SS 2006 *Name:*
Matrikel-Nr.:

PRÜFUNG STATISTIK I (2)

11. OKTOBER 2006
HILFSMITTEL: A4 BLATT MIT NOTIZEN, TR

Die Prüfung besteht aus zwei Teilen, die insgesamt 120 Punkte ergeben:
Teil 1: Multiple Choice (jede Aufgabe 2 Punkte)
Teil 2: 5 von 8 Aufgaben mit jeweils 20 Punkten

Bitte ankreuzen, welche Aufgaben bewertet werden sollen (5 von 8)!

Aufgabe:	1	2	3	4	5	6	7	8	MC	Gesamt
Bewerten:										
Punkte:										

Schreiben Sie bitte auf jedes Blatt oben rechts Ihren Namen!

Viel Erfolg!

Teil 1. : Multiple Choice

Zu jeder Frage ist genau eine richtige Antwortmöglichkeit vorgegeben. Tragen Sie Ihre Lösungen in die Kästchen auf der **übernächsten** Seite ein. Die Rückseite der Blätter können Sie für Berechnungen sowie zu Anmerkungen und Erläuterungen Ihrer Lösung verwenden.

(1) Am 10. August wurden zufällig 10 Fluggäste die am Münchener Flughafen ankamen nach ihren Security-Check Zeiten am Startflughafen gefragt. Es ergaben sich folgende Zeiten in Minuten: 7, 51, 14, 17, 14, 19, 36, 76, 23, 60. Welche Aussage über die Daten ist falsch?
  (a) Der Median ist 21.
  (b) Die Spannbreite der Daten ist 69.
  (c) Der Modus ist 17.
  (d) Im Schnitt mußte man über 30 Minuten warten.

(2) Welche der folgenden Aussagen trifft für einen Standard Mosaic Plot zu?
- (a) Alle Zellen haben dieselbe Fläche.
- (b) Die Fläche einer Zelle hängt von der Anzahl der selektierten Fällen ab.
- (c) Die Fläche einer Zelle hängt von der Anzahl der entsprechenden Fällen ab.
- (d) Alle Zellen sind quadratisch.

(3) Welche der Aussagen zu nicht-parametischen Tests ist falsch?
- (a) Für das Ergebnis des Wilcoxon-Rangsummentest ist es egal welche der beiden Rangsummen man betrachtet.
- (b) Der Wilcoxon Vorzeichen-Rangtest ist effizienter als der Vorzeichentest.
- (c) Für großes n kann man beim Wilcoxon-Rangsummentest die Normal-Approximation benutzen.
- (d) Die Teststatistik eines nicht-parametrischen Tests ist stets verteilungsfrei.

(4) Welche der folgenden Aussagen trifft für einen ML-Schätzer $\hat{\theta}$ zu?
- (a) $\hat{\theta}$ ist erwartungstreu.
- (b) $\hat{\theta}$ ist asymptotisch erwartungstreu.
- (c) $\hat{\theta}$ ist der Schätzer mit dem minimalsten MSE.
- (d) $\hat{\theta}$ ist stark konsistent.

(5) 1869 hat Galton eine Normalverteilung für die Größe von Männern, gemessen in Zoll, vorgeschlagen: $X \sim N(66, (3,2)^2)$. Mit diesem Modell, was ist die Wahrscheinlichkeit, dass ein zufällig ausgewählter Mann zwischen 70 und 80 Zoll groß ist (d.h. zwischen $177,8$ und $203,2$ cm)?
- (a) 0,212
- (b) 0,894
- (c) 0,106
- (d) 0,025

(6) Der p-Wert ist
- (a) ... unabhängig von der Entscheidung für einen 1-seitigen oder 2-seitigen Test.
- (b) ... abhängig von dem gewählten Signifikanz-Niveau.
- (c) ... abhängig von der Teststatistik.
- (d) ... unabhängig von der Verteilung unter der Nullhypothese.

In einer Studie wurde die maximale Muskelkraft im Oberschenkel gemessen und durch das Alter in Jahren und das Körpergewicht in kg erklärt. Die untersuchten Personen waren durchwegs Männer, die zwischen 20 und 50 Jahre alt waren.
Das Regressionsmodell lautet:

$$\text{Muskelkraft}_i = \alpha + \beta_1 \cdot \text{Körpergewicht}_i + \beta_2 \cdot \text{Alter}_i + E_i, \quad E_i \sim N(0, \sigma^2)$$

```
Coefficients: Estimate Std. Error t value
(Intercept)     115.00      42.78       ...
gewicht          3.00        0.47        ...
alter           -2.65        0.53        ...
Residual standard error: 30.44 on 27 degrees of freedom
F-statistic: 30.02 on 2 and 27 DF, p-value: 1.793e-08
```

(7) Mit wievielen Beobachtungen wurde diese Regression gerechnet?
(a) 30 (b) 29 (c) 28 (d) 27

(8) Ist es statistisch gesichert, dass in diesem Regressionmodell ein Zusammenhang zwischen dem Körpergewicht und der Muskelkraft besteht?
(a) ja
(b) nein
(c) keine Aussage möglich
(d) ja, aber nur zu einem Niveau von 5%!

(9) Wie groß ist die prognostizierte Muskelkraft eines 40-jährigen Mannes, der 80kg wiegt?
(a) 134 (b) 249 (c) 118.35 (d) 375

(10) Wie groß ist die Schätzung $\hat{\sigma}$ von σ?
(a) 25 (b) 27 (c) 30.02 (d) 30.44

Antworten:

1	2	3	4	5	6	7	8	9	10

Verteilungstabelle der Standardnormalverteilung

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.000	0.5000	0.750	0.7734	1.500	0.9332	2.250	0.9878
0.025	0.5100	0.775	0.7808	1.525	0.9364	2.275	0.9885
0.050	0.5199	0.800	0.7881	1.550	0.9394	2.300	0.9893
0.075	0.5299	0.825	0.7953	1.575	0.9424	2.325	0.9900
0.100	0.5398	0.850	0.8023	1.600	0.9452	2.350	0.9906
0.125	0.5497	0.875	0.8092	1.625	0.9479	2.375	0.9912
0.150	0.5596	0.900	0.8159	1.650	0.9505	2.400	0.9918
0.175	0.5695	0.925	0.8225	1.675	0.9530	2.425	0.9923
0.200	0.5793	0.950	0.8289	1.700	0.9554	2.450	0.9929
0.225	0.5890	0.975	0.8352	1.725	0.9577	2.475	0.9933
0.250	0.5987	1.000	0.8413	1.750	0.9599	2.500	0.9938
0.275	0.6083	1.025	0.8473	1.775	0.9621	2.525	0.9942
0.300	0.6179	1.050	0.8531	1.800	0.9641	2.550	0.9946
0.325	0.6274	1.075	0.8588	1.825	0.9660	2.575	0.9950
0.350	0.6368	1.100	0.8643	1.850	0.9678	2.600	0.9953
0.375	0.6462	1.125	0.8697	1.875	0.9696	2.625	0.9957
0.400	0.6554	1.150	0.8749	1.900	0.9713	2.650	0.9960
0.425	0.6646	1.175	0.8800	1.925	0.9729	2.675	0.9963
0.450	0.6736	1.200	0.8849	1.950	0.9744	2.700	0.9965
0.475	0.6826	1.225	0.8897	1.975	0.9759	2.725	0.9968
0.500	0.6915	1.250	0.8944	2.000	0.9772	2.750	0.9970
0.525	0.7002	1.275	0.8988	2.025	0.9786	2.775	0.9972
0.550	0.7088	1.300	0.9032	2.050	0.9798	2.800	0.9974
0.575	0.7174	1.325	0.9074	2.075	0.9810	2.825	0.9976
0.600	0.7257	1.350	0.9115	2.100	0.9821	2.850	0.9978
0.625	0.7340	1.375	0.9154	2.125	0.9832	2.875	0.9980
0.650	0.7422	1.400	0.9192	2.150	0.9842	2.900	0.9981
0.675	0.7502	1.425	0.9229	2.175	0.9852	2.925	0.9983
0.700	0.7580	1.450	0.9265	2.200	0.9861	2.950	0.9984
0.725	0.7658	1.475	0.9299	2.225	0.9870	2.975	0.9985

Quantile $\chi^2_{df;1-\alpha}$ der χ^2-Verteilung

df ╲ α	0.100	0.050	0.025	0.020	0.010	0.005	0.0025	0.001	0.0005
1	2.705	3.841	5.023	5.411	6.634	7.879	9.140	10.827	12.115
2	4.605	5.991	7.377	7.824	9.210	10.596	11.982	13.815	15.201
3	6.251	7.814	9.348	9.837	11.344	12.838	14.320	16.266	17.730
4	7.779	9.487	11.143	11.667	13.276	14.860	16.423	18.466	19.997
5	9.236	11.070	12.832	13.388	15.086	16.749	18.385	20.515	22.105
6	10.644	12.591	14.449	15.033	16.811	18.547	20.249	22.457	24.102
7	12.017	14.067	16.012	16.622	18.475	20.277	22.040	24.321	26.017
8	13.361	15.507	17.534	18.168	20.090	21.954	23.774	26.124	27.868
9	14.683	16.918	19.022	19.679	21.665	23.589	25.462	27.877	29.665
10	15.987	18.307	20.483	21.160	23.209	25.188	27.112	29.588	31.419
11	17.275	19.675	21.920	22.617	24.724	26.756	28.729	31.264	33.136
12	18.549	21.026	23.336	24.053	26.216	28.299	30.318	32.909	34.821
13	19.811	22.362	24.735	25.471	27.688	29.819	31.883	34.528	36.477
14	21.064	23.684	26.118	26.872	29.141	31.319	33.426	36.123	38.109
15	22.307	24.995	27.488	28.259	30.577	32.801	34.949	37.697	39.718

Quantile $t_{n;1-\alpha}$ der t-Verteilung

n ╲ α	0.100	0.050	0.025	0.020	0.010	0.005	0.0025	0.001	0.0005
1	3.077	6.313	12.706	15.894	31.820	63.656	127.321	318.308	636.619
2	1.885	2.919	4.302	4.848	6.964	9.924	14.089	22.327	31.599
3	1.637	2.353	3.182	3.481	4.540	5.840	7.453	10.214	12.923
4	1.533	2.131	2.776	2.998	3.746	4.604	5.597	7.173	8.610
5	1.475	2.015	2.570	2.756	3.364	4.032	4.773	5.893	6.868
6	1.439	1.943	2.446	2.612	3.142	3.707	4.316	5.207	5.958
7	1.414	1.894	2.364	2.516	2.997	3.499	4.029	4.785	5.407
8	1.396	1.859	2.306	2.448	2.896	3.355	3.832	4.500	5.041
9	1.383	1.833	2.262	2.398	2.821	3.249	3.689	4.296	4.780
10	1.372	1.812	2.228	2.359	2.763	3.169	3.581	4.143	4.586
11	1.363	1.795	2.200	2.328	2.718	3.105	3.496	4.024	4.436
12	1.356	1.782	2.178	2.302	2.680	3.054	3.428	3.929	4.317
13	1.350	1.770	2.160	2.281	2.650	3.012	3.372	3.851	4.220
14	1.345	1.761	2.144	2.263	2.624	2.976	3.325	3.787	4.140
15	1.340	1.753	2.131	2.248	2.602	2.946	3.286	3.732	4.072
16	1.336	1.745	2.119	2.235	2.583	2.920	3.251	3.686	4.014
17	1.333	1.739	2.109	2.223	2.566	2.898	3.222	3.645	3.965
18	1.330	1.734	2.100	2.213	2.552	2.878	3.196	3.610	3.921
19	1.327	1.729	2.093	2.204	2.539	2.860	3.173	3.579	3.883
20	1.325	1.724	2.085	2.196	2.527	2.845	3.153	3.551	3.849
25	1.316	1.708	2.059	2.166	2.485	2.787	3.078	3.450	3.725
50	1.298	1.675	2.008	2.108	2.403	2.677	2.936	3.261	3.496
100	1.290	1.660	1.983	2.080	2.364	2.625	2.870	3.173	3.390

Verteilungsfunktion der Binomialverteilung $B_{n;p}(k)$ für $p = 0.5$.

$n \diagdown k$	0	1	2	3	4	5	6	7	8	9	10
1	0.500	1.000									
2	0.250	0.750	1.000								
3	0.125	0.500	0.875	1.000							
4	0.062	0.313	0.687	0.938	1.000						
5	0.031	0.188	0.500	0.812	0.969	1.000					
6	0.016	0.109	0.344	0.656	0.891	0.984	1.000				
7	0.008	0.063	0.227	0.500	0.773	0.938	0.992	1.000			
8	0.004	0.035	0.145	0.363	0.637	0.855	0.965	0.996	1.000		
9	0.002	0.020	0.090	0.254	0.500	0.746	0.910	0.980	0.998	1.000	
10	0.001	0.011	0.055	0.172	0.377	0.623	0.828	0.945	0.989	0.999	1.000

Kritische Werte $W_{n;\alpha}$ des Vorzeichenrangtests von Wilcoxon

$n \diagdown \alpha$	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
4	0	0	0	1	9	10	10	10
5	0	0	1	3	12	14	15	15
6	0	1	3	4	17	18	20	21
7	1	3	4	6	22	24	25	27
8	2	4	6	9	27	30	32	34
9	4	6	9	11	34	36	39	41
10	6	9	11	15	40	44	46	49
11	8	11	14	18	48	52	55	58
12	10	14	18	22	56	60	64	68
13	13	18	22	27	64	69	73	78
14	16	22	26	32	73	79	83	89
15	20	26	31	37	83	89	94	100
16	24	30	36	43	93	100	106	112
17	28	35	42	49	104	111	118	125
18	33	41	48	56	115	123	130	138
19	38	47	54	63	127	136	143	152
20	44	53	61	70	140	149	157	166

Kritische Werte $W_{n;m;\alpha}$ des Rangsummentests von Wilcoxon

$\alpha = 0.05$	m																	
n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	6	7	8	9	9	10	10	11	12	12	13	14	14	15	16	16	17	18
4	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
5	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
6	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
7	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
8	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
9	49	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
10	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
11	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
12	84	88	92	96	100	105	109	113	117	121	126	130	134	139	143	147	151	156
13	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
14	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
15	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
16	145	151	156	162	167	173	179	185	191	197	202	208	214	220	226	232	238	244
17	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
18	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
19	201	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	321
20	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349

$\alpha = 0.01$	m																	
n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12
4	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21
5	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
6	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44
7	29	30	32	33	35	36	38	40	41	43	45	46	48	50	52	53	55	57
8	37	39	41	43	44	46	48	50	52	54	57	59	61	63	65	67	69	71
9	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86
10	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103
11	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120
12	81	84	87	90	93	96	100	103	107	110	114	117	121	125	128	132	135	139
13	94	97	101	104	108	112	115	119	123	127	131	135	139	143	147	151	155	159
14	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179
15	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201
16	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	213	219	224
17	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247
18	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272
19	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298
20	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325

Teil 2

Bearbeiten Sie 5 der 8 Aufgaben!

1. Diplomnoten

Zwischen den Studenten hält sich das Gerücht, dass die Informatikstudenten schlechtere Diplomnoten als die Mathematikstudenten haben. Student X behauptet, dass die durchschnittliche Diplomnote der Informatiker um 0,4 Punkte schlechter ist als die der Mathematiker. Um dies näher zu untersuchen befragt Student X 105 Diplom-Informatiker und 99 Diplom-Mathematiker nach ihrer exakten Diplomnote. Dabei ergab sich eine durchschnittliche Informatikdiplomnote von 2,58 und eine durchschnittliche Mathematikdiplomnote von 2,42. Die Stichprobenstandardabweichungen liegen bei $s_I = 1,5$ und $s_M = 1,6$

(a) Kann Student X mit diesen Werten seine Vermutung aufrechterhalten?

Student X weiß zudem, dass die Diplomnote annähernd normalverteilt ist.

(b) Wie lautet die Schlussfolgerung wenn $\sigma_I = \sigma_M$ sind?

(c) Was ändert sich am Ergebnis, wenn bekannt ist, dass $\sigma_I = 1,41$ und $\sigma_M = 1,23$ ist?

Wählen Sie jeweils ein geeignetes Signifikanzniveau.

2. Augsburg Universität Befragung

Im 2006 hat der Personalrat der Universität Augsburg eine Befragung der Mitarbeiter durchgeführt. 277 haben daran teilgenommen, davon 159 Frauen und 112 Männer. (6 Personen haben ihr Geschlecht nicht preisgegeben.)

Unter anderem sind die Teilnehmer nach ihrem Alter und wie lange sie an der Uni Augsburg gearbeitet haben befragt worden. Für die erste Frage gab es fünf mögliche Antworten (bis 30 Jahre, 31- 40 Jahre, 41 - 50 Jahre, 51 - 60 Jahre, 61 Jahre und mehr), für die zweite vier (bis 2 Jahre, 3-10 Jahre, 11 - 20 Jahre, 21 Jahre und mehr). Wie würden Sie die zwei Variablen gemeinsam graphisch darstellen? Zeichnen Sie ein Beispiel, wie die Daten aussehen könnten. Wäre es sinnvoll hier einen Chiquadrat-Test durchzuführen?

Falls die Teilnehmer um die genauen Angaben gebeten worden wären (d.h. Alter in Jahren, Länge der Berufstätigkeit an der Uni Augsburg in Jahren), wie hätten Sie die Daten dargestellt? Zeichnen Sie ein Beispiel für solche Daten. Wäre es hilfreich einen Korrelationskoeffizient zu berechnen?

3. Kugeln

Es wird oft gesagt, dass geschossene Kugeln identifiziert werden können. Einiges hängt von der Zusammensetzung der Kugeln ab. Konzentrationen von sieben Metallen werden gemessen. Studien zeigen, dass die log(Konzentration) von Kupfer in Kugeln innerhalb einem Produktionslauf als ungefähr normalverteilt betrachtet werden kann.

Nehmen wir an, dass wir eine Beobachtung x haben. Wir glauben apriori mit Wahrscheinlichkeit p_1, dass die Kugel von einem Lauf mit Mittelwert (auf dem Logskala) μ_1 kommt und dass sie sonst von einem Lauf mit Mittelwert μ_2 kommt. Wie sieht unser aposteriori Glauben aus?

Statt nur zwei Möglichkeiten für Mittelwerte einzuräumen, wird eine stetige apriori Verteilung vorgeschlagen. Warum soll vorzugsweise eine Normalverteilung angenommen werden? Skizzieren Sie, wie Sie Ihre Begründung mathematisch belegen würden. Was kann man tun, wenn das Gericht darauf besteht, dass eine andere Form von apriori-Verteilung verwendet werden muß?

(Hintergrund von *Chance* Spring 2006 Data Integrity and the Scientific Method: the Case of Bullet Lead Data as Forensic Evidence by C.H. Spiegelman and K. Kafadar)

4. LAPD

Zwischen Januar und Juni 2005 hat die Los Angeles Police Department (LAPD: Polizei) 5312 weiße Fahrer angehalten und ihre Autos durchsucht. In 3006 Fällen haben sie etwas gefunden. Für schwarze Fahrer waren die entsprechenden Zahlen 12016 und 5134. Sind die Fundraten signifikant unterschiedlich für die zwei Rassen?

Einige Beobachter sagen, dass alle Tests signifikant sind, wenn die Datenmengen groß genug sind. Zwischen Januar und Juni 2004 hat die LAPD 5849 weiße Fahrer angehalten und ihre Autos durchsucht. In 3221 Fällen haben sie etwas gefunden. Sind die Fundraten signifikant unterschiedlich für weiße Fahrer zwischen 2004 und 2005?

Welche Annahmen sind für den Einsatz der Tests notwendig? Sind sie gegeben?

(Daten von *Chance* Spring 2006 Driving While Black in the City of Angels by L.S. Khadjavi)

5. Oktoberfest

Resi und Rosi bedienen auf dem Oktoberfest im Schottenhammel-Zelt. Rosi ist schon seit 20 Jahren dabei und kennt sich bestens aus. Die hübsche Resi ist neu im Geschäft. Jede bedient ausschließlich an ihrem Tisch. An zwölf ausverkauften Abenden wurde für beide gemessen, wie viele Maß sie durchschnittlich pro 15 Minuten serviert haben:

Resi:	9.00	9.50	9.75	10.00	13.00	9.50	(Y)
Rosi:	11.50	12.00	9.00	11.50	13.25	13.00	(X)

Testen Sie mit dem Wilcoxon Rangsummentest, ob Rosi einen Vorteil hat durch ihre langjährige Erfahrung. D.h. für $X = Y + \Delta$ bzw. $F_X(x) = F_Y(x - \Delta)$ mit $\Delta \geq 0$ sei zu testen: $\mathrm{H}_0 : \Delta = 0$ vs. $\mathrm{H}_1 : \Delta > 0$. Als Niveau sei $\alpha = 5\%$ vorgegeben.

Zur Behandlung der auftretenden Bindungen benutzen Sie folgende Verfahren:

(a) Streichen Sie alle relevanten gebundenen Werte.
(b) Vergeben Sie Durchschnittsränge.
(c) Berechnen Sie die für H_0 günstigste, bzw. ungünstigste Zuweisung.

Entscheiden **und** diskutieren Sie jeweils das resultierende Testergebnis!
Wie entscheidet sich der entsprechende parametrische Test?

6. Erwartungstreue Schätzer: Binomial- und Exponential-Verteilung

(a) Für die Binomial-Verteilung $B(1, p)$ ist für die Erfolgswahrscheinlichkeit p bereits der erwartungstreue Schätzer $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ auf Basis einer u.i.v. Stichprobe $X_1, \ldots, X_n$ bekannt. Ein möglicher Schätzer für die Varianz ist $\hat{s} = \hat{p}(1 - \hat{p})$.

  i) Prüfen Sie ob $\hat{s}$ erwartungstreu bzw. asymptotisch erwartungstreu ist.

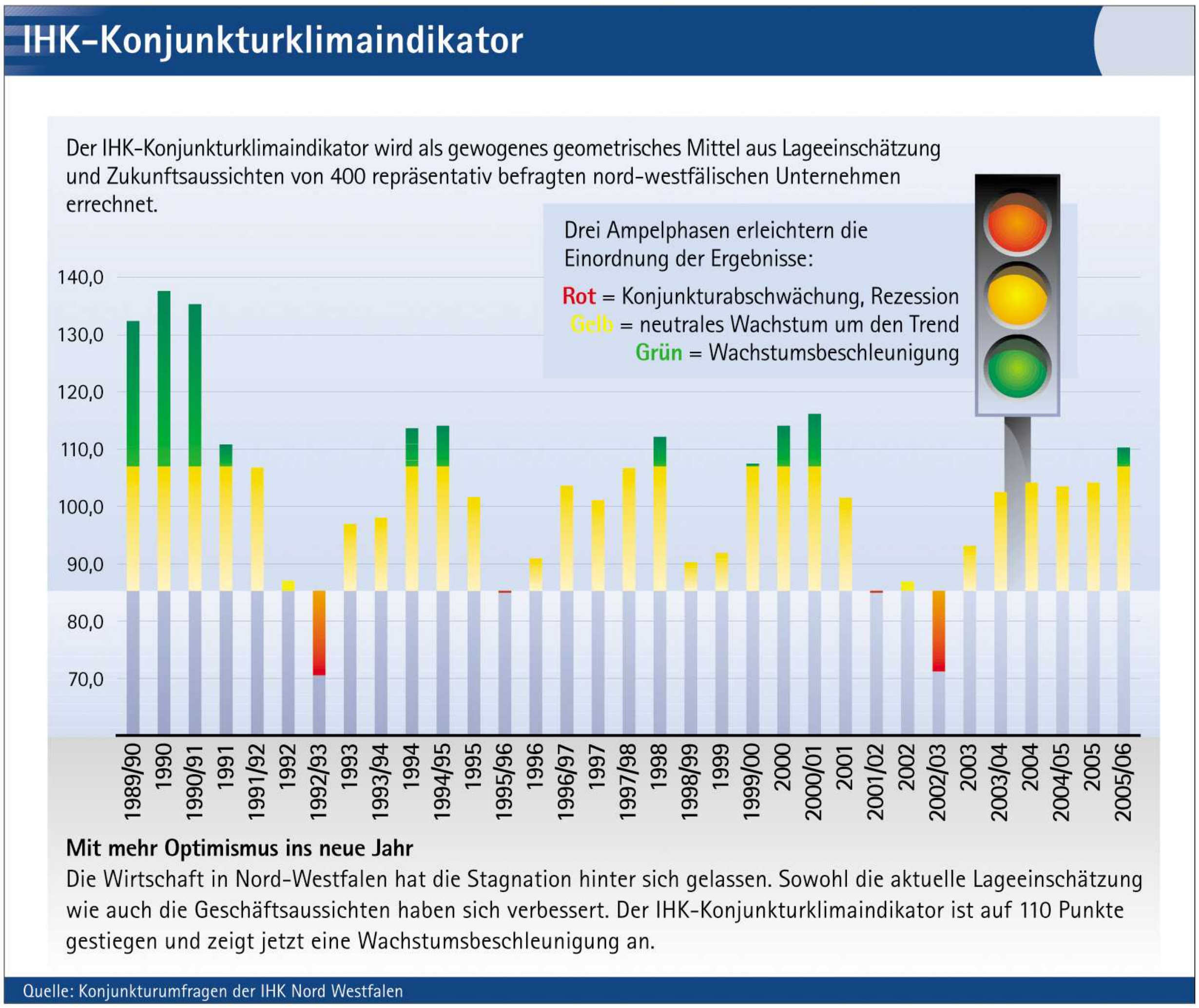
  ii) Wie muss der Schätzer $\hat{s}$ gegebenenfalls zu $\tilde{s}$ modifiziert werden, damit er erwartungstreu bzw. asymptotisch erwartungstreu wird?

(b) Sei nun $X_1, \ldots, X_n$ eine u.i.v Stichprobe aus einer Exponentialverteilung mit einem unbekannten $\lambda > 0$ und der Dichte $f_\lambda(x) = \lambda e^{-\lambda x}$
Für $\frac{1}{\lambda}$ sind die Schätzer $T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ und $T_2 = c_n \min(X_1, \ldots, X_n)$ gegeben.

  i) Ist T_1 erwartungstreu und für welche c_n ist T_2 erwartungstreu? Wie sieht es mit der Konsistenz dieser beiden Schätzer aus?

  ii) Untersuchen Sie mithilfe des MSE welcher dieser beiden Schätzer besser ist.

7. Graphik



Analysieren Sie die obenstehende Grafik.

a) Was genau wird dargestellt?
b) Erläutern Sie die beiden Achsen.
c) Wie beurteilen Sie Farbwahl und ihre Umsetzung?
d) Gibt es Probleme hinsichtlich der Interpretation der Grafik bzw. der eingefärbten Bereiche? Was würden Sie ändern?
e) Beurteilen Sie die Grafik hinsichtlich Eindeutigkeit, Übersichtlichkeit und Ästhetik.
f) Welche andere Darstellungsform der gegebenen Werte würden Sie bevorzugen? Warum? Skizzieren Sie (möglichst genau) einen geeigneten Plot.

8. Zehnkampf

Im Zehnkampf müssen die Athleten zehn Disziplinen in zwei Tagen bestreiten. Für jede Leistung werden Punkte vergeben, um die Resultate aus verschiedenen Disziplinen vergleichbar zu machen. Die ersten zwei Disziplinen am ersten Tag sind der 100m Sprint und der Weitsprung. Man könnte versuchen, die Leistungen der Athleten in der letzten Disziplin des ersten Tages (dem 400m Lauf) aus den ersten zwei mit einem Regressionsmodell zu schätzen.

(a) Wie würden Sie so ein Modell mathematisch formulieren?
(b) Welche Annahmen sind notwendig? Sind sie hier sinnvoll?
(c) Zeigen Sie warum die Kleinstquadratschätzer der Parameter mit dem Maximumlikelihoodschätzer genau übereinstimmen.
(d) Sie könnten entweder beide der ersten Disziplinen in Ihr Modell einbeziehen oder nur eins. Wie würden Sie das entscheiden?
(e) Welche Plots wären Ihnen hilfreich bei Ihren Entscheidungen?