

# FAQ-SITZUNG

---

Martin-Luther-Universität Halle-Wittenberg

Institut für Soziologie

Übung Einführung in die deskriptive Statistik

# WAS MAN AUS M2 MITNEHMEN SOLLTE....

---

# Wissenstest: Sitzung 2

- Wie ist eine Datenmatrix aufgebaut?
- Welche Skalenniveaus gibt es?  
Welche Eigenschaften haben diese?  
Was lässt sich berechnen?
- Wie sieht der idealtypische Forschungsablauf aus?
- Was besagt das per-fiat-Kriterium?

# Wissenstest: Sitzung 3

- Was versteht man unter Klassierung? Weshalb wird sie zum Teil verwendet? Welche Konsequenzen ergeben sich daraus?
- Welche Regeln gibt es zur Klassenbildung?
- Was sind Quantile? Ab wann sind diese sinnvoll bestimmbar?
- Wie ist eine empirische Verteilungsfunktion aufgebaut? Was kann in einer empirischen Verteilungsfunktion ablesen?
- Wie ist ein Histogramm aufgebaut? Was versteht man unter dem Prinzip der Flächentreue?

# Wissenstest: Sitzung 4

- Welche Maße der zentralen Tendenz gibt es? Was geben diese an? Ab welchem Skalenniveau sind diese bestimmbar? Welche Eigenschaften haben diese?
- Nennen Sie zwei wesentliche Eigenschaften des arithmetischen Mittels! Welche Probleme können mit der Verwendung des arithmetischen Mittels verbunden sein?
- Was versteht man unter einem getrimmten Mittel?
- Wie lautet die Daumenregel für Lagemaße?
- Wofür brauchen wir Streuungsmaße? Welche gibt es? Ab welchem Skalenniveau sind sie bestimmbar? Welchen Informationsgehalt haben sie?
- Was beinhaltet die 5-Punkte-Zusammenfassung?
- Was lässt sich alles in einem Boxplot ablesen?

# Wissenstest: Sitzung 5

- Nach welchen Kriterien lassen sich Häufigkeitsverteilungen noch charakterisieren, jenseits von zentraler Tendenz und Streuung?
- Was versteht man unter einer z-Transformation? Welchen Mittelwert, welche Varianz und welche Standardabweichung weist eine z-transformierte Variable auf?
- Wofür brauchen wir bivariate Analysen?
- Wie ist eine Kreuztabelle aufgebaut?
- Erläutern Sie den Aufbau eines Stapelbalkendiagramms.

# Wissenstest: Sitzung 6

- Definieren Sie den Begriff statistische Beziehung.
- Welche Anforderungen werden an Assoziationsmaße gestellt?
- Worin unterscheiden sich symmetrische und asymmetrische Maße?
- Wofür eignet sich die Prozentsatzdifferenz? Um was für ein Maß handelt es sich? Welche Vorteile und Nachteile sind mit ihr verbunden?
- Erläutern Sie kurz die Idee hinter dem Konzept der Odds.
- Was gibt der Odds-Ratio an? Wann wird es angewandt? Welche Vor- und Nachteile besitzt es gegenüber der Prozentsatzdifferenz?

# Wissenstest: Sitzung 7

- Erläutern Sie die Logik der stochastischen Unabhängigkeit.
- Warum können wir nicht  $\chi^2$  direkt als Zusammenhangsmaß verwenden?
- Welche Normierungen gibt es? Wofür eignen sich diese?
- Erläutern Sie kurz die Logik der proportionalen Fehlerreduktion allgemein. Welche Fehlertypen gibt es? Wodurch zeichnen sich diese aus? Wie werden solche Maße allgemein berechnet und interpretiert?
- Welches PRE-Maß kennen Sie für nominale Daten? Welche Probleme sind mit diesem verbunden?



# Wissenstest: Sitzung 8

- Erläutern Sie kurz die Logik des Paarvergleichs!
- Welche Paare werden unterschieden?
- Welche Zusammenhangsmaße kennen Sie? Sind diese symmetrisch oder asymmetrisch? Welche Besonderheiten haben diese? Wann sind sie besonders geeignet?
- Erläutern Sie die PRE-Logik hinter Gamma.
- Welches Maß für ordinale Zusammenhänge basiert nicht auf der Logik des Paarvergleichs? Wofür wird es hauptsächlich verwendet?
- Wie unterscheiden sich Tau-a und Spearman rho hinsichtlich der Auswirkung von Verknüpfungen?

# Wissenstest: Sitzung 9

- Worin unterscheiden sich Kovariation, Kovarianz und Pearsons  $r$ ?
- Welche Eigenschaften weist Pearsons  $r$  auf?
- Welche Grenzen bestehen hinsichtlich der Interpretation von Pearsons  $r$ ?
- Wann wird die lineare Regression verwendet? Wie bezeichnet man die beteiligten Variablen? Welches Ziel ist mit der Aufstellung eines Modells in der linearen Regression verbunden?
- Wie unterscheiden sich Vorhersagewert und tatsächlich realisierter Wert? Wie wird dieser Unterschied im Rahmen der linearen Regression benannt?
- Was versteht man unter einer stochastischen Gleichung? Was versteht man unter einer Schätzgleichung?

# Wissenstest: Sitzung 10

- Wofür steht OLS? Was wird bei der OLS-Methode minimiert? Warum?
- Welche Eigenschaften weist das Ergebnis der OLS-Methode auf?
- Wie lassen sich Regressionskonstante und Regressionsgewicht in der bivariaten Regression bestimmen? Was besagen diese?
- Wie lässt sich die Güte des Gesamtmodells bestimmen?
- Erläutern Sie die PRE-Logik anhand von  $R^2$ .

# Wissenstest: Sitzung 11

- Welche Besonderheit ermöglicht uns eine Vereinfachung der Berechnung von  $R^2$  im bivariaten Fall?
- Warum standardisiert man Regressionsgewichte? Wie verändert sich die Interpretation? Was ist bei der Interpretation von standardisierten Regressionsgewichten in der Dummyregression zu beachten?
- Welcher Zusammenhang besteht in der bivariaten Regression zwischen standardisiertem Regressionsgewicht und Pearsons  $r$ ?
- Was geschieht bei einer z-Transformation? Welche Eigenschaften weist die so entstandene Variable hinsichtlich Mittelwert, Standardabweichung und Varianz auf?

# Wissenstest: Sitzung 11 II

- Welche Vorteile haben standardisierte Regressionsgewichte? Welche Nachteile?
- Was sind Aggregatdaten?
- Welche Besonderheiten sind bei der Regression mit Aggregatdaten zu beachten?
- Was versteht man unter einem ökologischen Fehlschluss?

# Wissenstest: Sitzung 12

- Was sind Dumyvariablen? Welches Skalenniveau haben sie per definitionem? Wie lassen sich diese in der linearen Regression verwenden?
- Wie lautet die entsprechende stochastische Regressionsgleichung?
- Welche Besonderheiten gibt es hinsichtlich der Interpretation der Regressionskonstante und des unstandardisierten Regressionsgewichts?
- Was kann man mithilfe des standardisierten Regressionsgewichts aussagen? Warum?
- Welches Alternativverfahren gibt es, was sich nicht nur auf dichotome Variablen, sondern zusätzlich auch auf polytome Variablen anwenden lässt?

# Wissenstest: Sitzung 12 II

- Welche Voraussetzungen müssen gegeben sein, um eine einfaktorielle Varianzanalyse durchführen zu können?
- Erläutern Sie die Begriffe „Faktor“ und „Faktorstufe“.
- Welche Zusammenhangsmaße resultieren aus der Varianzanalyse? Wie sind diese zu deuten?
- Erläutern Sie das Grundprinzip der Varianzanalyse! Gehen Sie auf die unterschiedlichen Fehler ein.

# Wissenstest: Sitzung 13

- Welcher Zusammenhang besteht zwischen der Regression mit einer Dummyvariablen und der Varianzanalyse?
- Warum ist es nicht ohne Probleme möglich, nominale Variablen mit mehr als zwei Ausprägungen in der Regression verwenden? Inwiefern kann hier die Berechnung von  $\eta^2$  und  $\eta$  Abhilfe schaffen?
- Wie lassen sich polytome nominale Variablen dennoch in der Regression verwenden?
- Vergleichen Sie ANOVA und Regression hinsichtlich des Umgang mit nichtlinearen Zusammenhängen!
- Wann kann es bei metrischen Zusammenhängen sinnvoll anstelle  $R^2$   $\eta^2$  zu berechnen? Zeichnen Sie ein Entscheidungsdiagramm!



# ALLGEMEINES

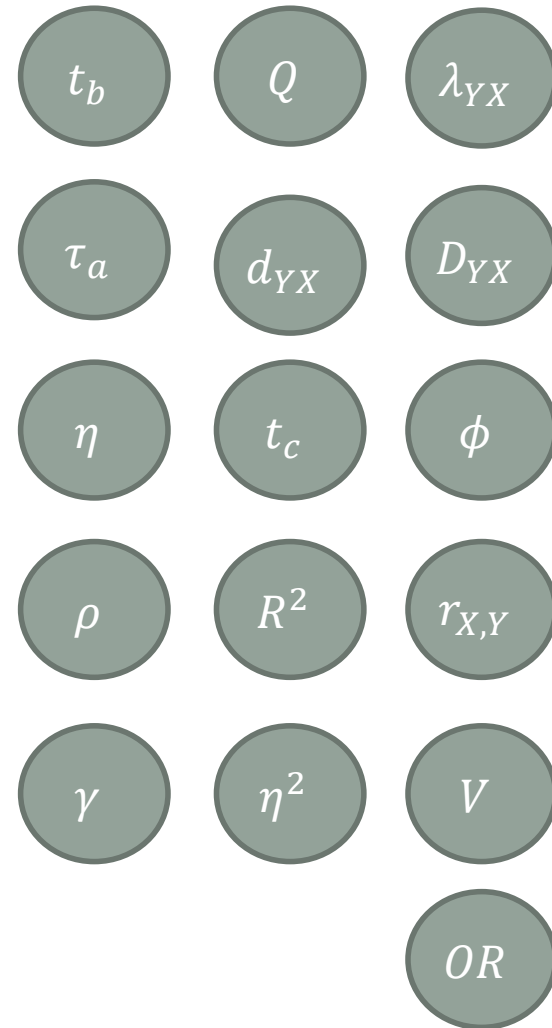
---

# Aufgabe 1: Skalenniveaus

- Welche Skalenniveaus kennen wir?
- Welche Eigenschaften haben diese?
- Welche Maße der zentralen Tendenz sind für die jeweiligen Skalenniveaus am informationsreichsten?
- Welche Eigenschaften haben die verschiedenen Maße der zentralen Tendenz?
- Welche Streuungsmaße lassen sich für welches Skalenniveau berechnen?

# Aufgabe 2.1: Beutebuch

Skalenniveau	symmetrisch	asymmetrisch
nominal		
ordinal		
metrisch		
gemischt		



# Aufgabe 2.2: Beutebuch

PRE-Maß	kein PRE-Maß

$$t_b$$

$$Q$$

$$\lambda_{YX}$$

$$\tau_a$$

$$d_{YX}$$

$$D_{YX}$$

$$\eta$$

$$t_c$$

$$\phi$$

$$\rho$$

$$R^2$$

$$r_{X,Y}$$

$$\gamma$$

$$\eta^2$$

$$V$$

$$OR$$

# Aufgabe 2: Lösung

- siehe Übersicht „Zusammenhangsmaße“ in stud.IP 😊

# Aufgabe 3

- Wie lautet die allgemeine Interpretation für ein normales Assoziationsmaß?
- Wie lautet die allgemeine Interpretation für ein PRE-Maß?
- Was gibt der E0- und der E1-Fehler bei einem PRE-Maß an?

# UNIVARIANTE VERTEILUNGEN

---

## Aufgabe 4: Soziale Not

In der folgenden Tabelle ist die Einstellung zu der Frage, ob Kirchen soziale Not lindern sollten, abgebildet:

**KIRCHEN: SOLLEN SOZIALE NOT LINDERN**

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozenze
Gültig	AUF JEDEN FALL	2448			
	EHER	797			
	EHER NICHT	141			
	AUF KEINEN FALL	49			
	Gesamtsumme	3435			
Fehlend	KEINE ANGABE	45			
Gesamtsumme		3480			

- Berechnen Sie die fehlenden Werte! (3)
- Zeichnen Sie die empirische Verteilungsfunktion. (4)

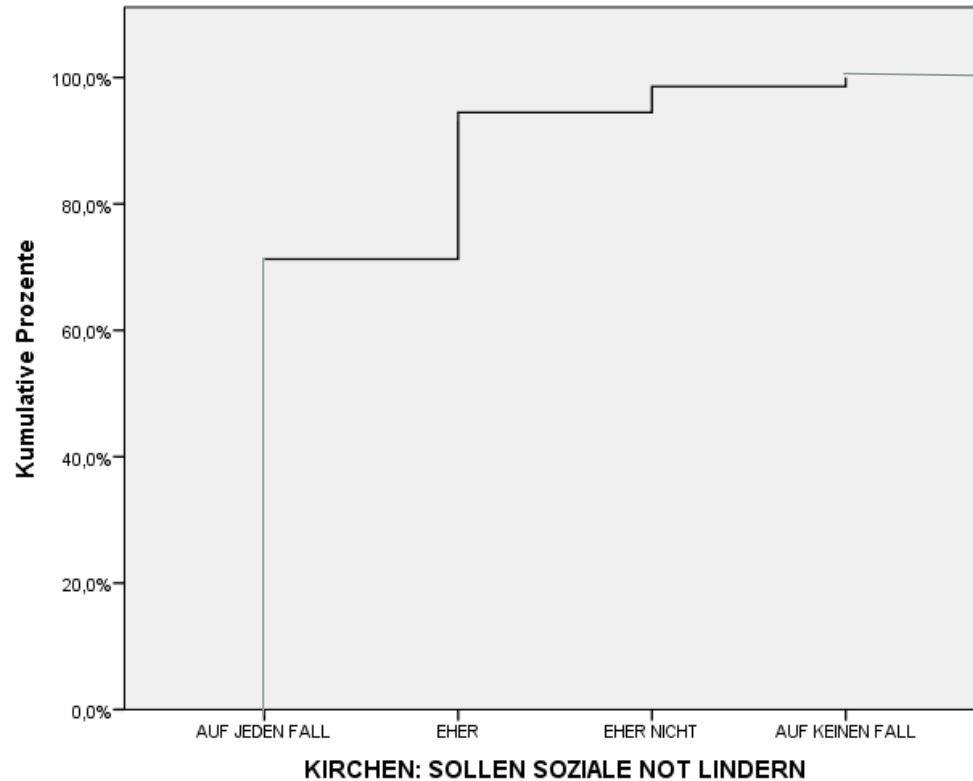


# Aufgabe 4a: Lösung

## KIRCHEN: SOLLEN SOZIALE NOT LINDERN

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	AUF JEDEN FALL	2448	70,3	71,3	71,3
	EHER	797	22,9	23,2	94,5
	EHER NICHT	141	4,1	4,1	98,6
	AUF KEINEN FALL	49	1,4	1,4	100,0
	Gesamtsumme	3435	98,7	100,0	
Fehlend	KEINE ANGABE	45	1,3		
Gesamtsumme		3480	100,0		

# Aufgabe 4b: Lösung



Beschriftung X-Achse (1)  
Beschriftung Y-Achse (1)  
„Treppenform“ (1)  
Werte „getroffen“ /  
sauber (1)

# Aufgabe 5

10 Personen wurden befragt, wie viele Kinder sie haben. Dabei kam folgende Häufigkeitstabelle heraus:

**Anzahl Kinder**

	Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig 0	4	40,0	40,0	40,0
1	2	20,0	20,0	60,0
2	2	20,0	20,0	80,0
3	1	10,0	10,0	90,0
4	1	10,0	10,0	100,0
Gesamtsumme	10	100,0	100,0	

- Bestimmen Sie Modus, Median und arithmetisches Mittel. (3)
- Interpretieren Sie die soeben bestimmten Maßzahlen. (3)
- Was können Sie mithilfe der Daumenregel für Lagemaße über die Schiefe der Verteilung sagen? (2)
- Berechnen Sie den Interquartilsabstand und interpretieren Sie diesen! (4)

# Aufgabe 5a: Lösung

- Modus (häufigste Kategorie):
  - $h = 0$
- Median (mittlerer Wert):
  - $\tilde{x} = 1$
- arithmetisches Mittel:
  - $\bar{x} = \frac{1}{n} \sum_{j=1}^J x_j * n_j$
  - $\bar{x} = \frac{(0*4)+(1*2)+(2*2)+(3*1)+(4*1)}{10} = 1,3$

		Anzahl Kinder			
		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	0	4	40,0	40,0	40,0
	1	2	20,0	20,0	60,0
	2	2	20,0	20,0	80,0
	3	1	10,0	10,0	90,0
	4	1	10,0	10,0	100,0
	Gesamtsumme	10	100,0	100,0	

# Aufgabe 5b: Lösung

- Modus:
  - $h = 0$
  - Die meisten Befragten haben keine Kinder. (1)
- Median (mittlerer Wert):
  - $\tilde{x} = 1$
  - Der mittlere Befragte hat 1 Kind. (1)
- arithmetisches Mittel:
  - $\bar{x} = 1,3$
  - Der durchschnittliche Befragte hat 1,3 Kinder (1)

# Aufgabe 5c: Lösung

- Daumenregel (2)
  - $Modus < Median < arithmetisches\ Mittel$
  - Die Verteilung ist rechtsschief
  - Viele Befragte haben eine geringe Kinderzahl, während nur eine Minderheit eine größere Kinderzahl hat.

# Aufgabe 5d: Lösung

- Interquartilsabstand berechnen:
  - $Q_{0,25} = 0$  (1)
  - $Q_{0,75} = 2$  (1)
  - $IQR = Q_{0,75} - Q_{0,25}$
  - $IQR = 2 - 0 = 2$  (1)
- Interquartilsabstand
  - Die mittlerer 50% der Fälle streuen über ein Bereich von 2 Einheiten (hier zwischen 0 und 2 Kindern) (1)

		Anzahl Kinder			
		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	0	4	40,0	40,0	40,0
	1	2	20,0	20,0	60,0
	2	2	20,0	20,0	80,0
	3	1	10,0	10,0	90,0
	4	1	10,0	10,0	100,0
	Gesamtsumme	10	100,0	100,0	

# Aufgabe 6: z-Transformation

- a) Wie verändert sich der Mittelwert einer Variablen, wenn man diese zentriert? Was geschieht mit der Standardabweichung?
- b) Wie verändert sich der Mittelwert einer Variablen, wenn man diese anhand der Standardabweichung normiert?
- c) Wie verändert sich der Mittelwert einer Variablen, wenn man diese mit Hilfe einer z-Transformation standardisiert? Was geschieht mit der Standardabweichung?



# Aufgabe 6: Lösung

- a) Durch die Zentrierung hat die neu gebildete Variable einen Mittelwert von 0, die Standardabweichung verändert sich nicht.
- b) Durch die Normierung erhält die neue Variable eine Standardabweichung von 1.
- c) Bei der z-Transformation (Standardisierung) erhält die neue Variable einen Mittelwert von 0 und eine Standardabweichung von 1.

# Aufgabe 7: Wartezeit

10 Menschen wurden Befragt, wie lange sie heute beim Brötchenkaufen in der Schlange standen.

Es ergaben sich folgende Werte:

Person	1	2	3	4	5	6	7	8	9	10
Wartezeit in Minuten	2	4	4	4	6	10	2	0	2	6

- Bestimmen Sie Modus, Median und arithmetisches Mittel!
- Bestimmen Sie das 25-%-, 75-%-Quantil und den Interquartilsabstand.
- Zeichnen Sie den entsprechenden Boxplot.

# Aufgabe 7a: Lösung

Person	1	2	3	4	5	6	7	8	9	10
Wartezeit in Minuten	2	4	4	4	6	10	2	0	2	6

- Modus:
  - Die Verteilung hat die zwei Modalwerte 2 und 4.
- arithmetisches Mittel:
  - $\bar{x} = \frac{2+4+4+4+6+10+2+0+2+6}{10}$
  - $\bar{x} = 4$

# Aufgabe 7a: Lösung II

- Für den Median müssen wir die Verteilung sortieren:

geordnete Position $x(k)$	1	2	3	4	5	6	7	8	9	10
Wartezeit in Minuten	0	2	2	2	4	4	4	6	6	10

- $\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$
- $\tilde{x} = \frac{x_{(5)} + x_{(6)}}{2}$
- $\tilde{x} = \frac{4+4}{2} = 4$

# Aufgabe 7b: Lösung

geordnete Position $x(k)$	1	2	3	4	5	6	7	8	9	10
Wartezeit in Minuten	0	2	2	2	4	4	4	6	6	10

- 25-%-Quantil:
  - $k = n * \alpha$
  - $k = 10 * 0,25 = 2,5 \rightarrow 3$
  - $x_{(k)} = 2$
- 75-%-Quantil:
  - $k = n * \alpha$
  - $k = 10 * 0,75 = 7,5 \rightarrow 8$
  - $x_{(k)} = 6$

# Aufgabe 7b: Lösung II

- IQR:

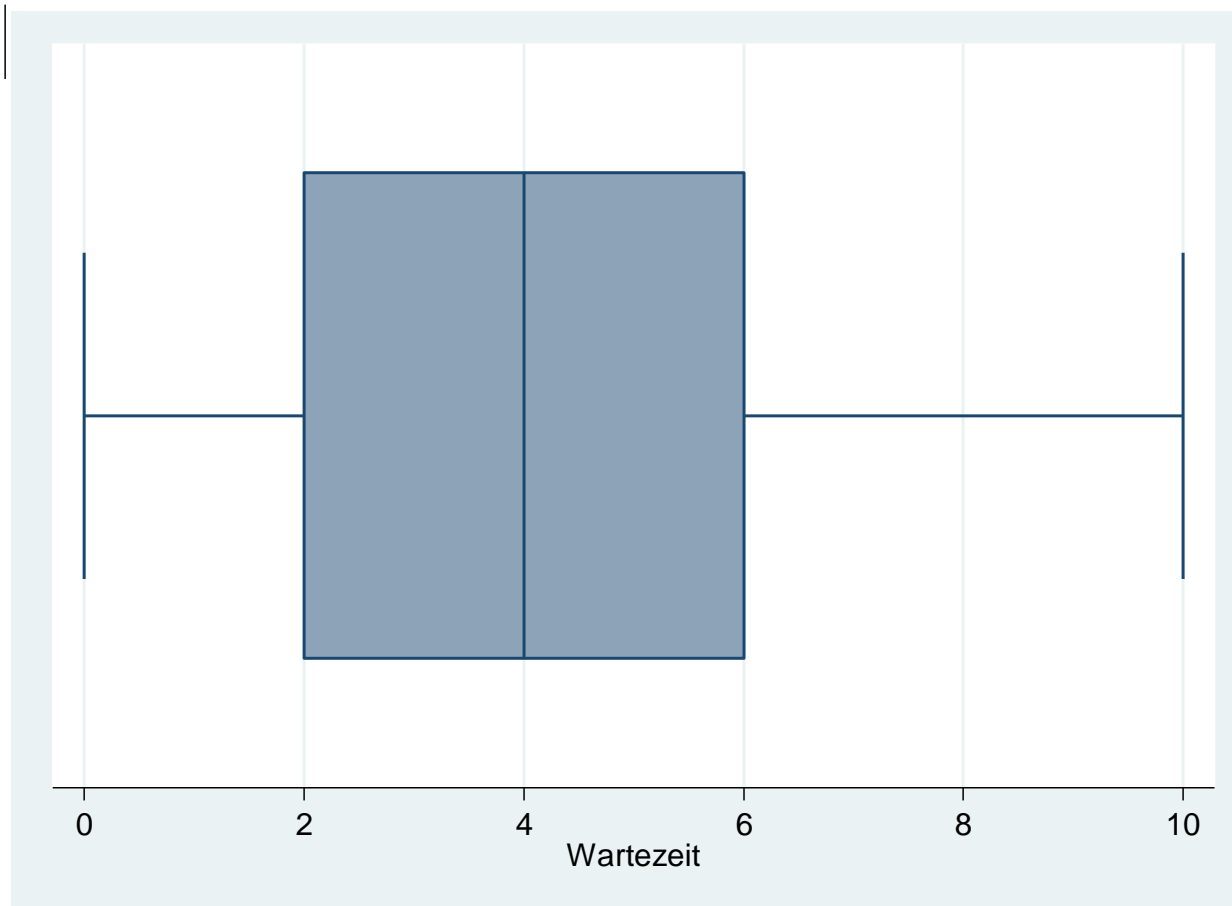
- $IQR = Q_{0,75} - Q_{0,25}$

- $IQR = 6 - 2 = 4$

# Aufgabe 7c: Lösung

- gegeben:
  - $Q_{0,75} = 6$
  - $Q_{0,25} = 2$
  - $IQR = 6 - 2 = 4$
  - $\tilde{x} = 4$
- unteres Barthaar:
  - minimal:  $Q_{0,25} - 1,5 * IQR = 2 - 4 * 1,5 = -4$
  - tatsächlicher unterster Wert im Intervall: 0 → wir zeichnen bis 0
- oberes Barthaar:
  - maximal:  $Q_{0,75} + 1,5 * IQR = 6 + 4 * 1,5 = 12$
  - tatsächlicher höchster Wert im Intervall: 10 → wir zeichnen bis 10
- Extremwerte / Ausreißer
  - liegen hier nicht vor!

# Aufgabe 7c: Lösung II





# KREUZTABELLEN

---

# Aufgabe 8: Wahlbeteiligung

Ein Soziologe interessiert sich dafür, ob einen Zusammenhang zwischen dem politischen Interesse eines Befragten und der Wahlbeteiligung gibt. Mithilfe von SPSS erhält er folgende Kreuztabelle:

wahlbeteiligung, letzte bundestagswahl?	Politisches Interesse			Total
	schwach	mittel	stark	
ja	284	544	424	1,252
nein	194	71	37	302
Total	478	615	461	1,554

- Welches Maß ist zur Untersuchung des Zusammenhangs geeignet, wenn Sie sich nicht festlegen wollen, was die x- und y-Variable ist?
- Berechnen Sie dieses Maß.
- Interpretieren Sie Ihr Ergebnis inhaltlich und statistisch.

# Aufgabe 8a: Lösung

symmetrische  
Fragestellung:  
Zusammenhang

Ein Soziologe interessiert sich dafür, ob einen Zusammenhang zwischen dem politischen Interesse eines Befragten und der Wahlbeteiligung gibt.

politisches Interesse:  
ordinal  
Wahlbeteiligung:  
nominal

wahlbeteiligung, letzte bundestagswahl?	Politisches Interesse			Total
	schwach	mittel	stark	
ja	284	544	424	1,252
nein	194	71	37	302
Total	478	615	461	1,554

geeignetes Maß:  
Cramér's V

# Aufgabe 8b: Lösung

<i>Zeile i</i>	<i>Spalte j</i>	$n_{ij}$	$e_{ij}$	$n_{ij} - e_{ij}$	$(n_{ij} - e_{ij})^2$	$\frac{(n_{ij} - e_{ij})^2}{e_{ij}}$
1	1	284	385,1	-101,1	10221,21	26,5417035
1	2	544	495,5	48,5	2352,25	4,74722503
1	3	424	371,4	52,6	2766,76	7,44954227
2	1	194	92,9	101,1	10221,21	110,023789
2	2	71	119,5	-48,5	2352,25	19,6841004
2	3	37	89,6	-52,6	2766,76	30,8790179
					$\chi^2 =$	<b>199,33</b>

# Aufgabe 8b: Lösung

- Berechnung Cramérs  $V$

- $V = \sqrt{\frac{\chi^2}{n \cdot \min(I-1, J-1)}}$

- $V = \sqrt{\frac{199,33}{1554 \cdot \min(3-1, 2-1)}}$

- $V = \sqrt{\frac{199,33}{1554 \cdot 1}}$

- $V = 0,358$

# Aufgabe 8c: Lösung

- Interpretation:
  - $V = 0,3582$
  - Es besteht ein mittlerer Zusammenhang zwischen der x-Variablen und der y-Variablen.
  - Es besteht ein mittlerer Zusammenhang zwischen dem politischen Interesse und der Wahlteilnahme.

# Aufgabe 9: Abtreibung

Ein Wissenschaftler interessiert sich dafür, ob die Zustimmung auf die Frage, dass eine Abtreibung möglich sein sollte, wenn die Frau es will, vom Erhebungsgebiet der Befragten abhängt. Mithilfe von SPSS erhält er folgende Kreuztabelle:

**Kreuztabelle ABTREIBUNG - WENN DIE FRAU ES WILL \*ERHEBUNGSGBIET <WOHNGBIET>: WEST - OST**

Anzahl

		ERHEBUNGSGBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLAE NDER	NEUE BUNDESLAE NDER	
ABTREIBUNG - WENN DIE FRAU ES WILL	JA,MOEGlich SEIN	409	313	722
	NEIN,NICHT MOEGlich	693	219	912
Gesamtsumme		1102	532	1634

- Welches Maß ist hier geeignet? Begründen Sie Ihre Wahl! (2)
- Berechnen Sie dieses Maß. (6)
- Interpretieren Sie ihr Ergebnis inhaltlich und statistisch. (2)

# Aufgabe 9a: Analyse

asymmetrische  
Fragestellung:  
Erhebungsgebiet  
(X) → Einstellung (Y)

Erhebungsgebiet:  
nominal  
Einstellung:  
nominal  
Vierfeldertafel

optimales Maß:  
Prozentsatzdifferenz  
 $d_{YX}\%$   
auch möglich:  
 $\lambda_{YX}$

Ein Wissenschaftler interessiert sich dafür, ob die Zustimmung auf die Frage, dass eine Abtreibung möglich sein sollte, wenn die Frau es will, vom Erhebungsgebiet der Befragten abhängt

Kreuztabelle ABTREIBUNG - WENN DIE FRAU ES WILL \*ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

Anzahl

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLAE NDER	NEUE BUNDESLAE NDER	
ABTREIBUNG - WENN DIE FRAU ES WILL	JA, MOEGlich SEIN	409	313	722
	NEIN, NICHT MOEGlich	693	219	912
Gesamtsumme		1102	532	1634



# Aufgabe 9b und 9c: Lösung

- Berechnung: Prozentsatzdifferenz

- $d_{YX}\% = 100 * \left( \frac{a}{a+c} - \frac{b}{b+d} \right) (2)$

- $d_{YX}\% = 100 * \left( \frac{409}{1102} - \frac{313}{532} \right) (2)$

- $d_{YX}\% = -21,72 (2)$

- Interpretation:

- Die Prozentsatzdifferenz beträgt -21,72 Prozentpunkte.

- Befragte, die aus Westdeutschland kommen, stimmen der These, dass Abtreibung möglich sein sollte, wenn die Frau es will, deutlich seltener zu als ostdeutsche Befragte.

# Aufgabe 9c: Lösung

Kreuztabelle ABTREIBUNG - WENN DIE FRAU ES WILL \*ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

			ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
			ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
ABTREIBUNG - WENN DIE FRAU ES WILL	JA,MOEGlich SEIN	Anzahl % in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST	409 37,1%	313 58,8%	722 44,2%
	NEIN,NICHT MOEGlich	Anzahl % in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST	693 62,9%	219 41,2%	912 55,8%
Gesamtsumme		Anzahl % in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST	1102 100,0%	532 100,0%	1634 100,0%

- Ostdeutsche stimmen um 21,72 Prozentpunkte häufiger der These zu, dass Abtreibung für Frauen möglich sein sollte, wenn die Frau es will, als Westdeutsche. (2)

# Aufgabe 9b: Alternative Lambda

- Berechnung  $\lambda_{YX}$ :
  - $\lambda_{YX} = PRE = \frac{E_0 - E_1}{E_0}$
  - $E_0 = 1634 - 912$  (1)
  - $E_0 = 722$  (1)
  - $E_1 = (1102 - 693) + (532 - 313)$  (1)
  - $E_1 = 628$  (1)
  - $\lambda_{YX} = \frac{722 - 628}{722}$  (1)
  - $\lambda_{YX} = 0,130$  (1)

# Aufgabe 9c: Lösung Lambda II

- Interpretation:
  - $\lambda_{YX} = 0,130$
  - Das bedeutet, dass sich durch Berücksichtigung des Erhebungsgebietes die Prognose der Einstellung zur Abtreibung, wenn die Frau es will, um 13,0% verbessern lässt im Gegensatz zur Nichtberücksichtigung.

# Aufgabe 10

Eine Wissenschaftlerin möchte wissen, ob sich ein Einfluss des Geschlechts auf das politische Interesse nachweisen lässt. Mithilfe von Stata erhält sie folgende Kreuztabelle:

**Kreuztabelle POLITSCHES INTERESSE (REK)\*GESCHLECHT, BEFRAGTE<R>**

Anzahl

		GESCHLECHT, BEFRAGTE<R>		Gesamtsum me
		MANN	FRAU	
POLITSCHE INTERESSE (REK)	STARK	659	325	984
	MITTEL	636	751	1387
	SCHWACH	430	679	1109
Gesamtsumme		1725	1755	3480

- Welches Maß ist hier geeignet? Begründen Sie Ihre Wahl! (2)
- Berechnen Sie dieses Maß. (6)
- Interpretieren Sie ihr Ergebnis inhaltlich und statistisch. (2)

# Aufgabe 10a: Analyse

asymmetrische  
Fragestellung:  
Geschlecht  
(X) → politisches  
Interesse (Y)

Eine Wissenschaftlerin möchte wissen, ob sich ein Effekt des Geschlechts auf das politische Interesse nachweisen lässt.

Geschlecht: nominal  
politisches Interesse:  
ordinal  
Mehrfeldertafel

Kreuztabelle POLITSCHES INTERESSE (REK)\*GESCHLECHT, BEFRAGTE<R>

Anzahl

		GESCHLECHT, BEFRAGTE<R>		Gesamtsumme
		MANN	FRAU	
POLITSCHE INTERESSE (REK)	STARK	659	325	984
	MITTEL	636	751	1387
	SCHWACH	430	679	1109
Gesamtsumme		1725	1755	3480

geeignetes Maß:  
 $\lambda_{YX}$

# Aufgabe 10b: Lösung

- Berechnung  $\lambda_{YX}$ :
  - $\lambda_{YX} = PRE = \frac{E_0 - E_1}{E_0}$
  - $E_0 = 3480 - 1387(1)$
  - $E_0 = 2093(1)$
  - $E_1 = (1725 - 659) + (1755 - 751) (1)$
  - $E_1 = 2070(1)$
  - $\lambda_{YX} = \frac{2093 - 2070}{2093} (1)$
  - $\lambda_{YX} = 0,0110 (1)$

# Aufgabe 10c: Lösung

- Interpretation:
  - $\lambda_{YX} = 0,0110$
  - Das bedeutet, dass sich durch Berücksichtigung des Geschlechts die Prognose des politischen Interesses um lediglich 1,1% verbessern lässt. Nicht sonderlich viel. (2)



# Aufgabe 11: Schichtzuordnung

Eine Wissenschaftlerin interessiert sich dafür, ob der Schulabschluss eines Befragten einen Einfluss auf die subjektive Schichteinstufung hat. Mithilfe von Stata erhält sie folgende Tabelle:

subjektive Schichteinstufung	höchster Schulabschluss			Total
	max. Haup	mittlere	Abitur	
Unterschicht / Arbeit	498	435	109	1,042
Mittelschicht	610	786	934	2,330
Oberschicht	3	1	16	20
Total	1,111	1,222	1,059	3,392

- Welches Maß ist hier geeignet? Begründen Sie Ihre Wahl! (2)
- Berechnen Sie dieses Maß. (6)
- Interpretieren Sie Ihr Ergebnis inhaltlich und statistisch. (2)

# Aufgabe 11a: Lösung

asymmetrische  
Fragestellung:  
Einfluss Schulabschluss  
(X) → Schicht(Y)

Eine Wissenschaftlerin interessiert sich dafür, ob der Schulabschluss eines Befragten einen Einfluss auf die subjektive Schichteinstufung hat.

beide Variablen sind  
ordinal

subjektive Schichteinstufung	höchster Schulabschluss			Total
	max. Haup	mittlere	Abitur	
Unterschicht / Arbeit	498	435	109	1,042
Mittelschicht	610	786	934	2,330
Oberschicht	3	1	16	20
Total	1,111	1,222	1,059	3,392

geeignetes Maß:  
Somers  $d_{YX}$

# Aufgabe 11b: Lösung

- Berechnung Paare:

- $C = 498 * (786 + 934 + 1 + 16) + 435 * (934 + 16) + 610 * (1 + 16) + 786 * 16 = 1.301.222$  (1)

- $D = 109 * (610 + 786 + 3 + 1) + 435 * (610 + 3) + 934 * (3 + 1) + 786 * 3 = 425.349$  (1)

- $T_Y = 498 * (435 + 109) + 435 * 109 + 610 * (786 + 934) + 786 * 934 + 3 * (1 + 16) + 1 * 16 = 2.101.718$

# Aufgabe 11b: Lösung II

- Berechnung Somers  $d_{YX}$ :

- $d_{YX} = \frac{C-D}{C+D+T_Y} \quad (1)$

- $d_{YX} = \frac{1301222-425349}{1301222+425349+2101718} \quad (1)$

- $d_{YX} = 0,2288 \quad (1)$

# Aufgabe 11c: Lösung

- Interpretation:
  - $d_{YX} = 0,2288$
  - Es besteht ein niedriger positiver Einfluss des Schulabschlusses auf die subjektive Schichteinstufung (1)
  - Leute mit höherem Schulabschluss stufen sich tendenziell etwas höher ein. (1)

# Aufgabe 12: Schulabschluss

Ein Wissenschaftler möchte wissen, ob es einen Zusammenhang zwischen dem höchsten Schulabschluss eines Befragten und der Einschätzung der persönlichen wirtschaftlichen Lage gibt. Hinsichtlich der Wirkrichtung möchte er sich nicht festlegen, da beide plausibel sind. Er erhält folgende Tabelle:

Wirtschaftslage Befragter heute	höchster Schulabschluss			Total
	max. Haupt	mittlere	Abitur	
eher gut	616	690	767	2,073
teils / teils	366	398	237	1,001
eher schlecht	152	143	63	358
Total	1,134	1,231	1,067	3,432

- Welches Maß ist zur Untersuchung des Zusammenhangs geeignet, wenn Sie an einer PRE-Aussage interessiert sind?
- Berechnen Sie dieses Maß.
- Interpretieren Sie Ihr Ergebnis inhaltlich und statistisch.

# Aufgabe 12a: Lösung

symmetrische  
Fragestellung:  
Zusammenhang

Ein Wissenschaftler möchte wissen, ob es einen Zusammenhang zwischen dem höchsten Schulabschluss eines Befragten und der Einschätzung der persönlichen wirtschaftlichen Lage gibt

politisches Interesse:  
ordinal  
Wahlbeteiligung:  
ordinal

Wirtschaftslage Befragter heute	höchster Schulabschluss			Total
	max. Haupt	mittlere	Abitur	
eher gut	616	690	767	2,073
teils / teils	366	398	237	1,001
eher schlecht	152	143	63	358
Total	1,134	1,231	1,067	3,432

PRE-Maß gesucht:  
Gamma

# Aufgabe 12b: Lösung

- Berechnung:

- $C = 616 * (398 + 237 + 143 + 63) + 690 * (237 + 63) + 366 * (143 + 63) + 398 * 63$

- $C = 825526$

- $D = 767 * (366 + 398 + 152 + 143) + 690 * (366 + 152) + 237 * (152 + 143) + 398 * 152$

- $D = 1300084$

- $\gamma = \frac{C-D}{C+D}$

- $\gamma = \frac{825526-1300084}{825526+1300084}$

- $\gamma = -0,223$



# Aufgabe 12c: Lösung

- Interpretation

- $\gamma = -0,223$
- Es besteht ein niedriger negativer Zusammenhang zwischen beiden Variablen. Leute mit höherem Schulabschluss schätzen ihre persönliche Lage tendenziell besser ein und umgekehrt
- Da Gamma zugleich ein PRE-Maß ist, lässt sich sagen, dass durch Kenntnis des Schulabschlusses die Vorhersage der persönlichen Lage um 22,3% verbessert werden kann und umgekehrt.

# KORRELATIONEN UND GEMISCHTE SKALENNIVEAUS

---

# Aufgabe 13

In der folgenden Tabelle ist die vergebene Punktzahl von zwei Gutachtern für insgesamt vier Statistikklausuren wiedergegeben:

Klausur	A	B	C	D
Punkte Gutachter 1	20	30	90	60
Punkte Gutachter 2	40	20	80	50

Die Arbeiten werden anschließend entsprechend der vergebenen Punkte in eine Rangreihenfolge gebracht. Inwieweit stimmen beide Gutachter hinsichtlich ihrer vergebenen Ränge überein?

- Welches Maß ist hier geeignet? Begründen Sie Ihre Wahl.
- Berechnen Sie dieses Maß! (6)
- Interpretieren Sie Ihr Ergebnis. (2)

# Aufgabe 13a: Lösung

symmetrische  
Fragestellung:  
Zusammenhang

. Inwieweit stimmen beide Gutachter hinsichtlich ihrer vergebenen Ränge überein?

zunächst metrisches  
Skalenniveau, soll in  
Ränge transformiert  
werden → ordinal

geeignetes Maß:  
 $r_{SP}$  → Kurzformel, da  
keine Verknüpfungen

Klausur	A	B	C	D
Punkte Gutachter 1	20	30	90	60
Punkte Gutachter 2	40	20	80	50

# Aufgabe 13b: Lösung

- Formel:

$$\bullet r_{SP} = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)}$$

- Ränge ermitteln (2):

$x_i$	$Rang(x_i)$	$y_i$	$Rang(y_i)$
20	1	40	2
30	2	20	1
90	4	80	4
60	3	50	3

# Aufgabe 13b: Lösung II

- Summe der quadrierten Abweichungen der Ränge ermitteln (1):

$x_i$	$Rang(x_i)$	$y_i$	$Rang(y_i)$	$d_i^2 = (Rang(x_i) - Rang(y_i))^2$
20	1	40	2	1
30	2	20	1	1
90	4	80	4	0
60	3	50	3	0
			$\sum d_i^2 =$	2

# Aufgabe 13b: Lösung III

- $r_{SP} = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)} \quad (1)$
- $r_{SP} = 1 - \frac{6 \cdot 2}{4 \cdot (16 - 1)} \quad (1)$
- $r_{SP} = 0,8 \quad (1)$

# Aufgabe 13c: Lösung

- Interpretation:
  - $r_{SP} = 0,8$
  - Es besteht eine hohe positive Rangkorrelation.
  - Kandidaten, die vom ersten Gutachter einen hohen Rang zugewiesen bekommen haben, haben tendenziell auch vom zweiten Gutachter einen hohen Rang zugewiesen bekommen und umgekehrt.(2)



# Aufgabe 14: Arbeitszeit

Ein Wissenschaftler interessiert sich dafür, ob das Geschlecht einen Einfluss auf die durchschnittliche Arbeitszeit pro Woche in Stunden ausübt. Mithilfe von SPSS erhält er folgende Tabelle:

## Bericht

BEFRAGTER: ARBEITSSTUNDEN PRO WOCHE

GESCHLECHT, BEFRAGTE<R>	Mittelwert	H	Varianz
MANN	44,085	1008	76,796
FRAU	35,173	857	134,022
Gesamtsumme	39,990	1865	122,769

- Welches Maß ist zur Untersuchung des Zusammenhangs geeignet?
- Berechnen Sie dieses Maß.
- Interpretieren Sie Ihr Ergebnis inhaltlich und statistisch.

# Aufgabe 14a: Lösung

asymmetrische  
Fragestellung:  
Geschlecht (X) →  
Arbeitszeit (Y)

Ein Wissenschaftler interessiert sich dafür, ob das Geschlecht einen Einfluss auf die durchschnittliche Arbeitszeit pro Woche in Stunden ausübt.

Geschlecht:  
nominal  
Arbeitszeit:  
metrisch

## Bericht

BEFRAGTER: ARBEITSSTUNDEN PRO WOCHE

GESCHLECHT, BEFRAGTE<R>	Mittelwert	H	Varianz
MANN	44,085	1008	76,796
FRAU	35,173	857	134,022
Gesamtsumme	39,990	1865	122,769

geeignetes Maß:  
Eta/Eta<sup>2</sup>

# Aufgabe 14b: Lösung

- Berechnung:

- $\eta^2 = \frac{E_0 - E_1}{E_1}$
- $E_0 = SAQ_Y = SAQ_{Gesamt}$
- $E_0 = (n_{Gesamt} - 1) * \hat{\sigma}_{Gesamt}^2$
- $E_0 = (1865 - 1) * 122,769$
- $E_0 = 228841,416$

## Bericht

BEFRAGTER: ARBEITSSTUNDEN PRO WOCHE

GESCHLECHT, BEFRAGTE<R>	Mittelwert	H	Varianz
MANN	44,085	1008	76,796
FRAU	35,173	857	134,022
Gesamtsumme	39,990	1865	122,769

# Aufgabe 14b: Lösung II

- Berechnung:

- $E_1 = SAQ_{Mann} + SAQ_{Frau}$

- $E_1 = \left( (n_{Mann} - 1) * \hat{\sigma}_{Mann}^2 \right) + \left( (n_{Frau} - 1) * \hat{\sigma}_{Frau}^2 \right)$

- $SAQ_{Mann} = (1008 - 1) * 76,7956$

- $SAQ_{Mann} = 77333,1692$

- $SAQ_{Frau} = (857 - 1) * 134,022$

- $SAQ_{Frau} = 114722,832$

- $E_1 = 77333,1692 + 114722,832$

- $E_1 = \mathbf{192056,0012}$

## Bericht

BEFRAGTER: ARBEITSSTUNDEN PRO WOCHE

GESCHLECHT, BEFRAGTE<R>	Mittelwert	H	Varianz
MANN	44,085	1008	76,796
FRAU	35,173	857	134,022
Gesamtsumme	39,990	1865	122,769

# Aufgabe 14b: Lösung III

- Berechnung  $\eta^2$

- $\eta^2 = PRE = \frac{E_0 - E_1}{E_1}$

- $\eta^2 = \frac{228841,416 - 192056,0012}{228841,416}$

- $\eta^2 = 0,1607$

- Berechnung  $\eta$

- $\eta = \sqrt{\eta^2}$

- $\eta = 0,4009$

- $E_0 = 228841,416$
- $E_1 = 192056,0012$

# Aufgabe 14c: Lösung

- Interpretation  $\text{Eta}^2$ 
  - Durch Kenntnis des Geschlechts lässt sich die Prognose der Wochenarbeitszeit um 16,1% verbessern.
- Interpretation  $\text{Eta}$ :
  - Es besteht ein mittlerer Einfluss des Geschlechts auf die Wochenarbeitszeit
- Interpretation Mittelwerte:
  - Anhand der Mittelwerte lässt sich ablesen, dass Männer mit 44h im Durchschnitt deutlich länger arbeiten als Frauen mit 35h.

# Aufgabe 15: Sportturnier

Ein Sportsoziologe möchte wissen, ob die Zahl der Trainingsstunden einen Einfluss auf die erzielten Turniersiege hat. Für 5 Sportler erhält er folgende Daten:

Sportler	A	B	C	D	E
Trainingsstunden	0	2	4	6	8
Turniersiege	2	4	3	4	10

- Welches Modell ist hier zur Untersuchung des Zusammenhangs geeignet? Geben Sie das allgemeine mathematische Modell an. (2)
- Wenden Sie das Verfahren an (6)
- Interpretieren Sie Ihr Ergebnis (2)

# Aufgabe 15: Analyse

asymmetrische  
Fragestellung:  
Einfluss  
Trainingsstunden  
(X) → Siege (Y)

Ein Sportsoziologe möchte wissen, ob die Zahl der Trainingsstunden einen Einfluss auf die erzielten Turniersiege hat.

beide Variablen metrisch

Sportler	A	B	C	D	E
Trainingsstunden	0	2	4	6	8
Turniersiege	2	4	3	4	10

geeignetes Maß: lineare  
Regression



# Aufgabe 15a: Lösung

- geeignetes Verfahren:
  - bivariate lineare Regression (1)
  - $y = b_0 + b_1 * x_1 + e$  (1)

# Aufgabe 15b: Lösung

$x$	$y$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) * (y_i - \bar{y})$
0	2	-4	16	-2,6	6,76	10,4
2	4	-2	4	-0,6	0,36	1,2
4	3	0	0	-1,6	2,56	0
6	4	2	4	-0,6	0,36	-1,2
8	10	4	16	5,4	29,16	21,6
$\bar{x} = 4$	$\bar{y} = 4,6$		$SAQ_X = 40$		$SAQ_Y = 39,2$	$SP_{XY} = 32$

# Aufgabe 15b: Lösung

- Berechnung:

- $b_1 = \frac{SP_{XY}}{SAQ_X}$

- $b_1 = \frac{32}{40} = 0,8 \text{ (1)}$

- $b_0 = \bar{y} - b_1 * \bar{x}$

- $b_0 = 4,6 - 0,8 * 4$

- $b_0 = 1,4 \text{ (1)}$

- $\bar{x} = 4$
- $\bar{y} = 4,6$
- $SAQ_X = 40$
- $SAQ_Y = 39,2$
- $SP_{XY} = 32$

# Aufgabe 15c: Lösung

- Interpretation Regressionsgewicht:
  - $b_1 = 0,8$
  - Wenn  $x$  um eine Einheit steigt, dann steigt  $y$  im Durchschnitt um 0,8 Einheiten. Das heißt also, wenn die Anzahl der Trainingsstunden um eine Stunde steigt, erhöht sich die Anzahl der Turniersiege um durchschnittlich 0,8 Siege.
- Interpretation Regressionskonstante:
  - $b_0 = 1,4$
  - Die Regressionskonstante liegt bei 1,4 und gibt den Schnittpunkt mit der  $y$ -Achsen an. Wenn  $x=0$  wäre, dann würde für  $y$  ein Wert von 1,4 vorhergesagt. Folglich würde bei 0 Trainingsstunden eine Anzahl von 1,4 Siegen vorhergesagt.

# Aufgabe 16: Kindersterblichkeit

Ein OECD-Forscher interessiert sich dafür, ob die wahrgenommene Wasserqualität in einem Land einen Einfluss auf die Kindersterblichkeit hat. Er greift auf Daten der Studie „Society at a Glance 2011“ für 34 Länder zurück und erhält folgende Tabelle:

Variable	Mittelwert	Standardabweichung
Zufriedenheit mit der Wasserqualität (in %)	86,074	10,0184
Kindersterblichkeit pro 1.000 Lebendgeburten	4,612	3,1834

Darüber hinaus ist bekannt, dass  $\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) = -16,511$  liegt.

- Berechnen Sie die Regressionskonstante  $b_0$  und das Regressionsgewicht  $b_1$ . Interpretieren Sie beide Koeffizienten.
- Berechnen Sie das standardisierte Regressionsgewicht und interpretieren Sie dieses.
- Welcher Wert wird für ein Land vorhergesagt, in dem die Zufriedenheit mit der Wasserqualität bei 90% liegt?
- Wie gut eignet sich das Modell zur Prognose der Kindersterblichkeit? Berechnen Sie  $R^2$  und interpretieren Sie ihr Ergebnis!

# Aufgabe 16a: Lösung

- gegeben:

- $\bar{x} = 86,074$

- $\bar{y} = 4,612$

- $s_x = 10,0184$

- $s_y = 3,1834$

- $s_{x,y} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) = -16,511$

- Berechnung:

- $b_1 = \frac{s_{x,y}}{s_y} = \frac{-16,511}{10,0184^2}$

- $b_1 = -0,165$

- $b_0 = \bar{y} - b_1 * \bar{x} = 86,074 - \frac{-16,511}{10,0184^2} * 4,612$

- $b_0 = 18,772$

# Aufgabe 16a: Lösung II

- Interpretation Regressionskonstante:
  - $b_0 = 18,772$
  - Für ein Land mit einer wahrgenommenen Zufriedenheit mit der Wasserqualität von 0% würde eine Kindersterblichkeit von 18,772 Kindern pro 1.000 Lebendgeburten vorhergesagt. Das ist hier natürlich ein hypothetischer Wert, da die Zufriedenheit mit der Wasserqualität wohl in kaum einem Land bei 0% liegt.
- Interpretation Regressionsgewicht:
  - $b_1 = -0,165$
  - Für jede zusätzliche Einheit von X sinkt Y im Durchschnitt um 1,648 Einheiten.
  - Für jedes zusätzliche Prozent Zufriedenheit mit der Wasserqualität in einem Land erfolgt eine durchschnittliche Verringerung der Kindersterblichkeit um 0,165 Kinder pro 1000 Lebendgeburten.

# Aufgabe 16b: Lösung

- gegeben:

- $s_x = 10,0184$

- $s_y = 3,1834$

- $b_1 = -0,165$

- Berechnung:

- $b_1^* = b_1 * \frac{s_x}{s_y}$

- $b_1^* = -0,165 * \frac{10,0184}{3,1834}$

- $b_1^* = -0,519$

- Interpretation:



# Aufgabe 16b: Lösung II

- Interpretation:
  - $b_1^* = -0,519$
  - Wenn die Zufriedenheit mit der Wasserqualität in einem Land um eine Standardabweichung steigt, so fällt die Kindersterblichkeit im Durchschnitt um 0,519 Standardabweichungen.
  - Da wir es mit einer bivariaten Regression zutun haben, entspricht dieser Wert auch Pearsons  $r$ . Dies bedeutet, dass ein hoher negativer Zusammenhang zwischen der Zufriedenheit mit Wasserqualität und der Kindersterblichkeit besteht. Je besser die Zufriedenheit mit der Wasserqualität in einem Land, umso niedriger ist tendenziell auch die Kindersterblichkeit.

# Aufgabe 16c: Lösung

- Berechnung:

- $b_1^* = r_{X,Y} = -0,519$
- $R^2 = r_{x,y}^2 = -0,519^2$
- $R^2 = 0,269$

- Interpretation:

- Durch Kenntnis der Zufriedenheit mit Wasserqualität in einem Land lässt sich die Prognose der Kindersterblichkeit um 26,9% verbessern.

# Aufgabe 16: Lösung SPSS

**Koeffizienten<sup>a</sup>**

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten
		B	Standardfehler	Beta
1	(Konstante)	18,772	4,164	
	Zufriedenheit mit Wasserqualität in Prozent	-,165	,048	-,518

a. Abhängige Variable: Kindersterblichkeit pro 1000 Lebendgeburten

**Modellübersicht**

Modell	R	R-Quadrat	Angepasstes R-Quadrat	Standardfehler der Schätzung
1	,518 <sup>a</sup>	,268	,245	2,7657

a. Prädiktoren: (Konstante), Zufriedenheit mit Wasserqualität in Prozent