

# Methoden der Politikwissenschaft

## Sommersemester 2013

### Teil „Einführung in die Statistik“

Die folgende Klausur umfasst 24 Fragen. Bitte bearbeiten Sie ALLE Fragen.

Bei den Multiple Choice Fragen können keine, eine, mehrere oder alle Antwortmöglichkeiten richtig sein. Kreuzen Sie die richtigen Aussagen an.

Die anderen Fragen sind mit Antwortkürzeln oder Zahlen zu beantworten. Tragen Sie diese in die vorgesehenen Felder ein.

Eine korrekt beantwortete Frage gibt zwei oder drei Pluspunkte. Insgesamt können 50 Punkte erreicht werden.

Korrekt beantwortet ist eine Frage nur, wenn Sie die richtigen Antwortmöglichkeiten vollständig angegeben haben. Bei den Multiple Choice Fragen dürfen Sie also weder zu wenige noch zu viele Antwortmöglichkeiten angekreuzt haben. Bei den übrigen Fragen gilt analog dazu, dass alle Antworten richtig gegeben sein müssen.

Andernfalls gilt eine Frage als falsch beantwortet und wird mit 0 Punkten bewertet.

Minuspunkte werden nicht vergeben.

**VIEL ERFOLG!**

#### Persönliche Angaben:

Name, Vorname: \_\_\_\_\_

Matrikelnummer: \_\_\_\_\_

Studiengang: \_\_\_\_\_

**1. Welche Funktion erfüllt Statistik in den Sozialwissenschaften? (2 Punkte)**

- ( X ) Statistik dient der Informationsverdichtung.
- ( X ) Statistik ermöglicht das Aufdecken von Gesetz- oder Regelmäßigkeiten, so dass wir Zusammenhänge erkennen und Aussagen über Ursache und Wirkung treffen können.
- ( F ) Statistik führt zur Verkomplizierung, da sich probabilistische Gesetzmäßigkeiten nur noch numerisch ausdrücken lassen.
- ( X ) Für die Sozialwissenschaften ist die Statistik eine Hilfswissenschaft.

**2. Für eine statistische Analyse wird eine Reihe von Merkmalen erfasst. Notieren Sie jeweils mit N, O, I bzw. R ob es sich um eine Nominal-, Ordinal-, Intervall-, oder Ratio-Skala handelt. (2 Punkte)**

- ( R ) Das Alter von Befragten
- ( N ) Die regionale Herkunft von Befragten (Ost- oder West-Deutschland)
- ( O ) Die Selbsteinschätzung auf einer Links-Rechts-Skala (von 0 bis 10)
- ( R ) Der Sitzanteil extremer Parteien im Parlament
- ( N ) Die Parteipräferenz eines Wählers/einer Wählerin bei der kommenden Bundestagswahl (SPD, CDU/CSU, Piraten, GRÜNE, FDP, LINKE, Andere)
- ( O ) Das Vertrauen in das Bundesverfassungsgericht auf einer 4er Skala (sehr hoch, hoch, niedrig, sehr niedrig)
- ( R ) Das Nettoeinkommen eines Befragten
- ( I ) Datum des Amtsantritts eines Präsidenten (bspw. Barack Obama am 20. Januar 2009)

**3. Welche Aussagen zu den Skalenniveaus sind korrekt? (2 Punkte)**

- ( X ) Nominalskalen ermöglichen es, „gleich“ und „ungleich“ zu unterscheiden.
- ( X ) Intervallskalen haben keinen natürlichen Nullpunkt.
- ( X ) Höhere Skalenniveaus enthalten auch immer die Informationen von niedrigeren Skalenniveaus.
- ( F ) Bei intervallskalierten Daten dürfen die Werte nicht mittels Addition und/oder Subtraktion transformiert werden.

**4. Mit einem Histogramm wird Folgendes dargestellt (2 Punkte):**

- ( F ) Häufigkeit von stetigen ordinalen Merkmalen
- ( F ) Die Signifikanz einer bivariaten Verteilung
- ( X ) Häufigkeit von stetigen metrischen Variablen
- ( X ) Die Schiefe von Verteilungen

**5. Welche Aussagen zu Lage- und Streuungsparametern sind richtig? (2 Punkte)**

- ( X ) Die quadrierten Abstände sind um den arithmetischen Mittelwert herum minimal.
- ( X ) Der Median teilt eine Verteilung in zwei gleich große Hälften.
- ( F ) Modus, Median und geometrisches Mittel lassen sich für nominalskalierte Daten berechnen.
- ( X ) Streuungsmaße geben Auskunft über die Unterschiedlichkeit / Variation der Werte in der Datenmenge.

**6. Welche der folgenden Aussagen sind richtig? (2 Punkte)**

- ( X ) Bei einer Standardnormalverteilung befinden sich Modus, Median und arithmetisches Mittel auf dem gleichen Erwartungswert.
- ( F ) Der Median ist sehr anfällig gegenüber Ausreißern und Extremwerten.
- ( F ) Bei einer rechtssteilen Verteilung befindet sich der Median rechts vom Modus.
- ( X ) Bei einer linksschiefen Verteilung befindet sich der Median links vom Modus.

**7. Welche Aussagen zur deskriptiven und induktiven Statistik sind richtig? (2 Punkte)**

- ( X ) Die deskriptive Statistik dient der Beschreibung beobachteter Ausprägungen von einem oder mehreren Merkmalen.
- ( X ) In den Sozialwissenschaften wird die induktive Statistik genutzt, um Hypothesen zu testen.
- ( F ) Deskriptive Statistik und Inferenzstatistik bezeichnen dasselbe.
- ( X ) Induktive Statistik und Inferenzstatistik bezeichnen dasselbe.

**8. Bringen Sie die Schritte eines Hypothesentests in die richtige Reihenfolge, indem Sie diese nummerieren (2 Punkte):**

- ( 3 ) Bestimmung der Irrtumswahrscheinlichkeit
- ( 2 ) Festlegung der Prüfgröße und Verteilung
- ( 4 ) Prüfgröße berechnen
- ( 1 ) Formulierung von Null- und Alternativhypothese
- ( 5 ) Über Annahme der Hypothese entscheiden

**9. Was gilt für statistische Hypothesentests? (2 Punkte)**

- ( F ) Es wird eine Hypothese betrachtet. Mittels einer Prüfstatistik kann das Signifikanzniveau zur Annahme dieser Hypothese bestimmt werden.
- ( F ) Getestet wird stets die Alternativhypothese  $H_1$ .
- ( X ) Wenn der p-Wert größer-gleich alpha ist ( $p \geq \alpha$ ), wird die Nullhypothese  $H_0$  angenommen.
- ( F ) Getestet werden können nur Unterschieds- nicht jedoch Zusammenhangshypothesen.

**10. Was sind alpha- und beta-Fehler? (2 Punkte)**

- ( X ) Der alpha-Fehler bezeichnet die Irrtumswahrscheinlichkeit, die für die Zuverlässigkeit eines statistischen Tests toleriert wird.
- ( F ) Der beta-Fehler entspricht dem Signifikanzniveau.
- ( F ) Der alpha-Fehler ist die fälschliche Ablehnung der Alternativhypothese  $H_1$ .
- ( F ) Der beta-Fehler ist die fälschliche Ablehnung der Nullhypothese  $H_0$ .

**11. Welche Aussagen zu Wahrscheinlichkeitsverteilungen sind richtig? (2 Punkte)**

- ( F ) Wahrscheinlichkeitsverteilungen für diskrete Zufallsvariablen sind Normalverteilungen.
- ( X ) Wahrscheinlichkeitsverteilungen für diskrete Zufallsvariablen sind Binomialverteilungen.
- ( X ) Wahrscheinlichkeitsverteilungen für stetige Zufallsvariablen sind Normalverteilungen.
- ( F ) Wahrscheinlichkeitsverteilungen für diskrete Zufallsvariablen sind Standardnormalverteilungen.

**12. Welche Aussage zum Konfidenzintervall sind richtig? (2 Punkte)**

- ( F ) Ein Konfidenzintervall ist eine Wertebereich, bei dem wir darauf vertrauen können, dass es den tatsächlichen Wert in der Population mit einer 95%igen Wahrscheinlichkeit (der Vertrauenswahrscheinlichkeit) überdeckt.
- ( X ) Für die Bestimmung des Konfidenzintervalls benötigen wir einen Kennwert aus der Stichprobe als Punktschätzer, als Intervallschätzer bestimmte Quantilwerte und den Standardfehler.
- ( F ) Konfidenzintervalle sind dazu da, die Güte einer Regression zu bewerten.
- ( F ) Konfidenzintervalle können nur für Intervallschätzer bestimmt werden.

**13. In einer Studie wurden das Geschlecht von Befragten und deren Teilnahme an der letzten Bundestagswahl erfasst. Welche Aussagen sind richtig? (2 Punkte)**

**Tabelle 1: Kontingenztabelle Wahlbeteiligung und Geschlecht**

	männlich	weiblich
Teilnahme an Wahl	800	700
Nicht-Teilnahme an Wahl	400	350

- ( X ) Es wurden 1200 Männer befragt. Davon gab die Mehrheit an, bei der letzten Bundestagswahl gewählt zu haben.
- ( X ) Die Kreuztabelle zeigt, dass keine statistische Abhängigkeit der beiden Merkmale besteht.
- ( F ) Männer gingen mit einer höheren Wahrscheinlichkeit wählen.
- ( X ) Nach einer Berechnung von  $\chi^2$  ( $\chi^2$ ) würde dieses den Wert von 0 annehmen.

**14. Welche Aussagen zu Zusammenhangsmaßen für dichotome und nominalskalierte Variablen sind richtig? (2 Punkte)**

- ( F ) Die Berechnung der Prozentsatzdifferenz sollte nicht für dichotome Variablen berechnet werden.
- ( X ) Cramér's V und der Kontingenzkoeffizient C sind beides  $\chi^2$ -/ $\chi^2$ -basierte Zusammenhangsmaße.
- ( F ) Lambda ( $\lambda$ ) ist ein Zusammenhangsmaß für nominale Variablen.
- ( X ) Kendall's Tau ( $\tau$ ) ist kein geeignetes Zusammenhangsmaß für dichotome Variablen.

**15. Welche Aussagen zu Zusammenhangsmaßen für ordinalskalierte Variablen sind richtig? (2 Punkte)**

- ( F ) Für ordinale Variablen kann keine Korrelation berechnet werden.
- ( F ) Cramér's V kann auch genutzt werden um ohne Informationsverlust zwischen ordinalskalierten Variablen die Stärke und Richtung des Zusammenhanges zu bestimmen.
- ( X ) Gamma ( $\gamma$ ) und Somer's D basieren beide auf der Logik des Paarvergleichs.
- ( F ) Somer's D ist ein asymmetrisches Zusammenhangsmaß für ordinalskalierte Variablen.

**16. Welche der folgenden Aussagen zu Korrelationen sind korrekt? (2 Punkte)**

- ( F ) Spearman's r basiert auf einem Paarvergleich.
- ( X ) Wenn Korrelation vorliegt, liegt nicht automatisch Kausalität vor.
- ( X ) Spearman's r und Pearson's r können Werte von -1 bis +1 annehmen.
- ( F ) Wenn ein Korrelationskoeffizient einen Wert  $> 0,5$  annimmt, kann von einem signifikanten Zusammenhang ausgegangen werden.

**17. Welche der folgenden Aussagen zur linearen Regression sind richtig? (2 Punkte)**

- ( F ) Ein negatives  $R^2$  deutet auf einen geringen Zusammenhang zwischen den Variablen hin.
- ( X ) Multikollinearität zwischen den unabhängigen Variablen ist bei der Schätzung einer linearen Regression problematisch.
- ( F ) Die Heteroskedastizität gibt die Güte der Regression an.
- ( X ) Dummy-Variablen können als unabhängige Variablen in das lineare Regressionsmodell integriert werden.

**18. Welche Aussagen zu den Prüfstatistiken sind korrekt? (2 Punkte)**

- ( X ) Der t-Test berechnet sich aus dem Koeffizienten dividiert durch seinen Standardfehler.
- ( F ) Bei der Regressionsanalyse wird der F-Test genutzt, um die Signifikanz des Effektes der einzelnen unabhängigen Variablen zu bestimmen.
- ( F ) F-Test und t-Test sind im Prinzip das Gleiche, nur von verschiedenen Personen erfunden.
- ( F ) Der t-Test vergleicht die erklärte Varianz eines Regressionsmodells mit der unerklärten Varianz und gibt somit Auskunft über die Signifikanz des Gesamtregressionsmodells.

**In einem Forschungsprojekt soll untersucht werden, wie sich die Höhe der Wahlbeteiligung bei Parlamentswahlen erklären lässt.**

Untersucht werden insgesamt 308 Parlamentswahlen in 28 Ländern. Für diese wird (als zu erklärendes Phänomen/abhängige Variable) die Wahlbeteiligung in Prozent erhoben.

Die Forscherinnen und Forscher nehmen zum einen an, dass die Wahlbeteiligung durch verschiedene institutionelle Faktoren bzw. Eigenschaften des politischen Systems beeinflusst wird: Als relevante Einflussfaktoren (unabhängige Variablen) werden identifiziert:

1. die Art des Wahlsystems (Verhältniswahl vs. Mehrheitswahl, kodiert 1/0),
2. die Fraktionalisierung des Parteiensystems,
3. die Tatsache, ob Wahlpflicht besteht oder nicht (kodiert 1/0), und
4. die Art des Regierungssystems (Präsidentialismus vs. Parlamentarismus, kodiert 1/0).

Zum anderen vermuten die Forscherinnen und Forscher, dass auch wirtschaftliche Faktoren die Wahlbeteiligung beeinflussen. Sie berücksichtigen daher

5. das Wirtschaftswachstum (Veränderungsrate des BIP) und
6. die Arbeitslosenrate

als weitere relevante Einflussfaktoren (unabhängige Variablen).

**Zunächst wollen die Forscherinnen und Forscher etwas über die Verteilung ihrer abhängigen Variable wissen und berechnen verschiedene univariate deskriptive Statistiken.**

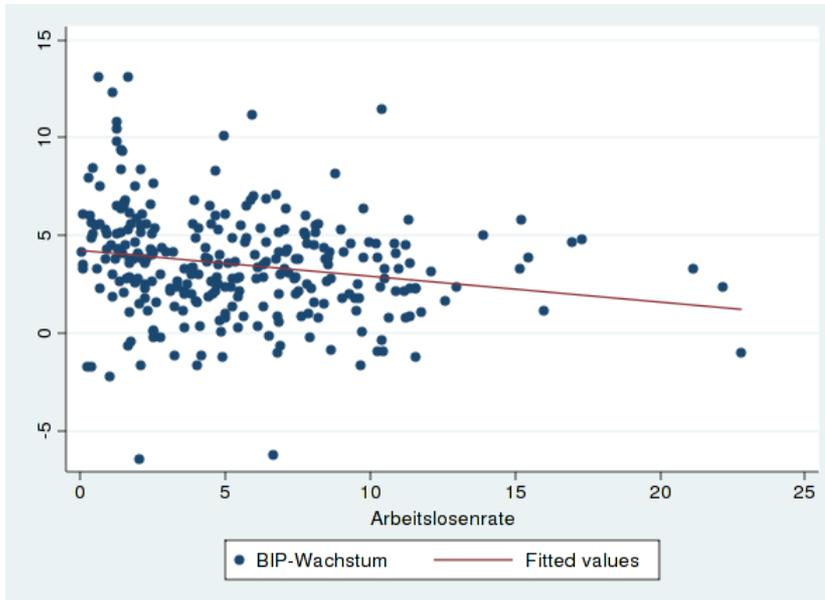
**Tabelle 2: Univariate Deskription**

Variable	N	Arithm. Mittel	Median	Standardabweichung	Minimum	Maximum
<b>Wahlbeteiligung</b>	308	78,95	81,9	13,53	35	95,8

**19. Welche Aussagen zur Tabelle 2: „Univariate Deskription“ sind richtig? (2 Punkte)**

- ( X ) Im Durchschnitt aller untersuchten Fälle liegt die Wahlbeteiligung bei 78,95%.
- ( F ) In das Intervall von 35% bis 95,8% Wahlbeteiligung fallen insgesamt 75% der Fälle.
- ( F ) In 50% der Fälle ist die Wahlbeteiligung höher als 78,95 %.
- ( X ) Die durchschnittliche Abweichung vom Mittelwert beträgt 13,53.

**Dann untersuchen die Forscherinnen und Forscher, wie sich die unabhängigen Variablen zueinander verhalten. Vor allem bei den wirtschaftlichen Faktoren vermuten Sie, dass diese in einem systematischen Zusammenhang miteinander stehen.**



**Grafik 1: Wirtschaftswachstum und Arbeitslosenrate**

**20. Welche Aussagen zur obigen Grafik 1 sind richtig? (2 Punkte)**

( F ) Bei Grafik 1 handelt es sich um ein Histogramm, in dem der Zusammenhang zwischen dem Wirtschaftswachstum und Arbeitslosenrate dargestellt wird.

( F ) Aufgrund der Skalenniveaus hätten Wirtschaftswachstum und Arbeitslosenrate nicht gemeinsam in Grafik 1 dargestellt werden dürfen.

( X ) Die Punkte in Grafik 1 streuen zwar stark. Die Gerade zeigt jedoch: Es kann von einem perfekten linearen Zusammenhang zwischen dem Wirtschaftswachstum und Arbeitslosenrate ausgegangen werden.

( F ) Aufgrund der Darstellung kann eindeutig zwischen unabhängiger und abhängiger Variable unterschieden werden: Die Gerade in Grafik 1 lässt sich daher so interpretieren, dass die Arbeitslosenrate einen leicht negativen kausalen Effekt auf das Wirtschaftswachstum hat.

**Anschließend führen die Forscherinnen und Forscher eine Korrelationsanalyse durch, um die Stärke des Zusammenhangs zwischen Wirtschaftswachstum und Arbeitslosenrate zu bestimmen.**

**Tabelle 3: Korrelation zwischen der Wirtschaftswachstum und Arbeitslosenrate**

	<b>Wirtschaftswachstum</b>	<b>Arbeitslosenrate</b>
<b>Wirtschaftswachstum</b>	1,0000	-0,2168***
<b>Arbeitslosenrate</b>	-0,2168***	1,0000

**21. Welche Aussagen zur obigen Korrelationsanalyse in Tabelle 3 sind richtig? (2 Punkte)**

( X ) Es besteht ein negativer Zusammenhang zwischen der Arbeitslosenrate und dem Wirtschaftswachstum.

( X ) Da der Korrelationskoeffizient nur einen Wert von -0,2168 annimmt und es sich um ein standardisiertes Maß mit einem Wertebereich von -1 bis +1 handelt, ist die Stärke des Zusammenhangs als eher schwach zubezeichnen.

( F ) Da die Korrelation ein negatives Vorzeichen hat, kann jedoch nicht angenommen werden, dass auch ein linearer Zusammenhang zwischen der Arbeitslosenrate und dem Wirtschaftswachstum besteht.

( F ) Da das Ergebnis höchst signifikant ist, kann auch von einem kausalen Effekt ausgegangen werden: Die Arbeitslosenrate beeinflusst das Wirtschaftswachstum.

**Anschließend führen die Forscherinnen und Forscher die Regressionsanalysen zur Erklärung der Wahlbeteiligung bei Parlamentswahlen durch. Sie formulieren zwei verschieden komplexe multivariate Regressionsmodelle, die Sie vergleichen wollen.**

**Tabelle 4: Der Einfluss ausgewählter institutioneller und ökonomischer Faktoren auf die Wahlbeteiligung bei Parlamentswahlen in 28 Ländern, 1960 - 2010.**

<b>Wahlbeteiligung</b>		
	Modell 1	Modell 2
Verhältniswahlsystem ja/nein	2,84* (1,32)	2,90* (1,28)
Fraktionalisierung des Parteiensystems	0,23*** (,007)	0,26*** (,007)
Wahlpflicht ja/nein	10,61*** (1,36)	10,91*** (1,32)
Präsidentialismus ja/nein	-20,27 *** (1,58)	-20,38 *** (1,55)
Arbeitslosigkeit		-0,52*** (0,13)
Wirtschaftswachstum (BIP)		0,25 (0,19)
Konstante	60,98*** (4,65)	60,92*** (4,74)
R <sup>2</sup>	0,523***	0,551***
adjusted R <sup>2</sup>	0,517	0,542
F-Test	83,003	61,665
N	308	308

Unstandardisierte Regressionskoeffizienten mit Standardfehlern in Klammern.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

**22. Stellen Sie die Regressionsformel für die Modell 1 (vgl. Tabelle 4) auf (2 Punkte):**

Modell 1: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Verwenden Sie dafür die folgenden Abkürzungen für die Variablen:

Verhältniswahl: VW

Fraktionalisierung: FP

Wahlpflicht: WP

Präsidentalismus: PR

Arbeitslosigkeit: AL

Wirtschaftswachstum: BIP

**23. Wie ist das Ergebnis der Regressionsanalyse (vgl. Tabelle 4) zu interpretieren? Welche der folgenden Aussagen zu den Regressionskoeffizienten und zur Modellgüte sind richtig? (3 Punkte)**

- ( F ) Auf Basis der obigen Regressionstabelle (mit den unstandardisierten Regressionkoeffizienten) lässt sich feststellen, dass im Vergleich zu den anderen unabhängigen Variablen die Art des Regierungssystems (Präsidentalismus vs. Parlamentarismus) den stärksten Einfluss auf die Wahlbeteiligung hat.
- ( X ) Laut Modell 2 hat eine Zunahme der Arbeitslosigkeit einen negativen Effekt auf die Wahlbeteiligung: Unter der Bedingung, dass die übrigen unabhängigen Variablen konstant gehalten werden, sinkt die Wahlbeteiligung um 5,2 Prozentpunkte, wenn die Arbeitslosigkeit um 10 Prozentpunkte steigt.
- ( X ) Laut Schätzung von Modell 1 liegt die Wahlbeteiligung in einem parlamentarischen (nicht-präsidentiellen) System mit Mehrheitswahlsystem (nicht Verhältniswahlsystem), ohne Fraktionalisierung des Parteiensystems und ohne Wahlpflicht (wenn also alle X-Variablen den Wert null annehmen) bei 60,98%.
- ( X ) Ein Vergleich der adjustierten (korrigierten)  $R^2$  von Modell 1 und Modell 2 zeigt: Die Hinzunahme wirtschaftlicher Faktoren steigert die Modellgüte.

**24. Welche der folgenden weiteren Aussagen zu den Teststatistiken der Regressionsanalyse (vgl. Tabelle 4) sind richtig? (3 Punkte)**

- ( F ) Nur das Modell 1 kann interpretiert werden. Das Modell 2 hingegen darf nicht interpretiert werden, weil eine der unabhängigen Variable keinen signifikanten Effekt auf die Wahlbeteiligung hat.
- ( X ) Eine Steigerung der Fraktionalisierung des Parteiensystems hat zwar nur eine geringe Erhöhung der Wahlbeteiligung um 0,26 (Modell 1) bzw. 0,23 (Modell 2) Prozentpunkte zur Folge (unter der Bedingung, dass die übrigen unabhängigen Variablen konstant gehalten werden). Allerdings ist dieser Effekt höchst signifikant und daher auf die Grundgesamtheit übertragbar.
- ( X ) Aufgrund des F-Tests ist die Nullhypothese, dass keine der gewählten unabhängigen Variablen einen Einfluss auf die Wahlbeteiligung hat, zurückzuweisen. Das Modell ist auf höchst signifikant.
- ( F ) Die Arbeitshypothese, die mit der Teststatistik für die Regressionskoeffizienten getestet wird lautet: Die gewählte unabhängige Variable hat keinen Effekt auf die abhängige Variable.