

Anwendungsorientierte Analyseverfahren

Mittelwertvergleiche mit kategoriellen Prädiktoren

Prof. Dr. Michael Scharkow

Sommersemester 2024

Fragen zur praktischen Übung?

WIEDERHOLUNG: TV TIME (H/TAG)

Parameter	Coefficient	95% CI	t(2101)	p	Std. Coef.	Fit
(Intercept)	3.91	(3.39, 4.43)	14.63	< .001	0.00	
age	-0.03	(-0.04, -0.01)	-4.62	< .001	-0.10	
AICc						11414.93
R2						0.01
R2 (adj.)						0.01
Sigma						3.65

MITTELWERTE VERGLEICHEN

- in Experimenten werden oft Mittelwerte einer oder mehrerer Outcome-Variablen zwischen verschiedenen Versuchsbedingungen verglichen.
- auch in nicht-experimentellen Analysen sind Mittelwertvergleiche häufig
- Nullhypothese, dass zwischen den Gruppen kein Unterschied besteht, d.h. die Mittelwerte sich nicht unterscheiden
- meistverwendet: t-Test oder einfaktorielle Varianzanalyse

DICHOTOME PRÄDIKTOREN

- Gruppierungsvariable wird in eine Dummy-Variablen recodiert (0 = Merkmal nicht vorhanden, 1 = Merkmal vorhanden)
- Dummy-Variablen werden in das Regressionsmodell aufgenommen:
- Referenzgruppe (0), ergibt (Intercept = Mittelwert der Kontrollgruppe) $Y_i = b_0 + b_1 X_i + \epsilon_i$
- als Differenz zwischen der Gruppe 1 und der Referenzgruppe $X_i = 1 \Rightarrow Y_i = b_0 + b_1$
- t-Wert (B/SE(B)) des Regressionskoeffizienten exakt wie beim t-Test

ALTERNATIVE CODIERUNG FÜR X

- Dummy-Codierung als meistverbreitete, aber nicht einzige Art, kategorielle Variablen abzubilden
- wichtig ist nur, dass
 - unterschiedliche Gruppen unterschiedliche Zahlenwerte erhalten
 - klar ist, was ein Unterschied von 1 bedeutet (Interpretation Regressionskoeffizient)
 - der Intercept sinnvoll interpretierbar ist
- Beispiel: **simple coding** mit -0.5 und 0.5 bei zwei Gruppen (vgl. Zentrierung) -
ist der Gesamtmittelwert, wieder der Unterschied zwischen den Gruppen
 b_0 b_1

MEHR ALS ZWEI MITTELWERTE VERGLEICHEN

- traditionell die einfaktorielle Varianzanalyse (ANOVA) als Standardauswertung
- F-Test der Varianzanalyse prüft die Nullhypothese, dass *alle* Mittelwerte gleich sind (keine Unterschiede zwischen den Gruppen)
- Alternativhypothese (“Irgendwelche Gruppen unterscheiden sich in ”) oft theoretisch sehr unbefriedigend Y
- weil der F-Test nicht sagt, *welche* Gruppen sich im Mittelwert von Y signifikant unterscheiden, ist oft ein zweiter Analyseschritt nötig
- Post-hoc-Tests oder Kontraste, d.h. (ausgewählte oder alle) paarweisen Vergleiche zwischen zwei Gruppen

GLM MIT MEHR ALS ZWEI GRUPPEN

- um Gruppen zu vergleichen, werden Prädiktor-Variablen erstellt (vor oder automatisch während der Analyse) $k - 1$
- die Variablen werden in das Regressionsmodell aufgenommen: .
- in der Referenzgruppe (alle $k - 1$) ergibt sich (Intercept = Mittelwert der Referenzgruppe)
$$Y_i = b_0 + b_1 X_1 + b_2 X_2$$
$$X_1 = X_2 = 0 \Rightarrow Y = b_0$$
- gibt (bei Dummy-Codierung) die Differenz zwischen der Gruppe 1 und der Referenzgruppe wider, die Differenz zwischen der Gruppe 2 und der Referenzgruppe, etc. b_2
- man kann mehrere, aber **nicht alle** paarweisen Vergleiche gleichzeitig modellieren

DUMMY-CODIERUNG 4 GRUPPEN

Gruppe A als Referenz

Zugehörigkeit	Gruppe B	Gruppe C	Gruppe D
Gruppe A	0	0	0
Gruppe B	1	0	0
Gruppe C	0	1	0
Gruppe D	0	0	1

Gruppe D als Referenz

Zugehörigkeit	Gruppe A	Gruppe B	Gruppe C
Gruppe D	0	0	0
Gruppe A	1	0	0
Gruppe B	0	1	0
Gruppe C	0	0	1

KONTRASTE DURCH GEZIELTE CODIERUNG

- spezifische Kontraste durch verschiedene Codierungen für die Prädiktorvariablen (vgl. Davis, 2010)
- einfache Alternative: Referenzgruppe ändern, Modell neu schätzen
- zahlreiche z.T. komplexe Codierungsverfahren, um z.B. ordinale Gruppenvariablen abzubilden
- Beispiel: **Helmert coding**, bei dem eine Gruppe mit jeweils allen nachfolgenden Gruppen verglichen werden,
 1. Kontrollgruppe mit allen Treatments
 2. Treatment 1 mit Treatment 2, etc.

POST-HOC-TESTS

- für kategoriale Prädiktoren kann man anhand des geschätzten Modells alle Mittelwerte paarweise vergleichen
- entspricht separaten T-Tests mit je zwei Gruppen
- aufgrund der Vielzahl einzelner Tests erhöht sich die Gefahr von Alpha-Fehlern (d.h. irrtümlich signifikante Ergebnisse)
- p-Werte (manchmal auch die CI) sollten daher korrigiert werden sollten (vgl. Bender & Lange, 2001)
- verschiedenste Korrekturverfahren möglich (Bonferroni, Hochberg), eines sollte gewählt werden

GLM VS. T-TEST/ANOVA

Vorteile

- keine unterschiedliche Nomenklatur und Testverfahren, egal ob 2 oder mehr Gruppen
- -Koeffizienten sind direkt als Mittelwertdifferenzen zwischen Gruppen interpretierbar, d.h. oft sind Post-Hoc-Tests unnötig
- beliebig erweiterbar durch weitere kategorielle und metrische Prädiktoren
- oft wird der globale F-Test sowie das R^2 als Effektstärkemaß zusätzlich ausgegeben

Nachteile

- Konvention und Fachgeschichte, d.h. GutachterInnen erwarten ANOVA oder t-Test
- Dummy-Codierung (oder andere Effekt-Codierungen) machen ggf. Zusatzaufwand

BEISPIELSTUDIE: KÜMPEL (2019)

Coming across news on social network sites (SNS) largely depends on news-related activities in one's network. Although there are many different ways to stumble upon news, limited research has been conducted on how distinct news curation practices influence users' intention to consume encountered content. In this mixed-methods investigation, using Facebook as an example, we first examine the results of an experiment (study 1, n = 524), showing that getting tagged in comments to news posts promotes news consumption the most.

DATEN

modus	rw	modus_tag
Tag	5	1
Chronik	2	0
Post	3	0
DM	1	0
Chronik	1	0
Chronik	2	0

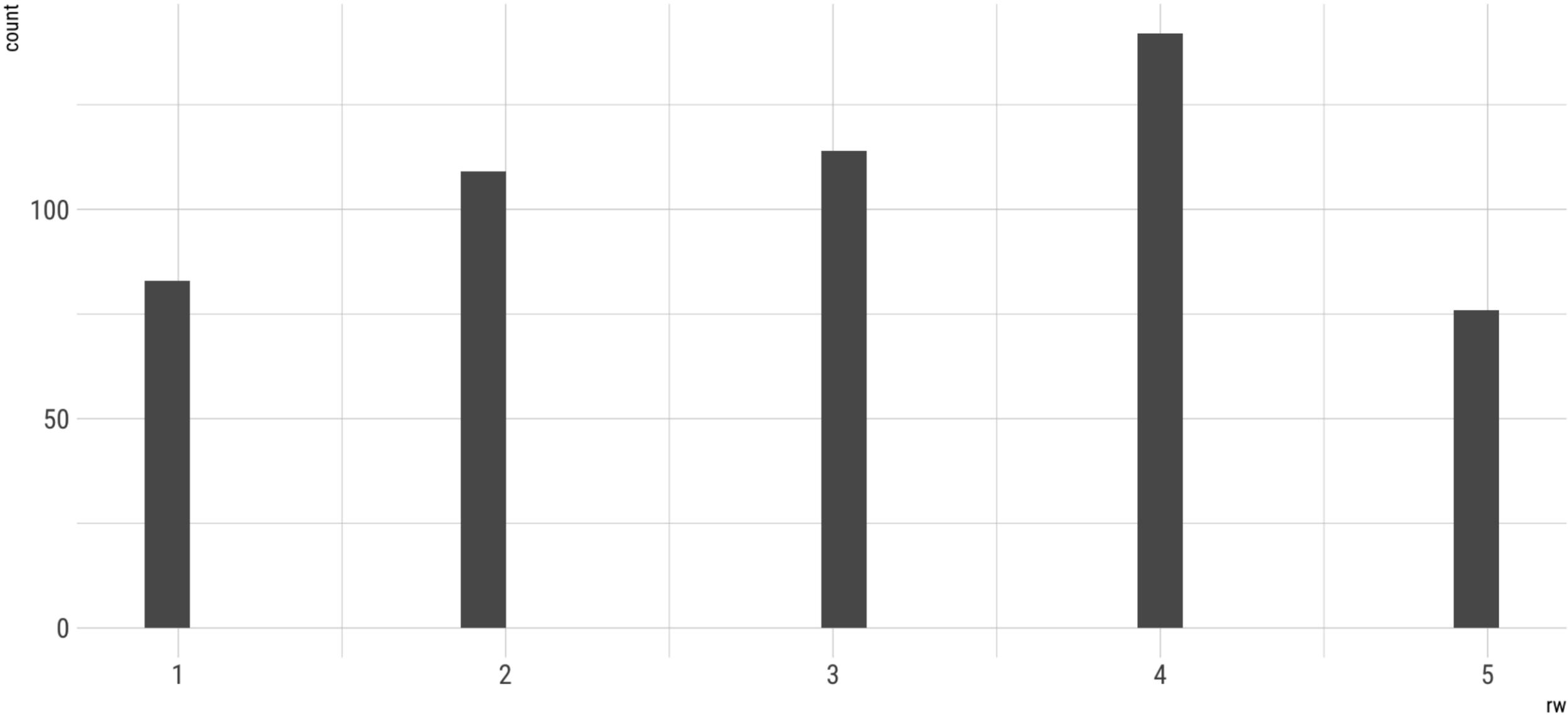
Outcome-Variable

Variable	Summary
Mean rw (SD)	3.04 (1.30)

Gruppenmittelwerte

modus	n	M	SD
Chronik	141	2.88	1.20
Post	97	2.79	1.25
Tag	152	3.51	1.33
DM	134	2.84	1.28

OUTCOME-VARIABLE



T-TEST

Difference	95% CI	t(522)	p	d
-0.67	(-0.91, -0.43)	-5.51	< .001	-0.48

GLM MIT ZWEI GRUPPEN

Parameter	Coefficient	95% CI	t(522)	p	Std. Coef.	Fit
(Intercept)	2.84	(2.71, 2.97)	43.26	< .001	0.00	
modus tag	0.67	(0.43, 0.91)	5.51	< .001	0.23	
AICc						1738.89
R2						0.05
R2 (adj.)						0.05
Sigma						1.27

ANOVA MIT VIER GRUPPEN

Parameter	Sum_Squares	df	Mean_Square	F	p	Eta2
modus	49.12	3	16.37	10.17	< .001	0.06
Residuals	837.19	520	1.61			

GLM MIT VIER GRUPPEN

Parameter	Coefficient	95% CI	t(520)	p	Std. Coef.	Fit
(Intercept)	2.88	(2.67, 3.09)	26.95	< .001	-0.12	
modus (Post)	-0.09	(-0.41, 0.24)	-0.51	0.609	-0.07	
modus (Tag)	0.63	(0.34, 0.93)	4.27	< .001	0.49	
modus (DM)	-0.04	(-0.34, 0.26)	-0.28	0.776	-0.03	
AICc						1742.69
R2						0.06
R2 (adj.)						0.05
Sigma						1.27

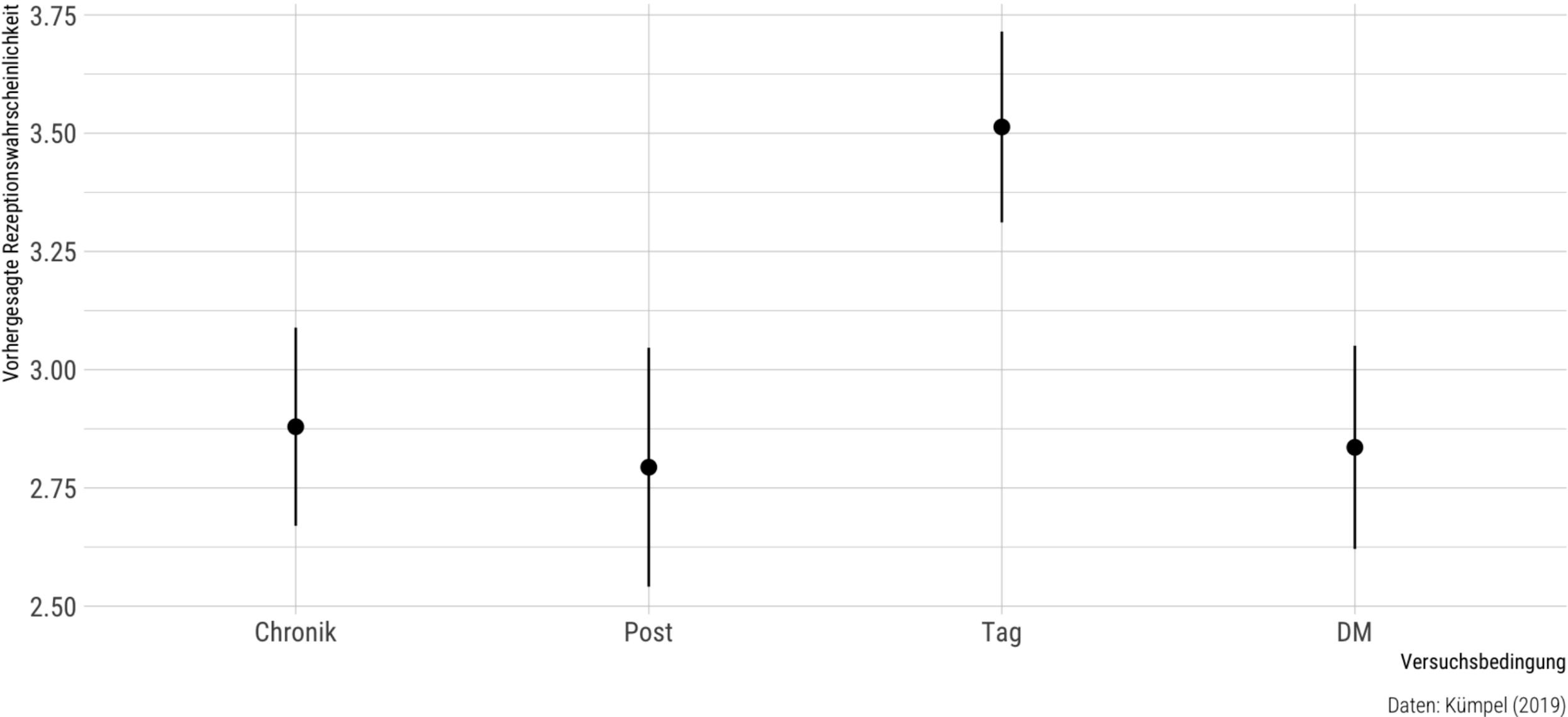
REFERENZKATEGORIE ÄNDERN

Parameter	Coefficient	95% CI	t(520)	p	Std. Coef.	Fit
(Intercept)	2.84	(2.62, 3.05)	25.87	< .001	-0.15	
modus dm (Chronik)	0.04	(-0.26, 0.34)	0.28	0.776	0.03	
modus dm (Post)	-0.04	(-0.37, 0.29)	-0.25	0.804	-0.03	
modus dm (Tag)	0.68	(0.38, 0.97)	4.50	< .001	0.52	
AICc						1742.69
R2						0.06
R2 (adj.)						0.05
Sigma						1.27

POST-HOC/KONTRASTE

term	contrast	estimate	std.error	statistic	p.value	s.value
modus	DM - Chronik	-0.04	0.15	-0.28	1	0.00
modus	DM - Post	0.04	0.17	0.25	1	0.00
modus	DM - Tag	-0.68	0.15	-4.50	0	14.62
modus	Post - Chronik	-0.09	0.17	-0.51	1	0.00
modus	Tag - Chronik	0.63	0.15	4.27	0	13.07
modus	Tag - Post	0.72	0.16	4.36	0	13.66

VISUALISIERUNGSVORSCHLAG



Fragen?

LITERATUR

Bender, R., & Lange, S. (2001). Adjusting for multiple testing—when and how?. *Journal of clinical epidemiology*, 54(4), 343-349.

Davis, M. J. (2010). Contrast coding in multiple regression analysis: Strengths, weaknesses, and utility of popular coding structures. *Journal of data science*, 8(1), 61-73.

Kümpel, A. S. (2019). Getting tagged, getting involved with news? A mixed-methods investigation of the effects and motives of news-related tagging activities on social network sites. *Journal of Communication*, 69(4), 373-395.

TAKE-HOME AUFGABE #1

Wir vergleichen die Tanzbarkeit (danceability) und musikalische Stimmung (valence) der Top 10-Hits über 4 Dekaden (1990er bis 2020er) auf Basis von Billboard und Spotify-Daten.

Beide Variablen sind von 0 (niedrig) - 100 (hoch) skaliert. Die Mittelwerte und Fallzahlen pro Dekade sind wie folgt:

decade	danceability	valence	n
1990s	64.72	56.09	588
2000s	67.34	57.98	558
2010s	67.31	51.93	499
2020s	66.10	51.38	69

TAKE-HOME AUFGABE #1

1. Interpretieren sie die Ergebnisse der **beiden** linearen Modelle, in denen die Mittelwertunterschiede getestet werden, Zeile für Zeile.
2. Welche Dekaden werden *nicht* miteinander verglichen, d.h. für diese bräuchten wir Post-Hoc Vergleiche?

Lösung bitte bis 29.05.2024, 12 Uhr in Moodle eintragen.

danceability			
<i>Predictors</i>	<i>Coefficient (B)</i>	<i>SE (B)</i>	<i>p</i>
(Intercept)	64.72	0.59	<0.001
decade [2000s]	2.62	0.85	0.002
decade [2010s]	2.59	0.88	0.003
decade [2020s]	1.38	1.83	0.452

valence		
<i>Predictors</i>	<i>Coefficient (B)</i>	<i>95% CI (B)</i>
(Intercept)	51.38	45.97 – 56.79
decade [1990s]	4.71	-1.01 – 10.43
decade [2000s]	6.60	0.87 – 12.34
decade [2010s]	0.56	-5.22 – 6.33