

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

## Klausur zur Vorlesung Statistik I und II, SS 2007

### Aufgabe 1

**(30 Punkte)**

- (a) Warum ist die Erhebung der zugelassenen Zahl von Fahrzeugen viel einfacher als die der Zahl der gefahrenen Kilometer?
- (b) Bei einer Mobilitätsbefragung sollen unter Anderem die Zahl der mit dem Kfz gefahrenen Kilometer sowie die Nutzung des öffentlichen Nahverkehrs erhoben werden. Um das Problem eines hohen Non-Response Anteils bei einem Meinungsumfrageinstitut zu umgehen, wird vorgeschlagen, doch die Mitglieder des ADAC (deutscher Automobilclub mit mehreren Millionen Mitgliedern) zu befragen, die ihrem Club sicher bereitwilliger Auskunft geben. Was ist dabei problematisch?
- (c) Bei Doppelinduktionsschleifen werden aus messtechnischen Gründen Geschwindigkeiten unter 10 km/h häufig nicht erfasst. Solche Geschwindigkeiten kommen auf den zu betrachtenden Strecke zu maximal 10% vor. Welche Lage- und Formmaße kann man trotz dieses Datenausfalls genau berechnen? Begründen Sie Ihre Aussagen!
- (d) Was könnte bei der Berechnung des arithmetischen Geschwindigkeitsmittels auf einer nichtverstauten Autobahn an Werktagen problematisch sein?
- (e) Mittels einer Zufalls-Stichprobe vom Umfang  $n > 6$  soll die mittlere Länge  $\mu$  des Weges von der Wohnung zur Arbeit erhoben werden. Geben Sie für die folgenden Schätzer für  $\mu$  an, ob sie erwartungstreu und/oder konsistent und/oder effizient sind:

(i)  $\bar{x}^{(i)} = \frac{1}{n-1} \sum_{i=1}^n x_i,$

(ii)  $\bar{x}^{(ii)} = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i,$

(iii)  $\bar{x}^{(iii)} = \sum_{i=1}^{n-1} c_i x_i$  mit  $c_i = \frac{1}{2n}$ , wenn  $i$  gerade und  $c_i = \frac{3}{n}$  sonst,

(iv)  $\bar{x}^{(iv)} = \frac{1}{n} \sum_{i=1}^n x_i,$

(v)  $\bar{x}^{(v)} = \frac{1}{3}(x_2 + x_4 + x_6).$

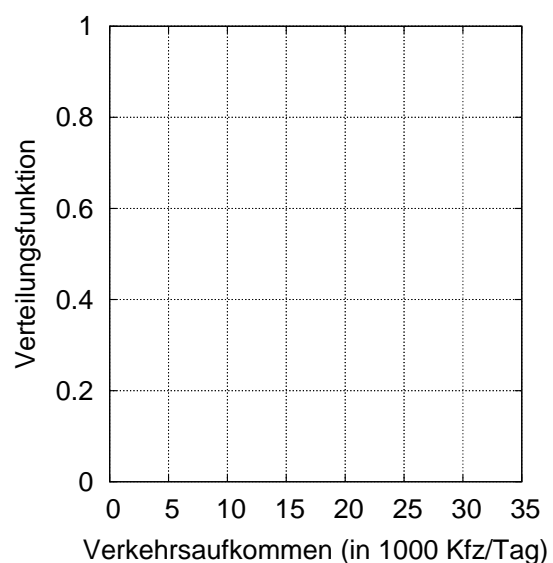
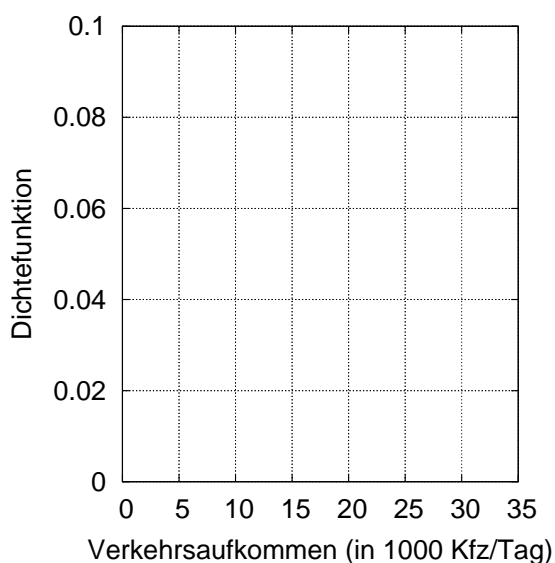
Name:	Vorname:	Matrikel-Nr.:
-------	----------	---------------

**Aufgabe 2****(50 Punkte)**

Die tägliche Verkehrsbelastung auf den Bundesstraßen in Deutschland teilte sich im Jahr 2005 wie folgt auf die Streckenkilometer auf:

Mittlere Verkehrsbelastung in 1000 Kfz/Tag	Streckenkilometer
0-5	8875
5-10	15380
10-20	11260
20-30	1860
> 30	420

- Wieviel Kilometer Bundesstraße wurden erfasst?
- Bestimmen Sie die relativen Häufigkeiten der einzelnen Klassen
- Bestimmen Sie die Dichtefunktion und die Verteilungsfunktion. Nehmen Sie dabei für die oberste Klasse eine Obergrenze von 35 000 Kfz/Tag an.
- Zeichnen Sie die Dichte- und Verteilungsfunktion in die Diagramme auf dieser Seite ein.
- Bestimmen Sie das arithmetische Mittel, den Median und den Modus (Dichtemittel).



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

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

Die Entwicklung der jährliche Verkehrsleistung in Milliarden Personenkilometern wird wie folgt angegeben:

Jahr	früheres Bundesgebiet	DDR	Deutschland
1950	30.7	10.0	–
1960	161.7	45.3	–
1970	350.6	81.1	–
1980	477.4	97.0	–
1990	601.8	114.2	715.8
2000	–	–	849.6
2004	–	–	907.0

- Diskutieren Sie, warum die Angabe von bis zu vier signifikanten Stellen hier höchst problematisch ist.
- Geben Sie den Index der Leistung für das frühere Bundesgebiet und die DDR von 1950-2000 zur Basis 1970 an!
- Basieren Sie nun auf 1950 um.
- Berechnen Sie nun einen Index für Gesamtdeutschland von 1950 bis 2004 zur Basis 1950.

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

Gegeben sind folgende Daten für die zeitliche Entwicklung des deutschen Luft-Güterverkehrs:

Jahr	transportierte Menge (Mio t)
1970	0.4
1980	0.7
1990	1.3
2000	2.2
2002	2.2
2004	2.7

- Führen Sie eine nichtlineare Regression mit der Exponentialfunktion  $\hat{y} = ae^{bx}$  durch!
- Rechnen Sie nun mit  $\hat{y}(2004) = 2.66$  Millionen t und einer Verdoppelung des Anstieges alle 12.6 Jahre. Prognostizieren Sie die transportierte Gütermenge im Luftverkehr für das Jahr 2010 anhand der Regressionsfunktion.
- Berechnen Sie das Bestimmtheitsmaß.

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

**Aufgabe 5****(40 Punkte)**

Da man “Druckfehler” bekanntlich selbst schlecht findet, lässt ein Diplomkandidat / eine Kandidatin die Diplomarbeit vor Abgabe noch von zwei Bekannten korrigieren. Bei der ersten Korrektur werden 75 Fehler gefunden, während der zweite Korrektor unabhängig davon 60 Fehler findet. Bei der Durchsicht findet der Kandidat, dass nur 30 Fehler von beiden gefunden wurden.

- Die Ereignisse “ein Fehler wurde vom ersten (bzw. zweiten) Korrektor entdeckt” seien nun mit  $A$  (bzw.  $B$ ) gekennzeichnet, während  $C$  das Ereignis kennzeichnet, dass beide den Fehler finden. Wie groß ist  $P(C|A)$  und  $P(C|B)$ ?
- Wie groß sind die unbedingten Wahrscheinlichkeiten  $P(A)$  und  $P(B)$ ? *Hinweis:* Verwenden Sie die Produktregel für Wahrscheinlichkeiten und berücksichtigen Sie die Unabhängigkeit!
- Rechnen Sie nun mit  $p(B) = 30/75$  weiter. Wie viele unentdeckte Fehler sind (im Mittel) in der Arbeit noch zu erwarten?
- Es wird nun angenommen, dass beim Schreiben der Arbeit die Wahrscheinlichkeit eines neuen Druckfehlers nicht von den bereits zuvor gemachten Fehlern abhängt. Welche der folgenden Verteilungen kann man zur Beschreibung der *Gesamtfehlerzahl* verwenden?
  - Gaußverteilung,
  - Exponentialverteilung,
  - Binomialverteilung
  - Poissonverteilung
  - hypergeometrische Verteilung
  - Gleichverteilung

*Hinweis:* Es können durchaus mehrere dieser Verteilungen anwendbar sein, teilweise mit Einschränkungen. Es kann außerdem sein, dass eine Anwendung zwar möglich, aber nicht sinnvoll ist. Geben Sie all dies an!

- Es seien nun pro Seite noch drei unentdeckter Fehler zu erwarten. Mit welcher Wahrscheinlichkeit ist die Kurzzusammenfassung (1.5 Seiten) fehlerfrei und mit welcher Wahrscheinlichkeit enthält sie mehr als 3 Fehler? Mit welcher Wahrscheinlichkeit gibt es in der gesamten Arbeit (50 Seiten) wenigstens eine Seite ohne Fehler?

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

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

Eine Schwangerschaft dauert, vom Zeitpunkt der letzten Menstruation aus gerechnet, im Mittel 280 Tage (Naegelesche Regel). Zum berechneten Termin kommen jedoch nur vier Prozent der Kinder zur Welt.

- (a) Die Schwangerschaftsdauer kann sehr gut mit einer Gaußverteilung beschrieben werden. Geben Sie deren Parameter an!
- (b) Kommt das Kind mehr als 3 Wochen vor dem mit der "Naegeleschen Regel" errechneten Mittelwert zur Welt, gilt dies als Frühgeburt. Wie groß ist die Wahrscheinlichkeit dafür? *Hinweis:* Nehmen Sie für diese und die folgenden Teilaufgaben eine Standardabweichung von 10 Tagen an.
- (c) Das Kind ist am errechneten Geburtstermin noch nicht da und es gibt auch keine Anzeichen, dass die Geburt unmittelbar bevorsteht. Warum ist es problematisch, die Verwandten auf einen "Besichtigungstermin" in zwei Wochen zu bestellen? Berechnen Sie dazu die Wahrscheinlichkeit dafür, dass die Geburt in den nächsten zwei Wochen stattfindet! (Achtung: Sie müssen mit bedingten Wahrscheinlichkeiten rechnen!)

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

## Tabellen

### Werte $\Phi(z)$ der Standardnormalverteilung

$z$	$\Phi(z)$	$z$	$\Phi(z)$	$z$	$\Phi(z)$	$z$	$\Phi(z)$
0.0	0.5000	1.00	0.8413	2.0	0.9772	3.0	0.9987
0.1	0.5398	1.1	0.8643	2.1	0.9821	3.1	0.9990
0.2	0.5793	1.2	0.8849	2.2	0.9861	3.2	0.9993
0.3	0.6179	1.3	0.9032	2.3	0.9893	3.3	0.9995
0.4	0.6554	1.4	0.9192	2.4	0.9918	3.4	0.9997
0.5	0.6915	1.5	0.9332	2.5	0.9938	3.5	0.9998
0.6	0.7257	1.6	0.9452	2.6	0.9953	3.6	0.9998
0.7	0.7580	1.7	0.9554	2.7	0.9965	3.7	0.9999
0.8	0.7881	1.8	0.9641	2.8	0.9974	3.8	0.9999
0.9	0.8159	1.9	0.9713	2.9	0.9981	3.9	1.0000

### Quantile $z_\alpha$ der Standardnormalverteilung

$\alpha$	$z_\alpha$	$\alpha$	$z_\alpha$	$\alpha$	$z_\alpha$	$\alpha$	$z_\alpha$
0.995	2.5758	0.945	1.5982	0.880	1.1750	0.680	0.4677
0.990	2.3263	0.940	1.5548	0.860	1.0803	0.660	0.4125
0.985	2.1701	0.935	1.5141	0.840	0.9945	0.640	0.3585
0.980	2.0537	0.930	1.4758	0.820	0.9154	0.620	0.3055
0.975	1.9600	0.925	1.4395	0.800	0.8416	0.600	0.2533
0.970	1.8808	0.920	1.4051	0.780	0.7722	0.580	0.2019
0.965	1.8119	0.915	1.3722	0.760	0.7063	0.560	0.1510
0.960	1.7507	0.910	1.3408	0.740	0.6433	0.540	0.1004
0.955	1.6954	0.905	1.3106	0.720	0.5828	0.520	0.0502
0.950	1.6448	0.900	1.2815	0.700	0.5244	0.500	0.0000

### Quantile $t_n^{(\alpha)}$ der Studentischen $t$ -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