

Einführung in die Statistik für Politikwissenschaftler

Sommersemester 2013

1. Welche Aussage zur Statistik (in den Sozialwissenschaften) sind richtig? (2 Punkte)

- Statistik ist die Lehre von Methoden zum Umgang mit quantitativen Informationen.
- Statistik wird in den Sozialwissenschaften als Hilfswissenschaft herangezogen.
- Die deskriptive Statistik dient der Beschreibung beobachteter Ausprägungen von einem oder mehreren Merkmalen.
- Die induktive Statistik wird (auch) genutzt um Hypothesen zu testen.

2. Für die 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 Rationalskala handelt. (2 Punkte)

- ___ Das Geburtsjahr
- ___ Das Gefühl, von Parteien politisch präsentiert zu werden, auf einer 5er-Skala (von sehr gut bis sehr schlecht)
- ___ Staatsangehörigkeit von Befragten
- ___ Stimmenanteil einer Partei bei der kommenden Bundestagswahl
- ___ Die Amtszeit einer Regierung in Tagen
- ___ Die Parteienpräferenz eines Wählers/einer Wählerin bei der kommenden Bundestagswahl (SPD, CDU/CSU, Piraten, GRÜNE, FDP, LINKE, Andere)

3. Welche der folgenden Aussagen zu Grafiken sind richtig? (2 Punkte)

- Ordinalskalierte Variablen lassen sich nur in Boxplot darstellen.
- Nominalskalierte Variablen eignen sich aufgrund des fehlenden Nullpunktes nicht für die Darstellung innerhalb eines Balkendiagramms.
- Ein Scatterplot oder Streudiagramm eignet sich, um zwei metrische Variablen grafisch darzustellen.
- Bei einem Boxplot bildet die mittlere Linie in der Box den Median ab, Anfang und Ende der Box die Grenzen von erstem und drittem Quartil. Zudem lassen sich die Extremwerte und Ausreißer darstellen.

4. Welche Aussagen zu Lageparametern sind richtig? (2 Punkte)

- Mittelwerte kennzeichnen die zentrale Lage einer Verteilung.
- Modus, Median und geometrisches Mittel lassen sich für metrischskalierte Daten berechnen.
- In einer Verteilung liegen 50% der Fälle unter und 50% der Fälle über dem Modus.
- Der Median bezeichnet den häufigsten Wert einer Verteilung.

5. Welche Aussagen zu Streuungsmaßen sind richtig? (2 Punkte)

- Als Variationsweite bezeichnet man die Spanne zwischen 0.25- und 0.75- Quartil.
- Für die Berechnung von Spannweite und Quatilsabstand benötigt man nur je zwei Werte.
- Variation, Varianz und Standardabweichung beziehen die Verteilung aller beobachteten Werte um das arithmetische Mittel mit ein.
- Die Standardabweichung ist die Wurzel der Varianz.

6. Bringen Sie die Schritte eines Hypothesentests in die richtige Reihenfolge, indem Sie diese nummerieren. (3 Punkte)

- ___ Festlegung der Prüfgröße und Verteilung
- ___ Über Annahme der Hypothese entscheiden
- ___ Bestimmung der Irrtumswahrscheinlichkeit
- ___ Prüfgröße berechnen
- ___ Formulierung von Null- und Alternativhypothese

7. Was gilt für statistische Hypothesentests? (3 Punkte)

- Getestet wird stets die Nullhypothese. Wenn diese abgelehnt wird, geht man üblicherweise von der Alternativhypothese aus.
- Die fälschliche Ablehnung der Nullhypothese wird in der Statistik als beta-Fehler bezeichnet
- Getestet werden können sowohl Unterschieds- als auch Zusammenhangshypothesen.
- Wenn der p-Wert kleiner alpha ist ($p < \alpha$), wird die Nullhypothese H_0 verworfen.

8. Welche Aussagen zum Konfidenzintervall sind richtig? (3 Punkte)

- Ein Konfidenzintervall ist ein Wertebereich, bei dem wir darauf vertrauen können, dass es den tatsächlichen Wert in der Population mit einer bestimmten Wahrscheinlichkeit (der Vertrauenswahrscheinlichkeit) überdeckt.
- Das Konfidenzintervall ist ein Signifikanztest für Punktschätzer.
- Das Konfidenzintervall ist nicht dasselbe wie das Vertrauensintervall.
- Es ist nicht möglich, Konfidenzintervalle für die Regressionskoeffizienten zu bestimmen.

9. 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

- Insgesamt wurden 2250 Personen befragt. Davon gab die Mehrheit an, bei der letzten Bundestagswahl gewählt zu haben.
- Die Kreuztabelle zeigt, dass eine statistische Abhängigkeit der beiden Merkmale besteht.
- Männer gingen mit einer höheren Wahrscheinlichkeit wählen.
- Nach einer Berechnung von χ^2 würde dieses dem Wert von 0 annehmen.

10. Welche Aussagen zu Zusammenhangsmaßen sind richtig? (3 Punkte)

- Cramér's V kann auch benutzt werden um den Zusammenhang zwischen ordinalskalierten Variablen zu bestimmen. Dann geht dies aber mit einem Informationsverlust einher.
- Kendall's Tau (τ) ist ein geeignetes Zusammenhangsmaß für dichotome Variablen.
- Gamma (γ) basiert auf der Logik des Paarvergleichs, Somer's D hingegen auf der Logik des Rangvergleichs.
- Somer's D ist ein symmetrische Zusammenhangs für ordinalskalierte Variablen.

11. Welche der folgenden Aussagen zu Korrelation sind korrekt? (2 Punkte)

- Wenn ein Korrelationskoeffizient einen Wert $> 0,5$ annimmt, kann davon ausgegangen werden, dass ein kausaler Zusammenhang besteht.
- Spearman's r und Pearson's r können Werte Werte von -1 bis +1 annehmen.
- Wenn $r < 0,05$, kann von einem signifikanten Zusammenhang ausgegangen werden.
- Spearman's r basiert auf dem Rangvergleich.

12. Welche Aussagen zur Regression sind richtig? (3 Punkte)

- Die Ausprägung mehrerer Variablen wird auf die Ausprägung einer unabhängigen Variable zurückgeführt.
- Alle Regressionsmodelle sind linear.
- Bei einer Regressionsanalyse muss kein Signifikanztest durchgeführt werden, weil das Modell an sich schon die Hypothese testet.
- Wenn abhängige und unabhängige Variablen hoch korrelieren, wird dies als Multikollinearität bezeichnet. Nimmt diese eine Wert von $> 0,5$ an, ist die Güte des Regressionsmodells positiv zu bewerten.

13. Was muss bei einer Durchführung einer linearen Regression beachtet werden? (3 Punkte)

- Es sollte Heteroskedastizität vorliegen.
- Die unabhängigen Variablen sollten möglichst hoch, nicht jedoch perfekt miteinander korrelieren.
- Die Beziehung zwischen abhängiger und unabhängiger Variablen soll linear sein.
- Die abhängige Variable muss mindestens ordinal skaliert sein.

14. Welche Aussagen zu Prüfstatistiken sind korrekt? (3 Punkte)

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

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 abhängige Variable) die Wahlbeteiligung in Prozent erhoben. Die Forscherinnen und Forscher nehmen zum einen an, dass die Wahlbeteiligung durch verschiedene institutionelle Faktoren (Eigenschaften des politischen Systems) beeinflusst wird:

1. die Art des Wahlsystems (Verhältnswahl 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:

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

Diese sechs Einflussfaktoren gehen als unabhängige Variablen in die Analyse ein.

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

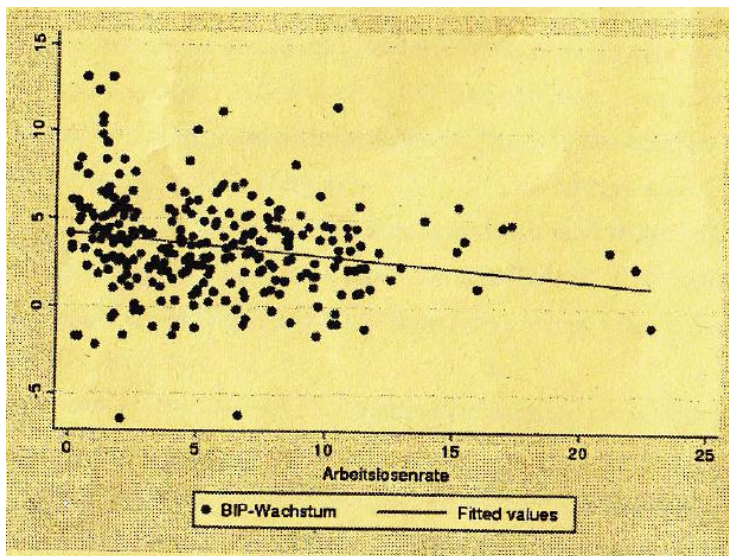
Tabelle 2: Univariate Deskription

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

15. Welche Aussagen zur Tabelle 2: "Univariate Deskription" sind richtig? (2 Punkte)

- () In den meisten Ländern liegt die Wahlbeteiligung bei 78,95%.
- () Die durchschnittliche Wahlbeteiligung liegt bei 81,9%.
- () In das Intervall von 35% bis 95,8% Wahlbeteiligung fallen insgesamt 75% der Fälle.
- () Die durchschnittliche Abweichung vom Mittelwert beträgt 13,53.

Dann untersuchen die Forscherinnen und Forscher zunächst mithilfe einer Grafik, ob zwischen den unabhängigen Variablen ein systematischer Zusammenhang besteht.



Grafik 1: Wirtschaftswachstum und Arbeitslosenrate

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

- () Bei Grafik 1 handelt es sich um ein Streudiagramm, in dem der Zusammenhang zwischen Wirtschaftswachstum und Arbeitslosenrate dargestellt wird.
- () Aufgrund des Skalenniveaus hätte Wirtschaftswachstum und Arbeitslosenrate nicht gemeinsam in Grafik 1 dargestellt werden dürfen.
- () Die Gerade in Grafik 1 stellt einen leicht negativen Zusammenhang zwischen Wirtschaftswachstum und Arbeitslosenrate dar.
- () Die Punkte in Grafik 1 streuen zwar stark. Die Gerade zeigt jedoch: Es kann von einem perfektem linearen Zusammenhang ausgegangen werden.

Dann analysieren die Forscherinnen und Forscher die Stärke des Zusammenhangs zwischen Wirtschaftswachstum und Arbeitslosenrate mithilfe einer Korrelationsanalyse.

Tabelle 3: Korrelation zwischen Wirtschaftswachstum und Arbeitslosenrate

	Wirtschaftswachstum	Arbeitslosenrate
Wirtschaftswachstum	1,0000	-0,2168***
Arbeitslosenrate	-0,2168***	1,0000

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

- () Es besteht ein kausaler Effekt der Arbeitslosenrate auf das Wirtschaftswachstum: Je geringer die Arbeitslosenrate, desto höher das Wirtschaftswachstum.
- () Es besteht ein schwacher negativer Zusammenhang zwischen der Arbeitslosenrate und dem Wirtschaftswachstum.
- () Die Korrelation ist höchst signifikant.
- () Da es sich bei der Korrelationsanalyse um ein symmetrisches Maß handelt, kann nicht zwischen abhängiger unabhängiger Variable unterschieden werden.

Anschließend führen die Forscherinnen und Forscher die Regressionsanalyse zur Erklärung der Wahlbeteiligung bei Parlamentswahlen durch. Sie formulieren zwei verschiedene 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.

die Wahlbeteiligung bei Parlamentswahlen in 28 Ländern, 1960 - 2010.		
Wahlbeteiligung		
	Modell 1	Modell 2
Verhältnisswahlssystem 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 ²	0,523***	0,551***
adjusted R ²	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

18. Stellen Sie die Regressionsformel für Modell 1 (vgl. Tabelle 4) auf. (3 Punkte)

Modell 1:

Verwenden Sie dafür folgende Abkürzungen für die Variablen:

Verhältnisswahl:	VW	Fraktionalisierung:	FP
Wahlpflicht:	WP	Präsidentialismus:	PR
Arbeitslosigkeit:	AL	Wirtschaftswachstum:	BIP

Wie ist das Ergebnis der Regressionsanalyse zu interpretieren?

19. Welche der folgenden Aussagen zu den Regressionskoeffizienten und zur Modellgüte (vgl. Tabelle 4) sind richtig? (3 Punkte)

- Die Effektgrößen können nicht direkt auf Basis der Werte der unstandardisierten Koeffizienten in der Regressionstabelle miteinander verglichen werden.
- Unter der Bedingung, dass die übrigen unabhängigen Variablen konstant gehalten werden, führt Wahlpflicht zur Erhöhung der Wahlbeteiligung um knapp 11 Prozentpunkte.
- Laut Modell 1 liegt die Wahlbeteiligung bei $(60,98 + 2,84 =) 63,82\%$ in einem parlamentarischen System ohne Wahlpflicht und ohne Fraktionalisierung des Parteiensystems, das ein Verhältniswahlssystem hat (wenn also - außer VW = 1 - alle X-Variablen den Wert 0 annehmen).
- An dem adjustierten (korrigierten) R^2 von Modell 2 sieht man, dass die unabhängigen Variablen 54,2% der Varianz der Wahlbeteiligung erklären.

20. Welche der folgenden Aussagen zu Teststatistiken der Regressionsanalysen (vgl. Tabelle 4) sind richtig? (3 Punkte)

- Der Effekt des Wirtschaftswachstums (vgl. Modell 2) ist nach etablierter Konvention in den Sozialwissenschaften nicht auf die Grundgesamtheit übertragbar, da er nicht mindestens auf dem 5%- Niveau signifikant ist.
- Von den institutionellen Variablen hat das Wahlsystem den geringsten Einfluss auf die Wahlbeteiligung, weil der Effekt nur auf dem 5%- Niveau signifikant ist.
- Der F-Test zeigt, dass die Nullhypothese für das Gesamtmodell nicht zurückzuweisen ist. Daher hat das Modell keine Aussagekraft.
- Die Nullhypothese, die mit der Teststatistik für den Regressionskoeffizienten getestet wird, lautet: Die unabhängige Variable hat keinen Effekt auf die abhängige Variable.