

Statistik und Wahrscheinlichkeitsrechnung

Übung 9

2. Teilprüfung am Dienstag, 20. Mai

Ort und Zeit:

- Start 8:15, Ende 9:45; Dauer: 90 Minuten.
- Studierende A - Eg im HIL E 6, Eh-Ky im HIL E 9
Studierende La - Z im HIL E 1

Erlaubte Hilfsmittel:

- Alle Unterlagen (Skripte, Bücher, andere Ausdrücke, etc.) erlaubt.
- Taschenrechner (ohne Kommunikationsmittel) erlaubt, auch programmierbare.
- Keine Kommunikationsmittel (z.B. Telefon) erlaubt.

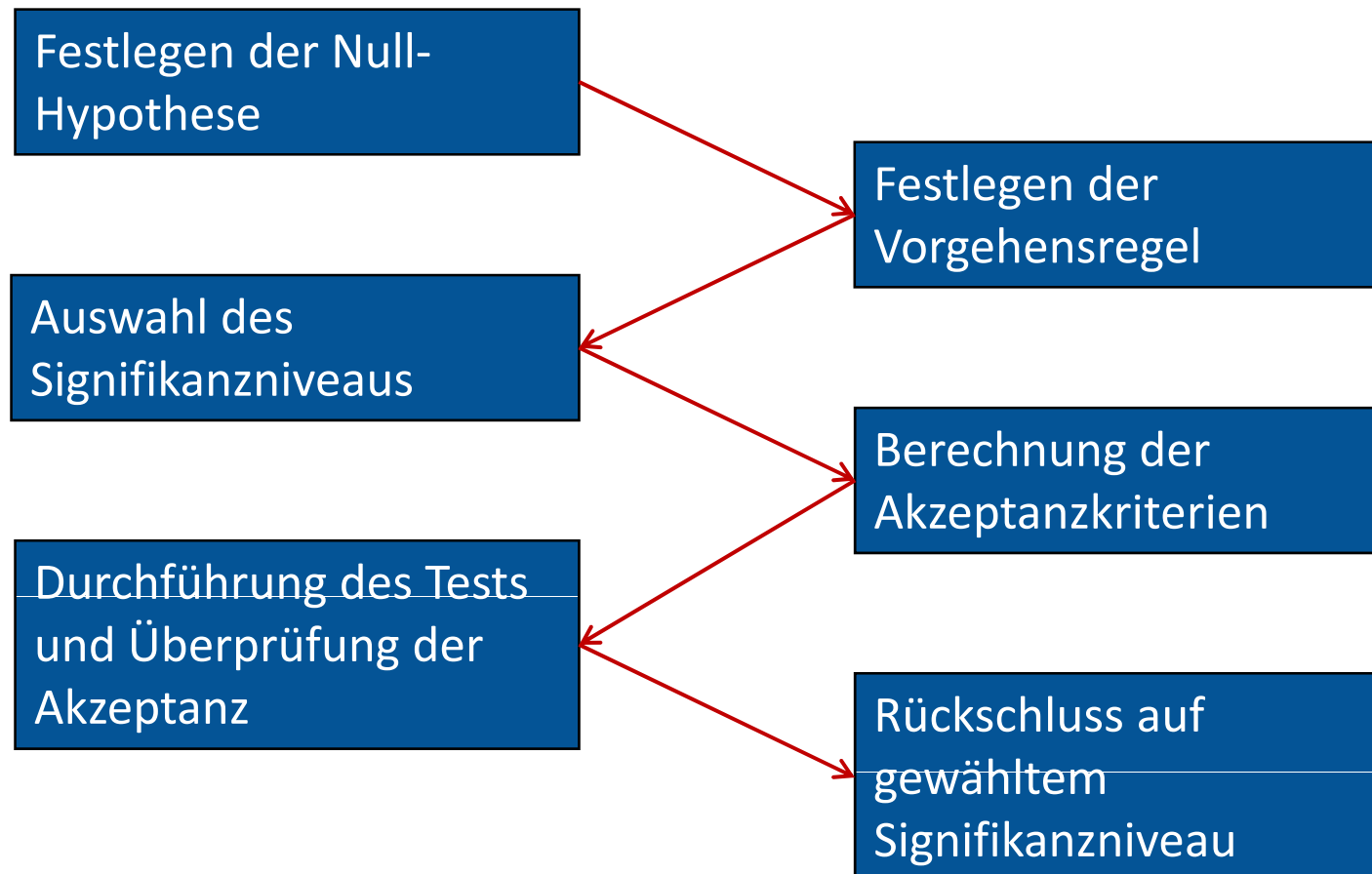
Hinweise:

- Jede Studentin/jeder Student erhält ein Couvert mit seinem Namen, darin enthalten sind die Aufgabenblätter und ein kariertes + gestempeltes Blatt für die Lösung der Rechenaufgabe.
- Legi mitnehmen!
- Falls vor 9:15 fertig: dürft ihr euch melden und hinausgehen.
- Danach müsst ihr bis zum Prüfungsende (9:45) sitzen bleiben.



Testen von Hypothesen

Generelles Vorgehen beim Hypothesentest



Aufgabe 9.1

Die wöchentliche Arbeitszeit in der Bauindustrie wurde um 2 Stunden reduziert.

Auf einer Baustelle soll untersucht werden, ob dies tatsächlich auch so durchgesetzt wird. Die Baugewerkschaft behauptet, dass die Arbeiter immer noch gleichviel arbeiten müssten wie vor der Reduktion der Arbeitszeit.

9 Arbeiter wurden zufällig ausgewählt und ihre Arbeitszeit in der Woche vor (X) und nach der Arbeitszeitverkürzung (Y) festgestellt.

Es wird angenommen, dass X und Y normalverteilt sind, und die Varianz der wöchentlichen Arbeitszeit sowohl vor wie auch nach der Reduktion $\sigma_X^2 = \sigma_Y^2 = 9.5 \text{ Stunden}^2$ beträgt.

Aufgabe 9.1

- a) Kann man anhand dieser Daten auf einem Signifikanzniveau von 5 % behaupten, dass der Mittelwert der wöchentlichen Arbeitszeit vor der Reduktion gleich 40 Stunden/Woche war?
- b) Teste die Hypothese der Baugewerkschaft auf einem Signifikanzniveau von 5 %.

Arbeiter (i)	Arbeitszeit vor der Reduktion (h)	Arbeitszeit nach der Reduktion (h)
1	38	38
2	41	39
3	40	41
4	42	39
5	43	40
6	40	40
7	39	39
8	37	38
9	43	40

Aufgabe 9.1 - Lösung

1. Formuliere die Null- und die Alternativhypothese:

$$H_0 : \mu_X = 40 \text{ Std./Woche}$$

$$H_1 : \mu_X \neq 40 \text{ Std./Woche}$$

2. Wähle ein Signifikanzniveau

$$\alpha = 0.05$$

3. Formuliere die Entscheidungsregel

Die Nullhypothese kann auf dem Signifikanzniveau von α nicht verworfen werden, wenn folgende Bedingungen erfüllt wird:

$$-k_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma_X \frac{1}{\sqrt{n}}} < k_{\alpha/2}$$

wobei n die Anzahl der Messungen ist.

Aufgabe 9.1 - Lösung

$$\begin{aligned} -k_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma_X \frac{1}{\sqrt{n}}} < k_{\alpha/2} &\Leftrightarrow -\Phi^{-1}\left(1 - \frac{0.05}{2}\right) < \frac{\bar{X} - 40}{\sqrt{9.5} \frac{1}{\sqrt{9}}} < \Phi^{-1}\left(1 - \frac{0.05}{2}\right) \Leftrightarrow \\ -1.96 < \frac{\bar{X} - 40}{\sqrt{9.5} \frac{1}{\sqrt{9}}} < 1.96 &\Leftrightarrow 37.99 \text{ Stunden} < \bar{X} < 42.01 \text{ Stunden} \end{aligned}$$

Die Nullhypothese kann auf dem Signifikanzniveau nicht verworfen werden, wenn der Stichprobenmittelwert der wöchentlichen Arbeitszeiten der 9 Arbeiter zwischen 37.99 und 42.01 Stunden liegt.

Der Mittelwert der wöchentlichen Arbeitszeit vor der Reduktion kann tatsächlich auf einem Signifikanzniveau von 5% als 40 Stunden/Woche bestätigt werden.

Aufgabe 9.1

b) Teste die Hypothese der Baugewerkschaft (dass die Arbeiter immer noch gleich viel arbeiten) auf einem Signifikanzniveau von 5 %.

→ Testen zweier oder mehrerer Datensätze

Die Gewerkschaft behauptet also: $\mu_X = \mu_Y$

Was gleichbedeutend ist mit: $\mu_X - \mu_Y = 0$

Wenn also $\mu_X - \mu_Y \leq \Delta$ dann könnte die Gewerkschaft Recht haben!

Aufgabe 9.1

b) Teste die Hypothese der Baugewerkschaft (dass die Arbeiter immer noch gleich viel arbeiten) auf einem Signifikanzniveau von 5 %.

→ Testen zweier oder mehrerer Datensätze

Beachte nun, dass die Differenz $\mu_X - \mu_Y$ der Mittelwerte der Grundgesamtheit der zwei Zufallsvariablen X und Y auch durch die Stichprobenstatistik $\bar{X} - \bar{Y}$ erhalten werden kann: $\mu_{\bar{X} - \bar{Y}} = \mu_X - \mu_Y$

Dies weil: $E[\bar{X}] = \mu_X$ (siehe Vorlesung 9)

Und: $E[X - Y] = E[X] - E[Y]$ Operator des Erwartungswertes
(siehe Aufgabe 6.1)

Aufgabe 9.1

b) Teste die Hypothese der Baugewerkschaft (dass die Arbeiter immer noch gleich viel arbeiten) auf einem Signifikanzniveau von 5 %.

→ Testen zweier oder mehrerer Datensätze

Wir können demnach als Stichprobenstatistik die Differenz $\bar{X} - \bar{Y}$ der Mittelwerte der beiden Zufallsvariablen X und Y heranziehen.

Für die Varianz gilt analog: $\sigma_{\bar{X}-\bar{Y}}^2 = \frac{\sigma_X^2}{k} + \frac{\sigma_Y^2}{l}$

Dies weil: $\text{Var}[\bar{X}] = \frac{1}{n} \sigma_X^2$

Und: $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$

Dies sind Operatoren für die Varianz, wobei wir in unserem Beispiel von unabhängigen Zufallsvariablen ausgehen (Cov=0) - Annahme, dass die Messung der Arbeitszeit zeitlich auseinanderliegt.

Aufgabe 9.1

1. Festlegen der Nullhypothese

$$H_0 : \mu_X = \mu_Y \Leftrightarrow \mu_X - \mu_Y = 0$$

$$H_1 : \mu_X \neq \mu_Y$$

2. Festlegen der Entscheidungsregel

Parameter für die Entscheidungsregel= Differenz der Mittelwerte

$$\bar{X} - \bar{Y} \leq \Delta$$

$$Z = \bar{X} - \bar{Y} \Rightarrow Z \leq \Delta$$

3. Auswahl des Signifikanzniveaus: $\alpha = 5 \%$

Aufgabe 9.1

4. Berechnung der Akzeptanzkriterien

Wir betrachten den kritischsten Fall, nämlich:

$$P[Z \leq \Delta \mid \mu_Z = 0] = 0.95 \text{ mit } Z = \bar{X} - \bar{Y}$$

Dann können wir die Wahrscheinlichkeit berechnen, indem wir standardisieren:

$$P(Z \leq \Delta) = 0.95$$

$$P\left(\frac{Z - \mu_Z}{\sigma_Z} \leq \frac{\Delta - \mu_Z}{\sigma_Z}\right) = 0.95$$

$$\text{Berechnung von } \mu \text{ und } \sigma: \mu_Z = \mu_X - \mu_Y = 0$$

$$\sigma_Z^2 = \frac{\sigma_X^2}{k} + \frac{\sigma_Y^2}{l} = \frac{9.5}{9} + \frac{9.5}{9} = 2.11$$

Aufgabe 9.1

4. Berechnung der Akzeptanzkriterien

Durch Einsetzen erhalten wir:

$$P\left(\frac{Z-0}{\sqrt{2.11}} \leq \frac{\Delta-0}{\sqrt{2.11}}\right) = 0.95$$

$$\Phi\left(\frac{\Delta}{\sqrt{2.11}}\right) = 0.95$$

$$\frac{\Delta}{\sqrt{2.11}} = 1.65$$

$$\Delta = 1.65\sqrt{2.11}$$

$$\Delta = 2.39$$

z	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
15	1.00	0.8413	1.50	0.9332	2.00	0.9772
50	1.01	0.8438	1.51	0.9345	2.10	0.9821356
85	1.02	0.8461	1.52	0.9357	2.20	0.9860966
19	1.03	0.8485	1.53	0.9370	2.30	0.9892759
54	1.04	0.8508	1.54	0.9382	2.40	0.9918025
88	1.05	0.8531	1.55	0.9394	2.50	0.9937903
23	1.06	0.8554	1.56	0.9406	2.60	0.9953388
57	1.07	0.8577	1.57	0.9418	2.70	0.9965330
90	1.08	0.8599	1.58	0.9429	2.80	0.9974449
24	1.09	0.8621	1.59	0.9441	2.90	0.9981342
57	1.10	0.8643	1.60	0.9452	3.00	0.9986501
91	1.11	0.8665	1.61	0.9463	3.10	0.9990324
24	1.12	0.8686	1.62	0.9474	3.20	0.9993129
57	1.13	0.8708	1.63	0.9484	3.30	0.9995166
89	1.14	0.8729	1.64	0.9495	3.40	0.9996631
22	1.15	0.8749	1.65	0.9505	3.50	0.9997674
54	1.16	0.8770	1.66	0.9515	3.60	0.9998409
86	1.17	0.8790	1.67	0.9525	3.70	0.9998922
17	1.18	0.8810	1.68	0.9535	3.80	0.9999277

Aufgabe 9.1

5. Durchführung des Tests

$$\bar{x} = 40.33 \text{ Stunden}$$

$$\bar{y} = 39.33 \text{ Stunden}$$

$$z = \bar{x} - \bar{y} = 1 \text{ Stunde}$$

6. Rückschluss auf gewähltem Signifikanzniveau

Da der Stichprobenmittelwert innerhalb $[z \leq 2.39 \text{ Stunden}]$ liegt, kann die Nullhypothese, dass die Bauarbeiter immer noch gleich viel arbeiten, auf einem Signifikanzniveau von 5% nicht verworfen werden.

Aufgabe 9.1

Man sollte stets aufpassen, dass man die Aussage dieser Tests nicht überschätzt, da die Hypothesen

- auf unterschiedlichen Wegen / von unterschiedlichen Richtungen und
- mit unterschiedlichen Signifikanzniveaus formuliert werden können.



Aufgabe 9.2

In einem Prüflabor werden **jeden Tag 30 Messungen** durchgeführt, um die Qualität des Wassers zu kontrollieren. Jedes Messergebnis folgt einer Normalverteilung mit einem Mittelwert $\mu = 23 \text{ ng / ml}$ und einer Standardabweichung $\sigma = 4.3 \text{ ng / ml}$.

- a) Mit welcher Wahrscheinlichkeit wird **eine** Messung einen Wert von höchstens 23 ng / ml aufweisen?

$$P[X \leq 23] = P\left[\frac{X - 23}{4.3} \leq \frac{23 - 23}{4.3}\right] = \Phi(0) = 0.5$$



Aufgabe 9.2

In einem Prüflabor werden **jeden Tag 30 Messungen** durchgeführt, um die Qualität des Wassers zu kontrollieren. Jedes Messergebnis folgt einer Normalverteilung mit einem Mittelwert $\mu = 23 \text{ ng / ml}$ und einer Standardabweichung $\sigma = 4.3 \text{ ng / ml}$.

- b) Mit welcher Wahrscheinlichkeit liegt **eine** Messung im Intervall $[19.5 \text{ ng/ml}; 20.5 \text{ ng/ml}]$?

$$\begin{aligned} P[19.5 < X \leq 20.5] &= P\left[\frac{19.5 - 23.0}{4.3} < \frac{X - 23.0}{4.3} \leq \frac{20.5 - 23.0}{4.3}\right] \\ &= \Phi(-0.58) - \Phi(-0.81) = 0.073 \end{aligned}$$



Aufgabe 9.2

In einem Prüflabor werden **jeden Tag 30 Messungen** durchgeführt, um die Qualität des Wassers zu kontrollieren. Jedes Messergebnis folgt einer Normalverteilung mit einem Mittelwert $\mu = 23 \text{ ng / ml}$ und einer Standardabweichung $\sigma = 4.3 \text{ ng / ml}$.

- c) Mit welcher Wahrscheinlichkeit ergibt sich ein **Tagesmittelwert** von kleiner als 20 ng / ml ?

$$E[\bar{X}] = \mu_X \quad \text{Var}[\bar{X}] = \frac{1}{n} \sigma_X^2$$

Der Stichprobenmittelwert der 30 Messergebnisse folgt einer Normalverteilung mit $\mu = 23 \text{ ng / ml}$ und $\sigma = 4.3 / \sqrt{30} = 0.79 \text{ ng / ml}$.

$$P(\bar{X} < 20) = P\left(\frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} < \frac{20 - \mu_X}{\sigma_X / \sqrt{n}}\right) = \Phi\left(\frac{20 - 23}{0.79}\right) = 7.3 \cdot 10^{-5}$$



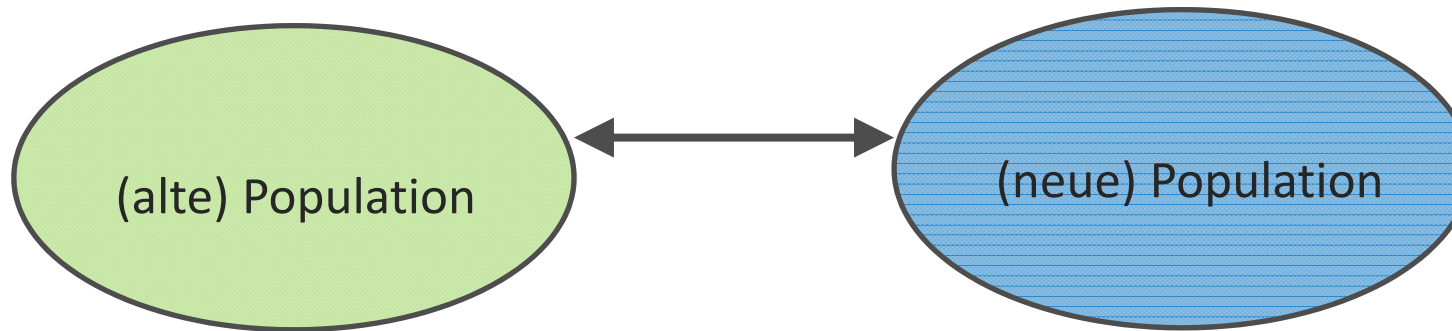
Aufgabe 9.2

In einem Prüflabor werden **jeden Tag 30 Messungen** durchgeführt, um die Qualität des Wassers zu kontrollieren. Jedes Messergebnis folgt einer Normalverteilung mit einem Mittelwert $\mu = 23 \text{ ng / ml}$ und einer Standardabweichung $\sigma = 4.3 \text{ ng / ml}$.

- d) Das Labor hat ein neues Instrument für die Messungen der Wasserqualität erworben und führt Eichmessungen durch, d.h. es vergleicht die Messwerte des neuen Instruments mit denen des alten Instruments.
- Es werden **15** Proben mit dem neuen Instrument gemessen, und deren Mittelwert von $\bar{x} = 19 \text{ ng / ml}$, sowie die Standardabweichung von $s = 5 \text{ ng / ml}$ berechnet.
- Teste auf einem 5% Signifikanzniveau, ob die neuen Messwerte zur gleichen Grundgesamtheit gehören wie die alten Messwerte.



t-Verteilung



$$\mu_A = 23 \text{ ng / ml}$$
$$\sigma_A = 4.3 \text{ ng / ml}$$

Normalverteilung

$$\bar{x}_N = 19 \text{ ng / ml}$$
$$s_N = 5 \text{ ng / ml}$$

Wir kennen weder μ_N noch σ_N der neuen Population!
Wir wissen aber, dass die Messergebnisse normalverteilt sind.



t-Verteilung

Ist die Varianz nicht bekannt, muss die folgende Stichprobenstatistik berücksichtigt werden (t-verteilt mit einem Freiheitsgrad von $n-1$):

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

$$S_{unbiased} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{der erwartungstreue Schätzer für die Varianz ist.}$$



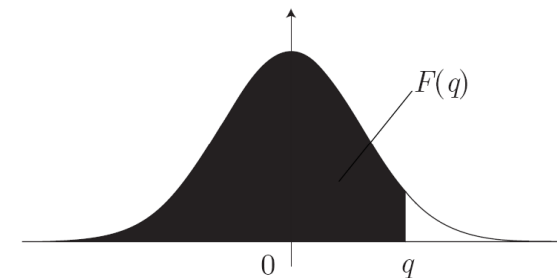
t-Verteilung

Man kann die t-Statistik berechnen oder den Wert aus einer Tabelle der t-Verteilung ablesen.

Die Wahrscheinlichkeit, dass eine t-verteilte Zufallsvariable T in einem bestimmten Intervall liegt, ist:

$$P[-t_{\alpha/2} \leq T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \leq t_{\alpha/2}]$$

Table T.2: Quantile values of the t-distribution q .



Probability density function of t-distribution.

v	F(q)=0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169

Aufgabe 9.2

$$\begin{array}{ll} \mu_A = 23 \text{ ng / ml} & \bar{x}_N = 19 \text{ ng / ml} \\ \sigma_A = 4.3 \text{ ng / ml} & s_N = 5 \text{ ng / ml} \end{array}$$



1. Festlegen der Nullhypothese

$$H_0 : \mu_N = 23 \text{ ng / ml}$$

$$H_1 : \mu_N \neq 23 \text{ ng / ml}$$

2. Festlegen der Entscheidungsregel

Der Stichprobenmittelwert muss in einem zu bestimmenden Intervall rund um μ_A liegen.

$$P[-t_{\alpha/2} \leq T = \frac{\bar{X}_N - \mu_A}{S_N / \sqrt{n}} \leq t_{\alpha/2}] = 1 - \alpha$$

Für die Stichprobe wird dies berechnet als:

$$\mu_A - t_{\alpha/2} \frac{s_N}{\sqrt{n}} \leq \bar{x}_N \leq \mu_A + t_{\alpha/2} \frac{s_N}{\sqrt{n}}$$

3. Auswahl des Signifikanzniveaus: $\alpha = 5 \%$

Aufgabe 9.2

4. Berechnung der Akzeptanzkriterien

Intervall für den Stichprobenmittelwert berechnen:

$$\mu_A - t_{\alpha/2} \frac{s_N}{\sqrt{n}} \leq \bar{x} \leq \mu_A + t_{\alpha/2} \frac{s_N}{\sqrt{n}}$$

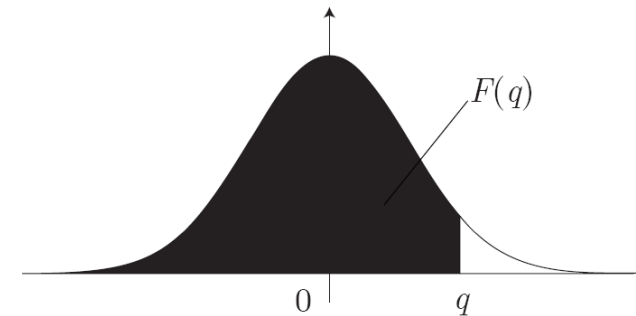
Untere Intervallgrenze:

$$\mu_A - t_{\alpha/2} \frac{s_N}{\sqrt{n}} = 23 - 2.14 \cdot \frac{5}{\sqrt{15}}$$

Obere Intervallgrenze:

$$\mu_A + t_{\alpha/2} \frac{s_N}{\sqrt{n}} = 23 + 2.14 \cdot \frac{5}{\sqrt{15}}$$

$$\begin{aligned} \mu_A &= 23 \text{ ng / ml} & \bar{x}_N &= 19 \text{ ng / ml} \\ \sigma_A &= 4.3 \text{ ng / ml} & s_N &= 5 \text{ ng / ml} \end{aligned}$$



Probability density function of t-distribution.

v	F(q)=0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947

Aufgabe 9.2

$$\begin{array}{ll} \mu_A = 23 \text{ ng / ml} & \bar{x}_N = 19 \text{ ng / ml} \\ \sigma_A = 4.3 \text{ ng / ml} & s_N = 5 \text{ ng / ml} \end{array}$$



5. Durchführung des Tests

Der Stichprobenmittelwert wurde berechnet zu: $\bar{x}_N = 19 \text{ ng / ml}$

Das Intervall der t-Statistik wurde berechnet als $[20.24 \text{ ng/ml}; 25.76 \text{ ng/ml}]$.

Der Stichprobenmittelwert liegt nicht im Intervall.

6. Rückschluss auf gewähltem Signifikanzniveau

Die Nullhypothese, dass es sich um die gleichen Populationen handelt, wird auf einem Signifikanzniveau von 5% verworfen.



Wahrscheinlichkeitspapier

Wir wollen wissen, ob eine Stichprobe mit einer bestimmten Verteilungsfamilie beschrieben werden kann –

Als erste Schätzung, ohne Parameter bestimmen zu müssen.

- ✓ Wenn die Punkte auf dem Wahrscheinlichkeitspapier auf oder sehr nahe an der linearisierten Funktion liegen, dann kann die Verteilungsfamilie akzeptiert werden.

Wie kann eine Punktmenge auf einer Linie liegen?

Wenn die Beziehung x - y linear ist → Linearisierung der y -Achse

Aufgabe 9.3

Aus Verkehrszählungen liegt eine Datenserie vor, die den täglichen Verkehrsfluss in der Rosengartenstrasse in Zürich beschreibt.



- a) Erstelle ein Wahrscheinlichkeitspapier für eine Dichtefunktion

$$f_X(x) = \begin{cases} \frac{2}{10000^2} x & 0 \leq x \leq 10000 \\ 0 & \text{sonst} \end{cases}$$

- b) Überprüfe mit Hilfe des erstellten Wahrscheinlichkeitspapiers, ob der tägliche Verkehrsfluss mit dieser Verteilung beschrieben werden kann.

Tag (i)	Anzahl Fahrzeuge
1	3600
2	4500
3	5400
4	6500
5	7000
6	7500
7	8700
8	9000
9	9500

Aufgabe 9.3

PDF
$$f_X(x) = \begin{cases} \frac{2}{10000^2} x & 0 \leq x \leq 10000 \\ 0 & \text{sonst} \end{cases}$$

CDF
$$F_X(x) = \begin{cases} 0 & 0 \leq x \\ \left(\frac{x}{10000}\right)^2 & 0 < x \leq 10000 \\ 1 & x > 10000 \end{cases}$$

LINEARISIERUNG
$$F_X(x) = \left(\frac{x}{10000}\right)^2 \Leftrightarrow \sqrt{F_X(x)} = \frac{x}{10000}$$

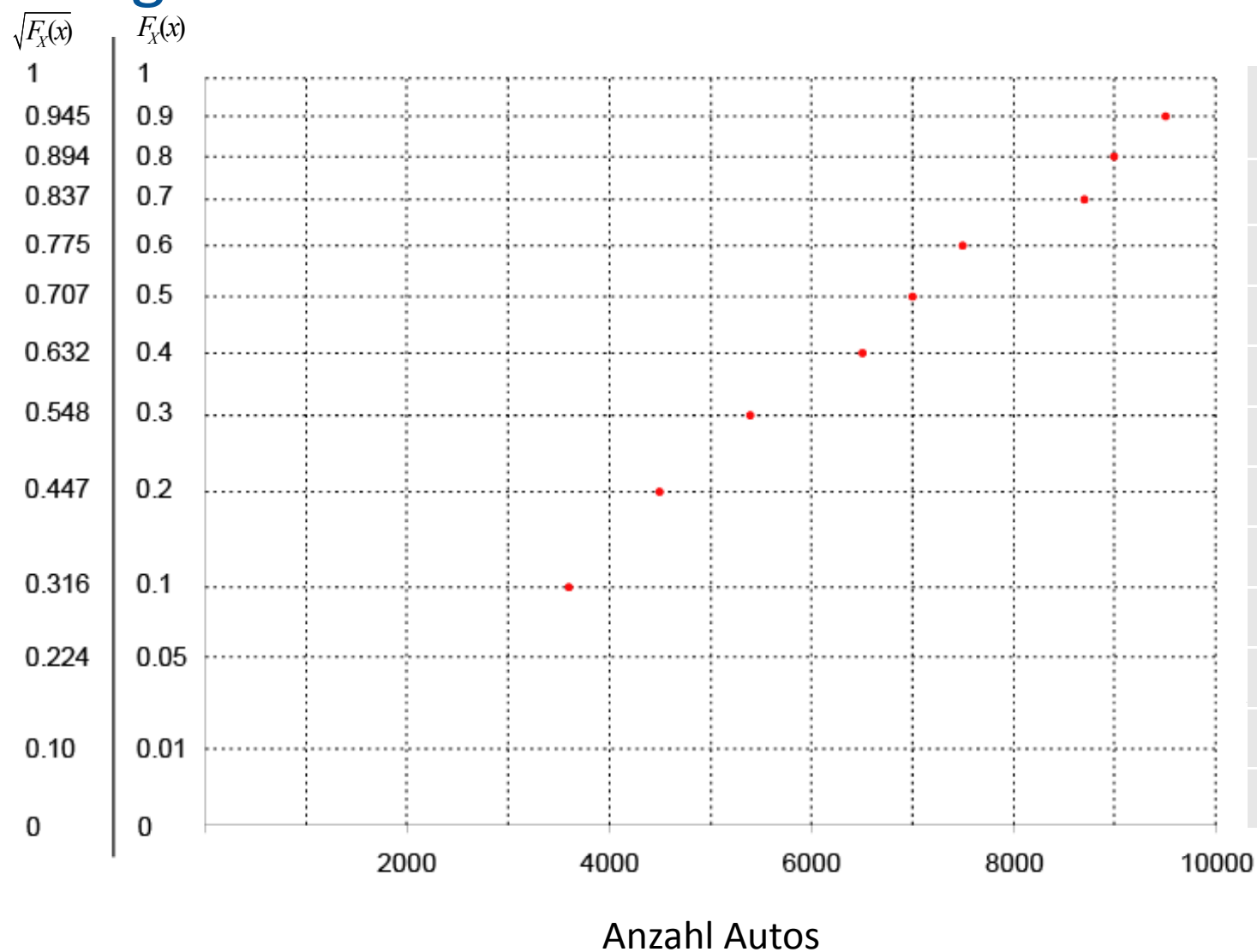


Aufgabe 9.3

Eine linearisierte y-Achse erhalten wir, indem wir für jede Quantile in bestimmten Abständen die linearisierte Form berechnen - beispielsweise zu jedem Datenpunkt.

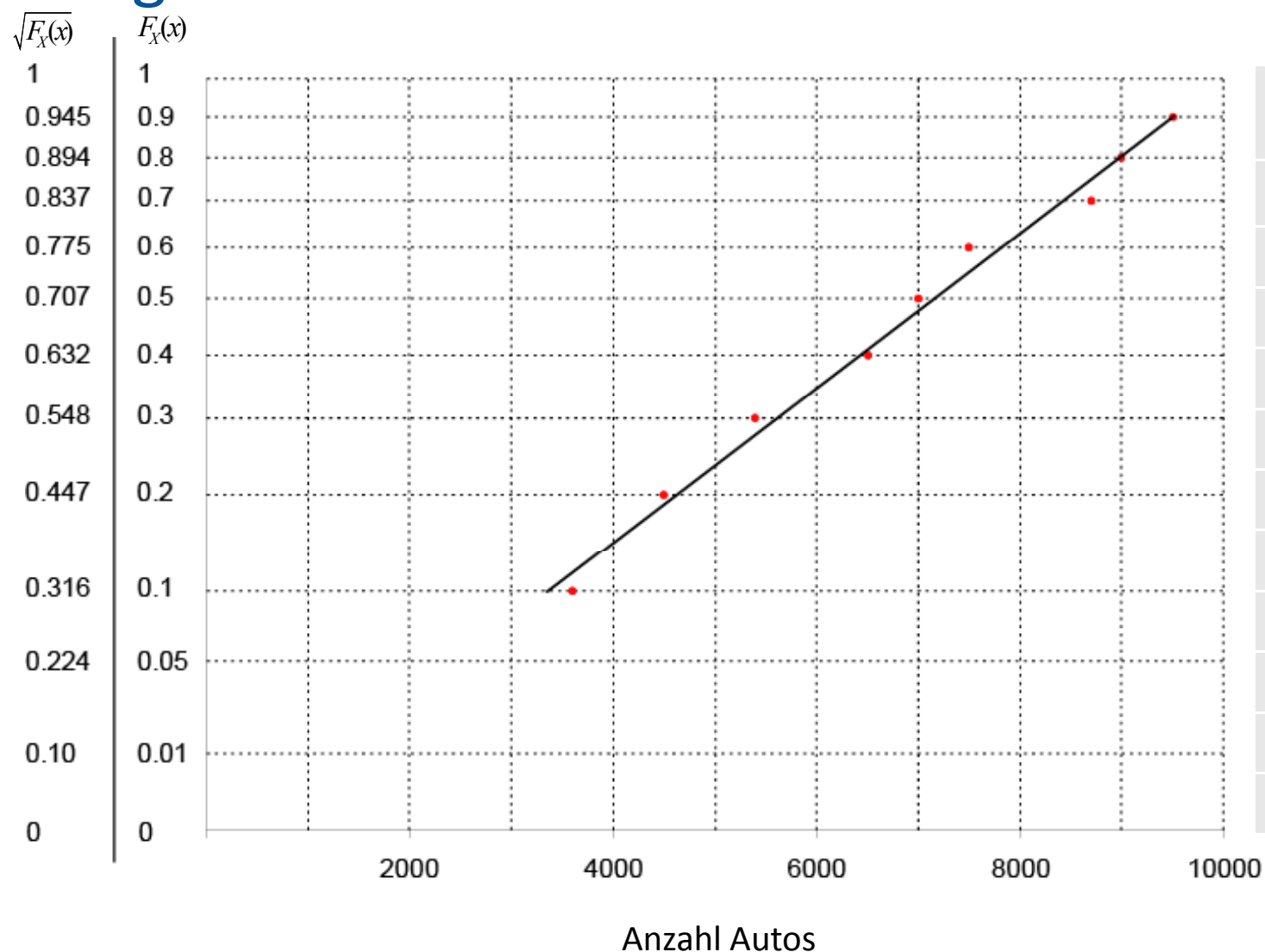
Anz. Autos (x-Achse)	$\sqrt{F_X(x)}$	$F_X(x)$
	0	0
3600	0.31	0.1
4500	0.45	0.2
5400	0.55	0.3
6500	0.63	0.4
7000	0.71	0.5
7500	0.77	0.6
8700	0.84	0.7
9000	0.89	0.8
9500	0.94	0.9
	1.0	1.0

Aufgabe 9.3



Anzahl Autos	$F_X(x)$	$\sqrt{F_X(x)}$
	0	0
3600	0.1	0.31
4500	0.2	0.45
5400	0.3	0.55
6500	0.4	0.63
7000	0.5	0.71
7500	0.6	0.77
8700	0.7	0.84
9000	0.8	0.89
9500	0.9	0.94
	1.0	1.0

Aufgabe 9.3



Anzahl Autos	$F_X(x)$	$\sqrt{F_X(x)}$
	0	0
3600	0.1	0.31
4500	0.2	0.45
5400	0.3	0.55
6500	0.4	0.63
7000	0.5	0.71
7500	0.6	0.77
8700	0.7	0.84
9000	0.8	0.89
9500	0.9	0.94
	1.0	1.0

Aufgabe 9.4

Um die Druckfestigkeit von Beton einer bestimmten Produktionsmarge zu modellieren, wurden 20 Stichproben gemessen. Die Ergebnisse sind in der Tabelle dargestellt. Es wird angenommen, dass die Grundgesamtheit der Stichproben einer Normalverteilung $N(\mu, \sigma)$ folgt.

Schätze die unbekannten Parameter (μ, σ) anhand der Momentenmethode.

No. of sample	Compressive strength [MPa]	No. of sample	Compressive strength [MPa]
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7

Aufgabe 9.4 – Lösung

Mit der Momentenmethode werden die Parameter wie folgt geschätzt:

$$m_1 = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \mu$$

$$m_2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sigma^2 + \mu^2$$

$$\left. \begin{aligned} \mu &= m_1 = \frac{1}{n} \sum_{i=1}^n \hat{x}_i \\ \sigma^2 &= m_2 - m_1^2 = \frac{1}{n} \sum_{i=1}^n \hat{x}_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \hat{x}_i \right)^2 \end{aligned} \right\} \begin{aligned} \mu &= 32.67 \\ \sigma &= 4.04 \end{aligned}$$



Maximum-Likelihood-Methode (MLM)

Die Likelihood-Funktion:

X ist eine Zufallsvariable mit der WDF: $f_X(x | \theta)$

Bei n Versuchen, beträgt die multivariate WDF der Ergebnisse

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n f_X(x_i | \theta) \quad \longleftarrow \text{vor den Beobachtungen}$$

Dabei handelt es sich um die Wahrscheinlichkeitsdichte für jeden beliebigen θ .

Wir können jedoch diesen Ausdruck auch mit der Funktion von θ nach den Beobachtungen $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ in Verbindung bringen...



Maximum-Likelihood-Methode (MLM)

Die Likelihood-Funktion:

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n \mid \theta) = \prod_{i=1}^n f_X(x_i \mid \theta) \quad \longleftarrow \quad \text{vor den Beobachtungen}$$

Wir können jedoch diesen Ausdruck auch mit der Funktion von θ nach den Beobachtungen $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ in Verbindung bringen...

... und erhalten dadurch die **Likelihood-Funktion**:

$$L(\theta \mid \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = \prod_{i=1}^n f_X(\hat{x}_i \mid \theta) \quad \longleftarrow \quad \text{nach den Beobachtungen}$$



Maximum-Likelihood-Methode (MLM)

Den **Maximum-Likelihood-Schätzwert** θ erhält man, indem man den Wert von θ wählt, der die Likelihood-Funktion $L(\theta)$ maximiert.

Dies ist gleichbedeutend mit der Maximierung der „**Log-Likelihood-Funktion**“ und wie folgt definiert:

$$l(\theta | \hat{\mathbf{x}}) = \ln L(\theta | \hat{\mathbf{x}}) = \ln \prod_{i=1}^n f_X(\hat{x}_i | \theta) = \sum_{i=1}^n \ln f_X(\hat{x}_i | \theta)$$

Aufgabe 9.5 Gruppenaufgabe

Aufgabenstellung aus 9.4.

- a) Beschreibe die Likelihood-Funktion.
- b) Schätze die unbekannten Parameter (μ, σ) anhand der MLM.

Was passiert, wenn die Normalverteilung durch eine Exponentialverteilung ersetzt wird?

- c) Schätze die Parameter der Exponentialverteilung mit der MLM und vergleiche die kumulative Verteilungsfunktion mit den beobachteten Werten.

No. of sample	Compressi ve strength [MPa]	No. of sample	Compressi ve strength [MPa]
1	24.4	11	33.3
2	27.6	12	33.5
3	27.8	13	34.1
4	27.9	14	34.6
5	28.5	15	35.8
6	30.1	16	35.9
7	30.3	17	36.8
8	31.7	18	37.1
9	32.2	19	39.2
10	32.8	20	39.7