Merkblatt Übersicht:

→ Wann welcher Ansatz?:

WK nicht Anbiett p ist immer gleich!

→ **Binominalverteilung:** → ① Wahrscheinlichkeit gegeben ② unabhängig ③ Feste Anzahl Versuche ④ nur Erfolg/Misserfolg → $X \sim B(n, p)$

85 Leute

→ Ziehen mit zurücklegen und nur zwei mögliche Ausgänge (Erfolg / Misserfolg)

↳ n = feste Anzahl Versuche, p = WK Erfolg

↳ WK für jeden Erfolg muss gleich bleiben und darf sich nicht verändern

→ **Normalverteilung:** → oft als Annäherung für Binominalverteilung verwendet

→ kontinuierlich Messergebnisse können beliebigen Wert annehmen (Körpergröße, Blutdruck)

→ Mittelwert μ und Varianz σ^2 gegeben, $N(\mu, \sigma^2)$ → zeigt Verteilung um Mittelwert

↳ Ansatz: TR → Normal CD mit Upper Lower → TR will immer Standardabweichung σ und nicht Varianz σ^2 immer darauf achten!!

→ **Poissonverteilung:** → Mittelwert pro Jahr

→ Mittelwert über Intervall nur angegeben (z.B. fällt im Mittel nur 18 mal pro Jahr aus)

↳ es gilt zu unterscheiden:

① → WK ("wie wahrscheinlich mehr als 20 mal pro Jahr")

→ EW ("wie hoch erwartete Anzahl Ausfälle pro Monat $(\frac{18}{12})$)

② → CD → Summe → ("WK mehr als 20 mal pro Jahr ausfällt")

→ PD → einzeln → ("WK im nächsten Monat genau 1 ausfällt")

③ → gegebener Zeitraum ("pro Jahr") → $X \sim P_0(18)$

→ anderer Zeitraum ("pro Monat") → $X \sim P_0(\frac{18}{12})$

→ **Hypergeometrische Verteilung:** → N = Grundgesamtheit gegeben, n = Anzahl Ziehungen, K = Erfolge in Grundgesamtheit ② abhängig ③ $X \sim H(n, N, K)$

→ Ziehen ohne zurücklegen ("Charge 50 Bauteile, Stichprobe 10 → man nimmt 10 raus und 3 nicht funktionsfähig")

↳ WK im vgl. zu Binominal ändert sich also mit Ziehen

↳ $X \sim H(10, 50, K)$

Gesamtheit N = 50

Erfolge K = 3

Stichprobe n = 10

Aufgaben in Wiwi:

- Aufgabe 1: Vertrauensbereich mit z bestimmen (einseitig / beidseitig)
- Aufgabe 2: Wahrscheinlichkeit für nicht eintreten Ereignis
- Aufgabe 3: Regression
- Aufgabe 4: Hypothesentest zur Normalverteilung mit t
- Aufgabe 5: zweidimensionale Verteilung
- Aufgabe 6: Binominalverteilung
- Aufgabe 7: Median

→ Hypothesentest:

→ $\boxed{\text{Allgemeiner Ansatz: → Zufallsstreubereich ausrechnen und mit Nullhypothese } H_0 \text{ vergleichen!}}$

→ t-Test:

↳ $\boxed{\text{immer wenn Stichprobe} \leq 30 \text{ ist}}$

↳ $\boxed{t_\Delta = H_0 \pm t \cdot \frac{s}{\sqrt{n}}}$ → $I = [H - t, H + t]$

↳ wichtig zu schauen:

→ ob einseitig / zweiseitig (entsprechend anders t ablesen)

→ ob ZSB untere / obere Grenze abdecken soll ("H_1: Flaschen weniger als 1 Liter → untere Grenze vergleichen mit H_0)

→ z-test:

↳ $\boxed{\text{bei größerer Stichprobe } n > 30}$

↳ $\boxed{\Delta p = z \cdot \frac{s}{\sqrt{n}} \pm \frac{1}{2n}}$ → $I = [p - \Delta p, p + \Delta p]$

↳ wichtig zu schauen:

→ ob einseitig / zweiseitig (entsprechend anders z ablesen)

→ ob ZSB untere / obere Grenze abdecken soll ("höchstens Anzahl Nebenwirkung")

Ergänzung: Aufstellen der Hypothesen für Parametertests

H_0 beinhaltet den für den unbekannten Parameter auszuschließenden Wert (bzw. die auszuschließenden Werte), H_1 das „Gegenteil“ von H_0, also die zu bestätigenden Werte.

Beispiele für Signalwörter im Text:

… gleich … $=$	… ungleich, weicht ab, … $\neq$
… höchstens, nicht mehr als … $\leq$	… größer, mehr als, … $>$
… mindestens, nicht weniger als … $\geq$	… kleiner, weniger als, … $<$
$\boxed{\mathbf{H_0}}$	$\boxed{\mathbf{H_1}}$

5.2.2 Tabellen für statistische Tests

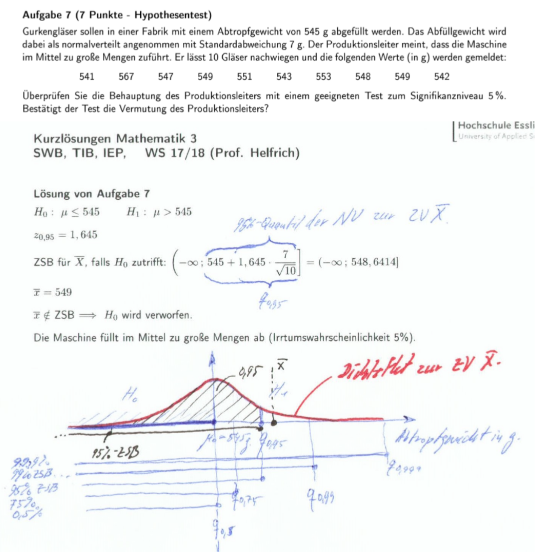
Allgemeiner Hinweis zu den Tabellen auf den folgenden Seiten: Bei den zweiseitigen Tests ist das $(1-\frac{\alpha}{2})$-Quantil der Normalverteilung (bzw. t-Verteilung) zu benutzen. Ist zum Beispiel $\alpha = 5\%$, so muss bei der Normalverteilung das Quantil $z_{1-\frac{\alpha}{2}} = z_{0{,}975} = 1{,}96$ benutzt werden.

1. Gauß-Test: Test für μ bei bekannter Standardabweichung σ zum Signifikanzniveau α

   Test (zum Signifikanzniveau α) einer Nullhypothese über den unbekannten Erwartungswert μ, z. B. (erster Eintrag in der Tabelle) die Hypothese, dass μ gleich einer vorgegebenen festen Zahl μ_0 (etwa einem Sollwert oder dem bisherigen Wert) ist.

   Gegeben: Stichprobe $x_1, x_2, \ldots, x_n$. Die Messwerte sind Realisierungen von n unabhängigen $N(\mu, \sigma^2)$- verteilten Zufallsvariablen mit unbekanntem Erwartungswert μ, aber bekannter Varianz σ^2.

$\mathbf{H_0}$	$\mathbf{H_1}$	**Zufallsstreubereich für $\overline{X}$, falls $\mathbf{H_0}$ zutrifft**	**$\mathbf{H_0}$ verwerfen, falls**
$\mu = \mu_0$	$\mu \neq \mu_0$	$[\mu_0 - \triangle\mu\ ;\ \mu_0 + \triangle\mu]$, mit $\triangle\mu = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$	$\overline{x} \notin$ ZSB
$\mu \geq \mu_0$	$\mu < \mu_0$	$\left[\mu_0 - z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\ ;\ \infty\right)$	$\overline{x} \notin$ ZSB
$\mu \leq \mu_0$	$\mu > \mu_0$	$\left(-\infty\ ;\ \mu_0 + z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right]$	$\overline{x} \notin$ ZSB


Aufgabe 7 (7 Punkte - Hypothesentest)

Gurkengläser sollen in einer Fabrik mit einem Abtropfgewicht von 545 g abgefüllt werden. Das Abfüllgewicht wird dabei als normalverteilt angenommen mit Standardabweichung 7 g. Der Produktionsleiter meint, dass die Maschine im Mittel zu große Mengen zuführt. Er lässt 10 Gläser nachwiegen und die folgenden Werte (in g) werden gemeldet:

541 567 547 549 551 543 553 548 549 542

Überprüfen Sie die Behauptung des Produktionsleiters mit einem geeigneten Test zum Signifikanzniveau 5 %. Bestätigt der Test die Vermutung des Produktionsleiters?

Hochschule Essli
University of Applied S

Kurzlösungen Mathematik 3
SWB, TIB, IEP, WS 17/18 (Prof. Helfrich)

Lösung von Aufgabe 7

$H_0:\ \mu \leq 545 \qquad H_1:\ \mu > 545$

$z_{0,95} = 1,645$

ZSB für $\overline{X}$, falls H_0 zutrifft: $\left(-\infty\,;\, \underbrace{\overset{\text{95\%-Quantil der NV zur ZV } \overline{X}.}{\overbrace{545 + 1,645 \cdot \frac{7}{\sqrt{10}}}}}_{q_{0,95}}\right] = (-\infty\,;\, 548,6414]$

$\overline{x} = 549$

$\overline{x} \notin$ ZSB $\Longrightarrow$ H_0 wird verworfen.

Die Maschine füllt im Mittel zu große Mengen ab (Irrtumswahrscheinlichkeit 5%).

Abbildung 11: Beispiel: Klausuraufgabe WiSe 17/18, Aufgabe 7

89. (Gauß-Test) Bei der Herstellung von Schokoladentafeln sei das Verpackungsgewicht X normalverteilt mit Standardabweichung $\sigma = 2$g. Der Sollwert der Tafeln liegt bei 100g. Der Hersteller möchte weder haben, dass $X < 100$g (dann müsste er Reklamationen der Verbraucher befürchten), noch dass $X > 100$g (unnötige Verschwendung und damit finanzielle Verluste). Eine Stichprobe vom Umfang $n = 10$ ergibt ein arithmetisches Mittel von $\bar{x} = 98.9$g.

(a) In einem ersten Schritt soll davon ausgegangen werden, dass der unbekannte Erwartungswert gleich dem Sollwert 100g ist. Welche Hypothesen müssen aufgestellt werden?

(b) Legen Sie unter der Annahme, dass die Hypothese H_0 zutrifft, einen symmetrischen Zufallsstreubereich um $\mu = 100$g fest, in dem $\bar{x}$ mit einer Wahrscheinlichkeit von 95% liegt.

(c) Kann die Hypothese $\mu = 100$g bei der vorgegebenen Stichprobe verworfen werden?

(d) Wie groß ist die Wahrscheinlichkeit die Hypothese $\mu = 100$g zu verwerfen, obwohl sie tatsächlich zutrifft?

(e) Wie lauten Nullhypothese H_0, Alternativhypothese H_1, Zufallsstreubereich und Ablehnungsbereich, wenn der Hersteller überprüfen will, ob der Erwartungswert der Schokoladentafeln größer oder gleich 100g ist?

(f) Kann an der Nullhypothese $\mu \geq 100$g bei der vorliegenden Stichprobe festgehalten werden?

2. t-Test: Test für μ bei *unbekannter* Standardabweichung σ zum Signifikanzniveau α

   Gegeben: Stichprobe $x_1, x_2, \ldots, x_n$. Die Messwerte sind Realisierungen von n unabhängigen $N(\mu, \sigma^2)$-verteilten Zufallsvariablen mit unbekanntem Erwartungswert μ, und **unbekannter** Varianz σ^2.

   $\longrightarrow$ Schätze σ durch s aus der Stichprobe. Deshalb müssen die Quantile der t-Verteilung (mit $n-1$ Freiheitsgraden) statt der Normalverteilung benutzt werden.

$\mathbf{H_0}$	$\mathbf{H_1}$	**Zufallsstreubereich für $\overline{X}$, falls $\mathbf{H_0}$ zutrifft**	**$\mathbf{H_0}$ verwerfen, falls**
$\mu = \mu_0$	$\mu \neq \mu_0$	$[\mu_0 - \triangle\mu\ ;\ \mu_0 + \triangle\mu]$, mit $\triangle\mu = t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$	$\overline{x} \notin$ ZSB
$\mu \geq \mu_0$	$\mu < \mu_0$	$\left[\mu_0 - t_{n-1;1-\alpha} \cdot \frac{s}{\sqrt{n}}\ ;\ \infty\right)$	$\overline{x} \notin$ ZSB
$\mu \leq \mu_0$	$\mu > \mu_0$	$\left(-\infty\ ;\ \mu_0 + t_{n-1;1-\alpha} \cdot \frac{s}{\sqrt{n}}\right]$	$\overline{x} \notin$ ZSB

Aufgabe 94

s: Standardabweichung der Stichprobe,
$t_{n-1;1-\alpha}$: $(1-\alpha)$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden

Au

3. Zweistichproben-t-Test

   Test (zum Signifikanzniveau α) über die Differenz zweier Erwartungswerte $\mu_X - \mu_Y$ zweier Grundgesamtheiten bei **unbekannter** aber gleicher Standardabweichung σ.

   Zum Test werden zwei Stichproben vom Umfang m und n mit den arithmetischen Mitteln $\overline{x}$ und $\overline{y}$ und mit den empirischen Standardabweichungen s_X und s_Y gezogen.

   Berechne zunächst die Hilfsgröße $s_d = \sqrt{(m-1) \cdot s_X^2 + (n-1) \cdot s_Y^2} \cdot \sqrt{\frac{m+n}{m \cdot n \cdot (m+n-2)}}$

$\mathbf{H_0}$	$\mathbf{H_1}$	**Zufallsstreubereich für $\overline{X} - \overline{Y}$, falls $\mathbf{H_0}$ zutrifft**	**$\mathbf{H_0}$ verwerfen falls**
$\mu_X - \mu_Y = 0$	$\mu_X - \mu_Y \neq 0$	$[-t_{m+n-2;1-\frac{\alpha}{2}} \cdot s_d\ ;\ t_{m+n-2;1-\frac{\alpha}{2}} \cdot s_d]$	$\overline{x} - \overline{y} \notin$ ZSB
$\mu_X - \mu_Y \geq 0$	$\mu_X - \mu_Y < 0$	$[-t_{m+n-2;1-\alpha} \cdot s_d\ ;\ \infty)$	$\overline{x} - \overline{y} \notin$ ZSB
$\mu_X - \mu_Y \leq 0$	$\mu_X - \mu_Y > 0$	$(-\infty\ ;\ t_{m+n-2;1-\alpha} \cdot s_d]$	$\overline{x} - \overline{y} \notin$ ZSB

Aufgabe 95

4. Test für eine unbekannte Wahrscheinlichkeit/einen unbekannten Anteil p einer Grundgesamtheit zum Signifikanzniveau α

   Bei einer Stichprobe vom Umfang n sei das gesuchte Ereignis k-mal eingetreten.

$\mathbf{H_0}$	$\mathbf{H_1}$	**Zufallsstreubereich für $\frac{\mathbf{k}}{\mathbf{n}}$, falls $\mathbf{H_0}$ zutrifft**	**$\mathbf{H_0}$ verwerfen falls**
$p = p_0$	$p \neq p_0$	$[p_0 - \triangle p; p_0 + \triangle p]$, mit $\triangle p = z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{p_0 \cdot (1-p_0)}}{\sqrt{n}} + \frac{0{,}5}{n}$	$\frac{k}{n} \notin$ ZSB
$p \geq p_0$	$p < p_0$	$\left[p_0 - z_{1-\alpha} \cdot \frac{\sqrt{p_0 \cdot (1-p_0)}}{\sqrt{n}} - \frac{0{,}5}{n}; 1\right]$	$\frac{k}{n} \notin$ ZSB
$p \leq p_0$	$p > p_0$	$\left[0\ ;\ p_0 + z_{1-\alpha} \cdot \frac{\sqrt{p_0 \cdot (1-p_0)}}{\sqrt{n}} + \frac{0{,}5}{n}\right]$	$\frac{k}{n} \notin$ ZSB

Begriffserklärung:

Varianz: $= \sigma^2$ (s² bei Stichproben)

↳ Streuung der Werte in Datensatz um den Mittelwert

→ Durchschnitt der quadrierten Abweichungen vom Mittelwert. Wird in ² gemessen

Standardabweichung: $= \sigma / s$ (bei Grundgesamtheit / bei Stichproben)

↳ Wurzel aus der Varianz und wird in gleichen Einheiten wie Daten gemessen

→ gibt bessere Vorstellung wie Daten um Mittelwert verteilt sind

Signifikanzniveau: $= z$

↳ Repräsentiert die Wahrscheinlichkeit Fehler 1. Art zu begehen. d.h. Nullhypothese fälschlicherweise abzulehnen wenn sie wahr ist

→ kann einseitig, beidseitig sein hier aufpassen nach was gefragt worden ist

→ Wert aus Tabelle von Vertrauensbereich ablesen

Vertrauensbereich / Vertrauensniveau:

↳ gibt an mit mit welcher Wahrscheinlichkeit wahre Wert eines Parameters in das berechnete Intervall fällt. z.B. 0,95 = 95% Sicherheit

Empirische Regressionsgerade:

↳ Gerade, die so durch Datenpunkte im Streudiagramm gelegt wird, dass Summe der quadrierten vertikalen Abstände der Punkte von der Geraden minimiert wird

↳ $\boxed{y = Bx + A}$, dafür in TR eingeben um B und A zu finden siehe Aufg. 2

Bestimmtheitsmaß $= R^2$:

↳ Maß dafür wie gut unabhängige Variable abhängige erklärt, z.B. wie nah Regressions-kurve an den echten Werten dran ist, Werte zwischen 0 und 1

↳ R^2 auch mit TR bestimmen siehe Aufg 2. unter Reg. Gute Werte je näher an 1 desto genauer.

Normalverteilung: $= X \sim N(\mu, \sigma^2)$ (Mittelwert / Varianz)

↳ kontinuierliche Wahrscheinlichkeitsverteilung, wird oft verwendet um reale Zufallsvariablen zu beschreiben die sich um einen Mittelwert gruppieren und symmetrisch abfallen, Werte weit weg von MW immer unwahrscheinlicher

↳ kontinuierlich bedeutet Messergebnisse können beliebige Werte innerhalb Intervall annehmen

→ Bsp. Körpergröße, Blutdruck etc.

Binominalverteilung: n = Anzahl Versuche p = Wahrscheinlichkeit Erfolg bei einzelnem Versuch

↳ Diskret, nur ganze, einzelne getrennte Einheiten

→ zwei mögliche Ergebnisse: beschreibt Anzahl Erfolge in fester Anzahl unabhängiger Versuchen, wobei jeder Versuch genau zwei mögliche Ergebnisse (Erfolg, Misserfolg)

log Normalverteilung:

↳ gleich wie normale nur log, wenn diese weit gestreut sind, auch gleiches verfahren wie bei NV in TR aber z.B anstatt $P(X > 14)$ $\ln(x) > \ln(14)$ und $\ln(14)$ in TR

Rest bleibt gleich unverändert also $X \sim N(15, 1)$

Poissonverteilung:

↳ WV die die Anzahl von Ereignissen in festem Intervall beschreibt, wobei Ereignisse mit konstanter Rate eintreten und unabhängig voneinander sind

Hypergeometrische Verteilung:

↳ WV, die Wk beschreibt, k Erfolge in einer Stichprobe ohne zurücklegen aus endlicher Grundgesamtheit mit N Obj. zu erhalten, von denen k eine bestimmte Eigenschaft haben

Diskrete Dichte:

↳ Wk, dass diskrete Zufallsvariable bestimmten Wert annimmt

Verteilungsfunktion:

↳ Wk, dass Zufallsvariable Wert kleiner oder gleich einem bestimmten Wert annimmt

Kovarianzmatrix:

↳ Matrix, die die Kovarianz zwischen verschiedenen Zufallsvariablen enthält. Die Diagonalelemente sind die Varianzen der einzelnen Variablen.

Korrelationskoeffizient: ρ/r

↳ Maß für die Stärke und Richtung des linearen Zusammenhangs zwischen zwei Variablen

Ergodisch:

↳ stochastischer Prozess ist ergodisch, wenn langfristige Zeitmittelwerte und Ensemble-Mittelwerte übereinstimmen

Stationär:

↳ stationär, wenn seine statistischen Eigenschaften nicht von der Zeit abhängen

Gausscher Prozess:

↳ jede Linearkombination von Zufallsvariablen normalverteilt

1. Vertrauensbereich bestimmen mit z:

Werte Sie für die Parameter einsetzen.

z hier ablesen

Vertrauensmean:	90%	95%	99%
z:	1,645	1,96	2,576
$z_{einseitig}$:	1,282	1,645	2,33

Aufgabe 1 (4 Punkte)

In einer klinischen Studie wurde $\underbrace{\text{2.400}}_{n}$ Personen mit einer bestimmten Erkrankung ein neues Medikament verabreicht. Bei 10 (k) der Personen trat Fieber als Nebenwirkung des Medikaments auf. Aus diesen Daten soll ermittelt werden, wie hoch der Anteil der Personen mit Fieber als Nebenwirkung höchstens ist. also $z = \alpha$ und nicht $z = \frac{\alpha}{2}$, da nur einseitig an höchstem interessiert
Berechnen Sie zur Beantwortung dieser Frage einen passenden Vertrauensbereich zum Vertrauensniveau 0,95.

↳ Intervall

$p = \frac{k}{n} = \frac{10}{2400} = \frac{1}{240}$, $\alpha = 1 - 0{,}95 = 5\%$

① z aus Tabelle ablesen, hier ist z einseitig, da nur nach höchstem gefragt wurde also
↳ $z_{einseitig} = 1{,}645$

nur +, da nur einseitig nach höchstem gesucht wird

② $\boxed{s = \sqrt{p(1-p)}} = \sqrt{\frac{1}{240}\left(1-\frac{1}{240}\right)}$

③ $\boxed{\Delta p = z \cdot \frac{s}{\sqrt{n}} \pm \frac{1}{2n}} = 0{,}05 \cdot \frac{\sqrt{\frac{1}{240}\cdot\left(1-\frac{1}{240}\right)}}{\sqrt{2400}} + \frac{1}{2 \cdot 2400}$

– wird nicht beachtet

④ $\boxed{I = [\, p - \Delta p,\ p + \Delta p \,]} = \left[\, 0,\ \frac{1}{240} + 1{,}645 \cdot \frac{\sqrt{\frac{1}{240}\cdot\left(1-\frac{1}{240}\right)}}{\sqrt{2400}} + \frac{1}{4800} \,\right] = [\, 0 ; 0{,}0065 \,]$

↳ Nebenwirkung Fieber bei höchstens 0,65%

← nur rechte Seite wurde beachtet

② Lineare Regression, Bestimmtheitsmaß, Argumente ob passend:

Aufgabe 2 (13 Punkte)

In einer Untersuchung des Zusammenhangs zwischen den Marketingausgaben pro Kunde und den Umsätzen pro Kunde wurden in einer Stichprobe bei 5 Unternehmen einer bestimmten Branche die folgenden Daten erhoben.

Marketingausgaben pro Kunde in 1.000 Euro	1,2	2,6	3,4	0,7	3,7
Umsatz pro Kunde in 1.000 Euro	5,5	7,5	12,0	1,0	15,0

Bitte beachten Sie: In dieser Aufgabe sind Ergebnisse und Zwischenergebnisse mit mindestens 4 Nachkommastellen anzugeben.

a) Zeichnen Sie das zugehörige Streudiagramm in das Koordinatensystem auf der nächsten Seite ein. einfach nur Punkte einzeichnen

b) Bestimmen Sie die Gleichung der empirischen Regressionsgeraden. ← Taschenrechner einfach

c) Zeichnen Sie die in b) berechnete empirische Regressionsgerade in das Koordinatensystem (nächste Seite) ein.

d) Ein weiteres Unternehmen dieser Branche hat 950 Kunden und plant für das kommende Jahr mit einem Marketingbudget von 2,3 Mio. Euro. Welchen Umsatz (in Mio. Euro) sagt die lineare Regression für dieses Unternehmen voraus? Marketingbudget : Kunden = x , x in Reg einsetzen

Hinweise: Hier ist nicht nur der Umsatz pro Kunde zu berechnen, sondern auch der (gesamte) Umsatz.
Geben Sie diesen in Mio. Euro an.

ß der linearen Regression.

f) Die lineare Regression ist hier gut geeignet, um den Zusammenhang zwischen den Marketingausgaben pro Kunde und dem Umsatz pro Kunde zu beschreiben. Woran kann man das feststellen? Geben Sie verschiedene Argumente an.

g) Für die logarithmische Regression ergibt sich die Gleichung $y = 7{,}1643\ln(x) + 3{,}4525$ sowie das Bestimmtheitsmaß $R^2 \approx 0{,}8906$.

Eignet sich die logarithmische Regression besser als die lineare Regression, um den Zusammenhang zwischen Marketingausgaben pro Kunde und Umsatz pro Kunde zu beschreiben?

Begründen Sie Ihre Antwort. Antworten ohne Begründung werden nicht bewertet.

b) bestimmen der Gleichung der empirischen Regressionsgeraden:

① In Taschenrechner eingeben:

1. Mode -> 2 -> Statistik modus
2. 2 auswählen: A + BX
3. in Tabelle alles eintragen, AC drücken um Tabelle zu verlassen
4. Shift + 1 -> 5 regression auswählen, dann A und B auswählen
   -> A = −1,0364
   -> B = 3,9812

↳ in diese Form einfügen: $\boxed{y = Bx + A}$ ← hier ist es vertauscht aufpassen nicht ax wie immer!!

↳ $y = 3{,}9812x - 1{,}0364$

d) Voraussage mit Regressionsgerade:

$\frac{2300000}{950} = \frac{46000}{19} \approx 2421$

↳ 2,421 in Regressionsgerade einsetzen:

↳ $y = 3{,}9812 \cdot 2421 - 1{,}0364$
$= 8{,}602$ pro Kunde Umsatz

Gesamtumsatz $= 8602 \cdot 950 = 8171980{,}94$

e) Bestimmtheitsmaß:

↳ r am TR: r = 0,9622 ← Ax+b in Tabelle x,y ← dann Reg: ← r

-> $R^2 = r^2 = 0{,}9622^2 = 0{,}9260$

f) Argumente wann Regressionsmodell passend?

-> Bestimmtheitsmaß ist hoch 0,926 sehr nah an 1
-> Im Diagramm Regressionslinie nah an echten Datenwerten

g) log Regression besser geeignet?:

Nein, das Bestimmtheitsmaß ist sogar noch geringer

③ WK ausrechnen mit Normalverteilung:

Aufgabe 3 (7 Punkte)

In der Versandabteilung eines Unternehmens werden elektronische Bauteile eines bestimmten Typs gereinigt und verpackt. Für die Zeit X (in Sekunden) für das Reinigen bzw. die Zeit Y (in Sekunden) für das Verpacken gelte $X \sim N(15;\ 1)$ bzw. $Y \sim N(27;\ 0{,}8)$. Die Zufallsvariablen X und Y sind dabei unabhängig. N bedeutet Normalverteilt mit Mittelwert $\mu = 15$ sek und Varianz $\sigma^2 = 1$

a) Mit welcher Wahrscheinlichkeit dauert das Reinigen eines Bauteils weniger als 14 Sekunden? CDF in TR $P(x < 14)$ will man hier rechnen

b) Mit welcher Wahrscheinlichkeit dauert das Verpacken eines Bauteils mehr als 29 Sekunden?

c) Mit welcher Wahrscheinlichkeit dauern das Reinigen und Verpacken eines Bauteils zusammen höchstens 44 Sekunden?

$P(x > 29)$ gesucht

a) Mit welcher Wahrscheinlichkeit dauert Reinigen Bauteil weniger als 14 sek?

↳ $P(X < 14)$ ← das hier ist gefragt Wahrscheinlichkeit P dass kleiner 14!

↳ in TR eingeben:

→ Mode → 4 (Dist) → 2 (Normal CD) in N(15, 1) wird immer N(μ, σ^2) angegeben aber es ist nur σ gefragt deshalb √

→ $\underbrace{\text{Lower} = 0}_{\text{Anfang} = 0\text{, da bei } 0 \text{ sek startet}}$ → $\underbrace{\text{Upper} = 14}_{\text{das ist das höchst mögliche Wert}}$ → $\underbrace{\sigma = \sqrt{1}}_{\text{Stdabw. } \sigma = 1 \text{, steht in Klammer}}$ → $\underbrace{\mu = 15}_{\text{Normalverteilung Mittelwert } \mu = 15}$

↳ $P(x < 14) \approx 0{,}1587$

b) Mit welcher Wahrscheinlichkeit Verpacken Bauteil mehr als 29 sek?

$P(Y > 29) \approx 0{,}1267$ beliebig hohen Wert z.B. 99 Hauptsache größer als MW und 29

↳ in TR: Lower = 29 → Upper = 99 → $\sigma = \sqrt{0{,}8}$ → $\mu = 27$

Wurzel nie vergessen

c) mit Welcher Wahrscheinlichkeit dauern verpacken und reinigen max 44 Sekunden?:

↳ einfach X + Y zusammenrechnen, also: $P(X+Y < 44)$

↳ hierfür auch die Normalenformen zusammenrechnen $X+Y \sim N(15+27;\ 1+0{,}8) = (42;\ 1{,}8)$

Wurzel nie vergessen!

↳ TR: Lower = 0 → Upper = 44 → $\sigma = \sqrt{1{,}8}$ → $\mu = 42$

↳ $P(X+Y > 44) \approx 0{,}932$

6. Poissonverteilung:

Poissonverteilung

X~Po(18) Poissonverteilung für WK in bekannten Intervall (18 mal pro Jahr) ← Erwartungswert 18

P(X>20) gesucht

Aufgabe 4 (8 Punkte)

Ein bestimmter Maschinentyp fällt im Mittel 18-mal pro Jahr aus.

① Unterscheiden WK →
a) Wie wahrscheinlich ist es, dass eine Maschine dieses Typs in einem Jahr mehr als 20-mal ausfällt?

EW →
b) Wie hoch ist die erwartete Anzahl von Ausfällen einer Maschine dieses Typs pro Monat?

② Unterscheiden: CD (Summe) PD (einzeln) →
c) Wie wahrscheinlich ist es, dass eine Maschine dieses Typs im nächsten Monat gar nicht ausfällt?

X~Po(18) kann nicht genutzt werden, da pro Monat nicht pro Jahr! Y~Po(1,5)

③ Unterscheiden: ↳ Zeitraum X~Po() (Jahr) / Y~Po() (Monat)
d) Ein Unternehmen betreibt unabhängig voneinander zwei Maschinen dieses Typs. Wie wahrscheinlich ist es, dass beide Maschinen innerhalb des nächsten Monats nicht ausfallen?

a) Wahrscheinlichkeit, dass Maschine mehr als 20 mal ausfällt:

↳ Es ist gegeben Anzahl Ausfälle pro Jahr: $X \sim Po(18)$

↳ $P(X>20) = 1 - P(X \leq 20) \approx 1 - 0{,}7307 \approx 0{,}2693$ ← man berechnet das gesuchte mit der Gegenwahrscheinlichkeit

↳ in TR:

CD wenn mehrere Dinge zusammengerechnet werden (cummulativ also Summe)

→ Mode → 4 (Dist) → 3 (Poisson CD) → 2 (Var) → X = 20 → μ = 18

Wann CD wann PD verwenden bei Taschenrechner? ← Wie PD/CD nutzen im TR?

↳ CD immer wenn "weniger/mehr" steht und PD wenn "genau" steht

b) Wie hoch ist die erwartete Anzahl von Ausfall Maschine pro Monat?

↳ "bestimmter Maschinentyp fällt im Mittel 18 mal pro Jahr aus"

↳ $X \sim Po(18)$ gibt bereits Erwartungswert!

↳ also Erwartete Anzahl Ausfälle pro Monat $= \frac{18}{12} = 1{,}5$

c) Wie wahrscheinlich ist es, dass Maschine im nächsten Monat garnicht ausfällt?

↳ $Y \sim Po(1{,}5)$ ← $X \sim Po(18)$ kann nicht mehr genutzt werden da nach pro monat gefragt

→ $P(Y=0) \approx 0{,}2231$ ← hier wurde gemacht wenn kein Fehler also $P(Y=0)$

↳ in TR:

PD, da nach einzelnem Wert gesucht

→ Mode → 4 (Dist) → 2 (Poisson PD) → 2 (Var)

d) Wahrscheinlichkeit dass beide Maschinen innerhalb von nächsten Monat nicht ausfallen:

↳ Pfadmultiplikation mit $P(Y=0)$:

↳ $P(\text{Beide Maschinen fallen nicht aus}) = P(Y_1=0) \cdot P(Y_2=0) = 0{,}2231 \cdot 0{,}2231 \approx 0{,}0498$

5.) t-Test Hypothesentest einseitig: --> t-Test immer wenn Stichprobe ≤ 30 ist!

Aufgabe 5 (8 Punkte)

Eine Maschine füllt Getränke in 1-Liter-Flaschen ab; dabei ist die Abfüllmenge normalverteilt.

Der Produktionsleiter vermutet allerdings, dass die Maschine die Flaschen im Mittel zu gering befüllt. Um dies zu überprüfen, entnahm er eine Stichprobe abgefüllter Flaschen und maß die Abfüllmengen nach. Dabei ergaben sich die folgenden Füllmengen (in ml):

1.000 1.005 992 998 995 999 999 994

Kann die Vermutung des Produktionsleiters auf Basis dieser Stichprobe bestätigt werden?

Führen Sie einen geeigneten statistischen Test zum Signifikanzniveau 1 % durch und geben Sie mit Hilfe der Testentscheidung eine Antwort auf diese Frage.

n \ α	0,10 0,20	0,05 0,10	0,025 0,05	0,01 0,02	0,005 0,01	0,001 0,002	0,0005 0,001
1	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	1,886	2,920	4,303	6,965	9,925	22,33	31,56
3	1,638	2,353	3,182	4,541	5,841	10,22	12,92
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	1,303	1,684	2,021	2,423	2,704	3,307	3,551
42	1,302	1,682	2,018	2,418	2,698	3,296	3,538
44	1,301	1,680	2,015	2,414	2,692	3,286	3,526
46	1,300	1,679	2,013	2,410	2,687	3,277	3,515
48	1,299	1,677	2,011	2,407	2,682	3,269	3,505
50	1,299	1,676	2,009	2,403	2,678	3,261	3,496
55	1,297	1,673	2,004	2,396	2,668	3,245	3,476
60	1,296	1,671	2,000	2,390	2,660	3,232	3,460
65	1,295	1,669	1,997	2,385	2,654	3,220	3,447
70	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	1,292	1,664	1,990	2,374	2,639	3,195	3,416
90	1,291	1,662	1,987	2,368	2,632	3,183	3,402
100	1,290	1,660	1,984	2,364	2,626	3,174	3,390
150	1,287	1,655	1,976	2,351	2,609	3,145	3,357
200	1,286	1,653	1,972	2,345	2,601	3,131	3,340

↳ Man soll schauen ob die Hypothese des Leiters richtig ist:

↳ $H_0: \mu_0 \geq 1000$ ← die Nullhypothese ist, denn ganz Normal 1 Liter

↳ $H_1: \mu_1 < 1000$ ← Leiter sagt Hypothese, dass kleiner als 1000 ist

↳ $\alpha = 0{,}0$ ← Signifikanzniveau α muss hier nicht zu z umgerechnet werden, da nicht z-test! (z-test immer wenn Stichprobe n > 30 ist)

→ Formel berechnen:

$\boxed{SZB = H_0 \pm t_{7,0{,}99} \cdot \frac{s}{\sqrt{n}}}$ → da nach unterer Grenze gesucht nur − verwenden!

↳ t aus Tabelle ablesen aber aufpassen, wenn nun für n−1 ablesen, also nicht bei t=8 sondern t=7 mit α = 0,01 → 2,998

↳ $\sqrt{n}$ ist die Anzahl der Werte in der Stichprobe also $\sqrt{8}$

s ist Signifikanzniveau α aber bei Stichproben wird s bezeichnet

↳ s ausrechnen: Mode ⇒ 2 → 2 (A+BX) → X Spalte auffüllen mit allen 8 Stichproben → Shift + 1 → 4 (Var) → 4 (sx) = 4,0620

→ $\bar{x}$ ausrechnen: gleich wie s aber am Ende statt 4 zwei auswählen für $\bar{x} = 997{,}75$

↳ nun kann man alles einsetzen und vergleichen:

Zufallsstreubereich; $\mu = 1000$ in H_0; ∞ da ja H_0 $\mu \geq 1000$

ZSB für $\bar{x}$, falls H_0 gilt: $\left[1000 - 2{,}998 \cdot \frac{4{,}0620}{\sqrt{8}}, \infty\right) \approx [995{,}6945, \infty)$ → bei oberer Grenze wäre es einfach 1000+6

↳ ZSB gibt an dass alles größer als 995,6945 noch akzeptiert werden kann, da Mittelwert $\bar{x} = 997{,}75$ darüber liegt, stimmt die Nullhypothese und wird nicht abgelehnt

⑥ Hypergeometrische Verteilung: –> Ziehen ohne zurücklegen, da man ein Gerät ja nicht nochmal testet nachdem es getestet wurde

Aufgabe 6 (3 Punkte)

In einem Unternehmen wurde eine Charge von 50 elektronischen Baugruppen gefertigt. In der Endkontrolle wird eine Stichprobe von 10 dieser Baugruppen auf Funktionsfähigkeit geprüft.

Wie groß ist die Wahrscheinlichkeit, dass in der Stichprobe mindestens 2 nicht funktionsfähige Baugruppen sind, wenn die Charge 3 nicht funktionsfähige Baugruppen enthält?

← Hypergeometrisch da endliche Gruppe von 50 und man 10 zieht zum testen, diese nicht zurücklegt

↳ WK berechnen für Stichprobe 10 mindestens 2 nicht funktionsfähige:

↳ Hypergeometrische Verteilung $X \sim H(\underbrace{10}_{\text{Stichprobengröße } n=10}, \underbrace{50}_{\text{Grundgesamtheit } N=50}, k)$ Erfolge $k=3$ nicht funktionsfähig

↳ Es ist also $P(X \geq 2)$ gesucht, also dass mindestens 2 kaputt sind, d.h. man könnte nun mit der Gegenwahrscheinlichkeit rechnen also $P(X \geq 2) = 1 - (P(X=1) + P(x=0))$ oder den Ansatz direkter Ansatz wie viele nicht funktionierende es sein können, da es ja nur 3 nicht funktionierende gibt kann man sagen $\boxed{P(X \geq 2) = P(X=2) + P(X=3)}$

↳ bei Hypergeometrischer Verteilung berechnet man einzelne Wahrscheinlichkeiten anders:

$$\boxed{P(X=k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}}$$

N = Grundgesamtheit, n = Stichprobe,
K = nicht funktionsfähig Gesamt, k = nicht funktionsfähig in Stichprobe (2;3)

↳ $P(x=2) = \frac{\binom{3}{2} \cdot \binom{50-3}{10-2}}{\binom{50}{10}} = \frac{\binom{3}{2} \cdot \binom{47}{8}}{\binom{50}{10}} = \frac{3C2 \cdot 47C8}{50C10} = \frac{9}{98}$

nicht vergessen als Verhältnis zu berechnen!

↳ $\boxed{\begin{array}{l}\text{Binominalkoeffizienten } \binom{n}{k} \text{ in TR eingeben!} \\ \text{–> Shift + nCR } (\div) \text{ –> man erhält C –> } \binom{n}{k} = nCk \text{, also z.B. } \binom{47}{8} = 47C8 \text{ in TR eingeben!}\end{array}}$

↳ $P(x=3) = \frac{\binom{3}{3} \cdot \binom{47}{7}}{\binom{50}{10}} = \frac{3C3 \cdot 47C7}{50C10} = \frac{3}{490}$

↳ vollständige Lösung:

WK min 2 defekt Hypergeometrisch –> $\boxed{P(X \geq 2) = P(x=2) + P(x=3) = \frac{\binom{3}{2} \cdot \binom{47}{8}}{\binom{50}{10}} + \frac{\binom{3}{3} \cdot \binom{47}{7}}{\binom{50}{10}} = \frac{9}{98} + \frac{3}{490} = \frac{24}{245} \approx 0{,}0980}$

7 Diskrete Dichte:

Diskrete Dichte gibt an mit welcher Wahrscheinlichkeit die Zufallsvariable X einen bestimmten Wert annimmt

Aufgabe 7 (6 Punkte)

Die untenstehende Tabelle zeigt die diskrete Dichte einer Zufallsvariablen X.

a) Berechnen Sie den Erwartungswert und die Varianz von X.

↳ Gibt Streuung um Mittelwert / Erwartungswert an

k	$P(X=k)$
-2	0,1
0	0,3
1	0,5
3	0,1

b) Berechnen Sie den Wert der Verteilungsfunktion von X an der Stelle 2.
Was besagt dieser Wert?

a) Erwartungswert und Varianz berechnen:

$$\boxed{\begin{aligned} E(X) &= \sum k \cdot P(X=k) = -2 \cdot 0{,}1 + 0 \cdot 0{,}3 + 1 \cdot 0{,}5 + 3 \cdot 0{,}1 = 0{,}6 \\ V(x) &= \left(\sum k^2 \cdot P(X=k)\right) - \mu^2 = \left((-2)^2 \cdot 0{,}1 + 0^2 \cdot 0{,}3 + 1^2 \cdot 0{,}5 + 3^2 \cdot 0{,}1\right) - 0{,}6^2 = 1{,}44 \end{aligned}}$$

b) Wert der Verteilungsfunktion von X an der Stelle 2 berechnen:

↳ Was ist Verteilungsfunktion: F(x) gibt Wk an dass Zufallsvariable X annimmt der ≤ x ist

↳ Wahrscheinlichkeiten werden bis zu diesem Punkt x aufsummiert

↳ $\boxed{F(x) = P(X \leq x) = \sum_{k \leq x} P(X=k)}$ → aufsummieren aller Wk $P(X=k)$ für $k \leq x$

→ also F(2):

$$\boxed{F(2) = P(X=-2) + P(X=0) + P(X=1) = 0{,}1 + 0{,}3 + 0{,}5 = 0{,}9}$$

→ Was sagt dieser Wert aus?:

F(2) ist Wk, dass Zufallsvariable X einen Wert kleiner oder gleich 2 annimmt oder:

F(2) ist Wk, dass Zufallsvariable X einen der Werte −2, 0, 1 annimmt

8) Kovarianzmatrix und Korrelationskoeffizienten:

Aufgabe 8 (4 Punkte)

Der Zufallsvektor $\binom{X}{Y}$ hat die Standardabweichungen $\sigma_X = 2$ und $\sigma_Y = 3$, und der Korrelationskoeffizient zwischen X und Y beträgt $\rho = 0{,}6$.
X ist eine normalverteilte Zufallsvariable.

a) Stellen Sie die Kovarianzmatrix für $\binom{X}{Y}$ auf.
b) Sind die Verteilungen von X und Y unabhängig? (Begründung)
c) Nun wird der Zufallsvektor $\binom{Z}{Y}$ betrachtet.
Für den Korrelationskoeffizienten zwischen Y und Z gilt $\rho = 0$. Z ist normalverteilt.
Kann davon ausgegangen werden, dass Z und Y unabhängig sind? (Begründung)

a) Kovarianzmatrix:

allgemeine Formel: $\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} = \begin{pmatrix} 2^2 & 3{,}6 \\ 3{,}6 & 3^2 \end{pmatrix} = \begin{pmatrix} 4 & 3{,}6 \\ 3{,}6 & 9 \end{pmatrix}$

σ_{xy} berechnen: $\sigma_{xy} = \rho_{xy} \cdot \sigma_x \cdot \sigma_y = 0{,}6 \cdot 2 \cdot 3 = 3{,}6$

b) sind X und Y unabhängig?:

↳ Nein, da die Kovarianz und damit auch ihre Korrelation nicht gleich null sind.

Der Korrelationskoeffizient ρ ist $\rho = 0{,}6$ und nicht $\rho = 0$, sind X und Y nicht unabhängig

c) Zwischen Y und Z gilt $\rho = 0$. Z ist normalverteilt. Sind Z und Y unabhängig?:

↳ Nur weil sie unkorreliert sind kann man nicht direkt davon ausgehen, dass sie unabhängig sind

Es herrscht keine lineare Beziehung, aber könnten immer noch nicht lineare Beziehung haben

Man weiß nicht ob Y normalverteilt ist.

↳ Nur wenn Y und Z normalverteilt und Korrelation = 0 ist kann man ausgehen dass unabhängig

9) ergodisch, stationär, gaussscher Prozess:

Aufgabe 9 (3 Punkte)

Wir betrachten den folgenden Stochastischen Prozess $u(t)$, für den gilt

$$u(0) = u_0 \qquad \text{und}$$
$$u(t) = u_0 \cdot X(t)^2 \qquad \text{für } t > 0, \text{wobei}$$

u_0 zum Zeitpunkt $t = 0$ zufällig gewählt wird mit $u_0 \sim N(0; 1)$ und $X(t) \sim N\left(0; \frac{1}{t}\right)$.

Mittelwert = 0 Varianz = 1

↳ Normalverteilung mit

Begründen Sie Ihre Antworten auf die folgenden Fragen. Antworten ohne Begründung werden nicht gewertet.

a) Ist der Prozess ergodisch?
b) Ist der Prozess stationär?
c) Nun wird $u_0 = 1$ konstant gesetzt und die Zufallsvariablen
$$Y_1 = u(1); \quad Y_2 = u(2); \quad \ldots \quad Y_n = u(n)$$
betrachtet.
Handelt es sich bei Y_i um einen Gauß'schen Prozess?

a) ergodisch?:

↳ Definition ergodisch: Zeitliche Mittelwerte und ensemble-Mittelwerte gleich. D.h. Statistische Eigenschaften des Prozesses bleiben über die Zeit konstant und hängen nicht vom Anfangszustand ab

↳ in diesem Fall:

↳ nicht ergodisch, die angenommenen Werte hängen vom Anfangswert ab, da u_0 verschiedene Werte annehmen kann, wird sich Prozess über die Zeit verändern.

b) Stationär?:

↳ Definition: Stationär, wenn statistische Eigenschaften wie Mittelwert und Varianz zeitunabhängig sind

↳ in diesem Fall:

↳ nicht stationär, da sich X(t) und somit auch u(t) im Laufe der Zeit ändert

Die Varianz wird kleiner, was sich direkt auf die Varianz des Prozesses auswirkt

c) Gaussscher Prozess:

↳ Definition: gaussscher (normalverteilter) Prozess sind Zufallsvariablen sind alle normalverteilt und jede lineare Kombination dieser Zufallsvariablen ist ebenfalls normalverteilt

↳ in diesem Fall:

u(t) ist nicht normalverteilt, da es sich um Produkt einer normalverteilten Variable u_0 mit dem Quadrat einer normalverteilten Variable X(t) handelt.

↳ Quadrat normalverteilter Variable ist nicht normalverteilt, sondern folgt Chi-Quadrat-Verteilung, sodass u(t) nicht normalverteilt ist

WS 22/23

① Korrelationskoeffizient, Bestimmtheitsmaß, lineare Regression:

Aufgabe 1 (10 Punkte)

Die Ernährungs- und Landwirtschaftsorganisation der Vereinten Nationen FAO veröffentlicht monatlich verschiedene Preisindizes. Zur Zeit sind die Indizes so normiert, dass der Gesamt-Preisindex für 2014 bis 2016 durchschnittlich 100 beträgt.

In dieser Aufgabe sollen Zusammenhänge zwischen dem Preisindex für Milchprodukte und dem Preisindex für Getreide untersucht werden.

Hinweis: Die beiden Aufgabenteile a) und b) können unabhängig voneinander bearbeitet werden.

a) Die nachstehende Tabelle enthält die Werte der beiden oben genannten Indizes für die Jahre 2015 bis 2019.

Jahr i	Preisindex Milchprodukte x_i	Preisindex Getreide y_i
2015	89,1	98,0
2016	87,9	94,0
2017	111,1	93,6
2018	105,4	99,0
2019	103,4	97,2

a1) Berechnen Sie den empirischen Korrelationskoeffizienten zwischen den beiden Preisindizes.
Hinweis: Hier genügt die Angabe des Wertes. Es ist nicht erforderlich, die Formel anzugeben, mit der dieser Wert berechnet werden kann.

a2) Berechnen Sie das zu a1) gehörende Bestimmtheitsmaß der linearen Regression.

a3) Ist es sinnvoll, den Zusammenhang zwischen den beiden Preisindizes im Zeitraum von 2015 bis 2019 durch eine lineare Regression zu beschreiben?

Begründen Sie Ihre Antwort. Antworten ohne Begründung werden nicht gewertet.

a) Korrelationskoeffizienten r berechnen:

in TR: Mode -> 2 -> 2 (A+BX) -> x Milchindizes eintragen -> y Preisindizes eintragen -> AC -> Shift+1 -> 5 -> 3 (r)

↳ $r \approx -0{,}0284$

a2) Bestimmtheitsmaß R^2 berechnen:

↳ $R^2 = r^2 = (-0{,}0284)^2 = 0{,}0008$

a3) Ist es sinnvoll lineare Regression zwischen Indizes?:

↳ Nein da Bestimmtheitsmaß $R^2 = 0{,}0008$ nahezu 0 ist

Die nachstehende Tabelle enthält die Werte der beiden Indizes für die Jahre 2020 bis 2022.

Jahr i	Preisindex Milchprodukte x_i	Preisindex Getreide y_i
2020	102,9	104,2
2021	118,5	130,5
2022	140,6	152,9

b1) Berechnen Sie die Gleichung der empirischen Regressionsgeraden, die angibt, wie im Zeitraum von 2020 bis 2022 der Index der Getreidepreise y von dem Index der Preise der Milchprodukte x abhing.
Hinweis: Hier genügt die Angabe des Wertes. Es ist nicht erforderlich, die Formel anzugeben, mit der dieser Wert berechnet werden kann.

b2) Angenommen, der Preisindex für Milchprodukte steigt im Jahr 2023 auf 160. Welchen Wert sagt die lineare Regression b1) dann für den Preisindex Getreide voraus?
Geben Sie auch an, wie Sie diesen Wert berechnet haben.

b3) Das Bestimmtheitsmaß für die lineare Regression b1) liegt bei 0,979. Welche der folgenden Aussagen über einen möglichen kausalen (= ursächlichen) Zusammenhang zwischen den Indizes im Zeitraum von 2020 bis 2022 ist am plausibelsten?

i) Der Preisanstieg für Milchprodukte hat den Preisanstieg für Getreide verursacht.

ii) Der Preisanstieg für Getreide hat den Preisanstieg für Milchprodukte verursacht.

iii) Es gibt eine dritte Einflussgröße (oder mehrere weitere Einflussgrößen), die sowohl den Preisanstieg für Milchprodukte als auch den Preisanstieg für Getreide verursacht hat.

iv) Keine der Antworten i) bis iii) ist richtig; das hohe Bestimmtheitsmaß ist reiner Zufall.

Wählen Sie die plausibelste Aussage aus und begründen Sie Ihre Wahl mit einer kurzen Erläuterung. Falls Sie z. B. iii) für richtig halten, nennen Sie mindestens eine gemeinsame Ursache für den Preisanstieg bei beiden Indizes.

b) Regressionsgerade bestimmen:

↳ TR: 2 -> 2 -> x,y eingabe -> AC -> Sh+1 -> 5 -> A/B

B kommt für m

↳ $y = mx + k$ ← A kommt für k!

$= Bx + A$ ← A und B vertauscht aufpassen!!!

$= 1{,}2732x - 24{,}4347$

b2) Preis für Milchprodukte 160 in 2023 wie sieht Preis für Getreide dann aus?

↳ $y(160) = 1{,}2732 \cdot 160 - 24{,}4347 = 179{,}3$

b3) iii am plausibelsten, Corona Krise könnte verantwortlich sein, also Abhängig von anderem Faktor

② Normaler Basic Ansatz ohne spezielle Verteilung, Pfadmultiplikation, Gegenwahrscheinlichkeit:

Aufgabe 2 (8 Punkte)

In einem Unternehmen wurde ein Gegenstand bisher auf einer alten Maschine gefertigt, deren Ausschussquote bei 5,7% lag. Diese Maschine soll künftig durch eine neue Maschine ersetzt werden, deren Ausschussquote nur 1,3% beträgt.

Derzeit stehen beide Maschinen zur Verfügung. Zu Kontrollzwecken hat die Produktionsleitung je einen Gegenstand aus der Fertigung der alten und der neuen Maschine entnommen.

Hinweis: Geben Sie die Ergebnisse der Aufgabenteile a) - c) auf 5 Nachkommastellen genau an.

a) Wie wahrscheinlich ist es, dass beide Gegenstände Ausschussstücke sind?

b) Wie wahrscheinlich ist es, dass mindestens einer der beiden Gegenstände Ausschuss ist?

c) Wie wahrscheinlich ist es, dass beide Gegenstände intakt sind?

d1) Welche statistische Annahme haben Sie bei a) - c) treffen müssen, um rechnen zu können?

d2) Ist die Annahme aus d1) auch plausibel? - Begründen Sie Ihre Antwort. Antworten ohne Begründung werden nicht gewertet.

↳ Ansatz: Wk schon gegeben, d.h. kein spezielle H, N, Po-Verteilung

a) Wk beide Stücke Ausschuss: ← einfach Pfadmultiplikation

P(beide Ausschuss) = P(ausschuss alte Maschine) · P(ausschuss neue Maschine) = $0{,}057 \cdot 0{,}013 = 0{,}00074$

b) Wk mindestens eine der beiden:

↳ Ansatz Gegenwahrscheinlichkeit:

Wk dass keins kaputt ist ist jeweils 94 %, also schon sehr hoch

Wk bei neuer keins kaputt auch einfach mit Gegenwahrscheinlichkeit berechnen

↳ P(mindestens 1 Ausschluss) = 1 − P(beide Intakt) = 1 − (1 − P(Ausschuss Alte)) · (1 − P(Ausschuss neue))

$= 1 - ((1-0{,}057) \cdot (1-0{,}013)) = 1 - \underbrace{0{,}930741}_{\text{Wk beide noch gut}} \approx 0{,}06926$ ← Wk min 1 kaputt

↳ Ansatz Inklusion-Exklusion:

Entweder altes oder neues kaputt einfach Formel dafür, funktioniert hier da ja min 1 gefragt ist

P(A oder B) = P(A) + P(B) − P(A und B) = $0{,}057 + 0{,}013 - 0{,}057 \cdot 0{,}013 \approx 0{,}06926$

c) Wk das beide Intakt sind:

↳ Gegenwahrscheinlichkeit: P(beide Intakt) = 1 − P(beide kaputt) = $1 - 0{,}057 \cdot 0{,}013 = 1 - 0{,}06926 = 0{,}9307$

d1) Welche Annahme musste man treffen um rechnen zu können?:

↳ Unabhängigkeit der beiden

d2) plausibel, entstehen Defektstücks bei der einen Maschine hat keinen Einfluss auf entstehen Defektstück bei anderer

3. Binominalverteilung:

Aufgabe 3 (5 Punkte)

Eine Fluggesellschaft hat in der Vergangenheit die Erfahrung gemacht, dass auf einer bestimmten Strecke nur 93% der Personen, die einen Flug gebucht haben, diesen tatsächlich auch antreten. → Wh = 93%

Für die nachfolgenden Rechnungen soll angenommen werden, dass der Flugantritt verschiedener Personen unabhängig voneinander erfolgt. } also Binominalverteilung

a) Für den 15.02.23 haben nur 85 Personen diesen Flug gebucht. → n = 85
Wie wahrscheinlich ist es, dass genau 80 Personen den Flug antreten? → PD in TR bei Binominalverteilung

b) Am 20.02.23 ist die Strecke stark nachgefragt; die Fluggesellschaft hat 208 Buchungen verkauft. Das an diesem Tag eingesetzte Flugzeug hat allerdings nur 200 Sitzplätze.
Wie wahrscheinlich ist es, dass alle Personen, die an diesem Tag den Flug antreten wollen, auch einen Sitzplatz erhalten? Mit anderen Worten: Wie wahrscheinlich ist es, dass höchstens 200 Personen den Flug antreten wollen?
↳ Binominal CD verwenden in TR

c) Wie hoch ist die erwartete Anzahl von Personen, die den Flug antreten, wenn die Fluggesellschaft 208 Buchungen verkauft? ↳ Erwartete Anzahl entspricht Erwartungswert / Mittelwert also μ

↳ Welche Ansatz verwenden?:

↳ Wh = 93% gegeben, sind unabhängig, Anzahl n = 85

↳ Es handelt sich um Binominalverteilung

a) Wh dass genau 80 Personen antreten $X \sim B(85; 0{,}93)$:

was gemeint ist: Anzahl "Versuche", also Anzahl der Leute

TR: Mode → 4 → 4 (Binominal PD) → 2 (Var) → X = 80 → N = 85 → p = 0,93 → 0,1660

↖ Wh in Kommaschreibweise

↳ $P(X=80) \approx 0{,}1660$

b) Wh dass von von 208 Buchungen höchstens 200 Personen antreten $X \sim B(208; 0{,}93)$:

TR: Mode → 4 → 1 (Binominal CD) → 2 (Var) → X = 200 → N = 208 → p = 0,93 → 0,9802

↳ $P(X \leq 200) \approx 0{,}9802$

c) Wie hoch erwartete Anzahl Personen wenn Fluggesellschaft 208 Buchungen verkauft:

↳ Erwartungswert gefragt!

→ $\boxed{E(x) = n \cdot p = 208 \cdot 0{,}93 = 193{,}44}$

6.) Bayes Theorem, Baumdiagramm:

Aufgabe 4 (7 Punkte)

In Deutschland sind 15 Prozent der Kinder und Jugendlichen zwischen 3 und 17 Jahren übergewichtig. 70 Prozent aller übergewichtigen Kinder sind später als junge Erwachsene ebenfalls übergewichtig. Von den nicht übergewichtigen Jugendlichen werden 25 Prozent als junge Erwachsene übergewichtig.

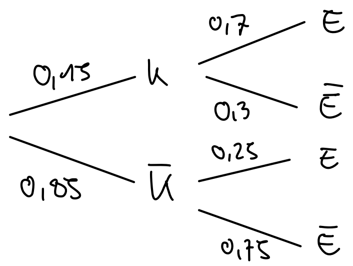
a) Veranschaulichen Sie den Sachverhalt in einem Baumdiagramm.

b) Berechnen Sie den Anteil übergewichtiger junger Erwachsener. –> Einfach Pfadmultiplikation und Addition

c) Sie treffen einen übergewichtigen jungen Erwachsenen. Wie groß ist die Wahrscheinlichkeit, dass er als Jugendlicher bereits übergewichtig war? –> Bayes Theorem

a) K = als Kind übergewichtig, E = als Erwachsener übergewichtig

$\bar{K}$ = nicht übergewichtig Kind $\bar{E}$ = nicht übergewichtig Erwachsener



b) Anteil übergewichtiger junger erwachsener berechnen:

↳ Einfach Pfadmultiplikation und Addition:

↳ $P(\text{übergewichtig Erwachsen}) = 0{,}15 \cdot 0{,}7 + 0{,}85 \cdot 0{,}25 = 0{,}3175$

c) Wie hoch WK dass jugendlicher schon als Kind übergewichtig war?

↳ Ansatz Bayes Theorem:

Pfad, dass Kind überg. auch als Erwachsener übergewichtig ist

↳ $P(K|E) = \frac{K \wedge E}{P(E)} = \frac{0{,}15 \cdot 0{,}7}{0{,}3175} = 0{,}33$

WK der erwachsenen übergewichtigen

⑤ Normalverteilung Anteile ausrechnen; Upper Lower anpassen!

Aufgabe 5 (11 Punkte)

0,45 kann man so in TR eingeben, da TR Standardabweichung will und nicht Varianz

In einem Unternehmen werden Schrauben produziert, deren Länge normalverteilt ist mit (*) Erwartungswert $\mu = 50$ mm und Standardabweichung $\sigma = 0{,}45$ mm. → Normalverteilung $X \sim N(50; 0{,}45^2)$

In N() Klammer wird immer Varianz angegeben, deshalb quadriert

a) Wie groß ist der Anteil der Schrauben, deren Länge 49,8 mm unterschreitet? Geben Sie Ihre Antwort in Prozent mit 2 Nachkommastellen an. → TR: NV mit CD!

b) Ein potentieller Kunde möchte einen größeren Fertigungsauftrag erteilen. Er verlangt jedoch, dass mindestens 95% der Schrauben eine Länge zwischen 49,4 mm und 51,4 mm haben. → Upper Lower anpassen!!!

b1) Weisen Sie nach, dass mit der derzeitigen Einstellung (siehe (*) oben) die Gesamtproduktion die Forderung des Kunden nicht erfüllt.

b2) Nach Angaben eines Technikers ist es aber möglich, die Fertigungsmaschine so neu einzustellen, dass die Fertigung mit (**) Erwartungswert $\mu = 50{,}5$ mm erfolgt. Die Standardabweichung $\sigma = 0{,}45$ mm würde sich dabei nicht ändern. Erfüllt die Gesamtproduktion mit der geänderten Einstellung (**) die Forderung des Kunden?

b3) Wenn man μ frei wählen könnte: Auf welchen Wert sollte man (bei unveränderter Fertigungsgenauigkeit $\sigma = 0{,}45$) den Erwartungswert μ der Fertigungsmaschine einstellen, damit möglichst viele Schrauben eine Länge zwischen 49,4 mm und 51,4 mm haben? → Mittelwert auf Mitte von 49,4 und 51,4 setzen, also $\mu = \frac{49{,}4 + 51{,}4}{2}$

c) Welche Länge wird von 90% der Schrauben mindestens erreicht? → 10% Quantil gesucht → z Wert aus Tabelle ablesen 90% einseitig = 1,282

a) Wie groß Anteil Schrauben, deren Länge 49,8 mm unterschreitet?

obwohl nach Anteil gefragt rechnet man trotzdem ganz normal wie wenn nach WK gesucht

↳ $P(X < 49{,}8) \approx 0{,}3284$

kann auch −1000 sein, hauptsache weit entfernt!

Std.abw. diese war schon gegeben, deshalb hier keine Wurzel ziehen

↳ In TR: Mode → 4 → 2 (NV CD) → Lower = 0 → Upper = 49,8 → $\sigma = 0{,}45$ → 0,3284

b) mindestens 95% sollen zwischen 49,4 mm und 51,4 mm sein. Ist das möglich?:

↳ Einfach nur Upper/Lower anpassen! (nicht Ansatz Vertrauensbereich mit Z berechnen!!)

↳ In TR: Mode → 4 → 2 → Lower = 49,4 → Upper = 51,4 → $\sigma = 0{,}45$ → $\mu = 50$ → 0,9079

↳ $\boxed{P(49{,}4 < X < 51{,}4) \approx 0{,}9079}$

↳ Lösung:

vergleich mit gegeben 95% zeigt, dass es nicht möglich ist, da nur 90,79% in dem vorgegebenen Bereich sein werden

b2) ist Bereich möglich, wenn Erwartungswert $\mu = 50{,}5$ mm?:

↳ In TR: Mode → 4 → 2 → Lower = 49,4 → Upper = 51,4 → $\sigma = 0{,}45$ → $\mu = 50{,}5$ → 0,967

↳ $P(49{,}4 < X < 51{,}4) \approx 0{,}967$

↳ Antwort: ja mit Umstellung wird es erfüllt

b3) Wie Mittelwert μ setzen, dass möglichst viele Schrauben Länge zwischen 49,4 mm und 51,4 mm haben?

↳ damit möglichst viele muss Mittelwert μ genau zwischen den beiden sein also:

↳ $\mu = \frac{49{,}4 + 51{,}4}{2} = 50{,}4$

Vertrauensmean:	90%	95%	99%
z:	1,645	1,96	2,576
z einseitig:	1,282	1,645	2,33

c) Welche Länge von mindestens 90% der Schrauben erreicht?:

↳ nach Quantil 90% gesucht, einseitig, da man nur unteren Bereich anschaut, deshalb $-z \cdot \sigma$ rechnen!

↳ $\boxed{q_{0,1} = \mu \pm z_{0,1} \cdot \sigma = 50 - 1{,}282 \cdot 0{,}45 = 49{,}421}$

⑥ Hypothesentest t-Test beidseitig:

Aufgabe 6 (6 Punkte)

Eine Maschine fertigt Stahlkugeln für Kugellager. Die Durchmesser der Kugeln können dabei als näherungsweise normalverteilt angesehen werden.

Bei einer Stichprobe vom Umfang 25 Stahlkugeln ergab sich als arithmetischer Mittelwert der Durchmesser 10,3 mm und als empirische Standardabweichung 0,6 mm.

H_1 ist die Hypothese, dass es eine Abweichung gibt

Lässt sich mit diesen Daten eine Abweichung des mittleren Durchmessers der gefertigten Kugeln vom Sollwert 10 mm nachweisen? Führen Sie zur Beantwortung dieser Frage einen passenden statistischen Test zum Signifikanzniveau 5% durch. ← Signifikanzniveau gibt den erlaubten Bereich an

H_0 μ = 10 mm so wie es bisher

n \ α	0,10	0,05	0,025	0,01	0,005	0,001	0,0005	einseitig
	0,20	0,10	0,05	0,02	0,01	0,002	0,001	zweiseitig
1	3,078	6,314	12,71	31,82	63,66	318,3	636,6	
2	1,886	2,920	4,303	6,965	9,925	22,33	31,56	
3	1,638	2,353	3,182	4,541	5,841	10,22	12,92	
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610	
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869	
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959	
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408	
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041	
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781	
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587	
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437	
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318	
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221	
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140	
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073	
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015	
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965	
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922	
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883	
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850	
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819	
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792	
23	1,319	1,714	2,069	2,500	2,807	3,485	3,768	
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745	
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725	
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707	
27	1,314	1,703	2,052	2,473	2,771	3,421	3,690	
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674	
29	1,311	1,699	2,045	2,462	2,756	3,396	3,659	
30	1,310	1,697	2,042	2,457	2,750	3,385	3,646	
32	1,309	1,694	2,037	2,449	2,738	3,365	3,622	
34	1,307	1,691	2,032	2,441	2,728	3,348	3,601	
36	1,306	1,688	2,028	2,434	2,719	3,333	3,582	
38	1,304	1,686	2,024	2,429	2,712	3,319	3,566	
40	1,303	1,684	2,021	2,423	2,704	3,307	3,551	
42	1,302	1,682	2,018	2,418	2,698	3,296	3,538	
44	1,301	1,680	2,015	2,414	2,692	3,286	3,526	
46	1,300	1,679	2,013	2,410	2,687	3,277	3,515	
48	1,299	1,677	2,011	2,407	2,682	3,269	3,505	
50	1,299	1,676	2,009	2,403	2,678	3,261	3,496	
55	1,297	1,673	2,004	2,396	2,668	3,245	3,476	
60	1,296	1,671	2,000	2,390	2,660	3,232	3,460	
65	1,295	1,669	1,997	2,385	2,654	3,220	3,447	
70	1,294	1,667	1,994	2,381	2,648	3,211	3,435	
80	1,292	1,664	1,990	2,374	2,639	3,195	3,416	
90	1,291	1,662	1,987	2,368	2,632	3,183	3,402	
100	1,290	1,660	1,984	2,364	2,626	3,174	3,390	
150	1,287	1,655	1,976	2,351	2,609	3,145	3,357	
200	1,286	1,653	1,972	2,345	2,601	3,131	3,340	

↳ Ansatz:

↳ H_0: $\mu_0 = 10$ mm ← denn hier ist wie es normal ist

↳ H_1: $\mu_1 \neq 10$ mm ← hier wird gesagt es ist nicht immer 10!

↳ α: 0,05

→ Formel berechnen:

$$\boxed{ZSB = \left[H_0 - t \cdot \frac{s}{\sqrt{n}} ; H_0 + t \cdot \frac{s}{\sqrt{n}}\right]}$$ ← Zufallsbereichsformel

↳ t beidseitig für 0,05 ablesen für $\boxed{n-1}$ → $t = 2{,}064$

Signif.niveau α bei Stichproben immer s!

↳ s ist das Signifikanzniveau $\sigma = 0{,}6$ schon gegeben

↳ $\sqrt{n}$ Anzahl von $\sqrt{25}$

↳ $\bar{x} = 10{,}3$ ist Mittelwert aus Stichprobe (μ ist MW von Grundgesamtheit)

↳ alles einsetzen:

$$\boxed{\begin{aligned} ZSB &= \left[10 - 2{,}064 \cdot \frac{0{,}6}{\sqrt{25}} ; 10 + 2{,}064 \cdot \frac{0{,}6}{\sqrt{25}}\right] \\ &= [9{,}7523 ; 10{,}2477] \end{aligned}}$$

↳ vergleichen mit $\bar{x}$:

↳ da $\bar{x} \notin ZSB$ wird die Nullhypothese verworfen,

nur wenn $\bar{x}$ innerhalb ZSB ist Nullhypothese gültig!

↳ mittlere Durchmesser der gefertigten Kugeln weicht vom Sollwert 10 mm ab

7 Vertrauensbereich berechnen mit z:

Aufgabe 7 (4 Punkte)

Bei der Fertigung von elektronischen Bauelementen fanden sich in einer Stichprobe von 400 Bauelementen 12 fehlerhafte.

Berechnen Sie aus diesen Zahlen den nach oben begrenzten einseitigen Vertrauensbereich zum Vertrauensniveau 90% für die Fehlerquote bei der Fertigung.

↳ Ansatz: Vertrauensbereich berechnen

Vertrauensniveau:	90%	95%	99%
z:	1,645	1,96	2,576
$z_{einseitig}$:	1,282	1,645	2,33

↳ $p = \frac{k}{n} = \frac{12}{400} = 0{,}03$

↳ z einseitig ablesen bei 90% → $z = 1{,}282$

1. $s = \sqrt{p(1-p)} = \sqrt{0{,}03 \cdot (1-0{,}03)}$

2. $\Delta p = z \cdot \frac{s}{\sqrt{n}} \pm \frac{1}{2n} = 1{,}282 \cdot \frac{\sqrt{0{,}03 \cdot (1-0{,}03)}}{\sqrt{400}} \pm \frac{1}{2 \cdot 400}$

3. $I = [p - \Delta p ; p + \Delta p] = \left[0 ; 0{,}03 + 1{,}282 \cdot \frac{\sqrt{0{,}03 \cdot (1-0{,}03)}}{\sqrt{400}} + \frac{1}{2 \cdot 400}\right] = [0 ; 0{,}0422]$

↳ Fehlerquote beträgt höchstens 4,22 %

8 Kovarianzmatrix, Kovarianz

Aufgabe 8 (5 Punkte)

Der Widerstandswert von temperaturabhängigen Widerständen in einer Fertigung hängt von zwei Zufallsvariablen ab:

R: Widerstandswert – gemessen bei einer konstanten Referenztemperatur (in Ω) und
C: Temperaturkoeffizient des Widerstandes (in $10^{-3}/K$).

Es ist bekannt, dass der Widerstandswert R durch eine Normalverteilung mit einem Mittelwert von $\mu_R = 1000\,\Omega$ und einer Standardabweichung von $\sigma_R = 5\,\Omega$ beschrieben werden kann.
Ebenso ist bekannt, dass auch der Temperaturkoeffizient C durch eine Normalverteilung abgebildet werden kann. Der mittlere Wert des Temperaturkoeffizienten liegt bei $\mu_C = 3{,}85 \cdot 10^{-3}/K$, die Standardabweichung beträgt $\sigma_C = 0{,}5 \cdot 10^{-3}/K$.
Der Korrelationskoeffizient zwischen Widerstand und Temperaturkoeffizient beträgt $0{,}65$.

a) Berechnen Sie die Kovarianz von R und C. → $\sigma_{RC} = \rho_{RC} \cdot \sigma_R \cdot \sigma_C$

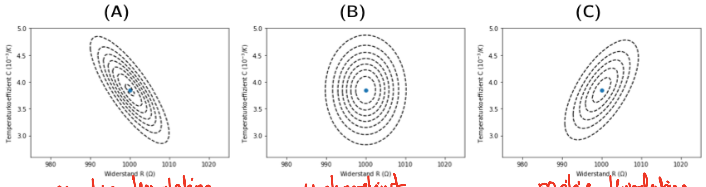
b) Geben Sie die gemeinsame Verteilung für R und C in der Form

$$\vec{X} = \begin{pmatrix} R \\ C \end{pmatrix} \sim N(\vec{\mu}; \underline{\Sigma})$$

an, wobei $\vec{\mu}$ der Mittelwertsvektor und $\underline{\Sigma}$ die Kovarianzmatrix von Widerständen und Temperaturkoeffizienten in der Fertigung ist. → also Mittelwert und Kovarianzmatrix suchen

c) Welches Schaubild zeigt die Höhenlinien der Dichte dieser zweidimensionalen Normalverteilung? → der Korrelationskoeffizient $\rho = 0{,}65$ gegeben bedeutet es positive Korrelation also Schaubild C

(A) (B) (C)



negative Korrelation — unkorreliert — positive Korrelation

a) berechnen Kovarianz von R und C:

↳ $\boxed{\underbrace{\sigma_{RC}}_{\text{Kovarianz RC}} = \overbrace{\rho_{RC}}^{\text{Korrelationskoeffizient zwischen RC}} \cdot \sigma_R \cdot \underbrace{\sigma_C}_{\text{Standardabweichung von C}} = 0{,}65 \cdot 25 \cdot 0{,}5 = 1{,}625 \ m\Omega/K}$

b) gemeinsame Verteilung für R und C in der Form $\vec{x} = \begin{pmatrix} R \\ C \end{pmatrix} \sim N(\vec{\mu}; \Sigma)$: ← dieser Term ist einfach nur denn dargestellt

Einfach Mittelwert der beiden

Kovarianzmatrix der beiden einfach nur

↳ Ansatz: einfach Werte für Normalverteilung von R und C in einen Vektor schreiben jeweils für μ und σ^2

↳ $R \sim N(1000; 5^2)$

↳ $C \sim N(0{,}00385; 0{,}5^2)$

↳ $\vec{X} = \begin{pmatrix} R \\ C \end{pmatrix} \sim N\left(\vec{\mu} = \begin{pmatrix} \mu_R \\ \mu_C \end{pmatrix} = \begin{pmatrix} 1000 \\ 0{,}00385 \end{pmatrix}; \ \Sigma = \begin{pmatrix} \sigma_R^2 & \sigma_{RC} \\ \sigma_{RC} & \sigma_C^2 \end{pmatrix} = \begin{pmatrix} 25 & 1{,}625 \\ 1{,}625 & 0{,}25 \end{pmatrix}\right)$

c)

Aufgabe 1: Vertrauensbereich mit z bestimmen (beidseitig):

1. Würfel (per Definition sechs Seiten) mit 3 grünen und 3 roten Seiten.
   Eintritt-WK für grün oder rot? $p = \frac{1}{2}$
   Wir wollen herausfinden, ob der Würfel fair ist? → fair, wenn alle Seiten gleiche Chance haben geworfen zu werden
   Annahme: Es gibt $n := 10000$ Würfe. Wie oft muss grün oder rot erscheinen, damit
   wir auf dem Signifikanzniveau 90% von Fairness sprechen können? ← da nicht explizit nur nach höchstem/niedrigstem gefragt
   beidseitig

   Wir suchen ein Intervall um den Mittelpunkt $\frac{1}{2}$, in dem dann der Quotient "Anzahl"

Was ist in der Aufgabenstellung gesucht:

↳ es ist der Zufallsbereich gesucht, dieser ist diskret da ganze Zahlen

→ also Binomialverteilung

↳ wie Zufallsbereich für Binomialverteilungen?

→ Zufallsbereich (Konfidenzintervall) gibt an, innerhalb welcher Grenzen, Anzahl Erfolge (grün, rot) mit einer bestimmten Wahrscheinlichkeit (entsprechend dem Signifikanzniveau) liegen sollte, wenn Würfel fair

↳ also Intervall berechnen:

① allgemeine Formel für Binomialverteilungen im Intervall:

$\boxed{I := [\,p - \Delta p\,,\; p + \Delta p\,]}$, $\boxed{\Delta p = z \cdot \frac{s}{\sqrt{n}} \pm \frac{1}{2n}}$ ← Stetigkeitskorrektur, da Binomialverteilung durch Normalverteilung approximiert wird

← n = Anzahl Versuche

$s = \sqrt{p(1-p)} = \sqrt{\frac{1}{2}(1-\frac{1}{2})} = \frac{1}{2} = 0{,}5$ ← s = Standardabweichung

$z_1 = \frac{\alpha}{2} = \frac{10\%}{2} = 1{,}645$ ← z ablesen für 90% und beidseitig → z = 1,645

$\Delta p = 1{,}645 \cdot \frac{0{,}5}{\sqrt{10000}} \pm \frac{1}{2 \cdot 10000}$

$\Delta p_+ = 1{,}645 \cdot \frac{0{,}5}{100} + \frac{1}{20000} = 0{,}008275$

$\Delta p_- = 1{,}645 \cdot \frac{0{,}5}{100} - \frac{1}{20000} = 0{,}008175$

$I = [\frac{1}{2} - 0{,}008175\,;\; \frac{1}{2} + 0{,}08275]$

$= [0{,}491825\,;\; 0{,}508275]$

Vertrauensniveau:	90%	95%	99%
z:	1,645	1,96	2,576
$z_{\text{einseitig}}$:	1,282	1,645	2,33

Aufgabe 2: Wk für nicht eintreten Ereignis?:

→ Wahrscheinlichkeiten für nicht eintreten Ereignis

→ Bsp: Mit welcher Wahrscheinlichkeit kann beim werfen von zwei „normalen" Würfeln keine 1 Auftreten

↳ Ansatz mit Wahrscheinlichkeit und Gegenwahrscheinlichkeit:

$p = \frac{1}{6}$ für 1, Gegenwahrscheinlichkeit also $\bar{p} = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$,

da zwei mal gewürfelt: $P(\text{keine 1 im Wurf}) = \bar{p} \cdot \bar{p} = \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$

Aufgabe 3: Regression:

Aufgabe 4. Hypothesentest:

Frage 4: → Normalverteilungs abfragen immer gleich, Hypothesentest, man stellt Hypothese auf, 0 Hypothese man behauptet gegenteil, das immer gleich abgefragt im Skript nachschauen

Normalverteilung → mit einer Variable arbeitet und mit anderem Wert multipliziert

→ Additionssatz für Normalverteilungen, Summenbildung: Summe Erwartungswerte Summe Varianzen

→ z.B. Bolzenfertigung und Toleranz — Differenzbildung: Differenz Erwartungswerte Summe Varianze

→ z.B. Schraube weiß man, man hat Toleranz = Standardabweichung

↳ wenn man Signifikanzniveau vorgibt

↳ was ist tolerierbar, welche größe ist noch ok von wo bis wo

→ wo bekommt man Quantile für Normalverteilung

↳ aus Tabelle im Skript findet man diese

Aus Vorlesung Notizen zu Aufgabe 5 (eigentlich nicht notwendig nur zur Sicherheit)

Zweidimensionale Normalverteilung

~~$X \sim NV(\mu_X, \sigma_X^2)$~~ oder $\boxed{X \sim NV(\mu_X, \sigma_X)}$ ← Für Prüfung wichtig, was will der Taschenrechner haben? Standardabweichung (σ_X) oder Varianz? Sehr wichtig! (Taschenrechner will Standardabweichung) TR

Normalverteilt; Normalverteilung; σ_X Standardabweichung; das hier nehmen

und $Y \sim NV(\mu_Y, \sigma_Y)$ seien eindimensionale

Normalverteilungen. Dann ist $\binom{X}{Y}$ eine zweidimensionale NV.

3. Regression.
4. Normalverteilung. Z.B. Bolzenfertigung und Toleranz.
5. Zweidimensionale Normalverteilung
   $X \sim NV(\mu_X, \sigma_X^2)$ *oder* $X \sim NV(\mu_X, \sigma_X)$? Was will der TR haben?
   σ_X Standardabweichung.
   Und $Y \sim NV(\mu_Y, \sigma_Y)$ seien eindimensionale Normalverteilungen.
   Dann ist
   $\binom{X}{Y}$ eine zweidimensionale NV. Wie lautet die Kovarianzmatrix Σ?
   $\Sigma := \begin{pmatrix} \mu_X & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \mu_Y \end{pmatrix}$ mit dem Korrelationskoeffizienten ρ
   Wann sind die **Normalverteilungen** – hier X und Y – unabhängig? X, Y unabh., g.d.w. $\rho = 0$ ist.

↳ Wie lautet die Kovarianzmatrix Σ? (Sigma)

↳ wie ist diese definiert?

↳ $\Sigma := \begin{pmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \sigma_Y^2 \end{pmatrix}$ ← Zwei kreuz zwei Matrix

mit Korrelationskoeffizienten ρ bei $\sigma_X \sigma_Y \rho$ (ρ = rho)

Matrix ist symmetrisch oben rechts unten links das gleiche

→ Fragestellung: wann sind die Normalverteilungen – hier X und Y – unabhängig? ← immer gleiche Fragestellung

↳ bei Normalverteilungen andersartiges Verfahren (nicht Normalverteilungen kompliziert)

↳ wenn $\rho = 0$ (rho) dann X, Y unabhängig

→ wenn man was ausrechnen will dann nicht Matrix Σ sondern Matrix Σ^{-1}, Hinweis!

→ wie bedeutet man Σ^{-1}?

normale

↳ $\Sigma := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\quad \Sigma^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ (inverse von Sigma-Matrix)

Probe $AA^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

$= \frac{1}{\det(A)} \begin{pmatrix} ad - bc & -ab + ab \\ cd - cd & -cb + ad \end{pmatrix} = E$ ← Einheitsmatrix

Wann sind die **Normalverteilungen** – hier X und Y – unabhängig? X, Y unabh., g.d.w. $\rho = 0$ ist.

$$A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Probe $AA^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} ad - bc & -ab + ab \\ cd - cd & -cb + ad \end{pmatrix} = E$

→ Für unabhängige Variablen X, Y gilt: $P(X \leq a \land Y \leq b) = P(X \leq a)\, P(Y \leq b)$

Aufgabe 5: Bivariate Normalverteilung:

Prüfungsaufgabe Statistik WS23/24 bivariate NV

Gegeben seien die normalverteilten Variablen X, Y mit den folgenden Parametern $\mu_X = 10, \mu_Y = 15, \sigma_X = 3, \sigma_Y = 5$ und dem Korrelationskoeffizienten $\rho_{XY} = -0.4$

a) Stellen Sie die Kovarianzmatrix Σ auf und geben Sie die Dichtefunktion $f(x, y)$ dieser Verteilung an.
b) Geben Sie die Randwahrscheinlichkeiten $f_X(x)$ und $f_Y(y)$ an.
c) Welche der zwei möglichen Ungleichungen ist richtig?
   a. $P(X \leq 12, Y \geq 18) > P(X \leq 12) \cdot P(Y \geq 18)$
   b. $P(X \leq 12, Y \geq 18) < P(X \leq 12) \cdot P(Y \geq 18)$
d) $P(X \leq 12, Y \geq 18)$
e) Berechnen Sie Erwartungswert μ_Z und Varianz der linearen Transformation
$$Z = 0.5X + 2Y - 1$$
f) Skizzieren Sie das Streudiagramm der bivariaten Normalverteilung, wobei die Werte von X auf der x-Achse und die Werte von Y auf der y-Achse abgetragen werden. Beachten Sie für die Ausrichtung die gegebenen Parameter.

Definition: Zweidimensionale Normalverteilung beschreibt die Verteilung zweier normalverteilter, möglicherweise korrelierter Zufallsvariablen mit spezifischen Mittelwerten, Standardabweichungen und einem Korrelationskoeffizienten

a) $\sigma_{xy} = \rho_{xy} \cdot \sigma_x \cdot \sigma_y = -0,4 \cdot 3 \cdot 5 = -6$

$\hookrightarrow \Sigma = \begin{pmatrix} 9 & -6 \\ -6 & 25 \end{pmatrix}$

$$f(x,y) = \frac{1}{2\pi \cdot \sigma_x \cdot \sigma_y \cdot \sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2\cdot(1-\rho^2)} \cdot \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho \cdot \left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$= \frac{1}{30\pi \cdot \sqrt{0,84}} \cdot e^{-\frac{25}{42} \cdot \left[\left(\frac{x-10}{3}\right)^2 + 0,8 \cdot \left(\frac{x-10}{3}\right)\left(\frac{y-15}{5}\right) + \left(\frac{y-5}{5}\right)^2\right]}$$

b) $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$

$$f_x(x) = \frac{1}{3\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-10}{3}\right)^2} = \frac{1}{3\sqrt{2\pi}} \cdot e^{-\frac{1}{18} \cdot (x-10)^2}$$

$$f_y(y) = \frac{1}{5\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{y-15}{5}\right)^2} = \frac{1}{5\sqrt{2\pi}} \cdot e^{-\frac{1}{50} \cdot (y-15)^2}$$

c) b ist richtig, da bei negativer Korrelation die gemeinsame Wahrscheinlichkeit zweier gegenläufiger Ereignisse geringer als das Produkt ihrer Einzelwahrscheinlichkeiten ist

e) $z = 0,5x + 2y - 1$

$\mu_z = 0,5 \cdot \mu_x + 2 \cdot \mu_y - 1$

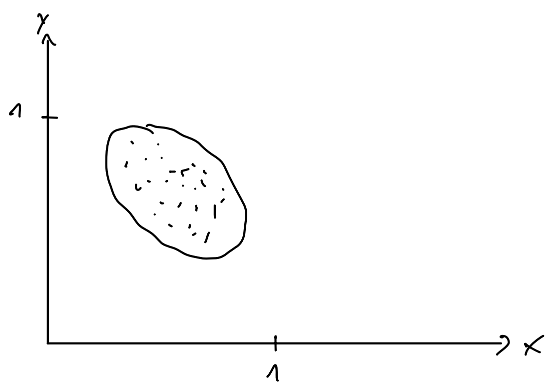
$\mu_z = 0,5 \cdot 10 + 2 \cdot 15 - 1 = 34$

$\sigma_z^2 = (0,5)^2 \cdot \sigma_x^2 + 2^2 \cdot \sigma_y^2 + 2 \cdot 0,5 \cdot 2 \cdot \sigma_{xy}$

$= 2,25 + 100 + 12$

$= 90,25$

f)



d) diese reduzierte WK spiegelt Tendenz wider, dass wenn X kleiner oder gleich 12 ist, Y weniger wahrscheinlich 18 oder mehr ist, was der negativen Korrelation entspricht

Aufgabe 6: Binominalverteilung:

1) Ein mäßig begabter Bäcker backt „Berliner" mit Himbeerfüllung. Jeder fünfte Berliner bleibt leider leer. Der mäßig begabte Bäcker verkauft seine mäßig gefüllten Berliner in Verpackungseinheiten zu 10 Stück/Karton.

   a) Welche Verteilung beschreibt das Backgeschick des mäßig begabten Bäckers?
   Lösung: Binomialverteilung $Berliner \sim B(p, n) = B(0.2, 10)$

   b) Wie groß ist die zu erwartende Anzahl an ungefüllten Berlinern pro Karton?
   Erwartungswert $\mu = pn = 0.2 \cdot 10 = 2$

   c) Wie wahrscheinlich ist es, dass in einem Karton 4 leere Berliner sind?
   Es gilt für Dichtefunktion $f(k) := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$
   $$f(4) := \binom{10}{4} p^k (1-p)^{4-k} = \binom{10}{4} 0.2^4 0.8^6 = 210 \cdot 0.0016 \cdot 0.262144$$
   $$= 0.0881 \approx 9\%$$

   d) Wie wahrscheinlich ist es, dass in einem Karton mehr als 6 leere Berliner sind?
   $$P(X > 6) = 1 - P(X \le 6) = 1 - 0.999 \approx 0\%$$

   e) Ein Kunde kauft 4 Kartons. Wie viele befüllte Berliner darf der Kunde erwarten?
   Neue ZV $X = 4Berliner \sim B(0.2, 40), \mu_{neu} = 32$
   Wie groß ist die Wahrscheinlichkeit, dass der Kunde 12 leere Berliner gekauft hat?
   $$f(X = 12) = \binom{40}{12} 0.2^{12} \cdot 0.8^{28} = 0.0443 \approx 4\%$$

→ Ansatz: welche Verteilung?:

↳ Binominalverteilung, da Wk immer $p = \frac{1}{5}$ ist, denn man einen leeren Berliner erhält und auch Anzahl verschiedene $n = 10$ vorhanden ist

a) welche Verteilung beschreibt das Backgeschick des mäßig begabten Bäckers?

↳ Binominalverteilung $X \sim B(n; p) = X \sim B(10; 0{,}2)$

b) wie groß ist die zu erwartende Anzahl an ungefüllten Berlinern pro Karton?:

↳ Erwartungswert berechnen:

↳ $\mu = p \cdot n = 0{,}2 \cdot 10 = 2$

Diskrete Dichte	$P(X = k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$
Erwartungswert	$E(X) = \mu = n \cdot p$
Varianz	$Var(X) = \sigma^2 = n \cdot p \cdot (1-p)$

c) Wie wahrscheinlich ist es, dass in einem Karton 4 leere Berliner sind?:

1. Ansatz mit Rechnung:

gesuchte defekte Berliner $k = 4$

↳ $\boxed{f(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \rightarrow P(X=4) = \binom{10}{4} \cdot 0{,}2^4 (1-0{,}2)^{10-4}} = \frac{10!}{4!(10-4)!} \cdot 0{,}2^4 \cdot 0{,}8^6 = 210 \cdot 0{,}0016 \cdot 0{,}262144 = 0{,}0881$

$n = 10$ Anzahl Berliner

2. Ansatz mit TR:

Mode → 4 → 4 (Binomial PD) → 2 → x = 4 → n = 10 → p = 0,2 → 0,0881

↳ $\boxed{P(X = 4) = 0{,}0881}$

d) Wie wahrscheinlich ist es dann in einem Karton mehr als 6 leere Berliner sind?

↳ TR berechnet immer nur die kumulative Wahrscheinlichkeit für die Erfolge deshalb $P(X > 6)$ mit Gegenwahrscheinlichkeit:

Wk für 6 leere Berliner kann in TR berechnet werden demnach Gleich

↳ $\boxed{P(X > 6) = 1 - P(X \le 6) = 1 - 0{,}999 \approx 0\%}$

→ TR: $P(X \le 6)$ = 4 → 1 (Binomial CD) → 2 → x = 6 → n = 10 → p = 0,2 → 0,999

e) Ein Kunde kauft 4 Kartons. wie viele befüllte Berliner darf der Kunde erwarten?:

↳ Neue ZV, da nun $n = 4 \cdot 10 = 40$ → $X \sim N(40; 0{,}2)$

↳ Erwartungswert = $n \cdot p = 8$ schlechte

↳ Also $\boxed{40 - 8 = 32 \text{ gute}}$ darf der Kunde erwarten

Wie groß Wk, dass der Kunde 12 leere Berliner gekauft hat?:

TR: 4 → 4 → 2 → x = 12 → n = 40 → p = 0,2 → 0,0443

$\boxed{f(X = 12) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{40}{12} \, 0{,}2^{12} \cdot 0{,}8^{28} \approx 0{,}0443}$

Aufgabe 7 Median:

2) Bei der Temperaturmessung in einem Krankenhauszimmer für Kassenpatienten ergeben sich folgende Messwerte:

$$37.0, 36.9, 37.5, 39.0, 42.0, 41.5, ?$$

a) Welchen Wert muss der Messwert ? besitzen, wenn das arithmetische Temperaturmittel per Krankenkassenvorgabe bei $\bar{T} := 30.8$ liegen soll? Ist dieser Wert eindeutig?

Lösung: ? =

b) Wie lautet der Median der gegebenen sechs Temperaturwerte?

c) Welchen Wert besitzt ?, wenn der Median aller sieben Messwerte per Krankenkassenvorgabe bei $\tilde{T} := 37.8$ liegen soll?

d) Welchen medizinisch plausiblen Wert besitzt ?, wenn die Spannweite der Temperaturen per Krankenkassenvorgabe bei $S := 6.0$ liegen soll?

a) Welchen Messwert muss ? besitzen, wenn arithmetische Temperaturmittel $\bar{T} = 30{,}8$ betragen soll?:

↳ Gleichung für Mittelwert $\bar{T} = 30{,}8$ aufstellen und umstellen nach ?:

$$30{,}8 = \frac{37 + 36{,}9 + 37{,}5 + 39 + 42 + 41{,}5 + ?}{7} \quad | \cdot 7 \quad | -(37 + \ldots + 41{,}5)$$

$$? = 30{,}8 \cdot 7 - 37 - 36{,}9 - 37{,}5 - 39 - 42 - 41{,}5 - ?$$

$$= -18{,}3$$

↳ $-18{,}3$ ist eindeutig, da nur dieser Wert arithmetisches Mittel von $\bar{T} = 30{,}8$ ermöglicht

b) Wie lautet der Median der gegebenen 6 Temperaturwerte?:

↳ einfach 2 mittlere Werte addieren und geteilt durch 2:

↳ $\frac{37{,}5 + 39}{2} = 38{,}25$

c) Welchen Wert besitzt ? wenn Median aller 7 Messwerte bei $\tilde{T} = 37{,}8$ liegen soll?

↳ $? = 37{,}8$ weil bei ungerader Anzahl von Werten genau der Mittlere Wert der Median ist

(Müsste man sozusagen in der Mitte einfügen)

d) Welchen medizinisch plausiblen Wert besitzt ? wenn Spannweite der Temperaturen bei $S = 6$ liegen soll?:

↳ Es gibt genau zwei mögliche Werte, weil ? = Minimum + S oder ? = Maximum − S

d.h. $? \in \{36{,}9 + 6, 42 - 6\} = \{42{,}9, 36\}$

↳ gute Wahl wäre 36 (bei 42,9 hätte man starkes Fieber)