Aufgabenblatt 8

Statistik für Wirtschaftsinformatiker, Übung, HTW Berlin

Martin Spott, Michael Heimann

Stand: 26.05.2024

Wiederholung

Diese Fragen beziehen sich auf qualitative Merkmale

- Was bedeutet es, wenn zwei Merkmale *unabhängig* voneinander sind?
- Was sind die *erwarteten Häufigkeiten im Falle von Unabhängigkeit* und wie berechnet man sie?
- Wie kann man den Grad der Unabhängigkeit messen?

Aufgabe 8.1

Wir benutzen die Daten von Aufgabe 7.1 (Aufgabenblatt 7):

```
library(knitr)
betriebe_daten <- matrix(c(639,64,41,487,131,41,203,153,33,54,91,17,46,112,18),
                nrow=5, ncol=3, byrow=T,
                dimnames=list(c("[0,50)", "[50, 180)",
                   "[180, 500)", "[500, 1000)", ">= 1000"),
                 c("Vollzeit", "Nebenerwerb", "Pacht")))

kable(betriebe_daten)
```

	Vollzeit	Nebenerwerb	Pacht
[0,50)	639	64	41
[50, 180)	487	131	41
[180, 500)	203	153	33
[500, 1000)	54	91	17
>= 1000	46	112	18

a) Berechnen Sie Pearsons χ^2-Statistik, den Φ-Koeffizienten, das Kontingenzmaß C nach Pearson und das Kontingenzmaß V nach Cramer aus der Tabelle mit R. Benutzen Sie dazu die Funktion `assocstats()` der Bibliothek `vcd` oder die Funktionen `Phi()`, `ContCoef()` oder `CramerV()` der Bibliothek `DescTools`.

b) Berechnen Sie die Kontingenztabelle für die erwarteten absoluten Häufigkeiten, die im Falle der Unabhängigkeit von Betriebsgröße und Betriebsführung auftreten würden.

c) Berechnen Sie die Kontingenztabelle für die erwarteten relativen Häufigkeiten, die im Falle der Unabhängigkeit von Betriebsgröße und Betriebsführung auftreten würden.

d) Erzeugen Sie einen Mosaikplot der Originaltabelle und der Tabelle mit den erwarteten Häufigkeiten und vergleichen Sie sie. Beschreiben und erklären Sie die Unterschiede.

e) Was sind die Werte von Pearsons χ^2-Statistik, dem Φ-Koeffizienten und des Kontingenzmaßes V nach Cramer für die Tabelle mit den erwarteten Häufigkeiten?

f) (Zusatzaufgabe) Berechnen Sie Pearsons χ^2-Statistik, den Φ-Koeffizienten und das Kontingenzmaß V nach Cramer aus Aufgabe c) händisch in R, in dem Sie die Formeln der Maße in R umsetzen.

Aufgabe 8.2

Benutzen Sie die Daten von Aufgabe 7.4 bezüglich der Wirkung einer Hautsalbe (Aufgabenblatt 7).

```r
daten <- matrix(c(223, 75, 107, 21), nrow=2, byrow=T,
                dimnames=list(c("mit Creme","ohne Creme"),
                              c("besser","schlechter")))

kable(daten)
```

	besser	schlechter
mit Creme	223	75
ohne Creme	107	21

a) Berechnen Sie die bedingten relativen Häufigkeiten $f(\text{Salbe verwendet}|\text{Ausschlag besser})$ und $f(\text{Salbe nicht verwendet}|\text{Ausschlag besser})$. Sagen diese beiden Häufigkeiten etwas darüber aus, ob die Anwendung der Salbe sinnvoll ist oder nicht? Interpretieren Sie die Ergebnisse.

b) Berechnen Sie die Kontingenztabelle für die erwarteten absoluten Häufigkeiten, die im Falle der Unabhängigkeit von Benutzung der Salbe und Hautausschlag auftreten würden.

c) Berechnen Sie Pearsons χ^2-Statistik, den Φ-Koeffizienten, das Kontingenzmaß C nach Pearson und das Kontingenzmaß V nach Cramer. Benutzen Sie dazu die Funktion `assocstats()` der Bibliothek `vcd`. Was sagen uns die berechneten Werte bezüglich des Grades der Unabhängigkeit?

d) Betrachten Sie folgende veränderte Kontingenztabelle:

```r
daten2 <- matrix(c(298, 0, 0, 128), nrow=2, byrow=T,
                 dimnames=list(c("mit Creme","ohne Creme"),
                               c("besser","schlechter")))

kable(daten2)
```

	besser	schlechter
mit Creme	298	0
ohne Creme	0	128

Interpretieren Sie die Tatsache, dass zwei der Werte Null sind. Berechnen Sie Pearsons χ^2-Statistik, den Φ-Koeffizienten und das Kontingenzmaß V nach Cramer und machen Sie sich die Werte klar.