

Anwendungsorientierte Analyseverfahren

Modellvorhersagen und -visualisierung

Prof. Dr. Michael Scharkow

Sommersemester 2024

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?

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

Fragen zur praktischen Übung?

MODELLVORHERSAGEN

- Modellvorhersagen helfen, komplexe Modelle besser verstehen zu können
- durch Einsetzen von Werten für die Prädiktoren in die Regressionsgleichung können wir Werte für Y vorhersagen, d.h. \hat{Y}
- es können real existierende oder fiktive Daten eingesetzt werden
- auf Grund der statistischen Unsicherheit in den Regressionskoeffizienten sind auch die Vorhersagen mit Unsicherheit behaftet
- daher erhalten wir Punktschätzer und Konfidenz bzw. Vorhersageintervalle für \hat{Y}

WELCHE DATEN VORHERSAGEN?

- *empirische* Daten, d.h. alle oder ausgewählte Fälle des Datensatzes, auf dem das Modell basiert
- *idealtypische* Daten, d.h. Beispieldaten, die auf (Kombinationen von) für uns relevanten Variablen basieren
- *counterfactual* Daten, d.h. nicht beobachtete Daten, die das Gegenteil der beobachteten in einer oder mehreren Variablen sind

- bei kategoriellen Prädiktoren werden die einzelnen Ausprägungen verwendet
- bei metrischen Prädiktoren werden typische Fälle (Min, Max, Median, Quartile) oder gezielte Einzelwerte eingesetzt

AGGREGATION

- oft sind neben einzelnen Vorhersagen auch aggregierte Vorhersagen für spezifische Gruppen von Interesse
- entweder (a) empirisch vorkommende Gruppen im Datensatz oder (b) kontrafaktische Gruppen
- Analysestrategie:
 1. für jede Gruppe je einen Datensatz auswählen (a) oder (b) generieren
 2. Vorhersagen für alle Fälle pro Datensatz berechnen
 3. Vorhersagen aggregieren, z.B. durch Berechnen des Mittelwertes für \hat{Y}

INTERVALLE

- bei Modellvorhersagen unterscheidet man zwischen *confidence* und *prediction intervals* für \hat{Y}
- in die Berechnung der Konfidenzintervalle fließt nur die Unsicherheit in den Regressionskoeffizienten ein
- in die Berechnung der Vorhersageintervalle fließt *zusätzlich* noch die Residualvarianz ein
- Vorhersageintervalle sind daher immer breiter (je nach R^2) als die Konfidenzintervalle der Vorhersagen
- Vorhersageintervalle werden meist nur für einzelne vorhergesagte Werte angegeben, ansonsten verwenden wir nur CI

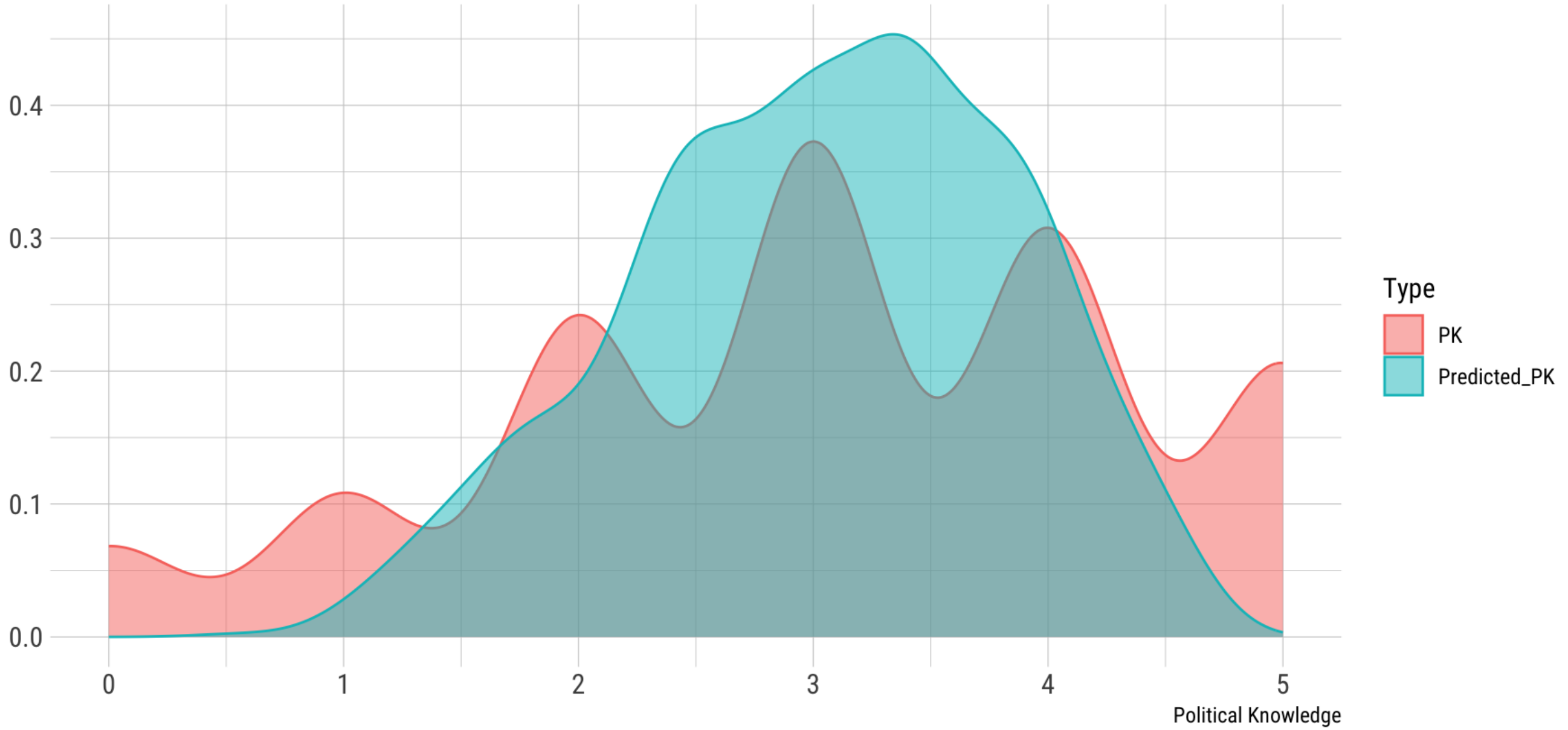
BEISPIEL: VAN ERKEL & VAN AELST, 2021

Parameter	Coefficient	95% CI	t(987)	p	Std. Coef.	Fit
(Intercept)	0.46	(0.09, 0.83)	2.45	0.014	-0.13	
Gender (female)	-0.51	(-0.65, -0.37)	-6.96	< .001	-0.37	
Age	0.02	(0.02, 0.03)	8.48	< .001	0.23	
Education (Middle)	0.35	(0.13, 0.56)	3.15	0.002	0.26	
Education (High)	0.60	(0.38, 0.82)	5.42	< .001	0.44	
Political interest	0.20	(0.18, 0.23)	14.76	< .001	0.40	
AICc						3021.49
R2						0.35
R2 (adj.)						0.34
Sigma						1.10

MODELLVORHERSAGEN FÜR DIE STICHPROBE

Gender	Age	Education	Political_interest	PK	Predicted_PK
female	45	Middle	3	2	1.91
female	59	High	7	4	3.29
female	52	High	7	4	3.13
female	23	High	4	1	1.88
female	23	High	3	1	1.67
female	36	Middle	0	2	1.10

VORHERGESAGTE VERTEILUNG



CONFIDENCE VS. PREDICTION INTERVALS

Confidence intervals

Gender	Age	Education	Political_interest	PK	fit	lwr	upr
female	45	Middle	3	2	1.91	1.76	2.05
female	59	High	7	4	3.29	3.15	3.42
female	52	High	7	4	3.13	3.01	3.26

Prediction intervals

Gender	Age	Education	Political_interest	PK	fit	lwr	upr
female	45	Middle	3	2	1.91	-0.26	4.08
female	59	High	7	4	3.29	1.12	5.46
female	52	High	7	4	3.13	0.96	5.30

KATEGORIELLE PRÄDIKTOREN: GESCHLECHT

- für jeden Fall im Datensatz wird jeweils jede Ausprägung von Geschlecht einmal eingesetzt (counterfactuals)
- alle anderen Prädiktoren bleiben, wie sie waren

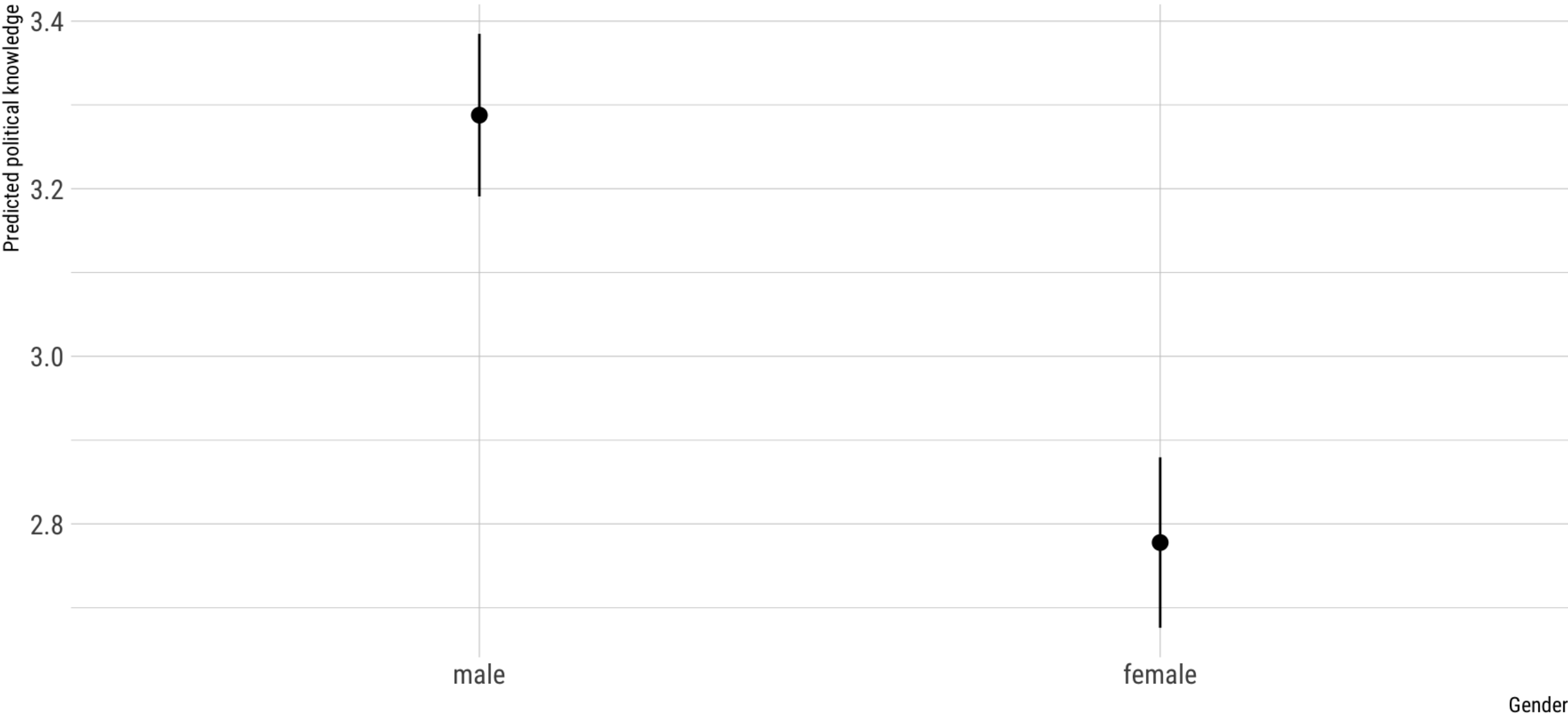
id	Age	Gender	Political_interest	PK	Predicted_PK
1	45	female	3	2	1.91
1	45	male	3	2	2.42
2	59	female	7	4	3.29
2	59	male	7	4	3.80
3	52	female	7	4	3.13
3	52	male	7	4	3.64

AGGREGIERTE VORHERSAGEN NACH GESCHLECHT

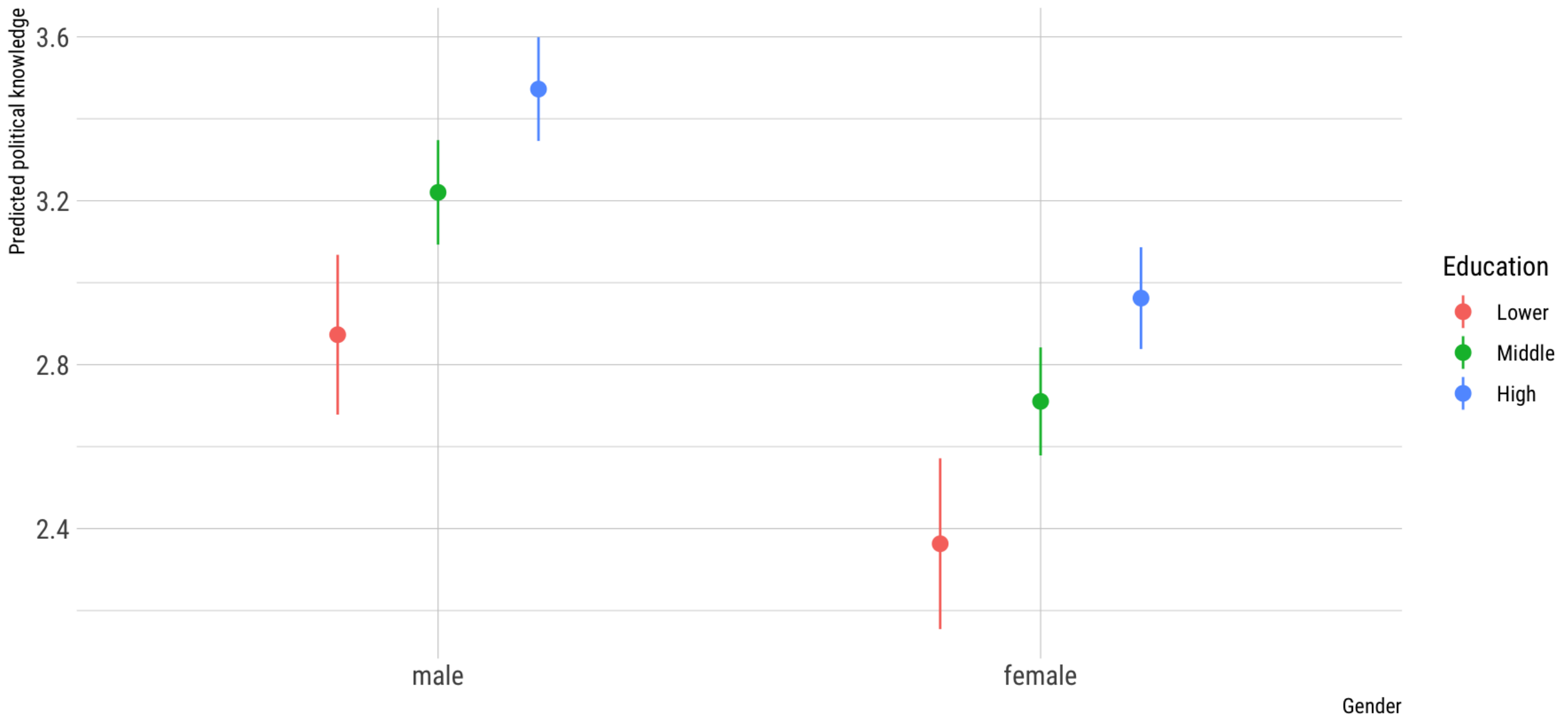
- der neu generierte Datensatz mit den counterfactuals wird nach Geschlecht geteilt
- pro Teildatensatz wird der Mittelwert sowie das CI von \hat{Y} berechnet
- Ergebnis sind die vorhergesagten Mittelwerte des politischen Wissens nach Geschlecht

Gender	estimate	std.error	conf.low	conf.high
female	2.78	0.05	2.68	2.88
male	3.29	0.05	3.19	3.38

VISUALISIERUNG DER VORHERSAGEN



VORHERSAGEN NACH GESCHLECHT UND BILDUNG



METRISCHE PRÄDIKTOREN: ALTER

- wir wählen spezifische Werte (z.B. 18, 40, 65) der Altersvariable
- für jeden Fall wird jeder Alterswert einmal eingesetzt (counterfactuals)
- alle anderen Prädiktoren bleiben, wie sie waren

id	Age	Gender	Political_interest	PK	Predicted_PK
1	18	female	3	2	1.31
1	40	female	3	2	1.80
1	65	female	3	2	2.35
2	18	female	7	4	2.38
2	40	female	7	4	2.87
2	65	female	7	4	3.42

FALLWEISE VORHERSAGEN (TYPISCHE WERTE)

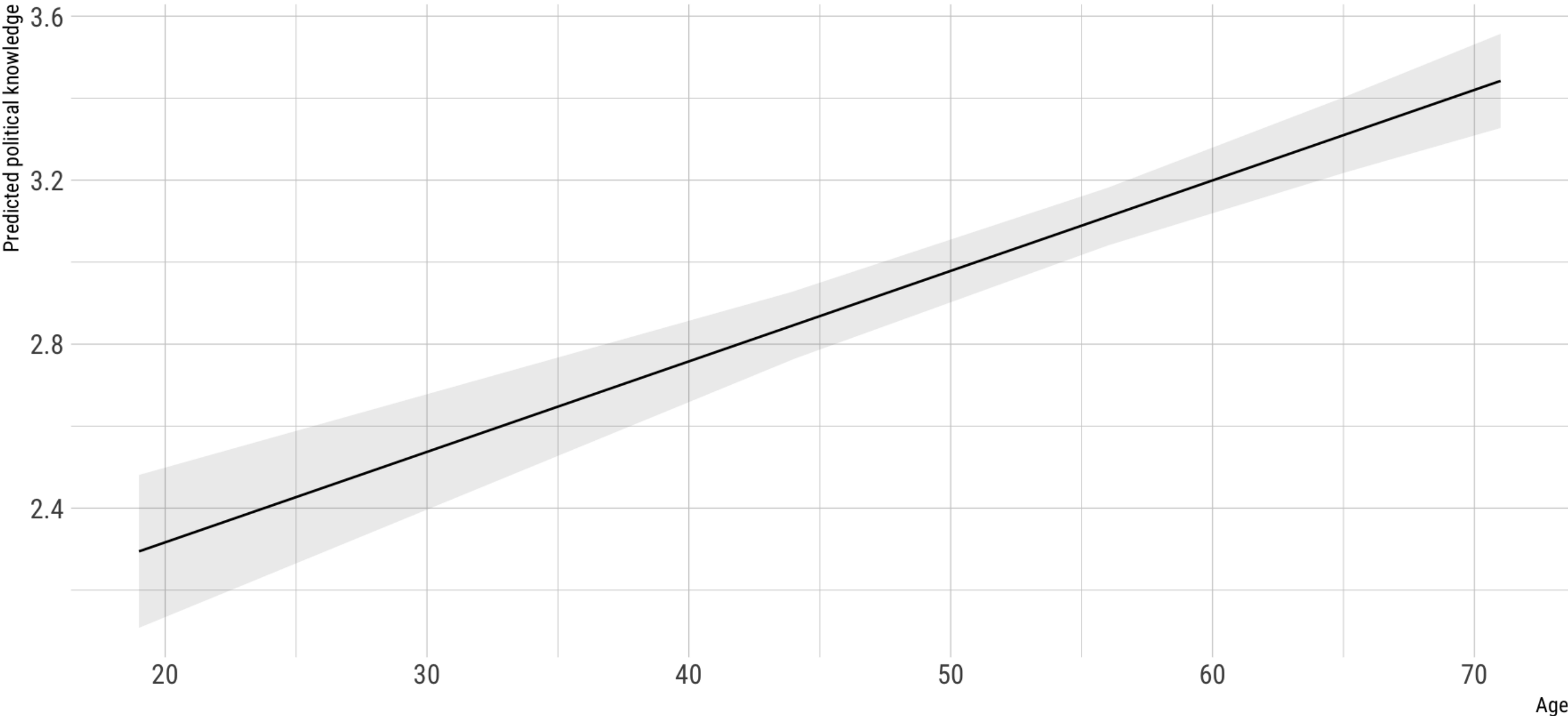
- statt spezifisch ausgewählten Werten verwenden wir Kennwerte
- Five Numbers: Minimum, 1. Quartil, Median, 3. Quartil, Maximum

id	Age	Gender	Political_interest	PK	Predicted_PK
1	19	female	3	2	1.33
1	44	female	3	2	1.89
1	56	female	3	2	2.15
1	65	female	3	2	2.35
1	71	female	3	2	2.48
2	19	female	7	4	2.40
2	44	female	7	4	2.96
2	56	female	7	4	3.22

AGGREGIERTE VORHERSAGEN NACH ALTER

Age	estimate	std.error	conf.low	conf.high
19	2.29	0.10	2.11	2.48
44	2.85	0.04	2.76	2.93
56	3.11	0.04	3.04	3.18
65	3.31	0.05	3.22	3.40
71	3.44	0.06	3.33	3.56

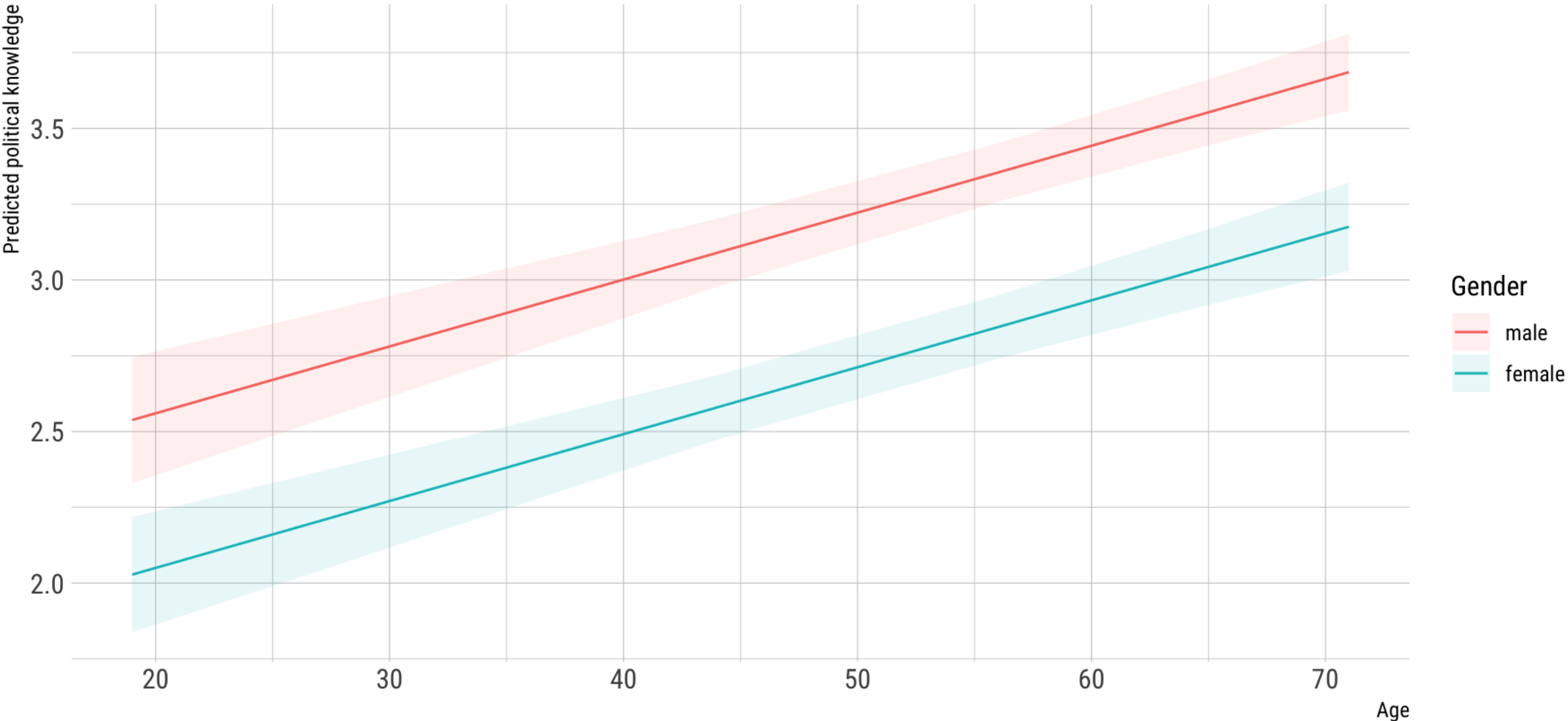
VISUALISIERUNG DER VORHERSAGEN



VORHERSAGEN NACH ALTER UND GESCHLECHT

Age	Gender	estimate	std.error	conf.low	conf.high
19	female	2.03	0.10	1.84	2.22
19	male	2.54	0.11	2.33	2.75
44	female	2.58	0.05	2.47	2.69
44	male	3.09	0.06	2.98	3.20
56	female	2.84	0.05	2.74	2.95
56	male	3.35	0.05	3.26	3.45
65	female	3.04	0.06	2.92	3.17
65	male	3.55	0.06	3.44	3.66
71	female	3.18	0.07	3.03	3.32
71	male	3.69	0.06	3.56	3.81

VISUALISIERUNG DER VORHERSAGEN



FAZIT

- mit Modellvorhersagen lassen sich vielfältige Fragen auf Basis desselben Modells beantworten
- für komplexe, nichtlineare Modelle intuitive(re) Grafiken statt schwer interpretierbarer Koeffizienten
- Wahl der passenden Prädiktorkombinationen nicht trivial (counterfactual, empirisch, typisch)
- subtile Unterschiede in der Interpretation und uneinheitliche Begrifflichkeiten (marginal, conditional, adjusted predictions)
- in der Kommunikationswissenschaft (leider) selten anzutreffen, außer in der Moderationsanalyse

TAKE HOME #2

Replizieren sie eine Regressionsanalyse aus van Erkel & van Aelst (2021) mit R oder SPSS oder anderer Software

- Studierende mit gerader Matrikelnummer: Tabelle 5
- Studierende mit ungerader Matrikelnummer: Tabelle 6
- Der Datensatz ist in [data/VanErkel_vanAelst2021.sav](#) und enthält alle nötigen Variablen.
- Machen sie einen Screenshot der Regressionstabelle als PNG oder JPG und laden Sie in in Moodle hoch.
- Deadline: 19.06.2024