

Prof. Antony Unwin, Alexander Pilhöfer
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
<http://stats.math.uni-augsburg.de/>

Statistik I

Übungsblatt 7

Abgabe: Dienstag 05. Juni 2012, bis spätestens 12.00 Uhr; Briefkasten: Statistik I oder per email an die Übungsleiter

Die Aufgaben können auch in 2er-Gruppen bearbeitet und abgegeben werden!

- In welchen Fällen kommt ein nichtparametrischer Test zum Einsatz? (1P)
 - Warum wird anstelle des Likelihood in der Regel das Loglikelihood eines Schätzers Θ maximiert?(1P)
 - Welche Annahme ist hierbei zentral? (1P)
 - Mit einem Wilcoxon-Rangsummentest wurde festgestellt, dass die Schüler der Klasse 11c signifikant schneller laufen, als die der 11d. (p-Wert 0.03). Der schnellste Schüler wurde falsch gemessen und darf noch einmal laufen, wobei er sich um 0.2 Sekunden verbessert. Was ist nun der p-Wert des Tests? (1P)

2. Leistungsindex (5P)

Für die Entscheidung, welche Spieler zur EM fahren dürfen zieht Trainer Löw auch deren Leistungsindizes hinzu, und wie sie sich entwickelt haben. Von 20 Spielern (einige stehen ohnehin schon fest) wurden je 10 nach unterschiedlichen Methoden trainiert und die Änderungen im Leistungsindex beobachtet.

n	1	2	3	4	5	6	7	8	9	10
Methode 1	-0.4	+0.6	+0.4	-0.5	-3.5	-2.1	-2.4	+0.9	-0.9	+0.1
Methode 2	+3.6	+1.2	+0.3	-0.8	+4.2	-1.3	-2.6	+3.2	+0.9	-0.3

- Testen Sie mit einem Vorzeichentest, ob die einzelnen Methoden einen positiven Einfluss auf den Leistungsindex haben! Was testen Sie dabei genau?
- Wählen Sie eine geeignete grafische Methode um evtl. vorhandene Unterschiede in den Methoden zu entdecken, und stellen Sie die Daten entsprechend dar!
- Testen Sie mit einem Rangsummentest nach Wilcoxon, ob die Methoden unterschiedlich wirken! Stimmt das Ergebnis des Tests mit dem Eindruck der grafischen Darstellung aus (b) überein?

3. Wine Quality (10P)

Laden Sie den Datensatz *winequality* von der Webseite der Vorlesung herunter. Dieser Datensatz soll zur Vorbereitung auf das Projekt intensiv untersucht werden. Verwenden Sie sowohl Mondrian als auch R um die nachfolgenden Fragen zu bearbeiten:

- Wie ist der Datensatz aufgebaut? Nennen Sie die wichtigsten Eckpunkte.

- Wie hängen die kontinuierlichen Variablen zusammen? Welche sind unabhängig, welche stark assoziiert? Verwenden Sie u.a. Grafiken und Korrelationen.
- Ändern sich Ihre Aussagen über die Zusammenhänge, wenn Sie rote und weiße Weine trennen?
- Sind rote und weiße Weine in den einzelnen Variablen signifikant unterschiedlich? Verwenden Sie geeignete parametrische sowie nicht-parametrische Tests in R!
- Welche Variablen haben Einfluss auf die Weinqualität?
- Erläutern Sie, ob man mit einem t-Test die Aussage, dass weiße Weine durchschnittlich besserer Qualität sind, als rote, überprüfen könnte.
- Untersuchen Sie unter Verwendung eines Parallele-Koordinaten-Plots (PCP) und weiterer interaktiver Grafiken, ob es in den Daten
 - auffällige Gruppen (Cluster)
 - Ausreißer
 zu finden gibt.

4. Wine Quality II (5P)

Der Assistent des Lehrstuhls meint, dass mehr Alkohol besser ist. Um seine These zu belegen, produziert er nachfolgende Grafik:

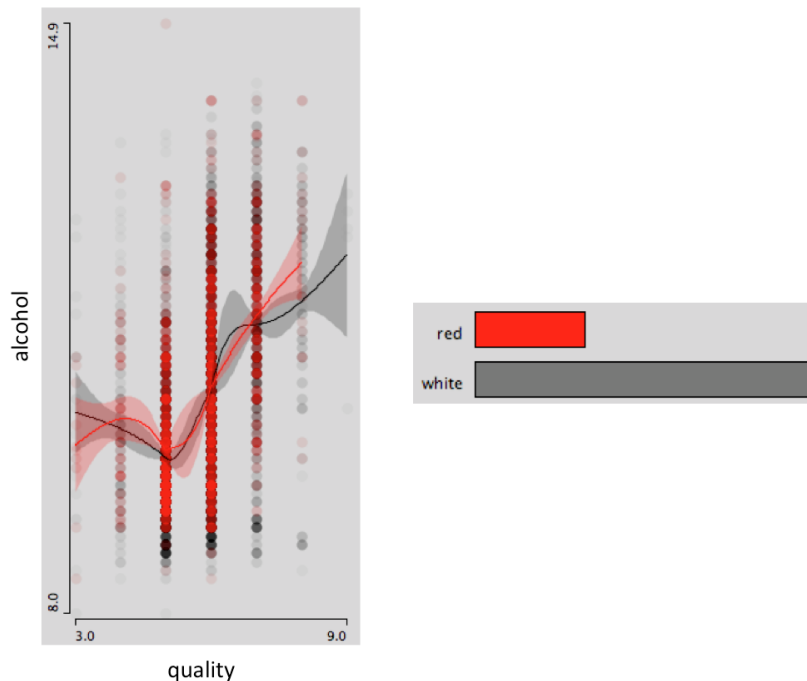


Abbildung 1:

- Bewerten Sie die Grafik im Hinblick auf Übersichtlichkeit, Suggestivität und Ästhetik.
- Kritisieren Sie seine Grafik anhand mehrerer wichtiger Punkte!
- Machen Sie es besser!
- Es wurden auch Glättungskurven eingezeichnet (das wird Stoff in Statistik III sein). Worum handelt es sich vermutlich bei den farblich hinterlegten Bereichen?