C1 - Stochastik

1. ▶ **Wahrscheinlichkeiten bestimmen**

In dieser Teilaufgabe sollst du die Wahrscheinlichkeiten für folgende Ereignisse bestimmen:

A: Eine zufällig ausgewählte männliche Person aus der Altersgruppe $14-24$ wird als internetabhängig eingestuft.

B: Unter 50 zufällig ausgewählten $14-16$-jährigen Mädchen sind genau 10, deren Internetnutzung als problematisch eingestuft wird.

C: Unter 100 zufällig ausgewählten $14-16$-jährigen Mädchen sind mehr als 20, aber weniger als 30, deren Internetnutzung als problematisch oder gar als Internetabhängigkeit eingestuft wird.

Du kannst hierbei die Wahrscheinlichkeiten mit der folgenden Formel berechnen:

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

In diesem Fall ist n die Anzahl der zufällig ausgewählten Personen, X die Anzahl der Personen, deren Internetnutzung als problematisch oder als Internetabhängig eingestuft wird und p die Wahrscheinlichkeit dafür, dass eine Person aus einer bestimmten Gruppe als Internetabhängig oder deren Internetnutzung als problematisch eingestuft wird. Die jeweiligen Wahrscheinlichkeiten für die entsprechenden Altersgruppen kann man hierbei aus der gegebenen Tabelle entnehmen.

Nun sind die Wahrscheinlichkeiten für die Ereignisse A, B und C gesucht.

Wahrscheinlichkeit für Ereignis A

Ereignis A beschreibt, dass eine zufällig ausgewählte männliche Person aus der Altersgruppe $14-24$ als internetabhängig eingestuft wird. Diese Wahrscheinlichkeit hast du bereits in der Aufgabenstellung gegeben.

In der Aufgabenstellung steht beschrieben, dass $2,4\,\%$ aller männlichen Personen aus der Altersgruppe $14-24$ internetabhängig sind, also folgt somit $P(A)=2,4\,\%$.

Wahrscheinlichkeit für Ereignis B

Ereignis B beschreibt, dass unter 50 zufällig ausgewählten $14-16$-jährigen Mädchen es genau 10 gibt, deren Internetnutzung als problematisch eingestuft wird. Gesucht ist somit die Wahrscheinlichkeit $P(B)=P(X=10)$. Hierbei gilt $n=50$ und für die Wahrscheinlichkeit folgt aus der Tabelle $p_B=0,172$. Somit folgt für $P(B)$:

$$\begin{aligned}
P(X=10) &= \binom{50}{10}\cdot 0,172^{10}\cdot(1-0,172)^{50-10}\\
&= \binom{50}{10}\cdot 0,172^{10}\cdot 0,828^{40}\\
&\approx 0,1225\\
&= 12,25\,\%
\end{aligned}$$

Somit tritt Ereignis B mit einer Wahrscheinlichkeit von $12,25\,\%$ auf.

Wahrscheinlichkeit für Ereignis C

Ereignis C beschreibt, dass unter 100 zufällig ausgewählten $14-16$-jährigen Mädchen mehr als 20, aber weniger als 30 sind, deren Internetnutzung als problematisch oder gar als Internetabhängigkeit eingestuft wird. Gesucht ist somit die Wahrscheinlichkeit $P(B) = P(20 < X < 30)$. Hierbei gilt $n = 100$ und für die Wahrscheinlichkeit muss man nun beachten, dass die Internetnutzung von den $14-16$-jährigen Mädchen entweder als Internetabhängig oder als problematisch eingestuft wird. In der Aufgabenstellung ist außerdem gegeben, dass die Internetnutzung einer Person nicht gleichzeitig als Internetabhängig und als problematisch eingestuft werden kann.

Somit kann man die einzelnen Wahrscheinlichkeiten wie folgt zusammenfassen:

$$\begin{aligned} p_C &= p_{Abh} + p_{probl} \\ &= 0,049 + 0,172 \\ &= 0,221 \end{aligned}$$

Deshalb gilt $p_C = 0,221$. $P(20 < X < 30)$ kannst du umschreiben in:

$$P(20 < X < 30) = P(X \leq 29) - P(X \leq 20)$$

Aus der beiliegenden Tabelle kannst du nun für $p_C = 0,221$ die einzelnen Wahrscheinlichkeiten bestimmen und daraus folgt für $P(B)$:

$$\begin{aligned} P(20 < X < 30) &= P(X \leq 29) - P(X \leq 20) \\ &= 0,9593 - 0,3572 \\ &= 0,6021 \\ &= 60,21\,\% \end{aligned}$$

Somit tritt Ereignis C mit einer Wahrscheinlichkeit von $60,21\,\%$ auf.

2.1 ▶ **Wahrscheinlichkeit berechnen**

In dieser Teilaufgabe sollst du die Wahrscheinlichkeit für das Ereignis R berechnen. Ereignis R beschreibt, dass wenn der Test eine Person als internetabhängig einstuft, die Person auch tatsächlich internetabhängig ist.

Hierbei bezeichnen wir, dass eine Person tatsächlich internetabhängig ist mit Ereignis A und dass der Test die Person als internetabhängig einstuft mit Ereignis B.

Somit ist die Wahrscheinlichkeit $P_B(A)$ gesucht.

In der Aufgabenstellung hast du bereits die Wahrscheinlichkeit dafür gegeben, dass der Test einen tatsächlich internetabhängigen Nutzer auch als internetabhängig einstuft und dass der Test einen nicht internetabhängigen Nutzer auch als nicht internetabhängig einstuft. Diese entsprechen der Wahrscheinlichkeit $P_A(B)$ und $P_{\overline{A}}(\overline{B})$. Diese sind mit $P_A(B) = 0,86$ und $P_{\overline{A}}(\overline{B}) = 0,75$ gegeben. Mit dem **Satz von Bayes** kannst du nun die Wahrscheinlichkeit $P_B(A)$ berechnen. Der Satz von Bayes lautet:

$$P_B(A) = \frac{P_A(B) \cdot P(A)}{P(B)}$$

Nun brauchst du noch die Wahrscheinlichkeiten dafür, dass die Person tatsächlich internetabhängig ist $P(A)$ und dass der Online-Test die Person als internetabhängig einstuft $P(B)$.

Erstelle hierzu beispielsweise eine **Vierfeldertafel** und trage die gegebenen Wahrscheinlichkeiten ein:

	A	$\overline{A}$	
B			
$\overline{B}$			
	0,024		1

Die Warscheinlichkeit dafür, dass die Person tatsächlich internetabhängig ist in der Aufgabenstellung bereits gegeben, da du aus der Tabelle entnehmen kannst ,dass aus der Altersgruppe $14-24$ eine Person mit einer Wahrscheinlichkeit von $P(A)=0,024$ tatsächlich internetabhängig ist und somit die Teilnehmer des Tests auch mit einer Wahrscheinlichkeit von $0,024$ internetabhängig sind.

Dadurch kannst du $P(\overline{A})$ mit der Gegenwahrscheinlichkeit berechnen:

$$\begin{aligned} P(\overline{A}) &= 1-P(A) \\ &= 1-0,024 \\ &= 0,976 \end{aligned}$$

Die Wahrscheinlichkeiten $P(A\cap B)$ und $P(\overline{A}\cap \overline{B})$ kannst du durch die bedingten Wahrscheinlichkeiten $P_A(B)$ und $P_{\overline{A}}(\overline{B})$ bestimmen. Für eine bedingte Wahrscheinlichkeit gilt die Formel:

$$P_A(B) = \frac{P(A\cap B)}{P(A)}$$

Somit kannst du nun die Wahrscheinlichkeiten $P(A\cap B)$ und $P(\overline{A}\cap \overline{B})$ folgendermaßen bestimmen:

$$\begin{aligned} P_A(B) &= \frac{P(A\cap B)}{P(A)} \qquad |\cdot P(A) \\ P(A\cap B) &= P_A(B)\cdot P(A) \\ &= 0,86\cdot 0,024 \\ &= 0,02064 \end{aligned}$$

Für $P(\overline{A}\cap \overline{B})$ folgt entsprechend:

$$\begin{aligned} P(\overline{A}\cap\overline{B}) &= P_{\overline{A}}(\overline{B})\cdot P(\overline{A}) \\ &= 0,75\cdot 0,976 \\ &= 0,7327 \end{aligned}$$

Nun kannst du die Vierfeldertafel vervollständigen:

	A	$\overline{A}$	
B	$0,02064$	$0,244$	$0,26464$
$\overline{B}$	$0,00336$	$0,732$	$0,73536$
	$0,024$	$0,976$	1

Anschließend kannst du mit dem Satz von Bayes die gesuchte Wahrscheinlichkeit $P_B(A)$ bestimmen:

$$\begin{aligned} P_B(A) &= \frac{P_A(B)\cdot P(A)}{P(B)} \\ &= \frac{0,86\cdot 0,024}{0,26464} \\ &= 0,078 \end{aligned}$$

Somit beträgt die Wahrscheinlichkeit dafür, dass wenn der Test die Person als als internetabhängig einstuft auch tatsächlich internetabhängig ist $7,8\,\%$.

2.2 ► **Qualität des Tests beurteilen**

In dieser Teilaufgabe hast du die Wahrscheinlichkeit dafür gegeben, dass eine Person aus der Altersgruppe $14-24$ Jahre tatsächlich nicht internetabhängig ist, wenn der Test sie als nicht internetabhängig einstuft. Diese Wahrscheinlichkeit beträgt ca. $99,5\,\%$. Nun sollst du die Qualität des Test anhand dieser Wahrscheinlichkeit und aufgrund der Wahrscheinlichkeit, welche du in Teilaufgabe 2.1 bestimmt hast, bewerten.

In Teilaufgabe 2.1 hast du die Wahrscheinlichkeit dafür berechnet, dass wenn der Test die Person als internetabhängig einstuft die Person auch tatsächlich internetabhängig ist. Diese Wahrscheinlichkeit beträgt $7,8\,\%$. Sie ist deshalb sehr gering und somit kann man aufgrund eines positiven Testes keine Aussage darüber machen, ob man tatsächlich internetabhängig ist.

Die Wahrscheinlichkeit dafür, dass eine Person tatsächlich nicht internetabhängig ist, wenn der Test sie als nicht internetabhängig einstuft beträgt ca. $99,6\,\%$. Diese Wahrscheinlichkeit ist sehr hoch und deswegen kann man ziemlich sicher sagen, wenn der Test negativ ausfällt, dass die Person nich internetabhängig ist.

Der Test ist somit bei positivem Ausgang sehr ungenau und du kannst dich somit nicht auf dessen Ergebnis verlassen. Bei negativem Ergebnis des Testes kann man ziemlich sicher sagen, dass die Person, die den Test durchgeführt hat nicht internetabhängig ist.

3.1 ► **Hypothesentest entwickeln und Entscheidungsregel im Sachzusammenhang angeben**

In dieser Teilaufgabe sollst du einen geeigneten Hypothesentest auf einem Signifikanzniveau von $2,5\,\%$ durchführen und die Entscheidungsregel im Sachzusammenhang angeben. Hierbei handelt es sich um eine Stichprobe mit 100 Schülern. Deshalb gilt $n=100$. Bezeichne hierbei die Anzahl der Schüler, die zur Gruppe der Wenignutzer gehören mit der Zufallsvariable X.

Die Nullhypothese kann man nun wie folgt formulieren:

$H_0: p \geq 0,459$ und

$H_1: p < 0,459$

In Worten lautet die Nullhypothese:Es sind mindestens $45,9\,\%$ der Schüler Wenignutzer.

Das heißt du führst einen linksseitigen Hypothesentest durch. Der Ablehnungsbereich $\overline{A}$ lautet:

$\overline{A} = 0, \dots, k$

Deshalb muss gelten:

$P(X \leq k) \leq 2,5\,\%$

Aus der Tabelle folgt:

$P(X \leq 35) = 0,0178$ und $P(X \leq 36) = 0,0289$.

Somit folgt für den Ablehungsbereich:

$\overline{A} = 0, \dots, 35$

Das bedeutet, dass wenn 35 oder weniger der befragten Nutzer zu der Gruppe der Wenignutzer gehören wird die Nullhypothese verworfen und somit wird die Vermutung des Mathematik-Leisungskurs als bestätigt betrachtet.

3.2 ▶ **Fehler 1. und 2. Art beschreiben**

In dieser Teilaufgabe sollst du die Fehler 1. und 2. Art im Sachzusammenhang beschreiben und anschließend die Wahrscheinlichkeit für den Fehler 2. Art berechnen, wenn der Anteil der Wenignutzer unter den Schülern tatsächlich $40\,\%$ beträgt.

Der Fehler 1. Art beschreibt, dass die Nullhypothese abgelehnt wird, obwohl diese in Wahrheit zutrifft. Im Sachzusammenhang bedeutet dies, dass man aufgrund des Hypothesentests davon ausgeht, dass der Anteil der Wenignutzer geringer als $45,9\,\%$ ist, also dass die Nullhypothese abgelehnt wird, obwohl der Anteil in Wahrheit $45,9\,\%$ oder mehr beträgt.

Der Fehler 2. Art beschreibt, dass die Nullhypothese angenommen wird, obwohl diese in Wahrheit nicht zutrifft. Im Sachzusammenhang bedeutet dies, dass man aufgrund des Hypothesentests davon ausgeht, dass der Anteil der Wenignutzer gleich oder größer als $45,9\,\%$ ist, also dass die Nullhypothese angenommen wird, obwohl der Anteil in Wahrheit kleiner als $45,9\,\%$ ist.

Nun sollst du die Wahrscheinlichkeit für den Fehler 2. Art bestimmen, wenn der Anteil der Wenignutzer unter den Schülern tatsächlich $40\,\%$ beträgt. Das bedeutet, du berechnest die Wahrscheinlichkeit dafür, dass X im

Annahmebereich liegt, wenn in Wahrheit die Wahrscheinlichkeit $40\,\%$ gilt:

$$P\,(\text{Fehler 2. Art}) = P(X > 35)\,.$$

Somit berechnet sich die Wahrscheinlichkeit für einen Fehler 2. Art wie folgt:

$$\begin{aligned} P\,(\text{Fehler 2. Art}) &= P(X > 35) \\ &= 1 - P(X \leq 35) \\ &= 1 - 0{,}1795 \\ &= 0{,}8205 \end{aligned}$$

Somit beträgt die Wahrscheinlichkeit für einen Fehler 2. Art $82{,}05\,\%$.

3.3 ► **Güte des Tests beurteilen**

In dieser Teilaufgabe sollst du anhand geeigneter Beispielwerte beurteilen, ob sich eine Erhöhung des Stichprobenumfangs n auf die Güte des Tests auswirkt. Die Graphen im Material zeigen, wie sich bei einem gleich bleibendem Signifikanzniveau von $2{,}5\,\%$ die Fehlerwahrscheinlichkeit β für den Fehler 2. Art in Abhängigkeit vom tatsächlichen Anteil p der Wenignutzer verhält.

Somit kannst du die Wahrscheinlichkeit für den Fehler 2. Art für ein festes p betrachten. Betrachte hierbei beispielsweise $\beta(0,4)$. Also die Wahrscheinlichkeit für einen Fehler 2. Art bei $p = 0{,}4$. Nun kannst du diesen Wert für unterschiedlich große Stichprobenumfänge n betrachten und die Werte miteinander vergleichen.

Für $n = 100$ gilt: $\beta(0,4) \approx 0{,}82$

Für $n = 350$ gilt: $\beta(0,4) \approx 0{,}43$

Für $n = 10000$ gilt: $\beta(0,4) \approx 0{,}04$

Somit kann man sagen, dass je größer der Stichprobenumfang n ist, desto kleiner die Wahrscheinlichkeit für einen Fehler 2. Art ist. Wenn man den Stichprobenumfang erhöht wird somit der Test besser, da die Wahrscheinlichkeit für ein Fehler 2. Art geringer wird.