MATHEMATIK

04.07.2019

WIEDERHOLUNG

Diese Lücken sollten nicht auch bei Ihnen vorhanden sein:

Aufgrund einer statistischen Untersuchung entsteht eine geordnete bzw. ungeordnete ___________, die durch Summierung je Ausprägung zur ___________Häufigkeit führt.
Werden diese Werte zur Gesamtheit in Beziehung gesetzt, so spricht man von der ___________Häufigkeit.
Im Bereich der Häufigkeiten wird eine kleiner/ gleich Frage als ___________und eine größer/ gleich Beziehung als __________bezeichnet.
Grafisch werden diese in Form einer ___________oder auch eines ___________dargestellt, während für die einfachen Häufigkeiten überwiegend _____________,_____________ bzw.___________ **-** oder ___________diagramme genutzt werden.

ZIELSETZUNG

Themen, die Sie nach dieser Veranstaltung kennen sollten:

- ✓ Aufgaben zu den behandelten Themen.
- ✓ Was sind statistische Streuungsmaße?
- ✓ Was versteht man unter der Spannweite?
- ✓ Wie berechnet man die Varianz?
- ✓ Was besagt die Standardabweichung?
- ✓ Worauf bezieht sich die mittlere absolute Abweichung?
- ✓ Was bedeutet der Variationskoeffizient?
- ✓ Aufgaben und Übungen zu den benannten Themen.

Aufgaben:

2) An einem Bankschalter werden die Kundenankünfte beobachtet.
(Anzahl der pro 10-Minuten-Zeitintervall ankommenden Kunden)

Für 40 derartige Zeitintervalle erhält man folgende Ergebnisse:

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

a) Berechnen Sie die absolute und relative Häufigkeit.
b) Bestimmen Sie die absolute und relative Summenhäufigkeiten.
c) Bestimmen Sie die absolute und relative Resthäufigkeiten.
d) Zeichnen Sie Ihre Ergebnisse in ein Schaubild Ihrer Wahl.

Statistische Maßzahlen

Bisher wurden die auftretenden Merkmalsausprägungen verdichtet, klassifiziert und zusammengefasst. Daraus entstanden verschiedene Häufigkeitsformen, Klassenbildungen und Darstellungsformen. Mittels der statistischen Maßzahlen werden diese Informationen noch weiter komprimiert und in knapper Form interpretiert.

✓ Mittelwert:

Das arithmetische Mittel sollte stets bei metrisch skalierten Merkmalen gebildet werden und entspricht dem Durchschnitt einer Untersuchung:

$$\mu = \frac{1}{N}(a_1 + a_2 + \cdots + a_N) = \frac{1}{N}\sum_{i=1}^{N} a_i$$

Oder, sofern ein Beobachtungswert mehrfach vorkommt, wird der Mittelwert mittels der absoluten Häufigkeiten gebildet:

$$\mu = \frac{1}{N}(x_1 h_1 + x_2 h_2 + \cdots + x_k h_k) = \frac{1}{N}\sum_{i=1}^{k} x_i h_i$$

Beispiel:

Verkaufte Zeitungen	0	1	2	3	4	5
Anzahl der Tage (h_i)	3	6	4	10	2	5
Anzahl der Zeitungen ($x_i h_i$)	0	6	8	30	8	25

Summe der Tage: $N = 30$

Summe der Zeitungen: $X = 77$

Arithmetischer Mittelwert: $\mu = \frac{X}{N} = \frac{77}{30} \cong 2{,}5$

Demzufolge wurden innerhalb der letzten 30 Tage im Durchschnitt 2,5 Zeitungen je Tag verkauft.

✓ <u>Modus / Modalwert:</u>

Dieser Wert gibt an, welcher Wert am häufigsten vorkommt und wird u.a. auch dichtester Wert genannt.
Zusätzlich kann er auch auf nominalen bzw. ordinalen Skalen bestimmt werden:

$$Modus = \overline{X_D} = \max\left(h(x_i)\right)$$

Beispiel:
In unserem Beispiel ergibt sich als Modus 3 Zeitungen je Tag.

✓ <u>Median / Zentralwert :</u>

Der Median ist die Merkmalsausprägung desjenigen Elements, das in der der Größe nach geordneten Beobachtungsreihe in der Mitte steht.
Daraus ergibt sich automatisch, dass mindestens eine ordinale Skala vorhanden sein muss.

<u>ungerade Anzahl N:</u> $\overline{X_Z} = x_{\left[\frac{n+1}{2}\right]}$

5,5,6,6,6,7,8,8,9,9,10 $\overline{X_Z} = x_{\left[\frac{11+1}{2}\right]} = x_6 = 7$

<u>gerade Anzahl N:</u> $\overline{X_Z} = \frac{1}{2}\left(x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2}+1\right]}\right)$

1,1,2,2,3,3,3,4,4,5,6,7 $\overline{X_Z} = \frac{1}{2}\left(x_{\left[\frac{12}{2}\right]} + x_{\left[\frac{12}{2}+1\right]}\right) = \frac{1}{2}(3+3) = 3$

Liegen die Daten in Form einer Häufigkeitsverteilung vor, dann ist der Median die Ausprägung, bei der die Summenhäufigkeit den Wert 0,5 überschreitet.

Beispiel:
In unserem Beispiel ergibt sich als Median 3 Zeitungen, da an der Stelle 3 der Wert 0,5 (mit 77%) zum ersten Mal überschritten wird.

- ✓ Mittelwert (geometrisch):
  Hat man es mit zeitlich aufeinanderfolgenden Wachstumsraten, Zuwächsen oder ähnlichen zu tun, ist das arithmetische Mittel nicht der sachlich korrekte Durchschnittswert, sondern das geometrische Mittel:

$$\overline{X_G} = \sqrt[N]{{x_1}^{h(x_1)} \cdot {x_2}^{h(x_2)} \cdot \ldots \cdot {x_k}^{h(x_k)}} = \sqrt[N]{\prod_{i=1}^{k} {x_i}^{h(x_i)}}$$

Beispiel:

Im Jahr 1 einer Untersuchung macht ein Unternehmen 1 Mio Umsatz. Im darauffolgenden Jahr 1,8 Mio und im 3. Jahr 1.98 Mio.

Gesamtzuwachsrate: $1{,}8 * 1{,}1 = 1{,}98$ demzufolge 98%

Geometrische Mittel: $\overline{X_G} = \sqrt[2]{1{,}8 * 1{,}1} = \sqrt[2]{1{,}98} \cong 1{,}4$

✓ Mittelwert (harmonisch):

Beziehen sich die Merkmalsausprägungen auf verschiedene Bezugssysteme wie z.B. die Durchschnittsgeschwindigkeit auf unterschiedlich lange Teilstrecken, liefert das harmonische Mittel den exakten Wert:

$$\overline{X_H} = \frac{N}{\sum_{i=1}^{k} \frac{h(x_i)}{x_i}}$$

Beispiel:

Teilstrecke Nr.	1	2	3	4
Länge der Teilstrecke [km]	30	10	40	20
Geschwindigkeit [km/h]	40	50	80	100

$$\overline{X_H} = \frac{30 + 10 + 40 + 20}{\frac{30}{40} + \frac{10}{50} + \frac{40}{80} + \frac{20}{100}} = \frac{100}{1{,}65} = 60{,}6$$

Die Durchschnittsgeschwindigkeit beträgt demzufolge 60,6 km/h.

- ✓ <u>Quantile:</u>

  Durch den Zentralwert (Median) wird eine Beobachtungsreihe in zwei gleich große Teile zerlegt.

  Wird nun eine solche Kette in k-Teile auf gespaltet, so spricht man von k-Quantilen (Quartile, Dezentile, Perzentile[%]).

Definition:

Die Merkmalsausprägungen $\overline{X_{p/k}}$ $(p = 1, \ldots, k-1)$ eines mind. ordinal messbaren Merkmals, die die geordnete Reihe der Beobachtungswerte in k gleiche Teile zerlegen, heißen k-Quantile.

Es gilt:

$$\sum_{x_i < x_{p/k}} f(x_i) \leq \frac{p}{k}$$

und

$$\sum_{x_i > x_{p/k}} f(x_i) \leq 1 - \frac{p}{k}, p = 1, \ldots, k-1$$

$\overline{x_{p/k}}$ heißt p-tes k-Quantil.

Beispiel:

a) Geordnete Reihe: 3,5,7,9,11 (n=5)

$$\overline{X_Z} = x_{\left[\frac{5+1}{2}\right]} = x_3 = 7$$

$$\overline{X_{0,3}} \rightarrow 5 * 0,3 = 1,5; k = 2 : \overline{X_{0,3}} = x_2 = 5$$

b) Geordnete Reihe: 3,5,7,9,11,25 (n=6)

$$\overline{X_Z} = \frac{1}{2}\left(x_{\left[\frac{6}{2}\right]} + x_{\left[\frac{6}{2}+1\right]}\right) = \frac{1}{2}(7+9) = 8$$

$$\overline{X_{0,4}} \rightarrow 6 * 0,4 = 2,4; k = 3 : \overline{X_{0,4}} = x_3 = 7$$

AUFGABE I

1. Geben Sie für die genannten Merkmale die mögliche Gesamtmasse und Einheit sowie deren Identifikationskriterien an.
   Welcher Skalentyp mit welchen möglichen Transformationen liegt je Merkmal vor?
   Handelt es sich um Bestands- oder Bewegungsmasse? (Begründung!)

   a) Bankleitzahl einer hessischen Filiale eines Kreditinstituts.
   b) Tägliche Anzahl neuer Anfragen von (neuen) Freunden auf facebook.
   c) Monatliche, eigene Handykosten (in EURO) für den SMS-Bereich.

2. Im Frankfurter Palmengarten wird das monatliche Wachstumsverhalten (in mm) einer Pflanze untersucht, wobei sich die folgenden Daten ergeben haben:

Monat	Jan	Feb	Mrz	Apr	Mai	Juni	Juli	Aug	Sept	Okt	Nov	Dez
2009	10	12	10	8	10	12	15	15	12	10	8	8

   a) Berechnen Sie die relativen und absoluten Häufigkeiten.
   b) Bestimmen Sie den monatlichen Wachstumsdurchschnitt und den Median.
   c) Wie groß ist die Standardabweichung des Wachstumsverhaltens der Pflanze?
   d) Berechnen Sie die monatlichen Wachstumsfaktoren und für diese einen geeigneten Mittelwert und begründen Sie Ihre Wahl.

Statistische Streuungsmaße I

Die zuvor beschriebenen statistischen Maßzahlen reichen oft zur Charakterisierung einer eindimensionalen Verteilung nicht aus.
Aus diesem Grund ist es wichtig den Abstand der Beobachtungswerte vom Mittelwert zusätzlich zu betrachten (Streuungsverhalten).

- ✓ Spannweite (Range):
  Ein einfaches Streuungsmaß, da es den lediglich aus der Differenz zwischen der größten und kleinsten Merkmalsausprägung besteht.
  Die Spannweite kann auch bei ordinalen Daten angewandt werden.

$$R = x_{max} - x_{min}$$

Nachteil:
Die Spannweite hängt nur von 2 Werten ab und ist aufgrund von „Ausreißern" wenig aussagekräftig.

Beispiel:
Bei dem Kaufverhalten von Zeitungen ergibt sich eine Spannweite von 6 Zeitungen (6-0).

Statistische Streuungsmaße II

- ✓ <u>Varianz:</u>
  Um dem Problem der positiven und negativen Zahlenwerte aus den Differenzen der Ausprägungen zum arithmetischen Mittel zu begegnen, werden daraus die Abweichungsquadrate gebildet und aus diesen Werten der Durchschnitt berechnet.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2$$

Im Fall der absoluten Häufigkeiten errechnet sich die Varianz aus:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \mu)^2 \cdot h(x_i) = \sum_{i=1}^{k}(x_i - \mu)^2 \cdot f(x_i)$$

bzw.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} {x_i}^2 \cdot h(x_i) - \mu^2 = \sum_{i=1}^{k} {x_i}^2 \cdot f(x_i) - \mu^2$$

Statistische Streuungsmaße III

- ✓ Standardabweichung:
  Die positive Quadratwurzel aus der Varianz bezeichnet man als Standardabweichung .
  Die Dimension der Standardabweichung entspricht der Einheit der Beobachtungswerte.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

- ✓ Variationskoeffizient (VC):
  Damit Streuungsmaße verschiedener Untersuchungen verglichen werden können, bezieht man die Standardabweichung auf das arithmetische Mittel und erhält das relative Streuungsmaß VC.

$$VC = \frac{\sigma}{\mu} \quad \text{bzw.} \quad VC = \frac{\sigma}{\mu} \cdot 100\ \%$$

Statistische Streuungsmaße IV

✓ <u>Mittlere absolute Abweichung (MAD):</u>
Bezieht man die Berechnung des Streuungsmaß nicht auf das arithmetischen Mittels μ, sondern hat $\overline{X_Z}$ (Median) als Bezugswert gewählt, so spricht man von der mittleren <u>absoluten</u> Abweichung.

$$d = \frac{1}{N}\sum_{i=1}^{N}|x_i - \overline{X_Z}|$$

bzw.

$$d = \frac{1}{N}\sum_{i=1}^{k}|x_i - \overline{X_Z}| \cdot h(x_i) = \sum_{i=1}^{k}|x_i - \overline{X_Z}| \cdot f(x_i)$$

Beispiel I:

Zeitungsverkauf bei einem Durchschnitt von 2,5 und einem Median von 2 Zeitungen

Verkaufte Zeitungen	x_i	0	1	2	3	4	5	Summe
absolute Häufigkeit	$h(x_i)$	3	6	4	10	2	5	**30**
Varianz I	$x_i^2 \cdot h(x_i)$	**0**	**6**	**16**	**90**	**32**	**125**	**269**
relative Häufigkeit	$f(x_i)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{15}$	$\frac{1}{3}$	$\frac{1}{15}$	$\frac{1}{6}$	$\frac{30}{30}$
Varianz II	$x_i^2 \cdot f(x_i)$	$\frac{0}{10}$	$\frac{1}{5}$	$\frac{8}{15}$	$\frac{9}{3}$	$\frac{16}{15}$	$\frac{25}{6}$	$\frac{269}{30}$
Abweichung (Betrag)	$\lvert x_i - \overline{X}_Z \rvert$	2	1	0	1	2	3	
MAD (absolut)	$\lvert x_i - \overline{X}_Z \rvert \cdot h(x_i)$	6	6	0	10	4	15	**41**
MAD (relativ)	$\lvert x_i - \overline{X}_Z \rvert \cdot f(x_i)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{0}{15}$	$\frac{1}{3}$	$\frac{2}{15}$	$\frac{1}{2}$	$\frac{41}{30}$

Beispiel II:

Varianz I:

$$\frac{1}{30}(0+6+16+90+32+125)-2{,}5^2$$

$$\frac{269}{30}-6{,}25 \approx 2{,}72$$

Varianz II:

$$\left(\frac{0}{10}+\frac{1}{5}+\frac{8}{15}+\frac{9}{3}+\frac{16}{15}+\frac{25}{6}\right)-2{,}5^2$$

$$\frac{269}{30}-6{,}25 \approx 2{,}72$$

Standardabweichung: $\sigma = \sqrt{2{,}72} \approx 1{,}5$ Stück

Variationskoeffizient: $VC = \frac{1{,}5}{2{,}5} = 0{,}6 = 60\%$

MAD (absolut): $\frac{1}{30}(6+6+0+10+4+15) = \frac{41}{30} = 1{,}37$

MAD (relativ):

$$\left(\frac{1}{5}+\frac{1}{5}+\frac{0}{15}+\frac{1}{3}+\frac{2}{15}+\frac{1}{2}\right)$$

$$\left(\frac{6}{30}+\frac{6}{30}+\frac{0}{30}+\frac{10}{30}+\frac{4}{30}+\frac{15}{30}\right) = \frac{41}{30} = 1{,}37$$

Aufgaben:

1. Für ein Waschpulver eines bestimmten Herstellers wurden in 10 Geschäften Fuldas folgende Preise für ein 1-kg-Paket ermittelt:
   1,40 – 1,60 – 1,70 – 1,50 – 1,40 – 1,80 – 1,70 – 1,60 – 1,50 – 1,80

   a) Berechnen den Durchschnittspreis und Zentralwert der Preise.
   b) Bestimmen Sie die Varianz und Standardabweichung.
   c) Geben Sie die mittlere absolute Abweichung an.

2. Berechnen Sie zur Aufgabe 2 (Seite 17) die Spannweite, Varianz, Standardabweichung, mittlere absolute Abweichung und den Variationskoeffizienten.

3. Die folgende Tabelle gibt die Körpergröße in cm von 5 Kindern an:

Name	Hugo	Fred	Peter	Paul	Maria
Größe	120	130	125	130	135

   a) Berechnen das arithmetische Mittel und den Median.
   b) Bestimmen Sie die Varianz und Standardabweichung.
   c) Geben Sie die mittlere absolute Abweichung an.
   d) Wie groß ist der Variationskoeffizient?