Klausur
Grundlagen der Statistik

18. Juli 2003

Prof. U. Kockelkorn, N. Krämer, A. Conrad

- Die Klausur besteht aus **5 Aufgaben**, die alle in **90 Minuten** zu lösen sind.
- Lösen Sie die Aufgaben **nur** auf den ausgeteilten **Lösungsbögen**.
- Vermerken Sie auf jedem Bogen Ihren **Namen** und Ihre **Matrikelnummer**.
- Schreiben Sie bitte **lesbar**.
- Achten Sie darauf, dass Ihr **Lösungsweg** nachvollziehbar ist.
- Als **Hilfsmittel** sind Taschenrechner sowie Bücher oder geheftete Unterlagen erlaubt (*keine* losen Blätter).

Aufgabe 1 *(10 Punkte)*

Nach Angaben des Statistischen Bundesamtes haben die Deutschen im letzten Jahr durchschnittlich $250\,g$ Tee pro Kopf der Bevölkerung verbraucht. Sie vermuten, dass der durchschnittliche Teeverbrauch in Friesland höher ist als der Bundesdurchschnitt und wollen Ihre Vermutung mit Hilfe eines statistischen Tests untermauern.

Eine Umfrage unter 10 zufällig ausgewählten Personen in Friesland ergab einen durchschnittlichen Jahresverbrauch von $285\,g$. Die Standardabweichung wird mit $\hat{\sigma} = 20\,g$ aus der Stichprobe geschätzt. Es kann davon ausgegangen werden, dass der Teeverbrauch normalverteilt ist.

1. Das Niveau Ihres Tests soll $1 - \alpha = 0,95$ betragen. Was bedeutet das Test-Niveau inhaltlich?
2. Welche beiden grundsätzlichen Fehler sind bei der Testentscheidung möglich? Erläutern Sie beide Fehler! Welche Wahrscheinlichkeitsaussage können Sie bezüglich des Fehlers 2.Art treffen?
3. Führen Sie nun Ihren Test durch.
   (a) Stellen Sie die Hypothesen auf und begründen Sie Ihre Wahl.
   (b) Berechnen Sie die Prüfgröße sowie den Annahmebereich.
4. Welches Ergebnis liefert Ihr Test? Interpretieren Sie Ihr Ergebnis.
5. Angenommen, die Menschen in Friesland trinken im Mittel tatsächlich genau $250\,g$ Tee pro Kopf und Jahr. Ist es dann möglich, dass Sie bei Ihrem Test aus 3a den Fehler 2. Art begehen? Begründen Sie Ihre Antwort!

Aufgabe 2 *(15 Punkte)*

Beim Backen entsteht in Waffeln der gesundheitsschädliche Schadstoff A. Die A-Menge jeder Waffel kann als unabhängig normalverteilt angesehen werden.

Für die A-Menge einer Schokoladenwaffel (A_S) in mg gilt $A_S \sim N(\mu_S = 10\,;\, \sigma_S^2 = 4)$.
Für die A-Menge einer Nusswaffel (A_N) in mg gilt $A_N \sim N(\mu_N = 10\,;\, \sigma_N^2)$.
Es ist bekannt, dass 95 % der Nusswaffeln einen A-Gehalt zwischen 5,1 und 14,9 mg aufweisen.

Am liebsten essen Sie die sogenannten Kombiwaffeln (K) die jeweils aus 1 Schokoladenwaffel und 2 Nusswaffeln bestehen.

1. Wie ist die A-Menge einer Kombiwaffel (A_K) verteilt, wenn die übrigen Zutaten kein A enthalten? Bestimmen Sie auch die beiden Parameter der Verteilung.

Auf Grund der gesundheitlichen Risiken durch A-Konsum wird die erlaubte Höchstmenge einer Kombiwaffel auf $40,44\,mg$ A festgelegt.

2. Wie groß ist die Wahrscheinlichkeit π_G, dass der A-Gehalt einer zufällig ausgewählten Kombiwaffel diesen Grenzwert übersteigt?
   (*Hinweis*: Wenn Sie diese Aufgabe nicht lösen können, rechnen Sie mit $\pi_G = 0,5\,\%$ weiter.)

In der Fabrik werden jeweils 500 Kombiwaffeln in einen Karton verpackt.

3. Wie ist die Anzahl der Kombiwaffeln pro Karton, die den Grenzwert überschreiten, verteilt?
4. Wie groß ist die Wahrscheinlichkeit, dass sich in einem Karton höchstens 2 Kombiwaffeln befinden, die mehr A enthalten als erlaubt ist?
   (*Hinweis*: Verwenden Sie eine sinnvolle Approximation.)
5. Wie viele Kombiwaffeln hat man im Schnitt bereits gegessen, bis man die erste Kombiwaffel aus dem Karton entnimmt, die den Grenzwert überschreitet?

Aufgabe 3 *(10 Punkte)*

In einem Kino laufen die beiden Filme „Forrest Gump II - Gump Again" (G) und „Director's Cut: Patch Adams" (D). Gehen Sie davon aus, dass sich alle Personen an der Kinokasse unabhängig voneinander für einen der beiden Filme entscheiden. Sei π_1 die Wahrscheinlichkeit, dass eine Person den Film G sehen möchte.

1. Die Wahrscheinlichkeit, dass von zwei aufeinanderfolgenden Personen in der Schlange vor der Kasse mindestens eine Film D sehen möchte, ist $93,75\,\%$. Wie groß ist π_1?
   (*Hinweis*: Falls Sie diesen Aufgabenteil nicht lösen können, rechnen Sie mit $\pi_1 = 0,25$ weiter.)
2. Es stehen 10 Personen in der Schlange vor der Kino-Kasse. Wie ist die Anzahl der Personen in der Schlange vor der Kasse, die Film G sehen wollen, verteilt?
3. Berechnen Sie die Wahrscheinlichkeiten folgender Ereignisse:
   (a) E1: „Genau 3 Personen wollen Film G sehen."
   (b) E2: „Mindestens 8 Personen wollen Film G sehen."
   (c) E3: „Die 10. Person in der Schlange ist die 3. Person, die Film G sehen möchte."
4. Aus Erfahrung ist bekannt, dass 60 % der Personen, die das Kino besuchen, männlich sind. 30 % der Männer möchten Film D sehen. Mit welcher Wahrscheinlichkeit ist eine Person, die sich Film G anschaut, weiblich?

Aufgabe 4 *(10 Punkte)*

Man nimmt an, dass das Einkommen X in € der BewohnerInnen von Berlin-Dahlem eine zufällige Variable mit Dichtefunktion

$$f(x||\theta) = \begin{cases} 0 & , \quad x < \theta \\ 2\frac{\theta^2}{x^3} & , \quad x \geq \theta \end{cases}$$

ist, wobei der Parameter $\theta > 0$ eine positive reelle Zahl ist.

1. Zeigen Sie, dass $f(x||\theta)$ eine Dichtefunktion ist.
2. Leiten Sie die Verteilungsfunktion von X her.
3. Leiten Sie allgemein den Maximum-Likelihoodschätzer für θ her.
   (*Hinweis*: Versuchen Sie **nicht** die Likelihoodfunktion abzuleiten!)

Durch die Befragung von 5 zufällig ausgewählten Personen aus Berlin-Dahlem erhält man die folgende Stichprobe:

Person	Einkommen in €
1	5.600
2	6.200
3	4.800
4	10.000
5	6.300

4. Wie lautet für diese Stichprobe die ML-Schätzung für θ?

Aufgabe 5 *(5 Punkte)*

Bewerten Sie die folgenden 5 Aussagen mit *Falsch* oder *Richtig* und **begründen** Sie Ihre Entscheidung. Für Entscheidungen ohne Begründung gibt es **keine** Punkte.

1. Sie wissen, dass es sich bei Y um eine stetige Zufallsvariable handelt. Dieses Wissen ist ausreichend, um die Wahrscheinlichkeit dafür angeben zu können, dass Y den Wert 6,66 annimmt.
2. Sei $K = [0,8\,;\,1,2]$ das Konfidenzintervall für den Erwartungswert μ einer beliebigen Verteilung zum Konfidenzniveau $1-\alpha = 0,9$. Dann gilt mit einer Wahrscheinlichkeit von 90 %, dass $0,8 \leq \mu \leq 1,2$.
3. Der Maximum-Likelihood-Schätzer $\hat{\theta}$ gibt denjenigen Wert für den Parameter θ an, der mit der größten Wahrscheinlichkeit der richtige ist.
4. Sei X eine zufällige Variable, die einen Erwartungswert und eine Varianz besitzt. Die Wahrscheinlichkeit, eine Realisation von X zu beobachten, die mehr als vier Standardabweichungen vom Erwartungswert abweicht, ist gering. Dabei kann die Zufallsvariable eine beliebige Verteilung besitzen.
5. Hat der Regressor x im Regressionsmodell $y_i = \beta_0 + \beta_1 x + \varepsilon_i$ in Wirklichkeit keinen Einfluss auf die Variable y, so muss $\hat{\beta}_1 = 0$ sein.

Tabellierte Werte:

Verteilungsfunktion $\Phi(u)$ *der Standard-Normalverteilung* $N(0;1)$ *für* $1 \leq u < 3$

u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Quantile $t^{(n)}_{(1-\alpha)}$ *der zentralen t-Verteilung mit n Freiheitsgraden für* $1 \leq n \leq 20$ *und* $n \in \{100, 200, 300, 400, 500\}$

$n \backslash 1-\alpha$	0.9	0.95	0.975	0.99	0.995	0.999
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145
4	1.5332	2.1318	2.7764	3.7470	4.6041	7.1732
5	1.4759	2.0150	2.5706	3.3649	4.0322	5.8934
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076
7	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247
12	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874
15	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518
100	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737
200	1.2858	1.6525	1.9719	2.3451	2.6006	3.1315
300	1.2844	1.6499	1.9679	2.3388	2.5923	3.1176
400	1.2837	1.6487	1.9659	2.3357	2.5882	3.1107
500	1.2832	1.6479	1.9647	2.3338	2.5857	3.1066