

Einführung in die Statistik für Politikwissenschaftler

Wintersemester 2011/2012

Es können von den Antwortmöglichkeiten alle, mehrere, eine oder keine Antwort(en) richtig sein. Nur bei einer korrekten Antwort (ohne Auslassungen und überflüssige Ankreuzungen) erhalten Sie die angegebene Punktzahl! Viel Erfolg!

1. 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 (3 Punkte).

- Das Geburtsjahr des Befragten
- Die Währung eines Landes
- Die durchschnittliche Lebenserwartung von Frauen in einem Land
- Die Stimmenanteil einer Partei in Prozent
- Das Vertrauen in den Bundespräsidenten (sehr hoch, hoch, gering, sehr gering)
- Die Arbeitslosenrate eines Landes
- Die Amtszeit einer Regierung in Tagen
- Das politische System eines Landes (Anarchie, Diktatur, Monarchie, Theokratie, Demokratie)

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

- Bei einer Standardnormalverteilung befinden sich Modus, Median und arithmetisches Mittel auf dem gleichen Erwartungswert.
- Der Median ist weniger anfällig gegenüber Ausreißern und Extremwerten als das arithmetische Mittel.
- Bei einer rechtssteilen Verteilung befindet sich der Median links vom Modus.
- Bei einer linksschiefen Verteilung befindet sich der Median links vom Modus.

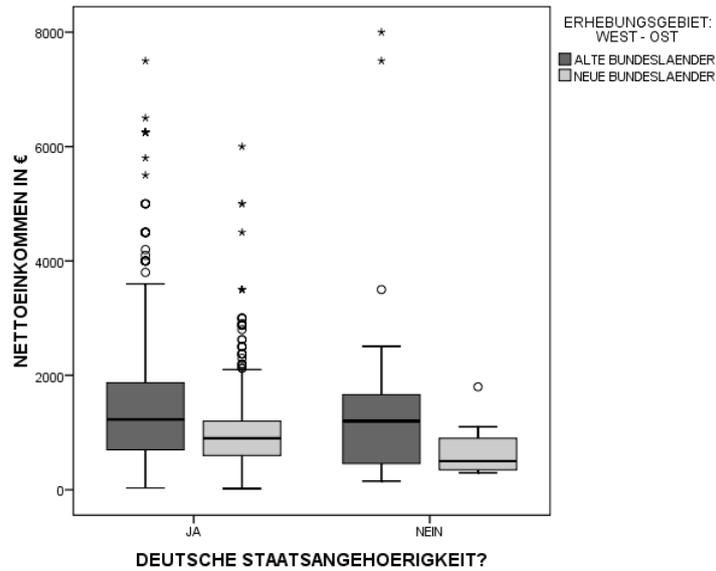
3. Welche der folgenden Aussagen zu Grafiken sind richtig? (2 Punkte)

- Ordinalskalierte Variablen lassen sich nur in Boxplots darstellen.
- Nominalskalierte Variablen eignen sich für die Darstellung innerhalb eines Balkendiagrammes.
- Man benötigt wenigstens zwei metrische Variablen, um ein Scatterplot zu erstellen.
- Ratioskalierte und nominalskalierte Variablen lassen sich innerhalb des gleichen Diagramms darstellen.

4. Ordnen Sie den Variablen eine passende Darstellungsform zu. Notieren Sie jeweils mit B, H, G oder S, ob es sich um ein Balkendiagramm, Histogramm, Gruppiertes Boxplot oder Streudiagramm handelt (3 Punkte):

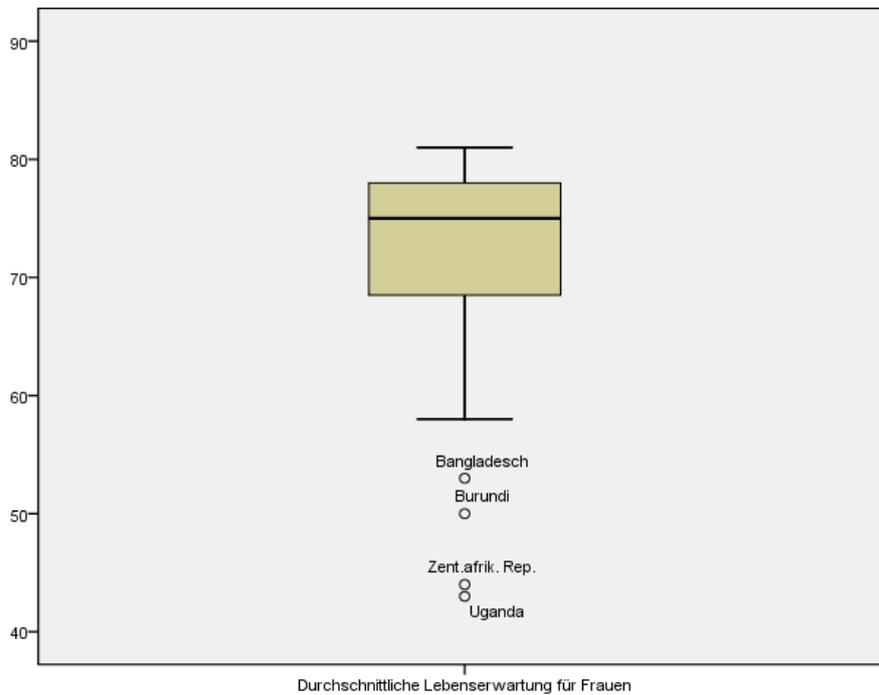
- Die Verteilung der Schulden der Befragten (in €), geordnet nach dem Bundesland der Befragten
- Die Gesundheitsausgaben verschiedener Länder (in US- $\text{\$}$) in Abhängigkeit von der Lebenserwartung
- Anzahl der Straftaten pro 100 Einwohner in verschiedenen Ländern in Abhängigkeit von der Arbeitslosenquote der Länder
- Die Wahlabsicht der Befragten in Prozent
- Anzahl der Polizisten pro 1.000 Einwohner in verschiedenen Ländern
- Bildungsniveau der Befragten

5. Erläutern Sie folgende Grafik (2 Punkte):



- Es wurde ein gruppiertes Boxplot für die Verteilung des Nettoeinkommens in den alten und neuen Bundesländern, jeweils für deutsche und andere Staatsangehörige erstellt.
- Weniger als 25% aller Fälle liegen über einem Nettoeinkommen von 2.000 Euro.
- Es kann von einer Normalverteilung der Nettoeinkommen ausgegangen werden.
- Die geringste Streuung besitzt die Verteilung des Nettoeinkommens der Befragten anderer Staatsangehörigkeit aus den neuen Bundesländern.

6. Interpretieren Sie folgende Grafik (2 Punkte):



- Aufgrund von Ausreißern hat die Grafik keine Erklärungskraft.
- Die Fälle Bangladesch, Burundi, Zentralafrikanische Republik und Uganda stellen Ausreißer dar.
- Der arithmetische Mittelwert liegt bei ungefähr 75.
- Der Median liegt bei ungefähr 75.

7. Bei einer Untersuchung sind die Merkmale Geschlecht und Schulabschluss erhoben worden. Welche Aussagen sind richtig? (2 Punkte)

	Männlich	Weiblich
Abitur	0	40
Kein Abitur	20	0

- Chi² sollte Null werden.
- Die Kreuztabelle spricht für vollkommene statistische Unabhängigkeit der beiden Merkmale.
- Phi sollte den Wert 1 annehmen.
- Die Kreuztabelle spricht für vollkommene statistische Abhängigkeit der beiden Merkmale.

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

- Häufigkeit von stetigen ordinalen Merkmalen
- Die Signifikanz einer bivariaten Verteilung
- Häufigkeit von stetigen metrischen Variablen
- Die Schiefe von Verteilungen

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

- Die Pearson-Korrelation basiert auf einem Rangvergleich.
- Es muss von Kausalität ausgegangen werden, wenn Pearson's r den Wert von 0,5 überschreitet.
- Wenn Korrelation vorliegt, liegt nicht automatisch Kausalität vor.
- Spearman's r und Pearson's r können Werte von -1 bis +1 annehmen.

10. Welche der folgenden Aussagen zu Ausreißern und Extremwerten sind richtig? (2 Punkte)

- Ausreißer und Extremwerte sind notwendig, um ein Boxplot zu erstellen.
- Die Spannweite ist aufgrund ihrer Berechnung nicht anfällig gegenüber Ausreißern.
- Das Vorhandensein von Extremwerten macht eine Stichprobe ungültig, um sie in weitere Berechnungen einfließen zu lassen.
- Extremwerte befinden sich 3 Kastenlängen oder mehr von der Box entfernt.

11. Was gilt für statistische Hypothesentests? (2 Punkte)

- Getestet wird stets die Nullhypothese. Wenn diese abgelehnt wird, kann man weiter von der Alternativhypothese ausgehen.
- Die fälschliche Ablehnung der Nullhypothese wird in der Statistik als beta-Fehler bezeichnet.
- Getestet werden können sowohl Unterschieds- als auch Zusammenhangshypothesen.
- Je größer die Stichprobe ist, desto höher ist die Wahrscheinlichkeit, dass der Mittelwert der Stichprobe dem der Grundgesamtheit entspricht.

12. Aus einem Länder-Datensatz ergibt sich folgende Korrelationsmatrix. Welche Aussagen sind richtig? (2 Punkte)

Korrelationen

		Alphabetisierungsrate (%)	Bevölkerungswachstum (% pro Jahr)	Mitgliedschaft OECD	Bruttoinlandsprodukt / Kopf
Alphabetisierungsrate (%)	Korrelation nach Pearson	1,000	-,699	-,419	,552
	Signifikanz (2-seitig)		,000	,000	,000
	N	107	107	107	107
Bevölkerungswachstum (% pro Jahr)	Korrelation nach Pearson	-,699	1,000	,480	-,521
	Signifikanz (2-seitig)	,000		,000	,000
	N	107	109	109	109
Mitgliedschaft OECD	Korrelation nach Pearson	-,419	,480	1,000	-,814
	Signifikanz (2-seitig)	,000	,000		,000
	N	107	109	109	109
Bruttoinlandsprodukt / Kopf	Korrelation nach Pearson	,552	-,521	-,814	1,000
	Signifikanz (2-seitig)	,000	,000	,000	
	N	107	109	109	109

- Zwischen Alphabetisierungsrate und Bevölkerungswachstum besteht ein signifikant negativer Zusammenhang.
- Ein geringes BIP pro Kopf geht mit einem hohen Bevölkerungswachstum einher.
- Es wurden mindestens 109 Länder untersucht.
- Die OECD-Mitgliedschaft hätte aufgrund des Skalenniveaus nicht in eine Pearson-Korrelation aufgenommen werden dürfen.

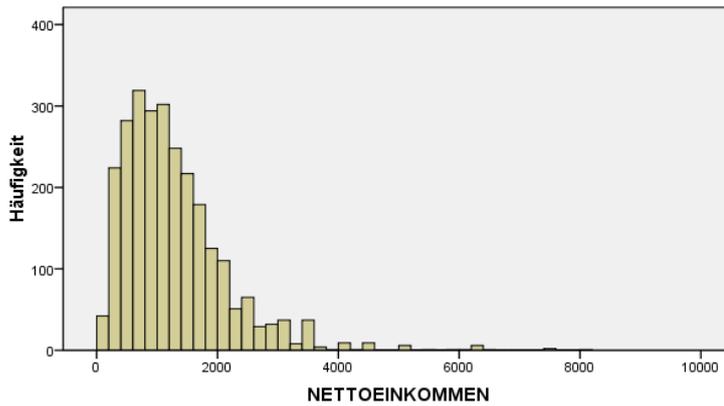
13. Bei einer Befragung von Studenten der Politikwissenschaft wurden unter anderem die Größe des Heimatortes (1= bis 1.999 EW, 2=2.000-4.999EW, ..., 7 = mehr als 499.999 EW), die politische Selbsteinschätzung auf einer Rechts-Links-Skala (1=Links, ..., 10=Rechts) sowie die Kirchengangshäufigkeit (1=Mehr als einmal die Woche, 2 = einmal die Woche, 3 = mehrmals im Monat, ..., 6=nie) abgefragt. Welche der folgenden Aussagen treffen zu? (2 Punkte)

Korrelationen

		Ortsgröße des Heimatortes	Links-Rechts	Kirchengangshäufigkeit
Spearman-Rho Ortsgröße des Heimatortes	Korrelationskoeffizient	1,000	-,093	,112
	Sig. (2-seitig)		,305	,205
	N	130	124	130
Links-Rechts	Korrelationskoeffizient	-,093	1,000	-,269
	Sig. (2-seitig)	,305		,003
	N	124	124	124
Kirchengangshäufigkeit	Korrelationskoeffizient	,112	-,269	1,000
	Sig. (2-seitig)	,205	,003	
	N	130	124	130

- Es wurden mindestens 130 Studenten befragt.
- Die durchgeführte Korrelation nach Spearman ist unzulässig, da die Variablen hierzu mindestens intervallskaliert sein müssen.
- Studenten, die häufiger in die Kirche gehen, schätzen sich auch als rechter ein.
- Es gibt keinen Zusammenhang zwischen der Ortsgröße und der Kirchengangshäufigkeit. Dies ist auf dem 10%-Niveau signifikant.

14. Erläutern Sie die folgende Grafik (2 Punkte):



- Es handelt sich um ein gruppiertes Balkendiagramm.
- Der Modus liegt bei ca. 5000 Euro Nettoeinkommen.
- Die Verteilung ist linkssteil.
- Die Verteilung ist linksschief.

15. Im Rahmen einer Wahlforschungsstudie zur Bundestagswahl 2005 wurden Bundesbürger zu Ihren Einstellungen (Sympathie) gegenüber Politikern und Parteien befragt, wobei 1 „überhaupt nicht mögen“ und 11 „sehr mögen“ entspricht. Ebenfalls wurde erhoben, ob der Befragte aus Ostdeutschland (kodiert mit 1) oder Westdeutschland (kodiert mit 0) stammt.

Welche der folgenden Aussagen treffen zu? (2 Punkte)

			Korrelationen				
			Sympathie CDU	Sympathie SPD	Gerhard Schröder	Angela Merkel	Ost-West
Spearman-Rho	Sympathie CDU	Korrelationskoeffizient	1,000	-,293	-,322	,749	-,102
		Sig. (2-seitig)	.	,000	,000	,000	,000
		N	2012	2009	2010	2010	2012
Sympathie SPD	Sympathie SPD	Korrelationskoeffizient	-,293	1,000	,742	-,232	-,033
		Sig. (2-seitig)	,000	.	,000	,000	,142
		N	2009	2011	2010	2010	2011
Gerhard Schröder (Kanzlerkandidat SPD)	Gerhard Schröder (Kanzlerkandidat SPD)	Korrelationskoeffizient	-,322	,742	1,000	-,226	-,023
		Sig. (2-seitig)	,000	,000	.	,000	,305
		N	2010	2010	2015	2014	2015
Angela Merkel (Kanzlerkandidat CDU)	Angela Merkel (Kanzlerkandidat CDU)	Korrelationskoeffizient	,749	-,232	-,226	1,000	-,053
		Sig. (2-seitig)	,000	,000	,000	.	,018
		N	2010	2010	2014	2015	2015
Ost-West	Ost-West	Korrelationskoeffizient	-,102	-,033	-,023	-,053	1,000
		Sig. (2-seitig)	,000	,142	,305	,018	.
		N	2012	2011	2015	2015	2018

- Ostdeutsche Befragte haben eine signifikant geringere Sympathie gegenüber der CDU als Westdeutsche.
- Ostdeutsche Befragte haben eine signifikant höhere Sympathie gegenüber der SPD als Westdeutsche.
- Die Sympathie gegenüber einem Kanzlerkandidaten korreliert signifikant mit der Sympathie zu seiner Partei.
- Angela Merkel wird von ostdeutschen Befragten signifikant sympathischer eingeschätzt als Gerhard Schröder.

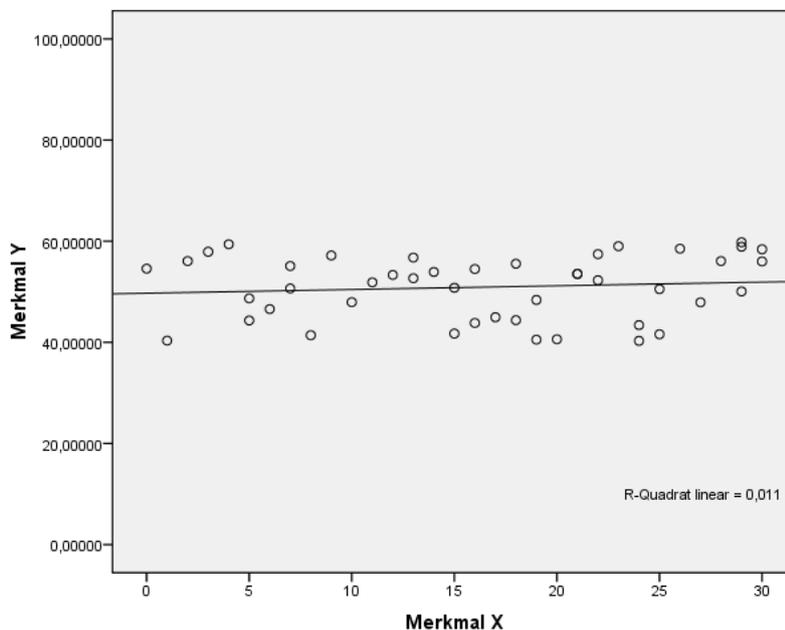
16. Welche Aussagen über den folgenden SPSS-Ausdruck sind richtig? (2 Punkte)

Gruppenstatistiken					
GESCHLECHT		N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
NETTOEINKOMMEN	MANN	1344	1526,95	982,734	26,806
	FRAU	1300	963,43	622,979	17,278

Test bei unabhängigen Stichproben							
	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit				
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz
NETTOEINKOMMEN Varianzen sind gleich	125,997	,000	17,544	2642	,000	563,516	32,120
NETTOEINKOMMEN Varianzen sind nicht gleich			17,669	2283,280	,000	563,516	31,892

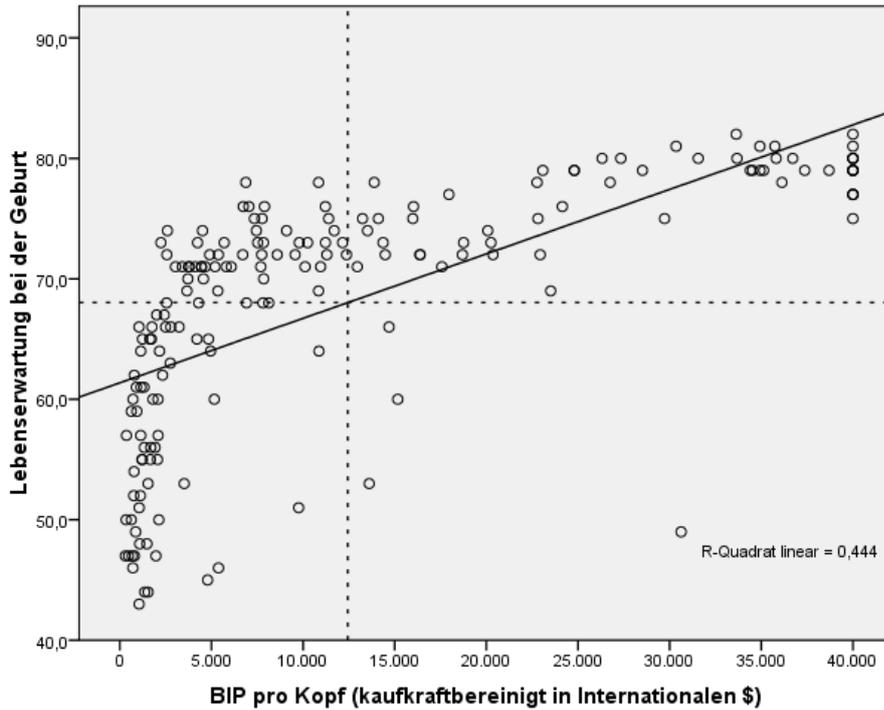
- Der Levene-Test hat bewiesen, dass die Varianzen der beiden Stichproben gleich sind.
- Es wird geprüft, ob die Einkommensunterschiede zwischen Männer und Frauen signifikant sind.
- Aufgrund des Ergebnisses Des Levene-Tests muss die zweite Zeile des T-Tests betrachtet werden.
- Der Test zeigt, dass Männer im Durchschnitt ein signifikant höheres Einkommen als Frauen haben.

17. Bitte interpretieren Sie nachfolgende Grafik! (2 Punkte)



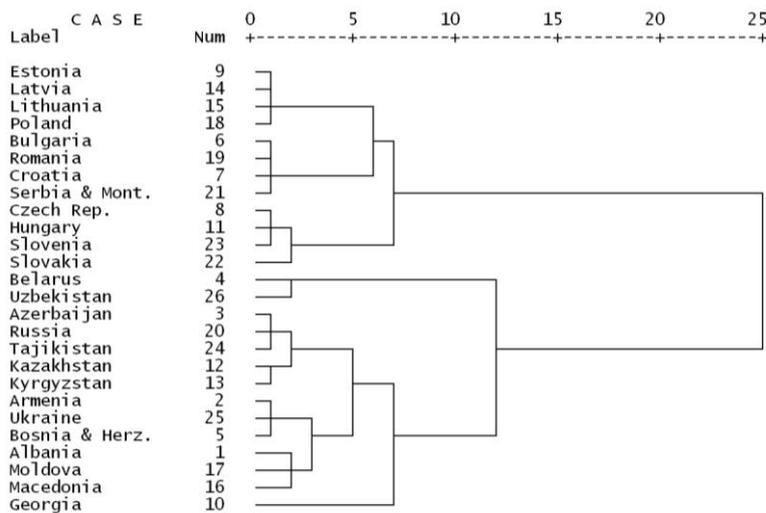
- Zwischen den Merkmalen X und Y besteht eine hohe positive Korrelation.
- Zwischen den Merkmalen X und Y besteht eine hohe negative Korrelation.
- Zwischen den Merkmalen X und Y besteht keine Korrelation.
- Ausreißer verzerren die Ergebnisse.

18. Bitte interpretieren Sie nachfolgende Graphik aus dem Datensatz des Human Development Index! (Gestrichelte Linien entsprechen den Mittelwerten der jeweiligen Variable.) (2 Punkte)



- Bis zu einem Bruttoinlandsprodukt von 5.000 Int.-\$ steigt die Lebenserwartung in der überwiegenden Zahl der dargestellten Fälle massiv an.
- Über 40% der Varianz der einen Variable werden durch die Varianz der anderen Variable erklärt.
- Aufgrund der Streuung der Variablen ist das Streudiagramm nicht geeignet.
- Extremwerte verzerren die Ergebnisse zu stark.

19. Die Ergebnisse welches Verfahrens zeigt die folgende Grafik: (2 Punkte)



- Faktorenanalyse
- Clusteranalyse
- Regression
- Multidimensionale Skalierung

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

- Das korrigierte R^2 unterscheidet sich vom normalen R^2 besonders bei multivariaten linearen Regressionen mit niedriger Fallzahl.
- Multikollinearität zwischen den unabhängigen Variablen ist anzustreben.
- Ein negatives R^2 deutet auf einen geringen Zusammenhang zwischen den Variablen hin.
- Die Heteroskedastizität gibt die Güte der Regression an.

21. Folgende Verfahren sind struktur-entdeckende Verfahren: (2 Punkte)

- Faktorenanalyse
- Clusteranalyse
- Regression
- Multidimensionale Skalierung

22. Welche Aussagen zur folgenden linearen Regression sind richtig? (3 Punkte)

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,598 ^a	,357	,355	8,9281

a. Einflußvariablen : (Konstante), BSP/Kopf

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	13693,672	1	13693,672	171,792	,000 ^a
	Residuen	24630,582	309	79,711		
	Gesamt	38324,254	310			

a. Einflußvariablen : (Konstante), BSP/Kopf

b. Abhängige Variable: Lebenserwartung

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	61,911	,599		103,436	,000
	BSP/Kopf	,001	,000	,598	13,107	,000

a. Abhängige Variable: Lebenserwartung

- Die Lebenserwartung soll durch das BSP pro Kopf erklärt werden.
- Die Signifikanz der F-Statistik lässt auf einen linearen Zusammenhang schließen.
- Der Determinationskoeffizient gibt den Anteil der durch die unabhängige Variable erklärten Varianz der abhängigen Variable an.
- Es besteht kein signifikanter Zusammenhang zwischen dem BSP pro Kopf und der Lebenserwartung.

23. Welche Aussagen zur folgenden linearen Regression sind richtig? (3 Punkte)

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,465 ^a	,216	,160	5,1174

a. Einflußvariablen : (Konstante), Linke Parteien in Prozent aller Ministerposten, Anteil der Gewerkschaftsmitglieder an allen Arbeitnehmern

ANOVA^b

Modell	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1 Regression	202,545	2	101,272	3,867	,033 ^a
Residuen	733,246	28	26,187		
Gesamt	935,791	30			

a. Einflußvariablen : (Konstante), Linke Parteien in Prozent aller Ministerposten, Anteil der Gewerkschaftsmitglieder an allen Arbeitnehmern

b. Abhängige Variable: Höhe der Sozialausgaben als Prozentsatz des BIP

Koeffizienten^a

Modell	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
	B	Standardfehler	Beta		
1 (Konstante)	11,330	2,053		5,518	,000
Anteil der Gewerkschaftsmitglieder an allen Arbeitnehmern	-,109	,043	-,421	-2,519	,018
Linke Parteien in Prozent aller Ministerposten	,027	,023	,196	1,174	,250

a. Abhängige Variable: Höhe der Sozialausgaben als Prozentsatz des BIP

- Der Anteil der linken Parteien an Ministerposten soll durch den Anteil der Gewerkschaftsmitglieder an allen Arbeitnehmern erklärt werden.
- Die Regressionsanalyse umfasst zwei unabhängige und eine abhängige Variable.
- Nur der Anteil der Gewerkschaftsmitglieder an allen Arbeitnehmern hat einen signifikanten Einfluss auf die abhängige Variable.
- Nur der Anteil der linken Parteien an allen Ministerposten hat einen signifikanten Einfluss auf die abhängige Variable.