Die Familie der $\chi^2(n)$-Verteilungen

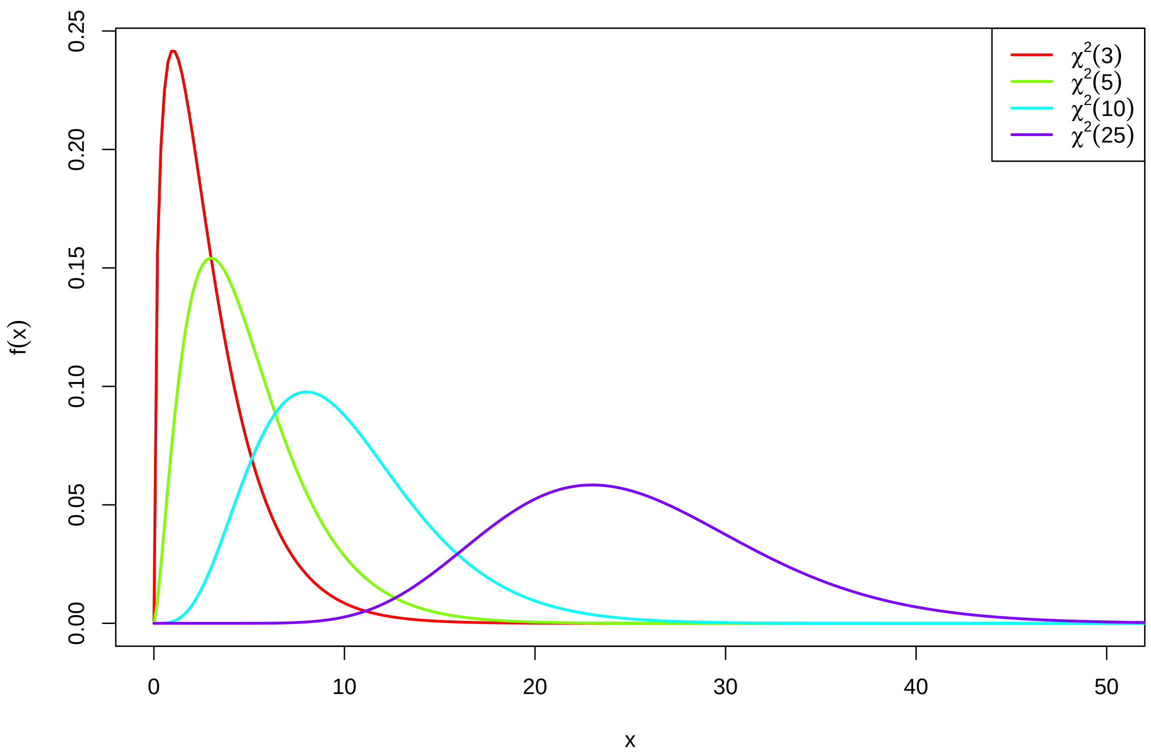
- Sind $Z_1, \ldots, Z_m$ für $m \geq 1$ unabhängig identisch standardnormalverteilte Zufallsvariablen, so genügt die Summe der *quadrierten* Zufallsvariablen

$$\chi^2 := \sum_{i=1}^{m} Z_i^2 = Z_1^2 + \ldots + Z_m^2$$

einer sog. **Chi-Quadrat-Verteilung mit** m **Freiheitsgraden**, in Zeichen $\chi^2 \sim \chi^2(m)$.
- Offensichtlich können $\chi^2(m)$-verteilte Zufallsvariablen nur nichtnegative Werte annehmen, der Träger ist also $[0, \infty)$.
- Ist $\chi^2 \sim \chi^2(m)$, so gilt $\mathrm{E}(\chi^2) = m$ sowie $\mathrm{Var}(\chi^2) = 2m$.
- Als Abkürzung für α-Quantile der $\chi^2(m)$-Verteilung verwenden wir (wie üblich) $\chi^2_{m;\alpha}$.

Grafische Darstellung einiger $\chi^2(m)$-Verteilungen

für $m \in \{3, 5, 10, 25\}$



Tests für die Varianz

- Für Aussagen über die Varianz von Y (als mittlere quadrierte Abweichung vom Erwartungswert) auf Basis einer einfachen Stichprobe $X_1, \ldots, X_n$ zu Y naheliegend: Untersuchung der quadrierten Abweichungen
$$(X_1 - \mu)^2, \ldots, (X_n - \mu)^2$$
bei bekanntem Erwartungswert $\mu = \mathrm{E}(Y)$ bzw. bei unbekanntem Erwartungswert der quadrierten Abweichungen vom Stichprobenmittelwert
$$(X_1 - \overline{X})^2, \ldots, (X_n - \overline{X})^2 \ .$$
- Man kann zeigen: Ist $Y \sim N(\mu, \sigma^2)$, so gilt
$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{(X_1 - \mu)^2}{\sigma^2} + \ldots + \frac{(X_n - \mu)^2}{\sigma^2} \sim \chi^2(n)$$
bzw. mit der Abkürzung $\widetilde{S}^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ für die mittlere quadratische Abweichung vom *bekannten* Erwartungswert aus der Stichprobe
$$\frac{n\widetilde{S}^2}{\sigma^2} \sim \chi^2(n) \ .$$

- Hieraus lassen sich analog zu den Tests für den Mittelwert Tests auf Abweichung der Varianz $\mathrm{Var}(Y)$ von einer vorgegebenen „Soll-Varianz" σ_0^2 entwickeln:
  - **Über**schreitet die tatsächliche Varianz von Y die (unter H_0 angenommene) „Soll-Varianz" σ_0^2, so verschiebt sich die Verteilung der Größe $\chi^2 := \frac{n\widetilde{S}^2}{\sigma_0^2}$ offensichtlich nach **rechts**.
  - **Unter**schreitet die tatsächliche Varianz von Y die (unter H_0 angenommene) „Soll-Varianz" σ_0^2, so verschiebt sich die Verteilung der Größe $\chi^2 := \frac{n\widetilde{S}^2}{\sigma_0^2}$ offensichtlich nach **links**.
- Gilt $Y \sim N(\mu, \sigma^2)$ und ist der Erwartungswert μ *unbekannt*, so kann weiter gezeigt werden, dass
$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} = \frac{(X_1 - \overline{X})^2}{\sigma^2} + \ldots + \frac{(X_n - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$
bzw. mit der bekannten Abkürzung $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ für die Stichprobenvarianz
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$
gilt, woraus ebenfalls Tests für die Varianz abgeleitet werden können.

Bemerkungen

- Bei der Konstruktion der kritischen Bereich ist zu beachten, dass die Testgrößen
$$\chi^2 = \frac{n\widetilde{S}^2}{\sigma_0^2} \qquad \text{bzw.} \qquad \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$
nur nichtnegative Wert annehmen können.
- Durch die fehlende Symmetrie sind viele von Gauß- und t-Tests bekannte Vereinfachungen nicht mehr möglich. Insbesondere
  - darf $\chi^2_{m;\alpha}$ nicht durch $-\chi^2_{m;1-\alpha}$ ersetzt werden,
  - kann die Berechnung des p-Werts im zweiseitigen Test nicht vereinfacht werden.

Wichtig!

Die Normalverteilungsannahme $Y \sim N(\mu, \sigma^2)$ ist für den Chi-Quadrat-Test für die Varianz wesentlich. Weicht die Verteilung von Y „deutlich" von einer Normalverteilung ab, unterscheidet sich die Verteilung der Testgröße χ^2 (auch unter H_0 für $\sigma^2 = \sigma_0^2$!) wesentlich von einer $\chi^2(n)$ bzw. $\chi^2(n-1)$-Verteilung.

Quantile der χ^2-Verteilungen: $\chi^2_{n;p}$

$n\backslash p$	0.01	0.025	0.05	0.50	0.90	0.95	0.975	0.99
1	0.000	0.001	0.004	0.455	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	1.386	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	2.366	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	3.357	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	4.351	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	5.348	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	6.346	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	7.344	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	8.343	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	9.342	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	10.341	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	11.340	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	12.340	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	13.339	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	14.339	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	15.338	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	16.338	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	17.338	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	18.338	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	19.337	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	20.337	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	21.337	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	22.337	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	23.337	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	24.337	34.382	37.652	40.646	44.314

Zusammenfassung: χ^2-Test für die Varianz

einer normalverteilten Zufallsvariablen mit **bekanntem** Erwartungswert

Anwendungs-voraussetzungen	exakt: $Y \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ **bekannt**, $\sigma^2 \in \mathbb{R}_{++}$ unbekannt $X_1, \ldots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$
Teststatistik Verteilung (H_0)	$\chi^2 = \frac{n \cdot \widetilde{S}^2}{\sigma_0^2}$ χ^2 (für $\sigma^2 = \sigma_0^2$) $\chi^2(n)$-verteilt		
Benötigte Größen	$\widetilde{S}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$		
Kritischer Bereich zum Niveau α	$[0, \chi^2_{n;\frac{\alpha}{2}})$ $\cup(\chi^2_{n;1-\frac{\alpha}{2}}, \infty)$	$(\chi^2_{n;1-\alpha}, \infty)$	$[0, \chi^2_{n;\alpha})$
p-Wert	$2 \cdot \min \left\{ F_{\chi^2(n)}(\chi^2), \right.$ $\left. 1 - F_{\chi^2(n)}(\chi^2) \right\}$	$1 - F_{\chi^2(n)}(\chi^2)$	$F_{\chi^2(n)}(\chi^2)$

Beispiel: Präzision einer Produktionsanlage

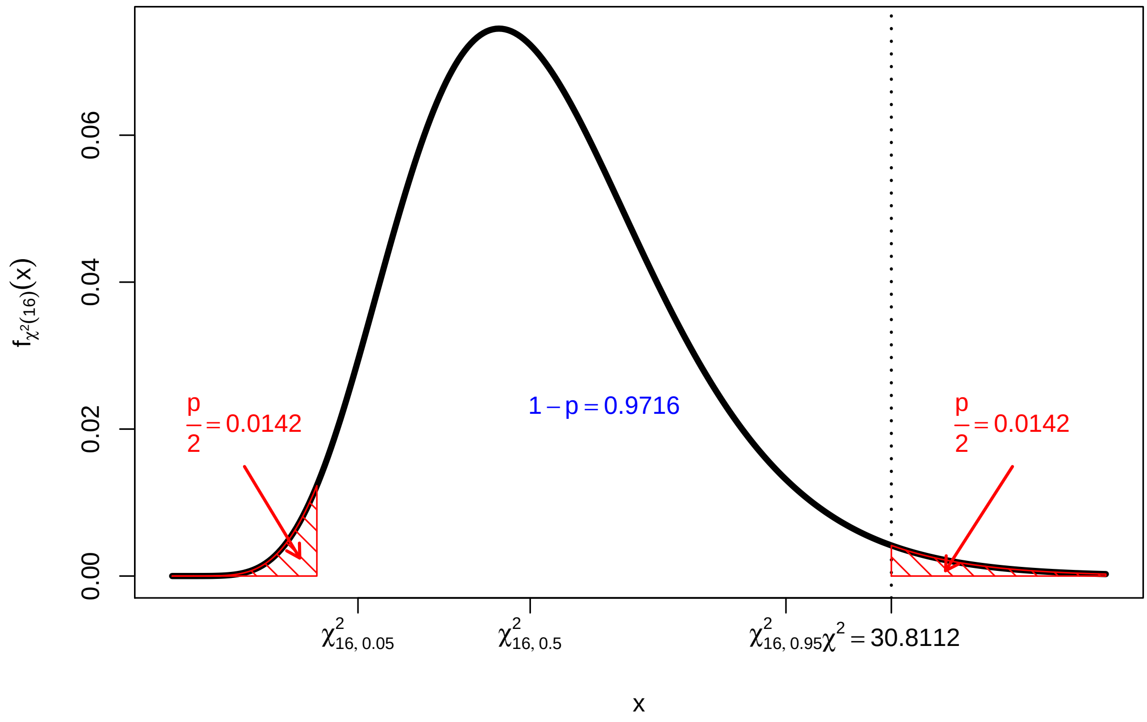
- Untersuchungsgegenstand: Bei einer Produktionsanlage für Maßbänder soll geprüft werden, ob die Herstellerangabe für die Produktionsgenauigkeit korrekt ist. Laut Hersteller ist die Länge der produzierten Maßbänder normalverteilt mit Erwartungswert 200 [mm] und Varianz $\sigma^2 = 0.1^2$. Der Betreiber der Anlage vermutet eine Abweichung der Präzision.
- Annahmen: Länge $Y \sim N(200, \sigma^2)$ mit σ^2 unbekannt.
- Stichprobeninformation: Realisation einer einfachen Stichprobe vom Umfang $n = 16$ zu Y liefert $\widetilde{S}^2 = \frac{1}{16} \sum_{i=1}^{16} (X_i - 200)^2 = 0.019257$.
- Gewünschtes Signifikanzniveau: $\alpha = 0.10$

Geeigneter Test:
Zweiseitiger Chi-Quadrat-Test für Varianz bei bekanntem Erwartungswert

1. Hypothesen: $H_0 : \sigma^2 = \sigma_0^2 = 0.1^2$ gegen $H_1 : \sigma^2 \neq \sigma_0^2 = 0.1^2$
2. Teststatistik: $\chi^2 = \frac{n \cdot \widetilde{S}^2}{\sigma_0^2} \sim \chi^2(16)$, falls H_0 gilt ($\sigma^2 = \sigma_0^2$)
3. Kritischer Bereich zum Niveau $\alpha = 0.10$:
   $K = [0, \chi^2_{16;0.05}) \cup (\chi^2_{16;0.95}, \infty) = [0, 7.962) \cup (26.296, \infty)$
4. Realisierter Wert der Teststatistik: $\chi^2 = \frac{16 \cdot 0.019257}{0.01} = 30.8112$
5. Entscheidung: $\chi^2 \in K \rightsquigarrow H_0$ wird abgelehnt; Test kommt zur Entscheidung, dass die Präzision von der Herstellerangabe abweicht.

Beispiel: p-Wert bei zweiseitigem χ^2-Test (Grafik)

Produktionsmaschinenbeispiel, realisierte Teststatistik $\chi^2 = 30.8112$, p-Wert: 0.0284



Zusammenfassung: χ^2-Test für die Varianz

einer normalverteilten Zufallsvariablen mit **unbekanntem** Erwartungswert

Anwendungsvoraussetzungen	exakt: $Y \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ **unbekannt**, $\sigma^2 \in \mathbb{R}_{++}$ unbekannt $X_1, \ldots, X_n$ einfache Stichprobe zu Y		
Nullhypothese Gegenhypothese	$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$H_0 : \sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$H_0 : \sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$
Teststatistik Verteilung (H_0)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ χ^2 (für $\sigma^2 = \sigma_0^2$) $\chi^2(n-1)$-verteilt		
Benötigte Größen	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$ mit $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$		
Kritischer Bereich zum Niveau α	$[0, \chi^2_{n-1;\frac{\alpha}{2}})$ $\cup(\chi^2_{n-1;1-\frac{\alpha}{2}}, \infty)$	$(\chi^2_{n-1;1-\alpha}, \infty)$	$[0, \chi^2_{n-1;\alpha})$
p-Wert	$2 \cdot \min\left\{F_{\chi^2(n-1)}(\chi^2),\right.$ $\left.1 - F_{\chi^2(n-1)}(\chi^2)\right\}$	$1 - F_{\chi^2(n-1)}(\chi^2)$	$F_{\chi^2(n-1)}(\chi^2)$

Beispiel: Präzision einer neuen Abfüllmaschine

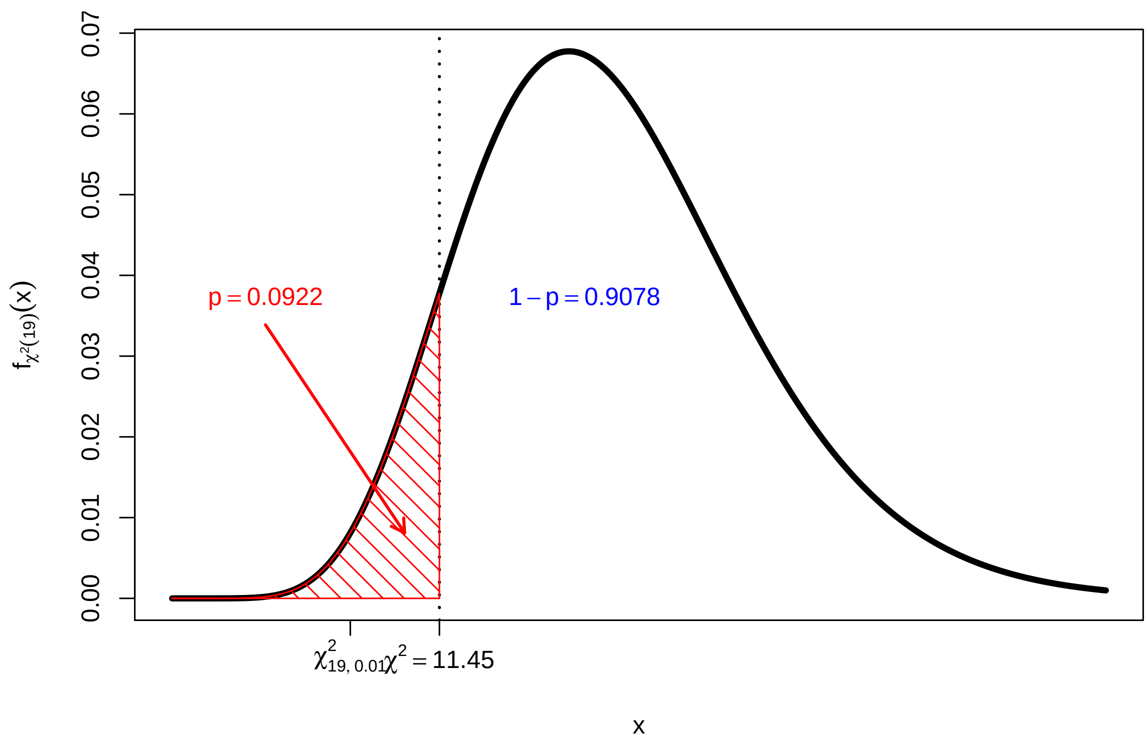
- Untersuchungsgegenstand: Für eine neue Abfüllmaschine wird geprüft, ob sie **präziser** als die alte Anlage arbeitet. Bei der alten Maschine beträgt die Standardabweichung des Füllgewichts um den eingestellten Wert 5 [g].
- Annahmen: Füllgewicht $Y \sim N(\mu, \sigma^2)$ mit μ, σ^2 unbekannt.
- Stichprobeninformation: Realisation einer einfachen Stichprobe vom Umfang $n = 20$ zu Y liefert Stichprobenmittel $\overline{x} = 25.8097$ und mittleres Quadrat $\overline{x^2} = 680.4535$, damit also $s^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right) = 15.066$.
- Gewünschtes Signifikanzniveau: $\alpha = 0.01$

Geeigneter Test: Linksseitiger Chi-Quadrat-Test für Varianz bei unbekanntem Erwartungswert

1. Hypothesen: $H_0 : \sigma^2 \geq \sigma_0^2 = 5^2 \quad$ gegen $\quad H_1 : \sigma^2 < \sigma_0^2 = 5^2$
2. Teststatistik: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(19)$, falls H_0 gilt $(\sigma^2 = \sigma_0^2)$
3. Kritischer Bereich zum Niveau $\alpha = 0.01$: $K = [0, \chi^2_{19;0.01}) = [0, 7.633)$
4. Realisierter Wert der Teststatistik: $\chi^2 = \frac{19 \cdot 15.066}{25} = 11.45$
5. Entscheidung: $\chi^2 \notin K \rightsquigarrow H_0$ wird **nicht** abgelehnt; Test kommt zur Entscheidung, dass es keine ausreichende statistische Evidenz für eine bessere Präzision der neueren Maschine gibt.

Beispiel: p-Wert bei linksseitigem χ^2-Test (Grafik)

Abfüllmaschinenbeispiel, realisierte Teststatistik $\chi^2 = 11.45$, p-Wert: 0.0922



Chi-Quadrat-Anpassungstest

- *Ziel:* Konstruktion eines Tests zur Überprüfung, ob Zufallsvariable Y einer bestimmten **Verteilung** (oder *später* allgemeiner: einer bestimmten Verteilungsklasse) folgt, **ohne** mögliche Verteilungen von Y bereits durch (parametrische) Verteilungsannahme eingrenzen zu müssen.
- *Eine* Möglichkeit: **Chi-Quadrat-Anpassungstest**
- *Grundlegende Idee:* Vergleich der **empirischen Häufigkeitsverteilung** aus der Stichprobenrealisation $(X_1, \dots, X_n)$ mit den **(theoretischen) Wahrscheinlichkeiten** der **hypothetischen** (d.h. unter H_0 angenommenen) Verteilung von Y.
- Hierzu nötig:
  - Erstellen der empirischen Häufigkeitsverteilung — bei diskreter hypothetischer Verteilung mit „vielen" Trägerpunkten bzw. stetiger hypothetischer Verteilung nach erfolgter Klassierung —
  - Berechnen der theoretischen Punkt- bzw. Klassenwahrscheinlichkeiten unter der hypothetischen Verteilung.
- Offensichtlich: **Große** Abweichungen der empirischen (in der Stichprobe beobachteten) Häufigkeiten von den theoretischen Wahrscheinlichkeiten sprechen eher **gegen** die hypothetische Verteilung von Y, **kleine** Abweichungen eher **dafür**.

- Noch nötig: Geeignete Testgröße zur Zusammenfassung der Abweichungen sowie Verteilungsaussage für die Testgröße bei Gültigkeit von H_0.
- $(X_1, \dots, X_n)$ sei (wie immer) eine einfache Stichprobe vom Umfang n zu Y.
- Bezeichnen
  - k die Anzahl der Ausprägungen bzw. Klassen der empirischen Häufigkeitsverteilung,
  - n_i für $i \in \{1, \dots, k\}$ die in der Stichprobe aufgetretenen (absoluten) Häufigkeiten für Ausprägung i bzw. Klasse i,
  - p_i^0 die bei Gültigkeit der hypothetischen Verteilung für Y tatsächlichen Wahrscheinlichkeiten für Ausprägung i bzw. Klasse i,

  so werden die Abweichungen $\frac{n_i}{n} - p_i^0$ (beim Vergleich relativer Häufigkeiten und Wahrscheinlichkeiten) bzw. $n_i - np_i^0$ (beim Vergleich absoluter Häufigkeiten und erwarteter Häufigkeiten) mit der Testgröße

$$\chi^2 := n \sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i^0\right)^2}{p_i^0} = \sum_{i=1}^{k} \frac{\left(n_i - np_i^0\right)^2}{np_i^0}$$

  zusammengefasst.
- Ist H_0 gültig, so konvergiert die Verteilung von χ^2 mit wachsendem n gegen die $\chi^2(k-1)$-Verteilung.

- Offensichtlich: Große Werte von χ^2 entstehen bei großen Abweichungen zwischen beobachteten Häufigkeiten und Wahrscheinlichkeiten bzw. erwarteten Häufigkeiten und sprechen damit gegen H_0.
- Sinnvoller kritischer Bereich zum Signifikanzniveau α also $(\chi^2_{k-1;1-\alpha};\infty)$.
- χ^2-Anpassungstest ist immer approximativer (näherungsweiser) Test. Vernünftige Näherung der Verteilung von χ^2 (unter H_0) durch $\chi^2(k-1)$-Verteilung kann nur erwartet werden, wenn $np_i^0 \geq 5$ für alle $i \in \{1,\dots,k\}$ gilt.
- Berechnung der p_i^0 zur Durchführung des Chi-Quadrat-Anpassungstest je nach Anwendung sehr unterschiedlich:
  - Bei diskreter hypothetischer Verteilung mit endlichem Träger in der Regel (falls $np_i^0 \geq 5$ für alle $i \in \{1,\dots,k\}$) besonders einfach, da keine Klassierung erforderlich ist und sich alle p_i^0 direkt als Punktwahrscheinlichkeiten ergeben.
  - Bei diskreter hypothetischer Verteilung mit unendlichem Träger bzw. bei Verletzung der Bedingung $np_i^0 \geq 5$ für alle $i \in \{1,\dots,k\}$ Klassierung (trotz diskreter Verteilung) erforderlich, so dass Bedingung erfüllt wird.
  - Bei stetiger hypothetischer Verteilung Klassierung stets erforderlich; Durchführung so, dass Bedingung $np_i^0 \geq 5$ für alle $i \in \{1,\dots,k\}$ erfüllt ist.
- Sobald p_i^0 (ggf. nach Klassierung) bestimmt sind, identische Vorgehensweise für alle Verteilungen.

Chi-Quadrat-Anpassungstest

zur Anpassung an eine hypothetische Verteilung

- Hypothesenformulierung z.B. über Verteilungsfunktion F_0 der hypothetischen Verteilung in der Form:

$$H_0 : F_Y = F_0 \qquad \text{gegen} \qquad H_1 : F_Y \neq F_0$$

- Allgemeine Vorgehensweise: Bilden von k Klassen durch Aufteilen der reellen Zahlen in k Intervalle

$$K_1 = (-\infty, a_1], K_2 = (a_1, a_2], \dots, K_{k-1} = (a_{k-2}, a_{k-1}], K_k = (a_{k-1}, \infty)$$

  und Berechnen der theoretischen Klassenwahrscheinlichkeiten p_i^0 als $p_i^0 = F_0(a_k) - F_0(a_{k-1})$ mit $a_0 := -\infty$ und $a_k := \infty$, also

$$\begin{aligned} p_1^0 &= F_0(a_1) - F_0(-\infty) = F_0(a_1), \\ p_2^0 &= F_0(a_2) - F_0(a_1), \\ &\vdots \\ p_{k-1}^0 &= F_0(a_{k-1}) - F_0(a_{k-2}), \\ p_k^0 &= F_0(\infty) - F_0(a_{k-1}) = 1 - F_0(a_{k-1}) . \end{aligned}$$

Zusammenfassung: Chi-Quadrat-Anpassungstest

zur Anpassung an **eine** vorgegebene Verteilung

Anwendungs-voraussetzungen	approximativ: Y beliebig verteilt $X_1, \ldots, X_n$ einfache Stichprobe zu Y $k-1$ Klassengrenzen $a_1 < a_2 < \ldots < a_{k-1}$ vorgegeben
Nullhypothese Gegenhypothese	$H_0 : F_Y = F_0$ $H_1 : F_Y \neq F_0$
Teststatistik Verteilung (H_0)	$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} = n \sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i^0\right)^2}{p_i^0} = \left(\frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_i^0}\right) - n$ χ^2 ist näherungsweise $\chi^2(k-1)$-verteilt, falls $F_Y = F_0$ (Näherung nur vernünftig, falls $np_i^0 \geq 5$ für $i \in \{1, \ldots, k\}$)
Benötigte Größen	$p_i^0 = F_0(a_i) - F_0(a_{i-1})$ 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	$(\chi^2_{k-1;1-\alpha}, \infty)$
p-Wert	$1 - F_{\chi^2(k-1)}(\chi^2)$

Vereinfachung bei diskreter hypothetischer Verteilung

mit k Trägerpunkten

- Einfachere „Notation" bei Anwendung des Chi-Quadrat-Anpassungstests meist möglich, falls hypothetische Verteilung diskret mit k Trägerpunkten $a_1, \ldots, a_k$.
- Bezeichnet p_0 die Wahrscheinlichkeitsfunktion der hypothetischen Verteilungen und gilt $n \cdot p_0(a_i) \geq 5$ für alle $i \in \{1, \ldots, k\}$, so ist keine „echte" Klassierung erforderlich (1 Trägerpunkt pro „Klasse").
- Man erhält dann die hypothetischen Punktwahrscheinlichkeiten p_i^0 als $p_i^0 = p_0(a_i)$.
- Hypothesen meist direkt über Vektor der Punktwahrscheinlichkeiten $p := (p_1, \ldots, p_k) := (p_Y(a_1), \ldots, p_Y(a_k))$ in der Form:
$$H_0 : p = (p_1, \ldots, p_k) = (p_1^0, \ldots, p_k^0) =: p^0 \quad \text{gegen} \quad H_1 : p \neq p^0$$
- Chi-Quadrat-Anpassungstest kann so auch auf „Merkmale" angewendet werden, deren Ausprägungen noch nicht „Zufallsvariablen-konform" durch (reelle) Zahlenwerte ausgedrückt (kodiert) worden sind, beispielsweise bei
  - Wochentagen: a_1=„Montag", a_2=„Dienstag", ...
  - Produktmarken: a_1=„Automarke A", a_2=„Automarke B", ...
  - Monaten: a_1=„Januar", a_2=„Februar", ...

Beispiel: Verteilung Auftragseingänge

auf 5 Wochentage Montag–Freitag (diskrete hypothetische Verteilung)

- Untersuchungsgegenstand: Sind die Auftragseingänge in einem Unternehmen gleichmäßig auf die 5 Arbeitstage Montag–Freitag verteilt, d.h, ist der Anteil der Auftragseingänge an jedem Wochentag gleich 0.2?
  [$\rightsquigarrow p^0 = (0.2, 0.2, 0.2, 0.2, 0.2)$]
- Stichprobeninformation: Einfache Stichprobe von 400 Auftragseingängen liefert folgende Verteilung auf Wochentage:

	Mo	Di	Mi	Do	Fr
n_i	96	74	92	81	57

- Gewünschtes Signifikanzniveau: $\alpha = 0.05$

Geeigneter Test: Chi-Quadrat-Anpassungstest

1. **Hypothesen:**
   $H_0 : p = p^0 = (0.2, 0.2, 0.2, 0.2, 0.2) \qquad H_1 : p \neq p^0$
2. **Teststatistik:**
   $$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} \text{ ist unter } H_0 \text{ approximativ } \chi^2(k-1)\text{-verteilt;}$$
   Näherung vernünftig, falls $np_i^0 \geq 5$ für alle i gilt.

3. **Kritischer Bereich zum Niveau** $\alpha = 0.05$**:**
   $K = (\chi^2_{k-1;1-\alpha}, +\infty) = (\chi^2_{4;0.95}, +\infty) = (9.488, +\infty)$
4. **Berechnung der realisierten Teststatistik:**

a_i	n_i	p_i^0	np_i^0	$\frac{(n_i-np_i^0)^2}{np_i^0}$
Mo	96	0.2	80	3.2000
Di	74	0.2	80	0.4500
Mi	92	0.2	80	1.8000
Do	81	0.2	80	0.0125
Fr	57	0.2	80	6.6125
Σ	400	1	400	$\chi^2 = 12.0750$

   Es gilt $np_i^0 \geq 5$ für alle $i \in \{1, \ldots, 5\} \rightsquigarrow$ Näherung ok.
5. **Entscheidung:**
   $\chi^2 = 12.075 \in (9.488, +\infty) = K \quad \Rightarrow \quad H_0$ wird abgelehnt!
   (p-Wert: $1 - F_{\chi^2(4)}(\chi^2) = 1 - F_{\chi^2(4)}(12.075) = 1 - 0.9832 = 0.0168$)

Test kommt zur Entscheidung, dass die Auftragseingänge nicht gleichmäßig auf alle 5 Arbeitstage (Montag-Freitag) verteilt sind.