

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

Induktive Statistik für Soziologinnen und Soziologen (Mariana Nold)

Thema: Grundlagen des statistischen Testens

Abgabe: Freitag 24. November 2017

Wichtige Definitionen:

1. Inferenzschluss
2. gerichtete Hypothese
3. ungerichtete Hypothese
4. Fehler 1. Art (α -Fehler)
5. Fehler 2. Art (β -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 Ablauf eines Hypothesentestes

(10 Punkte)

Schildern Sie wie der Hypothesentest in der klassischen Inferenz abläuft. Gehen Sie dabei auf die Frage ein, was ein Inferenzschluss bedeutet.

**Aufgabe
abgeben!**

5. Lesen Mädchen besser als Jungs: Doppelter t-Test und Welch-Test

Im Sommer-Semester haben wir uns in der letzten Vorlesung am 19.06.2017 mit der Frage beschäftigt, ob Mädchen besser lesen als Jungen. Diese Fragestellung sollen Sie nun mit dem doppelten t-Test untersuchen, den Sie in der Vorlesung kennengelernt haben.

- Der Index j wird genutzt um die Gruppen zu identifizieren:

$$j = \begin{cases} 1, & \text{“weiblich,”} \\ 2, & \text{“männlich”} \end{cases}$$

- Es sei $y_{i,j}$ die Anzahl der Kompetenzpunkte im Bereich Lesen, der i -ten Person in Gruppe j .

- So ist z. B. $y_{3,1}$ die Leseleistung der dritten Schülerin und
 - $y_{100,2}$ die Leseleistung des 100. Schülers.
 - Die mittlere Leseleistung der Schülerinnen ist \bar{y}_1 , die mittlere Leseleistung der Jungen ist \bar{y}_2 .
 - Es gibt $n_1 := 2462$ Schülerinnen und $n_2 := 2539$ Schüler im Datensatz.
- (a) Lesen Sie den Datensatz `pisa.2012.schueler.dta` ein. Nutzen Sie den STATA-Befehl
- ```
. keep if cnt== 'DEU'
```
- um alle Daten von nicht-deutschen Schülerinnen und Schülern aus dem Speicher zu nehmen. Geben sie den Befehl
- ```
. browse
```
- ein um sich den Datensatz anzusehen.
- (b) Die Information über das Geschlecht ist in der Variable `st04q01` enthalten. Nutzen Sie den Befehl
- ```
. tab st04q01
```
- um eine Häufigkeitstabelle zu erstellen.
- (c) Berechnen Sie für die Gruppe der Mädchen und Jungen jeweils die mittlere Anzahl der Lesepunkte und beschreiben Sie das Ergebnis jeweils in einem Satz. Nutzen Sie die dazu die Befehle `. summarize pvlread if st04q01==1` und `. summarize pvlread if st04q01==2`
- (d) Stellen Sie die Verteilung der Lesepunkte für Mädchen und Jungen in zwei Boxplots nebeneinander dar. Vergleichen Sie die Boxplots in zwei bis drei Sätzen. Nutzen Sie den Befehl
- ```
. graph box pvlread, over(st04q01).
```
- (e) Die Forschungshypothese lautet, dass Mädchen besser lesen als Jungen. Legen Sie das Signifikanzniveau auf 5% fest. Formulieren Sie diese Hypothese als statistische Hypothese und nutzen Sie den STATA-Befehl `. ttest pvlread, by(st04q01)` um den doppelten t-Test durchzuführen. Kann die Forschungshypothese nachgewiesen werden?
- (f) Wenn man davon ausgehen kann, dass die Varianzen in den beiden Gruppen gleich sind, dann verwendet man den doppelten t-Test. Wenn man nicht von der Varianzgleichheit ausgehen kann, dann rechnet man den Welch-Test. Rechnen Sie den Welch-Test und vergleichen Sie das Ergebnis mit dem des doppelten t-Test. (`. ttest pvlread, by(st04q01) welch`)

6. Die Füllmenge der Flaschen: Der einfach t-Test (14 Punkte)

Aufgabe
abgeben!

In einer Getränkefirma werden Flaschen von einer Maschine mit Mineralwasser gefüllt. Es ist bekannt, dass die Füllmenge normalverteilt ist. Eine Mitarbeiterin vermutet, dass die Maschine zu wenig Mineralwasser in die Flaschen füllt. Sie nimmt eine Stichprobe von 20 Flaschen und bestimmt die genaue Füllmenge.

- (a) Lesen Sie den Datensatz `flaschen.dta` ein.

- (b) Nutzen Sie den Befehl `sum, detail` und lesen Sie den Mittelwert, den Median, das 10%-Quantil ($q_{0.1}$) und das 90%-Quantil ($q_{0.9}$) ab. Interpretieren Sie das Intervall ($q_{0.1}, q_{0.9}$) inhaltlich. (5 Punkte)
- (c) Schlagen Sie ihre Formelsammlung auf S. 48 auf. Dort finden Sie den einfachen t-Test. Legen Sie das Signifikanzniveau auf 1% fest und formulieren Sie die Hypothesen der Mitarbeiterin. (2 Punkte)
- (d) Stellen Sie die Füllmenge mit Hilfe eines Boxplots dar und beschreiben Sie den Boxplot in zwei bis drei Sätzen. (. `graph box Volume`) (2 Punkte)
- (e) Nutzen Sie den Befehl `ttest Volume == 500` um den einfachen T-test zu rechnen. Interpretieren Sie das Ergebnis inhaltlich. Muss die Maschine neu eingestellt werden?
- (f) Wie würde die Testentscheidung ausfallen, wenn das Signifikanzniveau bei 10% liegen würde? (1 Punkte)
- (g) Formulieren Sie den Fehler 2. Art inhaltlich. Was bedeutet dieser Fehler für die Getränke-Firma. (2 Punkte)
- (h) Was könnte die Mitarbeiterin tun, um mehr Sicherheit im Bezug auf ihre Entscheidung zu haben? Begründen Sie ihre Antwort. (2 Punkte)
- (i) Welche Abweichung von der gewünschten Füllmenge würde ihrer Meinung nach gerade noch zulässig sein?