Statistik für Politikwissenschaftler
Klausur Sommersemester 2017

- Die folgende Klausur umfasst 21 Fragen. Bitte beantworten Sie ALLE Fragen.
- Bei den Single-Choice-Fragen ist jeweils nur eine der Antwortmöglichkeiten korrekt. Wollen Sie Ihre Antwort korrigieren, so schwärzen Sie das angekreuzte Kästchen komplett aus und kreuzen Sie die korrekte Antwort an. Bei Mehrfachkorrektur setzen Sie das korrekte Kreuz bitte neben das Kästchen.
- Die anderen Fragen sind mit Antwortkürzeln oder Zahlen zu beantworten. Tragen Sie diese in die vorgesehenen Felder ein. Wollen Sie Ihre Antwort korrigieren, so streichen Sie die falsche Antwort deutlich durch und geben Sie die korrekte Antwort neben oder über dem Antwortfeld an.
- Es gibt keinen Punktabzug für falsche Antworten.
- Insgesamt können 50 Punkte erreicht werden.

Persönliche Angaben:

Name, Vorname: _____

Matrikelnummer: _____

Studiengang: _____

Gesamtpunktzahl: _____ / Note: _____

Auswertungsbogen (nur für Korrigierende)

Aufgabe	erzielte Punkte	max. Punkte
1		2
2		2
3		0
4		2
5		2
6 (Standardabweichung)		6
7		2
8 (Gleichung)		3
9 (Schätzwert)		3
10		2
11		2
12		2
13		2
14		2
15		2
16		2
17		2
18		2
19 (Erw. Häufigkeiten)		4
20		2
21		2

Eine Forschergruppe untersucht die Ergebnisse der Bundestagswahl 2013. Dabei wird unter anderem die Wahlbeteiligung (in Prozent) in den 299 Wahlkreisen betrachtet. Bei einer ersten Auswertung der abhängigen Variable ergibt sich das folgende Bild.

Tabelle 1

. summarize wahlbeteiligung, detail format

wahlbeteiligung				
	Percentiles	Smallest		
1%	60.6	58.9		
5%	63.9	59.4		
10%	65.6	60.6	Obs	299
25%	69.4	61.1	Sum of Wgt.	299
50%	71.9		Mean	71.5
		Largest	Std. Dev.	3.9
75%	74.2	79.0		
90%	76.0	79.1	Variance	15.0
95%	77.5	79.8	Skewness	-0.6
99%	79.1	79.8	Kurtosis	3.4

1. Welche Aussage zu **Tabelle 1** ist korrekt?

- Es nur Lagemaße dargestellt.
- In allen Wahlkreisen lag die Wahlbeteiligung unter 80%.
- Die Wahlbeteiligung lag im Durchschnitt bei 71,9%.
- In 25% der Wahlkreise liegt die Wahlbeteiligung zwischen 61,1% und 69,4%.

2. Bestimmen Sie auf Basis der in **Tabelle 1** enthaltenen Informationen den Interquartilsabstand. (Bitte den Rechenweg angeben.)

3. Welche der folgenden Darstellungsformen ist geeignet, um die Verteilung der Wahlbeteiligung, wie sie sich aus Tabelle 1 ergibt, darzustellen?

- Boxplot
- Kreisdiagramm
- Balkendiagramm
- Streudiagramm

4. Eine mögliche Erklärung für Unterschiede in der Wahlbeteiligung, wie sie in Tabelle 1 dargestellt werden, liegt in der wirtschaftlichen Lage. Konkret wird die folgende Hypothese aufgestellt: Je höher die Arbeitslosenquote in einem Wahlkreis ist, umso geringer ist die Wahlbeteiligung. Wie lautet die dazugehörige Nullhypothese:

5. Abbildung 1 stellt für das Jahr 2012 die Verteilung der Arbeitslosenquote in den 299 Bundestagswahlkreisen dar. Welche der Aussagen bezüglich der Abbildung ist korrekt?

- Die Verteilung ist linkssteil und unimodal.
- Die Verteilung ist rechtsschief und bimodal.
- Die Verteilung ist rechtssteil und unimodal.
- Die Verteilung ist linksschief und bimodal.

6. Die Forscher untersuchten den Einfluss der wirtschaftlichen Lage auf die Wahlbeteiligung zunächst für eine ausgewählte Region in Deutschland. Diese Region setzt sich aus vier Wahlkreisen zusammen. Die Arbeitslosenquote in den einzelnen Wahlkreisen stellt sich wie folgt dar:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Tabelle 2

Wahlkreis	1	2	3	4
Arbeitslosenquote	4 %	10 %	12 %	10 %

Berechnen Sie auf Basis der Werte, die sich in Tabelle 2 finden, die Standardabweichung der Arbeitslosenquote. Verwenden Sie hierzu die folgenden Formeln.

Mittelwert:

Varianz
$$V(x) = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Standardabweichung:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Die Forschergruppe hat weitere Variablen identifiziert, die möglicherweise die Höhe der Wahlbeteiligung in den Wahlkreisen beeinflussen. In einem ersten Analyseschritt wird mit Hilfe einer multiplen linearen Regression der Einfluss der folgenden Variablen auf die Höhe der Wahlbeteiligung untersucht:

- *arbeitslosenquote*: Arbeitslosenquote in Prozent
- *ostdeutschland*: Dummy-Variable, unterscheidet ostdeutsche Bundesländer (1) von westdeutschen Bundesländern (0); Anmerkung: fehlende Werte für Berliner Wahlkreise
- *senioren*: prozentualer Anteil von Personen, die mindestens 60 Jahre alt sind, an der Gesamtbevölkerung

Die Analyse liefert das folgende Ergebnis:

Tabelle 3

. regress wahlbeteiligung arbeitslosenquote ostdeutschland senioren

Source	SS	df	MS			
Model	1388.13731	3	462.712437	Number of obs =	287	
Residual	2878.5026	283	10.1713873	F(3, 283) =	45.49	
				Prob > F =	0.0000	
				R-squared =	0.3253	
				Adj R-squared =	0.3182	
				Root MSE =	3.1893	
Total	4266.63991	286	14.9183214			

wahlbeteiligung	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
arbeitslosenquote	-0.4	0.1	-4.4	0.000	-0.5	-0.2
ostdeutschland	-3.5	0.7	-5.3	0.000	-4.8	-2.2
senioren	-0.1	0.1	-0.7	0.505	-0.2	0.1
_cons	76.2	2.4	31.8	0.000	71.5	80.9

7. Welche der folgenden Aussagen zum dargestellten Regressionsmodell ist **korrekt**?

- Der durchschnittliche Schätzfehler ist etwa 3.19 % groß.
- Alle Variablen üben einen signifikanten Effekt auf die Höhe der Wahlbeteiligung aus.
- Etwa 32,53 % der beobachteten Werte werden auf Basis des Modells korrekt vorhergesagt.
- Etwa 45,49 % der beobachteten Werte werden auf Basis des Modells korrekt vorhergesagt.

8. Stellen Sie auf Basis der Ergebnisse der Regressionsanalyse (**Tabelle 3**) die Regressionsgleichung auf, indem Sie die entsprechenden Werte einsetzen.

Hinweis: Für die Bezeichnung der unabhängigen Variablen können Sie die folgenden Abkürzungen verwenden: Arbeitslosenquote: ALQ; Ostdeutschland: OST; Senioren: SEN.

9. Berechnen Sie auf Basis der Ergebnisse der Regressionanalyse (**Tabelle 3**) - bzw. mit Hilfe der von Ihnen aufgestellten Regressionsgleichung - die geschätzte Wahlbeteiligung für einen Wahlkreis in Mecklenburg-Vorpommern, der die folgenden Merkmale aufweist.

10. Bei der Vorhersage auf Basis der Regressionsanalyse (**Tabelle 3**) ergibt sich für einen Wahlkreis ein Residuum von 15. Was bedeutet das?

- Die Wahlbeteiligung in diesem Wahlkreis lag bei 15%.
- Die Wahlbeteiligung in diesem Wahlkreis war um 15 Einheiten größer als der vorhergesagte Wert.
- Die Wahlbeteiligung in diesem Wahlkreis war um 15 Einheiten größer als der Mittelwert.
- Die Wahlbeteiligung in diesem Wahlkreis war um 15 Einheiten größer als die Konstante.

11. In einer späteren Analyse zeigt sich, dass in ostdeutschen Wahlkreisen die Wahlbeteiligung umso niedriger ist, je höher der Anteil der Senioren ist. Für westdeutsche Wahlkreise findet sich dieser Zusammenhang nicht. Wie wird ein derartiges Zusammenwirken von unabhängigen Variablen genannt?

- Transformation.
- Iteration
- Ökologischer Fehlschluss
- Interaktion

12. Damit eine lineare Regression durchgeführt werden darf, müssen bestimmte Annahmen erfüllt sein. Unter anderem darf keine hohe Multikollinearität vorliegen. Wann liegt diese vor?

- Wenn die abhängige Variable nur zwei Werte annehmen kann.
- Wenn unabhängige Variablen sehr stark miteinander korrelieren.
- Wenn sich das Modell nicht mit einer linearen Regression erfassen lässt.
- Wenn die Residuen nicht normalverteilt sind.

13. Bei einem Einstellungstest sind mehrere Puzzles schnellstmöglich zu lösen. Bei dem sind die Werte normalverteilt, der Mittelwert beträgt 140 Sekunden, die Standardabweichung sei 15 Sek. Wie viele Messwerte liegen dann zwischen 110 und 170 Sekunden?

- etwa 68%
- etwa 60%
- etwa 95%
- etwa 99%

Ausgehend von der Analyse der Ergebnisse der Bundestagswahl 2013 in den Wahlkreisen, führt die Forschergruppe eine weitere Analyse durch. In einem zweiten Schritt werden Umfragedaten aus dem Jahr 2014 genutzt, um zu untersuchen, welche Faktoren die Wahrscheinlichkeit der Beteiligung an der Bundestagswahl erhöhen.

Die abhängige Variable basiert auf der Frage, ob die/der Befragte bei der vorangegangenen Bundestagswahl teilgenommen hat. Die Dummy-Variable nimmt den Wert 1 an, wenn Befragte mit „Ja“ geantwortet haben.

Mit Hilfe einer logistischen Regression wird der Einfluss der folgenden Variablen auf die individuelle Wahlbeteiligung untersucht:

- *arbeitslos*: Dummy-Variable, unterscheidet Befragte, die in den letzten 10 Jahren arbeitslos waren (1) von Befragten, für die das nicht gilt (0)
- *ostdeutschland*: Dummy-Variable, unterscheidet Befragte, die in ostdeutschen Bundesländern leben (1) von Befragten, die in westdeutschen Bundesländern leben (0); fehlende Werte für Befragte, die in Berlin leben
- *senior*: Dummy-Variable, unterscheidet Befragte, die mindestens 60 Jahre alt sind (1) von Befragten, die jünger sind (0)

Die Analyse liefert das folgende Ergebnis:

Tabelle 4

```
. logit wahlbeteiligung_B arbeitslos ostdeutschland senior [iweight=V870
```

Iteration 0: log likelihood = -1326.0451
Iteration 1: log likelihood = -1299.7502
Iteration 2: log likelihood = -1299.034
Iteration 3: log likelihood = -1299.0334
Iteration 4: log likelihood = -1299.0334

Logistic regression

Log likelihood = -1299.0334	Number of obs =	=	3024
	LR chi2(3)	=	54.02
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.0204

wahlbeteiligung_B	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
arbeitslos	-0.4	0.1	-2.5	0.013	-0.6	-0.1
ostdeutschland	-0.5	0.1	-4.2	0.000	-0.7	-0.3
senior	0.6	0.1	4.9	0.000	0.4	0.8
_cons	1.7	0.1	24.5	0.000	1.5	1.8

14. Welche der folgenden Aussagen zum dargestellt Regressionsmodell ist **korrekt**?

- Die Wahrscheinlichkeit sich an der Wahl zu beteiligen, ist bei Befragten in Ostdeutschland um 0,5 Prozentpunkte geringer als bei Befragten in Westdeutschland.
- Valle Variablen üben einen signifikanten Effekt auf die Wahrscheinlichkeit, dass sich ein Befragter an der Wahl beteiligt, aus.
- Etwa 54,02% der beobachteten Werte werden auf Basis des Modells korrekt vorhergesagt.
- Etwa 2,04% der beobachteten Werte werden auf Basis des Modells korrekt vorhergesagt.

15. Werden die Ergebnisse der logistischen Regression (**Tabelle 4**) in Form von odd ratios dargestellt, ergibt sich für die Variable *senior* der Wert 1.8. Wie ist dieser zu interpretieren?
- Der Wert liegt unter dem Schwellenwert von 1,96. Die Variable übt also keinen signifikanten Einfluss auf die Wahrscheinlichkeit wählen zu gehen aus.
 - Der Wert liegt unter dem Schwellenwert von 1,96. Die Variable übt also einen signifikanten Einfluss auf die Wahrscheinlichkeit wählen zu gehen aus.
 - Befragte, die mindestens 60 Jahre alt sind, haben eine 1,8-mal so große Chance wählen zu gehen als jüngere Befragte.
 - Die Wahrscheinlichkeit wählen zu gehen, liegt bei Befragten, die mindestens 60 Jahre alt sind, um 1,8% Prozentpunkte über dem entsprechenden Wert für jüngere Befragte.

16. Wie wird durch den Rückgriff auf Individualdaten (wie bspw. in **Tabelle 4**) statt auf aggregierte Daten (wie bspw. in **Tabelle 3**) verhindert?
- Transformation
 - Iteration
 - Ökologischer Einfluss
 - Interaktion

17. Im Zusammenhang mit Hypothesentests ist von Fehlern erster und zweiter Art die Rede. Wodurch kennzeichnet sich der Fehler der erster Art?
- H_0 trifft in der Grundgesamtheit zu, auf Basis des Tests wird H_0 verworfen.
 - H_0 trifft in der Grundgesamtheit zu, auf Basis des Tests wird H_0 beibehalten.
 - H_1 trifft in der Grundgesamtheit zu, auf Basis des Tests wird H_0 verworfen.
 - H_1 trifft in der Grundgesamt zu, auf Basis des Tests wird H_0 beibehalten.

18. Was wird im Kontext der Schätzung auf Basis von Stichproben unter Erwartungstreue verstanden?
- Der Stichprobenfehler nimmt den Wert 1,96 an.
 - Die Schätzung schätzt den Stichprobenfehler im Durchschnitt richtig.
 - Der Stichprobenfehler entspricht der Standardabweichung in der Grundgesamtheit.
 - Die Schätzung schätzt den Wert in der Grundgesamtheit im Durchschnitt richtig.

19. Die Forscher wählen eine Reihe von Befragten zufällig aus, um den Zusammenhang zwischen Arbeitslosigkeit und der individuellen Wahlbeteiligung genauer zu betrachten. Dabei lassen sich die empirischen Ausprägungskombinationen wie folgt tabellarisch darstellen:

Tabelle 5

	Arbeitslos: Nein	Arbeitslos: Ja	Σ
Wahlbeteiligung: Nein	16	4	20
Wahlbeteiligung: Ja	74	6	80
Σ	90	10	100

Bestimmen Sie auf Basis der Werte, die sich in **Tabelle 5** befinden, die erwarteten Häufigkeiten. Sie können hierzu die folgende Formel verwenden:

Erwartete Häufigkeit:

$$e_{ij} = \frac{f_i \cdot f_j}{n}$$

Hinweis: Verwenden Sie im Folgenden so viele Zeilen, wie Sie benötigen. Ihr Rechenweg muss nachvollziehbar und Ihr Ergebnis deutlich als solches identifizierbar sein. Tragen Sie die ermittelten Werte in die untenstehende Tabelle ein.

Erwartete Häufigkeiten

	Arbeitslos: Nein	Arbeitslos: Ja	Σ
Wahlbeteiligung: Nein			
Wahlbeteiligung: Ja			
Σ			

20. Aus einem Abgleich von Tabelle 5 mit den berechneten Häufigkeiten ergibt sich für χ^2 der Stichprobenkennwert 2,78. Für das 95% - Quantil der entsprechenden χ^2 - Verteilung ergibt sich der Wert 3,84. Wie ist das zu interpretieren?

- In der Grundgesamtheit liegt ein Zusammenhang zwischen der Arbeitslosigkeit und der Wahlbeteiligung vor.
- In der Grundgesamtheit liegt kein Zusammenhang zwischen der Arbeitslosigkeit und der Wahlbeteiligung vor.
- In der Stichprobe haben arbeitslose Befragte eine 2,78-mal so große Chance wählen zu gehen wie andere Befragte. In der Grundgesamtheit ist die Chance 3,84-mal so groß.
- In der Stichprobe haben arbeitslose Befragte eine 3,84-mal so große Chance wählen zu gehen wie andere Befragte. In der Grundgesamtheit ist die Chance 2,78-mal so groß.

21. Die Größe des χ^2 hängt auch vom Format der Tabelle und der Fallzahl ab. Daher wird ergänzend auf folgendes Zusammenhangsmaß zurückgegriffen, das unempfindlich gegen unterschiedliche Tabellenformate und Fallzahlen ist:

- Cramer's V
- Pearsons' r
- Spearman's rho
- Helmdag's J