8. Verteilungstests

- χ^2-Unabhängigkeitstest
- χ^2-Anpassungstest
- Kolmogorov-Smirnov-Test

Verteilungstests

Bezeichnung für Tests, die einen Verteilungstyp oder eine Verteilungseigenschaft ohne parametrische Annahme prüfen. Hier besprochen:

- $\hookrightarrow$ χ^2-Unabhängigkeitstest (Kontingenztest),
  - □ Test mit zwei gebundenen Stichproben
  - □ Nullhypothese: Stichproben zu zwei „unabhängigen Grundgesamtheiten“
- $\hookrightarrow$ χ^2-Anpassungstest
  - □ Einstichprobentest
  - □ Vergleicht diskretes (bzw. diskretisiertes) Merkmal mit einer vorgegebenen endlich-diskreten Verteilung.
- $\hookrightarrow$ Kolmogorov-Smirnov-Anpassungstest
  - □ Einstichprobentest
  - □ Prüft, ob Daten zu einer vorgegebenen Verteilung passen
  - □ Vergleicht empirische und theoretische Verteilungsfunktion.

8.1 χ^2-Unabhängigkeitstest (Kontingenztest)

$\hookrightarrow$ Datenbasis ist verbundene u.i.v.-Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$.

$\hookrightarrow$ Es soll geprüft werden, ob X_ℓ, Y_ℓ st.u. sind

$\hookrightarrow$ Kann in dieser Allgemeinheit nicht getestet werden.

8.1 χ^2-Unabhängigkeitstest (Kontingenztest)

- $\hookrightarrow$ Datenbasis ist verbundene u.i.v.-Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$.
- $\hookrightarrow$ Es soll geprüft werden, ob X_ℓ, Y_ℓ st.u. sind
- $\hookrightarrow$ Kann in dieser Allgemeinheit nicht getestet werden.
- $\hookrightarrow$ Vereinfachende Annahmen (falls nicht erfüllt, durch Klassierung erzwingen):
  - □ $X_1, \ldots, X_n$ sind diskrete ZV mit Werten $A_1, \ldots, A_K$
  - □ $Y_1, \ldots, Y_n$ sind diskrete ZV mit Werten $B_1, \ldots, B_L$

8.1 χ^2-Unabhängigkeitstest (Kontingenztest)

$\hookrightarrow$ Datenbasis ist verbundene u.i.v.-Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$.

$\hookrightarrow$ Es soll geprüft werden, ob X_ℓ, Y_ℓ st.u. sind

$\hookrightarrow$ Kann in dieser Allgemeinheit nicht getestet werden.

$\hookrightarrow$ Vereinfachende Annahmen (falls nicht erfüllt, durch Klassierung erzwingen):

- $X_1, \dots, X_n$ sind diskrete ZV mit Werten $A_1, \dots, A_K$
- $Y_1, \dots, Y_n$ sind diskrete ZV mit Werten $B_1, \dots, B_L$

1. Hypothese

H_0 : X und Y sind st. unabhängig vs. H_1 : X und Y sind st. abhängig

$\hookrightarrow$ Unter H_0 gilt also

$$P(X_\ell = A_i, Y_\ell = B_j) = P(X_\ell = A_i)P(Y_\ell = B_j) \quad \forall i, j \qquad (*)$$

8.1 χ^2-Unabhängigkeitstest (Kontingenztest)

$\hookrightarrow$ Datenbasis ist verbundene u.i.v.-Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$.

$\hookrightarrow$ Es soll geprüft werden, ob X_ℓ, Y_ℓ st.u. sind

$\hookrightarrow$ Kann in dieser Allgemeinheit nicht getestet werden.

$\hookrightarrow$ Vereinfachende Annahmen (falls nicht erfüllt, durch Klassierung erzwingen):

- $X_1, \ldots, X_n$ sind diskrete ZV mit Werten $A_1, \ldots, A_K$
- $Y_1, \ldots, Y_n$ sind diskrete ZV mit Werten $B_1, \ldots, B_L$

1. Hypothese

H_0 : X und Y sind st. unabhängig vs. H_1 : X und Y sind st. abhängig

$\hookrightarrow$ Unter H_0 gilt also

$$P(X_\ell = A_i, Y_\ell = B_j) = P(X_\ell = A_i)P(Y_\ell = B_j) \quad \forall i, j \qquad (*)$$

$\hookrightarrow$ Idee des χ^2-Tests: Stelle den theoretischen WS geeignete empirische Kennzahlen gegenüber und formuliere $(*)$ approximativ mit diesen Kennzahlen.

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

$$
\begin{array}{c|cccc|c}
 & B_1 & B_2 & \dots & B_L & \Sigma \\
\hline
A_1 & H_{11} & H_{12} & \dots & H_{1L} & H_{1\bullet} \\
A_2 & H_{21} & H_{22} & \dots & H_{2L} & H_{2\bullet} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_K & H_{K1} & H_{K2} & \cdots & H_{KL} & H_{K\bullet} \\
\hline
\Sigma & H_{\bullet 1} & H_{\bullet 2} & \dots & H_{\bullet L} & n
\end{array}
$$

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

d.h. $$H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \qquad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \qquad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	H_{11}	H_{12}	H_{13}	$H_{1\bullet}$
Gym	H_{21}	H_{22}	H_{23}	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Rea	ja	Gym	vllt	Real	vllt	Gym	nein	Rea	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Rea	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Rea	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	H_{12}	H_{13}	$H_{1\bullet}$
Gym	H_{21}	H_{22}	H_{23}	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\ldots$	B_L	Σ
A_1	H_{11}	H_{12}	$\ldots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\ldots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\ldots$	$H_{\bullet L}$	n

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \qquad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \qquad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Rea	vllt
Real	ja	Gym	vllt	Rea	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Rea	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Rea	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Rea	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Rea	vllt	Rea	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Rea	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Rea	vllt	Rea	vllt	Rea	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Rea	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	H_{13}	$H_{1\bullet}$
Gym	H_{21}	H_{22}	H_{23}	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Rea	nein	Gym	vllt	Gym	nein	Rea	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Rea	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Rea	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Rea	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Rea	nein	Rea	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Rea	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	H_{21}	H_{22}	H_{23}	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

$$\text{d.h. } H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \qquad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \qquad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	8	H_{22}	H_{23}	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

$$\text{d.h. } H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \qquad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \qquad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	8	24	H_{23}	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gyr	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gyr	nein	Real	ja	Gyr	nein	Gyr	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gyr	nein	Gym	ja	Gyr	nein
Gym	vllt	Gym	vllt	Gyr	nein	Ges	vllt	Gym	ja	Gyr	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gyr	nein	Gyr	nein	Gym	ja	Gyr	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gyr	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gyr	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gyr	nein	Gyr	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gyr	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	8	24	16	$H_{2\bullet}$
Ges	H_{31}	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	8	24	16	$H_{2\bullet}$
Ges	2	H_{32}	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

	B_1	B_2	$\dots$	B_L	Σ
A_1	H_{11}	H_{12}	$\dots$	H_{1L}	$H_{1\bullet}$
A_2	H_{21}	H_{22}	$\dots$	H_{2L}	$H_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_K	H_{K1}	H_{K2}	$\cdots$	H_{KL}	$H_{K\bullet}$
Σ	$H_{\bullet 1}$	$H_{\bullet 2}$	$\dots$	$H_{\bullet L}$	n

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	8	24	16	$H_{2\bullet}$
Ges	2	6	H_{33}	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

$$
\begin{array}{c|cccc|c}
 & B_1 & B_2 & \dots & B_L & \Sigma \\
\hline
A_1 & H_{11} & H_{12} & \dots & H_{1L} & H_{1\bullet} \\
A_2 & H_{21} & H_{22} & \dots & H_{2L} & H_{2\bullet} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_K & H_{K1} & H_{K2} & \cdots & H_{KL} & H_{K\bullet} \\
\hline
\Sigma & H_{\bullet 1} & H_{\bullet 2} & \dots & H_{\bullet L} & n
\end{array}
$$

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	$H_{1\bullet}$
Gym	8	24	16	$H_{2\bullet}$
Ges	2	6	4	$H_{3\bullet}$
$H_{\bullet j}$	$H_{\bullet 1}$	$H_{\bullet 2}$	$H_{\bullet 3}$	n

$\hookrightarrow$ Erstelle **Kontingenztafel** (vgl. DuW)

$$
\begin{array}{c|cccc|c}
 & B_1 & B_2 & \dots & B_L & \Sigma \\
\hline
A_1 & H_{11} & H_{12} & \dots & H_{1L} & H_{1\bullet} \\
A_2 & H_{21} & H_{22} & \dots & H_{2L} & H_{2\bullet} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
A_K & H_{K1} & H_{K2} & \cdots & H_{KL} & H_{K\bullet} \\
\hline
\Sigma & H_{\bullet 1} & H_{\bullet 2} & \dots & H_{\bullet L} & n
\end{array}
$$

d.h. $H_{ij} = \sum_{\ell=1}^{n} 1_{\{X_\ell = A_i\}} 1_{\{Y_\ell = B_j\}}, \quad H_{i\bullet} = \sum_{j=1}^{L} H_{ij}, \quad H_{\bullet j} = \sum_{i=1}^{K} H_{ij}$

Beispiel (vgl. DuW) Sichtung von 84 Bewerbungsmappen:

Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus	Ab	Zus
Gym	vllt	Real	nein	Gym	vllt	Gym	nein	Real	nein	Ges	vllt	Real	vllt
Real	ja	Gym	vllt	Real	vllt	Gym	nein	Real	ja	Gym	nein	Gym	nein
Real	nein	Ges	ja	Real	vllt	Gym	ja	Gym	nein	Gym	ja	Gym	nein
Gym	vllt	Gym	vllt	Gym	nein	Ges	vllt	Gym	ja	Gym	nein	Gym	vllt
Real	vllt	Real	ja	Real	nein	Gym	vllt	Gym	vllt	Real	vllt	Ges	nein
Gym	vllt	Real	nein	Gym	ja	Gym	nein	Gym	nein	Gym	ja	Gym	nein
Real	vllt	Real	vllt	Real	nein	Real	nein	Ges	vllt	Ges	vllt	Ges	nein
Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	ja	Real	ja
Gym	vllt	Gym	vllt	Ges	ja	Gym	ja	Gym	vllt	Real	vllt	Gym	vllt
Gym	vllt	Gym	ja	Ges	nein	Gym	vllt	Real	vllt	Real	vllt	Real	vllt
Gym	vllt	Ges	vllt	Ges	vllt	Gym	vllt	Real	vllt	Real	nein	Ges	nein
Gym	nein	Gym	nein	Gym	vllt	Gym	vllt	Gym	vllt	Gym	nein	Gym	vllt

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

$\hookrightarrow$ Ansatz für eine Teststatistik:

- Unter H_0 gilt: $P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j)$

$\hookrightarrow$ Ansatz für eine Teststatistik:

- □ Unter H_0 gilt: $P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j)$
- □ ML-Schätzungen für die genannten WS sind $\frac{H_{ij}}{n}$, $\frac{H_{i\bullet}}{n}$ und $\frac{H_{\bullet j}}{n}$

$\hookrightarrow$ Ansatz für eine Teststatistik:

- Unter H_0 gilt: $P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j)$
- ML-Schätzungen für die genannten WS sind $\frac{H_{ij}}{n}$, $\frac{H_{i\bullet}}{n}$ und $\frac{H_{\bullet j}}{n}$
- Unter H_0 sollte deshalb für alle Häufigkeiten gelten:

$$\frac{H_{ij}}{n} \approx \frac{H_{i\bullet}}{n} \cdot \frac{H_{\bullet j}}{n}$$

$\hookrightarrow$ Ansatz für eine Teststatistik:

- Unter H_0 gilt: $P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j)$
- ML-Schätzungen für die genannten WS sind $\frac{H_{ij}}{n}$, $\frac{H_{i\bullet}}{n}$ und $\frac{H_{\bullet j}}{n}$
- Unter H_0 sollte deshalb für alle Häufigkeiten gelten:

$$\frac{H_{ij}}{n} \approx \frac{H_{i\bullet}}{n} \cdot \frac{H_{\bullet j}}{n} \qquad \Leftrightarrow \qquad H_{ij} \approx \frac{H_{i\bullet} \cdot H_{\bullet j}}{n} := \tilde{H}_{ij}$$

$\tilde{H}_{ij}$ sind die unter Unabhängigkeit erwarteten Häufigkeiten.

$\hookrightarrow$ Ansatz für eine Teststatistik:

- □ Unter H_0 gilt: $P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j)$
- □ ML-Schätzungen für die genannten WS sind $\frac{H_{ij}}{n}$, $\frac{H_{i\bullet}}{n}$ und $\frac{H_{\bullet j}}{n}$
- □ Unter H_0 sollte deshalb für alle Häufigkeiten gelten:

$$\frac{H_{ij}}{n} \approx \frac{H_{i\bullet}}{n} \cdot \frac{H_{\bullet j}}{n} \qquad \Leftrightarrow \qquad H_{ij} \approx \frac{H_{i\bullet} \cdot H_{\bullet j}}{n} := \tilde{H}_{ij}$$

  $\tilde{H}_{ij}$ sind die unter Unabhängigkeit erwarteten Häufigkeiten.
- □ Abweichungen zwischen H_{ij} und $\tilde{H}_{ij}$ sprechen gegen die Unabhängigkeit.
- □ Eine Teststatistik sollte die Abweichungen $(H_{ij} - \tilde{H}_{ij})^2$ verwenden.

↪ Ansatz für eine Teststatistik:

- ☐ Unter H_0 gilt: $P(X = A_i, Y = B_j) = P(X = A_i)P(Y = B_j)$
- ☐ ML-Schätzungen für die genannten WS sind $\frac{H_{ij}}{n}$, $\frac{H_{i\bullet}}{n}$ und $\frac{H_{\bullet j}}{n}$
- ☐ Unter H_0 sollte deshalb für alle Häufigkeiten gelten:

$$\frac{H_{ij}}{n} \approx \frac{H_{i\bullet}}{n} \cdot \frac{H_{\bullet j}}{n} \qquad \Leftrightarrow \qquad H_{ij} \approx \frac{H_{i\bullet} \cdot H_{\bullet j}}{n} := \tilde{H}_{ij}$$

$\tilde{H}_{ij}$ sind die unter Unabhängigkeit erwarteten Häufigkeiten.

- ☐ Abweichungen zwischen H_{ij} und $\tilde{H}_{ij}$ sprechen gegen die Unabhängigkeit.
- ☐ Eine Teststatistik sollte die Abweichungen $(H_{ij} - \tilde{H}_{ij})^2$ verwenden.

2. Teststatistik V und Verteilung unter H_0

$$V = \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{(H_{ij} - \tilde{H}_{ij})^2}{\tilde{H}_{ij}} = n \cdot \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \frac{H_{ij}^2}{H_{i\bullet} \cdot H_{\bullet j}} - 1 \right) \overset{approx.}{\sim} \chi^2_{(K-1)\cdot(L-1)}$$

↪ Faustregel: Die Approximation ist ausreichend genau, wenn $\min(\tilde{H}_{ij}) \geq 5$ (d.h. wenn jede Zelle der Kontingenztafel mindestens 5 Einträge enthält)

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

,

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

, erwartet:

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	$\frac{24\cdot 14}{84}$	$\frac{24\cdot 42}{84}$	$\frac{24\cdot 28}{84}$	24
Gym	$\frac{48\cdot 14}{84}$	$\frac{48\cdot 42}{84}$	$\frac{48\cdot 28}{84}$	48
Ges	$\frac{12\cdot 14}{84}$	$\frac{12\cdot 42}{84}$	$\frac{12\cdot 28}{84}$	12
$H_{\bullet j}$	14	42	28	

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

, erwartet:

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	

Approximation hier nicht ausreichend.

Ab	Zusage ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

, erwartet:

Ab	Zusage ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	

Approximation hier nicht ausreichend.

$$v = \sum_{i=1}^{K}\sum_{j=1}^{L}\frac{(H_{ij}-\tilde{H}_{ij})^2}{\tilde{H}_{ij}} = \frac{(4-4)^2}{4} + \frac{(12-12)^2}{12} + \cdots + \frac{(4-4)^2}{4} = 0$$

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

, erwartet:

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	

Approximation hier nicht ausreichend.

$$v = \sum_{i=1}^{K}\sum_{j=1}^{L}\frac{(H_{ij}-\tilde{H}_{ij})^2}{\tilde{H}_{ij}} = \frac{(4-4)^2}{4}+\frac{(12-12)^2}{12}+\cdots+\frac{(4-4)^2}{4} = 0$$

Rechenweg 2:

$$\begin{aligned} v &= n\cdot\left(\sum_{i=1}^{K}\sum_{j=1}^{L}\frac{H_{ij}^2}{H_{i\bullet}\cdot H_{\bullet j}}-1\right) \\ &= 84\cdot\left(\frac{4^2}{24\cdot 14}+\frac{12^2}{24\cdot 42}+\ldots+\frac{4^2}{12\cdot 28}-1\right) = 84\cdot(1-1) = 0 \end{aligned}$$

3. Entscheidungsregel (und p-Wert)

Entscheidungsregel:

$$H_0 \text{ wird abgelehnt} \quad \Longleftrightarrow \quad v > \chi^2_{(1-\alpha);(K-1)\cdot(L-1)}$$

p-Wert:

$$p^* = 1 - F_{\chi^2_{(K-1)\cdot(L-1)}}(v)$$

3. Entscheidungsregel (und *p*-Wert)

Entscheidungsregel:

$$H_0 \text{ wird abgelehnt} \quad \Longleftrightarrow \quad v > \chi^2_{(1-\alpha);(K-1)\cdot(L-1)}$$

p-Wert:

$$p^* = 1 - F_{\chi^2_{(K-1)\cdot(L-1)}}(v)$$

$\hookrightarrow$ In R: Quantil und *p*-Wert mit `qchisq(alpha, df = ...)` und `pchisq(v, df = ...)`

3. Entscheidungsregel (und *p*-Wert)

Entscheidungsregel:

$$H_0 \text{ wird abgelehnt} \quad \Longleftrightarrow \quad v > \chi^2_{(1-\alpha);(K-1)\cdot(L-1)}$$

p-Wert:

$$p^* = 1 - F_{\chi^2_{(K-1)\cdot(L-1)}}(v)$$

↪ In R: Quantil und *p*-Wert mit `qchisq(alpha, df = ...)` und `pchisq(v, df = ...)`

- ☐ geg.: zwei Merkmale X und Y, bestehend aus $K = 3$ und $L = 5$ Klassen
- ☐ hierzu ergibt sich der kritische Wert zum Signifikanzniveau $\alpha = 5\%$ durch: `qchisq(1 - 0.05, df = (3 - 1)*(5 - 1))` (Ergebnis: `15.50731`)
- ☐ entsprechend ergibt sich der *p*-Wert einer realisierten Teststatistik $v = 13$ zu `1 - pchisq(13, df = (3 - 1)*(5 - 1))` (Ergebnis: `0.1118496`)

↪ Durchführung in R: `chisq.test(...)` (auch für Anpassungstest, s.u.)

↪ Test hat i.d.R. geringe Güte.

Beispiel Bewerbungsmappe, $\alpha = 0.05$

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

, erwartet:

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	28	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	

$$\begin{aligned} v &= 0 \\ \chi^2_{(1-\alpha);(K-1)\cdot(L-1)} &= \chi^2_{0.95;2\cdot 2} = 9.48 > v \end{aligned}$$

Also kann H_0 nicht abgelehnt werden.

Extrembeispiel: Merkmale sind empirisch unabhängig (s. DuW)

Modifiziertes Beispiel „Bewerbungsmappe“

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	8	24	16	48
Ges	2	6	4	12
$H_{\bullet j}$	14	42	28	84

Modifiziertes Beispiel „Bewerbungsmappe“

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	20	24	4	48
Ges	2	6	4	12
$H_{\bullet j}$	26	42	16	84

$$
\begin{aligned}
v &= n \cdot \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \frac{H_{ij}^2}{H_{i\bullet} \cdot H_{\bullet j}} - 1 \right) \\
&= 84 \cdot \left(\frac{4^2}{24 \cdot 26} + \frac{12^2}{24 \cdot 42} + \frac{8^2}{24 \cdot 16} \cdots + \frac{4^2}{12 \cdot 16} - 1 \right) = 84 \cdot (0.13) = 10.9
\end{aligned}
$$

Modifiziertes Beispiel „Bewerbungsmappe“

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	20	24	4	48
Ges	2	6	4	12
$H_{\bullet j}$	26	42	16	84

$$
\begin{aligned}
v &= n \cdot \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \frac{H_{ij}^2}{H_{i\bullet} \cdot H_{\bullet j}} - 1 \right) \\
&= 84 \cdot \left(\frac{4^2}{24 \cdot 26} + \frac{12^2}{24 \cdot 42} + \frac{8^2}{24 \cdot 16} \cdots + \frac{4^2}{12 \cdot 16} - 1 \right) = 84 \cdot (0.13) = 10.9
\end{aligned}
$$

$\hookrightarrow \alpha = 0.05$: $\chi^2_{(1-\alpha);(K-1)\cdot(L-1)} = \chi^2_{0.95;\,4} = 9.48 < v \quad \Rightarrow H_0$ wird abgelehnt.

Modifiziertes Beispiel „Bewerbungsmappe“

	Zusage			
Ab	ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	20	24	4	48
Ges	2	6	4	12
$H_{\bullet j}$	26	42	16	84

$$
\begin{aligned}
v &= n \cdot \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \frac{H_{ij}^2}{H_{i\bullet} \cdot H_{\bullet j}} - 1 \right) \\
&= 84 \cdot \left(\frac{4^2}{24 \cdot 26} + \frac{12^2}{24 \cdot 42} + \frac{8^2}{24 \cdot 16} \cdots + \frac{4^2}{12 \cdot 16} - 1 \right) = 84 \cdot (0.13) = 10.9
\end{aligned}
$$

$\hookrightarrow \alpha = 0.05$: $\chi^2_{(1-\alpha);(K-1)\cdot(L-1)} = \chi^2_{0.95;\,4} = 9.48 < v$ $\Rightarrow H_0$ wird abgelehnt.

$\hookrightarrow \alpha = 0.01$: $\chi^2_{0.99\ 4} = 13.48 > v$ $\Rightarrow H_0$ wird nicht abgelehnt.

Modifiziertes Beispiel „Bewerbungsmappe“

Ab	Zusage: ja	vllt	nein	$H_{i\bullet}$
Real	4	12	8	24
Gym	20	24	4	48
Ges	2	6	4	12
$H_{\bullet j}$	26	42	16	84

$$
\begin{aligned}
v &= n \cdot \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \frac{H_{ij}^2}{H_{i\bullet} \cdot H_{\bullet j}} - 1 \right) \\
&= 84 \cdot \left(\frac{4^2}{24 \cdot 26} + \frac{12^2}{24 \cdot 42} + \frac{8^2}{24 \cdot 16} \cdots + \frac{4^2}{12 \cdot 16} - 1 \right) = 84 \cdot (0.13) = 10.9
\end{aligned}
$$

$\hookrightarrow$ $\alpha = 0.05$: $\chi^2_{(1-\alpha);(K-1)\cdot(L-1)} = \chi^2_{0.95;\,4} = 9.48 < v$ $\Rightarrow$ H_0 wird abgelehnt.

$\hookrightarrow$ $\alpha = 0.01$: $\chi^2_{0.99\;4} = 13.48 > v$ $\Rightarrow$ H_0 wird nicht abgelehnt.

$\hookrightarrow$ p-Wert: $p = 1 - F_{\chi^2_4}(10.9) = 0.027$ mit R: `1 - pchisq(10.9, df = 4)`

Für den Zusammenhang zwischen Geschlecht und Note bei einer Mathematikklausur ergibt sich folgende Häufigkeitstabelle:

	Note				
Geschlecht	1	2	3	4	5
m	8	55	87	40	122
w	4	13	42	34	66

Führen Sie den χ^2-Test auf Unabhängigkeit per Hand aus (Signifikanzniveau $\alpha = 0.01$). Verwenden Sie die Quantilstabelle der nachfolgenden Folie.

Quantile der χ_n^2-Verteilung zum Signifikanzniveau α:

										α										
n	0.005	0.01	0.02	0.025	0.03	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.97	0.975	0.98	0.99	0.995
1	0.000	0.000	0.001	0.001	0.001	0.004	0.016	0.064	0.148	0.275	0.455	0.708	1.074	1.642	2.706	3.841	4.709	5.024	5.412	6.635	7.879
2	0.010	0.020	0.040	0.051	0.061	0.103	0.211	0.446	0.713	1.022	1.386	1.833	2.408	3.219	4.605	5.991	7.013	7.378	7.824	9.210	10.597
3	0.072	0.115	0.185	0.216	0.245	0.352	0.584	1.005	1.424	1.869	2.366	2.946	3.665	4.642	6.251	7.815	8.947	9.348	9.837	11.345	12.838
4	0.207	0.297	0.429	0.484	0.535	0.711	1.064	1.649	2.195	2.753	3.357	4.045	4.878	5.989	7.779	9.488	10.712	11.143	11.668	13.277	14.860
5	0.412	0.554	0.752	0.831	0.903	1.145	1.610	2.343	3.000	3.655	4.351	5.132	6.064	7.289	9.236	11.070	12.375	12.833	13.388	15.086	16.750
6	0.676	0.872	1.134	1.237	1.330	1.635	2.204	3.070	3.828	4.570	5.348	6.211	7.231	8.558	10.645	12.592	13.968	14.449	15.033	16.812	18.548
7	0.989	1.239	1.564	1.690	1.802	2.167	2.833	3.822	4.671	5.493	6.346	7.283	8.383	9.803	12.017	14.067	15.509	16.013	16.622	18.475	20.278
8	1.344	1.646	2.032	2.180	2.310	2.733	3.490	4.594	5.527	6.423	7.344	8.351	9.524	11.030	13.362	15.507	17.010	17.535	18.168	20.090	21.955
9	1.735	2.088	2.532	2.700	2.848	3.325	4.168	5.380	6.393	7.357	8.343	9.414	10.656	12.242	14.684	16.919	18.480	19.023	19.679	21.666	23.589
10	2.156	2.558	3.059	3.247	3.412	3.940	4.865	6.179	7.267	8.295	9.342	10.473	11.781	13.442	15.987	18.307	19.922	20.483	21.161	23.209	25.188
11	2.603	3.053	3.609	3.816	3.997	4.575	5.578	6.989	8.148	9.237	10.341	11.530	12.899	14.631	17.275	19.675	21.342	21.920	22.618	24.725	26.757
12	3.074	3.571	4.178	4.404	4.601	5.226	6.304	7.807	9.034	10.182	11.340	12.584	14.011	15.812	18.549	21.026	22.742	23.337	24.054	26.217	28.300
13	3.565	4.107	4.765	5.009	5.221	5.892	7.042	8.634	9.926	11.129	12.340	13.636	15.119	16.985	19.812	22.362	24.125	24.736	25.472	27.688	29.819
14	4.075	4.660	5.368	5.629	5.856	6.571	7.790	9.467	10.821	12.078	13.339	14.685	16.222	18.151	21.064	23.685	25.493	26.119	26.873	29.141	31.319
15	4.601	5.229	5.985	6.262	6.503	7.261	8.547	10.307	11.721	13.030	14.339	15.733	17.322	19.311	22.307	24.996	26.848	27.488	28.259	30.578	32.801
16	5.142	5.812	6.614	6.908	7.163	7.962	9.312	11.152	12.624	13.983	15.338	16.780	18.418	20.465	23.542	26.296	28.191	28.845	29.633	32.000	34.267
17	5.697	6.408	7.255	7.564	7.832	8.672	10.085	12.002	13.531	14.937	16.338	17.824	19.511	21.615	24.769	27.587	29.523	30.191	30.995	33.409	35.718
18	6.265	7.015	7.906	8.231	8.512	9.390	10.865	12.857	14.440	15.893	17.338	18.868	20.601	22.760	25.989	28.869	30.845	31.526	32.346	34.805	37.156
19	6.844	7.633	8.567	8.907	9.200	10.117	11.651	13.716	15.352	16.850	18.338	19.910	21.689	23.900	27.204	30.144	32.158	32.852	33.687	36.191	38.582
20	7.434	8.260	9.237	9.591	9.897	10.851	12.443	14.578	16.266	17.809	19.337	20.951	22.775	25.038	28.412	31.410	33.462	34.170	35.020	37.566	39.997

8.2 χ^2-Anpassungstest

$\hookrightarrow$ Prüft die Hypothese:
Ist F_0 Stichprobenverteilung einer u.i.v.-Stichprobe $X_1, \dots, X_n$?

$\hookrightarrow$ z.B. Tests von Standardzufallszahlen.

8.2 χ^2-Anpassungstest

- ↪ Prüft die Hypothese:
  Ist F_0 Stichprobenverteilung einer u.i.v.-Stichprobe $X_1, \ldots, X_n$?
- ↪ z.B. Tests von Standardzufallszahlen.
- ↪ Test prüft nicht die Ausgangshypothese, sondern eine Vergröberung durch…
- ↪ …Klassierung des Stichprobenraums, z.B. Intervalle $A_1, \ldots, A_K$,
  Konkretisierung anhand der unter F_0 gegebenen Klassen-W'keiten $p_1, \ldots, p_K$.

8.2 χ^2-Anpassungstest

- $\hookrightarrow$ Prüft die Hypothese:
  Ist F_0 Stichprobenverteilung einer u.i.v.-Stichprobe $X_1, \ldots, X_n$?
- $\hookrightarrow$ z.B. Tests von Standardzufallszahlen.
- $\hookrightarrow$ Test prüft nicht die Ausgangshypothese, sondern eine Vergröberung durch...
- $\hookrightarrow$ ... Klassierung des Stichprobenraums, z.B. Intervalle $A_1, \ldots, A_K$,
  Konkretisierung anhand der unter F_0 gegebenen Klassen-W'keiten $p_1, \ldots, p_K$.

1. Hypothese:

$$H_0 : P(X = j) = p_j, \quad \forall j \in \{1, \ldots, K\}$$

8.2 χ^2-Anpassungstest

- ↪ Prüft die Hypothese:
  Ist F_0 Stichprobenverteilung einer u.i.v.-Stichprobe $X_1, \ldots, X_n$?
- ↪ z.B. Tests von Standardzufallszahlen.
- ↪ Test prüft nicht die Ausgangshypothese, sondern eine Vergröberung durch...
- ↪ ...Klassierung des Stichprobenraums, z.B. Intervalle $A_1, \ldots, A_K$,
  Konkretisierung anhand der unter F_0 gegebenen Klassen-W'keiten $p_1, \ldots, p_K$.

1. Hypothese:

$$H_0 : P(X = j) = p_j, \quad \forall j \in \{1, \ldots, K\}$$

- ↪ Vergleich erwarteter Häufigkeiten und beobachteter Häufigkeiten $H(A_j)$

2. Teststatistik und Nullverteilung

$$V = \sum_{j=1}^{K} \frac{(H(A_j) - n \cdot p_j)^2}{n \cdot p_j} \overset{approx}{\sim} \chi^2_{K-1} \text{ für } \min(n \cdot p_1, \ldots, n \cdot p_K) \geq 5$$

χ^2-Anpassungstest

3. Entscheidungsregel (und *p*-Wert)

Entscheidungsregel (Signifikanzniveau α):

$$H_0 \text{ wird abgelehnt} \qquad \Longleftrightarrow \qquad v > \chi^2_{(1-\alpha);\,K-1}$$

p-Wert:

$$p^* = 1 - F_{\chi^2_{K-1}}(v)$$

Prüfung d. Kassendaten einer Eisdiele durch Steuerfahndung ($n = 180$)

Für die führende Ziffer[a] wird bei einwandfreier Kassenführung die **Benford-Verteilung** angenommen: $p_j = \log_{10}(j+1) - \log_{10}(j),\ j = 1, 2, \ldots, 9.$

Die Kassendaten einer Eisdiele sollen darauf geprüft werden, ob sie einer Benford-Verteilung genügen:

	j	1	2	3	4	5	6	7	8	9	Σ
Daten	$H(A_j)$	45	36	26	20	17	11	13	6	6	180
theo-	p_j	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046	1
retisch	$n \cdot p_j$	54.19	31.70	22.49	17.44	14.25	12.05	10.44	9.21	8.24	180
Test-	$(H(A_j) - n \cdot p_j)^2$	84.37	18.52	12.33	6.53	7.55	1.10	6.56	10.29	5.00	
statistik	$\frac{(H(A_j)-n\cdot p_j)^2}{n\cdot p_j}$	1.56	0.58	0.55	0.37	0.53	0.09	0.63	1.12	0.61	$v = 6.038$

Prüfung d. Kassendaten einer Eisdiele durch Steuerfahndung ($n = 180$)

Für die führende Ziffer[a] wird bei einwandfreier Kassenführung die **Benford-Verteilung** angenommen: $p_j = \log_{10}(j+1) - \log_{10}(j),\ j = 1, 2, \ldots, 9.$

Die Kassendaten einer Eisdiele sollen darauf geprüft werden, ob sie einer Benford-Verteilung genügen:

	j	1	2	3	4	5	6	7	8	9	Σ
Daten	$H(A_j)$	45	36	26	20	17	11	13	6	6	180
theo-	p_j	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046	1
retisch	$n \cdot p_j$	54.19	31.70	22.49	17.44	14.25	12.05	10.44	9.21	8.24	180
Test-	$(H(A_j) - n \cdot p_j)^2$	84.37	18.52	12.33	6.53	7.55	1.10	6.56	10.29	5.00	
statistik	$\frac{(H(A_j)-n \cdot p_j)^2}{n \cdot p_j}$	1.56	0.58	0.55	0.37	0.53	0.09	0.63	1.12	0.61	$v = 6.038$

$\hookrightarrow$ Zum Signifikanzniveau $\alpha = 0.05$ gilt $v = 6.038 \leq \chi^2_{0.95;\,8} = 15.50731$,

$\hookrightarrow$ d.h. H_0: „einwandfreie Kassenführung“ kann nicht abgelehnt werden.

$\hookrightarrow$ Der p-Wert beträgt $p^* = 1 - F_{\chi^2_8}(6.038) \approx 0.643 > \alpha$

[a] bspw. hat 13,80 € die führende Ziffer $x = 1$ und -0,00057 die führende Ziffer $x = 5$

Übung: Schreiben Sie eine R-Funktion `leading.digit(x)`, welche aus einer beliebigen Zahl $x \neq 0$ die führende Ziffer der Mantisse berechnet.

Übung: Der Pseudozufallszahlen-Generator des TI-59 war der lineare Kongruenzgenerator $y_{i+1} \equiv (24298 \cdot y_i + 99991) \mod 199017$. Schreiben Sie ein R-Programm, welches mit Startwert $y_0 = 0$ die Pseudozufallszahlen $u_i = \frac{y_i}{199017}$, $i = 1, \ldots, 200$ erzeugt. Prüfen Sie, ob die Annahme einer Gleichverteilung nicht abgelehnt werden kann.

8.3 Kolomogorov-Smirnov Anpassungstest

- $\hookrightarrow$ Anwendung: Tests von Standardzufallszahlen („Diehard-Testbatterie“)
- $\hookrightarrow$ bei kleiner Stichprobe Alternative zum χ^2-Anpassungstest (höhere Güte)
- $\hookrightarrow$ keine Kategorisierung der Daten, dafür nur Vergleich mit stetiger Verteilung(sfunktion)

Gegeben eine u.i.v.-Stichprobe $X_1, \ldots, X_n$ mit (stetiger) Verteilungsfunktion F und angeordneter Stichprobe $X_{(1)} < \cdots < X_{(n)}$ (d.h. keine Bindungen)

1. Hypothese

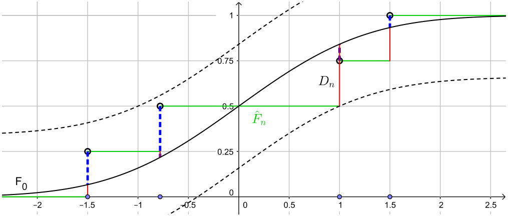
H_0: $F(x) = F_0(x)\ \forall x$ mit vorgegebener VF F_0.

2.1 Teststatistik

- basiert auf der empirischen VF $\hat{F}_n(x)$ der Realisationen $x_1, \ldots, x_n$
  $\hat{F}_n(x)$ ist der Anteil der Beobachtungen, die höchstens den Wert x aufweisen
- die Teststatistik misst den größten vertikalen Abstand zwischen hypothetischer und empirischer Verteilungsfunktion
  - $D_n = \sup\limits_{x \in \mathbb{R}} |F_0(x) - \hat{F}_n(x)| = \max\limits_{i \in \{0,1\},\, j \in \{1,\ldots,n\}} \left\{ \left| F_0(X_{(j)}) - \frac{j-i}{n} \right| \right\}$
  - $V = \sqrt{n} \cdot D_n$

Illustration $F_0(x) = \Phi(x)$

j	x_j	$F_0(x_j)$	$\lvert F_0(x_j) - \frac{j-1}{4} \rvert$	$\lvert F_0(x_j) - \frac{j}{4} \rvert$
1	-1.50	0.07	0.07	0.18
2	-0.78	0.22	0.03	0.28
3	1.00	0.84	$D_4 = 0.34$	0.09
4	1.50	0.93	0.18	0.07



2.2 Verteilung F_V von V unter $H_0 : F = F_0$

$\hookrightarrow$ für n klein: numerische Berechnung von F_V nach Durbin

$\hookrightarrow$ Approximativ ($n > 40$): $$P(V \leq t) = 1 + 2 \cdot \sum_{j=1}^{\infty} (-1)^j \cdot \exp(-2 \cdot j^2 \cdot t^2)$$

$\hookrightarrow$ Quantile $d_{n,\alpha}$ durch numerische Inversion der VF.

Tabelle der Quantile (bei fehlendem Eintrag nächstgrößerer Stichprobenumfang):

α \ n	5	6	7	8	9	10	15	20	25	30	35	40	> 40
0.90	1.14	1.15	1.15	1.16	1.16	1.17	1.18	1.18	1.19	1.19	1.19	1.20	1.22
0.95	1.26	1.27	1.28	1.28	1.29	1.29	1.31	1.32	1.32	1.32	1.33	1.33	1.36
0.99	1.49	1.51	1.52	1.53	1.54	1.55	1.57	1.58	1.58	1.59	1.59	1.59	1.63

(nach Durbin)

3. Entscheidungsregel

Entscheidungsregel (Signifikanzniveau α):

$$H_0 \text{ wird abgelehnt} \quad \Longleftrightarrow \quad v \geq d_{n,1-\alpha}$$

in R: `ks.test(x, y = "p<distr>",...)`

dabei ist `p<distr>` die zu testende VF

```
ks.test(x = rnorm(50), y = "pgamma", shape = 3, rate = 2)
```

testet, ob ein Vektor bestehend aus 50 standardnormalverteilten Beobachtungen einer Gamma-Verteilung mit Parametern `shape = 3` und `rate = 2` folgt

Bsp: Test auf Gleichverteilung ($\alpha = 0.05$, $n = 10$)

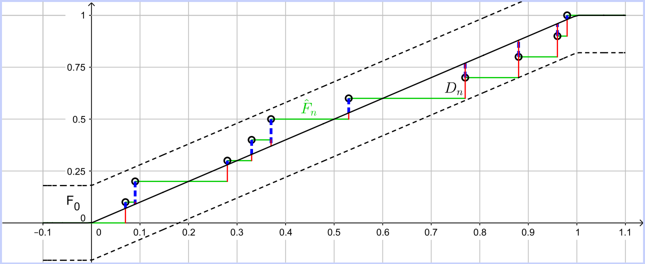
$H_0 : F_0 = Re(0,1)$ vs. $H_1 : F_0 \neq Re(0,1)$ (unter H_0 gilt $F_0(x) = x$)

j	1	2	3	4	5	6	7	8	9	10
$\left\lvert F_0(x_j) - \frac{j-1}{n} \right\rvert$	0.07	0.01	0.08	0.03	0.03	0.03	0.17	0.18	0.16	0.08
$\frac{j-1}{n}$	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
x_j	0.07	0.09	0.28	0.33	0.37	0.53	0.77	0.88	0.96	0.98
$\frac{j}{n}$	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
$\left\lvert F_0(x_j) - \frac{j}{n} \right\rvert$	0.03	0.11	0.02	0.07	0.13	0.07	0.07	0.08	0.06	0.02

in R:

```
x = c(0.07, 0.09, 0.28, 0.33, 0.37 0.53, 0.77, 0.88, 0.96, 0.98)
ks.test(x, "punif", min = 0, max = 1, alternative = "two.sided")
```

j	1	2	3	4	5	6	7	8	9	10
$\left\lvert F_0(x_j) - \frac{j-1}{n} \right\rvert$	0.07	0.01	0.08	0.03	0.03	0.03	0.17	0.18	0.16	0.08
$\frac{j-1}{n}$	0.00	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
x_j	0.07	0.09	0.28	0.33	0.37	0.53	0.77	0.88	0.96	0.98
$\frac{j}{n}$	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
$\left\lvert F_0(x_j) - \frac{j}{n} \right\rvert$	0.03	0.11	0.02	0.07	0.13	0.07	0.07	0.08	0.06	0.02



$\sqrt{10} \cdot D_n = \sqrt{10} \cdot 0.18 = 0.5691 < 1.29 = d_{10,\,0.95}$

$\Rightarrow$ Nichtablehnung von H_0

Gegeben sind die folgenden 5 Beobachtungen: -1.0, -0.2, 0.45, 1.05, 1.69.
Testen Sie, ob die Daten einer Normalverteilung mit $\mu = 0.5$ und $\sigma = 1$ folgen!

Nutzen Sie die folgende Tabelle zur Bestimmung von $F_0(x)$
(wählen Sie μ geeignet; $\sigma = 1$ ist bereits vorgegeben):

	μ								
x	-1.000	-0.950	-0.500	-0.050	0.000	0.050	0.500	0.950	1.000
-1.00	0.500	0.480	0.309	0.171	0.159	0.147	0.067	0.026	0.023
-0.20	0.788	0.773	0.618	0.440	0.421	0.401	0.242	0.125	0.115
0.45	0.926	0.919	0.829	0.691	0.674	0.655	0.480	0.309	0.291
1.05	0.980	0.977	0.939	0.864	0.853	0.841	0.709	0.540	0.520
1.69	0.996	0.996	0.986	0.959	0.954	0.949	0.883	0.770	0.755

Gegeben sind die folgenden 5 Beobachtungen: -1.0, -0.2, 0.45, 1.05, 1.69.

	μ								
x	-1.000	-0.950	-0.500	-0.050	0.000	0.050	0.500	0.950	1.000
-1.00	0.500	0.480	0.309	0.171	0.159	0.147	0.067	0.026	0.023
-0.20	0.788	0.773	0.618	0.440	0.421	0.401	0.242	0.125	0.115
0.45	0.926	0.919	0.829	0.691	0.674	0.655	0.480	0.309	0.291
1.05	0.980	0.977	0.939	0.864	0.853	0.841	0.709	0.540	0.520
1.69	0.996	0.996	0.986	0.959	0.954	0.949	0.883	0.770	0.755

Bestimmen Sie zunächst die Realisation der zugehörigen Teststatistik!

Gegeben sind die folgenden 5 Beobachtungen: -1.0, -0.2, 0.45, 1.05, 1.69.

				μ					
x	-1.000	-0.950	-0.500	-0.050	0.000	0.050	0.500	0.950	1.000
-1.00	0.500	0.480	0.309	0.171	0.159	0.147	0.067	0.026	0.023
-0.20	0.788	0.773	0.618	0.440	0.421	0.401	0.242	0.125	0.115
0.45	0.926	0.919	0.829	0.691	0.674	0.655	0.480	0.309	0.291
1.05	0.980	0.977	0.939	0.864	0.853	0.841	0.709	0.540	0.520
1.69	0.996	0.996	0.986	0.959	0.954	0.949	0.883	0.770	0.755

Entscheiden Sie mit Hilfe der kritischen Werte nach Durbin, ob H_0 zu einem Signifikanzniveau von $\alpha = 5\%$ verworfen werden kann!

Spezialfall für den Test auf Normalverteilung:

$\hookrightarrow$ für fixe Parameter μ_0, σ_0^2: $F_0(x) = \Phi\left(\frac{x-\mu_0}{\sigma_0}\right)$

$\hookrightarrow$ für nicht spezifierte Parameter ist der KS-Test nicht direkt anwendbar (hält das Signifikanzniveau strikt ein, lehnt H_0 zu selten ab: „zu konservativ")

$\hookrightarrow$ stattdessen Vorgehensweise mit modifizierten kritischen Werten:

- die Parameter μ, σ werden aus der Stichprobe geschätzt; die Verteilung hängt nicht mehr von den unbekannten Parametern ab, unterscheidet sich jedoch von der ursprünglichen Verteilung von D_n.

$$\sqrt{n} \cdot D_n^N = \sqrt{n} \cdot \sup_x \left| \hat{F}_n(x) - \Phi\left(\frac{x-\bar{x}}{s}\right) \right|$$

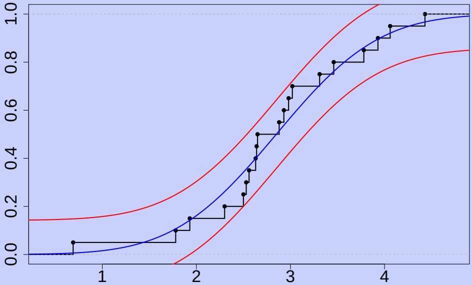
- H_0 wird abgelehnt, falls $\sqrt{n} \cdot D_n^N \geq d^N_{n,(1-\alpha)}$
- kritische Werte $d^N_{n,\alpha}$ aus Simulationsstudien (Lilliefors):

α \ n	5	8	10	20	30	> 30
0.90	0.72	0.74	0.76	0.79	0.80	0.81
0.95	0.76	0.81	0.82	0.85	0.88	0.89
0.99	0.91	0.94	0.94	1.03	1.03	1.04

Bsp: Test auf Normalverteilung

$x = (0.69, 1.78, 1.93, 2.30, 2.50, 2.53, 2.56, 2.63, 2.64, 2.65, 2.88, 2.93, 2.98, 3.02, 3.31, 3.46, 3.78, 3.93, 4.06, 4.43)$

$$H_0 : F_0 = \mathcal{N}(\bar{x}, \sigma_x^2) \quad \text{vs.} \quad H_1 : F_0 \neq \mathcal{N}(\bar{x}, \sigma_x^2)$$



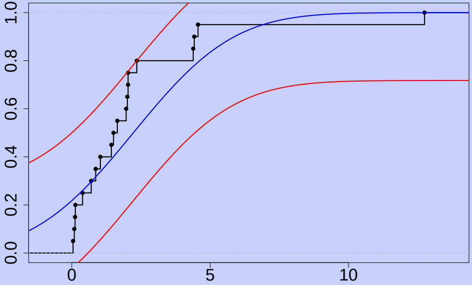
$\sqrt{20} \cdot D_n^N = \sqrt{20} \cdot 0.142 = 0.636 < 0.85 = d_{20,0.95}^N$

$\Rightarrow H_0$ kann nicht abgelehnt werden.

Bsp: Test auf Normalverteilung (II)

$x = (0.045, 0.089, 0.117, 0.127, 0.387, 0.700, 0.862, 1.031, 1.428, 1.506, 1.644, 1.963, 2.007, 2.032, 2.033, 2.347, 4.386, 4.423, 4.560, 12.743)$

$$H_0 : F_0 = \mathcal{N}(\bar{x}, \sigma_x^2) \quad \text{vs.} \quad H_1 : F_0 \neq \mathcal{N}(\bar{x}, \sigma_x^2)$$



$$\sqrt{20} \cdot D_n^N = \sqrt{20} \cdot 0.2824 = 1.263 > d_{20,0.95}^N = 0.85$$

$\Rightarrow H_0$ wird abgelehnt.