

**STATISTIK 2 (107.325) WS 2010**  
**COMPUTERSTATISTIK (107.258) WS 2010**

**Übung 4**

**30. Oktober 2010**  
**Dutter**

Analysieren Sie die Daten “Werner Chemistry Data“ mit dem Computer Program System  $\mathbb{R}$ , wobei gleich zu Beginn die beiden Ausreisser in der Variablen „CHOLSTR“ ausgeschlossen werden sollten (d.h. man nehme nur jene Zeilen der Matrix, bei denen  $\text{CHOLSTR} > 150$  und  $< 400$  ist).

13. Betrachten Sie die 2 Gruppen der Daten ‘eher jüngere’ und ‘eher ältere’ Patientinnen mit (willkürlich) ‘AGE  $\leq 35$ ’ und ‘AGE  $> 35$ ’. Gibt es einen signifikanten Unterschied der durchschnittlichen Cholesterinwerte zwischen den beiden Gruppen?
14. Betrachten Sie 4 Altersgruppen der Patientendaten “Werner Blood Chemistry Data“: (18,25], (25,32], (32,42], (42,55]. Die Anzahl pro Gruppe sollte sein: 50, 48, 42, 46.

Hinweis: Nützlich ist die Funktion `cut`.

Testen Sie, ob die mittleren Cholesterinwerte von Patientinnen dieser Altersgruppen als gleich angesehen werden können. D.h. machen Sie einen t-Test zwischen allen Gruppen mit allen Gruppen (paarweise) und kommentieren Sie das Resultat.

15. In der Inskriptionsstatistik der TU Wien werden Hörerzahlen der Studienrichtung „Technische Mathematik“ (SS2010) mit 811 männliche und 367 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 (auch in der Informatik mit 4727 männlichen und 855 weiblichen Inskribierten) signifikant. Wie ist es in der Studienrichtung Architektur, wo das Verhältnis 2335 zu 2205 heisst?

Bitte, rechnen Sie konkrete Signifikanztests für diese Fälle.

16. 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  $\alpha = .05$ .)

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?id=2604>

bis incl. 8. November 2010.

Bitte den Namen des pdf-File folgendermaßen:

**name\_exer\_4.pdf**

wobei 'name' für den Familiennamen steht. Bitte im pdf-File Name, Datum und Seitennumerierung nicht vergessen!

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