SPEZIELLE MAẞZAHLEN I: SPEARMANS RHO

Martin-Luther-Universität Halle-Wittenberg
Institut für Soziologie
Übung Einführung in die deskriptive Statistik

Agenda

- Wiederholung
  - Regression mit Zwischenergebnissen
  - Dummyregression
- Spearmans ρ
  - Langformel
  - Kurzformel

Aufgabe 1: Lebenserwartung

Eine OECD-Forscherin interessiert sich dafür, ob die Lebenserwartung (in Jahren) von den Pro-Kopf-Gesundheitsausgaben (in 1.000 USD PPP) abhängt. Für die 40 untersuchten Länder erhält sie folgende Prognosegleichung $\hat{y} = 72{,}334 + 1{,}977 * x$. Zusätzlich gegeben sind $\sum(x_i - \bar{x})^2 = 121{,}597$ und $\sum(y_i - \bar{y})^2 = 1342{,}656$.

a) Formulieren Sie einen Interpretationssatz für die Regressionsparameter b_0 und b_1.

b) Berechnen Sie das standardisierte Regressionsgewicht b_1^* und interpretieren Sie dieses.

c) Was würde mit dem standardisierten Regressionsgewicht b_1^* passieren, wenn das Pro-Kopf-Gesundheitsausgaben in USD PPP gemessen worden wäre und das Lebensalter in Monaten?

d) Bestimmen Sie die Güte des Modells und interpretieren Sie Ihr Ergebnis.

Aufgabe 1a: Lösung

- Interpretation Regressionskonstante:
  - $b_0 = 72{,}334$
  - Für einen Wert von x=0 würde für y ein Wert von 72,334 vorhergesagt.
  - Für ein Land, in dem die Gesundheitsausgaben bei 0 USD PPP liegen, würde folglich eine durchschnittliche Lebenserwartung von 72,334 Jahren mithilfe des aufgestellten Modells vorhergesagt.

Aufgabe 1a: Lösung II

- Interpretation Regressionsgewicht
  - $b_1 = 1{,}977$
  - Steigt x um eine Einheit wird für y im Durchschnitt eine Erhöhung um 1,977 Einheiten vorhergesagt.
  - Wenn die Pro-Kopf-Gesundheitsausgaben also um 1.000 USD PPP steigen, so erhöht sich die vorhergesagte Lebenserwartung im Durchschnitt um 1,977 Jahre.

Aufgabe 1b: Lösung

- gegeben:
  - $b_1 = 1{,}977$
  - $n = 40$
  - $\sum(x_i - \bar{x})^2 = 121{,}597$
  - $\sum(y_i - \bar{y})^2 = 1342{,}656$
- gesucht:
  - $b_1^* = b_1 * \frac{s_X}{s_Y}$

Aufgabe 1b: Lösung II

- Berechnung s_X:
  - $s_X = \sqrt{\frac{SAQ_X}{n}} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$
  - $s_X = \sqrt{\frac{121{,}597}{40}}$
  - $s_X = 1{,}744$

Aufgabe 1b: Lösung III

- Berechnung s_Y:
  - $s_Y = \sqrt{\frac{SAQ_Y}{n}} = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n}}$
  - $s_Y = \sqrt{\frac{1342{,}656}{40}}$
  - $s_Y = 5{,}794$

Aufgabe 1b: Lösung IV

- Berechnung b_1^*
  - $b_1^* = b_1 * \frac{s_X}{s_Y}$
  - $b_1^* = 1{,}977 * \frac{1{,}744}{5{,}794}$
  - $b_1^* = 0{,}595$

Aufgabe 1b: Lösung V

- Interpretation b_1^*:
  - $b_1^* = 0{,}595$
  - Statistisch gesehen bedeutet dies, dass wenn sich x um eine Standardabweichung erhöht, y im Durchschnitt um $b_1^* = 0{,}595$ Standardabweichungen steigt.
  - Wenn die Gesundheitsausgaben um eine Standardabweichung erhöhen, dann steigt die geschätzte Lebenserwartung im Durchschnitt um 0,595 Standardabweichungen.
  - Im bivariaten Fall entspricht das standardisierte Regressionsgewicht b_1^* der Produktmomentkorrelation nach Pearson $r_{X,Y}$. Dies bedeutet, dass man hier auch sagen kann, dass ein starker positiver Zusammenhang zwischen den Gesundheitsausgaben und der Lebenserwartung in den betrachteten Ländern besteht.

Aufgabe 1c: Lösung

- Bei dem standardisierten Regressionsgewicht wird der Einfluss der Maßeinheit entfernt.
- Dies bedeutet, dass das standardisierte Regressionsgewicht gleich bleibt, wenn die x-Variable in USD PPP und die y-Variable in Monaten gemessen wird.

Aufgabe 1d: Lösung

- Berechnung:
  - $R^2 = \left(r_{x,y}\right)^2$ im bivariaten Modell
  - $r_{x,y} = b_1^*$ im bivariaten Modell
  - $r_{x,y} = 0{,}595$
  - $R^2 = (0{,}595)^2$
  - $R^2 = 0{,}354$
- Interpretation:
  - Durch Kenntnis der Pro-Kopf-Ausgaben für die Gesundheit (in 1.000 USD PPP) lässt sich die Prognose der Lebenserwartung (in Jahren) um 35,4 % verbessern.

Aufgabe 1: Lösung SPSS

Modellübersicht

Modell	R	R-Quadrat	Angepasstes R-Quadrat	Standardfehler der Schätzung
1	,595[a]	,354	,337	4,77732

a. Prädiktoren: (Konstante), Pro-Kopf-Gesundheitsausgaben (in 1000 USD PPP)

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	72,334	1,480		48,872	,000
	Pro-Kopf-Gesundheitsausgaben (in 1000 USD PPP)	1,977	,433	,595	4,564	,000

a. Abhängige Variable: Lebenserwartung Geburt (in Jahren)

Aufgabe 2: Ost-West Einkommen 2014

Eine Ungleichheitsforscherin interessiert sich dafür, ob die Herkunft eines Befragten einen Einfluss auf das Nettoeinkommen in Euro hat. Das Herkunftsgebiet wurde hierbei so codiert, dass die ostdeutschen Befragten einen Wert von 0 und die westdeutschen Befragten einen Wert von 1 aufweisen. Mithilfe des ALLBUS 2014 ergab sich folgender Regressionsoutput:

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1338,094	48,637		27,512	,000
	Erhebungsgebiet	236,949	58,765	,071	4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

a) Stellen Sie die stochastische Regressionsgleichung auf!
b) Interpretieren Sie die unstandardisierten Regressionskoeffizienten.
c) Wie sollte das standardisierte Regressionsgewicht für das Erhebungsgebiet in der vorliegenden Dummyregression interpretiert werden?
d) Wie würde die Regressionsgleichung lauten, wenn Ostdeutsche den Code 1 und Westdeutsche den Code 0 hätten?

Aufgabe 2a: Lösung

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1338,094	48,637		27,512	,000
	Erhebungsgebiet	236,949	58,765	,071	4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

- Stochastische Regressionsgleichung:
  - $y_i = b_0 + b_1 * D_i + e_i$
  - $y_i = 1338{,}094 + 236{,}949 * D_i + e_i$

Aufgabe 2b: Lösung

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1338,094	48,637		27,512	,000
	Erhebungsgebiet	236,949	58,765	,071	4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

- Interpretation Regressionskonstante:
  - Die Regressionskonstante gibt in der Dummyregression den vorhergesagten Wert der Referenzkategorie (D=0) an, der dem durchschnittlichen Wert dieser Kategorie entspricht.
  - Hier ist $b_0 = 1338{,}094$. Das bedeutet, dass die Befragten aus Ostdeutschland im Durchschnitt etwa ein Nettoeinkommen von 1338,09 Euro aufweisen.

Aufgabe 2b: Lösung II

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1338,094	48,637		27,512	,000
	Erhebungsgebiet	236,949	58,765	,071	4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

- Interpretation unstandardisiertes Regressionsgewicht:
  - Das unstandardisierte Regressionsgewicht gibt die absolute Veränderung des Wertes an, wenn sich D von der Referenzkategorie (Ausprägung 0) zu Ausprägung 1 verändert.
  - Hier ist $b_1 = 236{,}949$. Das bedeutet, dass westdeutsche Befragte im Durchschnitt ein um etwa 236,95 Euro höheres Nettoeinkommen aufweisen als ihre ostdeutschen Kollegen.

Aufgabe 2c: Lösung II

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1338,094	48,637		27,512	,000
	Erhebungsgebiet	236,949	58,765	,071	4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

- Interpretation standardisiertes Regressionsgewicht:
  - Da es sich im vorliegenden Fall um eine Dummyregression handelt, sollte das standardisierte Regressionsgewicht nicht interpretiert werden. (Per Definition: nominale x-Variable!)

Aufgabe 2d: Lösung I

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1338,094	48,637		27,512	,000
	Erhebungsgebiet	236,949	58,765	,071	4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

Ost=0
West=1

- Bestimmung des Regressionsmodells (West=0, Ost=1)
  - $y_i = b_0 + b_1 * D_i + e_i$
  - b_0: Durchschnittsnettoeinkommen der Westdeutschen
  - $b_0 = 1338{,}094 + 236{,}949 = 1575{,}043$
  - b_1: „Wie verändert sich das Nettoeinkommen, wenn wir statt eines Westdeutschen einen Ostdeutschen haben?“
  - $b_1 = -236{,}949$
  - $y_i = 1575{,}043 - 236{,}949 * D_i + e_i$

Aufgabe 2d: Lösung II

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	t	Sig.
		B	Standardfehler	Beta		
1	(Konstante)	1575,042	32,982		47,755	,000
	Erhebungsgebiet Rek	-236,949	58,765	-,071	-4,032	,000

a. Abhängige Variable: BFR.:NETTOEINKOMMEN<OFFENE+LISTENANGABE>

West=0
Ost=1

- $y_i = 1575{,}043 - 236{,}949 * D_i + e_i$

Landesstipendium Sachsen-Anhalt (fiktiv)

Im Rahmen der Graduiertenförderung des Landes Sachsen-Anhalts haben Absolventen die Möglichkeit sich um ein einjähriges Promotionsstipendium zu bewerben. Für die eingehenden Anträge werden durch zwei Gutachter für die einzelnen Bewerbungen Punkte vergeben. Für fünf Absolventen ergaben sich hierbei folgende Punkte, die durch die einzelnen Gutachter vergeben wurden:

Person	1	2	3	4	5
X:Punkte Gutachter 1	15	40	60	70	75
Y:Punkte Gutachter 2	10	40	80	75	65

Nach der Punktevergabe wurden die Kandidaten von jedem Gutachter in eine Rangfolge gebracht, wobei dem jeweils beste Kandidat der Rangplatz 1, dem zweitbesten der Rangplatz 2 und dem schlechtesten Kandidaten der Rangplatz 5 zugewiesen wurde. Wie stark hängen die von den Gutachtern zugewiesenen Rangplätze nun zusammen?

Quiz: Welches Skalenniveau weisen die beiden Ausgangsvariablen auf?
Welches Skalenniveau haben sie nach der Transformation in Rangplätze?

Landesstipendium Sachsen-Anhalt: Analyse

Ausgangpunkt: zwei metrische Variablen

Person	1	2	3	4	5
X:Punkte Gutachter 1	15	40	60	70	75
Y:Punkte Gutachter 2	10	40	80	75	65

Wie stark hängen die von den Gutachtern zugewiesenen Rangplätze zusammen?

Transformation in Ränge: ordinales Skalenniveau

Person	1	2	3	4	5
Rg(X):Rang Gutachter 1	?	?	?	?	?
Rg(Y):Rang Gutachter 2	?	?	?	?	?

Möglichkeiten: 1., 2., 3., 4., 5.

Berechnung von Spearmans rho

Spearmans Rho (ρ)

- Verwendung:
  - **monotone** Zusammenhänge zwischen metrischen Variablen
  - Bestimmung der **Rangkorrelation** von zwei Variablen, besonders wenn zwei Gutachter eine Einschätzung abgeben
  - Alternativkonzept zur Paarvergleichslogik bei ordinalen Variablen (Es gibt gute Gründe, warum wir es da meist nicht verwenden…)

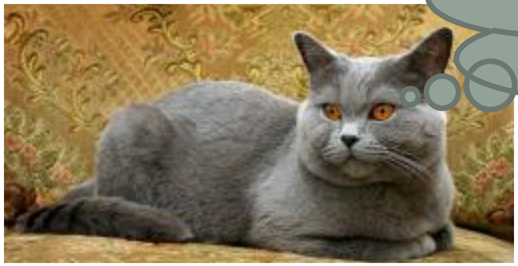
Spearmans Rho (ρ) II

- Berechnung:
  - $\rho = r_{sp} = \frac{SP_{(Rang(X),Rang(Y))}}{\sqrt{SAQ_{Rang(X)} * SAQ_{Rang(Y)}}}$

Kovariation der Ränge von X und Y

Variation des Rangs von X multipliziert mit Variation des Rangs von Y

  - Verwendung von Rängen statt Originaldaten
  - bester Wert Rang 1, zweitbester Wert Rang 2…
  - bei gleichen Rängen einer Variable wird Durchschnittsrang verwendet (arithmetisches Mittel der potenziellen Ränge)



Spearmans Rho wird genauso berechnet wie Pearsons r, bloß dass hier nicht die eigentlichen Werte, sondern die Ränge verwendet werden.

Aufgabe 3.1: Bestimmung der Rangplätze

Im Rahmen der Graduiertenförderung des Landes Sachsen-Anhalts haben Absolventen die Möglichkeit sich um ein einjähriges Promotionsstipendium zu bewerben. Für die eingehenden Anträge werden durch zwei Gutachter für die einzelnen Bewerbungen Punkte vergeben. Für fünf Absolventen ergaben sich hierbei folgende Punkte, die durch die einzelnen Gutachter vergeben wurden:

Person	1	2	3	4	5
X:Punkte Gutachter 1	15	40	60	70	75
Y:Punkte Gutachter 2	10	40	80	75	65

Bestimmen Sie jeweils für die Gutachter Rangreihenfolge der Kandidaten. Weisen Sie hierbei dem Kandidaten mit den meisten Punkten den ersten Rangplatz zu und dem Kandidaten mit den wenigsten Punkten den fünften Rangplatz.

Aufgabe 3.1: Lösung I

Person	x_i	Rang (x_i)	y_i	Rang (y_i)
1	15		10	
2	40		40	
3	60		80	
4	70		75	
5	75		65	

Aufgabe 3.1: Lösung II

Person	x_i	Rang (x_i)	y_i	Rang (y_i)
1	15	5	10	5
2	40	4	40	4
3	60	3	80	1
4	70	2	75	2
5	75	1	65	3

Aufgabe 3.2: Bestimmung der Rangkorrelation

Für die fünf Absolventen wurden folgende Rangplätze ermittelt:

Person	x_i	Rang (x_i)	y_i	Rang (y_i)
1	15	5	10	5
2	40	4	40	4
3	60	3	80	1
4	70	2	75	2
5	75	1	65	3

Inwieweit stimmen die vergebenen Ränge der beiden Gutachter überein? Berechnen Sie die Spearmans Rangkorrelationskoeffizient rho und interpretieren Sie ihr Ergebnis.

Hinweis: $\rho = r_{sp} = \frac{SP_{(Rang(X),Rang(Y))}}{\sqrt{SAQ_{Rang(X)} * SAQ_{Rang(Y)}}}$

Aufgabe 3.2: Lösung I

	Rg(x_i)	Rg(y_i)	$(Rg(x_i) - \overline{Rg(x)})$	$(Rg(x_i) - \overline{Rg(x)})^2$	$(Rg(y_i) - \overline{Rg(y)})$	$(Rg(y_i) - \overline{Rg(y)})^2$	$(Rg(x_i) - \overline{Rg(x)}) * ((Rg(y_i) - \overline{Rg(y)})$
1	5	5					
2	4	4					
3	3	1					
4	2	2					
5	1	3					
	$\overline{Rg(x)} =$	$\overline{Rg(y)} =$		$SAQ_{Rg(x)} =$		$SAQ_{Rg(y)} =$	$SP_{Rg(x),Rg(y)} =$

$$\rho = r_{sp} = \frac{SP_{(Rang(X),Rang(Y))}}{\sqrt{SAQ_{Rang(X)} * SAQ_{Rang(Y)}}}$$

Aufgabe 3.2: Lösung II

	Rg(x_i)	Rg(y_i)	$(Rg(x_i) - \overline{Rg(x)})$	$(Rg(x_i) - \overline{Rg(x)})^2$	$(Rg(y_i) - \overline{Rg(y)})$	$(Rg(y_i) - \overline{Rg(y)})^2$	$(Rg(x_i) - \overline{Rg(x)}) * ((Rg(y_i) - \overline{Rg(y)})$
1	5	5	2	4	2	4	4
2	4	4	1	1	1	1	1
3	3	1	0	0	-2	4	0
4	2	2	-1	1	-1	1	1
5	1	3	-2	4	0	0	0
	$\overline{Rg(x)} = 3$	$\overline{Rg(y)} = 3$		$SAQ_{Rg(x)} = 10$		$SAQ_{Rg(y)} = 10$	$SP_{Rg(x),Rg(y)} = 6$

$$\rho = r_{sp} = \frac{SP_{(Rang(X),Rang(Y))}}{\sqrt{SAQ_{Rang(X)} * SAQ_{Rang(Y)}}}$$

Aufgabe 3.2: Lösung III

- gegeben:
  - $SAQ_{Rg(x)} = 10$
  - $SAQ_{Rg(y)} = 10$
  - $SP_{Rg(x),Rg(y)} = 6$
- Berechnung:
  - $\rho = r_{sp} = \frac{SP_{(Rang(X),Rang(Y))}}{\sqrt{SAQ_{Rang(X)} * SAQ_{Rang(Y)}}}$
  - $\rho = \frac{6}{\sqrt{10*10}}$
  - $\rho = 0{,}6$

Aufgabe 3.3: Lösung

- Interpretation:
  - $\rho = 0{,}60$
  - Es besteht eine hohe positive Rangkorrelation zwischen dem ersten Gutachter und dem zweiten Gutachter. Kandidaten, die vom ersten Gutachter gut eingeschätzt werden, werden tendenziell auch vom zweiten Gutachter als gut eingeschätzt und umgekehrt

Aufgabe 3.3: Lösung SPSS

Symmetrische Maße

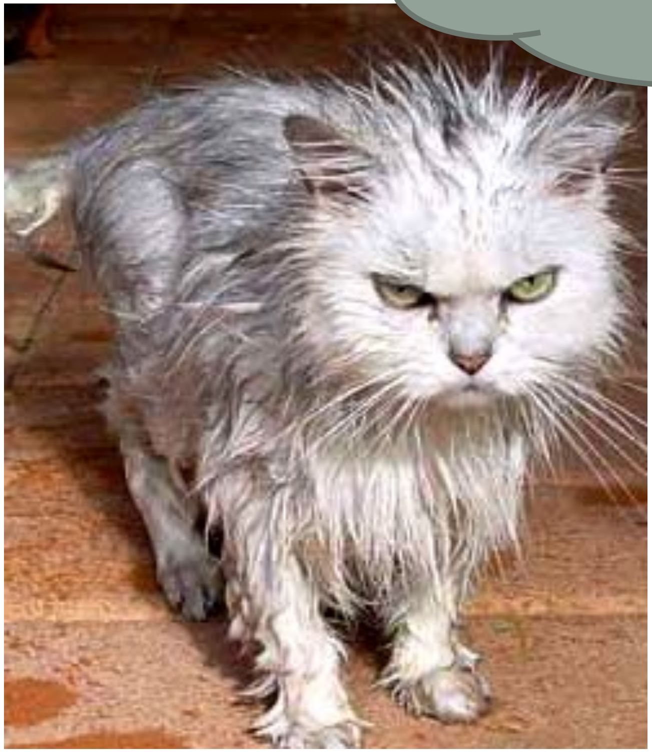
		Wert	Asymp. Standardfehler[a]	Näherungsweise A[b]	Näherungsweise Sig.
Intervall bezüglich Intervall	Pearson-R	,921	,063	4,089	,026[c]
Ordinal bezüglich Ordinal	Spearman-Korrelation	,600	,438	1,299	,285[c]
Anzahl der gültigen Fälle		5			

a. Die Nullhypothese wird nicht vorausgesetzt.

b. Unter Annahme der Nullhyphothese wird der asymptotische Standardfehler verwendet.

c. Basierend auf normaler Approximation.

Das dauert mir alles zu lang, gerade, wenn es viele Werte sind, geht das nicht auch einfacher?



Spearmans Rho (ρ) III: Kurzformel

- Berechnung über die Kurzformel:
  - wenn keine Verknüpfungen in X oder Y vorliegen, lässt sich die Kurzformel zur Bestimmung verwenden:
  - $\rho = 1 - \frac{6 * \sum_{i=1}^{n} d_i^2}{n * (n^2 - 1)}$

n: Fallzahl

d_i: Rang(x_i)-Rang(y_i)

Hinweis: In Normalsprache übersetzt bedeutet das, dass wir die Formel immer dann anwenden können wenn jeder Gutachter jeden Rangplatz nur genau einmal vergeben hat und wir keine Durchschnittsplätze berechnen mussten. ☺

Aufgabe 4.1: Graduiertenförderung II

Im Rahmen der Graduiertenförderung des Landes Sachsen-Anhalts haben Absolventen die Möglichkeit sich um ein einjähriges Promotionsstipendium zu bewerben. Für die eingehenden Anträge werden durch zwei Gutachter für die einzelnen Bewerbungen Punkte vergeben. Für fünf Absolventen ergaben sich hierbei folgende Punkte, die durch die einzelnen Gutachter vergeben wurden:

Person	1	2	3	4	5
X:Punkte Gutachter 1	15	40	60	70	75
Y:Punkte Gutachter 2	10	40	80	75	65

Inwieweit stimmen die beiden Gutachter in ihrer Rangfolge überein? Lässt sich im Vergleich zur Standardformel vereinfachen? Warum (nicht)?

Aufgabe 4.1: Analyse

Kandidat	Punkte Gutachter 1	Punkte Gutachter 2
1	15	10
2	40	40
3	60	80
4	70	75
5	75	65

kein doppelter Wert in X

kein doppelter Wert in Y

→ keine Verknüpfungen:

Nutzung der Kurzformel möglich: $r_{SP} = 1 - \frac{6 * \sum_{i=1}^{n} d_i^2}{n * (n^2 - 1)}$

Aufgabe 4.2

Die Ränge haben wir bereits bestimmt:

Person	x_i	Rang (x_i)	y_i	Rang (y_i)
1	15	5	10	5
2	40	4	40	4
3	60	3	80	1
4	70	2	75	2
5	75	1	65	3

a) Berechnen Sie nun die Differenz zwischen dem ersten Gutachter und dem zweiten Gutachter d_i =Rang(x_i)-Rang(y_i).

b) Quadrieren Sie Ihr Ergebnis für jeden einzelnen Kandidaten und bilden Sie daraus die Summe!

Aufgabe 4.2: Lösung

Person	Rang (x_i)	Rang (y_i)	$d_i = Rg_{(x_i)} - Rg_{(y_i)}$	d_i^2
1	5	5		
2	4	4		
3	3	1		
4	2	2		
5	1	3		
			$\sum_{i=1}^{n} d_i^2 =$	

Aufgabe 4.2a: Lösung

Person	Rang (x_i)	Rang (y_i)	$d_i = Rg_{(x_i)} - Rg_{(y_i)}$	d_i^2
1	5	5	0	
2	4	4	0	
3	3	1	2	
4	2	2	0	
5	1	3	-2	
			$\sum_{i=1}^{n} d_i^2 =$	

Aufgabe 4.2b: Lösung

Person	Rang (x_i)	Rang (y_i)	$d_i = Rg_{(x_i)} - Rg_{(y_i)}$	d_i^2
1	5	5	0	0
2	4	4	0	0
3	3	1	2	4
4	2	2	0	0
5	1	3	-2	4
			$\sum_{i=1}^{n} d_i^2 =$	8

Aufgabe 4.3

Person	Rang (x_i)	Rang (y_i)	$d_i = Rg_{(x_i)} - Rg_{(y_i)}$	d_i^2
1	5	5	0	0
2	4	4	0	0
3	3	1	2	4
4	2	2	0	0
5	1	3	-2	4
			$\sum_{i=1}^{n} d_i^2 =$	8

a) Berechnen Sie nun mithilfe der Kurzformel Spearmans $\rho = 1 - \frac{6 * \sum_{i=1}^{n} d_i^2}{n * (n^2 - 1)}$

b) Interpretieren Sie Ihr Ergebnis!

Aufgabe 4.3a und b: Lösung

- gegeben:
  - $n = 5$
  - $\sum_{i=1}^{n} d_i^2 = 8$
- Berechnung:
  - $\rho = 1 - \frac{6 * \sum_{i=1}^{n} d_i^2}{n * (n^2 - 1)}$
  - $\rho = 1 - \frac{6*8}{5*(5^2-1)}$
  - $\rho = 1 - \frac{48}{120}$
  - $\rho = 0{,}60$

Aufgabe 4.3b: Lösung II

- Interpretation:
  - $\rho = 0{,}60$
  - Es besteht ein hoher positiver Zusammenhang zwischen dem Rang, den Gutachter 1 vergeben hat, und dem Rang den Gutachter 2 vergeben hat.
  - Inhaltlich gesehen wurden die Kandidaten, die von Gutachter 1 besser eingestuft wurden, auch von Gutachter 2 besser eingestuft und umgekehrt.

Fragen?

Aufgabe 5: Kunstgutachter

Zwei Gutachter bewerten unabhängig voneinander die Qualität der Arbeiten von Künstlern auf einer Skala, die von 0 bis 10 reicht, wobei ein Wert von 0 für eine sehr niedrige und ein Wert von 10 für eine sehr hohe Qualität steht. Für vier Künstler erhalten sie folgende Werte:

Künstler	Gutachter 1	Gutachter 2
A	1	10
B	5	4
C	9	2
D	10	1

Ein Kunstwissenschaftler interessiert sich dabei, inwieweit die Rangfolge der Bewertung von beiden Gutachtern übereinstimmt.

a) Welches Maß ist hierbei zur Untersuchung des Zusammenhangs geeignet? Begründen Sie Ihre Wahl.
b) Berechnen Sie dieses Maß.
c) Interpretieren Sie Ihr Ergebnis.

Aufgabe 5: Analyse

Ein Kunstwissenschaftler interessiert sich dabei, inwieweit die **Rangfolge** der Bewertung von beiden **Gutachtern** übereinstimmt.

Rangfolgen:
ordinales Skalenniveau

Hinweis: Gutachter

geeignetes Maß:
Spearmans rho

Aufgabe 5: Analyse II

Künstler	Gutachter 1	Gutachter 2
A	1	10
B	5	4
C	9	2
D	10	1

kein doppelter Wert in X

kein doppelter Wert in Y

Nutzung der Kurzformel möglich: $r_{SP} = 1 - \frac{6 * \sum_{i=1}^{n} d_i^2}{n * (n^2 - 1)}$

Aufgabe 5: Lösung I

1. Bestimmung der Ränge:

Person	x_i	Rang (x_i)	y_i	Rang (y_i)
A	1	1	10	4
B	5	2	4	3
C	9	3	2	2
D	10	4	1	1

Aufgabe 5: Lösung II

2. Berechnung der quadrierten Differenzen:

Person	Rang (x_i)	Rang (y_i)	$d_i = Rg_{(x_i)} - Rg_{(y_i)}$	d_i^2
A	1	4	-3	9
B	2	3	-1	1
C	3	2	1	1
D	4	1	3	9
			$\sum_{i=1}^{n} d_i^2 =$	20

Aufgabe 5: Lösung III

- Berechnung:
  - $r_{SP} = 1 - \frac{6 * \sum_{i=1}^{n} d_i^2}{n * (n^2 - 1)}$
  - $r_{SP} = 1 - \frac{6 * 20}{4 * (4^2 - 1)} = 1 - \frac{120}{60}$
  - $r_{SP} = 1 - 2$
  - $r_{SP} = -1$
- Interpretation:
  - Es besteht ein perfekter negativer Zusammenhang zwischen der Rangfolge des ersten Gutachters und des zweiten Gutachters.
  - Konkret bedeutet dies hier, dass die Einschätzung diametral unterschiedlich ist. Alles, was Gutachter 1 als gut einschätzt wird von Gutachter 2 eher als Schund eingeschätzt und umgekehrt.

Literaturhinweise

- Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 243-253.
- Steffen-M. Kühnel / Dagmar Krebs (2012): Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. S. 450-455
- Hans Benninghaus (2007): Deskriptive Statistik. Eine Einführung für Sozialwissenschaftler. S. 177-184.