4. Stochastik

4.5 Signifikanztests

H. Wuschke

12. September 2022

Ziele der Sitzung

- Sigma-Regeln anwenden
- Begriffe *Nullhypothese*, *Gegenhypothese*, *Ablehnungsbereich*, *Vertrauensintervall*, *Signifikanz(-niveau)* bei konkreten Hypothesentests verwenden
- Links-, Rechts- und Beidseitige Hypothesentests durchführen können
- Fehler 1. Art und Fehler 2. Art deuten können

Wann kann man etwas glauben?
Woher weiß man, wie wahrscheinlich etwas ist?
„Wann ist es endlich richtig? Wann macht es einen Sinn?"

Der Traum von der idealen Münze?

Eine ideale Münze wird 100 Mal geworfen. Sei die binomialverteilte Zufallsgröße X: *Anzahl von Zahl*.

Der Erwartungswert ist $\mu = \frac{1}{2} \cdot 100 = 50$.

$$P(X = 50) \approx 0,0796$$

Die Münze wird 1.000 Mal geworfen.

Der Erwartungswert ist $\mu = \frac{1}{2} \cdot 1000 = 500$.

$$P(X = 500) \approx 0,0252$$

Wenn der Erwartungswert immer unwahrscheinlicher wird, je größer die Anzahl der Würfe wird, wie kann dann eine Münze ideal sein?

Der Traum von der idealen Münze?

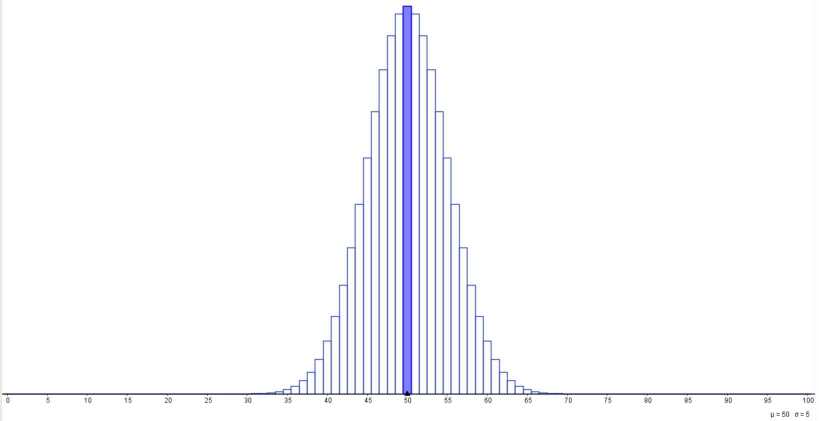


Abbildung: $P(X = 50)$ in der Binomialverteilung, GeoGebra HW 2018

$P(47 \leqq X \leqq 53) \approx 0,516 \qquad P(45 \leqq X \leqq 55) \approx 0,729$

$$P(40 \leqq X \leqq 60) \approx 0,965$$

Es liegt zu 96,5% von 40 und bis 60 Mal Zahl oben.

Sigma-Regeln, für $\sigma > 3$

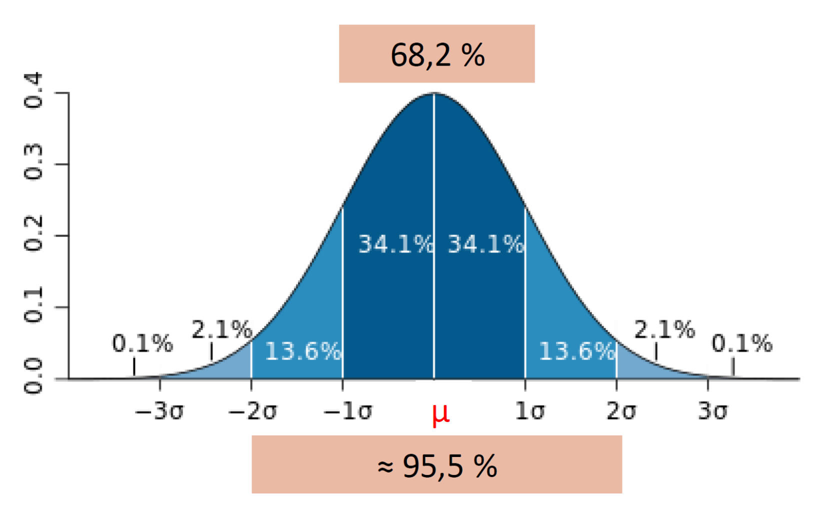


Abbildung: https://de.wikipedia.org/wiki/Normalverteilung#/media/File:Standard_deviation_diagram.svg

Sigma-Regeln

Sei $B_{n;p}$ eine Binomialverteilung einer Zufallsgröße X mit Erwartungswert μ und Standardabweichung $\sigma > 3$.
Dann gelten die **Sigma-Regeln** annähernd:

$$P(\mu - \sigma \leqq X \leqq \mu + \sigma) \approx 68,27\%$$

$$P(\mu - 1,645\sigma \leqq X \leqq \mu + 1,645\sigma) \approx 90\%$$

$$P(\mu - 1,96\sigma \leqq X \leqq \mu + 1,96\sigma) \approx 95\%$$

$$P(\mu - 2\sigma \leqq X \leqq \mu + 2\sigma) \approx 95,45\%$$

$$P(\mu - 2,576\sigma \leqq X \leqq \mu + 2,576\sigma) \approx 99\%$$

$$P(\mu - 3\sigma \leqq X \leqq \mu + 3\sigma) \approx 99,73\%$$

Bemerkung

Die Sigma-Regeln gelten für die Normalverteilung, welche für stetige Zufallsgrößen gilt. Bei diskreten Zufallsgrößen, weichen die gerundeten Werte ab, da nur natürliche Zahlen verwendet werden.

Beispiel Anwendung der Sigma-Regeln

Gegeben ist $B_{625;\frac{1}{5}}$. Dann ist $\mu = 125$ und $\sigma = 10$.

$P(\mu - \sigma \leqq X \leqq \mu + \sigma) = P(115 \leqq X \leqq 135) \approx 0,7064$
$P(\mu - 2\sigma \leqq X \leqq \mu + 2\sigma) = P(105 \leqq X \leqq 145) \approx 0,9598$
$P(\mu - 3\sigma \leqq X \leqq \mu + 3\sigma) = P(95 \leqq X \leqq 155) \approx 0,9977$

Es ist $P(\mu - 1,645\sigma \leqq X \leqq \mu + 1,645\sigma) = P(108,55 \leqq X \leqq 141,45)$. Dies ist für diskrete Zufallsgrößen nicht nutzbar. Es muss also gerundet werden.

$P(109 \leqq X \leqq 141) \approx 0,9013$ wäre eine legitime Rundung.
$P(108 \leqq X \leqq 142) \approx 0,9201$ erhöht die Wahrscheinlichkeit des Bereiches.

Prognoseintervalle bestimmen

In Deutschland haben ca. 41% der Einwohner:innen Blutgruppe 0. Es werden 1.000 Personen befragt. Bestimmen Sie für die Befragung das Prognoseintervall für eine Sicherheitswahrscheinlichkeit von mindestens 95%.

$\mu = 410$ und $\sigma \approx 15,55 > 3$, daher kann geschätzt werden, dass die Anzahl der Personen mit Blutgruppe 0 im Intervall $[410 - 1,96\sigma; 410 + 1,96\sigma]$ mit einer Sicherheitswahrscheinlichkeit von 95% liegen wird.
Rundet man die untere Grenze ab und die obere Grenze auf, erhält man $P(379 \leqq X \leqq 441) \approx 0,9572$.
Für die andere Rundung erhält man $P(380 \leqq X \leqq 440) \approx 0,9501$. Dies ist auch noch akzeptabel.
Für $P(381 \leqq X \leqq 439) \approx 0,9422$ ergibt sich erstmals ein Intervall, welches nicht mehr akzeptabel ist.
Das Prognoseintervall ist also $[380; 440]$.

Hypothese, Ablehnungsbereich, Vertrauensintervall

In der Beurteilenden Statistik untersucht man sogenannte **Hypothesen** (grch. Unterstellung). Die Ausgangsvermutung bzw. angezweifelte Aussage heißt **Nullhypothese** H_0. Wenn H_0 abgelehnt wird, geht man von der **Gegenhypothese** H_1 aus.

Für H_0 gibt es deshalb einen **Ablehnungsbereich** A, in dem die Stichprobenergebnisse enthalten sind, für die H_0 verworfen wird. Der restliche Bereich heißt **Vertrauensintervall** $\overline{A}$.

Unter einem **Test** versteht man eine Entscheidungsvorschrift zur Auswertung von Stichprobenergebnissen, aufgrund derer eine Nullhypothese entweder abgelehnt oder nicht abgelehnt wird.

Ideale Münze

Bei 100 Mal Werfen gehen wir davon aus, dass die Münze in Ordnung ist, wenn sie von 40 bis 60 Mal Zahl zeigt. Wenn sie weniger als 40 oder mehr als 61 Mal Zahl zeigt, stimmt etwas mit der Münze nicht.

Beispiele von Hypothesentests in der Wissenschaft (Auswahl)

Soziologie

- Frauen verdienen weniger als Männer (einseitiger Signifikanztest)
- Frauen und Männer haben unterschiedliche Einkommen (zweiseitiger Signifikanztest)

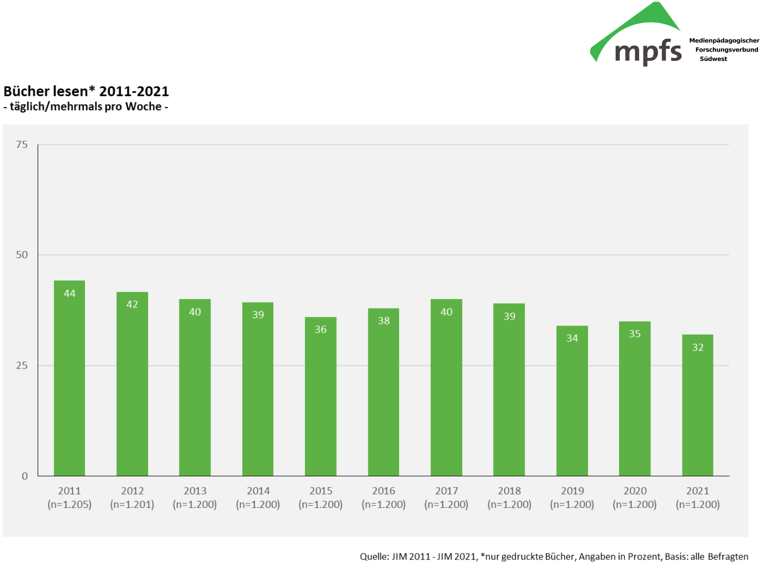
Psychologie

- Singles sind gewaltbereiter
- Ostdeutsche haben einen höheren Nationalstolz (einseitige Signifikanztests)

Medizin

- Es kommen mehr Jungen als Mädchen auf die Welt (einseitiger Signifikanztest)
- Babys werden zu 0,15% intersexuell geboren (zweiseitiger Signifikanztest)

Leseverhalten von Jugendlichen laut JIM-Studie 2021



Leseverhalten von Jugendlichen

Die JIM-Studie (Jugend Informationen Multimedia) hat 1.200 Jugendliche befragt und daraus Aussagen formuliert. Eine Aussage ist: 32% der Jugendlichen lesen täglich/mehrmals pro Woche Bücher.
These: $p < 0,32$
Durch soziale Medien, E-Books, Streaming-Dienste etc. ist das Leseverhalten von Printmedien geringer geworden.

Frage:
Ab wann gilt so eine These? Gesucht ist ein Gütekriterium.

Aufgabe:
Geben Sie für eine Stichprobe von 100 Jugendlichen an, bis zu welcher Anzahl von lesenden Jugendlichen diese These für Sie glaubwürdig wäre.

Leseverhalten (linksseitiger Hypothesentest)

$H_0 : p \geqq 0,32$ (Nullhypothese) $H_1 : p < 0,32$ (Gegenhypothese)

Binomialverteilung: $B_{100;0.32}$ Signifikanzniveau: $P(A) = \alpha \leqq 0,05$

Systematisches Probieren führt zu:

$P(X \leqq 24) \approx 0,0513 > 0,05$ und $P(X \leqq 23) \approx 0,0315 < 0,05$

Ablehnungsbereich: $A = \{0; 1; 2; \ldots; 22; 23\}$

Vertrauensintervall: $\overline{A} = \{24; 25; \ldots; 99; 100\}$

Lesen von 100 befragten Jugendlichen weniger als 24, so wird die Behauptung der JIM-Studie abgelehnt und die Gegenhypothese formuliert, dass $p < 0,32$ ist. Ab 24 lesenden Jugendlichen, darf diese Hypothese nicht verworfen werden und ihr muss vertraut werden. Dies geschieht mit einer Irrtumswahrscheinlichkeit von 5 %.

Irrtumswahrscheinlichkeit/Signifikanz

Der Ablehnungsbereich und das Vertrauensintervall werden in Abhängigkeit von einer Irrtumswahrscheinlichkeit α konstruiert. Diese Größe trifft Aussagen über die Signifikanz des Tests.

Leseverhalten (rechtsseitiger Hypothesentest)

$H_0 : p \leqq 0,32$ (Nullhypothese) $H_1 : p > 0,32$ (Gegenhypothese)

Binomialverteilung:$B_{100;0.32}$ Signifikanzniveau:$P(A) = \alpha \leqq 0,02$

Systematisches Probieren führt zu:
$P(X \geqq 42) \approx 0,0226 > 0,02$ und $P(X \geqq 43) \approx 0,0136 < 0,02$

Ablehnungsbereich: $A = \{43; 44; 45; \ldots; 99; 100\}$

Vertrauensintervall: $\overline{A} = \{0; 1; 2; \ldots; 41; 42\}$

Lesen von 100 befragten Jugendlichen mehr als 42, so wird die Behauptung der JIM-Studie abgelehnt und die Gegenhypothese formuliert, dass $p > 0,32$ ist. Bis zu 42 lesenden Jugendlichen, darf diese Hypothese nicht verworfen werden und ihr muss vertraut werden. Dies geschieht mit einer Irrtumswahrscheinlichkeit von 2 %.

Leseverhalten (beidseitiger Hypothesentest)

$H_0 : p = 0,32$ (Nullhypothese) $H_1 : p \neq 0,32$ (Gegenhypothese)

Binomialverteilung: $B_{100;0.32}$ Signifikanzniveau: $P(A) = \alpha \leqq 0,01$

Signifikanzniveau links: $\frac{\alpha}{2} \leqq 0,005$ Signifikanzniveau rechts: $\frac{\alpha}{2} \leqq 0,005$

Systematisches Probieren führt zu:

$P(X \leqq 20) \approx 0,0054 > 0,005$ und $P(X \leqq 19) \approx 0,0027 < 0,005$

$P(X \geqq 44) \approx 0,0079 > 0,005$ und $P(X \geqq 45) \approx 0,0044 < 0,005$

Ablehnungsbereich: $A = \{0; 1; 2; \ldots; 18; 19\} \cup \{45; 46; 47; \ldots; 99; 100\}$

Vertrauensintervall: $\overline{A} = \{20; 21; 22; \ldots; 43; 44\}$

Lesen von 100 befragten Jugendlichen mehr als 44 oder weniger als 20, so wird die Behauptung der JIM-Studie abgelehnt und die Gegenhypothese formuliert, dass $p \neq 0,32$ ist. Dies geschieht mit einem Signifikanzniveau von 1 %.

signifikante Hypothesentests

Ein Hypothesentests heißt **signifikant**, wenn für die Wahrscheinlichkeit des Ablehnungsbereiches gilt:

$$P(A) = \alpha < 0,05$$

Ein Hypothesentests heißt **hochsignifikant**, wenn für die Wahrscheinlichkeit des Ablehnungsbereiches gilt:

$$P(A) = \alpha < 0,01$$

Ein Hypothesentests heißt **höchstsignifikant**, wenn für die Wahrscheinlichkeit des Ablehnungsbereiches gilt:

$$P(A) = \alpha < 0,001$$

Wir sprechen dann von einem **Signifikanztest** mit Signifikanzniveau α und einer Güte von $1 - \alpha$.

Fehler bei der Testung von Hypothesen

In der Grundgesamtheit gilt: Wegen der Stichprobe wird	H_0 ist wahr	H_0 ist falsch
H_0 abgelehnt	Entscheidung falsch (Fehler 1. Art, α)	Entscheidung richtig
H_0 nicht abgelehnt	Entscheidung richtig	Entscheidung falsch (Fehler 2. Art, β)

Wird α immer kleiner, so wird β immer größer. Deshalb sollte α immer das „schlimmere Übel“ von beiden sein.

Beispiel – H_0: Die Milch ist abgelaufen

Fehler 1. Art: Die Milch ist abgelaufen, man nimmt aber an, dass sie gut ist.
Fehler 2. Art: Die Milch ist gut, man nimmt aber an, dass sie abgelaufen ist.

Im Vertrauensintervall

Interessiert mich nicht, weil ich dafür keine Kriterien habe, falls H_0 falsch ist.

Ablehnung

entweder alles O.K. und ich lehne ab
oder
α ist so klein, dass ich Rest verwerfen kann

Abbildung: Was heißt es, den Fehler 1. Art kleiner zu machen? CC0