



Mustererkennung in der Bioinformatik

Dr. Hans A. Kestler, Dipl.-Inf. André Müller

WS 2003/2004

Übungsblatt 6

Ergänzung zur Vorlesung

Test auf Gleichheit der Varianzen

Zwei Gruppen sollen auf die Gleichheit der Varianz getestet werden. Hierzu wird das folgende Hypothesenpaar verwendet:

$$H_0 : \sigma_1 = \sigma_2$$

$$H_1 : \sigma_1 \neq \sigma_2$$

Die Gruppe 1 besitzt n_1 Elemente und die Gruppe 2 entsprechend n_2 Elemente. Die empirischen Varianzen der Gruppen seien s_1^2 und s_2^2 . Auf die Teststatistik

$$T = \frac{s_1^2}{s_2^2}$$

wird getestet. T wird nun als F-verteilt angenommen, mit Freiheitsgraden $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$. Für das Signifikanzniveau α wird H_0 genau dann verworfen, wenn $T < F_{1-\alpha/2,(\nu_1,\nu_2)}$ oder $T > F_{\alpha/2,(\nu_1,\nu_2)}$.

t-Test für ungleiche Varianzen

Bei ungleichen Varianzen ($\sigma_1 \neq \sigma_2$) berechnet sich die t-Statistik zu

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

mit

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

Freiheitsgraden. Weiteres Vorgehen wie beim t-Test mit gleichen Varianzen (siehe Vorlesung).

Gepaarter t-Test

Bei gepaarten Messwerten (also z.B. vorher/nacher Vergleiche) berechnet sich die t-Statistik zu

$$t = \frac{\bar{D} - \mu_d}{\frac{s_d}{\sqrt{n}}},$$

wobei μ_d die mittlere Differenz zwischen beiden Messwerten und s_d die Standardabweichung über alle Differenzen ist ($D = X_1 - X_2$). Die Freiheitsgrade berechnen sich dann zu $n - 1$ (für n Messwertpaare).

1. Aufgabe (16): Analyse von Genexpressiondaten

Finden sie die statistisch signifikanten Gene auf einem Signifikanzniveau von 5% bei dem Genexpressionsdatensatz `Verona_Diag.xls` auf der Vorlesungswebseite (Zeilen=Gene, Spalten=Konditionen, letzte Zeile = Klassenlabels).

Hinweis: Testen Sie zuerst auf Gleichheit der Varianzen und wählen dann den passenden Einzeltest für jedes Gen. Bonferroni Korrektur!