit Ausreißern, können Mittelwerte und darauf basierende Tests deutlich verfälschte Ergebnisse liefern – dies gilt auch für deren Permutations- und Bootstrap-Varianten. In diesem Fall sollte man auf robuste Alternativen wie den Kruskal-Wallis H-Test (unabhängige Gruppen) bzw. den Quade oder den Friedman Test (abhängige Gruppen) ausweichen.

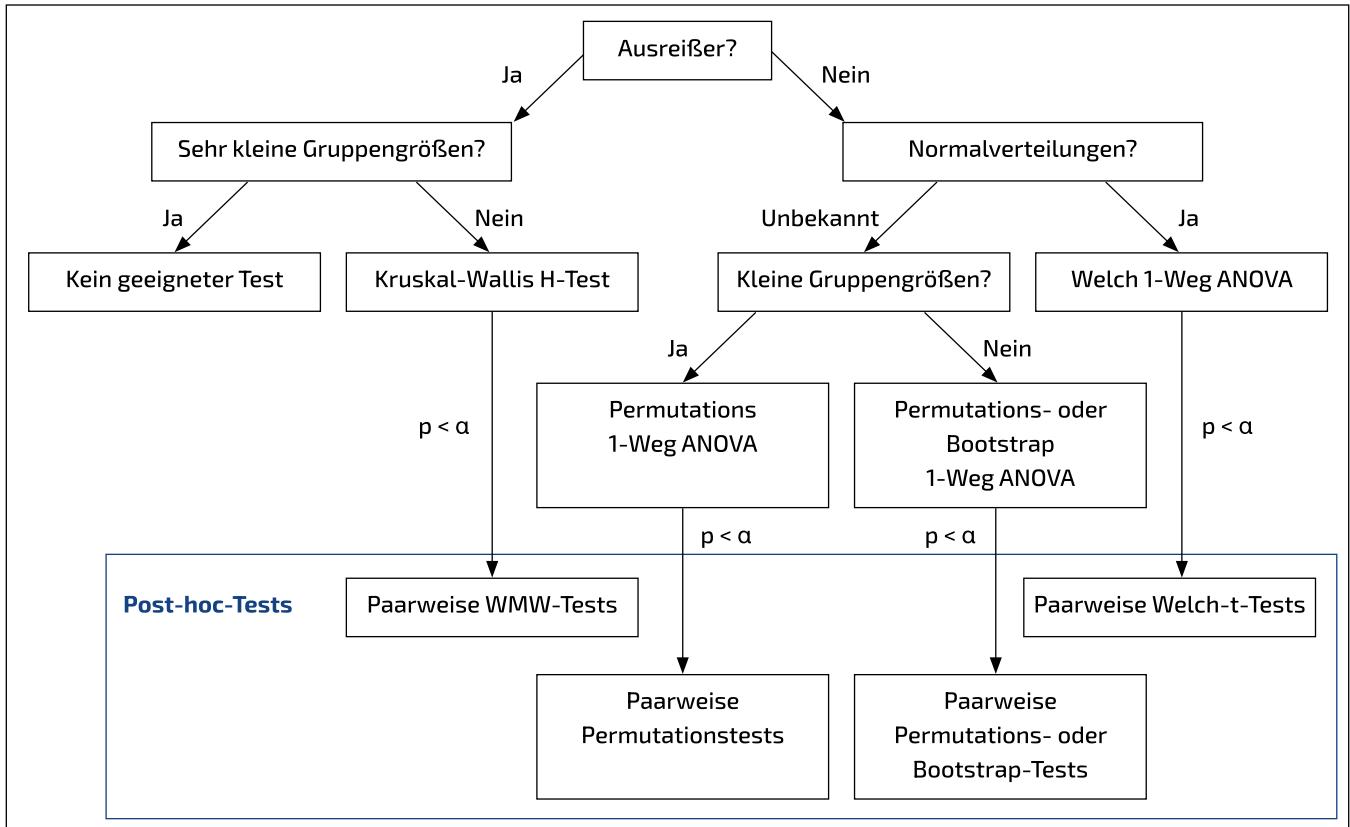


Abb. 1: Auswahl einer geeigneten 1-Weg ANOVA für unabhängige Gruppen unter der Annahme, dass die Varianzen verschieden sind (Heteroskedastizität)

miteinander verglichen werden. In der Regel werden dabei alle möglichen paarweisen Vergleiche durchgeführt. Damit bei diesen paarweisen Vergleichen, bei denen mehrere statistische Tests parallel durchgeführt werden, der Fehler 1. Art nicht unkontrolliert anwächst, werden die p-Werte aus den paarweisen Vergleichen üblicherweise einer Adjustierung unterzogen. Sehr gut geeignet ist hierfür insbesondere die **Methode von Holm** [3], welche die sogenannte **FWER** (family-wise error rate) kontrolliert. Bei der FWER handelt es sich um die Wahrscheinlichkeit, dass man mindestens einen Fehler 1. Art macht, also mindestens ein falsch positives Testergebnis erhält. Die adjustierten p-Werte können dann mit dem gewählten Signifikanzniveau α verglichen werden, wie dies bei einzelnen Tests üblich ist. Generell fällt die Adjustierung der p-Werte umso stärker aus, je mehr paarweise Tests durchgeführt werden. Daher ist es von Vorteil, wenn nicht alle paarweisen Vergleiche durchgeführt werden müssen. Dies kann zum Beispiel der Fall sein, wenn nur der Vergleich zu einer Referenzgruppe interessiert. Es ist bei der Anwendung von Post-hoc-Tests außerdem sehr wichtig, darauf zu achten, dass die angewendeten Post-hoc-Tests zur gewählten Variante der 1-Weg ANOVA passen. Andernfalls kann es zu inkonsistenten Ergebnissen kommen; d. h. trotz einer signifikanten 1-Weg ANOVA liefert keiner der Post-hoc-Tests ein signifikantes Ergebnis.

Unabhängige Gruppen

Folgen die Werte der unabhängigen Gruppen einer Normalverteilung mit identischer Varianz (Homoskedastizität), so führt dies auf die klassische **1-Weg ANOVA**. Sind die Varianzen verschieden (Heteroskedastizität), kann man stattdessen die **Welch 1-Weg ANOVA** verwenden [4]. Ist es unklar, ob

Normalverteilungen vorliegen, so können **Permutations- oder Bootstrap-Varianten** der 1-Weg ANOVA angewendet werden. Außerdem kann auch eine auf Rängen basierende Analyse (1-Weg ANOVA mit Rängen) durchgeführt werden, die als **Kruskal-Wallis H-Test** bekannt ist [5]. Insbesondere für Situationen mit ungleichen Gruppengrößen und/oder ungleichen Varianzen wurden eine Reihe unterschiedlicher Ansätze vorgeschlagen. Das Erweiterungspaket onewaytests [6] für die Statistiksoftware R [7] enthält zum Beispiel 22 verschiedene Methoden, um entsprechende Gruppenvergleiche durchzuführen. Im Wesentlichen handelt es sich dabei um Varianten der oben aufgeführten Möglichkeiten.

Für die Auswahl der am besten geeigneten 1-Weg ANOVA können Fragen herangezogen werden, die identisch zum 2-Stichprobenfall sind und bereits während der Planung einer Studie beantwortet werden sollten [2]. Die Auswahl der am besten geeigneten 1-Weg ANOVA erfolgt entsprechend analog zum 2-Stichprobenfall und ist in Abb. 1 grafisch dargestellt.

Zur Demonstration der Verfahren verwenden wir die glomeruläre Filtrationsrate (GFR) nach einer Operation mit Herz-Lungen-Maschine, wobei wir drei Gruppen unterscheiden. Eine Gruppe erhielt für eine bessere Verteilung des Blutes zum „venösen Pooling“ zusätzlich Nitroprussid, eine zweite Gruppe zusätzlich Nitroglycerin und eine Gruppe war ohne Intervention. Es liegen Daten von 614 Patient:innen vor (Nitroprussid: 219, Nitroglycerin: 176, ohne Intervention: 219). Wir gehen davon aus, dass die GFR-Werte für die drei Gruppen näherungsweise symmetrisch verteilt sind und ungleiche Varianzen aufweisen. Dies wird durch die Box- und Whisker-Plots in Abbildung 2 und die Tatsache bestätigt, dass Mittelwerte (MW) und Mediane (Med) der GFR ähnlich sind

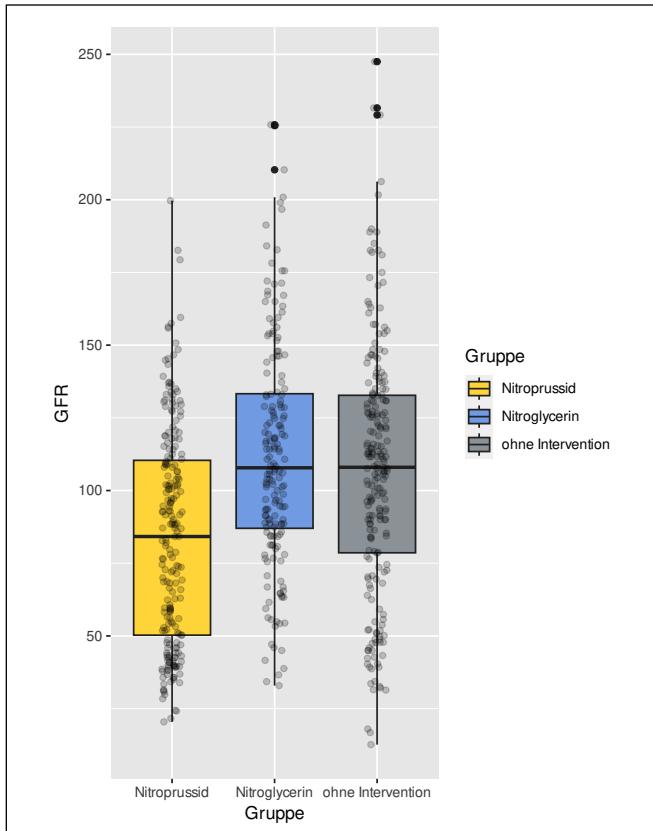


Abb. 2: Box-und-Whisker Plots der GFR-Werte für die drei Behandlungsgruppen (erstellt mit dem Paket ggplot2 [8] der Statistiksoftware R [7])

(Nitroprussid: MW = 84, Med = 84; Nitroglycerin: MW = 112, Med = 108; ohne Intervention: MW = 107, Med = 108), wobei die Standardabweichungen (SD) etwas voneinander abweichen (Nitroprussid: SD = 37; Nitroglycerin: SD = 39; ohne Intervention: SD = 43).

Bei Vorliegen einer symmetrischen Verteilung fallen arithmetisches Mittel, Median und Hodges-Lehmann-Schätzer (Median aller paarweisen Differenzen) zusammen. Im Folgenden verwenden wir daher bei Beschreibungen zur Vereinfachung jeweils nur den Begriff Mittelwert.

Wir berechnen neben der Welch 1-Weg ANOVA und dem Kruskal-Wallis H-Test eine 1-Weg ANOVA basierend auf Permutationen [9] und eine basierend auf Bootstrap [10]. Auch wenn keine Normalverteilung vorliegt, gehen wir aufgrund der recht großen Gruppen davon aus, dass auch die Welch 1-Weg ANOVA zumindest näherungsweise den Fehler 1. Art einhalten sollte. Für die Berechnungen verwenden wir die Statistiksoftware R [7] in Kombination mit dem Paket stats [7] für die Welch 1-Weg ANOVA und den Kruskal-Wallis H-Test. Für die Permutations-ANOVA verwenden wir das Paket GFD [11] und für die Bootstrap-ANOVA das Paket onewaytests [6]. Wir legen bei allen Tests ein Signifikanzniveau von 5 % zugrunde. Wir erhalten bei allen vier Tests ein signifikantes Ergebnis (Welch 1-Weg ANOVA: $p = 1,6 \cdot 10^{-13}$; Kruskal-Wallis H-Test: $p = 6,1 \cdot 10^{-12}$; Permutations-ANOVA: $p = 1,3 \cdot 10^{-12}$; Bootstrap-ANOVA: $p < 1 \cdot 10^{-5}$). Wir führen daher in einem nächsten Schritt paarweise Vergleiche zwischen den Gruppen mittels entsprechender Post-hoc-Tests durch. Im Fall der paarweisen Welch t-Tests und der WMW-Tests verwenden wir das R Paket MKinfer [12]. Im Fall der paarweisen Permutations- und Boot-

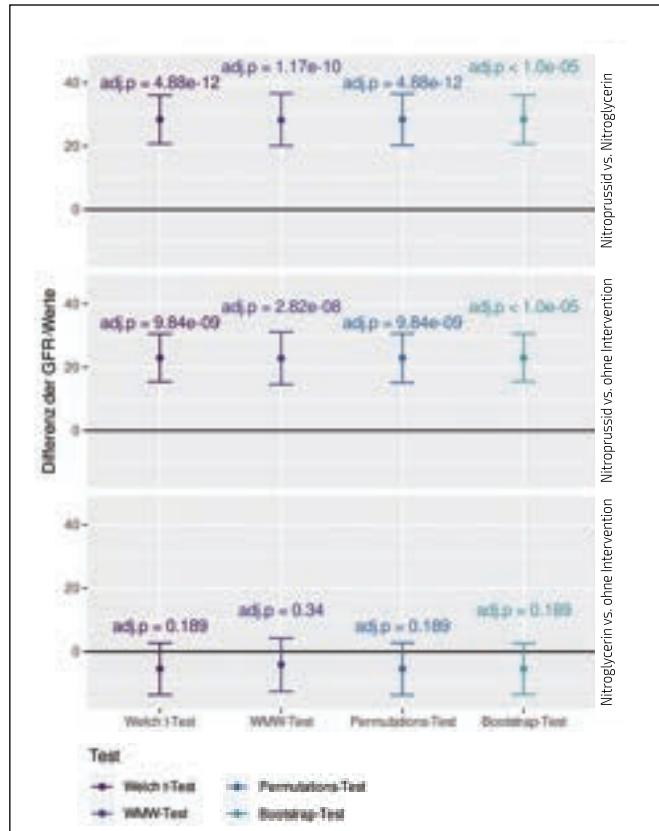


Abb. 3: adjustierte p-Werte (adj.p) und 95% Konfidenzintervalle (ohne Adjustierung) für die paarweisen Vergleiche der drei Behandlungsgruppen (erstellt mit dem Paket ggplot2 [8] der Statistiksoftware R [7]); die Schreibweise „Gruppe A vs. Gruppe B“ bedeutet hierbei, dass der Mittelwert von Gruppe A vom Mittelwert der Gruppe B subtrahiert wurde

strap-Tests erneut die Pakete GFD [11] und onewaytests [6]. Der Mittelwert der GFR-Werte ist in der Nitroprussid-Gruppe signifikant niedriger als in der Nitroglycerin-Gruppe und der Gruppe ohne Intervention. Der Mittelwert der GFR-Werte der Nitroglycerin-Gruppe unterscheidet sich nicht signifikant vom Mittelwert der Gruppe ohne Intervention. Die adjustierten p-Werte (adj.p) und die 95 % Konfidenzintervalle (ohne Adjustierung) sind in Abb. 3 dargestellt.

Wir erhalten für alle Tests recht ähnliche Ergebnisse. Dies dürfte daran liegen, dass die Gruppen relativ groß sind und keine wesentlichen Ausreißer vorliegen.

Abhängige Gruppen

Im Fall von abhängigen Gruppen führt die Annahme von Normalverteilungen auf die sogenannte **1-Weg ANOVA mit Messwiederholungen** (repeated measures) (rANOVA), wobei man zusätzlich voraussetzen muss, dass die Varianzen der Differenzen für alle möglichen Paare von Gruppen gleich sind. Diese Annahme wird auch **Sphärität** oder **Zirkularität** genannt. Heute wird aber bei Vorliegen von Normalverteilungen stattdessen oft eine **1-Weg ANOVA mit gemischten Effekten** (mixed effects) herangezogen. Man bezeichnet dies auch als ein **Mehrebenenmodell** (multi-level model). Hierbei werden die Messzeitpunkte als ein fester/deterministischer Effekt (fixed effect) betrachtet; d. h., es handelt sich um klar definierte Zeitpunkte, die man auch in einer neuen Studie wieder so wählen würde. Die Patient:innen werden als zufälliger Effekt (random effect) modelliert. Das bedeutet, dass es sich um eine Zufalls-

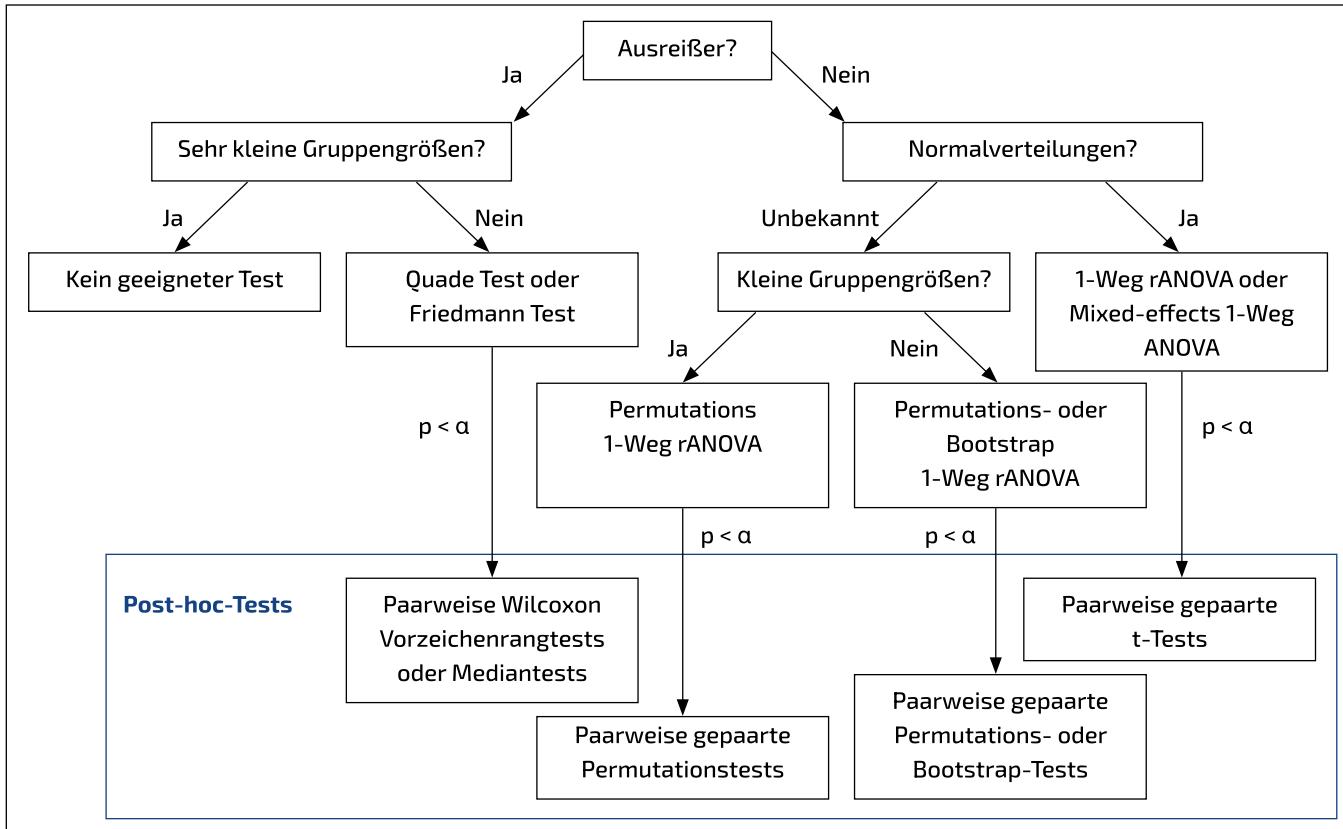


Abb. 4: Auswahl einer geeigneten 1-Weg ANOVA bei abhängigen Gruppen

stichprobe aus einer größeren Population handelt und man in einer neuen Studie andere Patient:innen zufällig wählen würde. Ist es unklar, ob Normalverteilungen vorliegen, so können **Permutations- oder Bootstrap-Varianten** der 1-Weg rANOVA angewendet werden. Außerdem kann auch eine auf Rängen basierende Analyse durchgeführt werden. Hierfür kommen der **Friedman Test** (Verallgemeinerung des Median Tests) [13] und der **Quade Test** (Verallgemeinerung des Wilcoxon Vorzeichenrangtests) [14] in Frage. Im Fall eines signifikanten Ergebnisses werden die Messzeitpunkte üblicherweise mit Hilfe geeigneter Post-hoc-Tests paarweise verglichen. Aufgrund der Abhängigkeit kommen hierfür gepaarte 2-Stichprobentests zum Einsatz. Die p-Werte der paarweisen Tests sollten zum Beispiel mit dem Verfahren von Holm [3] adjustiert werden.

Die Auswahl der am besten geeigneten 1-Weg ANOVA mit Messwiederholungen erfolgt ähnlich wie im Fall von unabhängigen Gruppen und ist in Abb. 4 dargestellt. Da bei einer Verletzung der Sphärizitäts-/Zirkularitäts-Annahme die Tests zu liberal werden (d. h., Fehler 1. Art ist größer als α), empfiehlt es sich, zusätzlich zum Beispiel die Huynh-Feldt Korrektur anzuwenden bzw. im Fall der ANOVA mit gemischten Effekten eine unstrukturierte Kovarianzmatrix zu verwenden [15]. Wir betrachten zur Demonstration wieder die glomeruläre Filtrationsrate (GFR) nach einer Operation mit Herz-Lungen-Maschine (HLM), wobei wir nur die Gruppe betrachten, die zusätzlich Nitroglycerin erhalten hat. Wir wollen den Verlauf der GFR-Werte vom Zeitpunkt vor der OP (prä OP) über den Zeitpunkt nach der extrakorporalen Zirkulation (post EKZ) bis hin zum ersten Tag nach der EKZ (24 h post EKZ) untersuchen. Wir schließen entsprechend nur Patient:innen ein, bei denen

Messwerte von allen drei Zeitpunkten vorliegen. Dies reduziert die Anzahl der Patient:innen von 176 auf 167. Wir gehen davon aus, dass die paarweisen Differenzen der GFR-Werte näherungsweise symmetrisch verteilt sind und ungleiche Varianzen aufweisen. Dies wird durch die Box- und Whisker-Plots in Abb. 5 und die Tatsache bestätigt, dass Mittelwerte (MW) und Mediane (Med) der paarweisen Differenzen ähnlich sind (prä OP vs. post EKZ: MW = -4,8, Med = -1,9; prä OP vs. 24 h post EKZ: MW = -14,9, Med = -12,6; post EKZ vs. 24 h post EKZ: MW = -10,1, Med = -9,2). Insbesondere können wir hieraus erkennen, dass die GFR-Werte zum Zeitpunkt 24 h post EKZ für die meisten Patient:innen am niedrigsten sind. Auch sind bei allen Vergleichen vereinzelte Ausreißer zu erkennen, wobei speziell bei den Ausreißern nach unten die Gefahr besteht, dass diese das Ergebnis hin zu einer verstärkten Abnahme verfälschen. Wir werden den Einfluss dieser Ausreißer auf das Ergebnis in der Analyse prüfen (Sensitivitätsanalyse).

Wir berechnen eine 1-Weg rANOVA mit Huynh-Feldt Korrektur, eine mixed effects 1-Weg ANOVA mit unstrukturierter Kovarianzmatrix sowie den Quade und den Friedman Test. Wir verwenden hierfür die R Pakete afex [16] und MKinfer [12]. Außerdem berechnen wir eine Permutations- und eine Bootstrap 1-Weg rANOVA mit den Paketen permuco [17] und boot [18]. Wir legen bei allen Tests ein Signifikanzniveau von 5 % zugrunde. In allen sechs Fällen ergibt sich ein signifikantes Ergebnis (1-Weg rANOVA: $p = 1,6 * 10^{-13}$; mixed effects 1-Weg ANOVA: $p = 3,0 * 10^{-14}$; Quade Test: $p = 6,9 * 10^{-15}$; Friedman Test: $p = 7,9 * 10^{-14}$; Permutations 1-Weg ANOVA: $p < 1,0 * 10^{-5}$; Bootstrap 1-Weg ANOVA: $p = 0,007$). Um den Einfluss der Ausreißer auf das Ergebnis zu prüfen, verwenden wir eine einfache Ausreißerverwerfungsregel. Wir schließen alle Patient:innen aus,

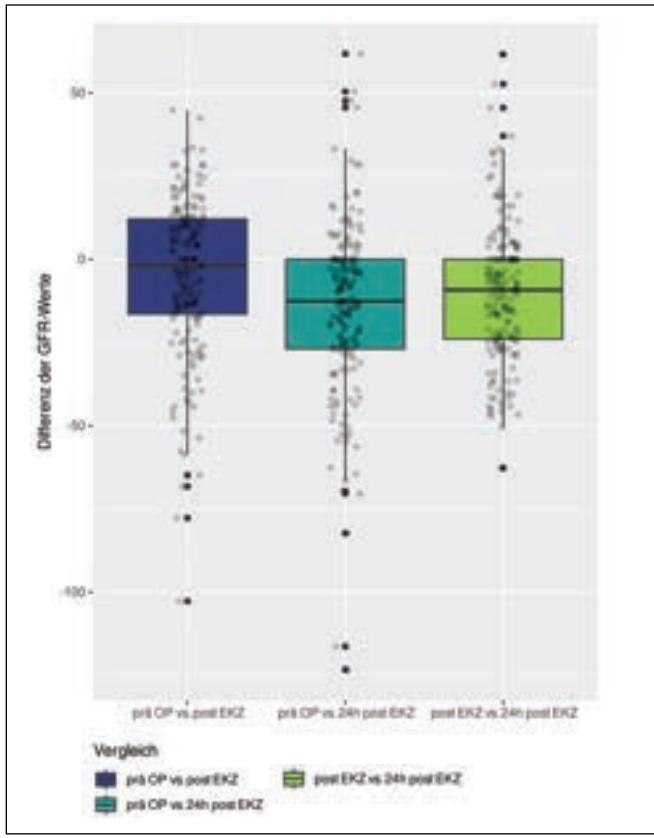


Abb. 5: Box-and-Whisker-Plots der paarweisen Differenzen der GFR-Werte für die Zeitpunkte prä OP, post EKZ und 24 h post EKZ (erstellt mit dem Paket ggplot2 [8] der Statistiksoftware R [6]); es handelt sich um die GFR-Werte der Nitroglycerin-Gruppe

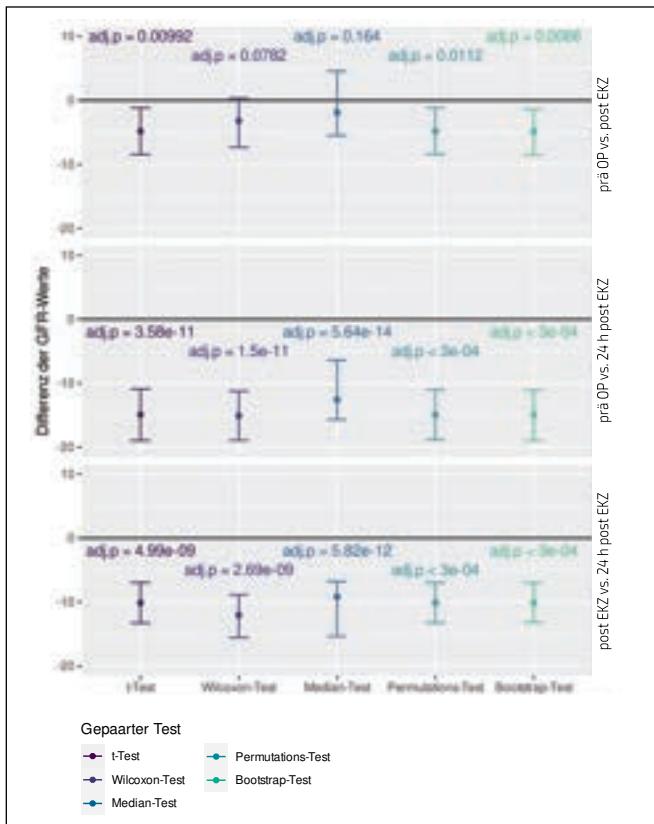


Abb. 6: adjustierte p-Werte (adj.p) und 95 % Konfidenzintervalle (ohne Adjustierung) für paarweise Vergleiche der drei Zeitpunkte (erstellt mit dem Paket ggplot2 [8] der Statistiksoftware R [7]); die Schreibweise „Zeitpunkt A vs. Zeitpunkt B“ bedeutet hierbei, dass die Werte von Zeitpunkt A von den Werten von Zeitpunkt B subtrahiert wurden und anschließend der Mittelwert berechnet wurde

die bei mindestens einem paarweisen Vergleich der Zeitpunkte eine Differenz aufweisen, die um mehr als das Dreifache des MAD (Median der absoluten Abweichungen vom Median) vom Median abweicht und wiederholen die Analyse. Dies führt zum Ausschluss von 9 Patient:innen und wir erhalten erneut bei allen Analysen ein signifikantes Ergebnis (1-Weg rANOVA: $p < 2,2 \cdot 10^{-16}$; mixed effects 1-Weg ANOVA: $p < 2,2 \cdot 10^{-16}$; Quade Test: $p = 2,5 \cdot 10^{-16}$; Friedman Test: $p = 3,5 \cdot 10^{-14}$; Permutations 1-Weg ANOVA: $p < 1,0 \cdot 10^{-5}$; Bootstrap 1-Weg ANOVA: $p < 1,0 \cdot 10^{-5}$). Wir führen folglich in einem nächsten Schritt paarweise Vergleiche zwischen den Zeitpunkten mittels entsprechender Post-hoc-Tests durch. Wir verwenden für die Berechnungen der gepaarten Tests die R Pakete stats [7] und MKinfer [12]. Der gepaarte t-Test ist hierbei sowohl für die 1-Weg rANOVA als auch für die mixed effects 1-Weg ANOVA ein geeigneter Post-hoc-Test. Bei allen Tests ergibt sich, dass der Mittelwert der GFR-Werte am Zeitpunkt 24 h post EKZ signifikant kleiner ist als an den Zeitpunkten prä OP und post EKZ. Der Vergleich zwischen den Zeitpunkten prä OP und post EKZ liefert kein so eindeutiges Ergebnis. Die adjustierten p-Werte (adj.p) und die 95 % Konfidenzintervalle (ohne Adjustierung) sind in Abb. 6 dargestellt.

Durch die etwas linksschiefe Verteilung und die Ausreißer nach unten (vgl. Abb. 5) dürften im Fall des Vergleiches der Zeitpunkte prä OP und post EKZ der Wilcoxon Vorzeichenrangtest und der gepaarte Median Test die verlässlichsten Ergebnisse darstellen. Der Unterschied zwischen den beiden

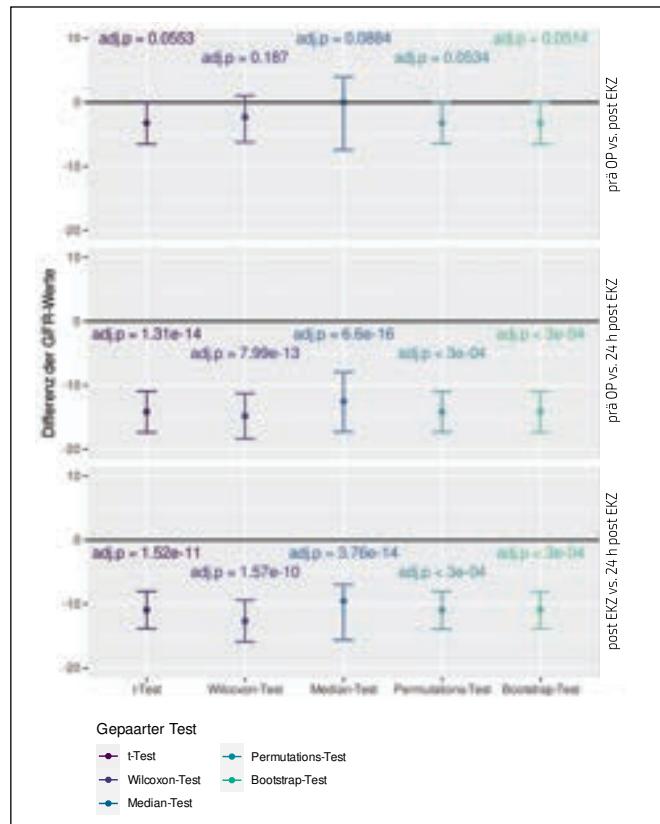


Abb. 7: adjustierte p-Werte (adj.p) und 95 % Konfidenzintervalle (ohne Adjustierung) für paarweise Vergleiche der drei Zeitpunkte nach der Entfernung von Ausreißern (erstellt mit dem Paket ggplot2 [8] der Statistiksoftware R [7]); die Schreibweise „Zeitpunkt A vs. Zeitpunkt B“ bedeutet hierbei, dass die Werte von Zeitpunkt A von den Werten von Zeitpunkt B subtrahiert wurden und anschließend der Mittelwert berechnet wurde

Zeitpunkten ist demnach zum Signifikanzniveau von 5 % nicht signifikant. Die anderen paarweisen Tests sind sensitiv gegenüber Ausreißern und schiefen Verteilungen und könnten daher ein verzerrtes und fälschlicherweise positives Ergebnis liefern. Wir wiederholen die Analyse ohne die als Ausreißer identifizierten Patient:innen. Die adjustierten p-Werte (adj.p) und die 95 % Konfidenzintervalle (ohne Adjustierung) sind in Abb. 7 dargestellt.

Die Entfernung der Ausreißer ändert die Signifikanz der Vergleiche zwischen prä OP und 24 h post EKZ sowie post EKZ und 24 h post EKZ nicht. Lediglich die 95 % Konfidenzintervalle verschieben sich etwas, was bei jeder Änderung der Daten zu erwarten ist. Beim Vergleich prä OP und post EKZ erhalten wir jetzt jedoch bei keinem Test ein signifikantes Ergebnis. Das Beispiel zeigt einmal mehr, wie wichtig es ist, auf Ausreißer zu achten und im Zweifelsfall besser robuste statistische Verfahren zu verwenden. Davon abgesehen empfiehlt es sich immer, eine Sensitivitätsanalyse durchzuführen und damit den Einfluss von auffälligen Werten auf das Ergebnis einer statistischen Analyse zu überprüfen.

Zusammenfassung

Da die Annahme einer Normalverteilung eine recht starke Voraussetzung darstellt, sollten in der Praxis verstärkt Permutations- und Bootstrap-Varianten der 1-Weg ANOVA verwendet werden. Außerdem führt die Verletzung der Annahme gleicher Varianzen (unabhängige Gruppen) bzw. von Sphärizität/Zirkularität (abhängige Gruppen) in der Regel zu liberalen Tests (d. h., Fehler 1. Art ist größer als α). Daher empfiehlt es sich, Varianten der 1-Weg ANOVA zu verwenden, die auch bei ungleichen Varianzen bzw. bei Nicht-Sphärizität/Nicht-Zirkularität zuverlässige Ergebnisse liefern. Beim Vorliegen von Aus-

reißern sollte auf robuste Tests wie den Kruskal-Wallis H-Test bzw. den Quade oder Friedman Test ausgewichen werden. Bei den Post-hoc-Tests, die üblicherweise durchgeführt werden, wenn die 1-Weg ANOVA ein signifikantes Ergebnis liefert, muss darauf geachtet werden, dass statistische Tests verwendet werden, die zur durchgeführten Variante der 1-Weg ANOVA passen. Andernfalls kann es zu inkonsistenten Ergebnissen kommen; d. h., trotz einer signifikanten 1-Weg ANOVA liefern die paarweisen Post-hoc-Tests kein signifikantes Ergebnis.

Literatur

1. Kohl M, Münch F. Statistik Teil 7: Statistische Signifikanztests. Kardiotechnik, 2023(3): 93-98.
2. Kohl M, Münch F. Statistik Teil 8: t-Tests und Alternativen. Kardiotechnik, 2023(4): 143-147.
3. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979, 6, 65-70.
4. Welch BL. On the comparison of several mean values: an alternative approach. Biometrika 1951, 38, 330-336.
5. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 1952, 47(160), 583-621.
6. Dag O, Kasikci M, Dolgun A, Konar NM, Weerahandi S, Ananda M, Sulekli HE. Onewaytests: one-way tests in independent groups designs. 2023. R package version 3.0.
7. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2023, Vienna, Austria. URL <https://www.R-project.org/>.
8. Wickham H. ggplot2: Elegant graphics for data analysis. 2016, Springer-Verlag New York.
9. Pauly M, Brunner E, Konietzschke F. Asymptotic permutation tests in general factorial designs. Journal of the Royal Statistical Society – Series B 2015, 77: 461-473.
10. Krishnamoorthy K, Lu F, Mathew T. A parametric bootstrap approach for ANOVA with unequal variances: fixed and random models. Computational Statistics and Data Analysis 2007, 51(12): 5731-5742.
11. Friedrich S, Konietzschke F, Pauly M. GFD: An R package for the analysis of general factorial designs. Journal of Statistical Software, Code Snippets, 2017, 79(1), 1-18.
12. Kohl M. Mkinfer: Inferential Statistics. 2023, R package version 1.2.
13. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association 1937, 32(200): 675-701.
14. Quade D. Using weighted rankings in the analysis of complete blocks with additive block effects. Journal of the American Statistical Association 1979, 74(367): 680-682.
15. Häverkamp N, Beauducel A. Violation of the sphericity assumption and its effect on type-I error rates in repeated measures ANOVA and multi-level linear models (MLM). 2017, Front. Psychol. 8:1841.
16. Singmann H, Bolker B, Westfall J, Aust F, Ben-Shachar M. Afex: Analysis of factorial experiments 2023. R package version 1.3-0.
17. Frossard J, Renaud O. Permutation tests for regression, ANOVA, and comparison of signals: The permuco package. Journal of Statistical Software 2021, 99(15): 1-32.
18. Canty A, Ripley B. Boot: Bootstrap R (S-Plus) functions. 2022. R package version 1.3-28.1.