Formelsammlung Statistik

Lageparameter und Streumaße:

Teil 1: Diskrete Merkmalsdarstellung (Einzelwerte)

Einfaches arithmetisches Mittel: $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \left[oder \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \right]$

Gewogenes arithmetisches Mittel:

$$\mu = \frac{\sum_{i=1}^{k}(x_i \cdot n_i)}{\sum_{i=1}^{k} n_i} = \frac{1}{n}\sum_{i=1}^{k}(x_i \cdot n_i) = \sum_{i=1}^{k}\left(x_i \cdot \frac{n_i}{n}\right)$$

Median (Zentralwert):

$$\overline{x_M} = \begin{cases} x_{\frac{n+1}{2}} & \text{für n ungerade} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{für n gerade} \end{cases}$$

Quantile:

$$\overline{x_p} = \begin{cases} x_{[n \cdot p]+1} & \text{für } (\mathrm{n} \cdot \mathrm{p}) \text{ nicht ganzzahlig} \\ \frac{1}{2}\left(x_{n \cdot p} + x_{n \cdot p+1}\right) & \text{für } (\mathrm{n} \cdot \mathrm{p}) \text{ ganzzahlig} \end{cases}$$

Varianz: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 - \mu^2$

Varianz bei Häufigkeitsverteilung: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \mu)^2 \cdot n_i$

Standardabweichung: $\sigma = \sqrt{\sigma^2}$

Modus: *Häufigster Wert einer Verteilung.*

Teil 2: Klassierte Merkmalsausprägungen

Arithmetisches Mittel bei klassierten Merkmalsausprägungen:

$$\mu = \frac{1}{n}\sum_{i=1}^{k}\left[(x_i)_m \cdot n_i\right] = \sum_{i=1}^{k}\left[(x_i)_m \cdot \frac{n_i}{n}\right] \quad mit\ (x_i)_m\ als\ Klassenmitte\ der\ Klasse\ i$$

Modus / Modalwert:

- ⇨ **Modale Klasse: Klasse mit max. Häufigkeitsdichte**
- ⇨ **Modus / Modalwert: Wert der Klassenmitte der modalen Klasse**

Median und Quartile:

Median (bei intervallskaliertem Merkmal): $\overline{X}_{0,5} \in [a;b]$, $\overline{X}_{0,5} = a + \frac{\Delta_i \cdot (0{,}5 - F(a))}{p_i}$

unteres Quartil (bei intervallskaliertem Merkmal): $\overline{X}_{0,25} \in [a;b]$, $\overline{X}_{0,25} = a + \frac{\Delta_i \cdot (0{,}25 - F(a))}{p_i}$

oberes Quartil (bei intervallskaliertem Merkmal): $\overline{X}_{0,75} \in [a;b]$, $\overline{X}_{0,75} = a + \frac{\Delta_i \cdot (0{,}75 - F(a))}{p_i}$

Regel(n) zur Klassenbildung in der Statistik

Rule of Sturges Erklärung, Anwendungen und Beispiele

Die **Sturges-Regel** ist ein Kriterium, das verwendet wird, um die Anzahl von Klassen oder Intervallen zu bestimmen, die notwendig sind, um einen Satz von statistischen Daten graphisch darzustellen. Diese Regel wurde 1926 vom deutschen Mathematiker Herbert Sturges verkündet.

Sturges schlug eine einfache Methode vor, die auf der Anzahl der Samples basiert, die es ermöglichen würden, die Anzahl der Klassen und ihre Reichweitenamplitude zu finden. Die Sturges-Regel wird insbesondere im Bereich der Statistik häufig verwendet, insbesondere um (Frequenz)Histogramme im Sinne von Säulendiagrammen zu erstellen.

Erklärung

Die Sturges-Regel ist eine empirische Methode, die in der deskriptiven Statistik weit verbreitet ist, um die Anzahl von Klassen zu bestimmen, die in einem (Frequenz)Histogramm existieren müssen, um eine Menge von Daten zu klassifizieren, die eine Stichprobe oder Population repräsentieren.

Grundsätzlich bestimmt diese Regel die Breite der Säulen der (Häufigkeits-)Histogramme.

Zur Herbeiführung seiner Regel betrachtete Herbert Sturges ein ideales Frequenzdiagramm, das aus k Intervallen besteht, wobei das i-te Intervall eine bestimmte Anzahl von Werten (i = 0, ... k - 1) enthält.
Diese Anzahl von Stichproben ergibt sich aus der Anzahl der Möglichkeiten, wie Anzahl von Teilmengen einer Menge dargestellt werden kann; das heißt, durch den Binomialkoeffizienten ausgedrückt wie folgt:

$$C_{(k-1;i)} = \binom{k-1}{i} = \frac{(k-1)!}{i! \cdot (k-i-1)!}$$

Dann konnte Sturges folgern, dass sich das Frequenzhistogramm einer Normalverteilung annähert, wenn die Anzahl der Intervalle (k) gemäß dem zentralen Grenzwertsatz zunimmt; zudem kann die Anzahl der Teilmengen jedes der Intervalle berechnet werden:

$$N = \sum_{i=0}^{k-1} \binom{k-1}{i} = 2^{k-1} \quad \left[Zeilensumme\ im\ Pacal'schen\ Dreieck \right]$$

Um den Ausdruck zu vereinfachen, verwendete er die Eigenschaften der Logarithmen in beiden Teilen der Gleichung:

$$\log_2(N) = \log_2\left(2^{k-1}\right) = k-1$$

$$\xrightarrow{\text{auflösen nach } k} k = \log_2(N)+1$$

$$\xrightarrow{\text{Basistransformation}} k = \frac{\log_{10}(N)}{\log_{10} 2}+1 = 1+3{,}3219 \cdot \log_{10}(N)$$

In diesem Ausdruck:
- k ist die Anzahl der Klassen und - N ist die Gesamtzahl der Beobachtungen in der Stichprobe

Um beispielsweise ein Häufigkeitshistogramm zu erstellen, das eine zufällige Stichprobe der Höhe von 142 Kindern darstellt, lautet die Anzahl der Intervalle oder Klassen, die die Verteilung haben soll:

k = 1 + 3,322* log (142) => k = 1 + 3,322* 2,1523 => k = 8.14 ÷ 8

Somit wird die Verteilung in 8 Intervallen erfolgen.
Die Anzahl der Intervalle sollte immer durch ganze Zahlen dargestellt werden. In Fällen, in denen der Wert dezimal ist, muss eine Annäherung an die nächste ganze Zahl vorgenommen werden.

Beispiel
Es ist notwendig, ein Histogramm gemäß den gegebenen Daten durchzuführen, die dem Alter entsprechen, das in einer Umfrage von Männern erhalten wird, die in einem örtlichen Fitnessstudio trainieren.

15	38	14	13	29	25
20	13	16	32	44	39
45	46	19	23	24	18
19	20	21	18	25	33
13	18	22	24	27	27

Um die Intervalle zu bestimmen, müssen Sie wissen, wie groß die Stichprobe oder die Anzahl der Beobachtungen ist. In diesem Fall hat man 30.
Dann gilt die Sturges-Regel:

k = 1 + 3,322* log (30) => k = 1 + 3,322* 1,4771 => k = 5,90 => 6 Intervalle.

Aus der Anzahl der Intervalle können Sie die Klassenbreiten berechnen, die diese haben werden;

das heißt, die Breite jedes Balkens im Histogramm: $$a = \frac{x_{\max} - x_{\min}}{k}$$

Die untere Grenze gilt als der niedrigste Wert der Daten und die obere Grenze ist der höchste Wert. Der Unterschied zwischen der oberen und unteren Grenze wird als Bereich der Variablen (R) bezeichnet.
Aus der Übersicht/Tabelle haben wir, dass die obere Grenze 46 und die untere Grenze 13 ist; auf diese Weise wird die Breite jeder Klasse sein:

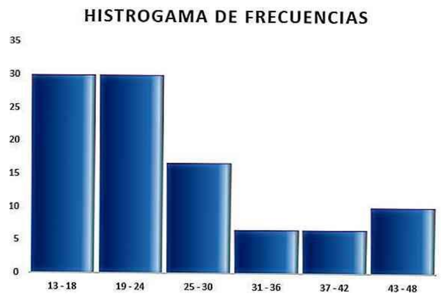
$$a = \frac{46-13}{6} = \frac{33}{6} = 5,5 \rightarrow c = 6\left[Klassenbreite\right]$$

Die Intervalle setzen sich aus einer oberen und unteren Grenze zusammen. Um diese Intervalle zu bestimmen, beginnen Sie, von der unteren Grenze aus zu zählen und addieren Sie dazu die Breite, die durch folgenden Algorithmus bestimmt wird, wie folgt:

Limite Inferior = Limite Superior + 1	Limite Superior = Li + amplitud - 1	Intervalos
13	13 + 6 - 1 = 18	13 – 18
18 + 1 = 19	19 + 6 - 1 = 24	19 – 24
24 + 1 = 25	25 + 6 - 1 = 30	25 – 30
30 + 1 = 31	31 + 6 - 1 = 36	31 – 36
36 + 7 = 37	37 + 6 - 1 = 42	37 – 42
42 + 1 = 43	43 + 6 – 1 = 48	43 – 48

Dann wird die absolute Häufigkeit berechnet, um die Anzahl der Männer zu bestimmen, die jedem Intervall entsprechen; in diesem Fall ist es:

Intervalo	Limites	Fa	Fi
1	13 - 18	9	0,30
2	19 - 24	9	0,30
3	25 - 30	5	0,1666
4	31 - 36	2	0,0666
5	37 - 42	2	0,0666
6	43 - 48	3	0,10



Auf diese Weise erlaubt die Sturges-Regel, die Anzahl der Klassen oder Intervalle zu bestimmen, in denen eine Probe geteilt werden kann, um eine Stichprobe von Daten durch die Erstellung von Tabellen und Graphen zusammenzufassen.

Referenzen

1. Alfonso Urquía, M. V. (2013). Modellierung und Simulation von diskreten Ereignissen. UNED,.
2. Altman Naomi, M. K. (2015). "Einfache lineare Regression." Nature Methods.
3. Antúnez, R.J. (2014). Statistiken in der Bildung. Digitale UNID.
4. Fox, J. (1997). Angewandte Regressionsanalyse, lineare Modelle und verwandte Methoden. SAGE-Veröffentlichungen.
5. Humberto Llinás Solano, C. R. (2005). Deskriptive Statistiken und Wahrscheinlichkeitsverteilungen. Universität des Nordens.
6. Panteleeva, O. V. (2005). Grundlagen der Wahrscheinlichkeit und Statistik.
7. O. Kühl, M. O. (2001). Design of Experiments: Statistische Prinzipien der Design und Analyse von Forschung. Thomson Verlag.

https://de.thpanorama.com/blog/ciencia/regla-de-sturges-explicacin-aplicaciones-y-ejemplos.html

Grundlegende Überlegungen: ***Regeln zur Klassenbildung in der Statistik***

Bei der **Klassenbildung** steht man vor diversen Problemen, die beachtet werden sollten. Ein Patentrezept gibt es hierzu nicht, grundsätzlich kann hier relativ willkürlich vorgegangen werden. Einige **Regeln** sollten wir aber – wenn möglich – beachten:

1. Bilde, wenn möglich, sinnvoll äquidistante Klassenbreiten.
2. Bilde nicht zu viele (gewünschte Informationsverdichtung wird nicht erreicht) aber auch nicht zu wenige Klassen (Struktur der ursprünglichen Daten geht evtl. verloren).

   Als Regeln für die Anzahl der Klassen k bei n voneinander verschiedenen Beobachtungswerten haben sich hierzu u.a. herausgebildet:
   - $k = \sqrt{n}$ für n ≤ 100 (Faustregel)
   - $k = 1 + 3,3 log_{10}$ n = $1 + 3,3 ln(n)/ln10$ (Sturges-Regel)
   - k = 10 bei n ≈ 100, k = 13 bei n ≈ 1.000 und k = 16 bei n ≈ 10.000 (DIN 55302, Blatt 1)
3. Vermeide es, Bereiche, in denen Merkmalsausprägungen gehäuft auftauchen, durch eine Klassengrenze zu zerschneiden oder sie am Rand einer Klasse gehäuft auftreten zu lassen.

   Hier kann es dann sinnvoll sein, nicht-äquidistante Klassenbreiten zu verwenden; es ist aber nicht immer möglich, sich an alle Regeln zu halten und in einigen Fällen (z.B. bei der Notenzuordnung) werden die Klassengrenzen auch im Vorhinein festgelegt, d.h. die Beobachtungen liegen noch gar nicht vor.
4. Versuche die Klassen homogen (= gleichmäßig) zu besetzen bzw. gehäufte Bereiche in die Klassenmitte zu bringen.
5. Fasse Bereiche mit sehr wenigen Merkmalsausprägungen zu einer einzigen Klasse zusammen.
6. Vermeide offene Randklassen, d.h. als untere Klasse eine Einteilung „weniger als …" bzw. als obere Randklasse „mehr als …" zu verwenden. Ist dies nicht möglich, sind streng genommen keine Histogramme darstellbar und keine statistischen Maßzahlen berechenbar. Hilfsweise kann man für die offenen Klassen
   - die sonst übliche Klassenbreite,
   - die benachbarte Klassenbreite oder
   - einen objektiv sinnvollen Wert

   verwenden.

   Offene Randklassen werden zumeist bei der Einkommensverteilung gebraucht, die Angabe einer Einkommenshöchstgrenze ist hier nicht möglich oder aus Datenschutzgründen sogar verboten.

https://www.wiwiweb.de/statistik/haeufigkeit/klassierte/klassenbild.html

Vorschriften zur Klassenbildung (heuristische Vorschriften):

$n \geq 25$!! (Mache keine statistische Analyse mit zu wenigen Daten!!)

- **Wahl der Klassenzahl k:**

1. „Faustregel":
   $k = \sqrt{n}$ $\quad 5 \leq k \leq 20 \quad$ ($\Rightarrow$ n > 25 !)
2. DIN-Normen für spezielle Anwendungen, z.B. DIN 55302

n	k
bis 100	>10
bis 1000	>13
bis 10000	>16

3. Informationstheoretisches Kriterium:
   $k = 1 + 3{,}32\log(n)$, $\quad 5 \leq k \leq 20 \quad k \leq 5\log(n)$

In jedem Fall wird k auf einen ganzzahligen Wert auf- oder abgerundet.

- **Wahl der Klassengrenzen/ Reduktionslage:**

Vorschriften, die besagen, wie mit Werten zu verfahren ist, die auf eine Klassengrenze fallen:

1. Die Klassengrenzen sollen so gewählt werden, daß kein Wert auf eine Klassengrenze fällt. (z.B., indem man sie zwischen zwei Werte legt (½-Genauigkeit!)).

   Das wird z.B. erreicht, indem man den unteren Wert a_1^u der ersten Klasse K_1 gemäß $a_1^u = x_{min} - \varepsilon$ festlegt, wobei ε die halbe Meßgenauigkeit ist und die Klassenbreite B auf die Meßgenauigkeit aufrundet.

   a_1^u wird als ***Reduktionslage*** bezeichnet.

2. Die Werte, die auf eine Klassengrenze fallen, werden zur Hälfte der einen und zur Hälfte der anderen zugeordnet. Bei ungeraden Anzahlen erhält die obere Klasse den „überschüssigen" Wert.

usw.

- **Wahl der Klassenbreite:**

Man wählt in der Regel eine konstante Klassenbreite $B = \Delta K_i$ für i=1,...,k.

Sei x_{min} – kleinster Wert und x_{max} – größter Wert der Stichprobe.

In Anbetracht, daß x_{min} und x_{max} nicht auf eine Klassengrenze fallen sollen, muß ein Bereich von x_{min} - a bis x_{max} + b in k Klassen zerlegt werden (a,b – vorgegebene Größe, z.B. (½-Maßeinheit)):

Klassenbreite = B = $\dfrac{(x_{max} + b - x_{min} + a)}{k}$

Man **rundet B** auf die Meßgenauigkeit **auf**!

> ***Vorschrift * :***
>
> Klassenzahl: $k = \sqrt{n}$, $5 \leq k \leq 20$, k wird auf einen ganzzahligen Wert gerundet
> Reduktionslage: $(= a_1^u)$ $\quad a_1^u = x_{min} - \varepsilon$, wobei ε die halbe Meßgenauigkeit ist
> Klassenbreite: B = $\dfrac{(x_{max} - x_{min} + 1)}{k}$, (a=b=ε), B wird auf die Meßgenauigkeit ***auf***gerundet.

Zu Beispiel 7:
Wir wollen eine Klasseneinteilung der in der Urliste Beispiel 7 aufgelisteten Längen der 40 Schrauben bilden:

1. Variante: Klassenzahl: Vorschrift 3: $k = 1 + 3{,}32\log(40) = 1 + 3{,}32 \cdot 1{,}602 = 6{,}32$
   $\Rightarrow$ k = 7 (wir runden hier auf).

   Reduktionslage: $x_{min} = 119$, $x_{max} = 176$
   Wir zerlegen den Bereich von 118 – 177 in 7 Klassen (a=b=1).
   $a_1^u = x_{min} - 1 = 118$

   Klassenbreite: $B = \left(\dfrac{177-118)}{7}\right) = 8{,}4 \Rightarrow B = 9$

   Wir erhalten die in Tabelle 1.9 gegebene Klasseneinteilung.

Es ist noch zu überprüfen, ob ein Wert auf eine Klassengrenze fällt. Das ist hier nicht der Fall.

2. Variante: Klasseneinteilung nach Vorschrift* :

   Klassenzahl: $\sqrt{40} = 6{,}32$ $\Rightarrow$ k = 6 (wir runden hier nach der üblichen Vorgehensweise).

   Reduktionslage: $x_{min} = 119$, $x_{max} = 176$
   Wir wählen: $a_1^u = x_{min} - 1/2$ und zerlegen den Bereich von 118,5 – 176,5 in 6 Klassen (a=b=ε=1/2).

Klassenbreite: $B = \left(\dfrac{176-119+1}{6}\right) = 9{,}67 \Rightarrow B = 10$

Ergänzung MW/Streuung

Mittelwerte und Streuungsmaße bei Einzelwerten, Klassen und stetigen Werten

	Einzelwerte (vorwiegend diskret)	Klassierte Werte
Arithmetisches Mittel	**Einzelwerte:** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ **Gewichtet auf Basis relativer Häufigkeit:** $\overline{x} = \sum_{i=1}^{k}\left(x_i \cdot \frac{n_i}{n}\right) = \sum_{i=1}^{k}(x_i \cdot p_i)$ $mit \quad p_i = \frac{n_i}{n} = \text{relative Häufigkeit}$	$\overline{x} = \sum_{i=1}^{k}\left[(x_i)_m \cdot \frac{n_i}{n}\right] = \sum_{i=1}^{k}\left[(x_i)_m \cdot p_i\right]$ $mit\ (x_i)_m$ *als Klassenmitte der Klasse i* $und\ \ p_i = \frac{n_i}{n} = \text{relative Häufigkeit der Klasse i}$
Median / Zentralwert	$\overline{x_M} = \begin{cases} x_{\frac{n+1}{2}} & \text{für n ungerade} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{für n gerade} \end{cases}$	$\overline{x_M} \Rightarrow \overline{x_{0,5}} = [a\,;b] \rightarrow \overline{x_{0,5}} = a + \frac{\Delta_i \cdot [0{,}5 - F(a)]}{p_i}$ $\Delta_i = Klassenbreite \quad p_i = rel.\ Häufigkeit$ $F(a) = \text{kum. rel. Häufigkeit bis Medianintervall } [a\,;b]$
Modus / Modalwert	***Häufigster Wert einer Verteilung.***	**Schritt 1: Modale Klasse bestimmen** ⇨ **Klasse mit max. Häufigkeitsdichte** **Schritt 2:** ⇨ **Wert der Klassenmitte der modalen Klasse**
Spannweite	$w = \max(x_i) - \min(x_i)$	$w = Klassengrenze_{\max} - Klassengrenze_{\min}$ **Differenz zwischen größter und kleinster Klassengrenze**
Interquartil-abstand	**IQA = Q3 – Q1**	**IQA = Q3 – Q1**

	Einzelwerte (vorwiegend diskret)	Klassierte Werte
Mittlere absolute Abweichung **(vom Median)**	$d = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - \overline{x_M}\right\|$	$d = \sum_{i=1}^{k}\left[\left\|(x_i)_m - \overline{x_M}\right\| \cdot \frac{n_i}{n}\right] = \sum_{i=1}^{k}\left[\left\|(x_i)_m - \overline{x_M}\right\| \cdot p_i\right]$ *mit* $(x_i)_m$ *als Klassenmitte der Klasse i* *und* $p_i = \frac{n_i}{n} =$ relative Häufigkeit der Klasse i
Varianz und Standard-abweichung **(Basis: arithmetisches Mittel)**	$\left.\begin{array}{l} V(X) \overset{Definitions}{\underset{formel}{=}} \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \\ V(X) \overset{Rechen}{\underset{formel}{=}} \left[\frac{1}{n}\sum_{i=1}^{n}(x_i)^2\right] - \overline{x}^2 \end{array}\right\}$ $\Rightarrow \quad S(X) = \sqrt{V(X)}$	$\left.\begin{array}{l} V(X) \overset{Definitions}{\underset{formel}{=}} \sum_{i=1}^{n}\left[(x_i)_m - \overline{x}\right]^2 \cdot p_i \\ V(X) \overset{Rechen}{\underset{formel}{=}} \left[\sum_{i=1}^{n}(x_i)_m^{\ 2} \cdot p_i\right] - \overline{x}^2 \end{array}\right\}$ $\Rightarrow \quad S(X) = \sqrt{V(X)}$ *mit* $(x_i)_m$ *als Klassenmitte der Klasse i* *und* p_i als relative Häufigkeit der Klasse i
Quartile **Q1 / Q3** **(Basis: Median)**	$\overline{x_p} = \begin{cases} x_{[n\cdot p]+1} & \text{für } (\text{n}\cdot\text{p}) \text{ nicht ganzzahlig} \\ \frac{1}{2}\left(x_{n\cdot p} + x_{n\cdot p+1}\right) & \text{für } (\text{n}\cdot\text{p}) \text{ ganzzahlig} \end{cases}$ $p = 0{,}25 \rightarrow q1 \quad und \quad p = 0{,}75 \rightarrow q3$	$q_1 \Rightarrow \overline{x_{0,25}} = [a\,;b] \rightarrow \overline{x_{0,25}} = a + \frac{\Delta_i \cdot [0{,}25 - F(a)]}{p_i}$ $q_3 \Rightarrow \overline{x_{0,75}} = [a\,;b] \rightarrow \overline{x_{0,75}} = a + \frac{\Delta_i \cdot [0{,}75 - F(a)]}{p_i}$ $\Delta_i = Klassenbreite \quad p_i = rel.\ Häufigkeit$ $F(a) =$ kum. rel. Häufigkeit bis q_1/q_3 - Intervall $[a\,;b]$
Variations-koeffizient	$v_{arith\_Mittel} = \frac{S(X)}{\overline{x}} \quad oder \quad v_{Median} = \frac{S(X)}{\overline{x_M}}$	

Gesetzmäßigkeiten zur linearen Transformation des Erwartungswertes und der Varianz diskreter Zufallsvariablen

⇨ **Erwartungswert:**

Beh. 1: $E(a \cdot X + b) = a \cdot E(X) + b$

Beh. 2: $E(X+Y) = E(X) + E(Y)$

Beh. 3: $E(b) = b$

Beh. 4: Monotonie I: $Ist\ X \le Y\ und\ E(X)+E(Y)\ existieren \Rightarrow E(X) \le E(Y)$

Beh. 5: Monotonie II $|E(X)| \le E(|X|)$

Beh. 6: Dreiecksungleichung I und II:

$E(|X+Y|) \le E(|X|)+E(|Y|)$ $\quad$ $|E(X+Y)| \le |E(X)|+|E(Y)|$

Beh. 7: $E(X \cdot Y) \ne E(X) \cdot E(Y)$

⇨ **Varianz:**

Beh. 8: $V(a \cdot X) = a^2 \cdot V(X) \quad und \quad V(X+b) = V(X)$

Zusammengefasst: $V(a \cdot X + b) = a^2 \cdot V(X)$

Beh. 9: $V(aX \pm bY) = a^2 \cdot V(X) + b^2 \cdot V(Y) + 2ab \cdot Cov(X;Y)$

$$\xrightarrow{f\ddot{u}r\ a=b=1} V(X \pm Y) = V(X) + V(Y) + 2 \cdot Cov(X;Y)$$

$mit\ Sonderfall:\ Cov(X;Y) = 0 \Leftrightarrow X;Y\ unabh\ddot{a}ngig$

Allgemeine Darstellung für n Summanden:

$$V\left(\sum_{k=1}^{n} a_k X_k\right) = V(a_1 X_1 + ... + a_n X_n)$$
$$= a_1^2 \cdot V(X_1) + ... + a_n^2 \cdot V(X_n) + 2a_1a_2 \cdot Cov(X_1;X_2) + 2a_1a_3 \cdot Cov(X_1;X_3) + ... + 2a_{n-1}a_n \cdot Cov(X_{n-1};X_n)$$
$$= \sum_{k=1}^{n} a_k^2 \cdot V(X_k) + 2\sum_{\substack{k=1;\\k<t}}^{n} a_k a_t \cdot Cov(X_k;X_t)$$

Allgemeine Darstellung für n Summanden:

$Cov(X;Y) = 0 \Leftrightarrow X;Y\ unabh\ddot{a}ngig:$

$$V\left(\sum_{k=1}^{n} a_k X_k\right) = V(a_1 X_1 + ... + a_n X_n) = a_1^2 \cdot V(X_1) + ... + a_n^2 \cdot V(X_n) = a_k^2 \cdot V\left(\sum_{k=1}^{n} X_k\right)$$

Bew. 1:

$$E(a\cdot X+b) = \frac{1}{n}\sum_{i=1}^{n}(a\cdot x_i+b) = \frac{1}{n}\sum_{i=1}^{n}(a\cdot x_i)+\frac{1}{n}\sum_{i=1}^{n}b$$

$$= a\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i)+\frac{1}{n}\cdot n\cdot b = a\cdot\bar{x}+b = a\cdot E(X)+b$$

Bew. 2:

$$E(X+Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i+y_i) = \frac{1}{n}\sum_{i=1}^{n}(x_i)+\frac{1}{n}\sum_{i=1}^{n}y_i$$

$$= \bar{x}+\bar{y} = E(X)+E(Y)$$

Bew. 3: $$E(b) = \frac{1}{n}\sum_{i=1}^{n}b \underset{n-mal}{\overset{b+b+\ldots+b}{=}} \frac{1}{n}\cdot n\cdot b = b$$

Bew. 7:

$$E(X\cdot Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i) = \frac{1}{n}(x_1\cdot y_1+x_2\cdot y_2+\ldots+x_n\cdot y_n)$$

$$\neq \frac{1}{n}(x_1+x_2+\ldots+x_n)\cdot\frac{1}{n}(y_1+y_2+\ldots+y_n) = \frac{1}{n}\sum_{i=1}^{n}x_i\cdot\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$= E(X)\cdot E(Y)$$

Bew. 8:

$$V(a\cdot X) = \frac{1}{n}\sum_{i=1}^{n}(a\cdot x_i)^2 - (a\cdot\bar{x})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}a^2\cdot x_i^2 - a^2\cdot\bar{x}^2 = a^2\cdot\frac{1}{n}\sum_{i=1}^{n}x_i^2 - a^2\cdot\bar{x}^2$$

$$= a^2\cdot\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2\right) = a^2\cdot V(X)$$

$$V(X+b) = \frac{1}{n}\sum_{i=1}^{n}(x_i+b)^2 - (\bar{x}+b)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2+2\cdot x_i\cdot b+b^2\right) - \left(\bar{x}^2+2\cdot\bar{x}\cdot b+b^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + 2\cdot b\cdot\frac{1}{n}\sum_{i=1}^{n}x_i + b^2\cdot\frac{1}{n}\sum_{i=1}^{n}1 - \bar{x}^2-2\cdot\bar{x}\cdot b-b^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + 2\cdot b\cdot\bar{x} + b^2\cdot\frac{1}{n}\cdot n - \bar{x}^2 - 2\cdot\bar{x}\cdot b - b^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + 2\cdot b\cdot\bar{x} + b^2 - \bar{x}^2 - 2\cdot\bar{x}\cdot b - b^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2 = V(X)$$

Bew. 9:

$$V(X\pm Y) = V(X)+V(Y)+2\cdot Cov(X;Y)$$

$$Sonderfall: \quad Cov(X;Y)=0 \quad \Leftrightarrow \quad X;Y\ unabhängig$$

$$V(X\pm Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i\pm y_i)^2 - (\bar{x}\pm\bar{y})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2\pm 2\cdot x_i\cdot y_i+y_i^2\right) - \left(\bar{x}^2\pm 2\cdot\bar{x}\cdot\bar{y}+\bar{y}^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2\pm 2\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i)+\frac{1}{n}\sum_{i=1}^{n}y_i^2 - \bar{x}^2\mp 2\cdot\bar{x}\cdot\bar{y}-\bar{y}^2$$

$$= \left[\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2\right]+\left[\frac{1}{n}\sum_{i=1}^{n}y_i^2-\bar{y}^2\right] \pm 2\cdot\frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i)\mp 2\cdot\bar{x}\cdot\bar{y}$$

$$= V(X) + V(Y) + 2\cdot\left[\frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i)-\bar{x}\cdot\bar{y}\right]$$

$$= V(X) + V(Y) + 2\cdot\left[E(x\cdot y)-E(X)\cdot E(Y)\right]$$

$$= V(X) + V(Y) + 2\cdot Cov(X;Y)$$

Diskrete Zufallsvariable - Varianz (einer Grundgesamtheit):

⇨ **Zusammenhang zwischen Definition und Berechnung**

Beh.: $$V(X) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 \overset{Berechnung}{=} \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

Bew. des Übergangs zwischen Definitions- und Berechnungsformel:

$$V(X) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2-2x_i\bar{x}+\bar{x}^2\right)$$

$$V(X) \overset{Binom}{=} \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}\left(-2x_i\bar{x}\right) + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2$$

$$V(X) \underset{\substack{\text{Summenzeichen}\\ \text{"ausklammern"}}}{\overset{\text{Konstanten aus}}{=}} \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\cdot\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}\bar{x}^2\cdot\sum_{i=1}^{n}1$$

$$V(X) \underset{\sum_{i=1}^{n}1=n}{\overset{\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}}{=}} \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}\cdot\bar{x} + \frac{1}{n}\bar{x}^2\cdot n = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\bar{x}^2 + \bar{x}^2$$

$$V(X) \overset{Berechnung}{=} \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

Begriff der Kovarianz / Co-Varianz (einer Grundgesamtheit)

Ausgang: Varianz $$V(X) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 \overset{Berechnung}{=} \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

Kovarianz:

$$Cov(X,Y) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right) \overset{Berechnung}{=} \frac{1}{n}\sum_{i=1}^{n}\left(x_i\cdot y_i\right) - \bar{x}\cdot\bar{y}$$

Herleitung:

$$Cov(X,Y) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(x_iy_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(x_i\bar{y}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}y_i\right) + \frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}\bar{y}\right)$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(x_iy_i\right) - \bar{y}\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}x_i - \bar{x}\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}y_i + \bar{x}\bar{y}\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}1$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(x_iy_i\right) - \bar{y}\cdot\bar{x} - \bar{x}\cdot\bar{y} + \bar{x}\bar{y}\cdot\frac{1}{n}\cdot n\cdot 1 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i\cdot y_i\right) - \bar{x}\cdot\bar{y}$$

Abgewandelte Variablen-Schreibweise (Erwartungswert anstelle von Mittelwert):

$$Cov(X,Y) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y)$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i)-\frac{1}{n}\sum_{i=1}^{n}(x_i\mu_y)-\frac{1}{n}\sum_{i=1}^{n}(\mu_x y_i)+\frac{1}{n}\sum_{i=1}^{n}(\mu_x\mu_y)$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i)-\mu_y\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}x_i-\mu_x\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}y_i+\mu_x\mu_y\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}1$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i)-\mu_y\cdot\mu_x-\mu_x\cdot\mu_y+\mu_x\mu_y\cdot\frac{1}{n}\cdot n\cdot 1$$

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i)-\mu_x\cdot\mu_y$$

Verschiebungssatz zur Berechnung:

$$Cov(X,Y) \overset{Def.}{=} \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_x)(y_i-\mu_y) = E\left[(X-E(X))\cdot(Y-E(Y))\right]$$

$$Cov(X,Y) = E\left[X\cdot Y-X\cdot E(Y)-E(X)\cdot Y+E(X)\cdot E(Y)\right]$$

$$Cov(X,Y) = E(X\cdot Y)-E(X)\cdot E(Y)-E(X)\cdot E(Y)+E(X)\cdot E(Y)$$

$$Cov(X,Y) = E(X\cdot Y)-E(X)\cdot E(Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i\cdot y_i)-\mu_x\cdot\mu_y$$

Kovarianz und Varianz im Vergleich:

$$Cov(X,Y) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y}) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i)-\bar{x}\cdot\bar{y}$$

$$Cov(X,Y) = \frac{1}{n}\cdot\sum_{i=1}^{n}\left[x_i-E(X)\right]\cdot\left[y_i-E(Y)\right] = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i)-E(X)\cdot E(Y)$$

$$Cov(X,Y) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\mu_x)\cdot(y_i-\mu_y) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i)-\mu_x\cdot\mu_y$$

$$V(X) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2 = \left[\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i)^2\right]-\bar{x}^2$$

$$V(X) = \frac{1}{n}\cdot\sum_{i=1}^{n}\left[x_i-E(X)\right]^2 = \left[\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i)^2\right]-E(X)^2$$

$$V(X) = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\mu_x)^2 = \left[\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i)^2\right]-\mu_x^2$$

Gini-Koeffizient:

Berechung der Flächen mit relativen kumulierten Häufigkeiten:

$$Gini(GK) = \frac{\text{Konzentrationsfläche K}}{\text{maximale Konzentrationsfläche } K_{\max}}$$

$$Gini(GK) = \frac{Fläche\ zw.\ Gleichvert.\ und\ Lorenzkurve}{Fläche\ unter\ Gleichverteilung}$$

Anmerkung: Als *„Konzentrationsfläche“ K* bezeichnet man die Fläche, zwischen der Gleichverteilungsdiagonalen und der LORENZ-Kurve: $0 \leq K \leq 1/2$.

$$Wegen \quad K_{\max} = \frac{1}{2} - \frac{1}{2n} \quad lautet\ der\ normierte\ Gini-Koeffizient:$$

$$norm.\ Gini = K \cdot \frac{2n}{n-1}$$

Der Wertebereich des Gini-Koeffizienten liegt zwischen 0 (= Gleichverteilung) und 1 (= vollständige Konzentration auf einen Merkmalsträger).

$$\left. GK = \frac{Fläche\ zw.\ Gleichvert.\ und\ Lorenzkurve}{0,5} = \frac{\to 0}{0,5} \to 0 \right\}$$

$$\to \text{nahezu Gleichverteilung}$$

$$\left. GK = \frac{Fläche\ zw.\ Gleichvert.\ und\ Lorenzkurve}{0,5} = \frac{\to 0,5}{0,5} \to 1 \right\}$$

$$\to vollständige\ Konzentration$$

$$keine\ Konzentration \Leftrightarrow GK \in [0;0,3]$$

$$(mäßige)\ Konzentration \Leftrightarrow GK \in]0,3;0,7[$$

$$vollständige\ Konzentration \Leftrightarrow GK \in [0,7;1]$$

Erläuterung zur Berechnung des Gini-Koeffizienten

Fläche unter der Gleichverteilungsgeraden: $A_{\Delta} = \frac{1}{2} \cdot g \cdot h = \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2}$

Fläche unterhalb der Lorenzkurve: Dreiecke $A_{Dreieck} = \frac{1}{2} \cdot g \cdot h$

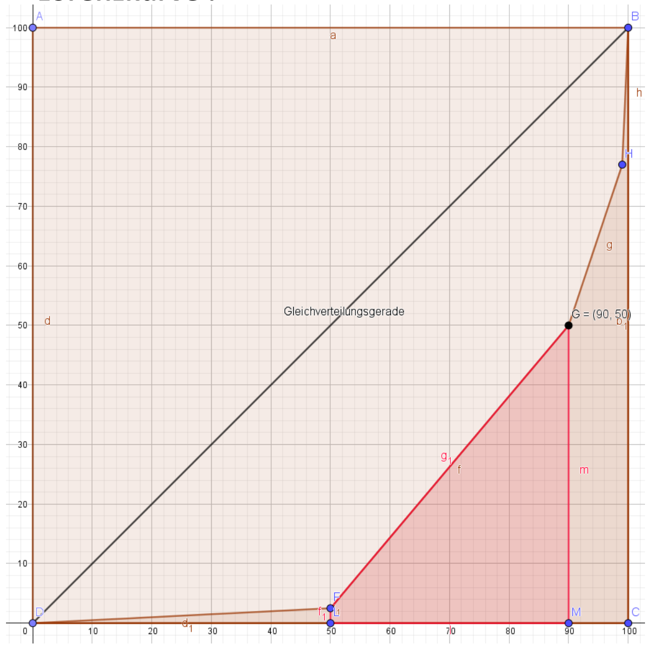
und Trapeze $A_{Trapez} = \frac{1}{2} \cdot (h_i + h_{i+1}) \cdot \Delta x_i$

Berechnung des Gini-Koeffizienten

Maximale Konzentrationsfläche => Fläche unterhalb der Gleichverteilungsgerade

$$A_\Delta \quad = \quad \frac{1}{2}\cdot g\cdot h \quad = \quad \frac{1}{2}\cdot 1\cdot 1 \quad = \quad \frac{1}{2}$$

Lorenzkurve :



Fläche unterhalb der Lorenzkurve:

$$A_1 = \frac{1}{2}\cdot g\cdot h \rightarrow \frac{1}{2}\cdot 0,5\cdot 0,025 = 0,00625$$

$$A_2 = \frac{1}{2}\cdot (h_1 + h_2)\cdot g \rightarrow \frac{1}{2}\cdot (0,025 + 0,5)\cdot 0,4 = 0,105$$

$$A_3 = \frac{1}{2}\cdot (h_1 + h_2)\cdot g \rightarrow \frac{1}{2}\cdot (0,5 + 0,77)\cdot 0,09 = 0,05715$$

$$A_4 = \frac{1}{2}\cdot (h_1 + h_2)\cdot g \rightarrow \frac{1}{2}\cdot (0,77 + 1)\cdot 0,01 = 0,00885$$

Fläche unterhalb der Lorenzkurve:
A1+A2+A3+A4 = 0,17725

Fläche zwischen Gleichverteilung und Lorenzkurve:
0,5 - 0,17725 = 0,32275

$$GK \quad = \quad \frac{\text{Fläche zw. Gleichvert. und Lorenzkurve}}{\text{Fläche unter Gleichverteilung}}$$

$$GK \quad = \quad \frac{0,5 - 0,17725}{0,5} \quad = \quad \frac{0,32275}{0,5} \quad = \quad 0,6455$$

Auswertung:

keine Konzentration $\Leftrightarrow \quad GK \in [0;0,3]$

(mäßige) Konzentration $\Leftrightarrow \quad GK \in \,]0,3;0,7[$

vollständige Konzentration $\Leftrightarrow \quad GK \in [0,7;1]$

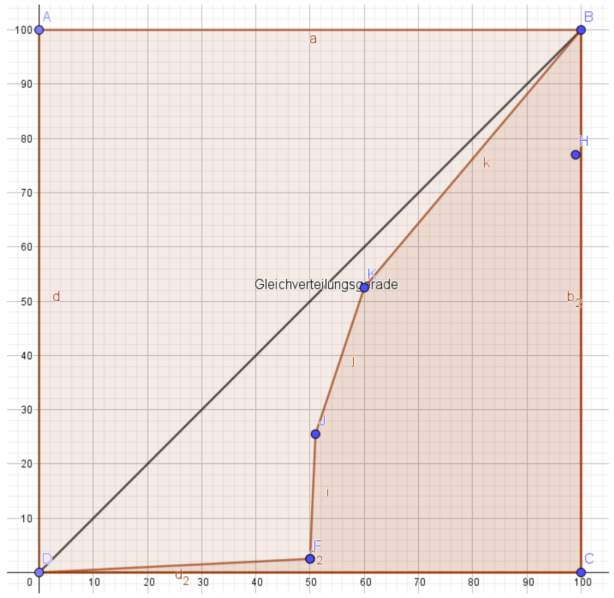
Grundlage: Verteilungstabelle

Bevölkerung			Vermögen			$\frac{\Delta y}{\Delta x}$ = *Steigung* → *Sekantensteigung* → *Differenzenquotient*
Prozentanteil pro Klasse	rel. Anteil	kumulierte Prozentwerte	Prozentanteil pro Klasse	rel. Anteil	kumulierte Prozentwerte	
0,00	0,00	0,00	0,00	0,00	0,000	
50,00	0,50	0,50	2,50	0,03	0,025	**0,05**
40,00	0,40	0,90	47,50	0,48	0,500	**1,1875**
9,00	0,09	0,99	27,00	0,27	0,770	**3**
1,00	0,01	1,00	23,00	0,23	1,000	**23**
100,00	1,00	x-Achse	100,00	1,00	y-Achse	

$$\left.\begin{array}{l}\text{Ordnen der Werte: (gemäß) Differenzenquotient}\\ \frac{\Delta y}{\Delta x} = (\text{durchschnittliche})\ \text{Steigung} = \text{Sekantensteigung}\end{array}\right\}\ \text{"von klein nach groß"}$$

richtige Reihenfolge der Anordnung

Lorenzkurve (fehlerhaft):



Verletzung des Ordnungsprinzips:

Ordnen der Werte gemäß Differenzenquotient:
$\rightarrow$ *"von klein nach groß"*

Werte:

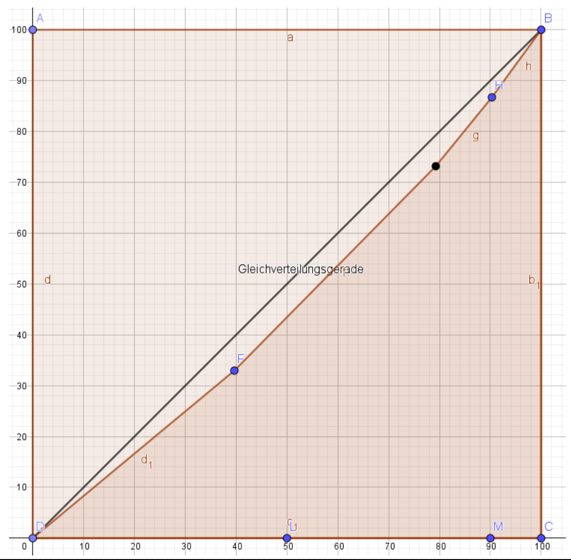
$$\frac{\Delta y}{\Delta x} = (\text{durchschn.})\ \text{Steigung} \leftrightarrow \text{Sekantensteigung}$$

Darstellung bei nicht berücksichtigtem Ordnungssystem

Bevölkerung			Vermögen			
Prozentanteil pro Klasse	rel. Anteil	kumulierte Prozentwerte	Prozentanteil pro Klasse	rel. Anteil	kumulierte Prozentwerte	$\frac{\Delta y}{\Delta x}$ = Steigung $\rightarrow$ Sekantensteigung $\rightarrow$ Differenzenquotient
0,00	0,00	0,00	0,00	0,000	0,000	
50,00	0,50	0,50	2,50	0,025	0,025	**0,05**
1,00	0,01	0,51	23,00	0,230	0,255	**23**
9,00	0,09	0,60	27,00	0,270	0,525	**3**
40,00	0,40	1,00	47,50	0,475	1,000	**1,1875**
100,00	1,00	x-Achse	100,00	1,000	y-Achse	

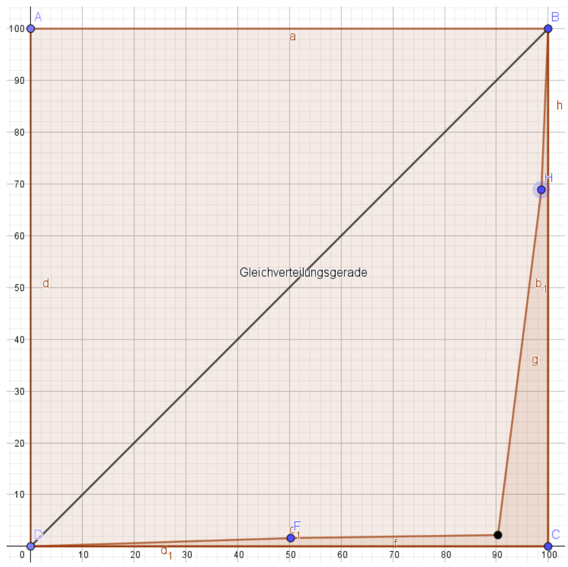
geordnet nach Prozentanteilen $\longleftrightarrow$ **fehlerhafte Reihenfolge der Anordnung**

Sonderformen:



$$GK = \frac{\text{Fläche zw. Gleichvert. und Lorenzkurve}}{0,5} = \frac{\rightarrow 0}{0,5} \rightarrow 0$$

$\rightarrow GK \rightarrow 0 \rightarrow$ nahezu Gleichverteilung



$$GK = \frac{\text{Fläche zw. Gleichvert. und Lorenzkurve}}{0,5} = \frac{\rightarrow 0,5}{0,5} \rightarrow 1$$

$\rightarrow GK \rightarrow 1 \rightarrow$ *vollständige Konzentration*

Preisindizes

nach Laspeyres: $$L_P = \frac{\sum p_{1i} \cdot q_{0i}}{\sum p_{0i} \cdot q_{0i}} = \sum \frac{p_{1i}}{p_{0i}} \cdot \frac{p_{0i} \cdot q_{0i}}{\sum p_{0i} \cdot q_{0i}} \quad mit\ i \in N$$

nach Paasche:

$$P_P = \frac{\sum p_{1i} \cdot q_{1i}}{\sum p_{0i} \cdot q_{1i}} = \sum \frac{p_{1i}}{p_{0i}} \cdot \frac{p_{0i} \cdot q_{1i}}{\sum p_{0i} \cdot q_{1i}} \quad mit\ i \in N$$

nach Fisher: $$F_P = \sqrt{L_P \cdot P_P} \qquad geometr.\ Mittel$$

Anmerkung: Geometrisches Mittel / Geometrischer Mittelwert

⇨ **Mittelwert aus einer Produktfolge**

Geometr. Mittel: Mittelwert einer Produktfolge

$$g_{MW} = \sqrt[n]{\prod_{i=1}^{n} q_i} = \sqrt[n]{q_1 \cdot q_2 \cdot \ldots \cdot q_n} \qquad mit\ q_i = 1 + \frac{p_i}{100}$$

Effektive Verzinsung einer tesaurierenden

(Zinsen werden kapitalisiert, d.h. dem Kapital hinzugerechnet) Anlage

=> Effektive Verzinsung bei Zinseszinseffekt

$$i_{eff} = \sqrt[n]{\prod_{i=1}^{n} q_i} - 1 = \sqrt[n]{q_1 \cdot q_2 \cdot \ldots \cdot q_n} - 1 = g_{MW} - 1$$

$$p_{eff} = (g_{MW} - 1) \cdot 100 = i_{eff} \cdot 100$$

Indizierung mit neuem Basisjahr

$$I_{Basisjahr\_neu;\, Zieljahr} = \frac{I_{Basisjahr\_alt;\, Zieljahr}}{I_{Basisjahr\_alt;\, Basisjahr\_neu}}$$

Beispiel: ***Index von 2000 auf der Basis von 2005 soll auf die Basis 2010 umgerechnet werden.***

$$I_{2010;\, 2000} = \frac{I_{2005;\, 2000}}{I_{2005;\, 2010}}$$

Verkettung

$$I_{Basisjahr\_alt;\, Zieljahr} = \frac{I_{Basisjahr\_neu;\, Basisjahr\_alt}}{I_{Basisjahr\_neu;\, Basisjahr\_neu}} \cdot I_{Basisjahr\_neu;\, Zieljahr}$$

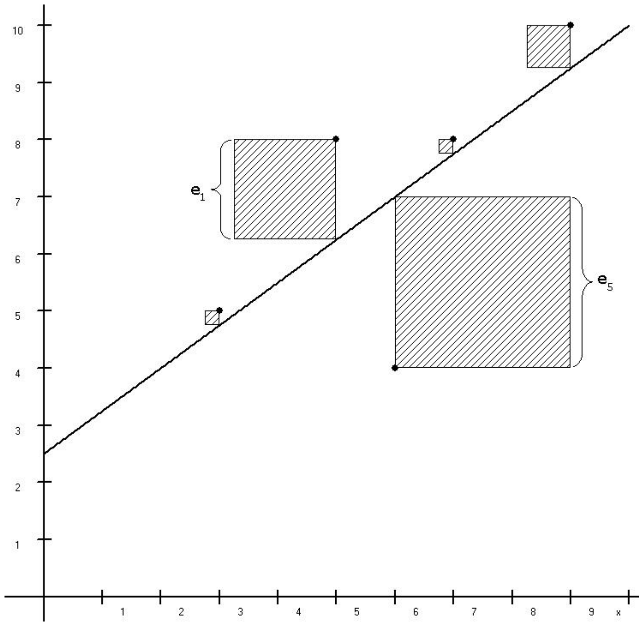
Beispiel: ***Indexwert von 2013 auf der Basis von 2010 soll auf die Basis 2005 umgerechnet werden.***

$$I_{2005;\, 2013} = \frac{I_{2010;\, 2005}}{I_{2010;\, 2010}} \cdot I_{2010;\, 2013}$$

Lineare Regression und Korrelation

Anwendung der partiellen Differentiation in der Statistik: Regressionsgerade

Gesucht: ***Eine Gerade, die zu allen Punkten den kleinsten Abstand hat; d.h. die Differenzen e_i müssen insgesamt möglichst klein sein (Optimierungsproblem)***



Ansatz: $$\sum_{i=1}^{n}(e_i)^2 \quad \rightarrow \quad \min.$$

$$\textit{Gesucht ist Gerade:} \quad y \;=\; b_0 + b_1 x$$

$\textit{Es gilt:}$

$(x_i \mid y_i)$ sind die Punkte des Streudiagramms

$\textit{ebenso gilt:}$

$(x_i \mid y_i)$ sind die Punkte auf der Geraden

daraus folgt: $\xrightarrow{y_i \leftrightarrow y_i}$ $\quad y_i \;=\; b_0 + b_1 x_i$

$\textit{und} \quad \Delta y \;=\; y_i - y_i \; [= \textit{Seitenlänge des Quadrats}]$

$$\rightarrow \sum_{i=1}^{n}(e_i)^2 \;=\; \sum_{i=1}^{n}\left(y_i - y_i\right)^2 \;=\; \sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2$$

$$E(b_1;b_0) \;=\; \sum_{i=1}^{n}\left(y_i^2 - y_i b_0 - y_i b_1 x_i - y_i b_0 + b_0^2 + b_0 b_1 x_i - y_i b_1 x_i + b_0 b_1 x_i + b_1^2 x_i^2\right)$$

$$E(b_1;b_0) \;=\; \sum_{i=1}^{n}\left(y_i^2 - 2y_i b_0 - 2y_i b_1 x_i + b_0^2 + 2b_0 b_1 x_i + b_1^2 x_i^2\right)$$

$$E(b_1;b_0) \;=\; \sum_{i=1}^{n} y_i^2 - 2b_0\sum_{i=1}^{n} y_i - 2b_1\sum_{i=1}^{n} x_i y_i + b_0^2\sum_{i=1}^{n} 1 + 2b_0 b_1\sum_{i=1}^{n} x_i + b_1^2\sum_{i=1}^{n} x_i^2$$

$$E(b_1;b_0) \;=\; \sum_{i=1}^{n} y_i^2 - 2b_0\sum_{i=1}^{n} y_i - 2b_1\sum_{i=1}^{n} x_i y_i + n b_0^2 + 2b_0 b_1\sum_{i=1}^{n} x_i + b_1^2\sum_{i=1}^{n} x_i^2$$

$$\frac{\partial E(b_1;b_0)}{\partial b_1} \;=\; -2\sum_{i=1}^{n} x_i y_i + 2b_0\sum_{i=1}^{n} x_i + 2b_1\sum_{i=1}^{n} x_i^2 \;=\; 0$$

$$\frac{\partial E(b_1;b_0)}{\partial b_0} \;=\; -2\sum_{i=1}^{n} y_i + 2nb_0 + 2b_1\sum_{i=1}^{n} x_i \;=\; 0$$

Lösung der Gleichungen der partiellen Ableitungen:

$$\text{I)}\quad -2\sum_{i=1}^{n} x_i y_i + 2b_0 \sum_{i=1}^{n} x_i + 2b_1 \sum_{i=1}^{n} x_i^2 \quad = \quad 0$$

$$\text{II)}\quad -2\sum_{i=1}^{n} y_i + 2nb_0 + 2b_1 \sum_{i=1}^{n} x_i \quad = \quad 0$$

Es gilt:

$$\sum_{i=1}^{n} x_i \quad = \quad n\cdot\overline{x} \qquad und \qquad \sum_{i=1}^{n} y_i \quad = \quad n\cdot\overline{y} \quad \text{(arithmetisches Mittel)}$$

$eingesetzt:$

$$\text{I)}\quad -2\sum_{i=1}^{n} x_i y_i + 2b_0 n\cdot\overline{x} + 2b_1 \sum_{i=1}^{n} x_i^2 \quad = \quad 0$$

$$\text{II)}\quad -2n\cdot\overline{y} + 2nb_0 + 2b_1 n\cdot\overline{x} \quad = \quad 0 \quad \xrightarrow{:(2n)} \quad -\overline{y} + b_0 + b_1\overline{x} \quad = \quad 0$$

$$\rightarrow \quad b_0 \quad = \quad \overline{y} - b_1\overline{x}$$

$$\xrightarrow{\text{II)}\ \ eingesetzt\ in\ \ \text{I)}} \text{I)}\quad -2\sum_{i=1}^{n} x_i y_i + 2\left(\overline{y} - b_1\overline{x}\right) n\cdot\overline{x} + 2b_1 \sum_{i=1}^{n} x_i^2 \quad = \quad 0$$

$$\xrightarrow{:2} \text{I)}\quad -\sum_{i=1}^{n} x_i y_i + n\cdot\overline{x}\cdot\overline{y} - b_1 n\cdot\left(\overline{x}\right)^2 + b_1 \sum_{i=1}^{n} x_i^2 \quad = \quad 0$$

$$\rightarrow \quad -b_1 n\cdot\left(\overline{x}\right)^2 + b_1 \sum_{i=1}^{n} x_i^2 \quad = \quad \sum_{i=1}^{n} x_i y_i - n\cdot\overline{x}\cdot\overline{y}$$

$$\rightarrow \quad b_1\left[\sum_{i=1}^{n} x_i^2 - n\cdot\left(\overline{x}\right)^2\right] \quad = \quad \sum_{i=1}^{n} x_i y_i - n\cdot\overline{x}\cdot\overline{y}$$

$$\rightarrow \quad b_1 \quad = \quad \frac{\sum_{i=1}^{n} x_i y_i - n\cdot\overline{x}\cdot\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\cdot\left(\overline{x}\right)^2} \quad \underset{\cdot\frac{1}{n}}{\overset{\cdot\frac{1}{n}}{=}} \quad \frac{\frac{1}{n}\cdot\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y}}{\frac{1}{n}\cdot\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2} \quad = \quad \frac{Cov(X,Y)}{V(X)}$$

Lineare Regression und Korrelation

Ansatz: $y = b_0 + b_1 x \qquad \mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i$

$$b_1 = \frac{\sigma_{xy}}{\sigma_{x^2}} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\cdot\mu_x\cdot\mu_y}{\sum_{i=1}^{n}(x_i^2) - n\cdot\mu_x^2} \overset{\cdot\frac{1}{n}}{\underset{\cdot\frac{1}{n}}{=}} \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - \mu_x\cdot\mu_y}{\frac{1}{n}\sum_{i=1}^{n}(x_i^2) - \mu_x^2}$$

$$b_1 = \frac{\sigma_{xy}}{\sigma_{x^2}} = \frac{Kov(X,Y)}{V(X)}$$

$$b_0 = \mu_y - b_1\cdot\mu_x$$

Korrelation nach Brevais-Pearson:

$$r \overset{Definition}{=} \frac{\sum_{i=1}^{n}(x_i-\mu_x)\cdot(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\mu_x)^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\mu_y)^2}} \overset{Berechnung}{=} \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i) - \mu_x\cdot\mu_y}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^2) - \mu_x^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^2) - \mu_y^2}}$$

$$r = \frac{Kov(X,Y)}{S(X)\cdot S(Y)} = \frac{\sigma(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

Rangkorrelation nach Spearman $\quad r_S = 1 - \frac{6\sum_{i=1}^{n}(d_i)^2}{n(n^2-1)}$

$r = [-1;1]$ möglicher Wertebereich

$r = [-1;-0,7[$ sehr starke negative Korrelation

$r = [-0,7;-0,3[$ starke negative Korrelation

$r = [-0,3;0,3[$ keine Korrelation (Punktwolke)

$r = [0,3;0,7[$ starke positive Korrelation

$r = [0,7;1]$ sehr starke positive Korrelation

Korrelation nach Brevais-Pearson – Herleitung:

$$r = \frac{\sum_{i=1}^{n}(x_i-\mu_x)\cdot(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i-\mu_x)^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\mu_y)^2}} = \frac{\sum_{i=1}^{n}(x_iy_i)-n\cdot\mu_x\cdot\mu_y}{\sqrt{\sum_{i=1}^{n}(x_i^2)-n\cdot\mu_x^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i^2)-n\cdot\mu_y^2}}$$

$$r \overset{\cdot\frac{1}{n}}{\underset{\cdot\frac{1}{n}}{=}} \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\mu_x\cdot\mu_y}{\frac{1}{n}\cdot\sqrt{\sum_{i=1}^{n}(x_i^2)-n\cdot\mu_x^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i^2)-n\cdot\mu_y^2}} = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\mu_x\cdot\mu_y}{\sqrt{\frac{1}{n}}\cdot\sqrt{\frac{1}{n}}\cdot\sqrt{\sum_{i=1}^{n}(x_i^2)-n\cdot\mu_x^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i^2)-n\cdot\mu_y^2}}$$

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\mu_x\cdot\mu_y}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^2)-\mu_x^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^2)-\mu_y^2}} = \frac{Kov(X,Y)}{S(X)\cdot S(Y)} = \frac{\sigma(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

Rangkorrelation nach Spearman – Herleitung:

Ausgangspunkt: Korrelation nach Brevais-Pearson

Problem: Die Merkmale sind nicht metrisch skaliert sondern ordinal skaliert

Lösungsansatz:
Umordnung der Merkmale in „metrische-skalierte" Ränge als Alternativmerkmale, damit die ordinale Skalierungsidee erhalten bleibt und gleichzeitig eine analytische Betrachtung und Verrechnung im Sinne einer statistischen Merkmalsauswertung (MW- und Varianzberechnung) ermöglicht werden.

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\bar{x}\cdot\bar{y}}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^2)-(\bar{x})^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^2)-(\bar{y})^2}}$$

Bemerkungen: Da hier für x und y Ränge/Rangfolgen von 1, ..., n betrachtet werden, gilt

$$\left.\begin{array}{l}\sum_{i=1}^{n}x_i = 1+2+3+...+n = \frac{n\cdot(n+1)}{2} \xrightarrow{Mittelwert} \frac{1}{n}\sum_{i=1}^{n}x_i = \frac{1}{n}\cdot(1+2+3+...+n) = \frac{1}{n}\cdot\frac{n\cdot(n+1)}{2} = \frac{n+1}{2} = \bar{x} \\ \sum_{i=1}^{n}y_i = 1+2+3+...+n = \frac{n\cdot(n+1)}{2} \xrightarrow{Mittelwert} \frac{1}{n}\sum_{i=1}^{n}y_i = \frac{1}{n}\cdot(1+2+3+...+n) = \frac{1}{n}\cdot\frac{n\cdot(n+1)}{2} = \frac{n+1}{2} = \bar{y}\end{array}\right\} \bar{x}=\bar{y}$$

zudem:

$$\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}y_i = \sum_{i=1}^{n}i = \frac{n\cdot(n+1)}{2} \quad und \quad \sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}y_i^2 = \sum_{i=1}^{n}i^2 = \frac{n\cdot(n+1)\cdot(2n+1)}{6}$$

Für Zähler folgt:

$$Z = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\bar{x}\cdot\bar{y} \overset{\bar{x}=\bar{y}}{\underset{\bar{x}=\frac{n+1}{2}}{=}} \frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\frac{n+1}{2}\cdot\frac{n+1}{2} = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\frac{(n+1)^2}{4}$$

Ergänzung für $\sum_{i=1}^{n}(x_i y_i)$: Binomischer Lehrsatz & Rechenregeln für Summen

$$\sum_{i=1}^{n}(x_i-y_i)^2 = \sum_{i=1}^{n}\left(x_i^2-2x_iy_i+y_i^2\right) = \sum_{i=1}^{n}x_i^2 - 2\sum_{i=1}^{n}(x_iy_i) + \sum_{i=1}^{n}y_i^2 \overset{\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}y_i^2}{\underset{\sum_{i=1}^{n}x_i^2=\frac{n\cdot(n+1)\cdot(2n+1)}{6}}{=}}$$

$$\sum_{i=1}^{n}(x_i-y_i)^2 = \frac{n\cdot(n+1)\cdot(2n+1)}{6} - 2\sum_{i=1}^{n}(x_iy_i) + \frac{n\cdot(n+1)\cdot(2n+1)}{6} = \frac{n\cdot(n+1)\cdot(2n+1)}{3} - 2\sum_{i=1}^{n}(x_iy_i)$$

$$\xrightarrow[\sum_{i=1}^{n}(x_iy_i)]{\textit{Auflösen nach}} \sum_{i=1}^{n}(x_iy_i) = \frac{n\cdot(n+1)\cdot(2n+1)}{6} - \frac{1}{2}\cdot\sum_{i=1}^{n}(x_i-y_i)^2$$

Für Nenner folgt:

$$N = \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i^2\right)-\left(\bar{x}\right)^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(y_i^2\right)-\left(\bar{y}\right)^2} \overset{\bar{x}=\bar{y}}{\underset{\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}y_i^2}{=}} \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i^2\right)-\left(\bar{x}\right)^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-\left(\bar{x}\right)^2}$$

$$N \overset{\textit{Wurzel}}{\underset{\textit{zusamemnfassen}}{=}} \frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i^2\right)-\left(\bar{x}\right)^2 \overset{\bar{x}=\frac{n+1}{2}}{\underset{\sum_{i=1}^{n}\left(x_i^2\right)=\frac{n\cdot(n+1)\cdot(2n+1)}{6}}{=}} \frac{1}{n}\cdot\frac{n\cdot(n+1)\cdot(2n+1)}{6}-\frac{1}{4}(n+1)^2 = \frac{(n+1)\cdot(2n+1)}{6}-\frac{(n+1)^2}{4}$$

$$N = \frac{2n^2+3n+1}{6}-\frac{n^2+2n+1}{4} = \frac{1}{3}n^2+\frac{1}{2}n+\frac{1}{6}-\frac{1}{4}n^2-\frac{1}{2}n-\frac{1}{4} = \frac{1}{12}n^2-\frac{1}{12} = \frac{1}{12}\left(n^2-1\right)$$

Daraus folgt:

$$r_S = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_iy_i)-\frac{(n+1)^2}{4}}{\frac{1}{12}\left(n^2-1\right)} = \frac{\frac{1}{n}\cdot\left[\frac{n\cdot(n+1)\cdot(2n+1)}{6}-\frac{1}{2}\cdot\sum_{i=1}^{n}(x_i-y_i)^2\right]-\frac{(n+1)^2}{4}}{\frac{1}{12}\left(n^2-1\right)}$$

$$r_S = \frac{\frac{(n+1)\cdot(2n+1)}{6}-\frac{1}{2n}\cdot\sum_{i=1}^{n}(x_i-y_i)^2-\frac{(n+1)^2}{4}}{\frac{1}{12}\left(n^2-1\right)} = \frac{\frac{\left(2n^2+3n+1\right)}{6}-\frac{1}{2n}\cdot\sum_{i=1}^{n}(x_i-y_i)^2-\frac{\left(n^2+2n+1\right)}{4}}{\frac{1}{12}\left(n^2-1\right)}$$

$$r_S = \frac{\frac{1}{12}\cdot\left[4n^2+6n+2-\frac{6}{n}\cdot\sum_{i=1}^{n}(x_i-y_i)^2-3n^2-6n-3\right]}{\frac{1}{12}\cdot\left(n^2-1\right)} = \frac{n^2-1-\frac{6}{n}\cdot\sum_{i=1}^{n}(x_i-y_i)^2}{\left(n^2-1\right)}$$

$$r_S = \frac{n^2-1}{\left(n^2-1\right)}-\frac{\frac{6}{n}\cdot\sum_{i=1}^{n}(x_i-y_i)^2}{\left(n^2-1\right)} = 1-\frac{6\cdot\sum_{i=1}^{n}(x_i-y_i)^2}{n\cdot\left(n^2-1\right)} \overset{x_i-y_i=d_i}{=} 1-\frac{6\cdot\sum_{i=1}^{n}d_i^2}{n\left(n^2-1\right)}$$

Stochastik

Informationen zu Erwartungswert, Varianz und Standardabweichung

Erwartungswert einer diskreten Zufallsvariablen X mit $X = \{x_1, x_2, ..., x_n\}$ **:**

$$\mu = E(X) = \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Varianz einer diskreten Zufallsvariablen:

$$\sigma^2 = V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot P(X = x_i) = \sum_{i=1}^{n} x_i^{\,2} \cdot P(X = x_i) - \mu^2$$

Standardabweichung einer diskreten Zufallsvariablen:

$$\sigma = S(X) = \sqrt{V(X)}$$

Pfadregeln
Vierfeldertafel + Baumdiagramm

Stochastische Unabhängigkeit:

$$A, B \text{ sind stochastisch unabhängig} \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$$

Bedingte Wahrscheinlichkeit:

$$P_B(A) = \frac{P(A \cap B)}{P(B)}$$ W'keit für A unter der Bedingung B (B ist bereits eingetreten)

Totale Wahrscheinlichkeit:

$$P(B) = \sum_{i=1}^{k} \left[P_{A_i}(B) \cdot P(A_i) \right]$$

Satz von Bayes:

$$P_B(A_j) = \frac{P_{A_j}(B) \cdot P(A_j)}{\sum_{i=1}^{k} \left[P_{A_i}(B) \cdot P(A_i) \right]}$$

Verteilungsfunktionen mit diskreten und stetigen Zufallsvariablen

Diskret:

Hypergeometrische Verteilung – Ziehen ohne Zurücklegen – ist abhängig von drei Parametern:

- ⇨ **N Anzahl der Elemente der Grundgesamtheit**
- ⇨ **M Anzahl der Elemente mit einer bestimmten Eigenschaft aus dieser Grundmenge**
- ⇨ **n Anzahl der Elemente der Stichprobe**
- ⇨ **k Anzahl der Elemente mit der der bestimmten Eigenschaft aus der Stichprobe**

Wahrscheinlichkeitsfunktion: $$H(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad mit \quad 0 \le k \le n \quad und \quad M \le N \quad und \quad n \le N \quad und \quad k \le M$$

Verteilungsfunktion: $$H(k_1 \le X \le k_2) = \sum_{X=k_1}^{k_2} \frac{\binom{M}{X}\binom{N-M}{n-X}}{\binom{N}{n}}$$

Erwartungswert: $$\mu_{HV}(X) = \sum_{k=0}^{n} k \cdot H(X=k) = n \cdot \frac{M}{N}$$

Varianz: $$\sigma_{HV}{}^2(X) = \sum_{k=0}^{n} k^2 \cdot H(X=k) - \mu_{HV}{}^2 = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$$

Binomialverteilung – Bernoulliexperiment/Bernoullikette (Ziehen mit Zurücklegen):

Wahrscheinlichkeitsfunktion: $$B_{n;p}(X=k) = \binom{n}{k} p^k \cdot (1-p)^{n-k} \quad mit \quad k \le n$$

Verteilungsfunktion: $$B_{n;p}(k_1 \le X \le k_2) = \sum_{X=k_1}^{k_2} \binom{n}{X} p^X \cdot (1-p)^{n-X}$$

Erwartungswert: $$\mu_{BV}(X) = \sum_{k=0}^{n} k \cdot B_{n;p}(X=k) = n \cdot p$$

Varianz: $$\sigma_{BV}{}^2(X) = \sum_{k=0}^{n} k^2 \cdot B_{n;p}(X=k) - \mu_{BV}{}^2 = n \cdot p \cdot (1-p) = \mu \cdot (1-p)$$

Poissonverteilung – Anwendung bei n groß und p $\leq$ 0,05:

Wahrscheinlichkeitsfunktion: $$P(X=k) = \frac{\mu^k}{k!}e^{-\mu} \quad mit \quad \mu = n \cdot p$$

Verteilungsfunktion: $$P(k_1 \leq X \leq k_2) = \sum_{X=k_1}^{k_2} \frac{\mu^X}{X!}e^{-\mu} \quad mit \quad \mu = n \cdot p$$

Erwartungswert: $$\mu_{PV}(X) = \sum_{k=0}^{n} k \cdot P(X=k) = \mu$$

Varianz: $$\sigma_{PV}{}^2(X) = \sum_{k=0}^{n} k^2 \cdot P(X=k) - \mu_{PV}{}^2 = \mu$$

Herleitung: Erwartungswert der Poissonverteilung

$$E(x) = \sum x \cdot f(x) = \sum_{x=0}^{\infty} x \cdot \frac{\mu^x e^{-\mu}}{x!} \;=$$

Der Erwartungswert ist die Summe der Produkte von x und der Wahrscheinlichkeit für das Auftreten von x. Da die Poissonverteilung für x von 0 bis unendlich definiert ist, läuft der Summationsindex von 0 bis unendlich.

$$= 0 + \sum_{x=1}^{\infty} x \cdot \frac{\mu^x e^{-\mu}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\mu^x e^{-\mu}}{x!} \;=$$

Der erste Summand ist 0, da das Produkt aus 0 und einer beliebigen Wahrscheinlichkeit 0 ist. Es verbleiben die Summanden für x=1 bis unendlich.

$$= \sum_{x=1}^{\infty} x \cdot \frac{\mu \cdot \mu^{x-1} e^{-\mu}}{x \cdot (x-1)!} = \mu \cdot \sum_{x=1}^{\infty} \frac{\mu^{x-1} e^{-\mu}}{(x-1)!} =$$

Die Exponentialfunktion im Zähler wird zerlegt, ebenso die Fakultät im Nenner. Das aus der Fakultät hervorgehende x wird mit dem x vor dem Bruch gekürzt, das My wird vor das Summenzeichen gezogen.

$$= \mu \cdot \underbrace{\sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!}}_{x=1} = \mu \cdot 1 = \mu$$

Nun wird x-1 durch x ersetzt. Die Summation läuft aufgrund der Ersetzung wieder ab 0. Ein Vergleich des zu summierenden Ausdrucks mit der Wahrscheinlichkeitsfunktion der Poissonverteilung zeigt Übereinstimmung. Die Summation über den gesamten Definitionsbereich muss deshalb 1 ergeben.

Herleitung: Varianz der Poissonverteilung

$$Var(X) = \sum x^2 \cdot f(x) - (E(X)^2) = \sum_{x=0}^{\infty} x^2 \cdot \frac{\mu^x \cdot e^{-\mu}}{x!} - \mu^2 =$$

Die Varianz der Poissonverteilung soll berechnet werden. Dazu wird die Wahrscheinlichkeitsfunktion der Poissonverteilung in die allgemeine Formel zur Berechnung der Varianz eingesetzt. Die Summation läuft über den gesamten Definitionsbereich der Poissonverteilung, also von 0 bis unendlich.

$$= 0 + \sum_{x=1}^{\infty} x^2 \cdot \frac{\mu^x \cdot e^{-\mu}}{x!} - \mu^2 = \sum_{x=1}^{\infty} x^2 \cdot \frac{\mu^x \cdot e^{-\mu}}{x!} - \mu^2 =$$

Der erste Summand ist 0, es verbleiben die Summanden für x von 1 bis unendlich.

$$= \sum_{x=1}^{\infty} x^2 \cdot \frac{\mu \cdot \mu^{x-1} \cdot e^{-\mu}}{x \cdot (x-1)!} - \mu^2 = \sum_{x=1}^{\infty} x^2 \cdot \frac{\mu \cdot \mu^{x-1} \cdot e^{-\mu}}{x \cdot (x-1)!} - \mu^2 =$$

$$= \mu \cdot \sum_{x=1}^{\infty} x \cdot \frac{\mu^{x-1} \cdot e^{-\mu}}{(x-1)!} - \mu^2 =$$

Die Exponentialfunktion im Zähler wird auseinandergezogen, ebenso die Fakultät im Zähler. Das My wird vor das Summenzeichen gezogen und das x im Nenner herausgekürzt.

$$= \mu \cdot \sum_{x=0}^{\infty} (x+1) \cdot \frac{\mu^x \cdot e^{-\mu}}{x!} - \mu^2 = \mu \cdot \sum_{x=0}^{\infty} x \cdot \frac{\mu^x \cdot e^{-\mu}}{x!} + \mu \cdot \sum_{x=0}^{\infty} \frac{\mu^x \cdot e^{-\mu}}{x!} - \mu^2 =$$

Das x wird durch x+1 ersetzt. Der Laufindex läuft wieder von 0 bis unendlich. „x-1“ wird zu „x“, „x“ wird zu „x+1“. Das „x+1“ vor dem Bruch wird ausmultipliziert und in zwei Summen aufgeteilt.

$$\mu \cdot \underbrace{\sum_{x=0}^{\infty} x \cdot \frac{\mu^x \cdot e^{-\mu}}{x!}}_{=E(X)=\mu} + \mu \cdot \underbrace{\sum_{x=0}^{\infty} \frac{\mu^x \cdot e^{-\mu}}{x!}}_{=1} - \mu^2 = \mu \cdot \mu + \mu \cdot 1 - \mu^2 = \mu$$

Es zeigt sich, dass die erste Summe dem Ausdruck zur Berechnung des Erwartungswertes entspricht. Dieser ist My [Beweis für Erwartungswert]. Die zweite Summe ist nichts anderes als die Summe der Wahrscheinlichkeiten der Poissonverteilung über den gesamten Definitionsbereich und ergibt von daher 1. Nach Vereinfachung ergibt sich My als Ergebnis.

Erwartungswert der Binomialverteilung

Behauptung:

$$\mu_{BV}(X) \overset{Def.}{=} n \cdot p$$

$$mit \quad E(X) = \sum_{k=0}^{n} k \cdot P(X=k) \overset{P=BV}{=} \sum_{k=0}^{n} k \cdot B_{n;p}(X=k)$$

Herleitung und Beweis:

Fall 1: n = 1

$k\,[Anzahl\ Treffer]$	0	1
$B_{1;p}(X=k)$	$\binom{1}{0} p^0 (1-p)^1 = (1-p)$	$\binom{1}{1} p^1 (1-p)^0 = p$

$$Erwartungswert: \quad E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$

Fall 2: n = 2

$k\,[Anzahl\ Treffer]$	0	1	2
$B_{2;p}(X=k)$	$\binom{2}{0} p^0 (1-p)^2$ $= (1-p)^2$	$\binom{2}{1} p^1 (1-p)^1$ $= 2p(1-p)$	$\binom{2}{2} p^2 (1-p)^0$ $= p^2$

$$Erwartungswert:$$

$$E(X) = 0 \cdot (1-p)^2 + 1 \cdot 2p(1-p) + 2 \cdot p^2 = 0 + 2p - 2p^2 + 2p^2 = 2p$$

Fall 3: n = 3

$k\,[Anzahl\ Treffer]$	0	1	2	3
$B_{3;p}(X=k)$	$\binom{3}{0} p^0 (1-p)^3$ $= (1-p)^3$	$\binom{3}{1} p^1 (1-p)^2$ $= 3p(1-p)^2$	$\binom{3}{2} p^2 (1-p)^1$ $= 3p^2(1-p)$	$\binom{3}{3} p^3 (1-p)^0$ $= p^3$

$$Erwartungswert:$$

$$E(X) = 0 \cdot (1-p)^3 + 1 \cdot 3p(1-p)^2 + 2 \cdot 3p^2(1-p) + 3 \cdot p^3$$

$$E(X) = 3p\left(1 - 2p + p^2\right) + 6p^2 - 6p^3 + 3p^3$$

$$E(X) = 3p - 6p^2 + 3p^3 + 6p^2 - 6p^3 + 3p^3 = 3p$$

Verallgemeinerung für $n \in \mathbb{N}$

$$E(X) = \sum_{k=0}^{n} k \cdot P(X=k) \overset{P=BV}{=} \sum_{k=0}^{n} k \cdot B_{n;p}(X=k) = \mu$$

$$\mu = \sum_{k=0}^{n} k \cdot B_{n;p}(X=k) = \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^{k} \cdot (1-p)^{n-k} = \sum_{k=0}^{n} k \cdot \frac{n!}{k! \cdot (n-k)!} \cdot p^{k} \cdot (1-p)^{n-k}$$

$$\mu \underset{k=0 \text{ und Rest}}{\overset{\text{Summe aufteilen}}{=}} 0 + \sum_{k=1}^{n} k \cdot \frac{n!}{k! \cdot (n-k)!} \cdot p^{k} \cdot (1-p)^{n-k} \underset{n! \text{ aufspalten}}{\overset{k \text{ kürzen}}{=}} \sum_{k=1}^{n} \frac{n \cdot (n-1)!}{(k-1)! \cdot (n-k)!} \cdot p^{k} \cdot (1-p)^{n-k}$$

$$\mu \overset{p \text{ aufspalten}}{=} \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} \cdot p \cdot p^{k-1} \cdot (1-p)^{n-k} = \sum_{k=1}^{n} np \cdot \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)}$$

$$\mu \underset{k=1,2,3 \text{ und } n}{\overset{\text{Auflösen Summe}}{=}} np \cdot \left[\binom{n-1}{0} \cdot p^{0} \cdot (1-p)^{(n-1)} + \binom{n-1}{1} \cdot p^{1} \cdot (1-p)^{(n-2)} + \binom{n-1}{2} \cdot p^{2} \cdot (1-p)^{(n-3)} + \ldots + \binom{n-1}{n-1} \cdot p^{n-1} \cdot (1-p)^{0} \right]$$

$$\textit{Binomischer Lehrsatz:} \quad (a+b)^{n} = \sum_{k=0}^{n} \binom{n}{k} \cdot a^{k} \cdot b^{n-k} \xrightarrow[b=1-p]{a=p}$$

$$\mu \underset{\text{Binomischer Lehrsatz}}{\overset{\text{Anwendung}}{=}} np \cdot \left[p + (1-p) \right]^{n-1} = np \cdot 1^{n-1} = np$$

Varianz der Binomialverteilung

⇨ ***Spezifische Formel der Varianz in der Binomialverteilung***

Behauptung:

$$\mu_{BV}(X) \overset{Def.}{=} n\cdot p \quad und \quad \sigma_{BV}{}^2(X) \overset{Def.}{=} n\cdot p\cdot(1-p)$$

$$mit \quad V(X)=\sum_{k=0}^{n}(k-\mu)^2\cdot P(X=k)$$

$$\rightarrow \quad V(X) \overset{P=BV}{=} \sum_{k=0}^{n}(k-\mu)^2\cdot B_{n;p}(X=k) \xrightarrow[vgl.\,Anmerkung\,1]{Rechenformel} \sum_{k=0}^{n}k^2\cdot B_{n;p}(X=k)-\mu^2$$

Anmerkung: ***Dieser Sachverhalt wurde bereits für die allgemeine Ermittlung der Varianz bewiesen und darf daher hier vorausgesetzt werden.***

Herleitung und Beweis:

Fall 1: n = 1

$k[Anzahl\ Treffer]$	0	1
$B_{1;p}(X=k)$	$\binom{1}{0}p^0(1-p)^1=(1-p)$	$\binom{1}{1}p^1(1-p)^0=p$

$$Erwartungswert: \quad E(X)=\mu=p$$

$$Varianz: \quad V(X)=0^2\cdot(1-p)+1^2\cdot p-p^2=1\cdot p-p^2=1\cdot p(1-p)$$

Fall 2: n = 2

$k[Anzahl\ Treffer]$	0	1	2
$B_{2;p}(X=k)$	$\binom{2}{0}p^0(1-p)^2$ $=(1-p)^2$	$\binom{2}{1}p^1(1-p)^1$ $=2p(1-p)$	$\binom{2}{2}p^2(1-p)^0$ $=p^2$

$$Erwartungswert: \quad E(X)=\mu=2p$$

$$Varianz: \quad V(X)=0^2\cdot(1-p)^2+1^2\cdot 2p(1-p)+2^2\cdot p^2-(2p)^2$$

$$V(X)=2p-2p^2+4p^2-4p^2=2p(1-p)$$

Fall 3: n = 3

$k[Anzahl\ Treffer]$	0	1	2	3
$B_{3;p}(X=k)$	$\binom{3}{0}p^0(1-p)^3$ $=(1-p)^3$	$\binom{3}{1}p^1(1-p)^2$ $=3p(1-p)^2$	$\binom{3}{2}p^2(1-p)^1$ $=3p^2(1-p)$	$\binom{3}{3}p^3(1-p)^0$ $=p^3$

$$Erwartungswert: \quad E(X)=\mu=3p$$

$$V(X)=0^2\cdot(1-p)^3+1^2\cdot 3p(1-p)^2+2^2\cdot 3p^2(1-p)+3^2\cdot p^3-(3p)^2$$

$$Varianz: \quad V(X)=3p\left(1-2p+p^2\right)+12p^2-12p^3+9p^3-9p^2$$

$$V(X)=3p-6p^2+3p^3+12p^2-12p^3+9p^3-9p^2$$

$$V(X)=3p-3p^2=3p(1-p)$$

Verallgemeinerung für $n \in \mathbb{N}$

$$V(X) \overset{P=BV}{=} \sum_{k=0}^{n} (k-\mu)^2 \cdot B_{n;p}(X=k) \xrightarrow[\text{vgl. Anmerkung 1}]{\text{Rechenformel}} \sum_{k=0}^{n} k^2 \cdot B_{n;p}(X=k) - \mu^2$$

$$\sigma^2 = \sum_{k=0}^{n} k^2 \cdot B_{n;p}(X=k) - \mu^2 = \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} - (np)^2 = \sum_{k=0}^{n} k^2 \cdot \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} - n^2 p^2$$

$$\sigma^2 \underset{k=0 \text{ und Rest}}{\overset{\text{Summe aufteilen}}{=}} 0 + \sum_{k=1}^{n} k^2 \cdot \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} - n^2 p^2 \underset{n! \text{ aufspalten}}{\overset{k \text{ kürzen}}{=}} \sum_{k=1}^{n} k^2 \cdot \frac{n}{k} \cdot \frac{(n-1)!}{(k-1)! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} - n^2 p^2$$

$$\sigma^2 \overset{p \text{ aufspalten}}{=} \sum_{k=1}^{n} n \cdot k \cdot \frac{(n-1)!}{(k-1)! \cdot (n-k)!} \cdot p \cdot p^{k-1} \cdot (1-p)^{n-k} - n^2 p^2 = \sum_{k=1}^{n} (n \cdot p) \cdot k \cdot \frac{(n-1)!}{(k-1)! \cdot (n-k)!} \cdot p^{k-1} \cdot (1-p)^{n-k} - n^2 p^2$$

$$\sigma^2 \underset{\text{ausklammern}}{\overset{\mu = n \cdot p}{=}} np \cdot \left[\sum_{k=1}^{n} k \cdot \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)} - np \right] \underset{\substack{\text{Grenze: } k \to k-1 \\ \text{Bildungsgesetz: } k \to k+1}}{\overset{\text{Indexverschiebung}}{=}} np \cdot \left[\sum_{k=0}^{n-1} (k+1) \cdot \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} - np \right]$$

$$\sigma^2 \underset{\substack{\text{faktorisieren/} \\ \text{ausmultiplizieren}}}{\overset{(k+1)}{=}} np \cdot \left[\sum_{k=0}^{n-1} k \cdot \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} + 1 \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} - np \right]$$

$\rightarrow$ *vgl. Anmerkung Schritt 1:* $\sigma^2 = np \cdot (1-p)$

Anmerkung Schritt 1:

$$\sigma^2 \quad = \quad np \cdot \left[\sum_{k=0}^{n-1} k \cdot \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} + 1 \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} - np \right]$$

$$NR(\textit{ für Klammerterm}): \quad \left[\sum_{k=0}^{n-1} k \cdot \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} + 1 \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot (1-p)^{(n-1)-k} - np \right]$$

Aufteilung der Summanden:

$$\left[0 \cdot \binom{n-1}{0} \cdot p^0 \cdot (1-p)^{(n-1)} + 1 \cdot \binom{n-1}{1} \cdot p^1 \cdot (1-p)^{(n-2)} + 2 \cdot \binom{n-1}{2} \cdot p^2 \cdot (1-p)^{(n-3)} + \ldots + 3 \cdot \binom{n-1}{n-1} \cdot p^{n-1} \cdot (1-p)^0 \right]$$

$$+ \quad \left[\binom{n-1}{0} \cdot p^0 \cdot (1-p)^{(n-1)} + \binom{n-1}{1} \cdot p^1 \cdot (1-p)^{(n-2)} + \binom{n-1}{2} \cdot p^2 \cdot (1-p)^{(n-3)} + \ldots + \binom{n-1}{n-1} \cdot p^{n-1} \cdot (1-p)^0 \right] \quad - \quad np$$

Anmerkung zu beiden Summanden:

Ausdruck 1 (Magenta)	$E(X)=\mu=(n-1)\,p$
Ausdruck 2 (blau) => Binomischer Lehrsatz	mit $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \xrightarrow[b=1-p]{a=p} \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = \left[p + (1-p) \right]^{n-1}$

$$\rightarrow \quad k \cdot B_{(n-1),p}(X=k) \quad + \quad \left[p + (1-p) \right]^{n-1} \quad - \quad np$$

$$\xrightarrow[\text{für Obergrenze (n-1)}]{\text{Erwartungswert}} \quad (n-1)\,p \quad + \quad \left[p + (1-p) \right]^{n-1} \quad - \quad np \quad = \quad (n-1)\,p + 1^{n-1} - np \quad = \quad np - p + 1 - np \quad = \quad -p + 1 \quad = \quad 1 - p$$

Stetig:

Reelle Zufallsvariable mit Dichtefunktion

Eine Funktion f(x) heißt Wahrscheinlichkeitsdichte (Dichtefunktion) genau dann, wenn gilt:

$$(1) \quad f(x) \geq 0 \quad \forall x \in \Re \qquad und \qquad (2) \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Ist X eine stetige Zufallsgröße mit der Dichtefunktion f(x), so heißt die reelle Zahl

(i) $\mu(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$ der Erwartungswert der Zufallsgröße X.

(ii) $\sigma^2(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$ die Varianz der Zufallsgröße X.

Stetige Wahrscheinlichkeitsverteilung - (Standard)Normalverteilung:

Wahrscheinlichkeitsdichte / Gauß-Glockenfunktion

$$\varphi_{\mu;\sigma}(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\cdot\left(\frac{z-\mu}{\sigma}\right)^2} \xrightarrow[und \quad \sigma=1]{mit \quad \mu=0} \varphi_{0;1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Umwandlung in die standardisierte ZV:

$$P(X \leq k) = \Phi_{\mu;\sigma}\left(\frac{k-\mu}{\sigma}\right) \quad und \quad z = \frac{k-\mu}{\sigma}$$

=> $\Phi(-z) = 1-\Phi(z) \quad und \quad \Phi(z)-\Phi(-z) = 2\Phi(z)-1$

Approximation der Binomialverteilung durch die Normalverteilung für n groß

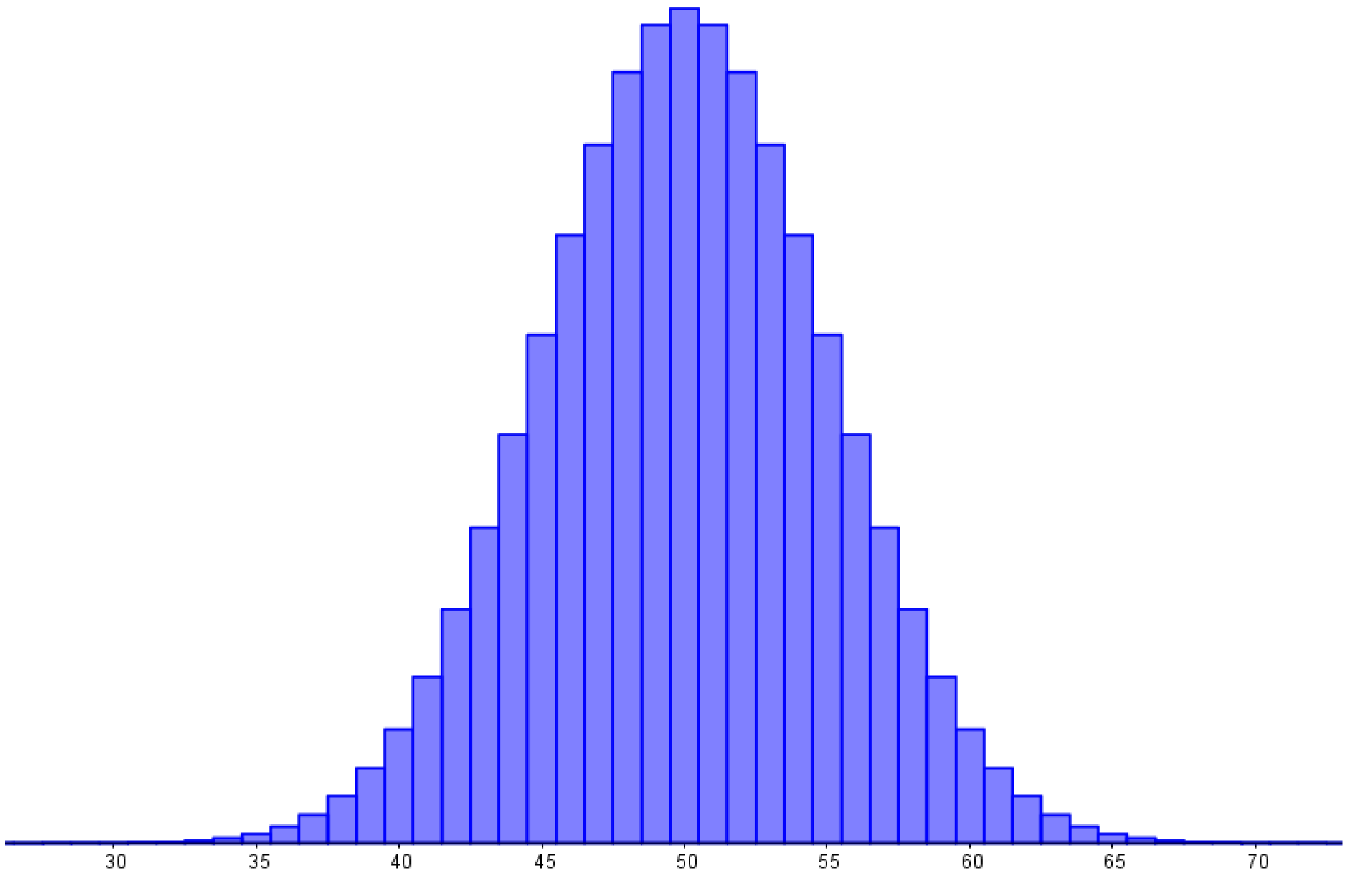
Problem: n groß => hoher Bedarf an Arbeitsspeicher für techn. Geräte
Klassierte / stetige Merkmale

Lösung: Generelle Verteilung zur Berechnung der Wahrscheinlichkeit
mit geringerem technischen Aufwand für stetige und diskrete Merkmale

Ausgangssituation **Binomialverteilung – Bernoulliexp./Bernoullikette (Ziehen mit Zurücklegen):**

$$B_{n;p}(X=k) = \binom{n}{k} p^k \cdot (1-p)^{n-k} \quad mit \quad k \leq n \quad hier: n = 100 \; und \; p = 0{,}5$$

Erwartungswert: $\mu = n \cdot p$ **Varianz:** $\sigma^2 = n \cdot p \cdot (1-p) = \mu \cdot (1-p)$



Link zur Geogebra-Datei

Aus der Graphik sind folgende Eigenschaften abzuleiten:

(1) Der Mittelwert/Erwartungswert ist der am häufigsten vertretene Wert mit der größten Wahrscheinlichkeit

Erwartungswert: $\mu = n \cdot p \rightarrow \mu = 100 \cdot 0,5 = 50$

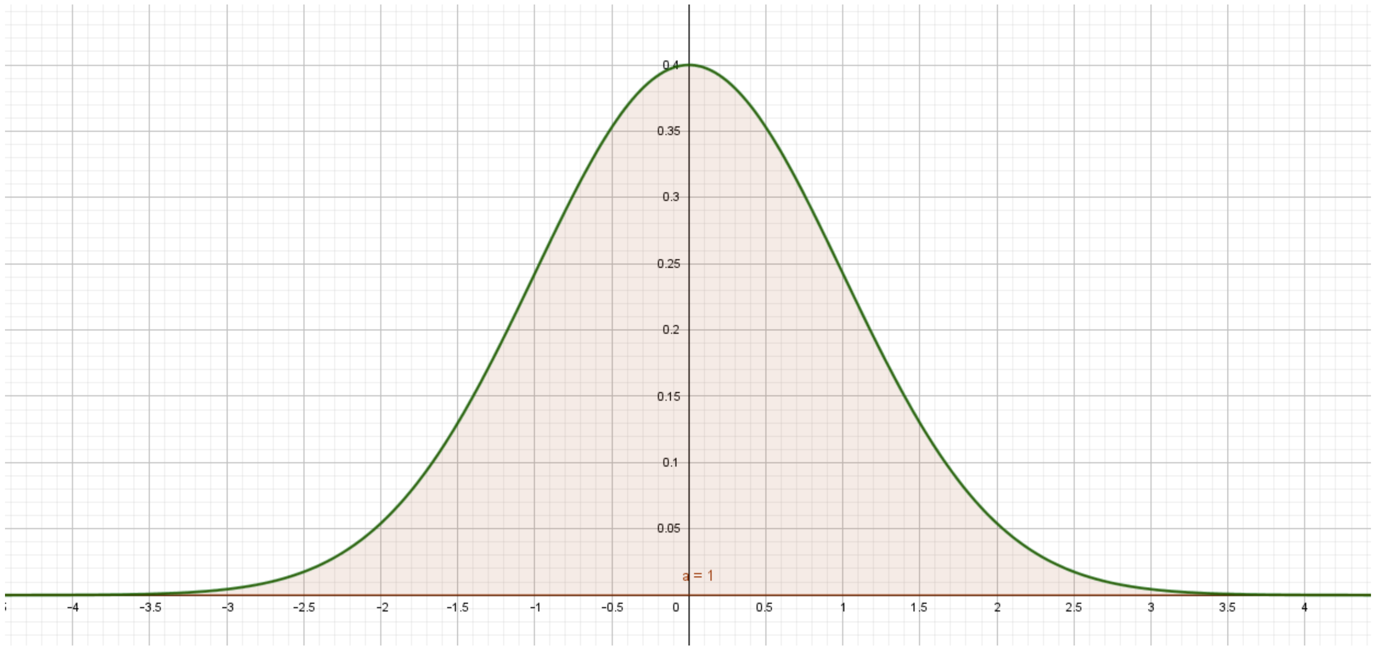
$$B_{100;0,5}(X=50) = \binom{100}{50} 0,5^{50} \cdot 0,5^{n} = \binom{100}{50} 0,5^{100} = 0,0796$$

(2) Die Wahrscheinlichkeitsverteilung bzw. der Graph ist symmetrisch zum Mittel-/Erwartungswert
=> Achsensymmetrie, wenn man die y-Achse auf dem Mittel-/Erwartungswert positionieren würde

(3) Die Häufigkeit bzw. Wahrscheinlichkeit eines Wertes ist um so kleiner, je weiter der Wert vom Mittelwert abweicht

Ziel: **Rekonstruktion einer Verteilungsfunktion, die diese Eigenschaften besitzt.**

⇨ **Gauß'sche Glockenkurve** $f(x) = c \cdot e^{-k \cdot x^2} \quad mit\ c, k \geq 0$



Kurvenuntersuchung mit dem Ziel der Bestimmung des Parameterwertes k, so dass die Verteilung entsprechend einfache Werte annimmt:

A) **Gauß'sche Glockenkurve hat keine Nullstellen**

$$f(x) = c \cdot e^{-k \cdot x^2} \quad mit\ c, k \geq 0$$

B) **Maximum auf der y-Achse:**

$$f'(x) = c \cdot e^{-k \cdot x^2} \cdot (-2k) \cdot x = 0 \xrightarrow[e^{-k \cdot x^2} \neq 0]{Satz: Nullprodukt} (-2k) \cdot x = 0 \rightarrow x = 0$$

$$f''(x) = c \cdot e^{-k \cdot x^2} \cdot (-2k) \cdot x \cdot (-2k) \cdot x + c \cdot e^{-k \cdot x^2} \cdot (-2k)$$

$$f''(x) = \left(4k^2x^2 - 2k\right) \cdot ce^{-k \cdot x^2} \rightarrow f''(0) = (-2k) \cdot c < 0 \rightarrow Max(0 \mid c)$$

C) **Wendepunkte:**

$$f''(x) = \left(4k^2x^2 - 2k\right) \cdot ce^{-k \cdot x^2} = 0 \xrightarrow[e^{-k \cdot x^2} \neq 0]{Satz: Nullprodukt} 4k^2x^2 - 2k = 0 \rightarrow |x| = \sqrt{\frac{1}{2k}}$$

$$f'''(x) = 8k^2x \cdot ce^{-k \cdot x^2} + \left(4k^2x^2 - 2k\right) \cdot ce^{-k \cdot x^2} \cdot (-2kx)$$

$$f'''(x) = \left(12k^2x - 8k^3x^3\right) \cdot ce^{-k \cdot x^2} \rightarrow f'''\left(\sqrt{\frac{1}{2k}}\right) = \left(12k^2 \cdot \sqrt{\frac{1}{2k}} - 8k^3 \sqrt{\frac{1}{2k}}^3\right) \cdot ce^{-k \cdot \frac{1}{2k}} \neq 0$$

$$f(x) = c \cdot e^{-k \cdot \sqrt{\frac{1}{2k}}^2} = c \cdot e^{-k \cdot \frac{1}{2k}} = c \cdot e^{-0,5} \rightarrow WP\left(\pm\sqrt{\frac{1}{2k}} \;\middle|\; c \cdot e^{-0,5}\right)$$

$$Es\ soll\ gelten: \sqrt{\frac{1}{2k}} = 1 \rightarrow k = \frac{1}{2}$$

D) Fläche unter der Gauß'schen Glockenkurve:

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} c\cdot e^{-k\cdot x^2}\,dx \overset{!}{=} 1$$

$$Es\ gilt:\quad c\cdot\int_{-\infty}^{\infty} e^{-k\cdot x^2}\,dx = c\cdot\sqrt{2\pi} \overset{!}{=} 1 \rightarrow c=\frac{1}{\sqrt{2\pi}}$$

Daraus resultiert für die Gauß'sche Glockenkurve folgender Term:

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\cdot x^2}$$

$$mit\ Max\left(0 \;\middle|\; \frac{1}{\sqrt{2\pi}}\right) \approx Max(0 \mid 0{,}3989) \quad und \quad WP\left(\pm 1 \;\middle|\; \frac{1}{\sqrt{2\pi}}\cdot e^{-0{,}5}\right) \approx WP(\pm 1 \mid 0{,}242)$$

Weiterhin soll gelten:

Erwartungswert: $\mu(X) = 0$

Varianz & Standardabweichung: $\sigma^2(X) = 1 \rightarrow \sigma(X) = 1$

Herleitung der Transformation der Zufallsvariablen:

WP => Varianz & Standardabweichung:

$$\sigma^2(X) = 1 \rightarrow \sigma(X) = 1$$

$$|x| = \sqrt{\frac{1}{2k}} \rightarrow x^2 = \frac{1}{2k} \xrightarrow{WP = Varianz} \sigma^2 = \frac{1}{2k} \rightarrow k = \frac{1}{2\sigma^2} = \frac{1}{2}\cdot\frac{1}{\sigma^2}$$

$$\Rightarrow \phi(z) = c\cdot e^{-\frac{1}{2}\cdot\frac{z^2}{\sigma^2}} = c\cdot e^{-\frac{1}{2}\cdot\left(\frac{z}{\sigma}\right)^2}$$

Verschiebung des Maximums an die Stelle $\mu(X) = 0$ **erreicht man durch die Transformation:**

$$z = x-\mu \rightarrow \phi(z) = c\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$$

Bestimmung der Formvariablen c:

Ausgangspunkt: Fläche unter der Gauß'schen Glockenkurve:

$$\int_{-\infty}^{\infty} e^{-k\cdot x^2}\,dx \;=\; \sqrt{2\pi} \quad und \quad z \;=\; \frac{x-\mu}{\sigma} \;\rightarrow\; \phi(z) \;=\; c\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$Es\ gilt: \int_{-\infty}^{\infty} c\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} dx \;=\; c\cdot\int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} dx \;\underset{z=\frac{x-\mu}{\sigma}\ \ und\ \ dz=\frac{1}{\sigma}dx}{\overset{Integration\ durch\ Substitution}{=}}\; c\cdot\sigma\cdot\int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot z^2} dz$$

$$\rightarrow\; c\cdot\sigma\cdot\sqrt{2\pi} \;\overset{!}{=}\; 1 \;\rightarrow\; c=\frac{1}{\sigma\cdot\sqrt{2\pi}}$$

Stetig:

Stetige Wahrscheinlichkeitsverteilung - (Standard)Normalverteilung:
Wahrscheinlichkeitsdichte / Gauß-Glockenfunktion:

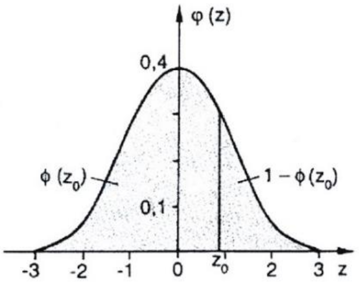
$$\varphi_{\mu;\sigma}(z) \;=\; \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\cdot\left(\frac{z-\mu}{\sigma}\right)^2} \;\xrightarrow[und\quad \sigma=1]{mit\quad \mu=0}\; \varphi_{0;1}(z) \;=\; \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

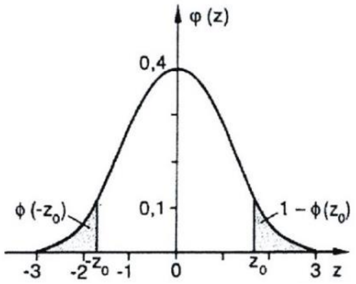
Umwandlung in die standardisierte ZV:

$$P(X\le k) \;=\; \Phi_{\mu;\sigma}\left(\frac{k-\mu}{\sigma}\right) \quad und \quad z=\frac{k-\mu}{\sigma}$$

=> $$\Phi(-z) \;=\; 1-\Phi(z) \quad und \quad \Phi(z)-\Phi(-z) \;=\; 2\Phi(z)-1$$

Tabelle zur Normalverteilung: $\textit{Lesen der Tabelle}: \quad \Phi(z) = 0,\ldots \quad \textit{und} \quad \Phi(-z) = 1 - \Phi(z)$





z	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	9987	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9996	9996	9996	9996	9996	9996	9996	9997
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9997	9998

Beispiele:

$$(1) \quad \Phi(2,37) \underset{ablesen}{\overset{direkt}{=}} 0,9911$$

$$(2) \quad \Phi(-2,37) \overset{Umwandlung}{=} 1 - \Phi(2,37) \underset{ablesen}{\overset{direkt}{=}} 1 - 0,9911 = 0,0089$$

$$(3) \quad \Phi(z) = 0,7910 \xrightarrow[ablesen]{\text{"umgekehrt"}} z = 0,81$$

$$(4) \quad \Phi(z) = 0,2090 \underset{1-x=0,2090}{\overset{x \ festlegen}{=}} 1 - 0,7910 \xrightarrow[ablesen\ bei\ 0,7910]{\text{"umgekehrt"}} z = -0,81$$

Einfaches Sigma-Intervall:

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - \sigma \le X \le \mu + \sigma)$$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 2 \cdot 0{,}8413 - 1 = 0{,}6826$$

$$z_2 = \frac{k_2 - \mu}{\sigma} = \frac{\mu + \sigma - \mu}{\sigma} = \frac{\sigma}{\sigma} = 1$$

$$z_1 = \frac{k_1 - \mu}{\sigma} = \frac{\mu - \sigma - \mu}{\sigma} = \frac{-\sigma}{\sigma} = -1$$

Zweifaches Sigma-Intervall:

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 2\sigma \le X \le \mu + 2\sigma)$$

$$= \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 2 \cdot 0{,}9772 - 1 = 0{,}9544$$

$$z_2 = \frac{k_2 - \mu}{\sigma} = \frac{\mu + 2\sigma - \mu}{\sigma} = \frac{2\sigma}{\sigma} = 2$$

$$z_1 = \frac{k_1 - \mu}{\sigma} = \frac{\mu - 2\sigma - \mu}{\sigma} = \frac{-2\sigma}{\sigma} = -2$$

Dreifaches Sigma-Intervall:

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 3\sigma \le X \le \mu + 3\sigma)$$

$$= \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 2 \cdot 0{,}9987 - 1 = 0{,}9974$$

90%, 95% und 99%-Intervalle => Grenzen

$\rightarrow$ Sigma-Intervall für 90%-Umgebung: $\pm 1{,}64\sigma\ [Tabelle: 0{,}95]$

$\rightarrow$ Sigma-Intervall für 95%-Umgebung: $\pm 1{,}96\sigma\ [Tabelle: 0{,}975]$

$\rightarrow$ Sigma-Intervall für 99%-Umgebung: $\pm 2{,}58\sigma\ [Tabelle: 0{,}995]$

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 1{,}64\sigma \le X \le \mu + 1{,}64\sigma) = 0{,}9$$

$$z_2 = \frac{k_2 - \mu}{\sigma} = \frac{\mu + 1{,}64\sigma - \mu}{\sigma} = \frac{1{,}64\sigma}{\sigma} = 1{,}64$$

$$z_1 = \frac{k_1 - \mu}{\sigma} = \frac{\mu - 1{,}64\sigma - \mu}{\sigma} = \frac{-1{,}64\sigma}{\sigma} = -1{,}64$$

$$\xrightarrow{neue\ Grenzen} \Delta k = 1{,}64\sigma \xrightarrow{\sigma=50} \Delta k = 1{,}64 \cdot 50 = 82 \rightarrow [98/262]$$

analog:

$$\xrightarrow{neue\ Grenzen (95\% \to 97{,}5\%)} \Delta k = 1{,}96\sigma \xrightarrow{\sigma=50} \Delta k = 1{,}96 \cdot 50 = 98 \rightarrow [82/278]$$

$$\xrightarrow{neue\ Grenzen (99\% \to 99{,}5\%)} \Delta k = 2{,}58\sigma \xrightarrow{\sigma=50} \Delta k = 2{,}58 \cdot 50 = 129 \rightarrow [51/309]$$

Sigma-Regeln:

Für eine binomialverteilte Zufallsgröße/-variable X mit den Parametern n und p,

dem Erwartungswert $\mu(X) = n \cdot p$

und der Standardabweichung $\sigma(X) = \sqrt{n \cdot p \cdot (1-p)}$ **mit der Bedingung** $\sigma(X) > 3$

erhält man folgende Näherungen:

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - \sigma \le X \le \mu + \sigma) = P(|X - \mu| \le \sigma) = 0,683$$

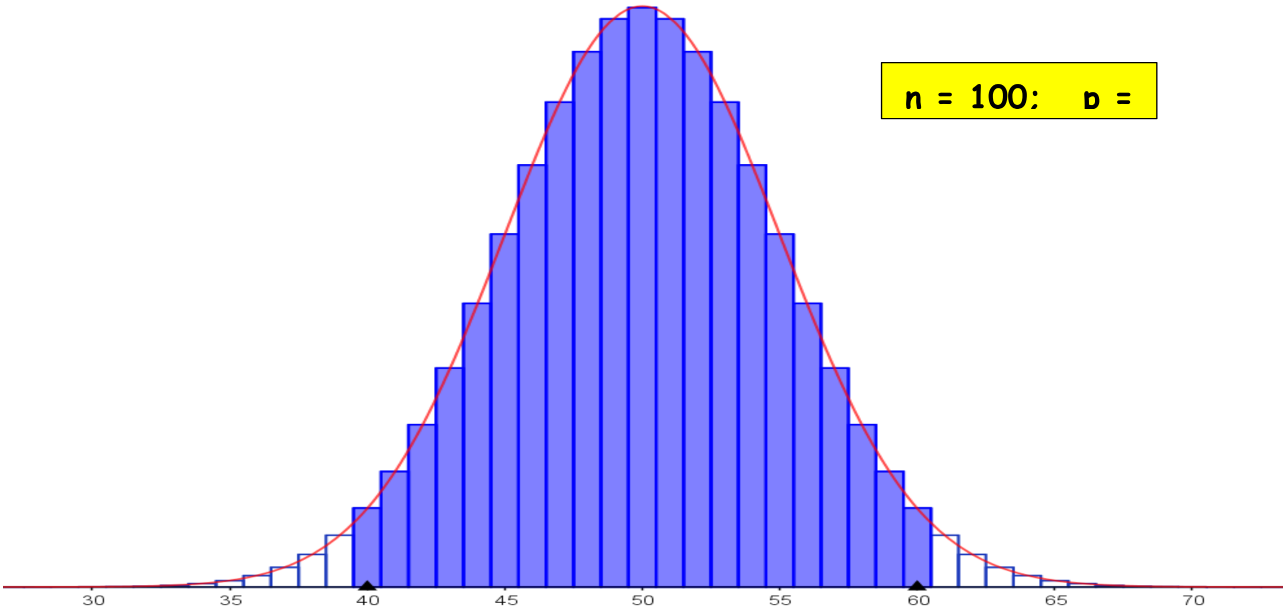
$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(|X - \mu| \le 2\sigma) = 0,954$$

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 3\sigma \le X \le \mu + 3\sigma) = P(|X - \mu| \le 3\sigma) = 0,997$$

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 1,64\sigma \le X \le \mu + 1,64\sigma) = P(|X - \mu| \le 1,64\sigma) = 0,90$$

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 1,96\sigma \le X \le \mu + 1,96\sigma) = P(|X - \mu| \le 1,96\sigma) = 0,95$$

$$P(k_1 \le X \le k_2) \overset{symmetrisch}{=} P(\mu - 2,58\sigma \le X \le \mu + 2,58\sigma) = P(|X - \mu| \le 2,58\sigma) = 0,99$$



$\sigma \leftrightarrow 68,3\,\%$

$2\sigma \leftrightarrow 95,4\,\%$

$3\sigma \leftrightarrow 99,7\,\%$

Bezogen auf die Standardnormalverteilung:

$\rightarrow$ Sigma-Intervall für 95%-Umgebung: $\pm 1,96\sigma\ [Tabelle: 0,975]$

$\rightarrow$ Sigma-Intervall für 90%-Umgebung: $\pm 1,64\sigma\ [Tabelle: 0,95]$

$\rightarrow$ Sigma-Intervall für 99%-Umgebung: $\pm 2,58\sigma\ [Tabelle: 0,995]$

Vertrauensintervall zur Ermittlung von p-Intervallen: Konfidenzintervall

Situation: ***Würfelexperiment***

Bei einem Würfelexperiment von 100 Würfen erzielt man 30 Einsen.
Handelt es sich um einen gezinkten Würfel?

Welche Erfolgswahrscheinlichkeiten sind mit dem Stichprobenergebnis von $\frac{X}{n} = h = 0{,}3$

unter einer vorgegebenen Sicherheitswahrscheinlichkeit von z.B. 95,4 % vereinbar?

⇨ **Gesucht ist Intervall für p**

Berechnung:

X = Anzahl der gewürfelten Einsen => $P(X=1) = \frac{1}{6}$

95,4 % bedeutet $2\sigma - \text{Intervall/Regel}$

$$P(k_1 \le X \le k_2) \overset{\text{symmetrisch}}{=} P(\mu - 2\sigma \le X \le \mu + 2\sigma) = P(|X-\mu| \le 2\sigma) = 0{,}954$$

Lösung Beispielsituation:

$$\mu - 2\sigma \le X \le \mu + 2\sigma \xrightarrow[X=30]{\text{Experiment / Stichprobe}} \mu - 2\sigma \le 30 \le \mu + 2\sigma$$

$$\Leftrightarrow \quad n \cdot p - 2\sqrt{n \cdot p \cdot (1-p)} \le 30 \le n \cdot p + 2\sqrt{n \cdot p \cdot (1-p)}$$

$$\xrightarrow[\text{für Intervallgrenzen}]{\text{Gleichung}} 30 = n \cdot p \pm 2\sqrt{n \cdot p \cdot (1-p)} \xrightarrow{\text{n = 100}} 30 = 100 \cdot p \pm 2\sqrt{100 \cdot p \cdot (1-p)}$$

$$\xrightarrow[\text{separieren}]{\text{Wurzel}} 30 = 100 \cdot p \pm 2\sqrt{100 \cdot p \cdot (1-p)} \rightarrow 100 \cdot p - 30 = \pm 2\sqrt{100 \cdot p \cdot (1-p)}$$

$$\xrightarrow{\text{quadrieren}} (100 \cdot p - 30)^2 = 4 \cdot 100 \cdot p \cdot (1-p) \rightarrow 10.000\,p^2 - 6000\,p + 900 = 400\,p - 400\,p^2$$

$$\rightarrow 10.400\,p^2 - 6400\,p + 900 = 0 \rightarrow p_1 = 0{,}2175 \wedge p_2 = 0{,}3978$$

$$\xrightarrow{\text{p-Intervall}} p \in [0{,}2175\,;\,0{,}3978]$$

Auswertung & Ergebnis:

Liegt $P(X=1) = \frac{1}{6}$ **im p-Intervall?** **=>** $P(X=1) = \frac{1}{6} \notin [0{,}2175\,;\,0{,}3978]$

(1) Der Würfel ist mit einer Sicherheitswahrscheinlichkeit von 95,4 % gezinkt.

(2) Die mit der Stichprobe gültigen/verträglichen Erfolgswahrscheinlichkeiten müssen im Intervall für p liegen.

Herleitung allgemein:

$$\mu - z_{\frac{\alpha}{2}} \cdot \sigma \le X \le \mu + z_{\frac{\alpha}{2}} \cdot \sigma \xrightarrow[X=T]{\text{Experiment / Stichprobe}} \mu - z_{\frac{\alpha}{2}} \cdot \sigma_S \le T \le \mu + z_{\frac{\alpha}{2}} \cdot \sigma_S$$

$$\Leftrightarrow \quad n \cdot p - z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p_S \cdot (1-p_S)} \le T \le n \cdot p + z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p_S \cdot (1-p_S)}$$

$$\xrightarrow[\text{für Intervallgrenzen}]{\text{Gleichung}} \quad T = n \cdot p \pm z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p_S \cdot (1-p_S)}$$

$$\xrightarrow[\text{separieren}]{\text{Wurzel}} \quad n \cdot p - T = \pm z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p_S \cdot (1-p_S)}$$

Option 1: Quadrieren des Ausdrucks und direkte Ermittlung des Intervalls per CAS

$$\xrightarrow{\text{Quadrieren}} \quad (n \cdot p_S - T)^2 = z_{\frac{\alpha}{2}}^{\ 2} \cdot n \cdot p_S \cdot (1-p_S) \quad \to \quad p - \mathit{Intervall}$$

Option 2: Direkte Ermittlung der Abweichungswerte => Intervallbildung

$$\xrightarrow[\text{durch } n]{\text{Division}} \quad p - \frac{T}{n} = \pm z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{n \cdot p_S \cdot (1-p_S)}}{n} \quad \xrightarrow[\frac{T}{n} = h = p_S]{\text{Kürzen } n} \quad p = p_S \pm z_{\frac{\alpha}{2}} \cdot \frac{\sqrt{p_S \cdot (1-p_S)}}{\sqrt{n}}$$

$$\left[\to \quad \sigma_S = \sqrt{\frac{p_S \cdot (1-p_S)}{n}} \right] \quad \to \quad p = p_S \pm z_{\frac{\alpha}{2}} \cdot \sigma_S \quad \to \quad p - \mathit{Intervall}$$

Allgemeine Ansätze:

Konfidenzintervall: $X = n \cdot p \pm z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p \cdot (1-p)} \quad \to \quad (n \cdot p - X)^2 = z_{\frac{\alpha}{2}}^{\ 2} \cdot n \cdot p \cdot (1-p)$

Konfidenzintervall (Abschätzung): $\to \quad p = p_S \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_S \cdot (1-p_S)}{n}} = p_S \pm z_{\frac{\alpha}{2}} \cdot \sigma_S$

Mindestumfang einer Stichprobe auf Basis einer gewünschten Sicherheitswahrscheinlichkeit

- ⇨ **Bei (un-)bekannter Wahrscheinlichkeit p**

(entsprechend den $\sigma-\text{Intervallen/Regeln}$ **)**

Allgemeine Fragestellung:
Wie groß muss der Stichprobenumfang n mindestens sein, damit mit vorgegebener Sicherheits-wahrscheinlichkeit (σ-Intervalle) die relative Häufigkeit $\frac{X}{n}=h$ ***in der ε-Umgebung d von p liegt?***

Konkrete Fragestellung am Beispiel:
Wie groß muss der Stichprobenumfang gewählt werden, um ein Konfidenzintervall für den unbekannten Anteil p eines Zufallsexperiments zum Konfidenzniveau von z.B. **95 %** und einem Schätzfehler von **0,05** zu bestimmen?

Wenn der Stichprobenumfang auf jeden Fall so groß gewählt wird, dass die Approximationsbedingung für eine Normalverteilung erfüllt ist (Faustregel: n > 30), kann man entsprechend den σ-Regeln und der NV **den Wert für z** aus der NV-Tabelle [N-0,1-Verteilung] ablesen.

Es wird deshalb der für die Festsetzung des Stichprobenumfanges ungünstigste Fall von p gewählt; dieser liegt bei p=0,5 (vgl. Herleitung NR)

Man befindet sich damit hinsichtlich des Stichprobenumfanges immer auf der "sicheren" Seite, denn für jeden anderen Wert von p würde sich ein kleinerer Stichprobenumfang ergeben.
Bestimmung des Stichprobenumfangs – MM*Stat (hu-berlin.de)

Lösung Beispielsituation:

$$n = \frac{z_{\frac{\alpha}{2}}^{2}}{d^2}\cdot p\cdot(1-p) \xrightarrow[p\ \text{unbekannt}]{vgl.\ NR} n = \frac{z_{\frac{\alpha}{2}}^{2}}{4d^2}$$

$$\rightarrow n = \frac{1{,}96^2}{4\cdot 0{,}05^2} = 384{,}16 \rightarrow n \geq 385$$

Herleitung allgemein:

$$\mu - z_{\frac{\alpha}{2}} \cdot \sigma \leq X \leq \mu + z_{\frac{\alpha}{2}} \cdot \sigma \xrightarrow[X=T]{Experiment\,/\,Stichprobe} \mu - z_{\frac{\alpha}{2}} \cdot \sigma_S \leq T \leq \mu + z_{\frac{\alpha}{2}} \cdot \sigma_S$$

$$\Leftrightarrow \quad n \cdot p - z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p \cdot (1-p)} \leq T \leq n \cdot p + z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p \cdot (1-p)}$$

$$\rightarrow \quad T = n \cdot p \pm z_{\frac{\alpha}{2}} \cdot \sqrt{n \cdot p \cdot (1-p)} \xrightarrow{:n} \frac{T}{n} = p \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

$$\xrightarrow{\frac{T}{n}=h} \quad h = p \pm z_{\frac{\alpha}{2}} \cdot \frac{1}{\sqrt{n}} \sqrt{p \cdot (1-p)} \quad \rightarrow \quad h - p = \pm z_{\frac{\alpha}{2}} \cdot \frac{1}{\sqrt{n}} \sqrt{p \cdot (1-p)}$$

$$\xrightarrow[Abweichung]{|h-p|=d} \quad d = \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{p \cdot (1-p)} \quad \rightarrow \quad \sqrt{n} = \frac{z_{\frac{\alpha}{2}}}{d} \sqrt{p \cdot (1-p)}$$

$$\xrightarrow{quadriert} \quad n = \frac{z_{\frac{\alpha}{2}}^{\;2}}{d^2} \cdot p \cdot (1-p) \xrightarrow[p\ unbekannt]{vgl.\,NR} \quad n = \frac{z_{\frac{\alpha}{2}}^{\;2}}{4d^2}$$

$NR:$

$$f(p) = \sqrt{p \cdot (1-p)} \xrightarrow[identisch]{Max\text{-}Stelle} f^2(p) = g(p) = p \cdot (1-p) = p - p^2$$

$$\xrightarrow{Extremum} \quad g'(p) = 1 - 2p = 0 \quad \rightarrow \quad p = 0{,}5 \quad \rightarrow \quad g''(p) = -2 < 0$$

$$\rightarrow \quad Max\ bei\ p = 0{,}5$$

Mindestumfang einer Stichprobe:

Allgemeiner Ansatz:

$$n = \frac{z_{\frac{\alpha}{2}}^{\;2}}{d^2} \cdot p_S \cdot (1 - p_S) \xrightarrow[p\ unbekannt]{p(\max) = \frac{1}{2}} n = \frac{z_{\frac{\alpha}{2}}^{\;2}}{4d^2}$$

$mit\ d = |h - p|$ bzw. zugelassener Abweichung

Zeitreihenanalyse - Tertiale

	A	B	C	D	E	F	G	H	I
1	Die Umsätze eine								
2	Ermitteln Sie		- Saisonkomonente und						
3			- irreguläre Komponente						
4									
5									
6	Jahr /	Ursprungs-	gleitende		$sk_i^* = y_i - y_i$			Saisonbereinigte	Irreguläre
7	Tertial	werte	Durchschnitte		Tertialdurchschnitte		sk_i	Werte	Komponente
8		y_i	$y_i \sim gk_i$	I	II	III		y_i - sk_i	ik= y_i $-sk_i$-gk_i
9									
10	**2003 / I**	**12**					**=\$D\$20**	**=B10-G10**	
11	**2003 / II**	**16**	**=(B10+B11+B12)/3**		**=B11-C11**		**=\$E\$20**	**=B11-G11**	**=B11-C11-G11**
12	**2003 / III**	**30**	**=(B11+B12+B13)/3**			**=B12-C12**	**=\$F\$20**	**=B12-G12**	**=B12-C12-G12**
13	**2004 / I**	**15**	**=(B12+B13+B14)/3**	**=B13-C13**			**=\$D\$20**	**=B13-G13**	**=B13-C13-G13**
14	**2004 / II**	**21**	**=(B13+B14+B15)/3**		**=B14-C14**		**=\$E\$20**	**=B14-G14**	**=B14-C14-G14**
15	**2004 / III**	**35**	**=(B14+B15+B16)/3**			**=B15-C15**	**=\$F\$20**	**=B15-G15**	**=B15-C15-G15**
16	**2005 / I**	**20**	**=(B15+B16+B17)/3**	**=B16-C16**			**=\$D\$20**	**=B16-G16**	**=B16-C16-G16**
17	**2005 / II**	**26**	**=(B16+B17+B18)/3**		**=B17-C17**		**=\$E\$20**	**=B17-G17**	**=B17-C17-G17**
18	**2005 / III**	**38**					**=\$F\$20**	**=B18-G18**	
19									
20			sk_i^*=	=(D13+D16)/2	=(E11+E14+E17)/3	=(F12+F15)/2			**=SUMME(I11:I17)**
21									
22									
23									
24									
25									
26			a = -7,00 + (-2,67) + 9,67						

Summe der Saisoneinflüsse müssen sich zum Wert 0 gegenseitig eliminieren. Ansonsten muss eine entsprechende Korrektur vorgenommen werden.

Durchschnitt = Saisonkomponent	23,000	-38,667	15,667

gleitender Durchschnitt
(x1+x2+x3)/3 = ???
(x2+x3+x4)/3 = ???
(30+15+21)/3 =

Differenz des gleitenden Durchschnittwertes vom Ursprungswert: **yi - gleitDurch**
Die Differenz wird dem Zeitraum zugewiesen, in dem der Ursprungswert liegt! => periodengerecht

Durchschnitte = Saisonkomponenten: Einordnen/Auflisten der Werte in ihrem jeweiligen Periodenbereich

Berechnung der saisonbereinigten Werte:
yi - ski
(Ursprungswert - Saisonkomponente)

Durchschnitte der Tertial-Abweichungen
Tertial I: (Summe der 1. Spaltenwerte)/n = ???
Tertial II: analog zu Tertial I

Zeitreihenanalyse - Quartale

	A	B	C	D	E	F	G	H	I	J
1	Jahr /	Ursprungs-	gleitende	$sk_i^* = y_i - y_i$				sk_i	Saisonbereinigte	Irreguläre
2	Quartal	werte	Durchschnitte	Quartalsdurchschnitte					Werte	Komponente
3		y_i	$y_i \sim gk$	I	II	III	IV		$y_i - sk_i$	$ik = y_i - sk_i - gk_i$
5	1 / I	20						= $D$33	=B5-H5	
6	1 / II	21						=$E$33	=B6-H6	
7	1 / III	28	=(B5/2+B6+B7+B8+B9/2)/4			=B7-C7		=$F$33	=B7-H7	=B7-C7-H7
8	1 / IV	29	=(B6/2+B7+B8+B9+B10/2)/4				=B8-C8	=$G$33	=B8-H8	=B8-C8-H8
9	2 / I	29	=(B7/2+B8+B9+B10+B11/2)/4	=B9-C9				= $D$33	=B9-H9	=B9-C9-H9
10	2 / II	32	=(B8/2+B9+B10+B11+B12/2)/4		=B10-C10			=$E$33	=B10-H10	=B10-C10-H10
11	2 / III	37	=(B9/2+B10+B11+B12+B13/2)/4			=B11-C11		=$F$33	=B11-H11	=B11-C11-H11
12	2 / IV	34	=(B10/2+B11+B12+B13+B14/2)/4				=B12-C12	=$G$33	=B12-H12	=B12-C12-H12
13	3 / I	35	=(B11/2+B12+B13+B14+B15/2)/4	=B13-C13				= $D$33	=B13-H13	=B13-C13-H13
14	3 / II	37	=(B12/2+B13+B14+B15+B16/2)/4		=B14-C14			=$E$33	=B14-H14	=B14-C14-H14
15	3 / III	44	=(B13/2+B14+B15+B16+B17/2)/4			=B15-C15		=$F$33	=B15-H15	=B15-C15-H15
16	3 / IV	45	=(B14/2+B15+B16+B17+B18/2)/4				=B16-C16	=$G$33	=B16-H16	=B16-C16-H16
17	4 / I	40	=(B15/2+B16+B17+B18+B19/2)/4	=B17-C17				= $D$33	=B17-H17	=B17-C17-H17
18	4 / II	42	=(B16/2+B17+B18+B19+B20/2)/4		=B18-C18			=$E$33	=B18-H18	=B18-C18-H18
19	4 / III	48	=(B17/2+B18+B19+B20+B21/2)/4			=B19-C19		=$F$33	=B19-H19	=B19-C19-H19
20	4 / IV	50	=(B18/2+B19+B20+B21+B22/2)/4				=B20-C20	=$G$33	=B20-H20	=B20-C20-H20
21	5 / I	47	=(B19/2+B20+B21+B22+B23/2)/4	=B21-C21				= $D$33	=B21-H21	=B21-C21-H21
22	5 / II	51	=(B20/2+B21+B22+B23+B24/2)/4		=B22-C22			=$E$33	=B22-H22	=B22-C22-H22
23	5 / III	57						=$F$33	=B23-H23	
24	5 / IV	58						=$G$33	=B24-H24	
27		Summe der Saisoneinflüsse müssen sich zum Wert 0 gegenseitig eliminieren. Ansonsten muss eine entsprechende Korrektur vorgenommen werden.	sk_i^*=	=(D9+D13+D17+D21)/4	=(E10+E14+E18+E22)/4	=(F7+F11+F15+F19)/4	=(G8+G12+G16+G20)/4			=SUMME(J7:J22)
28										
29										
30						a =	=D27+E27+F27+G27			
31			$^a/_4$ =	=G30/4						
33			sk_i=	=D27-$D$31	=E27-$D$31	=F27-$D$31	=G27-$D$31			

Durchschnitt = Saisonkomponente	23,375	-17,125	-25,875	19,500

gleitender Durchschnitt
(0,5*x1+x2+x3+x4+0,5*x5)/4 = ???
(0,5*x2+x3+x4+x5+0,5*x6)/4 = ???
usw.

Differenz des gleitenden Durchschnittwertes vom Ursprungswert:
yi - gleitDurch
Die Differenz wird dem Zeitraum zugewiesen, in dem der Ursprungswert liegt! => periodengerecht

Durchschnitte = Saisonkomponenten: Einordnen/Auflisten der Werte in ihrem jeweiligen Periodenbereich

Berechnung der saisonbereinigten Werte:
yi - ski
(Ursprungswert - Saisonkomponente)

Durchschnitte der Quartal-Abweichungen
Quartal I: (Summe der 1. Spaltenwerte)/n = ???
Quartal II: analog zu Quartal I