

Statistik – Methoden zum Mittelwertvergleich von mehreren Gruppen

Zweifaktorielle Varianzanalyse

Zweifaktorielle Varianzanalyse

- Bisher kennen wir die einfaktorielle Varianzanalyse, wir können Mittelwerte von Gruppen untersuchen, die sich durch einen Faktor unterscheiden
- Beispiel Fehlerrate für verschiedene Automobilemodelle
- Auf Grundlage der einfaktoriellen Analyse können wir entscheiden ob Unterschiede vorhanden und signifikant sind
- Was passiert, wenn wir zusätzliche Faktoren zur Unterscheidung unserer Gruppen einführen, zum Beispiel eine Motorvariante

Zweifaktorielle Varianzanalyse

- Im Fall von zwei Faktoren untersucht die zweifaktorielle Varianzanalyse ebenfalls die Mittelwerte der Gruppen
- Führt man weitere Faktoren ein, machen wir den Schritt von zweifaktorieller zur mehrfaktoriellen Varianzanalyse
- Das Analysekonzept wird sich dabei nicht ändern, die Auswertung wird aber komplexer

Zweifaktorielle Varianzanalyse

Einführung am Beispiel

Für die zweifaktorielle Varianzanalyse werden wir Daten über Haltbarkeit von Büroklammern nutzen

Sie finden die Daten in *Beispiel_Büroklammer.xlsx*

Wir werden für das Beispiel zwei Faktoren einführen

- Größe der Büroklammer (26 – 32mm)
- Eine mögliche Wärmebehandlung der Klammer (Nein/Ja)
- Das Ergebnis der Untersuchungen wird ein Biegeindex sein, der die Haltbarkeit der Klammern beschreibt.

Zweifaktorielle Varianzanalyse

Beispiel: Mittelwerte der Daten nach Faktoren

Wärme	Größe			
		26mm	29mm	32mm
	Nein	15,7495	13,7856	11,7813
	Ja	19,1823	20,6881	21,9831

- Es gibt erkennbare Unterschiede hinsichtlich Größe und Wärme
- Sind diese Unterschiede aber auch signifikant?

Zweifaktorielle Varianzanalyse

Voraussetzungen für die zweifaktorielle Varianzanalyse

- Mindestens intervallskalierte abhängige Variable
- Merkmalsausprägungen müssen unabhängig voneinander sein (falls nicht: ANOVA mit Messwiederholung...)
- Normalverteilung der abhängigen Variable innerhalb der einzelnen Gruppen (für den Gesamtdatensatz ist dies nicht erforderlich)
- Gleiche Varianz aller Gruppen

Zweifaktorielle Varianzanalyse

Hypothesen

- Wir wollen mit den Hypothesen unterschiedliche Fragestellungen beantworten und werden deshalb mehrere Hypothesen brauchen
1. Hat die Größe einen Einfluss auf die Haltbarkeit? (Haupteffekt)
 2. Hat die Wärme einen Einfluss auf die Haltbarkeit? (Haupteffekt)
 3. Haben Größe und Wärmebehandlung im Zusammenspiel einen Einfluss auf die Haltbarkeit? (Wechselwirkungseffekt)

Zweifaktorielle Varianzanalyse

Hypothesen

Größe

H_0 Es gibt **keinen** Unterschied der Haltbarkeit bei verschieden großen Büroklammern

H_1 Es gibt einen Unterschied der Haltbarkeit bei verschieden großen Büroklammern

Wärmebehandlung

H_0 Es gibt **keinen** Unterschied der Haltbarkeit durch die Wärmebehandlung

H_1 Es gibt einen Unterschied der Haltbarkeit durch die Wärmebehandlung

Zweifaktorielle Varianzanalyse

Hypothesen

Wechselwirkung Größe:Wärmebehandlung

- H_0 Die Haltbarkeit der verschieden großer Büroklammern wird nicht durch die Wärmebehandlung beeinflusst
- H_1 Die Haltbarkeit der verschieden großen Büroklammern ist abhängig von der Wärmebehandlung
- Das dritte Hypothesenpaar stellt sogenannte Interaktions-effekte dar
 - Gibt es eine Wechselwirkung zwischen den beiden Faktoren?

Zweifaktorielle Varianzanalyse

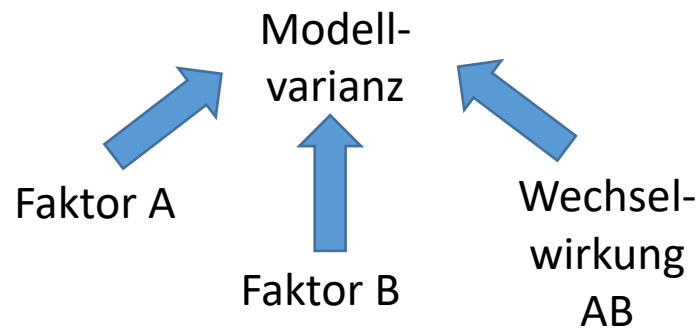
Funktionsprinzip

- Untersuchung der Daten auf systematische bzw. zufällige Einflüsse
- Systematisch: Ein Faktor wirkt auf die Haltbarkeit
- Zufällig: Innerhalb einer Gruppe streuen die Daten ohne das ein Grund erkennbar ist
- Aufteilung der gesamten Streuung in den Daten in Vorhersagevarianz (Modellvarianz) und Fehlervarianz

Zweifaktorielle Varianzanalyse

$$SS_{Gesamt} = SS_{zwischen} + SS_{innerhalb}$$

Fehler-
varianz



- Die Modellvarianz wird für die zweifaktorielle Varianzanalyse durch zwei Faktoren und deren Wechselwirkung gebildet

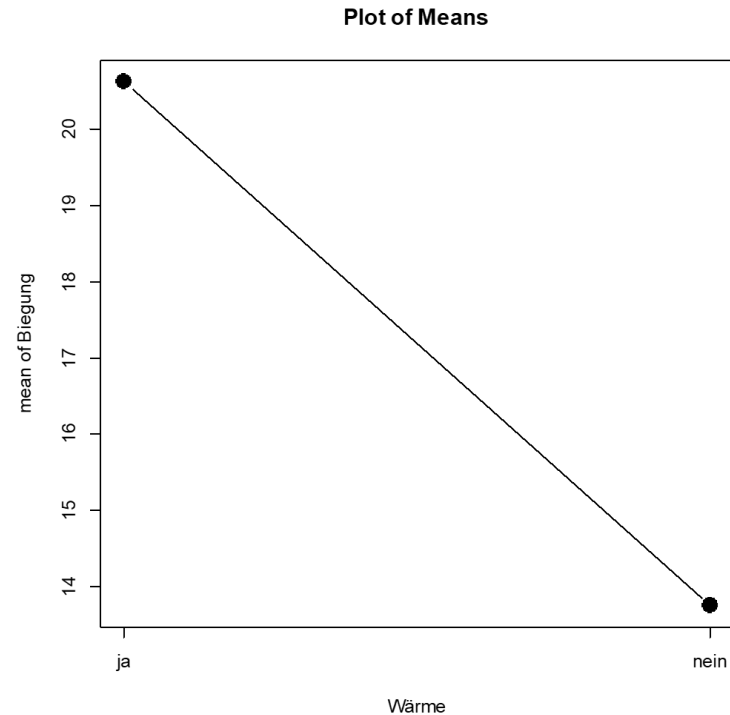
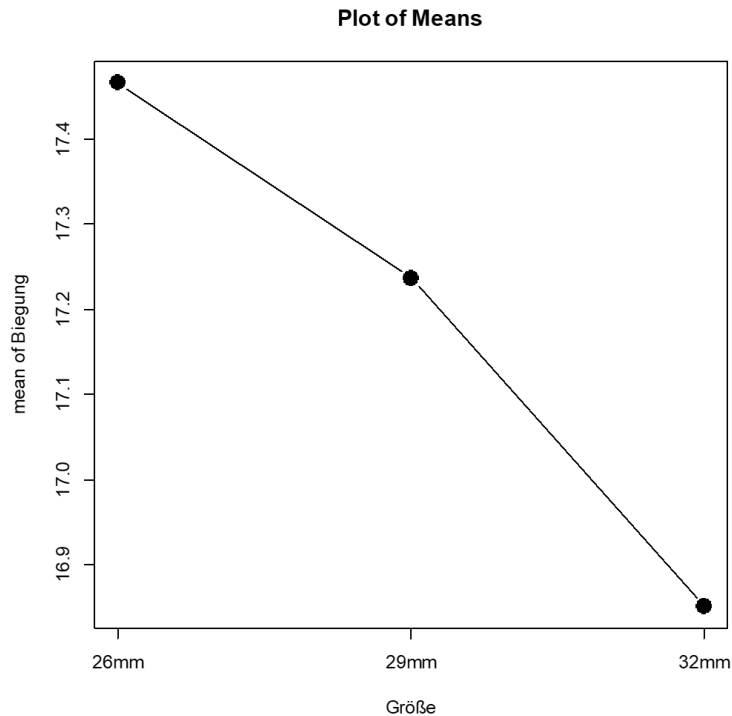
Zweifaktorielle Varianzanalyse

Haupteffekte

- Unmittelbarer Einfluss eines Faktors auf die abhängige Variable
- Hier: Größe und Wärme
- Die Wirkung der beiden Haupteffekte werden über die ersten beiden Hypothesenpaare untersucht
- Wechseln wir dort zur Alternativhypothese, ist der jeweilige Hauptfaktor signifikant

Zweifaktorielle Varianzanalyse

Haupteffekte



Darstellung im Plot of Mean ohne Konfidenzintervall!

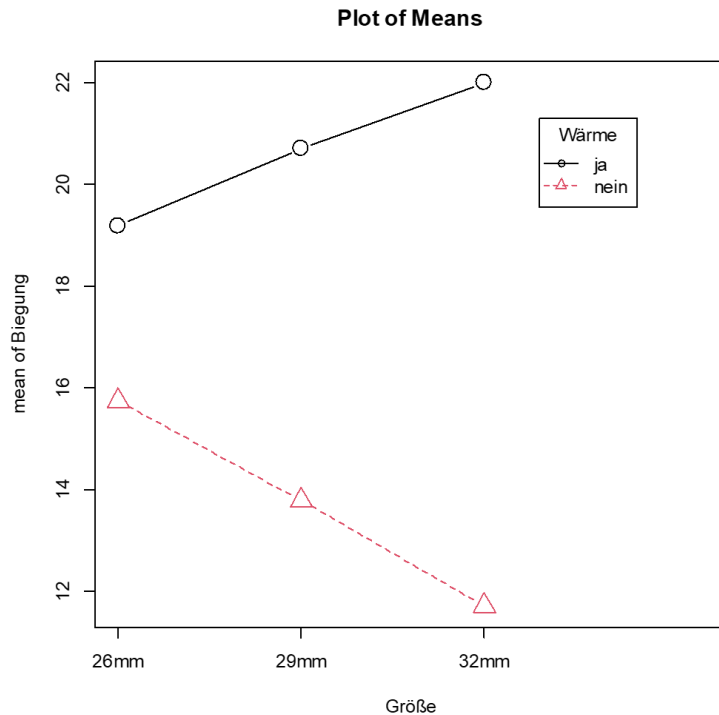
Zweifaktorielle Varianzanalyse

Wechselwirkungseffekte

- Die Wirkung eines Faktors hängt vom zweiten Faktor ab bzw. umgekehrt
- Die Wechselwirkung lässt sich durch sogenannte Wechselwirkungsdiagramme darstellen

Zweifaktorielle Varianzanalyse

Wechselwirkungsdiagramm

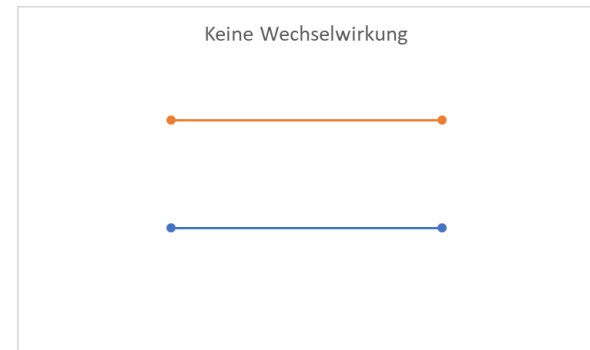


Im Diagramm verlaufen die Linien nicht parallel, wir können erwarten, dass die beiden Faktoren miteinander wechselwirken

Zweifaktorielle Varianzanalyse

Keine Wechselwirkung

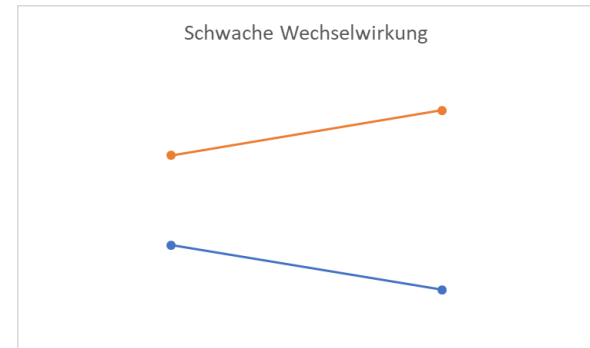
- Die Linien verlaufen parallel
- Es gibt keine Wechselwirkung zwischen den Faktoren
- Die Wirkung der einzelnen Hauptfaktoren kann unabhängig von einander beurteilt werden



Zweifaktorielle Varianzanalyse

Schwache Wechselwirkung

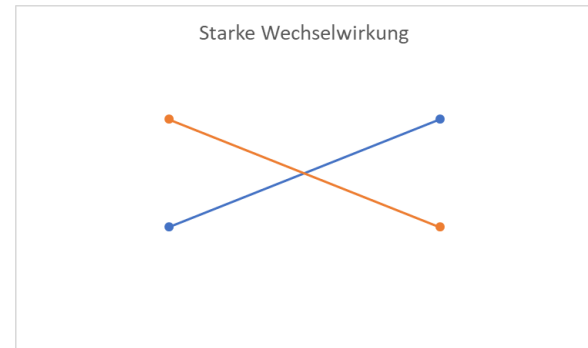
- Die Linien laufen aufeinander zu, kreuzen sich aber nicht
- Es gibt eine schwache Wechselwirkung zwischen den Faktoren
- Die Wirkung der einzelnen Hauptfaktoren kann nicht mehr unabhängig von einander beurteilt werden



Zweifaktorielle Varianzanalyse

Starke Wechselwirkung

- Die Linien laufen kreuzen sich
- Es gibt eine starke Wechselwirkung zwischen den Faktoren
- Die Wirkung der einzelnen Hauptfaktoren kann nicht unabhängig von einander beurteilt werden



Zweifaktorielle Varianzanalyse

Berechnung in R

- Wir können die Daten aus Beispiel_Büroklammer.xlsx in R laden und eine mehrfaktorielle Varianzanalyse durchführen
- Im Folgenden sehen Sie die Ergebnisse, die interpretiert werden müssen

Zweifaktorielle Varianzanalyse

Ergebnisse aus R

Anova Table (Type II tests)

Response: Biegung

	Sum Sq	Df	F value	Pr(>F)
Größe	11.60	2	211.84	< 2.2e-16 ***
Wärme	2121.80	1	77486.42	< 2.2e-16 ***
Größe:Wärme	350.10	2	6392.70	< 2.2e-16 ***
Residuals	4.76	174		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Zweifaktorielle Varianzanalyse

Ergebnisse aus R

means

Wärme

Größe	ja	nein
26mm	19.18231	15.74955
29mm	20.68810	13.78564
32mm	21.98309	11.71833

std. deviations

Wärme

Größe	ja	nein
26mm	0.2105476	0.1699336
29mm	0.1577139	0.1112106
32mm	0.2052480	0.1082631

counts

Wärme

Größe	ja	nein
26mm	30	30
29mm	30	30
32mm	30	30

Zweifaktorielle Varianzanalyse

Deutung der Ergebnisse aus R

Anova Table (Type II tests)

Response: Biegung

	Sum Sq	Df	F value	Pr(>F)
Größe	11.60	2	211.84	< 2.2e-16 ***
Wärme	2121.80	1	77486.42	< 2.2e-16 ***
Größe:Wärme	350.10	2	6392.70	< 2.2e-16 ***
Residuals	4.76	174		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Überprüfung der Faktoren Größe und Wärme
- Gibt es signifikante Unterschiede zwischen den Faktorstufen eines Faktors?

Zweifaktorielle Varianzanalyse

Deutung der Ergebnisse aus R

- Beide Haupteffekte (Größe und Wärme) sind durch sehr niedrige p-Werte gekennzeichnet
- Wechsel zur jeweiligen Alternativhypothese

H_1 Es gibt einen Unterschied der Haltbarkeit bei verschieden großen Büroklammern

H_1 Es gibt einen Unterschied der Haltbarkeit durch die Wärmebehandlung

- Größe und Wärme, jeweils für sich allein betrachtet, führen zu unterschiedlichen Mittelwerten der abhängigen Variable

Zweifaktorielle Varianzanalyse

Deutung der Ergebnisse aus R

Anova Table (Type II tests)

Response: Biegung

	Sum Sq	Df	F value	Pr(>F)
Größe	11.60	2	211.84	< 2.2e-16 ***
Wärme	2121.80	1	77486.42	< 2.2e-16 ***
Größe:Wärme	350.10	2	6392.70	< 2.2e-16 ***
Residuals	4.76	174		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Überprüfung der Wechselwirkung von Größe und Wärme
- Gibt es eine signifikante Wechselwirkung zwischen den Faktoren?

Zweifaktorielle Varianzanalyse

Deutung der Ergebnisse aus R

- Die Wechselwirkung Größe : Wärme ist durch einen sehr niedrigen p-Wert gekennzeichnet
- Wechsel zur Alternativhypothese

H_1 Die Haltbarkeit der verschieden großen Büroklammern ist abhängig von der Wärmebehandlung

- Größe und Wärme, in Kombination, führen zu unterschiedlichen Mittelwerten der abhängigen Variable
- Welcher funktionale Zusammenhang zwischen den beiden Faktoren besteht, ist aus der Varianzanalyse nicht erkennbar

Zweifaktorielle Varianzanalyse

Deutung der Ergebnisse aus R

Anova Table (Type II tests)

Response: Biegung

	Sum Sq	Df	F value	Pr(>F)
Größe	11.60	2	211.84	< 2.2e-16 ***
Wärme	2121.80	1	77486.42	< 2.2e-16 ***
Größe:Wärme	350.10	2	6392.70	< 2.2e-16 ***
Residuals	4.76	174		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ', 1

- Residuals beschreiben die Fehlervarianz
- Wie groß ist die Streuung durch unsystematische Einflüsse im System?

Zweifaktorielle Varianzanalyse

Deutung der Ergebnisse aus R

- Die beiden Faktoren Größe und Wärme sind für sich allein wichtig
- Sie führen, jeweils allein betrachtet, zu Mittelwertsunterschieden in Abhängigkeit von den Faktoreinstellungen
- Die Wechselwirkung Größe : Wärme ist wichtig.
- Je nach Faktorkombination kommt es zu Mittelwertsunterschieden der abhängigen Variable Biegung

Mehrfaktorielle Varianzanalyse

- Die mehrfaktorielle Varianzanalyse ist eine Weiterführung der zweifaktoriellen Varianzanalyse
- Es werden zusätzliche Faktoren in die Analyse aufgenommen
- Beispiel: Automobilmodell – Motorvariante – Getriebe

Mehrfaktorielle Varianzanalyse

- Es kommt zu einer Erhöhung von Hauptfaktoren, Wechselwirkungen und zum Auftreten höherer Wechselwirkungen
 - Erhöhung der Faktorzahl: $A - B - C$
 - Erhöhung der Wechselwirkungen: $A:B - A:C - B:C$
 - Auftreten höherer Wechselwirkungen: $A:B:C$
- Die Berechnung der Varianzanalyse ändert sich nicht

Mehrfaktorielle Varianzanalyse

- Für mehrfaktorielle Varianzanalysen steigt der Datenbedarf mit wachsender Anzahl der Faktoren
- Jede Gruppe, dargestellt durch die möglichen Faktorkombinationen, muss die Voraussetzungen für die Varianzanalyse erfüllen
- Bleibt die Datenmenge (N) gleich, reduziert sich die Stichprobengröße in den einzelnen Gruppen

Mehrfaktorielle Varianzanalyse

Anova Table (Type II tests)

Response: Biegung

	Sum Sq	Df	F value	Pr(>F)	
Größe	11.60	2	208.2551	<2e-16	***
Hersteller	0.01	1	0.5107	0.4758	
Wärme	2121.80	1	76175.4841	<2e-16	***
Größe:Hersteller	0.06	2	1.0005	0.3699	
Größe:Wärme	350.10	2	6284.5465	<2e-16	***
Hersteller:Wärme	0.00	1	0.0905	0.7639	
Größe:Hersteller:Wärme	0.01	2	0.2270	0.7972	
Residuals	4.68	168			

Faktoren

2-fach WW

3-fach WW

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Mehrfaktorielle Varianzanalyse

- Im vorliegenden Fall ist der Faktor Hersteller nicht signifikant ($p = 0,48$)
- Der Faktor Hersteller ist auch in den Zwei- und Dreifach-Wechselwirkungen nicht signifikant (alle p -Werte $> \alpha$)
- Der Faktor Hersteller kann also aus dem Modell entfernt werden