INHALT

4.3 Kontingenz => Chi^2-Unabhängigkeitstest (= Chi^2-Kontingenztest)
wir benutzen $h_i = n_i / n$ als Näherung für die Wahrscheinlichkeiten

ab 4.6
Die Ansätze für die Maßzahlen "Mittelwert" und "Varianz" werden schrittweise weiterentwickelt:
wir "übersetzen" h_i in f_i und dann "übersetzen" wir diskret → stetig

Mittelwert $\bar{x}$ und Erwartungswert μ

für einzelne Werte x_i in einer Datenreihe: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

für Werte in einer Häufigkeitstabelle, Häufigkeiten h_i $\sum_{i=1}^{k} x_i \cdot h_i$

für Wahrscheinlichkeiten f_i diskreter Ereignisse x_i $\mu = \sum_{i=1}^{k} x_i \cdot f_i$

für Wahrscheinlichkeiten f_i stetiger "Ereignisse" x $\mu = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$

Varianz s und erwartete Varianz σ

für einzelne Werte x_i ❶ $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ ❷ $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \cdot n_i - \bar{x}^2$

für Werte in einer Häufigkeitstabelle*) ❶ $\sigma^2 = \sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot h_i$ ❷ $\sigma^2 = \sum_{i=1}^{k} x_i^2 \cdot h_i - \bar{x}^2$

für Wahrscheinlichkeiten f_i diskreter Ereignisse x_i

❶ $\sigma^2 = \sum_{i=1}^{k}(x_i - \mu)^2 \cdot f_i$ ❷ $\sigma^2 = \sum_{i=1}^{k} x_i^2 \cdot f_i - \mu^2$

für Wahrscheinlichkeiten f_i stetiger "Ereignisse" x

❶ $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\,dx$ ❷ $\sigma^2 = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu^2$

ab 4.10
Bei stetigen Zufallsvariablen bestimmen wir die Wahrscheinlichkeiten, Erwartungswert
und Streuung mit Hilfe der Integralrechnung

Drei Grundsätze der Statistik:

1. Die relative Häufigkeit h_i konvergiert statistisch gegen die Wahrscheinlichkeit f_i.
   $$\lim_{n\to\infty} W\left(\lim_{n\to\infty}(h_n - p) = 0\right) = 1$$
2. Die Zufallsvariable X streut $\sqrt{n}$ mal stärker als das Stichprobenmittel $\bar{X}$.
   $$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$
3. Zentraler Grenzwertsatz: n Zufallsvariable überlagern sich: nX
   Die Verteilung von nX nähert sich mit zunehmendem n der Normalverteilung.

4.1 DEFINITIONEN

Drei Wahrscheinlichkeitsbegriffe

a) Wahrscheinlichkeit als Grenzwert der rel. Häufigkeit des Ereignisses A: $W(A) = \lim_{n\to\infty} h_n(A)$

[VON MISES, Richard, Berlin 1919]

b) Klassische Definition (LAPLACE): $W(A) = \dfrac{\text{Anzahl der für A günstigen Fälle}}{\text{Anzahl aller gleichmöglichen Fälle}}$

□ A sei das Ereignis, mit einem Würfel eine ungerade Zahl zu werfen: W(A) = 3/6 = 0,5 =50%. Die Elementarereignisse $\{1;2;3;4;5;6\} \in \Omega$ treten gleichwahrscheinlich ein, da physikalisch keines dieser Ereignisse bevorzugt ist. [LAPLACE, Pierre-Simon, Paris, 1773]

c) Axiomatisch wird die Wahrscheinlichkeit des Ereignisses A durch drei Grundsätze definiert:

(1) $0 \leq W(A) \leq 1$ oder $0\,\% \leq W(A) \leq 100\%$

(2) $W(\Omega) = 1$ sicheres Ereignis, wenn Ω alle Elementarereignisse enthält.

(3) $W(A \cup B) = W(A)+W(B)$ Additionseigenschaft von Wahrscheinlichkeiten

Aus diesen drei Grundsätzen lassen sich die Sätze der Wahrscheinlichkeits-Algebra herleiten.

[KOLMOGOROV, Andrei, N., 1933, Moskau]

4.2 WAHRSCHEINLICHKEITSRECHNUNG

1. Mit $p = W(A)$ bezeichnet man die Wahrscheinlichkeit des Treffer-Ereignisses A.
   $q = 1 - p$ ist dann die Wahrscheinlichkeit für das Nichteintreten des Treffers A. $q + p = 1$.
   □ Treffer-Ereignis beim Wurf eines Reißnagels sei die Spitzenlage ⋌. Nichttreffer: ⊥
   $p = 0{,}393 \quad q = 1 - 0{,}393 = 0{,}607$

2. W(A ∪B) ist die Wahrscheinlichkeit, dass das Ereignis A oder B eintritt, gleichgültig welches.
   Wenn es gleichgültig ist, ob A oder B, wächst die Wahrscheinlichkeit: Addition

   W(A∩B) ist die Wahrscheinlichkeit, dass sowohl A als auch B eintritt (saa);
   dadurch verringert sich die Wahrscheinlichkeit des Ereignisses: Multiplikation.
   □ 2 Würfe mit je einem Würfel: A: beim 1.Wurf eine [6] werfen, B: beim 2.Wurf eine [6]

   Spezieller Additionssatz:
   $W(A \cup B) = W(A) + W(B)$
   $W(A \cup B) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} = 33{,}3\%.$
   (Auf das Wesentliche vereinfacht)

   Spezieller Multiplikationssatz:
   $W(A \cap B) = W(A) \cdot W(B)$
   $W(A \cap B) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = 2{,}8\%$

   Diese beiden Sätze gelten nur, wenn A und B unabhängig voneinander eintreten.

3. Allgemeiner Additionssatz: $W(A \cup B) = W(A) + W(B) - W(A \cap B)$
   □ 1 Wurf mit 1 Würfel: A: gerade Augenzahl werfen B: Augenzahl größer als 4 werfen.
   A: {2; 4; 6} B: {5; 6} A ∩ B : {6}
   [6] ist sowohl Element des Ereignisses A als auch Element des Ereignisses B.
   Die Wahrscheinlichkeit W(A∩B) subtrahieren: $W(A \cup B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3} = 0{,}667$

4. Allgemeiner Multiplikationssatz: $W(A \cap B) = W(A) \cdot W(B|A)$
   B|A ist das Ereignis B, das eintritt, wenn das Ereignis A bereits eingetreten ist
   B|A heißt "Ereignis B unter der Bedingung A".
   □ Eine Urne enthält 2 weiße Kugeln und 3 schwarze Kugeln.
   Gesucht ist die Wahrscheinlichkeit, dass bei 2 Zügen 2 schwarze Kugeln gezogen werden.
   Das Ereignis A ist "schwarz beim 1.Zug", das Ereignis B ist "schwarz beim 2.Zug".
   Wir müssen unterscheiden, ob die Kugel nach dem 1.Zug zurückgelegt wird oder nicht.

Ziehen mit Zurücklegen *W* ist vom Ergebnis des 1.Zuges unabhängig	Ziehen ohne Zurücklegen *W* ist vom Ergebnis des 1.Zuges abhängig
W(A∩B) = W(A) · W(B) = $\frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25} = 0{,}36$	W(A∩B) = W(A) · W(B|A) = $\frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} = 0{,}30$

5. Unabhängigkeitssatz
   Die Umkehrung des allgemeinen Multiplikationssatzes gilt auch:
   Zwei Ereignisse A, B sind unabhängig, wenn W(A∩B) = W(A) · W(B).
   Wenn W(A), W(B), W(A∩B) bekannt sind, kann man prüfen, ob A und B unabhängig sind oder ob ein Zusammenhang zwischen A und B besteht.
   In der Realität werden W(A∩B) und W(A) · W(B) nicht exakt übereinstimmen, sondern es ergibt sich eine gewisse Abweichung.
   Mit einem geeigneten Prüfmaß wird man entscheiden, ob die Abweichung zufällig ist oder ausreicht um die Null-Hypothese der Unabhängigkeit zu verwerfen.

4.3 Kontingenztabellen

Beim Zusammenhang zwischen zwei quantitativen Zufallsvariablen spricht man von *Korrelation*, den Zusammenhang zweier nicht-quantitativen Zufallsvariable A, B nennt man *Kontingenz*.

Die Nullhypothese lautet: H_0 : es besteht kein Zusammenhang zwischen A und B oder: die Merkmale A und B sind voneinander unabhängig oder: $W(A \cap B) = W(A) \cdot W(B)$.

Beispiel 4.1

Man wirft eine Münze mit den möglichen Ereignissen "Zahl" oder "Wappen" und einen Würfel mit den möglichen Augenzahlen "1", "2", …, "6".

A ist "Augenzahl [6] des Würfels" B ist " [Zahl] auf der Münze".

Für die Wahrscheinlichkeiten gilt $W(\text{Zahl}) = \frac{1}{2}$, $W(6) = \frac{1}{6}$, $W(\text{Zahl saa } 6) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$

1. Kontingenztabelle mit den Wahrscheinlichkeiten (theoretisch) *Vierfeldertafel*

	"6"	"nicht-6"	
Zahl	$\frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12} = 8{,}33\%$	$\frac{5}{12}$	$W = \frac{1}{2}$
Wappen	$\frac{1}{12}$	$\frac{5}{12}$	$W = \frac{1}{2}$
	$W = \frac{1}{6}$	$W = \frac{5}{6}$	1

Hier gilt exakt $W(A \cap B) = W(A) \cdot W(B)$, also sind A und B unabhängig.

In der Realität sind die Wahrscheinlichkeiten nicht berechenbar, sondern werden durch relative Häufigkeiten $h_i = \frac{n_i}{n} = \frac{\text{einzelne Häufigkeiten}}{\text{Stichprobenumfang}}$ abgeschätzt. $W(A) = \lim_{n \to \infty} h_n(A)$

Für die erwarteten relativen Häufigkeiten gilt entsprechend $W(A \cap B)$?=? $W(A) \cdot W(B)$

$$h(A \cap B) = h_{ij} = h(A) \cdot h(B) \Rightarrow \frac{n_{ij}}{n} = \frac{n_{i\bullet}}{n} \cdot \frac{n_{\bullet j}}{n} \quad |\cdot n \quad \Rightarrow n_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$$

2. Kontingenztabelle mit den Anzahlen n_i, n_j, n_{ij} (theoretisch)

n, $n_{i\bullet}$, $n_{\bullet j}$, sind die Randhäufigkeiten: $n_{i\bullet}$ sind die Zeilensummen, $n_{\bullet j}$ die Spaltensummen

	A = "6"	"nicht-6"	
B="Zahl"	$\frac{150 \cdot 50}{300} = 25 = 8{,}33\%$	125	$n_{1\bullet} = 150$
Wappen	25	125	$n_{2\bullet} = 150$
	$n_{\bullet 1} = 50$	$n_{\bullet 2} = 250$	$n = 300$

3. Kontingenztabelle mit den beobachteten Anzahlen n_{ij} und den theoretischen $u_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$

	"6"		"nicht-6"		
	n_{ij}	u_{ij}	n_{ij}	u_{ij}	
Zahl	22	25	128	125	$n_{1\bullet} = 150$
Wappen	28	25	122	125	$n_{2\bullet} = 150$
	$n_{\bullet 1} = 50$		$n_{\bullet 2} = 250$		$n = 300$

Sind diese Abweichungen statistisch gesichert?

Kann man die Nullhypothese, dass A und B unabhängig sind, ablehnen?

Besteht ein statistisch gesicherter Zusammenhang zwischen den Ergebnissen von A und B ?

4.4 CHI² - UNABHÄNGIGKEITSTEST

Beispiel 4.2

200 zufällig ausgewählte Personen wurden nach ihrer Einstellung zur Abwrackprämie für PKW (Merkmal A) und zu ihrer präferierten Partei (Merkmal B) befragt.
125 Personen waren positiv für die Abwrackprämie eingestellt, 75 negativ.
77 präferierten Partei A, 50 die Partei B, 38 die Partei C und 35 präferierten sonstige Parteien.
Die folgende Tabelle zeigt diese Randhäufigkeiten und die beobachteten Häufigkeiten n_{ij} .

Parteipräferenz:	A		B		C		D		
	n_{ij}	u_{ij}	n_{ij}	u_{ij}	n_{ij}	u_{ij}	n_{ij}	u_{ij}	Summen $n_{i\bullet}$
positive Einstellung	45		40		20		20		125
negative Einstellung	32		10		18		15		75
Summen $n_{\bullet j}$	77		50		38		35		200

a) Freiheitsgrade

Wenn die Randhäufigkeiten der k Zeilen und l Spalten bekannt sind,
dann gibt es nur noch $\nu = (k - 1) \cdot (l - 1)$ Freiheitsgrade
In unserem Beispiel sind das $\nu = (2–1)\cdot(4–1) = 3$ Freiheitsgrade.

b) Die theoretischen Häufigkeiten

u_{ij} werden berechnet und den beobachteten Häufigkeiten n_{ij} gegenübergestellt: $u_{ij} = \dfrac{n_{i\bullet} \cdot n_{\bullet j}}{n}$

c) Abweichungsquadrate

Man berechnet die normierten Abweichungsquadrate $q_{ij} = \dfrac{(n_{ij} - u_{ij})^2}{u_{ij}}$

und bildet die Summe der normierten Abweichungsquadrate $chi^2 = \chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\dfrac{(n_{ij} - u_{ij})^2}{u_{ij}}$

oder $chi^2_{empirisch} = \chi^2_{empirisch} = \sum q_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{l}\dfrac{(n_{ij} - u_{ij})^2}{u_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{l}\dfrac{\left(n_{ij} - \dfrac{n_{i\bullet}\, n_{\bullet j}}{n}\right)^2}{\dfrac{n_{i\bullet}\, n_{\bullet j}}{n}}$

→ *Excel / Kontingenz*

d) Chi²-Unabhängigkeitstest

$u_{ij} \geq 5$, dann gilt:
Der Zusammenhang zwischen den Zufallsvariablen A und B gilt als statistisch gesichert
bei einer Irrtumswahrscheinlichkeit α, wenn $\chi^2_{empirisch} > \chi^2_{crit.}$.
andernfalls wird die Nullhypothese (A, B unabhängig, kein Zusammenhang) beibehalten.
Das Prüfmaß $\chi^2_{crit.}$ liest man für die Parameter ν und α in der *Tabelle 7.4* ab.
In unserem Beispiel ist $\chi^2_{crit.|3|0,05} = 7{,}815$, $\chi^2_{crit.|3|0,01} = 11{,}345$.
Die Rechnung in der Kontingenztabelle ergibt $\chi^2_{empirisch} = 9{,}082$.
Der Zusammenhang zwischen der Einstellung zur Abwrackprämie und der präferierten Partei ist auf dem Niveau α = 5% statistisch gesichert, weil $\chi^2_{empirisch} > \chi^2_{crit.|3|0,05}$.
Bei einem Sicherheitsgrad von 99% würde man die Einstellung zur Abwrackprämie und die präferierte Partei als voneinander unabhängig betrachten, weil $\chi^2_{empirisch} < \chi^2_{crit.|3|0,01}$.

4.5 KONTINGENZTEST

Chi^2 – Verteilung [Pearson, Karl, chi^2 , London, ca. 1914]

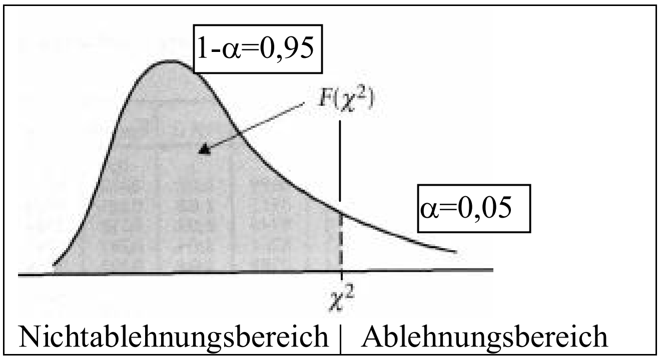
Das Schaubild zeigt den Funktionsgraphen der Dichtefunktion $f_\nu(\chi^2)$

Die Tabelle 7.4 enthält die Werte der Verteilungsfunktion $F_\nu(\chi^2)$

Der Zusammenhang zwischen den Merkmalen A und B ist statistisch gesichert, wenn gilt $\chi^2_{empirisch} > \chi^2_{crit}$.

d.h. wenn $\chi^2_{empirisch}$ im Ablehnungsbereich für die 95% oder 99% Aussagesicherheit liegt.,

andernfalls sind die Merkmale unabhängig.



Aufgabe Kontingenztest

gegeben: Randhäufigkeiten n, $n_{i\bullet}$, $n_{\bullet j}$, beobachteten Häufigkeiten n_{ij}, $u_{ij} \geq 5$
Tabelle als Vordruck aber ohne Farben und Überschriften
Maximal 3×4 beobachtete Häufigkeiten
Sicherheitsgrad 1–α oder Wahrscheinlichkeit α eventuell mehrere zum Vergleich.

gesucht:
1. Entscheidung, ob Nullhypothese beibehalten oder verworfen wird
… ob der Zusammenhang statistisch gesichert ist.
Es ist also auf Kontingenz der Zufallsvariablen A und B zu testen mit Hilfe des Chi^2-Unabhängigkeitstests.
Genaue verbale Formulierung unter expliziten Nennung der beiden Zufallsvariablen und einer kurzen Begründung.
2. wie Nr. 1 aber ein anderer Sicherheitsgrad 1–α

Schritte:

a) Kontingenztabelle erstellen bzw. vervollständigen

$$u_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \qquad q_{ij} = \frac{\left(n_{ij} - u_{ij}\right)^2}{u_{ij}}$$ berechnen und eintragen

b) $\Sigma\, q_{ij}$ bestimmen $= \chi^2_{empirisch}$

c) Freiheitsgrade bestimmen $\nu = (k - 1) \cdot (l - 1)$

d) $\chi^2_{crit.|\,\nu\,|\,\alpha}$ in der Tabelle ablesen

e) $\chi^2_{empirisch} > \chi^2_{crit.}$ prüfen und entscheiden mit einer kurzen Begründung, "weil…"

f) Für den 2. gegebenen Sicherheitsgrad Schritte d) und e) durchführen

4.6 DISKRETE ZUFALLSVARIABLE

Beispiel 4.3 Infektionskrankheit

X sei die Inkubationszeit (Zeit der Ansteckung bis zum Auftreten der ersten Krankheitssymptome). Die Zufallsvariable X werde in vollen Tagen gemessen, deshalb liegt ein diskretes Merkmal vor. Wir nehmen an, dass sich die Wahrscheinlichkeiten $f_i = W(X = x_i)$ aufgrund der Ausbreitungsdynamik des Erregers berechnen lassen.

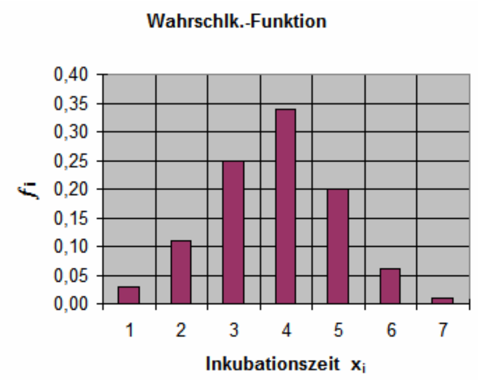
Wir kommen damit zu einer Arbeitstabelle, die denen empirischer Untersuchungen entspricht.

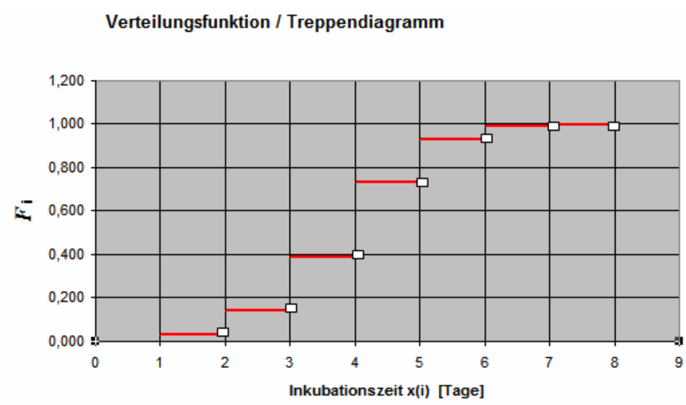
x_i	$f_i = W(X = x_i)$	$F_i = W(X \le x_i)$	$x_i * f_i$	$(x_i)^2 * f_i$
1	0,03	0,03	0,030	0,030
2	0,11	0,14	0,220	0,440
3	0,25	0,39	0,750	2,250
4	0,34	0,73	1,360	5,440
5	0,20	0,93	1,000	5,000
6	0,06	0,99	0,360	2,160
7	0,01	1,00	0,070	0,490
	1,00		3,790	15,810

1. Erwartungswert $\mu = \sum_{i=1}^{k} x_i \cdot f_i = 3{,}79$ [Tage]
2. Erwartete Varianz $\sigma^2 = \sum_{i=1}^{k} \left(x_i^2 \cdot f_i\right) - \mu^2 = 15{,}810 - 3{,}79^2 = 1{,}446$
3. Erwartete Standardabweichung $\sigma = +\sqrt{\sigma^2} = 1{,}20$ [Tage]
4. Mit Werten der Verteilungsfunktion F_i berechnet man Wahrscheinlichkeiten für Intervalle:
   a) Wahrscheinlichkeit für höchstens 3 Tage Inkubationszeit: $F_3 = 0{,}39$
   b) Wahrscheinlichkeit für eine Inkubationszeit von 4 bis 6 Tagen: $F_6 - F_3 = 0{,}60$

   allgemein: Wahrscheinlichkeit $W(X \in]\,a\,;\,b]) = W(a < X \le b) = F(b) - F(a - 1)$

   c) Wahrscheinlichkeit für eine Inkubationszeit von unter 6 Tagen: $F_5 = 0{,}93$
   d) Wahrscheinlichkeit für eine Inkubationszeit von mindestens 5 Tagen: $1 - F_4 = 0{,}27$
   e) Wahrscheinlichkeit für eine Inkubationszeit über 5 Tage: $1 - F_5 = 0{,}07$
5. Die Wahrscheinlichkeitsfunktion $f(x_i)$ ordnet jedem Wert x_i der Zufallsvariablen die Wahrscheinlichkeit $W(X=x_i)$ zu. Die Wertepaare x_i , $f(x_i)$ lassen sich als Stabdiagramm (oder auch als Histogramm) darstellen.
6. Die Verteilungsfunktion $F(x_i)$ ordnet jedem Wert x_i die aufsummierte Wahrscheinlichkeit $W(X \le x_i)$ zu . Die Wertepaare x_i , $F(x_i)$ lassen sich als Treppendiagramm (oder auch als Histogramm) darstellen





4.7 WÜRFEL-MODELL

<u>Beispiel 4.4</u>
Das Werfen eines idealen Würfels mit der Zufallsvariable X "Augenzahl beim Wurf eines Würfels" ist eines der einfachsten Modelle der Wahrscheinlichkeitsrechnung.
X kann die Werte {1; 2; …; 6} annehmen mit den Wahrscheinlichkeiten $W(X = x_i) = 1/6$.

Augen-zahl x(i)	$f_i = W(X = x_i)$	$F_i = W(X \le x_i)$	$x_i * f_i$	$(x_i)^2 * f_i$
1	0,167	0,167	0,167	0,167
2	0,167	0,333	0,333	0,667
3	0,167	0,500	0,500	1,500
4	0,167	0,667	0,667	2,667
5	0,167	0,833	0,833	4,167
6	0,167	1,000	1,000	6,000
	1,000		3,500	15,167

$$\mu = \sum_{i=1}^{k} x_i \cdot f_i = 3{,}5$$

$$\sigma^2 = \sum_{i=1}^{k} x_i^2 \cdot f_i - \mu^2 = 15{,}167 - 3{,}5^2 = 2{,}917$$

$$\sigma = +\sqrt{\sigma^2} = 1{,}708$$

Beim Wurf <u>eines</u> Würfels ist der Erwartungswert 3,5 und die Streuung $\sigma = 1{,}708$.
Wenn wir einen Zufallsversuch mit n Würfeln machen, überlagern sich die Zufallsvariablen additiv.
Wir erwarten weiterhin einen Mittelwert von 3,5 aber die Streuung müsste geringer sein – sie ist um das $\sqrt{n}$ -fache geringer, d.h. bei 4 Würfeln: $\sigma_{\bar{x},4} = 1{,}708 / 2 = 0{,}854$.

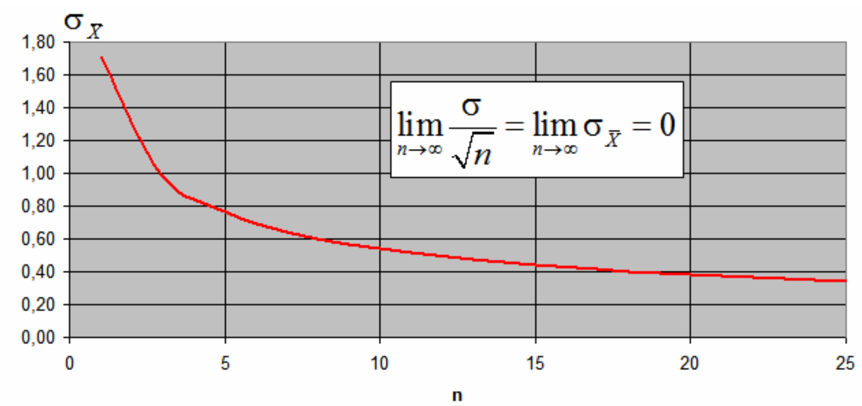
<u>Das Stichprobenmittel</u>
In vielen Situationen überlagern sich unabhängige Zufallsvariable $X_1, X_2, \ldots, X_n$ additiv, dabei entsteht die Zufallsvariable $U = X_1 + X_2 + \ldots + X_n$
und die Zufallsvariable "Stichprobenmittel" $\bar{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$.

Für die erwartete Standardabweichung des Stichprobenmittels gilt $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$ $\boxed{\sqrt{n} - \text{Gesetz}}$

Die Zufallsvariable X streut $\sqrt{n}$ mal stärker als das Stichprobenmittel $\bar{X}$.

- Die Inkubationszeit aus Beispiel 3.3 seien die Werte aus einem bestimmten Landkreis. Die Streuung war $\sigma = 1{,}2$ Tage. Bei 25 Landkreisen wäre $\sigma_{\bar{x}} = \dfrac{1}{\sqrt{25}} \cdot 1{,}2 = 0{,}24$ Tage.
- X sei die Zufallsvariable X "Summe der Augenzahl beim Wurf von n Würfeln" Die Streuung $\sigma_{n=1}$ war 1,708. Bei 25 Würfen oder 25 Würfeln: $\sigma_{n=25} = 0{,}34$.

Der Funktionsgraph zu $\sigma_{\bar{x}}$:



4.8 √N - GESETZ

1. Eine linear transformierte Zufallsvariable $aX + b$ *(linear: $y = m\,x + b$)*

- Ein Autoverkäufer verdient an jedem verkauften Auto 200 €, außerdem erhält er ein monatliches Grundgehalt von 1500 €.
  X sei die Zufallsvariable "Anzahl der pro Monat verkauften Autos".
  Der gesamte Monatsverdienst ist dann die Zufallsvariable 200 · X + 1500, allgemein a X + b.

Für den Erwartungswert gilt: $\mu\,(aX + b) = a \cdot \mu(X) + b$ ❶

$$\mu(a\,X+b)=\sum(a\,x_i+b)\cdot f(x_i)=\sum\big(a\,x_i\cdot f(x_i)+b\cdot f(x_i)\big)=\sum a\,x_i\cdot f(x_i)+\sum b\cdot f(x_i)$$
$$=a\sum x_i\cdot f(x_i)+b\sum f(x_i)=a\cdot\mu\big(X\big)+b\cdot 1=a\,\mu(X)+b$$

Für die erwartete Varianz gilt: $\sigma^2\,(aX + b) = a^2 \cdot \sigma^2(X)$ ❷

$$\sigma^2(aX+b)=\sum\big(a\,x_i+b-\mu(X)\big)^2\cdot f(x_i)=\sum\big(a\,x_i+b-a\,\mu(X)-b\big)^2\cdot f(x_i)$$
$$=\sum\big(a\,x_i-a\,\mu(X)\big)^2\cdot f(x_i)=\sum a^2\cdot\big(x_i-\mu(X)\big)^2\cdot f(x_i)=a^2\sum\big(x_i-\mu(X)\big)^2\cdot f(x_i)$$
$$=a^2\cdot\sigma^2(X)$$

2. Wir bestimmen nun die Maßzahlen μ und σ^2 für das Stichprobenmittel $\bar{X}$:

2.1 $\mu\big(\bar{X}\big)=\mu(X)$

aus ❶ $\mu\,(aX) = a \cdot \mu(X)$ $\Rightarrow$ $\mu(X)=\frac{1}{a}\mu(aX)$

mit n statt a: $\mu(X)=\frac{1}{n}\mu(nX)\ \Rightarrow\ \mu(\bar{X})=\frac{1}{n}\big(\mu\big(X_1\big)+\mu\big(X_2\big)+\ldots+\mu\big(X_n\big)\big)$

$$\Rightarrow\mu(\bar{X})=\frac{1}{n}n\,\mu(X)=\mu(X)$$

2.2 $\sigma^2\big(\bar{X}\big)=\frac{1}{n}\sigma^2\big(X\big)$

aus ❷ $\sigma^2(aX) = a^2 \cdot \sigma^2(X)$ $\Rightarrow$ $\sigma^2(X)=\frac{1}{a^2}\sigma^2(aX)$

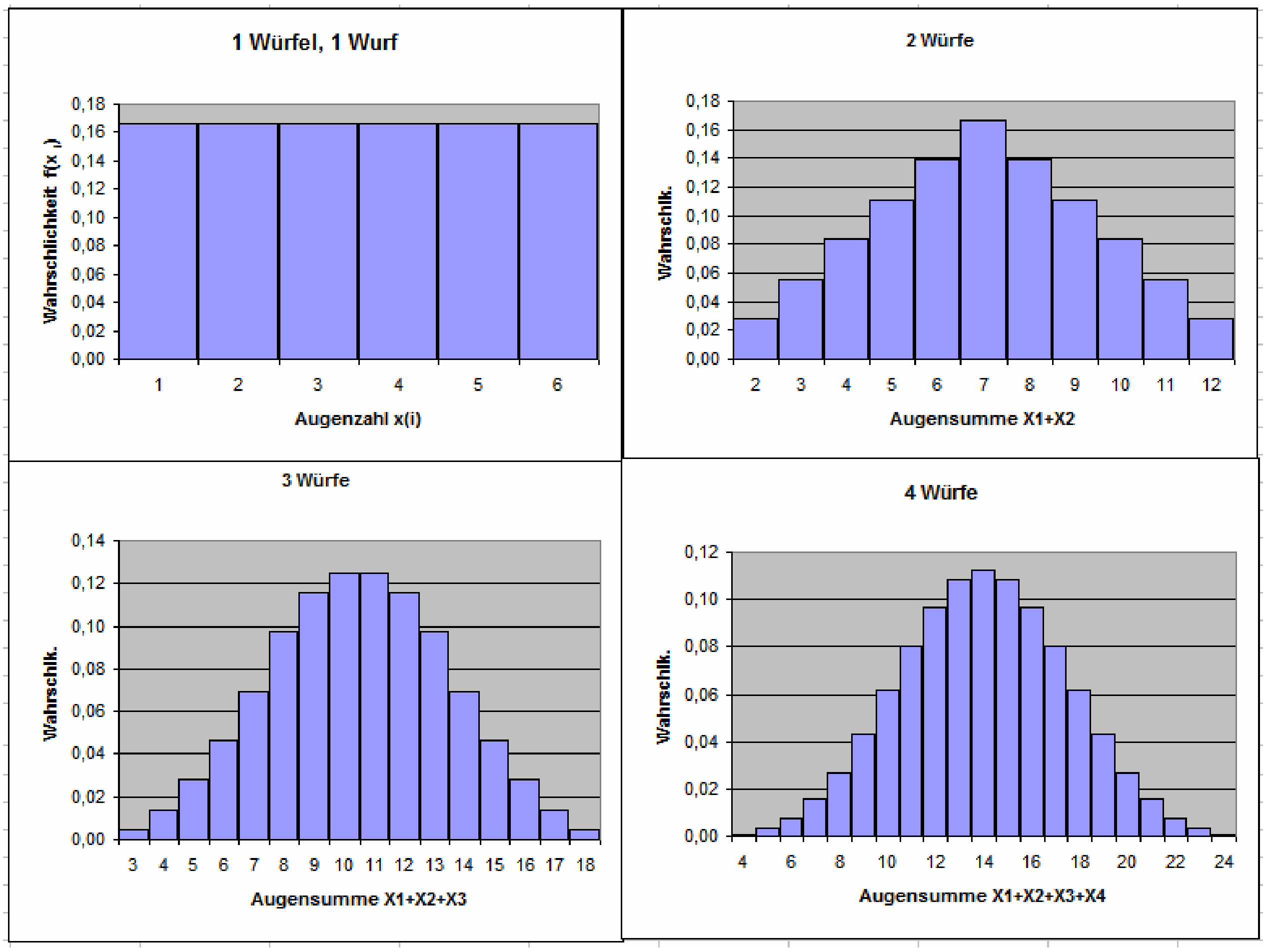
mit n statt a: $\sigma^2(X)=\frac{1}{n^2}\sigma^2(nX)\ \Rightarrow\ \sigma^2(\bar{X})=\frac{1}{n^2}\big(\sigma^2\big(X_1\big)+\sigma^2\big(X_2\big)+\ldots+\sigma^2\big(X_n\big)\big)$

$$\sigma^2(\bar{X})=\frac{1}{n^2}n\,\sigma^2(X)=\frac{\sigma^2(X)}{n}=\frac{1}{n}\sigma^2(X)\ \Rightarrow\ \sigma(\bar{X})=\frac{1}{\sqrt{n}}\sigma(X)$$

4.9 ZENTRALER GRENZWERTSATZ

Weiter Beispiel 4.4

X sei die Zufallsvariable X "Summe der Augenzahl beim Wurf von n Würfeln"
Die folgenden Histogramme zeigen die Wahrscheinlichkeiten für $n = 1, 2, 3, 4$.



Man erkennt anschaulich zwei grundsätzliche Gesetzmäßigkeiten der Statistik:

1. Mit zunehmendem n verringert sich die Streuung um die "zentrale Mitte" μ.

   exakt: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sqrt{n}$ – Gesetz

2. Obwohl die einzelne Zufallsvariable X eindeutig keine "Normalverteilung" zeigt,

   kommt die Zufallsvariable "Stichprobenmittel" $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$

   der Normalverteilung mit zunehmendem n immer näher.
   Diese Gesetzmäßigkeit nennt man Zentraler Grenzwertsatz.
   Dabei müssen die Zufallsvariablen X_i nicht völlig voneinander unabhängig sein.

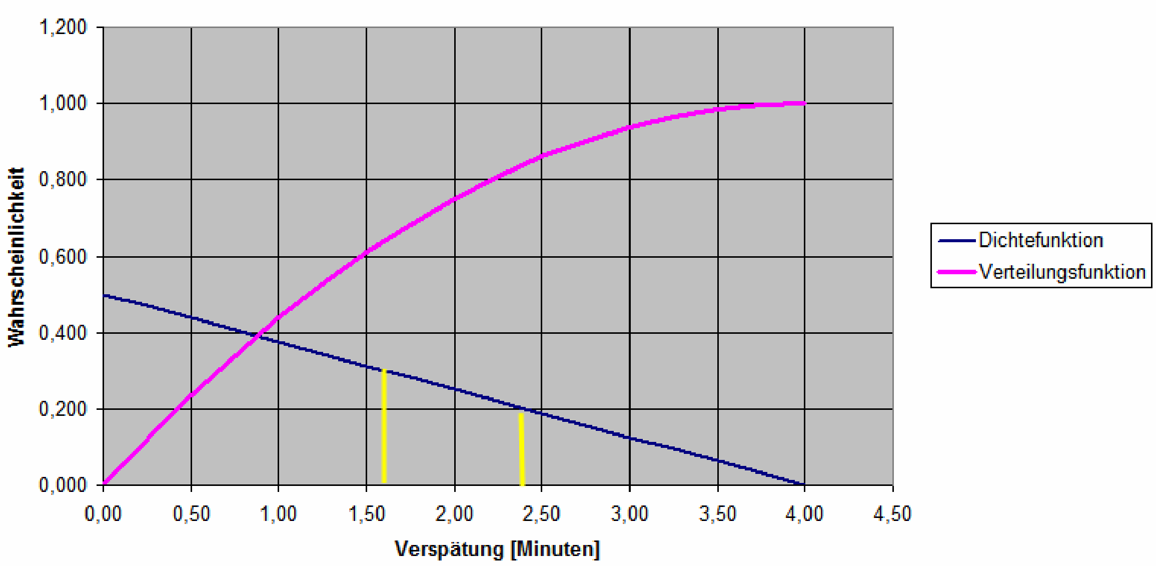
3. Für hinreichend große n kann man fast alle statistische Verteilungen mit Hilfe der Normalverteilung untersuchen. (Themen 5 und 6)

4.10 STETIGE ZUFALLSVARIABLE 1

Beispiel 4.5 *nach Bleymüller S. 41*

Die Verspätung der U-Bahn an einer bestimmten Haltestelle, ist eine stetige Zufallsvariable X. Die Verspätung ist in Minuten angegeben.

Die Dichtefunktion sei $f(x) = \begin{cases} 0{,}5 - 0{,}125\,x & \text{für } 0 \le x \le 4 \\ 0 & \text{für alle übrigen } x \end{cases}$



Dichte- und Verteilungsfunktion stetiger Zufallsvariable:

Allgemein gilt für die Dichten $f(x) \ge 0$ und $\int_{-\infty}^{+\infty} f(x)dx = 1 = 100\%$.

Die Verteilungsfunktion $F(x)$ ordnet jedem Intervall $-\infty < x < x_2$ die Wahrscheinlichkeit

$$W(X \le x_2) = \int_{-\infty}^{x_2} f(x)dx = F(x_2) \text{ zu.}$$

Die Verteilungsfunktion $F(x_2)$ ist stetig und monoton steigend mit $\lim_{x_2 \to \infty} F(x_2) = 1 = 100\%$

Die Wahrscheinlichkeit dafür, dass X im Intervall [a; b] liegt, ist

$$W(a \le X \le b) = \int_a^b f(x)dx = F(b) - F(a) = [F(x_2)]_a^b$$

Zum Beispiel 4.5

□ Verteilungsfunktion

$$\int (0{,}5 - 0{,}125x)dx = 0{,}5x - 0{,}0625x^2 \Rightarrow F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 0{,}5x - 0{,}0625x^2 & \text{für } 0 \le x \le 4 \\ 1 & \text{für } x > 4 \end{cases}$$

□ Wahrscheinlichkeit dafür, dass die U-Bahn zwischen 1,6 und 2,4 Minuten Verspätung hat:

$$W(1{,}6 \le X \le 2{,}4) = \int_{1{,}6}^{2{,}4} \left(\frac{1}{2} - \frac{1}{8}x\right)dx = \left[\frac{1}{2}x - \frac{1}{16}x^2\right]_{1{,}6}^{2{,}4} = 1{,}2 - 0{,}36 - (0{,}8 - 0{,}16) = 0{,}2 = 20\%$$

□ Wahrscheinlichkeit dafür, dass die U-Bahn genau 2 Minuten Verspätung hat

$W(X = 2) = 0 = F(2) - F(2)$

Bei stetigen Zufallsvariablen sind Wahrscheinlichkeiten nur für Intervalle definiert.

4.11 STETIGE ZUFALLSVARIABLE 2

Erwartungswert und Varianz für stetige Zufallsvariable ergeben sich in Analogie zu den diskreten Zufallsvariablen:

Erwartungswert $\mu = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$ entsprechend dem diskreten $\mu = \sum_i \mathrm{x_i} \cdot f_i$

Varianz (1.Formel) $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x)\,dx$ entsprechend: $\sum (\mathrm{x_i} - \mu)^2 \cdot f_i$

Varianz (2.Formel) $\sigma^2 = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu^2$ entsprechend: $\sum_i \mathrm{x_i^2} \cdot f_i - \mu^2$

Erwartete Standardabweichung $\sigma = +\sqrt{\sigma^2}$

Zum Beispiel 4.5

$$\mu = \int_0^4 x \cdot (0{,}5 - 0{,}125x)\,dx = \int_0^4 (0{,}5x - 0{,}125x^2)\,dx = \left[\frac{1}{4}x^2 - \frac{1}{24}x^3\right]_0^4 = 4 - 2{,}667 = 1{,}333\,[\text{Minuten}]$$

$$\sigma^2 = \int_0^4 x^2 \cdot (0{,}5 - 0{,}125x)\,dx - 1{,}333^2 = \int_0^4 (0{,}5x^2 - 0{,}125x^3)\,dx - 1{,}778 =$$

$$= \left[\frac{1}{6}x^3 - \frac{1}{32}x^4\right]_0^4 - 1{,}778 = 10{,}667 - 8 - 1{,}778 = 0{,}889$$

$$\sigma = +\sqrt{0{,}889} = 0{,}943\,[\text{Minuten}]$$

Aufgabe Stetige Zufallsvariable

gegeben: Dichtefunktion einer stetigen Zufallsvariablen, nur eine Gerade oder eine Parabel

gesucht / *Schritte:*

Geben Sie zu Ihren Ergebnissen die Einheiten an, wenn dies sinnvoll ist.

1. Skizze des Funktionsgraphen der Dichtefunktion

   einige Wertepaare bestimmen, Punkte zeichnen, verbinden

2. Funktionsgleichung der Verteilungsfunktion

   Integration, $W(X \le x_2) = \int_{-\infty}^{x_2} f(x)dx = F(x_2)$ *für die gegebene Dichtefunktion*

   Intervalle für $F(x_2)=0$ *und* $F(x_2)=1$ *nicht vergessen*

3. Wahrscheinlichkeiten für bestimmte Intervalle

   $W(a \le X \le b) = F(b) - F(a) = [F(x_2)]_a^b$ *Einheit: ohne oder %*

4. Erwartungswert

   $\mu = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx$ *, entsprechendes bestimmtes Integral berechnen, Einheit!*

5. Varianz

   $\sigma^2 = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu^2$ *, entsprechendes bestimmtes Integral berechnen*

6. Standardabweichung

   $\sigma = +\sqrt{\sigma^2}$ *berechnen, Einheit hinzufügen*