Wirtschaftsstatistik

Übungsblatt Modul 4 & 5

Lageparameter & Streuungsparameter

# Aufgabe 1

1. Worüber informiert …
   1. … die Standardabweichung?

**Die Standardabweichung informiert über die Streuung der Beobachtungswerte und über die Aussagekraft des Mittelwerts. Sie gibt also Auskunft, wie hoch die Varianz zum Mittelwert ist.**

**Bei einer kleinen Standardabweichung ist die Streuung gering und die Beobachtungswerte liegen nahe am Mittelwert.**

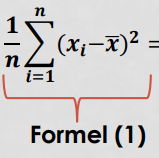
**Eine große Standardabweichung bedeutet, dass die Beobachtungswerte weit um den Mittelwert gestreut sind (große Streuung).**

**Bei normal verteilten Daten befinden sich 95% der Beobachtungswerte im Intervall** 

Die Standardabweichung ist die Wurzel aus der Varianz 

**Die Varianz berechnet sich aus der Summe der Quadrate zur Differenz aus der Merkmalsausprägung und dem arithmetischen Mittel zur Merkmalsausprägung geteilt durch die Anzahl der absoluten Häufigkeiten**  
(Varianz ist das arithmetische Mittel der Abweichungsquadrate).

(Merkmalsausprägung – arithmetische Mittel aus den Merkmalsausprägungen)² / n



Taschenrechner Casio:

* Mode 🡆 Taste „2“ für Menü „2: STAT“ 🡆 Taste „1“ für „1: 1- VAR“
* Eingabe der Werte zu den absoluten Häufigkeiten (Eingabe 1. Wert und „=“, Eingabe 2. Wert und „=“ usw.; mit Pfeil-Tasten kann zwischen den eingegebenen Werten gewechselt werden und die Werte können geprüft werden, ein falsch eingegeben Wert kann einfach durch Neueingabe und = überschrieben werden)
* Sind alle Werte eingegeben mit Taste „AC“ bestätigen
* Tasten „Shift“ und „1“
* Menü „4: Var“
* Neue Menüauswahl:  
  „1: n“   
  „2: x̅“ arithmet. Mittel (Mittelwert)  
  „3: Sigma x“ Standardabweichung  
  „4: sx“
* Tasten „3: Sigma x“ und „=“ zeigt die Standardabweichung
* Mit Taste „x²“ potenzieren ergibt die Varianz
  1. … das Quartil Q3?

Quantil ist ein Lageparameter (Lagemaß) der Statistik

Quantile teilen eine Verteilung in Abschnitte gleicher Häufigkeit.

Voraussetzung für die Nutzung von Quantilen ist eine geordnete statistische Reihe.

Das Quartil enthält 4 Intervalle (Abschnitte). Jeder dieser Abschnitte enthält 25% der Häufigkeit aus allen geordneten Beobachtungswerten.

**Das Quartil Q3 enthält 25% der Beobachtungswerte. 75% der Beobachtungswerte in einer geordneten statistischen Reihe sind kleiner als das 3. Quartil.**

* 1. … der Variationskoeffizient?

Der Variationskoeffizient ist ein relatives Streuungsmaß. Er ist das prozentuale Verhältnis der Standardabweichung s zum arithmetischen Mittel x̅



* 1. … der Interquartilsabstand?

Der (Inter)-Quartilsabstand ist ein Streuungsmaß und errechnet sich aus der Differenz zwischen dem oberen und unteren Quartil (Q3 – Q1). Er beinhaltet 50% der Verteilung.

Der Quartilsabstand gibt die Breite des mittleren Bereichs an und ist von Ausreißern unabhängig.

Zusätzliche Anmerkung:

Das zweite Quartil repräsentiert den Median, d. h. Median und das 2. Quartil sind identisch bzw. der Median ist ein Q2 Quantil (Q2 Quantil ist ein 50% Quantil und besteht aus 2 Intervallen)

* 1. … das 5%-Quantil?

Quantil ist ein Lageparameter (Lagemaß) der Statistik

Quantile teilen eine Verteilung in Abschnitte gleicher Häufigkeit.

Voraussetzung für die Nutzung von Quantilen ist eine geordnete statistische Reihe.

Das 5% Quantil (auch Quintil genannt) enthält 5% aller geordneten Beobachtungswerte, die kleiner als das 5% Quantil sind.

* 1. … der Median  (auch Zentralwert genannt)?

**Median ist ein Lageparameter.**

Median (Zentralwert) ist der Zentralwert. Vom Median **liegen mindestens 50% der Beobachtungswerte links des Medians** (unter dem Median) und **mindestens 50% der Beobachtungswerte rechts des Medians, d. h. über dem Median (den Medion ggf. mit eingerechnet)**.

Median ist ein **robustes Lokationsmaß** 🡆 robuste statistische Größen sind **wenig anfällig gegen Datenausreißer.** Die Hälfte der Daten (Stichprobe) muss gegen + ODERS gegen - verschoben werden, um Median selbst gegen +- wandern zu lassen

* 1. … der Modus  (auch Modalwert genannt)?

**Modus ist ein Lageparameter.**

Ist die am häufigsten auftretende Merkmalsausprägung (**größter Wert zur maximalen Häufigkeit**) Der Modus wird **hauptsächlich für nominale Merkmale** verwendet, ist **aber auch für andere diskrete Merkmalstypen sinnvoll**. Gibt es mehre Merkmalsausprägungen mit derselben maximalen Häufigkeit , dann existieren mehrere Modalwerte 🡆 wird Multimodale Verteilung genannt.

**unimodale Verteilung:** **Dichtekurve** hat **nur ein lokales Maximum**

**multimodale Verteilung: Dichtekurve** hat **mehrere lokale Maxima** (bimodale Verteilung, Trimodale Verteilung …)

* 1. … die Spannweite?

Die Spannweite (Variationsbreite) w ist das Maß für die Breite des Streubereichs einer Häufigkeitsverteilung. Sie repräsentiert die Gesamtbreite aller Werte in einer Verteilung.

**Die Spannweite kommt bei Ordinalwerten und metrischen Werten zur Anwendung und berechnet sich aus der Differenz aus dem höchsten Wert (Maximum) aller Beobachtungswerte und kleinsten (Minimum) aller Beobachtungswerte**

W = xmax - xmin

Die Spannweite ist somit vom größten und kleinsten Wert abhängig und damit kein robustes Maß, da anfällig gegenüber Ausreißern.

* 1. … das Quartil Q1?

Quantile ist ein Lagemaß der Statistik.

Quantile unterteilen eine Verteilung in einer geordneten statistischen Reihe in Abschnitte gleicher Häufigkeit.

Das Quartil beinhaltet 4 Intervalle mit gleiche großen Bereichen. Jedes Intervall enthält 25% der geordneten Beobachtungswerte.

**Das Quartil Q1 enthält 25% der geordneten Beobachtungswerte die kleiner sind als das Quartil Q1**.

1. Was haben die statistischen Parameter Varianz, Standardabweichung und Variationskoeffizient gemeinsam? Wodurch unterscheiden sich die drei Parameter?

Varianz, Standardabweichung und Variationskoeffizient

Alle drei Parameter sind Streuungsparameter, nur für metrische Merkmale anwendbar und basieren auf dem arithmetischen Mittel.

Die Varianz ist das arithmetische Mittel der Abweichungsquadrate und berechnet sich aus dem arithmetischen Mittel zur Summe der Quadrate aus der Differenz aus Merkmalsausprägung und dem arithmetischen Mittel zur Merkmalsausprägung.

s² = ∑ (xi – x̅i)² / n

Die Standardabweichung ist die Wurzel aus der Varianz.

Der Varianzkoeffizient ist das prozentuale Verhältnis aus Varianz und dem arithmetischen Mittel.

# Aufgabe 2

In einer Statistik über die Einkommen(Jahreseinkommen) von leitenden Angestellten im Rechnungswesen lesen Sie:

1. Quartil Q1: 75.000 €, 3. Quartil Q3: 150.000 €, Median 𝒙̅𝒁: 100.000 €, Mittelwert 𝒙̅: 140.000 €.

Welche der folgenden Aussagen über die Einkommensverteilung sind richtig? (Zutreffendes ankreuzen!)

|  |  |  |
| --- | --- | --- |
|  | **RICHTIG** | **FALSCH** |
| (1) Die Einkommensverteilung ist linksschief. |  | X |
| (2) 25% der Befragten verdienen weniger als 75.000 €. | X |  |
| (3) 50% der Befragten verdienen zwischen 75.000 und 150.000 €. | X |  |
| (4) 75% der Befragten verdienen mehr als 150.000 €. |  | X |
| (5) Die Einkommensverteilung ist symmetrisch. |  | X |
| (6) 50% der Befragten verdienen weniger als 140.000 €. |  | X |

**(1) Die Einkommensverteilung ist linksschief.**

*Skript Modul 4 – „Lageparameter“, Quartile auf Seite 36*

Diese Aussage ist falsch

Median 𝒙̅𝒁 = 100.000

Arithmetische Mittel 𝒙̅ = 140.000

Median 𝒙̅𝒁 < Arithmetische Mittel 𝒙̅ 🡆 rechtsschief (linkssteil)

**(2) 25% der Befragten verdienen weniger als 75.000 €.**

Diese Aussage ist korrekt

*Skript Modul 4 – „Lageparameter“, Quartile auf Seite 36*

Q1 = 25% aller geordneten Beobachtungswerte sind kleiner als das 1. Quartil.

**(3) 50% der Befragten verdienen zwischen 75.000 und 150.000 €**

*Skript Modul 5 – „Streuungsparameter“, (Inter-)Quartilsabstand, auf Seite 6*

Diese Aussage ist korrekt

Q1 = 75.000, Q3 = 150.000

Zwischen dem Q1. und Q3. Quartil liegen 50% aller Beobachtungswerte (auch Quartilsabstand genannt)  
Das bedeutet zwischen Q1 = 75.000 und Q3 = 100000 liegen 50% aller Beobachtungswerte 🡆 es verdienen 50% zwischen 75.000 und 150.000

**(4) 75% der Befragten verdienen mehr als 150.000 €**

Diese Aussage ist falsch.

*Skript Modul 4 – „Lageparameter“, Quartile auf Seite 36*

**Q3 = 75% aller geordneten Beobachtungswerte sind kleiner als das 3. Quartil.**

**(5) Die Einkommensverteilung ist symmetrisch**

Diese Aussage ist falsch.

Die Differenz zwischen Q1 = 75.000 und Q3 = 100.000 = 75.000 / 2 = 37.500 🡆 Q1 = 75.000 + 37.500 = 112.500

Eine symmetrische Verteilung liegt vor, wenn das 2. Quartil (= Median ZD ) einen Wert von 112.500 hätte.

Der Median ZD (das 2. Quartil ) hate jedoch einen Wert von 100.000

**(6) 50% der Befragten verdienen weniger als 140.000 €.**

Diese Aussage ist falsch

*Skript Modul 4 – „Lageparameter“, Quartile auf Seite 36*

**Q2 = 50% aller geordneten Beobachtungswerte sind kleiner als das 2. Quartil.**

Der **Median 𝒙̅𝒁 = 100.000** entspricht dem 2. Quartil Q2 und ist mit **100.000 < 140.000**

# Aufgabe 3

Erstellen Sie auf der Basis der folgenden Angaben eine eindimensionale, unklassierte Häufigkeitsverteilung:

In einer Stadt haben vier Taxi-Unternehmen jeweils drei Wagen, ein Taxi-Unternehmen hat 26 Wagen. Die übrigen Taxi-Unternehmen in der Stadt sind kleiner und haben weniger Wagen: acht Taxi-Unternehmen haben jeweils nur einen Wagen, sieben haben jeweils zwei Wagen.

**Anzahl der Fahrzeuge ist der Beobachtungswert (ein Absolutwert) 🡆 das ist der Wert, der gezählt wird**

**Anzahl der Unternehmen sind die Werte zur absoluten Häufigkeit** (zur Verteilung; Fragen zur Bestimmung der Werte zur absoluten Häufigkeit: **„Wieviel Unternehmen haben …“, „Welchen Anteil haben die Unternehmen an …“, „Wie groß ist die Verteilung an …?“** usw.)

Beantworten Sie auf der Basis der erstellten Tabelle die folgenden Fragen (**mit 1 Nachkommastelle**):

a) Wie viele Taxi-Unternehmen gibt es in dieser Stadt?

b) Wie viele Wagen bieten in dieser Stadt ihre Leistungen an?

c) Bestimmen / berechnen Sie für die obige Verteilung

a. den Modus

b. den Median

c. das arithmetische Mittel

d) Machen Sie Aussagen über die Schiefe der Verteilung

e) Berechnen Sie die Standardabweichung

f) Wie viel % der Taxi-Unternehmen haben mehr als einen Wagen

g) Wie viel % der Taxi-Unternehmen haben weniger als 3 Wagen

Tipp: Bitte beachten Sie, dass in dieser Aufgabe Unternehmen nach der Anzahl ihrer Wagen

(=Merkmalsausprägung) charakterisiert werden und somit werden die Unternehmen gezählt, nicht die Wagen.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| i | Beobachtungswert  Anzahl Wagen  xi | Absolute Häufigkeit  Anzahl Taxi-Unternehmen  h(xI) | Relative Häufigkeit  F(xi) in % | Modus  x̅D | Absolute Summen-  häufigkeit  H(xi) | Relative Summen-  häufigkeit  F(xi) in % | xi \* h(xi) |
| 1 | Modus **1** | **8** | 40,0% | **1** | 8 | 40,0% | **1** \* **8** =**8** |
| 2 | **2** | **7** | 35,0% |  | 15 | 75,0% | **2** = **7** = **14** |
| 3 | **3** | **4** | 20,0% |  | 19 | 95,0% | **3** \* **4** = **12** |
| 4 | **26** | **1** | 5,0% |  | **20** | **100,0%** | **26** \* **1** = **26** |
| ∑ |  | **20** | **100,0%** |  | **-** | **-** | **60** |
|  |  |  |  |  |  | **Arithm. Mittel** | **x̅ = 60 : 20 = 3,0** |

Hinweis: Wert aus Spalte i nicht beachten (i ist lediglich der Index zur jeweiligen Merkmalsausprägung xi bzw. der absoluten Häufigkeit h(xi) )

a) Wie viele Taxi-Unternehmen gibt es in dieser Stadt?

Anzahl Taxi-Unternehmen = Summe der absoluten Häufigkeiten (= Summe der Taxiunternehmen)  
 aus der Tabelle

**In dieser Stadt gibt es 20 Taxi-Unternehmen.**

b) Wie viele Wagen bieten in dieser Stadt ihre Leistungen an?

Summe Anzahl Wagen = ∑ (Anzahl Wagen \* Anzahl Unternehmen)

*Anzahl Wagen ist der Beobachtungswert xi,   
 Anzahl der Unternehmen ist* *die absolute Häufigkeit h(xi)*

Summe Anzahl Wagen = x1 \* h(x1) + x2 \* h(x2) + (xn \* h(xn))

Summe Anzahl Wagen = **1**\***8** + **2**\***7** + **3**\***4** + **26**\***1** = 60  
 **In dieser Stadt bieten 60 Wagen ihre Leistungen an**

c) Bestimmen / berechnen Sie für die obige Verteilung

a. den Modus

*Skript Modul 4 – „Lageparameter“, MODUS auf Seite 4*

Der **Modus (Modalwert) ist die am häufigsten auftretende Merkmalsausprägung** (maximale Häufigkeit), wird hauptsächlich für nominale Merkmale verwendet, ist auch für andere (diskreten) Merkmalstypen sinnvoll.

Bei **klassierten Daten ist der Modalwert die Mitte der Klasse mit den größten Häufigkeiten**. Diese Klasse nennt man die **Modalklasse**.Multimodale Verteilungen (bimodale Verteilung: zwei Modalwerte; trimodale Verteilung: drei Modalwerte) : mehrere Merkmalsausprägungen mit der gleichen maximalen Häufigkeit

**Kurz: Modus ist der Beobachtungswert xi mit der größten Häufigkeit aus Spalte h(xi)**

**In der Aufgabe hat die größte Häufigkeit h(xi) den Wert 8, der dazugehörige Beobachtungswert = 1**

**Modus x̅ D= 1**

Hinweis: Der **Modus bezieht sich immer auf den Beobachtungswert (Merkmalswert) von xi (hier die Anzahl der Wagen)**, nicht auf die absolute Häufigkeit h(xi)🡆 bzw. der daraus resultierenden Häufigkeitsverteilung

b. den Median

*Skript Modul 4 – „Lageparameter“, Median auf Seiten 6 - 16*

Für den Median (Zentralwert) x̅ Z werden die **Summe der absoluten Häufigkeiten aus der Summenzeile** (**Anzahl n der Merkmalsausprägungen = ) und eine geordnete Reihe von Beobachtungswerten** (Merkmalswerten) benötigt.

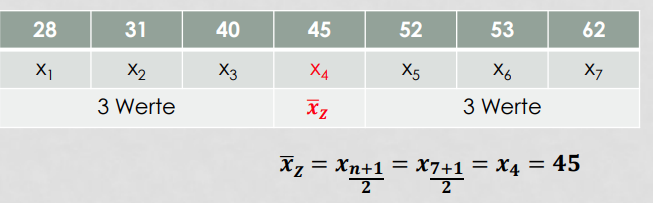
Der M**edian ist der Wert in der Mitte der geordneten Reihe.   
50% der Beobachtungswerte liegenunter dem Median, 50% liegen darüber**

**Bsp. aus dem Skript zu einer geraden Anzahl von Beobachtungswerten**   
*Skript Modul 4 – „Lageparameter“, Median, Seite 9)*

**Ungerade Anzahl der Merkmalsausprägungen (Summe der absoluten Häufigkeiten aus der Summenzeile) = 7**

x̅ Z = xn+1/2

**WICHTIG: Geordnete Reihe (im Bsp. zum Alter: 28, 31, 40, 45, 52, 53, 62)**

****

Wir rechnen **(7 + 1) / 2 = 4 🡆 4 ist die Mitte der geordneten Reihe**

* **der Median ist der Wert aus der Mitte der geordneten Reihe = 45**

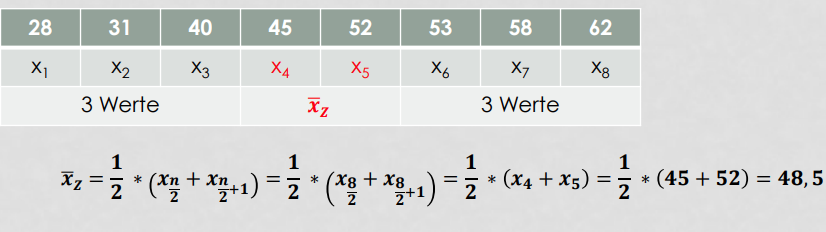
**Bsp. aus dem Skript zu einer geraden Anzahl von Beobachtungswerten**   
*Skript Modul 4 – „Lageparameter“, Median, Seite 10)*

Gerade Anzahl der Merkmalsausprägungen (Summe der absoluten Häufigkeiten aus der Summenzeile) = 8

Bei einer geraden Anzahl von Werten wird Median aus den beiden mittleren Werten berechnet.

x̅ Z = (xn/2 + xn/2+1) \* ½

**WICHTIG: Geordnete Reihe (im Bsp. zum Alter: 28, 31, 40, 45, 52, 53, 58, 62)**



Wir rechnen **(8 / 2) = 4 und (8 / 2) +1 = 5 🡆 4 und 5 sind die Mitte der geordneten Reihe**

🡆 der **Median ergibt sich aus der Summe der beiden Beobachtungswerte aus der Mitte dividiert durch 2**

Wir **summieren 45 aus der 4. Position (x4) und 52 auf der 5. Position (x5) und teilen die Summe durch 2**

* 45 + 52 = 97 / 2 = 48,5

**Lösung zur Aufgabe:**

Anzahl der Merkmalsausprägungen n = Summe der absoluten Häufigkeiten aus der Summenzeile

n = 8 + 7 + 4 + 1 = 20

Medianberechnung bei gerader Anzahl n der Merkmalsausprägungen (in der Aufgabe ist die Anzahl n der Merkmalsausprägungen = 20)

Bei einer ungeraden Anzahl von Beobachtungswerten (Merkmalswerten) wird die Anzahl der Beobachtungswerte +1 addiert und danach durch 2 dividiert (so dass die Mitte der Reihe erreicht wird), bei 9 ist die Mitte 9 + 1 / 2 = 5, bei 11 ist die Mitte 11 +1 / 2 = 6 usw.  
Der Median ist dann der Beobachtungswert xi zur Mitte.

Bei einer geraden Anzahl von Beobachtungswerten (Merkmalswerten) , hier 20) wird die Anzahl der Beobachtungswerte durch 2 geteilt.   
**Das ist der erste Teil des Medians = 10.   
Der 2. Teil des Medians ist der Nachfolger des ersten Teils = 10 + 1 = 11.**  
Die Beobachtungswerte xi beider Teilmediane werden addiert und danach durch 2 dividiert.

x̅ Z = (xn/2 + xn/2+1)

n = 20

**xn/2 = 10, xn/2 + 1 = 11 🡆 es sind die Beobachtungswerte zum 10. Und 11. Wert aus der geordneten Reihe zu addieren und anschließend durch 2 zu dividieren**

Die geordnete Reihe (in der Aufgabe ist die Reihe bereits geordnet) hat

8 \* den Beobachtungswert 1   
(Anzahl Unternehmen = absolute Häufigkeit ( h(xi) ) = 8 \* Anzahl Wagen (Beobachtungswert) (xi) = 1

7 \* den Beobachtungswert 2

4 \* den Beobachtungswert 3

1 \* den Beobachtungswert 26

Geordnete Reihe: **1 1 1 1 1 1 1 1** 2 2 2 2 2 2 2 3 3 3 3 26

Der **Beobachtungswert zum 10. und 11. Wert in der Reihe ist also= 2**  🡆 der 10. Und 11. Wert werden addiert und anschließend durch 2 dividiert 🡆 **2 + 2 = 4 : 2 = 2**

**Median** **x̅ Z = (2 + 2 ) : 2 = 2**

c. das arithmetische Mittel

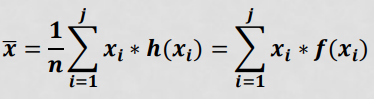
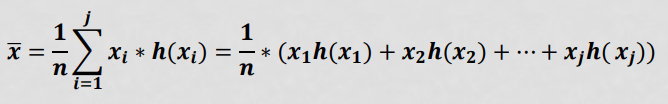
*Skript Modul 4 – „Lageparameter“, Arithmetisches Mittel auf Seiten 17 bis 28*

**N ist die Summe der absoluten Häufigkeiten aus der Summenzeile =   
Anzahl der Merkmalsausprägungen**

**Arithmetische Mittel bei absoluten Häufigkeiten:**

Summe aus dem Produkt (Beobachtungswert xi \* absolute Häufigkeit h(xi) ) / Summe der absoluten Häufigkeiten (im Bsp. = 20)

x̅ = (x1 \* h(x1)) + (x2 \* h(x2)) + … + (xn \* h(xn)) / n

𝒉(𝒙𝒊) = absolute Häufigkeit der Merkmalsausprägung xi

n = Summe der absoluten Häufigkeiten

j = Anzahl der Merkmalsausprägungen x

n = 8 + 7 + 4 + 1 = 20

x̅ = (x1 \* h(x1)) + (x2 \* h(x2)) + (xn \* h(xn)) / n

x̅ = (1\*8) + (2\*7) + (3\*4) + (26\*1) / 20 = 60 / 20 = 3

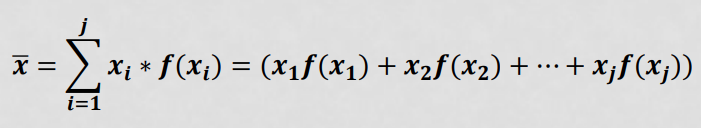
**arithmetische Mittel x̅ = 3**

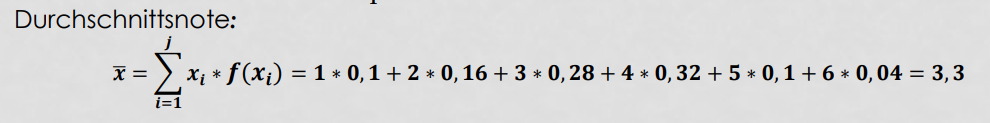
**Arithmetische Mittel bei relativen Häufigkeiten:**

(Summe aus dem Produkt Beobachtungswert xi \*relative Häufigkeit h(xi) )

x̅ = (x1 \* f(x1)) + (x2 \* f(x2)) + … + (xn \* f(xn))

**WICHTIG: Da bei Berechnung der relativen Häufigkeit bereits die absolute Häufigkeit durch die Summe der absoluten Häufigkeiten geteilt wird, darf zum arithmetischen Mittel bei relativer Häufigkeit NICHT durch die Summe n der absoluten Häufigkeiten geteilt werden.**





0,1 und 0,16 und 0,28 und 0,32 und 0,1 und 0,04 sind relative Häufigkeiten

d) Machen Sie Aussagen über die Schiefe der Verteilung

*Skript Modul 4 – „Lageparameter“, Arithmetisches Mittel auf Seiten 30 bis 33*

Für die Neigung / Schiefe werden die Werte für den Modus x̅D ,den Median x̅Z und das arithmetische Mittel x̅ benötigt

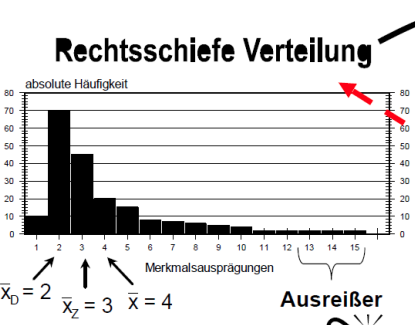
**rechtsschiefe (linkssteile) Häufigkeitsverteilung:**

**Modus (Beobachtungswert mit der größten Häufigkeit ist) < Median**

**Median < arithmetische Mittel**

**Modus x̅D < Median x̅Z < arithmetische Mittel x̅ x̅D < x̅Z < x̅**

**x̅Z < x̅**



**Lösung zur Aufgabe:**

Modus x̅D = 1

Median x̅Z = 2

arithmetische Mittel x̅ = 3

**Die Verteilung ist rechtschief (linkssteil).**

Es gibt Ausreißer bei den oberen Bereich der Verteilung.

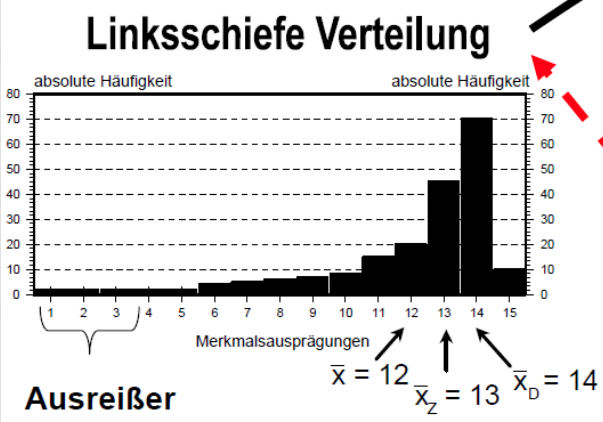
**linksschiefe (rechtssteile) Häufigkeitsverteilung:**

**Modus (Beobachtungswert mit der größten Häufigkeit ist) > Median**

**Median > arithmetische Mittel**

**Modus x̅D > Median x̅Z > arithmetische Mittel x̅ x̅D > x̅Z > x̅**

**x̅Z > x̅**



e) Berechnen Sie die Standardabweichung

*Skript Modul 5 – „Streuungsparameter“, STANDARDABWEICHUNG auf Seite 28*

* **Die Standardabweichung ist die Quadratwurzel aus der Varianz.** 
* Sie ist die Stärke zur Aussagekraft des Mittelwertes   
  (Maß dafür, wie hoch die Aussagekraft des Mittelwerts ist)
* kleine Standardabweichung: Alle Beobachtungswerte liegen nahe am Mittelwert (kleine Streuung).
* große Standardabweichung: Beobachtungswerte sind weit um den Mittelwert gestreut
* bei normalverteilten Daten liegen ca. 95% der Beobachtungswerte im Intervall [x̅ - 2s, x̅ + 2s]

**Varianz ist das arithmetische Mittel aus den Abweichungsquadraten** berechnet aus

- der Differenz zum Beobachtungswert xi und dem arithmetischen Mittel zum Quadrat multipliziert mit der absoluten Häufigkeit h(xi) pro Merkmalsausprägung dividiert durch die Summe der absoluten Häufigkeiten ∑ h(xi)

Für die Standardabweichung werden somit für jede absolute Häufigkeit

- das arithmetische Mittel x̅   
 Summe der Produkte aus (Beobachtungswert xi \* absolute Häufigkeit h(xi) pro Merkmalsausprägung)   
 dividiert durch Summe der absoluten Häufigkeiten h(xi)

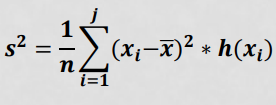
- die quadratischen Einzelabweichungen = Differenz aus Beobachtungswert xi und arithmetischen Mittel x̅   
 zum Quadrat

- die absolute Häufigkeit f(xi)

- Summe der absoluten Häufigkeiten n

benötigt.

Für jede absolute Häufigkeit wird die quadratischen Einzelabweichungen mit der absoluten Häufigkeit multipliziert. Die Summe der sich daraus ergeben Produkte dividiert durch die Summe der absoluten Häufigkeiten ist die Varianz.



Varianz s² = ( (x1 – x̅)² \* h(x1) + (x2 – x̅)² \* h(x2) + … + (xi – x̅)² \* h(xi) ) / n

Arithmetische Mittel x̅ = 3 (berechnet in Aufgabenteil c )

Summe der absoluten Häufigkeiten n = 20

Varianz s² = ( (1 – 3)² \* 8 + (2 – 3)² \* 7 + (3 – 3)² \* 4 + (26 – 3)² \* 1) / 20

Standardabweichung s =

**WICHTIG:**

**Bei Berechnung mit relativer Häufigkeit darf im letzten Schritt nicht durch die Summe der absoluten Häufigkeiten dividiert werden, da die relativen Häufigkeiten sich bereits aus der Summe der absoluten Häufigkeiten ergeben.**

Lösung zur Aufgabe:

|  |  |  |
| --- | --- | --- |
| Beobachtungswert  Anzahl Wagen  xi | Absolute Häufigkeit  Anzahl Taxi-Unternehmen  h(xI) | Arithmetisches Mittel  x̅ |
| **1** | **8** | **3** |
| **2** | **7** | **3** |
| **3** | **4** | **3** |
| **26** | **1** | **3** |
|  | **n = 20** |  |

arithmetisches Mittel x̅ = 3 (Lösung aus c)

Varianz s² = ( (1 – 3)² \* 8 + (2 – 3)² \* 7 + (3 – 3)² \* 4 + (26 – 3)² \* 1) / 20

Varianz s² = ( (-2)² \* 8 + (-1)² \* 7 + (-0)² \* 4 + (23)² \* 1) / 20

Varianz s² =( 4 \*8 + 1\*7 + 0 + 529 \* 1) / 20

Varianz s² = 32 + 7 + 0 + 529) / 20

Varianz s² = 568 / 20

**Varianz s² = 28,4**

**Standardabweichung s = = 5,33**

f) Wie viel % der Taxi-Unternehmen haben mehr als einen Wagen

relative Summenhäufigkeit zu Taxiunternehmen die einen Wagen haben = 40%

100 – 40 = 60

**60% der Taxi-Unternehmen haben mehr als einen Wagen**

g) Wie viel % der Taxi-Unternehmen haben weniger als 3 Wagen

Taxiunternehmen, die einen Wagen haben = 40%

Taxiunternehmen, die zwei Wagen haben = 35%

relative Summenhäufigkeit zu Taxiunternehmen, die weniger als 3 Wagen haben = 75%

**75% der Taxi-Unternehmen haben weniger als 3 Wagen.**

# Aufgabe 4

1. Sie lesen in einer Studie über die Altersverteilung in einer Gruppe, dass 𝒙̅𝒁= 32 Jahre und 𝒙̅ = 40 Jahre ist. Welche Schlüsse können Sie daraus über die Altersverteilung ziehen?

Median 𝒙̅𝒁= 32 Jahre

Arithmetisches Mittel 𝒙̅ = 40

Der Median x̅Z ist kleiner als das arithmetische Mittel x̅. Daher ist die **Altersverteilung rechtsschief bzw. linkssteil.** Das bedeutet, es befinden sich mehr Werte (Ausreißer), die größer als der Zentralwert (Median) x̅Z sind. Es gibt **mehr Ausreißer im oberen Bereich, die das arithmetische Mittel nach oben ziehen (nach oben beeinflussen),** aber auf den Median keinen Einfluss haben.

Die **Gruppe enthält als deutlich mehr Personen, die älter als 32 Jahre sind oder Ausreißer mit Personen, die älter als 32 Jahre sind.**

1. Sie lesen in einer Studie über die Einkommensverteilung einer Berufsgruppe: 𝒙̅𝒁 = 30.000 €, 𝒙̅ = 40.000 €, Q1 = 25.000 €, Q3 = 45.000 €.

Welche Informationen erhalten Sie aus diesen 4 statistischen Kennzahlen über die Einkommensverteilung? Erhalten Sie auch Informationen über die Streuung der Verteilung?

Der Median (Zentralwert) x̅Z hat den Wert = 30.000

Das arithmetische Mittel x̅ hat den Wert = 40.000

Das erste Quantil Q1 hat den Wert = 25.000

Das dritte Quantil Q3 hat den Wert = 45.000

**50% der Gruppe verdienen zwischen 25.000 und 45.000 EUR.**

**50% der Gruppe verdienen weniger als 30.000 EUR (= Median x̅Z), 50% der Gruppe verdienen mehr als 30.000 EUR.**

**Der Interquartilsabstand IQR (Q3 – Q1 = 45.000 – 25.000 = 20.000) gibt die Spannweite bei den mittleren 50% an und informiert über die Streuung zu 50% der Einkommen (der Einkommen im mittleren Bereich)**

**25% der Gruppe verdienen weniger als 25.000 EUR.**

**75% der Gruppe verdienen mehr als 25.000 EUR.**

**75% der Gruppe verdienen weniger als 45.000 EUR.**

**25% der Gruppe verdienen mehr als 45.000 EUR.**

**Die Verteilung ist ungleichmäßig** (Gleichmäßig wäre die Verteilung, wenn der Median x̅Z = 35.000 wäre 🡆 IQR = 20.000 / 2 = 10.000 🡆 Q1 = 25.000 + 10.000 = 35.000)

**Die Einkommensverteilung ist rechtsschief (linkssteil) da der Median x̅Z (30.000 EUR= < das arithmetische Mittel x̅ (40.000 ist). Es gibt Ausreißer bei den oberen Einkommen.**

# Aufgabe 5

Auf die Frage „Wie viel Stück des Produktes ABC haben Sie im letzten Monat gekauft?“ gab es bei der Hauptuntersuchung unterschiedliche Antworten zur Zahl der gekauften Stücke von ABC. Die statistische Reihe wurde zusammengefasst in der folgenden Häufigkeitsverteilung:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **i** | **gekaufte**  **Stückzahl von ABC xi** | **Anzahl der**  **Nennungen**  **n(xi)** | **Für x̅**  **xi \* n(xi)** | f(xi) in % | **x̅**  ∑(xI \* n(xi)) / n  600 / 500 | **xi - x̅** | **(xi - x̅ )²** | (**xi - x̅ ²) \* n(xi)** | H(xi) | F(xi) |
| **1** | xmin **0** | **100** | 0 | 20,0 % | **1,2** | -1,2 | 1,44 | 144,00 | 100 | 20,0% |
| **2** | Modus **1** | **300** | 300 | 60,0 % | **1,2** | -0,2 | 0,04 | 12,00 | 400 | 80,0% |
| **3** | **2** | **50** | 100 | 10,0 % | **1,2** | 0,8 | 0,64 | 32,00 | 450 | 90,0% |
| **4** | **3** | **20** | 60 | 4,0% | **1,2** | 1,8 | 3,24 | 64,80 | 470 | 94,0% |
| **5** | **4** | **20** | 80 | 4,0% | **1,2** | 2,8 | 7,84 | 156,80 | 490 | 98,0% |
| **6** | **5** | **5** | 25 | 1,0% | **1,2** | 3,8 | 14,44 | 72,20 | 495 | 99,0% |
| **7** | xmax **7** | **5** | 35 | 1,0% | **1,2** | 5,8 | 33,64 | 168,20 | 500 | 100,0% |
| **Σ** |  | **n = 500** | **600** | **100,0%** | **-** | - | - | **2.436,00** | - | - |

a) Bestimmen Sie den Modus, den Median und das arithmetische Mittel.

**Der Modus x̅D = 1**

Median: gerade Anzahl der Häufigkeiten

🡆 x̅Z = (xn/2 + xn/2+1) : 2

🡆 xn/2 = 500/2 =250 🡆 250. Wert = 1   
 🡆 xn/2+1 = 500/2+1 =251 🡆 251. Wert = 1

🡆 (1+1) / 2 = 1

**Median x̅Z = 1**

**arithmetische Mittel x̅**

es werden benötigt

* Produkt aus Wert xi zur Merkmalsausprägungen
* absolute Häufigkeit h(xi) pro Merkmalsausprägung
* ∑ (Wert xi zur Merkmalsausprägungen \* absolute Häufigkeit h(xi) pro Merkmalsausprägung)
* Summe der absoluten Häufigkeiten n = 500

**arithmetische Mittel x̅ = ∑xi \* h(xi) / n**

**arithmetische Mittel x̅ =600 / 500**  
**arithmetische Mittel x̅ = 1,2**

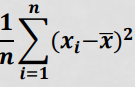
b) Bestimmen Sie die Streuungsparameter Spannweite w, Varianz s2, Standardabweichung s, den Variationskoeffizient v

**Spannweite w = x(max) – x(min) = 7 – 0 = 7**

**Varianz s2**

*Skript Modul 5 – „Streuungsparameter“, VARIANZ auf Seiten 15 bis 29*

* Varianz ist das arithmetische Mittel der Abweichungsquadrate



Varianz s² = ( (x1 – x̅)² \* h(x1) + (x2 – x̅)² \* h(x2) + … + (xi – x̅)² \* h(xi) ) / n

h(xi) = absolute Häufigkeit der Merkmalsausprägung xi

x̅ = arithmetische Mittel

n = Summe der absoluten Häufigkeiten

j = Anzahl der Merkmalsausprägungen x

es werden benötigt:

* Wert xi zur Merkmalsausprägungen
* arithmetisches Mittel x̅ = 1,2
* absolute Häufigkeit h(xi) pro Merkmalsausprägung
* Summe der absoluten Häufigkeiten n = 500

s2 = ( (0 – 1,2)² \* 100 + (1 – 1,2)² \* 300 + (2 – 1,2)² \* 50 + (3 – 1,2)² \* 20 + (4 – 1,2)² \* 20 + (5 – 1,2)² \* 5 + (7 – 1,2)² \* 5) / 500

s2 = (144 + 12 + 32 + 64,8 + 156,8 + 72,2 + 168,2) / 500

s2 = 650 / 500

**Varianz s2 = 1,30**

**Standardabweichung s =**

**Variationskoeffizient**

*Skript Modul 5 – „Streuungsparameter“, STREUUNGSPARAMETER auf Seite 31*

* prozentuale Verhältnis der Standardabweichung zum arithmetischen Mittel
* relatives Streuungsmaß und dimensionslose Größe

c) Welche Aussagen können Sie auf der Basis der Werte der Lageparameter über die Form der Verteilung machen?

*Skript Modul 4 – „Lageparameter“, NEIGUNG / SCHIEFE auf Seiten 30 bis 32*

**Es gibt nur einen Modus.**

**Modus x̅D = Median x̅Z (beide haben den Wert = 1).**

**Das arithmetische Mittel x̅ (= 1,2) > Median x̅Z 🡆 rechtsschiefe (linkssteile) Häufigkeitsverteilung**

**Es gibt Ausreißer bei den oberen Bereich der Verteilung.**

**arithmetische Mittel x̅ (= 1,2) liegt aber nahe dem Median, d. h. x̅ ≈ x̅Z**

x̅D ≈ x̅Z ≈ x̅

**Die Standardabweichung ist mit s = 1,14 sehr klein ≈ 1**

Eine kleine Standardabweichung bedeutet, alle Beobachtungswerte liegen nahe am Mittelwert (kleine Streuung).

Das heißt, die Streuung ist bei dieser Verteilung klein.

Die Häufigkeitsverteilung ist nahe der unimodalen symmetrischen Häufigkeitsverteilung.

# Aufgabe 6

Für 200 Unternehmen liegt für das Jahr 2006 die folgende Umsatzverteilung vor. Vervollständigen Sie die klassierte Häufigkeitsverteilung.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Klasse Nr. i** | **Umsatzklasse**  **(Mio €)** | **Anzahl**  **Unternehmen**  **hi** | **Klassen-mitte**  **mi** | **Für x̅**  **hi \* mi** | **Relative Häufigkeit**  **fi in %** | **x̅**  **arithmet. Mittel**  **(Mio €)** | **absolute**  **Summen-häufigkeit**  **Hi** | **relative**  **Summen-**  **Häufigkeit**  **Fi in %** |
| **1** | **0 b.u. 1** | **60** | **0,5** | 30 | 30,0% | 2,575 | 60 | 30,0% |
| Einfallsklasse k  **2** | **1 b.u.2** | **80** | Modus **1,5** | 120 | 40,0% | 2,575 | 140 | 70,0% |
| **3** | **2 b.u.5** | **40** | **3,5** | 140 | 20,0% | 2,575 | 180 | 90,0% |
| **4** | **5 b.u.10** | **10** | **7,5** | 75 | 5,0% | 2,575 | 190 | 95,0% |
| **5** | **10 b.u.20** | **10** | **15** | 150 | 5,0% | 2,575 | 200 | 100,0% |
| **Σ** |  | **n = 200** | **28** | **515** | **100,0%** |  | - | - |

**Bei klassierten Daten ist der Modus die Klassenmitte zur Klasse mit den größten Häufigkeiten**

**Bei klassierten Daten errechnet sich die Spannweite wie bei den unklassierten Daten aus**

**der Differenz aus Maximum (in der Aufgabe ist Maximum = 20 Mio EUR) – Minimum (in der Aufgabe ist Minimum = 0 Mio EUR).**

**Die Spannweite w ist somit = 20 Mio EUR.**

a) Bestimmen Sie den Modus, den Median und das arithmetische Mittel

**Modus x̅D**

*Skript Modul 4 – „Lageparameter“, MEDIAN BEI KLASSIERTEN DATEN auf Seite 4)*

Bei klassierten Daten ist der Modus die Klassenmitte zur Klasse mit den größten Häufigkeiten

**Modus x̅D = 1,5 Mio EUR (Mitte der Klasse 2 mit größtem hi)**

**Median x̅Z bei klassierten Daten**

*Skript Modul 4 – „Lageparameter“, MEDIAN BEI KLASSIERTEN DATEN auf Seiten 12 bis 16)*

Median x̅Z = (xk-1 + (xk – xk-1) \* (0,5 – Fk-1) / fk

xk-1 = Untergrenze der Einfallsklasse

xk = Obergrenze der Einfallsklasse

Fk-1 = relative Summenhäufigkeit der Klasse unterhalb der Einfallsklasse

fk = relative Häufigkeit der Einfallsklasse

**k = Einfallsklasse 🡆 Klasse mit F(x) = 50%**

**Einfallsklasse k = 2**

x̅Z = 1 + (2 – 1) \* (0,5 – 0,3) / 0,4

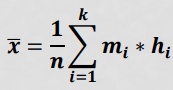
x̅Z = 1 + 1 \* (0,2 / 0,4)

**Median x̅Z = 1,5 Mio EUR**

**arithmetische Mittel x̅ bei klassierten Daten**

*Skript Modul 4 – „Lageparameter“, ARITHMETISCHES MITTEL BEI KLASSIERTEN DATEN auf Seiten 24 bis 28)*

**Bei absoluten Häufigkeiten**



x̅ = ∑(mi \* hi ) / n

mi = Klassenmitte der i-ten Klasse

hi = absolute Häufigkeit der i-ten Klasse

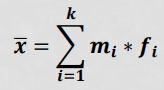
n = Summe der absoluten Häufigkeiten

k = Anzahl der Klassen

x̅ = (0,5 \* 60) + (1,5 \* 80) + (3,5 \* 40) + (7,5 \* 10) + (15 \* 10) / 200

**arithmetische Mittel x̅ = 515 / 200 = 2,575**

**Bei relativen Häufigkeiten**



mi = Klassenmitte der i-ten Klasse

fi = relative Häufigkeit der i-ten Klasse

k = Anzahl der Klassen

**WICHTIG:**

**Bei Berechnung mit relativer Häufigkeit darf im letzten Schritt nicht durch die Summe der absoluten Häufigkeiten dividiert werden, da die relativen Häufigkeiten sich bereits aus der Summe der absoluten Häufigkeiten ergeben.**

x̅ = ∑(mi \* fi )

x̅ = (0,5 \* 0,3) + (1,5 \* 0,4) + (3,5 \* 0,2) + (7,5 \* 0,05) + (15 \* 0,05)

x̅ = 015 + 0,6 + 0,7 + 0,375 + 0,75

**arithmetische Mittel x̅ = 2,575 Mio EUR**

b) Bestimmen Sie die Streuungsparameter Spannweite w, Varianz s2, Standardabweichung s und Variationskoeffizient v.

**Spannweite w bei klassierten Daten**

*Skript Modul 5 – „Streuungsparameter“, SPANNWEITE auf Seiten 9 bis 10*

**Spannweite w wird auch bei klassierten Daten wie bei unklassierten Daten berechnet:**

**Spannweite w = (xmax – xmin) 🡆 Maximum – Minimum**

**Spannweite w = 20 – 0**

**Spannweite w = 20 Mio EUR**

**Varianz s2 bei klassierten Daten**

*Skript Modul 5 – „Streuungsparameter“, VARIANZ BEI KLASSIERTEN DATEN AUS HÄUFIGKEITSTABELLEN auf   
 Seiten 25 bis 29*

**Varianz bei klassierten Daten mit absoluten Häufigkeiten**

**s² = ∑( (mi – x̅ )² \* hi ) / n**

**s² = ( (m1 – x̅ )² \* h1 + (m2 – x̅ )² \* h2 + … (mk – x̅ )² \* hk) / n**

mi = Klassenmitte der i-ten Klasse

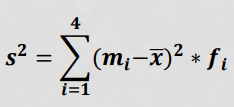
x̅ = arithmetisches Mittel

hi = absolute Häufigkeit der i-ten Klasse

k = Anzahl der Klassen

n = Summe der absoluten Häufigkeiten

**Varianz bei klassierten Daten mit relativen Häufigkeiten**

**s² = ∑( (mi – x̅ )² \* fi ) **

s² = ( (m1 – x̅ )² \* f1 + (m2 – x̅ )² \* f2 + … (mi – x̅ )² \* fi)

mi = Klassenmitte der i-ten Klasse

x̅ = arithmetisches Mittel

fi = relative Häufigkeit der i-ten Klasse

k = Anzahl der Klassen

**Varianz s² =**

(0,5 – 2,575)² \* 0,3 + (1,5 – 2,575)² \* 0,4 + (3.5 – 2,575)² \* 0,2 + (7,5 – 2,575)² \* 0,05 + (15 – 2,575)² \* 0,05

s² = 4,3 \* 0,3 + 1,156 \* 0,4 + 0,856 \* 0,2 + 24,26 \* 0,05 + 154,38 \* 0,05

s² = 1,29 + 0,46 + 0,17 + 1,21 + 7,719

**Varianz s² = 10,85**

**Standardabweichung s**

*Skript Modul 5 – „Streuungsparameter“, STANDARDABWEICHUNG auf Seite 28*

* Stärke zur Aussagekraft des Mittelwertes   
  (Maß dafür, wie hoch die Aussagekraft des Mittelwerts ist)
* kleine Standardabweichung: Alle Beobachtungswerte liegen nahe am Mittelwert (kleine Streuung).
* große Standardabweichung: Beobachtungswerte sind weit um den Mittelwert gestreut
* bei normalverteilten Daten liegen ca. 95% der Beobachtungswerte im Intervall [x̅ - 2s, x̅ + 2s]



**Standardabweichung**

**Variationskoeffizient v**

*Skript Modul 5 – „Streuungsparameter“, STREUUNGSPARAMETER auf Seite 31*

* prozentuale Verhältnis der Standardabweichung zum arithmetischen Mittel
* relatives Streuungsmaß und dimensionslose Größe

**Varianzkoeffizient**

Aufgabe 7 (Klausuraufgabe WS17/18 mit 18 Punkten):

In der folgenden Tabelle ist die Verteilung der männlichen Teilnehmer bei einer Umfrage auf Altersklassen dargestellt. Dabei wurde zwischen Personen mit und ohne Migrationshintergrund unterschieden.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Klassen  Nr | Alter  (von – b.u.) in Jahren | **mit Migrationshintergrund**  **f(x) in %** | **ohne Migrationshintergrund**  **f(x) in %** | Vergleichswert zur Entscheidung des Einfallsklasse  **mit Migrationshintergrund**  **F(x) in %** | Vergleichswert zur Entscheidung des Einfallsklasse  **ohne Migrationshintergrund**  **F(x) in %** |
| 1 | b.u. 15 | 22 | 12 | **22** | **12** |
| 2 | 15 – 35 | 31 | 23 | **53** | **35** |
| 3 | 35 – 55 | 29 | 33 | **82** | **68** |
| 4 | 55 – 75 | 15 | 25 | **97** | **93** |
| 5 | 75 und älter | 3 | 7 | **100** | **100** |

Berechnen Sie approximativ die Alters**quartile** (Q1, Q2, Q3) für die beiden Gruppen.

Alle Quartile auf ganze Zahlen runden.

TIPP: vervollständigen Sie zuerst die Häufigkeitstabelle.

**Berechnung der Quartile Q1, Q2, Q3 für die Gruppe „mit Migrationshintergrund“)**

Berechnung von Q1, Q2, Q3 wie Median für klassierte Daten! (s. Folie 13, 15, 16 Modul 4)



xk-1 = Untergrenze der Einfallsklasse

xk = Obergrenze der Einfallsklasse

Fk**-1** = relative Summenhäufigkeit der Vorklasse

fk = relative Häufigkeit

Einfallsklasse ist die zum Quartil betrachtete Klasse

**Einfallsklassen für Q1, Q2, Q3 für die Gruppe mit Migrationshintergrund**

für jedes Quartil muss die Einfallsklasse bestimmt werden.

Q1 **25% aller Beobachtungswerte** (Summe der absoluten oder relativen Häufigkeiten)   
 sind kleiner als 1. Quartil

Besser ausgedrückt:   
 Beobachtungswert des 1. Quartil muss ≥ 25% sein,   
 d. h. nehme die erste Klasse, die einen Wert ≥ 25% hat  
 **Wichtig:** Beobachtungswert muss der relativen Häufigkeit entsprechen  
 ist nur absolute Häufigkeit bekannt, muss zu jeder Klasse bzw. zu jeder Merkmalsausprägung die relative Häufigkeit ermittelt werden

Q2 **50% aller Beobachtungswerte** (Summe der absoluten oder relativen Häufigkeiten)   
 sind kleiner als 2. Quartil  
 Besser ausgedrückt:   
 Beobachtungswert des 2. Quartil muss ≥ 50% sein

d. h. nehme die erste Klasse, die einen Wert ≥ 50% hat

Q3 **75% aller Beobachtungswerte** (Summe der absoluten oder relativen Häufigkeiten)   
 sind kleiner als 3. Quartil

Besser ausgedrückt:   
 Beobachtungswert des 3. Quartil muss ≥ 75% sein  
 d. h. nehme die erste Klasse, die einen Wert ≥ 75% hat

**Einfallsklasse für Q1 ist K2 (Klasse „15 – 35“)**,  
da F(x) in % in dieser Klasse == 53  
F(x) in % in der Klasse K1 („b.u. 15“ ) = 22 🡆 22 < 25 (der Klassenwert muss aber ≥ 25 sein!),

Klasse K2 („15 – 35“) == 53 🡆53 ≥ 25, daher nehme Klasse 2 mit F(x) = 53

**Q1 (25%) k = 2 (15-35) Q1 = 15 + (35 – 15) \* (0,25 – 0,22) : 0,31 = 17**

**Einfallsklasse für Q2 ist K2 (Klasse „15 – 35“)**,  
da F(x) in % in dieser Klasse K2 („15 – 35“) == 53  
F(x) in % in Klasse K1 („b. u. 15“) == 22 🡆 22 ≤ 50 (der Klassenwert muss aber > = 50 sein!)

Klasse K2 („15 – 35“) == 53 🡆53 ≥ 50, daher nehme Klasse 2 mit F(x) = 53

**Q2 (50%) k = 2 (15-35) Q2 = 15 + (35 – 15) \* (0,50 – 0,22) : 0,31 = 33**

**Einfallsklasse für Q3 ist K3 (Klasse „35 – 55“)**,  
da F(x) in % in dieser Klasse K3 („35 – 55“) == 82  
F(x) in % in Klasse K2 („15 – 35“) == 53 🡆 53 ≤ 75 (der Klassenwert muss aber > = 75 sein!)

Klasse K3 („35 – 55“) == 82 🡆 82 ≥ 75, daher nehme Klasse 3 mit F(x) = 82

**Q3 (75%) k = 3 (35-55) Q3 = 35 + (55 – 35) \* (0,75 – 0,53) : 0,29 = 50**

**Berechnung der Quartile Q1, Q2, Q3 für die Gruppe „ohne Migrationshintergrund“)**

**Q1 (25%) k = 2 („15 – 35“)**  **Q1 = 15 + (35 – 15) \* (0,25 – 0,12) : 0,23 = 26**

**Q2 (50%) k = 3 („35 – 55“)** **Q2 = 35 + (55 – 35) \* (0,50 – 0,35) : 0,33 = 44**

**Q3 (75%) k = 4 („55 – 75“)**  **Q3 = 55 + (75 – 55) \* (0,75 – 0,68) : 0,25 = 61**

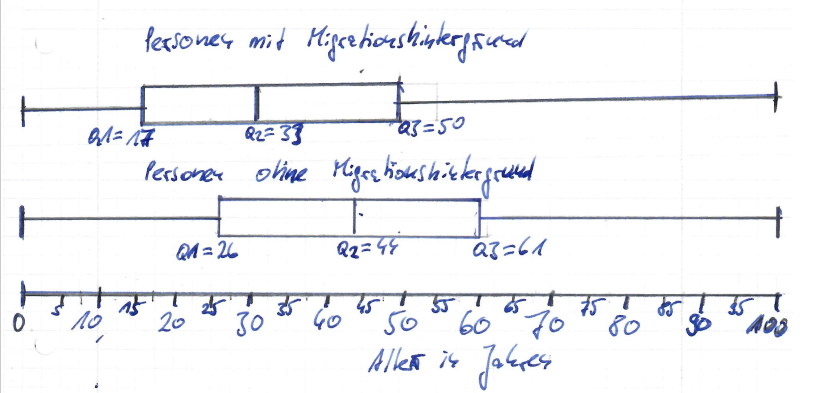
b)

Zeichnen Sie die Boxplots für beide Gruppen in einem Diagramm.

Bitte denken Sie an die „Lesbarkeit“ des Boxplots (Achsenbeschriftung und Legende nicht vergessen!)

Boxplot (Box-and-Whisker-Plot) beinhaltet

1. zur Spannweite w das Minimum
2. zur Spannweite w das Maximum
3. das Q1 Quartil
4. den Median (= das Q2 Quartil)
5. das Q3 Quartil



c)

Wie groß ist der Anteil der Personen im Alter zwischen 15 und 75 Jahre für jede Gruppe? (gemeint ist das Intervall [15;75))

**Anteil der Personen im Alter zwischen 15 und 75 für die Gruppe „mit Migrationshintergrund“**

Werte (= Anteil der Personen) aus der Spalte mit „Migrationshintergrund f(x) in %“ (relative Häufigkeit) zu den Klassierungen „15 – 35“, „35 – 55“, „55 – 75“ addieren

*oder (2. Lösungsweg)*

Wert (= Anteil Personen) aus der Spalte mit „Migrationshintergrund f(x) in %“ zur Klassierung „b. u. 15“ (Klasse 1) vom **kumulierten** Anteil Personen aus Klassierung „55 – 75“ (Klasse 4) subtrahieren:

Klasse Alter in Jahren Anteil in %

2 15 – 35 31

3 35 – 55 29

4 55 – 75 15

**∑ 75**

*oder (2. Lösungsweg)*

Klasse Alter in Jahren Anteil in %

1 b. u. 15 22

4 55 - 75 97

**97 – 22 = 75**

**Anteil der Personen im Alter zwischen 15 und 75 für die Gruppe „ohne Migrationshintergrund“**

Klasse Alter in Jahren Anteil in %

2 15 – 35 23

3 35 – 55 33

4 55 – 75 25

**∑ 81%**

**WICHTIG: % - Zeichen nicht vergessen!**

Klasse Alter in Jahren Anteil in %

1 b. u. 15 12

4 55 - 75 93

**93 – 12 = 81%**

**WICHTIG: % - Zeichen nicht vergessen!**

# Aufgabe 8

Die Tabelle zeigt die Zahl der Eheschließungen bzw. die Zahl der Ehescheidungen je 10.000 Ehen in Deutschland auf (Quelle: Statistisches Bundesamt):

a)

Bestimme den Modalwert, den Zentralwert und das arithmetische Mittel sowohl von den Eheschließungen als auch von den Ehescheidungen. Was fällt dir an den Ergebnissen auf?

b)

Bestimme die Varianz und die Standardabweichung sowohl von den Eheschließungen als auch von den Ehescheidungen.

c)

Welche Veranschaulichungsmöglichkeiten für solch einen tabellarischen Zusammenhang hast du bereits kennen gelernt?   
Wähle zwei davon aus und realisiere sie! Welche Visualisierungsform ist in diesem Fall besonders geeignet bzw. ungeeignet und warum?

|  |  |  |
| --- | --- | --- |
| **Jahr** | **Eheschließungen** | **Ehescheidungen je**  **10.000 Ehen** |
| 2001 | 389.000 | 198,2 |
| 2000 | 418.550 | 194 |
| 1999 | 430.674 | 187,7 |
| 1998 | 417.420 | 191,4 |
| 1997 | 422.776 | 181,2 |
| 1996 | 427.297 | 161 |
| 1995 | 430.534 | 153,8 |
| 1994 | 440.244 | 150 |
| 1993 | 442.605 | 135,6 |
| 1992 | 452.428 | 104,8 |
| 1991 | 454.291 | 104 |

**Eheschließungen**

**Modus (Modalwert) x̅D = 454.291**

*Die richtige Lösung zum Modalwert ist: Der Modalwert der Eheschließungen in Deutschland je Ehe existiert nicht, denn jeder Wert kommt nur ein einziges Mal vor. Kein Wert der Datenreihe ist ein Modalwert!*

**Median (Zentralwert) x̅Z = 430.534**

**arithmetische Mittel x̅ ≈ 429.620**

x̅ < x̅Z 🡺 linksschiefe (rechtssteile) Verteilung

Das arithmetische Mittel x̅ ≈ **429.619**,91 ist nah am Median x̅D = 430.534. Die Verteilung/Entwicklung ist gleichmäßig. Es gibt keine Ausreißer bzw. Ausreißer mit geringer Abweichung.

Varianz s² =

Varianz s² = (389.000 – **429.619**)² + (418.550 - **429.619**)² + (430.674 - **429.619**)² + (417.420 - **429.619**)² + (422.776 - **429.619**)² + (427.297 - **429.619**)² + (430.534 - **429.619**)² + (440.244 - **429.619**)² + (442.605 - **429.619**)² + (452.428 - **429.619**)² + (454.291 - **429.619**)² ) / n

**Varianz s² = 307.808.635**

**Standardabweichung s = 17.545**

**Ehescheidungen je 10.000 Ehen**

**Modus (Modalwert) x̅D = 198,20 je 10.000 Ehen**

*Die richtige Lösung zum Modalwert ist: Der Modalwert der Ehescheidungen in Deutschland je Ehe existiert nicht, denn jeder Wert kommt nur ein einziges Mal vor. Kein Wert der Datenreihe ist ein Modalwert!*

**Median (Zentralwert) x̅Z = 161,0 je 10.000 Ehen**

**arithmetische Mittel x̅ ≈ 160 je 10.000 Ehen**

x̅ < x̅Z 🡺 linksschiefe (rechtssteile) Verteilung

Das arithmetische Mittel x̅ = **160,15** ist nah am Median.

Das arithmetische Mittel x̅ ≈160 ist nah am Median x̅Z = 161,0.

Die Verteilung/Entwicklung ist gleichmäßig. Es gibt keine Ausreißer bzw. Ausreißer mit geringer Abweichung.

Varianz s² =

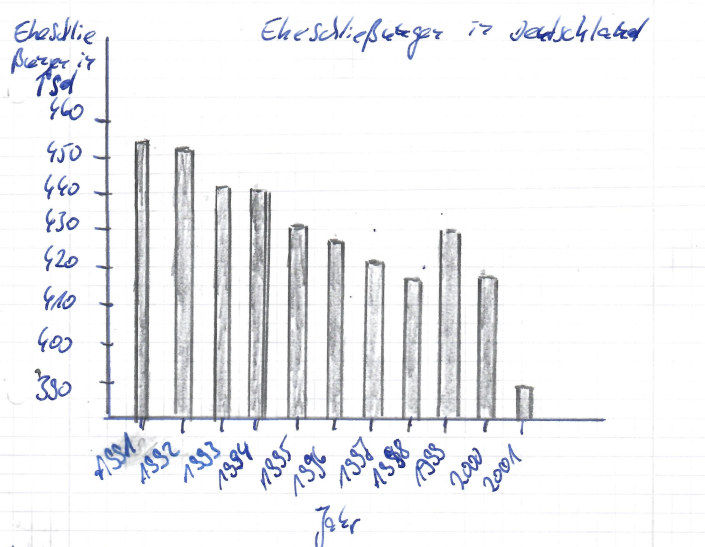
Varianz s² = (198,2 - **160,15**)² + (194 - **160,15**)² + (187,7 - **160,15**)² + (191,4 - **160,15**)² + (181,2- **160,15**)² + (161 - **160,15**)² + (153,8 - **160,15**)² + (150 - **160,15**)² + (135,6 - **160,15**)² + (104,8 - **160,15**)² + (104 - **160,15**)²

**Varianz s² = 1.066,89 je 10.000 Ehen**

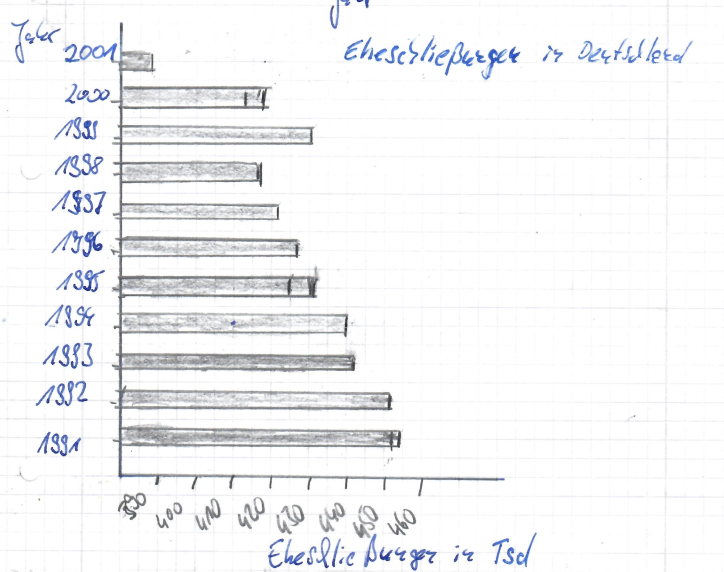
**Standardabweichung s = 32,66 je 10.000 Ehen**

Welche Veranschaulichungsmöglichkeiten für solch einen tabellarischen Zusammenhang hast du bereits kennen gelernt?   
Wähle zwei davon aus und realisiere sie! Welche Visualisierungsform ist in diesem Fall besonders geeignet bzw. ungeeignet und warum? Für die Veranschaulichung der Datenreihe eigenen sich Stabdiagramm, Säulendiagramm, Balkendiagramm und Graph sehr gut.

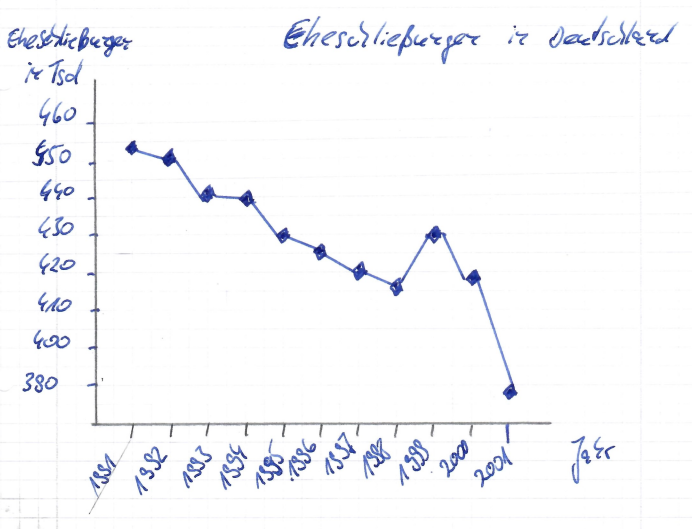
**Stabdiagramm**



**Balkendiagramm**



**Graph**



Das Kreisdiagramm ist ungeeignet, da das Kreisdiagramm zur besseren Übersichtlichkeit max. 7 Teilwerte enthalten sollte. Ferner ist die Darstellung von metrischen Werte im Kreisdiagramm unübersichtlich. Das Kreisdiagramm eignet sich besser für die Darstellung von diskreten Werten, besonders für das Nominal- und Ordinal-Skalenniveau. Die wichtigsten Informationen lassen sich zu einer Datenreihe mit metrischen Werten nicht gut veranschaulichen.

Aufgabe 9 (Klausuraufgabe WS18/19 mit 9 Punkten):

Um die Entwicklung der Telefonkosten der letzten 6 Monate des vergangenen Jahres zu analysieren, wird Claudia von ihrem Vater beauftragt, die mittleren Telefonkosten sowie deren Streuung zu berechnen. Die Telefonkosten (in €) sind in der folgenden Tabelle aufgeführt.

*Berechnung mit 2 Nachkommastellen!*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Monat** | Juli | August | September | Oktober | November | Dezember |
| **Kosten (€)** | **31,44** | **30,18** | **31,04** | **33,60** | **38,16** | **132,40** |

a) **(5 Punkte)**

Berechnen Sie das **arithmetische Mittel sowie die Streuung der Telefonkosten**.b) **(4 Punkte)**

**arithmetische Mittel x̅**

arithmetische Mittel x̅ = ∑xi / n

**n = 6**

arithmetische Mittel x̅ = 31,44 + 30,18 + 31,04 + 33,60 + 38,16 + 132,40 / 6

arithmetische Mittel x̅ = 296,82 / 6

**arithmetische Mittel x̅ = 49,47 €**

**Streuung der Telefonkosten**

Varianz s²

Varianz s² = ( (xi – x̅ )² \* hi  )/ n

**Varianz s² = ( (31,44 – 49,47)² \* 1 + (30,18 – 49,47)² \* 1 + (31,04 – 49,47)² \* 1 + (33,60 – 49,47)²   
 \* 1 + (38,16 – 49,47)² \* 1 + (132,40 – 49,47)² \* 1) / 6**

Varianz s² = (325,08 \* 1 + 372,10 \* 1 + 339,66 \*1 + 251,86 \* 1 +127,91 \* 1 +6877,38) / 6

**schnellerer Rechenweg:**

**alle Beobachtungswerte quadrieren und summieren 🡺 danach durch die Summe der   
 absoluten Häufigkeiten dividieren**

**im letzten Schritt das arithmetische Mittel x̅ quadrieren und von der Summe subtrahieren**

**Varianz s² = ( (x1² + x2² + … + xi²) / n ) – x̅²**

**Varianz s² = (31,44² + 30,18² + 31,04² + 33,60² + 38,16² + 132,40²) / 6 ) - 49,47²**

**Varianz s² = (988,47 + 910,83 + 963,48 + 1128,96 + 1456,19 + 17529,76) / 6) – 2447,28**

**Varianz s² = (22977,69 / 6) – 2447,28**

**Varianz s² = 3829,69 – 2447,28**

**Varianz s² = 1382,34 EUR**

**Varianz s² = 1382,34 EUR**

**Standradabweichung =**

Claudia, die im Dezember häufig bei teuren Hotlines angerufen hat, ist entsetztüber den hohen Mittelwert und befürchtet Taschengeldentzug durch ihren Vater.

Helfen Sie Claudia aus der Patsche, indem Sie ein alternatives Lageparameter, zu Claudias Gunsten, vorschlagen.

Begründen Sie Ihren Vorschlag kurz (maximal drei Sätze) und berechnen Sie den Wert Ihres vorgeschlagenen Lagemaßes.

Gedanken zu Lösung:

Welche Lageparameter stehen zur Wahl?

Modus x̅D

Der Modus ist Beobachtungswert mit der maximalen Häufigkeit (am häufigsten auftretenden Merkmalsausprägung).

**Der Modus x̅D = 132,40 EUR**

Der Modus ist mit **132,40 EUR** für Claudia zu hoch und für die Lösung der Aufgabenstellung ungeeignet.

arithmetische Mittel (Mittelwert ) x̅

Ist der Durchschnittswert zu allen Beobachtungswerten.

**Das arithmetische Mittel x̅ = 49,47 EUR**

Das arithmetische Mittel ist mit 49,47 EUR für Claudia zu hoch und für die Lösung der Aufgabenstellung ungeeignet.

Das Q1 Quantil beinhaltet 25% aller geordneten Beobachtungswerte einer geordneten Reihe

Werden die Telefonkosten nach Größe sortiert ergibt sich folgende Sortierung

30,18

31,04

31,44

33,60

38,16

132,40

Der Wert zum Q1 Quantil ergibt sich aus

Summe der absoluten Häufigkeiten = 6

Positionen zum Q2 Quantil: (6 / 2 =3) + (6 / 2 +1 = 4)

Positionen zum Q1 Quantil: (6/2 = 3)+1 / 4 =2

6 / 2 = 3 +1 / 2 =2

Wert zum Q1-Quantil aus der 2. Position = 31,04 EUR.

**Antwort zur Lösung:**

Der Lageparameter Q1 eignet sich zur Lösung.

25% aller geordneten Beobachtungswerte < das Q1 Quantil.

Der Wert zum Q1 Quantil = 31,04 EUR < als das arithmetische Mittel 49,47 EUR

Damit ist der Wert zum Q1 Quantil < als das arithmetische Mittel x̅ und der Median x̅Z.

**Lösung von Frau Dr. Merrins**

Der Median x̅ ist weniger empfindlich gegenüber Ausreißern und daher der bessere Vorschlag.

Zuvor ist aber die Reihe zu sortieren

30,18, 31,04, **31,44, 33,60**, 38,16, 132,40

Die Summe der Häufigkeit n = 6 🡺 Ermittlung des Medians:

**( 6 : 2 ) = 3 + ( 6 : 2 +1 ) = 4**

Der Median x̅Z ergibt sich aus dem 3. und 4. Wert:

**x̅Z = ½ \* (31,44 + 33,60 ) = ½ \* 65,04 = 32,52 EUR**