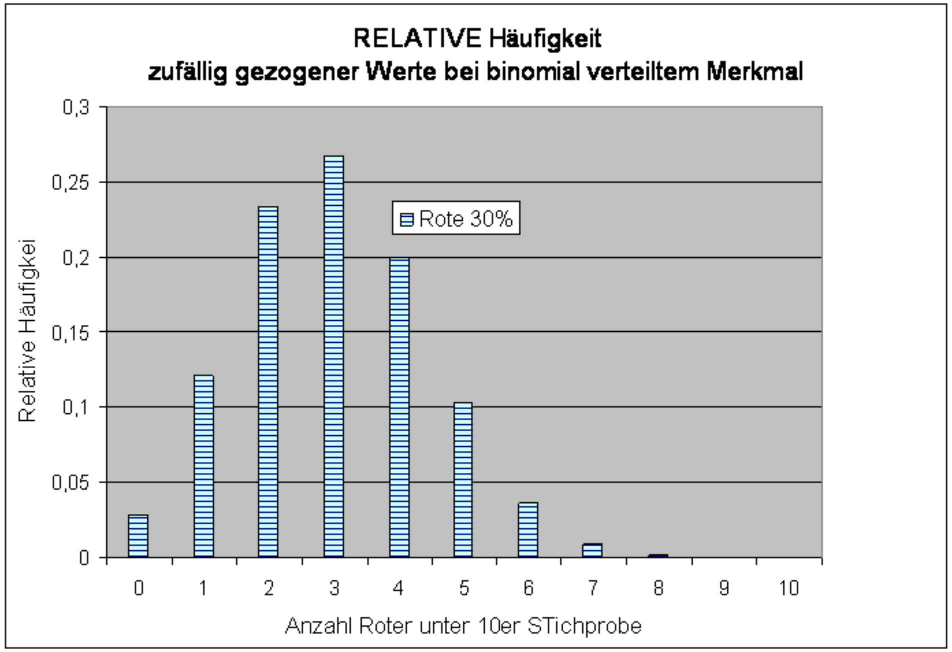
Erläutern Sie anhand des „Säckchen-Experiments“

## Grundgesamtheit - Stichprobe



# Grundgesamtheit - Stichprobe

Wie gewinnt man eine „qualitätsgerechte“ Stichprobe?

Es gibt zwei grundsätzlich verschiedene Wege zu repräsentativen Daten:

- **Zufällige und faire Auswahl** aus der Grundgesamtheit.
- **Unabhängige Wiederholung** identischer Einzelversuche.

Diskutieren Sie anhand des „Säckchen-Experiments“

## Datentypen, Daten-Skala

# Welche der folgenden Aussagen treffen zu?

- Die Skala bestimmt welche weiteren Verfahren angewendet werden sollten.
- Die Skala gibt Hinweise was in der weiteren Analyse beachtet werden sollte.
- Die Skala bestimmt, wie die Daten zusammengefaßt und beschrieben werden können.
- Die Bestimmung der Skala der Variablen ist daher der erste Schritt jeder Datenanalyse.

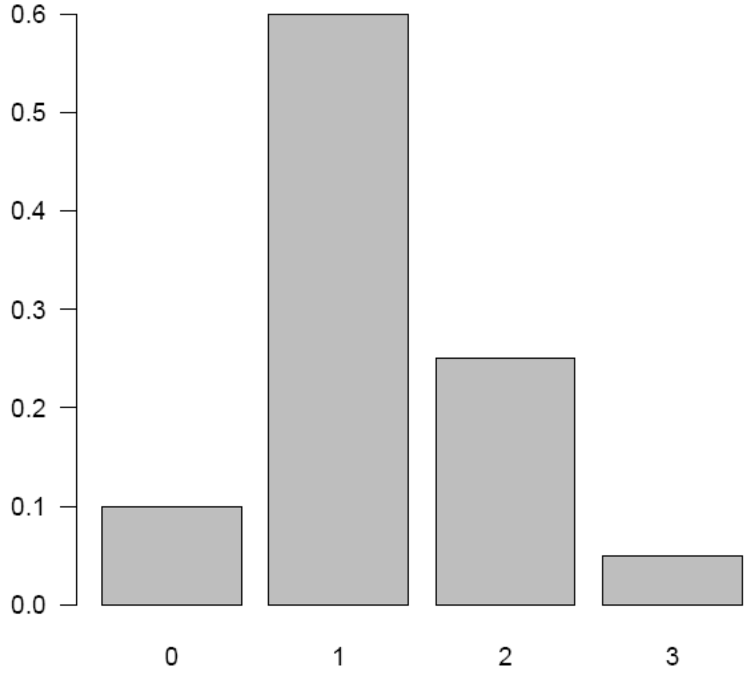
Aufgabe: Häufigkeiten

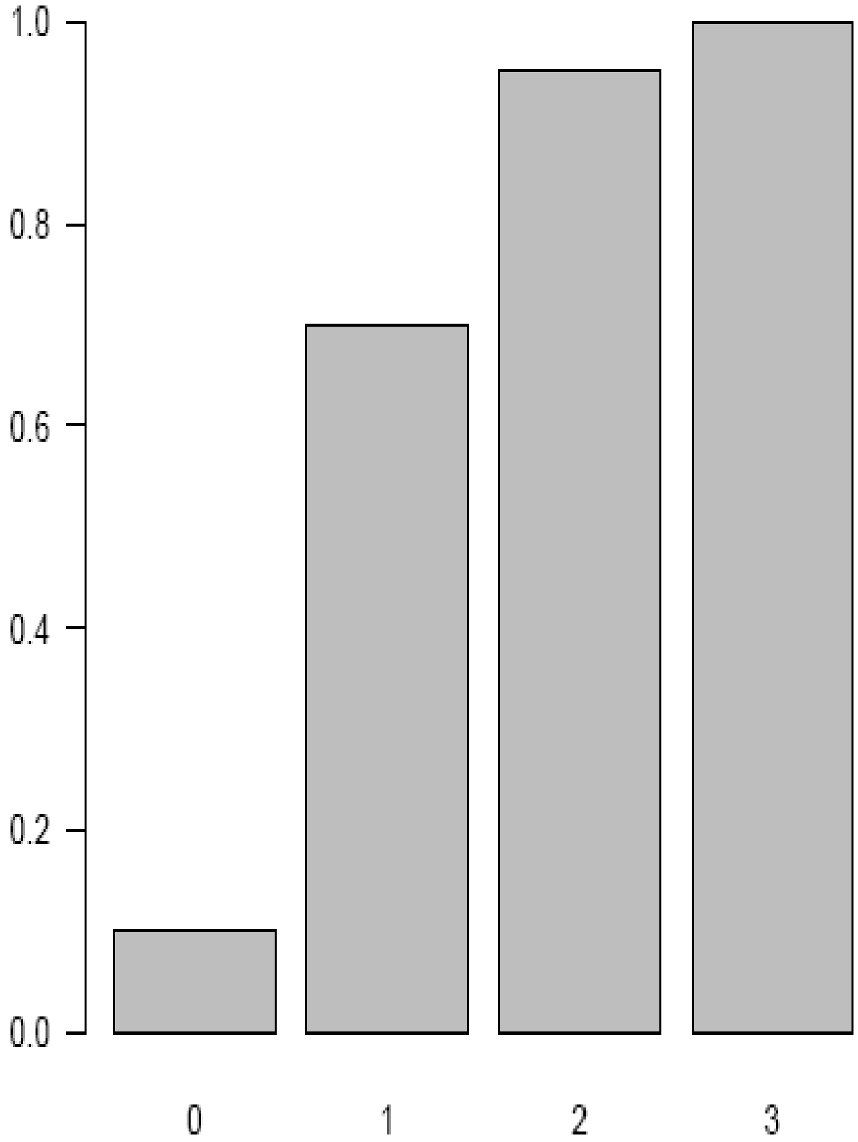
| Anzahl Geschwister | absolute Häufigkeit | relative Häufigkeit | kumulierte relative Häufigkeit |
|---|---|---|---|
| 0 | 2 | | |
| 1 | 12 | | |
| 2 | 5 | | |
| 3 | 1 | | |

$\Sigma = 20$

Fertigen Sie passende grafische Skizzen an.

## Aufgabe: Häufigkeiten



In einer Untersuchung sollte die medizinische Versorgung von Patienten mit Infektionskrankheiten überprüft werden.

| **Nr.** | Aufenthalt (Tage) | Alter | Geschlecht | Medikament | **Nr.** | Aufenthalt (Tage) | Alter | Geschlecht | Medikament |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 5 | 30 | w | n | **14** | 8 | 33 | w | j |
| **2** | 10 | 73 | w | n | **15** | 5 | 20 | w | n |
| **3** | 6 | 40 | w | n | **16** | 5 | 32 | m | n |
| **4** | 11 | 47 | w | n | **17** | 7 | 36 | m | j |
| **5** | 5 | 25 | w | n | **18** | 4 | 69 | m | n |
| **6** | 14 | 82 | m | j | **19** | 3 | 47 | m | j |
| **7** | 30 | 60 | m | j | **20** | 7 | 22 | m | n |
| **8** | 11 | 56 | w | n | **21** | 9 | 11 | m | n |
| **9** | 17 | 43 | w | n | **22** | 11 | 19 | m | j |
| **10** | 3 | 50 | m | n | **23** | 11 | 67 | w | n |
| **11** | 9 | 59 | w | n | **24** | 9 | 43 | w | n |
| **12** | 3 | 4 | m | n | **25** | 4 | 41 | w | n |
| **13** | 8 | 22 | w | j | **28** | 5 | 39 | m | j |
| **26** | 10 | 45 | W | J | **29** | 14 | 52 | w | n |
| **27** | 20 | 48 | m | n | **30** | 3 | 27 | m | j |

# Box-Whisker-Plot

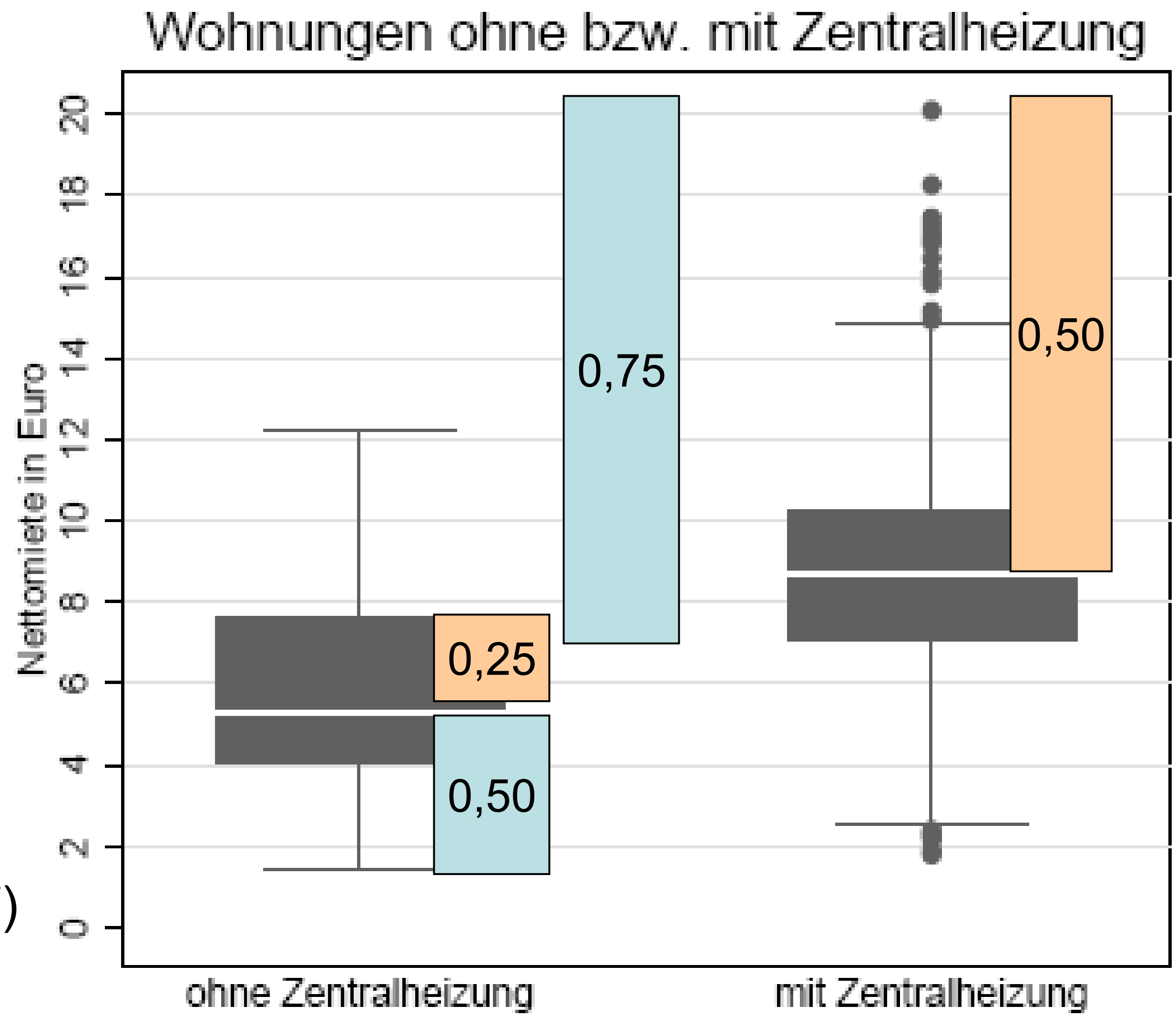
## Lesen des Diagramms

**Gehirnjogging**:

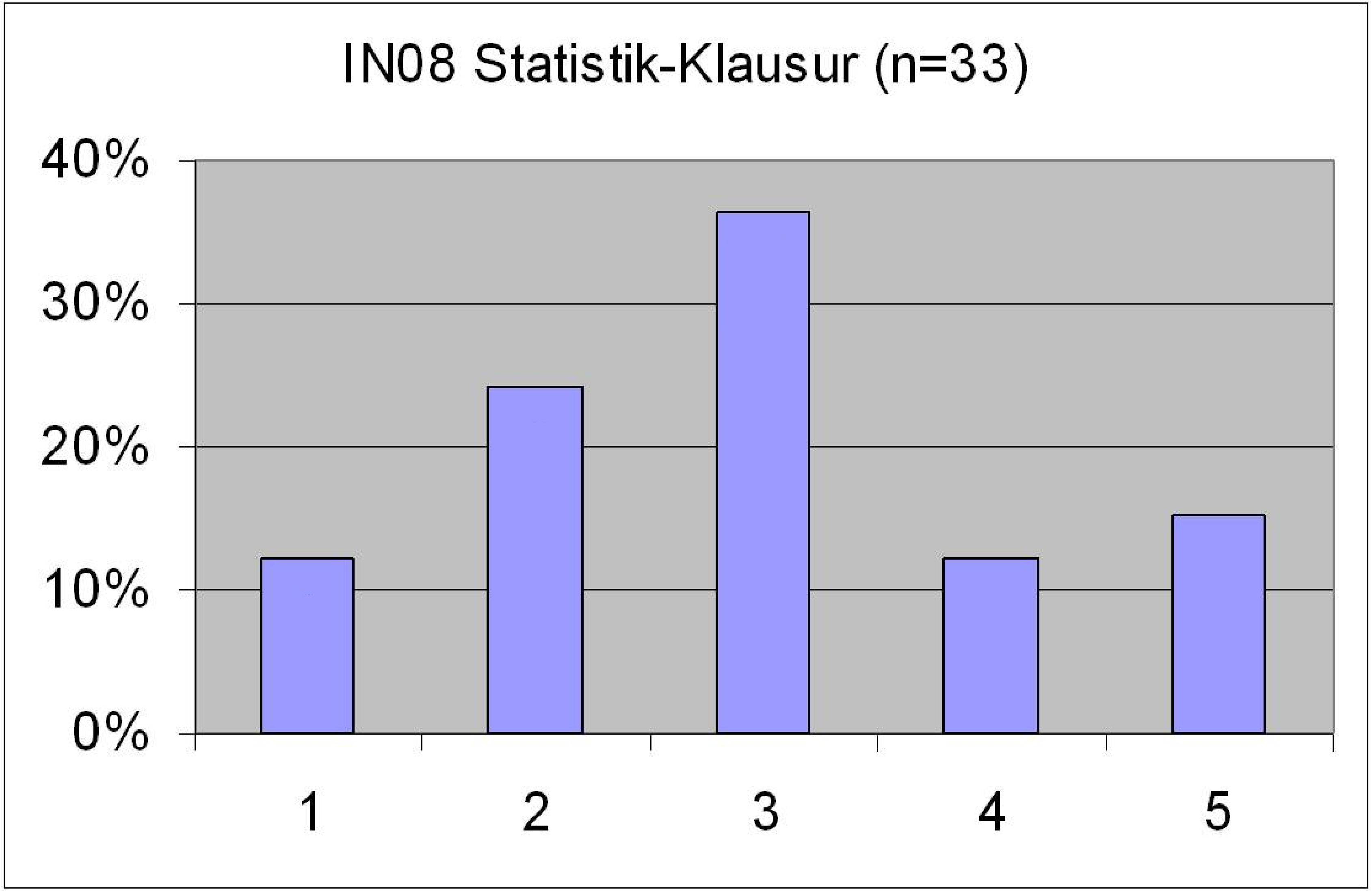
Wieso ist die Chance bei einer Wohnung ohne Zentralheizung günstiger zu fahren mindestens 50%?

**Paarweise Zufällig oZ & mZ. Was ist möglich?**

(50% oZ<5€ UND 75% mZ>7€) ODER (25% oZ 5€-8€ UND 50% mZ>8€) 50%*75%+25%*50% = 50% (mindestens)

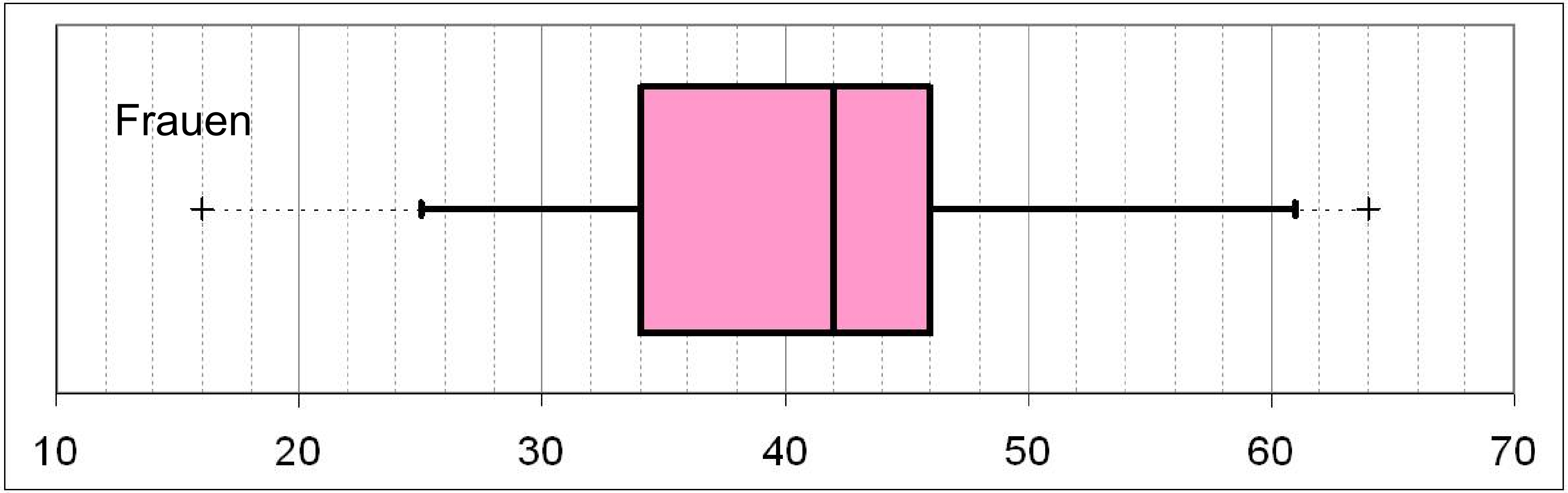


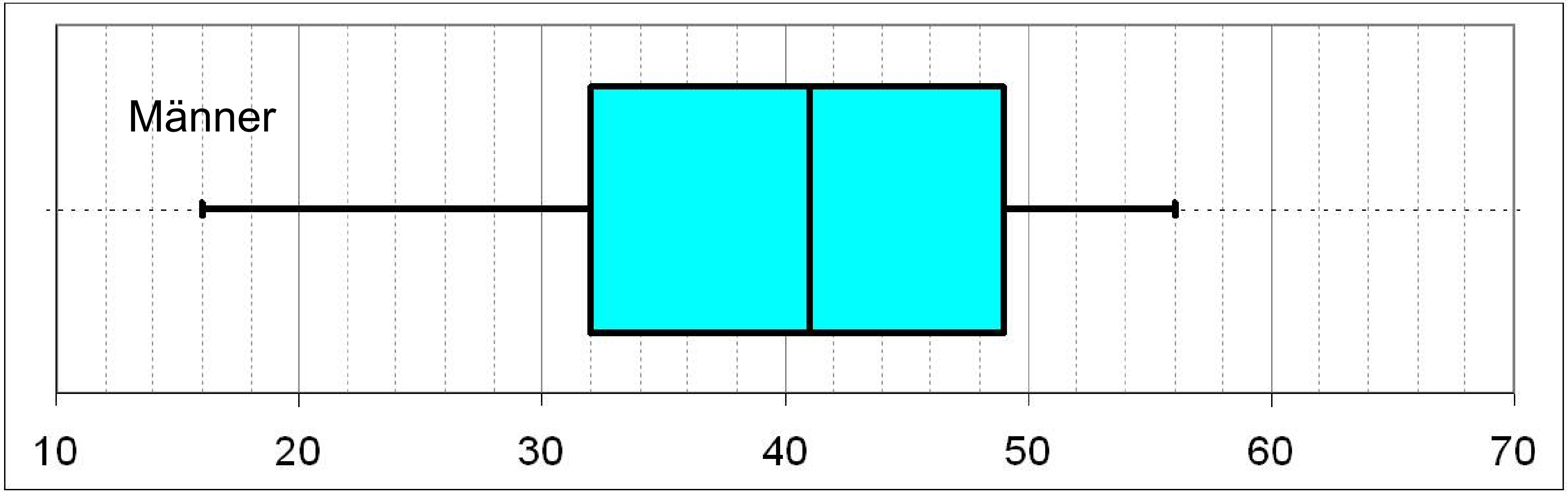
Klausurnoten 2009




Wieviele Studenten erreichten Note 3 und besser?

Klausurpunkte 2009



# Wer bin ich?

$$\frac{1}{n}\sum_{i=1}^{n} x_i$$

$$x_{0.75} - x_{0.25}$$

$$\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n - 1}$$

# Verteilungen

Die Wahrscheinlichkeitsfunktion der Lebensdauer einer gesamten Energiesparlampenserie lässt sich ziemlich genau durch folgende mathematische Formel beschreiben:

$$WS(x) = \frac{1}{2*\pi*10000}\, e^{-\left(\frac{(x-3000)^2}{2*10000}\right)}$$

- Skizzieren Sie die Wahrscheinlichkeitsfunktion.

- Nennen Sie die Theoretische Verteilung

- Ermitteln Sie Mittelwert und Standardabweichung der Gesamtheit aller Lampen

- Bestimmen Sie die Wahrscheinlichkeit mindestens 3000 Stunden Freude an Ihrer Lampe zu haben

- Ermitteln Sie die Wahrscheinlichkeit zwischen 2800 und 3200 Stunden Freude an Ihrer Lampe zu haben

- Berechnen Sie die Wahrscheinlichkeit höchstens 2750 Stunden Freude an Ihrer Lampe zu haben

# Standardisieren für Nutzung von Tabelle 9.1

**Seite 13**

lies: Seien die Zufallswerte X, normalverteilt mit Mittelwert μ und Varianz σ² …

Sei $X \sim N(\mu, \sigma^2)$

- $\frac{X-\mu}{\sigma} \sim N(0,1)$ (Standardisieren einer Zufallsvariable)

( <u>2750</u> – 3000 ) / 100 = **-** 2,5

$$\Phi(-z) = 1 - \Phi(z)$$ Warum?

Φ (-2,5) = 1 - Φ (2,5) ~ 1 – 0,994 = 0,006 = 0,6%

In weniger als 1% aller Käufe, wird die Lampe nicht länger als 2750 h leuchten.

8. Aufgabe [7]

Ein Hersteller garantiert, dass seine Holzbalkenlieferungen im Durchschnitt 300 cm ($\mu 0$) lang sind. Wobei er die Varianz des Längenmaßes mit 4 cm ($\sigma^2$) angibt. Weitere Angaben über die Verteilung des Merkmals macht der Hersteller nicht.

Der Abnehmer einer Lieferung von 1.000 derartigen Balken zweifelt an dieser Aussage und will die Einhaltung dieser Garantie prüfen. Er ermittelt die durchschnittliche Länge von 64 zufällig der Lieferung entnommenen Balken und führt einen geeigneten Hypothesentest durch.

a) Wie lautet die Null- und Alternativhypothese (mathematisch!- nicht in Worten, schreiben Sie $\mu$ für die mittlere Länge der 1.000 Balken)

b) Geben Sie die Formel der Prüfgröße an.

b) Wie heißt die Prüfverteilung?

c) Berechnen Sie mit Genauigkeit von 1 Kommastelle ab welcher aus der Stichprobe ermittelten Länge (oder identisch: ab welcher Abweichung des gemessenen Längenmittels von der Garantievorgabe), der Test zur Ablehnung der von Ihnen aufgestellten Nullhypothese führen würde. Legen Sie die Irrtumswahrscheinlichkeit sinnvoll fest.

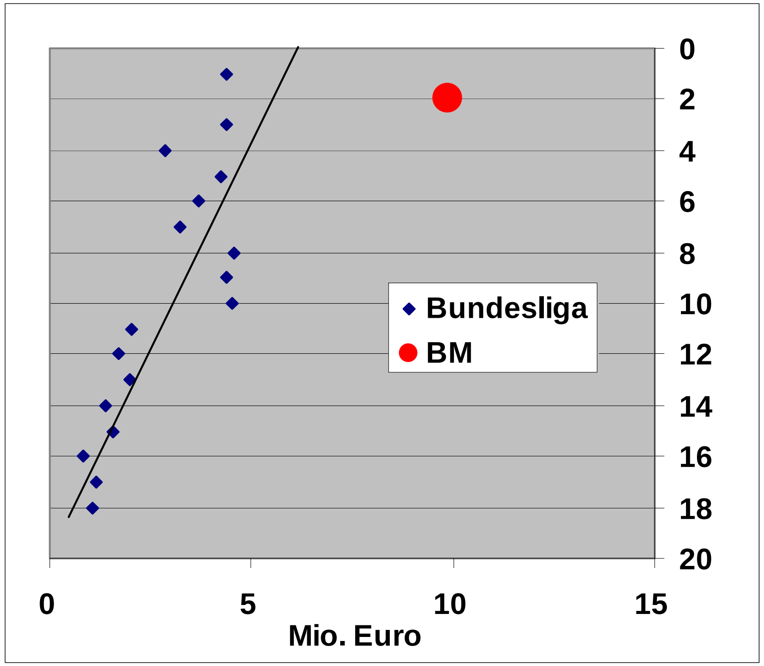
2 Beutel A, B

je 50 Einzelsteine mit zurücklegen

1. Konfidenzintervall

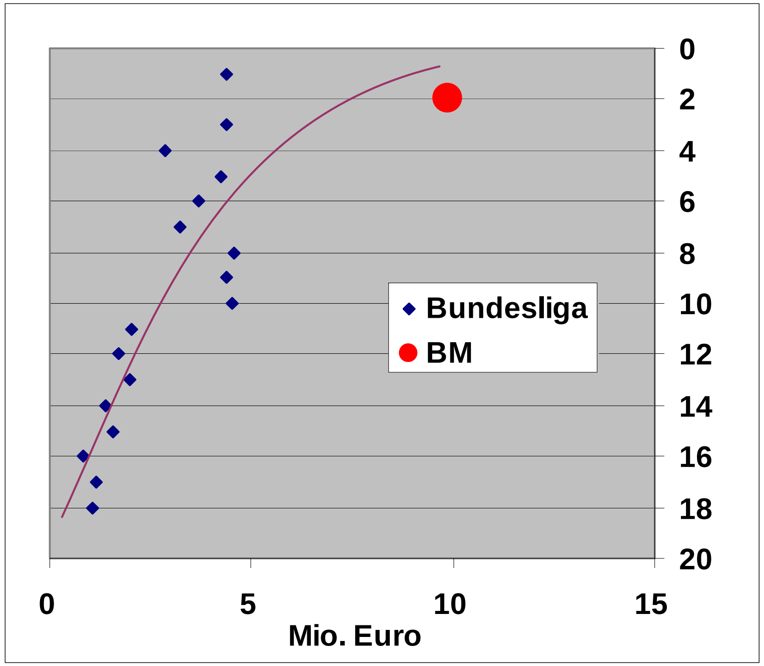
2. Test einer Vorgabe

| Team | Kohle | Platz |
| --- | --- | --- |
| Bayern München | 9.863.636 € | 2 |
| Schalke 04 | 4.582.609 € | 8 |
| Werder Bremen | 4.544.531 € | 10 |
| VfB Stuttgart | 4.405.769 € | 3 |
| Bayer Leverkusen | 4.396.154 € | 9 |
| VfL Wolfsburg | 4.377.500 € | 1 |
| Hamburger SV | 4.234.483 € | 5 |
| Borussia Dortmund | 3.713.043 € | 6 |
| TSG Hoffenheim | 3.212.500 € | 7 |
| Hertha BSC | 2.874.194 € | 4 |
| Hannover 96 | 2.008.929 € | 11 |
| Eintracht Frankfurt | 1.993.333 € | 13 |
| 1.FC Köln | 1.691.071 € | 12 |
| Borussia Mglb | 1.586.207 € | 15 |
| VfL Bochum | 1.380.172 € | 14 |
| Karlsruher SC | 1.144.444 € | 17 |
| Arminia Bielefeld | 1.038.889 € | 18 |
| Energie Cottbus | 847.656 € | 16 |




**Korrelation: - 0,77**

| Team | Kohle | Platz | LogKohle |
|---|---|---|---|
| Bayern München | 9,9 | 2 | 2,29 |
| Schalke 04 | 4,6 | 8 | 1,52 |
| Werder Bremen | 4,5 | 10 | 1,51 |
| VfB Stuttgart | 4,4 | 3 | 1,48 |
| Bayer Leverkusen | 4,4 | 9 | 1,48 |
| VfL Wolfsburg | 4,4 | 1 | 1,48 |
| Hamburger SV | 4,2 | 5 | 1,44 |
| Borussia Dortmund | 3,7 | 6 | 1,31 |
| TSG Hoffenheim | 3,2 | 7 | 1,17 |
| Hertha BSC | 2,9 | 4 | 1,06 |
| Hannover 96 | 2,0 | 11 | 0,70 |
| Eintracht Frankfurt | 2,0 | 13 | 0,69 |
| 1.FC Köln | 1,7 | 12 | 0,53 |
| Borussia Mglb | 1,6 | 15 | 0,46 |
| VfL Bochum | 1,4 | 14 | 0,32 |
| Karlsruher SC | 1,1 | 17 | 0,13 |
| Arminia Bielefeld | 1,0 | 18 | 0,04 |
| Energie Cottbus | 0,8 | 16 | -0,17 |




**Korrelation:**

**direkt: - 0,77**

**linearisiert: - 0,88**

**11. Aufgabe [14]**

Die Ausgaben deutscher Urlauber für Reisen nach Spanien betrugen in den Jahren 2005 bis 2008 (in Mill. Euro):

| | (2005) | (2006) | (2007) | (2008) | (2009) |
|---|---|---|---|---|---|
| Jahr | 1 | 2 | 3 | 4 | 5 |
| Umsatz | 156 | 171 | 228 | 246 | ?? |

a) Bestimmen Sie den mittleren Jahresumsatz der vergangenen vier Jahre. [2]

b) Skizzieren Sie die Zeitreihe als Punktediagramm. Skizzieren Sie Ihre Vermutung über den Zusammenhang zwischen Zeit und Umsatz. [2]

c) Unter der Annahme, dass der gefundene Trend anhält, treffen Sie eine Vorhersage über den vermutlichen Jahresumsatz in 2009. Stellen Sie dazu ein geeignetes Regressionsmodell auf. Berechnen Sie aus den Daten die Regressionskoeffizienten. Ermitteln Sie anhand der erhaltenen Trendgeraden den Prognosewert für den Umsatz im Jahr 5. [6]

d) Da der Prognosewert nur ein Punktschätzer ist, geben Sie nun das 95%-Konfidenzintervall zum $\boxed{\text{Prognosewert}}$ ($\hat{Y}_0 = \hat{\alpha} + \hat{\beta} * x_0$) an, wenn noch die Berechnung der Standardabweichung ($\hat{\sigma}$) folgenden Wert ergab: 12,85 (in Mill. Euro). [4]

# Rechenbeispiel

| Kind | W | S | Kind | W | S | Kind | W | S |
|---|---|---|---|---|---|---|---|---|
| 1 | 15,1 | 25,2 | 12 | 14,3 | 24,7 | 23 | 13,6 | 24,3 |
| 2 | 12,7 | 24,3 | 13 | 11,5 | 23,4 | 24 | 15,2 | 26,3 |
| 3 | 11,7 | 22,1 | 14 | 13,4 | 25,7 | 25 | 12,1 | 23,4 |
| 4 | 13,1 | 23,3 | 15 | 13,7 | 24,5 | 26 | 12,6 | 24,5 |
| 5 | 13,0 | 24,1 | 16 | 13,5 | 26,0 | 27 | 14,1 | 26,2 |
| 6 | 11,2 | 23,6 | 17 | 12,8 | 24,6 | 28 | 11,2 | 23,0 |
| 7 | 13,3 | 25,5 | 18 | 13,2 | 25,4 | 29 | 14,0 | 24,3 |
| 8 | 12,3 | 24,3 | 19 | 14,7 | 26,3 | 30 | 13,1 | 25,3 |
| 9 | 13,7 | 25,5 | 20 | 12,2 | 25,2 | 31 | 11,5 | 24,2 |
| 10 | 12,2 | 23,2 | 21 | 14,7 | 26,4 | 32 | 14,9 | 27,2 |
| 11 | 13,3 | 27,1 | 22 | 14,6 | 25,8 | 33 | 13,8 | 26,3 |

Machen Sie sich ein Bild der Daten (Häufigkeiten, Relation…)

Mittelwert + Standardabweichung

W & S

Konfidenzintervall

zu Mwert von S

Zu Varianz von W

# Konfidenzintervall zur Stichprobenvarianz s²

$$X \sim N(\mu, \sigma^2),$$
$$\mu \text{ beliebig}$$

$\sigma^2$ … aus Stichprobe geschätzt durch $S^2$

Wieso??

$$\left[ (n-1)\, S^2 \frac{1}{\chi^2_{1-\frac{\alpha}{2}}(n-1)},\ (n-1)\, S^2 \frac{1}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right]$$

1-α/2 – Quantil der Chi²-Verteilung

n ist Stichprobenumfang

# Korrelation nichtmetrischer Merkmale

| | Limousine | Kombi | Summen |
|---|---|---|---|
| Arbeiter | 19 | 18 | 37 |
| Angestellte | 43 | 20 | 63 |
| Summen | 62 | 38 | 100 |

Berechnen des Kontingenzkoeffizients.

Berechnen der Chi² Testgröße.

$$\frac{(19-\frac{37*62}{100})^2}{\frac{37*62}{100}}+\frac{(18-\frac{37*38}{100})^2}{\frac{37*38}{100}}+\frac{(43-\frac{63*62}{100})^2}{\frac{63*62}{100}}+\frac{(20-\frac{63*38}{100})^2}{\frac{63*38}{100}}$$

$$C = 0{,}166\sqrt{\frac{2}{2-1}} = 0{,}235$$

$$\chi^2 = 2{,}83$$

Test: Chi² $_{(2-1;2-1)}(1-\alpha) = 3{,}84$

Auf einem Automaten werden in sehr großer Anzahl Kaffeepakete zu 500 g Inhalt abgefüllt. Erfahrungsgemäß besitzt das Füllgewicht eine Standardabweichung von 3 g. Durch Stichproben vom Unfang 100 wird in regelmäßigen Abständen geprüft, ob die Abfüllaufomatik noch richtig justiert ist.

Innerhalb welches Intervalls [500-c; 500+c] muß mit einer Wahrscheinlichkeit von 95,44% das Durchschnittsgewicht der in die Stichproben gelangten Kaffeepakete liegen, wenn der Abfüllung richtig verläuft?

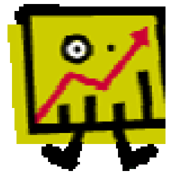
Bei der letzten Wahl erzielte eine Partei in einer Gemeinde einen Anteil von p=49%. Anhand einer Stichprobe aus den Wahlberchtigten soll geprüft werden, ob dieser Anteil bei den kommenden Wahlen übertroffen wird. Es genügt der Partei, die Entscheidung hierüber mit einer Irrtumswahrscheinlichkeit von höchstens $\alpha$= 0,1 zu treffen.

Von n=1000 Befragten wollen 510 Wahlberechtigte die Partei wählen. Wie lautet die Nullhypothese? Mit welcher Prüfverteilung findet der Chefstatistiker der Partei welchen Ablehnungsbereich? Zu welchem Schluß kommt er?

**Beispiel:** Ein Wirtschaftsforschungsinstitut soll das Wachstum für die sechs Branchen (I-VI) der Volkswirtschaft prognostizieren. Dabei werden fünf verschiedene Werte $x_i$ vorgegeben: Die **Prognose** des Instituts (Merkmal ***X***) ist in der folgenden Tabelle wiedergegeben. Darunter findet sich die **tatsächliche Entwicklung** in den einzelnen Branchen (Merkmal ***Y***).

| Ausprägungen | Code |
|---|---|
| „sehr niedrig" | 1 |
| „niedrig" | 2 |
| „mittelmäßig" | 3 |
| „hoch" | 4 |
| „sehr hoch" | 5 |

| Merkmal | Branche | | | | | |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI |
| **X** | **2** | **3** | **2** | **1** | **4** | **3** |
| **Y** | **3** | **3** | **1** | **2** | **4** | **5** |

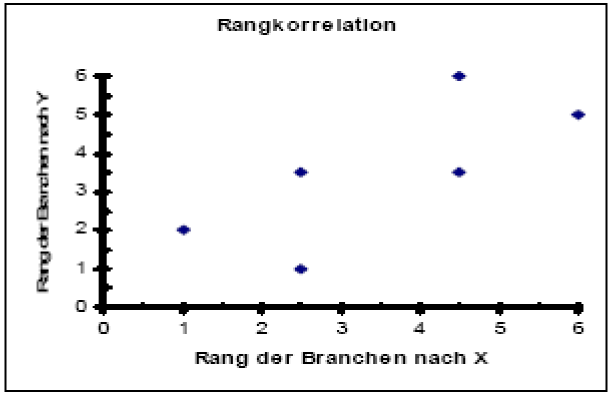


**Aufgabenstellung:**

Es interessiert, ob die Wachstumsprognose und die tatsächliche Entwicklung übereinstimmen, d. h. ob es ein Zusammenhang zwischen beiden Wertereihen besteht.

X: Wachstumsprognose eines Wirtschaftsinstituts
Y: die tatsächlichen Entwicklung von 6 Sektoren einer Volkswirtschaft

| Merkmal | Branche | | | | | |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI |
| X | 2 | 3 | 2 | 1 | 4 | 3 |
| Y | 3 | 3 | 1 | 2 | 4 | 5 |
| $r_X$ | 2,5 | 4,5 | 2,5 | 1 | 6 | 4,5 |
| $r_Y$ | 3,5 | 3,5 | 1 | 2 | 5 | 6 |



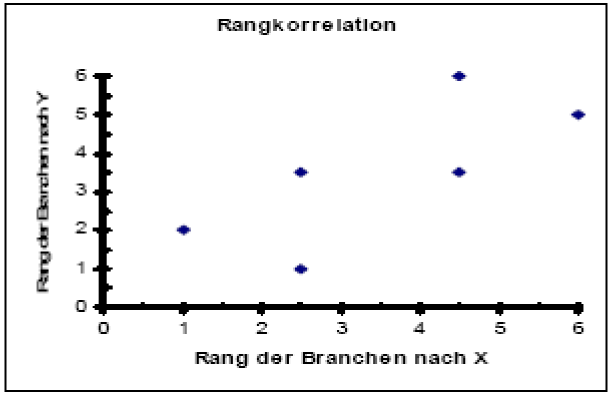

| Branche | $r_X$ | $r_Y$ | (rxi-ryi) | (rxi-ryi)² |
|---|---|---|---|---|
| I | 2,5 | 3,5 | -1 | 1 |
| II | 4,5 | 3,5 | 1 | 1 |
| III | 2,5 | 1 | 1,5 | 2,25 |
| IV | 1 | 2 | -1 | 1 |
| V | 6 | 5 | 1 | 1 |
| VI | 4,5 | 6 | -1,5 | 2,25 |
| Summe | 21 | 21 | | 8,5 |

$$1 - \frac{6\sum_{i=1}^{n}(r_{x,i} - r_{y,i})^2}{n(n^2-1)}$$

$$1 - \frac{6 * 8{,}5}{6 * (36-1)} = 0{,}757$$

X: Wachstumsprognose eines Wirtschaftsinstituts
Y: die tatsächlichen Entwicklung von 6 Sektoren einer Volkswirtschaft

| Merkmal | Branche | | | | | |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI |
| X | 2 | 3 | 2 | 1 | 4 | 3 |
| Y | 3 | 3 | 1 | 2 | 4 | 5 |
| $r_X$ | 2,5 | 4,5 | 2,5 | 1 | 6 | 4,5 |
| $r_Y$ | 3,5 | 3,5 | 1 | 2 | 5 | 6 |




| Branche | $r_X$ | $r_Y$ | $r^2_X$ | $r^2_Y$ | $r_X r_Y$ |
|---|---|---|---|---|---|
| I | 2,5 | 3,5 | 6,25 | 12,25 | 8,75 |
| II | 4,5 | 3,5 | 20,25 | 12,25 | 15,75 |
| III | 2,5 | 1 | 6,25 | 1,00 | 2,50 |
| IV | 1 | 2 | 1,00 | 4,00 | 2,00 |
| V | 6 | 5 | 36,00 | 25,00 | 30,00 |
| VI | 4,5 | 6 | 20,25 | 36,00 | 27,00 |
| Summe | 21 | 21 | 90,00 | 90,50 | 86,00 |

$$\rho_{SP} = \frac{\sum_{v=1}^{6} r_X(v) r_Y(v) - \frac{\left(\sum_{i=1}^{6} r_X(v)\right)\left(\sum_{i=1}^{6} r_Y(v)\right)}{6}}{\sqrt{\sum_{i=1}^{6} r_X^2(v) - \frac{\left(\sum_{i=1}^{6} r_X(v)\right)^2}{6}}\sqrt{\sum_{i=1}^{6} r_Y^2(v) - \frac{\left(\sum_{i=1}^{6} r_Y(v)\right)^2}{6}}}$$

$$= \frac{86{,}00 - \frac{21 \cdot 21}{6}}{\sqrt{90 - \frac{(21)^2}{6}}\sqrt{90{,}50 - \frac{(21)^2}{6}}} = 0{,}7464$$

Ein Professor hat in einem Semester Daten gesammelt (keiner der beiden, die diese Prüfung stellen, war's) und dabei festgestellt, daß die beste Klausur von dem Studenten geschrieben wurde, der die meisten Punkte bei den Vorleistungen erreicht hat. Die schlechteste Klausur wurde vom Studenten mit den wenigsten Vorleistungspunkten verfaßt. Er vermutet deshalb, daß Studenten, die mehr Vorleistungspunkte mitbringen, auch mehr Klausurpunkte erreichen. Von einem linearen Zusammenhang der beiden Punktzahlen möchte er aber nicht ausgehen. Wählen Sie *ein* (deskriptives) Abhängigkeitsmaß und berechnen dieses anhand der Punktzahlen der zehn Studenten A bis J.

| **STUDENT** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Klausurpunkte** | 76 | 44 | 32 | 53 | 25 | 58 | 26 | 59 | 29 | 65 |
| **Vorleistungspunkte** | 122 | 67 | 68 | 101 | 42 | 59 | 118 | 79 | 83 | 89 |

Für arbeitsmarktpolitische Entscheidungen in Deutschland sei die Fragestellung von Interesse, ob die Erwerbsstruktur geschlechterdifferenziert ist.

Ist-Situation Erwerbstätige (in 1000) im Jahr 2000

| Tätigkeit in | Erwerbstätige | |
|---|---|---|
| | Männer | Frauen |
| Land-, Forstwirtschaft, Fischerei | 639 | 348 |
| Produzierendes Gewerbe | 9230 | 2872 |
| Handel, Gastgewerbe | 4399 | 4018 |
| Sonstige Dienstleistungen | 6485 | 8612 |

| Tätigkeit in | bei Unabhängigkeit | |
|---|---|---|
| | M | F |
| Land-, Forstwirtschaft, Fischerei | 559,6 | 427,4 |
| Produzierendes Gewerbe | 6861,5 | 5240,5 |
| Handel, Gastgewerbe | 4772,2 | 3644,8 |
| Sonstige Dienstleistungen | 8559,6 | 6537,4 |

$$\boxed{n^*_{jk}} = \frac{n_{j.} \cdot n_{.k}}{n}$$

Kontingenzkoeffizient (**nominal-beliebig**)

| Tätigkeit in | bei Unabhängigkeit | | Differenzen $h^o_{ij} - h^e_{ij}$ | | $(h^o_{ij} - h^e_{ij})^2/h^e_{ij}$ | |
|---|---|---|---|---|---|---|
| | M | F | M | F | M | F |
| Land-, Forstwirtschaft, Fischerei | 559,6 | 427,4 | 79,4 | -79,4 | 11,26 | 14,75 |
| Produzierendes Gewerbes | 6861,5 | 5240,5 | 2368,5 | -2368,5 | 817,55 | 1070,44 |
| Handel, Gastgewerbes | 4772,2 | 3644,8 | -373,2 | 373,2 | 29,19 | 38,22 |
| Sonstige Dienstleistungen | 8559,6 | 6537,4 | -2074,6 | 2074,6 | 502,83 | 658,38 |
| | | | | Summe | **1360,84** | **1781,79** |

$0 \leq \chi^2 \leq n \cdot \min\{s-1, r-1\}$

r=4 und s=2

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(h^o_{ij} - h^e_{ij})^2}{h^e_{ij}} = 3142,63$$

$0 \leq \chi^2 \leq 36603 \cdot \min\{3, 1\}$

$0 \leq \chi^2 \leq 36603$

Kontingenzkoeffizient (**nominal-beliebig**)

Quadratischen Kontingenz

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(h_{ij}^{o}-h_{ij}^{e})^2}{h_{ij}^{e}} = 3142{,}63$$

Kontingenzkoeffizient von Pearson

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}} = \sqrt{\frac{3142{,}63}{36603+3142{,}63}} = 0{,}2812$$

$$0 \le C \le \sqrt{\frac{\min\{s-1, r-1\}}{1+\min\{s-1, r-1\}}}$$

$$0 \le C \le \sqrt{\frac{1}{2}}$$

$$0 \le C \le 0{,}7071$$

Korrigierter Kontingenzkoeffizient

$$C_{korr} = C\sqrt{\frac{1+\min\{s-1, r-1\}}{\min\{s-1, r-1\}}}$$

$$C_{korr} = 0{,}2812\sqrt{\frac{1+\min\{1,\ 3\}}{\min\{1,\ 3\}}} = 0{,}2812\sqrt{\frac{1+1}{1}} = 0{,}3977$$

$$0 \le C_{korr} \le 1$$

# Rechenbeispiel

In Börsenkreisen wird oft von einem Zusammenhang zwischen Rentenrenditen und Aktienkursen gesprochen. Zu 8 Zeitpunkten wurden folgende Werte für den Aktienindex der Frankfurter Allgemeinen Zeitung (FAZ-Index) und die Durchschnittsrendite der öffentlichen Anleihen mit 10 Jahren Laufzeit beobachtet.

| **Zeitpunkt** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|---|---|---|---|---|---|---|---|---|
| **FAZ-Index** | 221 | 251 | 346 | 376 | 401 | 421 | 471 | 481 |
| **Rendite in %** | 9,7 | 7,9 | 8,6 | 7,2 | 7,3 | 7,1 | 7,0 | 6,8 |

a) Berechnen Sie eine Regressionsgerade nach der Methode der kleinsten Quadrate..

b) Berechnen Sie das Bestimmtheitsmaß

Die Einwohnerzahl einer Kleinstadt entwickelte sich in fünf aufeinanderfolgenden Jahren wie folgt:

| **i** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| $t_i$ | 1997 | 1998 | 1999 | 2000 | 2001 |
| $y_i$ | 8,0 | 8,2 | 8,6 | 9,4 | 11 |

Dabei bezeichnen $t_i$ den Erhebungszeitpunkt und $y_i$ die Einwohnerzahl (in Tausend) zum Zeitpunkt $t_i$.

(1) Berechnen Sie für diese Zeitreihe die Parameter a und b der Trendgeraden $y = a + bt$ nach der Methode der kleinsten Quadrate.

(2) Ermitteln Sie die Korrelationskoeffizienten r von Bravais-Pearson und $r_s$ von Spearman.

(3) Interpretieren Sie r und $r_s$ vergleichend.

Von den in der ersten Vorlesung STATISTIK I abgegebenen Fragebögen wurden 20 zufällig ausgewählt. In der folgenden Liste ist das Alter der Studenten angegeben:

20 20 21 21 21 21 22 22 22 22 22 22 23 23 23 23 23 24 25 25

1.1 Erstellen Sie eine Tabelle mit den absoluten und relativen Häufigkeiten für das Merkmal Alter.

1.2 Geben Sie die empirische Verteilungsfunktion an.

1.3 Berechnen Sie den Median, das untere Quartil und das obere Quartil.

1.4 Skizzieren Sie den Box-Whisker-Plot

In einer Studie über das Lernverhalten der Studierenden werden zufällig 5 Studenten über die durchschnittliche Lernzeit und die Klausurnoten in Statistik befragt.

| Student | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Lernzeit (Std.) | 6 | 10 | 8 | 4 | 2 |
| Klausurnoten | 2 | 3 | 1 | 5 | 4 |

2.1 Zeichnen Sie das Streudiagramm für die Merkmale Lernzeit und Klausurnoten.

2.2 Welchen Zusammenhang zwischen den beiden Merkmalen entnehmen Sie aus dem Streudiagramm?

2.3 Berechnen Sie die Stichprobenvarianz für die Merkmale Lernzeit und Klausurnoten, die Stichprobenkovarianz zwischen den beiden Merkmalen und den Korrelationskoeffizienten nach Pearson. Interpretieren Sie den Koeffizienten.

5. Aufgabe

Die erzielten Einnahmen pro Werbeblock bei einem Fernsehsender können als unabhängig und normalverteilt angesehen werden, sie unterscheiden sich jedoch nach Wochentagen. Es liegen folgende Stichprobenergebnisse vor: Bei 64 werktags gesendeten Werbeblöcken wurden mittlere Einnahmen je Werbeblock von 72.500 Euro erzielt bei einer Standardabweichung von 16.000 Euro, die 25 am Wochenende gesendeten Werbeblöcke erbrachten im Mittel 183.000 Euro bei einer Standardabweichung von 26.500 Euro.

5.1 Ergänzen Sie die Punktschätzungen für die Erwartungswerte an Werktagen und am Wochenende jeweils durch eine Intervallschätzung auf 95%igem Konfidenzniveau. Berechnen Sie die Grenzen der beiden Konfidenzintervalle.

LSG:

$n1=64$; $\mu1=72.500$ $\sigma1=16.000$ ; $n2=23$; $\mu2= 183.00$ $\sigma2=26.500$

Schätzproblem: Intervallschätzer des Mittelwerte μ1 und μ2 (Formelsammlung Abschn.6.3.)

Mit den gelben Infos einsetzen und ausrechnen.

**5. Aufgabe**

Ein Automat füllt Packungen mit 50 Schrauben. Das Packungsgewicht soll 100g betragen. Es ist bekannt, dass die tatsächlichen Packungsgewichte normalverteilt sind und dass 99,73 % der Packungen ein Gewicht zwischen 98,5g und 101,5g haben.

5.1 Bestimmen Sie die Parameter $\mu$ und $\sigma$ der Normalverteilung.

5.2 Berechnen Sie die Wahrscheinlichkeit, dass eine zufällig ausgewählte Packung weniger als 100g wiegt.

5.3 Bestimmen Sie das Gewicht, welches 95 % der Packungen höchstens haben.

5.4 Der Produzent behauptet, dass mindestens 99 % der Schrauben einer Packung fehlerfrei sind. Ab welcher Anzahl fehlerhafter Schrauben je Packung kann die Behauptung des Produzenten bei 5%igem Signifikanzniveau abgelehnt werden? Nennen Sie die theoretische Verteilung, die als Modell genutzt wird und bestimmen Sie deren Parameter. Formulieren Sie die Null- und die Alternativhypothese für den statistischen Test.

5.5 Um die Arbeit des Automaten zu überwachen wird pro Schicht das Gewicht von 16 zufällig entnommenen Packungen festgestellt und das Stichprobenmittel des Packungsgewichtes berechnet. Bestimmen Sie einen symmetrischen Bereich um den Sollwert, in dem das berechnete Stichprobenmittel mit einer Sicherheit von 0,9545 liegt.

**Aufgabe 5**

Ein Widerstandshersteller produziert ein Standardprodukt mit einem Soll-Wert von 25 Ω. Durch die Installation einer neuen Produktionsanlage soll die Produktivität gesteigert werden. Zur Prüfung der Qualität der Produkte aus der neuen Produktionsanlage wurden 100 zufällig ausgewählte Widerstände gemessen. Der Mittelwert beträgt 25,2 Ω, die Varianz beträgt 0,0036.

5.1 Prüfen Sie zum Signifikanzniveau 0,01, ob sich der Erwartungswert der Widerstände von dem Soll-Wert des Produktes unterscheidet. Formulieren Sie eine Hypothese, geben Sie die Teststatistik und den kritischen Wert an. Treffen Sie eine Entscheidung.

5.2 Prüfen Sie, ob sich der Widerstandswert signifikant ($\alpha=0,01$) erhöht hat. Formulieren Sie eine Hypothese, geben Sie die Teststatistik und den kritischen Wert an. Treffen Sie eine Entscheidung.

Aufgabe
Sind die folgenden metrischen Merkmale diskret oder stetig:
Rechnungsbetrag
Wahlergebnis in Prozent einer Partei
Kraftstoffverbrauch eines PKW auf 100km
Summe der Zahlen einer Lottoziehung
Zeitspanne, die zur Verrichtung einer bestimmten Arbeit benötigt wird
Zahl der pro Stunde in einem Geschäft eintreffenden Kunden
Grundstücksgröße