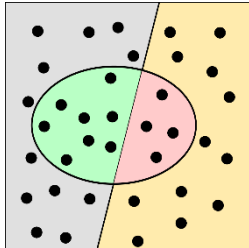


Ü03: Bivariate Deskriptive Statistik Aufgaben

Teil A

Bei einem diagnostischen Test (zur Überprüfung, ob eine bestimmte Krankheit vorliegt oder nicht) werden bei n Personen (schwarze Punkte im nachfolgenden Diagramm) zwei Merkmale mit folgenden Ausprägungen betrachtet:



K: „Krankheit“

K+: Krankheit liegt vor (links der schwarzen Trennlinie)
K-: Krankheit liegt nicht vor (rechts der schwarzen Trennlinie)

T: „Testergebnis“

T+: Test positiv (innerhalb des Ovals)
T-: Test negativ (außerhalb des Ovals)

- Charakterisieren Sie die vier Bereiche im Diagramm und stellen Sie den Zusammenhang zur entsprechenden Kontingenztafel her!

Es gelten weiterhin folgende Definitionen:

Prävalenz (Anteil der Kranken):	$p = f_k(K^+)$	Komplement:	$1 - p = f_k(K^-)$
Sensitivität (Richtig-positiv-Rate rpr):	$s = f_T(T^+ K^+)$		$1 - s = f_T(T^- K^+)$ (Falsch-negativ-Rate fnr)
Spezifität (Richtig-negativ-Rate nr):	$z = f_T(T^- K^-)$		$1 - z = f_T(T^+ K^-)$ (Falsch-positiv-Rate fpr)

- Tragen Sie Prävalenz, Sensitivität und Spezifität in die Kontingenztafel ein und ergänzen Sie die jeweiligen Komplemente!
- Ermitteln Sie in Abhängigkeit von n , p , s und z folgende Größen und tragen diese in die Kontingenztafel ein und ergänzen Sie die jeweiligen Komplemente:
 - die Anzahl der richtig positiv getesteten Personen r_p !
 - die Anzahl der richtig negativ getesteten Personen r_n !
 - die Anzahl der positiv getesteten Personen h_{T^+} !
 - die Anzahl der negativ getesteten Personen h_{T^-} !
 - den positiven prädiktiven Wert ppw , d.h. die relative bedingte Häufigkeit, dass die Krankheit vorliegt, wenn der Test positiv ist!
 - den negativen prädiktiven Wert npw , d.h. die relative bedingte Häufigkeit, dass die Krankheit nicht vorliegt, wenn der Test negativ ist!
 - die bei Unabhängigkeit erwarteten absoluten simultanen Häufigkeiten $\tilde{h}_{T,K}$!

- 4) Ermitteln Sie in Abhängigkeit von n , p , s und z folgende Größen:
- die Korrektklassifikationsrate kk_r ,
d.h. den Anteil der durch den Test korrekt klassifizierten Fälle!
 - die Falschklassifikationsrate fk_r ,
d.h. den Anteil der durch den Test falsch klassifizierten Fälle
 - die Chance krank zu sein vor Testdurchführung $odd(K^+)$!
 - die Chance krank zu sein, wenn das Testergebnis positiv ist $odd(K^+ | T^+)$!
 - den positiven Likelihood-Quotienten LQ_p ,
d.h. um welchen Faktor die Wahrscheinlichkeit eines positiven Tests größer ist,
wenn man infiziert ist als wenn man nicht infiziert ist!
Zeigen Sie, dass der positive Likelihood-Quotient gleich dem Chancenverhältnis
nach Testdurchführung mit positivem Ergebnis (d.) zu vor Testdurchführung (c.) ist.
 - die Chance krank zu sein, wenn das Testergebnis negativ ist $odd(K^+ | T^-)$!
 - den negativen Likelihood-Quotienten LQ_n ,
d.h. um welchen Faktor die Wahrscheinlichkeit eines negativen Tests größer ist,
wenn man infiziert ist als wenn man nicht infiziert ist!
Zeigen Sie, dass der negative Likelihood-Quotient gleich dem Chancenverhältnis
nach Testdurchführung mit negativem Ergebnis (f.) zu vor Testdurchführung (c.) ist.
 - das diagnostische Chancenverhältnis DOR (Diagnostic Odd Ratio),
d.h. um welchen Faktor der positive Likelihood-Quotient (e.) größer ist als der negative (g.)!

Ü03: Bivariate Deskriptive Statistik Aufgaben

Es werden bei $n = 100$ und $p = 0,10$ folgende Fälle betrachtet:

- a. Der Test liefert ausschließlich richtige Diagnosen.
- b. Der Test hat vollkommene Sensitivität, aber keine Spezifität.
- c. Der Test hat keine Sensitivität, aber vollkommene Spezifität.
- d. Der Test hat eine Sensitivität von $s = 0,80$ und eine Spezifität von $z = 0,90$.
- e. Der Test hat eine Sensitivität von $s = 0,50$ und eine Spezifität von $z = 0,50$.
- f. Der Test liefert ausschließlich falsche Diagnosen.

Bestimmen Sie jeweils für die Fälle a. bis f.

- 5) die Charakteristik des Tests!
- 6) die resultierende Kontingenztafel!
- 7) den normierten Kontingenzkoeffizient K^* nach Karl Pearson!
- 8) Cramérs V !
- 9) die Korrekturklassifikationsrate kk_r !
- 10) die Falschklassifikationsrate fk_r !
- 11) die Chance krank zu sein, wenn das Testergebnis positiv ist $\text{odd}(K^+ | T^+)!$
- 12) den positiven Likelihood-Quotienten LQ_p !
- 13) die Chance krank zu sein, wenn das Testergebnis negativ ist $\text{odd}(K^+ | T^-)!$
- 14) den negativen Likelihood-Quotienten LQ_n !
- 15) das diagnostische Chancenverhältnis DOR !

Ü03: Bivariate Deskriptive Statistik Aufgaben

Teil B

Bei 10 Personen werden folgende Merkmale erhoben:

X: „Anzahl der durchschnittlich pro Tag gerauchten Zigaretten“

Y: „Laufzeit in Sekunden auf 100 m“

Z: „Geschlecht“

Es ergaben sich folgende Werte:

i	x	y	z
1	3	12,2	m
2	8	12,6	m
3	12	13,3	m
4	15	14,5	m
5	16	14,8	m
6	0	15,2	w
7	3	15,8	w
8	5	16,2	w
9	6	17,5	w
10	7	19,2	w

Bei den Männern:

$$\sum_i (x_i | z_i = m) = 54 \quad \sum_i (y_i | z_i = m) = 67,4$$

$$\sum_i (x_i^2 | z_i = m) = 698 \quad \sum_i (y_i^2 | z_i = m) = 913,78 \quad \sum_i (x_i \cdot y_i | z_i = m) = 751,3$$

Bei den Frauen:

$$\sum_i (x_i | z_i = w) = 21 \quad \sum_i (y_i | z_i = w) = 83,9$$

$$\sum_i (x_i^2 | z_i = w) = 119 \quad \sum_i (y_i^2 | z_i = w) = 1.418,01 \quad \sum_i (x_i \cdot y_i | z_i = w) = 367,8$$

Korrelationen:

$$r_{xz} \approx -0,65 \quad r_{yz} \approx +0,80 \quad z_i = \begin{cases} 0 & z_i = m \\ 1 & z_i = w \end{cases}$$

- 1) Stellen Sie die Daten der Merkmale X und Y in einem Streudiagramm dar.
- 2) Berechnen und interpretieren Sie den Korrelationskoeffizienten nach Bravais-Pearson

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_i x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_i x_i^2 - n \cdot \bar{x}^2\right) \cdot \left(\sum_i y_i^2 - n \cdot \bar{y}^2\right)}}$$

für

- a) alle Personen insgesamt!
- b) bei den Männern!
- c) bei den Frauen!

- 3) Berechnen und interpretieren Sie die partielle Korrelation $r_{xy/z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2) \cdot (1 - r_{yz}^2)}}$!

Ü03: Bivariate Deskriptive Statistik Aufgaben

Teil C

In den Fällen I, II, III und IV ergeben sich folgende Datensätze für die Merkmale X und Y:

i	Fall I		Fall II		Fall III		Fall IV	
	x	y	x	y	x	y	x	y
1	4	4,26	4	3,10	4	5,39	8	5,25
2	5	5,68	5	4,74	5	5,73	8	5,56
3	6	7,24	6	6,13	6	6,08	8	5,76
4	7	4,82	7	7,26	7	6,42	8	6,58
5	8	6,95	8	8,14	8	6,77	8	6,89
6	9	8,81	9	8,77	9	7,11	8	7,04
7	10	8,04	10	9,14	10	7,46	8	7,71
8	11	8,33	11	9,26	11	7,81	8	7,91
9	12	10,84	12	9,13	12	8,15	8	8,47
10	13	7,58	13	8,74	13	12,74	8	8,84
11	14	9,96	14	8,10	14	8,84	19	12,50

- 1) Ermitteln Sie in den vier Fällen für X und Y
 - a) das arithmetische Mittel
 - b) die empirische Varianz
 - c) die empirische Standardabweichung
 - d) die Kovarianz
 - e) den Korrelationskoeffizient nach Bravais-Pearson
 - f) das Streudiagramm

- 2) Interpretieren Sie die Ergebnisse auf Basis
 - a) der Werten der Kennzahlen aus 1) a) bis e)
 - b) des Streudiagramms aus 1) f)