

# Wahrscheinlichkeitsrechnung und Statistik für Informatik

Grossmann/Neuwirth/Cenker

4. April 2006

Ausgearbeitet von corni, coke und the\_trinity

Aufgrund von den Prüfungen von:

<http://www.informatik-forum.at/showpost.php?p=321845&postcount=21>

<http://www.informatik-forum.at/showpost.php?p=322149&postcount=22>

## Beispiel 1 (20 Punkte)

In dem Hörsaal, in dem die Prüfung stattfindet, können maximal 200 Studierende die Prüfung ablegen. Die Prüfung ist ausgebucht, d.h., 200 Studierende haben sich zu dieser Prüfung angemeldet. Die Wahrscheinlichkeit, dass ein angemeldeter Studierender auch tatsächlich zur Prüfung erscheint, beträgt 0,99.

- a) **Wie groß ist die Wahrscheinlichkeit, dass alle, die sich angemeldet haben, auch zur Prüfung erscheinen? (4 Punkte)**

$$(200 \text{ über } 200) * 0,99^{200} * 0,01^0 = 0,99^{200} = 0,1339797$$

- b) **3 Studierende haben sich nicht zur Prüfung angemeldet, gehen aber trotzdem zum Prüfungstermin, weil sie hoffen, dass nicht alle Angemeldeten erscheinen und sie daher einen der freien Plätze zugewiesen erhalten (dies soll möglich sein). Wie groß ist die Wahrscheinlichkeit, dass alle 3 tatsächlich zur Prüfung antreten können? (6 Punkte)**

$$B(0): (200 \text{ über } 200) * 0,99^{200} * 0,01^0 = 0,13979$$

$$B(1): (200 \text{ über } 199) * 0,99^{199} * 0,01^1 = 0,27066$$

$$B(2): (200 \text{ über } 198) * 0,99^{198} * 0,01^2 = 0,27203$$

$$1 - (0,13979 + 0,27066 + 0,27203) = 0,31752$$

Also eine 31,75% Chance dass man antreten darf.

$$\text{Forumslösung: } (200 \text{ über } 197) * 0,99^{197} * 0,01^3 = 0,1814$$

Aber er hat bei mir hingeschrieben „B(0)!“ deshalb glaub ich eher das obere ist richtig

- c) **Die Wahrscheinlichkeit, die Prüfung bei zwei Wochen Vorbereitung zu bestehen, beträgt 0,9. Eine Gruppe von 10 Studierenden hat zusammen zwei Wochen gelernt. Wie groß ist die Wahrscheinlichkeit, dass höchstens einer der Studierenden die Prüfung nicht besteht? (6 Punkte)**

$$(10 \text{ über } 10) * 0,9^{10} * 0,1^0 = 0,3487$$

$$(10 \text{ über } 9) * 0,9^9 * 0,1^1 = 0,3874$$

$$0,3487 + 0,3874 = 0,7361$$

- d) Welche Wahrscheinlichkeitsverteilung haben sie in a) bis c) verwendet?  
Wieso ist das die passende Verteilung für diese Art von Problemen? (4 Punkte)**

Die Binomialverteilung, weil sie folgende 2 Eigenschaften hat, die bei diesem Beispiel relevant sind:

1. Das Experiment hat immer die gleiche Wahrscheinlichkeit (immer die gleiche Wahrscheinlichkeit dass ein Student zur Prüfung kommt) UND das der Ausgang des Experiment immer 2 Möglichkeiten hat: Erfolg oder Misserfolg (Student kommt zur Prüfung oder nicht.)
2. Das die Versuche voneinander unabhängig sind (Im Normalfall: Wenn Student A zur Prüfung kommt, hat das keinen Einfluss ob Student B zur Prüfung kommt.)

### **Beispiel 2 (20 Punkte)**

Folgender Test mit Hilfe einer Kennzahl zur Leistungsmessung wurde an einer Gruppe Sportler durchgeführt, um die mögliche Wirkung des leistungssteigernden Mittels WaPla zu überprüfen.

Zweistichproben t-test bei abhängigen Stichproben (Paarvergleichstest)		
	Vorher	Nachher
Mittelwerte	25,36	27,28
Stichprobenvarianzen	10,38	2,91
Varianz der Differenz der Mittelwerte	12,28	
Beobachtungen	?	?
Hypothetische Differenz der Mittelwerte	0	
Freiheitsgrade (df)	96	
t-Statistik	5,54	
P(T<=t) zweiseitig	2,69E-07	
Kritischer t-Wert bei zweiseitigen Test	2,62	

- a) Welchen Voraussetzungen mussten erfüllt sein, sodass dieser Test angewendet werden konnte? (möglichst kurz) (4 Punkte)**

Antwort die 2 Punkte brachte:

Die Zufallsvariablen müssen normalverteilt sein. Ausserdem macht der Test erst Sinn, wenn die Varianz unbekannt ist.

Es steht noch ein Fehlendzeichen da aber ich kann es nicht entziffern.

- b) Welche Nullhypothese und Alternative wurden angenommen? (mathematisch-statistisch exakte Formulierung) (4 Punkte)**

$H_0: \mu_D = 0$  bzw.  $H_0: (\mu_X - \mu_Y) = 0$

Als sinnvolle Nullhypothese ist anzunehmen, dass das Mittel keine Leistungssteigerung bewirkt.

$H_0: \mu_D \neq 0$  bzw.  $H_0: (\mu_X - \mu_Y) \neq 0$

Als sinnvolle Alternativhypothese, dass das Mittel eine Leistungssteigerung bewirkt

- c) **Wie viele Probanden wurden vor bzw. nach der Verabreichung des Mittels getestet? (2 Punkte)**

Beobachtungen = Freiheitsgrade - 2 -> Beobachtungen = 98, da bei jeder Testperson 2x mal der Test vollzogen wurde (vorher bzw. nachher) ->  $98/2 = 49$ .  
Es wurden jeweils 49 Probanden vor und nach der Einnahme getestet.

- d) **Welcher Standard-Alpha-Fehler wurde hier angenommen? (2 Punkte)**

Da der kritische Wert 2,62 ist und die Freiheitsgrade 96 kommt man zu der Formel:  
 $F_{96; 1-\alpha} = 2,62$  wenn man in der Tabelle der  $t_{n,y}$  der t-Verteilung in der Zeile 100 nach wo 2,62 steht -> 0,995 ->  $1 - 0,995 = 0,005$   
Das ganze noch mal multipliziert mit 2.  
Der Standard-Alpha-Fehler wird hier mit 0.01 angenommen.

- e) **Berechnen Sie das Konfidenz-Intervall für die Differenz der Mittelwerte! (4 Punkte)**

Keiner unserer Versuche hat auch nur einen Punkt gebracht.

- f) **Welches Ergebnis erzielte dieser Test? – Sowohl statistisch exakte als auch allgemein verständliche Formulierung! (4 Punkte)**

Da  $|T| > t\text{-kritisch} \rightarrow |5,54| > 2,62 \rightarrow$  Verwerfung der Nullhypothese  
Es wird die Alternativhypothese angenommen. Das Mittel hat eine Leistungssteigernde Wirkung.

### Beispiel 3 (20 Punkte)

	Quadratsummen	Freiheitsgrade	Mittl. Quadratsumme	F-Wert	p-Wert
Zwischen d. Gruppen	<b>180</b>	3	60	35,41	<b><math>7,7 \cdot 10^{-11}</math></b>
Innerhalb d. Gruppen	61	<b>36</b>	19,94		
Total	<b>241</b>	<b>39</b>			

(Fett markierte Zahlen waren vorgegeben, der Rest zum Vervollständigen)

- a) **Vervollständige die Tabelle und bestimme den Wert der F-Statistik zum Test der Nullhypothese, dass die Mittelwerte der Gruppen gleich sind. (3 Punkte)**

Siehe Tabelle

- b) **Erklären Sie das Prinzip der Varianzanalyse. (3 Punkte)**

Antwort die 1 Punkt brachte, was Besseres hatte leider keiner von uns:  
Die Varianzanalyse ist die Verallgemeinerung eines Vergleichs von zwei Reihen von gemessenen Stichproben auf mehrere Messreihen.  
Am besten vorher im Skriptum oder Wikipedia raussuchen und anschauen.

- c) **Welche Voraussetzungen müssen für die Daten gelten, damit das statistische Testmodell anwendbar ist. (4 Punkte)**

Die Berechnung der Varianzanalyse entwickelt aus dem Prinzip ein statistisches Testverfahren, das auf der Annahme der Normalverteilung der Beobachtungen beruht. (S. 3 Varianzanalyse) Da diese Tests gewisse Voraussetzungen haben, müssen diese überprüft werden. Es sind dies dieselben wie bei der Varianzanalyse.

Normalverteilte Fehler

Die Fehler haben alle dieselbe Varianz

Die Fehler sind im Mittel 0

Antwort die 3 von 4 Punkten brachte:

Die Untersuchungsbedingungen müssen in allen Gruppen die gleichen sein, um vergleichbare Ergebnisse zu erhalten. Wichtig sind die Homogenität der Varianzen und die Annahme der Normalverteilung für die Fehler.

- d) **Liegt aufgrund der Daten genügend Evidenz vor, dass zwischen den Gruppen ein Unterschied besteht (Signifikanzniveau  $\alpha = 0,01$ )? (3 Punkte)**

$$F_{f; SSE; 1-\alpha} = F_{3; 61; 0,99} = 4,129$$

$4,129 < 35,41 \rightarrow$  Alternativhypothese wird angenommen, das die F-Wert sehr von einander abweichen zeigt an das eine „größer“ Unterschied zwischen den Gruppen existiert.

- e) **Wie groß ist die Anzahl der Beobachtungen in jeder Gruppe unter der Annahme, dass die Anzahl der Beobachtungen in allen Gruppen gleich groß sind? (2 Punkte)**

Laut Formel:  $n/r = \text{Anzahl Element} / \text{Anzahl Gruppen} = \text{hier } 39+1 / 3+1 = 10$

Laut Gleichverteilung der Elemente sollten 10 Beobachtungen pro Gruppe gemacht worden sein

- f) **Welche deskriptiven Statistiken sollte man jedenfalls bei der Varianzanalyse betrachten? (3 Punkte)**

Man sollte die Mittelwerte und Standardabweichungen der einzelnen Gruppen beachten. Zur Darstellung könnten Whisker-Box-Plots oder Strip-Charts verwendet werden.

- g) **Mit welcher Analysetechnik kann man feststellen, welche Gruppen unterschiedlich und welche gleichartig sind? (2 Punkte)**

Post-Hoc-Analyse siehe Skriptum 8Varianzanalyse.pdf Seite 14