

TU DORTMUND

FALLSTUDIEN I - SOMMERSEMESTER 2025

Projekt 3: Vergleich von mehreren Verteilungen am Beispiel eines Experiments zum Ankereffekt

Dozenten:

JProf. Dr. Dennis Dobler

Loreen Sabel

Dr. Susanne Frick

Prof. Dr. Philipp Doeblen

Verfasser: Quang Vinh Nguyen

Gruppenmitglieder:

Alain Kauth, Ibrahim Rahman

2. Juni 2025

Inhaltsverzeichnis

1. Einleitung	1
2. Problemstellung	1
2.1. Datenmaterial	2
2.2. Ziele des Projekts	3
3. Statistische Methoden	4
3.1. Zweifaktorielle ANOVA	4
3.2. Levene-Test	7
3.3. Bonferroni-Korrektur	7
4. Statistische Auswertung	8
4.1. Deskriptive Statistik	8
4.2. Hypothesentests	9
5. Zusammenfassung	10
Literatur	12
A. Zusätzliche Grafiken	14

1. Einleitung

Bei der Urteilsbildung zeigen Menschen eine Tendenz, sich auf bereits bekannte Informationen zu verlassen. Dieses psychologische Phänomen wurde erstmals von Tversky und Kahneman (1974) entdeckt und ist allgemein als Ankereffekt bekannt. Janiszewski und Uy (2008) führten ein Experiment zum Ankereffekt durch, in dem Studierende aufgefordert werden, für jedes Szenario eine Zahl anzugeben, die kleiner als der vorgegebene Ankerwert ist. Dabei werden je nach Fragebogen des Experiments zwei Faktoren in den Szenarien variiert, nämlich die Präzision des Ankerwerts und die Unterschätzungsmotivation. Basierend auf diesem Experiment berichteten Janiszewski und Uy (2008) in der Fachzeitschrift *Psychological Science*, dass Schätzungen näher am Anker liegen, wenn der Anker präziser angegeben wird.

In Rahmen des vorliegenden Projekts erfolgt eine Auswertung des Datensatzes von Chandler (2015), der durch Replikation des obigen Experiments gewonnen wurde, mit Hinblick auf die Verifizierung der Befunde aus dem Forschungsartikel von Janiszewski und Uy (2008). Die beiden Faktoren, Präzision der Anker und Motivation, sowie deren Interaktion werden im zweifachen Varianzanalyse-Modell (ANOVA) zusammen betrachtet. Die Testergebnisse zeigen tatsächlich, dass Zahlenangaben weniger angepasst werden, wenn die Präzision der Anker höher ist.

Dieser Bericht besteht aus fünf Kapiteln. Kapitel 2 beschreibt die Versuchsplanung sowie die Struktur des Datensatzes. Kapitel 3 stellt die zweifaktorielle ANOVA und das Aligned-Rank-Transform-Verfahren (ART) sowie der Test auf Varianzhomogenität nach Levene. Außerdem wird auch in diesem Kapitel auf das Inflationsproblem des Fehlers 1. Art beim multiplen Testen und das Bonferroni-Korrekturverfahren eingegangen. Kapitel 4 untersucht den Einfluss von Präzision der Anker auf die Anpassung der Schätzungen unter Berücksichtigung der Motivationsbedingungen. Die Ergebnisse werden zuletzt in Kapitel 5 zusammengefasst und diskutiert.

2. Problemstellung

Dieses Kapitel beginnt mit der detaillierten Beschreibung der Versuchsplanung sowie der Erläuterung der Datenstruktur. Im Folgenden werden die Projektziele und die Vorgehensweise ausführlich dargelegt.

2.1. Datenmaterial

Chandler (2015) replizierte das Experiment von Janiszewski und Uy (2008) mit nur geringfügigen Abweichungen in Details und verwendete dazu die gleichen Fragebögen im ursprünglichen Experiment. Diese Fragebögen umfassen jeweils neun Szenarien. Jedes Szenario wird von einem Ankerwert begleitet, der als Referenz dient. Die Teilnehmenden müssen dann eine Schätzung unterhalb des genannten Ankerwerts anpassen, um eine Zahlenangabe zu tätigen. Die einleitenden Erklärungen zu den Szenarien sind in den unterschiedlichen Fragebögen identisch. Allerdings werden zwei Faktoren in den Formulierungen modifiziert. Der eine Faktor ist die Präzision der Anker, die die Darstellungsform der Ankerwerte als entweder runde Zahl oder Dezimalzahl determiniert. Der andere Faktor ist die Motivation, also Größenordnung der erfragten Unterschätzung, die spezifiziert, ob die Teilnehmenden ihre Schätzungen stark oder gering anpassen sollen.

Der Datensatz von Chandler (2015) enthält Informationen der 126 Studierenden, die am Experiment teilnahmen, sowie deren Schätzungen für die neun Szenarien. Die erste Variable im Datensatz dient der Anweisung, ob eine Beobachtung von der Analyse ausgeschlossen wurde. Zu den weiteren unabhängigen Variablen gehören die Teilnehmernummer, das Alter, das Geschlecht und die Studienphase dieser Studierenden sowie die bearbeitete Fragebogenvariante und ihr Verhalten im Experiment. Bei der Fragebogenvariante, die Kombination aus Anker- und Motivationstyp, handelt es sich um eine nominalskalierte Variable mit vier Ausprägungen¹. Da sechs Studierende aus der Analyse ausgeschlossen wurden, gibt es für jede Fragebogenvariante 30 Beobachtungen. Die abhängigen Variablen sind die Schätzungen der Studierenden für die neun Szenarien und weisen alle eine metrische Skala auf. Eine Übersicht über diese Variablen zusammen mit den zugehörigen Szenarien und Ankerwerten ist in Tabelle 1 zu finden.

Der Datensatz enthält neun Einträge mit fehlenden Daten bei mindestens einem der Szenarien sowie 21 Einträge, bei denen höhere Werte als die Anker trotz gegenteiliger Anweisung angegeben werden. Der Umgang mit diesen Einträge wird in Kapitel 4 im Detail erläutert. Abgesehen davon zeigen sich keine weiteren Auffälligkeiten. In Bezug auf die Versuchsplanung wird das Experiment ausgewogen konzipiert, mit 30 Beobachtungen pro Fragebogenvariante, was laut van Ginkel und Kroonenberg (2021, S. 39–40) eine eindeutige Interpretation und eine aussagekräftigere ANOVA ermöglicht als ein unausgewogenes Versuchsdesign.

¹1. Variante = präziser Anker und größere Motivation, 2. Variante = präziser Anker und kleinere Motivation, 3. Variante = runder Anker und größere Motivation bzw. 4. Variante = runder Anker und kleinere Motivation

Tabelle 1: Übersicht über abhängige Variablen im Datensatz

Variable	Szenario	Präziser Anker	Runder Anker
pen	Schreiblänge eines Stifts in km	3.998 km	4.000 km
Proteindrink	Proteingehalt eines Getränks in Gramm	9.8 g	10.0 g
lebron	Trefferquote von LeBron James	0.498	0.500
slidy	Preis des Spielzeugs Slidy in Dollar	\$39.75	\$40.00
Cheese	Preis von QuickMelt Cheese in Dollar	\$4.85	\$5.00
Figurine	Preis einer Figur auf dem Flohmarkt in Dollar	\$49	\$50
TV	Preis eines 50 Zoll Plasma-Fernsehers in Dollar	\$4,998	\$5,000
beachhouse	Preis eines Strandhauses in Dollar	\$799,800	\$800,000
number	Wert einer Zahl in einem Computer	9,989	10,000

Die Ankerwerte der Szenarien sind im Datensatz nicht verfügbar und werden aus den Fragebögen entnommen.

2.2. Ziele des Projekts

Das vorliegende Projekt hat zum Ziel, zentrale Befunde der Studie von Janiszewski und Uy (2008) zu verifizieren. Im Mittelpunkt steht die Frage, ob präzisere Anker dazu führen, dass Versuchspersonen näher am Ankerwert schätzen als bei unpräzisen Ankern. Darüber hinaus wird geprüft, ob sich dieser Effekt durch die Motivation beeinflussen lässt. Der Effekt der Motivation allein wird auch mit untersucht, ist jedoch von untergeordnetem Interesse.

Um die Vergleichbarkeit zu gewährleisten, werden die neun Schätzwerte zu einer neuen Variable aggregiert. Dieser ergibt sich als Mittelwert der relativen Differenzen (MRD) zwischen den individuellen Schätzungen und den jeweiligen Ankerwerten. Zur deskriptiven Analyse wird der MRD anhand statistischer Kennzahlen sowie grafisch mithilfe von Boxplots ausgewertet, um erste Einsichten in Lage, Streuung und potenzielle Ausreißer der Verteilung zu gewinnen. Zur Überprüfung der Effekte kommt ein mehrstufiges Testverfahren zum Einsatz. Zunächst werden die wichtigen Annahmen geprüft: Die Normalverteilung des MRD bei jeder Gruppe wird mithilfe von Q-Q-Plots beurteilt, die Homogenität der Varianzen mit dem Levene-Test getestet. Sind die Annahmen nicht vollständig erfüllt, wird das ART-Verfahren auf die Daten angewendet. Anschließend wird eine zweifaktorielle ANOVA mit den unabhängigen Variablen Präzision des Ankers und Motivation durchgeführt. Da im Rahmen der zweifaktoriellen ANOVA drei separate Hypothesen geprüft werden, wird zur Kontrolle der Inflation des Fehlers erster Art eine Bonferroni-Korrektur herangezogen. Abschließend wird die praktische Relevanz der Befunde anhand der Effektstärke beurteilt.

3. Statistische Methoden

Die Datenverarbeitung und Anwendung der hier aufgeführten Methoden erfolgte mittels der Software *R* mit Version 4.4.2 (R Core Team, 2024) und zusätzlichen Paketen für fortgeschrittene Methoden, wie z. B. *ARTool* (Kay et al., 2025) für das ART-Verfahren und *car* (Fox und Weisberg, 2019) für den Levene-Test. Zum Erstellen des Diagramms wurde das Paket *ggplot2* (Wickham et al., 2016) verwendet.

Der Kürze und Konsistenz halber werden in der Beschreibung aller Methoden folgenden Notationsregeln verwendet, ohne Erläuterungen explizit zu wiederholen. Zufallsvariablen (inkl. Teststatistik) werden in Großbuchstaben und Realisierungen in Kleinbuchstaben dargestellt. Die F-Verteilung einer Variable in Form eines Varianzverhältnisses wird als $F_{(df_1, df_2)}$ bezeichnet, wobei df_1 und df_2 die Freiheitsgrade von Zähler und Nenner sind.

3.1. Zweifaktorielle ANOVA

Die zweifaktorielle Varianzanalyse (engl. Analysis of Variance, ANOVA) ermöglicht die gleichzeitige Untersuchung zweier Einflussfaktoren sowie ihrer Interaktion auf eine abhängige Variable. Eine grundlegende Voraussetzung für die Anwendung dieses Verfahrens ist, dass die abhängige Variable kontinuierlich und numerisch skaliert vorliegt, während die beiden Einflussfaktoren kategorial sind und jeweils mindestens zwei Ausprägungen aufweisen.

Ein Versuchsplan mit J Ausprägungen des Faktors A und K Ausprägungen des Faktors B wird als $J \times K$ -Design bezeichnet. Jede Kombination der Ausprägungen $A_j, j = 1, \dots, J$ und $B_k, k = 1, \dots, K$ bildet eine Gruppe, die als Zelle des Versuchsplans bezeichnet wird. Die zweifaktorielle Varianzanalyse geht von drei Annahmen aus: Unabhängigkeit der Stichprobenvariablen, Normalverteilung der Variablen in jeder Gruppe und Homogenität der Varianz über die Gruppen hinweg (Holling und Gediga, 2016, S.256). Da die Aussagekraft von ANOVA-Methoden bei einem ausgewogenen Versuchsdesign am höchsten ist, wird im Folgenden zusätzlich angenommen, dass die Stichprobengrößen in allen Zellen gleich groß sind ($= I$).

Im Rahmen der folgenden Erläuterung der Hypothesentests, der Effektstärke sowie des ART-Verfahrens werden die nachstehenden Notationen verwendet:

- $\bar{Y}_{...} = \frac{1}{I \cdot J \cdot K} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K Y_{ijk}$ bezeichnet den Gesamtmittelwert aller Variablen.
- $\bar{Y}_{.j.} = \frac{1}{I \cdot K} \sum_{i=1}^I \sum_{k=1}^K Y_{ijk}$ ist der Randmittelwert für die Ausprägung A_j des Faktors A , über alle Ausprägungen des Faktors B .

- $\bar{Y}_{..k} = \frac{1}{I \cdot J} \sum_{i=1}^I \sum_{j=1}^J Y_{ijk}$ ist der Randmittelwert für die Ausprägung B_k des Faktors B , über alle Ausprägungen des Faktors A .
- $\bar{Y}_{.jk} = \frac{1}{I} \sum_{i=1}^I Y_{ijk}$ stellt den Zellmittelwert für die Kombination (A_j, B_k) .

Bei der zweifaktoriellen ANOVA lässt sich die Gesamtstreuung (SS_T) in eine Summe aus der Streuung bedingt durch Faktor A (SS_A), der Streuung bedingt durch Faktor B (SS_B), der Streuung bedingt durch die Interaktion zwischen den beiden Faktoren (SS_{AB}) und der Reststreuung aufgrund von Fehlern innerhalb jeder Gruppe (SS_W) zerlegen (Holling und Gediga, 2016, S.261):

$$SS_T = SS_A + SS_B + SS_{AB} + SS_W,$$

Wobei

- $SS_T = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{...})^2$ die Quadratsumme der Abweichung vom Gesamtmittelwert ist,
- $SS_A = I \cdot K \sum_{j=1}^J (\bar{Y}_{.j.} - \bar{Y}_{...})^2$ die Quadratsumme für den Effekt des Faktors A ist,
- $SS_B = I \cdot J \sum_{k=1}^K (\bar{Y}_{..k} - \bar{Y}_{...})^2$ die Quadratsumme für den Effekt des Faktors B ist,
- $SS_{AB} = I \sum_{j=1}^J \sum_{k=1}^K (\bar{Y}_{.jk} - \bar{Y}_{.j.} - \bar{Y}_{..k} + \bar{Y}_{...})^2$ die Quadratsumme für den Effekt der Interaktion ist und
- $SS_W = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (Y_{ijk} - \bar{Y}_{.jk})^2$ Quadratsumme der Fehler ist.

Hypothesentests: Um den Einfluss der beiden Faktoren A und B sowie ihrer Interaktion auf eine abhängige Variable zu untersuchen, werden drei Hypothesentests durchgeführt. Die Nullhypothese H_0^A besagt, dass die Mittelwerte der nach den Ausprägungen von Faktor A gruppierten Variablen gleich sind, während die Alternativhypothese H_1^A annimmt, dass mindestens zwei Mittelwerte unterschiedlich sind. Die Teststatistik wird als $F_A := \frac{SS_A/(J-1)}{SS_W/(I \cdot J \cdot K - J \cdot K)} \stackrel{H_0^A}{\sim} \mathcal{F}_{(J-1, I \cdot J \cdot K - J \cdot K)}$ formuliert. Analog wird für Faktor B geprüft. Hier werden die Nullhypothese H_0^B so formuliert, dass die Mittelwerte der Variablen bei allen Ausprägungen des Faktors B gleich sind, und die Alternativhypothese H_1^B , dass mindestens zwei Mittelwerte unterschiedlich sind. Die Teststatistik lautet $F_B := \frac{SS_B/(K-1)}{SS_W/(I \cdot J \cdot K - J \cdot K)} \stackrel{H_0^B}{\sim} \mathcal{F}_{(K-1, I \cdot J \cdot K - J \cdot K)}$. Schließlich wird die Interaktion zwischen Faktor A und Faktor B untersucht. Die Nullhypothese H_0^{AB} nimmt an, dass keine Wechselwirkung besteht, die Alternativhypothese H_1^{AB} eine signifikante Interaktion. Die Teststatistik ist $F_{AB} := \frac{SS_{AB}/((J-1)(K-1))}{SS_W/(I \cdot J \cdot K - J \cdot K)} \stackrel{H_0^{AB}}{\sim} \mathcal{F}_{((J-1)(K-1), I \cdot J \cdot K - J \cdot K)}$.

Hier erfolgt die Formulierung der obigen Teststatistiken gemäß den Formeln von Fahrmeier et al. (2016) (Siehe S.546). Demnach sind die Teststatistiken zur Prüfung der drei Testprobleme jeweils so konzipiert, dass die Streuung bedingt durch den Effekt (des Faktors A, des Faktors B bzw. der Interaktion) mit der Streuung bedingt durch die Fehler verglichen werden kann. Die entsprechende Quadratesumme wird dabei durch einen Wert dividiert, der als zugehöriger Freiheitsgrad bezeichnet wird. Ein hoher Wert der Teststatistik weist darauf hin, dass die Streuung eher auf systematische Effekte als auf Zufall zurückzuführen ist. Da unter der jeweiligen Nullhypothese die Teststatistiken einer F-Verteilung folgen, werden die Nullhypothesen zum Niveau α abgelehnt, falls jeweils gilt

$$(H_0^A) \Pr(F_A > f_A) < \alpha; (H_0^B) \Pr(F_B > f_B) < \alpha \text{ bzw. } (H_0^{AB}) \Pr(F_{AB} > f_{AB}) < \alpha.$$

Aligned-Rank-Transform-Verfahren: Sofern die Annahmen der Normalverteilung oder der Varianzhomogenität in einer zweifaktoriellen ANOVA nicht erfüllt sind, besteht die Möglichkeit, die Daten vor der Durchführung von Hypothesentests durch das Aligned-Rank-Transform-Verfahren (ART) zu transformieren. In Anlehnung an Higgins et al. (1990, S.187) werden die Daten zunächst je nach Testproblem wie folgt ausgerichtet.

$$\begin{aligned} \text{Zum Testen des Effekts von Faktor } A: \quad Y_{ijk}^{(A)} &= Y_{ijk} - \bar{Y}_{.jk} + \bar{Y}_{.j.} - \bar{Y}_{...}, \\ \text{zum Testen des Effekts von Faktor } B: \quad Y_{ijk}^{(B)} &= Y_{ijk} - \bar{Y}_{.jk} + \bar{Y}_{..k} - \bar{Y}_{...}, \\ \text{und zum Testen des Interaktionseffekts: } Y_{ijk}^{(AB)} &= Y_{ijk} - \bar{Y}_{.j.} - \bar{Y}_{..k} - \bar{Y}_{...}. \end{aligned}$$

Nach der Ausrichtung werden die Daten in eine Rangfolge gebracht und zur Berechnung der Statistiken für das entsprechende Problem herangezogen. Die Anwendung des ART-Verfahrens scheint eine robuste zweifaktorielle ANOVA gegenüber der zugrunde liegenden Fehlerverteilung zu ermöglichen (Higgins et al., 1990, S.190). Higgins et al. (1990) hat anhand eines Beispiels (Siehe S.194) dargelegt, dass selbst im Falle einer schiefen Verteilung, wie der Exponentialverteilung, die Fehlerrate erster Art zum Niveau von 5% nur leicht ansteigt.

Effektstärke: Neben der statistischen Signifikanz ist in der Regel auch die praktische Bedeutung eines Effekts von Interesse. Zur Beurteilung der Effektstärken in zweifaktorielle ANOVA kann man die sogenannte partielle Effektstärken betrachtet, die wie folgt definiert werden (Holling und Gediga, 2016, S.265-266):

$$\eta_p^2(\text{Effekt}) = \frac{SS_{\text{Effekt}}}{SS_{\text{Effekt}} + SS_W}.$$

Dabei ist SS_{Effekt} ein Platzhalter für eine der Quadratsummen SS_A, SS_B, SS_{AB} . Gemäß der Klassifikation von Cohen (1988) wird der Wert von η_p^2 als kleiner (0.010), mittlerer (0.059) oder großer (0.138) Effekt interpretiert.

3.2. Levene-Test

Der Levene-Test wird verwendet, um die Varianzhomogenität einer kontinuierlichen, numerischen Variable zwischen mehreren Stichprobengruppen zu prüfen. Angenommen sind J unabhängige Gruppen. Diese Gruppen haben jeweils den Umfang von $n_j, j = 1, \dots, J$ und die Variablen $X_{ij}, i = 1, \dots, n_j$ sind auch innerhalb jeder Gruppe unabhängig. Die Nullhypothese H_0 lautet, dass die Varianz über alle Gruppen hinweg gleich ist, gegen die Alternative H_1 , dass mindestens zwei Varianzen unterschiedlich sind. Die Teststatistik ist definiert als $F = \frac{N-J}{J-1} \cdot \frac{\sum_{j=1}^J n_j (\bar{Z}_{\cdot j} - \bar{Z}_{\cdot\cdot})^2}{\sum_{j=1}^J \sum_{i=1}^{n_j} (Z_{ij} - \bar{Z}_{\cdot j})^2} \stackrel{H_0}{\approx} \mathcal{F}_{(J-1, N-J)}$ (Hedderich und Sachs, 2020, S.624, 627), wobei

- $Z_{ij} = |X_{ij} - \tilde{X}_{\cdot j}|$ die absolute Abweichung von der Variable X_{ij} zur Gruppenlage $\tilde{X}_{\cdot j}$ (Mittelwert oder Median) bezeichnet,
- $\bar{Z}_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Z_{ij}$ der Mittelwert der Abweichungen in Gruppe j ist und
- $\bar{Z}_{\cdot\cdot} = \frac{1}{N} \sum_{j=1}^J \sum_{i=1}^{n_j} Z_{ij}$ der Gesamtmittelwert der Abweichungen ist.

Da F unter H_0 näherungsweise der F-Verteilung mit $J - 1$ und $N - J$ Freiheitsgraden folgt, wird H_0 zum Niveau α verworfen, falls $Pr(F > f) < \alpha$.

3.3. Bonferroni-Korrektur

Bei multiplen Tests steigt das Risiko, dass mindestens eine Nullhypothese fälschlich abgelehnt wird. Um diese Fehlerrate zu begrenzen, wird die Bonferroni-Korrektur angewendet. Sei eine Familie mit m Hypothesen H_1, \dots, H_m zu testen, wird das Signifikanzniveau α für die gesamte Familie gleichmäßig auf die m Hypothesen verteilt. Dementsprechend wird jede Hypothese $H_i, i = 1, \dots, m$ auf das Niveau $\alpha_i = \frac{\alpha}{m}$ getestet (Holling und Gediga, 2016, S.241-242). Die Bonferroni-Korrektur hält die Fehlerrate erster Art über alle Tests hinweg unter α , erschwert aber bei vielen Hypothesen die Ablehnung der Nullhypothesen.

4. Statistische Auswertung

Zur Vergleichbarkeit der neun Schätzungen wird für jede Versuchsperson die relative Differenz berechnet, indem der Schätzwert vom jeweiligen Ankerwert subtrahiert und durch den Ankerwert geteilt wird. Nur positive Werte (d.h. tatsächliche Unterschätzungen) werden berücksichtigt, da die Instruktion eine Schätzung unter dem Anker verlangt. Der Mittelwert der relativen Differenzen (MRD) über alle Aufgaben ergibt eine einheitliche Kennzahl pro Person, die als abhängige Variable in der Analyse dient.

4.1. Deskriptive Statistik

Die deskriptiven Kennwerte in Tabelle 2 zeigen deutliche Unterschiede im MRD zwischen den vier Bedingungen. Die höchsten Mittelwerte treten bei hoher Motivation zur Unterschätzung auf (Bedingung 1: 33.20%, Bedingung 3: 41.00%), während die Werte bei geringer Motivation deutlich niedriger liegen (Bedingung 2: 9.92%, Bedingung 4: 16.80%). Die Medianwerte folgen einem ähnlichen Muster wie die Mittelwerte, wobei die Unterschiede zwischen den Bedingungen konsistent bleiben.

Tabelle 2: Vergleich des MRD nach Fragebogenvariante

Bedingung	Min	Median	Mittelwert	Max	SD	IQR
1	3.84	32.10	33.20	59.50	10.80	14.00
2	1.74	7.21	9.92	36.90	7.12	6.32
3	15.30	42.20	41.00	63.30	10.90	13.00
4	1.66	15.20	16.80	47.50	10.50	10.20

Anmerkung: SD = Standardabweichung, IQR = Interquartilsabstand.

Bedingung: 1. Variante = präziser Anker und größere Motivation, 2. Variante = präziser Anker und kleinere Motivation, 3. Variante = runder Anker und größere Motivation bzw. 4. Variante = runder Anker und kleinere Motivation

Die Standardabweichung ist in allen Bedingungen vergleichbar (zwischen 10.50% und 10.90%), mit Ausnahme von Bedingung 2, die mit 7.12% eine geringere Streuung aufweist. Der Boxplot in Abbildung 1 bestätigt diese Muster und zeigt zusätzlich einzelne Ausreißer in den Bedingungen 1, 3 und 4, während Bedingung 2 eine kompakte Verteilung ohne extreme Werte aufweist.

4.2. Hypothesentests

Für die Anwendung einer zweifaktoriellen ANOVA müssen drei Annahmen berücksichtigt werden: Unabhängigkeit der Beobachtungen, Normalverteilung innerhalb jeder Gruppe und Homogenität der Varianz. Die Unabhängigkeit der Beobachtungen ist gegeben, da jede Versuchsperson ausschließlich einer Bedingung zugeordnet ist und keine Messwiederholungen vorliegen. Die Normalverteilung innerhalb jeder Gruppen wird visuell anhand von Q-Q-Plots beurteilt (vgl. Abbildung 2 bis Abbildung 5). In den Bedingungen 1 und 3 folgen die Datenpunkte weitgehend der Hauptdiagonalen, was auf eine annähernde Normalverteilung hindeutet. In den Bedingungen 2 und 4 zeigen sich hingegen deutliche Abweichungen, insbesondere in den Randbereichen. Zusätzlich ist zu beachten, dass der MRD die Werte ausschließlich im Bereich von 0% bis 100% annehmen kann und somit eine natürliche Begrenzung aufweist. Eine exakte Normalverteilung ist daher theoretisch nicht zu erwarten. Die Homogenität der Varianzen wird mithilfe des Levene-Tests geprüft. Der Test ergibt einen p-Wert von 0.168. Da dieser über dem Signifikanzniveau von 0.05 liegt, kann die Annahme der Varianzgleichheit beibehalten werden. Da die Annahme einer Normalverteilung nicht erfüllt ist, wird zur Überprüfung der Effekte zusätzlich das ART-Verfahren angewendet.

Tabelle 3: Ergebnisse der zweifaktoriellen Varianzanalyse mit Aligned Rank Transform für die Faktoren Ankertyp und Motivation

Nullhypothese	Teststatistik	p-Wert	H_0
H_0^A : Kein Effekt des Faktors Ankertyp	19.57	$2.20 \cdot 10^{-5}$	abgelehnt
H_0^B : Kein Effekt des Faktors Motivation	156.07	$< 2.22 \cdot 10^{-16}$	abgelehnt
H_0^{AB} : Keine Interaktion	0.36	0.54	beibehalten

Anmerkung: Testniveau: $\alpha = 0.05$; Bonferroni-korrigiert: $\alpha = 0.05/3 \approx 0.0017$ pro Hypothese.

Die Ergebnisse der zweifaktoriellen Varianzanalyse (Siehe Tabelle 3) zeigen signifikante Haupteffekte für beide untersuchten Faktoren. Der Faktor Ankertyp weist eine Teststatistik von 19.5 mit einem p-Wert von $2.20 \cdot 10^{-5}$ auf, wodurch die Nullhypothese eines fehlenden Effekts klar abgelehnt wird. Noch deutlicher fällt der Effekt des Faktors Motivation aus, mit einer Teststatistik von 156.07 und einem p-Wert kleiner als $2.22 \cdot 10^{-16}$, was ebenfalls zur Ablehnung der entsprechenden Nullhypothese führt. Im Gegensatz dazu erweist sich die Interaktion zwischen Ankertyp und Motivation als nicht signifikant. Mit einer Teststatistik von 0.36 und einem p-Wert von 0.54 wird die Nullhypothese einer fehlenden Interaktion beibehalten.

Tabelle 4: Partielle Effektstärken von Haupt- und Interaktionseffekten in der zweifaktoriellen Varianzanalyse

Effekt	Effektstärke	Interpretation
Ankertyp	0.1236	mittlere
Motivation	0.5957	groß
Interaktion	0.0006	unerheblich

Anmerkung: kleiner Effekt(0.010), mittlerer Effekt (0.059), großer (0.138) Effekt.

Die Analyse der partiellen Effektstärken (Siehe Tabelle 4) zeigt deutliche Unterschiede zwischen den untersuchten Faktoren. Der Faktor Motivation weist mit $\eta_p^2 = 0.5957$ eine große partielle Effektstärke nach Klassifikation von Cohen auf. Der Ankertyp zeigt mit $\eta_p^2 = 0.1236$ eine mittlere partielle Effektstärke, während die Interaktion zwischen beiden Faktoren mit $\eta_p^2 = 0.0006$ als unerheblich einzustufen ist und deutlich unter der Schwelle für kleine Effekte liegt.

5. Zusammenfassung

Die vorliegende Studie untersuchte als Replikation von Janiszewski und Uy (2008) den Einfluss von Anker-Präzision (präzise vs. runde Zahlen) und Motivationsstärke (hoch vs. niedrig zur Unterschätzung) auf das Schätzverhalten. Basierend auf dem Datensatz von Chandler (2015) wurden 120 Versuchspersonen in einem 2×2 -faktoriellen Design analysiert, die jeweils neun Schätzaufgaben zu verschiedenen Szenarien bearbeiteten. Die zentrale Forschungsfrage war, ob präzise Anker zu geringeren Abweichungen führen als runde Anker und ob dieser Effekt von der Motivation zur Unterschätzung moderiert wird.

Die deskriptive Analyse der mittleren prozentualen Abweichungen zeigt deutliche Unterschiede zwischen den Bedingungen: Die geringsten Abweichungen traten bei präzisen Ankern mit niedriger Motivation auf (9.92%), die höchsten bei runden Ankern mit hoher Motivation (41.0%). Aufgrund verletzter Normalverteilungsannahmen wurde eine Aligned Rank Transform (ART) angewendet, um robuste nichtparametrische Analysen der Faktoreffekte zu ermöglichen. Die anschließende zweifaktorielle Varianzanalyse auf den transformierten Daten bestätigte signifikante Haupteffekte für beide Faktoren: Motivation ($F = 156.07, p < 2.22 \cdot 10^{-16}, \eta_p^2 = 0.5957$) und Ankertyp ($F = 19.57, p = 2.20 \cdot 10^{-5}, \eta_p^2 = 0.1236$). Eine Interaktion zwischen beiden Faktoren

war nicht nachweisbar ($F = 0.36, p = 0.54, \eta_p^2 = 0.0006$). Alle Ergebnisse blieben nach Bonferroni-Korrektur für Mehrfachtestung robust. Die Befunde stützen die ursprünglichen Ergebnisse von Janiszewski und Uy (2008) bezüglich der Überlegenheit präziser Anker. Die vorliegende Replikation zeigt additive Effekte beider Faktoren, wobei die Motivation den dominanten Einfluss ausübt. Die robuste statistische Absicherung durch Bonferroni-Korrektur stärkt die Verlässlichkeit der vorliegenden Ergebnisse

Die praktische Relevanz der Anker-Präzision für Verhandlungs- und Entscheidungssituationen rechtfertigt weitere Replikationsstudien mit größeren Stichproben und kulturell diverseren Populationen. Zukünftige Untersuchungen sollten die kognitiven Mechanismen der differenziellen Ankereffekte systematischer erforschen und dabei moderne statistische Verfahren zur Analyse wiederholter Messungen einsetzen. Die Diskrepanzen zur Originalstudie unterstreichen die Notwendigkeit präregistrierter Replikationen in der psychologischen Forschung.

Literatur

- J. J. Chandler. Reproducibility Project: Psychology/Replication of Janiszewski & Uy (2008, PS, Study 4b). <https://osf.io>, 2015. Besucht am: 2025-05-23.
- J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1988.
- L. Fahrmeier, C. Heumann, R. Künstler, I. Pigeot, und G. Tutz. *Statistik: Der Weg zur Datenanalyse*. Springer-Verlag, New York, 2016. (8. Auflage),.
- J. Fox und S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition, 2019. URL <https://www.john-fox.ca/Companion/>.
- J. Hedderich und L. Sachs. *Angewandte Statistik: Methodensammlung mit R*. Springer Spektrum, Heidelberg, 2020. (17. Auflage).
- J. J. Higgins, R. C. Blair, und S. Tashtoush. The aligned rank transform procedure. In *Conference on Applied Statistics in Agriculture*. Kansas State University, New Prairie Press, 1990. URL <https://doi.org/10.4148/2475-7772.1443>. 2nd Annual Conference Proceedings.
- H. Holling und G. Gediga. *Statistik – Testverfahren*. Bachelorstudium Psychologie. Hogrefe, 1 edition, 2016.
- C. Janiszewski und D. Uy. Anchor precision influences the amount of adjustment. *Psychological Science*, 19(2):121–127, 2008.
- M. Kay, L. A. Elkin, J. J. Higgins, und J. O. Wobbrock. *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs*, 2025. URL <https://github.com/mjskay/ARTool>. R package version 0.11.2.
- R Core Team. R: A language and environment for statistical computing, 2024. URL <https://www.R-project.org/>.
- A. Tversky und D. Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, 1974.
- J. R. van Ginkel und P. M. Kroonenberg. Multiple Imputation to Balance Unbalanced Designs for Two-Way Analysis of Variance. *Methodology*, 17(1):39–57, 2021.

H. Wickham, D. Navarro, und T. L. Pedersen. *ggplot2: Elegant Graphics for Data Analysis (3e)*. Springer-Verlag New York, work-in-progress 3rd edition, 2016. URL <https://ggplot2-book.org/>.

A. Zusätzliche Grafiken

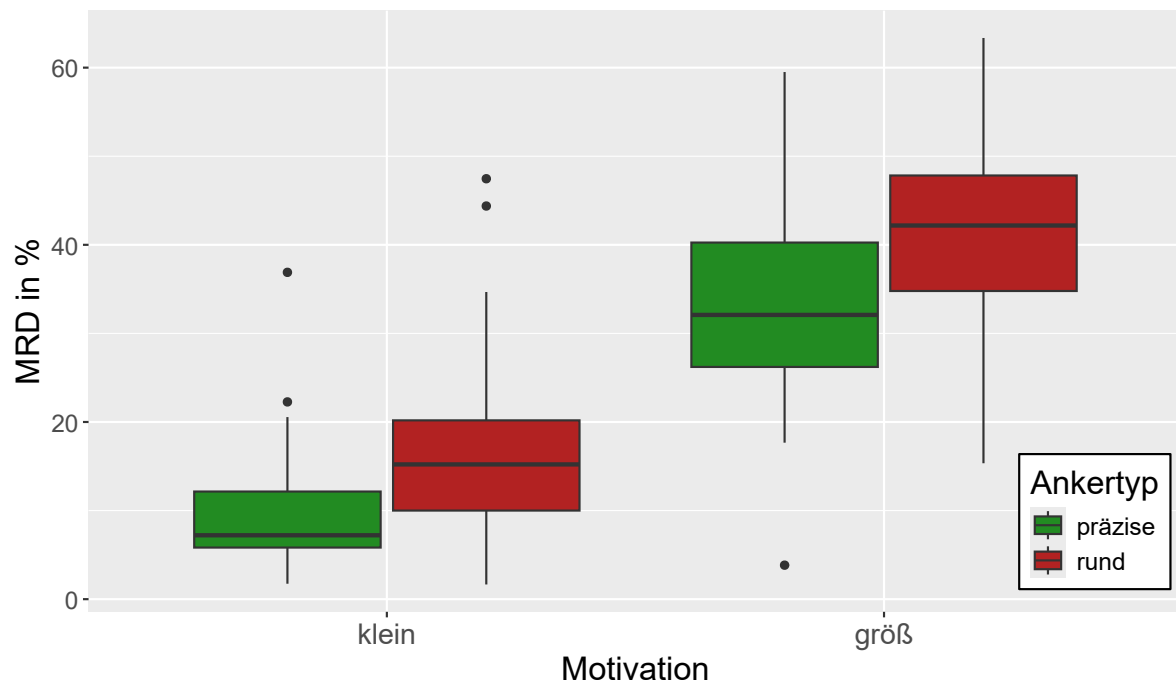


Abbildung 1: Boxplot des MRD nach Fragebogenvariante

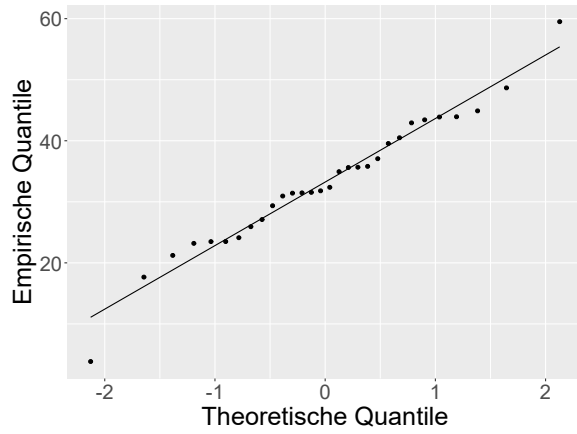


Abbildung 2: Q-Q-Plot des MRD unter Bedingung 1

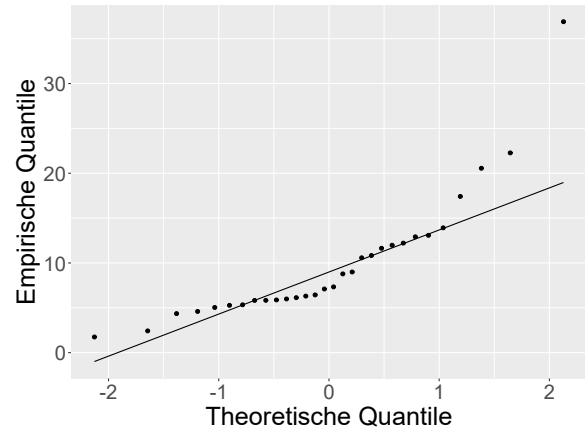


Abbildung 3: Q-Q-Plot des MRD unter Bedingung 2

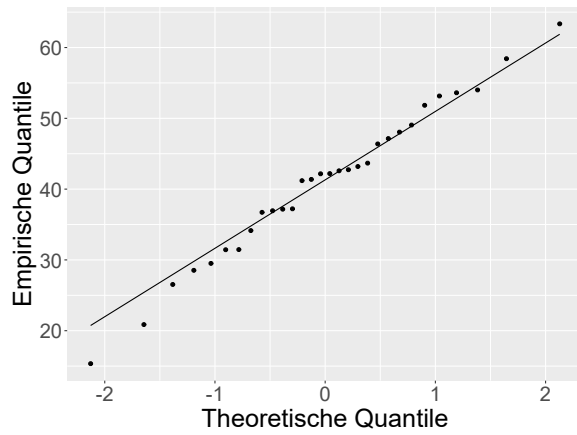


Abbildung 4: Q-Q-Plot des MRD unter Bedingung 3

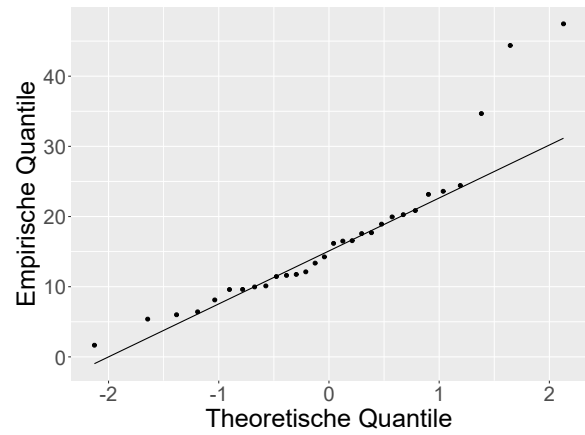


Abbildung 5: Q-Q-Plot des MRD unter Bedingung 4