Ü03: Bivariate Deskriptive Statistik
Musterlösung

Teil A

1) im grünen Bereich: richtig als infiziert klassifiziert (r_p: richtig positiv)
im roten Bereich: falsch als infiziert klassifiziert (f_p: falsch positiv)
im grauen Bereich: falsch als nicht infiziert klassifiziert (f_n: falsch negativ)
im gelben Bereich: richtig als nicht infiziert klassifiziert (r_n: richtig negativ)

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$	$h_{T+,K-} = f_p$	$h_{T+} = r_p + f_p$
T-	$h_{T-,K+} = f_n$	$h_{T-,K-} = r_n$	$h_{T-} = f_n + r_n$
Σ	$h_{K+} = r_p + f_n$	$h_{K-} = f_p + r_n$	n

absolute Häufigkeiten h_T bzw. h_K
absolute simultane Häufigkeiten: $h_{T,K}$

2) Prävalenz p, Sensitivität s und Spezifität:

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$ $f_T(T^+ \mid K^+)$ $= r_p / h_{K+}$ = **rpr** = **s**	$h_{T+,K-} = f_p$	$h_{T+} = r_p + f_p$
T-	$h_{T-,K+} = f_n$	$h_{T-,K-} = r_n$ $f_T(T^- \mid K^-)$ $= r_n / h_{K-}$ = **rnr** = **z**	$h_{T-} = f_n + r_n$
Σ	$h_{K+} = r_p + f_n$ = **p***n	$h_{K-} = f_p + r_n$	n

Komplemente:

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$ $f_T(T^+ \mid K^+)$ $= r_p / h_{K+}$ = **rpr** = **s**	$h_{T+,K-} = f_p$ $f_T(T^+ \mid K^-)$ $= f_p / h_{K-}$ = **fpr** = **(1-z)**	$h_{T+} = r_p + f_p$
T-	$h_{T-,K+} = f_n$ $f_T(T^- \mid K^+)$ $= f_n / h_{K+}$ = **fnr** = **(1-s)**	$h_{T-,K-} = r_n$ $f_T(T^- \mid K^-)$ $= r_n / h_{K-}$ = **rnr** = **z**	$h_{T-} = f_n + r_n$
Σ	$h_{K+} = r_p + f_n$ = **p***n	$h_{K-} = f_p + r_n$ = **(1-p)***n	n

3) a) $s = \frac{r_p}{h_{K^+}} \Leftrightarrow r_p = s \cdot h_{K^+} = s \cdot p \cdot n$

$$h_{K^+} = r_p + f_n \Leftrightarrow f_n = h_{K^+} - r_p = p \cdot n - s \cdot p \cdot n = (1-s) \cdot p \cdot n$$

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$	$h_{T+,K-} = f_p$	$h_{T+} = r_p + f_p$
	$f_T(T^+ \mid K^+)$ $= r_p / h_{K+}$ $=$ **rpr = s**	$f_T(T^+ \mid K^-)$ $= f_p / h_{K-}$ $=$ **fpr = (1-z)**	
T-	$h_{T-,K+} = f_n$ $= (1-s)*p*n$	$h_{T-,K-} = r_n$	$h_{T-} = f_n + r_n$
	$f_T(T^- \mid K^+)$ $= f_n / h_{K+}$ $=$ **fnr = (1-s)**	$f_T(T^- \mid K^-)$ $= r_n / h_{K-}$ $=$ **rnr = z**	
Σ	$h_{K+} = r_p + f_n =$ **p***n	$h_{K-} = f_p + r_n =$ **(1-p)***n	n

b) $z = \frac{r_n}{h_{K^-}} \Leftrightarrow r_n = z \cdot h_{K^-} = z \cdot (1-p) \cdot n$

$$h_{K^-} = f_p + r_n \Leftrightarrow f_p = h_{K^-} - f_p = (1-p) \cdot n - z \cdot (1-p) \cdot n = (1-z) \cdot (1-p) \cdot n$$

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$	$h_{T+,K-} = f_p$ $(1-z)*(1-p)*n$	$h_{T+} = r_p + f_p$
	$f_T(T^+ \mid K^+)$ $= r_p / h_{K+}$ $=$ **rpr = s**	$f_T(T^+ \mid K^-)$ $= f_p / h_{K-}$ $=$ **fpr = (1-z)**	
T-	$h_{T-,K+} = f_n$ $= (1-s)*p*n$	$h_{T-,K-} = r_n$ $= z*(1-p)*n$	$h_{T-} = f_n + r_n$
	$f_T(T^- \mid K^+)$ $= f_n / h_{K+}$ $=$ **fnr = (1-s)**	$f_T(T^- \mid K^-)$ $= r_n / h_{K-}$ $=$ **rnr = z**	
Σ	$h_{K+} = r_p + f_n =$ **p***n	$h_{K-} = f_p + r_n =$ **(1-p)***n	n

c) $h_{T^+} = r_p + f_p = s \cdot n \cdot p + (1-z) \cdot (1-p) \cdot n = (s \cdot p + (1-z) \cdot (1-p)) \cdot n$

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$	$h_{T+,K-} = f_p$ $(1-z)*(1-p)*n$	$h_{T+} = r_p + f_p =$ $(s*p+(1-z)*(1-p))*n$
	$f_T(T^+ \mid K^+)$ $= r_p / h_{K+}$ $=$ **rpr = s**	$f_T(T^+ \mid K^-)$ $= f_p / h_{K-}$ $=$ **fpr = (1-z)**	
T-	$h_{T-,K+} = f_n$ $= (1-s)*p*n$	$h_{T-,K-} = r_n$ $= z*(1-p)*n$	$h_{T-} = f_n + r_n$
	$f_T(T^- \mid K^+)$ $= f_n / h_{K+}$ $=$ **fnr = (1-s)**	$f_T(T^- \mid K^-)$ $= r_n / h_{K-}$ $=$ **rnr = z**	
Σ	$h_{K+} = r_p + f_n =$ **p***n	$h_{K-} = f_p + r_n =$ **(1-p)***n	n

d) $$h_{T^-} = f_n + r_n = (1-s)\cdot n\cdot p + z\cdot(1-p)\cdot n = \left((1-s)\cdot p + z\cdot(1-p)\right)\cdot n$$

T \ K	K+	K-	Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$ $f_T(T^+\|K^+)$ $= r_p / h_{K+}$ **= rpr = s**	$h_{T+,K-} = f_p$ $(1-z)*(1-p)*n$ $f_T(T^+\|K^-)$ $= f_p / h_{K-}$ **= fpr = (1-z)**	$h_{T+} = r_p + f_p =$ $(s*p+(1-z)*(1-p))*n$
T-	$h_{T-,K+} = f_n$ $= (1-s)*p*n$ $f_T(T^-\|K^+)$ $= f_n / h_{K+}$ **= fnr = (1-s)**	$h_{T-,K-} = r_n$ $= z*(1-p)*n$ $f_T(T^-\|K^-)$ $= r_n / h_{K-}$ **= rnr = z**	$h_{T-} = f_n + r_n =$ $((1-s)*p+z*(1-p))*n$
Σ	$h_{K+} = r_p + f_n =$ **p***n	$h_{K-} = f_p + r_n =$ **(1-p)***n	n

e) $$ppw = f_K\left(K^+ \middle| T^+\right) = \frac{r_p}{h_{T^+}} = \frac{s\cdot p\cdot \cancel{n}}{\left(s\cdot p + (1-z)\cdot(1-p)\right)\cdot \cancel{n}} = \frac{s\cdot p}{s\cdot p + (1-z)\cdot(1-p)}$$

$$1-ppw = f_K\left(K^- \middle| T^+\right) = \frac{f_p}{h_{T^+}} = \frac{(1-z)\cdot(1-p)\cdot \cancel{n}}{\left(s\cdot p + (1-z)\cdot(1-p)\right)\cdot \cancel{n}} = \frac{(1-z)\cdot(1-p)}{s\cdot p + (1-z)\cdot(1-p)}$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$	$f_K(K^+\|T^+)$ $= r_p / h_{T+}$ **= ppw** $= s*p / (s*p+(1-z)*(1-p))$	$h_{T+,K-} = f_p$ $(1-z)*(1-p)*n$	$f_K(K^-\|T^+)$ $= f_p / h_{T+}$ **= (1-ppw)** $= (1-z)*(1-p) / (s*p+(1-z)*(1-p))$	$h_{T+} = r_p + f_p =$ $(s*p+(1-z)*(1-p))*n$
	$f_T(T^+\|K^+)$ $= r_p / h_{K+}$ **= rpr = s**		$f_T(T^+\|K^-)$ $= f_p / h_{K-}$ **= fpr = (1-z)**		
T-	$h_{T-,K+} = f_n$ $= (1-s)*p*n$		$h_{T-,K-} = r_n$ $= z*(1-p)*n$		$h_{T-} = f_n + r_n =$ $((1-s)*p+z*(1-p))*n$
	$f_T(T^-\|K^+)$ $= f_n / h_{K+}$ **= fnr = (1-s)**		$f_T(T^-\|K^-)$ $= r_n / h_{K-}$ **= rnr = z**		
Σ	$h_{K+} = r_p + f_n =$ **p***n		$h_{K-} = f_p + r_n =$ **(1-p)***n		n

f)

$$npw = f_K\left(K^-\middle|T^-\right) = \frac{r_n}{h_{T^-}} = \frac{z\cdot(1-p)\cdot\cancel{n}}{\left((1-s)\cdot p + z\cdot(1-p)\right)\cdot\cancel{n}} = \frac{z\cdot(1-p)}{(1-s)\cdot p + z\cdot(1-p)}$$

$$1-npw = f_K\left(K^+\middle|T^-\right) = \frac{f_n}{h_{T^-}} = \frac{(1-s)\cdot p\cdot\cancel{n}}{\left((1-s)\cdot p + z\cdot(1-p)\right)\cdot\cancel{n}} = \frac{(1-s)\cdot p}{(1-s)\cdot p + z\cdot(1-p)}$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$	$f_K(K^+\|T^+)$ $= r_p / h_{T+}$ = **ppw** = s*p / (s*p+(1-z)*(1-p))	$h_{T+,K-} = f_p$ (1-z)*(1-p)*n	$f_K(K^-\|T^+)$ $= f_p / h_{T+}$ = **(1-ppw)** = (1-z)*(1-p) / (s*p+(1-z)*(1-p))	$h_{T+} = r_p + f_p =$ (s*p+(1-z)*(1-p))*n
	$f_T(T^+\|K^+)$ $= r_p / h_{K+}$ = **rpr = s**		$f_T(T^+\|K^-)$ $= f_p / h_{K-}$ = **fpr = (1-z)**		
T-	$h_{T-,K+} = f_n$ = (1-s)*p*n	$f_K(K^+\|T^-)$ $= f_n / h_{T-}$ = **(1-npw)** = (1-s)*p / ((1-s)*p+z*(1-p))	$h_{T-,K-} = r_n$ = z*(1-p)*n	$f_K(K^-\|T^-)$ $= r_n / h_{T-}$ = **npw** = z*(1-p) / ((1-s)*p+z*(1-p))	$h_{T-} = f_n + r_n =$ ((1-s)*p+z*(1-p))*n
	$f_T(T^-\|K^+)$ $= f_n / h_{K+}$ = **fnr = (1-s)**		$f_T(T^-\|K^-)$ $= r_n / h_{K-}$ = **rnr = z**		
Σ	$h_{K+} = r_p + f_n =$ **p*n**		$h_{K-} = f_p + r_n =$ **(1-p)***n		n

g)

$$\tilde{h}_{T^+,K^+} = \frac{h_{T^+}\cdot h_{K^+}}{n} = \frac{\left(\left(s\cdot p+(1-z)\cdot(1-p)\right)\cdot n\right)\cdot\left(p\cdot\cancel{n}\right)}{\cancel{n}} = \left(s\cdot p+(1-z)\cdot(1-p)\right)\cdot p\cdot n$$

$$\tilde{h}_{T^+,K^-} = \frac{h_{T^+}\cdot h_{K^-}}{n} = \frac{\left(\left(s\cdot p+(1-z)\cdot(1-p)\right)\cdot n\right)\cdot\left((1-p)\cdot\cancel{n}\right)}{\cancel{n}} = \left(s\cdot p+(1-z)\cdot(1-p)\right)\cdot(1-p)\cdot n$$

$$\tilde{h}_{T^-,K^+} = \frac{h_{T^-}\cdot h_{K^+}}{n} = \frac{\left(\left((1-s)\cdot p+z\cdot(1-p)\right)\cdot n\right)\cdot\left(p\cdot\cancel{n}\right)}{\cancel{n}} = \left((1-s)\cdot p+z\cdot(1-p)\right)\cdot p\cdot n$$

$$\tilde{h}_{T^-,K^-} = \frac{h_{T^-}\cdot h_{K^-}}{n} = \frac{\left(\left((1-s)\cdot p+z\cdot(1-p)\right)\cdot n\right)\cdot\left((1-p)\cdot\cancel{n}\right)}{\cancel{n}} = \left((1-s)\cdot p+z\cdot(1-p)\right)\cdot(1-p)\cdot n$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = r_p$ $= s*p*n$	$f_K(K^+\|T^+)$ $= r_p / h_{T+}$ = **ppw** = s*p / (s*p+(1-z)*(1-p))	$h_{T+,K-} = f_p$ (1-z)*(1-p)*n	$f_K(K^-\|T^+)$ $= f_p / h_{T+}$ = **(1-ppw)** = (1-z)*(1-p) / (s*p+(1-z)*(1-p))	$h_{T+} = r_p + f_p =$ (s*p+(1-z)*(1-p))*n
	$f_T(T^+\|K^+)$ $= r_p / h_{K+}$ = **rpr = s**	$h\sim_{T+,K+} =$ (s*p+(1-z)*(1-p))*p*n	$f_T(T^+\|K^-)$ $= f_p / h_{K-}$ = **fpr = (1-z)**	$h\sim_{T+,K-} =$ (s*p+(1-z)*(1-p))*(1-p)*n	
T-	$h_{T-,K+} = f_n$ = (1-s)*p*n	$f_K(K^+\|T^-)$ $= f_n / h_{T-}$ = **(1-npw)** = (1-s)*p / ((1-s)*p+z*(1-p))	$h_{T-,K-} = r_n$ = z*(1-p)*n	$f_K(K^-\|T^-)$ $= r_n / h_{T-}$ = **npw** = z*(1-p) / ((1-s)*p+z*(1-p))	$h_{T-} = f_n + r_n =$ ((1-s)*p+z*(1-p))*n
	$f_T(T^-\|K^+)$ $= f_n / h_{K+}$ = **fnr = (1-s)**	$h\sim_{T-,K+} =$ ((1-s)*p+z*(1-p))*p*n	$f_T(T^-\|K^-)$ $= r_n / h_{K-}$ = **rnr = z**	$h\sim_{T-,K-} =$ ((1-s)*p+z*(1-p))*(1-p)*n	
Σ	$h_{K+} = r_p + f_n =$ **p*n**		$h_{K-} = f_p + r_n =$ **(1-p)***n		n

4) a)

$$kkr = \frac{r_p + r_n}{n} = \frac{s \cdot p \cdot \cancel{n} + z \cdot (1-p) \cdot \cancel{n}}{\cancel{n}} = s \cdot p + z \cdot (1-p)$$

b)

$$fkr = \frac{f_p + f_n}{n} = \frac{(1-z) \cdot (1-p) \cdot \cancel{n} + (1-s) \cdot p \cdot \cancel{n}}{\cancel{n}} = (1-z) \cdot (1-p) + (1-s) \cdot p$$

c)

$$odd(K^+) = \frac{h_{K^+}}{h_{K^-}} = \frac{p \cdot \cancel{n}}{(1-p) \cdot \cancel{n}} = \frac{p}{(1-p)}$$

d)

$$odd(K^+ | T^+) = \frac{h_{T^+,K^+}}{h_{T^+,K^-}} = \frac{s \cdot p \cdot \cancel{n}}{(1-z) \cdot (1-p) \cdot \cancel{n}} = \frac{s \cdot p}{(1-z) \cdot (1-p)}$$

e)

$$LQ_p = \frac{f_T\left(T^+ | K^+\right)}{f_T\left(T^+ | K^-\right)} = \frac{s}{1-z}$$

$$\frac{odd(K^+ | T^+)}{odd(K^+)} = \frac{\frac{s \cdot p}{(1-z) \cdot (1-p)}}{\frac{p}{(1-p)}} = \frac{\cancel{1-p}}{\cancel{p}} \cdot \frac{s \cdot \cancel{p}}{(1-z) \cdot \cancel{(1-p)}} = \frac{s}{1-z} = LQ_p$$

f)

$$odd(K^+ | T^-) = \frac{h_{T^-,K^+}}{h_{T^-,K^-}} = \frac{(1-s) \cdot p \cdot \cancel{n}}{z \cdot (1-p) \cdot \cancel{n}} = \frac{(1-s) \cdot p}{z \cdot (1-p)}$$

g)

$$LQ_n = \frac{f_T\left(T^- | K^+\right)}{f_T\left(T^- | K^-\right)} = \frac{1-s}{z}$$

$$\frac{odd(K^+ | T^-)}{odd(K^+)} = \frac{\frac{(1-s) \cdot p}{z \cdot (1-p)}}{\frac{p}{(1-p)}} = \frac{\cancel{1-p}}{\cancel{p}} \cdot \frac{(1-s) \cdot \cancel{p}}{z \cdot \cancel{(1-p)}} = \frac{1-s}{z} = LQ_n$$

h)

$$DOR = \frac{LQ_p}{LQ_n} = \frac{\frac{s}{1-z}}{\frac{1-s}{z}} = \frac{s}{1-z} \cdot \frac{z}{1-s} = \frac{s \cdot z}{(1-s) \cdot (1-z)}$$

a. 5) Der Test liefert ausschließlich richtige Diagnosen.
Perfekte Diagnose:
Fehldiagnosen sind ausgeschlossen.
Alle Kranken werden als solches erkannt und
alle Gesunden werden als solches erkannt.

6)
$$s = 1{,}00$$
$$z = 1{,}00$$
$$p \cdot n = 0{,}10 * 100 = 10$$
$$h_{T^+,K^+} = s \cdot n \cdot p = 1{,}00 \cdot 100 \cdot 0{,}10 = 10$$
$$h_{T^-,K^-} = z \cdot (1-p) \cdot n = 1{,}00 \cdot (1-0{,}10) \cdot 100 = 90$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = 10$				
	s = 1,00				
T-			$h_{T-,K-} = 90$		
			z = 1,00		
Σ	p*n = 10				n = 100

T \ K	K+		K-		Σ
T+	$r_p = 10$	ppw = 10/10 = 1,00	f_p = 90 - 90 = 0	= 0/10 = 0,00	= 10 + 0 = 10
	s = 1,00	= 10*10/100 = 1,00	= 0/90 = 0	= 90*10/100 = 9,00	
T-	f_n = 10 - 10 = 0	= 0/90 = 0	$r_n = 90$	npw = 90/90 = 1,00	= 0 + 90 = 90
	= 0/10 = 0	10*90/100 = 9,00	z = 1,00	= 90*90/100 = 81,00	
Σ	10		= 100 - 10 = 90		100

T \ K	K+		K-		Σ
T+	10	1,00	0	0,00	10
	1,00	1,00	0	9,00	
T-	0	0,00	90	1,00	90
	0,00	9,00	1,00	81,00	
Σ	10		90		100

7) $\chi^2 = \sum_{i,j} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$ $\qquad K_{max} = \sqrt{\frac{m-1}{m}} \overset{m=\min\{I=2,J=2\}=2}{=} \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}}$

$$= \frac{(10-1{,}00)^2}{1{,}00} + \frac{(0-9{,}00)^2}{9{,}00} + \frac{(0-9{,}00)^2}{9{,}00} + \frac{(90-81{,}00)^2}{81{,}00}$$

$$= \frac{81 \cdot 9^2 + 9 \cdot (-9)^2 + 9 \cdot (-9)^2 + 1 \cdot 9^2}{81} = \frac{\cancel{81} \cdot (81 + 9 + 9 + 1)}{\cancel{81}} = 100$$

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{100}{100 + 100}} = \sqrt{\frac{1}{2}}$$

$$K^* = \frac{K}{K_{max}} = \frac{\cancel{\sqrt{\frac{1}{2}}}}{\cancel{\sqrt{\frac{1}{2}}}} = 1$$

8) $V = \sqrt{\frac{\chi^2}{n \cdot (m-1)}} = \sqrt{\frac{100}{100 \cdot (2-1)}} = 1$

9) $kkr = s \cdot p + z \cdot (1-p) = 1{,}00 \cdot 0{,}10 + 1{,}00 \cdot (1-0{,}10) = 0{,}10 + 0{,}90 = 1{,}00$

10) $fkr = (1-z) \cdot (1-p) + (1-s) \cdot p = (1-1{,}00) \cdot (1-0{,}10) + (1-1{,}00) \cdot 0{,}10 = 0{,}00 + 0{,}00 = 0{,}00$

11) $odd(K^+ | T^+) = \frac{s \cdot p}{(1-z) \cdot (1-p)} = \frac{1{,}00 \cdot 0{,}10}{(1-1{,}00) \cdot (1-0{,}10)} = \frac{0{,}10}{0{,}00} = n.def.$

12) $LQ_p = \frac{s}{1-z} = \frac{1{,}00}{1-1{,}00} = \frac{1{,}00}{0{,}00} = n.def.$

13) $odd(K^+ | T^-) = \frac{(1-s) \cdot p}{z \cdot (1-p)} = \frac{(1-1{,}00) \cdot 0{,}10}{1{,}00 \cdot (1-0{,}10)} = \frac{0{,}00}{0{,}90} = 0{,}00$

14) $LQ_n = \frac{1-s}{z} = \frac{1-1{,}00}{1{,}00} = \frac{0{,}00}{1{,}00} = 0{,}00$

15) $DOR = \frac{LQ_p}{LQ_n} = \frac{n.def.}{0{,}00} = n.def.$

b. 5) Der Test hat vollkommene Sensitivität, aber keine Spezifität
Liberale Diagnose:
Alle Kranken werden als solches erkannt,
aber auch alle Gesunden werden für krank gehalten.

6)
$$s = 1{,}00$$
$$z = 0{,}00$$
$$p \cdot n = 0{,}10 * 100 = 10$$
$$h_{T^+,K^+} = s \cdot n \cdot p = 1{,}00 \cdot 100 \cdot 0{,}10 = 10$$
$$h_{T^-,K^-} = z \cdot (1-p) \cdot n = 0{,}00 \cdot (1-0{,}10) \cdot 100 = 0$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+}$ = 10				
	s = 1,00				
T-			$h_{T-,K-}$ = 0		
			z = 0,00		
Σ	p*n = 10				n = 100

T \ K	K+		K-		Σ
T+	r_p = 10	ppw = 10/100 = 0,10	f_p = 90 - 0 = 90	= 90/100 = 0,90	= 10 + 90 = 100
	s = 1,00	= 10*100/100 = 10,00	= 90/90 = 1,00	= 90*100/100 = 90,00	
T-	f_n = 10 - 10 = 0	= 0/0 = n.def.	r_n = 0	npw = 0/0 = n.def.	= 0 + 0 = 0
	= 0/10 = 0,00	10*0/100 = 0,00	z = 0,00	= 90*0/100 = 0,00	
Σ	10		= 100 - 10 = 90		100

T \ K	K+		K-		Σ
T+	10	0,10	90	0,90	100
	1,00	10,00	1,00	90,00	
T-	0	n.def.	0	n.def.	0
	0,00	0,00	0,00	0,00	
Σ	10		90		100

7) $$\chi^2 = \sum_{i,j} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}} \qquad K_{max} = \sqrt{\frac{m-1}{m}} \overset{m=\min\{I=2,J=2\}=2}{=} \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}}$$

$$= \frac{(10-10{,}00)^2}{10{,}00} + \frac{(90-90)^2}{90{,}00} + \frac{(0-0{,}00)^2}{0{,}00} + \frac{(0-0{,}00)^2}{0{,}00}$$

$$= 0 + 0 + \text{n.def.} + \text{n.def.} = \text{n.def.}$$

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{\text{n.def.}}{\text{n.def} + 100}} = \text{n.def}$$

$$K^* = \frac{K}{K_{max}} = \frac{\text{n.def.}}{\sqrt{\frac{1}{2}}} = \text{n.def.}$$

8) $$V = \sqrt{\frac{\chi^2}{n \cdot (m-1)}} = \sqrt{\frac{\text{n.def}}{100 \cdot (2-1)}} = \text{n.def.}$$

9) $$kkr = s \cdot p + z \cdot (1-p) = 1{,}00 \cdot 0{,}10 + 0{,}00 \cdot (1-0{,}10) = 0{,}10 + 0{,}00 = 0{,}10$$

10) $$fkr = (1-z)\cdot(1-p) + (1-s)\cdot p = (1-0{,}00)\cdot(1-0{,}10) + (1-1{,}00)\cdot 0{,}10 = 0{,}90 + 0{,}00 = 0{,}90$$

11) $$odd(K^+ \mid T^+) = \frac{s \cdot p}{(1-z)\cdot(1-p)} = \frac{1{,}00 \cdot 0{,}10}{(1-0{,}00)\cdot(1-0{,}10)} = \frac{0{,}10}{0{,}90} = 0{,}\overline{11}$$

12) $$LQ_p = \frac{s}{1-z} = \frac{1{,}00}{1-0{,}00} = \frac{1{,}00}{1{,}00} = 1{,}00$$

13) $$odd(K^+ \mid T^-) = \frac{(1-s)\cdot p}{z \cdot (1-p)} = \frac{(1-1{,}00)\cdot 0{,}10}{0{,}00 \cdot (1-0{,}10)} = \frac{0{,}00}{0{,}00} = \text{n.def.}$$

14) $$LQ_n = \frac{1-s}{z} = \frac{1-1{,}00}{0{,}00} = \frac{0{,}00}{0{,}00} = \text{n.def.}$$

15) $$DOR = \frac{LQ_p}{LQ_n} = \frac{1{,}00}{\text{n.def.}} = \text{n.def.}$$

c. 5) Der Test hat keine Sensitivität, aber vollkommene Spezifität
Konservative Diagnose:
Alle Gesunden werden als solches erkannt,
aber auch alle Kranken werden für gesund gehalten.

6)
$$s = 0{,}00$$
$$z = 1{,}00$$
$$p \cdot n = 0{,}10 * 100 = 10$$
$$h_{T^+,K^+} = s \cdot n \cdot p = 0{,}00 \cdot 100 \cdot 0{,}10 = 0$$
$$h_{T^-,K^-} = z \cdot (1-p) \cdot n = 1{,}00 \cdot (1-0{,}10) \cdot 100 = 90$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = 0$				
	s = 0,00				
T-			$h_{T-,K-} = 90$		
			z = 1,00		
Σ	p*n = 10				n = 100

T \ K	K+		K-		Σ
T+	$r_p = 0$	ppw = 0/0 = n.def.	f_p = 90 - 90 = 0	= 0/0 = n.def.	= 0 + 0 = 0
	s = 0,00	= 10*0/100 = 0,00	= 0/90 = 0,00	= 90*0/100 = 0,00	
T-	f_n = 10 - 0 = 10	= 10/100 = 0,10	$r_n = 90$	npw = 90/100 = 0,90	= 10 + 90 = 100
	= 10/10 = 1,00	10*100/100 = 10,00	z = 1,00	= 90*100/100 = 90,00	
Σ	10		= 100 - 10 = 90		100

T \ K	K+		K-		Σ
T+	0	n.def.	0	n.def.	0
	0,00	0,00	0,00	0,00	
T-	10	0,10	90	0,90	100
	1,00	10,00	1,00	90,00	
Σ	10		90		100

7) $\chi^2 = \sum_{i,j} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$ $\qquad K_{max} = \sqrt{\frac{m-1}{m}} \overset{m=\min\{I=2,J=2\}=2}{=} \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}}$

$$= \frac{(0-0{,}00)^2}{0{,}00} + \frac{(0-0{,}00)^2}{0{,}00} + \frac{(10-10{,}00)^2}{10{,}00} + \frac{(90-90{,}00)^2}{90{,}00}$$

$$= \text{n.def.} + \text{n.def} + 0 + 0 = \text{n.def.}$$

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{\text{n.def.}}{\text{n.def.} + 100}} = \text{n.def.}$$

$$K^* = \frac{K}{K_{max}} = \frac{\text{n.def.}}{\sqrt{\frac{1}{2}}} = \text{n.def.}$$

8) $V = \sqrt{\frac{\chi^2}{n \cdot (m-1)}} = \sqrt{\frac{\text{n.def}}{100 \cdot (2-1)}} = \text{n.def.}$

9) $kkr = s \cdot p + z \cdot (1-p) = 0{,}00 \cdot 0{,}10 + 1{,}00 \cdot (1-0{,}10) = 0{,}00 + 0{,}90 = 0{,}90$

10) $fkr = (1-z) \cdot (1-p) + (1-s) \cdot p = (1-1{,}00) \cdot (1-0{,}10) + (1-0{,}00) \cdot 0{,}10 = 0{,}00 + 0{,}10 = 0{,}10$

11) $odd(K^+ | T^+) = \frac{s \cdot p}{(1-z) \cdot (1-p)} = \frac{0{,}00 \cdot 0{,}10}{(1-1{,}00) \cdot (1-0{,}10)} = \frac{0{,}00}{0{,}00} = \text{n.def.}$

12) $LQ_p = \frac{s}{1-z} = \frac{0{,}00}{1-1{,}00} = \frac{0{,}00}{0{,}00} = \text{n.def.}$

13) $odd(K^+ | T^-) = \frac{(1-s) \cdot p}{z \cdot (1-p)} = \frac{(1-0{,}00) \cdot 0{,}10}{1{,}00 \cdot (1-0{,}10)} = \frac{0{,}10}{0{,}90} = 0{,}\overline{11}$

14) $LQ_n = \frac{1-s}{z} = \frac{1-1000}{1{,}00} = \frac{1{,}00}{1{,}00} = 1{,}00$

15) $DOR = \frac{LQ_p}{LQ_n} = \frac{\text{n.def.}}{1{,}00} = \text{n.def.}$

d. 5) Der Test hat eine Sensitivität von s = 0,80 und eine Spezifität von z = 0,90
Test mit eingeschränkter Sensitivität:
Nicht alle Kranken werden als solches erkannt,
also einige davon werden fälschlicherweise für gesund gehalten.
und eingeschränkter Spezifität:
Nicht alle Gesunden werden als solches erkannt,
also einige davon werden fälschlicherweise für krank gehalten.

6)
$$s = 0{,}80$$
$$z = 0{,}90$$
$$p \cdot n = 0{,}10 * 100 = 10$$
$$h_{T^+,K^+} = s \cdot n \cdot p = 0{,}80 \cdot 100 \cdot 0{,}10 = 8$$
$$h_{T^-,K^-} = z \cdot (1-p) \cdot n = 0{,}90 \cdot (1-0{,}10) \cdot 100 = 81$$

K / T	K+		K-		Σ
T+	$h_{T+,K+} = 8$				
	s = 0,80				
T-			$h_{T-,K-} = 81$		
			z = 0,90		
Σ	p*n = 10				n = 100

K / T	K+		K-		Σ
T+	$r_p = 8$	ppw = 8/17 = 0,47	$f_p = 90 - 81 = 9$	= 9/17 = 0,53	= 8 + 9 = 17
	s = 0,80	= 10*17/100 = 1,70	= 9/90 = 0,10	= 90*17/100 = 15,30	
T-	$f_n = 10 - 8 = 2$	= 2/83 = 0,02	$r_n = 81$	npw = 81/83 = 0,98	= 2 + 81 = 83
	= 2/10 = 0,20	10*83/100 = 8,30	z = 0,90	= 90*83/100 = 74,70	
Σ	10		= 100 - 10 = 90		100

K / T	K+		K-		Σ
T+	8	0,47	9	0,53	17
	0,80	1,70	0,10	15,30	
T-	2	0,02	81	0,98	83
	0,20	8,30	0,90	74,70	
Σ	10		90		100

7) $\chi^2 = \sum_{i,j} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$ $\qquad K_{max} = \sqrt{\frac{m-1}{m}} \overset{m=\min\{I=2,J=2\}=2}{=} \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}}$

$$= \frac{(8-1{,}70)^2}{1{,}70} + \frac{(9-15{,}30)^2}{15{,}30} + \frac{(2-8{,}30)^2}{8{,}30} + \frac{(81-74{,}70)^2}{74{,}70}$$

$$\approx 23{,}35 + 2{,}59 + 4{,}78 + 0{,}53 = 31{,}25$$

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}} \approx \sqrt{\frac{31{,}25}{31{,}25 + 100}} = 0{,}48795$$

$$K^* = \frac{K}{K_{max}} \approx \frac{0{,}48795}{\sqrt{\frac{1}{2}}} = \frac{0{,}48795}{0{,}70711} \approx 0{,}69$$

8) $V = \sqrt{\frac{\chi^2}{n \cdot (m-1)}} = \sqrt{\frac{31{,}25}{100 \cdot (2-1)}} \approx 0{,}5590$

9) $kkr = s \cdot p + z \cdot (1-p) = 0{,}80 \cdot 0{,}10 + 0{,}90 \cdot (1-0{,}10) = 0{,}08 + 0{,}81 = 0{,}89$

10) $fkr = (1-z) \cdot (1-p) + (1-s) \cdot p = (1-0{,}90) \cdot (1-0{,}10) + (1-0{,}80) \cdot 0{,}10 = 0{,}09 + 0{,}02 = 0{,}11$

11) $odd(K^+ | T^+) = \frac{s \cdot p}{(1-z) \cdot (1-p)} = \frac{0{,}80 \cdot 0{,}10}{(1-0{,}90) \cdot (1-0{,}10)} = \frac{0{,}08}{0{,}09} = 0{,}\overline{88}$

12) $LQ_p = \frac{s}{1-z} = \frac{0{,}80}{1-0{,}90} = \frac{0{,}80}{0{,}10} = 8$

13) $odd(K^+ | T^-) = \frac{(1-s) \cdot p}{z \cdot (1-p)} = \frac{(1-0{,}80) \cdot 0{,}10}{0{,}90 \cdot (1-0{,}10)} = \frac{0{,}02}{0{,}81} \approx 0{,}024691$

14) $LQ_n = \frac{1-s}{z} = \frac{1-0{,}80}{0{,}90} = \frac{0{,}20}{0{,}90} = 0{,}\overline{22}$

15) $DOR = \frac{LQ_p}{LQ_n} = \frac{8}{0{,}\overline{22}} = 36$

e. 5) Der Test hat eine Sensitivität von s = 0,50 und eine Spezifität von z = 0,50.
Der Test bringt keinen Erkenntnisgewinn über das Vorliegen der Krankheit und entspricht dem Wurf einer idealen Münze.

6)
$$s = 0{,}50$$
$$z = 0{,}50$$
$$p \cdot n = 0{,}10 * 100 = 10$$
$$h_{T^+,K^+} = s \cdot n \cdot p = 0{,}50 \cdot 100 \cdot 0{,}10 = 5$$
$$h_{T^-,K^-} = z \cdot (1-p) \cdot n = 0{,}50 \cdot (1-0{,}10) \cdot 100 = 45$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = 5$				
	s = 0,50				
T-			$h_{T-,K-} = 45$		
			z = 0,50		
Σ	p*n = 10				n = 100

T \ K	K+		K-		Σ
T+	$r_p = 5$	ppw = 5/50 = 0,10	f_p = 90 - 45 = 45	= 45/50 = 0,90	= 4 + 45 = 50
	s = 0,50	= 10*50/100 = 5,00	= 45/90 = 0,50	= 90*50/100 = 45,00	
T-	f_n = 10 - 5 = 5	= 5/50 = 0,10	$r_n = 45$	npw = 45/50 = 0,90	= 5 + 45 = 50
	= 5/10 = 0,50	10*50/100 = 5,00	z = 0,50	= 90*50/100 = 45,00	
Σ	10		= 100 - 10 = 90		100

T \ K	K+		K-		Σ
T+	5	0,10	45	0,53	50
	0,50	5,00	0,50	45,00	
T-	5	0,10	45	0,90	50
	0,50	5,00	0,50	45,00	
Σ	10		90		100

7) $$\chi^2 = \sum_{i,j} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}} \qquad K_{max} = \sqrt{\frac{m-1}{m}} \overset{m=\min\{I=2,J=2\}=2}{=} \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}}$$

$$= \frac{(5-5,70)^2}{5,00} + \frac{(45-45,00)^2}{45,00} + \frac{(5-5,00)^2}{5,00} + \frac{(45-45,00)^2}{45,00}$$

$$= 0 + 0 + 0 + 0 = 0$$

$$K = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{0}{0+100}} = 0$$

$$K^* = \frac{K}{K_{max}} = \frac{0}{\sqrt{\frac{1}{2}}} = 0$$

8) $$V = \sqrt{\frac{\chi^2}{n \cdot (m-1)}} = \sqrt{\frac{0}{100 \cdot (2-1)}} = 0$$

9) $$kkr = s \cdot p + z \cdot (1-p) = 0,50 \cdot 0,10 + 0,50 \cdot (1-0,10) = 0,05 + 0,45 = 0,50$$

10) $$fkr = (1-z) \cdot (1-p) + (1-s) \cdot p = (1-0,50) \cdot (1-0,10) + (1-0,50) \cdot 0,10 = 0,45 + 0,05 = 0,50$$

11) $$odd(K^+ | T^+) = \frac{s \cdot p}{(1-z) \cdot (1-p)} = \frac{0,50 \cdot 0,10}{(1-0,50) \cdot (1-0,10)} = \frac{0,05}{0,45} = 0,\overline{11}$$

12) $$LQ_p = \frac{s}{1-z} = \frac{0,50}{1-0,50} = \frac{0,50}{0,50} = 1$$

13) $$odd(K^+ | T^-) = \frac{(1-s) \cdot p}{z \cdot (1-p)} = \frac{(1-0,50) \cdot 0,10}{0,50 \cdot (1-0,10)} = \frac{0,05}{0,45} = 0,\overline{11}$$

14) $$LQ_n = \frac{1-s}{z} = \frac{1-0,50}{0,50} = \frac{0,50}{0,50} = 1$$

15) $$DOR = \frac{LQ_p}{LQ_n} = \frac{1}{1} = 1$$

f. 5) Der Test liefert ausschließlich falsche Diagnosen:
Imperfekte Diagnose:
Alle Kranken werden für gesund gehalten und
alle Gesunden für krank.

6)

$$s = 0{,}00$$

$$z = 0{,}00$$

$$p \cdot n = 0{,}10 * 100 = 10$$

$$h_{T^+,K^+} = s \cdot n \cdot p = 0{,}00 \cdot 100 \cdot 0{,}10 = 0$$

$$h_{T^-,K^-} = z \cdot (1-p) \cdot n = 0{,}00 \cdot (1-0{,}10) \cdot 100 = 0$$

T \ K	K+		K-		Σ
T+	$h_{T+,K+} = 0$				
	s = 0,00				
T-			$h_{T-,K-} = 0$		
			z = 0,00		
Σ	p*n = 10				n = 100

T \ K	K+		K-		Σ
T+	$r_p = 0$	ppw = 0/90 = 0,00	f_p = 90 - 0 = 90	= 90/90 = 1,00	= 0 + 90 = 90
	s = 0,00	= 10*90/100 = 9,00	= 90/90 = 1,00	= 90*90/100 = 81,00	
T-	f_n = 10 - 0 = 10	= 10/10 = 1,00	$r_n = 0$	npw = 0/10 = 0,00	= 10 + 0 = 10
	= 10/10 = 1,00	10*10/100 = 1,00	z = 0,00	= 90*10/100 = 9,00	
Σ	10		= 100 - 10 = 90		100

T \ K	K+		K-		Σ
T+	0	0,00	90	1,00	90
	0,00	9,00	1,00	81,00	
T-	10	1,00	0	0,00	10
	1,00	1,00	0,00	9,00	
Σ	10		90		100

7) $\chi^2 = \sum_{i,j} \frac{\left(h_{ij} - \tilde{h}_{ij}\right)^2}{\tilde{h}_{ij}}$ $\qquad K_{max} = \sqrt{\frac{m-1}{m}} \overset{m=\min\{I=2,J=2\}=2}{=} \sqrt{\frac{2-1}{2}} = \sqrt{\frac{1}{2}}$

$$= \frac{(0-9{,}00)^2}{9{,}00} + \frac{(90-81{,}00)^2}{81{,}00} + \frac{(10-1{,}00)^2}{1{,}00} + \frac{(0-9{,}00)^2}{9{,}00}$$

$$= \frac{9\cdot(-9)^2 + 1\cdot 9^2 + 81\cdot 9^2 + 9\cdot(-9)^2}{81} = \frac{\cancel{81}\cdot(9+1+81+9)}{\cancel{81}} = 100$$

$$K = \sqrt{\frac{\chi^2}{\chi^2+n}} = \sqrt{\frac{100}{100+100}} = \sqrt{\frac{1}{2}}$$

$$K^* = \frac{K}{K_{max}} = \frac{\cancel{\sqrt{\frac{1}{2}}}}{\cancel{\sqrt{\frac{1}{2}}}} = 1$$

8) $V = \sqrt{\frac{\chi^2}{n\cdot(m-1)}} = \sqrt{\frac{100}{100\cdot(2-1)}} = 1$

9) $kkr = s\cdot p + z\cdot(1-p) = 0{,}00\cdot 0{,}10 + 0{,}00\cdot(1-0{,}10) = 0{,}00 + 0{,}00 = 0{,}00$

10) $fkr = (1-z)\cdot(1-p) + (1-s)\cdot p = (1-0{,}00)\cdot(1-0{,}10) + (1-0{,}00)\cdot 0{,}10 = 0{,}90 + 0{,}10 = 1{,}00$

11) $odd(K^+ \mid T^+) = \frac{s\cdot p}{(1-z)\cdot(1-p)} = \frac{0{,}00\cdot 0{,}10}{(1-0{,}00)\cdot(1-0{,}10)} = \frac{0{,}00}{0{,}90} = 0{,}00.$

12) $LQ_p = \frac{s}{1-z} = \frac{0{,}00}{1-0{,}00} = \frac{0{,}00}{1{,}00} = 0{,}00$

13) $odd(K^+ \mid T^-) = \frac{(1-s)\cdot p}{z\cdot(1-p)} = \frac{(1-0{,}00)\cdot 0{,}10}{0{,}00\cdot(1-0{,}10)} = \frac{0{,}10}{0{,}00} = \text{n.def.}$

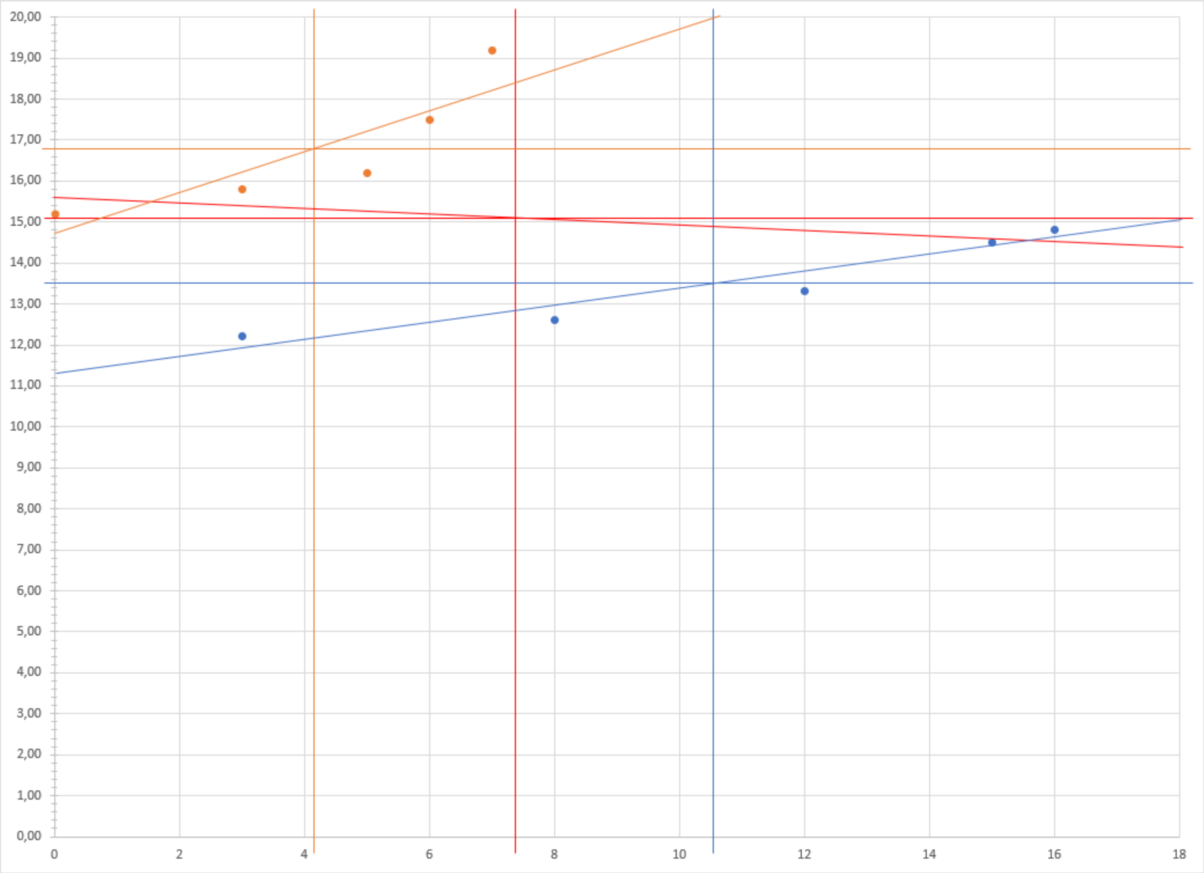
14) $LQ_n = \frac{1-s}{z} = \frac{1-0{,}00}{0{,}00} = \frac{1{,}00}{0{,}00} = \text{n.def.}$

15) $DOR = \frac{LQ_p}{LQ_n} = \frac{0{,}00}{\text{n.def.}} = \text{n.def.}$

Teil B

1)

Y: „Laufzeit in Sekunden auf 100 m“



X: „Anzahl der durchschnittlich pro Tag gerauchten Zigaretten“

Z: „Geschlecht“

- z_i = m: Männer
- z_i = w: Frauen

2) a)

$$\sum_i x_i \cdot y_i = \sum_i \left(x_i \cdot y_i | z_i = m\right) + \sum_i \left(x_i \cdot y_i | z_i = w\right) = 751{,}3 + 367{,}8 = 1.119{,}1$$

$$\bar{x} = \frac{\sum_i \left(x_i | z_i = m\right) + \sum_i \left(x_i | z_i = w\right)}{n} = \frac{54 + 21}{10} = 7{,}5$$

$$\bar{y} = \frac{\sum_i \left(y_i | z_i = m\right) + \sum_i \left(y_i | z_i = w\right)}{n} = \frac{67{,}4 + 83{,}9}{10} = 15{,}13$$

$$\sum_i x_i^2 = \sum_i \left(x_i^2 | z_i = m\right) + \sum_i \left(x_i^2 | z_i = w\right) = 698 + 119 = 817$$

$$\sum_i y_i^2 = \sum_i \left(y_i^2 | z_i = m\right) + \sum_i \left(y_i^2 | z_i = w\right) = 913{,}78 + 1.418{,}01 = 2.331{,}79$$

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

$$= \frac{1.119{,}1 - 10 \cdot 7{,}5 \cdot 15{,}13}{\sqrt{\left(817 - 10 \cdot 7{,}5^2\right) \cdot \left(2.331{,}79 - 10 \cdot 15{,}13^2\right)}} = \frac{-15{,}65}{\sqrt{254{,}5 \cdot 42{,}621}} \approx -0{,}15$$

Insgesamt liegt ein schwacher negativer linearer Zusammenhang zwischen X: „Anzahl gerauchter Zigaretten pro Tag“ und Y: „Laufzeit auf 100 m“ vor, d.h. je mehr Zigaretten pro Tag, desto geringer die Laufzeit auf 100 (also schneller). Das widerspricht der Erwartung, denn man würde vermuten, dass sich Rauchen eher negativ auf die sportliche Fitness auswirken würde.

b)

$$\bar{x}_m = \frac{\sum_i \left(x_i \middle| z_i = m\right)}{n_m} = \frac{54}{5} = 10{,}8$$

$$\bar{y}_m = \frac{\sum_i \left(y_i \middle| z_i = m\right)}{n_m} = \frac{67{,}4}{5} = 13{,}48$$

$$r_m = \frac{s_{xy_m}}{s_{x_m} \cdot s_{y_w}} = \frac{\sum_i \left(x_i \cdot y_i \middle| z_i = m\right) - n_m \cdot \bar{x}_m \cdot \bar{y}_m}{\sqrt{\left(\sum_i \left(x_i^2 \middle| z_i = m\right) - n_m \cdot \bar{x}_m^{\ 2}\right) \cdot \left(\sum_i \left(y_i^2 \middle| z_i = m\right) - n_m \cdot \bar{y}_m^{\ 2}\right)}}$$

$$= \frac{751{,}3 - 5 \cdot 10{,}8 \cdot 13{,}48}{\sqrt{\left(698 - 5 \cdot 10{,}8^2\right) \cdot \left(913{,}78 - 5 \cdot 13{,}48^2\right)}} = \frac{23{,}38}{\sqrt{114{,}8 \cdot 5{,}228}} \approx +0{,}95$$

Bei den Männern liegt ein sehr hoher positiver linearer Zusammenhang zwischen X: „Anzahl gerauchter Zigaretten pro Tag“ und Y: „Laufzeit auf 100 m“ vor, d.h. je mehr Zigaretten pro Tag, desto höher die Laufzeit auf 100 (also langsamer)

c)

$$\bar{x}_w = \frac{\sum_i \left(x_i \middle| z_i = w\right)}{n_w} = \frac{21}{5} = 4{,}2$$

$$\bar{y}_w = \frac{\sum_i \left(y_i \middle| z_i = w\right)}{n_w} = \frac{83{,}9}{5} = 16{,}78$$

$$r_w = \frac{s_{xy_w}}{s_{x_w} \cdot s_{y_w}} = \frac{\sum_i \left(x_i \cdot y_i \middle| z_i = w\right) - n_w \cdot \bar{x}_w \cdot \bar{y}_w}{\sqrt{\left(\sum_i \left(x_i^2 \middle| z_i = w\right) - n_w \cdot \bar{x}_w^{\ 2}\right) \cdot \left(\sum_i \left(y_i^2 \middle| z_i = w\right) - n_w \cdot \bar{y}_w^{\ 2}\right)}}$$

$$= \frac{367{,}8 - 5 \cdot 4{,}2 \cdot 16{,}78}{\sqrt{\left(119 - 5 \cdot 4{,}2^2\right) \cdot \left(1.418{,}01 - 5 \cdot 16{,}78^2\right)}} = \frac{15{,}42}{\sqrt{30{,}8 \cdot 10{,}168}} \approx +0{,}87$$

Bei den Frauen liegt ein sehr hoher positiver linearer Zusammenhang zwischen X: „Anzahl gerauchter Zigaretten pro Tag“ und Y: „Laufzeit auf 100 m“ vor, d.h. je mehr Zigaretten pro Tag, desto höher die Laufzeit auf 100 (also langsamer)

Hier liegt das sogenannte Simpson-Paradoxon vor.
Es gibt eine „Störvariable“ (Z: „Geschlecht“),

welche bei Nicht-Beachtung zu einer Scheinkorrelation führt.
Betrachtet man den Zusammenhang getrennt nach den Ausprägungen der Störvariablen, wird die Erwartung, dass sich Rauchen negativ auf die sportliche Fitness auswirkt, bestätigt.

Die Störvariable Z (Geschlecht) wirkt sich auf beide Variablen aus:
X: Männer rauchen mehr als Frauen und
Y: Männer sind schneller auf 100 m als Frauen.

Der Einfluss der Störvariablen ist dabei so groß,
dass insgesamt betrachtet sich der Effekt umdreht
und scheinbar Rauchen sich positiv auf die sportliche Fitness auswirkt.

3) $$r_{xy/z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{\left(1-r_{xz}^2\right)\cdot\left(1-r_{yz}^2\right)}} \approx \frac{-0{,}15-(-0{,}65)\cdot 0{,}80}{\sqrt{\left(1-(-0{,}65)^2\right)\cdot\left(1-0{,}80^2\right)}} = \frac{0{,}37}{\sqrt{0{,}2079}} \approx 0{,}81$$

Bei der partiellen Korrelation wird die Störvariable (Z: „Geschlecht“) herausgerechnet. Dann ergibt sich, wie erwartet, dass sich das Rauchen negativ auf die sportliche Fitness auswirkt.

Teil C

1) **Fall I**

	x	x²	y	y²	x·y
	4	16	4,26	18,1476	17,04
	5	25	5,68	32,2624	28,40
	6	36	7,24	52,4176	43,44
	7	49	4,82	23,2324	33,74
	8	64	6,95	48,3025	55,60
	9	81	8,81	77,6161	79,29
	10	100	8,04	64,6416	80,40
	11	121	8,33	69,3889	91,63
	12	144	10,84	117,5056	130,08
	13	169	7,58	57,4564	98,54
	14	196	9,96	99,2016	139,44
Σ	**99**	**1.001**	**82,51**	**660,1727**	**797,60**
Σ/n	**9**	**91**	**7,50**	**60,0157**	**72,51**

$\bar{x}_I = 9$ $\qquad$ $\bar{y}_I = 7{,}5$

$s^2_{X_I} = 91 - 9^2 = 91 - 81 = 10$ $\qquad$ $s^2_{Y_I} = 60{,}0157 - 7{,}5^2 = 60{,}0157 - 56{,}25 = 3{,}7657$

$s_{X_I} = \sqrt{10} \approx 3{,}1623$ $\qquad$ $s_{Y_I} \approx \sqrt{3{,}7657} \approx 1{,}9405$

$s_{XY_I} \approx 72{,}51 - 9 \cdot 7{,}5 = 5{,}01$ $\qquad$ $r_I \approx \dfrac{5{,}01}{3{,}1623 \cdot 1{,}9405} \approx 0{,}8164$

Fall II

	x	x²	y	y²	x·y
	4	16	3,10	9,6100	12,40
	5	25	4,74	22,4676	23,70
	6	36	6,13	37,5769	36,78
	7	49	7,26	52,7076	50,82
	8	64	8,14	66,2596	65,12
	9	81	8,77	76,9129	78,93
	10	100	9,14	83,5396	91,40
	11	121	9,26	85,7476	101,86
	12	144	9,13	83,3569	109,56
	13	169	8,74	76,3876	113,62
	14	196	8,10	65,6100	113,40
Σ	**99**	**1.001**	**82,51**	**660,1763**	**797,59**
Σ/n	**9**	**91**	**7,50**	**60,0160**	**72,51**

$\bar{x}_{I} = 9$

$s^2_{X_{II}} = 91 - 9^2 = 91 - 81 = 10$

$s_{X_{II}} = \sqrt{10} \approx 3{,}1623$

$s_{XY_{II}} \approx 72{,}51 - 9 \cdot 7{,}5 = 5{,}01$

$\bar{y}_{II} = 7{,}5$

$s^2_{Y_{II}} = 60{,}0160 - 7{,}5^2 = 60{,}0160 - 56{,}25 = 3{,}7660$

$s_{Y_{II}} \approx \sqrt{3{,}7660} \approx 1{,}9406$

$r_{II} \approx \frac{5{,}01}{3{,}1623 \cdot 1{,}9406} \approx 0{,}8164$

Fall III

	x	x²	y	y²	x·y
	4	16	5,39	29,0521	21,56
	5	25	5,73	32,8329	28,65
	6	36	6,08	36,9664	36,48
	7	49	6,42	41,2164	44,94
	8	64	6,77	45,8329	54,16
	9	81	7,11	50,5521	63,99
	10	100	7,46	55,6516	74,60
	11	121	7,81	60,9961	85,91
	12	144	8,15	66,4225	97,80
	13	169	12,74	162,3076	165,62
	14	196	8,84	78,1456	123,76
Σ	**99**	**1.001**	**82,51**	**659,9762**	**797,47**
Σ/n	**9**	**91**	**7,50**	**59,9978**	**72,50**

$\bar{x}_{III} = 9$

$s^2_{X_{III}} = 91 - 9^2 = 91 - 81 = 10$

$s_{X_{III}} = \sqrt{10} \approx 3{,}1623$

$s_{XY_{III}} \approx 72{,}50 - 9 \cdot 7{,}5 = 5$

$\bar{y}_{III} = 7{,}5$

$s^2_{Y_{III}} = 59{,}9978 - 7{,}5^2 = 59{,}9978 - 56{,}25 = 3{,}7478$

$s_{Y_{III}} \approx \sqrt{3{,}7478} \approx 1{,}9359$

$r_{III} \approx \frac{5}{3{,}1623 \cdot 1{,}9359} \approx 0{,}8167$

Fall IV

	x	x²	y	y²	x·y
	8	64	5,25	27,5625	42,00
	8	64	5,56	30,9136	44,48
	8	64	5,76	33,1776	46,08
	8	64	6,58	43,2964	52,64
	8	64	6,89	47,4721	55,12
	8	64	7,04	49,5616	56,32
	8	64	7,71	59,4441	61,68
	8	64	7,91	62,5681	63,28
	8	64	8,47	71,7409	67,76
	8	64	8,84	78,1456	70,72
	19	361	12,50	156,2500	237,50
Σ	**99**	**1.001**	**82,51**	**660,1325**	**797,58**
Σ/n	**9**	**91**	**7,50**	**60,0120**	**72,51**

$\bar{x}_{IV} = 9$

$s^2_{X_{IV}} = 91 - 9^2 = 91 - 81 = 10$

$s_{X_{IV}} = \sqrt{10} \approx 3{,}1623$

$s_{XY_{IV}} \approx 72{,}51 - 9 \cdot 7{,}5 = 5{,}01$

$\bar{y}_{IV} = 7{,}5$

$s^2_{Y_{IV}} = 60{,}0120 - 7{,}5^2 = 60{,}0120 - 56{,}25 = 3{,}7620$

$s_{Y_{IV}} \approx \sqrt{3{,}7620} \approx 1{,}9396$

$r_{IV} \approx \frac{5}{3{,}1623 \cdot 1{,}9396} \approx 0{,}8152$

a) $\bar{x}_{I} = 9 \quad = \quad \bar{x}_{II} = 9 \quad = \quad \bar{x}_{III} = 9 \quad = \quad \bar{x}_{IV} = 9$

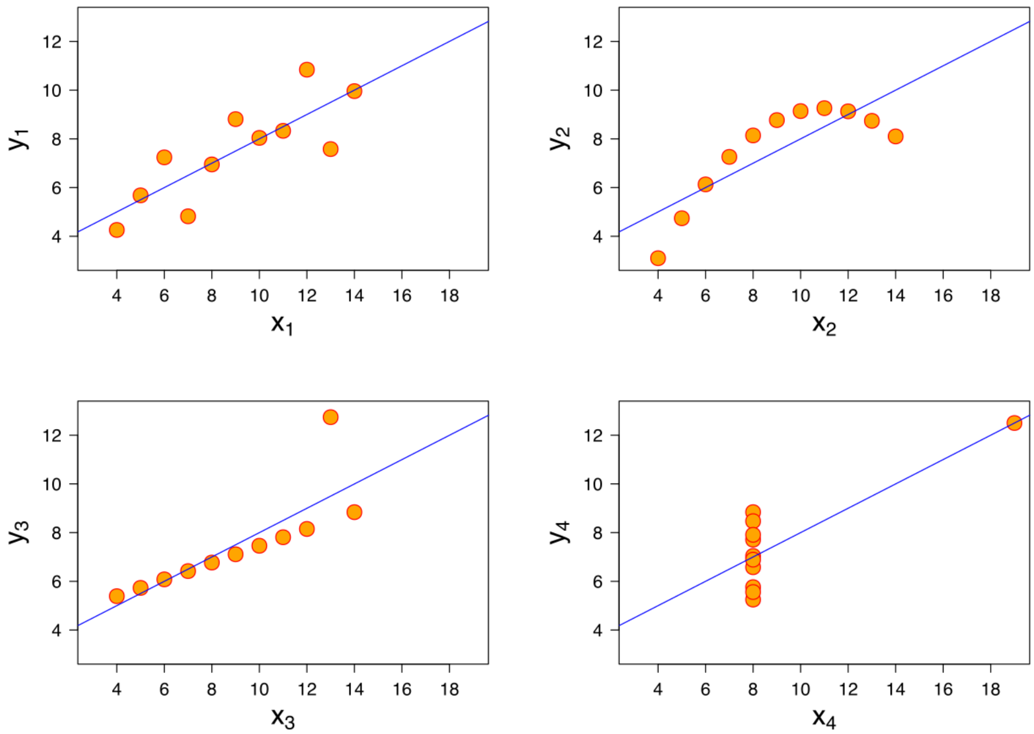
b) $s^2_{X_I} = 10 \quad = \quad s^2_{X_{II}} = 10 \quad = \quad s^2_{X_{III}} = 10 \quad = \quad s^2_{X_{IV}} = 10$

c) $s_{X_I} \approx 3{,}1623 \quad = \quad s_{X_{II}} \approx 3{,}1623 \quad = \quad s_{X_{III}} \approx 3{,}1623 \quad = \quad s_{X_{IV}} \approx 3{,}1623$

d) $s_{XY_I} \approx 5{,}01 \quad = \quad s_{XY_{II}} \approx 5{,}01 \quad \approx \quad s_{XY_{III}} \approx 5 \quad \approx \quad s_{XY_{IV}} \approx 5{,}01$

e) $r_I \approx 0{,}8164 \quad \approx \quad r_{II} \approx 0{,}8164 \quad \approx \quad r_{III} \approx 0{,}8167 \quad \approx \quad r_{IV} \approx 0{,}8152$

f)



2) a) Es liegt in allen 4 Fällen ein starker positiver linearer Zusammenhang vor ($r \approx 0{,}82$).

b) In **Fall I** liegt in der Tat ein starker positiver linearer Zusammenhang vor.
In **Fall II** gibt es zwar einen starken Zusammenhang, aber nicht linear (sondern eher parabolisch).
In **Fall III** liegt auch ein starker positiver linearer Zusammenhang vor, aber dieser wird durch den Ausreißer verzerrt.
In **Fall IV** schein es gar keinen (linearen) Zusammenhang zu geben, nur der Ausreißer führt dazu, dass dies rechnerisch so erscheint.

Es handelt sich hierbei um das sogenannte **Anscombe-Quartett**. Dies beinhaltet vier künstlich erzeugte Datensätze, die zwar zu nahezu identischen Werten bei den betrachteten Kennzahlen führen, aber dennoch unterschiedlich bewertet werden müssen. Dies erschließt sich erst aus der zusätzlichen Betrachtung des Streudiagramms.

Dies verdeutlicht die Bedeutung der graphischen Datenanalyse, die erfolgen sollte, bevor man aufgrund einer Annahme über die statistischen Eigenschaften der Daten mit der Analyse beginnt. Weiterhin zeigt es, dass einfache statistische Maßzahlen zur Beschreibung der Daten nicht immer ausreichen.