Quatember, A. (2014). *Statistik ohne Angst vor Formeln*. 4. Auflage, Pearson Studium, München.

Lösungen zu den Übungsaufgaben:

Kapitel 1 Beschreibende Statistik

Aufgabe 1: [Merkmalstypen]

Welchen Merkmalstypen (nach den Einteilungskriterien "metrisch - ordinal - nominal" beziehungsweise "diskret- stetig") gehören die folgenden Merkmale an?

a) Länge von Videobändern einer Produktion in cm
b) Reiseziel von befragten Urlaubsbuchern
c) Güteklassen von Obst am Markt
d) Inflationsrate verschiedener Länder
e) Religion von Befragten
f) Heizungsart von Mietwohnungen in einer Stadt
g) Anzahl an Kinobesuchen von Schülern einer Schule in den Ferien
h) Einstellung von Befragten zur Einführung einer Berufsarmee (Antwortalternativen: sehr skeptisch, eher skeptisch, unentschieden, eher positiv, sehr positiv)

Lösung

a) m s
b) n d
c) o d
d) m s
e) n d
f) n d
g) m d
h) o d

Aufgabe 2: [Merkmalstypen]

Welchen Merkmalstypen gehören die folgenden Merkmale an und wie sind eventuelle Kodierungen der Merkmalsausprägungen vorzunehmen?

a) Fußballinteresse von Befragten (Merkmalsausprägungen: sehr groß, groß, mittel, schwach, gar keines)
b) Einkommen von Erwerbstätigen in ganzen EURO
c) Einstellung der Bevölkerung zu einem EU-Beitritt eines Kandidatenlandes (Merkmalsausprägungen: dafür, teils-teils, dagegen)
d) Temperatur um 12 Uhr
e) Anzahl an in einem Monat in einem Konzern produzierten Autos
f) Militärische Dienstgrade in der deutschen Bundeswehr
g) Lieblingsschauspieler/in aus einer Liste von 20 vorgeschlagenen Personen
h) Meereshöhe

Lösung:

a)	o	d	z.B.: sehr groß = 1, groß = 2, mittel = 3, …
b)	m	d	(da nicht in Cent)
c)	o	d	z.B. dafür = 1, teils-teils = 2, dagegen = 3
d)	m	s	
e)	m	d	
f)	o	d	z.B. General = 1, Generalleutnant = 2, …, Gefreiter = 28, Soldat = 29
g)	n	d	z.B. Brad Pitt = 1, Angelina Jolie = 2, …, Daniel Craig = 20
h)	m	s	

Aufgabe 3: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Am 18. März 2012 wählte die deutsche Bundesversammlung einen Bundespräsidenten. Recherchieren Sie im Internet, wie viele Stimmen der Delegierten auf die einzelnen Kandidaten gefallen sind und wie viele ungültig gewählt beziehungsweise sich enthalten haben. Berechnen Sie danach mit dem Taschenrechner die relativen Häufigkeiten und Prozentzahlen des Stimmverhaltens der **anwesenden** Delegierten.

Lösung:

Zum Beispiel in Wikipedia:

Stimmverhalten	Anzahl	Relative Häufigkeit	Prozent
Joachim Gauck	991	0,804	80,4
Beate Klarsfeld	126	0,102	10,2
Olaf Rose	3	0,002	0,2
Enthaltungen	108	0,088	8,8
Ungültige Stimmen	4	0,003	0,3
SUMME	**1232**	**1**	**100**

Es wird empfohlen, bei der Berechnung der relativen Häufigkeiten erst an der 3. Stelle zu runden. Das ergibt Prozentzahlen, die immerhin noch eine Stelle nach dem Komma aufweisen. Eine solche Genauigkeit beim Runden wird meistens noch als Informationsgewinn empfunden.

Berechnung: $p_i = \frac{h_i}{N}$
z.B.: Relative Häufigkeit für Joachim Gauck
$\mathrm{p_i} = \mathrm{h_i}/\mathrm{N}$ => 991 : 1232 = 0,804 (schafft eine Relation zur Größe der Grundgesamtheit)
z.B.: Prozentzahl für Joachim Gauck => 0,804 · 100 = 80,4 % (ist anschaulicher als die relative Häufigkeit, weil sie sich auf eine Gesamtheit von 100 Personen bezieht)
Durch das Runden entsteht ein Rundungsfehler, der sich hier dadurch äußert, dass die Summe der fünf relativen Häufigkeiten (bzw. der fünf Prozentzahlen) nicht 1 (bzw. nicht 100), sondern 1,001 (bzw. 100,1 %) beträgt. Dennoch schreibt man unten 1 (bzw. 100), weil die Summe ja 1 (bzw. 100) betragen muss und die Abweichung nur durch den Fehler des Rundens entstanden ist.

Interpretation:
z.B.: 80,4 % (= 80,4 von 100) der anwesenden Delegierten der deutschen Bundesversammlung wählten bei der Wahl des Bundespräsidenten den Kandidaten Joachim Gauck.

Aufgabe 4*: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

An N= 100 Kindern derselben Schulstufe wird in einer Schule im Rahmen einer Studie ihre Fähigkeit im Verstehen eines einfachen Textes durch sechs sich auf den Text beziehende Fragen überprüft. Für jedes Kind liegt die Anzahl der falsch beantworteten Fragen vor:

0	2	0	2	0	0	1	2	0	0	2	1	2	1	2	1	1	1	1	6
1	2	0	2	0	0	1	0	0	1	0	0	1	0	0	1	1	0	2	0
1	1	1	1	2	3	2	3	2	0	0	2	1	2	5	1	1	2	4	5
0	1	1	1	2	3	2	3	0	3	5	1	3	3	2	2	1	1	2	1
0	4	1	0	2	3	0	3	1	0	0	0	3	1	0	0	1	2	1	0

Verwenden Sie die im Internet bereitstehende Excel-Lerndatei und stellen Sie darin der Anleitung folgend

a) die Häufigkeiten,
b) die relativen Häufigkeiten,
c) die Prozentzahlen,
d) die relativen Summenhäufigkeiten tabellarisch dar.

Lösung:

Anzahl an Fehlern	Häufigkeit	Rel. Häufigkeit	Prozent	Rel. Summenh.
0	30	0,30	30	0,30
1	32	0,32	32	0,62
2	22	0,22	22	0,84
3	10	0,10	10	0,94
4	2	0,02	2	0,96
5	3	0,03	3	0,99
6	1	0,01	1	1
SUMME	**100**	**1**	**100**	

Berechnung:
Relative Häufigkeiten z.B. für die Anzahl an Fehlern 2: Die Summe der relativen Häufigkeiten von 0, 1 und 2 Fehlern ist 0,30 + 0,32 + 0,22 = 0,84
Interpretation:
Relative Summenhäufigkeit z.B.: 84 % der Kinder beantworteten höchstens 2 (= 2 oder weniger) Fragen falsch.

Aufgabe 5: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Die erste Nationalratswahl der 2. Republik in Österreich am 25.11.1945 ergab folgende Verteilung der gültigen Stimmen auf die damals zur Wahl stehenden Parteien:

Partei	Stimmen
ÖVP	1.602.227
SPÖ	1.434.898
KPÖ	174.257
Sonstige	5.972

Berechnen Sie mit dem Taschenrechner und zur Kontrolle Ihrer Ergebnisse auch in Excel die relativen Häufigkeiten und die Prozentzahlen der einzelnen Parteien.

Lösung:

Partei	Anzahl	Rel. Häufigkeit	Prozent
ÖVP	1,602.227	0,498	49,8
SPÖ	1,434.898	0,446	44,6
KPÖ	174.257	0,054	5,4
Sonstige	5.972	0,002	0,2
SUMME	**3,217.354**	**1**	**100**

Berechnung: $p_i = \frac{h_i}{N}$
z.B. ÖVP: 1,602.227 : 3,217.354 = 0,498=49,8%;

Interpretation:
z.B.: Nicht ganz die Hälfte (49,8 von 100) aller Stimmen entfielen auf die ÖVP.

Aufgabe 6: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Auf der Internetseite *www.welt-in-zahlen.de/laenderinformation.phtm/* finden sich folgende Angaben über die Häufigkeitsverteilung von Staaten und Territorien bezüglich des Merkmals Kontinent [Stand Januar 2014].

Kontinent	Anzahl der Staaten
Afrika	58
Asien	50
Australien und Ozeanien	24
Europa	51
Nord-, Mittel- und Südamerika	51

Berechnen Sie mit dem Taschenrechner die relativen Häufigkeiten und die Prozentzahlen der Aufteilung aller Staaten und Territorien auf die einzelnen Kontinente.

Lösung:

Kontinent	Anzahl	Rel. Häufigkeit	Prozent
Afrika	58	0,248	24,8
Asien	50	0,214	21,4
Australien/Ozeanien	24	0,103	10,3
Europa	51	0,218	21,8
Nord-, Mittel-, Südamerika	51	0,218	21,8
SUMME	**234**	**1**	**100**

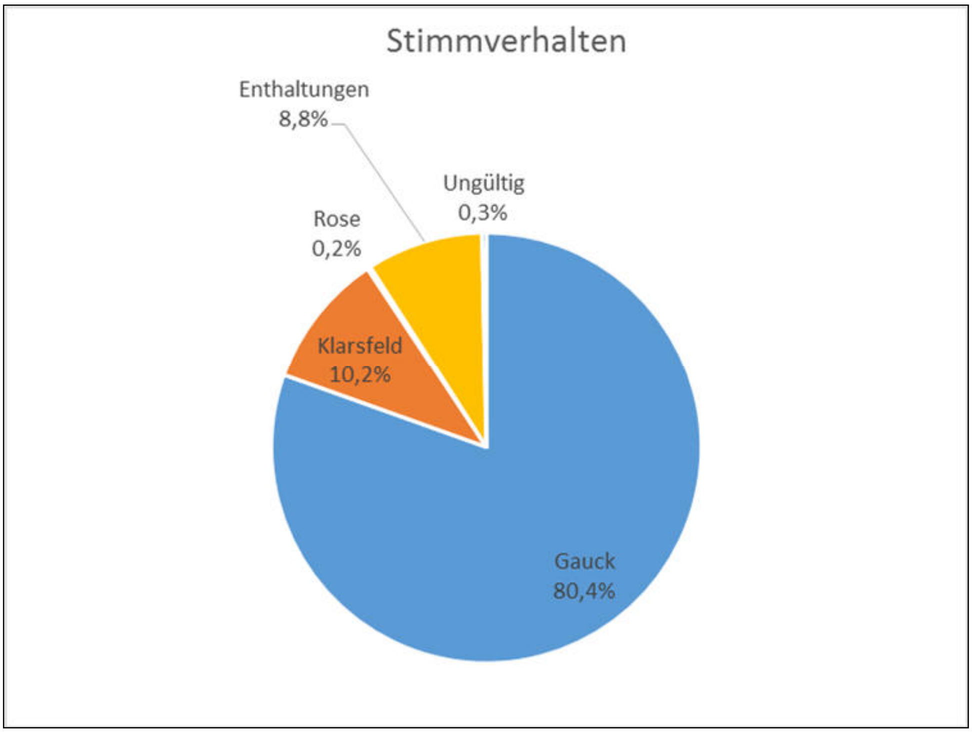
Berechnung: $p_i = \frac{h_i}{N}$
z.B. Afrika: 58 : 234 = 0,248; 0,248 · 100 = 24,8 %

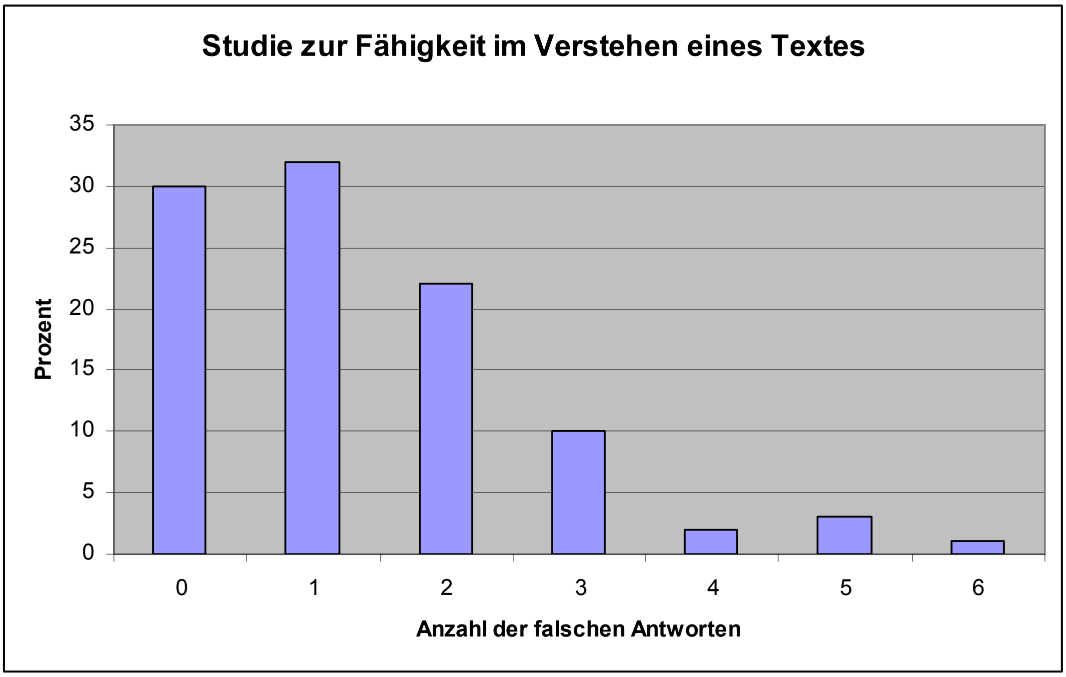
Interpretation:
z.B.: Nicht ganz ein Viertel aller Staaten und Territorien befindet sich in Afrika. 21,4 von 100 Staaten befinden sich in Asien usw.

Aufgabe 7*: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Lösen Sie folgende Aufgaben unter Verwendung der dafür im Internet bereitstehenden Excel-Lerndatei: Stellen Sie die Häufigkeitsverteilung aus Aufgabe 3 in Excel in einem Kreisdiagramm dar. Verwenden Sie die Daten aus Aufgabe 4 und stellen Sie diese Häufigkeitsverteilung in Excel in einem Säulendiagramm dar.

Lösung:

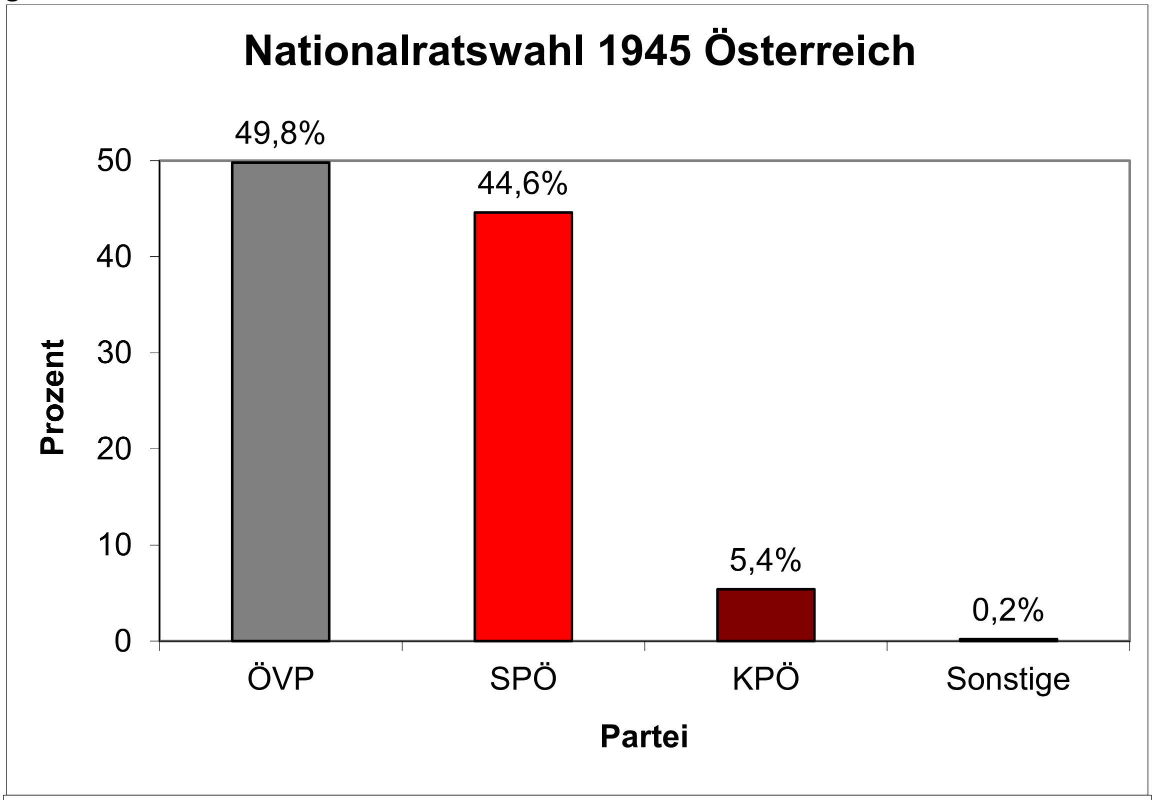


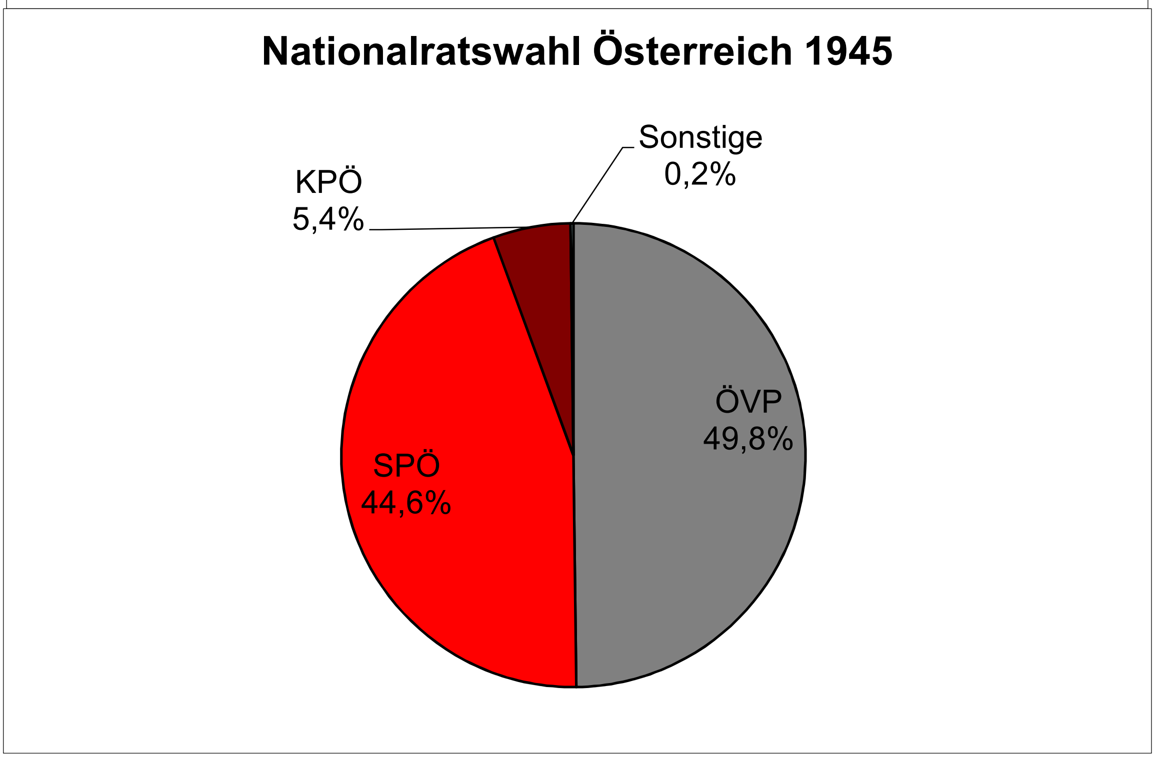


Aufgabe 8*: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Verwenden Sie die Angaben aus Aufgabe 5 und stellen Sie diese Häufigkeitsverteilung in Excel in einem Säulendiagramm und einem Kreisdiagramm dar.

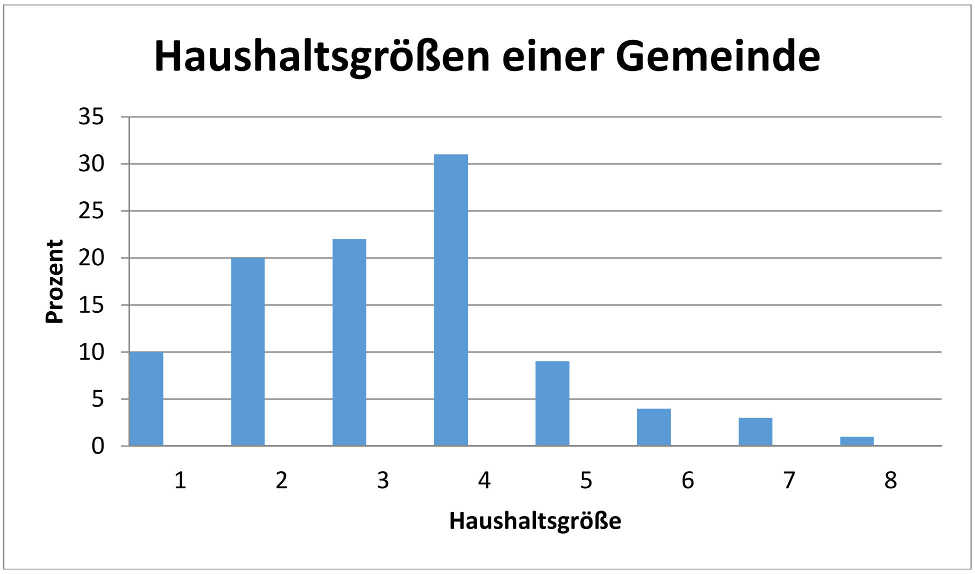
Lösung:





Aufgabe 9: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Das unten folgende Säulendiagramm beschreibt die Verteilung des Merkmals Anzahl der Personen in einem Haushalt (= Haushaltsgröße) in einer Gemeinde:

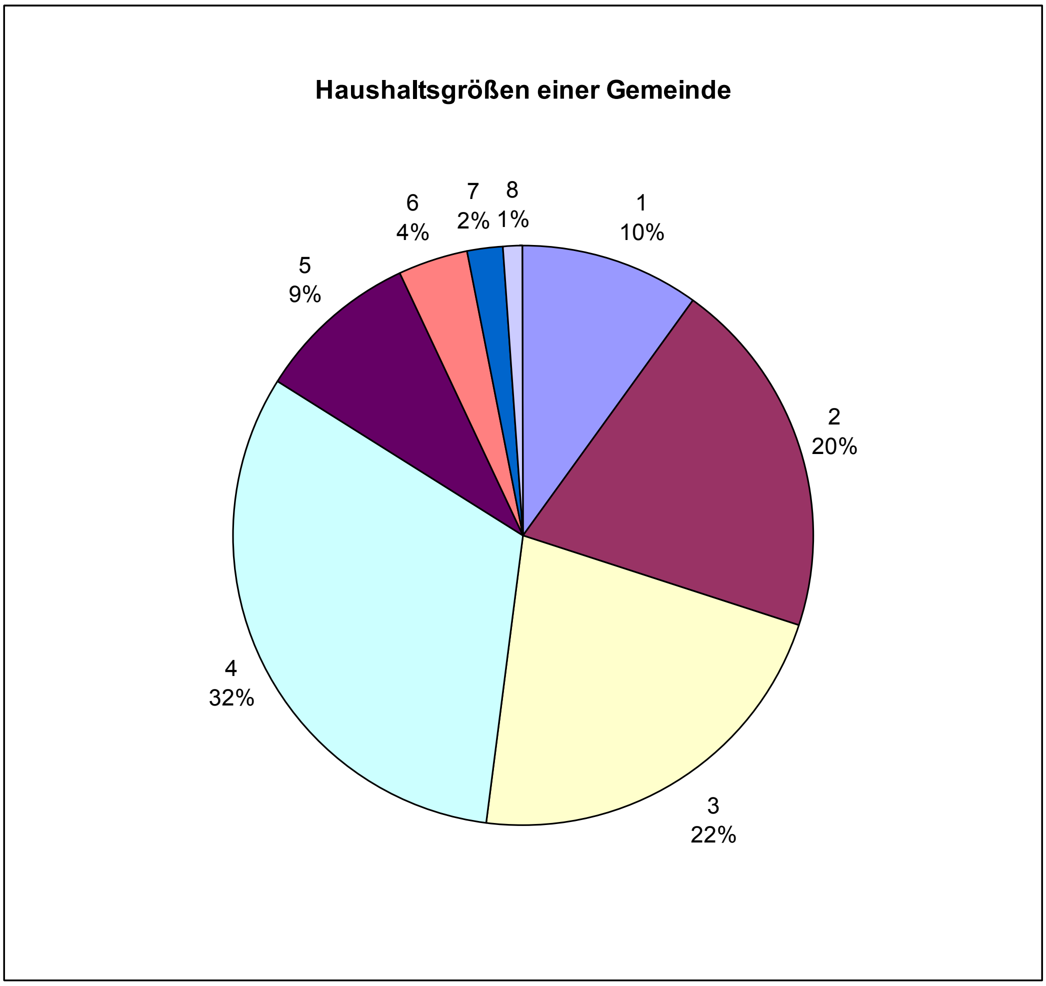


Lesen Sie die relativen Häufigkeiten der einzelnen Merkmalsausprägungen (ungefähr) ab und erstellen Sie damit in Excel eine Tabelle mit den Prozentzahlen und den relativen Summenhäufigkeiten. Zeichnen Sie ferner in Excel das dazu gehörende Kreisdiagramm und versuchen Sie darin die relativen Summenhäufigkeiten abzulesen.

Lösung:

Haushaltsgröße	Prozent	Rel. Summenh.
1	10	0,10
2	20	0,30
3	22	0,52
4	32	0,84
5	9	0,93
6	4	0,97
7	2	0,99
8	1	1,00
Summe	**100**	

Interpretation:
z.B. die relative Summenhäufigkeit der Merkmalsausprägung 3: In 52% der Haushalte leben höchstens 3 Personen.



Relative Summenhäufigkeiten im Kreisdiagramm: Aus den von 12 Uhr beginnend im Uhrzeigersinn angeordneten Merkmalsausprägungen lassen sich neben den Prozentzahlen der einzelnen Merkmalsausprägungen auch die relativen Summenhäufigkeiten ablesen. Von 12 Uhr nach rechts gehend ergeben die ersten 3 Kreissegmente etwas mehr als die Hälfte (52 %) des Kreises, weil 52 % der Haushalte von höchstens 3 Personen bewohnt werden.

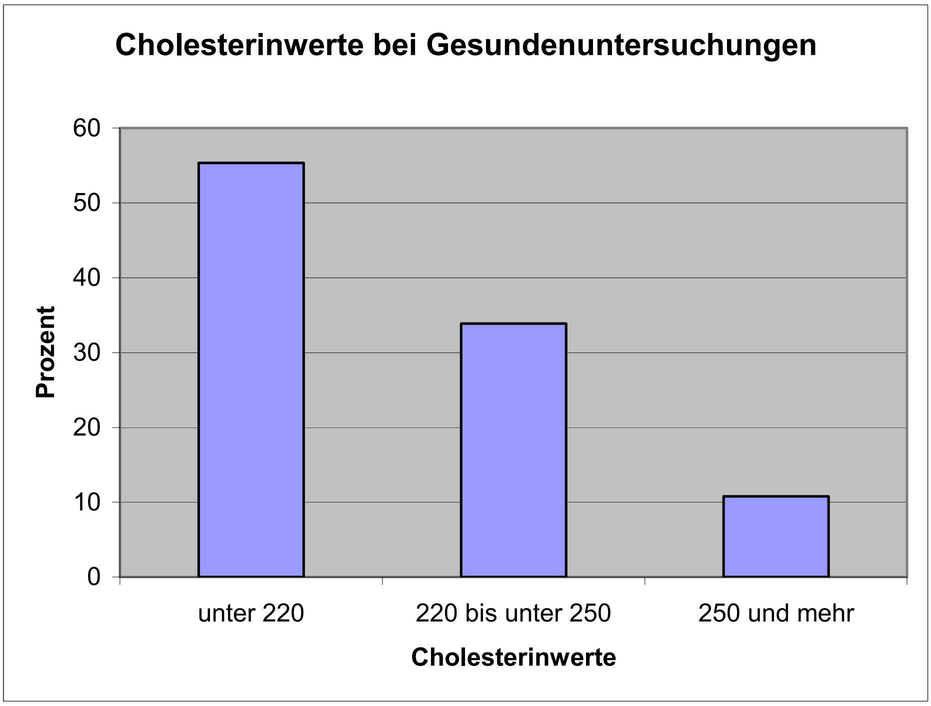
Aufgabe 10*: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

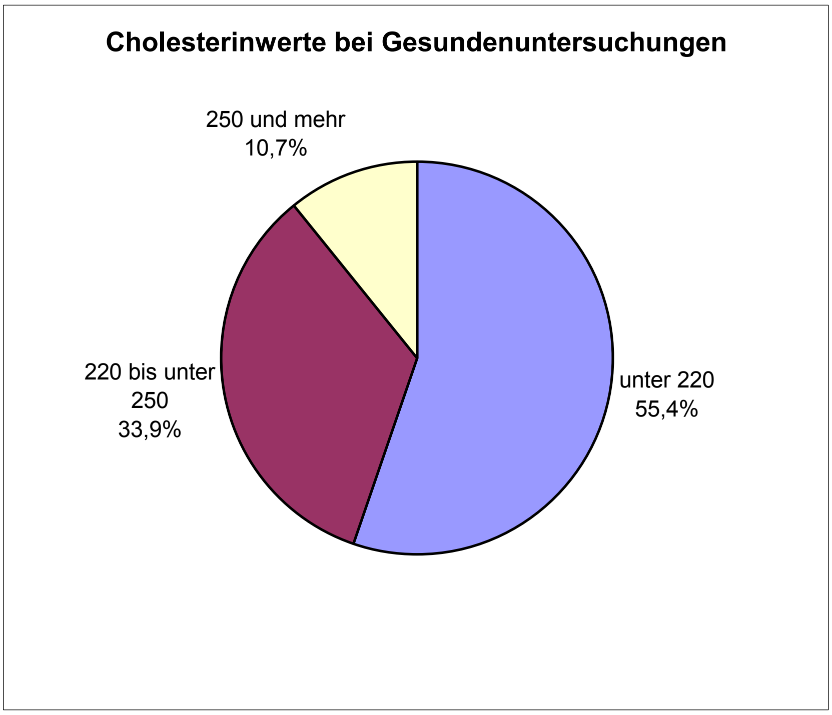
Lösen Sie folgende Aufgabe unter Verwendung der dafür im Internet bereit stehenden Excel-Lerndatei. In dieser Datei befinden sich die durch einen Arzt aufgezeichneten Ergebnisse von Messungen des Cholesterinwertes (in mg/dl) von Personen bei Gesundenuntersuchungen in einem bestimmten Monat. Erstellen Sie eine Tabelle mit den Häufigkeiten, den relativen Häufigkeiten, den Prozentzahlen und den relativen Summenhäufigkeiten des Merkmals Cholesterin, wobei der Arzt die Merkmalsausprägungen in folgende Intervalle einteilen möchte: "unter 220" (normal), ,,220 bis unter 250" (unter Beobachtung) und ,,250 und höher" (erhöht). Stellen Sie diese Häufigkeitsverteilung auch grafisch in einem Säulen- und in einem Kreisdiagramm dar (Anleitungen dafür entnehmen Sie den Tabellenblättern der EXCEL-Lerndatei zu Aufgabe 7).

Lösung:

Cholesterinwerte	**Häufigkeiten**	**Rel. Häufigkeiten**	**Prozent**	**Rel. Summenhäufigkeiten**
unter 220	67	0,554	55,4	0,554
220 bis unter 250	41	0,339	33,9	0,893
250 und mehr	13	0,107	10,7	1,000

Interpretationen:
z.B. 33,9 % (also 33,9 von 100 Personen) weisen einen Cholesterinwert von 220 bis unter 250 auf.
z.B. 89,3 % (also 89,3 von 100 Personen) weisen einen Cholesterinwert unter 250 auf..





Aufgabe 11: [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Bei einer Befragung von 794 zufällig ausgewählten Personen gaben 241 Befragte auf dem Fragebogen an, dass Sie männlich und mit der EURO-Währung zufrieden sind, 198, dass Sie weiblich und zufrieden sind, 122, dass Sie männlich und unzufrieden sind, 173, dass Sie weiblich und unzufrieden sind. Von den restlichen 60 Personen waren 40 männlich und in der Euro-Frage neutral und 20 weiblich und neutral.

a) Erstellen Sie je eine Tabelle mit den Häufigkeiten und mit den relativen Häufigkeiten der gemeinsamen Verteilung auf dem Merkmal Geschlecht und Euro-Einstellung.
b) Berechnen Sie die bedingten relativen Häufigkeiten der Verteilungen des Merkmals Euro-Einstellung unter den Männern und unter den Frauen.

Lösung:
a)
Häufigkeiten:

		Einstellung zur Euro-Währung			Randverteilung $N_{i.}$
		zufrieden	unzufrieden	neutral	
Geschlecht	männlich	241	122	40	403
	weiblich	198	173	20	391
	Randverteilung $N_{.j}$	439	295	60	794

Interpretation:
z.B. 173 von 794 Befragten sind weiblich und mit der Euro-Währung unzufrieden.
z.B. 439 von 794 Befragten sind mit der Euro-Währung zufrieden oder 403 der 794 befragten Personen sind männlich.

Relative Häufigkeiten:
Berechnung: $p_{ij} = \frac{h_{ij}}{N}$
z.B. für die Kombination „weiblich“ und „unzufrieden“: $p_{22} = \frac{h_{22}}{N} = \frac{173}{794} = 0{,}218$

		Einstellung zur Euro-Währung			Randverteilung $p_{i.}$
		zufrieden	unzufrieden	neutral	
Geschlecht	männlich	0,304	0,154	0,050	0,508
	weiblich	0,249	0,218	0,025	0,492
	Randverteilung $p_{.j}$	0,553	0,372	0,076	1

Interpretation:
z.B. relative Häufigkeiten: 21,8 % (= 21,8 von 100) der Befragten sind weiblich und mit der Euro-Währung unzufrieden.
z.B. Randverteilung: 37,2 % aller Befragten sind mit der Eurowährung unzufrieden bzw. 49,2 % aller Befragten sind weiblich.

b)
Relative Häufigkeiten der Verteilungen des Merkmals Euro-Einstellung unter den Männern und unter den Frauen:

Berechnung:

$p_{j|i=k} = \frac{h_{kj}}{N_{k.}}$ mit k als einer bestimmten Zeile (männlich oder weiblich)

Bedingte rel. Häufigkeiten unter den Männern und dann unter den Frauen(in der 1. Zeile wird jede Häufigkeit auf 403 Männer, in der 2. auf 391 Frauen bezogen):

z.B.: bedingte Häufigkeit einer unzufriedenen Einstellung unter Frauen: $p_{2|i=2} = \frac{h_{22}}{N_{2.}} = \frac{173}{391} = 0{,}442$

		Einstellung zur Euro-Währung			
		zufrieden	unzufrieden	neutral	
Geschlecht	männlich	0,598	0,303	0,099	1
	weiblich	0,506	0,442	0,051	1

Interpretation:

z.B.: 44,2 % der Frauen sind mit der Euro-Währung unzufrieden, im Gegensatz dazu sind nur 30,3 % der Männer mit der Euro-Währung unzufrieden.

Aufgabe 12 [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Vergleichen Sie nun die Ergebnisse der beiden Verteilungen aus Aufgabe 11 b) und beantworten Sie damit die Frage: Sind die Frauen in dieser Erhebung zufriedener mit der Euro-Währung als die Männer oder verhält es sich umgekehrt?

Lösung:
Unter den befragten Männern ist der Zufriedenenanteil deutlich höher als unter den befragten Frauen (59,8 % gegen 50,6 %). Der Unzufriedenenanteil ist unter den Frauen deutlich höher. In dieser Befragung äußerten sich die Männer also zufriedener.

Aufgabe 13* [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Seit Jahrzehnten verschiebt sich in den Industrieländern das Geburtenverhältnis von Mädchen und Jungen. Dass heute mehr Mädchen geboren werden als früher, soll auch auf den Einfluss von Giftstoffen aus der Umwelt auf das männliche Fortpflanzungssystem zurückzuführen sein. In einer wissenschaftlichen Studie an 400 Kindern wurden das Geschlecht der Kinder (Kodierung: männlich = 1; weiblich = 2) und das Rauchverhalten der Eltern in den drei Monaten vor der Zeugung (beide Nichtraucher =1; ein Elternteil Raucher, der andere Nichtraucher =2; beide Raucher =3) erhoben. Verwenden Sie die im Internet bereitstehende Excel-Lerndatei und berechnen Sie darin den Anweisungen folgend die relativen Häufigkeiten

a) der gemeinsamen Verteilung der 400 Erhebungseinheiten auf den Merkmalen Geschlecht
b) und Rauchverhalten der Eltern,
c) der bedingten Verteilungen des Geschlechts der Kinder innerhalb der drei verschiedenen Rauchergruppen !

Lösung:

Häufigkeiten		**Rauchverhalten der Eltern**			Randverteilung $N_{i.}$
		2 x NR (=1)	1 NR, 1 R (=2)	2 x R (=3)	
Geschlecht	männlich (=1)	97	55	43	195
des Kindes	weiblich (=2)	86	61	58	205
	Randverteilung $N_{.j}$	183	116	101	400

relative Häufigkeiten		**Rauchverhalten der Eltern**			Randverteilung $p_{i.}$
		2 x NR (=1)	1 NR, 1 R (=2)	2 x R (=3)	
Geschlecht	männlich (=1)	0,243	0,138	0,108	0,488
des Kindes	weiblich (=2)	0,215	0,153	0,145	0,513
	Randverteilung $p_{.j}$	0,458	0,290	0,253	1,000

Bedingte relative Häufigkeiten		**Rauchverhalten der Eltern**			
		2 x NR (=1)	1 NR, 1 R (=2)	2 x R (=3)	
Geschlecht	männlich (=1)	0,530	0,474	0,426	
des Kindes	weiblich (=2)	0,470	0,526	0,574	
		1,000	1,000	1,000	

Interpretation:

Häufigkeiten: z.B. in 97 Fällen war das Kind männlich und beide Elternteile vor der Zeugung Nichtraucher, 195 Kinder waren männlich, 183 Eltern rauchten beide nicht.

Relative Häufigkeiten (in Prozent): z.B. in 21,5 % der Fälle war das Kind weiblich und beide Elternteile vor der Zeugung Nichtraucher, in 51,3 % der Fälle war das Kind weiblich, in 45,8 % der Fälle waren beide Elternteile vor der Zeugung Nichtraucher.

Relative Häufigkeit unter den verschiedenen Rauchergruppen der Eltern: z.B. unter den nichtrauchenden Eltern waren 53,0 % der geborenen Kinder Knaben, unter Eltern mit gemischtem Rauchverhalten waren dies nur 47,4 % der Geborenen und unter den Eltern, von denen beide Teile Raucher waren, kamen sogar nur 42,6 % männliche Kinder zur Welt.

Aufgabe 14 [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

In einer Befragung unter 1.000 Personen wurden in diesem Jahr die Merkmale Geschlecht und Einstellung zum Ankauf neuer Abfangjäger für die Landesverteidigung erhoben.

	Ankauf von Abfangjägern			
Geschlecht	**dafür**	**unentschlossen**	**dagegen**	**Summe**
Weiblich	10	40	250	300
Männlich	190	160	350	700
Summe	200	200	600	1000

Berechnen Sie mit dem Taschenrechner die relativen Häufigkeiten in dieser Tabelle.

Lösung:

Berechnung: $p_{ij} = \frac{h_{ij}}{N}$

Jede Häufigkeit der 1. Tabelle wird für die 2. Tabelle durch 1.000 dividiert. Zum Beispiel die relative Häufigkeit von Männern die gegen den Kauf von Abfangjägern ist (2 Zeile, 3 Spalte) → $p_{23} = \frac{h_{23}}{N} = \frac{350}{1.000} = 0{,}35$

Häufigkeiten:

		Ankauf von Abfangjägern			Randverteilung $N_{i.}$
		dafür	unentschlossen	dagegen	
Geschlecht	weiblich	10	40	250	300
	männlich	190	160	350	700
	Randverteilung $N_{.j}$	200	200	600	1000

Relative Häufigkeiten:

		Ankauf von Abfangjägern			Randverteilung $p_{i.}$
		dafür	unentschlossen	dagegen	
Geschlecht	weiblich	0,01	0,04	0,25	0,30
	männlich	0,19	0,16	0,35	0,70
	Randverteilung $p_{.j}$	0,20	0,20	0,60	1

Aufgabe 15 [Tabellarische und grafische Darstellung von Häufigkeitsverteilungen]

Berechnen Sie mit dem Taschenrechner die bedingten relativen Häufigkeiten des Merkmals Ankauf von Abfangjägern aus Aufgabe 14 unter den Frauen und auch jene unter den Männern.

Lösung:

Berechnung:

$p_{j|i=k} = \frac{h_{kj}}{N_{k.}}$ mit k als einer bestimmten Zeile (männlich oder weiblich)

Unter Frauen (1. Zeile → k=1): $p_{j|i=1} = \frac{h_{1j}}{N_{1.}}$

Unter Männern (2. Zeile → k=2): $p_{j|i=2} = \frac{h_{2j}}{N_{2.}}$

Jede Häufigkeit der 1. Zeile wird durch 300, jede der 2. durch 700 dividiert.
z.B. die bedingte relative Häufigkeit von gegen den Kauf zu sein unter Männern (j=3 und k=2) → $p_{3|i=2} = \frac{h_{23}}{N_{2.}} = \frac{350}{700} = 0{,}5$

Relative Häufigkeit der bedingten Verteilung des Merkmals Ankauf von Abfangjägern unter den Männern und unter den Frauen:

		Ankauf von Abfangjägern			
		dafür	unentschlossen	dagegen	
Geschlecht	weiblich	0,033	0,133	0,833	1
	männlich	0,271	0,229	0,500	1

Interpretation:
z.B.: Unter befragten Männern sind 50% gegen den Kauf, 27,1% für den Kauf und 22,9% unentschlossen.

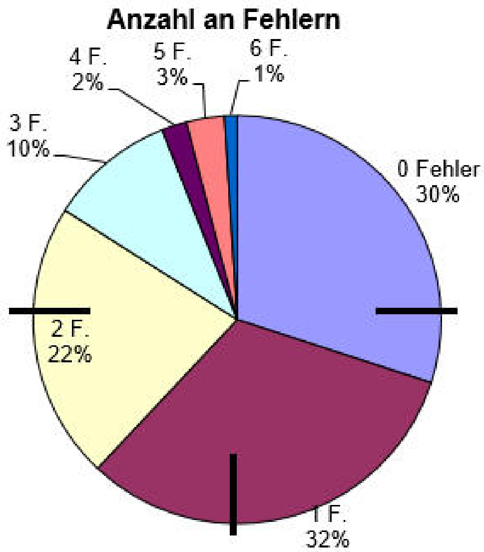
Aufgabe 16* [Kennzahlen der Lage]

Berechnen Sie mit dem Taschenrechner mit Hilfe der in Aufgabe 4 erzeugten Tabelle und den dort eingetragenen Häufigkeiten und in Excel unter Verwendung der Anweisungen in der dafür im Internet bereitstehenden Excel-Lerndatei Modus, Median, unteres und oberes Quartil sowie den Mittelwert für die Häufigkeitsverteilung von Aufgabe 4. Lesen Sie ferner Median, unteres und oberes Quartil auch in Excel aus einem Kreisdiagramm ab.

Lösung:

Modus:	1
Median:	1
Mittelwert:	1,35
Unteres Quartil:	0
Oberes Quartil:	2

Berechnung:
Modus ist die am Häufigsten vorkommende Merkmalsausprägung: 1
Median: Bei gerader Anzahl an Erhebungseinheiten (N = 100) stehen 2 Werte in der Mitte der sortierten Liste; das 50. und 51. Element der Liste besitzen jeweils die Ausprägung 1 → Median = (1+1) : 2 = 1;
Auch aus den relativen Summenhäufigkeiten ablesbar, weil der Median die Merkmalsausprägung ist, bei der die relative Summenhäufigkeit von 0,5 überschritten wird. Würde 0,5 genau erreicht ist der Median der Mittelwert dieser und der nächstfolgenden Merkmalsausprägung.
Mittelwert: $(0 \cdot 30 + 1 \cdot 32 + 2 \cdot 22 + 3 \cdot 10 + 4 \cdot 2 + 5 \cdot 3 + 6 \cdot 1) : 100 = 1{,}35$
Unteres Quartil (25%-Perzentil) ist die Merkmalsausprägung, bei der in der Tabelle eine relative Summenhäufigkeit von 0,25 überschritten wird: 0
Oberes Quartil (75%-Perzentil) ist die Ausprägung dort ablesen, bei der in der Tabelle eine relative Summenhäufigkeit von 0,75 überschritten wird: 2



Die Quartile lassen sich auch aus dem Kreisdiagramm ablesen. Die Merkmalsausprägung, die das untere Quartil ist, findet man um 3 Uhr (also 0), die des Medians um 6 Uhr (also 1) und die des oberen Quartils um 9 Uhr (also 2). Dazu braucht man sich die Uhrzeiten eigentlich nur zu denken und muss diese nicht einzeichnen.

Interpretationen:
Modus: Von allen Fehleranzahlen kam die Fehlerzahl 1 am Häufigsten vor.
Median: Mindestens eine Hälfte der Kinder beantwortete höchstens 1 Frage falsch, mindestens eine Hälfte beantwortete mindestens 1 Frage falsch.
Mittelwert: Würde man die Gesamtzahl der falsch beantworteten Fragen gleichmäßig auf die alle Kinder aufteilen, so hätte jedes Kind 1,35 Fragen falsch beantwortet.
Unteres Quartil: Mindestens 25% der Kinder beantworteten keine (höchstens null) der Fragen falsch und mindestens 75% der Kinder gaben 0 oder mehr falsche Antworten.
Oberes Quartil: Mindestens 75% der Kinder beantworteten höchstens 2 Fragen falsch und mindestens 25% der Kinder gaben mindestens 2 falsche Antworten.

Aufgabe 17* [Kennzahlen der Lage]

Die 28 alphabetisch geordneten EU-Länder des Jahres 2014 besaßen zu Beginn des Jahres 2013 folgende Ausprägungen beim Merkmal Einwohnerzahl in Millionen (*Quelle: Eurostat*):

Staat	Einwohner (in Mio.)	Staat	Einwohner (in Mio.)
Belgien	11,2	Luxemburg	0,5
Bulgarien	7,3	Malta	0,4
Dänemark	5,64	Niederlande	16,8
Deutschland	80,5	Österreich	8,5
Estland	1,3	Polen	38,5
Finnland	5,4	Portugal	10,5
Frankreich	65,6	Rumänien	20,0
Griechenland	11,1	Schweden	9,6
Großbritannien	63,9	Slowakei	5,4
Irland	4,6	Slowenien	2,1
Italien	59,7	Spanien	46,7
Kroatien	4,3	Tschechien	10,5
Lettland	2,0	Ungarn	9,9
Litauen	3,0	Zypern	0,9

a)

b) Berechnen Sie in Excel Median und Mittelwert sowie das obere und untere Quartil dieses Merkmals.
c) Tippen Sie nun in den Daten Ihrer Excel-Datei statt des korrekten Werts für Deutschland die Zahl 805 (Mio.) ein und verfolgen Sie, wie unterschiedlich sich dieser "Tippfehler" auf Ihre Ergebnisse für Median und Mittelwert auswirkt.
d) Ausgehend von a) verändern Sie den Wert für Luxemburg auf 50 Mio. Verfolgen Sie die Auswirkungen diese Tippfehlers.
e) Erstellen Sie ferner für dieses Merkmal einen Boxplot, indem Sie folgendermaßen vorgehen: Verwenden Sie die Anweisungen in der Excel-Lerndatei, um den Boxplot auf der Internetseite *http://www.wessa.net/rwasp_notchedbox1.wasp* zu zeichnen. Versuchen Sie in diesem Boxplot den Median, das untere und obere Quartil und die Whiskers ungefähr abzulesen.

Lösung

a)

Staat	**EW (in Mio)**			**Staat**	**EW (in Mio)**		
Belgien	11,2			Belgien	11,2		
Bulgarien	7,3			Bulgarien	7,3		
Dänemark*	5,64			Dänemark	5,64		
Deutschland	80,5			**Deutschland**	**805**		
Estland	1,3			Estland	1,3		
Finnland	5,4	**Median =**	9,1	Finnland	5,4	**Median =**	9,05
Frankreich	65,6	**MW =**	18,1	Frankreich	65,6	**MW =**	43,9
Griechenl.	11,1	$\mathbf{Q_{0,75}}$ **=**	17,6	Griechenl.	11,1	$\mathbf{Q_{0,75}}$ **=**	17,6
Großbrit.	63,9	$\mathbf{Q_{0,25}}$ **=**	4,0	Großbrit.	63,9	$\mathbf{Q_{0,25}}$ **=**	4,0
Irland	4,6			Irland	4,6		
Italien	59,7			Italien	59,7		
Kroatien	4,3			Kroatien	4,3		
Lettland	2,0			Lettland	2,0		
Litauen	3,0			Litauen	3,0		
Luxemburg	0,5			Luxemburg	0,5		
Malta	0,4			Malta	0,4		
Niederlande	16,8			Niederlande	16,8		
Österreich	8,5			Österreich	8,5		
Polen	38,5			Polen	38,5		
Portugal	10,5			Portugal	10,5		
Rumänien	20,0			Rumänien	20,0		
Schweden	9,6			Schweden	9,6		
Slowakei	5,4			Slowakei	5,4		
Slowenien	2,1			Slowenien	2,1		
Spanien	46,7			Spanien	46,7		
Tschechien	10,5			Tschechien	10,5		
Ungarn	9,9			Ungarn	9,9		
Zypern	0,9			Zypern	0,9		

*) Es wird mit der Einwohnerzahl 5,64 Mio. gerechnet. Diese ist jedoch ein Fehler im Buch! Eigentlich sollte auch für Dänemark nur eine Stelle nach dem Komma, also 5,6 Mio., angegeben werden. In einem Nachdruck zur 4. Auflage wird dies korrigiert werden. Die auf eine Stelle nach dem Komma gerundeten Resultate für Mittelwert und Median werden dadurch jedoch nicht verändert.

Interpretationen:

Mindestens eine Hälfte aller EU-Staaten hat höchstens 9,05 Millionen Einwohner, mindestens die andere Hälfte hat mindestens 9,05 Millionen Einwohner.

Würde man die Einwohner aller 28 EU-Staaten des Jahres 2014 gleichmäßig auf alle Staaten aufteilen, würden auf jeden einzelnen Staat 18,1 Millionen Einwohner fallen.

Tippfehler bei Deutschland: 80,5 Mio. durch 805 Mio. ersetzt → Der Median bleibt unverändert, da bei der Sortierung der Einwohnerzahlen nach ihrer Größe weiterhin die Länder Österreich und Schweden in der Mitte bleiben. Der Median berechnet sich, indem man die Einwohnerzahlen dieser beiden Länder addiert, und anschließend halbiert: (8,5 + 9,6) / 2 = 9,05.

Der Mittelwert verändert sich durch den Tippfehler sehr stark, da die Summe der Einwohnerzahlen aller Länder wesentlich größer wird und somit auch die gleichmäßige Aufteilung dieser Summe auf alle Länder einen viel höheren Wert liefert.

Der Mittelwert ist also weit sensibler gegenüber starke Ausreißer, der Median kann trotz eines Ausreißers unverändert bleiben.

c)

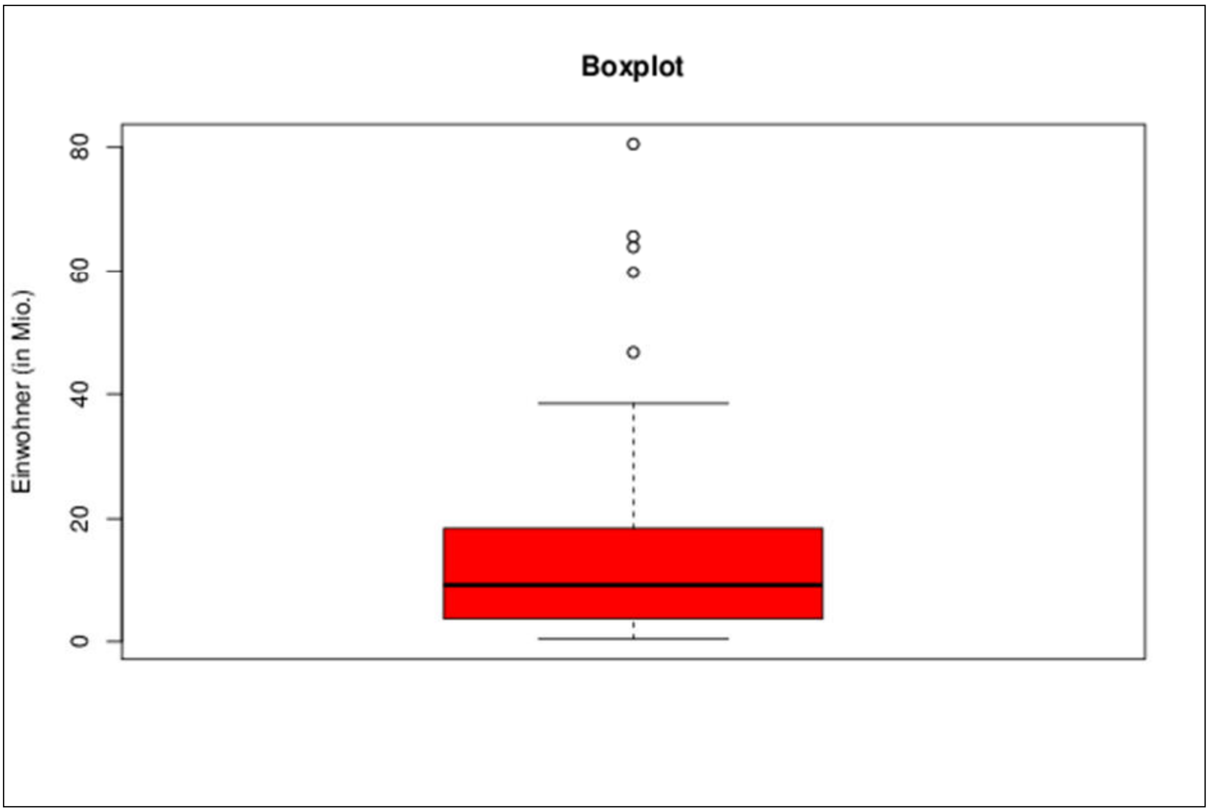
Staat	**EW (in Mio)**			**Staat**	**EW (in Mio)**		
Belgien	11,2			Belgien	11,2		
Bulgarien	7,3			Bulgarien	7,3		
Dänemark	5,64			Dänemark	5,64		
Deutschland	80,5			Deutschland	80,5		
Estland	1,3			Estland	1,3		
Finnland	5,4	**Median =**	9,05	Finnland	5,4	**Median =**	9,75
Frankreich	65,6	**MW =**	18,1	Frankreich	65,6	**MW =**	19,8
Griechenl.	11,1	**$Q_{0,75}$ =**	19,6	Griechenl.	11,1	**$Q_{0,75}$ =**	24,6
Großbrit.	63,9	**$Q_{0,25}$ =**	4,0	Großbrit.	63,9	**$Q_{0,25}$ =**	4,5
Irland	4,6			Irland	4,6		
Italien	59,7			Italien	59,7		
Kroatien	4,3			Kroatien	4,3		
Lettland	2,0			Lettland	2,0		
Litauen	3,0			Litauen	3,0		
Luxemburg	0,5			**Luxemburg**	**50**		
Malta	0,4			Malta	0,4		
Niederlande	16,8			Niederlande	16,8		
Österreich	8,5			Österreich	8,5		
Polen	38,5			Polen	38,5		
Portugal	10,5			Portugal	10,5		
Rumänien	20,0			Rumänien	20,0		
Schweden	9,6			Schweden	9,6		
Slowakei	5,4			Slowakei	5,4		
Slowenien	2,1			Slowenien	2,1		
Spanien	46,7			Spanien	46,7		
Tschechien	10,5			Tschechien	10,5		
Ungarn	9,9			Ungarn	9,9		
Zypern	0,9			Zypern	0,9		

Interpretationen:

Tippfehler bei Luxemburg: 0,5 Mio. durch 50 Mio. ersetzt → Der Median verändert sich, da bei der Sortierung der Einwohnerzahlen nach ihrer Größe nun die Länder Schweden und Ungarn in der Mitte stehen. Die Einwohnerzahl Luxemburgs „wandert" sozusagen von der linken Seite des Medians auf die rechte Seite.

Der Mittelwert verändert sich in ähnlicher Weise wie beim Tippfehler a). Die Summe der Einwohnerzahlen aller Länder wird durch den Ausreißer erhöht, somit ergibt auch eine gleichmäßige Aufteilung dieser Summe auf alle Länder einen höheren Wert als ursprünglich.

d)



Interpretation:

Aus einem Boxplot-Diagramm lassen sich einige wichtige statistische Kennzahlen ablesen. Die rot gefärbte Box wird von zwei Quartilen begrenzt, dem unteren Quartil (hier: 3,975) und dem oberen Quartil (hier: 17,6). Das heißt, dass mindestens 25% der EU-Länder bis zu vier Millionen Einwohner haben und mindestens 75% eine Einwohnerzahl von 17,6 Mio. aufweisen. Zwischen den beiden Quartilen liegt die mittlere Hälfte der erhobenen Daten.
Geteilt wird das Intervall durch den Median, welcher 9,05 beträgt (siehe oben). Die Kreise am oberen Rand des Diagramms zeigen, dass Ausreißer vorhanden sind. Ausreißer sind Daten, die deutlich außerhalb der Box liegen. Der oberste Kreis gibt zugleich die höchste erhobene Einwohnerzahl an, in diesem Fall die von Deutschland mit 80,5 Mio. Einwohnern.
Man sieht, dass die Datenhälfte links vom Median eine geringere Spannweite aufweist als die Daten rechts vom Median. Weiters treten fünf Ausreißer auf, die bei Bedarf identifiziert werden können (die fünf höchsten Datenwerte: Deutschland, Frankreich, Großbritannien, Italien, Spanien). Ausreißer können auf fehlerhafte Daten hinweisen, in diesem Kontext lassen sie sich aber auf die sehr unterschiedliche Größe der EU-Staaten zurückführen.

Aufgabe 18 [Kennzahlen der Lage]

Bei einer Radtour notierte einer der Teilnehmer die täglich am Tachometer angezeigte Kilometerleistung:

66	85	55	32	73	55

a) Berechnen Sie Mittelwert und Median dieses Merkmals.
b) Berechnen Sie $Q_{0,25}$ und $Q_{0,75}$ für dieses Merkmal.

Lösung:
Berechnung:
Mittelwert:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{66+85+55+32+73+55}{6} = \frac{366}{6} = 61$$

Interpretation: Würde man die gefahrenen Kilometer gleichmäßig auf alle Etappen aufteilen, so hätte jede Etappe eine Länge von 61 km.

Median: N = 6 (gerade)→ Sortierte Kilometerleistungen für 6 Etappen:
32 55 55 **66** **73** 55 85

$$\tilde{x} = \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) = \frac{1}{2} \cdot \left(x_{\frac{6}{2}} + x_{\frac{6}{2}+1}\right) = \frac{1}{2} \cdot (x_3 + x_4) = \frac{1}{2} \cdot (66+73) = 69{,}5$$

Bei gerader Zahl stehen 2 Merkmalsausprägungen in der Mitte der sortierten Liste, nämlich die des 3. und 4. Elements.
Interpretation: (mindestens) die Hälfte der Etappen hatte eine Länge kleiner gleich 69,5 km. Die andere Hälfte der Etappen ist länger.

$Q_{0,25}$ (unteres Quartil): N·p=6·0,25=1,5 (nicht ganzzahlig)
$Q_p = x_{\lceil n \cdot p \rceil} \rightarrow Q_{0,25} = x_{\lceil 6 \cdot 0{,}25 \rceil} = x_{\lceil 1{,}5 \rceil} = x_2 = 55$

$Q_{0,75}$ (oberes Quartil): N·p=6·0,75=1,5 (nicht ganzzahlig)
$Q_p = x_{\lceil n \cdot p \rceil} \rightarrow Q_{0,75} = x_{\lceil 6 \cdot 0{,}25 \rceil} = x_{\lceil 4{,}5 \rceil} = x_5 = 73$

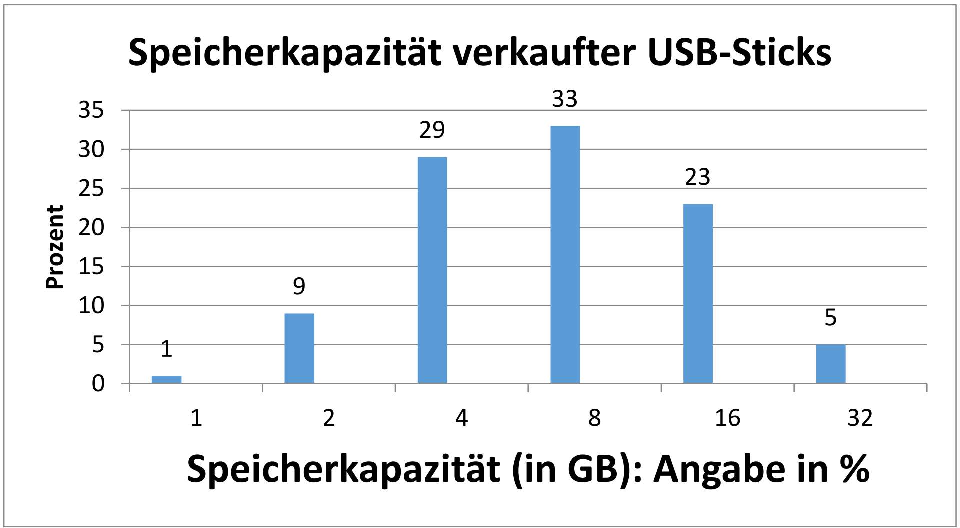
Ergebnisübersicht:

Median:	69,5
Mittelwert:	61
$Q_{0,25}$ (unteres Quartil)	48
$Q_{0,75}$ (oberes Quartil)	73

Interpretation: (mindestens) 25% der Etappen haben eine Länge von höchstens (kleiner gleich) 55 km. (Mindestens) 75% der Etappen haben eine Länge von höchstens (kleiner gleich) 73 km.

Aufgabe 19 [Kennzahlen der Lage]

Das nachfolgende Säulendiagramm beschreibt die relative Häufigkeitsverteilung des Merkmals *Speicherkapazität von verkauften USB-Sticks* (weltweit im Jahr 2009; aus: *CHIP Test &Kauf*; Juni 2010, S. 8):



Berechnen Sie Median, 20%-Perzentil, Mittelwert und Modus für dieses Merkmal.

Lösung:

Berechnung:

Modus: x_{mod} ist die am Häufigsten vorkommende Ausprägung → $x_{mod} = x_4$ Speicherchips mit 8 GB

Median → Ausprägung dort ablesen, wo eine relative Summenhäufigkeit von 0,5 überschritten wird: 8 GB

Mittelwert: $\bar{x} = \sum_{i=1}^{k} x_i \cdot p_i = 1 \cdot 0{,}01 + 2 \cdot 0{,}09 + 4 \cdot 0{,}29 + 8 \cdot 0{,}33 + 16 \cdot 0{,}23 + 32 \cdot 0{,}05 =$ 9,27 GB

20%-Perzentil → Ausprägung dort ablesen, wo in der Tabelle eine relative Summenhäufigkeit von 0,20 überschritten wird: 4 GB

Interpretationen:

Modus: Die am Häufigsten vorkommende Speichergröße von 2009 verkauften USB-Sticks ist 8 GB.

Median: Mindestens eine Hälfte der USB-Sticks hatte höchstens 8, mindestens die andere Hälfte mindestens 8 GB Speicher.

Mittelwert: Würde man die Speicherkapazität der USB-Sticks gleichmäßig auf alle USB-Sticks aufteilen, so würden in jedem USB-Sticks 9,27 GB Platz haben.

20%-Perzentil: In mindestens 20% der USB-Sticks befinden sich höchstens 4 GB Speicher und in mindestens 80% mindestens 4 GB.

Aufgabe 20* [Kennzahlen der Lage]

Die Aktie eines Unternehmens wuchs innerhalb von fünf Jahren im 1. Jahr um 1,5 Prozent, im 2. um 28,2 Prozent, im 3. um 80,7 Prozent, im 4. Um 14,5 Prozent und fiel im 5. Jahr um 0,7 Prozent. Berechnen Sie

a) die fünf Wachstumsfaktoren und damit ausgehend von einem Basiskurs von 100 Euro vor Beginn dieser Zeitspanne den Aktienkurs am Ende der fünf Jahre unter Berücksichtigung des jährlichen Wachstums,
b) mit den fünf Wachstumsfaktoren das durchschnittliche jährliche prozentuelle Kurswachstum mit dem Taschenrechner und überprüfen Sie Ihr Ergebnis in Excel mit Hilfe der im Internet bereitstehenden Excel-Lerndatei.
c) Wie hoch hätte demnach der Prozentsatz eines festverzinslichen Wertpapiers sein müssen, um einen höheren Gewinn als die Aktie zu erzielen?

Lösung:

Jahr	jährl. Wachstum in %	Wachstumsfaktoren
1	1,5	1,015
2	28,2	1,282
3	80,7	1,807
4	14,5	1,145
5	-0,7	0,993

Produkt der Wachstumsfaktoren 2,673

Basiskurs	100
Endkurs	267,342
geometr. Mittel	1,217

a)

Berechnung:
Wachstumsfaktor des 1. Jahres: 1 + 1,5 : 100 = 1,015
Wachstumsfaktor des 2. Jahres: 1 + 28,2 : 100 = 1,282
Wachstumsfaktor des 3. Jahres: 1 + 80,7 : 100 = 1,807
Wachstumsfaktor des 4. Jahres: 1 + 14,5 : 100 = 1,145
Wachstumsfaktor des 5. Jahres: 1 + (-0,7) : 100 = 0,993
Kurs nach dem 1. Jahr: 100 · 1,105 = 101,5
Kurs nach dem 2. Jahr: 101,5 · 1,282 = 130,123
Kurs nach dem 3. Jahr: 130,123 · 1,807 = 235,132
Kurs nach dem 4. Jahr: 235,132 · 1,145 = 269,226
Endkurs ist der Kurs nach dem letzten Jahr: 269,226 · 0,993 = 267,342

b)
Berechnung: $g = \sqrt[n]{x_1 \cdot x_2 \cdot, \dots, \cdot x_n}$
Geometrischer Mittelwert ist hier die 5. Wurzel aus dem Produkt der Wachstumsfaktoren:
$g = \sqrt[5]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5} = \sqrt[5]{1{,}015 \cdot 1{,}282 \cdot 1{,}807 \cdot 1{,}145 \cdot 0{,}993}$=1,217

Interpretation:
Würde das jährliche prozentuelle Wachstum der Aktie jeweils 21,7 % betragen, dann ergäbe sich (bis auf Rundungsfehler) derselbe Endkurs 267,342 wie bei der tatsächlichen jährlichen, prozentuellen Entwicklung.

c)
Jedes festverzinsliche Wertpapier mit einem Prozentsatz von mehr als 21,7 % pro Jahr würde einen höheren Gewinn erzielen als die Aktie (z.B.: 100 · $1{,}218^5$ = 268,063)

Aufgabe 21 [Kennzahlen der Lage]

Der Preisindex für die Lebenshaltung lag vor genau drei Jahren bei einem Wert von 204. Nach diesen drei Jahren beträgt er 262. Wie groß war
a) der gesamte prozentuelle Zuwachs in diesen drei Jahren?
b) die durchschnittliche jährliche Inflationsrate?

Lösung:

gesamter prozentueller Zuwachs	28,43%
durchschnittl. prozentuelle Inflationsrate	8,70%

Berechnung:
gesamter prozentueller Zuwachs: 262 : 204 = 1,2843 (= 28,4 %)
durchschnittliche prozentuelle Inflationsrate: $g = \sqrt[3]{1{,}2843} = 1{,}087\ (= 8{,}7\ \%)$

Aufgabe 22 [Kennzahlen der Streuung]

Berechnen Sie mit dem Taschenrechner Varianz, Standardabweichung und Variationskoeffizient des Merkmals mit folgenden Ausprägungen 66, 85 und 55. (=Länge der ersten 3 Etappen aus Aufgabe 18)

Lösung:

Varianz:	153,56
Standardabweichung:	12,39

Variationskoeffizient: 0,18

Berechnung:

Mittelwert: $\bar{x} = \frac{66+85+55}{3} = 68{,}67$

Varianz: $s^2 = \frac{(66-68{,}67)^2+(85-68{,}67)^2+(55-68{,}67)^2}{3} = \frac{(-2{,}67)^2+(16{,}33)^2+(-13{,}67)^2}{3} = \frac{7{,}11+266{,}78+186{,}67}{3} = 153{,}56$

Standardabweichung: $s = \sqrt{s^2} = 12{,}39$

Variationskoeffizient: $v = \frac{s}{\bar{x}} = \frac{12{,}39}{68{,}67} = 0{,}18$

Interpretationen:

Varianz und **Standardabweichung** sind nicht sehr anschaulich interpretierbar (Die Varianz wäre etwa die durchschnittliche quadrierte Abweichung der einzelnen km-Zahlen von ihrem Mittelwert 68,67). Eine Interpretation ist aber auch nicht nötig, da diese Kennzahlen meist nur Mittel zum Zweck sind. Der **Variationskoeffizient**, der zum Vergleich von Streuungen verschiedener Verteilungen verwendet wird, gibt die Größe der Standardabweichung in Relation zum Mittelwert an. Hier beträgt die Standardabweichung 18 % des Mittelwerts.

Aufgabe 23* [Kennzahlen der Streuung]

Berechnen Sie für das Merkmal aus Aufgabe 4 unter Verwendung der im Internet bereitstehenden Excel-Lerndatei und der darin befindlichen Anweisungen Varianz, Standardabweichung und Variationskoeffizient.

Lösung:

Varianz:	1,708
Standardabweichung:	1,307
Mittelwert:	1,350
Variationskoeffizient:	0,968

Aufgabe 24* [Kennzahlen der Streuung]

Berechnen Sie für das Merkmal Einwohner aus Aufgabe 17 in Excel Varianz, Standardabweichung und Variationskoeffizient. Halbieren Sie nun die Einwohnerzahl jedes Landes aus Aufgabe 17 und berechnen Sie in Excel (Excelblatt „Ü17“) neuerlich Varianz, Standardabweichung und Variationskoeffizient. Wie wirkt sich diese "Transformation" auf die Kennzahlen aus?

Lösung:

Staat	**EW (in Mio)**			**Staat**	**EW (in Mio)**		
Belgien	11,2			Belgien	5,6		
Bulgarien	7,3			Bulgarien	3,65		
Dänemark*	5,64			Dänemark	2,82		
Deutschland	80,5			Deutschland	40,25		
Estland	1,3			Estland	0,65		
Finnland	5,4	**s^2 =**	520,5	Finnland	2,7	**s^2 =**	130,1
Frankreich	65,6	**s =**	22,8	Frankreich	32,8	**s =**	11,4
Griechenl.	11,1	**MW=**	18,1	Griechenl.	5,55	**MW=**	9,0
Großbrit.	63,9	**v =**	1,3	Großbrit.	31,95	**v =**	1,3
Irland	4,6			Irland	2,3		
Italien	59,7			Italien	29,85		
Kroatien	4,3			Kroatien	2,15		
Lettland	2,0			Lettland	1,0		
Litauen	3,0			Litauen	1,5		
Luxemburg	0,5			Luxemburg	0,25		
Malta	0,4			Malta	0,2		
Niederlande	16,8			Niederlande	8,4		
Österreich	8,5			Österreich	4,25		
Polen	38,5			Polen	19,25		
Portugal	10,5			Portugal	5,25		
Rumänien	20,0			Rumänien	10,0		
Schweden	9,6			Schweden	4,8		
Slowakei	5,4			Slowakei	2,7		
Slowenien	2,1			Slowenien	1,05		
Spanien	46,7			Spanien	23,35		
Tschechien	10,5			Tschechien	5,25		
Ungarn	9,9			Ungarn	4,95		
Zypern	0,9			Zypern	0,45		

*) Es wird mit der Einwohnerzahl 5,64 Mio. gerechnet. Diese ist jedoch ein Fehler im Buch! Eigentlich sollte auch für Dänemark nur eine Stelle nach dem Komma, also 5,6 Mio., angegeben werden. In einem Nachdruck zur 4. Auflage wird dies korrigiert werden. Die auf eine Stelle nach dem Komma gerundeten Resultate für Varianz, Standardabweichung, Mittelwert und Variationskoeffizient werden dadurch jedoch nicht verändert.

Interpretation:

Der **Mittelwert** wird genauso wie die einzelnen Einwohnerzahlen einfach halbiert, da die Summe der Einwohnerzahlen nur noch die Hälfte der ursprünglichen Summe ist (abgesehen von Rundungsfehler beim Runden an der 1. Stelle nach dem Komma).

Die **Varianz** s^2 wird nach Halbierung der Einwohnerzahlen nicht halbiert, sondern sie ist nur noch ein Viertel der ursprünglichen Varianz, da bei der Berechnung der Varianz die Differenzen quadriert werden.

Die **Standardabweichung** s ist die Wurzel aus der Varianz. Wurde Letztere durch die „Transformation“ geviertelt, so muss die Standardabweichung halbiert sein.

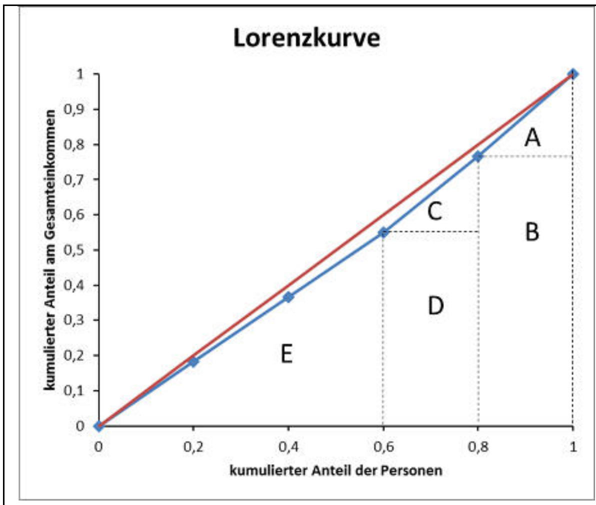
Der **Variationskoeffizient** v bleibt unverändert, da Standardabweichung und Mittelwert jeweils halbiert werden und somit das Verhältnis der beiden Werte zueinander gleich bleibt.

Aufgabe 25* [Konzentration]

Erstellen Sie unter Verwendung der im Internet bereitstehenden Excel-Lerndatei bei Einkommen von fünf Personen in der Höhe von 11.000, 13.000, 14.000, 11.000 und nochmals 11.000 Euro in Excel die für das Zeichnen der Lorenzkurve der Konzentration der Merkmalssumme auf die einzelnen Erhebungseinheiten nötige Tabelle und zeichnen Sie diese anschließend der Anleitung folgend. Berechnen Sie ferner mit dem Taschenrechner den dazugehörigen normierten Ginikoeffizienten.

Lösung:

Person	Anteile an der Grundge-samtheit	Kumulierte Anteile an der Grundge-samtheit	Einkommen	Anteile am Gesamt=einkommen	Kumulierte Anteile am Gesamt-einkommen
		0			0
A	0,2	0,2	11000	0,183	0,183
D	0,2	0,4	11000	0,183	0,367
E	0,2	0,6	11000	0,183	0,550
B	0,2	0,8	13000	0,217	0,767
C	0,2	1	14000	0,233	1,000
			60000		



Fläche unterhalb der (blaufarbenen) Lorenzkurve:
$A: (1-0{,}8)\cdot(1-0{,}767)/2 = 0{,}0233$
$B: (1-0{,}8)\cdot 0{,}767 = 0{,}1534$
$C: (0{,}8-0{,}6)\cdot(0{,}767-0{,}550)/2 = 0{,}0217$
$D: (0{,}8-0{,}6)\cdot 0{,}550 = 0{,}11$
$E: (0{,}6-0)\cdot 0{,}550/2 = 0{,}165$
Summe A-E: 0,4734

Fläche unter Diagonale = 0,5
Fläche zwischen Lorenzkurve und Diagonale =
$0{,}5 - 0{,}4734 = 0{,}0266$

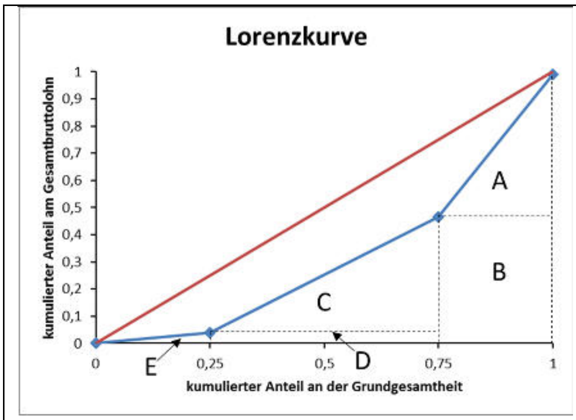
Fläche zwischen Lorenzkurve und Diagonale bei Maximalkonzentration (N=5): $\frac{1}{2}\cdot\left(1-\frac{1}{N}\right) = \frac{1}{2}\cdot\left(1-\frac{1}{5}\right) = 0{,}4$
Normierter Gini-Koeffizient: $\frac{0{,}0266}{0{,}4} = 0{,}067$

Interpretation:
Ein normierter Ginikoeffizient von 0,067 bedeutet, dass die Konzentration des Gesamteinkommens (inklusive Prämien) gemessen an der Lorenzkurve nur 6,7 % der Maximalkonzentration beträgt. Somit liegt sie nahe der Nullkonzentration, bei der das Gesamteinkommen völlig gleichmäßig auf alle Erhebungseinheiten aufgeteilt wäre.

Aufgabe 26 [Konzentration]
Laut Statistik Austria (Lohnsteuerstatistik 2005) verdienen die 25 Prozent der am schlechtesten verdienenden österreichischen Arbeitnehmer 3,9 Prozent des Gesamtbruttolohns, die 50 Prozent der mittleren Verdienenden 42,6 Prozent des Gesamtbruttolohns und die 25 Prozent der am besten Verdienenden insgesamt 53,5 Prozent des Gesamten. Skizzieren Sie die Lorenzkurve der Konzentration des Gesamtbruttolohns auf die drei Gruppen von Arbeitnehmer und berechnen Sie den normierten Ginikoeffizienten. Die Anzahl der österreichischen Arbeitnehmer beträgt 4 Mio.

Lösung:

	Anteile an der Grundge-samtheit	Kumulierte Anteile an der Grundge-samtheit	Anteile am Gesamt-bruttolohn	Kumulierte Anteile am Gesamt-bruttolohn
		0		0
schlechte Verdiener	0,25	0,25	0,039	0,039
mittlere Verdiener	0,5	0,75	0,426	0,465
beste Verdiener	0,25	1	0,535	1



Fläche unterhalb der (blaufarbenen) Lorenzkurve:
$A: (1 - 0{,}75) \cdot (1 - 0{,}465)/2 = 0{,}066875$
$B: (1 - 0{,}75) \cdot 0{,}465 = 0{,}11625$
$C: (0{,}75 - 0{,}25) \cdot (0{,}465 - 0{,}039)/2 = 0{,}1065$
$D: (0{,}75 - 0{,}25) \cdot 0{,}039 = 0{,}0195$
$E: (0{,}25 - 0) \cdot 0{,}039/2 = 0{,}004875$
Summe A-E: 0,314

Fläche unter Diagonale = 0,5
Fläche zwischen Lorenzkurve und Diagonale =
$0{,}5 - 0{,}314 = 0{,}186$

Fläche zwischen Lorenzkurve und Diagonale bei Maximalkonzentration (N=4 Mio): $\frac{1}{2} \cdot \left(1 - \frac{1}{4\,Mio}\right) =$ $\approx 0{,}5$ → Die Lorenzkurve würde im Fall der Maximalkonzentration praktisch von der Koordinate (0/0) über (1/0) zu (1/1) gehen.

Normierter Gini-Koeffizient: $\frac{0{,}0186}{0{,}5} = 0{,}372$

Interpretation:
Die Konzentration des Gesamtbruttolohns auf die Arbeitnehmenden beträgt gemessen an der Lorenzkurve 37,2 % der Maximalkonzentration.

Aufgabe 27 [statistischer Zusammenhang]
Gemessen wurde der Zusammenhang zwischen Prüfungserfolg (Merkmalsausprägungen: bestanden, nicht bestanden) und selbstständiger Lernfrequenz (regelmäßig, nicht regelmäßig) bei 200 Studierenden:

	Lernfrequenz	
Erfolg	regelmäßig	Nicht regelmäßig
Bestanden	152	8
Nicht bestanden	8	32

Berechnen Sie mit dem Taschenrechner eine geeignete Kennzahl zur Messung des statistischen Zusammenhanges zwischen diesen beiden Merkmalen.

Lösung:

Berechnung:
… der **beobachteten relativen Häufigkeiten**: $p_{ij} = \frac{h_{ij}}{N}$

z.B. für die 1. Zelle (i=1 und j=1): $p_{11} = \frac{h_{11}}{N} = \frac{152}{200} = 0{,}76$
alle anderen Häufigkeiten (auch die der Randverteilungen) ebenfalls durch 200 dividieren (Relation zur Größe der Grundgesamtheit)

Beobachtete relative Häufigkeiten:

		Lernfrequenz regelmäßig	Lernfrequenz nicht regelmäßig	Randverteilung $N_{i.}$
Erfolg	bestanden	0,76	0,04	0,8
	nicht bestanden	0,04	0,16	0,2
	Randverteilung $N_{.j}$	0,8	0,2	1

… der **bei Fehlen eines Zusammenhangs erwarteten relativen Häufigkeiten**:

$p_{ij}^{e} = p_{.j}^{b} \cdot p_{i.}^{b}$ (Produkt der beobachteten Randhäufigkeiten)

Multiplikation der jeweiligen Randhäufigkeiten → z.B. für die 1. Zelle (i=1 und j=1): $p_{11}^{e} = p_{.1}^{b} \cdot p_{1.}^{b} = 0{,}8 \cdot 0{,}8 = 0{,}64$

		Lernfrequenz regelmäßig	Lernfrequenz nicht regelmäßig	Randverteilung $p_{i.}$
Erfolg	bestanden	0,64	0,16	0,8
	nicht bestanden	0,16	0,04	0,2
	Randverteilung $p_{.j}$	0,8	0,2	1

$$\chi^2 = N \cdot \sum \frac{\left(p_{ij}^b - p_{ij}^e\right)^2}{p_{ij}^e} \quad \text{mit b=beobachtet und e=erwartet.}$$

$$\chi^2 = 200 \cdot \left(\frac{(0{,}76 - 0{,}64)^2}{0{,}64} + \frac{(0{,}04 - 0{,}16)^2}{0{,}16} + \frac{(0{,}04 - 0{,}16)^2}{0{,}16} + \frac{(0{,}16 - 0{,}04)^2}{0{,}04} \right) = 112{,}5$$

k … Anzahl der Ausprägungen des Merkmals Erfolg: k = 2
l … Anzahl der Ausprägungen des Merkmals Lernfrequenz: l = 2 → min(k,l) = 2

Cramers V: $V = \sqrt{\frac{\chi^2}{N \cdot (\min(k,l) - 1)}} = \sqrt{\frac{112{,}5}{200 \cdot (2-1)}} = 0{,}75$

Interpretation:
Faustregel für V zwischen 0,6 und unter 1: Es besteht ein starker statistischer Zusammenhang zwischen den Merkmalen Lernfrequenz und Erfolg.

Zusätzliche Erklärung für Aufgabe 27:
Da es häufig am Anfang schwierig fällt, den Zusammenhang zwischen den Randhäufigkeiten und den erwarteten relativen Häufigkeiten zu erfassen, hier eine ausführlichere Lösungsskizze die dem Verständnis dienen soll und nicht in dieser umfassenden Form in einer Klausur abgefragt wird.

Beobachtete relative Häufigkeiten:
Relative und (in Klammern) absolute beobachtete Häufigkeiten

		Lernfrequenz		
		regelmäßig	nicht regelmäßig	
Erfolg	bestanden	0,76 (152)	0,04 (8)	0,8 (160)
	nicht bestanden	0,04 (8)	0,16 (32)	0,2 (40)
		0,8 (160)	0,2 (40)	1 (200)

Interpretation einiger (aber nicht aller) Randhäufigkeiten: 80% der Studenten haben bestanden, 80% haben regelmäßig gelernt. Allerdings bedeutet das nicht, dass 80% der Studenten die regelmäßig gelernt haben, auch bestehen. Es ist einleuchtend, dass dieser Anteil wesentlich höher als 80% ist, da Lernen tatsächlich zum Klausurerfolg beiträgt. Also wie hoch ist die Wahrscheinlichkeit zu bestehen, wenn man lernt? Diese Frage kann man mit Hilfe von bedingten relativen Häufigkeiten beantworten, die in der nächsten Tabelle abgebildet sind. Von den 160 Studierenden, die regelmäßig gelernt haben, bestand 152, also 95% die Klausur und nur 8 Studenten (5%) bestanden die Klausur nicht. Diese hohe relative Häufigkeit des Bestehens unter der Bedingung, dass Studenten regelmäßig lernen liegt natürlich darin begründet, dass regelmäßiges Lernen beim Bestehen der Klausur hilft. Mit umgekehrten Vorzeichen lassen sich die bedingten relativen Häufigkeiten der nicht regelmäßig Lernenden interpretieren.

Bedingte relative Häufigkeit (in Klammern) absolute Häufigkeiten nach Lernfrequenz

		Lernfrequenz regelmäßig	nicht regelmäßig	
Erfolg	bestanden	0,95 (152)	0,2 (8)	
	nicht bestanden	0,05 (8)	0,8 (32)	
		0,8 (160)	0,2 (20)	

Die Vermutung eines statistischen Zusammenhangs zwischen regelmäßigem Lernen und Bestehen sowie nicht regelmäßigen Lernens und Scheitern in der Klausur liegt also nahe. Wie kann man das statistisch sichtbar machen? Hierzu brauchen wir das Konzept der erwarteten Häufigkeiten. Erwartete Häufigkeiten bezeichnen die Verteilung der Studenten auf die einzelnen Zellen wenn es keinen statistischen Zusammenhang zwischen Lernen und Bestehen geben würde, also wenn beides unabhängig voneinander wäre. Dazu ein Gedankenexperiment: 80% der Studierenden haben regelmäßig gelernt, die andere 20% haben es nicht. Wenn jetzt das Bestehen der Klausur unabhängig vom Lernaufwand ist, müsste die Bestehensquote für die Klausur in beiden Gruppen gleich sein. Wie hoch müsste sie sein? Ein Blick zurück auf die vorhergehende Tabelle zeigt, dass 80% aller Studenten die Klausur bestehen. Bei Unabhängigkeit zwischen Lernen und Klausurerfolg, müssten also 80% der regelmäßig Lernenden wie auch 80% der nicht regelmäßig Lernenden die Klausur bestehen.
Wenn das schwierig vorzustellen ist, kann man einfach das Merkmal Lernfrequenz mit dem Merkmal ersetzen mit welchem Fuß man aufgestanden ist (mit dem linken Fuß oder mit dem rechten Fuß). Bei Unabhängigkeit zwischen dem Aufstehen mit einem bestimmten Fuß und bestehen der Klausur müssten 80% der mit dem rechten wie mit dem linken Fuß aufgestandenen Studenten die Klausur bestehen.
In der Realität ist die Bestehensquote sicherlich unabhängig von dem Fuß mit dem man aufsteht, aber nicht unabhängig von der Lernfrequenz. Aber wenn die Lernfrequenz nichts mit dem Bestehen der Klausur zu tun **hätte**, von wie vielen Studenten **hätten** wir denn dann erwartet die Klausur zu bestehen, bzw. durchzufallen? Um die Antwort auf diese Frage braucht es nur noch einen kleinen logischen Dreh. Wenn 80% (160) der Studenten gelernt haben und von diesen würden 80% bestehen, dann wären dies natürlich 80% von 160 Studenten: also 128 (=160 × 0,08). Dies entspräche genau 64% aller Studenten (=128 : 200). Auf diese Zahl von 64% könnte man auch einfacher ohne Rückgriff auf die erwartete absolute Häufigkeit von 128 Studenten kommen, in dem man einfach die Randhäufigkeiten „bestehen“ und „regelmäßig Lernen“ miteinander multipliziert. 80% haben regelmäßig gelernt und davon bestehen 80%: dies entspricht 0,8 × 0,8 = 0,64, also 64% der gesamten Studierenden. Diesen Gedankengang wenden wir jetzt natürlich auch für die Gruppe der Nichtregelmäßiglernenden an. Wenn 20% nicht regelmäßig gelernt haben, aber Lernen unabhängig vom Klausurerfolg ist, bestehen 80% dieser Gruppe ebenso die Klausur. Dies sind also 0,2 × 0,8 = 0,16, sprich 16% aller 200 Studierenden, also 32 Studierende.
Den gleichen Gedankengang wiederholt man jetzt für die Gruppe der 20% Studierenden, welche die Klausur nicht bestanden haben. Wenn 20% der Studierenden nicht bestehen, aber dies unabhängig vom Lernen ist, sollten 20% der regelmäßig Lernenden wie auch 20% der nicht regelmäßig Lernenden die Klausur nicht bestehen. In der Gruppe der 80% Studenten, die regelmäßig gelernt haben, sind dies also 0,8 × 0,2 = 0,16, sprich 16% aller 200 Studierenden, also 32 Studierende. In der Gruppe der 20% Studenten, die nicht regelmäßig gelernt haben, sind dies also 0,2 × 0,2 = 0,04, sprich 4% aller 200 Studierenden, also 8 Studierende.
Nochmal zusammengefasst mit den Randhäufigkeiten:
80% haben regelmäßig gelernt: Davon bestehen 80% → 0,8 × 0,8 = 0,64
20% haben nicht regelmäßig gelernt: Davon bestehen 80% → 0,2 × 0,8 = 0,16
80% haben regelmäßig gelernt: Davon bestehen 20% nicht → 0,8 × 0,2 = 0,16
20% haben nicht regelmäßig gelernt: Davon bestehen 20% nicht → 0,2 × 0,2 = 0,04
In tabellarischer Form ergibt sich:

Bei Fehlen eines Zusammenhangs erwartete relative (in Klammern) absolute Häufigkeiten:

		Lernfrequenz		
		regelmäßig	nicht regelmäßig	
Erfolg	bestanden	0,64 (128)	~~0,16 (32)~~	0,8 (160)
	nicht bestanden	0,16 (32)	0,04 (8)	0,2 (40)
		0,8 (160)	0,2 (40)	1

Der Unterschied zwischen den erwarteten relativen Häufigkeiten und den beobachteten relativen Häufigkeiten ist dann die Basis für die Berechnung des Chi-Quadrats. Je größer die Differenz, desto größer ist der Einfluss von Lernen bzw. Nichtlernen auf den Klausurerfolg und desto größer wird Chi-Quadrat.

Aufgabe 28* [statistischer Zusammenhang]

Verwenden Sie die Angaben aus Aufgabe 11 und berechnen Sie mit Hilfe der im Internet bereitgestellten Excel-Lerndatei die Tabelle der bei Fehlen eines statistischen Zusammenhangs erwarteten relativen Häufigkeiten. Darauf basierend berechnen Sie eine geeignete Kennzahl, um festzustellen, wie stark der statistische Zusammenhang zwischen den beiden Merkmalen in dieser Grundgesamtheit ist.

Lösung:

beobachtete absolute Häufigkeiten:

		EURO-Einstellung			
		zufrieden	unzufr.	neutral	
Geschlecht	männlich	241	122	40	403
	weiblich	198	173	20	391
		439	295	60	794

beobachtete relative Häufigkeiten:

		EURO-Einstellung			
		zufrieden	unzufr.	neutral	
Geschlecht	männlich	0,304	0,154	0,050	0,508
	weiblich	0,249	0,218	0,025	0,492
		0,553	0,372	0,076	1

Bei Fehlen eines Zusammenhangs erwartete rel. Häufigkeiten:

		EURO-Einstellung			
		zufrieden	unzufr.	neutral	
Geschlecht	männlich	0,281	0,189	0,038	0,508
	weiblich	0,272	0,183	0,037	0,492
		0,553	0,372	0,076	1

		EURO-Einstellung			
		zufrieden	unzufr.	neutral	
Geschlecht	männlich	0,0019	0,0065	0,0038	
	weiblich	0,0019	0,0067	0,0039	

Chiquadrat	19,519
Cramers V	0,157

Interpretation:
Faustregel für V über 0 und bis 0,2: Es besteht ein schwacher statistischer Zusammenhang zwischen den Merkmalen Geschlecht und EURO-Einstellung.

Aufgabe 29* [statistischer Zusammenhang]

Fortsetzung von Aufgabe 13: In einer Studie dänischer und japanischer Wissenschaftler an 400 Kindern wurden das Geschlecht der Kinder und das Rauchverhalten der Eltern in den drei Monaten vor der Zeugung erhoben. Berechnen Sie in Excel die Stärke des statistischen Zusammenhanges zwischen den beiden Merkmalen.

Lösung:

beobachtete absolute Häufigkeiten:

		Rauchverhalten der Eltern			
		2 x NR (=1)	1 NR, 1 R (=2)	2 x R (=3)	
Geschlecht	männlich	97	55	43	195
	weiblich	86	61	58	205
		183	116	101	400

Beobachtete rel. Häufigkeiten:

		Rauchverhalten der Eltern			
		2 x NR (=1)	1 NR, 1 R (=2)	2 x R (=3)	
Geschlecht	männlich	0,243	0,138	0,108	0,488
	weiblich	0,215	0,153	0,145	0,513
		0,458	0,290	0,253	1,000

Erwartete relative Häufigkeiten:

		Rauchverhalten der Eltern			
		2 x NR (=1)	1 NR, 1 R (=2)	2 x R (=3)	
Geschlecht	männlich	0,223	0,141	0,123	0,488
	weiblich	0,234	0,149	0,129	0,513
		0,458	0,290	0,253	1,000

Chiquadrat:	2,96
Cramers V:	0,086

Berechnung:
wie in Ü27
Interpretation:
Faustregel für V über 0 und bis 0,2: Es besteht ein schwacher statistischer Zusammenhang zwischen den Merkmalen Geschlecht und EURO-Einstellung.

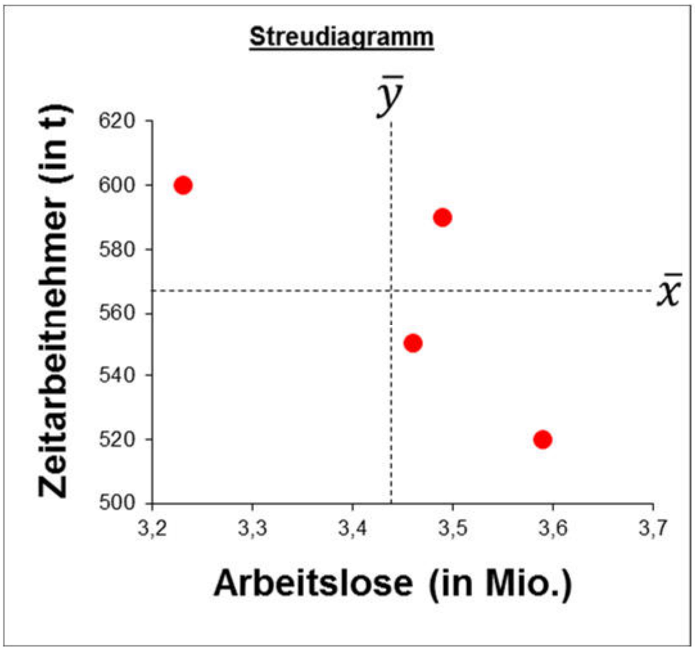
Aufgabe 30 [statistischer Zusammenhang]

"Sie werden als Erste entlassen, aber offenbar auch als Erste wieder eingestellt: Die Zahl der Zeitarbeitnehmer in Deutschland steigt nach einem Einbruch Anfang des Jahres 2009 kontinuierlich an.... Die Kunden der Zeitarbeitsfirmen nutzen die Möglichkeit, Arbeiter zu beschäftigen, von denen sie sich auch schnell wieder trennen können" *(DIE ZEIT,* Nr.51, 10. Dezember 2009, S.35):

Anfang des Quartals	Arbeitslose (in Mio.)	Zeitarbeitnehmer (in T.)
I	3,49	590
II	3,59	520
III	3,46	550
IV	3,23	600

Skizzieren Sie per Hand das Streudiagramm dieser Verteilung und berechnen Sie mit dem Taschenrechner den Korrelationskoeffizienten der beiden Merkmale.

Lösung:



Kovarianz: - 3,288
Korrelationskoeffizient: - 0,779

Berechnung:

x ... Arbeitslose: $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{3{,}49+3{,}59+3{,}46+3{,}23}{4} = 3{,}44$

y ... Zeitarbeitnehmer: $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{590+520+550+600}{4} = 565$

Standardabweichung von x: $s_x^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N} = \frac{(3{,}49-3{,}44)^2+(3{,}59-3{,}44)^2+(3{,}46-3{,}44)^2+(3{,}23-3{,}44)^2}{4} =$
$0{,}017 \;\rightarrow s_x = \sqrt{s_x^2} = 0{,}132$

Standardabweichung von y: $s_y^2 = \frac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N} = \frac{(590-565)^2+(520-565)^2+(550-565)^2+(600-565)^2}{4} = 1025 \;\rightarrow$
$s_y = \sqrt{s_y^2} = 32{,}016$

Kovarianz von x und y: $s_{xy} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})\cdot(y_i-\bar{y})}{N} =$

$$\frac{(3{,}49-3{,}44)\cdot(590-565)+(3{,}59-3{,}44)\cdot(520-565)+(3{,}46-3{,}44)\cdot(550-565)+(3{,}23-3{,}44)\cdot(600-565)}{4} = -3{,}288$$

Korrelationskoeffizient: $r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-3{,}288}{0{,}132\cdot 32{,}06} = -0{,}779$

Interpretation:

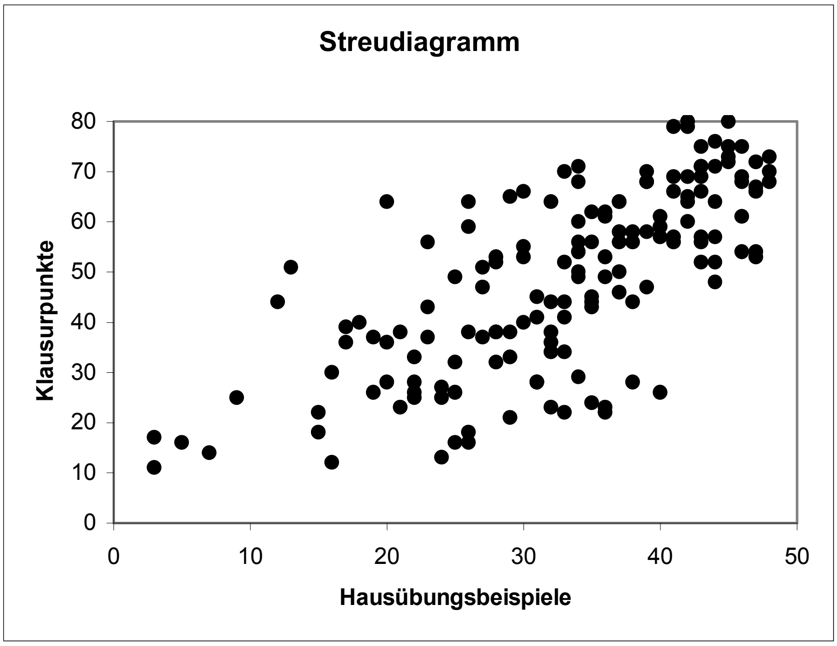
Zwischen den Merkmalen Arbeitslosenzahl und Zahl der Zeitarbeitnehmerinnen und -arbeitnehmer besteht ein starker, negativer linearer statistischer Zusammenhang.

Aufgabe 31* [statistischer Zusammenhang]

Bei der Erhebung des statistischen Zusammenhangs zwischen der Anzahl an vorbereiteten Hausübungsbeispielen in einem Mathematikkurs (x:0- 50 Beispiele) und den Punkten bei der Abschlussklausur (y:0 - 80Punkte) ergab sich für 155 Prüflinge eine Datenreihe, die im Internet als Excel-Lerndatei abgespeichert ist. Folgen Sie dort den Anweisungen und

a) stellen Sie die Daten in Excel in einem Streudiagramm dar,
b) berechnen Sie in Excel zuerst die beiden Mittelwerte, dann die beiden Standardabweichungen und die Kovarianz und zuletzt den Korrelationskoeffizienten zur Messung des statistischen Zusammenhangs zwischen den beiden Merkmalen.

Lösung:



	x	y
Mittelwerte	32,8	48,6
Standardabweichungen	10,103	17,912

Kovarianz	129,46
Korrelationskoeffizient	0,716

Interpretation:

Faustregel für einen Korrelationskoeffizient zwischen +0,6 und unter +1: Zwischen den Merkmalen Hausübungen und Klausurpunkte besteht ein starker, positiver linearer statistischer Zusammenhang.

Aufgabe 32 [statistischer Zusammenhang]

Verwenden Sie Ihre Ergebnisse von Aufgabe 31 b) und

a) berechnen Sie mit dem Taschenrechner a) die Gleichung der Regressionsgeraden,
b) mit dieser Regressionsgeraden eine Schätzung für die Punktezahl bei der Klausur bei x = 40 Hausübungsbeispielen,
c) das Bestimmtheitsmaß zur Messung der Güte dieses Regressionsmodells.

Die zusammengefassten Ergebnisse aus Aufgabe 31 sind:
$\bar{x} = 32{,}8; \bar{y} = 48{,}6;\ s_x = 10{,}103;\ s_y = 17{,}912;\ s_{xy} = 129{,}46; r = 0{,}716$

Lösung:

a)
Regressionsgerade: $y = 1{,}27 \cdot x + 6{,}95$

Berechnung:
Geradengleichung: $y = b_1 \cdot x + b_2$
Mit $b_1 = \frac{s_{xy}}{s_x^2} = \frac{129{,}46}{10{,}103^2} = 1{,}27$
Und $b_2 = \bar{y} - b_1 \cdot \bar{x} = 48{,}6 - 1{,}27 \cdot 32{,}8 = 6{,}95$

Schätzung für y bei x = 40:
Berechnung:
Regressionsgerade: $y = 1{,}27 \cdot x + 6{,}95$
für x = 40 einsetzen → $y = 1{,}27 \cdot 40 + 6{,}95 = 57{,}75$
Interpretation:
Auf Basis der Daten wird bei einer Hausübungszahl von x = 40 ein Klausurergebnis von 57,75 Punkte geschätzt.

c)
Bestimmtheitsmaß B: 0,513

Berechnung:
Bestimmtheitsmaß: $B = r^2 = 0{,}716^2 = 0{,}513$

Interpretation:
51,3 % der Streuung des zu schätzenden Merkmals y wird durch das Regressionsmodell erklärt. Das ist nach unserer Faustregel (Korrelation sollte im Bereich des starken statistischen Zusammenhangs liegen) ein ausreichend hoher Erklärungswert und das Vertrauen in die Schätzung der Klausurpunktezahl ist (zumindest als ungefährer Trend) groß.

Aufgabe 33 [statistischer Zusammenhang]

Berechnen Sie mit den Daten aus Aufgabe 30 mit dem Taschenrechner

a) die Gleichung der Regressionsgeraden,

b) die Schätzung für die Zahl der Zeitarbeitnehmer bei einer Arbeitslosenzahl von 3,5 Millionen,

c) das Bestimmtheitsmaß dieses Regressionsmodells.

Die Daten aus Aufgabe 30 zusammengefasst sind:
$\bar{x} = 3{,}44; \bar{y} = 565;\ s_x = 0{,}132;\ s_y = 32{,}016;\ s_{xy} = -3{,}288; r = -0{,}779$

Lösung:

a)
Regressionsgerade: $y = -\ 188{,}68 \cdot x + 1.214{,}63$

Berechnung:

Geradengleichung: $y = b_1 \cdot x + b_2$

Mit $b_1 = \frac{s_{xy}}{s_x^2} = \frac{3{,}443}{0{,}132^2} = -188{,}68$

Und $b_2 = \bar{y} - b_1 \cdot \bar{x} = 565 - (-188{,}68) \cdot 3{,}443 = 1.214{,}63$

b)
Schätzung für y bei x = 3,5: $y = 554{,}25$

Berechnung:

Regressionsgerade: $y = -\ 188{,}68 \cdot x + 1214{,}63$

für x = 3,5 einsetzen → $y = -\ 188{,}68 \cdot 3{,}5 + 1214{,}63 = 554{,}25$

c)
Bestimmtheitsmaß B: 0,607

Berechnung:

Bestimmtheitsmaß: $B = r^2 = (-0{,}7791)^2 = 0{,}607$

Interpretation:

Durch das hohe Bestimmtheitsmaß ist unser Vertrauen in die Schätzung in b) groß.

Aufgabe 34 [statistischer Zusammenhang]

Bei 4 Schülern (A-D) wurden die Mathematiknoten und die Ränge bei einem Sportwettbewerb gemessen:

Schüler	Mathematiknote	Rang im Sport
A	4	1
B	1	2
C	4	4
D	5	3

Berechnen Sie in Excel den Rangkorrelationskoeffizienten. Dazu müssen Sie zuerst auch aus den Mathematiknoten der 4 Schüler die *Ränge* in Mathematik bestimmen.

Lösung

	Note (Math)	Rang (Sport)
A	4	1
B	1	2
C	4	4
D	5	3
	Rang (Math)	Rang (Sport)
A	2,5	1
B	1	2
C	2,5	4
D	4	3

Berechnung:
Da die Daten bei der Note noch nicht als Ränge gegeben sind, muss in Mathematik eine Rangliste wie beim Sport erstellt werden. B mit Note 1 ist bester und erhält den Rang 1. A und C mit Noten 4 sind beide 2. bzw. 3.-beste und erhalten den geteilten Rang 2,5. D ist in Mathematik Vierter. Jetzt ist so vorzugehen wie für den Korrelationskoeffizienten bei metrischen Merkmalen.

u … Mathematikränge: $\bar{u} = \frac{\sum_{i=1}^{N} u_i}{N} = \frac{2{,}5+1+2{,}5+4}{4} = 2{,}5$
y … Sportränge: $\bar{v} = \frac{\sum_{i=1}^{N} v_i}{N} = \frac{1+2+3+4}{4} = 2{,}5$

Standardabweichung von u: $s_u^2 = \frac{\sum_{i=1}^{N}(u_i-\bar{u})^2}{N} = \frac{(2{,}5-2{,}5)^2+(1-2{,}5)^2+(2{,}5-2{,}5)^2+(4-2{,}5)^2}{4} = 1{,}125 \rightarrow s_u = \sqrt{s_u^2} = 1{,}061$

Standardabweichung von v: $s_v^2 = \frac{\sum_{i=1}^{N}(v_i-\bar{v})^2}{N} = \frac{(1-2{,}5)^2+(2-2{,}5)^2+(3-2{,}5)^2+(4-2{,}5)^2}{4} = 1{,}25 \rightarrow s_v = \sqrt{s_v^2} = 1{,}118$

Kovarianz von x und y: $s_{xy} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})\cdot(y_i-\bar{y})}{N} = \frac{(2{,}5-2{,}5)\cdot(1-2{,}5)+(1-2{,}5)\cdot(2-2{,}5)+(2{,}5-2{,}5)\cdot(4-2{,}5)+(4-2{,}5)\cdot(3-2{,}5)}{4} = 0{,}375$

Korrelationskoeffizient: $r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{0{,}375}{1{,}061 \cdot 1{,}118} = 0{,}316$

Interpretation:
Faustregel für einen Spearmanscher Korrelationskoeffizienten der Rangzahlen zwischen +0,2 und +0,6: Zwischen den Rängen der beiden sportlichen Bewerbe besteht ein mittlerer, positiver linearer statistischer Zusammenhang.

Aufgabe 35* [statistischer Zusammenhang]

Bei 4 Studierenden (A-D) wurden ihre Beurteilung der Qualität einer Lehrveranstaltung (nach Schulnoten) und ihre Klausurnoten in dieser Lehrveranstaltung beobachtet:

Studierende	Lehrveranstaltung	Klausur
A	1	1
B	2	2
C	2	1
D	4	4

Berechnen Sie in Excel den Rangkorrelationskoeffizienten. Dazu müssen Sie zuerst auch aus beiden Benotungen der 4 Studierenden die *Ränge* bestimmen.

Lösung:

	Lehrveranstaltung	Klausurnote
A	1	1
B	2	2
C	2	1
D	4	4
	Rang (LV)	Rang (Klausur)
A	1	1,5
B	2,5	3
C	2,5	1,5
D	4	4

Berechnung:
Siehe Excelanleitung oder wie in Übung 34

	x	y
Mittelwerte Ränge	2,5	2,5
Standardabweichungen Ränge	1,061	1,061
Kovarianz Ränge	0,938	
Rangkorrelationskoeffizient	0,833	

Interpretation:
Faustregel für einen Spearmanscher Korrelationskoeffizienten der Rangzahlen zwischen +0,6 und +1: Zwischen den Rängen der beiden sportlichen Bewerbe besteht ein starker, positiver linearer statistischer Zusammenhang.

Kapitel 2 Wahrscheinlichkeitsrechnung

Aufgabe 36 [Wahrscheinlichkeitsrechnung: Grundbegriffe]

Bei einem Rouletteversuch wird zufällig eine der ganzen Zahlen von 0 bis 36 realisiert. Wie groß sind die Wahrscheinlichkeiten dafür, dass die Kugel

a) auf die Zahl 22
b) auf eine gerade Zahl
c) eine Zahl zwischen 1 und 12 fällt (jeweils einschließlich),
d) eine Zahl zwischen 1 und 12 (jeweils einschließlich), gefallen ist, wenn man weiß, dass eine Zahl zwischen 1 und 18 gekommen ist (jeweils einschließlich)?

Lösung:

x … Zahl beim Roulette
Alle Zahlen sind gleich wahrscheinlich ➔ Abzählregel ist anwendbar
a)
37 mögliche Fälle, 1 günstiger Fall
→ $Pr(x = 22) = 1 : 37 = 0{,}027$

b)
37 mögliche Fälle, 18 günstige Fälle („0“ ist per Definition weder gerade noch ungerade)
→ $Pr(x \text{ gerade}) = 18 : 37 = 0{,}486$

c)
37 mögliche Fälle, 12 günstige Fälle
→ $Pr(1 \leq x \leq 12) = 12 : 37 = 0{,}324$

d)
Nur mehr 18 mögliche Fälle, 12 günstige Fälle
→ $Pr(1 \leq x \leq 12 \mid 1 \leq x \leq 18) = 12 : 18 = 0{,}666$

Interpretationen:
z.B. c) Bei 100 Versuchen (Wiederholungen) wird dies *durchschnittlich* in ca. 32 Versuchen passieren.
d) Durch die Information, dass eine Zahl von 1 bis 18 gewürfelt wurde, hat sich die Wahrscheinlichkeit für eine Zahl zwischen 1 und 12 mehr als verdoppelt und dies wird durchschnittlich in 2/3 Drittel der Fälle passieren.

Aufgabe 37 [Wahrscheinlichkeitsrechnung: Grundbegriffe]

Wie viele gleich wahrscheinliche Gruppen zu 5 Personen lassen sich aus 20 Personen (A-T) ziehen und wie wahrscheinlich ist es, dass

a) die Gruppe {A, B, C, D, E}
b) die Gruppe {A, C, F, H, M} gezogen wird,
c) die Person A in der 5-Personen-Gruppe vorkommt?

Lösung:

Die Anzahl an Möglichkeiten, die es gibt, um 5 Personen aus 20 Personen zu ziehen:

$$\binom{20}{5} = 15.504$$

Alle sind gleich wahrscheinlich, weshalb man wieder die Abzählregel verwenden kann.

a)
15.504 mögliche Fälle, 1 günstiger Fall
→ Pr({A,B,C,D,E}) = 1 : 15.504 = 0,000064

b)
Auch hier: 15.504 mögliche Fälle, 1 günstiger Fall
→ Pr({A,C,F,H,M}) = 1 : 15.504 = 0,000064

c)
Wie viele der 15.504 Gruppen enthalten A? Es sind dies alle Gruppen, die außer aus A aus irgendeiner 4er Kombination der restlichen 19 Elemente bestehen. Die Anzahl an Möglichkeiten, die es gibt, um 4 Personen aus 19 Personen zu ziehen:

$$\binom{19}{4} = 3.876$$

15.504 mögliche Fälle, 3.876 günstige Fälle
→ Pr(A enthalten) = 3.876 : 15.504 = 0,25
Alternativer Lösungsweg Es ist einfach die Wahrscheinlichkeit dafür, dass - wenn 5 aus 20 Elementen gezogen werden - ein bestimmtes Element wie „A" gezogen wird, gleich 5 : 20 = 0,25.
Interpretation:
Bei 100 Versuchen wird dies *durchschnittlich* in 25 Versuchen passieren.

Aufgabe 38 [Wahrscheinlichkeitsrechnung: Anwendung]

In einem Vorrat von neun Glühlampen befinden sich vier mit einer Leistung von 40 Watt und fünf mit einer Leistung von 60 Watt. Sie wählen zufällig drei Stück aus. Berechnen Sie mit dem Taschenrechner, wie wahrscheinlich es ist, dass sich darunter

a) keine,
b) eine,
c) mehr als eine

40-Watt-Glühlampe befindet.

Lösung:
Hypergeometrische Verteilung muss verwendet werden
N = 9, A = 4, n = 3
x … Anzahl der gezogenen 40 Watt-Glühlampen

Allgemein: Hypergeometrische Verteilung nach

$$\Pr(x = a) = \frac{\binom{A}{a} \cdot \binom{N-A}{n-a}}{\binom{N}{n}}$$

a)

$$\Pr(x=0)=\frac{\binom{4}{0}\cdot\binom{9-4}{3-0}}{\binom{9}{3}}=\frac{\binom{4}{0}\cdot\binom{5}{3}}{\binom{9}{3}}=\frac{1\cdot 10}{84}=0{,}119$$

Interpretation:
Bei 100 Versuchen wird dies *durchschnittlich* in 11,9 Versuchen passieren.

b)

$$\Pr(x=1)=\frac{\binom{4}{1}\cdot\binom{9-4}{3-1}}{\binom{9}{3}}=\frac{\binom{4}{1}\cdot\binom{5}{2}}{\binom{9}{3}}=\frac{4\cdot 10}{84}=0{,}476$$

c)

$$\Pr(x>1)=1-\Pr(x\leq 1)=1-0{,}119-0{,}476=0{,}405$$

Die Wahrscheinlichkeit mehr als eine 40-Watt-Glühbirne zu ziehen ist natürlich 1-die Wahrscheinlichkeit keine oder nur eine 40-Watt-Glühbirne zu ziehen.

Aufgabe 39 [Wahrscheinlichkeitsrechnung: Anwendung]

Alle vier Jahre werden "seit Generationen" Abziehbilder für das Fußballer-WM-Sammelalbum der Firma Panini gesammelt. Nehmen Sie an, dass jedes Paket mit fünf Abziehbilder bei der Produktion im Jahr 2010 völlig zufällig mit fünf verschiedenen aus den insgesamt 640 Fotos gefüllt wurde. Wie wahrscheinlich war es dann, dass man durch Kauf eines Paketes sein Album vervollständigt hat, wenn nur noch 5 Abziehbilder gefehlt haben?

Lösung:

Es müssen also genau die fehlenden fünf Bilder im gekauften Paket sein;
Anzahl aller Möglichkeiten an verschiedenen 5er Gruppen aus 640 Bildern :

$$\binom{640}{5}=\frac{640!}{5!\cdot 635!}=\frac{640\cdot 639\cdot 638\cdot 637\cdot 636}{120}=8{,}8088\cdot 10^{11}$$

Eine dieser 5er Gruppen ist die günstige. Die Wahrscheinlichkeit ist also enorm klein: 1 : 880 Milliarden!

Aufgabe 40 [Wahrscheinlichkeitsrechnung: Anwendung]

Aus einer Gruppe von 20 Personen mit acht Frauen und zwölf Männern werden fünf Personen zufällig ausgewählt. Wie wahrscheinlich ist es, dass in einer so ausgewählten Gruppe mehr Frauen als Männer sind?

Lösung:
Hypergeometrische Verteilung muss verwendet werden.

$$\Pr(x=a)=\frac{\binom{A}{a}\cdot\binom{N-A}{n-a}}{\binom{N}{n}}$$

N = 20, A = 8, n = 5
x … Anzahl der gezogenen Frauen

$$\Pr(x \geq 3) = \Pr(x = 3) + \Pr(x = 4) + \Pr(x = 5) = \frac{\binom{8}{3}\cdot\binom{12}{2}}{\binom{20}{5}} + \frac{\binom{8}{4}\cdot\binom{12}{1}}{\binom{20}{5}} + \frac{\binom{8}{5}\cdot\binom{12}{0}}{\binom{20}{5}} = 0,296$$

Interpretation:
Bei 100 Versuchen wird dies durchschnittlich in ca. 30 Versuchen passieren.

Aufgabe 41 [Wahrscheinlichkeitsrechnung: Anwendung]

Beim Elfmeterschießen treten im Fußball 5 Schützen je Mannschaft an. Angenommen die Trefferwahrscheinlichkeit aller Schützen einer Mannschaft ist jeweils 70%: Wie wahrscheinlich ist es, dass vier Schützen einer Mannschaft treffen?

Lösung:
Binomialverteilung muss verwendet werden.

$Pr(x = a) = \binom{n}{a} \cdot \pi^a \cdot (1-\pi)^{n-a}$ mit α=4, n=5, π=0,7

$Pr(x = 4) = \binom{5}{4} \cdot 0{,}7^4 \cdot (1-0{,}7)^{5-4} = 5 \cdot 0{,}2401 \cdot 0{,}3 = 0{,}36$

Interpretation:
Die Wahrscheinlichkeit dass genau vier Schützen treffen ist 36%.

Aufgabe 42 [Wahrscheinlichkeitsrechnung: Anwendung]

Erfahrungsgemäß bedarf ein neuer "Tablet - PC" einer bestimmten Marke innerhalb der ersten beiden Jahre mit einer Wahrscheinlichkeit von 0,2 einer Reparatur. Wie wahrscheinlich ist es, dass von 5 lagernden Geräten

a) keines
b) eines
c) mindestens eines

in den ersten beiden Jahren nach dem Verkauf repariert werden muss?

Lösung:
Binomialverteilung muss verwendet werden

$Pr(x = a) = \binom{n}{a} \cdot \pi^a \cdot (1-\pi)^{n-a}$ mit α=4, n=5, π=0,7

x … Anzahl der zu reparierenden Geräte
n = 5
π= 0,2

a) $\Pr(x = 0) = \binom{5}{0} \cdot 0{,}2^0 \cdot 0{,}8^5 = 0{,}328$

b) $\Pr(x = 1) = \binom{5}{1} \cdot 0{,}2^1 \cdot 0{,}8^4 = 0{,}410$

c) $\Pr(x \geq 1) = 1 - \Pr(x = 0) = 1 - 0{,}328 = 0{,}672$

Interpretation:
Zum Beispiel c) Die Wahrscheinlichkeit dafür, dass man mindestens ein (eins oder mehr) solches Gerät reparieren muss, beträgt etwas mehr als zwei Drittel.

Aufgabe 43 [Wahrscheinlichkeitsrechnung: Anwendung]

In einer Bevölkerung vom Umfang N=2 Millionen besitzen 40% eine bestimmte Eigenschaft. Es werden 40 Erhebungseinheiten zufällig (und natürlich ohne Zurücklegen) gezogen. Wie wahrscheinlich ist es, dass 16 davon diese Eigenschaft besitzen, dass also in der Stichprobe derselbe Anteil dieser Eigenschaft vorkommt, wie in der Grundgesamtheit? Sollten Sie Probleme mit den großen Zahlen bei der Hypergeometrischen Verteilung bekommen, versuchen Sie es mit der Binomialverteilung. Warum ist dies hier erlaubt?

Lösung:
Eigentlich hypergeometrisch

$\Pr(x = a) = \frac{\binom{A}{a} \cdot \binom{N-A}{n-a}}{\binom{N}{n}}$ mit N = 2 Mio, A = 800.000, N - A = 1,2 Mio, n = 40, a = 16, n - a = 24

$$\Pr(x = 16) = \frac{\binom{800.000}{16} \cdot \binom{1{,}2\ Mio}{24}}{\binom{2\ Mio}{40}} = 0{,}1279$$

Aber für diese großen Zahlen lassen sich schwer Binomialkoeffizienten ausrechnen

Daher alternativ Binomialverteilung verwenden, das diese bei großen Grundgesamtheiten und kleinen Stichproben sehr ähnlich verteilt sind.

$Pr(x = a) = \binom{n}{a} \cdot \pi^a \cdot (1 - \pi)^{n-a}$
x … Anzahl der Erhebungseinheiten mit der Eigenschaft unter den 40 gezogenen Erhebungseinheiten
n = 40
π = 0,4

$$\Pr(x = 16) = \binom{40}{16} \cdot 0{,}4^{16} \cdot 0{,}6^{24} = 0{,}128$$

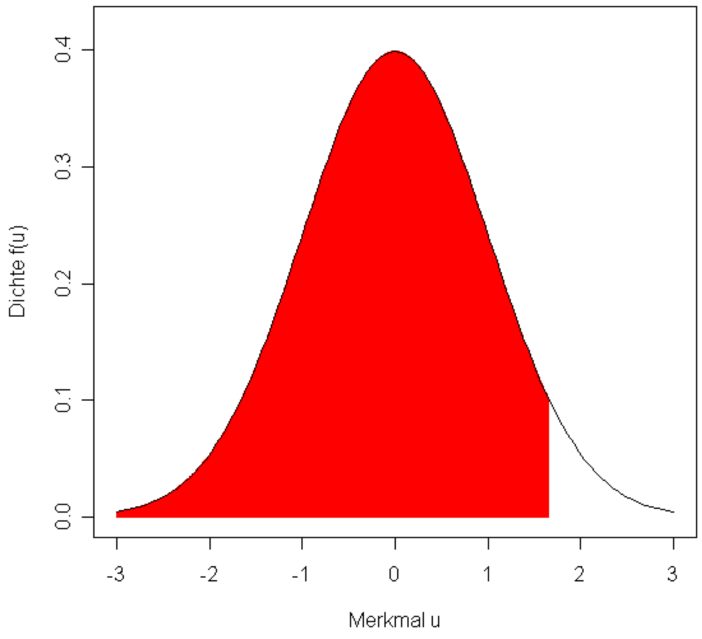
Dies ist erlaubt - obwohl ohne Zurücklegen gezogen wird - da bei großen Grundgesamtheiten und vergleichsweise kleinen Stichproben der Unterschied zwischen dem Ziehen ohne und dem Ziehen mit Zurücklegen praktisch verschwindet. Daher sind die Lösungen der (falschen) Binomialverteilung jenen der (eigentlich richtigen) Hypergeometrischen Verteilung sicherlich sehr ähnlich.

Aufgabe 44 [Normalverteilung]

Ein Merkmal sei standardnormalverteilt. Wie wahrscheinlich ist es, dass ein Messwert von

a) höchstens 1,65,
b) höchstens 1,96,
c) mindestens 1,65,
d) größer als - 1,65,
e) kleiner als -1,96 auftritt?

Lösung:
a)

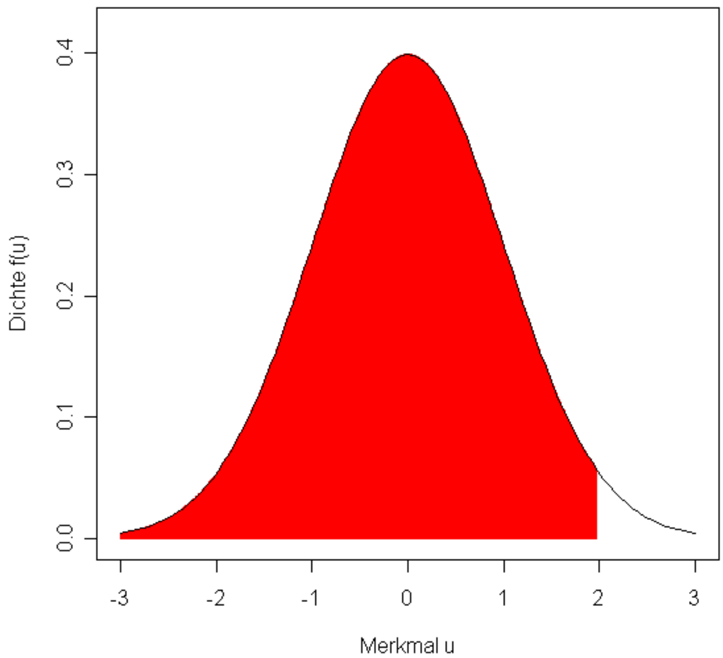


→ Pr(u ≤ 1,65) = 0,9505

Interpretation:

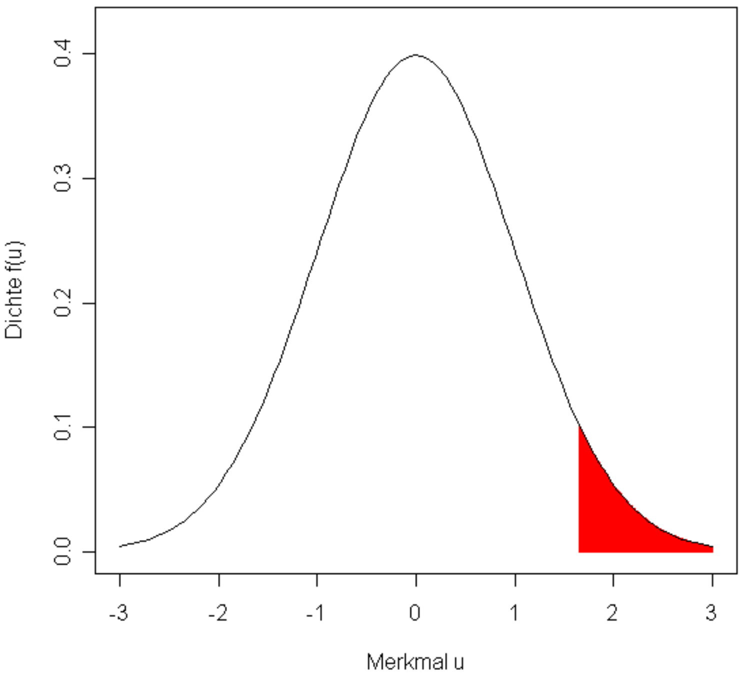
Bei 100 Versuchen wird dies *durchschnittlich* in ca. 95 Versuchen passieren.

b)



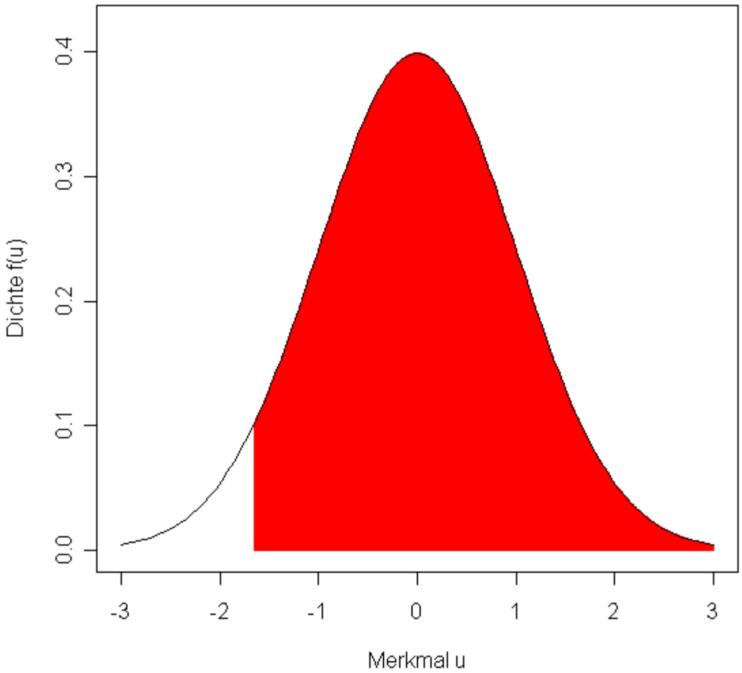
→ Pr(u ≤ 1,96) = 0,9750

c)



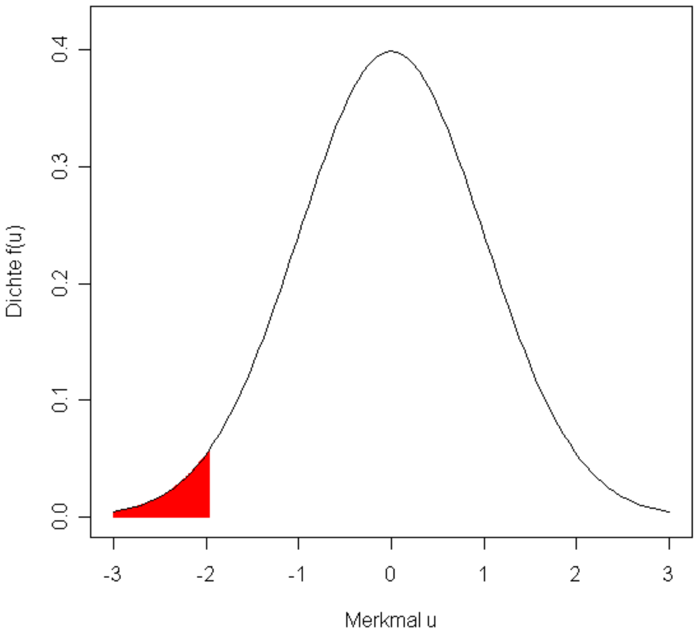
→ $Pr(u \geq 1{,}65) = 1 - Pr(u < 1{,}65) = 1 - 0{,}9505 = 0{,}0495$

d)



→ Pr(u > -1,65) = Pr(u < +1,65) = 0,9505

e)



→ Pr(u < -1,96) = Pr(u > +1,96) = 1 - Pr(u ≤ +1,96) = 1 - 0,9750 = 0,0250

Aufgabe 45 [Normalverteilung]

Die Länge von Schrauben einer Produktion ist normalverteilt mit Erwartungswert μ = 6 cm und theoretische Varianz σ^2 = 0,01 (cm²). Wie wahrscheinlich ist es, dass eine Schraube

a) höchstens 6,15 cm,
b) höchstens 6,196 cm lang ist?
c) mindestens 5,85 cm,
d) höchstens 5,83 cm lang,
e) mindestens 5,9 cm und höchstens 6,1 cm lang ist?

Lösung :

a)
Standardisierung: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{6{,}15-6}{0{,}1} = 1{,}5$
$Pr(x \leq 6{,}15) = Pr(u \leq 1{,}5) = 0{,}9332$

b)
Standardisierung: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{6{,}196-6}{0{,}1} = 1{,}96$
$Pr(x \leq 6{,}196) = Pr(u \leq 1{,}96) = 0{,}9750$

c)
Standardisierung: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{5{,}85-6}{0{,}1} = -1{,}5$
$Pr(x \geq 5{,}85) = Pr(u \geq -1{,}5) = Pr(u \leq +1{,}5) = 0{,}9332$

d)
Standardisierung: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{5{,}83-6}{0{,}1} = -1{,}7$
$Pr(x \leq 5{,}83) = Pr(u \leq -1{,}7) = Pr(u \geq 1{,}7) = 1 - Pr(u \leq 1{,}7) = 1 - 0{,}9554 = 0{,}0446$

e)
Standardisierung: $u_u = \frac{x_u-\mu}{\sigma} = \frac{5{,}9-6}{0{,}1} = -1;\ u_o = \frac{x_o-\mu}{\sigma} = \frac{6{,}1-6}{0{,}1} = 1$

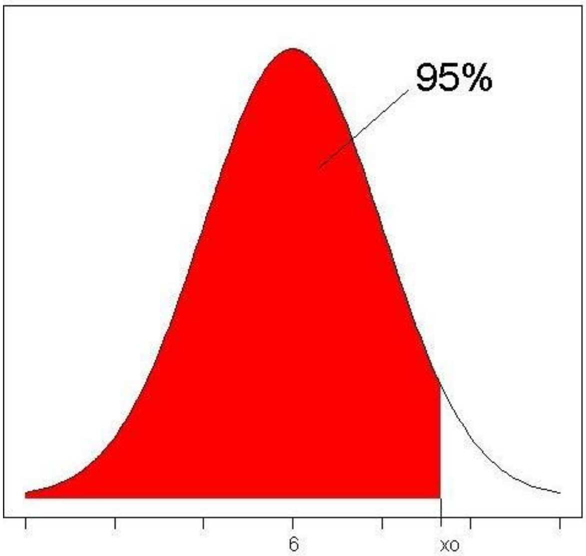
$$\begin{aligned} Pr(5{,}9 \leq x \leq 6{,}1) &= Pr(-1 \leq u \leq 1) = Pr(u \leq 1) - Pr(u \leq -1) \\ &= Pr(u \leq 1) - (1 - (Pr(u \leq 1)) = 0{,}8413 - (1 - 0{,}8413) = 0{,}6826 \end{aligned}$$

Interpretation:
Circa 2/3 der Schrauben liegen im Intervall -1 bis +1 Standardabweichung um den Mittelwert herum.

Aufgabe 46 [Normalverteilung]

Fortsetzung von Aufgabe 45: Berechnen Sie nun jene Länge der Schrauben, die mit einer Wahrscheinlichkeit von 0,95 unterschritten wird.

Lösung :



1. Ablesen von u_0 aus Standardnormalverteilung
$Pr(u \leq u_o) = 0{,}95 \rightarrow u_o = 1{,}65$ (eigentlich zwischen 1,64 und 1,65, aber man nimmt immer den größeren Wert, da bei 1,64 die Wahrscheinlichkeit eben kleiner 95% wäre)

2. Destandardisieren $\rightarrow x_0$
Standardisierungsformel: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{x_0-6}{0,1}$ umstellen nach x_0

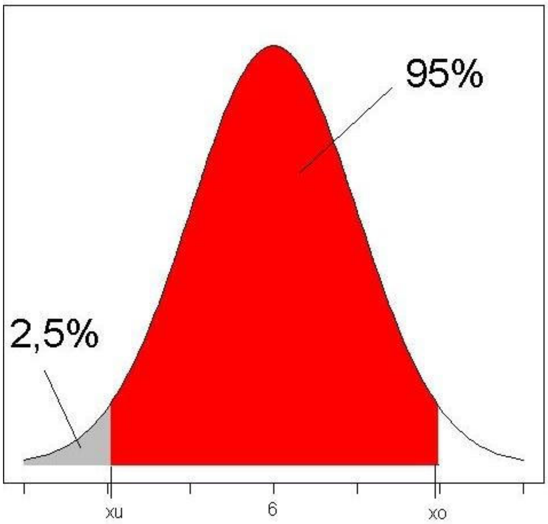
$x_o = 1{,}65 \cdot 0{,}1 + 6 = 6{,}165$

Interpretation:
95% der Schrauben sind 6,165 cm oder kürzer (diese Länge von 6,165 cm wird von 95% der Schrauben unterschritten)

Aufgabe 47 [Normalverteilung]

Fortsetzung von Aufgabe 45 bis Aufgabe 46: Berechnen Sie das symmetrische Intervall, in welchem die Länge der Schrauben mit einer Wahrscheinlichkeit von 0,95 liegt.

Lösung:



1. Ablesen von u_0 aus Standardnormalverteilung
$Pr(u \leq u_o) = 0{,}975$ (rote und graue Fläche) → $u_o = 1{,}96$
1,96 ist der Wert des standardnormalverteilten Merkmals bei dem die Wahrscheinlichkeit, dass dieser Wert unterschritten wird, bei 97,5% liegt. Ablesen aus Standardnormalverteilung den Wert bei dem die Wahrscheinlichkeit 0,975 das erste Mal erreicht oder überschreitet.

2. Destandardisieren → x_0
Standardisierungsformel: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{x_0-6}{0{,}1}$ umstellen nach x_0

$x_o = 1{,}96 \cdot 0{,}1 + 6 = 6{,}196$

3. Verteilung symmetrisch um den Mittelwert → x_u
Differenz zwischen Mittelwert und oberer Grenze = Differenz zwischen Mittelwert und unterer Grenze → Also wenn man die Differenz zwischen Mittelwert und obere Grenze kennt, kann man die untere Grenze ausrechnen
$x_o - \mu = 6{,}196 - 6 = 0{,}196$
$\mu - x_u = 6 - x_u = 0{,}196$ → $x_u = 5{,}804$

oder kürzer und zusammengefasst: $x_u = \mu - (x_o - \mu) = 6 - (6{,}196 - 6) = 5{,}804$
Interpretation:
Das Intervall [5,804; 6,196] ist jenes symmetrisch um den Mittelwert angeordnete Intervall der Länge der Schrauben, in dem 95% der Schrauben sich befinden.

Aufgabe 48 [Normalverteilung]

Die Länge von Drehbleistiftminen (in cm) verteilt sich annähernd normal mit Erwartungswert μ =5 und theoretische Varianz $\sigma^2 = 0{,}25$.
Wie wahrscheinlich ist es dann, dass eine solche Mine

a) kürzer als 5,2 cm,
b) länger als 5,98 cm ist?

Lösung:
a)
Standardisierung: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{5,2-5}{0,5} = 0,4$
Pr(x < 5,2) = Pr(u < 0,4) = Pr(u ≤ 0,4) = 0,655
Zur Erinnerung : Pr(u < 0,4) = Pr(u ≤ 0,4), da Pr (u = 0,4) = 0

Interpretation:
Die Wahrscheinlichkeit dass eine Mine kürzer als 5,2 cm ist beträgt 65,5%.

b)
Standardisierung: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{5,98-5}{0,5} = 1,96$
Pr(x > 5,98) = Pr(u > 1,96) = 1 - Pr(u ≤ 1,96) = 1 - 0,975 = 0,025
Interpretation:
Die Wahrscheinlichkeit dass eine Mine länger als 5,98 cm ist beträgt 2,5%.

Aufgabe 49 [Normalverteilung]

Fortsetzung von Aufgabe 48: Wie wahrscheinlich ist es, dass eine solche Mine zwischen 4,75 und 5,5 cm lang ist?

Lösung:
Achtung: Intervall nicht symmetrisch um den Mittelwert 5 cm
Standardisierungen:

$$u_o = \frac{x_o - \mu}{\sigma} = \frac{5,5 - 5}{0,5} = 1$$

$$u_u = \frac{x_u - \mu}{\sigma} = \frac{4,75 - 5}{0,5} = -0,5$$

Pr(4,75 ≤ x ≤ 5,5) = Pr(x ≤ 5,5) - Pr(x ≤ 4,75) = Pr(u ≤ 1) - Pr(u ≤ -0,5) = Pr(u ≤ 1) - Pr(u ≥ 0,5) = Pr(u ≤ 1) - [1 - Pr(u ≤ 0,5)] = 0,8413 - [1- 0,6915] = 0,5328

Aufgabe 50 [Normalverteilung]

Die Psychologen Stanford, Binet und Wechsler haben festgestellt, dass die Intelligenz - definiert als geistige Leistungsfähigkeit bei der Lösung von Testaufgaben - unter der Bevölkerung normalverteilt ist. Für Mittelwert und Standardabweichung des Intelligenzquotienten (= IQ) gilt:
μ = 100 und σ = 15.
In welchem symmetrischen Intervall liegt der IQ einer zufällig ausgewählten Testperson mit einer Wahrscheinlichkeit von

a) von 0,95,
b) von 0,5?

Lösung:
a)

1. Ablesen von u_0 aus Standardnormalverteilung
Pr(u ≤ u_o) = 0,975 → u_o = 1,96
1,96 ist der Wert des standardnormalverteilten Merkmals bei dem die Wahrscheinlichkeit, dass dieser Wert unterschritten wird, bei 97,5% liegt. Ablesen aus Standardnormalverteilung den Wert bei dem die Wahrscheinlichkeit 0,975 das erste Mal erreicht oder überschreitet.

2. Destandardisieren → x_0

Standardisierungsformel: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{x_0-100}{15}$ umstellen nach x_0

$x_o = 1{,}96 \cdot 15 + 100 = 129{,}4$

3. Verteilung symmetrisch um den Mittelwert → x_u

Differenz zwischen Mittelwert und oberer Grenze = Differenz zwischen Mittelwert und unterer Grenze → Also wenn man die Differenz zwischen Mittelwert und obere Grenze kennt, kann man die untere Grenze ausrechnen

$x_o - \mu = 129{,}4 - 100 = 29{,}4$

$\mu - x_u = 100 - x_u = 29{,}4 \rightarrow x_u = 70{,}6$

oder kürzer und zusammengefasst: $x_u = \mu - (x_o - \mu) = 100 - (129{,}4 - 100) = 70{,}6$

Interpretation:

Das Intervall [70,6; 129,4] ist jenes symmetrisch um den Mittelwert angeordnete Intervall, in dem der IQ von einer ausgewählten Testperson mit 95% Wahrscheinlichkeit liegt.

b)

1. Ablesen von u_0 aus Standardnormalverteilung

$Pr(u \leq u_o) = 0{,}75 \rightarrow u_o = 0{,}68$

0,68 ist der Wert des standardnormalverteilten Merkmals bei dem die Wahrscheinlichkeit, dass dieser Wert unterschritten wird, bei 75,0% liegt. Ablesen aus Standardnormalverteilung den Wert bei dem die Wahrscheinlichkeit 0,750 das erste Mal erreicht oder überschreitet.

2. Destandardisieren → x_0

Standardisierungsformel: $u_0 = \frac{x_0-\mu}{\sigma} = \frac{x_0-100}{15}$ umstellen nach x_0

$x_o = 0{,}68 \cdot 15 + 100 = 110{,}2$

3. Verteilung symmetrisch um den Mittelwert → x_u

Differenz zwischen Mittelwert und oberer Grenze = Differenz zwischen Mittelwert und unterer Grenze → Also wenn man die Differenz zwischen Mittelwert und obere Grenze kennt, kann man die untere Grenze ausrechnen

$x_o - \mu = 110{,}2 - 100 = 10{,}2$

$\mu - x_u = 100 - x_u = 10{,}2 \rightarrow x_u = 89{,}8$

oder kürzer und zusammengefasst: $x_u = \mu - (x_o - \mu) = 100 - (110{,}2 - 100) = 89{,}8$

Interpretation:

Das Intervall [89,8; 110,5] ist jenes symmetrisch um den Mittelwert angeordnete Intervall, in dem der IQ einer ausgewählten Testperson mit 50% Wahrscheinlichkeit liegt.

Im Vergleich zur Aufgabe a) sieht man, dass wenn man eine größere Wahrscheinlichkeit/Sicherheit haben will, das Intervall größer ist. Im Extremfall könnte man sagen, dass mit 100% Wahrscheinlichkeit der IQ einer Testperson zwischen 0 und dem Maximalwert der Bevölkerung (sagen wir mal 200) liegen würde.

Kapitel 3 Schließende Statistik

Aufgabe 51* [Die Handlungslogik der schließenden Statistik]

Im Internet finden Sie in der Excel-Datei zum Buch eine Grundgesamtheit mit allen 2.500 an einer Universität inskribierten Studentinnen und eine Anleitung zur Ziehung einer uneingeschränkten Zufallsauswahl. Die einzelnen Erhebungseinheiten sind durch eine Identifikationsnummer charakterisiert (und sie wären darüber kontaktierbar). Sie sollen aus dieser Grundgesamtheit eine Zufallsstichprobe vom Umfang n= 250 ziehen, in die jede dieser Studentinnen mit gleicher Wahrscheinlichkeit gelangen soll.

Lösung:
Verwendung der Anleitung in der EXCEL-Lerndatei und Ziehung der Stichprobe durch die Funktion „ZUFALLSZAHL()“

Aufgabe 52* [Die Handlungslogik der schließenden Statistik]

Berechnen Sie in Ihrer in Aufgabe 51 gezogenen Stichprobe vom Umfang n = 250 den Punktschätzer für

a) die relative Häufigkeit der Studentinnen, die noch kinderlos sind,
b) den Mittelwert der Kinderzahl,
c) den Korrelationskoeffizienten zur Messung des statistischen Zusammenhangs zwischen den beiden Merkmalen Kinderzahl und Anzahl belegter Kurse und vergleichen Sie die drei Stichprobenergebnisse mit den Parametern. Letztere lassen sich aus der ursprünglichen Grundgesamtheit von 2.500 Studentinnen aus Aufgabe 51 errechnen.

Lösung:

Hier können keine richtigen Lösungen angegeben werden, da jeder Anwender eine eigene Stichprobe zieht und sich die Stichprobenergebnisse p, $\bar{x}$ und r für die Aufgaben a) bis c) dementsprechend von Anwender(in) zu Anwender(in) unterscheiden werden. Sie sollten aber alle nahe zu den Parametern liegen:
Die relative Häufigkeit π in der Grundgesamtheit beträgt 0,6, der Mittelwert μ 0,54 und der Korrelationskoeffizient ρ ist -0,7715.
Sollten mehrere Personen diese durchführen und ihre Ergebnisse vergleichen, so erfährt man – vorausgesetzt, dass die Stichprobenziehung und die Berechnungen von allen korrekt durchgeführt wurden – was darunter zu verstehen ist, wenn man sagt, dass Stichprobenergebnisse schwanken.

Aufgabe 53* [Die Handlungslogik der schließenden Statistik]

Berechnen Sie in Ihrer in Aufgabe 51 gezogenen Stichprobe den Punktschätzer für die relative Häufigkeit der Studentinnen, die noch kinderlos sind. Verwenden Sie dazu jedoch nur die antwortenden Studentinnen, die Sie daran erkennen, dass bei Ihnen in der Spalte "Responseverhalten" eine " 1" steht. Vergleichen Sie die relative Häufigkeit an Kinderlosen unter den Antwortenden dieser Stichprobe mit der relativen Häufigkeit unter allen 2.500 Studentinnen der Grundgesamtheit aus Aufgabe 51 und mit jener unter den 250 Studentinnen aus Aufgabe 52.

Lösung:

Die relative Häufigkeit an Kinderlosen sollte deutlich unter p aus Ü52 und auch unter dem diesbezüglichen Parameter π liegen. Dies passiert also, wenn man zu viele Nichtantwortende hat. Trotz der Ziehung einer uneingeschränkten Zufallsstichprobe nach dem Urnenmodell ist die Stichprobe verzerrt, weil der

Untersuchungsgegenstand mit dem Antwortverhalten zusammenhängt. Es muss durch verschiedene Methoden dafür gesorgt werden, dass die Erhebungseinheiten, die ausgewählt werden, auch tatsächlich Auskunft geben. Mit diesen Methoden (Motivationsschreiben, Interviewerverhalten, Datenerhebungstechnik, …) beschäftigt sich die empirische Sozialforschung. Die Qualität der Ergebnisse von Stichproben hängt also – wie gesehen – stark von der vorhandenen Datenqualität ab.

Aufgabe 54 [Schätzen einer relativen Häufigkeit]

Bei der Abstimmung über den EU-Beitritt Österreichs stimmten im Jahr 1994 66,6 Prozent der Bevölkerung dafür. In welchem Bereich wäre an diesem Tag mit einer Wahrscheinlichkeit von 0,95 ein diesbezüglicher Anteil in einer zufällig aus dieser Grundgesamtheit gezogenen uneingeschränkten Zufallsauswahl gelegen, wenn man dazu

a) 100,
b) 400,
c) 800

Personen ausgewählt hätte? Was ist hier gegeben: die relative Häufigkeit in der Grundgesamtheit oder jene in der Stichprobe?

Lösung:

π … relative Häufigkeit in der Grundgesamtheit (gegeben)
p … relative Häufigkeit in der Stichprobe
π gegeben, Frage bezieht sich auf p: π = 0,666, α = 0,05 → $u_{1-\alpha/2} = 1{,}96$

a) n = 100
Berechnung mit (15a,b):

$$p_o = \pi + u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}666 + 1{,}96 \cdot \sqrt{\frac{0{,}666 \cdot (1-0{,}666)}{100}} = 0{,}758$$

$$p_u = \pi - u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}666 - 1{,}96 \cdot \sqrt{\frac{0{,}666 \cdot (1-0{,}666)}{100}} = 0{,}574$$

Interpretation:
Wäre an diesem Tag eine zufällige Stichprobe von n = 100 Personen aus der Grundgesamtheit gezogen worden, dann wäre der Anteil der EU-Beitrittsbefürworter in der Stichprobe mit einer Wahrscheinlichkeit von 0,95 im Intervall [0,574; 0,758] gelegen.

b) n = 400
Berechnung:

$$p_o = \pi + u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}666 + 1{,}96 \cdot \sqrt{\frac{0{,}666 \cdot (1-0{,}666)}{400}} = 0{,}712$$

$$p_u = \pi - u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}666 - 1{,}96 \cdot \sqrt{\frac{0{,}666 \cdot (1-0{,}666)}{400}} = 0{,}620$$

c) n = 800
Berechnung:

$$p_o = \pi + u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}666 + 1{,}96 \cdot \sqrt{\frac{0{,}666 \cdot (1-0{,}666)}{800}} = 0{,}699$$

$$p_u = \pi - u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}666 - 1{,}96 \cdot \sqrt{\frac{0{,}666 \cdot (1-0{,}666)}{800}} = 0{,}633$$

Aufgabe 55* [Schätzen einer relativen Häufigkeit]

Eine Woche vor der letzten Wahl wurden zufällig ausgewählte wahlberechtigte Staatsbürgerinnen und Staatsbürger gefragt, welche Partei sie wählen würden. Die kodierten Antworten finden Sie in einer zu diesem Beispiel im Internet bereitstehenden Excel-Datei. Bestimmen Sie die relative Häufigkeit der Partei A in der Stichprobe mit Hilfe der dafür vorgesehenen Excel-Funktion (siehe Aufgabe 4) und berechnen Sie davon ausgehend mit dem Taschenrechner das Konfidenzintervall zur Sicherheit 1 - α = 0,95 für die relative Häufigkeit an Wählern dieser Partei in der Gesamtbevölkerung.

Lösung:

π … relative Häufigkeit in der Grundgesamtheit
p … relative Häufigkeit in der Stichprobe (gegeben)
p gegeben, Frage bezieht sich auf π: n = 400, p = 0,4 (aus den Daten), α = 0,05 → $u_{1-\alpha/2} = 1,96$

Berechnung mit (16):

$$\pi_o = p + u_{1-\alpha/2} \cdot \sqrt{\frac{p\cdot(1-p)}{n}} = 0,4 + 1,96 \cdot \sqrt{\frac{0,4\cdot(1-0,4)}{400}} = 0,448$$

$$\pi_u = p + u_{1-\alpha/2} \cdot \sqrt{\frac{p\cdot(1-p)}{n}} = 0,4 - 1,96 \cdot \sqrt{\frac{0,4\cdot(1-0,4)}{400}} = 0,352$$

Interpretation:
Mit 95%iger Sicherheit wird die relative Häufigkeit π (der Wähler von Partei A in der Grundgesamtheit) vom Intervall [0,352; 0,448] überdeckt.

Aufgabe 56 [Schätzen einer relativen Häufigkeit]

Ein Meinungsforschungsinstitut veröffentlichte das Umfrageergebnis einer Zufallsstichprobe von *n* = 500 Personen aus der Grundgesamtheit aller Wahlberechtigten. Es gaben 80 Prozent der Befragten an, gegen den Einkauf von neuen Flugzeugen für die Landesverteidigung zu sein. Was ist hier gegeben: die relative Häufigkeit in der Grundgesamtheit oder jene in der Stichprobe? Berechnen Sie

a) das Konfidenzintervall zur Sicherheit 1 - *α* = 0,95,
b) das Konfidenzintervall zur Sicherheit 1 - *α = 0,9,*
c) das Konfidenzintervall zur Sicherheit 1 - *α* = 0,99

für diese relative Häufigkeit in der Gesamtheit aller Wahlberechtigten.

Lösung:

p gegeben, Frage bezieht sich auf π (so auch in den nachfolgenden Beispielen): n = 500, p = 0,8

a)
α = 0,05 → $u_{1-\alpha/2} = 1,96$

Berechnung:

$$\pi_o = p + u_{1-\alpha/2} \cdot \sqrt{\frac{p\cdot(1-p)}{n}} = 0,8 + 1,96 \cdot \sqrt{\frac{0,8\cdot(1-0,8)}{500}} = 0,835$$

$$\pi_u = p - u_{1-\alpha/2} \cdot \sqrt{\frac{p\cdot(1-p)}{n}} = 0,8 - 1,96 \cdot \sqrt{\frac{0,8\cdot(1-0,8)}{500}} = 0,765$$

Interpretation:
Mit 95%iger Sicherheit wird die relative Häufigkeit π (der Gegner des Einkaufs neuer Flugzeuge in der Grundgesamtheit) vom Intervall [0,765; 0,835] überdeckt.

b)
α = 0,1 → $u_{1-\alpha/2}$ = 1,65

Berechnung:

$$\pi_o = p + u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0,8 + 1,65 \cdot \sqrt{\frac{0,8 \cdot (1-0,8)}{500}} = 0,829$$

$$\pi_u = p - u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0,8 + 1,65 \cdot \sqrt{\frac{0,8 \cdot (1-0,8)}{500}} = 0,771$$

c)
α = 0,01 → $u_{1-\alpha/2}$ = 2,58

Berechnung:

$$\pi_o = p + u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0,8 + 2,58 \cdot \sqrt{\frac{0,8 \cdot (1-0,8)}{500}} = 0,846$$

$$\pi_u = p - u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0,8 - 2,58 \cdot \sqrt{\frac{0,8 \cdot (1-0,8)}{500}} = 0,754$$

Höhere Sicherheit = niedrigere Genauigkeit (und umgekehrt).

Aufgabe 57 [Schätzen einer relativen Häufigkeit]

Eine Zufallsstichprobe von n = 300 Werkstücken aus einer Produktion ergab einen Ausschussanteil p = 0,14. Berechnen Sie daraus das Konfidenzintervall zur Sicherheit 1 - α = 0,95 für den Ausschussanteil π in der gesamten Produktion.

Lösung:

n = 300, p = 0,14, α = 0,05 → $u_{1-\alpha/2}$ = 1,96

Berechnung:

$$\pi_o = p + u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0,14 + 1,96 \cdot \sqrt{\frac{0,14 \cdot (1-0,14)}{300}} = 0,179$$

$$\pi_u = p - u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0,14 - 1,96 \cdot \sqrt{\frac{0,14 \cdot (1-0,14)}{300}} = 0,101$$

Aufgabe 58 [Schätzen einer relativen Häufigkeit]

In einer Meinungsumfrage soll festgestellt werden, wie hoch der derzeitige Stimmenanteil einer bestimmten Partei wäre. Wie viele Wahlberechtigte sind in einer uneingeschränkten Zufallsauswahl aus der Grundgesamtheit mindestens zu befragen, wenn das Konfidenzintervall zur Sicherheit 1 - α = 0,95 eine Schwankungsbreite ε von 0,03 besitzen soll und man davon ausgehen kann, dass diese relative Häufigkeit

a) bei etwa 42 Prozent,
b) zwischen 5 und 15 Prozent,
c) bei mindestens 30 Prozent liegen wird?

Lösung:

$$\begin{matrix}\pi_o\\ \pi_u\end{matrix} = p \pm u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Schwankungsbreite um die relative Häufigkeit herum ist $\varepsilon = u_{1-\alpha/2} \cdot \sqrt{\frac{p\cdot(1-p)}{n}}$

ε = 0,03

$1-\alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow 1-\alpha/2 = 0{,}975 \rightarrow u_{1-\alpha/2}$ (durch Ablesen in Standardnormalverteilung) $= 1{,}96$

a) Für π: 0,42.

Berechnung:

$$\varepsilon = u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \; \Big| \; Quadrat \rightarrow \varepsilon^2 = u^2_{1-\frac{\alpha}{2}} \cdot \frac{p \cdot (1-p)}{n} \; \Big| : \varepsilon^2 \;\; | : n$$

$$n = \frac{u^2_{1-\alpha/2}}{\varepsilon^2} \cdot p \cdot (1-p) = \frac{1{,}96^2}{0{,}03^2} \cdot 0{,}42 \cdot (1-0{,}42) = 1.040$$

(*bei dieser Berechnung immer aufrunden, damit genügend Personen in der Stichprobe sind!*)

Beweis, dass bei dieser Stichprobengröße die Schwankungsbreite bei 0,03 liegt:

$$\begin{matrix}\pi_o\\ \pi_u\end{matrix} = p \pm u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = 0{,}42 \pm 1{,}96 \cdot \sqrt{\frac{0{,}42 \cdot (1-0{,}42)}{1.040}} = 0{,}42 \pm 1{,}96 \cdot \sqrt{\frac{0{,}42 \cdot (1-0{,}42)}{1.040}}$$
$$= 0{,}42 \pm 0{,}03$$

b) Für π: jenen Wert zwischen 0,05 und 0,15, der 0,5 am Nächsten liegt. Das ist 0,15. Grund ist die Schwankungsbreite nimmt zu je näher man an die relative Häufigkeit von 0,5 kommt. Dort wäre das Maximum.

Berechnung:

$$n = \frac{u^2_{1-\alpha/2}}{\varepsilon^2} \cdot p \cdot (1-p) = \frac{1{,}96^2}{0{,}03^2} \cdot 0{,}15 \cdot (1-0{,}15) = 545$$

c) Für π: jenen Wert zwischen 0,3 und 1, der 0,5 am Nächsten liegt. Das ist 0,5 selbst! Denn bei 0,5 ist die Schwankungsbreite am größten!

Berechnung:

$$n = \frac{u^2_{1-\alpha/2}}{\varepsilon^2} \cdot p \cdot (1-p) = \frac{1{,}96^2}{0{,}03^2} \cdot 0{,}5 \cdot (1-0{,}5) = 1.068$$

Aufgabe 59 [Schätzen einer relativen Häufigkeit]

Die relative Häufigkeit des Auftretens einer bestimmten Eigenschaft in der Gesamtbevölkerung soll in einer Stichprobe geschätzt werden. Welcher Stichprobenumfang ist zu wählen, wenn keinerlei Abschätzung des tatsächlichen Anteils existiert und das Konfidenzintervall zur Sicherheit $1 - \alpha = 0{,}95$ eine von Ihnen festzulegende Schwankungsbreite aufweisen soll?

Lösung:

$$\frac{\pi_o}{\pi_u} = p \pm u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Schwankungsbreite um die relative Häufigkeit herum ist $\varepsilon = u_{1-\alpha/2} \cdot \sqrt{\frac{p\cdot(1-p)}{n}}$

$\varepsilon = 0{,}02$

$1 - \alpha = 0{,}95 \rightarrow \alpha = 0{,}05 \rightarrow 1 - \alpha/2 = 0{,}975 \rightarrow u_{1-\alpha/2}$ (durch Ablesen in Standardnormalverteilung) $= 1{,}96$

$\pi = 0{,}5$ da keine Infos vorliegen wie hoch die relative Häufigkeit sein könnte, muss man vom schlimmsten denkbaren Fall ausgehen. Die Schwankungsbreite ist immer am größten bei $\pi = 0{,}5$

Berechnung:

$$\varepsilon = u_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \;\Big|\; Quadrat \rightarrow \varepsilon^2 = u^2_{1-\frac{\alpha}{2}} \cdot \frac{p \cdot (1-p)}{n} \;\Big|: \varepsilon^2 \;\; | : n$$

$$n = \frac{u^2_{1-\alpha/2}}{\varepsilon^2} \cdot p \cdot (1-p) = \frac{1{,}96^2}{0{,}02^2} \cdot 0{,}5 \cdot (1 - 0{,}5) = 2.401$$

Aufgabe 60 [Testen von Hypothesen über eine relative Häufigkeit]

Eine politische Partei will feststellen, ob sich die relative Häufigkeit an österreichischen EU-Beitritts-Befürwortern gegenüber dem Ergebnis bei der Volksabstimmung im Jahr 1994 inzwischen verändert hat. Bei der Volksabstimmung hatten 66,6 Prozent zugestimmt.

a) Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
b) In einer aktuellen Umfrage unter $n = 800$ zufällig ausgewählten Wahlberechtigten befürworten die EU 43,6%. Entscheiden Sie sich auf Basis dieses Stichprobenergebnisses auf einem Signifikanzniveau $\alpha = 0{,}05$ für eine der beiden Hypothesen.
c) Wie würde Ihre Entscheidung bei $p = 0{,}536$ lauten?
d) Wie würde Ihre Entscheidung bei $p = 0{,}636$ lauten?

Lösung:

a)

π … Anteil an EU-Beitritts-Befürwortern in der Grundgesamtheit

Die Regierung will herausfinden ob sich der Anteil der EU-Befürworter gegenüber den 66,6% verändert hat → zweiseitiger Test. Es ist etwas anders also ≠ wie vorher. Was einen interessiert ist immer die Einshypothese. Die Nullhypothese ist immer das Gegenteil von dem man zuerst ausgeht → der Anteil der EU-Befürworter hat sich nicht verändert, ist also gleich geblieben.

$H_1: \pi \neq 0{,}666$

$H_0: \pi = 0{,}666$

b)
$n = 800$, $\alpha = 0,05 \rightarrow u_{1-\alpha/2} = 1,96$
$p = 349 : 800 = 0,436$

Bereich der schwachen Indizien gegen H_0:

$$p_o = \pi + u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0,666 + 1,96 \cdot \sqrt{\frac{0,666 \cdot (1-0,666)}{800}} = 0,699$$

$$p_u = \pi - u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0,666 - 1,96 \cdot \sqrt{\frac{0,666 \cdot (1-0,666)}{800}} = 0,633$$

Interpretation:
Da p = 0,436 nicht im Bereich [0,633; 0,699] der schwachen Indizien gegen die Nullhypothese enthalten ist, liegt ein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0,05$ verwerfen, Entscheidung für H_1. Relative Häufigkeit der EU-Befürworter ist signifikant anders als 66,6%.

c)
p = 0,536

Interpretation:
Da p = 0,536 nicht im Bereich [0,633; 0,699] der schwachen Indizien gegen die Nullhypothese enthalten ist, liegt ein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0,05$ verwerfen, Entscheidung für H_1. Relative Häufigkeit der EU-Befürworter ist signifikant anders als 66,6%.

d)
p = 0,636

Interpretation:
Da p = 0,636 Element des Bereichs [0,633; 0,699] der schwachen Indizien gegen die Nullhypothese ist, liegt nur ein schwaches Indiz gegen H_0 vor → H_0 wird auf dem Signifikanzniveau $\alpha = 0,05$ beibehalten. Relative Häufigkeit der EU-Befürworter ist NICHT signifikant anders als 66,6%.

Aufgabe 61 [Testen von Hypothesen über eine relative Häufigkeit]

Ein TV-Sender möchte feststellen, ob die Einschaltquote (= relative Häufigkeit der zugeschalteten Haushalte von allen Haushalten mit TV-Anschluss) einer TV-Show unter 10 Prozent gefallen ist.

a) Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
b) In einer Stichprobe unter n = 1.200 Haushalten hatten sich 102 Haushalte zugeschaltet. Entscheiden Sie sich auf Basis dieses Stichprobenergebnisses auf einem Signifikanzniveau $\alpha = 0,05$ für eine der beiden Hypothesen.
c) Wie würde Ihre Entscheidung bei $p = 0,09$ lauten?
d) Wie würde Ihre Entscheidung bei $p = 0,12$ lauten?

Lösung:

a)

π ... Einschaltquote in der Grundgesamtheit

Der Fernsehsender interessiert sich, ob die Einschaltquote geringer ist als 10%. Was einen interessiert ist immer die Einshypothese. Die Nullhypothese ist immer das Gegenteil davon, also in dem Fall dass die Einschaltquote mindestens 10% betragen würde.

H_1: $\pi < 0{,}1$
H_0: $\pi \geq 0{,}1$

b)

n = 1200, $\alpha = 0{,}05$ → $u_{1-\alpha} = 1{,}65$ (einseitiger Test; siehe Einshypothese)
p = 102 : 1.200 = 0,085

Berechnung:

Untere Schranke der schwachen Indizien gegen H_0:

$$p_u = \pi - u_{1-\alpha} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}1 - 1{,}65 \cdot \sqrt{\frac{0{,}1 \cdot (1-0{,}1)}{1.200}} = 0{,}086$$

Interpretation:

Da p = 0,085 kleiner als die untere Schranke der schwachen Indizien gegen H_0 ist, liegt ein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ verwerfen, Entscheidung für H_1. Die Einschaltquote ist signifikant geringer als 10%.

c)

p = 0,09

Interpretation:

Da p = 0,09 größer als die untere Schranke der schwachen Indizien gegen H_0 ist, liegt nur ein schwaches Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ beibehalten. Die Einschaltquote ist zwar geringer als 10%, aber nicht signifikant (also deutlich) geringer als 10%

d)

p = 0,12

Interpretation:

Da p = 0,12 größer ist als die untere Schranke der schwachen Indizien gegen H_0 ist, liegt natürlich kein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ beibehalten. Die Einschaltquote ist nicht signifikant geringer als 10%.

Man könnte sogar hier ganz auf die tatsächliche Berechnung der unteren Schranke verzichten. Die Formel sagt aus

$$p_u = \pi - u_{1-\alpha} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} \rightarrow p_u = \pi - \text{Sicherheitsabschlag}$$

$$p_u = 0{,}1 - \text{Sicherheitsabschlag} < 0{,}1$$

Dass von der Einschaltquote von 10% immer noch ein bestimmter Sicherheitsabstand abgezogen wird. Die untere Schranke ist deshalb in diesem Fall immer kleiner 10% ab wann man sagen könnte, dass die Einschaltquote deutlich unter 10% gefallen wäre. Jetzt ist die Einschaltquote in der Stichprobe mit 12% automatisch immer über dieser unteren Schranke p_u=10%-X. Als geschulter Statistiker würden sie daher die Nullhypothese immer ablehnen müssen. Wie sollte auch die Einschaltquote mit hoher Sicherheit unter 10% sein, wenn in der Stichprobe die Einschaltquote schon mit 12% deutlich höher als 10% ist?!!!

Aufgabe 62 [Testen von Hypothesen über eine relative Häufigkeit]

In einer Zufallsstichprobe vom Umfang $n = 350$ aus der wahlberechtigten Bevölkerung stellte man Ende des vergangenen Jahres fest, dass 192 der Befragten einem EU-Beitritt eines bestimmten Kandidatenlandes skeptisch gegenüberstanden. Konnte man aus dem Stichprobenergebnis auf dem Signifikanzniveau $\alpha = 0{,}05$ folgern, dass eine Mehrheit der Gesamtbevölkerung in dieser Frage skeptisch eingestellt war?

Lösung:

π … Anteil an Skeptikern in der Grundgesamtheit
$n = 350$, $p = 0{,}549$, $\alpha = 0{,}05$ → $u_{1-\alpha} = 1{,}65$ (einseitiger Test; siehe H_1)

H_0: $\pi \leq 0{,}5$
H_1: $\pi > 0{,}5$

Berechnung:
Obere Schranke der schwachen Indizien gegen H_0:

$$p_o = \pi + u_{1-\alpha} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}5 + 1{,}65 \cdot \sqrt{\frac{0{,}5 \cdot (1-0{,}5)}{350}} = 0{,}544$$

Interpretation:
Da $p = 0{,}549$ größer als die obere Schranke der schwachen Indizien gegen H_0 ist, liegt ein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ verwerfen, Entscheidung für H_1. Die relative Häufigkeit der Skeptiker ist signifikant größer als 50%.

Aufgabe 63 [Testen von Hypothesen über eine relative Häufigkeit]

Von Werkstücken, die ein Jahr lang gelagert wurden, sind 40 Prozent unbrauchbar. Nach einer Änderung der Lagerbedingungen wird überprüft, ob sich die relative Häufigkeit an unbrauchbaren Werkstücken verringert hat.

a) Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
b) Entscheiden Sie sich auf dem Signifikanzniveau $\alpha = 0{,}05$ für eine der beiden Hypothesen, wenn unter 100 zufällig ausgewählten Stücken nunmehr 36 Prozent unbrauchbar waren.

Lösung:
π … Anteil an unbrauchbaren Werkstücken in der Grundgesamtheit
$n = 100$, $p = 0{,}36$, $\alpha = 0{,}05$ → $u_{1-\alpha} = 1{,}65$ (einseitiger Test; siehe H_1)
a)
H_0: $\pi \geq 0{,}4$
H_1: $\pi < 0{,}4$
b)
Berechnung:
Untere Schranke der schwachen Indizien gegen H_0:

$$p_u = \pi - u_{1-\alpha} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n}} = 0{,}4 - 1{,}65 \cdot \sqrt{\frac{0{,}4 \cdot (1-0{,}4)}{100}} = 0{,}319$$

Interpretation:
Da $p = 0{,}36$ größer als die untere Schranke der schwachen Indizien gegen H_0 ist, liegt nur ein schwaches Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ beibehalten. Die relative Häufigkeit von unbrauchbaren Werkstücken ist nicht signifikant kleiner als 40%.

Aufgabe 64 [Testen von Hypothesen über eine relative Häufigkeit]

Entscheiden Sie sich bei den Hypothesen in Aufgabe 60 bis Aufgabe 63 auf dem Signifikanzniveau $\alpha = 0{,}05$, wenn Sie erfahren, dass der zweiseitige p-Wert des Tests

a) 0,225,
b) 0,064,
c) 0,021 betragen hat.

Zur Erinnerung: Einseitige Tests wurden in Aufgabe 61-63 durchgeführt. Nur in Aufgabe 60 wurde ein zweiseitiger Test durchgeführt. Weitere Infos aus diesen Aufgaben benötigen Sie für die Aufgabe 64 nicht.

Lösung:

Aufgabe 60 (Befürwortung EU geändert): zweiseitiger Test, $\alpha = 0{,}05$
Aufgabe 61 (Einschaltquote < 10%): einseitiger Test $\alpha = 0{,}05$
Aufgabe 62 (Ablehnung EU Beitritt): einseitiger Test $\alpha = 0{,}05$
Aufgabe 63 (Anteil unbrauchbarer Werkzeuge verringert): einseitiger Test, $\alpha = 0{,}05$

Nullhypothese beibehalten, wenn gilt: p-Wert $\alpha_2 > \alpha$ (zweiseitiger Test) bzw. p-Wert $\alpha_1 > \alpha$ (einseitiger Test) mit der Beziehung $\alpha_1 = \alpha_2/2$

Bei Aufgabe 60 Vergleich p-Wert α_2 mit zugestandener Irrtumswahrscheinlichkeit α (2-seitiger Test)
Bei Aufgabe 61-63 Vergleich p-Wert α_1 mit zugestandener Irrtumswahrscheinlichkeit α (1-seitiger Test)

In der Ergebnistabelle ist dann aufgeführt für welche Hypothese H_0 oder H_1 man sich entscheidet.

Entscheidung:	A60	A61	A62	A63
a) $\alpha_2 = 0{,}225 \rightarrow$ $\alpha_1 = 0{,}1125$	H_0	H_0	H_0	H_0
b) $\alpha_2 = 0{,}064 \rightarrow$ $\alpha_1 = 0{,}032$	H_0	H_1	H_1	H_1
c) $\alpha_2 = 0{,}021 \rightarrow$ $\alpha_1 = 0{,}0195$	H_1	H_1	H_1	H_1

Aufgabe 65 [Schätzen eines Mittelwertes]

a) Zur Erprobung der Wirksamkeit eines Schlafmittels wurde bei allen 20.000 Angestellten (große Grundgesamtheit) eines sehr großen pharmazeutischen Unternehmens die Schlafdauer gemessen. Der Mittelwert betrug 6,36 Stunden mit einer Varianz von 1,561. Wenn man 200 Personen dieser Testgruppe per Zufallsstichprobe auswählen würde, in welchem Bereich wäre mit einer Wahrscheinlichkeit von 0,99 der Mittelwert der Schlafdauer für diese Personengruppe?
b) Die Ergebnisse der Tests mit allen Angestellten sollen nun auf die gesamte deutsche Bevölkerung hochgerechnet werden. Berechnen Sie für die Schlafdauer das Konfidenzintervall zur Sicherheit von $1 - \alpha = 0{,}99$. Achtung: Die 2.000 Angestellte sind nun als Stichprobe aus der Grundgesamtheit aller Deutschen zu behandeln.
c) Das Gesundheitsministerium ist unzufrieden mit der großen Schwankungsbreite des Konfidenzintervalls aus Aufgabe b). Wie hoch müsste die Stichprobe insgesamt sein um die Schwankungsbreite des Konfidenzintervalls um den Mittelwert herum auf 0,05 zu reduzieren bei sonst gleichbleibenden Daten (Mittelwert, Varianz, Sicherheitsniveau)

Lösung:

a)

Gegeben sind

$\mu = 6{,}36$ als Mittelwert der Grundgesamtheit (ALLE Angestellten)

$\sigma^2 = 1{,}561$ als Varianz der Schlafdauer

$1-\alpha = 0{,}99 \rightarrow \alpha = 0{,}01 \rightarrow 1-\alpha/2 = 0{,}995 \rightarrow u_{1-\alpha/2}$ (durch Ablesen in Standardnormalverteilung) $= 2{,}58$

$$\bar{x}_o = \mu + u_{1-\alpha/2} \cdot \sqrt{\frac{\sigma^2}{n}} = 6{,}36 + 2{,}58 \sqrt{\frac{1{,}561}{200}} = 6{,}36 + 2{,}58 \cdot \sqrt{\frac{1{,}561}{200}} = 6{,}36 + 0{,}228 = 6{,}588$$

$$\bar{x}_u = \mu - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} = 6{,}36 - 2{,}58 \cdot \sqrt{\frac{1{,}561}{200}} = 6{,}36 + 2{,}58 \cdot \sqrt{\frac{1{,}561}{200}} = 6{,}36 - 0{,}228 = 6{,}132$$

Interpretation:

Mit 99%-iger Sicherheit wird der Mittelwert des Merkmals Schlafdauer in der Stichprobe im Intervall vom Intervall [6,588; 6,132] liegen.

b)

$\bar{x} = 6{,}36$ als Mittelwert der Stichprobe der 2.000 Angestellten

$s^2 = 1{,}561$ als Varianz für die Schlafdauer in der Stichprobe

$1-\alpha = 0{,}99 \rightarrow \alpha = 0{,}01 \rightarrow 1-\alpha/2 = 0{,}995 \rightarrow u_{1-\alpha/2}$ (durch Ablesen in Standardnormalverteilung) $= 2{,}58$

$$\mu_o = \bar{x} + u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 6{,}36 + 2{,}58 \sqrt{\frac{1{,}561}{200}} = 6{,}36 + 2{,}58 \cdot \sqrt{\frac{1{,}561}{200}} = 6{,}36 + 0{,}228 = 6{,}588$$

$$\mu_u = \bar{x} - u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 6{,}36 - 2{,}58 \sqrt{\frac{1{,}561}{200}} = 6{,}36 - 2{,}58 \cdot \sqrt{\frac{1{,}561}{200}} = 6{,}36 - 0{,}228 = 6{,}132$$

Interpretation:

Mit 99%-iger Sicherheit wird der Mittelwert des Merkmals Schlafdauer in der Grundgesamtheit vom Intervall [6,588; 6,132] liegen.

Zu c)

$$\begin{matrix}\mu_o\\ \mu_u\end{matrix} = \bar{x} \pm u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}}$$

Schwankungsbreite um den Mittelwert herum ist $\varepsilon = u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}}$

$$\varepsilon = u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} \quad | \, Quadrat$$

$$\varepsilon^2 = u_{1-\alpha/2}^2 \cdot \frac{s^2}{n} \quad | \cdot n \quad |: \varepsilon^2$$

$$n = u_{1-\frac{\alpha}{2}}^2 \cdot \frac{s^2}{\varepsilon^2} = 2{,}58^2 \cdot \frac{1{,}561}{0{,}05^2} = 4.156{,}26 \approx 4.157$$

Interpretation:
Bei einer Stichprobengröße von 4.157 Personen (aufrunden) wäre die Schwankungsbreite des Konfidenzintervalls um den Mittelwert herum bei 0,05. Beweis:

$$\begin{matrix}\mu_o \\ \mu_u\end{matrix} = \bar{x} \pm u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 6{,}36 \pm 2{,}58 \cdot \sqrt{\frac{1{,}561}{4.157}} = 6{,}36 \pm 0{,}05$$

Aufgabe 66 [Schätzen eines Mittelwertes]

Bei der Abfüllung von Mineralwasser in Literflaschen wird der Magnesiumgehalt je Liter gemessen. n = 116 Kontrollmessungen ergaben folgende Werte für den Gehalt an Magnesium (in mg/l): $\bar{x} = 25{,}452$, s^2 = 0, 850. Berechnen Sie das Konfidenzintervall zur Sicherheit 1 – α = 0,95 für den mittleren Magnesiumgehalt in der Gesamtproduktion.

Lösung:

$\bar{x}$ ... Mittelwert in der Stichprobe
µ ... Mittelwert in der Grundgesamtheit
$\bar{x}$ gegeben, Auskunft über µ gesucht.

$\bar{x}$ = 25,452, s^2 = 0,850, n = 116, α = 0,05 → $u_{1-\alpha/2}$ = 1,96

Berechnung:

$$\mu_o = \bar{x} + u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 25{,}452 + 1{,}96 \cdot \sqrt{\frac{0{,}850}{116}} = 25{,}620$$

$$\mu_u = \bar{x} - u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 25{,}452 - 1{,}96 \cdot \sqrt{\frac{0{,}850}{116}} = 25{,}284$$

Interpretation:
Mit 95%-iger Sicherheit wird der Mittelwert des Merkmals Magnesiumgehalt in der Grundgesamtheit vom Intervall [25,284; 25,620] überdeckt.

Aufgabe 67 [Schätzen eines Mittelwertes]

Eine Zufallsstichprobe vom Umfang n = 836 ergab hinsichtlich eines Merkmals x einen Mittelwert x = 22,5 bei einer Standardabweichung von s = 3,2. Bestimmen Sie das Konfidenzintervall zur Sicherheit
1 – α = 0,95 für den wahren Mittelwert von *x.*

Lösung:

$\bar{x}$ ... Mittelwert in der Stichprobe
µ ... Mittelwert in der Grundgesamtheit
$\bar{x}$ gegeben, Auskunft über µ gesucht.

$\bar{x}$ = 22,5, s = 3,2 → s^2 = 10,24, n = 836, α = 0,05 → $u_{1-\alpha/2}$ = 1,96

Berechnung:

$$\mu_o = \bar{x} + u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 22{,}5 + 1{,}96 \cdot \sqrt{\frac{10{,}24}{836}} = 22{,}717$$

$$\mu_u = \bar{x} - u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 22,5 - 1,96 \cdot \sqrt{\frac{10,24}{836}} = 22,283$$

Interpretation:
Mit 95%-iger Sicherheit wird der Mittelwert des Merkmals Magnesiumgehalt in der Grundgesamtheit vom Intervall [22,283; 22,717] überdeckt.

Aufgabe 68* [Testen von Hypothesen über einen Mittelwert]

Eine Zufallsstichprobe vom Umfang n = 86 ergab hinsichtlich eines normalverteilten Merkmals x Messergebnisse, die in der im Internet bereitgestellten Excel-Datei zu finden sind. Testen Sie zum Signifikanzniveau $\alpha = 0{,}05$ folgende Hypothesen:

a) $H_0: \mu = 4.500$ gegen $H_1: \mu \neq 4.500$.
b) $H_0: \mu \geq 4.500$ gegen $H_1: \mu < 4.500$.

Lösung:
μ … Mittelwert des Merkmals in der Grundgesamtheit
$\bar{x} = 4.464{,}63$, $s^2 = 12.037{,}089$, $n = 86$

a)
$H_1: \mu \neq 4500$
$H_0: \mu = 4500$
$\alpha = 0{,}05 \rightarrow u_{1-\alpha/2} = 1{,}96$ (zweiseitiger Test)
Berechnung:
Bereich der schwachen Indizien gegen H_0:

$$\bar{x}_o = \mu + u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 4.500 + 1,96 \cdot \sqrt{\frac{12.037,089}{86}} = 4.523,19$$

$$\bar{x}_u = \mu - u_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{n}} = 4.500 - 1,96 \cdot \sqrt{\frac{12.037,089}{86}} = 4.476,81$$

Interpretation:
Da $\bar{x} = 4.464{,}63$ nicht im Bereich [4.476,81; 4.523,19] der schwachen Indizien gegen die Nullhypothese enthalten ist, liegt ein starkes Indiz gegen H_0 vor $\rightarrow$ H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ verwerfen, Entscheidung für H_1. Das Messergebnis ist signifikant anders als 4.500.

b)
$H_1: \mu < 4.500$
$H_0: \mu \geq 4.500$
$\alpha = 0{,}05 \rightarrow u_{1-\alpha} = 1{,}65$ (einseitiger Test)
Berechnung:
Untere Schranke der schwachen Indizien gegen H_0:

$$\bar{x}_u = \mu - u_{1-\alpha} \cdot \sqrt{\frac{s^2}{n}} = 4.500 - 1,65 \cdot \sqrt{\frac{12.037,089}{86}} = 4.480,48$$

Interpretation:
Da $\bar{x} = 4.464{,}63$ unter der unteren Schranke der schwachen Indizien gegen die Nullhypothese liegt, liegt ein starkes Indiz gegen H_0 vor $\rightarrow$ H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ verwerfen, Entscheidung für H_1. Das Messergebnis ist signifikant kleiner als 4.500.

Aufgabe 69 [Testen von Hypothesen über einen Mittelwert]

Die (stetige) Punktezahl bei einem Aufnahmetest sei annähernd normalverteilt mit $\mu = 75$. Nach Einführung verpflichtender vorbereitender Kurse soll überprüft werden, ob sich der Mittelwert der Punktezahlen erhöht hat.

a) Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
b) Testen Sie diese Hypothesen auf einem Signifikanzniveau von $\alpha = 0{,}05$, wenn in einer Zufallsstichprobe vom Umfang $n = 120$ nach Einführung der Kurse ein Mittelwert von 78,4 Punkten erzielt wird und in dieser Stichprobe eine Standardabweichung von s = 6 gemessen wird.

Lösung:

μ … Mittelwert der Punktezahlen in der Grundgesamtheit
$\bar{x} = 78{,}4$, $s = 6 \rightarrow s^2 = 36$, $n = 120$, $\alpha = 0{,}05 \rightarrow u_{1-\alpha} = 1{,}65$ (einseitiger Test; siehe H_1)

a)
Es interessiert ob sich die Punktezahl erhöht hat. Daher ist die Einshypothese >
H_1: $\mu > 75$
H_0: $\mu \leq 75$

b)
Berechnung:
Obere Schranke der schwachen Indizien gegen H_0:

$$\bar{x}_o = \mu + u_{1-\alpha} \cdot \sqrt{\frac{s^2}{n}} = 75 + 1{,}65 \cdot \sqrt{\frac{36}{120}} = 75{,}90$$

Interpretation:
Da $\bar{x} = 78{,}4$ über der oberen Schranke der schwachen Indizien gegen die Nullhypothese liegt, liegt ein starkes Indiz gegen H_0 vor $\rightarrow$ H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ verwerfen, Entscheidung für H_1. Die Punktezahl ist signifikant höher als 75.

Aufgabe 70 [Testen von Hypothesen über einen Mittelwert]

Die Psychologen Stanford, Binet und Wechsler haben festgestellt, dass der Intelligenzquotient in der Bevölkerung normalverteilt ist. In der Bevölkerung besitzt der IQ einen Mittelwert von 100. Überprüfen Sie auf einem Signifikanzniveau von $\alpha = 0{,}05$, ob dieser Mittelwert unter Studierenden höher ist. Dazu steht eine Zufallsstichprobe vom Umfang n = 100 zur Verfügung, in der ein durchschnittlicher IQ von 108 bei einer Standardabweichung von $s = 19$ gemessen wurde.

Lösung:
Es interessiert ob der IQ bei Studenten größer ist. Daher ist die Einshypothese >

H_1: $\mu > 100$
H_0: $\mu \leq 100$

μ … Mittelwert der Intelligenzquotienten in der Grundgesamtheit der Studierenden
$\bar{x} = 108$, $s = 19 \rightarrow s^2 = 361$, $n = 100$, $\alpha = 0{,}05 \rightarrow u_{1-\alpha} = 1{,}65$ (einseitiger Test; siehe H_1)

Berechnung:

Obere Schranke der schwachen Indizien gegen H_0:

$$\bar{x}_o = \mu + u_{1-\alpha} \cdot \sqrt{\frac{s^2}{n}} = 100 + 1{,}65 \cdot \sqrt{\frac{361}{100}} = 103{,}135$$

Interpretation:

Da $\bar{x}$ = 108 über der oberen Schranke der schwachen Indizien gegen die Nullhypothese liegt, liegt ein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau α = 0,05 verwerfen, Entscheidung für H_1. Der IQ von Studenten ist signifikant höher als 100.

Aufgabe 71* [einfache Regressionsanalyse]

Mit den Daten der im Internet bereitgestellten Excel-Datei aus Aufgabe 87 soll auf einem Signifikanzniveau α = 0,05 überprüft werden, ob die Punkte in der Hausübung einen Einfluss auf die Punkte in der Statistikklausur haben.

Lösung:

H_1: $\beta_1 \neq 0$
H_0: $\beta_1 = 0$ (kein Effekt bedeutet Koeffizient=0, horizontale Regressionsgerade)

n = 155, r = 0,788 (aus Ü87), s_y^2 = 3,109 (in der Excel-Datei mit der Funktion „Varianzen" berechnet), s_x^2 = 33,622 (ebenso), s_{xy} = 8,059 (in der Excel-Datei mit der Funktion „Kovar" berechnet), α = 0,05 → $u_{1-\alpha/2}$ = 1,96 (zweiseitiger Test; siehe H_1)

Berechnung:

$$s_{b_1}^2 = \frac{1-r^2}{n-2} \cdot \frac{s_y^2}{s_x^2} = \frac{1-0{,}788^2}{155-2} \cdot \frac{3{,}109}{33{,}622} = 0{,}00023$$

Bereich der schwachen Indizien gegen H_0:

$$b_{1o} = u_{1-\alpha/2} \cdot s_{b_1} = 1{,}96 \cdot \sqrt{0{,}00023} = 0{,}0297$$

$$b_{1u} = -u_{1-\alpha/2} \cdot s_{b_1} = -1{,}96 \cdot \sqrt{0{,}00023} = -0{,}0297$$

Steigung der Regressionsgeraden in der Stichprobe:

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{8{,}059}{33{,}622} = 0{,}240$$

Interpretation:

Da b_1 = 0,240 nicht im Bereich [-0,0297; 0,0297] enthalten ist, liegt ein starkes Indiz gegen H_0 vor → H_0 auf dem Signifikanzniveau α = 0,05 verwerfen, Entscheidung für H_1. Es gibt einen Einfluss der Hausübung auf die Klausur der signifikant anders als 0 ist.

Aufgabe 72* [einfache Regressionsanalyse]

Mit den Daten der im Internet bereitgestellten Excel-Datei aus Aufgabe 87 soll auf einem Signifikanzniveau $\alpha = 0{,}05$ überprüft werden, ob der Anstieg der Regressionsgerade (UV=Punkte in Übung, AV=Punkte in der Klausur) größer als 0 ist.

Lösung:

H_1: $\beta_1 > 0$ (bedeutet 1 Punkt mehr in Übung schlägt sich nieder in höheren Punkten in einer höheren Punktzahl in der Statistikklausur)
H_0: $\beta_1 \leq 0$

n = 155, r = 0,788 (aus Ü87), $s_y^2 = 3{,}109$ (in der Excel-Datei mit der Funktion „Varianzen" berechnet), $s_x^2 = 33{,}622$ (ebenso), $s_{xy} = 8{,}059$ (aus der Datei mit der Funktion „Kovar"), $\alpha = 0{,}05 \rightarrow u_{1-\alpha} = 1{,}65$ (einseitiger Test)

Obere Schranke der schwachen Indizien gegen H_0:

$$b_{1o} = u_{1-\alpha/2} \cdot s_{b1} = u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{1-r^2}{n-2} \cdot \frac{s_y^2}{s_x^2}} = 1{,}65 \cdot \sqrt{\frac{1-0{,}788^2}{155-2} \cdot \frac{3{,}109}{33{,}622}} = 0{,}025$$

Steigung der Regressionsgeraden in der Stichprobe:

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{8{,}059}{33{,}622} = 0{,}240$$

Interpretation:
Da $b_1 = 0{,}240$ oberhalb der oberen Schranke 0,025 liegt, liegt ein starkes Indiz gegen H_0 vor $\rightarrow$ H_0 wird auf dem Signifikanzniveau $\alpha = 0{,}05$ abgelehnt und die H_1 angenommen. Der Einfluss der Hausübung auf die Statistikklausur ist signifikant größer als 0.

Aufgabe 73 [einfache Regressionsanalyse]

Nehmen Sie an, es wurden die Tests von Aufgabe 71 bis Aufgabe 72 durchgeführt. Sie erfahren als Testergebnis von jedem dieser Tests lediglich, dass der zweiseitige p – Wert 0,075 beträgt. Für welche der beiden Hypothesen aus Aufgabe 71 bis Aufgabe 72 sollten Sie sich demnach jeweils entscheiden?

Lösung:

Nullhypothese beibehalten, wenn gilt: p-Wert $\alpha_2 > \alpha$ (zweiseitiger Test) bzw. p-Wert $\alpha_1 > \alpha$ (einseitiger Test) mit der Beziehung $\alpha_1 = \alpha_2/2$

Bei Aufgabe 91 Vergleich p-Wert α_2 mit zugestandener Irrtumswahrscheinlichkeit α (2-seitiger Test)
Bei Aufgabe 92 Vergleich p-Wert α_1 mit zugestandener Irrtumswahrscheinlichkeit α (1-seitiger Test)

In der Ergebnistabelle ist dann aufgeführt für welche Hypothese H_0 oder H_1 man sich entscheidet.
$\alpha = 0{,}05$

	Ü71	Ü72
$\alpha_2 = 0{,}075 \rightarrow$ $\alpha_1 = 0{,}0375$	H_0	H_1

Aufgabe 74 [multivariate Regressionsanalyse]

Interpretieren Sie bitte folgendes Ergebnistableau der Regressionsanalyse mit $y = ß_1 \cdot x_1 + ß_2 \cdot x_2 + ß_3 \cdot x_3 + ß_4$ mit

y = Arbeitseinkommen je Monat
x_1 = Alter in Jahren
x_2 = Dummy Humankapital (1=hat mindestens Bachelor, 0=hat keinen Bachelor)
x_3 = Größe der Firma in Anzahl Mitarbeiter
wenn ihre zugestandene Irrtumswahrscheinlichkeit α=0,05 beträgt und zweiseitige Hypothesentests auf H_0: $ß_i = 0$ durchgeführt wurden.

Unabhängige Variable UV	Regressionskoeffizient	p-Wert (2 Stellen gerundet)
Alter in Jahren	100	0,07
Dummy Humankapital	500	0,00
Größe der Firma	0,5	0,23
Achsenabschnitt $ß_4$	300	0,00

Lösung:

Interpretation des p-Werts (Frage ob Regressionskoeffizienten signifikant von 0 verschieden sind.
p-Wert bei Alter in Jahren ist größer als 5% (nämlich 7%) → nicht signifikant → H_0: $ß_1$=0 kann nicht verworfen werden.
p-Wert bei Dummy Humankapital <5% → signifikant. H_0: $ß_2$=0 kann verworfen werden.
p-Wert bei Größe der Firma ist größer als 5% → nicht signifikant → H_0: $ß_3$=0 kann nicht verworfen werden.
p-Wert bei Achsenabschnitt ist kleiner als 5% → signifikant → H_0: $ß_4$=0 kann verworfen werden.

Interpretation der Größe der Koeffizienten
b_1=100 → 1 Jahr älter führt zu 100 Euro mehr Einkommen
b_2=500 → Ein Bachelor-Abschluss führt zu 500 Euro mehr Einkommen
b_3=0,5 → Ein Mitarbeiter mehr in der Firma führt zu 50 Cent mehr Einkommen
b_4=300 → Achsenabschnitt ist das Einkommen was man verdient wenn man 0 Jahre ist, keinen Bachelorabschluss hat und die Firma 0 Mitarbeiter hat. Im Grunde ist dies der Teil des Einkommens der nicht mit dem Regressionsmodell erklärt werden kann.