

QAF: Quantitative Forschungsmethoden

Tarek Carls

14. November 2023

Agenda

- **Session 1:** Grundlagen, induktive Statistik, Konfidenzintervalle
- **Session 2:** t-Tests, einfaktorielle ANOVA
- **Session 3:** Mehrfaktorielle ANOVA
- **Session 4:** Lineare Regression, logistische Regression
- **Session 5:** Fragen und Wiederholung

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke

Einführung

- Mehrfaktorielle ANOVA: Analyse von Datensätzen mit mehr als einem Faktor.
- Erweiterung der einfaktoriellen ANOVA.
- Ermöglicht die Untersuchung von Wechselwirkungen zwischen Faktoren.
- Beispiel: Einfluss von Diät (Faktor 1) und Sport (Faktor 2) auf das Körpergewicht.

Unterschiede zur einfaktoriellen ANOVA

- Einfaktorielle ANOVA: Analysiert den Einfluss eines einzigen Faktors.
- Mehrfaktorielle ANOVA: Betrachtet gleichzeitig mehrere Faktoren.
- Möglichkeit, Wechselwirkungen zwischen Faktoren zu analysieren.
- Beispiel: Nicht nur der Effekt von Diät oder Sport allein, sondern auch, wie sie sich gegenseitig beeinflussen.

Annahmen der mehrfaktoriellen ANOVA

- Unabhängigkeit der Beobachtungen: Keine Korrelation zwischen den Beobachtungen.
- Normalverteilung der Residuen: Daten sollten normalverteilt sein, wenn man die Gruppen vergleicht.
- Homogenität der Varianzen (Homoskedastizität): Die Varianzen in den verschiedenen Gruppen sollten ähnlich sein.
- Bei Verletzung dieser Annahmen können alternative Methoden oder Transformationen der Daten notwendig sein.

Grundlagen der mehrfaktoriellen ANOVA

- Faktoren und Stufen: Jeder Faktor kann verschiedene Stufen oder Kategorien haben.
- Haupteffekte: Der Einfluss eines einzelnen Faktors auf die abhängige Variable.
- Wechselwirkungen: Wie sich die Kombination von Faktoren auf die abhängige Variable auswirkt.
- Beispiel: Haupteffekt von Diät, Haupteffekt von Sport, Wechselwirkung zwischen Diät und Sport.

Interpretation der Ergebnisse

- Signifikanz der Haupteffekte und Wechselwirkungen.
- ANOVA-Tabelle: Quelle der Variation, Summe der Quadrate, Freiheitsgrade, F-Wert, p-Wert.
- Post-hoc-Analysen bei signifikanten Ergebnissen, um spezifische Gruppenunterschiede zu identifizieren.
- Beispiel: Wenn die Wechselwirkung signifikant ist, könnte dies bedeuten, dass der Effekt von Sport auf das Gewicht von der Art der Diät abhängt.

Beispielfragestellung

Untersuchung des Einflusses der Arbeitsumgebung (Büro vs. Homeoffice) und der Arbeitszeitflexibilität (flexibel vs. fest) auf die Arbeitszufriedenheit.

- Faktor 1: Arbeitsumgebung
 - Faktorstufen: Büro, Homeoffice
- Faktor 2: Arbeitszeit:
 - Faktorstufen: Fest, Flexibel
- Abhängige Variable: Zufriedenheit

Datengrundlage

Teilnehmer	Arbeitsumgebung	Arbeitszeit	Zufriedenheit
1	Büro	Flexibel	7
2	Homeoffice	Flexibel	8
3	Büro	Fest	6
4	Homeoffice	Fest	5
5	Büro	Flexibel	7
6	Homeoffice	Flexibel	9
7	Büro	Fest	6
8	Homeoffice	Fest	7
9	Büro	Flexibel	8
10	Homeoffice	Flexibel	9

Tabelle: Beispieldaten zur Arbeitszufriedenheit

Datengrundlage

	Büro		Homeoffice	
	Flexibel	Fest	Flexibel	Fest
Teilnehmer 1	7	6	8	5
Teilnehmer 2	7	6	9	5
Teilnehmer 3	8	6	8	6
Teilnehmer 4	7	5	9	6
Teilnehmer 5	8	6	9	5
Mittelwert	7.4	5.8	8.6	5.4
Varianz	0.24	0.16	0.24	0.24

Tabelle: Kreuztabelle der Arbeitszufriedenheit nach Arbeitsumgebung und Arbeitszeit

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global**
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke

Globaler Erwartungswert in der mehrfaktoriellen ANOVA

- Der globale Erwartungswert (μ) ist der durchschnittliche Wert der abhängigen Variablen über alle Stufen aller Faktoren hinweg.
- In unserem Beispiel: Durchschnittliche Arbeitszufriedenheit über alle Kombinationen von Arbeitsumgebung und Arbeitszeit.

$$\mu = \frac{\sum_{i=1}^n Y_i}{n} \quad (1)$$

- Y_i repräsentiert die Arbeitszufriedenheitswerte.
- n ist die Gesamtzahl der Beobachtungen.
- Beispiel: Wenn wir 20 Beobachtungen haben, ist μ der Durchschnitt dieser 20 Werte.

Bedeutung des globalen Erwartungswertes

- Der globale Erwartungswert dient als Referenzpunkt, um die Effekte der einzelnen Faktoren zu beurteilen.
- In der ANOVA wird untersucht, wie stark die Werte von diesem globalen Durchschnitt abweichen.
- Diese Abweichungen werden auf die Einflüsse der Faktoren (Arbeitsumgebung, Arbeitszeit) und deren Wechselwirkungen zurückgeführt.

Lokale Erwartungswerte

- Lokale Erwartungswerte beziehen sich auf den Durchschnittswert der abhängigen Variablen innerhalb einer Faktorstufe, über alle Stufen des anderen Faktors hinweg.
- Beispiel: Durchschnittliche Arbeitszufriedenheit im Büro, unabhängig von der Arbeitszeit.

$$\mu_{\text{Faktor A, Stufe } i} = \frac{\sum Y_{\text{Faktor A, Stufe } i}}{n_{\text{Faktor A, Stufe } i}} \quad (2)$$

- $Y_{\text{Faktor A, Stufe } i}$ sind die Werte der abhängigen Variablen für eine bestimmte Stufe des Faktors A.
- $n_{\text{Faktor A, Stufe } i}$ ist die Anzahl der Beobachtungen in dieser Stufe.

Bedeutung lokaler Erwartungswerte

- Lokale Erwartungswerte helfen, den spezifischen Einfluss einer Faktorstufe zu verstehen.
- Sie zeigen, wie sich eine bestimmte Stufe eines Faktors auf die abhängige Variable auswirkt, unabhängig von den Stufen des anderen Faktors.
- Beispiel: Vergleich der Arbeitszufriedenheit zwischen Büro und Homeoffice, unabhängig davon, ob die Arbeitszeit flexibel oder fest ist.

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte**
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke

Haupteffekte in der mehrfaktoriellen ANOVA

- Haupteffekte beschreiben den Einfluss eines einzelnen Faktors auf die abhängige Variable, unabhängig von anderen Faktoren.
- In unserem Beispiel: Der Einfluss der Arbeitsumgebung und der Arbeitszeit auf die Arbeitszufriedenheit, jeweils betrachtet ohne Berücksichtigung des anderen Faktors.

Berechnung der Haupteffekte

- Haupteffekte messen den durchschnittlichen Einfluss jeder Faktorstufe auf die abhängige Variable.
- Effekt einer Faktorstufe A (z.B. Arbeitsumgebung):

$$\alpha_i = \bar{Y}_{i.} - \bar{Y}$$

- Effekt einer Faktorstufe B (z.B. Arbeitszeit):

$$\beta_j = \bar{Y}_{.j} - \bar{Y}$$

- Hierbei ist $\bar{Y}_{i.}$ der Mittelwert der i-ten Stufe des Faktors A, $\bar{Y}_{.j}$ der Mittelwert der j-ten Stufe des Faktors B und \bar{Y} der Gesamtmittelwert.

Das Haupteffektmodell in der mehrfaktoriellen ANOVA

- Das Haupteffektmodell konzentriert sich auf die Analyse der Haupteffekte der Faktoren, ohne Interaktionseffekte zu berücksichtigen.
- Modellgleichung:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad (3)$$

- Y_{ijk} : Beobachteter Wert der abhängigen Variablen.
 - μ : Globaler Erwartungswert.
 - α_i : Effekt der i-ten Stufe des Faktors A.
 - β_j : Effekt der j-ten Stufe des Faktors B.
 - ϵ_{ijk} : Zufälliger Fehlerterm.
- Dieses Modell geht davon aus, dass es keine Interaktion zwischen den Faktoren gibt.
- Es ermöglicht eine vereinfachte Analyse, die sich auf die isolierten Effekte jedes Faktors konzentriert.

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte**
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke

Interaktionseffekte in der mehrfaktoriellen ANOVA

- Interaktionseffekte treten auf, wenn der Effekt eines Faktors auf die abhängige Variable von der Stufe eines anderen Faktors abhängt.
- In unserem Beispiel: Wie die Kombination von Arbeitsumgebung und Arbeitszeit die Arbeitszufriedenheit beeinflusst.

Berechnung der Interaktionseffekte

- Interaktionseffekte zeigen, wie die Kombination der Faktorstufen die abhängige Variable beeinflusst.
- Interaktionseffekt zwischen Faktorstufen A und B:

$$(\alpha\beta)_{ij} = \bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}$$

- \bar{Y}_{ij} ist der Mittelwert der Kombination der i-ten Stufe des Faktors A und der j-ten Stufe des Faktors B.
- Diese Berechnung zeigt, ob die Kombination der Faktorstufen einen anderen Effekt hat als die Summe der individuellen Haupteffekte.

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung**
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke

Modellgleichung der mehrfaktoriellen ANOVA

- Die allgemeine Modellgleichung:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad (4)$$

- Y_{ijk} : Beobachteter Wert der abhängigen Variablen (z.B. Arbeitszufriedenheit).
- μ : Globaler Erwartungswert der abhängigen Variablen.
- α_i : Effekt der i-ten Stufe des Faktors A (z.B. Arbeitsumgebung).
- β_j : Effekt der j-ten Stufe des Faktors B (z.B. Arbeitszeit).
- $(\alpha\beta)_{ij}$: Interaktionseffekt zwischen der i-ten Stufe des Faktors A und der j-ten Stufe des Faktors B.
- ϵ_{ijk} : Zufälliger Fehlerterm, der die unerklärte Variabilität repräsentiert.
- Diese Gleichung ermöglicht es, die Effekte der einzelnen Faktoren sowie ihre Interaktionen auf die abhängige Variable zu analysieren.

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung**
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke

Quadratsummenzerlegung in der mehrfaktoriellen ANOVA

- Die Quadratsummenzerlegung teilt die Gesamtvariabilität in der abhängigen Variablen in verschiedene Komponenten auf.
- Diese Komponenten umfassen die Effekte der einzelnen Faktoren (Haupteffekte), die Interaktionseffekte zwischen den Faktoren und den Fehler (unerklärte Variabilität).

Komponenten der Quadratsummenzerlegung

- Gesamtquadratsumme (Total Sum of Squares, TSS): Die gesamte Variabilität in den Daten.
- Quadratsumme der Haupteffekte (Sum of Squares for Factor A, SS_A ; Sum of Squares for Factor B, SS_B).
- Quadratsumme der Interaktionseffekte (Sum of Squares for Interaction, SS_{AB}).
- Fehlerquadratsumme (Error Sum of Squares, SS_{Error}): Die Variabilität, die nicht durch die Faktoren oder ihre Interaktion erklärt wird.

Berechnung der Quadratsummen

- $TSS = SS_A + SS_B + SS_{AB} + SS_{Error}$.
- Jede Quadratsumme wird durch Vergleich der Mittelwerte der entsprechenden Gruppen oder Kombinationen von Gruppen berechnet.
- Beispiel: SS_A wird berechnet, indem die Mittelwerte der verschiedenen Arbeitsumgebungen verglichen werden.

Formeln für die Quadratsummen

- Gesamtquadratsumme (TSS):

$$\text{TSS} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

- Quadratsumme für Faktor A (SS_A):

$$\text{SS}_A = \sum_{i=1}^a n_i (\bar{Y}_{i\cdot} - \bar{Y})^2$$

- Quadratsumme für Faktor B (SS_B):

$$\text{SS}_B = \sum_{j=1}^b m_j (\bar{Y}_{\cdot j} - \bar{Y})^2$$

Formeln für Interaktions- und Fehlerquadratsummen

- Quadratsumme der Interaktionseffekte (SS_AB):

$$SS_{AB} = \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y})^2$$

- Fehlerquadratsumme (SS_Error):

$$SS_{Error} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij})^2$$

- Hierbei ist \bar{Y} der Gesamtmittelwert, $\bar{Y}_{i.}$ und $\bar{Y}_{.j}$ sind die Mittelwerte der Faktorstufen und \bar{Y}_{ij} ist der Mittelwert der Interaktion zwischen den Faktorstufen.

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen**
- 8 F-Bruch und Effektstärke

Bestimmung der Freiheitsgrade in der mehrfaktoriellen ANOVA

- Die Freiheitsgrade df sind ein Maß für die Anzahl der unabhängigen Werte, die in der Berechnung einer Statistik verwendet werden.
- Freiheitsgrade für Haupteffekte:
 - df für Faktor A (df_A): $a - 1$, wobei a die Anzahl der Stufen von Faktor A ist.
 - df für Faktor B (df_B): $b - 1$, wobei b die Anzahl der Stufen von Faktor B ist.
- Freiheitsgrade für Interaktionseffekte:
 - df für die Interaktion AxB (df_{AB}): $(a - 1)(b - 1)$.
- Freiheitsgrade für den Fehler (df_{Error}):
 - df_{Error} : $n - ab$, wobei n die Gesamtzahl der Beobachtungen ist.
- Gesamtfreiheitsgrade (df_{Total}): $n - 1$.

Berechnung der mittleren Quadratsummen

- Die mittleren Quadratsummen (Mean Squares, MS) sind ein Maß für die durchschnittliche Variabilität, die jeder Effekt erklärt.
- Mittlere Quadratsumme für Faktor A (MS_A):

$$MS_A = \frac{SS_A}{df_A}$$

- Mittlere Quadratsumme für Faktor B (MS_B):

$$MS_B = \frac{SS_B}{df_B}$$

- Mittlere Quadratsumme für Interaktionseffekte (MS_{AB}):

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$$

- Mittlere Quadratsumme für den Fehler (MS_{Error}):

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{df_{\text{Error}}}$$

Agenda - Session 3

- 1 Einführung
- 2 Erwartungswerte - Lokal und Global
- 3 Haupteffekte
- 4 Interaktionseffekte
- 5 Modellgleichung
- 6 Quadratsummenzerlegung
- 7 Freiheitsgrade und mittlere Quadratsummen
- 8 F-Bruch und Effektstärke**

Berechnung der F-Brüche

- Die F-Brüche (F-Werte) werden verwendet, um die statistische Signifikanz der Effekte zu testen.

- F-Wert für Faktor A:

$$F_A = \frac{MS_A}{MS_{\text{Error}}}$$

- F-Wert für Faktor B:

$$F_B = \frac{MS_B}{MS_{\text{Error}}}$$

- F-Wert für Interaktionseffekte AxB:

$$F_{AB} = \frac{MS_{AB}}{MS_{\text{Error}}}$$

- Diese Werte werden mit kritischen F-Werten aus F-Verteilungstabellen verglichen, um die Signifikanz zu bestimmen.

Effektstärke: Partielles Eta-Quadrat

- Das partielle Eta-Quadrat (η^2) misst die Effektstärke, d.h. den Anteil der Gesamtvarianz, der durch einen Faktor erklärt wird.
- Berechnung für Faktor A:

$$\eta_A^2 = \frac{SS_A}{SS_A + SS_{\text{Error}}}$$

- Berechnung für Faktor B:

$$\eta_B^2 = \frac{SS_B}{SS_B + SS_{\text{Error}}}$$

- Höhere Werte von η^2 deuten auf eine stärkere Effektstärke hin.