Fachbereich Mathematik und Informatik
der Universität Marburg
Dr. Helga Lohöfer

Sommersemester 2012

Mathematik für Natur- und Sozialwissenschaftler

- Blatt 11 -
Abgabe: Mittwoch, den 004.07.2012, 12:10 Uhr vor der Vorlesung

Skript und Aufgabenblätter: Im Internet unter www.mathematik.uni-marburg.de/~lohoefer/
Lektürehinweis: Skript Kap 8.8 und 8.9.2
Wichtige Begriffe: Normalverteilung, t-Test, Einfache Varianzanalyse

1. Von zwei Weizenanbaugebieten A und B kennt man den Ernteertrag (in 100kg/ha) seit mehreren Jahren. Es ergaben sich zwei normalverteilte Messreihen ohne Ausreißer wie folgt:

Anbaugebiet	Anzahl Jahre	Mittelwert	Standardabweichung s
A	7	52,3	5,1
B	15	58,2	5,6

Sind die Unterschiede der Ertragfähigkeit als rein zufällig anzusehen? Oder sind die Anbaugebiete mit benennbarer Sicherheit unterschiedlich ertragreich? (2)

2. Beim Zentralabitur schrieben zwei Kurse im Großstadtgymnasium A und dem Kleinstadtgymnasium B dieselbe Klausur. Es ergaben sich bei den Noten (1 bis 15) zwei normalverteilte Messreihen ohne Ausreißer wie folgt:

Schule	Anzahl Teilnehmer	Mittelwert	Standardabweichung s
A	25	10,6	3,2
B	19	8,9	2,1

Lässt sich mit einer benennbaren Sicherheit behaupten, dass der Kurs A bessere Leistungen lieferte? (2)

3. Ein gewisser Werkstoff enthält unvermeidlich eine gewisse gesundheitsschädliche Substanz. Zu seiner Herstellung sind drei Verfahren bekannt. Untersucht werden soll an jeweils vier bis fünf Proben, ob die Schadstoffkonzentration (in mg pro 100 g) unabhängig vom benutzten Herstellungsverfahren ist.

A	61	58	60	60	
B	62	59	61	61	60
C	65	63	63	62	

a) Berechnen Sie die statistischen Daten n_i , $\overline{x}_i$ und s_i der drei Messreihen (3 Nachkommastellen). (6)

b) Man erkennt anhand der Tabelle schon bei genauerem Hinsehen, dass bei Verfahren A die kleinsten, bei Verfahren C die größten Konzentrationen gemessen wurden. Aber bleiben diese geringen Unterschiede nicht doch im Rahmen des rein Zufälligen? Überprüfen Sie diese Nullhypothese mit dem Verfahren der Einfachen Varianzanalyse, wenn möglich, mit Angabe der Sicherheit ($\overline{x}$, s_{mitt}^2 und $s_{\overline{x}}^2$ mit 3 Nachkommastellen). (6)

Bedingungen für die erfolgreiche Teilnahme:

- Regelmäßige Teilnahme an der Vorlesung (maximal 2 mal Fehlen).
- Auf mindestens 9 Übungsblätter jeweils mindestens 5 Punkte erhalten, insgesamt jedoch mindestens 75 Punkte. Es sind 11 Übungsblätter mit jeweils mindestens 15 Punkten geplant.
- Dazu das Bestehen einer Klausur.

Klausurtermine:
1. Klausur: Do, 19.07.12, 11:00 – 12:45, Hörsaalgebäude Chemie Lahnberge, Hans-Meerwein-Str., HS A + B
2. Klausur: Do, 11.10.12, 11:00 – 12:45, Hörsaalgebäude Chemie Lahnberge, Hans-Meerwein-Str., HS B

Abhängigkeit einer normalverteilten Variablen X von verschiedenen Nebenbedingungen

Gegeben mehrere Messreihen zu X, normalverteilt, ohne Ausreißer (s. Skript Kap. 8.3), aber unter verschiedenen Nebenbedingungen gemessen. Längen n_i , Mittelwerte $\overline{x}_i$, Standardabweichungen s_i $(i = 1,\dots,m)$. Die Standardabweichungen mögen sich nur im Rahmen des Zufälligen unterscheiden (s. Skript, F-Test (Kap.8.6) bzw. Bartlett-Test (Kap. 8.9.1))

1. Fall: m = 2 Messreihen: **<u>Der t-Test</u>**: Nullhypothese H_o = „Es ist kein Unterschied der Mittelwerte $\overline{x}_i$ feststellbar, d.h. die unterschiedlichen Nebenbediungen haben keinen Einfluss auf X.“

- berechne $s_d = \sqrt{\dfrac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2}}$
- berechne $T = \dfrac{|\overline{x}_1 - \overline{x}_2|}{s_d} \cdot \sqrt{\dfrac{n_1 \cdot n_2}{n_1 + n_2}}$ und $f = n_1 + n_2 - 2$
- schlage für dieses f in der **t-Tabelle** für 95% nach und vergleiche mit T:
  Falls T < tab-t(95%), wird die H_0 angenommen, ohne mögliche Sicherheitsangabe.
  Falls T ≥ tab-t(95%), so wird die H_0 mit 95% Sicherheit (5 % Irrtumswahrscheinlichkeit) abgelehnt, analog für höhere Prozentzahlen (99%, 99,9%). Mit entsprechender Sicherheit liegt ein Einfluss der Nebenbedingungen auf X vor.

2. Fall: m ≥ 3 Messreihen: **Die Einfache Varianzanalyse:** Nullhypothese H_o = „Es ist kein Unterschied der Mittelwerte $\overline{x}_i$ feststellbar, d.h. die unterschiedlichen Nebenbediungen haben keinen Einfluss auf X.“

<u>1. Schritt:</u> Berechne für jede einzelne Messreihe ihre statistischen Daten n_i, $\overline{x}_i$ und s_i (i = 1, … , m)

<u>2. Schritt:</u> Betrachte die Gesamtmenge aller Messdaten als eine einzige große Messreihe und berechne für sie die beiden statistischen Daten n und $\overline{x}$ mittels folgender Hilfsformeln:

$$n = \sum_{i=1}^{m} n_i \quad \text{und} \quad \overline{x} = \frac{\sum_{i=1}^{m} n_i \cdot \overline{x}_i}{n}$$

<u>3. Schritt:</u> Berechne die **mittlere Varianz innerhalb der Messreihen** (= ein gemittelter s_i -Wert):

$$s_{mitt}^2 = \frac{\sum_{i=1}^{m} n_i \cdot s_i^2}{n - m}$$

<u>4. Schritt:</u> Berechne die **mittlere Varianz zwischen den Messreihen** (= misst die Abweichung der Mittelwerte $\overline{x}_i$ voneinander):

$$s_{\overline{x}}^2 = \frac{\sum_{i=1}^{m} n_i \cdot \left(\overline{x}_i - \overline{x}\right)^2}{m - 1}$$

<u>5. Sc hritt:</u> Berechne die Prüfgröße

$F = \dfrac{s_{\overline{x}}^2}{s_{mitt}^2}$ sowie die „Freiheitsgrade“ $f_1 = m - 1$ sowie $f_2 = n - m$.

<u>6. Schritt</u>: Schlage für f_1 und f_2 in der F-Tabelle nach und vergleiche F mit dem Tabellenwert tab-F.

Auswertung:

- Ist F ≤ 1 oder F < tab-F(95%), so ist ein mehr als zufälliger Unterschied zwischen den $\overline{x}_i$ nicht feststellbar, die Nullhypothese ist anzunehmen (ohne dass man eine Sicherheit für ihre Richtigkeit angeben kann).
- Ist F ≥ tab-F(95%), so sind die $\overline{x}_i$ **wahrscheinlich** zum Teil verschieden (= nicht alle gleich), die Nullhypothese ist mit 95% Sicherheit, also mit 5% Irrtumswahrscheinlichkeit, abzulehnen.
- Ist F ≥ tab-F(99%), so sind die $\overline{x}_i$ **signifikant** verschieden (= nicht alle gleich), die Nullhypothese ist mit 99% Sicherheit, also mit 1% Irrtumswahrscheinlichkeit, abzulehnen.
- Ist F ≥ tab-F(99,9%), so sind die $\overline{x}_i$ **hochsignifikant** verschieden (= nicht alle gleich), die Nullhypothese ist mit 99,9% Sicherheit, also mit 0,1% Irrtumswahrscheinlichkeit, abzulehnen.

<u>Hinweis:</u> Ablehnung der Nullhypethese bedeutet hier nicht, dass sich mit der genannten Sicherheit alle einzelnen Mittelwerte $\overline{x}_i$ untereinander mehr als nur zufällig unterscheiden, sondern lediglich, dass wenigstens (!) einer sich von den übrigen derart unterscheidet.

<u>Hinweis zum Internetskript</u>: Die Standardabweichung wurde dort etwas anders definiert (Kap. 7.2, im Nenner n – 1 statt n), entsprechend sehen dort die Formeln z.T. leicht anders aus.