Univ.-Prof. DI Dr. Ernst Stadlober

1.) [T] Einfache Varianzanalyse mit r Gruppen. 8P

Sei $Y_{ij} = \mu_0 + \alpha_i + \epsilon_{ij},\ i = 1, \dots, r\,; j = 1, \dots, n_i;\ \epsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma),\ \sum_{i=1}^{r} n_i \alpha_i = 0.$

(a) Man zeige, dass die Linearkombinationen $\sum_{i=1}^{r} \mathbf{e_i} \alpha_i$ mit $\sum_{i=1}^{r} n_i \alpha_i = 0$ den linearen Unterraum $L_1 \ominus L_2$ aufspannen.

(b) Zeigen Sie, dass (siehe **Satz 4.2.1**)

$$E(MSA) = \sigma^2 + \frac{1}{r-1} \sum_{i=1}^{r} n_i \alpha_i^2 \quad \text{mit} \quad MSA = \frac{1}{r-1} \sum_{i=1}^{r} n_i \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2$$

$$E(MSR) = \sigma^2 \quad \text{mit} \quad MSR = \frac{1}{n-r} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_{i.}\right)^2 .$$

(c) **Kruskal–Wallis–Test.**
Seien $Y_{i1}, \dots, Y_{in_i} \overset{\text{iid}}{\sim} F_i$, F_i stetig, $i = 1, \dots, r$, und $R_{11}, R_{12}, \dots, R_{rn_r}$ die Ränge der kombinierten geordneten Stichprobe, $R_{i.} = \sum_{j=1}^{n_i} R_{ij}$.

i. Weisen Sie nach, dass unter $H_0 :\ F_i(z) = F(z),\ i = 1, \dots, r$ gilt:

$$E(R_{i.}) = n_i \frac{n+1}{2}, \quad Var(R_{i.}) = \frac{n_i(n+1)(n-n_i)}{12}, \quad E(H) = r - 1 .$$

ii. Zeigen Sie, dass für $r = 2$ gilt (mit W_N Wilcoxon–Statistik nach Formel (2.22)):

$$H = \frac{(W_N - E(W_N))^2}{Var(W_N)} \ \text{mit}\ E(W_N) = \frac{n(N+1)}{2},\ Var(W_N) = \frac{mn(N+1)}{12} .$$

2.) [T] Varianzstabilisierende Transformationen. 8P

Sei X eine Zufallsvariable mit $E(X) = \mu$ und $Var(X) = \sigma^2(\mu)$. Man suche eine *varianzstabilisierende Transformation* $Y = T(X)$ mit der Eigenschaft $Var(Y) \approx c =$ constant. Sei $T(x)$ eine zweimal differenzierbare Funktion.

(a) Zeigen Sie mit Hilfe der Taylor–Formel, dass

$$Var(Y) \approx E\left((T'(\mu)(X-\mu)\right)^2 = \left(T'(\mu)\right)^2 \sigma^2(\mu) .$$

(b) Wie lautet $T(X)$ für $X \sim Poisson(\lambda)$ und $X \sim Bin(n, p)$ bei kleinem p?

(c) Man zeige:
Für $X = S/n$ mit $S \sim Bin(n, p)$, ist die varianzstabilisierende Transformation gegeben durch

$$T(x) = 2\sqrt{n} \arcsin(\sqrt{x}) \quad \text{(Arcussinus-Wurzel–Transformation)} .$$

(d) Eine Zufallsvariable X heißt **log-normalverteilt**, wenn $\log(X) \sim N(m, s)$. Erwartungswert und Varianz von X sind gegeben durch

$$E(X) = \mu = e^{m+\frac{1}{2}s^2} \quad Var(X) = \sigma^2 = e^{2\,m+s^2}\left(e^{s^2} - 1\right) .$$

Man zeige, dass die varianzstabilisierende Transformation von x gegeben ist durch $T(x) = \log(x)$, die hier auch die normalisierende Transformation darstellt.

3.) Einfache Varianzanalyse; `aimu_85.dat` [`R 2.8`] 4P

Man untersuche das Merkmal `fev1` (i) in Abhängigkeit vom Faktor `al_kl`, (ii) in Abhängigkeit vom Faktor `gr_kl`, welche in **Aufgabe 1.1** definiert wurden. Analysieren Sie die Daten mit graphischen und varianzanalytischen Methoden.

(a) Explorative Analyse mit Boxplotserien und Fehlerbalken.

(b) Führen Sie eine einfache Varianzanalyse in `R` mit den Befehlen gemäß *Hand-Out Abschnitt 4.1-4.3* durch.

(c) Welche Parametrisierung liefert der Aufruf `summary(lm(fev1~al_kl))`? Rufen Sie auch die Prozedur `oneway.test()` auf, die keine Homogenität der Varianzen voraussetzt.

(d) Mit dem Befehl `TukeyHSD()` kann eine Post-hoc-Analyse durchgeführt werden. Durch `plot(TukeyHSD())` werden die Konfidenzintervalle geplottet. Man führe auch die paarweisen t–Tests `pairwise.t.test()` durch und vergleiche die Ergebnisse.

(e) Welches Ergebnis liefert der Kruskal–Wallis-Test `kruskal.test()`?

(f) Fassen sie die Ergebnisse in Form eines Reports zusammen.

4.) Einfache Varianzanalyse, Zement [`R 2.8`] 4P

Die Zugfestigkeit von Portland–Zement soll untersucht werden. Vier verschiedene Mischtechniken liefern ein zufrieden stellendes Ergebnis. Folgende Daten wurden für die Untersuchung gesammelt (aus D.C. Montgomery, S.117, Bsp. 3-1 und 3-3).

Mischtechnik	Zugfestigkeit (lb/in^2)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

(a) Geben Sie die Daten in `R` ein und speichern Sie diese als File `zement.dat` ab.

(b) Hat die Mischtechnik einen Einfluss auf die Zugfestigkeit des Portland–Zements?

(c) Führen Sie eine Post Hoc Analyse mit dem TukeyHSD–Test durch.

(d) Analysieren Sie die Residuen des Experiments.

(e) Interpretieren Sie die Ergebnisse auch mit Hilfe von Boxplots.

(f) Wie lauten die 95%–Konfidenzintervalle für die Mittelwerte jeder Mischtechnik?

(g) Wie lautet das 95%–Konfidenzintervall der Differenz der Mittelwerte der Mischtechniken 1 und 3?

(h) Führen Sie den Kruskal–Wallis–Test durch.

5.) Fallbeispiel Luftschadstoffdaten (3. Teil) `grazluft.dat`; [`R 2.8`]. 6P

(a) Führen Sie eine einfache Varianzanalyse für `pm10` mit dem Faktor `ort` durch. Hat der Faktor Messort einen Einfluss auf den PM10–Gehalt der Luft? An welchen Messorten ist der PM10–Gehalt auffällig hoch?

(b) Analysieren Sie die Residuen bzgl. `pm10`. Sind die Voraussetzungen für eine einfache ANOVA gegeben? Was liefern die nichtparametrischen Verfahren?

(c) Versuchen Sie gegebenenfalls eine Transformation für `pm10` zu finden, welche die Varianz stabilisiert. Was kann man über die Verteilung der Residuen für die transformierte Variable sagen?

(d) Welche Methoden sind für den Vergleich der `pm10`–Daten bzgl. des Faktors `periode` adäquat? Vergleichen Sie die transformierten `periode`–Daten in ähnlicher Weise.

(e) Verfassen Sie einen kurzen Bericht Ihrer Analysen.

Herunter laden der Daten über die HomePage des Instituts: `www.statistics.tugraz.at`

Speichern Sie die **gesamten Übungen in einem pdf-File** mit folgendem Namen ab:
`Angstat_Nachname1*` z.B. `Angstat_Schiefer3.pdf`

und übermitteln Sie **einen File pro Gruppe** mit *Subject:* **Angstat** an die e-mail-Adresse `statistik@tugraz.at`.

Transfer der Files bis spätestens: Fr. 11. 12. 2009, 18.00 Uhr

BESPRECHUNGSTERMIN: MO. 14. 12. 2009, 9.00–10.45, SR STATISTIK