

UNIVARIANTE VERTEILUNGEN V UND KREUZTABELLEN

Martin-Luther-Universität Halle-Wittenberg

Institut für Soziologie

Übung Einführung in die deskriptive Statistik

Was machen wir heute?

- Univariate Verteilungen V:
 - Schiefe
 - Steilheit
 - Histogramm
 - Boxplot
- Von der univariaten zur bivariaten Statistik
 - Kreuztabellen und Prozentuierungen

Übersicht: univariate Kennwerte

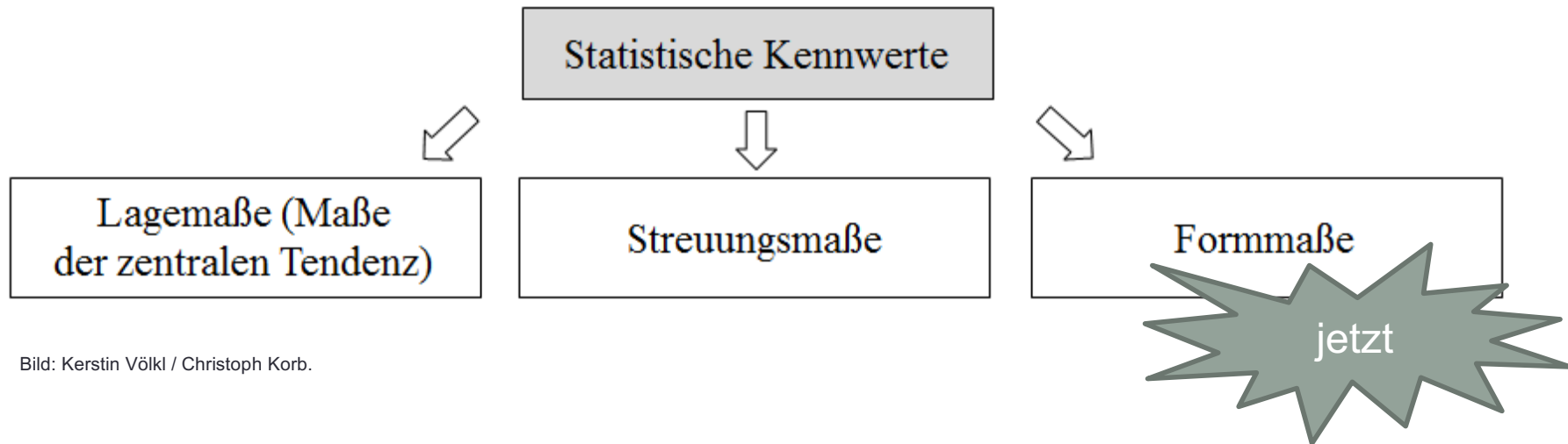
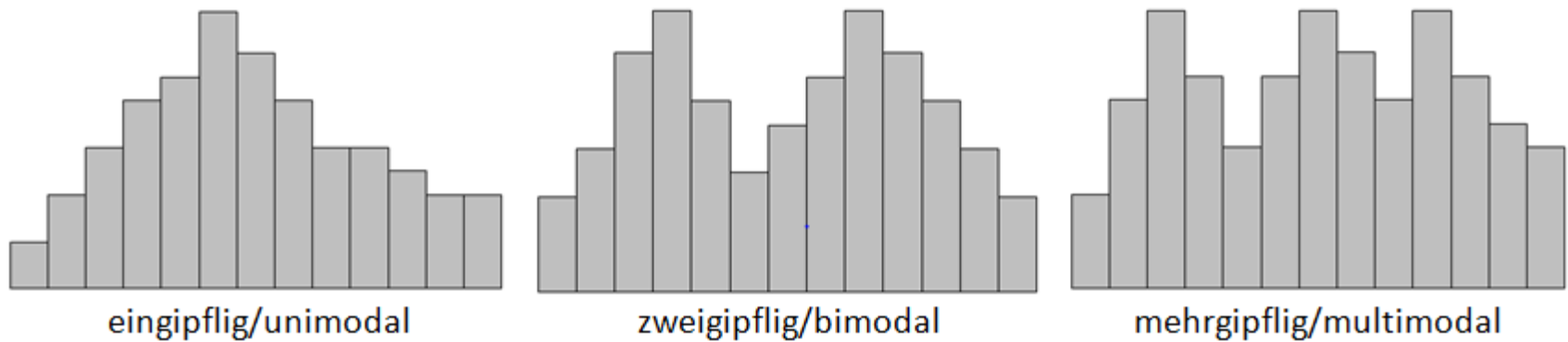


Bild: Kerstin Völkl / Christoph Korb.

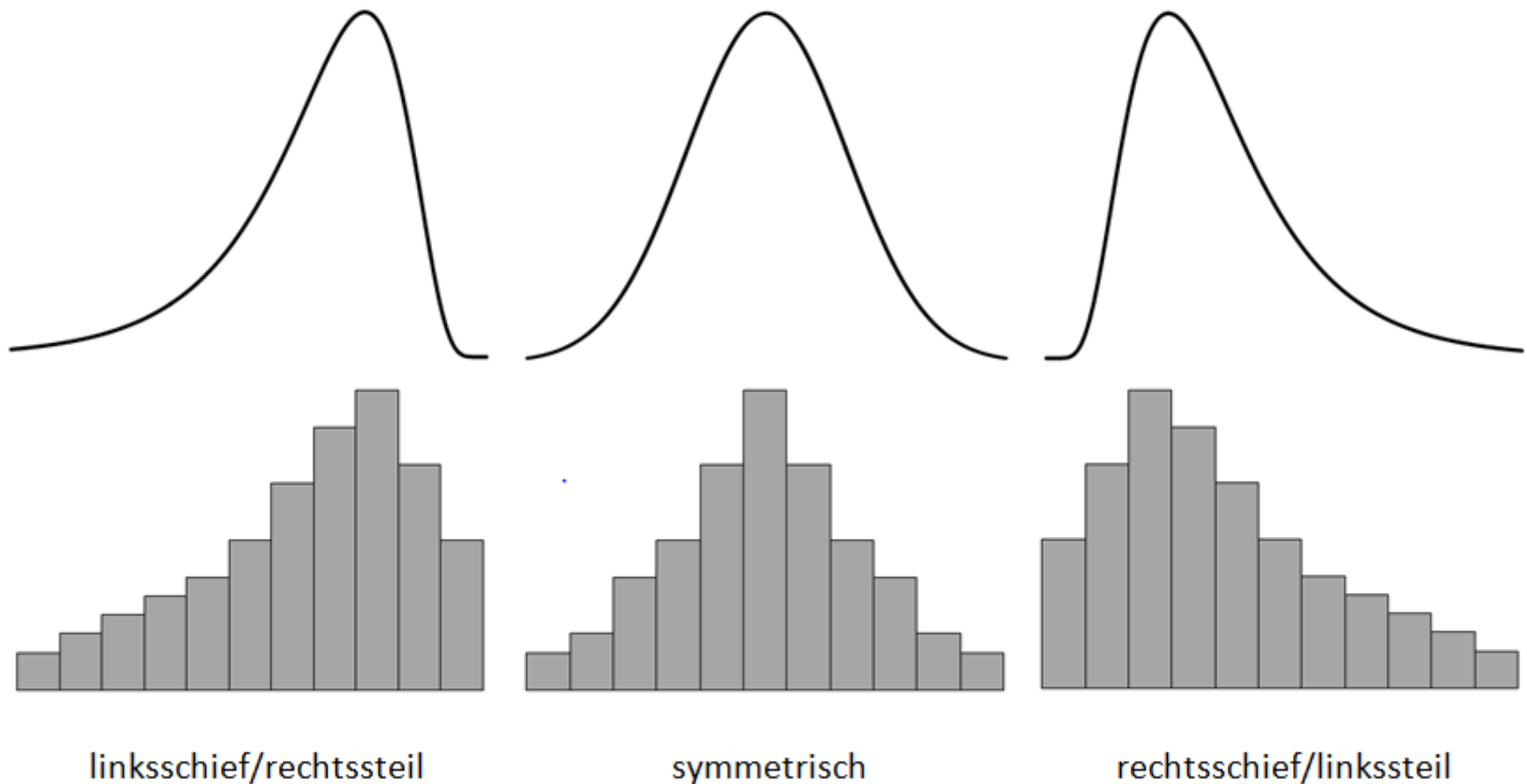
Verteilungsformen: Anzahl der Gipfel

- Unterscheidung nach Anzahl der Gipfel:



Verteilungsformen: Schiefe (Visualisierung)

- Unterscheidung nach Schiefe:



Verteilungsformen: Schiefe (Lageregel)

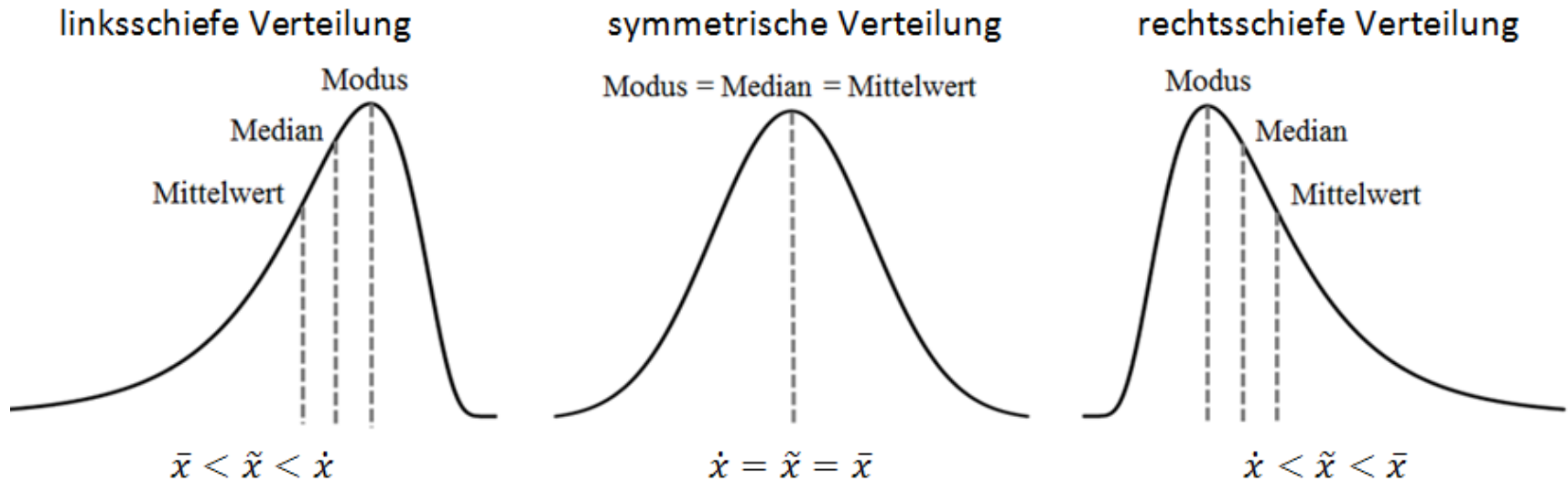


Bild: Kerstin Völkl / Christoph Korb (2018: 122).

Hinweis: Es handelt sich hierbei lediglich um eine Daumenregel, die nicht immer zutreffen muss. Bei mehrgipfligen Verteilungen wird der Modus ignoriert.

Verteilungsformen: Schiefe (Formel)

- Formel:

- $Schiefe = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s_X^3}$

- Interpretation:

- $Schiefe < 0 \rightarrow$ linksschiefe (rechtssteile) Verteilung
 - $Schiefe = 0 \rightarrow$ symmetrische Verteilung
 - $Schiefe > 0 \rightarrow$ rechtsschiefe (linkssteile) Verteilung

Aufgabe 1: Haushaltseinkommen 2014

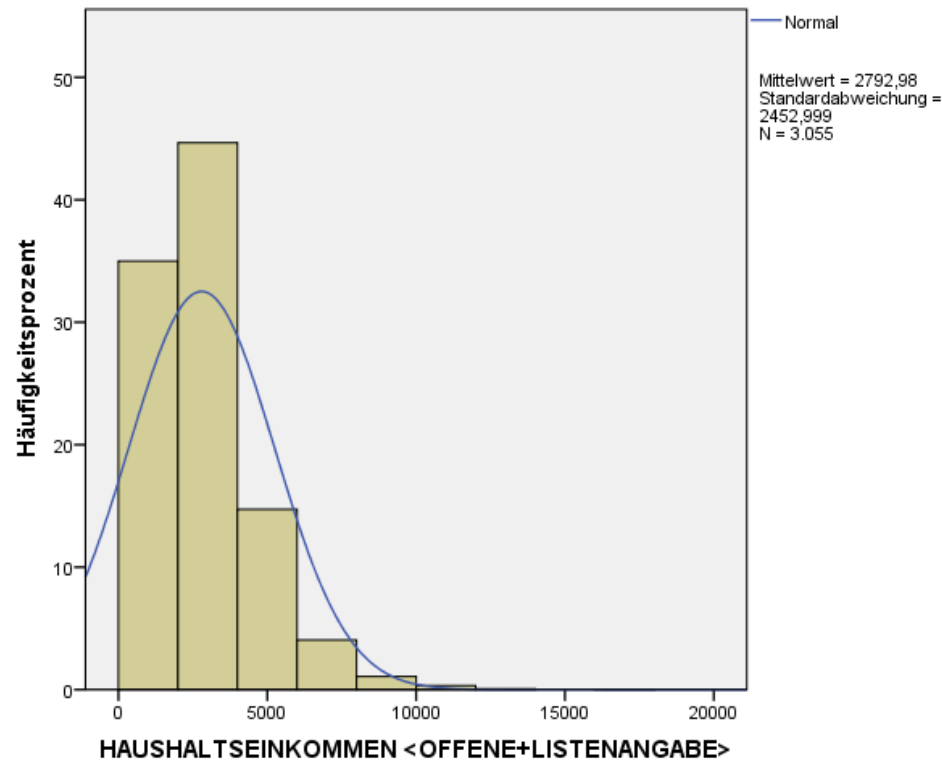
Für das Haushaltseinkommen im ALLBUS 2014 wurden folgende Kennwerte mithilfe von SPSS bestimmt:

N	Gültig	3055
	Fehlend	416
Schiefe		16,170
Standardfehler der Schiefe		,044
Kurtosis		462,971
Standardfehler der Kurtosis		,089

Welche Aussage können Sie mithilfe der Angaben über die Schiefe der Verteilung treffen?

Aufgabe 1: Lösung

- Die Verteilung ist rechtsschief.



Verteilungsformen: Wölbung

- Unterscheidung nach Wölbung:

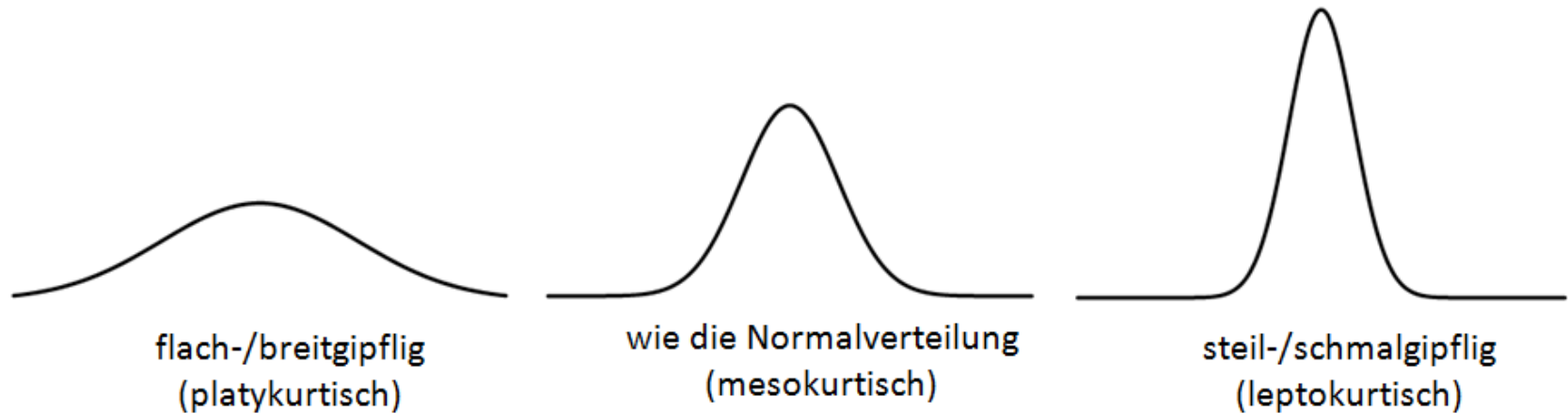


Bild: Kerstin Völkl / Christoph Korb (2018): 125.

Verteilungsformen: Wölbung (Kurtosis)

- Berechnung:

- $$Kurtosis = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s_X^4} - 3$$

- Interpretation:

- >0 → Verteilung steiler als Normalverteilung (leptokurtisch)
 - $=0$ → Verteilung genauso steil wie Normalverteilung (mesokurtisch)
 - <0 → Verteilung flacher als Normalverteilung (platykurtisch)

Achtung: Manche Statistikprogramme (z.B. Stata) verwenden eine andere Formel, bei der Wert 3 nicht abgezogen wird. Dies entspricht der ursprünglichen Formulierung nach Pearson / Fisher.

Aufgabe 2: Haushaltseinkommen

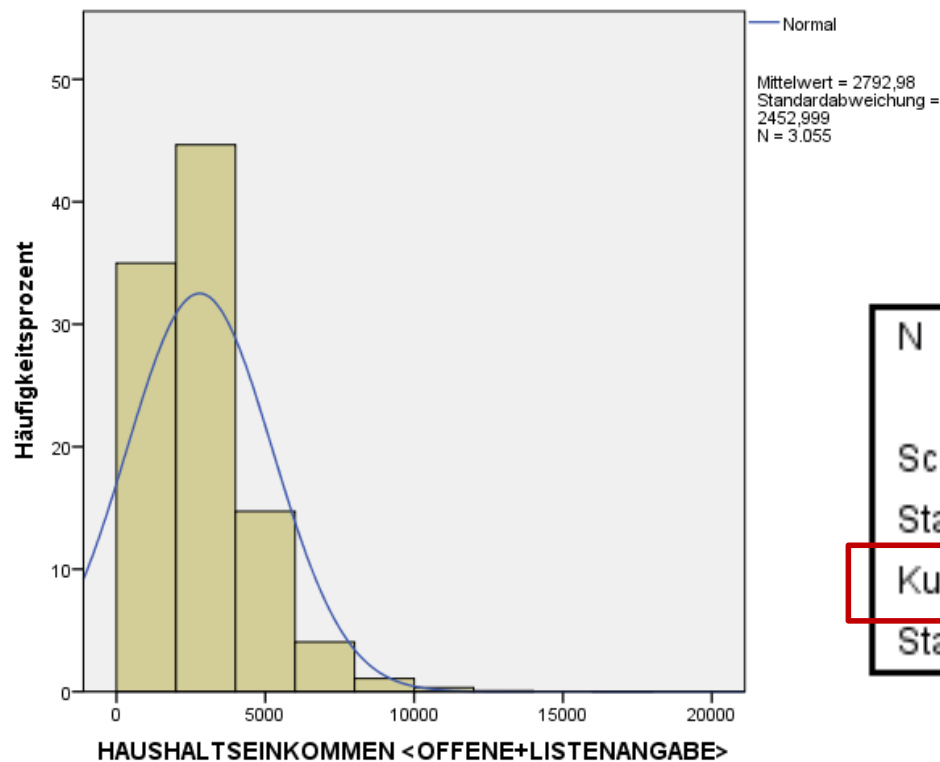
Für das Haushaltseinkommen im ALLBUS 2014 erhalten Sie folgende Tabelle.

N	Gültig	3055
	Fehlend	416
Schiefe		16,170
Standardfehler der Schiefe		,044
Kurtosis		462,971
Standardfehler der Kurtosis		,089

Welche Aussage können Sie zur Wölbung treffen? Ist die Kurve steiler oder flacher als die Normalverteilung?

Aufgabe 2: Lösung

- $Kurtosis > 0 \rightarrow$ Verteilung ist steiler als Normverteilung



N	Gültig	3055
	Fehlend	416
Schiefe		16,170
Standardfehler der Schiefe		,044
Kurtosis		462,971
Standardfehler der Kurtosis		,089

Zurück zu Judith und Clemens...

Wie wir bereits wissen, verbringt Judith viel Zeit damit auf ihren Freund Clemens zu warten.

Für die letzten 15 Verabredungen hat Sie sich die Zeit genommen, ihre Wartezeit in Minuten zu notieren dabei entstand folgende Liste:

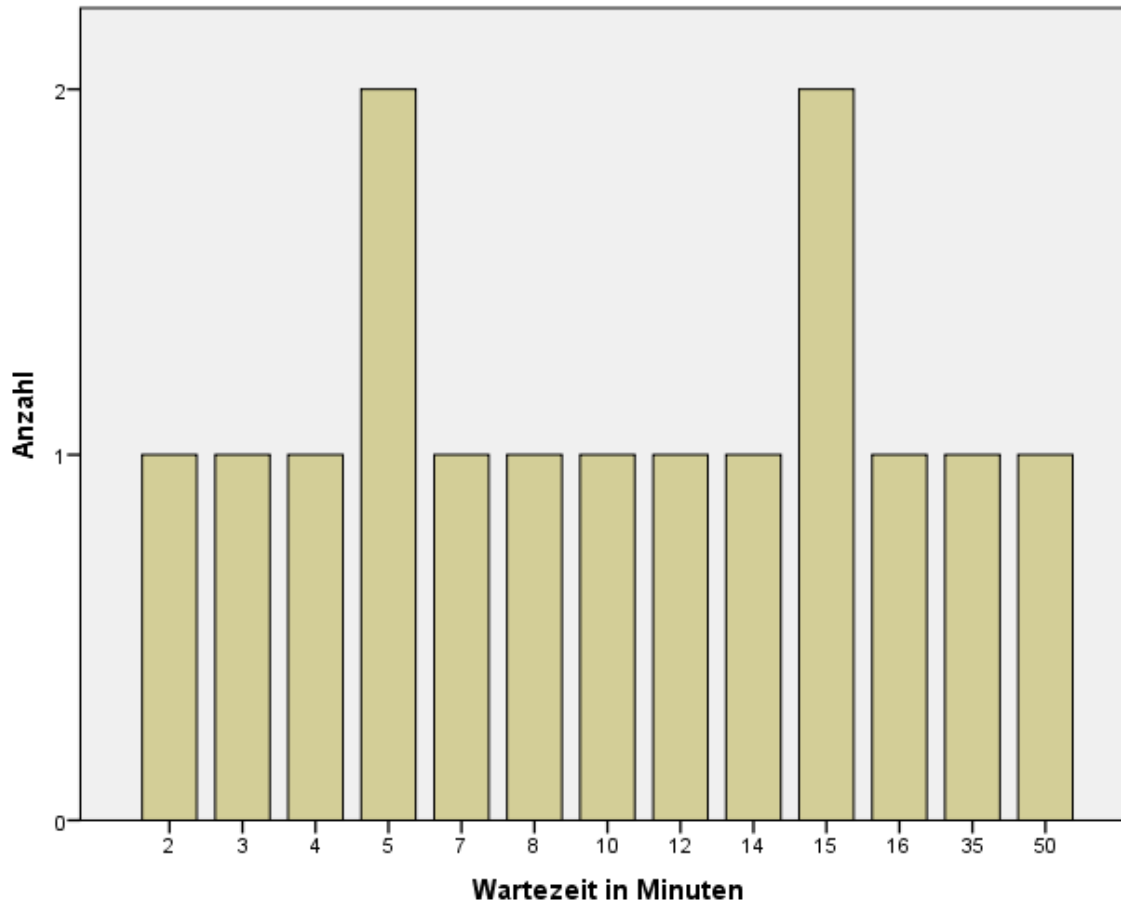
2,3,4,5,5,7,8,10,12,14,15,15,16,35,50



Bild: „Angry Cat“ von Orias1978. Lizenz: CC-BY-2.0.

Wie lassen sich solche Daten sinnvoll darstellen?

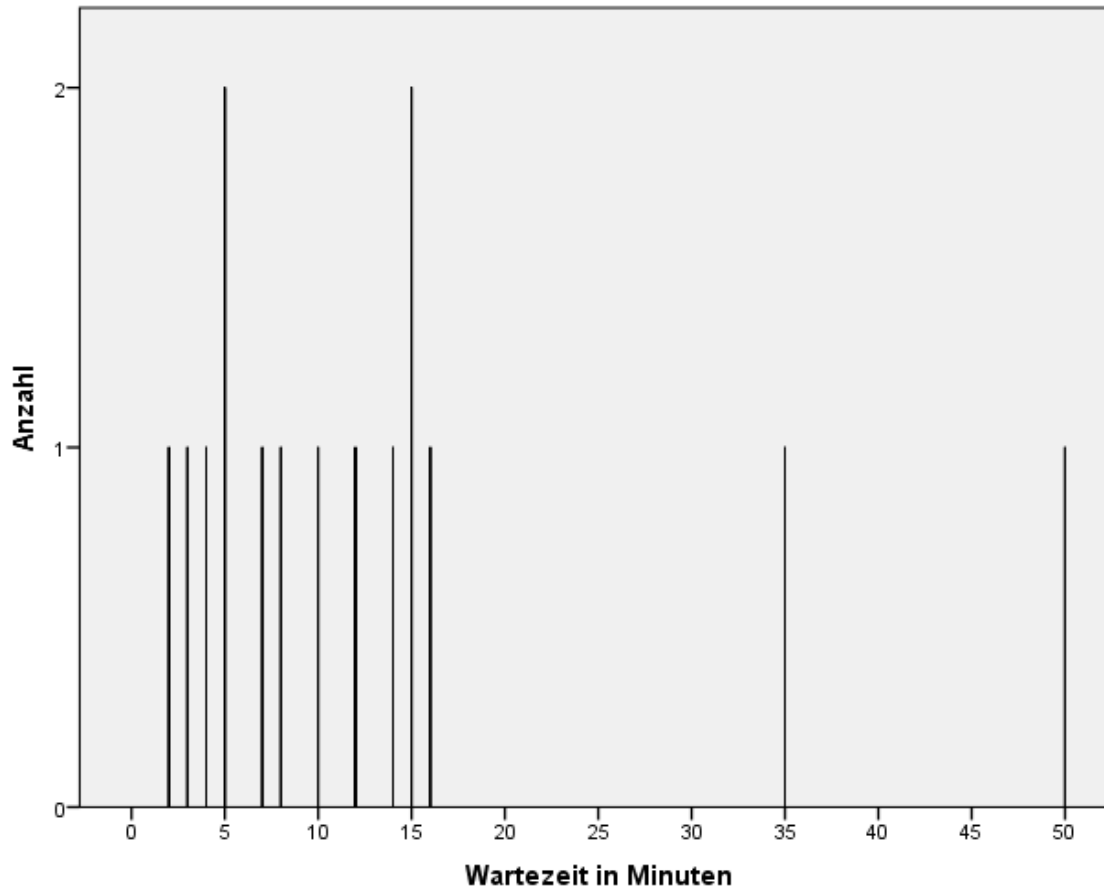
Balkendiagramme für metrische Daten?



Quiz: Handelt es sich um eine adäquate Darstellung für die Daten? Warum bzw. warum nicht?

Quiz: Wofür eignen sich Balkendiagramme besser?

Stabdiagramm



Quiz: Was ist im Vergleich zum Balkendiagramm anders? Was ist besser?

Wie lassen sich klassierte Häufigkeiten darstellen?

Ausgangsbeispiel: Wartezeit

Wartezeit in Minuten

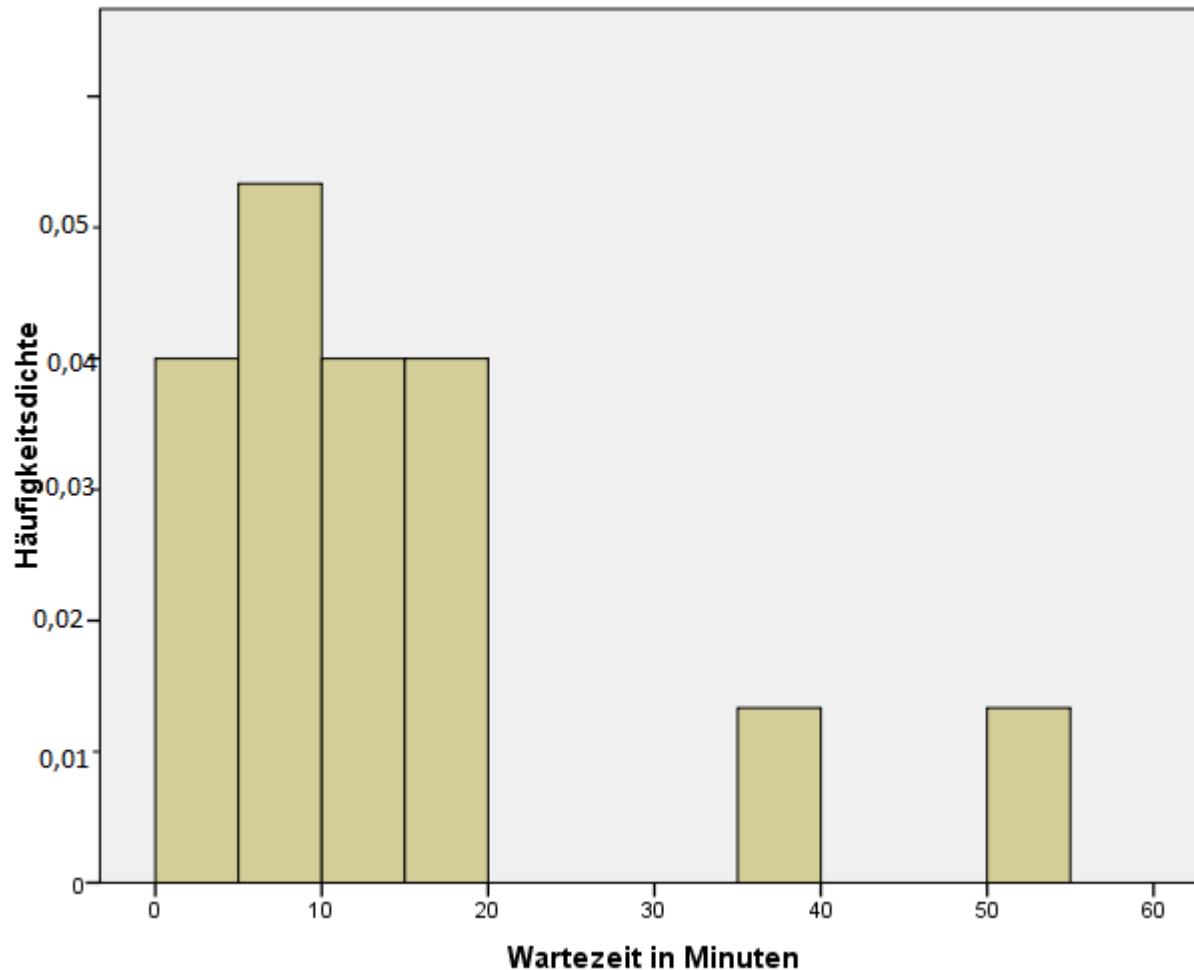
		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente	
Gültig	2	1	6,3	6,7	6,7	
	3	1	6,3	6,7	13,3	
	4	1	6,3	6,7	20,0	
	5	2	12,5	13,3	33,3	
	7	1	6,3	6,7	40,0	
	8	1	6,3	6,7	46,7	
	10	1	6,3	6,7	53,3	
	12	1	6,3	6,7	60,0	
	14	1	6,3	6,7	66,7	
	15	2	12,5	13,3	80,0	
	16	1	6,3	6,7	86,7	
	35	1	6,3	6,7	93,3	
	50	1	6,3	6,7	100,0	
	Gesamtsumme		15	93,8	100,0	
	Fehlend System		1	6,3		
Gesamtsumme		16	100,0			

Wartezeit in Minuten(Klassiert)

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	0 - <5	3	20,0	20,0	20,0
	5 - <10	4	26,7	26,7	46,7
	10 - <15	3	20,0	20,0	66,7
	15 - <20	3	20,0	20,0	86,7
	35 - <40	1	6,7	6,7	93,3
	50 - <55	1	6,7	6,7	100,0
	Gesamtsumme		15	100,0	100,0

Klassierung

Histogramm



Klassierung der Daten

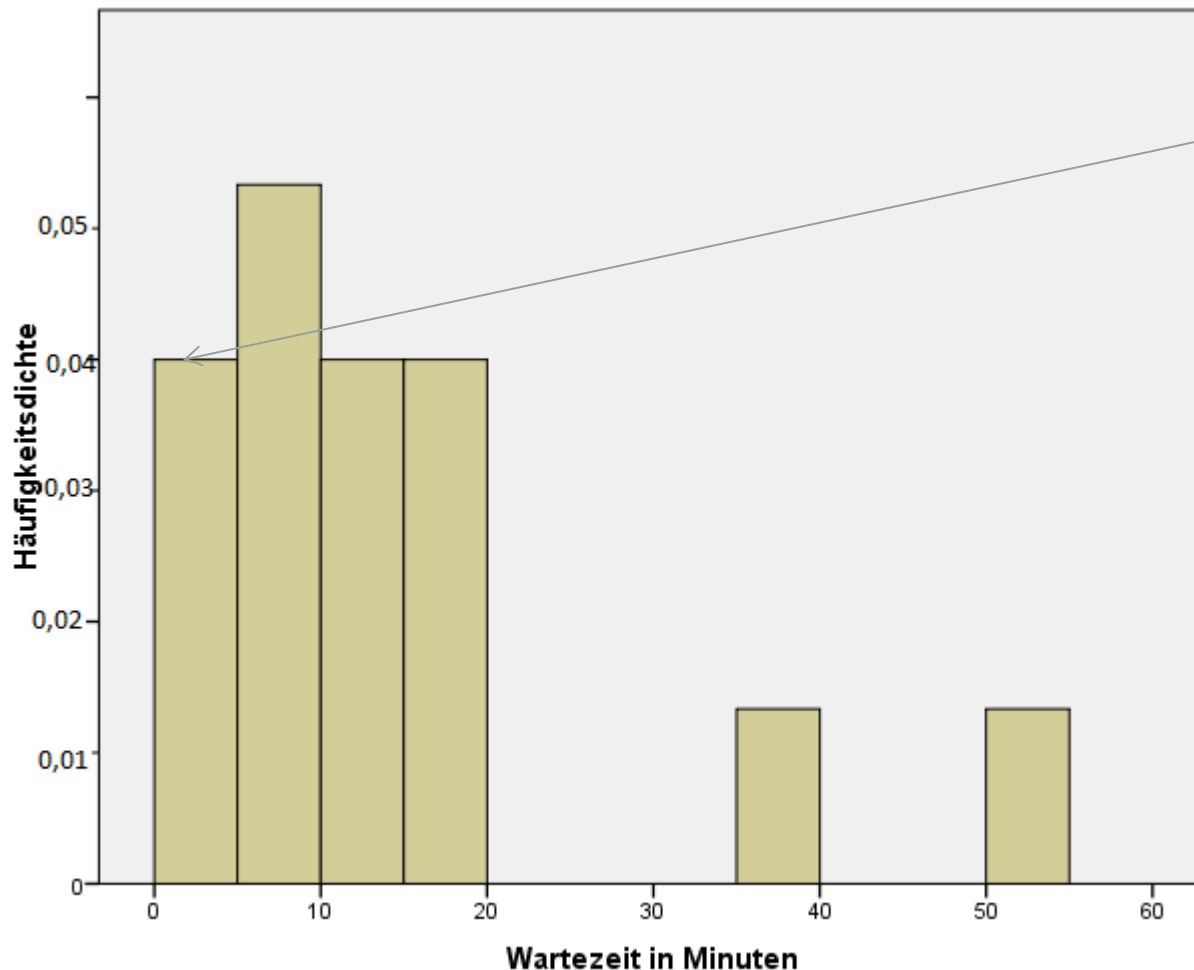
Idee:
Prinzip der
Flächentreue

Breite = Klassenbreite

Höhe =
Häufigkeitsdichte

$$\hat{f}_j = \frac{p_j}{(o_j - u_j)}$$

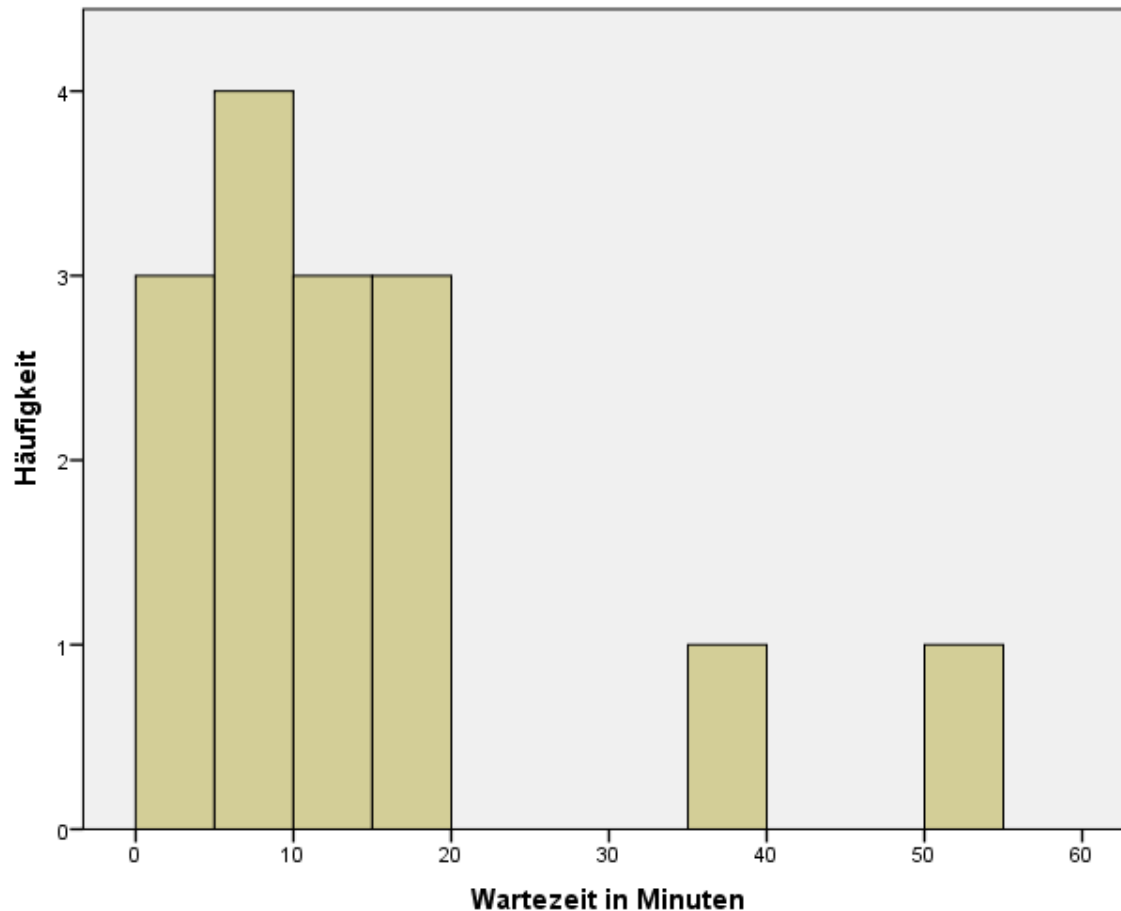
Histogramm II



$$\hat{f}_j = \frac{p_j}{(o_j - u_j)}$$
$$\hat{f}_j = \frac{0,20}{(5 - 0)} = 0,04$$

Durch Verwendung von Histogrammen lassen sich auch Häufigkeiten von Klassen darstellen, die unterschiedlich groß sind. Die Fläche der Rechtecke ist hierbei entscheidend.

Histogramm III

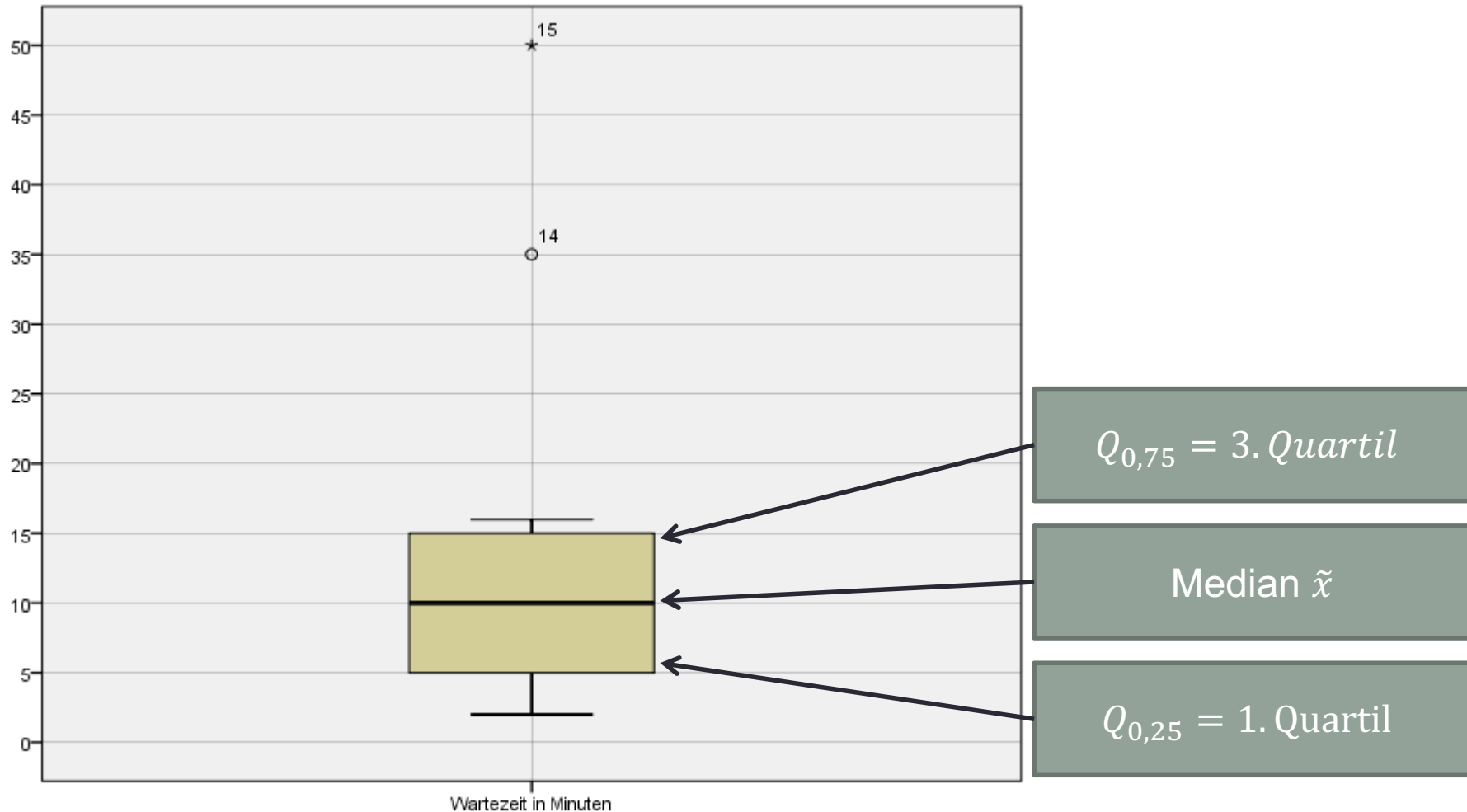


Da Statistikprogramme, die Intervalle gleich groß gestalten, lassen sich bei solchen Histogrammen die Häufigkeiten direkt ablesen. Dies an sich ist jedoch keine Eigenschaft des Histogramms!

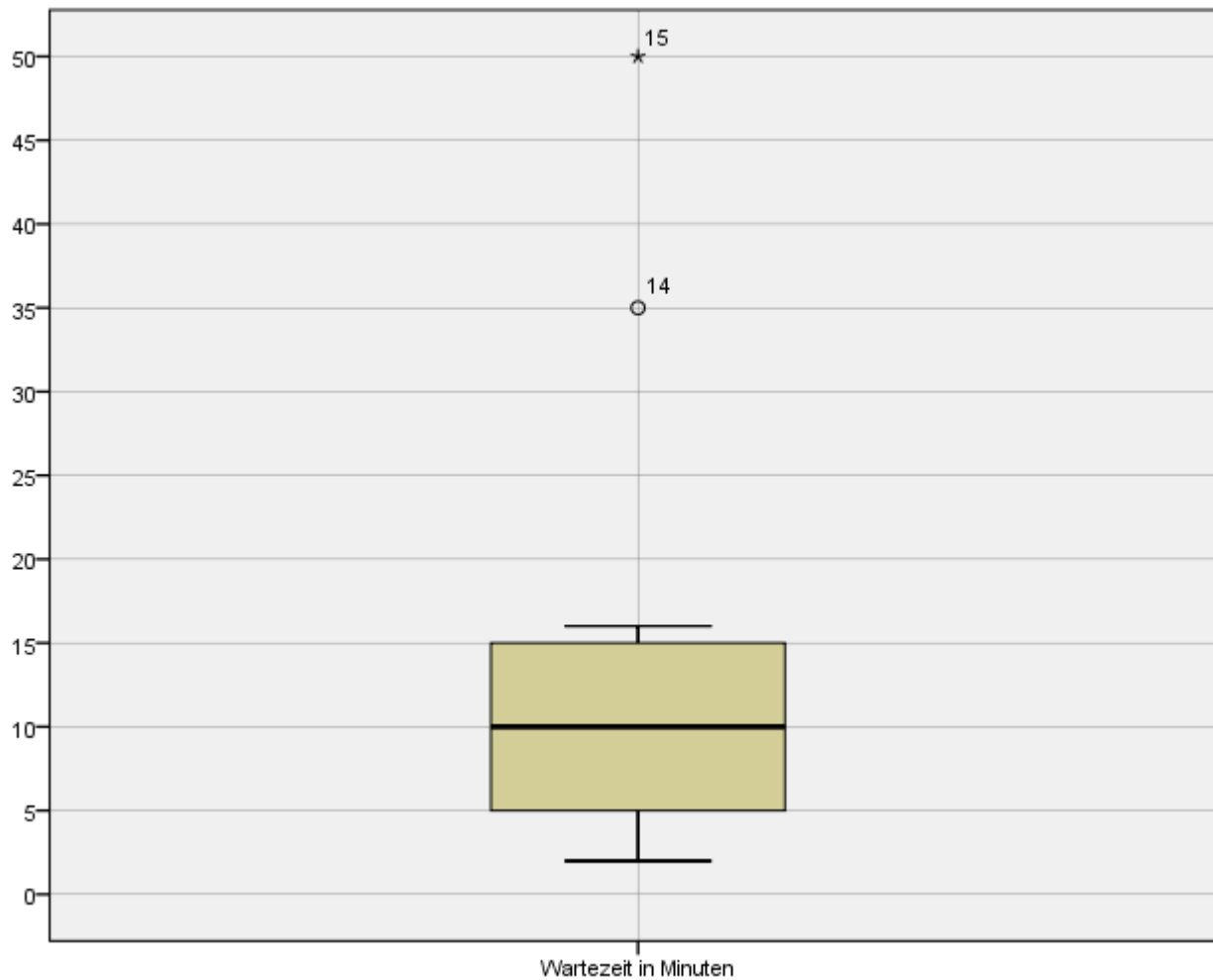
Histogramm IV: Eigenschaften

- geeignet zur Darstellung metrischer Variablen, wenn die Variable eine Vielzahl von Ausprägungen aufweist
- einzige sinnvolle Möglichkeit klassierte Daten darzustellen
- Klassengrenzen müssen nicht zwangsweise gleich groß sein, sondern können eine Variable breite aufweisen
- Fläche der Boxen proportional zu den relativen Anteilen, nicht die Höhe der Balken

Box-Whisker-Plot: Bestimmung der Box



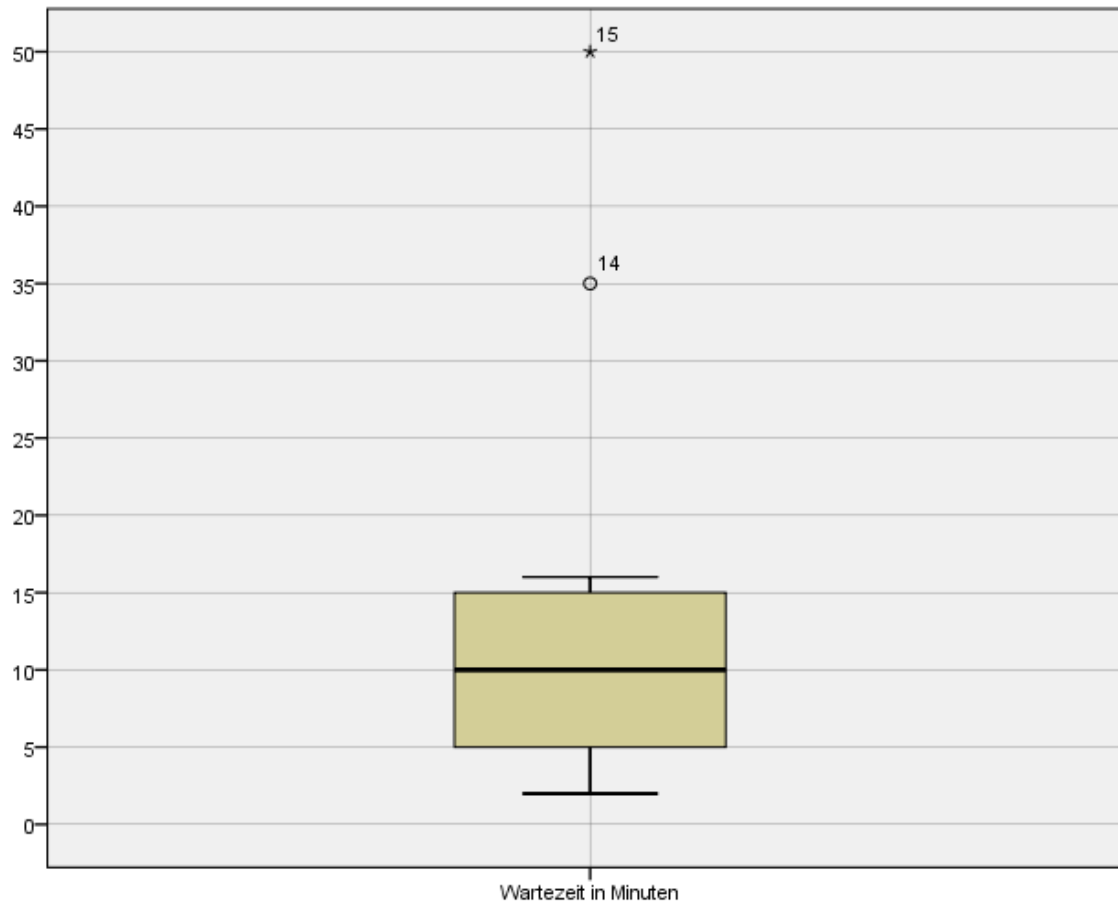
Quiz



Bestimmen Sie aus der Grafik das 25%-Quantil, den Median sowie das 75%-Quantil.

Was besagen diese Werte?

Quiz: Lösung

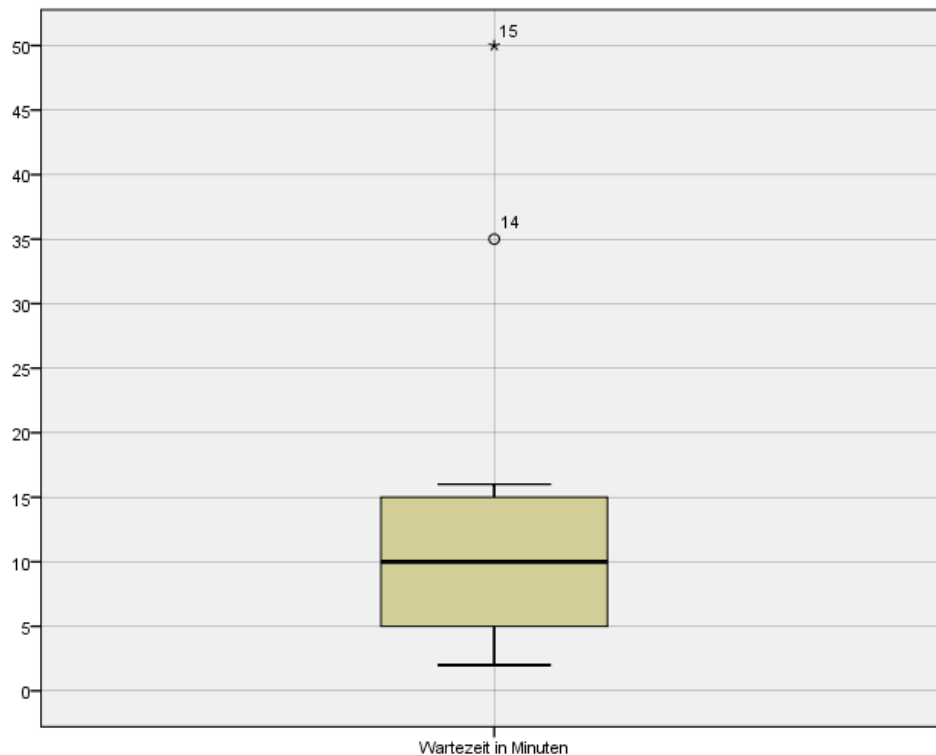


Statistiken

Wartezeit in Minuten

N	Gültig	15
	Fehlend	0
Median		10,00
Perzentile	25	5,00
	75	15,00

Quiz II



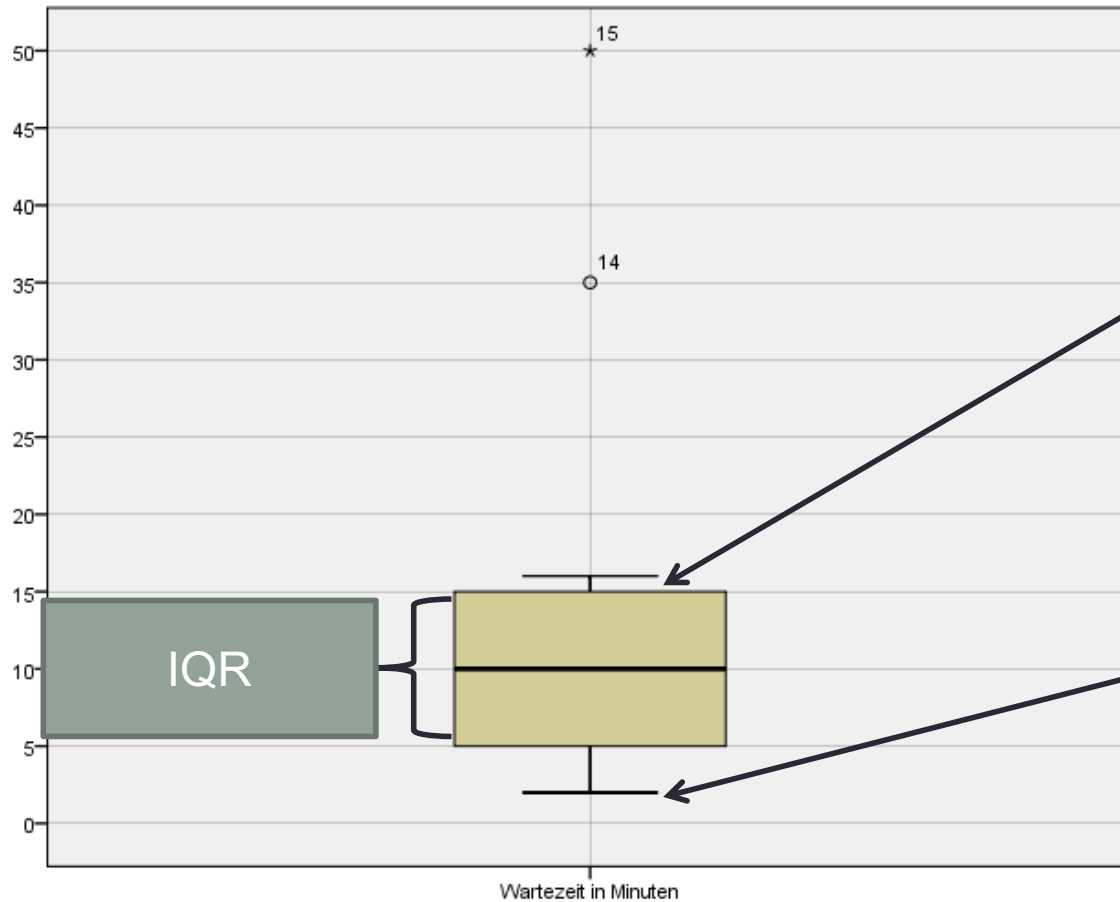
Statistiken

Wartezeit in Minuten

N	Gültig	15
	Fehlend	0
Median		10,00
Perzentile	25	5,00
	75	15,00

Bestimmen Sie den Interquartilsabstand.
Wo findet sich dieser im Box-Plot wieder?

Quiz II: Lösung



größter realisierter Wert, der kein Ausreißer oder Extremwert ist

kleinster realisierter Wert, der kein Ausreißer oder Extremwert ist

Box-Whisker-Plot: Länge der Barthaare

- maximale Länge des oberen Barthaares:
 - $Q_{0,75} + 1,5 * IQR$
 - gezeichnet wird bis zum höchsten tatsächlich realisierten Wert im Intervall von $Q_{0,75}$ bis $Q_{0,75} + 1,5 * IQR$
- maximale Länge des unteren Barthaares:
 - $Q_{0,25} - 1,5 * IQR$
 - gezeichnet wird bis zum niedrigsten tatsächlich realisierten Wert im Intervall von $Q_{0,25}$ bis $Q_{0,25} - 1,5 * IQR$

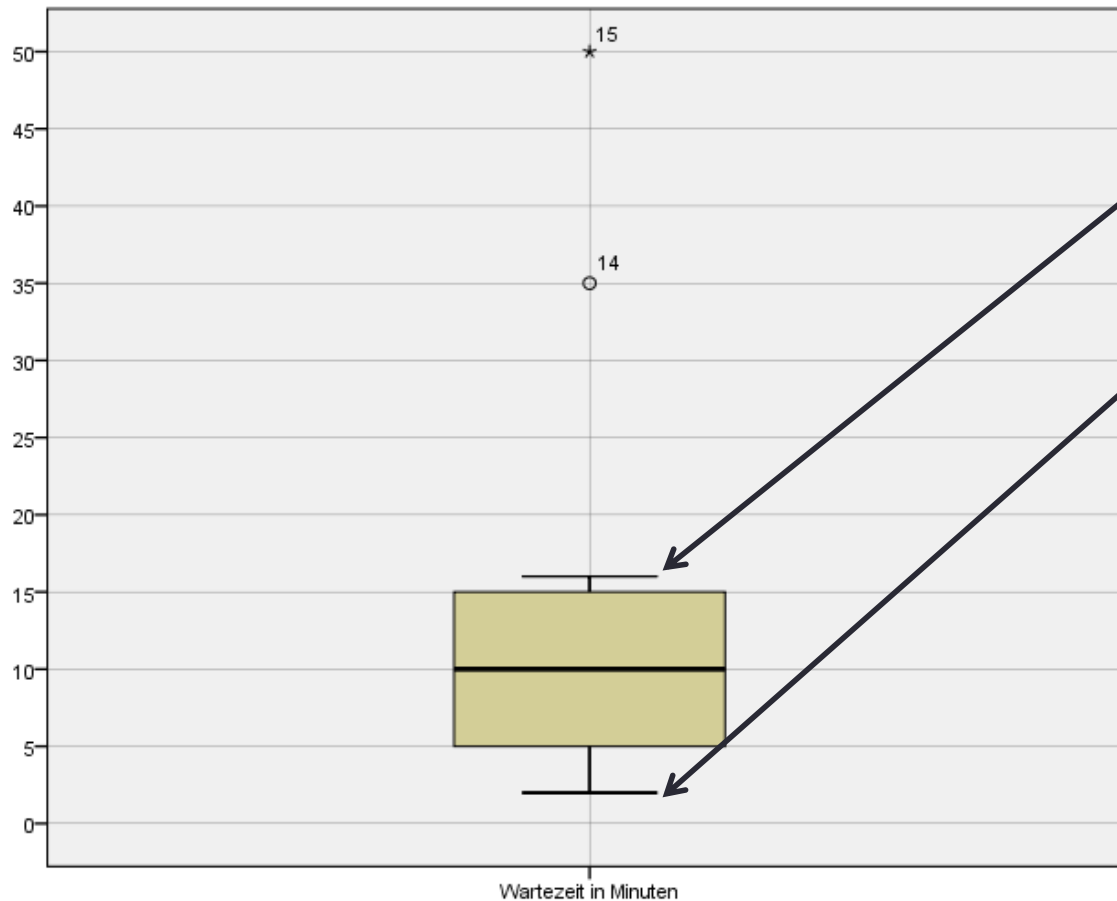
Box-Whisker-Plot: Beispiel

- $IQR = Q_{0,75} - Q_{0,25} = 15 - 5 = 10$
- maximale Länge des unteren Barthaars:
 - $Q_{0,25} - 1,5 * IQR = 5 - 1,5 * 10 = -10$
- maximale Länge des oberen Barthaars:
 - $Q_{0,75} + 1,5 * IQR = 15 + 1,5 * 10 = 30$

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	2	3	4	5	5	7	8	10	12	14	15	15	16	35	50

Bis zu welchen Werten werden die Whisker im vorliegenden Beispiel gezeichnet?

Box-Whisker-Plot: Beispiel II



16

2

Achtung: Barthaare werden nur bis zu tatsächlich realisierten Werten gezeichnet, die in besagten Intervallen liegen!

Box-Whisker-Plot: Ausreißer und Extremwerte

- **Ausreißer:**
 - alle Werte im Intervall von $[Q_{0,75} + 1,5 * IQR ; Q_{0,75} + 3 * IQR]$ und im Intervall von $[Q_{0,25} - 1,5 * IQR ; Q_{0,25} - 3 * IQR]$
 - etwas größer / kleiner als die üblichen Werte der Verteilung sind
 - SPSS: Kreis
- **Extremwerte:**
 - $> (Q_{0,75} + 3 * IQR)$ bzw. $< (Q_{0,25} - 3 * IQR)$
 - sehr weit weg von den üblichen Werten der Verteilung
 - SPSS: Sternchen

Bestimmen Sie die Ausreißer und Extremwerte!

Box-Whisker-Plot: Beispiel III

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	2	3	4	5	5	7	8	10	12	14	15	15	16	35	50

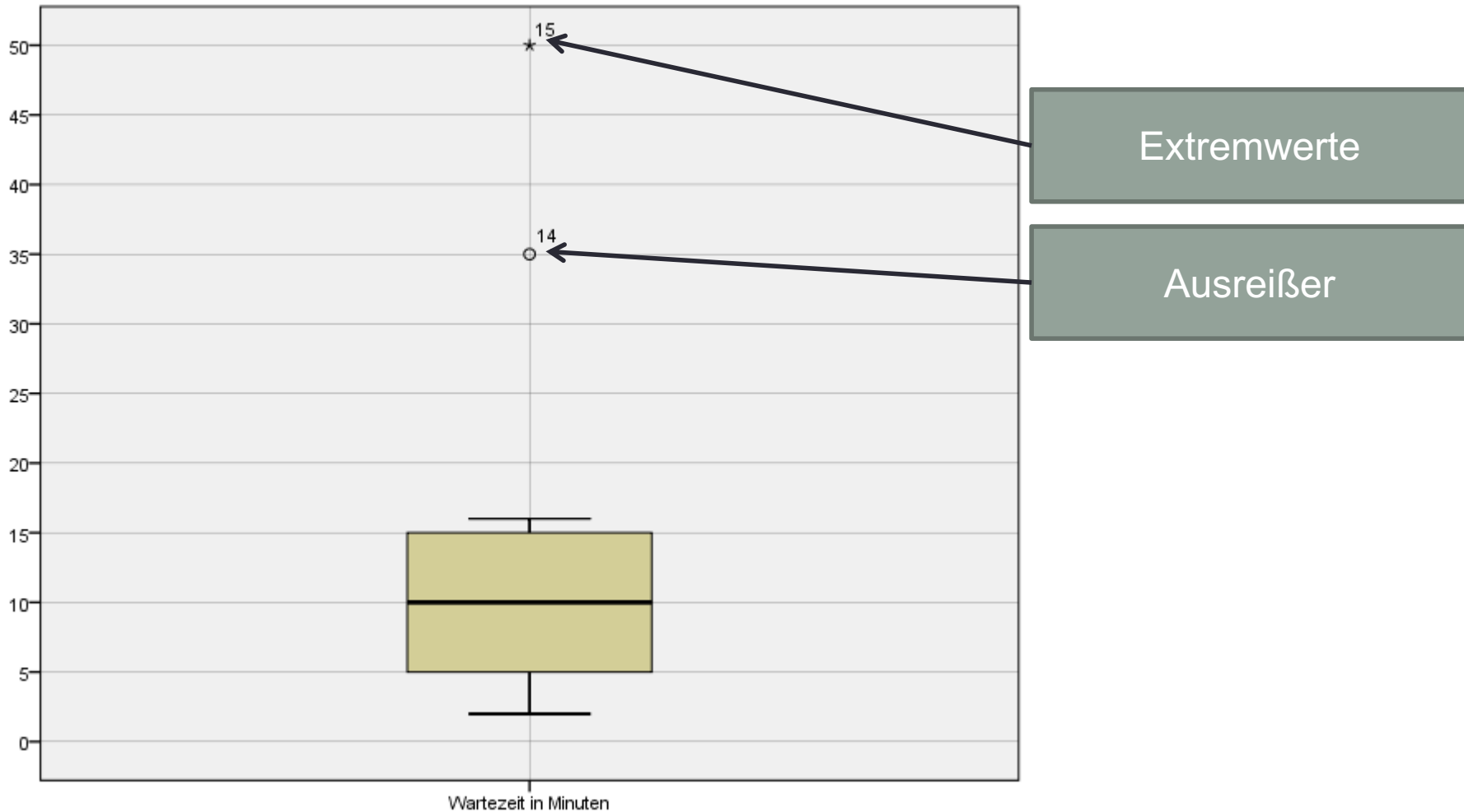
- Ausreißer nach oben
 - alle Werte im Intervall von $[Q_{0,75} + 1,5 * IQR ; Q_{0,75} + 3 * IQR]$
 - alle Werte im Intervall von $[15 + 15 = 30; 15 + 30 = 45]$
 - im angegebenen Intervall befindet sich der Wert 35
- Ausreißer nach unten:
 - alle Werte im Intervall von $[Q_{0,25} - 3 * IQR ; Q_{0,25} - 1,5 * IQR]$
 - alle Werte im Intervall von $[5 - 30 = -25; 5 - 15 = -10]$
 - kein Wert im Intervall → keine Ausreißer nach unten

Box-Whisker-Plot: Beispiel IV

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	2	3	4	5	5	7	8	10	12	14	15	15	16	35	50

- Extremwerte nach oben:
 - Werte $> Q_{0,75} + 3 * IQR$
 - Werte > 45
 - diese Bedingung trifft auf den Wert 50 zu \rightarrow ein Extremwert nach oben
- Extremwerte nach unten:
 - Werte $< Q_{0,25} - 3 * IQR$
 - Werte < -25
 - keine Extremwerte nach unten

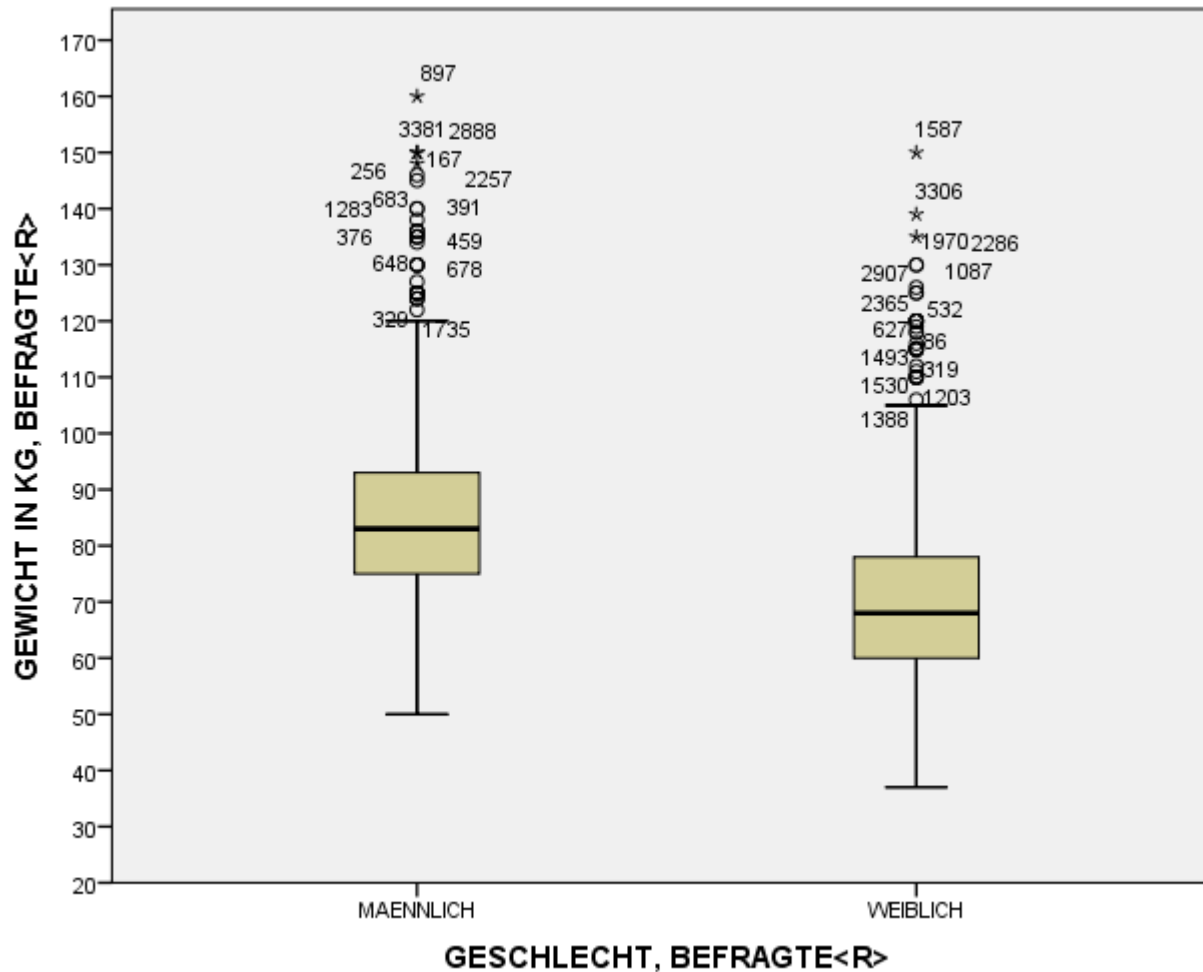
Box-Whisker-Plot: Beispiel V



Box-Whisker-Plot: Anwendungszweck

- geeignet zur schnellen Darstellung zentraler Merkmale einer Verteilung:
 - Interquartilabstans
 - 25-%- und 75-%-Quantil
 - Median
 - kleinster und größter Wert in den Daten, die keine Extremwerte oder Ausreißer sind
 - kleinster und größter Wert generell in Daten (Spannweite ablesbar)
- besonders geeignet, um schnell Verteilungen von metrischen Merkmalen zu vergleichen

Quiz III



Quiz: Was lässt sich anhand des Boxplots über Körpergewicht von Männern und Frauen aussagen? Inwiefern unterscheiden sich beide voneinander?

Aufgabe 3: Bachelorarbeiten

Bei einem Methodendozenten wurden in den letzten zwei Jahren neun Bachelorarbeiten abgegeben, wobei hierfür nachfolgende Seitenzahlen ermittelt wurden:

20, 36, 44, 48, 50, 40, 55, 40, 85

- Bestimmen Sie das 25-%- und 75-%-Quantil sowie Median und Interquartilsabstand.
- Zeichnen Sie den entsprechenden Boxplot!

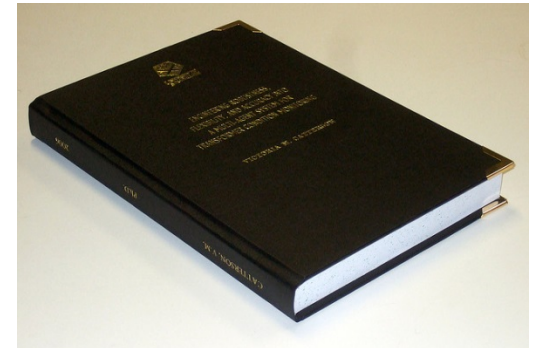


Bild: flickr.com (User:Victoria Catterson Lizenz: CC-BY-2.0) Abrufbar unter <https://flic.kr/p/xn1VC>.

Aufgabe 3a: Lösung

(i)	1	2	3	4	5	6	7	8	9
$x_{(i)}$	20	36	40	40	44	48	50	55	85

- Median:
 - für ungerade Fallzahl: $\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{9+1}{2}\right)} = x_{(5)} = 44$
- 75-%-Quantil:
 - $k = 9 * 0,75 = 6,75$
 - $x_{(7)} = 50$

Aufgabe 3a: Lösung II

(i)	1	2	3	4	5	6	7	8	9
$x_{(i)}$	20	36	40	40	44	48	50	55	85

- 25-%-Quantil:
 - $k = 9 * 0,25 = 2,25 \rightarrow 3$
 - $x_{(3)} = 40$
- IQR:
 - $IQR = Q_{0,75} - Q_{0,25}$
 - $IQR = 50 - 40 = 10$

Aufgabe 3b: Lösung

(i)	1	2	3	4	5	6	7	8	9
$x_{(i)}$	20	36	40	40	44	48	50	55	85

- Länge des Barthaares nach oben:
 - höchster tatsächlicher Wert bis $(Q_{0,75} + 1,5 * IQR)$
 - $50 + 1,5 * 10 = 65$
 - letzter Wert unterhalb von 65 \rightarrow 55
- Länge des Barthaares nach unten:
 - niedrigster tatsächlicher Wert bis $(Q_{0,25} - 1,5 * IQR)$
 - $40 - 1,5 * 10 = 25$
 - letzter Wert oberhalb von 25 \rightarrow 36

Aufgabe 3b: Lösung II

(i)	1	2	3	4	5	6	7	8	9
$x_{(i)}$	20	36	40	40	44	48	50	55	85

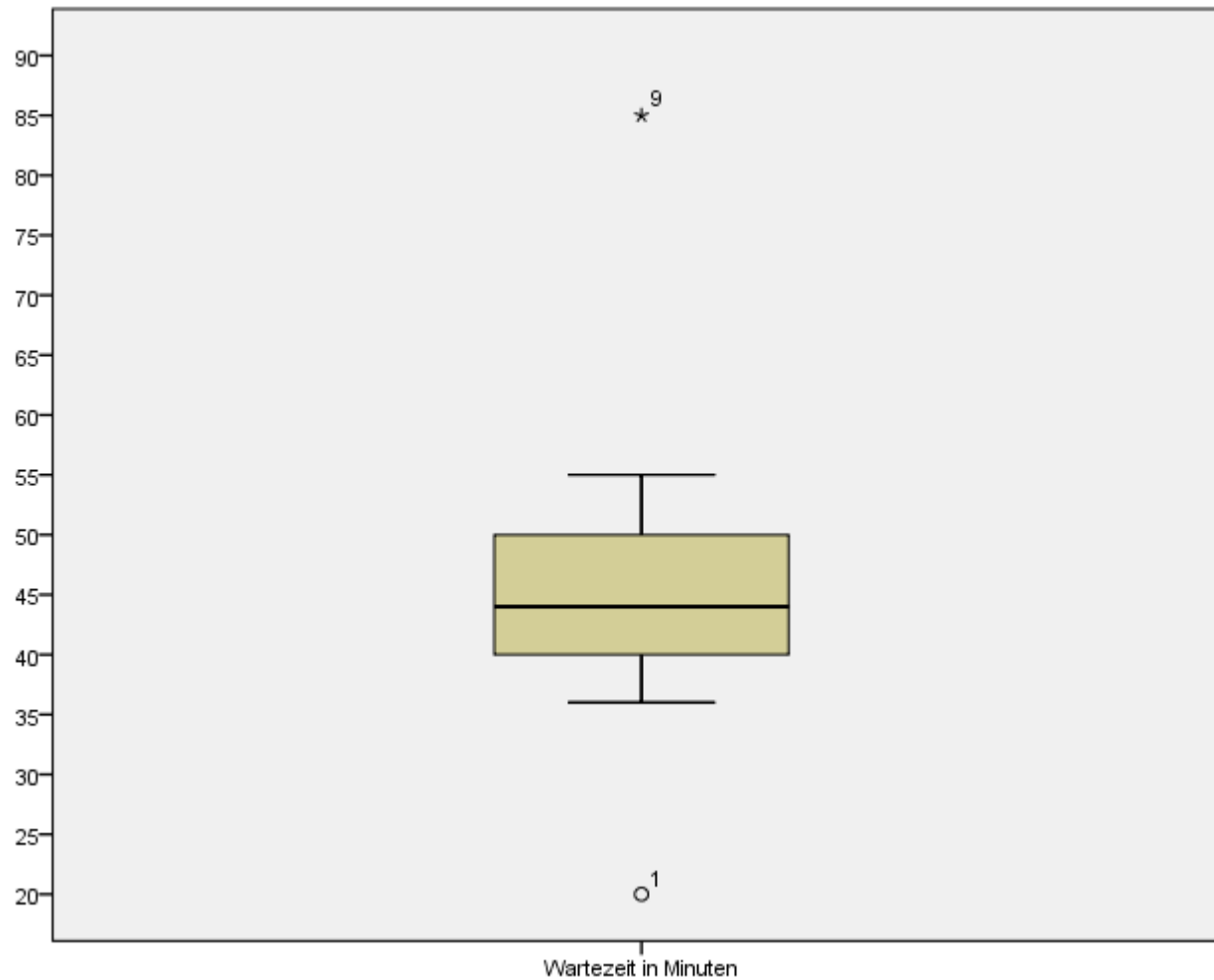
- Ausreißer nach oben:
 - alle Werte im Intervall $[Q_{0,75} + 1,5 * IQR; Q_{0,75} + 3 * IQR]$
 - $50 + 1,5 * 10 = 65$ bis $50 + 3 * 10 = 80$
 - kein Wert zwischen 65 und 80 vorhanden → keine Ausreißer
- Ausreißer nach unten
 - alle Werte im Intervall $[Q_{0,25} - 1,5 * IQR; Q_{0,25} - 3 * IQR]$
 - $40 - 1,5 * 10 = 25$ bis $40 - 3 * 10 = 10$
 - tatsächliche Werte zwischen 10 und 25: 20 → Ausreißer bei 20 (Kreis)

Aufgabe 3b: Lösung III

(i)	1	2	3	4	5	6	7	8	9
$x_{(i)}$	20	36	40	40	44	48	50	55	85

- Extremwerte nach oben:
 - $> (Q_{0,75} + 3 * IQR)$
 - $ab\ 50 + 3 * 10 = 80$
 - tatsächliche Werte 80: 85 \rightarrow Extremwert bei 85 (Stern)
- Extremwerte nach unten
 - $< (Q_{0,25} - 3 * IQR)$
 - kleiner $40 - 3 * 10 = 10$
 - Werte kleiner 10: keine \rightarrow kein Extremwert nach unten

Aufgabe 3b: Lösung III



Vom Fragebogen zur Kreuztabelle I

Wie zufrieden sind Sie mit Ihrer eigenen wirtschaftlichen Lage?

1 unzufrieden

2 zufrieden

99 Antwort verweigert

Haben Sie im letzten Jahr an einer Demonstration teilgenommen?

1 nein

2 ja

99 Antwort verweigert

Vom Fragebogen zur Kreuztabelle II

Person	1	2	3	4	5	6	7	8	9	10	11	12
X: Zufriedenheit wirtschaftliche Lage	1	1	1	1	2	1	2	2	2	99	2	2
Y: Demonstrationsteilnahme	1	1	1	1	99	2	1	1	1	1	2	2

X: 1 „unzufrieden“, 2 „zufrieden“, 99 „Antwort verweigert“

Y: 1 „nein“, 2 „ja“, 99 „Antwort verweigert“

Wie kommen wir zur Kreuztabelle?

1. Auszählen aller Kombinationen, wo die Untersuchungseinheiten bei beiden Variablen gültige Werte aufweisen
2. Ergänzung um Zeilen- und Spaltensummen.

Kreuztabelle: Beispiel

		Zufriedenheit wirtschaftliche Lage		Gesamt
		1 unzufrieden	2 zufrieden	
Demonstrationsteilnahme	1 nein	4	3	7
	2 ja	1	2	3
Gesamt		5	5	10

- 4 Personen sind unzufrieden mit ihrer wirtschaftlichen Lage und haben im letzten Jahr an keiner Demonstration teilgenommen. ($X = 1, Y = 1$)
- 1 Person ist unzufrieden, hat sich jedoch an einer Demonstration beteiligt. ($X = 1, Y = 2$)
- 3 Personen sind zufrieden und haben im letzten Jahr nicht demonstriert. ($X = 2, Y = 1$)
- 2 Personen sind sowohl zufrieden mit der Wirtschaftslage und Demonstrationsteilnehmer gewesen. ($X = 2, Y = 2$)

Kreuztabelle: Beispiel II

		Zufriedenheit wirtschaftliche Lage		Gesamt
		1 unzufrieden	2 zufrieden	
Demonstrationsteilnahme	1 nein	4	3	7
	2 ja	1	2	3
Gesamt		5	5	10

7 Personen haben beide Fragen mit gültigen Werten beantwortet und an keiner Demonstration teilgenommen

3 Personen haben beide Fragen mit gültigen Werten beantwortet an und einer Demonstration teilgenommen

10 Personen haben beide Fragen mit gültigen Werten beantwortet

5 Personen haben beide Fragen mit gültigen Werten beantwortet und waren unzufrieden mit der wirtschaftlichen Lage

5 Personen haben beide Fragen mit gültigen Werten beantwortet und waren zufrieden mit der wirtschaftlichen Lage

Kreuztabelle: allgemeine Konvention

Kreuztabelle mit I Zeilen und J Spalten		Spaltenvariable X				Zeilen- summen
		X = 1	X = 2	X = 3	X = j	
Zeilen- variable Y	Y = 1	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,j}$	$n_{1,+}$
	Y = 2	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,j}$	$n_{2,+}$
	Y = i	$n_{i,1}$	$n_{i,2}$	$n_{i,3}$	$n_{i,j}$	$n_{i,+}$
Spaltensummen		$n_{+,1}$	$n_{+,2}$	$n_{+,3}$	$n_{+,j}$	$n = n_{+,+}$

gemeinsame Verteilungen

Randverteilungen

- i: Index der Zeile
- j: Index der Spalte
- $n_{i,j}$: absolute Häufigkeit der Ausprägung in der i-ten Zeile und j-ten Spalte
- $n_{i,+}$: Summe der i-ten Zeile
- $n_{+,j}$: Summe der j-ten Spalte

Aufgabe 4: Musikunterricht

Im ALLBUS 2014 wurden die Befragten in Ost- und Westdeutschland gebeten anzugeben, ob sie jemals privaten Musikunterricht erhalten haben. Dabei entstand folgende Kreuztabelle:

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	
	NEIN	1519	813	
Gesamtsumme				

Ergänzen Sie die fehlenden Randverteilungen.

Aufgabe 4: Lösung

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	1138
	NEIN	1519	813	2332
Gesamtsumme		2361	1109	3470

Quiz

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	1138
	NEIN	1519	813	2332
Gesamtsumme		2361	1109	3470

Was lässt sich mithilfe der Tabelle aussagen?
Lässt sich auf einfache Weise ein Vergleich zwischen Ost- und Westdeutschland vornehmen? Warum (nicht)?

Quiz: Lösung

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	1138
	NEIN	1519	813	2332
Gesamtsumme		2361	1109	3470

Problem der unterschiedlichen Fallzahl in beiden Gruppen

Lösungsansatz: Prozentuierung

Kreuztabelle: Prozentuierungen

- Spaltenprozentente:
 - Basis: Spaltensumme
 - $p_{i|j}\% = \frac{n_{i,j}}{n_{+,j}} * 100$
- Zeilenprozentente:
 - Basis: Zeilensumme
 - $p_{j|i}\% = \frac{n_{i,j}}{n_{i,+}} * 100$
- Gesamtprozentente:
 - Basis: Gesamtsumme
 - $p_{i,j}\% = \frac{n_{i,j}}{n} * 100$

Legende:

- $n_{i,j}$: absolute Häufigkeit der in Zeile i und Spalte j stehenden Merkmalskombination
- $n_{+,j}$: Summe der Spalte j
- $n_{i,+}$: Summe der Zeile i
- n : Anzahl der Fälle, bei denen beide Untersuchungseinheiten gültige Werte aufweisen

Spaltenprozent I

Fragestellung: Wie viel Prozent der Westdeutschen bzw. Ostdeutschen (X-Variablen) hatten privaten Musikunterricht (Y-Variablen)?

		ERHEBUNGSGEBIET <WOHN- & GEBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	1138
	NEIN	1519	813	2332
Gesamtsumme		2361	1109	3470

Bezugsgröße: Spaltensumme

Spaltenprozent II

		Erhebungsgebiet		
		alte Bundesländer	neue Bundesländer	Gesamt
Privaten Musikunterricht	ja	$\frac{842}{2361} * 100$	$\frac{296}{1109} * 100$	$\frac{1138}{3470} * 100$
	nein	$\frac{1519}{2361} * 100$	$\frac{813}{1109} * 100$	$\frac{2332}{3470} * 100$
Gesamt		2361(100%)	1109 (100%)	3470(100%)

Berechnen Sie die Spaltenprozent!

Spaltenprozent III

% in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	35,7%	26,7%	32,8%
	NEIN	64,3%	73,3%	67,2%
Gesamtsumme		100,0%	100,0%	100,0%

35,17% der Befragten aus den alten Bundesländern hatten privaten Musikunterricht, während es in den neuen Bundesländern nur 26,7% der Befragten waren.

Quiz

% in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	35,7%	26,7%	32,8%
	NEIN	64,3%	73,3%	67,2%
Gesamtsumme		100,0%	100,0%	100,0%

Wie viel Prozent der befragten Ostdeutschen hatten keinen privaten Musikunterricht?

Zeilenprozent

Fragestellung: Wie viel Prozent der Befragten, die die privaten Musikunterricht hatten (Y-Variable), kommen aus den neuen Bundesländern (X-Variable)?

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	1138
	NEIN	1519	813	2332
Gesamtsumme		2361	1109	3470

Bezugsgröße: Zeilensummen

Zeilenprozent II

		Erhebungsgebiet		Gesamt
		alte Bundesländer	neue Bundesländer	
Privaten Musikunterricht	ja	$\frac{842}{1138} * 100$	$\frac{296}{1138} * 100$	1138(100%)
	nein	$\frac{1519}{2332} * 100$	$\frac{813}{2332} * 100$	2332(100%)
Gesamt		$\frac{2361}{3470} * 100$	$\frac{1109}{3470} * 100$	3470(100%)

Berechnen Sie die Zeilenprozent!

Zeilenprozent III

% in PRIVATER MUSIKUNTERRICHT (AUCH GESANG)

		ERHEBUNGSGEBIET <WOHNGBEIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	74,0%	26,0%	100,0%
	NEIN	65,1%	34,9%	100,0%
Gesamtsumme		68,0%	32,0%	100,0%

Quiz: Wie viel Prozent derer, die privaten Musikunterricht hatten, kamen aus den alten Bundesländern?

Gesamtprozente

Fragestellung: Wie viel Prozent **aller Befragten** kommen aus den alten Bundesländern **und** hatten privaten Musikunterricht?

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	842	296	1138
	NEIN	1519	813	2332
Gesamtsumme		2361	1109	3470

Bezugsgröße: Gesamtsumme

Gesamtprozente II

		Erhebungsgebiet		Gesamt
		alte Bundesländer	neue Bundesländer	
Privaten Musikunterricht	ja	$\frac{842}{3470} * 100$	$\frac{296}{3470} * 100$	$\frac{1138}{3470} * 100$
	nein	$\frac{1519}{3470} * 100$	$\frac{813}{3470} * 100$	$\frac{2332}{3470} * 100$
Gesamt		$\frac{2361}{3470} * 100$	$\frac{1109}{3470} * 100$	3470(100%)

Berechnen Sie die Gesamtprozente!

Gesamtprozente III

% des Gesamtergebnisses

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
PRIVATER MUSIKUNTERRICHT (AUCH GESANG)	JA	24,3%	8,5%	32,8%
	NEIN	43,8%	23,4%	67,2%
Gesamtsumme		68,0%	32,0%	100,0%

Quiz: Wie viel Prozent der Befragten kamen aus den alten Bundesländern und hatten privaten Musikunterricht?

Übungsaufgabe 1: Reisende

15 Reisende wurden befragt, wie lange Sie am Hauptbahnhof Halle auf ihren Anschlusszug gewartet haben. Dabei wurde folgende Häufigkeitstabelle erzielt:

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- Bestimmen Sie Spannweite und Interquartilsabstand sowie den Median.
- Interpretieren Sie den Median sowie das 1. Quartil und 3. Quartil inhaltlich.
- Zeichnen Sie für die vorliegende Verteilung einen Boxplot!

Übungsaufgabe 1a: Lösung

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- **Spannweite:**

- $Range = x_{(n)} - x_{(1)}$
- $Range = 120 \text{ Minuten} - 10 \text{ Minuten}$
- $Range = 110 \text{ Minuten}$

- **3. Quartil:**

- $k = n * \alpha$
- $k = 15 * 0,75 = 11,25 \rightarrow 12$
- $x_{(12)} = 45$

Übungsaufgabe 1a: Lösung II

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- 1. Quartil:
 - $k = n * \alpha$
 - $k = 15 * 0,25 = 3,75 \rightarrow 4$
 - $x_{(4)} = 15$
- Interquartilsabstand:
 - $IQR = Q_{0,75} - Q_{0,25}$
 - $IQR = 45 - 15 = 30$

Übungsaufgabe 1a: Lösung II

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- Median:
 - $\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{(8)}$
 - $\tilde{x} = 25$

Übungsaufgabe 1b: Lösung

- Median:
 - $\tilde{x} = 25$
 - Der mittlere Befragte hat 25 Minuten auf einen Anschlusszug gewartet.
- 1. Quartil:
 - In mindestens 25 % der Fälle lag die Wartezeit bei höchstens 15 Minuten für einen Anschlusszug.
- 3. Quartil:
 - In mindestens 75 % der Fälle warteten die Reisenden höchstens 45 Minuten auf ihren Anschlusszug.

Übungsaufgabe 1c: Lösung

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- Länge des unteren Barthaars:
 - theoretisch bis mindestens $Q_{0,25} - 1,5 * IQR = 15 - 1,5 * 30 = -30$
 - letzter tatsächlicher Wert in dem Bereich: 10
- Länge des oberen Barthaars:
 - theoretisch bis höchstens $Q_{0,75} + 1,5 * IQR = 45 + 1,5 * 30 = 90$
 - letzter tatsächlicher Wert in dem Bereich: 80

Übungsaufgabe 1c: Lösung II

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

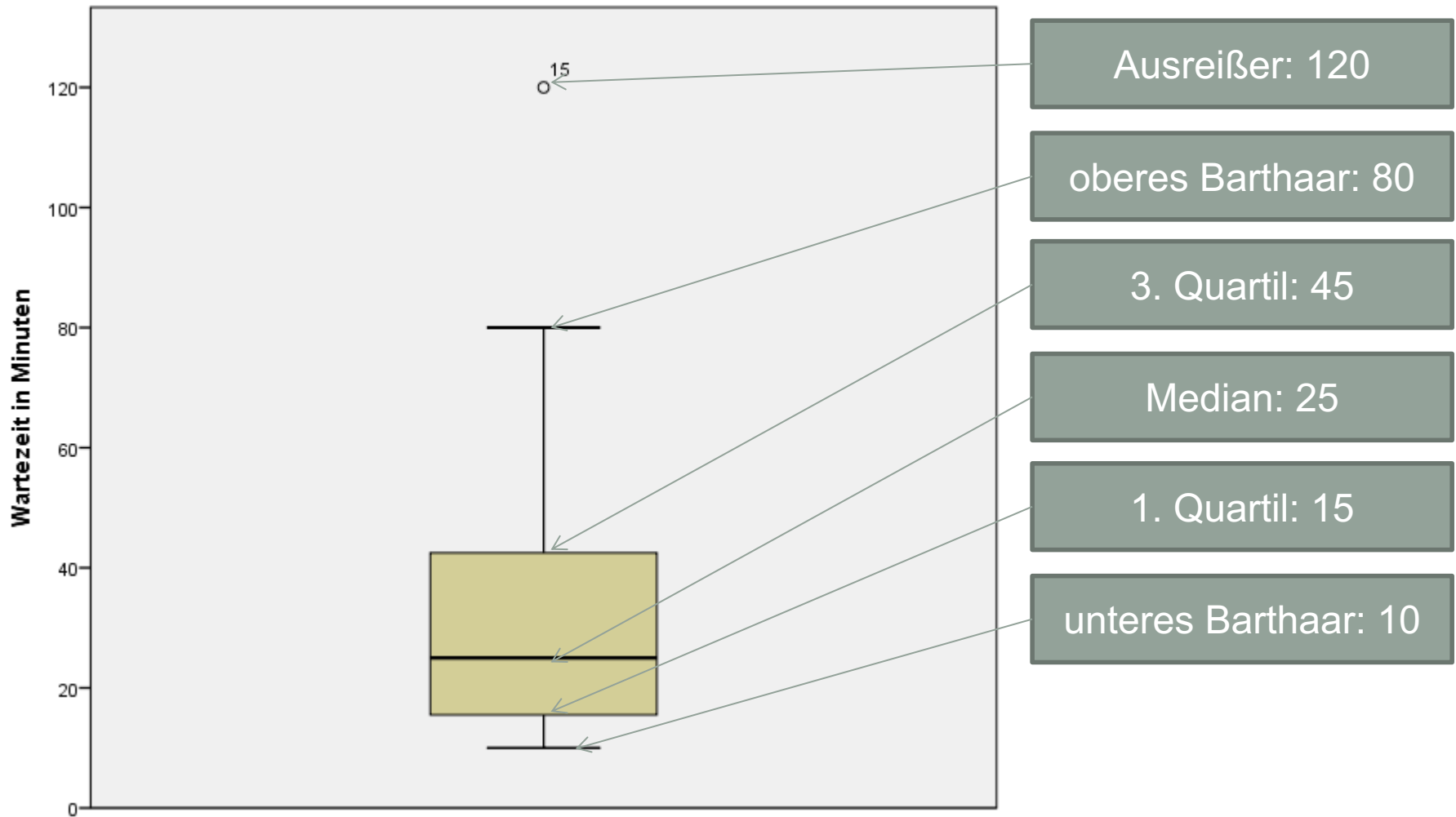
- Ausreißer nach unten:
 - alle Werte im Intervall $[Q_{0,25} - 3 * IQR; Q_{0,25} - 1,5 * IQR]$
 - alle Werte im Intervall von $[-75; -15]$
 - keine Werte \rightarrow keine Ausreißer
- Ausreißer nach oben:
 - alle Werte im Intervall $[Q_{0,75} + 1,5 * IQR; Q_{0,75} + 3 * IQR;]$
 - alle Werte im Intervall von $[75; 135]$
 - Wert in Bereich: 120 \rightarrow Ausreißer bei 120 Minuten

Übungsaufgabe 1c: Lösung III

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$x_{(k)}$	10	12	12	15	16	17	20	25	27	27	40	45	60	80	120

- Extremwerte nach unten:
 - alle Werte $< (Q_{0,25} - 1,5 * IQR)$
 - alle Werte < -75
 - keine Werte \rightarrow keine Extremwerte nach unten
- Extremwerte nach oben:
 - alle Werte $> (Q_{0,75} + 3 * IQR)$
 - alle Werte > 135
 - keine Werte \rightarrow keine Extremwerte nach oben

Übungsaufgabe 1c: Lösung IV



Übungsaufgabe 2: öffentlicher Dienst

Im ALLBUS 2014 wurden die Befragten, die arbeiten gehen, gebeten anzugeben, ob sie im öffentlichen Dienst beschäftigt sind. Mithilfe von SPSS konnte folgende Kreuztabelle erstellt werden:

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
IM ÖFFENTLICHEN DIENST TÄTIG?	JA	315	123	438
	NEIN	870	392	1262
Gesamtsumme		1185	515	1700

- Bestimmen Sie Spalten-, Zeilen- und Gesamtprozent!
- Wie viel Prozent der Befragten, die aus den alten Bundesländern kommen, sind im öffentlichen Dienst tätig?
- Wie viel Prozent der Befragten, die im öffentlichen Dienst tätig sind, kommen aus den neuen Bundesländern?
- Wie viel Prozent aller Befragten mit gültigen Werten kommen aus Westdeutschland und sind nicht im öffentlichen Dienst tätig?

Übungsaufgabe 2a: Lösung (Spaltenprozent)

% in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
IM OEFFENTLICHEN DIENST TAETIG?	JA	26,6%	23,9%	25,8%
	NEIN	73,4%	76,1%	74,2%
Gesamtsumme		100,0%	100,0%	100,0%

Übungsaufgabe 2a: Lösung (Zeilenprozent)

% in IM OEFFENTLICHEN DIENST TAETIG?

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLAE NDER	NEUE BUNDESLAE NDER	
IM OEFFENTLICHEN DIENST TAETIG?	JA	71,9%	28,1%	100,0%
	NEIN	68,9%	31,1%	100,0%
Gesamtsumme		69,7%	30,3%	100,0%

Übungsaufgabe 2a: Lösung (Gesamtprozent)

% des Gesamtergebnisses

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
IM OEFFENTLICHEN DIENST TAETIG?	JA	18,5%	7,2%	25,8%
	NEIN	51,2%	23,1%	74,2%
Gesamtsumme		69,7%	30,3%	100,0%

Übungsaufgabe 2b: Lösung

% in ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
IM OEFFENTLICHEN DIENST TAETIG?	JA	26,6%	23,9%	25,8%
	NEIN	73,4%	76,1%	74,2%
Gesamtsumme		100,0%	100,0%	100,0%

Spaltenprozent: 26,6% der Befragten, die aus den alten Bundesländern kommen, sind im öffentlichen Dienst tätig.

Übungsaufgabe 2c: Lösung

% in IM OEFFENTLICHEN DIENST TAETIG?

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLAE NDER	NEUE BUNDESLAE NDER	
IM OEFFENTLICHEN DIENST TAETIG?	JA	71,9%	28,1%	100,0%
	NEIN	68,9%	31,1%	100,0%
Gesamtsumme		69,7%	30,3%	100,0%

Zeilenprozent: Von den Befragten, die im öffentlichen Dienst tätig sind, kommen 28,1 % aus den neuen Bundesländern.

Übungsaufgabe 2d: Lösung (Gesamtprozent)

% des Gesamtergebnisses

		ERHEBUNGSGEBIET <WOHNGBIET>: WEST - OST		Gesamtsumme
		ALTE BUNDESLÄNDER	NEUE BUNDESLÄNDER	
IM OEFFENTLICHEN DIENST TAETIG?	JA	18,5%	7,2%	25,8%
	NEIN	51,2%	23,1%	74,2%
Gesamtsumme		69,7%	30,3%	100,0%

Gesamtprozent: Von allen Befragten mit gültigen Werten kommen 51,2% aus Westdeutschland und sind nicht im öffentlichen Dienst tätig.

Übungsaufgabe 3: Private Internetnutzung

Im ALLBUS 2012 wurden die Befragten gebeten anzugeben, ob sie das Internet auch privat nutzen. Im Folgenden ist für den Geschlechterunterschied eine Kreuztabelle wiedergegeben:

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	1303	1222	2525
	NEIN	422	533	955
Gesamt		1725	1755	3480

- Berechnen Sie Spalten-, Zeilen- und Gesamtprozent.
- Wie hoch ist der prozentuale Anteil derer, die das Internet privat nutzen unter den Männern?
- Wie viel Prozent sind Frauen unter den privaten Internetnutzern?
- Wie viel Prozent sind Männer und nutzen privat das Internet nicht?

Übungsaufgabe 3a: Lösung Spaltenprozent

NUTZEN SIE PRIVAT DAS INTERNET? * GESCHLECHT, BEFRAGTE<R> Kreuztabelle

% innerhalb von GESCHLECHT, BEFRAGTE<R>

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	75,5%	69,6%	72,6%
	NEIN	24,5%	30,4%	27,4%
Gesamt		100,0%	100,0%	100,0%

Übungsaufgabe 3a: Lösung Zeilenprozent

NUTZEN SIE PRIVAT DAS INTERNET? * GESCHLECHT, BEFRAGTE<R> Kreuztabelle

% innerhalb von NUTZEN SIE PRIVAT DAS INTERNET?

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	51,6%	48,4%	100,0%
	NEIN	44,2%	55,8%	100,0%
Gesamt		49,6%	50,4%	100,0%

Übungsaufgabe 3a:Lösung Gesamtprozente

NUTZEN SIE PRIVAT DAS INTERNET? * GESCHLECHT, BEFRAGTE<R> Kreuztabelle

% der Gesamtzahl

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	37,4%	35,1%	72,6%
	NEIN	12,1%	15,3%	27,4%
Gesamt		49,6%	50,4%	100,0%

Übungsaufgabe 3b: Lösung

- Spaltenprozentage sind gesucht:

NUTZEN SIE PRIVAT DAS INTERNET? ^ GESCHLECHT, BEFRAGTE<R>
Kreuztabelle

% innerhalb von GESCHLECHT, BEFRAGTE<R>

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	75,5%	69,6%	72,6%
	NEIN	24,5%	30,4%	27,4%
Gesamt		100,0%	100,0%	100,0%

Übungsaufgabe 3c: Lösung

Zeilenprozentage waren gesucht.

NUTZEN SIE PRIVAT DAS INTERNET? * GESCHLECHT, BEFRAGTE<R> Kreuztabelle

% innerhalb von NUTZEN SIE PRIVAT DAS INTERNET?

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	51,6%	48,4%	100,0%
	NEIN	44,2%	55,8%	100,0%
Gesamt		49,6%	50,4%	100,0%

Übungsaufgabe 3d: Lösung

Gesamtprozente waren gesucht.

NUTZEN SIE PRIVAT DAS INTERNET? * GESCHLECHT, BEFRAGTE<R> Kreuztabelle

% der Gesamtzahl

		GESCHLECHT, BEFRAGTE<R>		Gesamt
		MANN	FRAU	
NUTZEN SIE PRIVAT DAS INTERNET?	JA	37,4%	35,1%	72,6%
	NEIN	12,1%	15,3%	27,4%
Gesamt		49,6%	50,4%	100,0%

Übungsaufgabe 4: Altersverteilung

Für die Altersverteilung einer Stichprobe von 15 Befragten an einer sozialwissenschaftlichen Studie wurde folgende Häufigkeitstabelle erstellt.

ALTER: BEFRAGTE<R>

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 20	1	6,7	6,7	6,7
21	1	6,7	6,7	13,3
24	1	6,7	6,7	20,0
25	1	6,7	6,7	26,7
30	2	13,3	13,3	40,0
37	1	6,7	6,7	46,7
38	1	6,7	6,7	53,3
43	1	6,7	6,7	60,0
46	2	13,3	13,3	73,3
50	1	6,7	6,7	80,0
61	1	6,7	6,7	86,7
75	1	6,7	6,7	93,3
86	1	6,7	6,7	100,0
Gesamt	15	100,0	100,0	

- Bestimmen Sie die benötigten Maßzahlen für einen Boxplot.
- Zeichnen Sie einen Boxplot für die Verteilung.

Literaturhinweis

- Kerstin Völkl / Christoph Korb (2018): Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler. S. 116-149.