Klausur
Grundlagen der Statistik
(WS 2002/03)

Bemerkungen:

- Die Klausur besteht aus **5 Aufgaben**, die alle zu lösen sind.
- Lösen Sie die Aufgaben **nur** auf den ausgeteilten Lösungsbögen.
- Vermerken Sie auf **jedem Blatt** Ihren Namen und Ihre Matrikelnummer.
- Schreiben Sie bitte **lesbar**.
- Achten Sie darauf, dass Ihr **Lösungsweg nachvollziehbar** ist.

Aufgabe 1 *(15 Punkte)*

Die Aufgabe besteht aus vier Teilaufgaben.

Bis vor einem Jahr hat Bauer M. immer die Getreidesorte X angebaut. Der erwartete Hektarertrag waren 20 GE (Gewichtseinheiten) und der Hektarertrag wies eine Varianz von 4 GE^2 auf. Der Großhändler über den Bauer M. sein Saatgut bezieht, hat Bauer M. im vergangenen Jahr probehalber auch Saatgut der Getreidesorte Y verkauft, mit der Anmerkung, dass der Hektarertrag zwar unterschiedlich ist, aber die Varianz des Hektarertrages gegenüber der Getreidesorte X unverändert ist.

Bauer M. geht aus der Erfahrung davon aus, daß die Erträge je Hektar unabhängig voneinander sind.

Bauer M. erntet dieses Jahr auf der Fläche von 4 Hektar, die er versuchsweise mit Getreidesorte Y bepflanzt hat, insgesamt 96 GE.

1. Geben Sie einen Wertebereich an in dem der 4-Hektar-Ertrag für Saatgut X mit einer Wahrscheinlichkeit von etwa 90 Prozent liegt. *(6 Punkte)*
2. Bewerten Sie die Aussage, daß der Erwartungswert des Hektarertrages der Sorte Y im Bereich von 21 bis 27 GE liegt. *(6 Punkte)*
3. Kann Bauer M. die Varianz des durchschnittlichen Hektarertrages reduzieren, wenn er auf den insgeamt 100 Hektar Anbaufläche die beiden Getreidesorten mischt? *(2 Punkte)*
4. Ist es für Bauer M. hinsichtlich der Prognose des durchschnittlichen Hektarertrages sinnvoll die 100 Hektar Anbaufläche nur teilweise zu bebauen? *(2 Punkte)*

Aufgabe 2 *(15 Punkte)*

Die Aufgabe besteht aus 6 Aufgabenteilen. In den Aufgabenteilen 5 und 6 werden keine Angaben oder Lösungen aus 1–4 benötigt.

Neben wichtiger elektronischer Post erhalten Sie leider auch unerwünschte (Werbe-)E-Mail, sogenannte Spam-Mail. Die Wahrscheinlichkeit, dass eine E-Mail, die in Ihrer Mail-Box landet, eine Spam-Mail ist, sei $\mathbf{P}_1$. Gehen Sie davon aus, dass an Sie adressierte E-Mails unabhängig voneinander versendet werden.

1. Die Wahrscheinlichkeit, dass unter zwei aufeinander folgenden E-Mails an Sie *mindestens eine* keine Spam-Mail ist, sei **0,9375**. Berechnen Sie die Wahrscheinlichkeit $\mathbf{P}_1$. *(1 Punkt)*

(*Hinweis:* Wenn Sie Aufgabeteil 1 nicht lösen können, rechnen Sie mit $\mathbf{P}_1 = \mathbf{0{,}25}$ weiter).

Am Morgen finden Sie **10** E-Mails in Ihrer Mail-Box vor. Berechnen Sie die Wahrscheinlichkeit der folgenden Ereignisse:

2. $\mathbf{P}_2$: Genau 3 Spam-Mails sind in Ihrer Mail-Box. *(3 Punkte)*
3. $\mathbf{P}_3$: Die zehnte E-Mail ist die dritte Spam-Mail. *(3 Punkte)*

Auf Grund der nicht endenden Belästigung durch Spam-Mail programmieren Sie ein E-Mail-Filter-Programm. Dabei wird der Text jeder eingehenden E-Mail automatisch überprüft. Wenn dieser bestimmte Wörter enthält, wird die Nachricht als "Spam-Mail" deklariert. Im anderen Fall wird die E-Mail als "erwünschte Nachricht" deklariert.

1 % der gefilterten E-Mails werden von Ihrem Programm falsch deklariert. Von den falsch deklarierten Nachrichten waren **20 %**Spam-Mail.

4. Wie groß ist demnach die Wahrscheinlichkeit, dass Ihr Programm eine Spam-Mail richtig deklariert. *(4 Punkte)*

Sie schreiben ein anderes Filter-Programm, dass nur den Absender **(A)** und die Betreffzeile **(B)** jeder E-Mail überprüft. Leider tritt dabei in einigen Fällen ein Programmfehler auf. Die Wahrscheinlichkeit, dass *mindestens eines* der Merkmale **A** oder **B** einen Fehler verursacht sei **11,5 %**. Gehen Sie davon aus, dass die Fehler unabhängig voneinander auftreten.

5. Berechnen Sie die Wahrscheinlichkeit für einen Fehler bei der Überprüfung von **A,** wenn bekannt ist, dass bei der Überprüfung von **B** mit einer Wahrscheinlichkeit **8,5 %** ein Fehler auftritt. *(4 Punkte)*
6. Mit welcher Wahrscheinlichkeit verursacht *nur eines* der Merkmale (**A** oder **B**) einen Fehler? *(3 Punkte)*

Aufgabe 3 *(10 Punkte)*

Die Aufgabe besteht aus zwei Teilaufgaben

Eddy und Gerd würfeln gegeneinander. Jeder hat einen eigenen Würfel und wer die höhere Augenzahl bei einem Wurf würfelt gewinnt die Runde. Sie haben schon 99 Runden gespielt und Eddy, der gerade am verlieren ist, hat sich die letzten zehn Augenzahlen von Gerd notiert: 6, 5, 4, 6, 3, 6, 1, 6, 6, 2.

Da Eddy darunter fünf Sechser zählt stellt er die Behauptung auf, daß Gerd mit einem unfairen Würfel spielt. Eddy, der in der Schule bei der Wahrscheinlichkeitsrechnung und Statistik gut aufgepasst hat, rechnet Gerd vor:

> Bezeichne π die Wahrscheinlichkeit mit Gerd's Würfel ein Sechs zu würfeln. Dann ist $\pi = \frac{1}{6}$, falls der Würfel fair ist. Unter der Annahme, daß der Würfel fair ist, beträgt die Wahrscheinlichkeit für fünf Sechsen bei zehn Würfen nur 1.3 Prozent. Deshalb kann der Würfel mit einer Wahrscheinlichkeit von über 98 Prozent nicht fair sein. Damit ist Gerd als Falschspieler entlarvt.

1. Stellen Sie basierend auf zehn Beobachtungen einen korrekten Test (Annahmebereich, Entscheidungsregel, Hypothesen, Kritische Region, Modell, Prüfgröße, Signifikanzniveau) zur Überprüfung der Hypothesen $\pi = \frac{1}{6}$ auf. *(7 Punkte)*
2. Zählen Sie drei Fehler auf, die Eddy bei seinem "Test" begeht. *(3 Punkte)*

Aufgabe 4 *(10 Punkte)*

Die Aufgabe besteht aus zwei Teilaufgaben.

Die Wartezeit bei der Abarbeitung von Druckaufträgen wird als exponentialverteilt angenommen. Die Wartezeiten der einzelnen Druckaufträge werden als unabhängig angesehen. Um eine Schätzung des Erwartungswertes für die Wartezeit zu berechnen, liegen zwei Beobachtungen vor:

- Wartezeit des Druckauftrags I war kürzer als drei Sekunden,
- Wartezeit des Druckauftrags II war länger als drei Sekunden.

1. Stellen Sie mit Hilfe der beiden Beobachtungen eine geeignete Likelihoodfunktion zur Schätzung des Erwartungswertes der Wartezeit auf. *(6 Punkte)*
2. Schätzen Sie den Erwartungswert der Wartezeit basierend auf den beiden Beobachtungen. *(4 Punkte)*

Aufgabe 5 *(10 Punkte)*

Überprüfen Sie den Wahrheitsgehalt folgender Aussagen und begründen Sie ihre Antwort.

1. Bei standardisierten Daten fallen das arithmetische Mittel und der empirische Median immer zusammen, falls der Median eindeutig ist. *(2 Punkte)*
2. Die Summe zweier unabhängiger Zufallsvariablen X, Y hat immer eine kleinere Varianz als die Summe zweier abhängiger Zufallsvariablen U, T, falls alle Zufallsvariablen X, Y, U und T die gleiche Varianz haben. *(2 Punkte)*

Bearbeiten Sie die beiden folgenden Aufgabenstellungen:

3. Erläutern Sie, warum im Zusammenhang mit dem Histogramm von der "Häufigkeitsdichte" gesprochen wird. *(2 Punkte)*
4. Als Schätzer für den Erwartungswert der Normalverteilung betrachten wir die zwei Schätzfunktionen $\widehat{\theta}_1(X_1, \ldots, X_n) = X_1$ und $\widehat{\theta}_2(X_1, \ldots, X_n) = \overline{X} - \frac{1000}{n}$.
   (a) Welcher der beiden Schätzer ist effizienter? *(1 Punkt)*
   (b) Welcher Schätzer ist im Sinne des MSE besser? *(2 Punkte)*
   (c) Welcher Schätzer ist im asymptotischen Sinne besser? *(1 Punkt)*

Tabelle: Verteilungsfunktion einer Binomialverteilung mit $n = 10$ und $\pi = \frac{1}{6}, \frac{1}{2}$

k	0	1	2	3	4	5	6	7	8	9
$P(X \leq k)$	0.1615	0.4845	0.7752	0.9303	0.9845	0.9975	0.9997	1.000	1.000	1.000
$P(X \leq k)$	0.0010	0.0107	0.0547	0.1719	0.3770	0.6330	0.8281	0.9453	0.9893	0.9990

z	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

Table 1: Verteilungsfunktion $\phi(z)$ der Standard-Normalverteilung $N(0;1)$