Universität Bremen

FB3, A. Mändle

WS 2019/20, 13.12.2019

Einführung in R
10. Aufgabenblatt

Präsenzaufgabe 1

In R ist folgender Datensatz `sleep` enthalten, welcher die den Einfluss eines Schlafmittels auf die Schlafdauer von 10 Probanden darstellt. Die Schlafdauer ist nicht absolut, sondern als Abweichung von der Schlafdauer einer Kontrollgruppe angegeben.

Patient	1	2	3	4	5	6	7	8	9	10
vorher	0.7	-1.6	-0.2	-1.2	-0.1	3.4	3.7	0.8	0.0	2.0
nachher	1.9	0.8	1.1	0.1	-0.1	4.4	5.5	1.6	4.6	3.4

a. Können Sie dieschlafverlängernde Wirkung des Medikaments zum Signifikanzniveau $\alpha = 0.05$ betätigen? Nutzen Sie den t-Test für paarweise Daten. Vergleichen Sie die Therapien zusätzlich mit einem Boxplot und beschrften Sie diesen geeignet.

b. Berechnen Sie die Power des Tests, um die beobachtete Differenz der Mittelwerte als Signifikant zu erkennen.

c. Würde bereits eine geringere Fallzahl genügen, um den beobachteten Effekt (d.h. die bei den Stichproben beobachtete Differenz) mit einer Power von mindestens 0.95 abzusichern?

Hinweis: Verwenden Sie die empirische Standardabweichung der Differenz in b. und c.

Präsenzaufgabe 2

Es soll gezeigt werden, dass der mittlere diastolische Blutdruck bei Patienten mit Bluthochdruck mittels einer neue Therapie gegenüber der Standardtherapie um mindestens 10 mmHg sinkt. Berechnen Sie die benötigte Fallzahl zur Absicherung dieser Differenz für eine Studie mit

a. zwei unabhängigen gleich großen Gruppen (*Standard* und *neue Therapie*).

b. nur einer Gruppe, bei der zuerst die Standardtherapie angewendet wird und dann nach einer angemessenen Zeitdauer zur Vermeidung von Übertragungseffekten die neue Therapie.

Setzen Sie jeweils $\alpha = 0.05$, $\beta = 0.2$, $\sigma = 20$ mmHg.

Präsenzaufgabe 3

Generieren Sie je 10, 12, 15 und 20 Zufallszahlen X_i, Y_i, Z_i und W_i:

$X_1, .., X_{10} \overset{i.i.id.}{\sim} \mathcal{N}(\mu_1, 1)$ mit $\mu_1 = 1$
$Y_1, .., Y_{12} \overset{i.i.id.}{\sim} \mathcal{N}(\mu_2, 1)$ mit $\mu_2 = 1.4$
$Z_1, .., Z_{15} \overset{i.i.id.}{\sim} \mathcal{N}(\mu_3, 1)$ mit $\mu_3 = 1.8$
$W_1, .., W_{20} \overset{i.i.id.}{\sim} \mathcal{N}(\mu_4, 1)$ mit $\mu_4 = 2.4$

Führen Sie den paarweisen t-Test mit einem (globalen) Signifikanzniveau von $\alpha = 0.05$ aus, um die Alternativhypothese $H_1 : \exists i, j \in \{1, ..4\}, \mu_i \neq \mu_j, i \neq j$ zu testen!

Präsenzaufgabe 4

Generieren Sie 2 Stichproben: $(x_1, .., x_9)$ und $(y_1, .., y_8)$ aus:

$X_i \overset{i.i.id.}{\sim} \mathcal{N}(1, 1), \quad Y_i \overset{i.i.id.}{\sim} \mathcal{N}(1.2, 1.4^2)$

a. Testen Sie die Alternativhypothese, dass die Varianz der ersten Grundgesamtheit kleiner ist als die Varianz der zweiten Grundgesamtheit.

b. Testen Sie die Nullhypothese, dass die Erwartungswerte beider Grundgesamtheiten gleich sind.

Verwenden Sie ein Signifikanzniveau von $\alpha = 0.1$ für Ihren Test.

Hausaufgabe 1 (6 Punkte)

In einer Studie soll gezeigt werden, dass der Blutzuckerwert unter einer bestimmten Therapie um $10 mg/dl$ gesenkt werden kann. Aus Voruntersuchungen ist $\sigma = 20 mg/dl$ bekannt. Der Test soll mit einem Signifikanzniveau $\alpha = 0.05$ und einer Power von 0.90 durchgeführt werden. Berechnen Sie die benötigte Fallzahl zur Absicherung dieses Effekts.

Eine Studie mit 40 Patienten wird durchgeführt. Wie groß soll der tatsächliche Effekt (Blutzuckerwertsenkung) sein, damit er mit einem Signifikanzniveau von $\alpha = 0.05$ und einer Power von 0.90 aufgedeckt wird. Nehmen Sie an, dass $\sigma = 10 mg/dl$ ist.
Hinweis: Verwenden Sie die Funktion power.t.test().

Hausaufgabe 2 (8 Punkte)

Laden Sie den Datensatz *normtemp* aus dem Paket *UsingR*. Der Datensatz enthält die Variablen Körpertemperatur, Geschlecht (1 = Mann, 2 = Frau) und Herzschlagfrequenz (hr).

a. Testen Sie anhand dieser Daten (mit dem Signifikanzniveau $\alpha = 0.05$) die Nullhypothese, dass (unter Annahme normalverteilter Daten) die Körpertemperatur unabhängig vom Geschlecht ist. Vergleichen Sie die Methoden zusätzlich mit einem Boxplot.

b. Bestimmen Sie die Power des Tests, um die beobachtete Differenz der Mittelwerte als signifikant erkennen zu können.

c. Berechnen Sie die benötigte Fallzahl zur Absicherung dieser (bei den Stichproben beobachten) Differenz, wenn eine Power von mindestens 0.80 erreicht werden soll.

Hinweis: verwenden Sie power.t.test() Funktion.

Hausaufgabe 3 (9 Punkte)

a. Laden Sie den Datensatz *iris* aus dem Paket *datasets*. Anhand dieses Datensatzes wollen wir die Nullhypothese testen, dass die Mittelwerte von *Sepal.Length* bei verschiedenen *Species* gleich sind, d.h. ($\mu_{versicolor} = \mu_{setosa} = \mu_{virginica}$). Nehmen Sie an, dass *Sepal Length* normalverteilt ist. Benutzen Sie den paarweisen t-Test mit Bonferroni und Holm-Korrekturen mit $\alpha = 0.05$ für Ihren Test.

b. Führen Sie die gleiche Prozedur für weitere metrisch-skalierte Merkmale (*Sepal.Width, Petal.Length, Petal.Width*) aus.

Zusatzaufgabe (12 Punkte) ***Eine Vollmondnacht...*** *(basierend auf: Tino Werner, Tutorial zu Einführung in die Angewandte Statistik, WiSe18/19, Uni Oldenburg)*

Sie wohnen in einem Landsitz hinter einem Wald und wissen, dass sich ein Sendbote auf dem Weg zu Ihnen befindet, welcher den Wald in dieser Vollmondnacht durchqueren wird. Es gibt drei Wege durch den Wald, die er nehmen kann: Der erste Weg ist sehr unwegsam, sodass Sie davon ausgehen können, dass man nachts mit 25 Prozent Wahrscheinlichkeit vom Weg abkommt. Ist dies der Fall, so gelangt man aber mit 20 Prozent Wahrscheinlichkeit auf den zweiten Weg, welcher vom Eingang bis zum Waldrand problemlos passierbar ist. Allerdings befindet man sich dort am Flusstal, wo sich mit 10 Prozent Wahrscheinlichkeit so dichter Nebel bildet, dass ein Weiterkommen unmöglich ist. Der Fluss trennt ein Gebiet des Waldes ab, durch welches der dritte Weg verläuft. Auch dieser ist so einfach, dass man sich nicht verirren kann, jedoch lebt dort ein Einsiedler, dem man nachsagt, ein Werwolf zu sein. Sofern keine Behinderungen auftreten, ist noch am Morgen mit einer Ankunft zu rechnen.

Sie dürfen davon ausgehen, dass Ihr Sendbote bei dichtem Nebel rastet und erst mittags seinen Weg fortsetzen kann. Ebenso findet er, sofern er sich verlaufen hat und nicht den zweiten Weg gefunden hat, den Ausgang des Waldes erst am nächsten Mittag. Sollte der Einsiedler ein Werwolf sein, so erreicht der Sendbote beim Benutzen des dritten Weges mit 90-prozentiger Wahrscheinlichkeit sein Ziel nicht. Der Werwolf, sofern er einer ist, bleibt immer in seinem Waldgebiet. Ist der Einsiedler ein Mensch, so schläft er nachts und hindert Ihren Sendboten nicht am Voranschreiten.

Ihr Sendbote hat einen Dodekaeder dabei, der darüber entscheidet, welchen der drei Wege er einschlägt. Er wirft diesen zwei Mal und berechnet das Produkt der gesehenen Augenzahlen. Ist dieses ungerade, nimmt er den dritten Weg. Ist dieses gerade und durch 7 teilbar, so entscheidet er sich für den den zweiten Weg, ansonsten für den ersten Weg.

Simulieren Sie 1000 Mal die Ereignisse dieser Vollmondnacht unter der Bedingung, dass der Einsiedler ein Werwolf ist. Lassen Sie am Ende ausgeben, wie oft der Sendbote welchen Weg nahm, wie oft er sich verlaufen hat bzw. im Nebel gerastet hat bzw. dem Werwolf begegnete und wie oft er insgesamt morgens oder später am Ziel ankam. Veranschaulichen Sie die Ergebnisse mit Kontingenztafeln und durch geeignete Plots. Ist die Wahrscheinlichkeit noch am Morgen anzukommen größer als 50%? Wenn möglich, widerlegen oder bestätigen Sie die Ihre Antwort mithilfe eines Binomialtests.

Abgabe der Lösungen: bis ***Montag 06.01.2020,*** *an maendle@uni-bremen.de*