

2. Übungsblatt vom Montag, den 13. November 2017 zur Vorlesung

## Induktive Statistik für Soziologinnen und Soziologen (Mariana Nold)

**Thema:** Grundlagen des statistischen Testens

Abgabe: keine Abgabe, wird in der Übung besprochen

Wichtige Definitionen:

1. Inferenzschluss
2. gerichtete Hypothese
3. ungerichtete Hypothese
4. Fehler 1. Art ( $\alpha$ -Fehler)
5. Fehler 2. Art ( $\beta$ -Fehler)
6. Teststatistik des Zwei-Stichproben t-Test
7. Ablauf des Zwei-Stichproben t-Test
8. kritischer Wert und Ablehnbereich
9. p-Wert

### 4. Der $\chi^2$ -Test: Statusverlustangst und relative Deprivation im Thüringen-Monitor 2015

(vgl. Aufgabe 18, des 5. Übungsblattes im letzten Semester.)

Leseempfehlung: LWLG: S. 212-218

Schlagen Sie Ihre Formelsammlung auf S. 51 auf. Dort finden Sie unter 8.5.1 den  $\chi^2$ -Unabhängigkeitstest. Das Wort Prüfgröße ist ein Synonym für Teststatistik. Die Teststatistik des  $\chi^2$ -Unabhängigkeitstest ist mit dem  $\chi^2$ -Koeffizienten identisch. Die Prüfgröße lautet

$$\chi^2 = \sum_{j=1}^m \sum_{k=1}^q \frac{(h(x_j, y_k) - h_e(x_j, y_k))^2}{h_e(x_j, y_k)}.$$

- Es ist  $h(x_j, y_k)$  die Häufigkeit mit der die Kombination  $(x_j, y_k)$  beobachtet wurde,  $j \in \{1, 2\}$  und  $k \in \{1, 2\}$ .

- Es ist  $h_e(x_j, y_k)$  die bei Unabhängigkeit erwartete Häufigkeit der Kombination  $(x_j, y_k)$ .

Zur Erinnerung an Definitionen aus der Vorlesung (im letzten Semester):

Die Frage nach der Statusverlustangst: “Es macht mir Sorgen, durch die gesellschaftliche Entwicklung immer mehr auf die Verliererseite des Lebens zu geraten.”

- Die Antwortkategorien: “lehne ab,” “stimme zu”
- Im Datensatz: “lehne ab = nein” ( $y_1$ ), “stimme zu = ja” ( $y_2$ )

Die Frage zur Deprivation: “Im Vergleich dazu, wie andere hier in Deutschland leben: Glauben Sie, dass Sie Ihren gerechten Anteil erhalten, mehr als Ihren gerechten Anteil, etwas weniger oder sehr viel weniger?”

- Die Antwortkategorien: “erhalte sehr viel weniger,” “erhalte etwas weniger,” “erhalte gerechten Anteil,” “erhalte mehr als gerechten Anteil,” “weiß nicht,” “keine Angabe”
- Im Datensatz:

$$x_1 : \text{“erhalte weniger”} = \begin{cases} \text{“erhalte sehr viel weniger,”} \\ \text{“erhalte etwas weniger”} \end{cases}$$

$$x_2 : \text{“erhalte genug”} = \begin{cases} \text{“erhalte gerechten Anteil,”} \\ \text{“erhalte mehr als gerechten Anteil”} \end{cases}$$

Im Thüringen-Monitor 2015 haben 940 Personen sowohl die Frage nach der Statusverlustangst als auch nach der Deprivation beantwortet. Die anderen 70 Personen haben mit “weiß nicht” oder “keine Angabe” geantwortet oder die Antwort vollständig verweigert. Die Vierfeldertafel 1 zeigt die bivariaten Häufigkeiten der 940 Personen.

Tabelle 1: Bivariate Häufigkeit: Statusverlustangst  $Y$  und Deprivation  $X$

	erhalte weniger	erhalte genug
nein	289	362
ja	214	75

Lösen Sie die folgenden Aufgaben mit Hilfe von **STATA 14**:

- Lesen Sie die beobachtete Häufigkeit  $h(x_1, y_2)$  aus Tabelle 1 und interpretieren Sie diese Häufigkeit.
- Lesen Sie mit dem Befehl `. use ‘monitor_15.dta’, clear` den Datensatz ein und erstellen Sie die Vierfeldertafel 1.
- Geben Sie den Befehl `. tab2 Angst Deprivation, chi2` in das Command-Window ein und erklären Sie den Werte des  $\chi^2$ -Koeffizienten und das ab.

- (d) Erklären Sie mit Bezug auf den  $\chi^2$ -Koeffizient warum hohe Werte der Teststatistik gegen eine Unabhängigkeit der beiden Merkmale  $X$  und  $Y$  sprechen.
- (e) Erklären sie mit Bezug auf die angegeben  $2 \times 2$  Kreuztabelle, was man unter einem Inferenzschluss versteht.
- (f) Wir möchten nachweisen, dass die beiden Merkmale  $X$  und  $Y$  (stochastisch) abhängig sind, wie können entsprechend die Nullhypothese und die Forschungshypothese formuliert werden?
- (g) Was versteht man unter dem Signifikanzniveau bzw. der Irrtumswahrscheinlichkeit?
- (h) Legen Sie das Signifikanzniveau fest (auf 5%) und suchen Sie im **STATA-output** des Befehls `. tab2 Angst Deprivation, chi2` nach dem p-Wert. Kann die Forschungshypothese nachgewiesen werden?
- (i) Es lässt sich zeigen, dass der  $\chi^2$ -Koeffizient unter  $H_0$  einer  $\chi^2$ -Verteilung mit einem Freiheitsgrad folgt. Diese ist in Abbildung 1 dargestellt. Es ist auch der kritische Wert  $k$  eingezeichnet. Der kritische Wert ist 3.84.
  - i) Der  $\chi^2$ -Koeffizient entspricht im  $\chi^2$ -Test der Teststatistik. Vergleichen Sie den Wert der Teststatistik mit dem kritischen Wert  $k = 3.84$ .
  - ii) Schlagen Sie die Formelsammlung auf S. 62 auf. Wie können sie den kritischen Wert ablesen?
  - iii) Markieren Sie den Ablehnbereich farbig in Abbildung 1 und malen Sie die Fläche zwischen Ablehnbereich und Dichtkurve aus. Wie hoch ist der Anteil der ausgemalten Fläche an der Fläche unter der Dichtkurve.
  - iv) Quantifizieren Sie die Unsicherheit des Inferenzschlusses indem Sie Auskunft über den  $\alpha$ - und  $\beta$ -Fehler geben.

## 5. Approximativer und exakter Binomialtest

Ein Politiker behauptet in einer Pressemitteilung mit Bezug auf den Thüringen-Monitor 2015, dass der Anteil der Thüringer Wahlbevölkerung der angibt weniger als seinen gerechten Anteil zu erhalten mittlerweile auf über 50% gestiegen ist. Er behauptet weiter, dass dieser Anteil unter den Arbeitslosen sogar mehr als 60% beträgt. Schlagen Sie Ihre Formelsammlung auf S. 49 auf. Der dort beschriebene Test heißt approximativer Binomialtest.

- (a) Berechnen Sie

$$\frac{9}{p(1-p)}$$

für  $p = 0.5$  um zu entscheiden, ob der Test für die Überprüfung der ersten Behauptung des Politikers geeignet ist.

- (b) Um den Test zu rechnen, muss die Variable **Deprivation** in einer binären Kodierung vorliegen. Erzeugen Sie mit dem **STATA-Befehl** `. gen depri01 =.` zunächst eine Variable, die nur fehlende Werte enthält. Mit den Befehlen

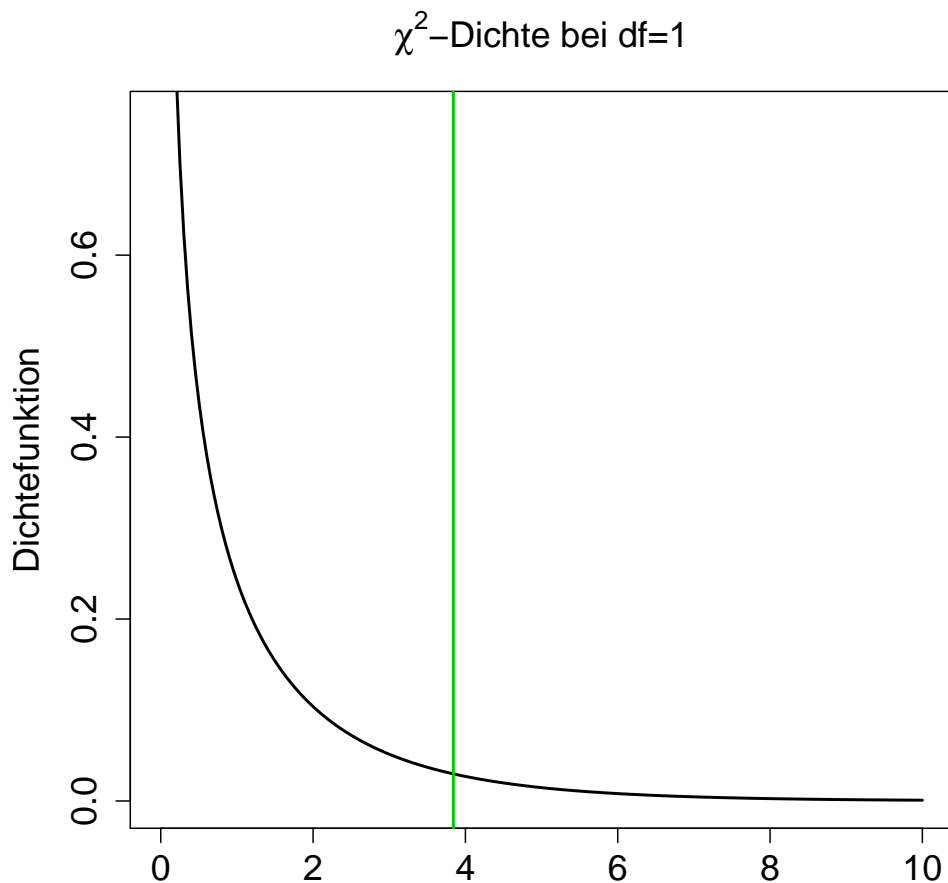


Abbildung 1: Dichte der  $\chi^2$ -Verteilung mit einem Freiheitsgrad

```
. replace depri01 = 0 if(Deprivation==2)
und
```

```
. replace depri01 = 1 if(Deprivation==1)
```

haben Sie eine binäre erzeugen Sie eine binäre Variable. Überprüfen Sie mit den Befehlen `. tab Deprivation` und `. tab depri01` ob die neue Variable `depri01` die gleiche Häufigkeitsverteilung haben.

- (c) Wie hoch ist in der Stichprobe der Anteil derer, die die Frage nach der Deprivation bejahen?
- (d) Nutzen Sie den Befehl `. prtest depri01 = 0.5` um die Teststatistik des auf Seite 49 beschriebenen Tests zu rechnen. Interpretieren Sie das Ergebnis und nehmen Sie Stellung zu der ersten Behauptung des Politikers.
- (e) Wie viele Arbeitslose sind in der Stichprobe? Berechnen Sie

$$\frac{9}{p(1-p)}$$

für  $p = 0.6$ . Ist der approximative Binomialtest für die Testentscheidung geeignet?

- (f) Als Alternative bei kleinen Stichprobenumfängen gibt es den exakten Binomialtest. Dieser Test kann in **STATA** mit Hilfe des Befehls **bittest** aufgerufen werden. Nutzen Sie den **STATA**-Befehl
- ```
. bittest depri01 = 0.6 if(EWTS==4)
```
- um mit Hilfe des exakten Tests die Behauptung des Politikers zu überprüfen. Wie hoch ist der Anteil der Personen, die die Frage nach der Deprivation bejahen unter den Arbeitslosen in der Stichprobe? Ist es beruhend auf diese Stichprobe sinnvoll davon auszugehen, dass der Anteil der Personen die depriviert in der Grundgesamtheit sind über 60% liegt? Begründen Sie ihre Antwort.
- (g) Es ist erlaubt, den exakten Test bei großen Stichproben zu verwenden. Umgekehrt ist es problematisch den approximativen Test bei kleinen Stichproben zu verwenden. Nutzen Sie den exakten Test um die erste Behauptung des Politikers zu überprüfen und vergleichen Sie das Ergebnis mit dem Ergebnis unter Teilaufgabe d).