Basisprüfung B. Sc.

Statistik und Wahrscheinlichkeitsrechnung

FS 2010

Prof. Dr. M. H. Faber

ETH Zürich

Mittwoch, 18. August 2010

9:00 – 11:00

Vorname: ....

Name: ....

Stud. Nr.: ....

Studienrichtung: ....

Basisprüfung B. Sc.: Statistik und Wahrscheinlichkeitsrechnung
Bau-, Umwelt- und Geomatikingenieurwissenschaften

Datum und Dauer:

Mittwoch, 18. August 2010
Beginn: 9:00 Uhr
Zeitdauer: 120 Minuten

Hilfsmittel:

- Alle Unterlagen (Skripte, Bücher, Mitschriften, Ausdrucke etc.) sind erlaubt.
- Taschenrechner (ohne Kommunikationsmittel) sind erlaubt.
- Kommunikationsmittel (z.B. Telefon, Laptop) sind nicht erlaubt.

Administratives:

- Bitte kontrollieren Sie zuerst die Vollständigkeit ihrer Unterlagen:
    - Aufgabenstellung inkl. genereller Informationen und Anhängen 12 Seiten.
    - Papierbogen kariert, gestempelt 4 mal.
- Während der 15-minütigen Einlesezeit dürfen die Lösungsbögen nicht beschrieben werden!
- Bitte legen Sie Ihre Legi vor sich auf den Tisch.
- Alle Lösungsblätter müssen mit Namen und Vornamen versehen werden.
- Gewertet werden nur diejenigen Lösungswege und Ergebnisse, welche eindeutig gemäss Aufgabenblatt nummeriert sind und entweder auf die Aufgabenblätter selbst oder auf die karierten, gestempelten Bögen geschrieben werden.
- Nur die zur Verfügung gestellten Blätter dürfen verwendet werden. Verwenden Sie für jede Aufgabe einen neuen Papierbogen.
- Legen Sie am Ende der Prüfung alle Aufgaben- und Lösungsblätter in das Couvert zurück und lassen Sie dieses am Platz liegen.

Hinweise:

- Die Prüfung ist so konzipiert, dass alle Aufgaben gelöst werden sollen.
- Wenn Ihnen für einen Aufgabenteil ein Zwischenresultat fehlt, treffen Sie eine sinnvolle Annahme und markieren Sie diese deutlich. Sie können die Aufgabe mit Ihrer Annahme zu Ende lösen.
- Bitte geben Sie Resultate auf mindestens 3 Nachkommastellen genau an.

Inhalt der Prüfung:

Inhalt	Aufgaben	Seite	Punkte
Aufgabe 1	Beschreibende Statistik, Modellvergleich	3-5	35
Aufgabe 2	Regressionsanalyse, Zuverlässigkeitstheorie	6-7	30
Aufgabe 3	Diskrete Zufallsprozesse	8	25
Aufgabe 4	Entscheidungsanalyse	9-10	30
Anhang	Tabellen	11-12	-

Aufgabe 1: Beschreibende Statistik, Modellvergleich (35 Punkte)

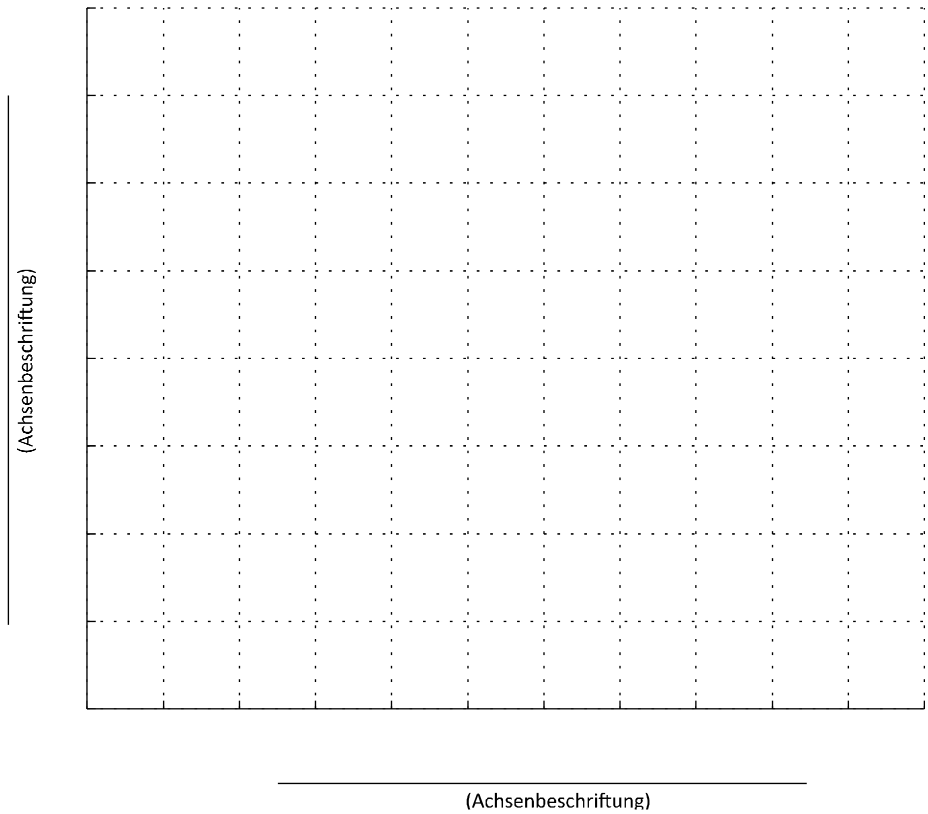
In Tabelle 1.1 sind Beobachtungen der jährlichen Unfallrate und des durchschnittlichen Tagesverkehrs gegeben.

Tabelle 1.1: Unfallrate und durchschnittlicher Tagesverkehr 2005-2009.

Jahr	Unfallrate pro 100 km und pro Jahr (y)	Durchschnittlicher Tagesverkehr (x)	Hilfsspalte	Hilfsspalte
2005	1.88	9582		
2006	1.72	8582		
2007	1.99	9907		
2008	2.19	10951		
2009	2.34	11834		

a) *(4 Punkte)*

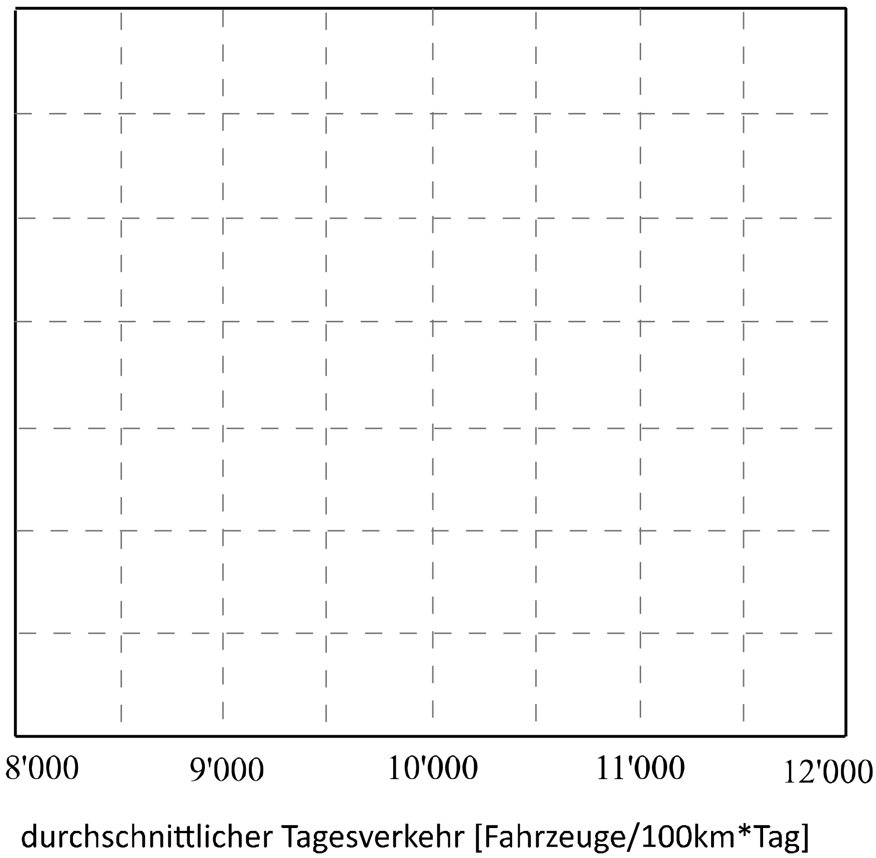
Zeichnen Sie ein zweidimensionales Streudiagramm der jährlichen Unfallrate und des durchschnittlichen Tagesverkehrs aus Tabelle 1.1. Zeichnen Sie zusätzlich die von Ihnen vermutete Lage einer Regressionsgeraden in das Streudiagramm.



b) *(10 Punkte)*

Überprüfen Sie anhand eines Wahrscheinlichkeitspapiers und unter Verwendung der Tabelle 1.1, ob die Beobachtungen des durchschnittlichen Tagesverkehrs (Spalte 3) anhand einer Normalverteilung repräsentiert werden können.

Benutzen Sie für die Lösung dieser Aufgabe den folgenden Vordruck und erstellen Sie eine sinnvolle Skalierung der y-Achse. Eine Tabelle für die Standardnormalverteilung finden Sie im Anhang 1.



c) *(5 Punkte)*

Schätzen Sie mit Hilfe der Methode der Momente die Parameter der Normalverteilung **und** Lognormalverteilung zur Modellierung des durchschnittlichen Tagesverkehrs (Tabelle 1.1, Spalte 3).

ACHTUNG:
Verwenden Sie für die folgenden Teilaufgaben d) und e) die folgenden Parameter für die Lognormalverteilung: $LN(\lambda = 9, \zeta = 0.2)$.

d) *(10 Punkte)*

Überprüfen Sie anhand eines χ^2-Tests auf einem Signifikanzniveau von 10%, ob eine Lognormalverteilung zur Repräsentation der Daten (Tabelle 1.1, Spalte 3) geeignet ist. Die gegebenen Parameter $LN(\lambda = 9, \zeta = 0.2)$ wurden nicht anhand der Daten geschätzt. Verwenden Sie die folgende Tabelle 1.2 mit den vorgegebenen Häufigkeiten in den bereits definierten Intervallen und die Tabelle mit den χ^2-Werten im Anhang 2.

Tabelle 1.2: Chi-Quadrat-Test.

Intervall	Häufigkeit			
0-8000	10			
8000-9000	9			
9000-10‘000	8			
10‘000-11‘000	8			
>11‘000	15			

e) *(5 Punkte)*

Berechnen Sie für den durchschnittlichen Tagesverkehr (Tabelle 1.1, Spalte 3) die Stichproben-Log-Likelihood für eine Lognormalverteilung unter Verwendung der Parameter $LN(\lambda = 9, \zeta = 0.2)$.

f) *(1 Punkt)*

Die Log-Likelihood der Normalverteilung für dieselben Daten beträgt -50.1674.

Für welche Verteilungsfamilie würden Sie sich aufgrund aller vorherigen Berechnungen entscheiden, um die Beobachtungen des durchschnittlichen Tagesverkehrs zu repräsentieren?

Verteilungsfamilie: ________________________________

Aufgabe 2: Regressionsanalyse, Zuverlässigkeitstheorie (30 Punkte)

Sie sollen für einen Autobahnabschnitt eine Unfallrisikoabschätzung durchführen. Dafür möchten Sie ein Modell erstellen, welches die jährliche Unfallrate pro 100 km in Abhängigkeit des durchschnittlichen Tagesverkehrs (DTV) abschätzt. Daten zum DTV und zur jährlichen Unfallrate pro 100 km sind in Tabelle 1.1 gegeben.

Erstellen Sie ein lineares Regressionsmodell, welches die Unfallrate pro km (y) in Abhängigkeit des DTV (x) darstellt. Verwenden Sie hierfür die Werte aus Tabelle 1.1 und folgende Regressionsgleichung: $y = \beta_0 + \beta_1 x + \varepsilon$

a) *(10 Punkte)*

Ermitteln Sie die Regressionskoeffizienten β_0 und β_1.

b) *(4 Punkte)*

Berechnen Sie den Residualwert ε_1 für das erste Datenpaar und geben Sie den Mittelwert der Residualwerte an.

Verwenden Sie für die Teilaufgaben c, d und e die folgenden Werte:

$\beta_0 = 0.01$, $\beta_1 = 0.0003$ und $\sigma_\varepsilon = 0.01$

c) *(4 Punkte)*

Ermitteln Sie die Kovarianzmatrix der Regressionskoeffizienten β_0 und β_1.

Tipp: Die Inverse einer 2x2-Matrix berechnet sich folgendermassen:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

d) *(4 Punkte)*

Der betrachtete Autobahnabschnitt hat eine Länge $L = 10\ km$ und einen durchschnittlicher Tagesverkehr $DTV = 10'000$ Fahrzeuge pro Tag. Verwenden Sie nun das Regressionsmodell, um die durchschnittliche jährliche Unfallrate für den Autobahnabschnitt zu ermitteln.

e) *(8 Punkte)*

Die akzeptierbare Unfallrate R und der durchschnittliche Tagesverkehr DTV sind normalverteilt mit den folgenden Mittelwerten und Standardabweichungen:

Akzeptierbare Unfallrate R	$\mu_R = 1$	$\sigma_R = 0.2$
durchschnittliche Tagesverkehr DTV	$\mu_{DTV} = 10'000$	$\sigma_{DTV} = 2'000$

Verwenden Sie die folgende lineare Grenzzustandsfunktion, um die Wahrscheinlichkeit abzuschätzen, dass die tatsächliche Unfallrate die akzeptierbare übersteigt:

$M = R - L$, wobei $L = \left(\beta_0 + \beta_1 DTV + \varepsilon\right)/10$ und $\sigma_\varepsilon = 0.01$

Aufgabe 3: Diskrete Zufallsprozesse (25 Punkte)

Ein Autobahnabschnitt soll in Bezug auf seine Unfallhäufigkeit bewertet werden.

a) *(7 Punkte)*

Sie möchten die Anzahl Unfälle auf dem Autobahnabschnitt mit einem homogenen Poissonprozess modellieren. Basierend auf Erfahrungswerten von vergleichbaren Strecken nehmen Sie die folgende Intensität an: $\nu = 2 / Monat$.
Berechnen Sie die Wahrscheinlichkeiten der folgenden Ereignisse:

$A1$ Im August 2010 kommt es zu genau 2 Unfällen.

$A2$ In der zweiten Augusthälfte 2010 gibt es mindestens einen Unfall.

$A3$ Die Wartezeit zwischen zwei Unfällen beträgt weniger als einen Monat.

b) *(9 Punkte)*

Die Wahrscheinlichkeit, dass es bei einem Unfall zu Todesopfern kommt, beträgt $p = 0.2$.
Sie nehmen an, dass verschiedene Unfälle voneinander unabhängig sind.
Berechnen Sie die Wahrscheinlichkeiten der folgenden Ereignisse:

$B1$ Bei insgesamt 10 Unfällen gibt es keine Todesopfer.

$B2$ Von insgesamt 10 Unfällen kommt es bei mindestens 2 Unfällen zu Todesopfern.

$B3$ Bis es bei einem Unfall zu Todesopfern kommt, gibt es mindestens 10 Unfälle ohne Todesopfer.

c) *(4 Punkte)*

Berechnen Sie die erwartete Anzahl Unfälle mit Todesopfern während eines Jahres. Verwenden Sie hierzu dieselben Annahmen wie in den vorherigen Teilaufgaben:

Anzahl Unfälle: Homogener Poissonprozess, Intensität $\nu = 2 / Monat$
Bedingte Todesfall-Wahrscheinlichkeit bei einem Unfall: $p = 0.2$

d) *(4 Punkte)*

Auf Basis der Daten aus dem Jahr 2009 in Tabelle 3.1 möchten Sie die Intensität ν des homogenen Poissonprozesses für die Anzahl Unfälle neu schätzen. Verwenden Sie hierfür die Methode der Momente.

Tabelle 3.1: Unfallstatistik 2009.

Monat	Jan.	Feb.	März	April	Mai	Juni	Juli	Aug.	Sep.	Okt.	Nov.	Dez.
Unfälle	7	6	4	3	2	0	1	0	2	5	7	8

e) *(1 Punkt)*
Ist die Annahme eines homogenen Poissonprozess sinnvoll? Begründen Sie Ihre Antwort auf Basis der Daten (1 Satz!).

Aufgabe 4: Entscheidungsanalyse (30 Punkte)

Ein Zementwerk möchte die Anbindung an das öffentliche Strassennetz mit Hilfe einer kleinen Brücke verbessern. Voruntersuchungen haben gezeigt, dass die Brücke zu einer Kostenreduktion von 120‘000 CHF führen würde. Die Baukosten sind abhängig von den Bodenverhältnissen. Bei einem massiven Felsuntergrund kostet die Brücke 100‘000 CHF. Ist der Untergrund nicht massiv (kein Fels), sind die Kosten 150‘000 CHF. Ein Experte schätzt die Wahrscheinlichkeit, dass es sich um einen Felsuntergrund handelt, auf 40%.

a) *(6 Punkte)*

Sollte die Brücke aufgrund der vorhandenen Information, also ohne Untersuchung der Bodenverhältnisse, gebaut werden? Führen Sie eine a-priori-Entscheidungsanalyse durch und benützen Sie hierzu den oberen Teil des Entscheidungsbaumes in Abbildung 4.1.

Durch eine Sondierung kann auf die Bodenverhältnisse geschlossen werden. Diese Inspektion ist allerdings mit Unsicherheiten verbunden, siehe Tabelle 4.1:

Tabelle 4.1: Bedingte Wahrscheinlichkeiten für eine Indikation des Zustands des Untergrundes.

	Tatsächlicher Zustand	
Inspektionsresultat	F Fels	$\overline{F}$ Kein Fels
I (Inspektion indiziert Fels)	$P(I\|F)=0.9$	$P(I\|\overline{F})=$
$\overline{I}$ (Inspektion indiziert kein Fels)	$P(\overline{I}\|F)=$	$P(\overline{I}\|\overline{F})=0.8$

b) *(16 Punkte)*

Berechnen Sie anhand einer prä-posteriori-Entscheidungsanalyse, wie viel eine Sondierung kosten dürfte, damit sie sich lohnt. Ergänzen Sie die fehlenden Angaben zu den Wahrscheinlichkeiten und den Kosten im Entscheidungsbaum in Abbildung 4.1 sowie in Tabelle 4.1.

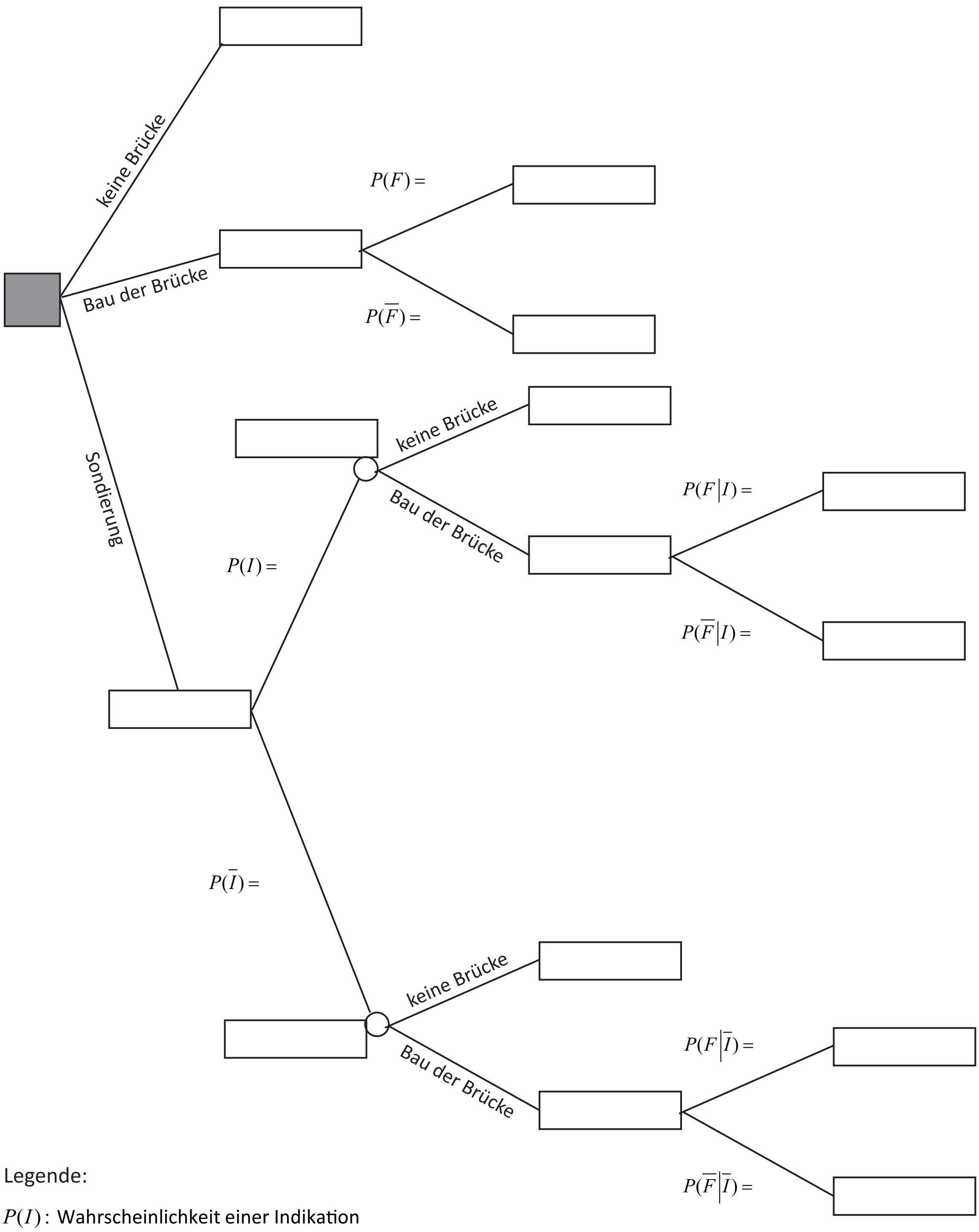
c) *(1 Punkt)*

Wie gross ist die Wahrscheinlichkeit, dass tatsächlich Fels vorliegt, wenn die Inspektion diese Aussage liefert?

d) *(7 Punkte)*

Die Sondierung indiziert Fels. Wie gross müsste die Wahrscheinlichkeit $P(I|F)$ mindestens sein, damit Sie sich bei dieser Indikation für den Bau der Brücke entscheiden würden?

Annahme: $P(I|\overline{F})=0.2$



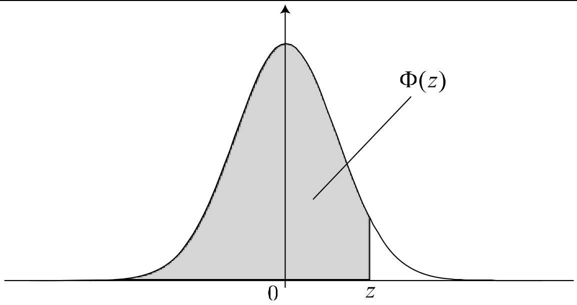
Legende:

$P(I)$: Wahrscheinlichkeit einer Indikation von Fels

$P(F)$: Wahrscheinlichkeit von Fels

Abbildung 4.1: Entscheidungsbaum für die Prä-Posteriori-Entscheidungsanalyse.

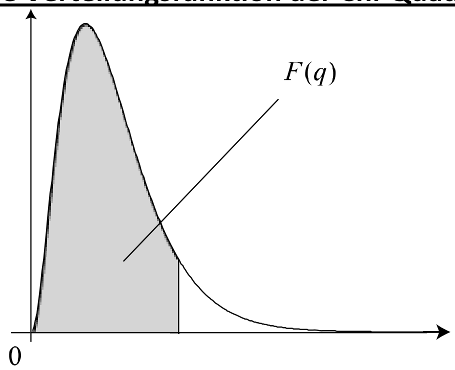
Anhang 1: Kumulative Verteilungsfunktion der Standardnormalverteilung.



Wahrscheinlichkeitsdichtefunktion der Standardnormalverteilung.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.10	0.9821356
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.20	0.9860966
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.30	0.9892759
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.40	0.9918025
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.50	0.9937903
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.60	0.9953388
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.70	0.9965330
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.80	0.9974449
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.90	0.9981342
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	3.00	0.9986501
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	3.10	0.9990324
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	3.20	0.9993129
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	3.30	0.9995166
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	3.40	0.9996631
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	3.50	0.9997674
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	3.60	0.9998409
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	3.70	0.9998922
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	3.80	0.9999277
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	3.90	0.9999519
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	4.00	0.9999683
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	4.10	0.9999793
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	4.20	0.9999867
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	4.30	0.9999915
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	4.40	0.9999946
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	4.50	0.9999966
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	4.60	0.9999979
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	4.70	0.9999987
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	4.80	0.9999992
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	4.90	0.9999995
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	5.00	0.9999997
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649		
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656		
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664		
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671		
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678		
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686		
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693		
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699		
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706		
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713		
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719		
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726		
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732		
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738		
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744		
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750		
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756		
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761		
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767		

Anhang 2: Kumulative Verteilungsfunktion der Chi-Quadrat-Verteilung.



Wahrscheinlichkeitsdichtefunktion für die Chi-Quadrat-Verteilung

v \ $F(q)$	0.75	0.90	0.95	0.98	0.99	0.995	0.999
1	1.3233	2.7055	3.8415	5.4119	6.6349	7.8794	10.8276
2	2.7726	4.6052	5.9915	7.8240	9.2103	10.5966	13.8155
3	4.1083	6.2514	7.8147	9.8374	11.3449	12.8382	16.2662
4	5.3853	7.7794	9.4877	11.6678	13.2767	14.8603	18.4668
5	6.6257	9.2364	11.0705	13.3882	15.0863	16.7496	20.5150
6	7.8408	10.6446	12.5916	15.0332	16.8119	18.5476	22.4577
7	9.0371	12.0170	14.0671	16.6224	18.4753	20.2777	24.3219
8	10.2189	13.3616	15.5073	18.1682	20.0902	21.9550	26.1245
9	11.3888	14.6837	16.9190	19.6790	21.6660	23.5894	27.8772
10	12.5489	15.9872	18.3070	21.1608	23.2093	25.1882	29.5883
11	13.7007	17.2750	19.6751	22.6179	24.7250	26.7568	31.2641
12	14.8454	18.5493	21.0261	24.0540	26.2170	28.2995	32.9095
13	15.9839	19.8119	22.3620	25.4715	27.6882	29.8195	34.5282
14	17.1169	21.0641	23.6848	26.8728	29.1412	31.3193	36.1233
15	18.2451	22.3071	24.9958	28.2595	30.5779	32.8013	37.6973
16	19.3689	23.5418	26.2962	29.6332	31.9999	34.2672	39.2524
17	20.4887	24.7690	27.5871	30.9950	33.4087	35.7185	40.7902
18	21.6049	25.9894	28.8693	32.3462	34.8053	37.1565	42.3124
19	22.7178	27.2036	30.1435	33.6874	36.1909	38.5823	43.8202
20	23.8277	28.4120	31.4104	35.0196	37.5662	39.9968	45.3147
21	24.9348	29.6151	32.6706	36.3434	38.9322	41.4011	46.7970
22	26.0393	30.8133	33.9244	37.6595	40.2894	42.7957	48.2679
23	27.1413	32.0069	35.1725	38.9683	41.6384	44.1813	49.7282
24	28.2412	33.1962	36.4150	40.2704	42.9798	45.5585	51.1786
25	29.3389	34.3816	37.6525	41.5661	44.3141	46.9279	52.6197
26	30.4346	35.5632	38.8851	42.8558	45.6417	48.2899	54.0520
27	31.5284	36.7412	40.1133	44.1400	46.9629	49.6449	55.4760
28	32.6205	37.9159	41.3371	45.4188	48.2782	50.9934	56.8923
29	33.7109	39.0875	42.5570	46.6927	49.5879	52.3356	58.3012
30	34.7997	40.2560	43.7730	47.9618	50.8922	53.6720	59.7031

v: Freiheitsgrade