ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik
Musterlösung 12

1. Eine Klimaanlage schafft es, die Raumtemperatur bis auf eine Standardabweichung von einem halben Grad Celsius konstant zu halten. Die angestrebte Raumtemperatur beträgt 20.00 Grad Celsius. An zehn aufeinanderfolgenden Tagen wurden die folgenden Temperaturen gemessen:

20.71 19.76 20.56 21.39 21.00 19.67 20.92 20.31 20.39 20.72.

Führe einen Vorzeichentest zum 5%-Niveau durch, um zu beurteilen, ob die Klimaanlage richtig geeicht ist. Wie lautet der zugehörige P-Wert? Wie lautet der Testentscheid?

Lösung: Beim Vorzeichentest beobachten wir die Zufallsvariable Q welche angibt, wie viele der gemessenen Werte über (bzw. unter, je nach Geschmack) 20 Grad Celsius liegen. Q ist dann binomialverteilt mit den Parametern $n = 10$ und p. Bei richtiger Eichung (Nullhypothese) sollte p gleich $\frac{1}{2}$ sein. Die Hypothesen lauten daher

$$H_0 : p = p_0 = 0.5 \qquad\qquad H_A : p \neq p_0.$$

Da die $\mathrm{Bin}(n, \frac{1}{2})$-Verteilung symmetrisch ist, können wir die Wahrscheinlichkeiten einfach berechnen:

$$\begin{aligned} P_{p=p_0}(Q = 0) &= P_{p=p_0}(Q = 10) = 0.5^{10} = 9.8 \cdot 10^{-4} \\ P_{p=p_0}(Q = 1) &= P_{p=p_0}(Q = 9) = 10 \cdot 0.5^{10} = 9.8 \cdot 10^{-3} \\ P_{p=p_0}(Q = 2) &= P_{p=p_0}(Q = 8) = \frac{10 \cdot 9}{2} \cdot 0.5^{10} = 0.0439 \end{aligned}$$

Durch aufsummieren erhalten wir:

$$\begin{aligned} P_{p=p_0}(Q \leq 1) + P_{p=p_0}(Q \geq 9) &= 0.0215 \\ P_{p=p_0}(Q \leq 2) + P_{p=p_0}(Q \geq 8) &= 0.1094 \end{aligned}$$

Damit der Fehler 1. Art kleiner als 5% ist, ergibt sich also ein Verwerfungsbereich von $K = \{0, 1\} \cup \{9, 10\}$. Die Realisierung von Q ist $8 \notin K$ weshalb die Nullhypothese nicht verworfen wird.

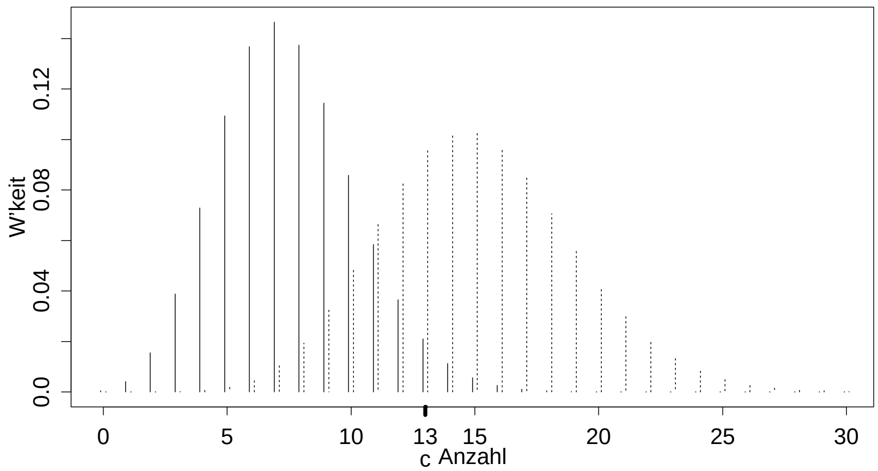
Der P-Wert errechnet zur gegebenen Beobachtung ist wegen der Symmetrie der $\mathrm{Bin}(n, \frac{1}{2})$-Verteilung $2P_{p=p_0}(Q \leq 2) = 0.109$. Wegen $5\% = 0.05 < 0.109$ wird, konsistent zur obigen Entscheidung, die Nullhypothese nicht verworfen.

Anmerkung: Die kumulativen Wahrscheinlichkeiten können in Matlab auch mit dem Befehl `binocdf(x,n,p)` berechnet werden.

2. Wir bezeichnen mit X die Anzahl Asbestfasern in einem bestimmten Volumen und wählen eine Poisson-Verteilung $X \sim Pois(\lambda)$.

In der untenstehenden Abbildung ist die Verteilung von X für die Nullhypothese $\lambda_0 = 7.5$ respektive für die Alternativhypothese $\lambda = 15$ und die kritische Grenze $c = 13$ angegeben ($\alpha = 0.05$, einseitiger Test und c gehört zum Verwerfungsbereich).

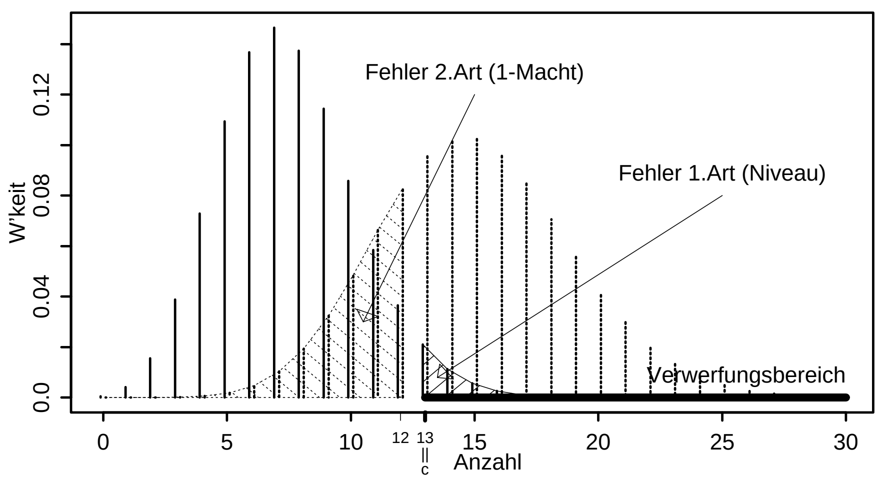
a) Zeichne in diese Grafik den Verwerfungsbereich, die Wahrscheinlichkeit des Fehlers 1. Art und des Fehlers 2. Art ein.



b) Beurteile die folgenden Aussagen mit "richtig" oder "falsch". Gib eine kurze Begründung an.

1. Die Macht wird kleiner, wenn man das Niveau des Tests vergrössert.
2. Die Alternative mit $\lambda = 20$ hat einen grösseren Fehler 2. Art als die Alternative mit $\lambda = 15$.
3. Falls die kritische Grenze auf $c = 14$ erhöht wird, so wird die Macht kleiner.

Lösung:



a)

b)
1. Falsch: Wenn das Niveau α des Tests vergrössert wird, so wird H_0 häufiger zugunsten von H_A verworfen.
2. Falsch: Die Verteilung mit $\lambda = 20$ ist weiter von der Verteilung mit λ_0 entfernt. Die Macht nimmt zu und damit der Fehler 2.Art ab.
3. Richtig: Wenn die kritische Grenze erhöht wird, so verkleinert sich automatisch der Verwerfungsbereich. Die Macht wird dadurch kleiner.

3. Die Schmelzwärme von Wasser kann mit zwei Methoden gemessen werden. In verschiedenen Versuchen ergaben sich die folgenden Messwerte (in cal/g)

Methode 1: 12 Werte, $\overline{x} = 80.02$, $\sum_{i=1}^{12}(x_i - \bar{x})^2 = 0.00691$.

Methode 2: 8 Werte, $\overline{y} = 79.98$, $\sum_{i=1}^{8}(y_i - \overline{y})^2 = 0.00673$.

Teste unter Annahme der Normalverteilung auf dem 1%-Niveau, ob sich die beiden Messmethoden systematisch unterscheiden.

Lösung:

Es handelt sich hier um ungepaarte Stichproben. Ungepaarter t-Test:

- **Modellannahmen**: $X_1, \ldots, X_{12}$ sind i.i.d. normalverteilt mit Erwartungswert μ_X und Standardabweichung σ. $Y_1, \ldots, Y_8$ sind i.i.d. normalverteilt mit Erwartungswert μ_Y und Standardabweichung σ. Ausserdem sollen alle Zufallsvariablen $X_1, \ldots, X_{12}, Y_1, \ldots, Y_8$ unabhängig voneinander sein.
- **Nullhypothese** H_0: $\mu_X = \mu_Y$.
- **Alternativhypothese** H_A: $\mu_X \neq \mu_Y$.
- **Teststatistik** T:

$$\begin{aligned} T &= \frac{\bar{X} - \bar{Y}}{S_{\text{pool}}\sqrt{1/12 + 1/8}} \quad \text{wobei} \\ S_{\text{pool}}^2 &= \frac{1}{12+8-2}\left(\sum_{i=1}^{12}(X_i - \bar{X})^2 + \sum_{i=1}^{8}(Y_i - \bar{Y})^2\right) \end{aligned}$$

- **Verteilung von T unter H_0**: Falls H_0 gilt, ist T t-verteilt mit $12+8-2 = 18$ Freiheitsgraden.
- **Verwerfungsbereich**: H_0 wird verworfen, falls $|T|$ grösser ist als 2.878.
- **Beobachtung**: Es wurde

$$t = \frac{80.02 - 79.98}{\sqrt{1/18(0.00691 + 0.00673)(1/8 + 1/12)}} = 3.18354$$

  beobachtet, was im Verwerfungsbereich liegt.
- **Testentscheid**: H_0 wird verworfen, d.h. die beiden Messmethoden weichen systematisch voneinander ab.

4. Zwei Tiefen-Messgeräte messen für die Tiefe einer Gesteins-Schicht an 9 verschiedenen Orten die folgenden Werte:

Messgerät A	120	265	157	187	219	288	156	205	163
Messgerät B	127	281	160	185	220	298	167	203	171
Differenz x_i	-7	-16	-3	2	-1	-10	-11	2	-8

Kennzahlen für die Differenz: $\bar{x}$ beträgt -5.78, die Standardabweichung $s = 6.2$.

Es wird vermutet, dass Gerät B systematisch grössere Werte misst. Bestätigen die Messwerte diese Vermutung oder ist eine zufällige Schwankung als Erklärung plausibel?

a) Handelt es sich um verbundene (gepaarte) oder um unabhängige Stichproben?

b) Führe einen t-Test auf dem Niveau $\alpha = 0.05$ durch. Formuliere explizit: Modellannahmen, Nullhypothese, Alternative, Teststatistik, Verwerfungsbereich und Testergebnis.

Lösung:

a) Es handelt sich um **gepaarte** Stichproben. Am gleichen Ort wird mit beiden Geräten gemessen.

b)
- **Modell:** $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, σ wird durch S geschätzt.
- **Nullhypothese:** $H_0 : \mu = \mu_0 = 0$.
  Alternative: $H_A : \mu < 0$.
- **Teststatistik:**

$$T = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}$$

  Verteilung der Teststatistik unter H_0: $T \sim t_{n-1}$.
- **Signifikanzniveau:** $\alpha = 5\%$.
- **Verwerfungsbereich:**

$$K = (-\infty, -t_{9-1,0.95}] = (-\infty, -1.86]$$

- **Testentscheid:**

$$t = \frac{\bar{x} - 0}{S/\sqrt{9}} = -2.8$$

  Der realisierte Wert t der Teststatistik liegt im Verwefungsbereich, d.h. eine neue Eichung der Geräte ist angezeigt.

5. Alte Prüfungsaufgabe: Winter 2013

Am kommenden Samstag findet bei der alpinen Ski WM 2013 in Schladming die Abfahrt der Herren statt. Pro Nation dürfen jeweils 4 Läufer starten. Der Trainer der Schweizer Abfahrtsmannschaft hat bereits 3 Athleten fix nominiert. Für den letzten noch zu vergebenden Startplatz kann er sich nicht zwischen Carlo J. oder Patrick K. entscheiden. In weiser Voraussicht hatte er bereits zu Beginn der Saison angekündigt, dass in solchen Fällen getestet werden soll, ob die bis zur WM erreichten Saisonresultate statistisch signifikant unterschiedlich sind oder nicht. Falls ja, so darf der entsprechende Läufer (für den die Daten sprechen) starten, sonst entscheidet eine faire Münze. Die von den beiden Läufern in den bisherigen 8 Saisonabfahrten erzielten Zeiten (in Minuten:Sekunden) sind dabei wie folgt:

Abfahrt Nr. (i)	1	2	3	4	5	6	7	8
Carlo (a_i)	1:52.43	2:02.37	1:20.78	1:38.67	2:34.23	1:57.13	1:43.14	1:50.88
Patrick (b_i)	1:53.20	2:03.40	1:20.66	1:39.13	2:33.41	1:58.58	1:42.82	1:51.04

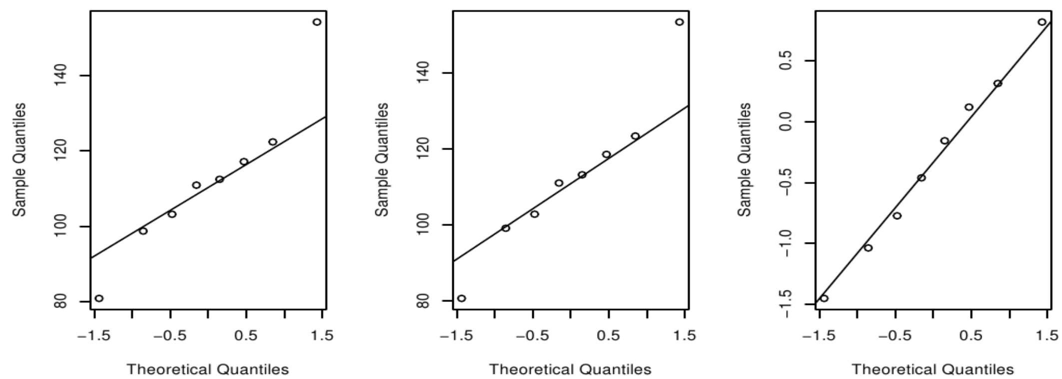
Wenn wir die Stichprobendifferenz mit $d_i := a_i - b_i$ $(i = 1, \ldots, 8)$ bezeichnen, dann ergeben sich daraus die folgenden empirischen Kennwerte (in Sekunden): Mittelwert: $\overline{a}_8 = 112.45$, $\overline{b}_8 = 112.78$, $\overline{d}_8 = -0.33$. Streuung: $s_a = 21.21$, $s_b = 21.1$, $s_d = 0.75$. Gepoolte Streuung von a und b: $s_{pool} = 21.16$.

Es soll nun ein entsprechender Test entwickelt werden um zu prüfen, ob einer der beiden Skifahrer signifikant schneller ist. Und zwar so, dass für jeden einzelnen der beiden Kontrahenten, unter der

Annahme er sei gleich schnell wie sein Konkurrent, die Wahrscheinlichkeit dass bereits durch den Test (also ohne dass es zu einem fairen Münzwurf kommt) gegen ihn entschieden wird, maximal 10% beträgt.

a) Ist die Stichprobe gepaart oder ungepaart? Ist der Test ein- oder zweiseitig? Wie lautet das Niveau α? Mit welcher Wahrscheinlichkeit kommt es zu einem Münzwurf wenn beide Sportler gleich schnell sind?

b) Die folgenden Graphiken sind QQ-Plots der Stichproben $(a_1, \ldots, a_8)$, $(b_1, \ldots, b_8)$ sowie $(d_1, \ldots, d_8)$ (in Sekunden).



Ist die Annahme der Normalverteilung für die für den Test relevanten Daten gerechtfertigt? Geben Sie eine Begründung an. Welcher Test ist daher (unter den üblichen i.i.d. Annahmen) am besten geeignet?

c) Entwickeln Sie den Test, d.h. geben Sie Null- und Alternativhypothese, die Teststatistik sowie den Verwerfungsbereich an.

d) Für welche Werte der Teststatistik würde (ohne dass es zu einem Münzwurf kommt) gegen Carlo entschieden werden?

e) Führen Sie den Test durch. Wie entscheidet der Test? Kommt es zu einem Münzwurf oder nicht?

Lösung:

a) Die Stichprobe ist gepaart. Der Test ist zweiseitig zum Niveau $\alpha = 20\%$. Zu einem Münzwurf kommt es wenn der Test, unter der Annahme beide seien gleich schnell, nicht verwirft, die Wahrscheinlichkeit dafür ist definitionsgemäß $1 - \alpha = 1 - 0.2 = 80\%$.

b) Die für den Test relevanten Daten ist die Stichprobe $(d_1, \ldots, d_8)$, die Punkte des entsprechenden QQ-Plots liegen für die verhältnismässig kleine Stichprobengrösse von $n = 8$ sehr schön auf einer Geraden, somit ist die Annahme der Normalverteilung gerechtfertigt. Da die Varianz unbekannt ist, werden wir unter den üblichen Annahmen, dass die $D_1, \ldots, D_8$ i.i.d. sind, einen t-Test durchführen.

c) Wir haben $D \sim \mathcal{N}(\mu, \sigma^2)$ mit μ unbekannt (wird getestet) und σ unbekannt (wird geschätzt). Wir haben $\mu_0 := 0$ mit den Hypothesen

$$H_0 : \mu = \mu_0 \qquad\qquad H_A : \mu \neq \mu_0.$$

Die Teststatistik ist $T = \sqrt{8}\frac{\overline{D}_8}{S_d}$, sie ist unter der Nullhypothese t_7-verteilt. Der Verwerfungsbereich ist daher gegeben durch $VB_{20\%} = (-\infty, -t_{7,0.9}] \cup [t_{7,0.9}, \infty) = (-\infty, -1.415] \cup [1.415, \infty)$.

d) Je größer die d_i's, desto schlechter für Carlo. Liegt die Teststatistik in $[1.415, \infty)$, dann wird ohne Münzwurf gegen ihn entschieden.

e) Die Realisierung der Teststatistik ist $t = \sqrt{8}\frac{\overline{d}_8}{s_d} = \sqrt{8}\frac{-0.33}{0.75} = -1.24$. Wegen $-1.24 \notin VB_{20\%}$ kann die Nullhypothese nicht verworfen werden. Es muss die Münze entscheiden.