Wichtige Rechenregeln zur Statistik I und II Wintersemester 2020/2021

Prof. Dr. C. Becker
C. Herrmann
S. Kirschstein-Barczewski
D. Lukas
S. Rhein

Diese Unterlagen sind nicht Teil der Formelsammlung und dürfen daher nicht während der Klausur genutzt werden!

1. Indizes

Bei Umfragen liegen (in einer Urliste) oft je ein Wert (eine Merkmalsausprägung) zu jeder Beobachtungseinheit vor. Diese Werte werden mit einer Variable und einem Index bezeichnet. So wird z.B. die Merkmalsausprägung der ersten befragten Person mit x_1, die der zweiten Person mit x_2, usw. bezeichnet. Werden die Merkmalsausprägungen der Größe nach geordnet, schreibt man die Indizes in Klammern. Die kleinste Ausprägung heißt dann $x_{(1)}$, die zweitkleinste $x_{(2)}$ usw. Mit n wird die Anzahl der Merkmalsträger und daher auch die Anzahl der beobachteten Merkmalsausprägungen bezeichnet. Der letzte beobachtete Wert ist also x_n, der größte Wert ist $x_{(n)}$.

Beispiel: 5 Personen wurden nach ihrer Körpergröße gefragt ($n = 5$). Die Antworten waren: 175cm, 161cm, 185cm, 156cm und 181cm Damit ist:

$$\begin{aligned} x_1 &= 175 & x_{(1)} &= 156 \\ x_2 &= 161 & x_{(2)} &= 161 \\ x_3 &= 185 & x_{(3)} &= 175 \\ x_4 &= 156 & x_{(4)} &= 181 \\ x_n = x_5 &= 181 & x_{(n)} = x_{(5)} &= 185 \end{aligned}$$

In der Urliste wird meist der Buchstabe i als Index verwendet. Wenn Daten zusammengefasst werden (klassiert oder unklassiert), wird oft der Index j verwendet. Mit j werden bei unklassierten Daten die Merkmalsausprägungen nummeriert und bei klassierten Daten die Klassen. Die Anzahl aller Klassen wird oft mit k bezeichnet.

Beispiel:
Zu den fünf Personen, die im vorherigen Beispiel befragt werden, kommen nun noch weitere 3 Personen dazu:

$$\begin{aligned} x_6 &= 175 \\ x_7 &= 156 \\ x_8 &= 162 \end{aligned}$$

Da Ausprägungen nun mehrmals auftreten, können die Daten zusammengefasst werden. a_j bezeichnet die Merkmalsausprägung; h_j oder $h(a_j)$ die Häufigkeit des Auftretens der Merkmalsausprägung

a_j.

$$
\begin{aligned}
a_1 &= 156 & h_1 &= 2 \\
a_2 &= 161 & h_2 &= 1 \\
a_3 &= 162 & h_3 &= 1 \\
a_4 &= 175 & h_4 &= 2 \\
a_5 &= 181 & h_5 &= 1 \\
a_6 &= 185 & h_6 &= 1
\end{aligned}
$$

Die Daten können auch in Klassen zusammengefasst werden, z.B. in drei Klassen ($k = 3$):

Klasse 1: $155 < x \leq 165$
Klasse 2: $165 < x \leq 175$
Klasse 3: $175 < x \leq 185$

Ist der Durchschnitt in einer Klasse unbekannt, dann wird mit einem repräsentativen Wert gerechnet (meist ist das die Klassenmitte), die mit m_j bezeichnet wird. n_j ist die Anzahl an Beobachtungen in der jeweiligen Klasse.

$$
\begin{aligned}
m_1 &= 160 & n_1 &= 4 \\
m_2 &= 170 & n_2 &= 2 \\
m_3 &= 180 & n_3 &= 2
\end{aligned}
$$

Bei der Berechnung von Lage- und Streuungsmaßen mit der Klassenmitte sind die Ergebnisse nur näherungsweise richtig.

2. Rechnen mit dem Summen- und Produktzeichen

Bei Summierung vieler Messwerte verwendet man häufig das Summensymbol $\sum$. Für eine feste Anzahl n, eine beliebige Konstante c und Beobachtungen x_i, y_i für $i = 1, ..., n$ bedeutet

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n.$$

Wenn klar ist, wie viele Beobachtungen vorliegen und dass alle aufsummiert werden sollen, schreibt man oft auch kurz

$$\sum x_i \quad \text{statt} \quad \sum_{i=1}^{n} x_i.$$

Sollen nur die ersten j Beobachtungen summiert werden, schreibt man

$$\sum_{i=1}^{j} x_i = \sum_{i \leq j} = x_1 + x_2 + ... + x_j.$$

Summationsregeln

S1 : $\sum(x_i + y_i) = \sum x_i + \sum y_i$

S2 : $\sum c \cdot x_i = c \cdot \sum x_i$

S3 : $\sum c = n \cdot c$

Beispiele

1. Gegeben sind vier Beobachtungen $(n = 4)$: $x_1 = 4$, $x_2 = -3$, $x_3 = 2$, $x_4 = 1$. Damit ergibt sich

$$\sum x_i = x_1 + x_2 + x_3 + x_4 = 4 + (-3) + 2 + 1 = 4$$

$$\sum_{i \leq 3} x_i = \sum_{i=1}^{3} x_i = x_1 + x_2 + x_3 = 4 + (-3) + 2 = 3$$

$$\sum_{i=3}^{4} x_i = x_3 + x_4 = 2 + 1 = 3$$

$$\sum_{i=2}^{2} x_i = x_2 = -3$$

$$\sum x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 = 16 + 9 + 4 + 1 = 30$$

2. Es gilt: $\sum x_i \cdot (y_i + c) = \sum x_i \cdot y_i + c \cdot \sum x_i$, denn

$$\begin{aligned} \sum x_i \cdot (y_i + c) &= \sum(x_i \cdot y_i + c \cdot x_i) && \text{(ausmultiplizieren)} \\ &= \sum x_i \cdot y_i + \sum c \cdot x_i && \text{(S1)} \\ &= \sum x_i \cdot y_i + c \cdot \sum x_i && \text{(S2)} \end{aligned}$$

An einem Zahlenbeispiel:

Es seien $n = 4$ und $c = 2$, und gegeben sind die folgenden Daten:

i	x_i	y_i
1	4	3
2	1	-2
3	-6	0
4	2	-1

Rechnet man $\sum x_i \cdot (y_i + c)$ direkt:

i	x_i	$y_i + c$	$x_i \cdot (y_i + c)$
1	4	5	20
2	1	0	0
3	-6	2	-12
4	2	1	2
$\sum$			10

Also $\sum x_i \cdot (y_i + c) = 10$.

Rechnet man $\sum x_i \cdot y_i + c \cdot \sum x_i$:

i	x_i	$x_i \cdot y_i$
1	4	12
2	1	-2
3	-6	0
4	2	-2
$\sum$	1	8

Also $\sum x_i \cdot y_i + c \cdot \sum x_i = 8 + 2 \cdot 1 = 10$.

3. Vereinfachen des folgenden Ausdrucks:

$$\begin{aligned}
&\frac{\sum(x_i - y_i)^2 + \sum 2 \cdot x_i \cdot y_i}{\sum 3 \cdot (x_i^2 + y_i^2)} \\
&= \frac{\sum(x_i^2 - 2 \cdot x_i \cdot y_i + y_i^2) + \sum 2 \cdot x_i \cdot y_i}{\sum 3 \cdot (x_i^2 + y_i^2)} && ((c-b)^2 = c^2 - 2 \cdot b \cdot c + b^2) \\
&= \frac{\sum x_i^2 - 2 \cdot \sum x_i \cdot y_i + \sum y_i^2 + 2 \cdot \sum x_i \cdot y_i}{3 \cdot \sum(x_i^2 + y_i^2)} && (\mathrm{S1}, \mathrm{S2}) \\
&= \frac{\sum(x_i^2 + y_i^2)}{3 \cdot \sum(x_i^2 + y_i^2)} && (\text{subtrahieren}) \\
&= \frac{1}{3} && (\text{kürzen})
\end{aligned}$$

Produktzeichen

Das Produktzeichen funktioniert ähnlich wie das Summenzeichen, hier werden die Werte jedoch multipliziert anstatt addiert.

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_3 \cdot x_3 \cdot \ldots. \cdot x_n$$

3. Weitere Rechenregeln

Die **Abrundungsfunktion** oder **Gaußklammer** ordnet einer reelen Zahl die nächste, nicht größere, ganze Zahl zu.

$[x] = max\{k \in \mathbb{Z} | k \leq x\}$

Beispiele:
$[3.4] = 3;\quad [9.8] = 9;\quad [-7.4] = -8;\quad [6] = 6$

Logarithmus und Exponentialfunktion

Wenn $a^x = b$, dann ist $x = log_a(b)$
Sprich: *x ist der Logarithmus von b zur Basis a.*

Oft wird der natürliche Logarithmus benötigt, die Basis ist hier die Eulersche Zahl:
Wenn $e^x = b$, dann ist $x = log_e(b) = ln(b)$
Sprich: *x ist der Logarithmus von b zur Basis e oder x ist der natürliche Logarithmus von b.*
$e =$ Eulersche Zahl ≈ 2.71828

Schreibweise der Exponentialfunktion: $e^x = exp(x)$, diese Schreibweise wird vor allem verwendet, wenn der Exponent selbst aus mehreren Termen besteht. (Beispiel: Formelsammlung 15.2, Dichtefunktion der Normalverteilung)

Rechenregeln für Logarithmus und Exponentialfunktion

$ln(b \cdot c) = ln(b) + ln(c)$
$ln(\frac{b}{c}) = ln(b) - ln(c)$
$ln(b^c) = c \cdot ln(b)$

$e^a \cdot e^b = e^{a+b}$
$(e^a)^b = e^{a \cdot b}$
$e^{-a} = \frac{1}{e^a} = (\frac{1}{e})^a$
$ln(e^a) = e^{ln(a)} = a$

Fakultät und Binomialkoeffizient

Das Produkt aller natürlichen Zahlen (ohne Null) bis zu einer Zahl n wird als $n!$ (n Fakultät) bezeichnet.

$$n! = \prod_{i=1}^{n} i = 1 \cdot 2 \cdot 3 \cdot .... \cdot n$$

Beispiele:
$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$
$7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$

Per Definition ist:
$1! = 1$
$0! = 1$

Der Binomialkoeffizient zweier nicht-negativer ganzer Zahlen ist:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Kürzen von Brüchen

Da $n! = n \cdot (n-1)!$, ist:

$$\frac{n!}{(n-1)!} = \frac{n \cdot (n-1)!}{(n-1)!} = n \quad \text{und} \quad \frac{n!}{(n-2)!} = \frac{n \cdot (n-1) \cdot (n-2)!}{(n-2)!} = n \cdot (n-1) \quad \text{usw.}$$

Beispiele:

$$\frac{8!}{7!} = \frac{8 \cdot 7!}{7!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 8$$

$$\frac{40!}{36!} = \frac{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36!}{36!} = 40 \cdot 39 \cdot 38 \cdot 37 = 2193360$$