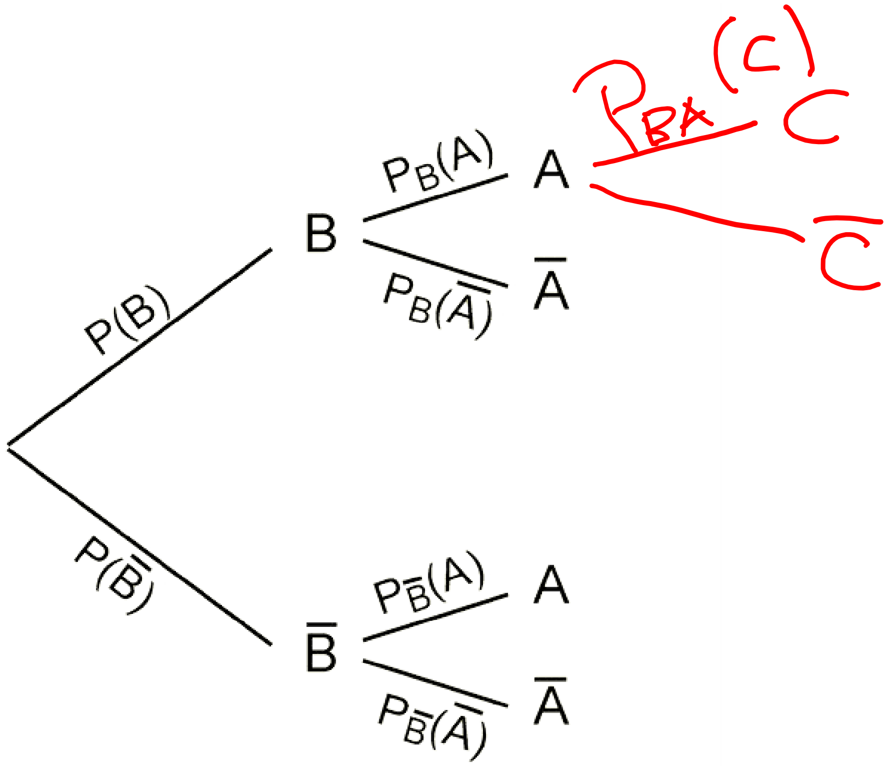
Mathematik - Stochastik

Stochastik - Komplettübersicht

- Baumdiagramm und Vierfeldertafel
- Absolute / bedingte Wahrscheinlichkeiten
- Absolute / relative Häufigkeiten
- Binomialverteilung, Kumulierte Verteilung („genau…, mindestens…, höchstens…“)
- Kombinatorik
- Erwartungswert, Standardabweichung (Laplace-Bedingung), Sigma-Umgebung, Konfidenzintervall
- Hypothesentests (linksseitig, rechtsseitig, beidseitig)

Baumdiagramm und Vierfeldertafel



Baumdiagramm und Vierfeldertafel

Absolute und relative Häufigkeiten

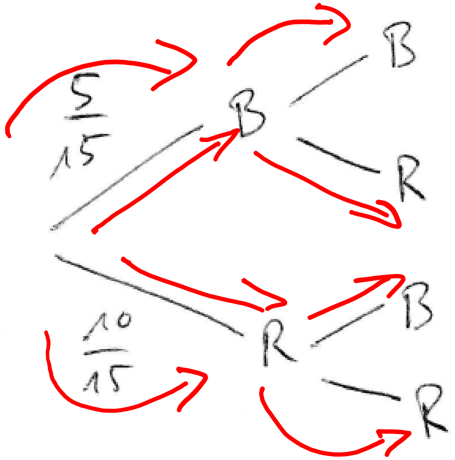
- Absolute Häufigkeit: „Wie oft kommt etwas vor?“
  - Beispiel: Notenspiegel
- Relative Häufigkeit = Absolute Häufigkeit / Anzahl aller Häufigkeiten
  - Beispiel: Notenverteilung

Baumdiagramm und Vierfeldertafel

5 BLAUE Kugeln

10 ROTE Kugeln (15)

mit Zurücklegen



$(BB) = \frac{5}{15} \cdot \frac{5}{15}$

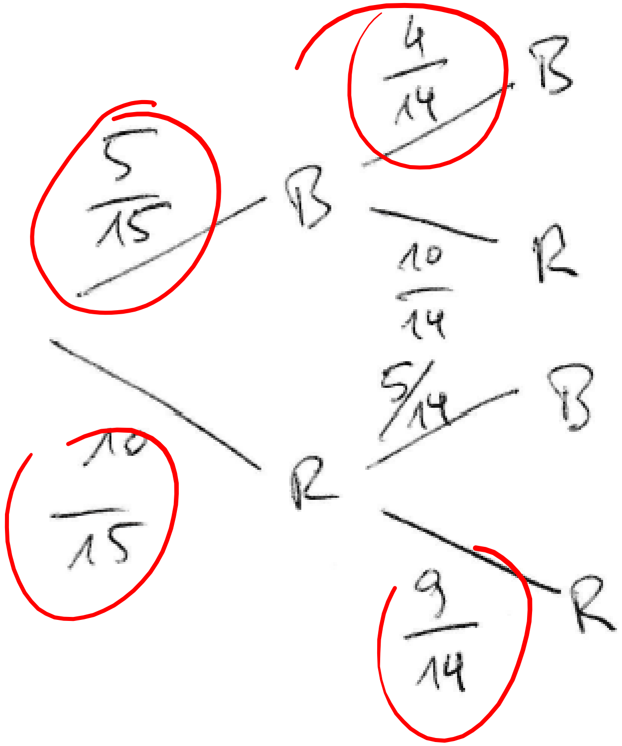
$+$

$(RR) \ \frac{10}{15} \cdot \frac{10}{15}$

$P(\text{gleiche Farbe}) = \Sigma$

Baumdiagramm und Vierfeldertafel

$P_B(B)$, $P(BB)$)



ohne Zurücklegen

$P(BB) = \frac{5}{15} \cdot \frac{4}{14}$

$P(RR) = \frac{10}{15} \cdot \frac{9}{14}$

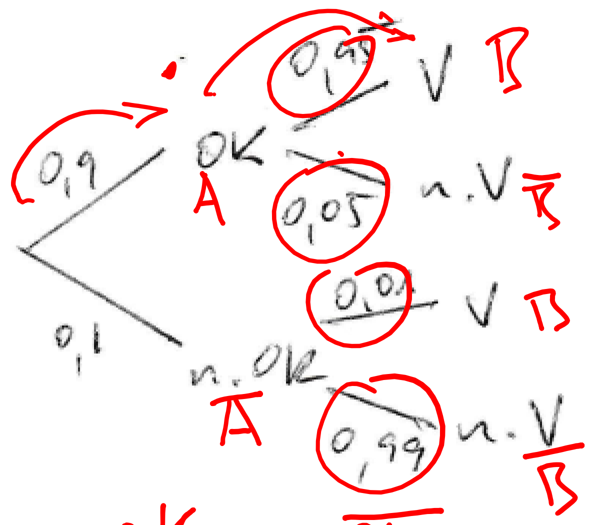
$P_{Ges} = P(BB) + P(RR)$

Baumdiagramm und Vierfeldertafel

90% der Schrauben sind ok

→ Von diesen werden 95% verkauft

→ Von den kaputten werden 1% verkauft



0,855	$P(OK \cap V)$
0,045	$P(OK \cap n.V)$
0,001	$P(n.OK \cap V)$
0,099	$P(n.OK \cap n.V)$

n.OK = $\overline{OK}$

Baumdiagramm und Vierfeldertafel

	OK	n.OK	
	$P_A(B)$	$P_A(B)$	
V	0,855	0,001	0,856
n.V.	0,045	0,099	0,144
	0,9	0,1	1

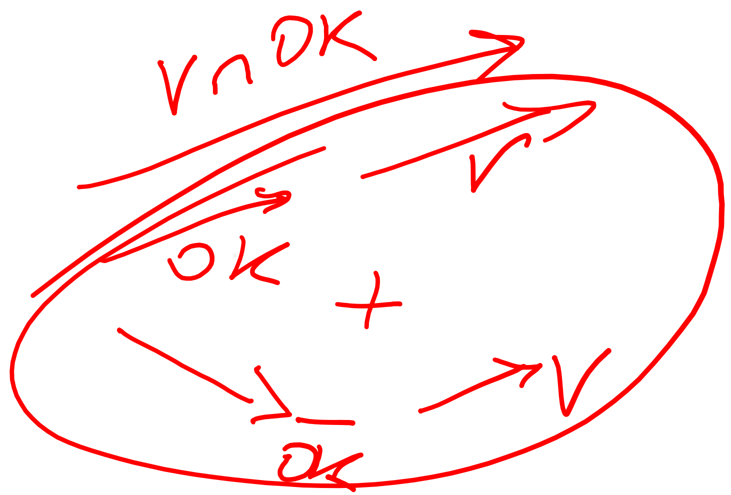
$P(V)$ Σ $P(n.V.)$

Σ

$P(OK)$ $P(n.OK)$

$\cap$ = „und" $\vee$ = „oder"

Baumdiagramm und Vierfeldertafel

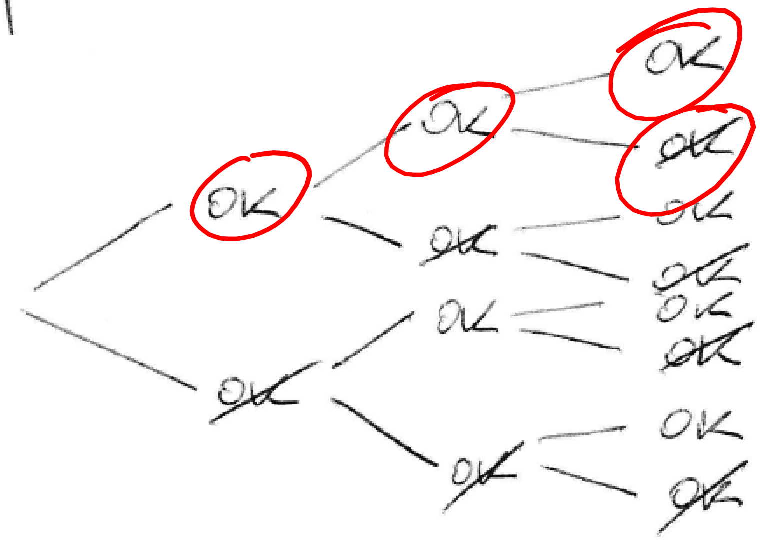


1. Schraube ist verkauft und OK
2. Schraube ist verkauft
3. Eine verkaufte Schraube ist OK

$$P_v(OK) = \frac{P(v \cap OK)}{P(v)} = \frac{0{,}855}{0{,}856} = \underline{0{,}9988}$$

Baumdiagramm und Vierfeldertafel

Schraubenproduktion:



- mit Zurücklegen
- 2 Wahrscheinlichkeiten (WK)

Baumdiagramm und Vierfeldertafel

WK, dass von 3... genau 2 ok sind

I	→ OK	→ OK	→ ~~OK~~	:	$0{,}9 \cdot 0{,}9 \cdot 0{,}1$
II	→ OK	→ ~~OK~~	→ OK	:	$0{,}9 \cdot 0{,}1 \cdot 0{,}9$
III	→ ~~OK~~	→ OK	→ OK	:	$0{,}1 \cdot 0{,}9 \cdot 0{,}9$

$$3 \cdot 0{,}9^2 \cdot 0{,}1^1$$

Baumdiagramm und Vierfeldertafel

$$P(x=k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$\rightarrow P(x=2) = \binom{3}{2} \cdot 0{,}9^2 \cdot (\underbrace{1-0{,}9}_{0{,}1})^{3-2}$$

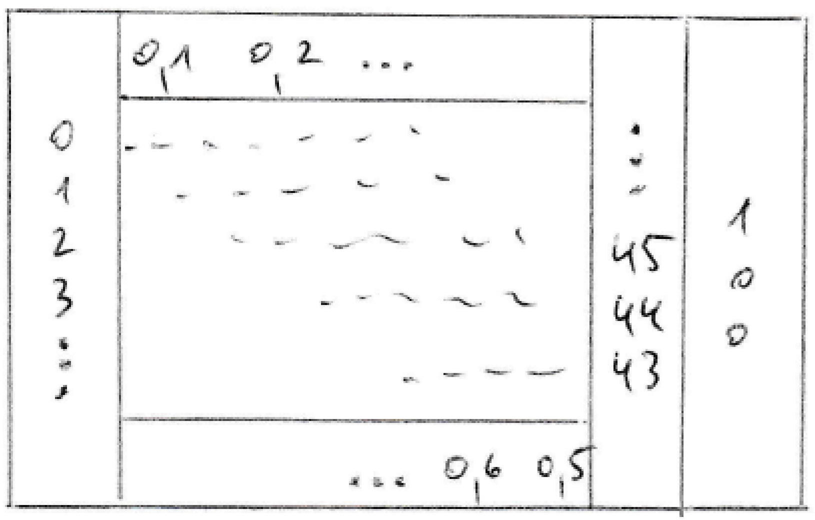
k = Trefferanzahl

n = Gesamtanzahl aller Ereignisse

Binomialverteilungen, kumulierte Verteilungen

$B(n;p;k)$

„genau"



$F(n;p;k)$

„höchstens", „mindestens"

Gleiches Schema, aber kumuliert!

Also von 0 bis zum k

kumuliert = aufsummiert

Binomialverteilungen, kumulierte Verteilungen

Beispiel: Von 100 Leuten ($p = 0{,}2$ f. Raucher) sind

a) genau 3 Raucher

b) höchstens 3 Raucher

Binomialverteilt

a) $P(X=3) = \binom{100}{3} \cdot 0{,}2^{3} \cdot 0{,}8^{97} = B(100;\, 0{,}2;\, 3)$

kumuliert

b) $P(X \leq 3) = F(100;\, 0{,}2;\, 3)$

Binomialverteilungen, kumulierte Verteilungen

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

Binomialverteilungen, kumulierte Verteilungen

WK für eine defekte Schraube: 5%

100 Schrauben werden untersucht...

a) genau 7 defekte

b) höchstens 7 defekte

c) mindestens 7 defekte

d) mehr als 2, aber weniger als 10 defekte

n = 100

p = 0,05

k = ?

Binomialverteilungen, kumulierte Verteilungen

a) $P(X=7) = B(100;\, 0{,}05;\, 7) = \binom{100}{7} \cdot 0{,}05^{7} \cdot 0{,}95^{93}$

b) $P(X \leq 7) = F(100;\, 0{,}05;\, 7)$

oder $\sum_{x=0}^{7} \binom{100}{x} \cdot 0{,}05^{x} \cdot 0{,}95^{100-x}$

c) $P(X \geq 7) = 1 - P(X \leq 6) = 1 - F(100;\, 0{,}05;\, 6)$

oder $\sum_{x=7}^{100} \binom{100}{x} \cdot 0{,}05^{x} \cdot 0{,}95^{100-x}$

d) $P(3 \leq X \leq 9) = F(100;\, 0{,}05;\, 9) - F(100;\, 0{,}05;\, 2)$

Binomialverteilungen, kumulierte Verteilungen

Genau die 7. ist defekt:

$$P = 0{,}95^{99} \cdot 0{,}05^{1}$$

$$= 0{,}95^{6} \cdot 0{,}05^{1} \cdot 0{,}95^{93}$$

$$\frac{P(7.\ \text{defekt})}{P(\text{alle defekt})}$$

bis runter auf P(nur 7. defekt)

Binomialverteilungen, kumulierte Verteilungen

$n = 100$ $p = 0{,}05$

0,95

1 OK — 2 OK ... — 6 OK — 7 $\overline{OK}$

~~P = 0,98 100/100~~

Kombinatorik

1) Zahlenschloss: 4 Ziffern, 0-9

$$\underline{2}\ \underline{8}\ \underline{0}\ \underline{9}$$

$$10 \cdot 10 \cdot 10 \cdot 10$$

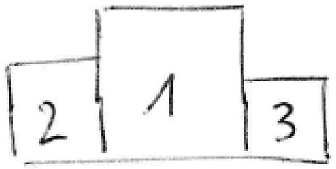
→ Reihenfolge wichtig? Ja

→ Wdh.? Ja

→ $n^k = 10^4$ Möglichkeiten

Kombinatorik

2) 10 Rennläufer auf Plätzen 1–3



$9 \cdot 10 \cdot 8$

→ Reihenfolge? Ja

→ Wdh. Nein

$$\frac{n!}{(n-k)!} = \frac{10!}{(10-3)!} = \frac{10 \cdot 9 \cdot 8 \cdot \cancel{7 \cdot \ldots \cdot 3 \cdot 2 \cdot 1}}{\cancel{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}$$

$$= \boxed{nPR}$$

Kombinatorik

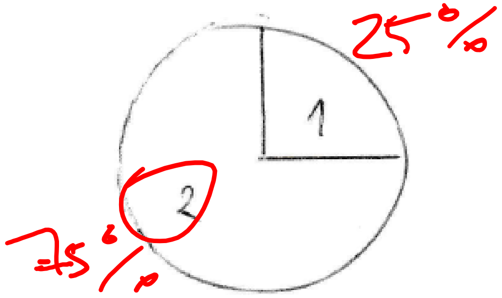
3) Von 10 Personen werden 3 zufällig ausgewählt:

→ Reihenfolge? Nein

→ Wdh.? Nein

$$\binom{n}{k} = \binom{10}{3} = \frac{n!}{(n-k)! \cdot k!} = \boxed{nCR}$$

Erwartungswert



Einsatz: 5 €

E_1: 20 € Auszahlung

E_2: 3 € — " —

→ Erwarteter Gewinn?

→ X: Gewinn in €

x_i	+15 €	−2 €
$P(X=x_i)$	0,25	0,75

Erwartungswert

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

$$E(X) = 0{,}25 \cdot 15 + 0{,}75 \cdot (-2) = 2{,}25$$

→ Fair ist ein Spiel nur, wenn $E(X) = 0$!

Erwartungswert

Beispiel:

Tasche OK : 90 %
Tasche mit Farbfehler : 5 %
Tasche mit Verschlussfehler : 3 %
Tasche mit V. und F. : 2 %

Produktions-
Kosten je Tasche : 80 €
Erlös je Tasche : 150 €
Farb- od. Verschl.f. : +30 € auf Kosten
Beides : neue Tasche

Erwartungswert

Erwarteter Gewinn? (X: Gewinn in €)

x_i	70	40	40	−10
$P(x_i)$	0,9	0,05	0,03	0,02

$$E(x) = 0{,}9 \cdot 70 + 0{,}05 \cdot 40 + 0{,}03 \cdot 40 + 0{,}02 \cdot (-10)$$

$$= 66\,€$$

➢ „Spiel“ ist **nicht** fair. Ein Unternehmen möchte ja einen Erwartungswert > 0…

Standardabweichung / Sigma-Umgebung

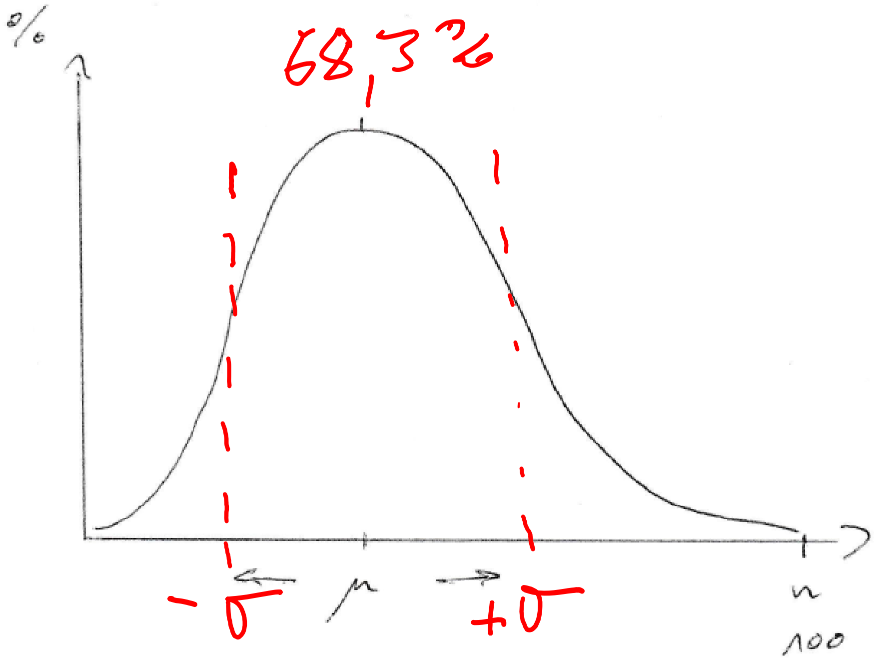
$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 68{,}3\,\%$$

$$P(\mu - 1{,}64\,\sigma \leq X \leq \mu + 1{,}64\,\sigma) = 90\,\%$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 95{,}4\,\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99{,}7\,\%$$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)}$$



Konfidenzintervall

Würfel 4000-mal werfen. 700-mal soll eine 6 fallen. Sicherheit 99,7 %

Relative Häufigkeit $h_n = \frac{X}{n} = \frac{700}{4000} = 0{,}175$

$$\frac{X}{n} \in \left[p - 3\frac{\sigma}{n} \,;\, p + 3\frac{\sigma}{n} \right]$$

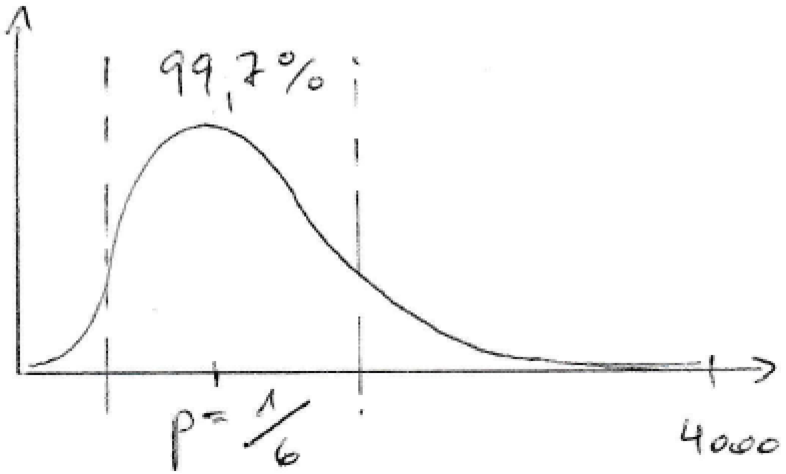
Konfidenzintervall

$$\sigma = \sqrt{n \cdot p \cdot (1-p)}$$

$$\left| \frac{X}{n} - p \right| \overset{!}{\leq} 3 \frac{\sigma}{n}$$

$$|0{,}175 - p| \leq 3 \cdot \frac{\sqrt{4000\, p(1-p)}}{4000} \quad | ()^2$$

$$\underbrace{(0{,}175 - p)^2}_{\text{2. bin. Formel}} \leq 9 \cdot \frac{p(1-p)}{4000}$$



Konfidenzintervall

$$0{,}030625 - 0{,}35p + p^2 \leq 9 \cdot \frac{p \cdot (1-p)}{4000} \quad \Big| \cdot 4000$$

$$\rightarrow \; 122{,}5 - 1400p + 4000p^2 \leq 9p - 9p^2$$

$$\Rightarrow \; 4009p^2 - 1409p + 122{,}5 \leq 0$$

$$\rightarrow \; p_1 = 0{,}1937 \qquad p_2 = 0{,}1577$$

Konfidenzintervall

- Wahrscheinlichkeit p mittels pq-Formel / Mitternachtsformel / quadratischer Ergänzung bestimmen ist die **genauste** Variante
- *Weitere Möglichkeit*: Statt p auszurechnen kann man auch die **relative Häufigkeit** einsetzen (in unserem Beispiel $h_n = 0{,}175$) und vergleichen, meist hinreichend genau!

Hypothesentests

Was sind Hypothesentests?

- Anhand einer **Stichprobe** herausfinden, ob eine Vermutung (Hypothese) wahr ist
  - Man stellt zuerst eine Nullhypothese N_0 auf
  - Dann gibt man ein Signifikanzniveau (Irrtumswahrscheinlichkeit) vor und bestimmt den Ablehnungsbereich
  - Stichprobe ziehen
  - Nach Test festlegen, ob Aussage **angenommen** oder **verworfen** werden muss

Hypothesentests

Fehler beim Testen von Hypothesen

- Nach gezogener Stichprobe ist man nun zu einem Ergebnis gekommen…
  - Es kann passieren, dass eine Hypothese als falsch angesehen wird, obwohl sie von Anfang an richtig war (**Fehler 1. Art**, α-Fehler)
  - Oder eine Hypothese ist falsch, wurde aber irrtümlich **nicht** verworfen, weil das Stichprobenergebnis noch im Annahmebereich liegt (**Fehler 2. Art**, β-Fehler)

Hypothesentests

Linksseitiger Hypothesentest

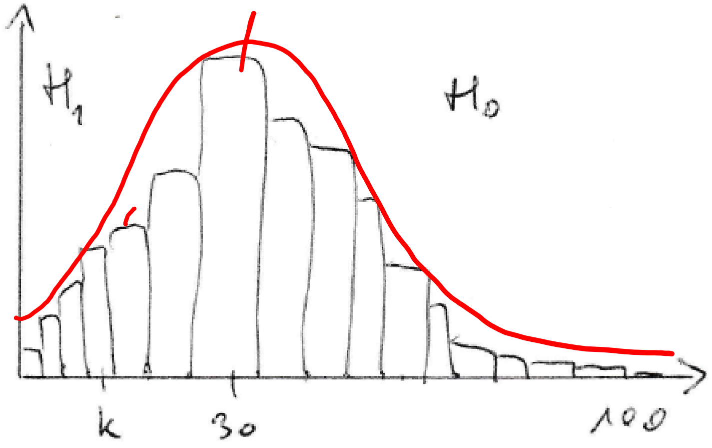
Mindestens 30% der Schüler lernen mit Videos. 100 werden befragt ($\alpha = 5\%$)

$H_0: p_0 \geq 0{,}3$

$H_1: p_1 < 0{,}3$

$\alpha = 0{,}05$

$n = 100$



Hypothesentests

Linksseitiger Hypothesentest

$$P(X \leq k_{links}) \leq \overbrace{0{,}05}^{\alpha}$$

$$F(100;\, 0{,}3;\, k_{links}) \leq 0{,}05$$

Test in Tabelle:

$$F(100;\, 0{,}3;\, 21) = 0{,}0288$$

$$F(100;\, 0{,}3;\, 22) = 0{,}0479 \quad ✓$$

$$F(100;\, 0{,}3;\, 23) = 0{,}0755 \quad ✗$$

Ablehnen / Verwerfen $\bar{A} = \{0 \ldots 22\}$

Hypothese bestätigt $A = \{23 \ldots 100\}$

Hypothesentests

Rechtsseitiger Hypothesentest

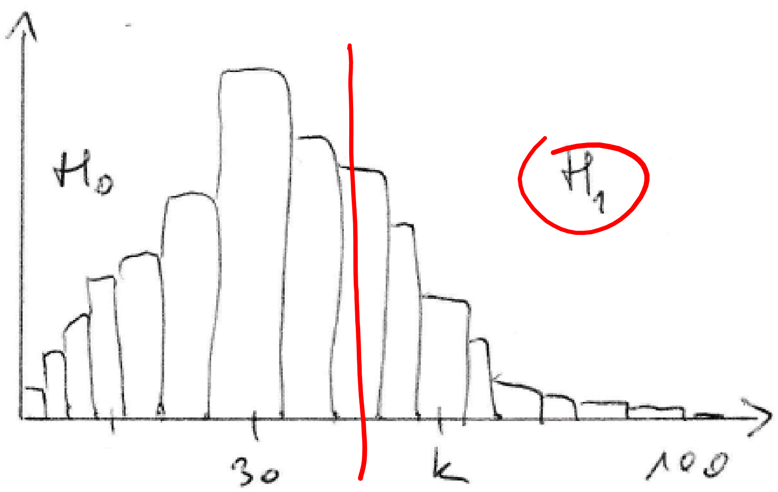
Höchstens 30% der Schüler lernen mit Videos. 100 werden befragt ($\alpha = 5\%$)

$H_0: p_0 \leq 0{,}3$

$H_1: p_1 > 0{,}3$

$\alpha = 0{,}05$

$n = 100$



Hypothesentests

Rechtsseitiger Hypothesentest

$$P(X \geq k_{rechts}) \leq \overset{\alpha}{\widetilde{0{,}05}}$$

$$1 - P(X \leq k_{rechts} - 1) \leq 0{,}05$$

$$0{,}95 \leq P(X \leq k_{rechts} - 1)$$

$$0{,}95 \leq F(100;\, 0{,}3;\, k_{rechts} - 1)$$

$$0{,}947 = F(100;\, 0{,}3;\, 37)$$

$$0{,}966 = F(100;\, 0{,}3;\, 38) \quad \checkmark$$

→ Ablehnen / Verwerfen : $\bar{A} = \{39 \ldots 100\}$

Annehmen : $A = \{0 \ldots 38\}$

Hypothesentests

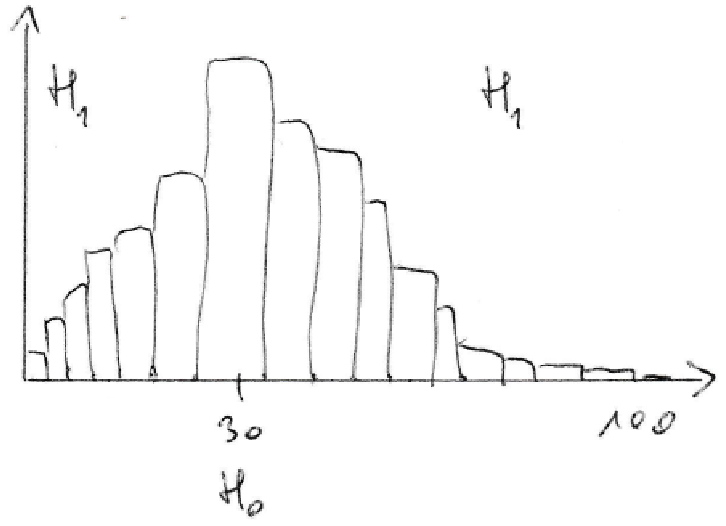
Zweiseitiger Hypothesentest

30 % der Schüler lernen mit Videos.
100 werden befragt ($\alpha = 5\%$)

H_0: $p_0 = 0{,}3$

H_1: $p_1 \neq 0{,}3$

$\alpha = 0{,}05$
$n = 100$



Hypothesentests

Zweiseitiger Hypothesentest

$0{,}05$

$\frac{1}{2}$ $\frac{1}{2}$

$\rightarrow P(X \leq k_l) \leq 0{,}025$! $\qquad P(X \geq k_r) \leq 0{,}025$

$F(100; 0{,}3; k_l) \leq 0{,}025 \qquad 0{,}975 \leq F(100; 0{,}3; k_r - 1)$

aus Tabelle $\quad F(100; 0{,}3; 20) = 0{,}0165 \qquad 0{,}979 = F(100; 0{,}3; 39)$

$\bar{A} = \{0 \dots 20\} \qquad \bar{A} = \{40 \dots 100\}$

$A = \{21 \dots 39\}$

Hypothesentests

Beispiel:

Mindestens 96 % der Ware ist OK.

→ Aussage wird bezweifelt und soll getestet werden.

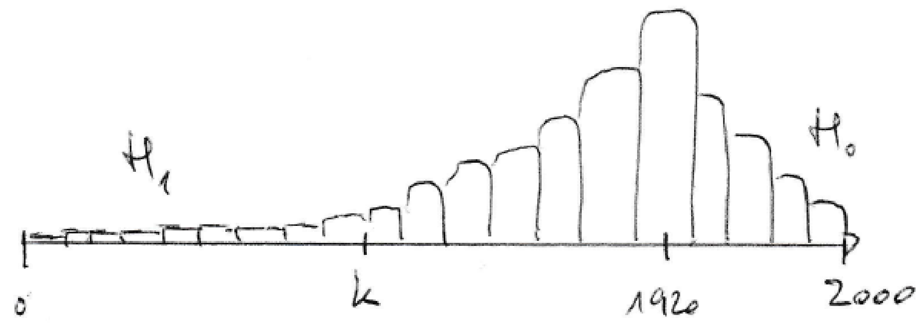
→ 2000 Stück , Irrtumswahrscheinl. 5 %

$H_0: \; p_0 \geq 0{,}96$ $\qquad n = 2000$

$H_1: \; p_1 < 0{,}96$ $\qquad \alpha = 0{,}05$

$\sigma = \sqrt{2000 \cdot 0{,}96 \cdot 0{,}04} = 8{,}76 > 3$ ✓

Hypothesentests



α	1 %	5 %	10 %
Wert	2,33	1,64	1,28

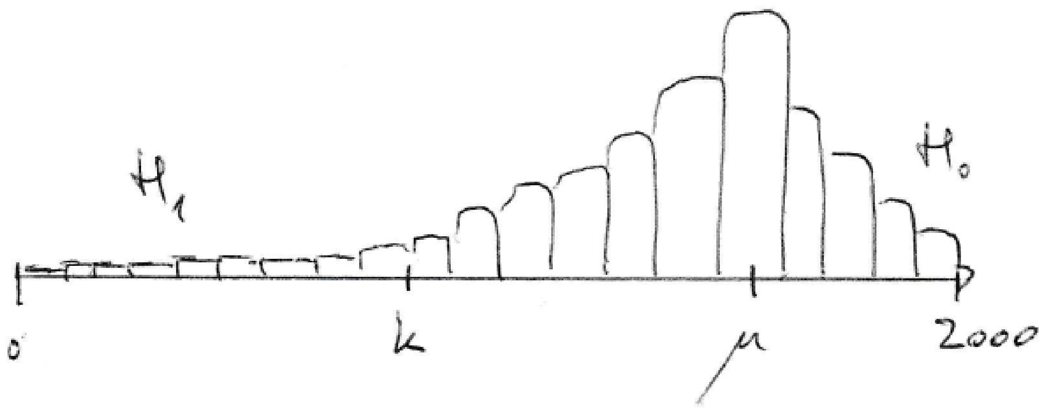
$k = \mu \pm \text{Wert} \cdot \sigma$

$\rightarrow \; k = 1920 - 1{,}64 \cdot 8{,}76 = 1905{,}63$

$\rightarrow \; \bar{A} = \{0 \ldots\ldots 1905\} \qquad A = \{1906 \ldots 2000\}$

Hypothesentests

70% wählen Mathe als ihr Lieblingsfach. Ihr behauptet es seien weniger. 2000 Schüler werden gefragt. $\alpha = 5\%$

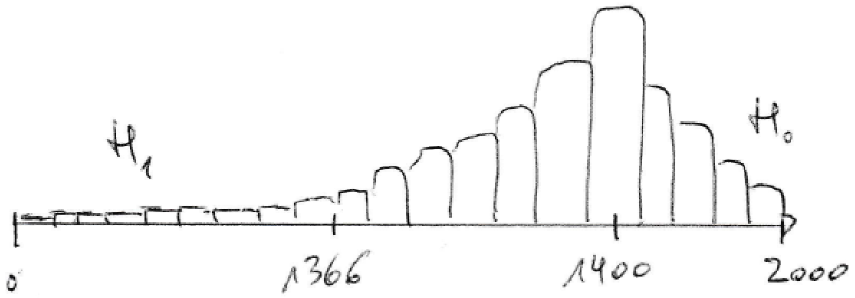


Hypothesentests

$k = \mu \pm \text{Wert} \cdot \sigma$

$\sigma = \sqrt{2000 \cdot 0{,}7 \cdot 0{,}3} = 20{,}49 > 3$ ✓

$\bar{A} = \{0 \ldots 1366\}$ $A = \{1367 \ldots 2000\}$



→ $N = \frac{1366{,}39}{2000} = 0{,}683$

→ Tatsächlich nur 68 %, die Mathe wählen.

Hypothesentests

$$\beta = P(X \geq 1367) = 1 - P(X \leq 1366)$$

$$= 1 - \phi\left(\frac{1366 - 2000 \cdot 0{,}68 + 0{,}5}{\sqrt{2000 \cdot 0{,}68 \cdot 0{,}32}}\right)$$

$$= 1 - \phi(1{,}26) = 0{,}1038 = 10{,}38\,\%$$

- Bestimmung Fehler 2. Art mittels ϕ-Funktion

Hypothesentests

$$n \geq \frac{\ln(1-a)}{\ln(1-p)}$$

a = Mindestwahrsch.

p = Trefferwahrsch.

$$n \geq \frac{\ln(1-0{,}9)}{\ln\left(1-\frac{1}{6}\right)} \approx 12{,}63 \rightarrow \text{mind. } \underline{\underline{13}}$$

Stochastische Matrix

	von V	von H
nach ~~von~~ V	0,69	0,45
nach H	0,31	0,55

	~~nach V~~	~~nach H~~

$$P = \begin{pmatrix} 0{,}69 & 0{,}45 \\ 0{,}31 & 0{,}55 \end{pmatrix}$$

Dichtefunktion

- Eine Funktion f heißt Wahrscheinlichkeitsdichte über einem Intervall I, wenn
  - $f(x) \geq 0 \; für \; alle \; x \; aus \; I$
  - $\int_a^b f(x)dx = 1$
- Zufallsgröße X ist im Intervall I **stetig verteilt (überall definiert)** mit der Wahrscheinlichkeitsdichte f, wenn gilt
  - $P(r \leq X \leq s) = \int_r^s f(x)dx$

Dichtefunktion

- Die Länge von Natursteinen weicht nach Angaben eines Herstellers vom Sollmaß um maximal $\pm$1cm ab
- Wahrscheinlichkeiten von Abweichungen sind durch eine Wahrscheinlichkeitsdichte mit $f(x) = k \cdot (1 - x^2)$ über dem Intervall [-1;1] beschrieben
- Abweichungen sind Werte einer Zufallsgröße
- a) k so bestimmen, dass f eine Wahrscheinlichkeitsdichte wird
- b) Wie groß ist die Wahrscheinlichkeit P(0,4$\leq X \leq$0,9) einer Abweichung zwischen 0,4 und 0,9 cm, wenn wir den Wert für k verwenden?

Dichtefunktion

a) $f(x) \geq 0$ für $-1 \leq x \leq 1$ und $k > 0$

wegen $\int_{-1}^{1} k(1-x^2)\,dx = k\int (1-x^2)\,dx$

$= k \cdot \left(x - \frac{1}{3}x^3\right)\Big|_{-1}^{1} = k \cdot \frac{4}{3} = 1 \iff k = \frac{3}{4}$

Dichtefunktion

b) $P(0{,}4) = 0$

$$P(0{,}4 \le X \le 0{,}9) = \int_{0{,}4}^{0{,}9} \frac{3}{4}(1-x^2)\,dx$$

$$= \frac{3}{4} \cdot \left[x - \frac{1}{3}x^3\right]_{0{,}4}^{0{,}9} = 0{,}20875 \approx 20{,}88\,\%$$