11 Chancen

Lösung zu Aufgabe 19

Weibliche Teilpopulation:

$\gamma_w(1,2|MaschBau) = \frac{275}{6} = 45.83$
$\gamma_w(1,2|BWL) = \frac{314}{79} = 3.97$
$\gamma_w(1,2|KoWi) = \frac{202}{247} = 0.82$

Männliche Teilpopulation:

$\gamma_m(1,2|MaschBau) = \frac{342}{20} = 17.10$
$\gamma_m(1,2|BWL) = \frac{368}{252} = 1.46$
$\gamma_m(1,2|KoWi) = \frac{22}{30} = 0.73$

Odds Ratio:

$\gamma(1,2|weiblich, männlich) = \frac{\gamma(1,2|weiblich)}{\gamma(1,2|männlich)} =$

Beobachtung & Erklärung des Phänomens:

12 χ^2, K und K^*

Lösung zu Aufgabe 20

$$\chi^2_w = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(h_{ij} - \frac{h_{i\cdot}h_{\cdot j}}{n})^2}{\frac{h_{i\cdot}h_{\cdot j}}{n}} = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(h_{ij} - \frac{h_{i\cdot}h_{\cdot j}}{n})^2}{\frac{h_{i\cdot}h_{\cdot j}}{n}} =$$

$$= \frac{(275 - \frac{281\cdot 791}{1123})^2}{\frac{281\cdot 791}{1123}} + \frac{(314 - \frac{393\cdot 791}{1123})^2}{\frac{393\cdot 791}{1123}} + ... + \frac{(247 - \frac{449\cdot 332}{1123})^2}{\frac{449\cdot 332}{1123}} = 258.0473$$

$$\chi^2_m = \sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(h_{ij} - \frac{h_{i\cdot}h_{\cdot j}}{n})^2}{\frac{h_{i\cdot}h_{\cdot j}}{n}} = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{(h_{ij} - \frac{h_{i\cdot}h_{\cdot j}}{n})^2}{\frac{h_{i\cdot}h_{\cdot j}}{n}} =$$

$$= \frac{(342 - \frac{362\cdot 732}{1034})^2}{\frac{362\cdot 732}{1034}} + \frac{(368 - \frac{620\cdot 732}{1034})^2}{\frac{620\cdot 732}{1034}} + ... + \frac{(30 - \frac{52\cdot 302}{1034})^2}{\frac{52\cdot 302}{1034}} = 157.8287$$

$$K_w = \sqrt{\frac{\chi^2_w}{n + \chi^2_w}} = \sqrt{\frac{258.0473}{1034 + 258.0473}} = 0.4323$$

$$K_m = \sqrt{\frac{\chi^2_m}{n + \chi^2_m}} = \sqrt{\frac{157.8287}{1034 + 157.8287}} = 0.3639$$

Für beide Teilpopulationen gilt, dass $M = min\{k, m\} = min\{3, 2\} = 2$.

$$K^*_w = \frac{K_w}{\sqrt{\frac{M-1}{M}}} = \frac{0.4323}{\sqrt{\frac{2-1}{2}}} = 0.6113$$

$$K^*_m = \frac{K_w}{\sqrt{\frac{M-1}{M}}} = \frac{0.3639}{\sqrt{\frac{2-1}{2}}} = 0.5146$$

Interpretation:

Die Kontingenzmaße χ^2, K, K^* besitzen allesamt folgende Eigenschaften:

1. Gemessen wird die **Stärke** des Zusammenhangs, nicht aber dessen Richtung!
2. Die Maße sind **vergleichender Art** $\Rightarrow$ Eindeutige Interpretation nur in Spezialfällen ($K^* \in 0, 1$)!
3. **Invarianz** gegenüber Zeilen-/Spalten-Vertauschungen $\Rightarrow$ Eignung ab Nominalskala!

13 Korrelation und Kausalität

Lösung zu Aufgabe 21

Ein (perfekt) linearer Zusammenhang zwischen zwei Merkmalen X und Y lässt sich beschreiben durch die Gleichung $Y = aX + b$, wobei $a \in \mathbb{R}\backslash\{0\}, b \in \mathbb{R}$ fest, aber beliebig.
Einsetzen in die rechengünstige Formel des Korrelationskoeffizienten ergibt unter Anwendung der Transformationsregel:

$$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n y_i^2 - n\bar{y}^2)}} =$$

$$= \frac{\sum_{i=1}^n x_i(ax_i + b) - n\bar{x}(a\bar{x} + b)}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(\sum_{i=1}^n (ax_i + b)^2 - n(a\bar{x} + b)^2)}} =$$

$$= \frac{a\sum_{i=1}^n x_i^2 + b\sum_{i=1}^n x_i - an\bar{x}^2 - bn\bar{x}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(a^2\sum_{i=1}^n x_i^2 + b^2 n + 2ab\sum_{i=1}^n x_i - a^2 n\bar{x}^2 - b^2 n - 2abn\bar{x})}} =$$

$$= \frac{a(\sum_{i=1}^n x_i^2 - n\bar{x}^2)}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)(a^2(\sum_{i=1}^n x_i^2 - n\bar{x}^2))}} = \frac{a}{|a|} = sign(a) \in \{-1, 1\}$$

Lösung zu Aufgabe 22

Scheinkorrelation:
Von einer Scheinkorrelation spricht man, wenn man eine hohe Korrelation zwischen zwei Merkmalen X und Y beobachtet, zwischen denen sich aber kein kausaler Zusammenhang herstellen lässt. Solche scheinbaren Kausalzusammenhänge können dadurch bewirkt werden, dass ein mit X und Y hochkorreliertes Merkmal Z, die so genannte **Drittvariable**, übersehen wird und somit unberücksichtigt bleibt.

Beispiel : "Wer auf großem Fuß lebt, verdient mehr."

Zwischen der Schuhgröße (X) und dem Einkommen (Y) [von Personen im erwerbsfähigen Alter] besteht eine positive Korrelation...

<u>Verdeckte Korrelation:</u> Wenn bei der Betrachtung zweier Merkmale ein wichtiger Faktor übersehen wird, kann es sein, dass sich anhand des Korrelationskoeffizienten keine Korrelation festgestellen lässt, obschon diese tatsächlich besteht (und oft allein schon aus sachlogischen Gründen zu erwarten wäre). In diesem Fall zerfällt die untersuchte Population hinsichtlich dieses wichtigen Einflussfaktors (Drittvariable) in Teilpopulationen. (Im ungünstigsten Fall ist es dann sogar möglich, dass die Korrelationskoeffizienten einen negativen Zusammenhang anzeigen, obwohl der Zusammenhang nachweislich positiv ist.)
Um verdeckte Korrelationen zu ent-decken, sind **Streudiagramme** ein geeignetes Hilfsmittel!

Beispiel :

Lösung zu Aufgabe 23

Diese Aufgabe wird im Tutorium diskutiert!