Chi-Quadrat-Anpassungstest auf parametrisches Verteilungsmodell

- Chi-Quadrat-Anpassungstest kann auch durchgeführt werden, wenn statt (einzelner) hypothetischer Verteilung eine parametrische Klasse von Verteilungen als hypothetische Verteilungsklasse fungiert.
- Durchführung des Chi-Quadrat-Anpassungstests dann in zwei Schritten:
  1. Schätzung der Verteilungsparameter innerhalb der hypothetischen Verteilungsklasse mit der ML-Methode.
  2. Durchführung des (regulären) Chi-Quadrat-Anpassungstest mit der hypothetischen Verteilung zu den geschätzen Parametern.
- Zu beachten:
  - **Verteilung der Testgröße χ^2 ändert sich!** Bei ML-Schätzung auf Basis der für die Durchführung des Chi-Quadrat-Anpassungstest maßgeblichen Klassierung der Stichprobe gilt unter H_0 näherungsweise $\chi^2 \sim \chi^2(k - r - 1)$, wobei r die Anzahl der per ML-Methode geschätzten Parameter ist.
  - Werden die Verteilungsparameter nicht aus den klassierten Daten, sondern aus den ursprünglichen Daten mit ML-Methode geschätzt, gilt diese Verteilungsaussage so nicht mehr (Abweichung allerdings moderat).

Zusammenfassung: Chi-Quadrat-Anpassungstest

zur Anpassung an parametrische Verteilungs**familie**

Anwendungsvoraussetzungen	approx.: Y beliebig verteilt, $X_1, \ldots, X_n$ einf. Stichprobe zu Y Familie von Verteilungsfunktionen F_θ für $\theta \in \Theta$ vorgegeben $k-1$ Klassengrenzen $a_1 < a_2 < \ldots < a_{k-1}$ vorgegeben
Nullhypothese Gegenhypothese	$H_0 : F_Y = F_\theta$ für ein $\theta \in \Theta$ $H_1 : F_Y \neq F_\theta$ (für alle $\theta \in \Theta$)
Teststatistik	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} = n \sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i^0\right)^2}{p_i^0} = \left(\frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_i^0}\right) - n$
Verteilung (H_0)	χ^2 ist unter H_0 näherungsweise $\chi^2(k-r-1)$-verteilt, wenn $\widehat{\theta}$ ML-Schätzer des r-dim. Verteilungsparameters θ auf Basis klassierter Daten ist (Verwendung von $\widehat{\theta}$ siehe unten). (Näherung nur vernünftig, falls $np_i^0 \geq 5$ für $i \in \{1, \ldots, k\}$)
Benötigte Größen	$p_i^0 = F_{\widehat{\theta}}(a_k) - F_{\widehat{\theta}}(a_{k-1})$ mit $a_0 := -\infty, a_k := \infty$, $n_i = \#\{j \in \{1, \ldots, n\} \mid x_j \in (a_{i-1}, a_i]\}, i \in \{1, \ldots, k\}$
Kritischer Bereich zum Niveau α	$(\chi^2_{k-r-1;1-\alpha}, \infty)$
p-Wert	$1 - F_{\chi^2(k-r-1)}(\chi^2)$

Beispiel: Chi-Quadrat-Anpassungstest auf $H_0 : Y \sim \text{Geom}(p)$ für $p \in (0, 1)$

- Stichprobeninformation: Häufigkeitsverteilung aus vorangegangenem Beispiel:

i	1	2	3	4	5	6
a_i	0	1	2	3	4	≥ 5
n_i	32	19	16	16	6	11

- **Erster Schritt:**
  ML-Schätzung von p mit Hilfe der klassierten Stichprobeninformation:
  - Man kann zeigen, dass der ML-Schätzer auf Basis der klassierten Stichprobe durch
  $$\widehat{p} = \frac{n - n_k}{n - n_k + \sum_{i=1}^{k}(i-1) \cdot n_i}$$
  gegeben ist.
  - Hier erhält man also die Realisation
  $$\widehat{p} = \frac{100 - 11}{100 - 11 + 0 \cdot 32 + 1 \cdot 19 + 2 \cdot 16 + 3 \cdot 16 + 4 \cdot 6 + 5 \cdot 11} = \frac{89}{267} = 0.3333$$

- **Zweiter Schritt:**
  Durchführung des Chi-Quadrat-Anpassungstest für $H_0 : F_Y = F_{0.3333}$ (mit $F_p := F_{\text{Geom}(p)}$) gegen $H_1 : F_Y \neq F_{0.3333}$ **unter Berücksichtigung der ML-Schätzung von p durch geänderte Verteilung von χ^2 unter H_0!**

Insgesamt: Chi-Quadrat-Anpassungtest für Verteilungsfamilie:

1. **Hypothesen:**

   $H_0 : F_Y = F_p$ für ein $p \in (0,1)$ (mit $F_p := F_{\text{Geom}(p)}$) gegen $H_1 : F_Y \neq F_p$

2. **Teststatistik:**

   $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}$ ist unter H_0 approximativ $\chi^2(k-1-r)$-verteilt, falls $np_i^0 \geq 5$ für alle i gilt und r-dimensionaler Verteilungsparameter per ML-Methode aus den klassierten Daten geschätzt wurde.

3. **Kritischer Bereich zum Niveau $\alpha = 0.10$:**

   $K = (\chi^2_{k-1-r;1-\alpha}, +\infty) = (\chi^2_{4;0.90}, +\infty) = (7.779, +\infty)$

4. **Berechnung der realisierten Teststatistik:**

   Eine ML-Schätzung aus den klassierten Daten liefert den Schätzwert $\widehat{p} = 0.3333$ für den unbekannten Verteilungsparameter p.

K_i	n_i	p_i^0	np_i^0	$\frac{(n_i - np_i^0)^2}{np_i^0}$
$(-\infty, 0]$	32	$(1 - 0.3333)^0 \cdot 0.3333 = 0.3333$	33.33	0.0531
$(0, 1]$	19	$(1 - 0.3333)^1 \cdot 0.3333 = 0.2223$	22.23	0.4693
$(1, 2]$	16	$(1 - 0.3333)^2 \cdot 0.3333 = 0.1481$	14.81	0.0956
$(2, 3]$	16	$(1 - 0.3333)^3 \cdot 0.3333 = 0.0988$	9.88	3.7909
$(3, 4]$	6	$(1 - 0.3333)^4 \cdot 0.3333 = 0.0658$	6.58	0.0511
$(4, +\infty)$	11	$1 - \sum_{i=1}^5 p_i^0 = 0.1317$	13.17	0.3575
Σ	100	1	100	$\chi^2 = 4.8175$

   Es gilt $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, 6\} \rightsquigarrow$ Näherung ok.

5. **Entscheidung:**

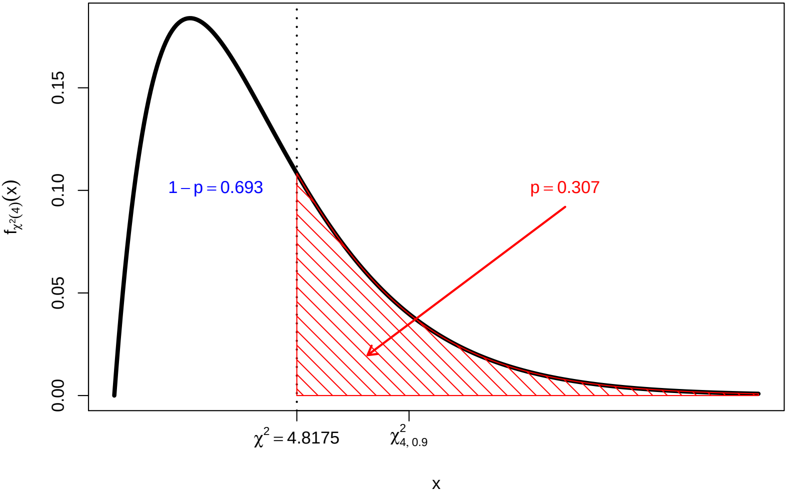
   $\chi^2 = 4.8175 \notin (7.779, +\infty) = K \quad \Rightarrow \quad H_0$ wird nicht abgelehnt!

   (p-Wert: $1 - F_{\chi^2(4)}(\chi^2) = 1 - F_{\chi^2(4)}(4.8175) = 1 - 0.6935 = 0.3065$)

Test kommt zum Ergebnis, dass $Y \sim \text{Geom}(p)$ nicht verworfen werden kann.
(ML-Schätzung von p: $\widehat{p} = 0.3333$)

Beispiel: *p*-Wert bei Chi-Quadrat-Anpassungstest (Grafik)

Test auf geometrische Verteilung, realisierte Teststatistik $\chi^2 = 4.8175$, *p*-Wert: 0.307



Chi-Quadrat-Unabhängigkeitstest (Kontingenztest)

- *Bisher:* Einfache Stichprobe $X_1, \dots, X_n$ zu **einer** Zufallsvariablen Y.
- *Im Folgenden:* Betrachtung von einfachen Stichproben zu mehrdimensionalen Zufallsvariablen bzw. (später) mehreren (unabhängigen) einfachen Stichproben zu mehreren Zufallsvariablen.
- Erste Problemstellung: **Untersuchung** von zwei Zufallsvariablen Y^A, Y^B **auf stochastische Unabhängigkeit**.
- Erforderliche Stichprobeninformation: Einfache Stichprobe

$$(X_1^A, X_1^B), (X_2^A, X_2^B), \dots, (X_n^A, X_n^B)$$

  vom Umfang n zu zweidimensionaler Zufallsvariable (Y^A, Y^B).
- *Testidee:* den bei Unabhängigkeit von Y^A, Y^B bestehenden Zusammenhang zwischen Randverteilungen von Y^A und Y^B sowie gemeinsamer Verteilung von (Y^A, Y^B) ausnutzen:
  - Gemeinsame Wahrscheinlichkeiten stimmen bei Unabhängigkeit mit Produkt der Randwahrscheinlichkeiten überein (falls (Y^A, Y^B) diskret).
  - Daher sprechen geringe Abweichungen zwischen gemeinsamen (relativen) Häufigkeiten und Produkt der (relativen) Randhäufigkeiten für Unabhängigkeit, große Abweichungen dagegen.

- Betrachtete Anwendungssituationen:
  1. Sowohl Y^A als auch Y^B sind diskret mit „wenigen" Ausprägungen, in der Stichprobe treten die Ausprägungen $a_1, \ldots, a_k$ von Y^A bzw. $b_1, \ldots, b_l$ von Y^B auf.
  2. Y^A und Y^B sind diskret mit „vielen" Ausprägungen oder stetig, die Stichprobeninformation wird dann mit Hilfe von Klassierungen $A_1 = (-\infty, a_1], A_2 = (a_1, a_2], \ldots, A_k = (a_{k-1}, \infty)$ von Y^A bzw. $B_1 = (-\infty, b_1], B_2 = (b_1, b_2], \ldots, B_l = (b_{l-1}, \infty)$ von Y^B zusammengefasst.
  3. Mischformen von 1 und 2.
- Der Vergleich zwischen (in der Stichprobe) **beobachteten** gemeinsamen absoluten Häufigkeiten n_{ij} und **bei Unabhängigkeit** (auf Basis der Randhäufigkeiten) **zu erwartenden** gemeinsamen absoluten Häufigkeiten $\widetilde{n}_{ij}$ erfolgt durch die Größe

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}} ,$$

wobei n_{ij} die beobachteten gemeinsamen Häufigkeiten für (a_i, b_j) bzw. (A_i, B_j) aus der Stichprobenrealisation und $\widetilde{n}_{ij} = n \cdot \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ die erwarteten gemeinsamen Häufigkeiten aus den Randhäufigkeiten $n_{i\cdot}$ von a_i bzw. A_i und $n_{\cdot j}$ von b_j bzw. B_j sind ($i \in \{1, \ldots, k\}$, $j \in \{1, \ldots, l\}$).

- Für wachsenden Stichprobenumfang n konvergiert die Verteilung der Testgröße χ^2 bei Gültigkeit von

$$H_0 : Y^A, Y^B \text{ sind stochastisch unabhängig}$$

gegen die $\chi^2((k-1)\cdot(l-1))$-Verteilung.
- Die Näherung der Verteilung von χ^2 unter H_0 ist für endlichen Stichprobenumfang n vernünftig, falls gilt:

$$\widetilde{n}_{ij} \geq 5 \text{ für alle } i \in \{1,\ldots,k\}, j \in \{1,\ldots,l\}$$

- Wie beim Chi-Quadrat-Anpassungstest sprechen **große** Werte der Teststatistik χ^2 **gegen** die Nullhypothese „Y^A und Y^B sind stochastisch unabhängig", während kleine Werte für H_0 sprechen.
- Als kritischer Bereich zum Signifikanzniveau α ergibt sich also entsprechend:

$$K = (\chi^2_{(k-1)\cdot(l-1);1-\alpha}, \infty)$$

- Die Testgröße χ^2 ist eng verwandt mit der bei der Berechnung des korrigierten Pearsonschen Kontingenzkoeffizienten benötigten Größe χ^2.
- Analog zum Chi-Quadrat-Anpassungstest kann der Chi-Quadrat-Unabhängigkeitstest ebenfalls auf „Merkmale" Y^A bzw. Y^B angewendet werden, deren Ausprägungen $a_1,\ldots,a_k$ bzw. $b_1,\ldots,b_l$ noch nicht „Zufallsvariablen-konform" als reelle Zahlen „kodiert" wurden.

- Darstellung der Stichprobeninformation üblicherweise in Kontingenztabelle der Form

$Y^A \setminus Y^B$	b_1	b_2	$\cdots$	b_l
a_1	n_{11}	n_{12}	$\cdots$	n_{1l}
a_2	n_{21}	n_{22}	$\cdots$	n_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}

bzw.

$Y^A \setminus Y^B$	B_1	B_2	$\cdots$	B_l
A_1	n_{11}	n_{12}	$\cdots$	n_{1l}
A_2	n_{21}	n_{22}	$\cdots$	n_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_k	n_{k1}	n_{k2}	$\cdots$	n_{kl}

.

- Benötigte Größen $\widetilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ können dann — nach Ergänzung der Kontingenztabelle um ihre Randhäufigkeiten $n_{i\cdot} = \sum_{j=1}^{l} n_{ij}$ und $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$ — in weiterer Tabelle mit analogem Aufbau

$Y^A \setminus Y^B$	B_1	B_2	$\cdots$	B_l	$n_{i\cdot}$
A_1	$\widetilde{n}_{11} = \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n}$	$\widetilde{n}_{12} = \frac{n_{1\cdot} \cdot n_{\cdot 2}}{n}$	$\cdots$	$\widetilde{n}_{1l} = \frac{n_{1\cdot} \cdot n_{\cdot l}}{n}$	$n_{1\cdot}$
A_2	$\widetilde{n}_{21} = \frac{n_{2\cdot} \cdot n_{\cdot 1}}{n}$	$\widetilde{n}_{22} = \frac{n_{2\cdot} \cdot n_{\cdot 2}}{n}$	$\cdots$	$\widetilde{n}_{2l} = \frac{n_{2\cdot} \cdot n_{\cdot l}}{n}$	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_k	$\widetilde{n}_{k1} = \frac{n_{k\cdot} \cdot n_{\cdot 1}}{n}$	$\widetilde{n}_{k2} = \frac{n_{k\cdot} \cdot n_{\cdot 2}}{n}$	$\cdots$	$\widetilde{n}_{kl} = \frac{n_{k\cdot} \cdot n_{\cdot l}}{n}$	$n_{k\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot l}$	n

(hier für 2. Variante) oder (falls genügend Raum vorhanden) direkt in der Kontingenztabelle berechnet werden.

Zusammenfassung: Chi-Quadrat-Unabhängigkeitstest

Anwendungsvoraussetzungen	approximativ: (Y^A, Y^B) beliebig verteilt $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ einfache Stichprobe zu (Y^A, Y^B) Ausprägungen $\{a_1, \ldots, a_k\}$ von Y^A, $\{b_1, \ldots, b_l\}$ von Y^B *oder* Klassengrenzen $a_1 < \ldots < a_{k-1}$ zu Y^A, $b_1 < \ldots < b_{l-1}$ zu Y^B
Nullhypothese Gegenhypothese	$H_0 : Y^A, Y^B$ stochastisch unabhängig $H_1 : Y^A, Y^B$ nicht stochastisch unabhängig
Teststatistik	$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}} = \left(\sum_{i=1}^{k} \sum_{j=1}^{l} \frac{n_{ij}^2}{\widetilde{n}_{ij}} \right) - n$
Verteilung (H_0)	χ^2 ist näherungsweise $\chi^2((k-1) \cdot (l-1))$-verteilt, falls H_0 gilt (Näherung nur vernünftig, falls $\widetilde{n}_{ij} \geq 5$ für alle i, j)
Benötigte Größen	$n_{ij} = \#\{m \in \{1, \ldots, n\} \mid (x_m, y_m) \in A_i \times B_j\}$ für alle i, j mit $A_i = \{a_i\}$, $B_j = \{b_j\}$ bzw. Klassen A_i, B_j nach vorg. Grenzen, $\widetilde{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$ mit $n_{i\cdot} = \sum_{j=1}^{l} n_{ij}$, $n_{\cdot j} = \sum_{i=1}^{k} n_{ij}$,
Kritischer Bereich zum Niveau α	$(\chi^2_{(k-1)\cdot(l-1);1-\alpha}, \infty)$
p-Wert	$1 - F_{\chi^2((k-1)\cdot(l-1))}(\chi^2)$

Beispiel: Zusammenhang Geschlecht/tägl. Fahrzeit (PKW)

- Untersuchungsgegenstand: Sind die beiden Zufallsvariablen „Geschlecht" (Y^A) und „täglich mit PKW zurückgelegte Strecke" (Y^B) stochastisch unabhängig?
- Stichprobeninformation: (Kontingenz-)Tabelle mit gemeinsamen (in der Stichprobe vom Umfang $n = 2000$ beobachteten) Häufigkeiten, wobei für Y^B eine Klassierung in die Klassen „kurz", „mittel" und „lang" durchgeführt wurde:

	Fahrzeit (Y^B)		
Geschlecht (Y^A)	kurz	mittel	lang
Männlich	524	455	221
Weiblich	413	263	124

- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: **Chi-Quadrat-Unabhängigkeitstest**

1. **Hypothesen:**

   $H_0 : Y^A, Y^B$ stochastisch unabhängig gegen $H_1 : Y^A, Y^B$ stoch. abhängig

2. **Teststatistik:**

   $$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(n_{ij} - \widetilde{n}_{ij})^2}{\widetilde{n}_{ij}}$$ ist unter H_0 approximativ

   $\chi^2((k-1)\cdot(l-1))$-verteilt, falls $\widetilde{n}_{ij} \geq 5$ für alle $1 \leq i \leq k$ und $1 \leq j \leq l$.

3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**

   $K = (\chi^2_{(k-1)\cdot(l-1);1-\alpha}, +\infty) = (\chi^2_{2;0.95}, +\infty) = (5.991, +\infty)$

4. **Berechnung der realisierten Teststatistik:**

   Um Randhäufigkeiten $n_{i\cdot}$ und $n_{\cdot j}$ ergänzte Tabelle der gemeinsamen Häufigkeiten:

$Y^A \setminus Y^B$	kurz	mittel	lang	$n_{i\cdot}$
Männlich	524	455	221	1200
Weiblich	413	263	124	800
$n_{\cdot j}$	937	718	345	2000

Tabelle der $\widetilde{n}_{ij} = \dfrac{n_{i\cdot} \cdot n_{\cdot j}}{n}$:

$Y^A \setminus Y^B$	kurz	mittel	lang	$n_{i\cdot}$
Männlich	562.2	430.8	207.0	1200
Weiblich	374.8	287.2	138.0	800
$n_{\cdot j}$	937	718	345	2000

Es gilt $\widetilde{n}_{ij} \geq 5$ für alle $1 \leq i \leq 2$ und $1 \leq j \leq 3$ $\rightsquigarrow$ Näherung ok.

4. *(Fortsetzung: Berechnung der realisierten Teststatistik)*

$$
\begin{aligned}
\chi^2 &= \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(n_{ij}-\widetilde{n}_{ij})^2}{\widetilde{n}_{ij}} \\
&= \frac{(524-562.2)^2}{562.2}+\frac{(455-430.8)^2}{430.8}+\frac{(221-207)^2}{207} \\
&\quad +\frac{(413-374.8)^2}{374.8}+\frac{(263-287.2)^2}{287.2}+\frac{(124-138)^2}{138} \\
&= 2.5956+1.3594+0.9469 \\
&\quad +3.8934+2.0391+1.4203 \\
&= 12.2547
\end{aligned}
$$

5. **Entscheidung:**

$\chi^2 = 12.2547 \in (5.991, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

(p-Wert: $1 - F_{\chi^2(2)}(\chi^2) = 1 - F_{\chi^2(2)}(12.2547) = 1 - 0.9978 = 0.0022$)

Der Test kommt also zum Ergebnis, dass die beiden Zufallsvariablen „Geschlecht“ und „tägliche Fahrzeit (PKW)“ stochastisch **abhängig** sind.

Mittelwertvergleiche

- *Nächste Anwendung:* Vergleich der Mittelwerte zweier *normalverteilter* Zufallsvariablen Y^A und Y^B
  1. auf **derselben** Grundgesamtheit durch Beobachtung von Realisationen $(x_1^A, x_1^B), \ldots, (x_n^A, x_n^B)$ einer (gemeinsamen) einfachen Stichprobe $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ zur **zweidimensionalen** Zufallsvariablen (Y^A, Y^B), insbesondere von Realisationen von Y^A und Y^B für **dieselben** Elemente der Grundgesamtheit („verbundene Stichprobe"),
  2. auf **derselben oder unterschiedlichen** Grundgesamtheit(en) durch Beobachtung von Realisationen $x_1^A, \ldots, x_{n_A}^A$ und $x_1^B, \ldots, x_{n_B}^B$ zu zwei **unabhängigen** einfachen Stichproben $X_1^A, \ldots, X_{n_A}^A$ und $X_1^B, \ldots, X_{n_B}^B$ (möglicherweise mit $n_A \neq n_B$) zu den beiden Zufallsvariablen Y^A und Y^B.
- Anwendungsbeispiele für beide Fragestellungen:
  1. Vergleich der Montagezeiten zweier unterschiedlicher Montageverfahren auf Grundlage von Zeitmessungen beider Verfahren *für dieselbe (Stichproben-)Auswahl von Arbeitern*.
  2. Vergleich der in Eignungstests erreichten Punktzahlen von männlichen und weiblichen Bewerbern (auf Basis zweier unabhängiger einfacher Stichproben).

t-Differenzentest bei verbundener Stichprobe

- Idee für Mittelwertvergleich bei verbundenen Stichproben:
  - Ein Vergleich der Mittelwerte von Y^A und Y^B kann anhand des Mittelwerts $\mu := \mathrm{E}(Y)$ der Differenz $Y := Y^A - Y^B$ erfolgen, denn mit $\mu_A := \mathrm{E}(Y^A)$ und $\mu_B := \mathrm{E}(Y^B)$ gilt offensichtlich $\mu = \mu_A - \mu_B$ und damit:

    $$\mu < 0 \iff \mu_A < \mu_B \qquad \mu = 0 \iff \mu_A = \mu_B \qquad \mu > 0 \iff \mu_A > \mu_B$$

  - Mit $x_1 := x_1^A - x_1^B, \ldots, x_n := x_n^A - x_n^B$ liegt eine Realisation einer einfachen Stichprobe $X_1 := X_1^A - X_1^B, \ldots, X_n := X_n^A - X_n^B$ vom Umfang n zu $Y = Y^A - Y^B$ vor.
  - Darüberhinaus gilt: Ist (Y^A, Y^B) gemeinsam (zweidimensional) normalverteilt, so ist auch die Differenz $Y = Y^A - Y^B$ normalverteilt.
- Es liegt also nahe, die gemeinsame Stichprobe zu (Y^A, Y^B) zu „einer“ Stichprobe zu $Y = Y^A - Y^B$ zusammenzufassen und den bekannten t-Test für den Mittelwert einer (normalverteilten) Zufallsvariablen bei unbekannter Varianz auf der Grundlage der einfachen Stichprobe $X_1, \ldots, X_n$ zu Y durchzuführen.
- Prinzipiell wäre bei bekannter Varianz von $Y = Y^A - Y^B$ auch ein entsprechender Gauß-Test durchführbar; Anwendungen hierfür sind aber selten.

Zusammenfassung: *t*-Differenzentest

Anwendungs-voraussetzungen	exakt: (Y^A, Y^B) gemeinsam (zweidimensional) normalverteilt, $\mathrm{E}(Y^A) = \mu_A, \mathrm{E}(Y^B) = \mu_B$ sowie Varianzen/Kovarianz unbekannt approx.: $\mathrm{E}(Y^A) = \mu_A, \mathrm{E}(Y^B) = \mu_B, \mathrm{Var}(Y^A), \mathrm{Var}(Y^B)$ unbek. $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ einfache Stichprobe zu (Y^A, Y^B)		
Nullhypothese Gegenhypothese	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A \neq \mu_B$	$H_0 : \mu_A \leq \mu_B$ $H_1 : \mu_A > \mu_B$	$H_0 : \mu_A \geq \mu_B$ $H_1 : \mu_A < \mu_B$
Teststatistik	$t = \frac{\overline{X}}{S}\sqrt{n}$		
Verteilung (H_0)	t für $\mu_A = \mu_B$ (näherungsweise) $t(n-1)$-verteilt		
Benötigte Größen	$X_i = X_i^A - X_i^B$ für $i \in \{1, \ldots, n\}$, $\quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)}$		
Kritischer Bereich zum Niveau α	$(-\infty, -t_{n-1;1-\frac{\alpha}{2}}) \cup (t_{n-1;1-\frac{\alpha}{2}}, \infty)$	$(t_{n-1;1-\alpha}, \infty)$	$(-\infty, -t_{n-1;1-\alpha})$
p-Wert	$2 \cdot (1 - F_{t(n-1)}(\lvert t \rvert))$	$1 - F_{t(n-1)}(t)$	$F_{t(n-1)}(t)$

Beispiel: Montagezeiten von zwei Verfahren

- Untersuchungsgegenstand: Ist ein neu vorgeschlagenes Montageverfahren besser (im Sinne einer im Mittel kürzeren Bearbeitungsdauer Y^B) als das zur Zeit eingesetzte Montageverfahren (mit Bearbeitungsdauer Y^A)?
- Stichprobeninformation: Zeitmessungen der Montagedauern x_i^A für Verfahren A und x_i^B für Verfahren B bei **denselben** $n = 7$ Arbeitern:

Arbeiter i	1	2	3	4	5	6	7
x_i^A	64	71	68	66	73	62	70
x_i^B	60	66	66	69	63	57	62

- Annahme: (Y^A, Y^B) gemeinsam normalverteilt, $(X_1^A, X_1^B), \ldots, (X_n^A, X_n^B)$ einfache Stichprobe zu (Y^A, Y^B).
- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: Exakter t-**Differenzentest** für verbundene Stichproben

1. **Hypothesen:**

   $H_0 : \mu_A \leq \mu_B \quad$ gegen $\quad H_1 : \mu_A > \mu_B$

2. **Teststatistik:**

   $t = \frac{\overline{X}}{S}\sqrt{n}$ ist unter H_0 $t(n-1)$-verteilt (für $\mu_A{=}\mu_B$).

3. **Kritischer Bereich zum Niveau $\alpha = 0.05$:**

   $K = (t_{n-1;1-\alpha}, +\infty) = (t_{6;0.95}, +\infty) = (1.943, +\infty)$

4. **Berechnung der realisierten Teststatistik:**

Arbeiter i	1	2	3	4	5	6	7
x_i^A	64	71	68	66	73	62	70
x_i^B	60	66	66	69	63	57	62
$x_i = x_i^A - x_i^B$	4	5	2	-3	10	5	8

   Mit $\overline{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 4.4286$ und $s = \sqrt{\frac{1}{7-1}\sum_{i=1}^{7}(x_i - \overline{x})^2} = 4.1975$:

   $$t = \frac{\overline{x}}{s}\sqrt{n} = \frac{4.4286}{4.1975}\sqrt{7} = 2.7914$$

5. **Entscheidung:**

   $t = 2.7914 \in (1.943, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!

   (p-Wert: $1 - F_{t(6)}(t) = 1 - F_{t(6)}(2.7914) = 1 - 0.9842 = 0.0158$)

Der Test kommt also zur Entscheidung, dass das neue Montageverfahren eine im Mittel signifikant kürzere Montagedauer aufweist.