

4.4 Prognosen mithilfe der Sigma-Regeln

- Die Abbildung zeigt die Ergebnisse der Bundestagswahl 2021. Im Wahlkreis Berlin-Mitte wurden 152 513 Stimmen abgegeben.
- Berechnen Sie die Wahrscheinlichkeit dafür, dass Union und SPD jeweils zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$ Stimmen erhielten.
 - In Berlin-Mitte bekam die Union 16 856 und die SPD 32 133 Stimmen. Beurteilen Sie diese Ergebnisse mithilfe von a). ■



Sigma-Regeln

Ein Bereich der beurteilenden Statistik beschäftigt sich mit dem **Schluss von der Gesamtheit auf die Stichprobe**. Der Anteil p der Objekte mit einer bestimmten Eigenschaft innerhalb einer Gesamtheit ist bekannt. Es werden Prognosen über den Anteil innerhalb einer Stichprobe abgegeben. Dazu wird die Situation mit einer Bernoulli-Kette beschrieben.

Gesamtheit	Eigenschaft	Anteil	Stichprobe
Bevölkerung Deutschlands	weiblich	51%	1000 Personen

Eine plausible Prognose für den zu erwartenden Wert von X liefert der Erwartungswert μ . Die Wahrscheinlichkeit für einzelne Werte wie μ ist jedoch sehr gering. Aussagekräftiger sind Intervalle um den Erwartungswert, in denen der Wert von X mit einer vorgegebenen Mindestwahrscheinlichkeit liegt (hier 80 %). Mit einer Wahrscheinlichkeit von mindestens 80 % liegt die Zahl der Frauen also zwischen 490 und 530. Mithilfe der Standardabweichung σ lassen sich diese Intervalle zu den Mindestwahrscheinlichkeiten von 68 %, 95,5 % und 99,7 % direkt angeben.

X : Anzahl der Frauen innerhalb der Stichprobe ($X \sim B_{1000; 0,51}$)

Prognosewert:

$$\mu = 1000 \cdot 0,51 = 510 \quad P(X = \mu) \approx 0,025$$

Intervall als Prognose:

k	$[\mu - k; \mu + k]$	$P(\mu - k \leq X \leq \mu + k)$
1	[509; 511]	0,0756
2	[508; 512]	0,1256
3	[507; 513]	0,1752
...	...	...
19	[491; 529]	0,7826
20	[490; 530]	0,8053

Satz: σ -Regeln

X sei eine binomialverteilte Zufallsgröße mit den Parametern n und p , dem Erwartungswert $\mu = n \cdot p$ und der Standardabweichung $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$.

Falls die **Laplace-Bedingung** $\sigma > 3$ erfüllt ist, liegt die Trefferzahl näherungsweise mit den Wahrscheinlichkeiten 68 %, 95,5 % und 99,7 % in den angegebenen Intervallen.

σ -Umgebung	Wahrscheinlichkeit
$[\mu - \sigma; \mu + \sigma]$	68 %
$[\mu - 2\sigma; \mu + 2\sigma]$	95,5 %
$[\mu - 3\sigma; \mu + 3\sigma]$	99,7 %

Die Intervalle, die sich aus den σ -Regeln ergeben, heißen **1 σ -Umgebung**, **2 σ -Umgebung** bzw. **3 σ -Umgebung** zum Erwartungswert μ . Die Wahrscheinlichkeit, dass die Trefferanzahl X in einer σ -Umgebung liegt, ändert sich nicht, wenn man die Intervallgrenzen „nach innen“ auf ganze Zahlen rundet, denn X ist immer ganzzahlig.

Beispielsweise gilt für $n = 100$, $p = 0,3$, $\mu = 30$ und $\sigma \approx 4,58$:

$$1\sigma\text{-Umgebung: } P(\mu - \sigma \leq X \leq \mu + \sigma) = P(25,4 \leq X \leq 34,6) = P(26 \leq X \leq 34)$$

$$2\sigma\text{-Umgebung: } P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(20,8 \leq X \leq 39,2) = P(21 \leq X \leq 39)$$

$$3\sigma\text{-Umgebung: } P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(16,3 \leq X \leq 43,7) = P(17 \leq X \leq 43)$$

Hinweis:

Beim „Runden nach innen“ wird die untere Grenze auf- und die obere Grenze abgerundet. Das Intervall wird kleiner.

Beispiel 1: Laut Umfrage fahren durchschnittlich 40% der 1100 Schüler eines Gymnasiums mit dem Bus zur Schule. Die Zufallsgröße X zählt an einem ausgewählten Tag, wie viele Schüler mit dem Bus kommen. Stellen Sie mit einer Wahrscheinlichkeit von 95,5% eine Prognose dafür auf, wie viele Schüler an diesem Tag mit dem Bus kommen.

Lösung:

Berechnen Sie μ und σ . Prüfen Sie, ob die Laplace-Bedingung erfüllt ist.
Die Trefferanzahl liegt nach der 2σ -Regel mit etwa 95,5% Wahrscheinlichkeit im Intervall $[\mu - 2\sigma; \mu + 2\sigma]$.

$$\begin{aligned}\mu &= 1100 \cdot 0,4 = 440 \\ \sigma &= \sqrt{1100 \cdot 0,4 \cdot 0,6} \approx 16,25 > 3 \\ \mu - 2\sigma &\approx 407,5; \quad \mu + 2\sigma \approx 472,5 \\ 2\sigma\text{-Umgebung: } &[407,5; 472,5]\end{aligned}$$

Runden Sie $\mu - 2\sigma$ auf die nächste ganze Zahl auf und $\mu + 2\sigma$ ab.

Mit etwa 95,5% Wahrscheinlichkeit kommen zwischen 408 und 472 Schüler mit dem Bus.

Basisaufgaben

1. Die Zufallsgröße X ist binomialverteilt mit:

① $n = 250; p = 0,8$

② $n = 300; p = 0,5$

③ $n = 1000; p = 0,05$

④ $n = 2500; p = 0,75$

- Bestimmen Sie jeweils die 1σ -, die 2σ - und die 3σ -Umgebung.
- Berechnen Sie die Wahrscheinlichkeiten dafür, dass X in diesen Umgebungen liegt.
- Zeigen Sie jeweils, dass die Laplace-Bedingung erfüllt ist. Vergleichen Sie die Ergebnisse aus b) mit den σ -Regeln.

2. Eine 1-Cent-Münze wird 300-mal (400-mal; 500-mal) geworfen.

- Bestimmen Sie ein um den Erwartungswert symmetrisches Intervall, sodass die Anzahl der Würfe mit „Zahl“ mit einer Wahrscheinlichkeit von etwa 68% (99,7%) in diesem Intervall liegt.
- Bestimmen Sie die 2σ -Umgebung um den Erwartungswert sowie die Wahrscheinlichkeit dafür, dass die Anzahl der Würfe mit „Zahl“ in dieser Umgebung liegt.

3. Bei einer Lotterie auf einem Jahrmarkt sind 10% aller Lose Gewinne.

- An einem normalen Tag werden etwa 1000 Lose verkauft. Bestimmen Sie die 1σ - und die 3σ -Umgebung und erklären Sie die Bedeutung dieser Intervalle.
- Ein Los kostet 0,50 €. 2% aller Lose sind Hauptgewinne im Wert von 15 € und 8% aller Lose sind normale Gewinne im Wert von 2 €. Der Veranstalter rechnet in einer Woche mit etwa 5000 verkauften Losen.

Bestimmen Sie die 2σ -Umgebungen mit ganzzahligen Grenzen für die Anzahl der Hauptgewinne und die Anzahl der normalen Gewinne in dieser Woche.

- Berechnen Sie mithilfe von b), wie hoch die Ausgaben für Gewinne für den Betreiber im besten Fall und im schlechtesten Fall zu etwa 95,5% sind.



Hinweis:

In Sachzusammenhängen rundet man σ -Umgebungen meist nach innen auf ganzzahlige Grenzen, da die Zufallsgröße nur ganzzahlige Werte annehmen kann.

4. Die Zufallsgröße X ist binomialverteilt mit $p = 0,5$.

- Berechnen Sie die Wahrscheinlichkeit $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$ der 2σ -Umgebung für
① $n = 10$, ② $n = 40$, ③ $n = 400$, ④ $n = 4000$.
- Prüfen Sie, ob der Wert 95,5% der 2σ -Regel die Ergebnisse aus a) mit steigendem n besser annähert. Prüfen Sie für jedes n auch, ob die Laplace-Bedingung erfüllt ist.

Schluss von der Stichprobe auf die Gesamtheit

Der genaue Wert der Trefferwahrscheinlichkeit p ist häufig nicht bekannt. Oft aber gibt es für p eine Angabe aus früheren Untersuchungen oder Umfragen. Dieser Wert von p kann mithilfe einer Stichprobe beurteilt werden.

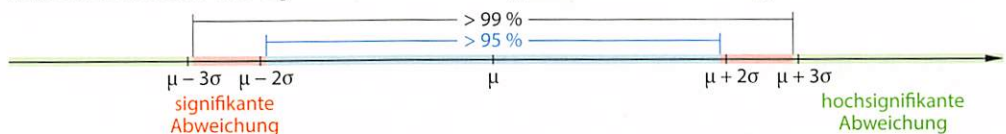
Wissen: Signifikante und hochsignifikante Abweichungen

Liegt ein Stichprobenergebnis außerhalb eines um μ symmetrischen Intervalls, in dem die Trefferzahlen mit einer Wahrscheinlichkeit von etwa 95 % bzw. 99 % liegen, so spricht man von einer **signifikanten** bzw. **hochsignifikanten Abweichung** vom Erwartungswert.

Hinweis:

Der Umkehrschluss gilt nicht. Liegt ein Stichprobenergebnis innerhalb der 2σ -Umgebung, kann es sich dennoch um eine signifikante Abweichung handeln (siehe Aufgabe 11).

In diesen Fällen kann der Wert von p angezweifelt werden. Mit den σ -Regeln lässt sich dies schnell überprüfen: liegt ein Stichprobenergebnis außerhalb der 2σ - bzw. 3σ -Umgebung, so handelt es sich um eine signifikante bzw. hochsignifikante Abweichung.



Beispiel 2: Bei der letzten Erhebung lag der Anteil an Vegetariern bei $p = 0,06$. In einer aktuellen Umfrage von 300 Personen geben 28 Personen an, Vegetarier zu sein. Beurteilen Sie die Angabe für p anhand des aktuellen Umfrageergebnisses.

Lösung:

Berechnen Sie μ und σ . Prüfen Sie, ob die Laplace-Bedingung erfüllt ist.
Bestimmen Sie die 2σ -Umgebung.

$$\begin{aligned}\mu &= 300 \cdot 0,06 = 18 \\ \sigma &= \sqrt{16,92} \approx 4,11 > 3 \\ [\mu - 2\sigma; \mu + 2\sigma] &= [9,78; 26,22]\end{aligned}$$

Prüfen Sie, ob das Stichprobenergebnis außerhalb der 2σ -Umgebung liegt.

$X = 28$ liegt nicht im Intervall $[9,78; 26,22]$ und weicht somit signifikant von μ ab.
Der Wert für p kann angezweifelt werden.

Basisaufgaben

5. Gegeben sind ein Stichprobenumfang n und eine Trefferwahrscheinlichkeit p .

① $n = 250; p = 0,4$

② $n = 1000; p = 0,25$

③ $n = 100; p = 0,8$

- Zeigen Sie, dass die Laplace-Bedingung für ①, ② und ③ erfüllt ist.
- Bestimmen Sie für ①, ② und ③ jeweils die 2σ - und die 3σ -Umgebung.
- Geben Sie jeweils eine Trefferanzahl an, welche vom Erwartungswert signifikant bzw. hochsignifikant abweicht.

6. Auf einem Jahrmarkt wirbt eine Losbude damit, dass 20 % aller Lose Gewinne seien. Eine Schülergruppe beobachtet den Stand eine Stunde lang. In dieser Zeit werden 255 Lose verkauft und dabei 34 Gewinne ausgegeben.

- Bestimmen Sie unter der Annahme von $p = 0,2$ und $n = 255$ die 2σ -Umgebung.
- Beurteilen Sie $p = 0,2$ anhand des konkreten Stichprobenergebnisses.
- Weisen Sie nach, dass die Beobachtungen in den folgenden drei Stunden signifikant vom Erwartungswert für $p = 0,2$ abweichen.

① 205 Lose, 26 Gewinne

② 310 Lose, 42 Gewinne

③ 280 Lose, 71 Gewinne

7. Eine Süßigkeitenfirma produziert Bonbons in unterschiedlichen Geschmacksrichtungen. Zum Firmenjubiläum gibt es eine Sondermischung mit 50 % Erdbeerbonbons. Thomas kauft zwei Packungen mit je 40 Bonbons und zählt die Anzahl der Erdbeerbonbons.
- Berechnen Sie $P(X = 40)$ sowie $P(X \geq 40)$ für eine geeignete binomialverteilte Zufallsvariable X und deuten Sie die Ergebnisse im Sachzusammenhang.
 - In beiden Familienpackungen zusammen findet Thomas 54 Erdbeerbonbons. Begründen Sie, dass er mithilfe dieser Stichprobe den Anteil von 50 % Erdbeerbonbons aufgrund einer signifikanten Abweichung anzweifeln sollte.
 - Untersuchen Sie, ob Thomas den Anteil von 50 % Erdbeerbonbons auch aufgrund einer hochsignifikanten Abweichung anzweifeln sollte.
8. Es wird angenommen, dass 30 % der Bevölkerung Deutschlands an einer Allergie leiden. In einer aktuellen Umfrage werden in verschiedenen Bundesländern jeweils 1000 Personen befragt.
- Begründen Sie, dass die Anzahl der Allergiker in einer Stichprobe als binomialverteilt angenommen werden kann.
 - Ermitteln Sie, welche Umfrageergebnisse außerhalb der 3σ -Umgebung um den Erwartungswert zum vermuteten Anteil an Allergikern liegen.

Bundesland	Allergiker in der Umfrage
Bayern	321
Berlin	617
Hamburg	235
Hessen	262
Niedersachsen	317
Saarland	285
Thüringen	305

Weiterführende Aufgaben

9. Der Anteil an Rechtshändern beträgt in Deutschland etwa 88 %. Ein Handballverein hat insgesamt 250 Mitglieder. Gesucht sind Prognosen über die zu erwartende Anzahl der Rechtshänder im Verein.
- Berechnen Sie den Erwartungswert für die Anzahl der Rechtshänder unter den Mitgliedern und erläutern Sie seine Bedeutung für eine Prognose über die Stichprobe.
 - Geben Sie als Prognose ein möglichst kleines Intervall um den Erwartungswert an, in dem die Anzahl der Rechtshänder im Verein mit einer Wahrscheinlichkeit von mindestens 95,5 % bzw. 99,7 % liegt.
 - Im Verein gibt es 205 Rechtshänder. Beurteilen Sie diesen Wert mithilfe von b).



10. **Stolperstelle:** Mark berechnet für Binomialverteilungen mit $n = 20$ in der Tabelle Wahrscheinlichkeiten für 1σ -Umgebungen und behauptet: „Die Werte liegen nicht in der Nähe von 0,68. Die Sigma-Regeln gelten nicht für p nahe 0 oder 1.“

- Erklären Sie, warum diese Behauptung nicht stimmt. Was hat Mark nicht beachtet?
- Zeigen Sie mit geeigneten Werten, dass die Sigma-Regeln auch für $p = 0,05$ gelten.

	A p	B μ	C σ	D wkeit
1	0.05	1.	0.974679	0.377354
2	0.1	2.	1.34164	0.74547
3	0.9	18.	1.34164	0.74547
4	0.95	19.	0.974679	0.377354
5				

11. Ein Würfel wird 800-mal geworfen. Dabei fallen insgesamt 154 Sechsen.
- Prüfen Sie, ob das Ergebnis in der 2σ -Umgebung für die Anzahl an Sechsen liegt.
 - Weisen Sie nach, dass es sich bei dem Ergebnis um eine signifikante Abweichung vom Erwartungswert handelt. Zeigen Sie hierzu, dass die Anzahl an Sechsen mit einer Wahrscheinlichkeit von etwa 95 % im Intervall $[\mu - 1,96\sigma; \mu + 1,96\sigma]$ liegt.

- 12. Ein LED-Lampen-Hersteller produziert und verkauft Lampen für 3 € pro Stück. Etwa 5 % seiner Lampen sind fehlerhaft. Die Lieferungen an Großhändler erfolgen immer in 500er-Paketen. Die Händler überprüfen die Qualität der Lampen stichprobenartig.



- a) Großhändler und Lieferant vereinbaren: „Wenn mehr als 4 % der LEDs in einer Stichprobe defekt sind, dann verringert sich der Preis auf 2,50 € pro Stück.“
Bestimmen Sie die Wahrscheinlichkeit einer Preisreduzierung für einen Stichprobenumfang von ① $n = 250$ und ② $n = 400$.
- b) Bestimmen Sie den Erwartungswert für die Kosten einer Lieferung bei einer Stichprobe von ① $n = 250$ und ② $n = 400$.
- c) Der Großhändler nimmt Stichproben von $n = 300$ Stück und prüft die Lampen. Bestimmen Sie eine 2σ -Umgebung für die Anzahl der defekten LEDs in dieser Stichprobe. Erläutern Sie die Bedeutung dieses Intervalls.
- d) Begründen Sie, dass eine Stichprobe mindestens den Umfang $n = 190$ haben sollte.

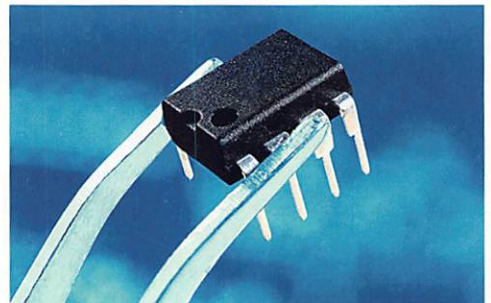
- 13. Der Zeitungsartikel beschreibt die Situation um Schwarzfahrer in Deutschland. Kontrolleure können im Tagesdurchschnitt etwa 750 Personen in einer Schicht kontrollieren.

Erhöhtes Bußgeld ohne Effekt

Nach einer Umfrage liegt der Anteil der Personen, die den öffentlichen Personennahverkehr ohne gültiges Ticket nutzen, in den größten deutschen Städten bei etwa 2,5 %. Diese „Schwarzfahrerquote“ hat sich in den letzten Jahren nicht verändert und auch die letzte Erhöhung des Bußgeldes von 40 auf 60 Euro hat die Zahlungsmoral der Schwarzfahrer scheinbar nicht erhöht.

- a) Bestimmen Sie ein Intervall für die Anzahl der Schwarzfahrer, mit denen die Kontrolleure in einer Schicht zu etwa 95,5 % rechnen können.
- b) Bestimmen Sie zu a) die Einnahmen, mit denen ein Kontrolleur in einer Schicht zu etwa 95,5 % rechnen kann.
- c) Ein Kontrolleur sagt in einem Interview: „An einem normalen Tag nehmen wir zwischen 900 € und 1380 € durch Bußgelder ein“. Untersuchen Sie, auf welche Sigma-Umgebung sich der Kontrolleur bezieht.

- 14. Jeder von der Firma Untec in Massenproduktion hergestellte Mikrochip ist mit einer Wahrscheinlichkeit von 7,5 % fehlerhaft. In Stichproben von jeweils 300 Chips wird geprüft, ob sich der Anteil fehlerhafter Chips verändert.



- a) Geben Sie eine Prognose für den Mittelwert der Anzahl der fehlerhaften Chips innerhalb von Stichproben von 300 Exemplaren.
- b) Simon sagt: „Wenn in einer Stichprobe nicht genau 7,5 % der Chips fehlerhaft sind, dann deutet das auf eine Veränderung des Anteils hin.“ Beurteilen Sie diese Aussage.
- c) Bestimmen Sie das kleinstmögliche Intervall mit dem Mittelpunkt 22,5, in dem bei einer Stichprobe von 300 Chips die Anzahl der fehlerhaften Chips mit einer Wahrscheinlichkeit von mindestens 95,5 % liegt.
- d) Beschreiben Sie, wie Sie das Intervall aus c) nutzen können, um Stichproben von 300 Chips zu beurteilen. Nutzen Sie dabei auch den Begriff „Streuung“.

15. Nehmen Sie begründet Stellung zur Aussage.

- Die 2σ -Umgebung ist immer in der 3σ -Umgebung enthalten.
- Wenn ein Ergebnis signifikant vom Erwartungswert abweicht, dann gilt dies auch für alle Ergebnisse, die noch weiter vom Erwartungswert entfernt liegen.
- Wenn ein Stichprobenergebnis signifikant vom Erwartungswert eines vermuteten Werts für p abweicht, dann bedeutet das, dass der Wert von p falsch ist.
- Wird ein Wert von p anhand einer Stichprobe verworfen, so entspricht der Anteil der Treffer beim Stichprobenergebnis dem richtigen Wert von p .

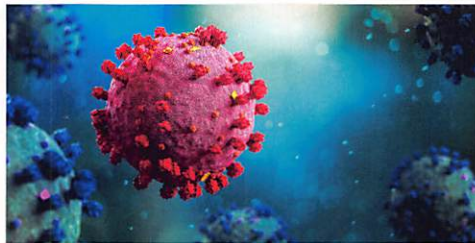
16. **Kritische Bewertung von Stichprobenergebnissen:** In einer Gemeinde sind laut einer Umfrage 32 % aller Schüler aktiv in einem Sportverein. Schülerin Sarah ist in einem Schwimmverein und sagt: „Das Ergebnis der Umfrage kann nicht stimmen, in meinem Freundeskreis sind sogar 27 von 42 Schülern in einem Sportverein.“

- Begründen Sie mithilfe einer geeigneten Rechnung, warum Sarah das Umfrageergebnis anzweifelt.
- Erläutern Sie, warum Sarahs Freundeskreis als Stichprobe ungeeignet ist.

17. Laut der Statistik des Robert-Koch-Instituts vom September 2021 mussten insgesamt 285 019 der 4 046 112 bestätigten Covid-Fälle hospitalisiert werden.

Berkan: „An meiner Schule und der Schule eines Freundes gab es insgesamt 367 Fälle. Aber keiner von denen ist deshalb im Krankenhaus gelandet.“

Katrin: „In meiner großen Familie gab es 14 Covid-Fälle und 5 davon mussten deswegen im Krankenhaus behandelt werden.“



- Begründen Sie mithilfe einer geeigneten Rechnung, dass beide Stichprobenergebnisse sehr unwahrscheinlich sind.
- Diskutieren Sie mögliche Gründe für diese Ergebnisse unter der Annahme, dass die statistischen Daten korrekt sind.

18. In einem Supermarkt kann jeder Kunde, der für mindestens 20 € einkauft, ein Los ziehen und mit etwas Glück ein Getränk oder einen 10-€-Gutschein gewinnen. Der Supermarkt behauptet, dass die Chancen für ein Getränk und einen Gutschein gleich sind. An einem Tag gibt es insgesamt 48 Gewinne, darunter sind nur 15 Gutscheine. Prüfen Sie, ob man die Behauptung des Supermarkts aufgrund dieses Ergebnisses anzweifeln kann.

● 19. **Ausblick: Testen von p**

Es wird vermutet, dass eine Reißzwecke mit einer Wahrscheinlichkeit von $p = 0,4$ mit der Spitze nach oben landet. Dieser Wert soll anhand einer Stichprobe mit 200 Würfeln überprüft werden. Es wird eine **Entscheidungsregel** formuliert, die den Wert von p verwirft, wenn das Stichprobenergebnis außerhalb der 2σ -Umgebung des Erwartungswerts liegt.

- Vervollständigen Sie die folgende Entscheidungsregel: „Der angenommene Wert $p = 0,4$ wird genau dann verworfen, wenn in der Stichprobe ...“
- Bei den 200 Würfeln landet die Reißzwecke 74-mal (94-mal) mit der Spitze nach oben. Entscheiden Sie mit der Regel aus a), ob $p = 0,4$ verworfen werden sollte.
- Erläutern Sie, welche Fehler auftreten können, wenn man sich nach der Regel in a) anhand einer konkreten Stichprobe für bzw. gegen den Wert von p entscheidet.
- Beim Testen von p liefert eine Stichprobe ein Ergebnis, das außerhalb der 2σ -Umgebung liegt. Geben Sie zwei mögliche Gründe für solch ein Ergebnis an.