Florian Wickelmaier
Forschungsmethoden und Mathematische Psychologie
Fachbereich Psychologie
Universität Tübingen

Vorlesung Statistik II
SS 2026

Übungsaufgaben

Aufgabe 28. Bei einer Befragung soll die Einstellung zu Auslandseinsätzen der Bundeswehr ermittelt werden. Dazu sollen 12 Personen zufällig einem von zwei Interviewern zugewiesen werden, so dass Interviewer A genau 7 Personen und Interviewer B genau 5 Personen befragt. Jeder Person werden 10 Fragen gestellt und die Anzahl der in Schlüsselrichtung (also z. B. Zustimmung) beantworteten Fragen wird bestimmt. Interviewer A erhält die Ergebnisse

$$x_1 = 8, x_2 = 5, x_3 = 7, x_4 = 3, x_5 = 9, x_6 = 5, x_7 = 6,$$

während bei Interviewer B die Befragung die Ergebnisse

$$y_1 = 1, y_2 = 7, y_3 = 2, y_4 = 6, y_5 = 4$$

liefert. Die Hypothese, dass sich die Ergebnisse der beiden Interviewer voneinander unterscheiden, soll mit einem Permutationstest überprüft werden. Als Teststatistik wird die Differenz $d = \bar{x} - \bar{y}$ der Mittelwerte verwendet.

a) Erläutern Sie die Nullhypothese dieses Tests und die grundlegende Vorgehensweise zur Bestimmung der Verteilung der Teststatistik.

b) Berechnen Sie die Teststatistik für die vorliegenden Daten, sowie deren kleinsten und größten möglichen Wert.

c) Wie viele Fälle sind für die Bestimmung des p-Wertes insgesamt zu betrachten?

d) Welche Situation müsste ausgehend von der Gesamtzahl der Fälle vorliegen, damit die Nullhypothese bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.10$ und einer zweiseitigen Fragestellung abgelehnt wird?

Aufgabe 29. Die folgenden Daten entstammen einer Untersuchung zur Wirkung der Verabreichung von Kalzium auf den Blutdruck (Experimentalgruppe, $n = 10$), im Vergleich zur Gabe eines Placebos (Kontrollgruppe, $m = 11$). Aus der Messung des Blutdrucks in mmHg vor und nach der Verabreichung des Verums/Placebos wird für jede Versuchsperson die Differenz berechnet.

Experimentalgruppe			Kontrollgruppe		
Beginn	Ende	Differenz	Beginn	Ende	Differenz
107	100	7	123	124	-1
110	114	-4	109	97	12
123	105	18	112	113	-1
129	112	17	102	105	-3
112	115	-3	98	95	3
111	116	-5	114	119	-5
107	106	1	119	114	5
112	102	10	114	112	2
136	125	11	110	121	-11
102	104	-2	117	118	-1
			130	133	-3

Überprüfen Sie unter Verwendung von R die Hypothese der Gleichheit der Varianzen in den beiden Stichproben (Irrtumswahrscheinlichkeit $\alpha = 0.05$).

a) Führen Sie zunächst einen F-Test zum Vergleich der Varianzen zweier Normalverteilungen durch. Welche Schlussfolgerungen sind aus den Ergebnissen zu ziehen?

b) Führen Sie einen Randomisierungstest zum Vergleich der Varianzen durch. Benutzen Sie hierzu folgendes R-Skript:

```
> dat <- data.frame(
+     y = c(7, -4, 18, 17, -3, -5, 1, 10, 11, -2,       # EG
+           -1, 12, -1, -3, 3, -5, 5, 2, -11, -1, -3),  # CG
+   grp = factor(rep(1:2, c(10, 11)), labels = c("EG", "CG"))
+ )
> f0 <- Reduce("/", tapply(dat$y, dat$grp, var))  # original ratio
> f  <- replicate(5000,                           # 5000 replicates for f0
+                 Reduce("/", tapply(sample(dat$y), dat$grp, var))
+ )
> p <- mean(f <= min(f0, 1/f0) | f >= max(f0, 1/f0))
```

Hinweis: `tapply()` berechnet die Varianzen pro Gruppe. `Reduce("/", ...)` bestimmt das Verhältnis der beiden Varianzen. Der Vektor `f` enthält dann 5000 unter Randomisierung der Gruppenzugehörigkeit berechnete Werte der Teststatistik.

Welche Schlussfolgerungen sind auf der Grundlage des hier berechneten p-Wertes zu ziehen?

c) Vergleichen Sie die Verteilung der Teststatistik des F-Tests mit der im Randomisierungstest erhaltenen Verteilung der Teststatistik (d. h. Verteilung der Werte des Vektors `f`).

   (a) Erstellen Sie ein Histogramm der Verteilung der Teststatistik im Randomisierungstest (auf der Ordinate Dichtewerte abtragen!) und zeichnen Sie in diese

Grafik die Dichtefunktion der F-Verteilung mit der entsprechenden Anzahl von Freiheitsgraden ein.

(b) Erstellen Sie einen Quantil-Quantil-Plot, in dem die Verteilung der Teststatistik im Randomisierungstest (auf der Ordinate) gegen die F-Verteilung mit der entsprechenden Anzahl von Freiheitsgraden (auf der Abszisse, erzeugt durch Ziehung von 5000 Werten aus der entsprechenden F-Verteilung) abgetragen wird.

Aufgabe 30. Vor zehn Jahren wurde in einer Untersuchung des Rauchverhaltens von amerikanischen Männern im Alter von 40-50 Jahren der tägliche Zigarettenkonsum über die durchschnittliche Anzahl Packungen (aufgerundet zur nächsten ganzen Zahl) erhoben. Dabei wurden folgende Wahrscheinlichkeiten ermittelt:

keine	1	2	3	≥ 4
0.43	0.17	0.24	0.10	0.06

a) Bestimmen Sie per Simulation den Stichprobenumfang für eine Studie, in der ein Rückgang des Zigarettenkonsums mit ca. 80% Power ($\alpha = 0.01$) gezeigt werden kann. Gehen Sie dabei von folgender Verteilung aus:

keine	1	2	3	≥ 4
0.45	0.20	0.24	0.07	0.04

b) Es wurde eine neuen Stichprobe von 863 amerikanischen Männern derselben Altersgruppe erhoben. Überprüfen Sie anhand der nachfolgenden Werte, ob der Zigarettenkonsum tatsächlich zurückgegangen ist ($\alpha = 0.01$):

keine	1	2	3	≥ 4
406	164	189	78	26

Führen Sie alle Berechnungen von Hand durch und überprüfen Sie diese dann durch Verwendung der entsprechenden Funktionen in R.