**Wirtschaftsstatistik**

**WiSe 2023/2024**

Dozentin: Frau Dr. Merrins

**Zusammenfassung der Präsenzen**

**Inhaltsverzeichnis**

[1. Präsenz 20.10.2023 - Häufigkeiten und Häufigkeitsverteilungen 3](#_Toc149125155)

# 1. Präsenz 20.10.2023 - Häufigkeiten und Häufigkeitsverteilungen

**Aufgabe:**

Den aktuellen Kurs „Wirtschaftsstatistik“ belegen 42 Studierende

Davon belegen den Kurs 18 Studierende das erste Mal und 16 Studierende das zweite Mal. 8 Studierende haben der Kurs bereits mehr als zweimal belegt.

a)

Was ist das Merkmal?

Begründen Sie!

b)

Was ist die Merkmalsausprägung?

Begründen Sie!

c)

Erstellen Sie eine eindimensionale Häufigkeitsverteilung mit Summenhäufigkeitsverteilung zur Merkmalsausprägung mit der dazugehörigen Häufigkeit.

Alle Werte in der Häufigkeitsverteilung sind auf 2 NK-Stellen zu runden.

d)

Erstellen Sie ein Diagramm zur eindimensionalen Häufigkeitsverteilung.

Wählen Sie zur Darstellung einen geeigneten Diagrammtyp.

e)

Erstellen Sie ein Diagramm zur empirischen Verteilungsfunktion.

Wählen Sie zur Darstellung einen geeigneten Diagrammtyp.

**Lösungen:**

**a)**

**Was ist das Merkmal?**

**Begründen Sie!**

Das Merkmal ist die Anzahl der Studierenden (Teilnehmer) im Kurs.

Dabei handelt es sich um ein quantitatives Merkmal. Quantitative Merkmale sind metrische Merkmale, d. h. Merkmale die mit Zahlen gemessen werden. Die Anzahl der Studierenden (Teilnehmer) ist ein diskretes Merkmal: Die Menge der Merkmalsausprägungen ist endlich/abzählbar (i.d.R. ganze Zahlen).

**b)**

**Was ist die Merkmalsausprägung?**

**Begründen Sie!**

Die Belegungen des Kurses sind die Merkmalsausprägungen.

Merkmalsausprägungen entsprechen den möglichen Ausformungen eines Merkmals (Wert der Variable oder Beobachtungswert). In der Aufgabenstellung sind die Kursbelegungen die möglichen Varianten bzw. der Beobachtungswert.

**c)**

**eindimensionale Häufigkeitsverteilung mit Summenhäufigkeitsverteilung**

**(zur Merkmalsausprägung mit der dazugehörigen Häufigkeit)**

**Erstellen Sie eine eindimensionale Häufigkeitsverteilung mit Summenhäufigkeitsverteilung zur Merkmalsausprägung mit der dazugehörigen Häufigkeit.**

**Alle Werte in der Häufigkeitsverteilung sind auf 2 NK-Stellen zu runden.**

**Schritt 1:**

Sortieren der Daten → geordnete Reihe nach irgendeiner Ordnung, z. B. alphabetische Ordnung der Merkmalsträger oder Größenordnung der Merkmalsausprägung

**Schritt 2:**

Verdichten der sortierten Daten auf Merkmalsausprägungen und zählen, wie oft diese vorkommen

→ **geordnete Menge von Wertepaaren (Merkmalsausprägung und zugehörige Häufigkeit) heißt Häufigkeitsverteilung**

**Schritt 3:**

Darstellen tabellarisch von nach Merkmalsausprägungen sortierten Häufigkeitsverteilungen

→ die Häufigkeitstabelle

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Merkmalsaus-**  **Prägung: Anzahl der Belegungen** | | **Merkmal: Anzahl der Teilnehmer** | | | |
|  | xi | absolute Häufigkeit  h(xi) | relative Häufigkeit  f(xi) | absolute  Summenhäufigkeit  H(xi) | relative Summenhäufigkeit  F(xi) |
| 1 | 1. Belegung | 18 | 0,43 | 18 | 0,43 |
| 2 | 2. Belegung | 16 | 0,38 | 34 | 0,81 |
| 3 | 3. Belegung | 8 | 0,19 | 42 | 1 |
| Summe | | 42 | 1 | - | - |

**WICHTIG:**

* **Summenzeile nicht vergessen!**
* Die Summe zur absoluten Häufigkeit h(xi) ist gleich Anzahl der Beobachtungen (42).   
  Das entspricht der Urliste (Werteliste) von 42 Studenten).

**relative Häufigkeit f(xi) = absolute Häufigkeit h(xi) / Summe der Merkmalsausprägungen**

z. B. relative Häufigkeit f(xi) für die 1. Belegung: relative Häufigkeit f(xi) = 18 / 42 🡺 auf 2 NK-Stellen gerundet

WICHTIG:

Die Werte zur relativen Häufigkeit h(xi) können als Dezimalwerte (z. B. 0,43) oder Prozentwerte angegeben werden.

Frau Merrins empfiehlt Dezimalwerte, da diese in Diagrammen einfacher darzustellen sind.

Die Summe in der Spalte „relative Häufigkeit f(xi)“ muss 1 (oder bei Angabe von Prozentwerten 100) ergeben!

**absolute Summenhäufigkeit H(xi) = absolute Häufigkeit h(xi) + absolute Häufigkeit h(xi) aus der vorausgehenden Zeile**

z. B. absolute Summenhäufigkeit H(xi) für die 2. Belegung: 18 + 16 (absolute Summenhäufigkeit aus der 1. Zeile (zur 1. Belegung) + absolute Summenhäufigkeit aus der 2. Zeile (zur 2. Belegung)

WICHTIG:

Die absolute Summenhäufigkeit H(xi)zur letzten Merkmalsausprägung ist gleich der Summe aller Merkmalsausprägungen (in der Aufgabe = 42)

In der Summenzeile wird zur absoluten Summenhäufigkeit H(xi) ist kein Wert bzw. „-“ eingetragen.

**relative Summenhäufigkeit F(xi) = relative Häufigkeit f(xi) + relative Häufigkeit f(xi) aus der vorausgehenden Zeile**

z. B. relative Summenhäufigkeit F(xi) für die 2. Belegung: 0,43 + 0,38 (relative Summenhäufigkeit aus der 1. Zeile (zur 1. Belegung) + relative Summenhäufigkeit aus der 2. Zeile (zur 2. Belegung) = 0,81

WICHTIG:

Die relative Summenhäufigkeit F(xi)zur letzten Merkmalsausprägung muss 1 (oder bei Angabe von Prozentwerten 100) ergeben!

In der Summenzeile wird zur relative Summenhäufigkeit F(xi) ist kein Wert bzw. „-“ eingetragen.

**d)**

**Diagramm zur eindimensionalen Häufigkeitsverteilung**

**Erstellen Sie ein Diagramm zur eindimensionalen Häufigkeitsverteilung.**

**Wählen Sie zur Darstellung einen geeigneten Diagrammtyp.**

Die Merkmalsausprägungen sind absolut skalierte Merkmale (Anzahlen und Stückzahlen)

Allgemein: Häufigkeiten oder alles, was man zählen kann.

z. B. Anzahl der Beschäftigten oder wie im Bsp. Anzahl der Studenten.

**Die Anzahl der Studenten hat somit das Skalenniveau „metrische Skala“ und die Skalenart „Absolutskala“, da die Anzahl nur ganzen Werte, also keinen Fließkommawerten, entspricht.**

Zur grafischen Darstellung der Skalenart „Absolutskala“ eignet sich der diagrammtyp „Säulendiagramm“, da diese für die Präsentation von Zählwerten (ganze Zahlen) gut geeignet ist.

Unterschied zwischen qualitativen und quantitativen Merkmalen und 🡺 innerhalb der quantitativen Merkmale den Unterschied zwischen stetig und diskret lernen.

zeitlich, sachlich, räumlich

(Modul\_2\_Skalen\_und\_Klassierung.pdf, Folie 4)

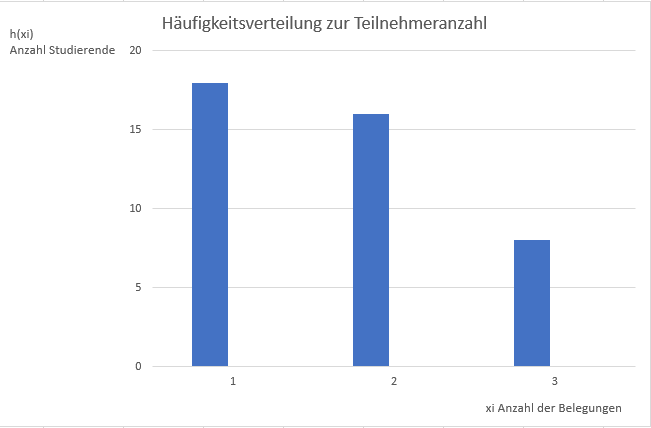
metrische Skalen und nicht metrische Skalen lernen

(Modul\_2\_Skalen\_und\_Klassierung.pdf, Folien 5 bis 7)

Bezeichnungen für Diagramm können sein:

„grafische Darstellung“ oder „Diagramm“ „Veranschaulichung“ etc.

Diagrammtypen und Darstellung lernen, kommen auf jeden Fall in der Klausur vor!



**WICHTIG:**

* Diagrammtitel
* Achsenbeschriftung
* zur absoluten Häufigkeit passende Skaleneinteilung (0, 5, 15, 20) (y-Achse)  
  beste Wahl zur Skaleneinteilung:  
  arithmetisches Mittel aus der Summe aller Differenzen aus den Werten in der Spalte „ absolute Häufigkeit f(xi)“ 🡆 im Bsp. (18 – 16) + (16 – 8) / 2 = 5
* Skaleneinteilung zu den Merkmalsausprägungen (0, 1, 2, 3) (x-Achse)
* 0-Punkt nicht vergessen!
* An den Achsenenden Pfeile -> (im Bsp. nicht enthalten)

**e)**

**Diagramm mit Treppenfunktion zur empirischen Verteilungsfunktion**

**Erstellen Sie ein Diagramm zur empirischen Verteilungsfunktion.**

**Wählen Sie zur Darstellung einen geeigneten Diagrammtyp.**

Die Verteilungsfunktion enthält die gesamte Information, die in den Daten steckt. Die ursprüngliche Reihenfolge geht verloren.

Die **empirische Verteilungsfunktion F(x) ist eine relative Summenhäufigkeitskurve**

**relative Summenfunktion:**

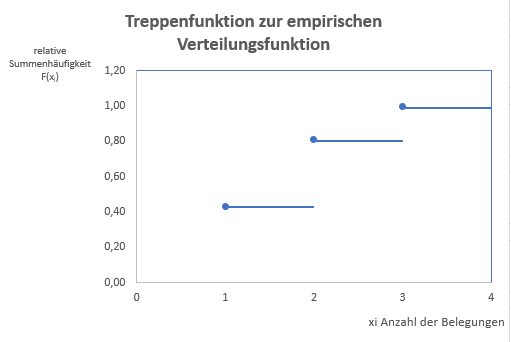
* Die empirische Verteilungsfunktion F(x) gibt für jede beliebige reelle Zahl x den **Anteil der Merkmalsträger** an, für die das Merkmal X einen Wert xi annimmt, der kleiner oder gleich x ist  
  F(x) ist ein Anteil von X 🡆 z. B. ist 0,81 der relative Summenhäufigkeitsanteil des Merkmals X (Gesamtanzahl der Studenten) = 42
* Wertebereich: 0 ≤ F(x) ≤ 1
* F(x) ist monoton nichtfallend (steigt oder ist konstant)
* F(x) ist eine Treppenfunktion mit Sprungstellen bei x1 , x2 , ..., xi
* Die Größe der Sprünge beträgt fi = F(xi ) - F(xi-1 )

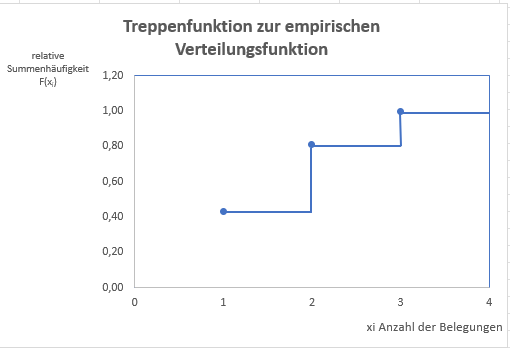
|  |  |  |  |
| --- | --- | --- | --- |
| **Merkmalsaus-**  **Prägung: Anzahl der Belegungen** | | **Merkmal: Anzahl der Teilnehmer** | |
| i | xi | absolute  Summenhäufigkeit  H(xi) | **relative Summenhäufigkeit**  **F(xi)** |
| 1 | 1. Belegung | 18 | **0,43** |
| 2 | 2. Belegung | 34 | **0,81** |
| 3 | 3. Belegung | 42 | **1,00** |
| Summe | | - | - |

**WICHTIG:**

* **Summenzeile nicht vergessen!**
* **relative Summenhäufigkeit F(xi) = relative Häufigkeit f(xi) + relative Häufigkeit f(xi) aus der vorausgehenden Zeile**

z. B. relative Summenhäufigkeit F(xi) für die 2. Belegung: 0,43 + 0,38 (relative Summenhäufigkeit aus der 1. Zeile (zur 1. Belegung) + relative Summenhäufigkeit aus der 2. Zeile (zur 2. Belegung) = 0,81



****

**WICHTIG:**

* Diagrammtitel
* Achsenbeschriftung
* zur relativen Summenhäufigkeit passende Skaleneinteilung (0, 0,20, 0,40, 0,60, 1,00) (y-Achse);   
  beste Wahl zur Skaleneinteilung:  
  geringste Differenz aus den Werten in der Spalte „ relative Summenhäufigkeit F(xi)“ 🡆 im Bsp. AUFRUNDEN(1,00 – 0,81) = 0,20 oder   
  arithmetisches Mittel aus der Summe aller Differenzen aus den Werten in der Spalte „ relative Summenhäufigkeit F(xi)“ 🡆 im Bsp. ( 1,00 - 0,81) + (0,81 - 0,43) / 2 = 0,3
* Skaleneinteilung zu den Merkmalsausprägungen (0, 1, 2, 3) (x-Achse)  
  letzter Wert (4) in der Skaleneinteilung ist nicht anzugeben (ist durch Excel vorgegeben)
* Die oberste Linie zum Wert 1 muss horizontal noch über den Wert 3 im Diagramm gezeichnet werden, d. h. die x-Achse muss etwas länger als 3 gezeichnet werden
* 0-Punkt nicht vergessen!
* An den Achsenenden Pfeile -> (im Bsp. nicht enthalten)
* 2 Arten zur Darstellung der Treppenfunktion möglich: wie im Bsp. 1 als Strichpunkt oder als Treppe, d. h. vertikale Verbindung zwischen den Werten (sieh Bsp. 2)

**2-dimensionale und 3-dimensionale Häufigkeitsverteilung**

**(2-dimensionale und 3-dimensionale Kreuztabellen)**

Aufgabe:

Im Rahmen einer Marktforschungsstudie wurden n = 12 Personen u.a. gefragt nach den **drei Merkmalen**

* Geschlecht G (w = weiblich, m = männlich),
* Alter A (Alter in Jahren) und
* Markenpräferenz M (A = Produkt A, B = Produkt B).

Die Erhebung ergab die folgenden 12 Befragungsergebnisse (Beobachtungswertekombinationen):

(w, 37, A), (m, 65, A), (w, 26, A), (m, 37, B), (w, 21, B), (m, 29, A),

(w, 52, B), (m, 43, A), (w, 48, A), (m, 58, B), (w, 24, A), (m, 58, B).

Lesebeispiel:

Die 1. Person ist weiblich, 37 Jahre alt und bevorzugt Produkt A.

a)

Erstellen Sie die zwei folgenden zweidimensionalen Kreuztabellen:

Geschlecht x Markenpräferenz

klassiertes Alter x Markenpräferenz

(2 Altersklassen: 1. Klasse: bis unter 40 Jahre, 2. Klasse: 40 Jahre und älter)

b)

Versuchen Sie eine dreidimensionale Kreuztabelle zu erstellen für die drei Merkmale:

Geschlecht x klassiertes Alter x Markenpräferenz

c)

Welche der drei Merkmale kann man als unabhängige bzw. abhängige Merkmale betrachten?

d)

Sie erheben bei einer Grundgesamtheit Daten für 3 Merkmale.

Was liefert mehr Information: die dreidimensionale Häufigkeitsverteilung oder alle drei möglichen ein- und zweidimensionalen Häufigkeitsverteilungen zusammen?

**Lösungen:**

**a)**

**Erstellen Sie die zwei folgenden zweidimensionalen Kreuztabellen:**

* **Geschlecht x Markenpräferenz**
* **klassiertes Alter x Markenpräferenz**

**(2 Altersklassen: 1. Klasse: bis unter 40 Jahre, 2. Klasse: 40 Jahre und älter)**

**zweidimensionalen Kreuztabelle: Geschlecht x Markenpräferenz**

relative Spaltenhäufigkeit

Zellenwert / Spaltensumme

z. B. 3,00 / 6,00 = 0,50

Die Summe aller relativen Spaltenhäufigkeiten = 1

z. B. 0,50 + 0,50 = 1

relative Zeilenhäufigkeit

Zellenwert / Zeilensumme

z. B. 3,00 / 7,00 = 0,43

Die Summe aller relativen Zeilenhäufigkeiten = 1

z. B. 0,43 + 0,57 = 1

relative Häufigkeit der Merkmalsausprägungskombination

Zellenwert / Gesamtanzahl aller Ausprägungen

z. B. 3,00 / 12,00 = 0,25

Die Summe aller relativer Häufigkeiten zur Merkmalsausprägungs-kombination= 1

z. B. 0,25 + 0, 33 + 0,25 + 0,17

|  |  |  |  |
| --- | --- | --- | --- |
| Merkmal Markenpräferenz M | Merkmal  Geschlecht G | | |
| m | w | **Summe** |
| **Produkt A**  relative Spaltenhäufigkeit  relative Zeilenhäufigkeit  relative Häufigkeit der Merkmalsausprägungskombination | **3,00**  0,50  0,43  0,25 | **4,00**  0,67  0,57  0,33 | **7,00**  0,58 |
| **Produkt B**  relative Spaltenhäufigkeit  relative Zeilenhäufigkeit  relative Häufigkeit der Merkmalsausprägungskombination | **3,00**  0,50  0,60  0,25 | **2,00**  0,33  0,40  0,17 | **5,00**  0,42 |
| **Summe** | **6,00**  0,50 | **6,00**  0,50 | **12,00**  1,00 |

**zweidimensionalen Kreuztabelle: klassiertes Alter x Markenpräferenz**

2 Altersklassen:

* Klasse: bis unter 40 Jahre
* Klasse: 40 Jahre und älter)

|  |  |  |  |
| --- | --- | --- | --- |
| Merkmal Markenpräferenz M | Merkmal  Klassiertes Alter A | | |
| b. u. 40 | 40+ | Summe |
| **Produkt A**  relative Spaltenhäufigkeit  relative Zeilenhäufigkeit  relative Häufigkeit der Merkmalsausprägungskombination | **4,00**  0,67  0,57  0,33 | **3,00**  0,50  0,43  0,25 | **7,00**  0,58 |
| **Produkt B**  relative Spaltenhäufigkeit  relative Zeilenhäufigkeit  relative Häufigkeit der Merkmalsausprägungskombination | **2,00**  0,33  0,40  0,17 | **3,00**  0,50  0,60  0,25 | **5,00**  0,42 |
| Summe | **6,00**  0,50 | **6,00**  0,50 | **12,00**  1,00 |

**WICHTIG:**

* **Summenzeile und Summenspalte nicht vergessen!**

**b)**

**Versuchen Sie eine dreidimensionale Kreuztabelle zu erstellen für die drei Merkmale:**

**Geschlecht x klassiertes Alter x Markenpräferenz**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Merkmal Markenpräferenz M | Geschlecht m | | Geschlecht w | | Summe |
| klass. Alter  b. u. 40 | klass. Alter  40+ | klass. Alter  b. u. 40 | klass. Alter  40+ |  |
| **Produkt A**  relative Spaltenhäufigkeit  relative Zeilenhäufigkeit  relative Häufigkeit der Merkmalsausprägungskombination | **1**  0,50  0,14  0,08 | **2**  0,50  0,29  0,17 | **3**  0,75  0,43  0,25 | **1**  0,50  0,14  0,08 | **7,00**  0,58 |
| **Produkt B**  relative Spaltenhäufigkeit  relative Zeilenhäufigkeit  relative Häufigkeit der Merkmalsausprägungskombination | **1**  0,50  0,20  0,08 | **2**  0,50  0,40  0,17 | **1**  0,25  0,20  0,08 | **1**  0,50  0,20  0,08 | **5,00**  0,42 |
| Summe | **2,00**  0,17 | **4,00**  0,33 | **4,00**  0,33 | **2,00**  0,17 | **12,00**  1,00 |

**c)**

**Welche der drei Merkmale kann man als unabhängige bzw. abhängige Merkmale betrachten?**

Die Merkmale Alter und Geschlecht sind unabhängige Merkmale (Regressor), da diese Merkmale die Markenpräferenz (Wahl des Produktes A oder B) beeinflussen können.

Die Markenpräferenz (Produktwahl) ist somit abhängig vom Alter bzw. Geschlecht. Das Merkmal Markenpräferenz ist also die abhängige Variable (Regressand).

Es wird beobachtet, ob dieses Merkmaldurch andere (sog. unabhängige) Merkmale beeinflusst.

Regression:

Aufteilung einer Variablen in einen systematischen und einen zufälligen Teil zur angenäherten Beschreibung einer Variablen als Funktion anderer.

**d)**

**Sie erheben bei einer Grundgesamtheit Daten für 3 Merkmale.**

**Was liefert mehr Information: die dreidimensionale Häufigkeitsverteilung oder alle drei möglichen ein- und zweidimensionalen Häufigkeitsverteilungen zusammen?**

In der dreidimensionalen Häufigkeitsverteilung werden die Informationen zur Markenpräferenz dediziert nach Geschlecht und Alter (männlich bis unter 40 Jahre, männlich 40+, weiblich bis unter 40 Jahre, weiblich 40+) aufgezeigt. Die ein- und zweidimensionale Häufigkeitsverteilung enthält diese detaillierte Aufschlüsselung nicht. In der ein- und zweidimensionale Häufigkeitsverteilung ist mit einem Informationsverlust behaftet. Die dreidimensionale Häufigkeitsverteilung liefert also mehr Informationen als die ein- und zweidimensionale Häufigkeitsverteilung.

Aufgabe:

Folgende vier Umsatzklassen und dazugehörigen Häufigkeiten sind gegeben:

Umsatzklasse absolute Häufigkeit

1,50 bis unter 2,00 4,00

2,00 bis unter 2,50 6,00

2,50 bis unter 3,00 8,00

3,00 bis unter 3,50 1,00

Ermitteln Sie in einer eindimensionalen Häufigkeitsverteilung die absolute und relative Häufigkeit und die Klassenbreite.

Stellen Sie die **absolute Häufigkeiten der klassierten Daten flächenproportional** in einer Grafik dar.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Klasse i | Umsatz in  Mio EURO  pro Branche | Absolute  Häufigkeit h(xi) | Relative  Häufigkeit f(xi) | Klassenbreite  B(xi) ober Klassengrenze - untere Klassengrenze | Rechteckhöhe  (Klassenhäufigkeit)  ri = h(xi) / b(xi) |
| 1 | [1,50; 2,00) | 6 | 0,32 | 0,50 | **12** |
| 2 | [2,00; 2,50) | 8 | 0,42 | 0,50 | **16** |
| 3 | [2,50; 3,00) | 4 | 0,21 | 0,50 | **8** |
| 4 | [3,00; 3,50) | 1 | 0,05 | 0,50 | **2** |
| **Summe** |  | **19** | **1,00** |  |  |

Offene runde Klammer **(** bedeutet „ab“ (entspricht der unteren Klassengrenze)

Geschlossene eckige Klammer **]** bedeutet „bis unter“ (entspricht der oberen Klassengrenze)

Es wird die Verteilung des Umsatzes auf Branchen betrachtet:

In vier Branchen wurde ein Umsatz zwischen 1,50 Millionen bis unter 2,00 Millionen Euro erzielt

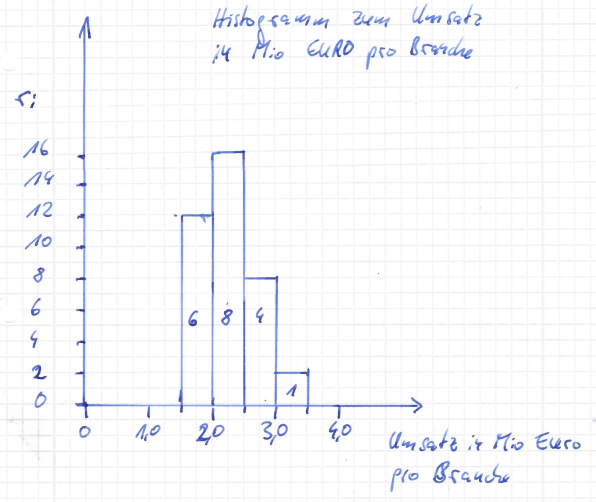
In sechs Branchen wurde ein Umsatz zwischen 2,00 Millionen bis unter 2,50 Millionen Euro erzielt

In acht Branchen wurde ein Umsatz zwischen 2,50 Millionen bis unter 3,00 Millionen Euro erzielt

In einer Branche wurde ein Umsatz zwischen 3,00 Millionen bis unter 3,50 Millionen Euro erzielt

**Histogramm**

* **grafische flächenproportionale Darstellung der Häufigkeiten von klassierten Daten**
* Absolute oder relative Häufigkeiten der Klassen werden durch die Flächen der Rechtecke dargestellt: Fläche = Breite x Höhe
* x-Achse muss eine Skala mit geordneten Werten sein. Diese Werte haben gleiche Abstände.
* direkt nebeneinanderliegende Rechtecke (ohne Abstände)
* Breite der Rechtecke entspricht der Klassenbreite
* Höhe der Rechtecke entspricht den Klassenhäufigkeiten (Häufigkeit / Klassenbreite)
* Die Fläche eines Rechtecks = c \* f(xj )  
  wobei f(xj ) ist die relative Klassenhäufigkeit der Klasse j   
  c ist ein Proportionalitätsfaktor   
  Ist c gleich dem Stichprobenumfang (c = n), so ist die Fläche eines jeden Rechtecks gleich der absoluten Klassenhäufigkeit h(xj).
* Wenn Summe der Flächeninhalte aller Rechtecke = n, wird das Histogramm absolut genannt.
* Summe der Flächeninhalte aller Rechtecke = 1 ( Verwendung der relativen Klassenhäufigkeiten, wird das Histogramm relativ oder normiert genannt (



**WICHTIG:**

* Diagrammtitel
* Achsenbeschriftung (Beschriftung muss die Häufigkeitsverteilung beschreiben, also welche Häufigkeiten zu klassierten Daten werden verteilt)
* Aufgabenstellung beachten! soll die absolute oder relative Häufigkeit verteilt und grafisch dargestellt werden
* Histogramm muss x-Achse mit Skala, deren Werte geordnet sind und die gleiche Abstände haben, enthalten
* Rechtecke dürfen auf der x-Achse keine Abstände (Zwischenräume) haben
* Häufigkeiten der Klassen werden durch Flächen der Rechtecke dargestellt
* Rechteckbreite = Klassenbreite (obere Klassengrenze – untere Klassengrenze, z. B.   
  Klassenbreite = 2,00 – 1,50 = 0,50)  
  Hinweis: im Bsp. ist die Klassenbereite über zu allen Verteilungen gleich groß. Das ist in der Regel nicht der Fall
* Rechteckhöhe ri = absolute oder relative Häufigkeit / durch Summe aller Häufigkeiten
* Obere Klassengrenze ist maßgeblich für das Histogramm (man zeichnet auf der x-Achse von der oberen Klassengrenze zur unteren Klassengrenze, in der Aufgabe z. B. von 2,00 zu 1,50)
* 0-Punkt nicht vergessen!
* An den Achsenenden Pfeile ->

**3. Präsenz am 08.12.2023**

*Wichtige Informationen von Frau Merrins zur Klausur*

*Aufgabenstellung genau lesen.  
Nur das beantworten bzw. lösen, was in der Aufgabenstellung gefragt ist, nicht mehr und nicht weniger*

*präzise antworten*

*Bei Diagrammen die Beschriftungen für die x-Achse und y-Achse nicht vergessen.*

*Im Streudiagramm die Regressionsgerade über den gesamten Bereich des Streudiagramms (bzw. über den letzten Wert auf der x-Achse bzw. y-Achse zeichnen.  
Es muss ersichtlich sein, dass die Regressionsgerade eine lineare Funktion visualisiert und diese Funktion über die Beispielwerte aus der Aufgabenstellung hinaus geht (Die lineare Funktion ist unendlich)*

*Zum Zeichnen des Streudiagramms ein separates Blatt nehmen*

*Wenn* ***im Streudiagramm oberhalb der Regressionsgerade f(x)*** *geschrieben und die Gerade somit als lineare Funktion gekennzeichnet wird, gibt das einen Zusatzpunkt*

*Die Skala der Achsen im Streudiagramm muss nicht bei 0 beginnen, sondern kann auch mit dem kleinsten y- oder x-Wert beginnen (wenn sich das Diagramm dadurch besser zeichnen lässt.) Dann muss am Achsen-Schnittpunkt (x,y) ein Viertelkreis gezeichnet werden (siehe Beispiel)*

*Für die Regressionsgerade müssen mittels der Regressionsgleichung ŷ = a + b \* x nur 2 Punkte ermittelt werden.  
Für den ersten Punkt kann für x = 0 eingesetzt werden. Dann entspricht der Startpunkt den Koordinaten (x-Wert = 0, y-Wert = Wert zum Regressionskoeffizienten a) 🡆 Regressionskoeffizient a wird anhand der dazugehörigen Formel berechnet.  
Für den 2. Punkt einen x-Wert nehmen, der einigen Abstand zum x-Wert des ersten Punktes hat.  
Das kann, muss aber kein Wert aus der Lösungstabelle sein. Es sollte ein Wert sein, mit dem man gut die Regressionsfunktion berechnen kann.   
Beide Punkte im Streuungsdiagramm einzeichnen und mit der Geraden verbinden (Gerade aber den gesamten Bereich des Streudiagramms zeichnen. Die Gerade muss über alle in der Aufgabenstellung genannten x-Werte und über den gesamten Bereich des Streudiagramms gezeichnet werden.  
Die Punkte im Streuungsdiagramm als Punkte oder kleine Kreise, kleine Quadrate oder Rauten zeichnen, WICHTIG die Punkte müssen exakt den Koordinaten x, y entsprechen (Präzision ist ihr wichtig!)*

*programmierbare Taschenrechner sind erlaubt (da alle Schulrechner heute programmierbar sind)  
Es dürfen auch die Programmierfunktionen genutzt werden  
WICHTIG   
Der Rechenweg muss ersichtlich sein 🡆 Tabelle ausfüllen 🡆 Funktion schreiben 🡆 Werte aus Tabelle einsetzen 🡆 Werte aus Zwischenrechnung 🡆 Ergebnis*

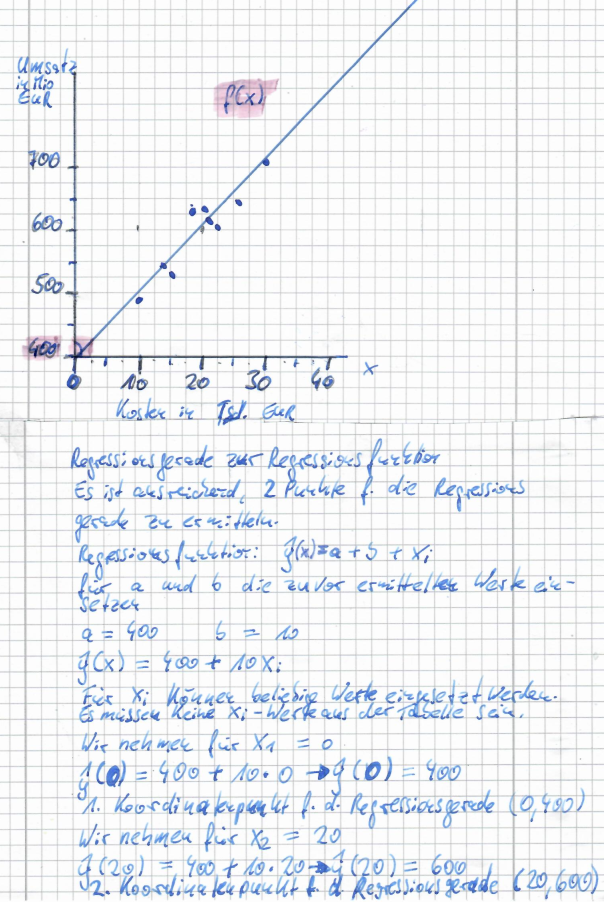
**Übungsaufgabe in der Präsenz**

**Aufgabenstellung**

Ein Filialleiter testet für neun bezüglich des Standortes, der Verkaufsfläche und des Zeitraums vergleichbare Filialen den Einfluss der kosten auf den Umsatz.

1. Zeichnen Sie das Streudiagramm und die Regressionsgerade zur Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!
2. Ermitteln Sie die Korrelationskoeffizienten
3. Berechnen Sie das Bestimmtheitsmaß
4. Interpretieren sie das Bestimmtheitsmaß
5. Prognostizieren Sie die geschätzten Kosten bei einem Umsatz von 900 Mio EUR
6. Prognostizieren Sie den geschätzten Umsatz bei Kosten von 30.000 EUR
7. Zeichnen Sie das Streudiagramm und die Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!

Zeichnen Sie das Streudiagramm und die Regressionsgerade zur Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!



Zum Zeichnen des Streudiagramms ein separates Blatt nehmen

Bei Diagrammen die Beschriftungen für die x-Achse und y-Achse nicht vergessen.

Im Streuungsdiagramm die Punkte zu den x,y-Koordinaten als Punkte oder kleine Kreise, kleine Quadrate oder Rauten zeichnen.

WICHTIG die Punkte müssen exakt den Koordinaten x, y entsprechen (Präzision ist ihr wichtig! Evtl. millimeter-Papier verwenden)

Im Streudiagramm die Regressionsgerade über den gesamten Bereich des Streudiagramms (bzw. über den letzten Wert auf der x-Achse bzw. y-Achse zeichnen.

Es muss ersichtlich sein, dass die Regressionsgerade eine lineare Funktion visualisiert und diese Funktion über die Beispielwerte aus der Aufgabenstellung hinaus geht (Die lineare Funktion ist unendlich)

Wenn im Streudiagramm oberhalb der Regressionsgerade f(x) geschrieben und die Gerade somit als lineare Funktion gekennzeichnet wird, gibt das einen Zusatzpunkt

Die Skala der Achsen im Streudiagramm muss nicht bei 0 beginnen, sondern kann auch mit dem kleinsten y- oder x-Wert beginnen (wenn sich das Diagramm dadurch besser zeichnen lässt.) Dann muss am Achsen-Schnittpunkt (x,y) ein Viertelkreis gezeichnet werden (siehe Beispiel)

Für die Regressionsgerade müssen mittels der Regressionsgleichung ŷ(x) = a + b \* xi nur 2 Punkte ermittelt werden.

Für den ersten Punkt kann für x = 0 eingesetzt werden.

Dann entspricht der Startpunkt den Koordinaten (x-Wert = 0, y-Wert = Wert zum Regressionskoeffizienten a)

Regressionskoeffizient a wird anhand der dazugehörigen Formel berechnet.

Für den 2. Punkt einen x-Wert nehmen, der einigen Abstand zum x-Wert des ersten Punktes hat.

Das kann, muss aber kein Wert aus der Lösungstabelle sein.

Es sollte ein Wert sein, mit dem man gut die Regressionsfunktion berechnen kann.

Beide Punkte im Streuungsdiagramm einzeichnen und mit der Geraden verbinden (Gerade aber den gesamten Bereich des Streudiagramms zeichnen. Die Gerade muss über alle in der Aufgabenstellung genannten x-Werte und über den gesamten Bereich des Streudiagramms gezeichnet werden.

In diesem Diagramm wurden für die Bestimmung des **ersten x.y-Koordinatenpunkt für x der Wert 0 eingesetzt**.

**x = 0**

ŷ(x) = a + b \* xi **a = 400, b = 10**

**ŷ(0) = 400 + 10 \* 0**

**x = 0**

**ŷ = 400**

Der **erste x.y-Koordinatenpunkt ist somit 0, 400 (x =0, y = 400)**

In diesem Diagramm wurden für die Bestimmung des **zweiten x.y-Koordinatenpunkt für x der Wert 20 eingesetzt**.

**x = 20**

ŷ(x) = a + b \* xi **a = 400, b = 10**

**ŷ(20) = 400 + 10 \* 20**

**x = 20**

**ŷ = 600**

Der **zweite x.y-Koordinatenpunkt ist somit 20, 600 (x =20, y = 600)**

Beide Koordinatenpunkte (0, 400) und (20, 600) im Diagramm einzeichnen und mit der Geraden verbinden.

Die Gerade jedoch über die gesamte Breite des Diagramms zeichnen“

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i  Filial-Nr. | xi  Kosten  in Tsd. € | yi  Umsatz  in Mio. € | xi \* yi | xi² | yi² |
| 1 | 10 | 480 | 4.800 | 100 | 230.400 |
| 2 | 18 | 632 | 11.376 | 324 | 399.424 |
| 3 | 30 | 702 | 21.060 | 900 | 492.804 |
| 4 | 22 | 630 | 13.860 | 484 | 396.900 |
| 5 | 26 | 645 | 16.770 | 676 | 416.025 |
| 6 | 14 | 545 | 7.630 | 196 | 297.025 |
| 7 | 24 | 606 | 14.544 | 576 | 367.236 |
| 8 | 21 | 630 | 13.230 | 441 | 396.900 |
| 9 | 15 | 530 | 7.950 | 225 | 280.900 |
| **SUMMEN** | **180** | **5.400** | **111.220** | **3.922** | **3.277.614** |

x̅ = 180 : 9 = 20

y̅ = 5.400 : 9 = 600

**Regressionsfunktion**

**ŷ(x) = a + b \* xi** | a und b sind die Regressionskoeffizienten   
 | und müssen berechnet werden

*Hinweis:*

*Warum schreibt man ŷ (y-Dach)? und nicht einfach y?   
ŷ kennzeichnet Schätzwerte (Das Ergebnis der Regressionsrechnung ist somit immer ein Schätzwert (Prognosewert)).*

***Berechnung der Regressionskoeffizienten a und b***

*Hinweis:*

*Der Nenner in nachfolgenden Formeln (hier 2.898) muss nur einmal berechnet werden, da er in beiden Formeln identisch ist!  
Formel zur Berechnung des Nenners zu beiden Regressionskoeffizienten*

***Berechnung des Regressionskoeffizienten a***

**a = 400**

***Berechnung des Regressionskoeffizienten b***

**b = 10**

**Regressionsfunktion mit Regressionskoeffizienten a und b**

**ŷ(x) = 400 + 10xi**

*Hinweis:*

*Die Regressionskoeffizienten müssen in Worten interpretiert werden können. 🡆 welche Abhängigkeit zwischen abhängiger Variable b und unabhängiger Variable a?  
Was bedeuten in der Regressionsrechnung die Regressionskoeffizienten b und a?   
b ist ein Faktor im veränderlichen Term der Regressionsrechnung,   
nur der Regressionskoeffizient a ist konstant.*

*Wenn z. B. x den Wert 0 hat (keine Kosten (unabhängige Variable x) entstehen bzw. keine Investitionen getätigt werden, wird immer noch ein Umsatz erwirtschaftet   
ŷ hat dann den Wert des Regressionskoeffizient a 🡆 ŷ = a + b \* x = a + b \* 0 🡆 ŷ = a*

**Korrelationskoeffizient rx,y**

**Ermitteln Sie die Korrelationskoeffizienten**

**Korrelationskoeffizient rx,y = 0,925**

**Bestimmtheitsmaß R²**

Berechnen Sie das Bestimmtheitsmaß R²

*Hinweis:*

*Bei einfacher linearer Regression (Regression mit nur 2 Variablen) entspricht das Bestimmtheitsmaß R² dem Quadrat des Korrelationskoeffizienten rx,y..*

Bestimmtheitsmaß R² = 0,925² = 0,856 = 85,6 %

*Hinweis:*

*Das Bestimmtheitsmaß in immer auch in % anzugeben.*

**Interpretation des Bestimmtheitsmaßes R².**

*Hinweis:*

*Des Bestimmtheitsmaß R² muss in Worten interpretiert werden können   
 z. B.   
Bestimmtheitsmaß ist ein Gütemaß des Modells, beschreibt wie gut das Modell der Realität entspricht.  
Bestimmtheitsmaß ist Anteil der Varianz zur abhängigen Variable der sich durch den Anteil der Varianz der unabhängigen variable erklären lässt.  
Z. B.   
52% der Varianz der Umsätze (Unterschiede zu den Umsätzen (abhängiges Merkmal)) lassen sich durch die* *Varianz der Kosten (Kostenunterschiede (unabhängiges Merkmal)) erklären. Die übrigen 48% der Varianz zu den Kosten werden durch andere Einflussgrößen (Faktoren) erklärt.*

Interpretieren Sie das Bestimmtheitsmaß zur Aufgabenstellung.

85,6% der Varianz zum Umsatz (Umsatzunterschiede) lassen sich durch die Kostenvarianz erklären. Die übrigen 14,4% der Varianz zum Umsatz werden durch andere Einflussgrößen (Faktoren) erklärt.

Zu den übrigen Einflussgrößen können die Marktgröße, die Lage des Marktes, die Kaufkraft der Kunden u. a. sein.

Ein Bestimmtheitsmaß von 0,856 bzw. 85,6% kennzeichnet eine starke positive Korrelation zwischen dem abhängigen Merkmal (der abhängigen Variablen) „Umsatz“ und dem unabhängigen Merkmal (der unabhängigen Variablen) „Kosten“.

**Prognosewerte**

Prognostizieren Sie die geschätzten Kosten bei einem Umsatz von 900 Mio EUR

y (Umsatz) = 900 €

ŷ(x) = a + b \* xi | ŷ = 900, a = 400, b = 10

| Werte für ŷ und a und b in die Gleichung einsetzen

900 = 400 + 10 \* x | -400, : 10

500 : 10 = x

x = 50 Tsd. €

Hinweis:

Für den Wert der unabhängigen Variable immer auch die Einheit (hier „Tsd. €“ ) angeben.

Bei einem Umsatz von 900 Mio. € betragen die geschätzten Kosten voraussichtlich 50 Tsd. €.

Prognostizieren Sie den geschätzten Umsatz bei Kosten von 30.000 EUR.

x (Kosten) = 30 Tsd. €

ŷ(x) = a + b \* xi | x = 30, a = 400, b = 10

| Werte für a und b und x in die Gleichung einsetzen

ŷ(30) = 400 + 10 \* 30

ŷ(30) = 700 Mio €

Hinweis:

Für den Wert der abhängigen Variable immer auch die Einheit (hier „Mio. €“ ) angeben.

Bei Kosten von 30 Tsd. € beträgt der geschätzte Umsatz voraussichtlich 700 Mio. €.