

STATISTIK 2 (107.325) WS 2011
COMPUTERSTATISTIK (107.258) WS 2011

Übung 4

16. November 2011
Dutter

13. In der Inskriptionsstatistik der TU Wien werden Hörerzahlen der Studienrichtung „Technische Mathematik“ (SS2011) mit 559 männliche und 252 weiblich Inskribierte ausgewiesen.

Frage: Sind diese Zahlen voneinander signifikant verschieden (oder rein zufällig verschieden).

Lösungsansatz: Annahme: Ein/e Studierender/Studierende ist bezüglich Geschlecht als zufällig zuzuordnen. D.h. die Zufallsvariable X liefert 0 oder 1, entsprechend Geschlecht männlich oder weiblich. Angenommen, die Wahrscheinlichkeit von weiblich, d.h. $P(X = 1)$ sei gleich p . Unter der Annahme der Unabhängigkeit der Studierenden ist $\sum_{i=1}^n X_i$ binomialverteilt $Bi(n, p)$, wobei n die Gesamtanzahl der betrachteten Studierenden bezeichnet. Die Binomialverteilung kann für große n und p nicht allzu verschieden von .5 mit der Normalverteilung mit Mittel np und Varianz $p(1 - p)n$ gut approximiert werden. (Damit kann der übliche t-Test `t.test` in \mathbb{R} verwendet werden.)

Hier ist der Unterschied wohl offensichtlich signifikant. Wie ist es in der Studienrichtung Architektur, wo das Verhältnis 1618 zu 1635 heisst?

Bitte, rechnen Sie konkrete Signifikanztests für diese Fälle.

14. Nehmen wir an, dass in einer Übungsgruppe nur ein Student und mehrere Studentinnen sitzen. Wie gross muss die Anzahl der weiblichen Studentinnen sein, damit statistisch ein signifikanter Unterschied in der Anzahl erscheint, genauer, dass die Anzahl weiblicher Teilnehmer signifikant größer als die der männlichen ist? (Wie so häufig, nehmen wir $\alpha = .05$.)
15. Analysieren Sie die Daten “Werner Chemistry Data“ mit dem Computer Program System \mathbb{R} , wobei gleich zu Beginn die beiden Ausreisser in der Variablen „CHOLSTRL“ ausgeschlossen werden sollten (d.h. man nehme nur jene Zeilen der Matrix, bei denen $CHOLSTRL > 150$ und < 400 ist).

Betrachten Sie 4 Altersgruppen der Patientendaten “Werner Blood Chemistry Data“: (18,25], (25,32], (32,42], (42,55]. (Hinweis: Die Funktion ‘cut’ ist recht nützlich zum Zerschneiden des Vektors ‘AGE’. Die Anzahl pro Gruppe sollte dann sein: 50, 48, 42, 46.)

Testen Sie auf Einfluss des Alters (der Altersgruppen) auf Cholesterinwerte.

Erklären Sie die Voraussetzungen zu diesem Test. Kommentieren Sie das Resultat.

16. Verwenden Sie von den Altersgruppen der Patientendaten nur die ersten 3 bzw. die mittleren 2. Wie ändern sich die Signifikanzen? Diskutieren Sie Ihr Resultat.

Bitte, stellen Sie die Ausarbeitung in Form eines pdf-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Kurz-Listing des Programmkodes (Funktion) in die TUWEL-Seite

<https://tuwel.tuwien.ac.at/course/view.php?idnumber=107258-2011W>
bis zum 21. November 2011.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_4.pdf

wobei 'name' für den Familiennamen steht.

Empfehlenswert ist es, den Bericht mit „Sweave“ zu erstellen. (Vorlage siehe <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>.)

Kreuzen Sie außerdem bitte im Internet jede Nummer des Übungsbeispiels an, das Sie dann in der Übungsstunde an der Tafel (mit Beamer-Unterstützung) vorrechnen wollen und können. Der Termin ist üblicherweise 2 Stunden vor der Übung, d.h. Mi., 12:00. Siehe <http://www.statistik.tuwien.ac.at/cgi-bin/uebbsp/bspstart.cgi> .