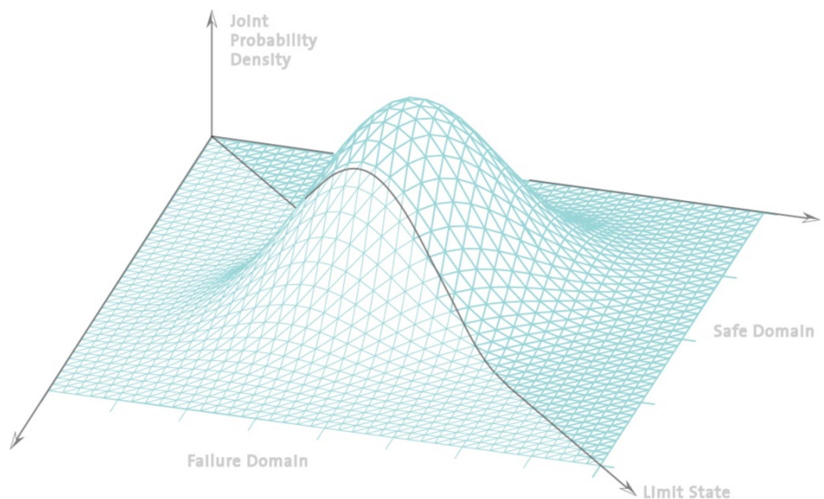


Statistik und Wahrscheinlichkeitsrechnung

Dr. Jochen Köhler

Inhalt der heutigen Vorlesung

- Kurze Zusammenfassung der letzten Vorlesung
- Schätzung und Modellentwicklung – Überblick
- Statistische Signifikanztests
  - Vorgehen beim Hypothesentest
  - Mittelwerttest bei bekannter Varianz
  - Mittelwerttest bei unbekannter Varianz
  - Varianztest
  - Test von zwei oder mehreren Datensätzen
- Wahl der Verteilungsfunktion
  - Modellwahl anhand von Wahrscheinlichkeitspapier

Kurze Zusammenfassung der letzten Vorlesung

Daten / Beobachtungen

n	x_n	$F_X(x_n)$
1	24.4	0.047619048
2	27.6	0.095238095
3	27.8	0.142857143
4	27.9	0.19047619
5	28.5	0.238095238
6	30.1	0.285714286
7	30.3	0.333333333
8	31.7	0.380952381
9	32.2	0.428571429
10	32.8	0.476190476
11	33.3	0.523809524
12	33.5	0.571428571
13	34.1	0.619047619
14	34.6	0.666666667
15	35.8	0.714285714
16	35.9	0.761904762
17	36.8	0.80952381
18	37.1	0.857142857
19	39.2	0.904761905
20	39.7	0.952380952

→ Mittelwert

→ Varianz

→ Median

⋮

→ etc.

Stichproben-
charakteristik

oder

Stichprobenstatistik

Stichprobenstatistiken beschreiben die Stichproben

Kurze Zusammenfassung der letzten Vorlesung

Stichprobenstatistiken sind zum Beispiel:

Stichprobenmittelwert $\bar{X}$

Stichprobenvarianz S^2

Stichprobenstatistiken sind Variabel!

Die Variabilität hängt ab vom Umfang der Stichprobe und von der Variabilität der zugrundeliegenden Zufallsvariablen.

$$E\left[\bar{X}\right] = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad E\left[S^2_{unbiased}\right] = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$Var\left[\bar{X}\right] = \frac{1}{n}\sigma_X^2 \qquad Var\left[S^2_{unbiased}\right] = \frac{2(n-1)}{n^2}\sigma_X^4$$

Kurze Zusammenfassung der letzten Vorlesung

Konfidenzintervall:

Z.B. in was für einem Bereich liegt der wahre Mittelwert mit einer Konfidenz von 95%? (bei gegebener Varianz von $\sigma_X^2 = 9$)

Schrittweises Vorgehen:

1. Identifikation der Entsprechenden Fraktile der Standardnormalverteilung.

$$95\% \text{ Konfidenz} \;\Rightarrow\; 1-\alpha = 0.95 \;\Rightarrow\; \alpha = 0.05 \;\Rightarrow\; \alpha/2 = 0.025$$

Kurze Zusammenfassung der letzten Vorlesung

Konfidenzintervall:

Z.B. in was für einem Bereich liegt der Wahre Mittelwert mit einer Konfidenz von 95%? (bei gegebener Varianz von $\sigma_X^2 = 9$)

Schrittweises Vorgehen:

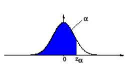
2. Ablesen der Fraktilwerte.

Quantile z_α der Standardnormalverteilung $N(0,1)$

Ablesebeispiel: $z_{0.95} = 1.6449$.

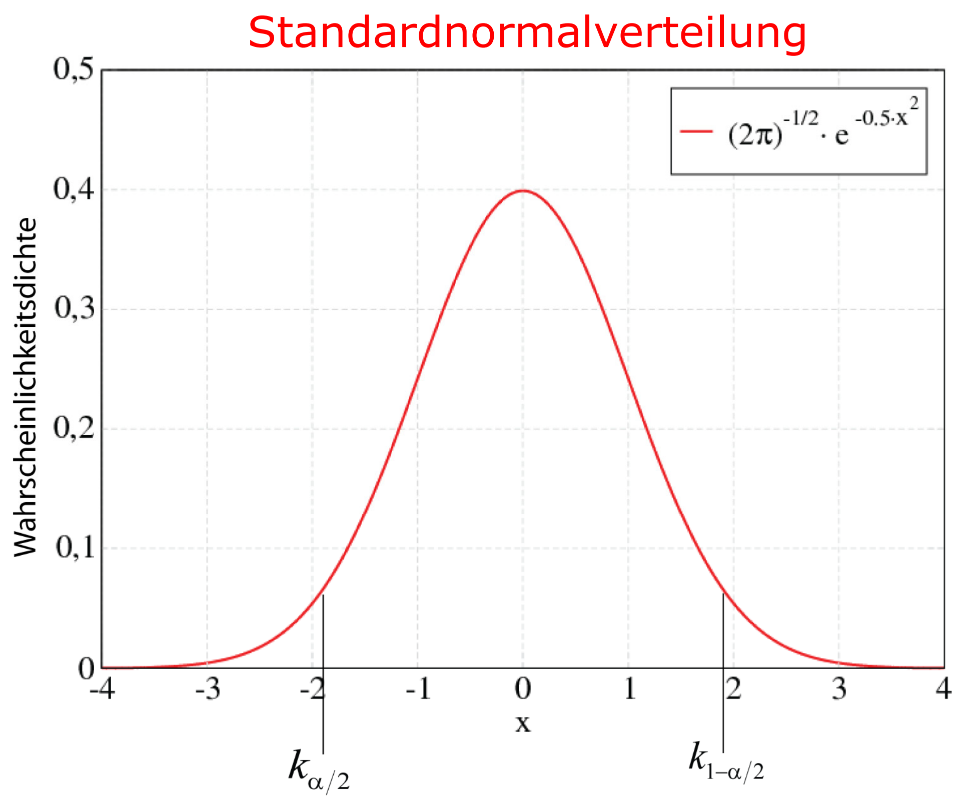
Erweiterung der Tafel: $z_{1-\alpha} = -z_\alpha$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} \mathrm{d}z \quad \text{Verteilungsfunktion}$$



α	z_α	α	z_α	α	z_α	α	z_α
0.9999	3.7190	0.9955	2.6121	0.975	1.9600	0.780	0.7722
0.9998	3.5401	0.9950	2.5758	0.970	1.8808	0.770	0.7388
0.9997	3.4316	0.9945	2.5427	0.965	1.8119	0.760	0.7063
0.9996	3.3528	0.9940	2.5121	0.960	1.7507	0.750	0.6745
0.9995	3.2905	0.9935	2.4838	0.955	1.6954	0.740	0.6433
0.9994	3.2389	0.9930	2.4573	0.950	1.6449	0.730	0.6128
0.9993	3.1946	0.9925	2.4324	0.945	1.5982	0.720	0.5828
0.9992	3.1559	0.9920	2.4089	0.940	1.5548	0.710	0.5534
0.9991	3.1214	0.9915	2.3867	0.935	1.5141	0.700	0.5244
0.9990	3.0902	0.9910	2.3656	0.930	1.4758	0.690	0.4959
0.9989	3.0618	0.9905	2.3455	0.925	1.4395	0.680	0.4677
0.9988	3.0357	0.9900	2.3263	0.920	1.4051	0.670	0.4399
0.9987	3.0115	0.9895	2.3080	0.915	1.3722	0.660	0.4125
0.9986	2.9889	0.9890	2.2904	0.910	1.3408	0.650	0.3853
0.9985	2.9677	0.9885	2.2734	0.905	1.3106	0.640	0.3585
0.9984	2.9478	0.9880	2.2571	0.900	1.2816	0.630	0.3319
0.9983	2.9290	0.9875	2.2414	0.895	1.2536	0.620	0.3055
0.9982	2.9112	0.9870	2.2262	0.890	1.2265	0.610	0.2793
0.9981	2.8943	0.9865	2.2115	0.885	1.2004	0.600	0.2533
0.9980	2.8782	0.9860	2.1973	0.880	1.1750	0.590	0.2275
0.9979	2.8627	0.9855	2.1835	0.875	1.1503	0.580	0.2019
0.9978	2.8480	0.9850	2.1701	0.870	1.1264	0.570	0.1764
0.9977	2.8338	0.9845	2.1571	0.865	1.1031	0.560	0.1510
0.9976	2.8202	0.9840	2.1444	0.860	1.0803	0.550	0.1257
0.9975	2.8070	0.9835	2.1321	0.855	1.0581	0.540	0.1004
0.9974	2.7944	0.9830	2.1201	0.850	1.0364	0.530	0.0753
0.9973	2.7821	0.9825	2.1084	0.845	1.0152	0.520	0.0502
0.9972	2.7703	0.9820	2.0969	0.840	0.9945	0.510	0.0251
0.9971	2.7589	0.9815	2.0858	0.835	0.9741	0.500	0.0000
0.9970	2.7478	0.9810	2.0749	0.830	0.9542		
0.9969	2.7370	0.9805	2.0642	0.825	0.9346		
0.9968	2.7266	0.9800	2.0537	0.820	0.9154		
0.9967	2.7164	0.9795	2.0435	0.815	0.8965		
0.9966	2.7065	0.9790	2.0335	0.810	0.8779		
0.9965	2.6968	0.9785	2.0237	0.805	0.8596		
0.9964	2.6874	0.9780	2.0141	0.800	0.8416		
0.9963	2.6783	0.9775	2.0047	0.795	0.8239		
0.9962	2.6693	0.9770	1.9954	0.790	0.8064		
0.9961	2.6606	0.9765	1.9863	0.785	0.7892		
0.9960	2.6521	0.9760	1.9774	0.780	0.7722		

$$k_{\alpha/2} = -k_{1-\alpha/2}$$



Kurze Zusammenfassung der letzten Vorlesung

Konfidenzintervall:

Z.B. in was für einem Bereich liegt der Wahre Mittelwert mit einer Konfidenz von 95%? (bei gegebener Varianz von $\sigma_X^2 = 9$)

Schrittweises Vorgehen:

3. Intervall skalieren mit der Standardabweichung des Mittelwertes: $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$

$$\Rightarrow \mu_X = \bar{X} \pm k_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

4. Intervall verschieben um die Realisation von $\bar{X}$: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i$

$$\Rightarrow \mu_X = \bar{x} \pm k_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

Kurze Zusammenfassung der letzten Vorlesung

Konfidenzintervall:

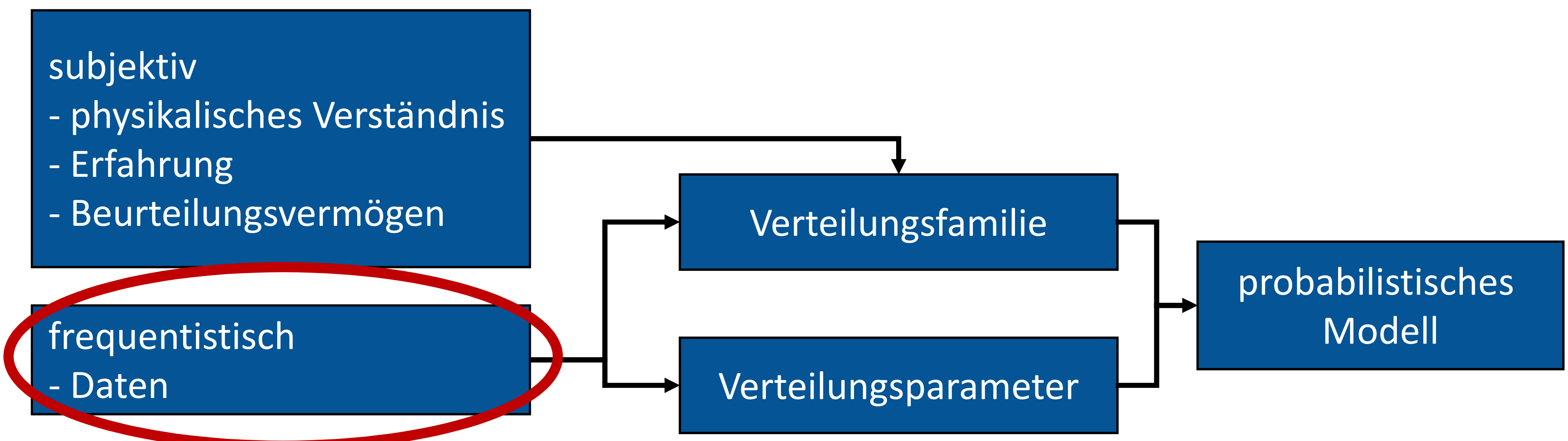
Typische Konfidenzintervalle und die dazugehörigen Fraktilwerte der Standardnormalverteilung:

Konfidenz	α	$k_{\alpha/2}$	$k_{1-\alpha/2}$
99%	0.01	-2.586	2.586
95%	0.05	-1.960	1.960
90%	0.1	-1.645	1.645

Schätzung und Modellentwicklung - Übersicht

Wenn man Modelle im Ingenieurbereich entwickeln möchte, müssen unterschiedliche Typen von Informationen herangezogen werden.

- subjektive Informationen
- frequentistische Informationen



statistische Signifikanztests

Bei Konfidenzintervallen haben wir die Unsicherheit einer Stichprobenstatistik ausgedrückt.

Statistische Signifikanztest werden durchgeführt um Daten (repräsentiert durch Stichprobenstatistiken) und bestehende Modelle zu vergleichen.

Beispiele:

Es sollen ein paar "vor Ort" – Tests durchgeführt werden, um ein Modell für die Bodenfestigkeiten zu verifizieren.

Das Verkehrsaufkommen auf einer Brücke soll beobachtet werden, um zu überprüfen, ob die vorherigen Annahmen diesbezüglich zutreffend sind.

Grundwasserproben sollen gezogen werden, um die Trinkwasserqualität zu bestätigen.

statistische Signifikanztests

Es ist wichtig, dass diese Rückschlüsse anhand konsistenter und transparenter Grundlagen gezogen werden. So sollten die Rückschlüsse alle Beweismittel (Daten) berücksichtigen und nach vergleichbaren Kriterien durchgeführt werden.

Ein häufig angewandter und hilfreicher Weg, um solche Rückschlüsse zu unterstützen ist folgender:

- Formulieren einer Hypothese
- Hypothese Testen
- Die Hypothese wird akzeptiert oder verworfen (mit einer gewissen Signifikanz)

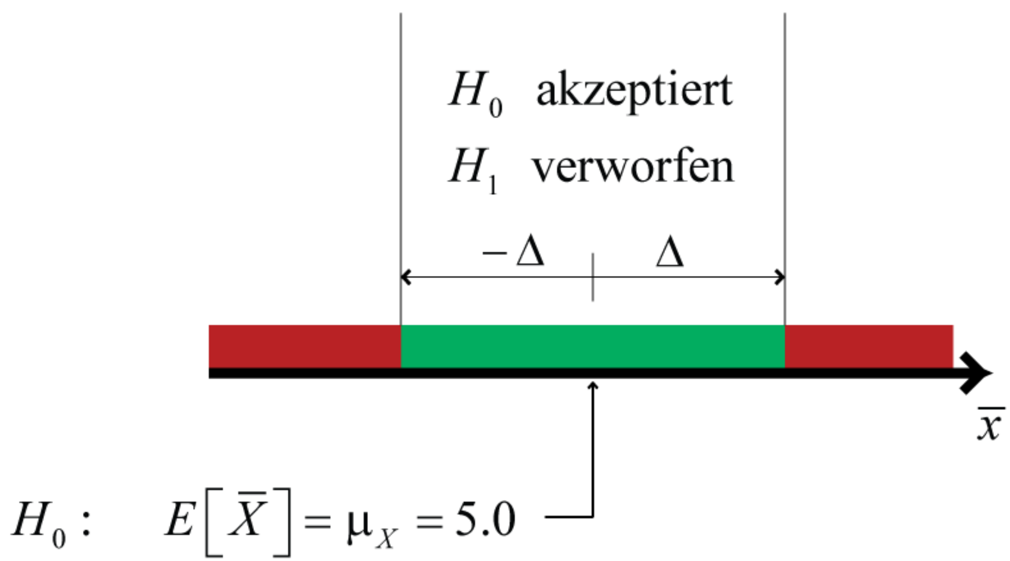
Im Folgenden wird ein genauer Blick auf diesen Ansatz geworfen…

statistische Signifikanztests

1. Der erste Schritt liegt in der Formulierung einer Null-Hypothese H_0.
   Z.B. man legt fest, dass ein Erwartungswert einer Stichprobenstatistik einem bestimmten Wert entspricht.

$$H_0 : \quad E\left[\overline{X}\right] = \mu_X = 5.0$$

2. Der Folgeschritt beinhaltet die Formulierung einer Vorgehensregel, nach der die Null-Hypothese entweder akzeptiert oder abgelehnt wird.
   Die Vorgehensregel muss den Stichprobenraum der Stichprobenstatistik in zwei sich gegenseitig ausschliessende Bereiche unterteilen.
   Eine Ablehnung der Null-Hypothese beinhaltet die Akzeptanz der Ersatzhypothese H_1.



statistische Signifikanztests

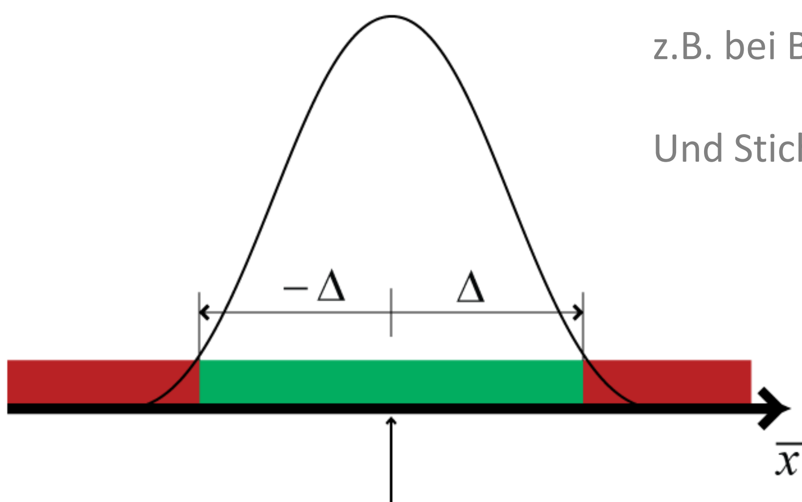
3. Nun wählt man ein Signifikanzniveau für die Testdurchführung, wobei α die Wahrscheinlichkeit ist, dass die Hypothese abgelehnt wird, obwohl sie zutrifft (Fehler 1. Art). Auf diesem Weg beeinfluss α ebenso die Wahrscheinlichkeit, dass die Null-Hypothese akzeptiert wird, obwohl sie nicht zutrifft (Fehler 2. Art).

4. Nun muss man den zu α gehörigen Wert von Δ berechnen.

statistische Signifikanztests

3. Nun wählt man ein Signifikanzniveau für die Testdurchführung, wobei α die Wahrscheinlichkeit ist, dass die Hypothese abgelehnt wird, obwohl sie zutrifft (Fehler 1. Art). Auf diesem Weg beeinfluss α ebenso die Wahrscheinlichkeit, dass die Null-Hypothese akzeptiert wird, obwohl sie nicht zutrifft (Fehler 2. Art).

4. Nun muss man den zu α gehörigen Wert von Δ berechnen.



$H_0: \quad E\left[\bar{X}\right] = \mu_X = 5.0$

z.B. bei Bekannter Standardabweichung σ_X

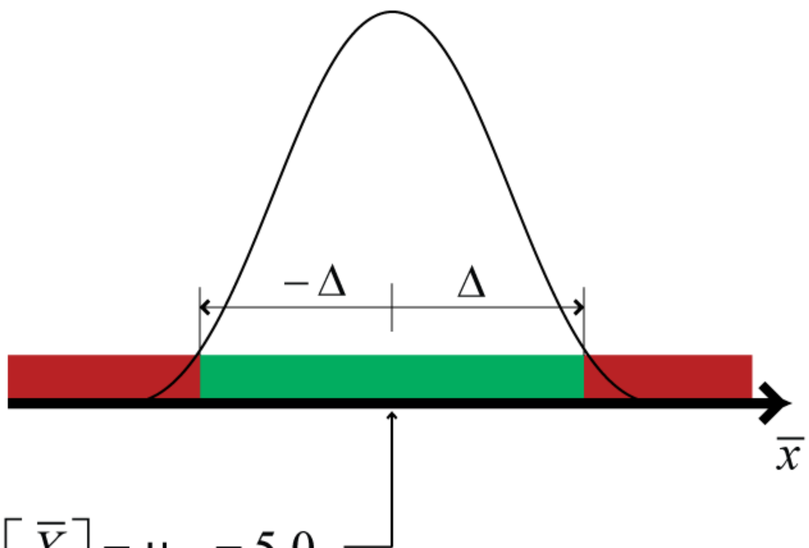
Und Stichprobenumfang n

$$\Delta = -k_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$$

statistische Signifikanztests

5. Die geplanten Tests werden durchgeführt und die beobachteten Stichprobenstatistiken beurteilt. Eine Ablehnung oder Akzeptanz der Null-Hypothese wird geprüft.

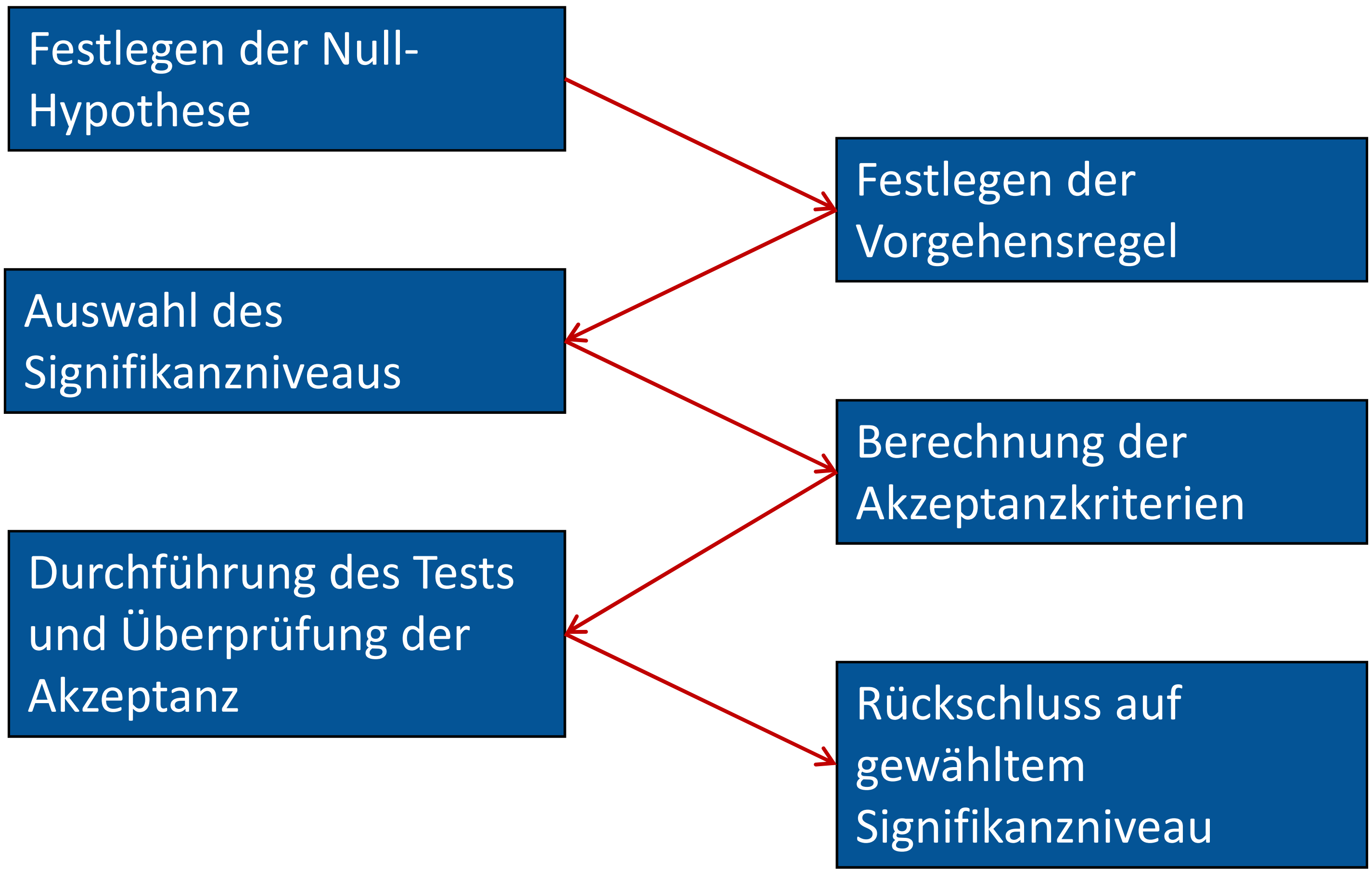
6. Sollte die Null-Hypothese von den Ergebnissen der Stichprobe nicht gestützt werden, wird sie auf dem Signifikantniveau von α abgelehnt – sonst wird sie akzeptiert.



$$H_0: \quad E\left[\bar{X}\right] = \mu_X = 5.0$$

statistische Signifikanztests

Grafische Darstellung des Vorgehens bei Hypothesentests:



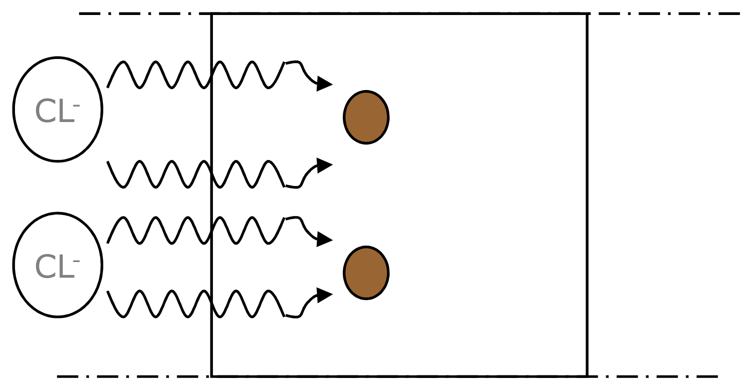
statistische Signifikanztests

Typische Anwendungen im Ingenieurbereich:

- Testen des Mittelwertes – mit bekannter Varianz
- Testen des Mittelwertes – mit unbekannter Varianz
- Testen der Varianz
- Testen zweier oder mehrerer Datensätze

statistische Signifikanztests

Beispiel: Chlorid bedingte Korrosion an Betonkonstruktionen



Man betrachte beispielsweise die Situation, in der beurteilt werden soll, ob die Chlorid-Konzentration an der Oberfläche einer Betonkonstruktion mit den vorher angenommenen Bemessungsgrundlagen übereinstimmt.

statistische Signifikanztests

Testen des Mittelwertes – mit bekannter Varianz

Bemessungsgrundlage: Mittelwert der Oberflächen-Chlorid-Konzentration: 0,3%

Null-Hypothese

Die Standardabweichung der Oberflächen-Chlorid-Konzentration sei bekannt und entspricht 0,04%.

Festlegung des Annahmebereichs:

Die Null-Hypothese wird auf dem Signifikanzniveau α akzeptiert, falls

$$0.3 - \Delta \leq \bar{X} \leq 0.3 + \Delta$$

statistische Signifikanztests

Testen des Mittelwertes – mit bekannter Varianz

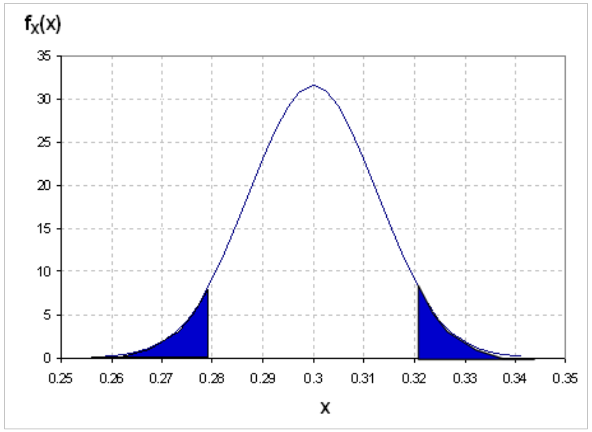
Das Akzeptanzkriterium bei gegebenem α kann wie folgt berechnet werden:

$$P\left(0.3-\Delta \leq \bar{X} \leq 0.3+\Delta\right)=1-\alpha$$

Wählt man $\alpha = 0.1$ und $n=10$ Versuche und nimmt man an, dass der Stichprobenmittelwert normalverteilt ist, erhält man:

$$\Delta = -k_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} = 1.645\frac{0.04}{\sqrt{10}} = 0.0208$$

$$\Rightarrow \qquad P\left(0.2792 \leq \bar{X} \leq 0.3208\right) = 0.9$$



statistische Signifikanztests

Testen des Mittelwertes – mit bekannter Varianz

Liegt die Stichprobenvarianz innerhalb von $\left[0.2792 \leq \bar{x} \leq 0.3208\right]$, sollte die Null-Hypothese H_0 akzeptiert werden.

Nimmt man an, dass 10 Versuche durchgeführt wurden und die folgenden Ergebnisse daraus entstanden:

$$\mathbf{x} = (0.33, 0.32, 0.25, 0.31, 0.28, 0.27, 0.29, 0.3, 0.27, 0.28)^T$$

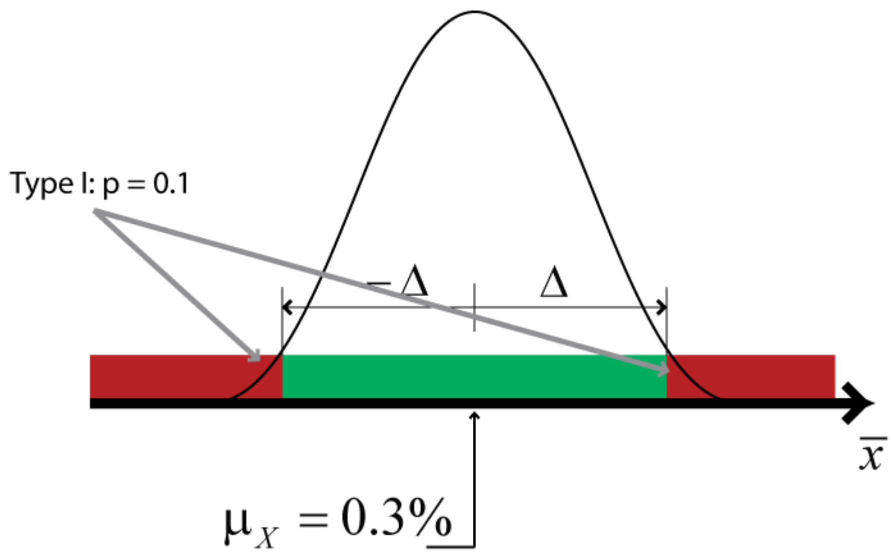
Bei einem Stichprobenmittelwert von $\bar{x} = 0.29$, kommt man zu dem Schluss, dass die Null-Hypothese auf einem Signifikanzniveau von 0,1 akzeptiert werden sollte.

statistische Signifikanztests

Testen des Mittelwertes – mit bekannter Varianz

Bei diesem Beispiel beträgt die Wahrscheinlichkeit einen Typ I Fehler zu begehen 0.1 %.

Aber was ist die Wahrscheinlichkeit die Null Hypothese fälschlicherweise zu akzeptieren ???

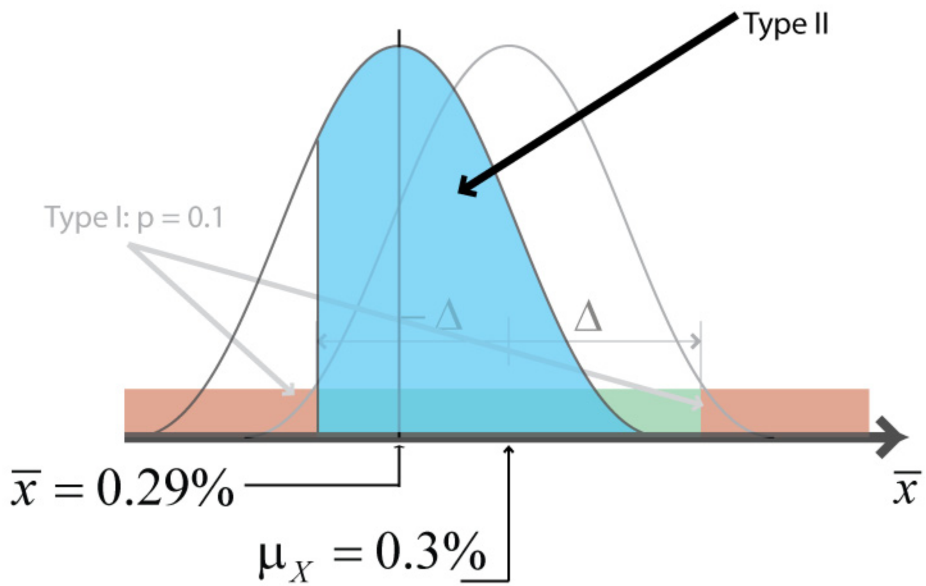


statistische Signifikanztests

Testen des Mittelwertes – mit bekannter Varianz

Bei diesem Beispiel beträgt die Wahrscheinlichkeit einen Typ I Fehler zu begehen 0.1 %.

Aber was ist die Wahrscheinlichkeit die Null Hypothese fälschlicherweise zu akzeptieren ???



statistische Signifikanztests

Testen des Mittelwertes – mit unbekannter Varianz

Geht man nun davon aus, dass die Varianz nicht bekannt ist, muss die folgende Stichprobenstatistik berücksichtigt werden

$$T = \frac{\bar{X} - \mu}{\frac{S_{unbiased}}{\sqrt{n}}}$$

T ist t-verteilt mit einem Freiheitsgrad von $n\text{-}1$.

Die Vorgehensregel wäre dem entsprechend: $P(-\Delta \leq T \leq \Delta) = 1 - \alpha$

statistische Signifikanztests

Testen des Mittelwertes – mit unbekannter Varianz

Mit den gleichen Versuchsergebnissen wie zuvor erhält man nun den gleichen Stichprobenmittelwert, und die Varianz ist:

$$s_{unbiased} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2} = 0.025$$

Für Δ ergibt sich : $\Delta = -t_{\alpha/2,n-1} \frac{s_{unbiased}}{\sqrt{n}} = 1.83 \frac{0.025}{\sqrt{10}} = 0.015$

$t_{\alpha/2,n-1}$ ist der α/2-Fraktilwert der t-Verteilung mit 9 Freiheitsgraden.

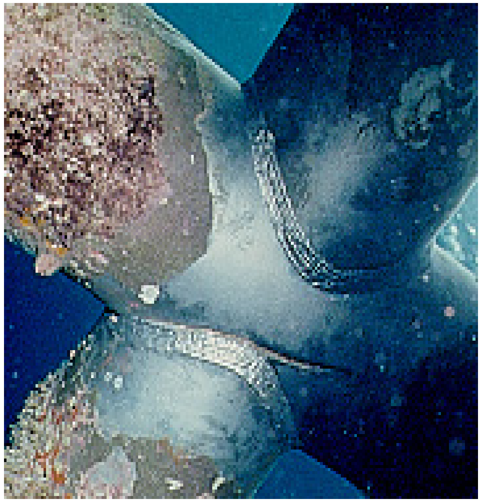
$$\left[0.285 \le \bar{x} \le 0.315\right]$$

Folglich sollte die Null-Hypothese nicht abgelehnt werden($\bar{x} = 0.29$).

statistische Signifikanztests

Testen der Varianz

Man betrachte beispielsweise den Fall, bei dem versucht wird, die Varianz der Lebensdauern von geschweißten Verbindungen zu reduzieren, indem die Schweißnahtoberfläche behandelt wird.



Weil Experimente sehr teuer sind, stehen nur wenige Daten zur Verfügung, um den Effekt der Schweißnahtbehandlung zu bestätigen.

statistische Signifikanztests

Testen der Varianz

Als Null-Hypothese nehmen wir an, dass die Varianz der Lebensdauern der Schweißnähte mit einer Oberflächenbehandlung geringer ist, als diejenige ohne eine Behandlung: $\sigma^2_{neu} \leq \sigma^2_{alt}$

Die Vorgehensregel besagt demzufolge, dass die Null-Hypothese akzeptiert wird, wenn $P\left[V_{test} \leq \Delta\right] = 1-\alpha$

In der Regel ist S^2 Chi-Quadrat verteilt mit n bzw. n-1 Freiheitsgraden.

$$V_{test} = \frac{\sum_{i=1}^{n}\left(x_i - \mu\right)^2}{\sigma_0^2} \rightarrow$$ Wenn μ bekannt: $$\Delta = \chi^2_n\left(1-\alpha\right)$$

$$V_{test} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{\sigma_0^2} \rightarrow$$ Wenn μ unbekannt: $$\Delta = \chi^2_{n-1}\left(1-\alpha\right)$$

statistische Signifikanztests

Testen von mehr als einem Datensatz

Typischer Weise befinden wir uns in einer Situation, in der uns mehr als nur ein Datensatz vorliegen, von denen jeder nicht sonderlich umfangreich ist. Dabei entsteht die Frage, wie man die Daten hinsichtlich folgender Grössen vergleichen kann:

- Mittelwerte — Test für gleiche Mittelwerte
- Varianzen — Test für gleiche Varianzen
- Korrelation — Test für Null-Korrelation

statistische Signifikanztests

Testen für gleiche Mittelwerte (bei bekannter Varianz)

Wir nehmen an, dass uns zwei Datensätze vorliegen:

$$\mathbf{x} = \left(x_1, x_2, .., x_k\right)^T \qquad \text{und} \qquad \mathbf{y} = \left(y_1, y_2, .., y_l\right)^T$$

Diese sind Realisationen der Zufallsvariablen X und Y, die beide als normalverteilt gelten, mit den Mittelwerten μ_x, μ_y und den Varianzen σ_x, σ_y.

Die Statistik $T = \overline{X} - \overline{Y}$ ist normalverteilt mit einem Mittelwert von:

$$\mu_{\overline{X}-\overline{Y}} = \mu_X - \mu_Y$$

und einer Varianz von:

$$\sigma^2_{\overline{X}-\overline{Y}} = \frac{\sigma_X^2}{k} + \frac{\sigma_Y^2}{l}$$

statistische Signifikanztests

Testen für gleiche Mittelwerte (bei bekannter Varianz)

Wenn α gleich 0.1 ist, kann Δ wie folgt berechnet werden:

$$P\left(\bar{X} - \bar{Y} \leq \Delta\right) = 1 - \alpha \qquad \Delta = 1.28\sqrt{\frac{\sigma_X^2}{k} + \frac{\sigma_Y^2}{l}}$$

statistische Signifikanztests

Testen für gleiche Mittelwerte (bei bekannter Varianz)

$$P\left(\overline{X}-\overline{Y}\leq\Delta\right)=1-\alpha \qquad \Delta=-k_{\alpha}\sqrt{\frac{\sigma_X^2}{k}+\frac{\sigma_Y^2}{l}}$$

Einseitig

statistische Signifikanztests

Testen für gleiche Varianzen

Für gleiche Varianzen kann auf folgende Weise ein Test durchgeführt werden:

$$T = \frac{S^2_{X,unbiased}}{S^2_{Y,unbiased}}$$

was als Verhältnis von Chi-Quadrat verteilten Zufallsvariablen verstanden werden kann. Daraus leitet sich ab, dass T **F-verteilt** ist mit den entsprechenden Parametern k und l.

Die Null-Hypothese H_0 ist: $\sigma_X^2 = \sigma_Y^2$

Die Vorgehensregel, um H_0 zu akzeptieren, ist: $T \leq \Delta$

wobei Δ aus $P(T \leq \Delta) = 1 - \alpha$ berechnet wird -> $\Delta = F_{\alpha,k-1,l-1}$

statistische Signifikanztests

Besondere Bemerkungen:

Statistische Signifikanztests können für eine Vielzahl unterschiedlicher Problemstellungen formuliert werden.

Man sollte aufpassen, dass man die Aussage dieser Tests nicht überschätzt, da die Hypothesen auf unterschiedlichen Wegen und mit unterschiedlichen Signifikanzniveaus formuliert werden können. Es sollte deshalb zum Prinzip werden, alles zu überprüfen.

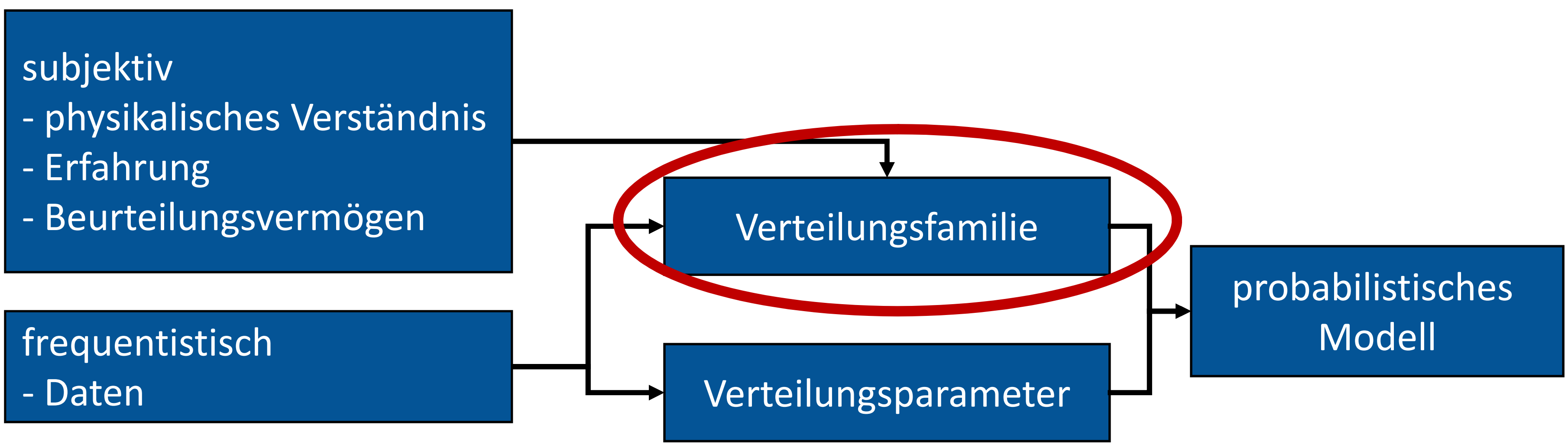
Unterschiedliche Herangehensweisen verursachen einen direkten Effekt auf die Wahrscheinlichkeit, dass Fehler der 1. oder 2. Art entstehen. Dies kann signifikante ökonomische Konsequenzen nach sich ziehen.

Die Formulierung der Hypothese und die Wahl des Signifikanzniveaus sollte als Entscheidungsproblem behandelt werden – dazu später mehr…

Schätzung und Modellentwicklung - Übersicht

Wenn man Modelle im Ingenieurbereich entwickeln möchte, müssen unterschiedliche Typen von Informationen herangezogen werden.

- subjektive Informationen
- frequentistische Informationen



Schätzung und Modellentwicklung

Auswahl von Wahrscheinlichkeitsverteilungsfunktionen

Grundsätzlich müssen Verteilungsfunktionen für Zufallsvariablen oder –Prozessen auf Basis folgender Punkte ausgewählt werden:

Frequentistische Information:	Daten
Physikalische Argumente:	Verständnis Ingenieurproblemstellungen

Der klassische Ansatz ist folgender:

1. Bestimmen einer Hypothese für eine Wahrscheinlichkeitsverteilungsfamilie.
2. Schätzen der Funktionsparameter der bestimmten Wahrscheinlichkeitsverteilung.
3. Durchführen eines statistischen Tests, um die Hypothese abzulehnen oder zu akzeptieren.

Schätzung und Modellentwicklung

Auswahl von Wahrscheinlichkeitsverteilungsfunktionen

Im Ingenieurwesen tritt häufig der Fall ein, dass die verfügbaren Daten zu spärlich sind, um einen Hypothesentest für eine gegebene Wahrscheinlichkeitsverteilung durchzuführen – zumindest mit einer vernünftigen Signifikanz.

Deshalb ist ein einheitliches Vorgehen sehr wichtig:

Zuerst werden physikalische Argumente herangezogen, um eine passende Verteilung zu identifizieren.

Darauf aufbauend wird überprüft, ob die zur Verfügung stehenden Daten der gewählten Verteilungsfunktion widersprechen.

Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Ein Wahrscheinlichkeitspapier ist so skaliert, dass eine bestimmte Funktion beim Aufzeichnen auf dieses Papier die Form einer geraden Linie erhält.

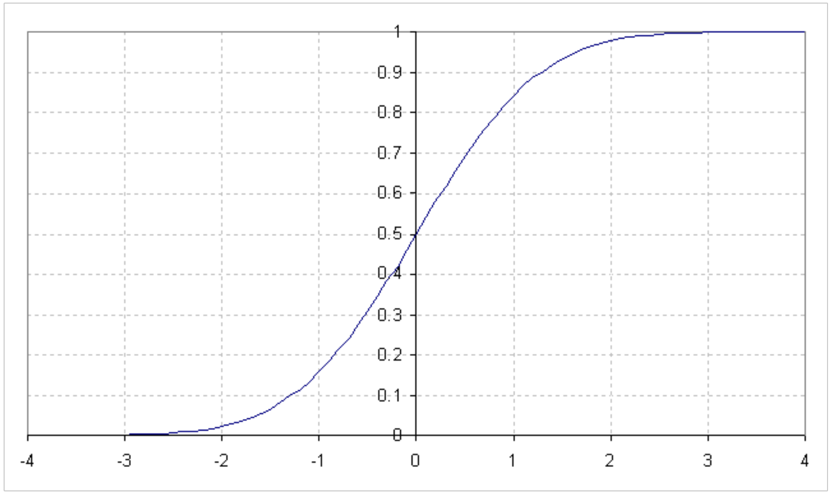
-> die Form eine geraden Linie ist z.B. gegeben durch:

$$x = a + by$$

Schätzung und Modellentwicklung

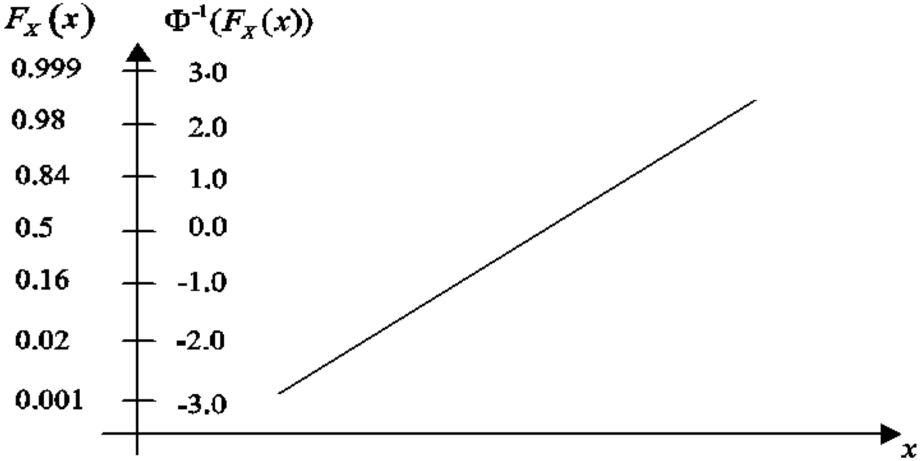
Modellauswahl anhand von Wahrscheinlichkeitspapier

Beispiel:
Wahrscheinlichkeitspapier für eine normalverteilte Wahrscheinlichkeits-verteilungsfunktion



$$F_X(x) = \Phi\left(\frac{x-\mu_X}{\sigma_X}\right)$$

$$x = \Phi^{-1}(F_X(x)) \cdot \sigma_X + \mu_X$$



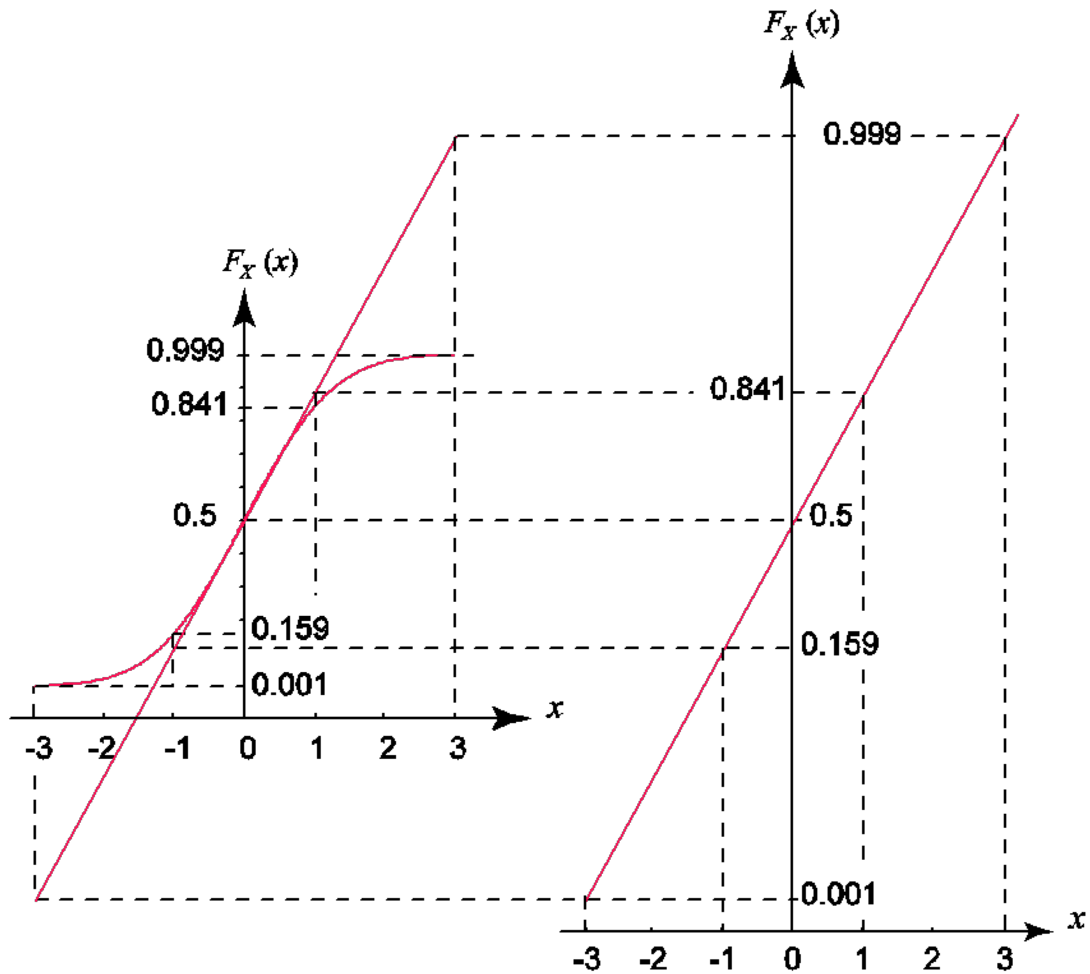
Die y-Achse ist nicht-linear skaliert.

Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Grafischer Ansatz:

Normalverteilung



Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Die Stichproben-Verteilungsfunktion kann anhand einer sortierten Messreihe abgeleitet werden:

$$F_X(x_i) = \frac{i}{N+1}$$

Beispiel: Druckfestigkeit von Beton

Lösung:
Normalverteilungs-Wahrscheinlichkeitspapier

i	x_i	$F_X(x_i)$	$\Phi^{-1}(F(x_i))$
1	24.4	0.047619	-1.668391
2	27.6	0.095238	-1.309172
3	27.8	0.142857	-1.067571
4	27.9	0.190476	-0.876143
5	28.5	0.238095	-0.712443
6	30.1	0.285714	-0.565949
7	30.3	0.333333	-0.430727
8	31.7	0.380952	-0.302981
9	32.2	0.428571	-0.180012
10	32.8	0.47619	-0.059717
11	33.3	0.52381	0.059717
12	33.5	0.571429	0.180012
13	34.1	0.619048	0.302981
14	34.6	0.666667	0.430727
15	35.8	0.714286	0.565949
16	35.9	0.761905	0.712443
17	36.8	0.809524	0.876143
18	37.1	0.857143	1.067571
19	39.2	0.904762	1.309172
20	39.7	0.952381	1.668391

Schätzung und Modellentwicklung

Modellauswahl anhand von Wahrscheinlichkeitspapier

Zeichnet man die Stichproben-Verteilungsfunktion in das Wahrscheinlichkeitspapier ein, erhält man