8 Experimentieren: Sigma-Regeln

Didaktische Hinweise

Mit dieser Station wird ein Unterrichtsbeispiel zur Einführung der Sigma-Regeln vorgestellt, die von den Schülerinnen und Schülern an mehreren Beispielen angewendet und überprüft werden sollen. Die Aufgaben können in individualisierten Lernphasen oder in Gruppen bearbeitet und präsentiert werden.

Bei einigen Aufgaben wird auch die Frage nach der notwendigen Stichprobengröße aufgeworfen. Die Station stellt daher auch eine Vorbereitung auf die Einführung des „Zentralen Grenzwertsatzes für die Binomialverteilung" dar, also der Approximation der Binomialverteilung durch die Normalverteilung, für die die Sigma-Regeln exakt gelten.

Zur Lösung der untenstehenden Aufgaben kann das vorgegebene GeoGebra-Arbeitsblatt verwendet werden oder ein neues GeoGebra-Arbeitsblatt erstellt werden. Tipps dazu werden im Anschluss an den Schülerarbeitsauftrag gegeben.

σ-Regeln für eine binomialverteilte Zufallsvariable

Für genügend große Stichproben (*) gilt:

Die Wahrscheinlichkeit, dass die Anzahl der „Treffer"

- in der **1σ-Umgebung** um den Erwartungswert liegt, beträgt etwa **68 %**
- in der **2σ-Umgebung** um den Erwartungswert liegt, beträgt etwa **95,5 %**
- in der **3σ-Umgebung** um den Erwartungswert liegt, beträgt etwa **99,7 %.**

Die Wahrscheinlichkeit beträgt

- **90 %**, wenn die Zahl der Treffer im Intervall **[$\mu - 1{,}64\,\sigma$; $\mu + 1{,}64\,\sigma$]** liegt
- **95 %**, wenn die Zahl der Treffer im Intervall **[$\mu - 1{,}96\,\sigma$; $\mu + 1{,}96\,\sigma$]** liegt
- **99 %**, wenn die Zahl der Treffer im Intervall **[$\mu - 2{,}58\,\sigma$; $\mu + 2{,}58\,\sigma$]** liegt
- **99,9 %**, wenn die Zahl der Treffer im Intervall **[$\mu - 3{,}29\,\sigma$; $\mu + 3{,}29\,\sigma$]** liegt.

(*) Die Stichprobe ist insbesondere dann genügend groß, wenn $\sigma > 3$ ist.

Beispiel

In einer Population tritt eine bestimmt Krankheit mit der Wahrscheinlichkeit 5 % auf. Aus dieser Population werden zufällig 200 Individuen ausgewählt und untersucht.

$n = 200$ .... $p = 0{,}05$.................... $\mu = 10$...................... $\sigma = 3{,}082$.................

2σ-Umgebung: [4; 16] → Mit etwa 95,5 % Wahrscheinlichkeit sind mindestens 4 und höchstens 16 Individuen krank.

1,64 σ-Umgebung: [5; 15] → Mit etwa 90 % Wahrscheinlichkeit sind 5 bis 15 Individuen krank.

Arbeitsauftrag für SchülerInnen

Sie wählen in der Gruppe eine der Aufgaben, die Sie gemeinsam bearbeiten. Mit dem GeoGebra-Arbeitsblatt *Sigma_Regeln.ggb* können Sie Ihre Ergebnisse überprüfen.

Wenn Sie Hilfe brauchen, können Sie sich Tipps zur Aufgabe oder Tipps zu GeoGebra anschauen.

1. Aufgabe

1.1. Für einen Flug eines Airbus A300 der Lufthansa mit 270 Plätzen rechnet das Unternehmen erfahrungsgemäß mit Stornierungen von 10 % der gebuchten Plätze.
Wie groß ist die Wahrscheinlichkeit, dass bei 280 Buchungen nicht alle Plätze besetzt sind?

1.2. Stellen Sie mit Hilfe der Sigma-Regeln eine Prognose auf, wie viele Plätze mit 99,9 % Wahrscheinlichkeit benötigt werden, wenn 300 Buchungen angenommen werden.
Berechnen Sie die Wahrscheinlichkeit, dass Ihre Prognose zutrifft.

1.3. Wie viele Buchungen höchstens dürfen angenommen werden, wenn die vorhandenen Plätze in einem Airbus A300 mit 99 % Wahrscheinlichkeit ausreichen sollen.

2. Aufgabe

2.1. Die Wahrscheinlichkeit, dass beim Lottospiel „6 aus 49" eine bestimmte Zahl gezogen wird, beträgt $\frac{6}{49}$. Mit welcher Wahrscheinlichkeit, wird bei 52 Lottospielen höchstens fünf Mal die Zahl 13 gezogen?

2.2. Seit 1955 wurden insgesamt 3858 Lottospiele „6 aus 49" durchgeführt. In wie viel Spielen war mit der Ziehung der Zahl 13 zu rechnen. Geben Sie das Intervall an, in dem mit 95 % Wahrscheinlichkeit die Ziehungshäufigkeit der Zahl 13 liegt.

2.3. Aus der unter www.lottozahlenonline.de veröffentlichten Statistik geht hervor, dass die 13 in den 3858 Lottospielen am seltensten gezogen wurde, nämlich nur 410 Mal. Kann man aufgrund dieses Ergebnisses mit 99,9 % Sicherheit behaupten, dass die Ziehungshäufigkeit der 13 zu niedrig ist?

3. Aufgabe

In Deutschland erkranken von insgesamt etwa 13 Millionen Kindern unter 15 Jahren jährlich etwa 610 Kinder an Leukämie. In den Elbmarschen, der Landschaft, in der das Kernkraftwerk Krümmel und der Forschungsreaktor Geesthacht liegen, sind in den Jahren 1990 bis 2005 19 Kinder unter 15 Jahren an Leukämie erkrankt.

3.1. Mit welcher Wahrscheinlichkeit erkranken jährlich höchstens 50 von den zurzeit rund 1,5 Millionen Kindern unter 15 Jahren in Baden-Württemberg an Leukämie?

3.2. Wie viele Neuerkrankungen an Leukämie waren nach der bundesweiten Statistik in den Jahren von 1990 bis 2005 bei den rund 11000 Kindern unter 15 Jahren in den Elbmarschen zu erwarten. Geben Sie nach den Sigma-Regeln ein Intervall an, für das Ihre Aussage mit 99 % Wahrscheinlichkeit zutrifft.

3.3. Prüfen Sie, ob die tatsächliche Leukämiehäufigkeit bei den Kindern unter 15 Jahren in den Elbmarschen so hoch ist, dass man mit 99,9 % Sicherheit behaupten kann, dass das Leukämie-Risiko in den Elbmarschen höher ist als im Bundesgebiet.

Tipps zu den Aufgaben

1. Aufgabe

1.1. Definieren Sie die binomialverteilte Zufallsvariable X: Anzahl der stornierten Plätze und die dazugehörenden Parameter n und p.
Berechnen Sie $P(X \geq 10)$

1.2. Stellen Sie mit Hilfe der Regeln $\mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ als Prognose-Intervall die $3{,}29\sigma$-Umgebung auf.
Berechnen Sie anschließend $P(\mu - 3{,}29\sigma \leq X \leq \mu + 3{,}29\sigma)$.

1.3. Gesucht ist die Zahl n, sodass $P(X \geq n - 270) \geq 0{,}99 \quad \Leftrightarrow \quad P(X \leq n - 271) \leq 0{,}01$

Lösen Sie die Ungleichung näherungsweise

- durch Experimentieren mit GeoGebra
- mit der Normalverteilung $\Phi(z) \leq 0{,}01 \quad \Leftrightarrow \quad z \leq -2{,}326$, wobei $z = \dfrac{X - \mu + 0{,}5}{\sigma}$.
- mit der 99 %-Sigma-Regel.
  Anmerkung:
  Mit der Regel berechnet man ein zum Mittelwert symmetrisches Intervall.
  In der Aufgabe ist die Wahrscheinlichkeit $P(0 \leq X \leq x)$ gegeben. Der ermittelte Wert wird also etwas zu groß sein.

2. Aufgabe

2.1. Definieren Sie die binomialverteilte Zufallsvariable X: Anzahl der Ziehungen der 13 und die dazugehörenden Parameter n und p.
Berechnen Sie $P(X \leq 5)$

2.2. Berechnen Sie den Erwartungswert $\mu = n \cdot p$
und mit Hilfe der Regel $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ die $1{,}96\sigma$-Umgebung.

2.3. Sie nehmen an, dass die Ziehungshäufigkeit der unverändert $\frac{6}{49}$ ist, und berechnen die Wahrscheinlichkeit $P(X \leq 409)$. Wenn das Ergebnis höchstens 0,1 % beträgt, liegt die Zahl 410 im Bereich, dessen Wahrscheinlichkeit mindestens 99,9 % beträgt.
Wenn nicht, dann trifft die Behauptung mit 99,9 % Sicherheit zu.

3. Aufgabe

3.1. Definieren Sie die binomialverteilte Zufallsvariable
X: Zahl der Kinder, die jährlich in Baden-Württemberg an Leukämie erkranken.
Berechnen Sie $P(X \leq 50)$

3.2. Berechnen Sie den Erwartungswert $\mu = n \cdot p$ und damit die Zahl für die Jahre 1990 bis 2005.
Geben Sie mit Hilfe der Regel $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ die $2{,}58\sigma$-Umgebung an.

3.3. Sie nehmen an, dass die Leukämiehäufigkeit genau so groß ist wie im ganzen Bundesgebiet und berechnen die Wahrscheinlichkeit $P(X \leq 19)$. Wenn das Ergebnis weniger als 99,9 % beträgt, trifft die Behauptung mit 99,9 % Sicherheit zu.

Tipps zu GeoGebra

- Zu Beginn erstellen Sie einen Schieberegler für n und geben den Werte für p ein.
  Die Werte für μ und σ berechnen Sie mit den Formeln $\mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$
- Das vollständige Histogramm der Binomialverteilung entsteht mit der Funktion
  Binomial[n, p]
- Um die Wahrscheinlichkeit p mit Hilfe eines Eingabefeldes zu verändern geben Sie mit dem Werkzeug a=1 das Eingabefeld ein und ordnen es p zu.
- Für die Darstellung der kumulierten Binomialverteilung definiert man zunächst einen Schieberegler für k. Zur Darstellung des Histogramms verwendet man den Befehl
  Balkendiagramm[L_1, L_2].

  Die beiden Listen L_1 und L_2 werden definiert durch:
  L_1 = Folge[m, m, 0, k]
  L_2 = Folge[Binomial[n, p, m, false], m, 0, k]].
- Für die Darstellung eines $a \cdot \sigma$-Intervalls muss zuerst die Liste L_3 der natürlichen Zahlen k mit $\mu - a \cdot \sigma \leq k \leq \mu + a \cdot \sigma$ definiert werden. In GeoGebra lautet die Funktion
  L_3 = Folge[k, floor(μ – a σ + 1), floor(m + a σ)].

  Die zugehörige Liste L_4 der Wahrscheinlichkeiten wird definiert durch
  L_4 = Folge[Binomial[n, p, m, false],m, Element[L_3, 1], Element[Letztes[L_3, 1]]

  Das Histogramm für das $a \cdot \sigma$-Intervall wird mit der Funktion
  Balkendiagramm[L_3, L_4]
  gezeichnet.
- Die Ausgabe des Textes auf dem Bildschirm erreicht man mit dem Textwerkzeug ABC.
  Zur Anzeige der Werte von Variablen fügt man an der entsprechenden Stelle im Text die Variable aus der Liste der Objekte ein. Man kann auch Rechenoperationen mit Variablen als Objekt eingeben, z.B. $P(X \leq \boxed{n\ p + \sigma})$.

Lösungsvorschlag zu den Aufgaben

1. Aufgabe

1.1. X: Anzahl der stornierten Plätze, $n = 280$, $p = 0{,}1$.
$P(X \geq 10) = 1 - P(X \leq 9) \approx 1 - 0{,}0000142 \approx 0{,}999986$

1.2. $n = 300$, $\mu = 30$, $\sigma \approx 5{,}20$
$3{,}29\sigma$-Umgebung: [13; 47]
Berechnen Sie anschließend $P(13 \leq X \leq 47) \approx 0{,}999$.

1.3. Gesucht ist die Zahl n, sodass $P(X \geq n - 270) \geq 0{,}99 \quad \Leftrightarrow \quad P(X \leq n - 271) \leq 0{,}01$

$n \leq 287$

Mit der $2{,}58\sigma$-Sigma-Regel:
$n - 270 \geq 0{,}1n - 2{,}58\sqrt{0{,}09\,n}$
$\Leftrightarrow \quad 300 - n \leq 0{,}86\sqrt{n}$
$\Leftrightarrow \quad n^2 - 600{,}74\,n + 90000 \leq 0$
$\Leftrightarrow \quad n \leq 285{,}5 \; (\vee \; n \geq 315{,}27)$

2. Aufgabe

2.1. X: Anzahl der Ziehungen der 13, $n = 52$, $p = \frac{6}{49}$.

$P(X \leq 5) \approx 0{,}375$

2.2. $\mu = 6{,}37$, $\sigma = 2{,}36$
$1{,}96\sigma$-Umgebung: [2; 11]

2.3. $P(X \leq 409) \approx 0{,}00828$
d.h. man kann die Behauptung nicht mit 99,9 % Sicherheit behaupten, dass die 13 zu selten gezogen wird.

3. Aufgabe

3.1. X: Zahl der Kinder, die jährlich in Baden-Württemberg an Leukämie erkranken.

$n = 1\,500\,000$, $p = \frac{610}{13\,000\,000} \approx 0{,}0000469$

$P(X \leq 50) \approx 0{,}00663$

3.2. $n = 11000$, $p = 16 \cdot \frac{610}{13\,000\,000} \approx 0{,}000751$

$\mu \approx 8{,}25$, $\sigma \approx 2{,}873$

In den Jahren 1990 bis 2005 war mit etwa 8 Neuerkrankungen zu rechnen.
$2{,}58\sigma$-Umgebung für die 16 Jahre: [1; 15]

3.3. $P(X \leq 19) \approx 0{,}9996$
Man kann nicht mit 99,9 % Sicherheit behaupten, dass die Leukämierate in den Elbmarschen höher ist als im Bundesgebiet.