Übungen zur Vorlesung „Stochastik für die Informatik"

Abgabe der Lösungen zu den S-Aufgaben: Freitag, 25. Januar 2013

37. S An 11 Individuen wurde die Messung einer bestimmen Größe vor und nach einer Behandlung durchgeführt. Das Ergebnis der Messung (die *Daten*) waren 11 Zahlenpaare $(x_1, y_1), \dots, (x_{11}, y_{11})$.

Unser Modell ist, dass sich die Messung vor der Behandlung bei Individuum i als Summe aus einem systematischen und einem zufälligen Beitrag ergibt: $X_i = \mu_i + \sigma Z_{0i}$, i=1,..., 11. Analog ergebe sich die Messung nach der Behandlung als $Y_i = \mu_i + \delta + \sigma Z_{1i}$, dabei ist δ der *Effekt* der Behandlung. Die Zufallsvariablen Z_{0i}, Z_{1j} werden dabei als unabghängig und standard-normalverteilt angenommen. Damit sind die Zufallvariablen $D_i := Y_i - X_i$ unabhängige Kopien einer Zufallvariablen D.

a) Geben Sie Schätzwerte (i) für δ und (ii) für σ_D an, ausgedrückt durch die Daten.
b) Die geschätzte Standardabweichung s_D ist 1.853, der Mittelwert $\bar{x}$ der x_i ist 182.33, der Mittelwert $\bar{y}$ der y_i ist 183.69.
(i) Mit welcher Zahl c muss man $\bar{y} - \bar{x}$ multiplizieren, damit man unter der Hypothese "$\delta = 0$" den Wert $c \cdot (\bar{y} - \bar{x})$ als Realisierung einer t-verteilten Zufallsvariablen ansehen kann?
(ii) Zu welchem p-Wert kann man die Hypothese "$\delta = 0$" ablehnen? Verwenden Sie dazu die angegebene Wertetabelle von $\mathrm{pt}(q, \mathrm{df}) := \mathbf{P}_{\mathrm{df}}(T \le q)$, der Verteilungsfunktion einer t-Verteilung mit df Freiheitsgraden.

df\q	2	2.2	2.4	2.6
10	0.963	0.973	0.981	0.987
11	0.965	0.977	0.982	0.988
12	0.966	0.976	0.983	0.988

38 S. Wir betrachten zwei Stichproben A: 4.4, 3.5, 6.3, 11.5, 7.5, 5.2, 6.7, 10.3,
B: 4.0, 3.1, 5.0.
Es geht um die Frage, ob die Verteilungen, aus denen A und B gezogen wurden, dieselben sind, oder ob sie gegeneinander verschoben sind.

a) Was sind die Ränge von B in der Gesamtheit von A und B? (Hinweis: Gefragt ist hier nach drei Zahlen zwischen 1 und 11.)

b) Wie groß ist die Rangsumme von B?

c) Wieviele Möglichkeiten gibt es, drei Werte aus der Gesamtheit von A und B auszuwählen, deren Rangsumme nicht größer als die unserer Stichprobe B ist?
(Hinweis: Eine dieser Möglichkeiten ist xxxoooooooo. Listen Sie auch die anderen auf.)

d) Wie wahrscheinlich ist es, dass die Rangsumme von drei zufällig aus der Gesamtheit von A und B gewählten Werte mindestens so weit an den beiden Rändern ("oben" und "unten") liegt wie die unserer Stichprobe B ?

39. In einer Stichprobe aus 500 österreichischen Männern betrug das durchschnittliche Körpergewicht 81,0 kg, in einer Stichprobe aus 1000 deutschen Männern betrug es 82,4 kg. Angenommen die Standardabweichungen sind $\sigma_A = \sigma_G = 12$ kg. Zu welchem p-Wert kann man die Hypothese "deutsche und österreichische Männer sind im Schnitt gleich schwer" ablehnen? Welche Teststatistik schlagen Sie dafür vor? Wieso darf man mit approximativer Normalverteiltheit dieser Teststatistik rechnen, auch wenn es der Fall sein sollte, dass das Gewicht eines rein zufällig gewählten österreichischen Mannes nicht einmal annähernd normalverteilt ist?

40. In zwei Regionen, A und B, beide gespalten in Befürworter und Gegener einer Gesetzesinitiative, ergaben sich bei einer Meinungsumfrage mit Stichprobengrößen $n_A = 500$ und $n_B = 1000$ die geschätzten Anteile $\hat{p}_A = 0,495$ bzw $\hat{p}_B = 0,492$ für die Befürworter.
a) Finden Sie approximative 95%-Konfidenzintervalle für (i) p_A und (ii) $p_A - p_B$.
b) Mit welchem p-Wert lässt sich die Hypothese (i) $p_A \ge 0,5$ bzw. (ii) $p_A = p_B$ ablehnen?
Hinweis: Verwenden Sie die Normalapproximation der Binomialverteilung.

FROHE WEIHNACHTEN UND EIN GLÜCKLICHES JAHR 2013 !