

Name:	Vorname:	Matrikel-Nr.:
-------	----------	---------------

**Klausur zur Vorlesung Statistik I und II,  
WS 2008/09  
für Diplom-Studenten**

**Aufgabe 1****(40 Punkte)**

Um die wesentlichen Beeinflussungsfaktoren für die Passagier- und Güterverkehrsleistung deutscher Flughäfen zu bestimmen, wird eine Bestandsaufnahme relevanter Eigenschaften der verschiedenen Flughäfen aufgrund der letzten verfügbaren Statistiken erstellt. Insbesondere Zahl der Start- und Landebahnen und ihre Länge, ob es ein Nachtflugverbot gibt, die Wirtschaftskraft und Zahl der Hotelbetten im Einzugsgebiet (Region im 100 km bzw 1:30 Kfz-h-Umkreis), Größe und Entfernung des nächsten Konkurrenz-Flughafens, abgefertigte Passagierzahlen, abgefertigte Gütermenge, ob sich eine Hauptstadt im Einzugsgebiet befindet, ob eine Fluggesellschaft den Flughafen als Heimatflughafen benutzt und wie hoch die Start/Landegebühren sind

- (a) Geben Sie die statistische Gesamtheit und den Merkmalsträger an. Grenzen Sie die statistische Gesamtheit ab.
- (b) Geben Sie von allen oben genannten Merkmalen die Skalierung an.
- (c) Die Analyse der Beeinflussungsfaktoren soll mit einer multivariaten Regressionsanalyse durchgeführt werden. Welche zwei abhängigen Variablen sind dabei sinnvoll?
- (d) Stellen Sie nun für den Passagierverkehr eine multivariate lineare Regressionsfunktion auf, welche mindestens fünf unabhängige Variable enthält.
- (e) Wie könnte man nominalskalierte Variable wie die Existenz eines Nachtflugverbots in die Schätzfunktion der linearen Regression einbringen?

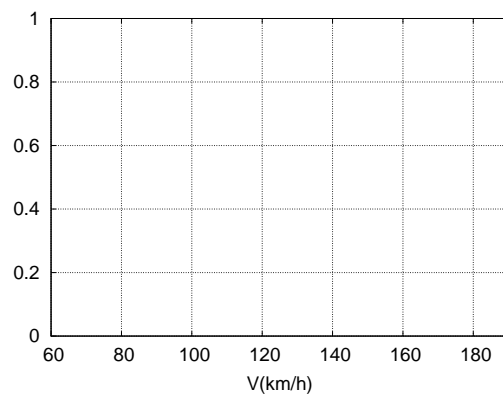
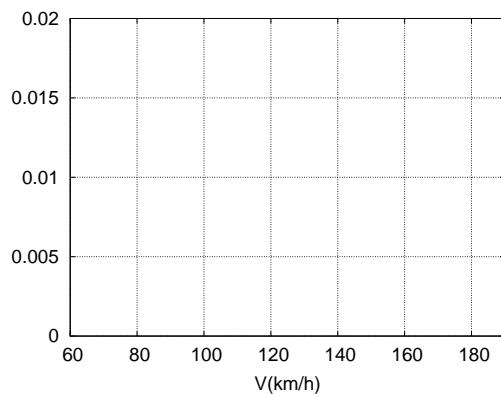
Name:	Vorname:	Matrikel-Nr.:
-------	----------	---------------

**Aufgabe 2****(60 Punkte)**

Auf einem mit Tempolimit 120 versehenen Autobahnabschnitt ergaben Induktionsschleifendetektoren während eines Tages folgende Zählergebnisse:

Geschwindigkeit (km/h)	60 - 80	80 - 90	90 - 100	100 - 110	110 - 120	120 - 140	140 - 160	160 - 180	180 - 200
Messungen	800	3300	2200	2500	3400	4400	1350	350	50

- (a) Wieviel Fahrzeuge wurden insgesamt erfasst? Welchem Verkehrsfluss (Kfz/h) entspricht das im Mittel?
- (b) Bestimmen Sie die relativen Häufigkeiten, Summenhäufigkeiten und die empirische Dichtefunktion für jede Klasse.
- (c) Zeichnen Sie die Dichte- und Verteilungsfunktion in die Diagramme ein. Ist die Verteilung bimodal? Wenn ja, was könnte der Grund sein?



- (e) Bestimmen Sie das arithmetische Mittel, den Median und den bzw. die Modi (mit Feinberechnung). Zeichnen Sie in die Grafik das 80. Perzentil ein.
- (f) Bestimmen Sie Varianz und die Standardabweichung und zeichnen Sie einen Boxplot dieser Verteilung.
- (g) Am Messquerschnitt soll nun ein "Blitzer" installiert werden. Mit wieviel Einnahmen kann man (zumindest anfangs) pro Tag rechnen, wenn nach folgender (vereinfachter aber aktueller) Gebührentabelle "abgerechnet" wird und jeweils 3 km/h Toleranz gegeben wird? Warum werden durch die Gleichverteilungsannahme innerhalb der Klassen die Einnahmen überschätzt?

Überschreitung um (km/h)	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	> 70
€	30	80	120	160	240	440	600

Name:	Vorname:	Matrikel-Nr.:
-------	----------	---------------

**Aufgabe 3****(30 Punkte)**

Zwischen Deutschland und China soll die Kaufkraftparität (Yuán pro Euro) anhand folgender deutscher und chinesischer monatlicher Warenkörbe ermittelt werden:

Gut	Sack Reis	Big Mac	Miete	Laptop	Bahnfahrt
Preis D	8 €	3.50 €	400 €/Mon.	1 000 €	0.20 €/km
Menge D	0.1	5	1 Monat	0.05	400 km
Preis China	15 Yuán	25 Yuán	700 Yuán/Mon.	5 000 Yuán	0.20 Yuán/km
Menge China	1	1	1 Monat	0.02	1 000 km

- Geben Sie die Kaufkraftparität bezüglich des deutschen Warenkorbs an und vergleichen Sie mit dem offiziellen Wechselkurs (5.5 Yuán/€)
- Ermitteln Sie nun die Kaufkraftparität bezüglich des chinesischen Warenkorbs.
- Nach einer These des Wirtschaftswissenschaftlers Balassa haben in Entwicklungs- und Schwellenländern nicht-handelbare Güter einen niedrigeren Preis als es dem Wechselkurs entspricht, während er bei ideal (ohne Kosten) handelbaren Gütern in etwa gleich dem Wechselkurs ist. Trifft diese These hier zu?
- Um die Wechselkurse aller Länder beurteilen zu können, hat der "New Economist" den *Big-Mac-Index* eingeführt, bei dem der Warenkorb zur Berechnung der Kaufkraftparität einheitlich aus einem (überall erhältlichen) *Big Mac* besteht. Nennen Sie einen Vor- und einen Nachteil dieses Index!

Name:	Vorname:	Matrikel-Nr.:
-------	----------	---------------

**Aufgabe 4****(50 Punkte)**

In dieser Aufgabe werden Fahrzeuge betrachtet, welche an einen festen Streckenquerschnitt vorbeifahren und dort durch Induktionsschleifen detektiert werden.

- Bei geringem Verkehrsaufkommen ist die Zeitlücke zwischen zwei Fahrzeugen angenähert exponentialverteilt. Bestimmen Sie den Parameter dieser Verteilung als Funktion des Verkehrsfusses  $Q$  (Fahrzeuge pro Zeiteinheit).
- Gegeben ist nun ein Fluss von 1 Fahrzeug pro Minute. Mit welcher Wahrscheinlichkeit kommen in den nächsten 5 Minuten höchstens zwei Fahrzeuge vorbei? Wie hoch ist der Median des zeitlichen Abstandes zwischen zwei Fahrzeugen?
- Nehmen Sie nun an, dass die Geschwindigkeit aller Fahrzeuge genau 72 km/h bzw. 20 m/s beträgt. Wie lautet die Verteilungsfunktion der räumlichen Fahrzeugabstände, wenn das Verkehrsaufkommen nach wie vor 1 Fahrzeug pro Minute beträgt?
- Für höhere Verkehrsaufkommen (dichter Verkehr) und Richtungsfahrbahnen mit nur einem Fahrstreifen ist die Exponentialverteilung zur Beschreibung der Abstände nicht geeignet. Ermitteln Sie dazu den aus ihr folgenden wahrscheinlichsten (zeitlichen) Abstand und erläutern Sie den Sachverhalt.
- Um einen zeitlichen Mindestabstand  $x_{\min} = 1 \text{ s} = 1/60 \text{ min}$  zu berücksichtigen, wird die Verteilung modifiziert. Die Dichtefunktion der neuen Verteilung ist gegeben durch

$$f(x) = \begin{cases} 0 & x < x_{\min} \\ \tilde{\lambda} e^{-\tilde{\lambda}(x-x_{\min})} & x \geq x_{\min} \end{cases} .$$

Ermitteln Sie  $\tilde{\lambda}$  in Abhängigkeit des Verkehrsaufkommens  $Q$  so, dass der mittlere zeitliche Abstand  $E(X)$  nach wie vor durch den Kehrwert  $1/Q$  des Verkehrsfusses gegeben ist.

*Hinweis:*  $\int_0^{\infty} x e^{-\tilde{\lambda}x} dx = \frac{1}{\tilde{\lambda}^2} .$

- Beachten Sie nun die beiden Werte  $Q_1 = 1 \text{ Kfz/min}$  und  $Q_2 = 30 \text{ Kfz/min}$  des Verkehrsaufkommens (geringer bzw. starker Verkehr). Für den Parameter  $\tilde{\lambda}_1$  der Verteilung von Aufgabenteil (e) gilt dann  $\tilde{\lambda}_1 = 1.01695$  (geringer Verkehr) bzw.  $\tilde{\lambda}_2 = 60$  (starker Verkehr). Beurteilen Sie die Auswirkung der Modellierung des Mindestabstandes für beide Fälle, indem Sie jeweils
  - den Wert des Parameters  $\tilde{\lambda}$  mit dem entsprechenden Wert aus der Exponentialverteilung vergleichen,
  - den Median der beiden Verteilungen vergleichen.

Für welches der beiden Verkehrsaufkommen  $Q_1$  und  $Q_2$  fällt der Unterschied zwischen der ursprünglichen und der modifizierten Exponentialverteilung geringer aus?

Name:	Vorname:	Matrikel-Nr.:
-------	----------	---------------

**Aufgabe 5****(30 Punkte)**

Bei einer Erhebung zum Mobilitätsverhalten soll auch das Einkommen als relevante Einflussgröße ermittelt werden. Da es heikel ist, danach direkt zu fragen, wird das Einkommen indirekt durch Erhebung von mit dem Einkommen korrelierten, aber unproblematischeren Merkmalen abgeschätzt. Konkret wird nach der Zahl der im Haushalt verfügbaren Autos gefragt. Aus Volkszählendaten ist folgendes bekannt:

- In der Einkommensklasse 1 (unter 30 000 €/Jahr) besitzen 40% kein Kfz, 55% ein Kfz und nur 5% mehr als 1 Kfz. In der Einkommensklasse 2 (30 000 bis 50 000 €/Jahr) ist die Aufteilung 20% (kein Kfz), 60% (ein Kfz) und 20% (mehrere), während in der Einkommensklasse 3 (>50 000 €/Jahr) 5% (kein Kfz), 45% (ein Kfz) und 50% (mehrere) gilt.
- 20% der Haushalte gehören zu Einkommensklasse 1, 60% zu Klasse 2 und der Rest zu Klasse 3.

Wie hoch sind die Wahrscheinlichkeiten dafür, dass (i) Haushalte ohne Kfz, (ii) Haushalte mit einem Kfz, (iii) Haushalte mit mehreren Kfz jeweils zu den Einkommensklassen 1, 2 oder 3 gehören?

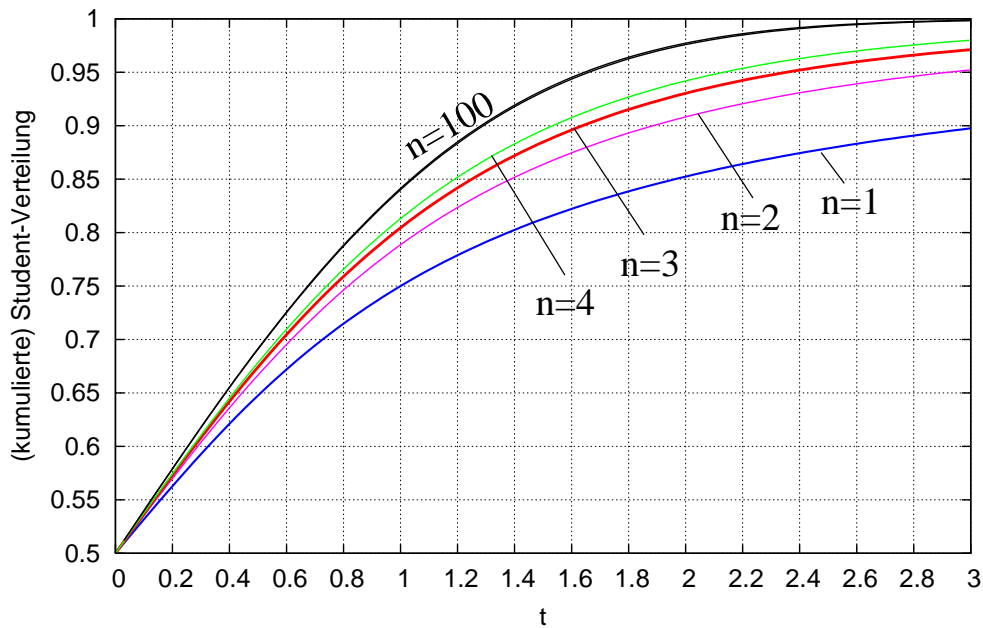
**Aufgabe 6****(40 Punkte)**

Auf einem Motorprüfstand wird die Leistung vier gleichartiger Motoren gemessen:

Motoren-Nr	1	2	3	4
Leistung (kW)	98.5	99.0	95.5	98.0

- Berechnen Sie zu einer Fehlerwahrscheinlichkeit von 5% das Konfidenzintervall der mittleren Leistung
- Der Fahrzeughersteller gibt die mittlere Leistung mit "mindestens 100 kW" an. Kann man ihn bei einer Fehlerwahrscheinlichkeit von 5% der Falschaussage bezichtigen?
- Das in (a) errechnete Konfidenzintervall enthält den Wert der Nullhypothese des Aufgabenteils (b), dennoch kann man in (b) die Nullhypothese verwerfen. Erläutern Sie diese scheinbar widersprüchlichen Ergebnisse.
- Zu welcher Fehlerwahrscheinlichkeit kann man die Nullhypothese "Leistung mindestens 100 kW" gerade noch verwerfen? Lesen Sie das Ergebnis von der Abbildung am Ende dieser Aufgabe ab! Falls Sie (b) nicht gerechnet haben, nehmen Sie einen Stichprobenwert der Test-Statistik von  $-2.8$  an.

- (e) Wie groß ist die Wahrscheinlichkeit für einen “Fehler zweiter Art”, wenn der tatsächliche Mittelwert der (gaußverteilten) Motorleistung  $\mu = 98 \text{ kW}$  beträgt und der Test wie bei (b) durchgeführt wird? *Hinweis:* Sie müssen hier anhand der Test-Statistik bezüglich der ursprünglichen Nullhypothese  $\mu_0 = 100 \text{ kW}$  den Wert  $\bar{x}_c$  von  $\bar{X}$  finden, oberhalb dem der Test fälschlicherweise angenommen wird. Tatsächlich gehorcht aber die “wahre” Statistik  $(\bar{X} - \mu)/S$  und nicht  $(\bar{X} - \mu_0)/S$  der Studentverteilung. Setzen Sie nun  $\bar{x}_c$  in die wahre Statistik ein. Verwenden Sie wieder die Grafik



## Tabellen

Quantile  $t_n^{(\alpha)}$  der Studentschen  $t$ -Verteilung

$n$	$\alpha = 0.60$	0.70	0.80	0.90	0.95	0.975	0.990	0.995	0.999	0.9995
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
$\infty$	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090	3.291