Schätzen von Parametern – Prognose- und Konfidenzintervalle

Material zum beispielhaften SiLP GOSt Mathematik NRW 2023

Juni 2023

# Kurzbeschreibung

Das Schätzen von Parametern erlaubt es, die Modellbildung auf Grundlage einer Stichprobe vorzunehmen und zu beurteilen. Dieses Unterrichtsvorhaben schließt dabei an die Behandlung von Sigma-Regeln an, die jeweils auf den Erwartungswert und die Standardabweichung einer Binomialverteilung zurückzuführen sind.

Im vorliegenden Unterrichtsvorhaben geht es zunächst darum, welche Vorhersagen über Ergebnisse von binomialverteilten Zufallsexperimenten mit bekannter Erfolgswahrscheinlichkeit möglich sind. Mithilfe der Sigma-Regeln werden dabei mit Prognoseintervallen Häufigkeiten als Ergebnisse von Zufallsexperimenten geschätzt.

Danach werden ausgehend von den Ergebnissen von Stichproben mögliche Rückschlüsse auf die Gesamtheit gezogen. Ausgehend von relativen Häufigkeiten werden dabei mit Konfidenzintervallen die zugrundeliegenden Wahrscheinlichkeiten geschätzt.

Ellipsendiagramme werden zur graphischen Darstellung von Prognose- und Konfidenzintervallen eingesetzt und vernetzen diese miteinander.

Abschließend soll auch die Mindestgröße einer Stichprobe für ein Konfidenzintervall vorgegebener Länge abgeschätzt werden.

# [Das Unterrichtsvorhaben](file:///C%3A%5CUsers%5Cdenishusemann%5CDownloads%5CTestergebn#_Das_Unterrichtsvorhaben_") im Überblick

Zeitbedarf: ca. 12-14 Unterrichtsstunden

Die Unterrichtsreihe schließt an die Behandlung von Erwartungswert und Standardabweichung einer Binomialverteilung sowie der Sigma-Regeln an, die nun als bekannt vorausgesetzt werden.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Thema** | **Unterrichtsverlauf** | Stützpunktaufgaben / Material |
| 1 | Prognoseintervalle für absolute Häufigkeiten | * Erstellen von Prognosen über die zu erwartende Anzahl der Erfolge bei Zufallsexperimenten mit Münzen mithilfe der Sigma-Regeln
* **Einführung des Begriffs „Prognoseintervall für absolute Häufigkeiten“**
* Anwendung des Prognoseintervalls für das Merkmal „Augenfarbe blau“ bei verschiedenen Stichprobengrößen
 | A1 |
| 2 | Prognoseintervalle für relative Häufigkeiten | * Prognosen für absolute und relativen Häufigkeiten für eine Untersuchung zum „Zunge Rollen“
* **Einführung des Begriffs „Prognoseintervall für relative Häufigkeiten“**
* Herleitung einer Formel für die Intervallgrenzen von Prognoseintervallen für relative Häufigkeiten ausgehend von Prognoseintervallen für absolute Häufigkeiten mittels Division durch den Stichprobenumfang *n*
 | A2 |
| 3 | Ellipsendiagramm – graphische Darstellung der Prognoseintervalle | * Berechnung von Prognoseintervallen für relative Häufigkeiten für verschiedene Werte von *p*
* Darstellung in einem gemeinsamen Koordinatensystem – die „Ellipse“
* weitergehende Übung zum Ablesen des Prognoseintervalls im Ellipsendiagramm (zu gegebenem *p*) mit den GeoGebra-Dateien „Prognoseintervall-Ellipse“ und „Ellipse\_n\_variabel“
 | A3.1, A3.2undGeoGebra-Dateien |
| 4 | Ellipsendiagramm „quer“: Das Konfidenzintervall | Schätzung der unbekannten Wahrscheinlichkeit in der Grundgesamtheit bei bekannter Häufigkeit:* graphisches Bestimmen eines Intervalls für die Wahrscheinlichkeit *p* von Paket-Rücksendungen, wenn aufgrund einer Stichprobe eine relative Häufigkeit *h* bekannt ist
* Ablesen weiterer waagerechter Intervalle im Ellipsendiagramm
* **Einführung des Begriffs „Konfidenzintervall“**
 | A4 und GeoGebra-Datei |
| 5 | Konfidenzintervall rechnerisch | Berechnung der Grenzen von Konfidenzintervallen mithilfe des MMS:* Berechnung der Grenzen eines Konfidenzintervalls im Anwendungskontext einer Corona-Studie
* Vergleich der Größe der Konfidenzintervalle für Stichproben mit unterschiedlichem Umfang *n*
 | A5 |
| 6 | Konfidenzintervalle in medizinischen Tests | * Anwendung der Konfidenzintervalle zur Angabe der Genauigkeit von medizinischen Tests
* Untersuchung des Beipackzettels eines gängigen Corona-Schnelltests
 | A6 |
| 7 | Stichprobenumfang abschätzen | * Ermitteln der Mindestgröße einer Stichprobe, um den Anteil der vollständig geimpften Personen für eine vorgegebene Genauigkeit zu bestimmen
* Vergleichen der notwendigen Mindestgrößen einer Stichprobe bei unterschiedlicher Genauigkeit
 | A7 |

# Hinweise zu den Materialien:

Im Folgenden werden mit Stützpunktaufgaben mögliche unterrichtliche Zugänge dargestellt, die grundsätzlich und insbesondere in Bezug auf Differenzierungen und die jeweilige Sozialform und (kooperative) Arbeitsform an die Bedürfnisse der jeweiligen Lerngruppe angepasst werden sollten. Mit Kommentaren und Hinweisen werden im Folgenden mögliche Schwerpunkte und Perspektiven für die Unterrichtsplanung aufgezeigt. Weiterführende detaillierte Unterrichtsplanungen werden hier nicht dargestellt. Auf der letzten Seite folgen weitere Hinweise.

# Lehrplanbezug

Dieses Unterrichtsvorhaben konkretisiert eine mögliche Umsetzung des beispielhaften schulinternen Lehrplans Mathematik, der auf dem Kernlehrplan der gymnasialen Oberstufe Mathematik (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2023) basiert.

Diese Unterrichtsreihe kann als letzter Teil des Unterrichtsvorhabens LK-S3 und als Unterrichtsvorhaben LK-S4 des beispielhaften schulinternen Lehrplans dienen. Die aufgeführten Kompetenzerwartungen des Kernlehrplans sind Schwerpunkte der Kompetenzentwicklung in diesen Unterrichtsvorhaben.

|  |
| --- |
| **Kompetenzerwartungen**Die Schülerinnen und Schüler …**Inhaltsbezogene Kompetenzen**LK-S(16) ermitteln mithilfe der σ-Regeln Prognoseintervalle für die absoluten und relativen Häufigkeiten in einer Stichprobe und interpretieren diese im Sachkontext,LK-S(17) ermitteln auf Grundlage einer relativen Häufigkeit ein Konfidenzintervall für den Parameter p einer binomialverteilten Zufallsgröße und interpretieren das Ergebnis im Sachkontext (Schluss von der Stichprobe auf die Grundgesamtheit),LK-S(18) schätzen den für ein Konfidenzintervall vorgegebener Länge erforderlichen Stichprobenumfang ab**Prozessbezogene Kompetenzen**Ope-(2) übersetzen formale Sprache in natürliche Sprache und umgekehrt,Ope-(12) verwenden im Unterricht ein modulares Mathematiksystem (MMS) zum …- Lösen von Gleichungen und Gleichungssystemen auch abhängig von Parametern,- Ermitteln der Kennzahlen statistischer Daten und von Wahrscheinlichkeitsverteilungen,- Variieren der Parameter von Wahrscheinlichkeitsverteilungen,- Berechnen von Wahrscheinlichkeiten bei binomialverteilten und im Leistungskurs auch normalverteilten Zufallsgrößen,- Berechnen der Grenzen von Konfidenzintervallen im Leistungskurs,Ope-(13) entscheiden situationsangemessen über den Einsatz mathematischer Hilfsmittel und digitaler Mathematikwerkzeuge und wählen diese begründet aus,Ope-(14) reflektieren die Möglichkeiten und Grenzen digitaler Mathematikwerkzeuge.Mod-(5) erarbeiten mithilfe mathematischer Kenntnisse und Fertigkeiten Lösungen innerhalb des mathematischen Modells,Mod-(6) beziehen erarbeitete Lösungen wieder auf die reale Situation und interpretieren diese als Antwort auf die Fragestellung, Mod-(7) reflektieren die Abhängigkeit der Lösungen von den getroffenen Annahmen,Mod-(8) benennen Grenzen aufgestellter mathematischer Modelle und vergleichen Modelle bzgl. der Angemessenheit, Mod-(9) verbessern aufgestellte Modelle mit Blick auf die Fragestellung, Pro-(6) wählen geeignete Begriffe, Zusammenhänge, Verfahren sowie Medien und Werkzeuge zur Problemlösung aus, Pro-(10) überprüfen die Plausibilität von Ergebnissen und interpretieren diese vor dem Hintergrund der Fragestellung, Pro-(12) vergleichen und beurteilen verschiedene Lösungswege und optimieren diese mit Blick auf Schlüssigkeit und Effizienz, Pro-(14) variieren und verallgemeinern Fragestellungen vor dem Hintergrund einer Lösung,Arg-(5) begründen Lösungswege und nutzen dabei mathematische Regeln und Sätze sowie sachlogische Argumente,Arg-(12) beurteilen Argumentationsketten hinsichtlich ihres Geltungsbereichs und ihrer Übertragbarkeit, Arg-(13) überprüfen, inwiefern Ergebnisse, Begriffe und Regeln verallgemeinert werden können,Kom-(7) wählen begründet geeignete digitale und analoge Medien und mathematische Darstellungsformen (graphisch-visuell, algebraisch-formal, numerisch-tabellarisch, verbal-sprachlich) aus, Kom-(8) wechseln flexibel zwischen mathematischen Darstellungsformen. |

# Stützpunktaufgaben und didaktische Kommentare

## A1. Prognoseintervalle für absolute Häufigkeiten

**Aufgaben:**

a) Wenn wir 200-mal eine Münze werfen, erwarten wir, dass 100-mal die Seite „Wappen“ geworfen wird. Geben Sie mithilfe der Sigma-Regeln ein zum Erwartungswert symmetrisches Intervall an, in dem mit einer Wahrscheinlichkeit von ca. 95% die Anzahl der „Wappen“ liegt.

b) Führen Sie das beschriebene Zufallsexperiment möglichst oft (z.B. in Ihrem Kurs) durch und erheben Sie die Anzahl der „Wappen“. Vergleichen Sie die Ergebnisse des Zufallsexperiments mit Ihrer Prognose aus a). Ermitteln Sie den Anteil an Ergebnissen, bei denen die Anzahl der „Wappen“ im berechneten Prognoseintervall liegt.

c) In einem wissenschaftlichen Artikel wird angegeben, dass in Deutschland ca. 30% der Menschen blaue Augen haben. Bei 100 zufällig ausgewählten Personen wird die Augenfarbe bestimmt. Geben Sie auf Grundlage der genannten 30% ein Prognoseintervall an, in dem mit einer Wahrscheinlichkeit von 90% die Anzahl der Personen liegt, die blaue Augen haben.
Berechnen Sie auch das 90%-Prognoseintervall für die Größe Ihres Kurses und beurteilen Sie, inwieweit das Prognoseintervall eine Aussagekraft für Ihren Kurs hat.

**Lösungen/Hinweise:**

1. 95%-Prognoseintervall: für *n* = 200 und *p* = 0,5:

$\left[np-1,96∙\sqrt{np\left(1 - p\right)} ; np + 1,96∙\sqrt{np\left(1 - p\right)} \right]≈\left[86;114\right]$

Kontrolle: $P\left(86\leq X\leq 114\right)≈95,9\%$

*Bei der Berechnung der Prognoseintervalle muss auf das Runden der Werte geachtet werden, damit das Intervall das kleinstmögliche Intervall darstellt, das symmetrisch um den Erwartungswert liegt. Es bietet sich eine Kontrolle des Ergebnisses an.*

1. *Hier können Abweichungen von dem berechneten Prognoseintervall und die Bedeutung der Sicherheitswahrscheinlichkeit* 95% *thematisiert werden.*
2. 90%-Prognoseintervall für *n* = 100 und *p* = 0,3:

$$\left[np-1,64∙\sqrt{np\left(1 - p\right)} ; np + 1,64∙\sqrt{np\left(1 - p\right)} \right]≈\left[22;38\right]$$

Kontrolle: $P\left(22\leq X\leq 38\right)≈93,7\%$
Es kann thematisiert werden, inwieweit der Kurs eine zufällige Auswahl darstellt.

**Kommentar:**

*Für die Unterrichtseinheit wurde ein experimenteller Zugang gewählt, damit die Schülerinnen und Schüler handlungsorientiert und mit authentischen Daten an die neuen Inhalte herangeführt werden. Bei der Augenfarbe handelt es sich um ein Merkmal, zu dem im Kurs eine Häufigkeit bestimmt werden kann und zu der die Schülerinnen und Schüler einen unmittelbaren Zugang haben. Mathematik wird dadurch in ihrem Alltag erfahrbar. Es ist wichtig an dieser Stelle den Begriff „Prognoseintervall für absolute Häufigkeiten“ und im Zusammenhang damit die Sicherheitswahrscheinlichkeit einzuführen. In der nächsten Unterrichtseinheit kann daraus der Begriff „Prognoseintervall für relative Häufigkeiten“ entwickelt werden. Im weiteren Verlauf der Reihe wird der Begriff des Konfidenzintervalls hinzukommen, der von den Prognoseintervallen abzugrenzen ist.*

## A2. Prognoseintervalle für relative Häufigkeiten

**Aufgaben:**

1. Der Anteil der Menschen, die ihre Zunge rollen können, beträgt 70%. Untersuchen Sie mithilfe der 1,96σ-Regel, in welchem Intervall voraussichtlich die relative Häufigkeit für das Merkmal „Zunge rollen“ in einer zufälligen Stichprobe von 1000 Menschen liegen wird.
2. Entwickeln Sie eine Formel, mit der die Grenzen eines 95%-Prognoseintervalls (90%- Prognoseintervalls) für relative Häufigkeiten für eine Stichprobengröße n und eine Trefferwahrscheinlichkeit p mithilfe der Sigma-Regeln berechnet werden können.
3. (1) Ein Würfel wird geworfen und wir betrachten den Anteil der geworfenen 6en.
Berechnen Sie das 95%-Prognoseintervall für relative Häufigkeiten für 100 Würfe und für 1000 Würfe.

(2) Bei einer Stichprobe von 100 bzw. 1000 Würfen wurde jeweils eine Häufigkeit von *h* = 0,12 beobachtet. Erläutern Sie die Bedeutung dieses Stichprobenergebnisses für das jeweilige Prognoseintervall.

1. Simulieren Sie mit einer Tabellenkalkulation mehrfach das Experiment mit 100 Würfen und überprüfen Sie, wie oft die relative Häufigkeit im Prognoseintervall aus b)(1) liegt.

**Weitergedacht***:* Ermitteln Sie die maximale Stichprobengröße, so dass *h* = 0,12 im 95%- Prognoseintervall liegt.

**Lösungen:**

1. Für die relativen Häufigkeiten ergibt sich wegen *n* = 1000 und *p* = 0,7 das 95 %-Prognoseintervall $\left[p-1,96∙\frac{\sqrt{p∙\left(1-p\right)}}{\sqrt{n}} ; p+1,96∙\frac{\sqrt{p∙(1-p)}}{\sqrt{n}}\right]≈$ [0,671; 0,729].
2. $\left[p-1,96∙\frac{\sqrt{p∙\left(1-p\right)}}{\sqrt{n}} ; p+1,96∙\frac{\sqrt{p∙(1-p)}}{\sqrt{n}}\right]$ (bzw. $\left[p-1,64∙\frac{\sqrt{p∙\left(1-p\right)}}{\sqrt{n}} ; p+1,64∙\frac{\sqrt{p∙(1-p)}}{\sqrt{n}}\right])$
3. (1) Für *n* = 100 und *p* = 1/6 ergibt sich das 95%-Prognoseintervall $\left[0,09 ;0,24\right]$.
 Für *n =* 1000 und *p* = 1/6 ergibt sich 95%-Prognoseintervall $\left[0,14 ;0,19\right]$.

(2) Die Häufigkeit *h* = 0,12 liegt im 95%-Prognoseintervall für 100 Würfe.
Bei einer Stichprobe von 1000 Würfen weicht eine relative Häufigkeit von *h* = 0,12 signifikant vom 95%-Prognoseintervall ab. Auf Grundlage einer solchen Abweichung könnte die Modellierung mit p = 1/6 hinterfragt werden.

1. Individuelle Ergebnisse

**Lösungen Weitergedacht**: Durch systematisches Variieren der Stichprobengröße kann der gesuchte Wert ermittelt werden:

|  |  |  |
| --- | --- | --- |
| Stichprobengröße n | Untere Grenze des Intervalls | Obere Grenze des Intervalls |
| 100 | 0,0936 | 0,2397 |
| 1000 | 0,1436 | 0,1897 |
| 500 | 0,134 | 0,1993 |
| 250 | 0,1205 | 0,2129 |
| 245 | 0,12 | 0,2133 |

**Kommentar:**

*In einem ersten Zugang bestimmen die Schülerinnen und Schüler selbstständig ein Prognoseintervall für relative Häufigkeiten, z.B. indem sie zunächst ein Prognoseintervall für die absoluten Häufigkeiten berechnen und dies durch n teilen.*

*Im Anschluss daran sollte der Begriff des Prognoseintervalls für relative Häufigkeiten im Zusammenhang mit der Sicherheitswahrscheinlichkeit eingeführt und die Formel für die Grenzen des Prognoseintervalls formal hergeleitet werden. Hier bietet sich auch ein Vergleich mit der Formel im* [Dokument mit mathematischen Formeln (vgl. standardsicherung.nrw.de)](https://www.standardsicherung.schulministerium.nrw.de/cms/zentralabitur-gost/faecher/getfile.php?file=5705) *an.*

*In Teilaufgabe c) wird das zuvor eingeführte Prognoseintervall für relative Häufigkeiten für weitere Rechnungen genutzt. Seine Bedeutung wird thematisiert.*

*In Teilaufgabe d) verwenden die Schülerinnen und Schüler Simulationen zur Untersuchung stochastischer Situationen und nutzen dabei auch digitale Mathematikwerkzeuge, hier insbesondere die Tabellenkalkulation.*

*Die Aufgabe „Weitergedacht“ bietet sich als Differenzierungsaufgabe für leistungsstarke Schülerinnen und Schüler an. Es handelt sich um eine explorative Aufgabe, bei der die maximale Stichprobengröße durch die Variationen von Parametern entdeckt und begründet werden soll.*

## A3.1 Visualisierung von Prognoseintervallen durch ein Ellipsendiagramm[[1]](#footnote-1)

Im Diagramm ist das 95%-Prognoseintervall für *p* = 0,3 und *n* = 100 graphisch dargestellt.



1. Im Folgenden wird von der Stichprobengröße *n* = 100 ausgegangen. Berechnen Sie (ggf. arbeitsteilig) für verschiedene Werte von *p* die Grenzen der 95%-Prognoseintervalle für relative Häufigkeiten und notieren Sie Ihre Ergebnisse in der Tabelle.
Ergänzen Sie die fehlenden Darstellungen von 95%-Prognoseintervallen im Koordinatensystem, indem Sie zu jedem Wert von *p* die vertikalen Strecken zwischen den Intervallgrenzen einzeichnen.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Wahrscheinlichkeit p | 0 | 0,05 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 | 0,95 | 1 |
| untere Grenze für *h*:$$p-1,96∙\frac{\sqrt{p∙(1-p)}}{\sqrt{100}}$$ |  |  |  |  | 0,21 |  |  |  |  |  |  |  |  |
| obere Grenze für *h*:$$p+1,96∙\frac{\sqrt{p∙(1-p)}}{\sqrt{100}}$$ |  |  |  |  | 0,39 |  |  |  |  |  |  |  |  |

1. Verbinden Sie die unteren Grenzen der Prognoseintervalle mit einer grünen Linie und die oberen Grenzen mit einer blauen Linie und beschreiben Sie die sich ergebende Darstellung.

**Lösungen:**

 

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Wahrscheinlichkeit p | 0 | 0,05 | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 | 0,95 | 1 |
| untere Grenze für h:$$p-1,96∙\frac{\sqrt{p∙(1-p)}}{\sqrt{100}}$$ | 0 | 0,01 | 0,04 | 0,12 | 0,21 | 0,30 | 0,40 | 0,50 | 0,61 | 0,72 | 0,84 | 0,91 | 1 |
| obere Grenze für h:$$p+1,96∙\frac{\sqrt{p∙(1-p)}}{\sqrt{100}}$$ | 0 | 0,09 | 0,16 | 0,28 | 0,39 | 0,50 | 0,60 | 0,70 | 0,79 | 0,88 | 0,96 | 0,99 | 1 |

**Kommentar:**

*Der Wechsel der Darstellungsformen vom Prognoseintervall zum Ellipsendiagramm steht bei dieser Aufgabe im Zentrum. Die Tabelle sollte dabei in Gruppenarbeit arbeitsteilig gefüllt werden, sodass Schülerinnen und Schüler nicht jeden Wert berechnen müssen, ihnen insgesamt aber genügend Informationen für die Zeichnung zur Verfügung stehen. Somit ist sowohl die individuelle Auseinandersetzung mit fachlichen Strukturen als auch die wechselseitige Verständigung und Kooperation gefordert. Auch der Einsatz einer Tabellenkalkulation bzw. eines MMS ist eine mögliche Alternative.*

*Die hier verwendete Abbildung ist mithilfe der GeoGebra-Datei „Prognoseintervall-Ellipse“ erstellt worden. Im Material A3.2 wird mit der Datei weitergearbeitet.*

## A3.2 Graphisches Ablesen von Prognoseintervallen

**Aufgaben:**

1. In der GeoGebra-Datei „Prognoseintervall-Ellipse“ können Sie mit einem Schieberegler zu verschiedenen Wahrscheinlichkeiten *p* die 95%-Prognoseintervalle für relative Häufigkeiten für *n* = 100 zeichnen.
Bestimmen Sie mithilfe der Grafik in der Datei das 95%-Prognoseintervall
für *p* = 0,75 und *n* = 100 und geben Sie die Grenzen möglichst genau an.
2. Zeichnen Sie mit Hilfe des Schiebereglers das vollständige Ellipsendiagramm.
3. Beschreiben Sie die Bedeutung der blauen und grünen Linien.
4. Erklären Sie, wie das 95%-Prognoseintervall für *n* = 100 zu einer gegebenen Wahrscheinlichkeit *p* an der Ellipse abgelesen werden kann.
5. Beurteilen Sie mit der Grafik, ob eine beobachtete Häufigkeit *h* = 0,36 im 95%-Prognoseintervall für *n* = 100 und *p* = 0,25 liegt.

**Lösungen:**

1. Mit der GeoGebra-Datei kann das Intervall $\left[0,67 ;0,83\right]$ abgelesen werden.
2. (1) Die blaue Linie beschreibt die von der Wahrscheinlichkeit *p* abhängige obere Grenze der 95%-Prognoseintervalle für Stichproben mit *n* = 100, die grüne Linie die untere Grenze. Ihre Funktionsgleichungen lauten:
$f\_{-}\left(p\right)=p-1,96\sqrt{\frac{p∙(1-p)}{n}}$ und $f\_{+}\left(p\right)=p+1,96\sqrt{\frac{p∙(1-p)}{n}}$.

(2) individuelle Lösung

(3) Die Häufigkeit *h* = 0,35 liegt oberhalb des „blauen“ Punktes (0,25 | 0,33), der die obere Grenze des Prognoseintervalls zu *p* = 0,25 darstellt.
Der Wert *h* = 0,36 stellt eine signifikante Abweichung vom Prognoseintervall dar.

**Kommentar:**

*An dieser Stelle kann das Ablesen von Prognoseintervallen in Ellipsendiagrammen weiter geübt werden. Es bietet sich an Ellipsendiagramme zu verschiedenen Werten von n zu betrachten und ihre Darstellungen zu vergleichen, z.B. mit der GeoGebra-Datei „Ellipse\_n\_variabel“.*

*Das Arbeiten mit GeoGebra macht dabei auch komplexere Sachverhalte zugänglich und ermöglicht, mathematische Zusammenhänge zu visualisieren und zu dynamisieren.*

## A4. Ellipsendiagramme „quer“

**Aufgaben:**

Ein Onlinehändler möchte wissen, in welchem Bereich die Wahrscheinlichkeit liegt, dass ein Paket wieder zurückgeschickt wird. Bei einer Stichprobe von 100 Bestellungen wurden 60 Pakete wieder zurückgesandt.

1. Nehmen Sie Stellung zu der Aussage:
*Die Wahrscheinlichkeit, dass ein Paket zurückgeschickt wird beträgt genau* 60%.

Die Wahrscheinlichkeit, dass ein Paket zurückgeschickt wird, soll nun abgeschätzt werden indem ein Bereich angegeben wird, in dem diese Wahrscheinlichkeit voraussichtlich liegen wird.

1. Erklären Sie, wie mithilfe des Ellipsendiagramms (zur Sicherheitswahrscheinlichkeit 95%) ein Bereich bestimmt werden kann, in dem die Wahrscheinlichkeit für eine Paket-Rücksendung voraussichtlich liegen wird.

Mit welcher Wahrscheinlichkeit sollte der Onlinehändler kalkulieren? Erläutern Sie Ihre Vorgehensweise und formulieren Sie eine Antwort für den Onlinehändler.

****

**Lösungen:**

1. Die genaue Wahrscheinlichkeit für eine Rücksendung ist nicht bekannt. 60% ist die Häufigkeit der Rücksendungen in dieser einen Stichprobe. Daraus kann keine zuverlässige Aussage für die genaue Wahrscheinlichkeit abgeleitet werden. Die Häufigkeit in der Stichprobe kann aber als erste ungefähre Schätzung („Punktschätzung“) dieser unbekannten Wahrscheinlichkeit genutzt werden, die voraussichtlich „in der Nähe“ von 60% liegen wird.
2. Anders als bei den vorangegangenen Aufgaben ist hier die Häufigkeit *h* gegeben und die Wahrscheinlichkeit *p* gesucht.
Daher muss im Ellipsendiagramm ausgehend von *h* = 0,6 ein Intervall für *p* bestimmt werden. Durch Einzeichnen einer waagerechten Strecke ausgehend von
*h* = 0,6 erhält man ungefähr $0,50 \leq p\leq 0,70. $
Die Schülerinnen und Schüler formulieren individuelle Lösungen zur Vorgehensweise und Antworten für den Onlinehändler.

**Kommentar:**

1. *Der Fokus der ersten Teilaufgabe liegt darauf, dass die bei einer Stichprobe erhobene relative Häufigkeit als Punktschätzung der unbekannten zugrundeliegenden Wahrscheinlichkeit interpretiert werden kann.
Die Kontrollrechnung* $P\_{0,6}^{100}\left(X=60\right)≈0,0812$ *zeigt, dass bei Annahme einer Wahrscheinlichkeit von p* = 0,6 *eine solche Punktschätzung in ca.* 91,88% *der Fälle nicht zur zugrundeliegenden Häufigkeit führen würde. Dies zeigt, dass die Punktschätzung nur in wenigen Fällen die genaue Wahrscheinlichkeit p liefert.
Die Frage, in welcher Umgebung der Punktschätzung die unbekannte Wahrscheinlichkeit voraussichtlich liegt, führt zur Intervallschätzung und damit zum Konfidenzintervall.*
2. *Bei dieser Teilaufgabe kann auch der Einfluss der Sicherheitswahrscheinlichkeit (hier 95%) auf die Form der Ellipse und damit auf die Intervallschätzung thematisiert werden. In Bezug auf die Bedeutung der Intervallschätzung kann an der Ellipse erkannt werden, dass diese Intervallschätzung alle Wahrscheinlichkeiten enthält, für die die beobachtete Häufigkeit (hier 60%) im zugehörigen Prognoseintervall liegt. Somit liefert diese Intervallschätzung alle Wahrscheinlichkeiten, denen man auf Grundlage dieser Sicherheitswahrscheinlichkeit und dieser einen Stichprobe (inkl. Stichprobengröße) vertrauen kann.
Bei der Bearbeitung dieser Aufgabe werden Informationen aus einem Ellipsendiagramm entnommen und zur fundierten Kommunikation genutzt. Die Schülerinnen und Schüler erfassen und erläutern die mathematische Darstellung (rezipieren) und formulieren eigene Lösungswege (produzieren).*

***Im Anschluss sollte der Begriff des Konfidenzintervalls eingeführt*** *und weitere Konfidenzintervalle grafisch mithilfe von Ellipsendiagrammen ermittelt* ***werden.****Die Abgrenzung und der Zusammenhang zwischen dem Prognoseintervall (p bekannt) und dem Konfidenzintervall (h bekannt) wird durch die unterschiedlichen Arten des Ablesens (vertikal bzw. horizontal) im Ellipsendiagramm verdeutlicht.*

## A5. Konfidenzintervalle rechnerisch

**Aufgaben:**

1. In einer Studie im Ruhrgebiet im Februar 2021 wurde untersucht, wie viele Kinder bereits mit dem Corona-Virus infiziert waren. Dazu wurde bei 3000 Kindern eine Blutuntersuchung durchgeführt. Bei 8% der untersuchten Kinder wurden Antikörper nachgewiesen, die Kennzeichen einer bestehenden oder durchgemachten Corona-Infektion sind.
	1. Für diesen Sachzusammenhang könnte das zugehörige 95%-Konfidenzintervall mit einem „Ellipsendiagramm“ grafisch ermittelt werden. Geben Sie jeweils einen Funktionsterm für die obere und untere Linie der Ellipse an.
	2. Bestimmen Sie mithilfe Ihres MMS rechnerisch die beiden Grenzen des 95%-Konfidenzintervalls für die zugrundeliegende Wahrscheinlichkeit dieser Stichprobe und beschreiben Sie Ihre Vorgehensweise.
	3. Erläutern Sie die Bedeutung des Konfidenzintervalls im Sachzusammenhang.
2. Angenommen in der Studie wären nur 300 Kinder untersucht worden, wobei die Häufigkeit von 8% für Kinder mit Antikörpern gleichgeblieben wäre. Berechnen Sie das zugehörige 95%-Konfidenzintervall und vergleichen Sie Ihr Ergebnis mit den Werten aus a).
3. Wir nehmen an, es wären nur 30 Kinder untersucht worden, von denen 8% Antikörper im Blut gehabt hätten. Stellen Sie zunächst eine Vermutung für das 95%-Konfidenzintervall auf und berechnen Sie anschließend die exakten Werte.
Erläutern Sie anschließend, welche Bedeutung die Größe der Stichprobe für das zugehörige Konfidenzintervall hat.

**Lösungen:**

1. Es wird eine Stichprobe mit *n* = 3000 und *h* = 0,08 beschrieben.
	1. $f\_{-}\left(p\right)=p-1,96\sqrt{\frac{p∙(1-p)}{3000}}$ und $f\_{+}\left(p\right)=p+1,96\sqrt{\frac{p∙(1-p)}{3000}}$.
	2. Die Grenzen des Konfidenzintervall erhält man, indem man die folgenden Gleichungen mithilfe des MMS nach *p* auflöst:
	$0,08=p-1,96\sqrt{\frac{p∙(1-p)}{3000}}$ und $0,08=p+1,96\sqrt{\frac{p∙(1-p)}{3000}}$
	Mithilfe des MMS ergibt sich [0,07; 0,09] als 95%-Konfidenzintervall.
	3. Im 95 %–Konfidenzintervall [0,07; 0,09] liegen alle Wahrscheinlichkeiten *p*, dass ein Kind im Ruhrgebiet zu der Zeit Antikörper im Blut hat, denen man auf Grundlage der Stichprobe vertrauen kann, weil die beobachtete relative Häufigkeit in den 95 %–Prognoseintervallen dieser Wahrscheinlichkeiten *p* liegt.
2. Für *n* = 300 und *h* = 0,08 ergibt sich das 95%-Konfidenzintervall [0,05 ; 0,12].
3. Für *n* = 30 und *h* = 0,08 ergibt sich das 95%-Konfidenzintervall [0,02 ; 0,23].

Bei kleiner werdender Stichprobengröße nimmt die Länge der berechneten Konfidenzintervalle zu.

Dieser Zusammenhang kann auch mithilfe von Ellipsendiagrammen veranschaulicht werden:



**Kommentar:**

*Ausgehend von den bekannten Ellipsen wird in einem ersten Schritt die rechnerische Bestimmung selbst entdeckt und in weiteren Aufgaben vertieft. Zur Differenzierung kann bei sehr leistungsstarken Schülerinnen und Schülern ggf. auf die Teilaufgabe a)1. verzichtet werden, während bei erhöhtem Unterstützungsbedarf abweichend vom hier vorgestellten Vorgehen zunächst eine grafische Lösung angestrebt werden sollte, die dann auf ein rechnerisches Vorgehen übertragen wird.*

*In den dargestellten Lösungen wurde das Konfidenzintervall mit einem MMS berechnet. Als weiterer optionaler Lösungsweg für die Berechnung von Konfidenzintervallen ist das Umstellen der Gleichung(en) nach p möglich. Dies stellt eine mögliche zusätzliche Vertiefung für algebraisch sehr versierte Lerngruppen dar.*

*Neben dem rechnerischen Ansatz zur Bestimmung von Konfidenzintervallen wird bei diesen Aufgaben der Einfluss des Stichprobenumfangs auf die Größe der Konfidenzintervalle betrachtet.*

*Dabei sollen die Schülerinnen und Schüler begründete Vermutungen über die Existenz und Art von Zusammenhängen aufstellen. Diese zu prüfen und ggf. zu verallgemeinern ist wesentlicher Bestandteil des mathematischen Argumentierens.*

*Quelle der Daten von Teilaufgabe a) ist die CorKid Studie der Ruhr-Universität Bochum. Diese und weitere Studienergebnisse zum Coronavirus sind beim Robert Koch Institut zu finden:*

[*https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\_Coronavirus/AK-Studien/Ergebnisse.html*](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/AK-Studien/Ergebnisse.html%20) *(zuletzt aufgerufen am 02.06.2023)*

## A6. Konfidenzintervalle in medizinischen Tests

**Aufgaben:**

Auf dem Beipackzettel eines Corona-Schnelltests\* finden sich Informationen zur Güte des Tests: In einer klinischen Studie wurden 508 teilnehmende Personen mit diesem Corona-Schnelltest getestet. Zusätzlich wurde bei jeder Person ein PCR-Test gemacht. Das Ergebnis des PCR-Tests wird als zuverlässiger Nachweis für die tatsächlich vorliegende Erkrankung angenommen.

Die Ergebnisse der klinischen Studie sind in folgender Tabelle des Beipackzettels zusammengefasst.

|  |  |
| --- | --- |
| Corona-Schnelltest | **PCR-Test** |
| **positiv** | **negativ** |
| positiv | 231 | 1 |
| negativ | 12 | 264 |
| Diagnostische Sensitivität | 95,06%(95%CI: 91,57% ~ 97,15%) | - |
| Diagnostische Spezifität | - | 99,62%(95%CI: 97,89% ~ 99,93%) |

1. Geben Sie an, wie viele teilnehmende Personen der Studie tatsächlich mit Corona infiziert waren, also einen positiven PCR-Test hatten.
2. Kontrollieren Sie durch eine Rechnung die angegebene Sensitivität des Tests, d.h. den Anteil der vom Corona-Schnelltest positiv getesteten Personen an der Gesamtzahl der (mit dem PCR-Test nachgewiesenen) Corona-positiven.
3. Ermitteln Sie das 95%-Konfidenzintervall für eine Sensitivität von 95,06% und vergleichen Sie dieses mit den entsprechenden Angaben des Herstellers, die mit „95%CI“ in der Tabelle des Beipackzettels zu finden sind.
4. Erläutern Sie, warum auf dem Beipack-Zettel zusätzlich zur Sensitivität auch das Konfidenzintervall angegeben ist.

*\* Bei den Angaben in dieser Aufgabe handelt es sich um authentischen Werte, die in dieser Form auf dem Beipackzettel eines handelsüblichen Corona-Schnelltests aus dem Jahr 2022 angegeben sind.*

**Lösungen:**

1. Es waren 243 Personen in der Studie positiv.
2. $\frac{231}{243}≈0,9506$. Die Sensitivität des Corona-Schnelltests liegt bei ca. 95,06%.
3. $ p-1,96∙\sqrt{\frac{p∙(1-p)}{243}}\leq \frac{231}{243}\leq p+1,96∙\sqrt{\frac{p∙(1-p)}{243}}$

Es folgt $0,9157\leq \frac{231}{243}\leq 0,9715$ und somit [0,9157; 0,9715] als 95%-Konfidenzintervall.

1. Die Sensitivität entspricht der in dieser einen Stichprobe ermittelten relativen Häufigkeit. Diese relative Häufigkeit kann aufgrund der zufälligen Schwankung von der zugrundeliegenden Wahrscheinlichkeit abweichen.

Daher wird zusätzlich ein 95%-Konfidenzintervall angeben, in dem alle Wahrscheinlichkeiten liegen, für die die in dieser Stichprobe ermittelte Häufigkeit im zugehörigen 95% Prognoseintervall liegt.

**Kommentar:**

*Bei den Angaben in dieser Aufgabe handelt es sich um die authentischen Studienergebnisse eines handelsüblichen Corona-Schnelltests, der in den Jahren 2021 – 2022 die Lebenswirklichkeit der Schülerinnen und Schüler stark geprägt hat. Der Umgang mit den angegebenen Informationen aus dem Beipackzettel trägt zur vertieften allgemeinen Bildung der Schülerinnen und Schüler bei. Die Schülerinnen und Schüler vergleichen und beurteilen mathematikhaltige Informationen und Darstellungen in Alltagsmedien unter mathematischen Gesichtspunkten und nehmen zu mathematikhaltigen Aussagen und Darstellungen begründet und konstruktiv Stellung. Durch das Überprüfen und Bewerten der Informationen wird die persönliche Urteilsfähigkeit weiter ausgebildet.*

*Hinweis für die Lehrkraft: Die für die vorliegenden Berechnungen geforderte Laplace-Bedingung ist erfüllt:* $σ=\sqrt{243∙0,9506∙(1-0,9506)}≈3,38$

## A7. Stichprobenumfang abschätzen

**Aufgaben:**

1. Die Wahrscheinlichkeit *p* einer binomialverteilten Zufallsgröße *X* soll mit einer Stichprobe vom Umfang *n* geschätzt werden. Ermitteln Sie, wie groß der Stichprobenumfang bei einer Sicherheitswahrscheinlichkeit von 90% mindestens sein muss, damit das Konfidenzintervall die Länge 0,05 nicht überschreitet.
2. Mit einer repräsentativen Studie möchte man herausfinden, wie groß der Anteil an Erwachsenen ist, die vollständig gegen das Corona-Virus geimpft sind. Die Forschenden möchten für eine Sicherheitswahrscheinlichkeit von 95% den Anteil der vollständig geimpften Erwachsenen auf einen Prozentpunkt nach oben und unten genau schätzen. Ermitteln Sie die dafür notwendige Mindestgröße der Umfrage.
3. Man möchte mit einer (großen) repräsentativen Studie die Sensitivität eines Corona-Schnelltests für ein Konfidenzniveau von 95% schätzen. Aufgrund von Vorstudien kann davon ausgegangen werden, dass die zugrundeliegende Wahrscheinlichkeit nicht unter 92% liegt. Ermitteln Sie die notwendige Stichprobengröße, wenn man bei der Sensitivität eine Genauigkeit von 1% erreichen möchte.

Bestimmen Sie, wie viele Personen weniger in die Studie einbezogen werden müssten, wenn die Genauigkeit der Sensitivität von 2% genügen würde.

Begründen Sie, welche Auswirkungen der Unterschied von 1% bei der Genauigkeit für eine Studie bedeutet.

**Lösungen:**

1. Aus der Ungleichung $p-1,64\sqrt{\frac{p∙\left(1-p\right)}{n}}\leq h\leq p+1,64\sqrt{\frac{p∙\left(1-p\right)}{n}} $für das 90%-Konfidenzintervall ergibt sich $-1,64\sqrt{\frac{p∙(1-p)}{n}}\leq h-p\leq +1,64\sqrt{\frac{p∙(1-p)}{n}}$.

Da die Länge des Konfidenzintervalls 0,05 nicht überschreiten darf, folgt:

$\left|h-p\right|\leq +1,64∙\sqrt{\frac{p∙\left(1-p\right)}{n}}\leq \frac{0,05}{2}$ [= halbe max. Intervalllänge]

Um die Mindestgröße der Stichprobe zu ermitteln, wird die Ungleichung nach n umgeformt:

$$\left(\frac{1,64}{0,025}\right)^{2}∙p∙(1-p)\leq n$$

Da keine Schätzung von p bekannt ist, berücksichtig man den ungünstigsten Fall, also den Fall, dass $p∙(1-p)$ maximal ist. Dies gilt für $p=0,5$.

Daraus ergibt sich für den Stichprobenumfang:

$$n\geq \left(\frac{1,64}{0,025}\right)^{2}∙0,5∙(1-0,5)=1075,84$$

1. Aus der Ungleichung $p-1,96\sqrt{\frac{p∙\left(1-p\right)}{n}}\leq h\leq p+1,96\sqrt{\frac{p∙\left(1-p\right)}{n}} $für das 95%-Konfidenzintervall ergibt sich $-1,96\sqrt{\frac{p∙(1-p)}{n}}\leq h-p\leq +1,96\sqrt{\frac{p∙(1-p)}{n}}$. Da die Abweichung aus dem Ergebnis der Stichprobe und dem tatsächlichen Anteil nicht mehr als ein Prozent nach oben und unten betragen darf, folgt:

$\left|h-p\right|\leq +1,96∙\sqrt{\frac{p∙\left(1-p\right)}{n}}\leq 0,01$ [=> max. ein Prozentpunkt nach oben u. unten]]

Um die Mindestgröße der Stichprobe zu ermitteln, wird die Ungleichung nach n umgeformt:

$$\left(\frac{1,96}{0,01}\right)^{2}∙p∙(1-p)\leq n$$

Wie bei a) wird der ungünstigste Fall $p=0,5$ angenommen.

Daraus ergibt sich für den Stichprobenumfang:

$$\left(\frac{1,96}{0,01}\right)^{2}∙0,5∙\left(1-0,5\right)=9604\leq n$$

Der Stichprobenumfang muss mindestens 9604 betragen.

1. In dieser Aufgabe geht man aufgrund der vorliegenden Information von $p\geq 0,92$ aus:

Der ungünstigste Fall (vgl. a)) ist damit nicht mehr p = 0,5 sondern p = 0,92.
Daraus ergibt sich für den Stichprobenumfang:

$$\left(\frac{1,96}{0,01}\right)^{2}∙0,92∙0,08=2827,4176\leq n$$

Bei einer Genauigkeit von 1% müssen mindestens 2828 Personen für die Studie berücksichtigt werden.

 Bei einer Genauigkeit von 2% müssen mindestens $\left(\frac{1,96}{0,02}\right)^{2}∙0,92∙0,08≈707$ Personen berücksichtigt werden. Um die Genauigkeit von 2% auf 1% zu verändern, muss die Stichprobengröße hier vervierfacht werden.

**Kommentar:**

*In dieser Aufgabe wird für eine vorgegebene Länge des Konfidenzintervalls der erforderliche
Stichprobenumfang abgeschätzt.* *Während in der ersten Teilaufgabe direkt die Länge des Konfidenzintervalls vorgegeben ist, wird in den beiden folgenden Teilaufgaben jeweils die Genauigkeit der Intervallschätzung (z.B. plus/minus 1%) vorgegeben ist. Dies wird durch den Faktor ½ berücksichtig, der in der ersten Teilaufgabe (gesamte Länge vorgegeben) enthalten ist und in den folgenden Teilaufgaben (plus/minus vorgegeben) nicht.*

*Bei der letzten Teilaufgabe wird für die Sicherheitswahrscheinlichkeit die ebenfalls gebräuchliche Formulierung „Konfidenzniveau“ verwendet.*

*Die Schülerinnen und Schüler erfassen, strukturieren und formalisieren bei diesen Aufgaben Informationen aus zunehmend komplexen, mathematikhaltigen Texten. Insbesondere die Angaben zur Sicherheitswahrscheinlichkeit „95%“ und zur Genauigkeit „auf ein Prozentpunkt“ bzw. „Länge des Konfidenzintervalls“ sind hier zu unterscheiden.*

*Die verschiedenen Genauigkeiten bei Teilaufgabe c) können auch arbeitsteilig berechnet werden. Dann bietet es sich an, die großen Unterschiede der Ergebnisse zu diskutieren, wobei die Schülerinnen und Schüler begründet und konstruktiv Stellung nehmen.*

# Weitere Hinweise:

Anregungen für vernetzende Übungsaufgaben können auch den Beispielaufgaben für das Abitur nach dem KLP GOSt 2023 (vgl. Standardsicherung.nrw.de) sowie den Prüfungsaufgaben aus dem länderübergreifenden Pool des IQB entnommen werden. Exemplarisch: <https://www.iqb.hu-berlin.de/abitur/pools2020/abitur/pools2020/mathematik/erhoeht/2020_M_erhoeht_B_Stochastik_CAS_2.pdf>

Ein Quiz zu den Inhalten der Unterrichtsreihe, das sich auch im Multiplayer-Modus spielen lässt, befindet sich unter: <https://learningapps.org/watch?v=p2jdu0ihj23> (Stand: Juni 2023)

1. Angelehnt an: Wolfgang Riemer: Statistik unterrichten, Klett 2023, Kapitel 5.2 Konfidenzellipse kooperativ, S. 41. [↑](#footnote-ref-1)