Klausur Statistik I WS 05/06 Ersttermin

- Beginnen Sie jede Aufgabe auf einem neuen Blatt.
- Schreiben Sie zu jeder Lösung den Verweis in eckigen Klammern [**X1**], so daß die Zuordnung möglich ist. Fehlt die Zuordnung, können keine Punkte vergeben werden.
- Machen Sie zu jeder Lösung den Lösungsweg deutlich, ansonsten können keine Punkte vergeben werden.

Aufgabe 1: An einer Universität wurden bei der Einschreibung zu den Studienfächern Jura, Philosophie und Naturwissenschaft Daten zu dem Geschlecht der Bewerber erhoben. Die gemeinsame Verteilung der statistischen Variablen "Studienfach" (X) und "Geschlecht" (Y) ist in der folgenden Kontingenztabelle dargestellt:

.

		Studienfach: X			Gesamt
		Jura	Philosophie	Naturwissenschaft	
Geschlecht: Y	Mann	25	82	250	357
	Frau	155	135	65	355
Gesamt		180	217	315	712

[**O1**] Berechnen Sie die relativen Häufigkeiten des Merkmals "Studienfach", d.h. $relH(X = x_i)$, und stellen Sie diese in einer geeigneten Grafik dar (Achsenbeschriftungen nicht vergessen!). Wie bezeichnet man die von Ihnen verwendete grafische Repräsentation? (5 Punkte)
[**O2**] Sind die beiden Merkmale "Studienfach" und "Geschlecht" statistisch unabhängig? Argumentieren Sie warum (oder warum nicht) dies so ist. Welches Skalenniveau haben die beiden statistischen Variablen? Welchen Wert für den Kontingenzkoeffizienten würden Sie erwarten (Vorzeichen, Größenordnung)? (5 Punkte)
[**O3**] Berechnen Sie die gemeinsamen relativen Häufigkeiten der beiden Merkmale unter der Annahme der Unabhängigkeit der beiden Merkmale und schreiben Sie die Ergebnisse in eine geeignete Tabelle. Berechnen sie die ausserdem die bedingten relativen Häufigkeiten $relH(Y = y_j|X = x_i)$. Benutzen Sie auch hier die Tabellenform zur Präsentation Ihrer Resultate. (8 Punkte)

Aufgabe 2: Die GK2006-Bank investiert in zwei Finanzanlagen. X (für Finanzanlage 1) und Y (für Finanzanlage 2) bezeichnen zwei statistische Variablen, welche die (Quartals-)Renditen der Finanzanlagen bezeichnen. Die Bank teilt das Anlagevermögen in die zwei Finanzanlagen so auf, dass sie $a\cdot$ 100% in Finanzanlage 1 und $(1-a)\cdot$ 100% in die Finanzanlage 2 investiert ($0 \leq a \leq 1$). Die Rendite des resultierenden Portfolios Z ergibt sich mit:

$$Z = a \cdot X + (1-a) \cdot Y$$

Aus den Zeitreihen der historischen Renditen der Finanzanlagen wurden die arithmetischen Mittel, die empirischen Varianzen und die empirische Kovarianz der beiden Renditen berechnet: $\bar{x} = 0.03$, $\bar{y} = -0,02$, $s_X^2 = 0,008$, $s_Y^2 = 0,004$, $c_{XY} = -0,005$

[**R1**] Nehmen Sie an, die Bank hätte das Anlagevermögen zu 10% in die Finanzanlage 1 und zu 90 % in die Finanzanlage 2 investiert. Berechnen Sie das arithmetische Mittel $\bar{z}$ und die Standardabweichung s_Z der Rendite dieses Portfolios. (8 Punkte)

[**R2**] Wie hätte die Bank das Anlagevermögen in die beiden Finanzanlagen aufteilen müssen (durch geeignete Wahl von a), damit die Standardabweichung der Rendite des resultierenden Portfolios so klein wie möglich gewesen wäre. Welches arithmetische Mittel hätte sich für die Rendite eines solchen Portfolios ergeben? Welche Wahl von a, ($0 \leq a \leq 1$), hätte die Varianz des Portfolios maximiert? (8 Punkte)

[**R3**] Tragen Sie in ein Diagramm, in dem auf der Abszisse die Standardabweichung der (Portfolio-)Rendite und auf der Ordinate das arithmetische Mittel der (Portfolio-)Rendite abgetragen wird, die Standardabweichung-Mittelwert-Kombinationen für das Portfolio mit der kleinsten Standardabweichung sowie die der beiden Finanzanlagen ein (haben Sie das Portfolio in [R2] nicht berechnen können, nehmen Sie einen sinnvollen Punkt an). Skizzieren (nicht berechnen) Sie in diesem Diagramm die Menge aller möglichen Standardabweichung-Mittelwert-Kombinationen, welche die Bank durch eine Variation von a erzielen kann. (8 Punkte)

Aufgabe 3:

Sie beobachten den Verlauf einer Zeitreihe für den Auslandsumsatz eines Produktes (Y), der in Abbildung 1 dargestellt ist. Ihre Aufgabe ist es, mit Hilfe einer Trendfunktion zukünftige Werte des Auslandsumsatzes zu prognostizieren. Eine mögliche geeignete Spezifikation ist:

$$y_t = b_0 + b_1 \cdot \sqrt{t} + u_t \quad .$$

[**S1**] Bestimmen Sie die Parameter des Trendmodells mit der Kleinstquadrate-Methode. Dafür stehen Ihnen folgende Angaben zur Verfügung: $\sum_{t=1}^{70} \sqrt{t} = 394$, $\sum_{t=1}^{70} t = 2485$, $\sum_{t=1}^{70} y_t = 406$, $\sum_{t=1}^{70} \sqrt{t} \cdot y_t = 2566$. Berechnen Sie die Umsatzprognose für $t = 71$ unter Verwendung der geschätzten Parameter. Das Be-

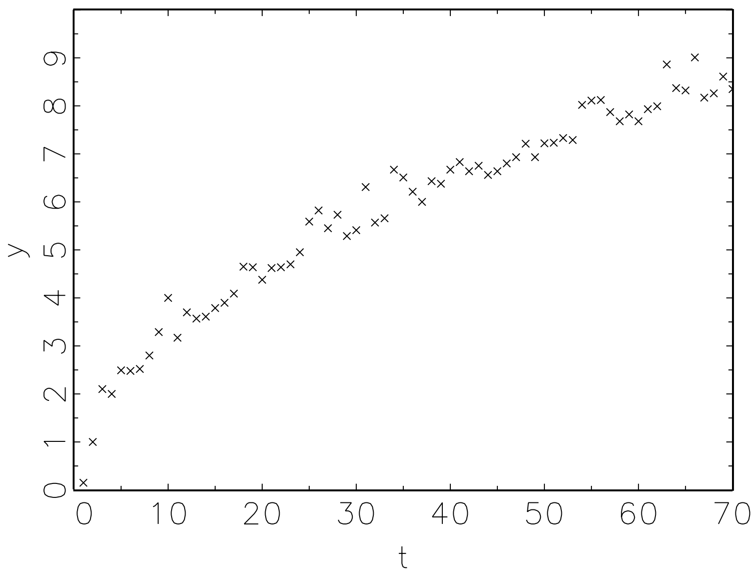

Abbildung 1: Auslandsumsatz

stimmtheitsmaß der Regression beträgt 0,85. Interpretieren Sie dieses Ergebnis. (8 Punkte)
[**S2**] Schlagen Sie eine weitere geeignete Trendfunktion vor mit der Sie den in der Abbildung 1 dargestellten Verlauf der Zeitreihe des Auslandsumsatzes beschreiben und prognostizieren können. Argumentieren Sie, warum Ihr Vorschlag Vorteile hätte und wie Sie die Parameter ihres Modells schätzen würden. (5 Punkte)

Aufgabe 4: Die Leiterin einer Presseabteilung eines Unternehmens möchte den Erfolg einer neu eingerichteten Hompage analysieren. Die Anzahl der Besucher auf der Web-Seite sollen Aufschluß über die Nutzungsintensität geben. Die statistische Variable X bezeichnet die Anzahl der Besucher pro Tag. In den ersten 15 Tagen wurden die folgenden Besucher pro Tag gezählt:
$\{11\ 10\ 13\ 12\ 14\ 9\ 9\ 9\ 10\ 10\ 12\ 13\ 12\ 10\ 11\}$
[**K1**] Stellen Sie die Häufigkeitsverteilung der statistischen Variable X in einer geeigneten Grafik dar. Charakterisieren Sie die statistische Variable X nach Merkmalstyp und Skalierung. Welches Konzept würden Sie wählen, um die Lage der empirischen Verteilung zu bestimmen? Nennen Sie mindestens zwei sinnvolle Lageparameter und benennen Sie kurz einen Nachteil und einen Vorteil derselben. (8 Punkte)
[**K2**] Zeichnen Sie die empirische Verteilungsfunktion der statistischen Variablen X. (5 Punkte)
[**K3**] Zeichnen Sie in der empirischen Verteilungsunktion das 0,75 Quantil ein und berechnen Sie den Modus und Median der empirischen Verteilung von X.

(5 Punkte)
[**K4**] Eine Kollegin schlägt vor, diesen Daten generierenden Prozess mit einer Poissonverteilung zu modellieren. Genauer gesagt, schlägt Sie die folgende Spezifikation vor: $f_{Po}(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$ wobei $\lambda = 11$. Welche Gründe könnten die Kollegin bewogen haben, dies vorzuschlagen? Halten Sie dies für eine gute Lösung? Argumentieren Sie! (5 Punkte)
[**K5**] Sie nehmen den Rat der Kollegin an und verwenden die in [K4] angegebene Verteilung. Wie groß ist die Wahrscheinlichkeit, dass genau 12 Besucher die Homepage besuchen und wie groß ist die Wahscheinlichkeit, dass 3 oder mehr Besucher die Homepage besuchen? (10 Punkte)

Aufgabe 5: Die Praktikantin schlägt vor den Erfolg einer neu eingerichteten Homepage an der Nutzungsdauer der Besucher zu messen, also wie lange ein Nutzer auf der Homepage surft. Dafür wird ein Programm entwickelt, das in der Lage ist die Nutzungsdauer der Besucher in Minuten aufzuzeichen. Sie erhalten folgende (aufsteigend georndete) Datenreihe, welche die Minuten pro Besuch wiedergibt. {0,442 0,657 0,746 0,780 1,113 1,430 1,527 1,619 2,390 2,795 3,339 3,530 3,968 3,978 5,405 5,862 5,876 6,683 18,990 22,013}
[**J1**] Stellen Sie die Häufigkeitsverteilung der statistischen Variable "Verweildauer in Minuten" in einer geeigneten Grafik dar (Achsenbeschriftungen nicht vergessen!). Dokumentieren Sie Ihre Vorgehensweise. (10 Punkte)
[**J2**] Für die statistische Variable "Verweildauer in Minuten" ergeben sich folgender Mittelwert und Median: $\bar{x} = 4,657$, $x_{Med} = 3,067$. Charakterisieren sie diese empirische Verteilung als (eher) symmetrisch, links- oder rechts-schief. Begründen Sie Ihre Aussage und geben Sie eine mögliche (psychologische/ökonomische oder sonst geartete) Erklärung für die Differenz von arithmetischem Mittel und Median. (5 Punkte)
[**J3**] Die Praktikantin schlägt vor, die Wahrscheinlichkeitsverteilung der Zufallsvariable X ="Verweildauer auf der Homepage in Minuten" mit einer Exponentialverteilung zu modellieren. Konkret schlägt sie vor, die folgende Dichtefunktion zu verwenden: $f_{Ex}(x;\lambda) = \lambda e^{-\lambda x}$ wobei $\lambda = 2$. Man entschließt sich, diesem Vorschlag zu folgen. Approximieren Sie zunächst mit Hilfe dieser Dichtefunktion die Wahrscheinlichkeit für das Ereignis, dass ein Besucher zwischen 6 und 8 Minuten auf der Homepage surft. Berechnen Sie dann die exakte Wahrscheinlichkeit für das Ereignis, dass ein Besucher zwischen 6 und 8 Minuten auf der Homepage surft (siehe Lösungshinweis). Wie hoch ist die Wahrscheinlichkeit für einen Besuch auf der Homepage, der weniger als 12 Minuten dauert und wie wahrscheinlich ist es, dass ein Besucher genau 12 Minuten auf der Homepage surft? (10 Punkte)
Lösungshinweis: $\int \lambda \cdot e^{-\lambda x} = -e^{-\lambda x} + C$.