Frage: Ist die Reaktionszeit in einem bestimmten Experiment vom Geschlecht unabhängig?

Gegebene Objekte:

Gesunde Personen zwischen 30 und 40 Jahre.

Merkmale:

T = Reaktionszeit (ms)
S = Geschlecht (m, w)

Zeit: Stetig

Geschlecht: Diskret, nominal / skaliert

Stetig ist die Zeit

↓

Klassen

$K_1 = [100; 130)$

$K_2 = [130; 160)$

$K_3 = [160; 190)$

200 Personen

$$Z = (T, S)$$

P.P.1

$$z_i = (t_i, s_i)$$

Merkmalsausprägung eines Probande i, $i = 1, \ldots, 200$

T \ S	0 m	1 w	Σ
[100, 130)	20	40	60
[130, 160)	50	60	110
[160, 190)	10	20	30
Σ	80	120	200

Absolute Zell-Häufigkeiten

T \ S	0 m	1 w	Σ
[100, 130)	0,1	0,2	0,3
[130, 160)	0,25	0,3	0,55
[160, 190)	0,05	0,1	0,15
Σ	0,4	0,6	1

Relative Zell-Häufigkeiten

Tab.: 3.1: Zweidimensionale Häufigkeitstabellen von Reaktionszeit T und Geschlecht S

Verteilungen in Gruppen, m und w bezogen auf Beobachtungszahlen aus der Gruppen:

T :	$[100, 130)$ K_1	$[130, 160)$ K_2	$[160, 190)$ K_3
m :	20/80	50/80	10/80
w :	40/120	60/120	20/120
Σ :	60/200	110/200	30/200

T:	K_1	K_2	K_3
u:	0,25	0,625	0,125
w:	0,33	0,5	0,17
Σ	0,3	0,55	0,15

Etwas Hilfe

? Mäßig ?

Zweidimensionale Häufigkeitsverteilungen

Sei $(X, Y) = Z$,

$x \in \mathbb{X}$, $y \in \mathbb{Y}$

ein 2-dim Merkmal

Fall 1: X, Y diskret

$\mathbb{X} = \{a_1, \ldots, a_k\}$

$\mathbb{Y} = \{b_1, \ldots, b_e\}$

$Z = \{(x, y)\}$

Wertebereich von Z:

$\not{Z} = \not{X} \times Y = \{c_{ij} = (a_i, b_j) \mid a_i \in \not{X},\ b_j \in Y\}$

Stichprobe vom Umfang n:

$z_i = (x_i, y_i), \quad i = 1, \ldots, n$

.

Bezeichnung:

$C_{ij} = (a_i, b_j)$ – Zelle

H_{ij} := # Aller Paare (x_k, y_k) in der Stichprobe mit

$x_k = a_i$

$y_k = b_j$

Absolute Zellenhäufigkeit

$h_{ij} := \frac{H_{ij}}{n}$ relative Zellenhäufigkeit,

Fall 2: X stetig $\vee$ Y stetig.

Klassenteilung von X bzw. Y (bzw. X und Y).

K_i^X $i = 1, \dots, k$ Klassenteilung von X

K_j^Y $j = 1, \dots, \ell$ Klassenteilung von Y

(K_i^X, K_j^Y) – (Klassen-) Zelle

H_{ij} – Anzahl aller (x_ν, y_ν) mit $x_\nu \in K_i^X$, $y_\nu \in K_j^Y$

h_{ij} – Anteil aller (x_ν, y_ν) mit $x_\nu \in K_i^X$, $y_\nu \in K_j^Y$

Def: Die Gesamtheit aller absoluten bzw. relativen Häufigkeiten:

$$\left((h_{ij})_{i=1,\dots,k}^{j=1,\dots,\ell} \right)$$

bzw.

$$\left((h_{ij})_{i=1,\dots,k}^{j=1,\dots,\ell} \right)$$

heißt zweidimensionale absolute bzw. relative Häufigkeitsverteilung von Z.

Kontingenztabelle (k x l –Tabelle):

X \ Y	K_1^Y b_1	K_2^Y b_2	...	K_j^Y b_j	...	K_l^Y b_l	Zeilensumme Σ
K_1^X a_1	H_{11}	H_{12}	...	H_{1j}	...	H_{1l}	$H_{1\bullet}$
K_2^X a_2	H_{21}	H_{22}	...	H_{2j}	...	H_{2l}	$H_{2\bullet}$
⋮	⋮	⋮		⋮		⋮	⋮
K_i^Y a_i	H_{i1}	H_{i2}	...	H_{ij}	...	H_{il}	$H_{i\bullet}$
⋮	⋮	⋮		⋮		⋮	⋮
K_k^X a_k	H_{k1}	H_{k2}	...	H_{kj}	...	H_{kl}	$H_{k\bullet}$
Σ Spaltensumme	$H_{\bullet 1}$	$H_{\bullet 2}$	...	$H_{\bullet j}$	...	$H_{\bullet l}$	n

Tab. 3.2

$$H_{i\bullet} = \sum_{j=1}^{l} H_{ij}$$ – i-te Zeilensumme

$$H_{\bullet j} = \sum_{i=1}^{k} H_{ij}$$ – j-te Spaltensumme

$H_{i\bullet}, H_{\bullet j}$ – Randsummen

$$H_{\bullet\bullet} = \sum_{i=1}^{k} \sum_{j=1}^{l} H_{ij} = n$$

– Stichprobenumfang

Def.: Die Gesamtheit der Randsummen bilden die Häufigkeitsverteilung von X bzw. Y:

$$\underbrace{\underbrace{\left((H_{i\cdot})_{i=1,\dots,k}\right)}_{\text{Häufigkeitsverteilung von } X}\ \underbrace{\left((H_{\cdot j})^{j=1,\dots,\ell}\right)}_{\text{Häufigkeitsverteilung von } Y}}_{\text{Randverteilung}}$$

Bem.: Wenn wir statt

H_{ij} die relative Häuf.

h_{ij} betrachten, so erhalten

wir die Randsummen

$h_{i\cdot}$, $h_{\cdot j}$. Diese sind

relative Häufigkeitsverteilungen

von X bzw. Y.

Häufigkeitstabelle der 1000 Beobachtungen:

Y / X	$2 < Y \leq 3$	$3 < Y \leq 4$	$4 < Y \leq 5$	$5 < Y \leq 6$	Summe
$4 < X \leq 5$	10	66	0	0	76
$5 < X \leq 6$	5	250	157	0	412
$6 < X \leq 7$	0	47	380	29	456
$7 < X \leq 8$	0	0	15	41	56
Summe	15	363	552	70	1000

Zweidimensionale Häufigkeitsverteilung für (X_g, Y_g)

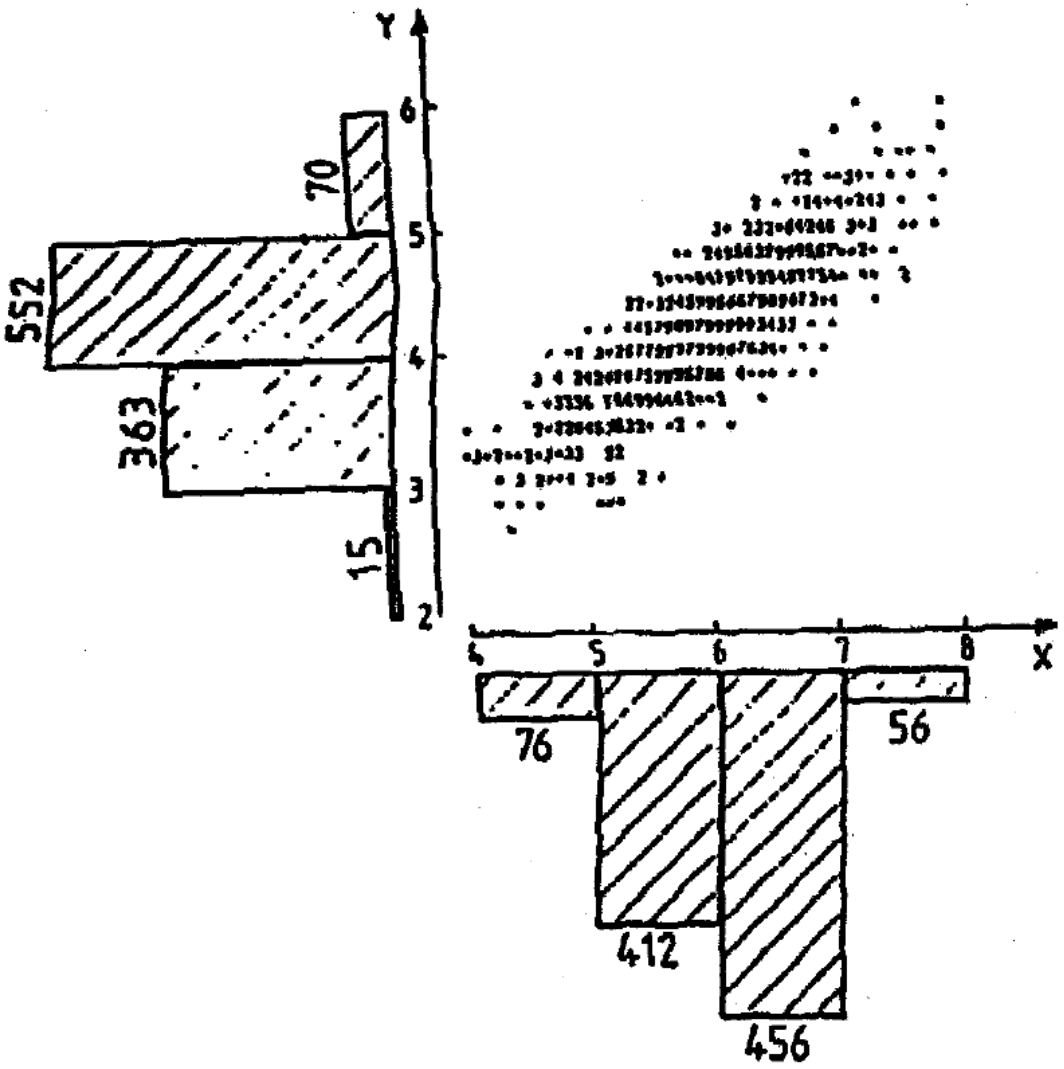
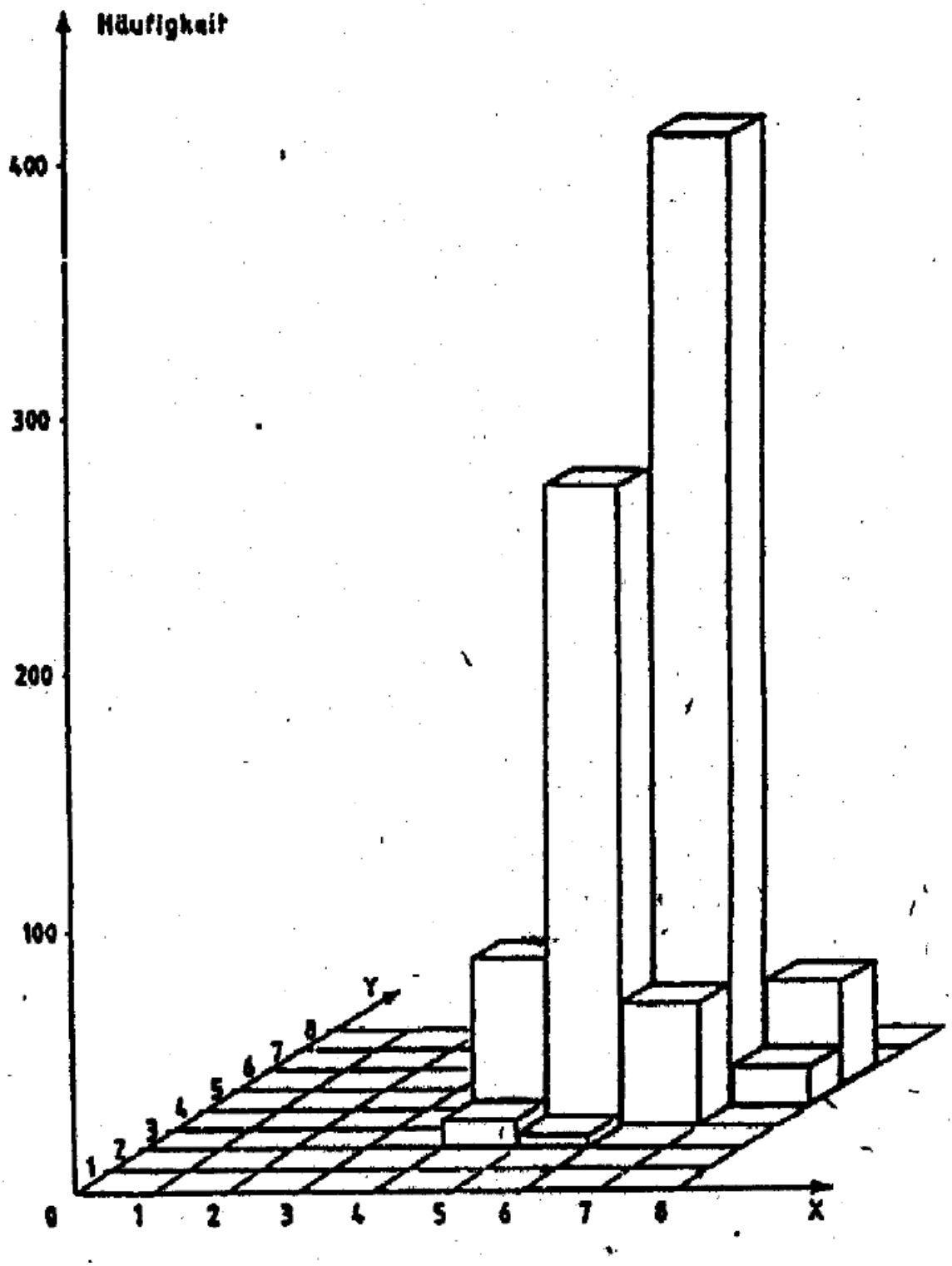


Abb.: Zweidimensionaler Plot


Häufigkeit
400
300
200
100
0
1
2
3
4
5
6
7
8
X
Y