Statistik I
WS 2013/2014
Musterlösung 6
Aufgaben 19–22

Aufgabe 19

Bei Unabhängigkeit erwartet man ($\tilde{n}_{ij}$):

Ausbildungsbereich	männlich	weiblich	gesamt
Industrie und Handel	542,17	290,83	833,00
Handwerk	401,52	215,38	616,90
öffentlicher Dienst	30,92	16,58	47,50
Summe	974,60	522,80	1497,40

Dabei ist beispielsweise:

$$
\begin{aligned}
\tilde{n}_{1,1} &= \frac{n_{1,\bullet} \cdot n_{\bullet,1}}{n} \\
&= \frac{833,00 \cdot 974,60}{1497,40} \\
&= 542,17 \\
\tilde{n}_{3,2} &= \frac{n_{3,\bullet} \cdot n_{\bullet,2}}{n} \\
&= \frac{47,50 \cdot 522,80}{1497,40} \\
&= 16,58
\end{aligned}
$$

Damit berechnet man mit $\tilde{n}_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$:

$$
\begin{aligned}
\chi^2 =& \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{i,j} - \tilde{n}_{i,j})^2}{\tilde{n}_{i,j}} \\
=& \frac{(471,50 - 542,17)^2}{542,17} + \frac{(361,50 - 290,83)^2}{290,83} + \\
& \ldots + \frac{(29,90 - 16,58)^2}{16,58} \\
=& 9,2110 + 17,1711 + \ldots + 10,6918 \\
=& 93,1231
\end{aligned}
$$

Somit ergibt sich für den Kontingenzkoeffizienten bzw. den korrigierten Kontingenzkoeffizienten:

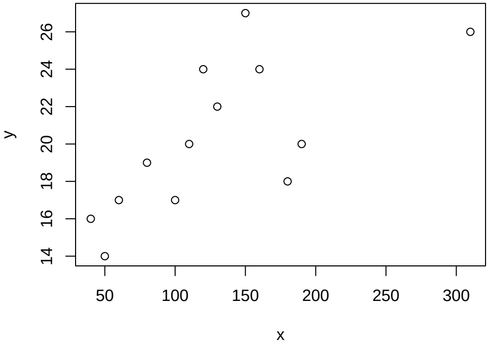
$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$
$$= \sqrt{\frac{93,1231}{93,1231 + 1497,40}}$$
$$= 0,2420$$

$$
\begin{aligned}
C^* &= \frac{C}{\sqrt{\frac{\min\{k,l\}-1}{\min\{k,l\}}}} \\
&= \sqrt{\frac{\chi^2}{\chi^2+n}\cdot\frac{\min\{k,l\}}{\min\{k,l\}-1}} \\
&= \sqrt{\frac{93,1231}{93,1231+1497,40}\cdot\frac{2}{2-1}} \\
&= 0,3422
\end{aligned}
$$

Das Ergebnis zeigt, dass die Berufswahl nicht unabhängig vom Geschlecht ist.

Aufgabe 20

(a)



(b)

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	40	16	1600	256	640
2	50	14	2500	196	700
3	60	17	3600	289	1020
4	80	19	6400	361	1520
5	100	17	10000	289	1700
6	110	20	12100	400	2200
7	120	24	14400	576	2880
8	130	22	16900	484	2860
9	150	27	22500	729	4050
10	160	24	25600	576	3840
11	180	18	32400	324	3240
12	190	20	36100	400	3800
13	310	26	96100	676	8060
$\sum$	1680	264	280200	5556	36510

$$
\begin{aligned}
\overline{x} &= \frac{1}{13} \cdot 1680 \\
&\approx 129,23 \\
\overline{y} &= \frac{1}{13} \cdot 264 \\
&\approx 20,31
\end{aligned}
$$

Für den Korrelationskoeffizienten berechnet man:

$$
\begin{aligned}
s_x^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\
&= \frac{1}{13} \cdot 280200 - 129,23^2 \\
&= 4853,25 \\
s_y^2 &= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{y}^2 \\
&= \frac{1}{13} \cdot 5556 - 20,31^2 \\
&= 14,98 \\
s_{x,y} &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y} \\
&= \frac{1}{13} \cdot 36510 - 129,23 \cdot 20,31 \\
&= 184,08
\end{aligned}
$$

Damit erhält man für den Korrelationskoeffizienten:

$$\begin{aligned} r_{x,y} &= \frac{s_{x,y}}{s_x \cdot s_y} \\ &= \frac{184,08}{\sqrt{4853,25} \cdot \sqrt{14,98}} \\ &= 0,6827 \end{aligned}$$

Aufgabe 21

Man verwendet nachfolgende Tabelle.

Universität	Studenten (x_i)	Professoren (y_i)	$rg(x_i)$	$rg(y_i)$
Köln	3,0	2,3	3,5	4
Frankfurt	3,1	2,4	6	5
Münster	3,2	2,1	8	2,5
Hamburg	3,0	3,1	3,5	9
Mannheim	3,1	1,6	6	1
Göttingen	2,9	2,8	2	8
München	2,6	2,6	1	7
Saarbrücken	3,1	2,1	6	2,5
Erlangen	3,3	2,5	9	6

Mit diesen Werten errechnet man ($\overline{rg(x)} = \overline{rg(y)} = 5$):

$$
\begin{aligned}
s^2_{rg(x)} &= \frac{1}{n}\sum_{i=1}^{n} rg(x_i)^2 - \overline{rg(x)}^2 \\
&= \frac{1}{9}(3,5^2 + 6^2 + \ldots + 9^2) - 5^2 \\
&= \frac{1}{9}\cdot 282,5 - 25 \\
&= 6,3\overline{8} \\
s^2_{rg(y)} &= \frac{1}{n}\sum_{i=1}^{n} rg(y_i)^2 - \overline{rg(y)}^2 \\
&= \frac{1}{9}(4^2 + 5^2 + \ldots + 6^2) - 5^2 \\
&= \frac{1}{9}\cdot 284,5 - 25 \\
&= 6,6\overline{1}
\end{aligned}
$$

$$
\begin{aligned}
s_{rg(x),rg(y)} &= \frac{1}{n}\sum_{i=1}^{n} rg(x_i)\cdot rg(y_i) - \overline{rg(x)}\cdot\overline{rg(y)} \\
&= \frac{1}{9}(3,5\cdot 4 + 6\cdot 5 + \ldots + 9\cdot 6) - 25 \\
&= \frac{1}{9}\cdot 193,5 - 25 \\
&= -3,5
\end{aligned}
$$

Damit erhält man:

$$
\begin{aligned}
r^{\mathrm{Sp}}_{x,y} &= \frac{s_{rg(x),rg(y)}}{s_{rg(x)}\cdot s_{rg(y)}} \\
&= \frac{-3,5}{\sqrt{6,3\overline{8}}\cdot\sqrt{6,6\overline{1}}} \\
&= -0,5385
\end{aligned}
$$

Die beiden Merkmale sind somit negativ korreliert.

Aufgabe 22

Man verwendet die Ergebnisse aus Aufgabe 22. Diese sind:

$$\begin{aligned} \bar{x} &= 129,23 \\ \bar{y} &= 20,31 \\ s_x^2 &= 4853,25 \\ s_{x,y} &= 184,08 \end{aligned}$$

Damit erhält man für den Steigungsparameter b bzw. für den y-Achsenabschnitt a:

$$\begin{aligned}
\hat{b} &= \frac{s_{x,y}}{s_x^2} \\
&= \frac{184,08}{4853,25} \\
&= 0,0379 \\
\hat{a} &= \bar{y} - \hat{b} \cdot \bar{x} \\
&= 20,31 - 0,0379 \cdot 129,23 \\
&= 15,406
\end{aligned}$$

Somit erhält man als Regressionsgerade:

$$\begin{aligned}\hat{y} &= \hat{a} + \hat{b} \cdot x \\ &= 15,406 + 0,0379 \cdot x\end{aligned}$$