Abiturvorbereitung

Sportbegeisterung, bedingte Wahrscheinlichkeit, Hypothesentest

Aufgabenblatt

Aufgabe 11

11. Die Befragung an einem Berufskolleg ergab, dass 75% aller weiblichen Schüler (W) und 65% aller männlichen Schüler (M) gerne Sport (S) treiben.
54% aller Schüler sind weiblich.

a) Stellen Sie diesen Sachverhalt in einer Vierfeldtafel dar.

b) Wie viel Prozent aller Schüler treiben gerne Sport?

c) Zeichnen Sie das Baumdiagramm und den inversen Baum.
Berechnen Sie alle Pfadwahrscheinlichkeiten.

d) Berechnen Sie für die zufällige Auswahl eines Schülers die Wahrscheinlichkeit folgender Ereignisse:
- A: Der zufällig ausgewählte Schüler ist männlich und treibt gerne Sport.
- B: Der zufällig ausgewählte Schüler treibt gerne Sport.
- C. Der zufällig ausgewählte Schüler ist männlich. Wie groß ist die Wahrscheinlichkeit, dass dieser ungern Sport treibt?
- D: Der zufällig ausgewählte Schüler treibt gerne Sport. Mit welcher Wahrscheinlichkeit ist er weiblich?

Im Folgenden wird davon ausgegangen, dass 70% aller Schüler, gerne Sport treiben
Weiterhin wird angenommen, dass die Anzahl der Schüler, die gerne Sport treiben einer Binomialverteilung genügt.
Eine Tabelle der Binomialverteilung für n = 100 und p = 0,7 ist beigefügt.

e) Mit welcher Wahrscheinlichkeit findet man in einer Zufallsstichprobe unter 100 ausgewählten Schülern:
- (1) genau 70 Sportbegeisterte
- (2) weniger als 75 Sportbegeisterte
- (3) mindestens 60 höchstens 71 Sportbegeisterte
- (4) mehr als 75 Sportbegeisterte

f) Die Annahme p = 0,7 soll auf einem Signifikanzniveau von höchstens 10% getestet werden. Bestimmen Sie den Annahme und den Ablehnungsbereich. Überprüfen Sie die für den gewählten Ablehnungsbereich den Fehler 1. Art und kommentieren Sie das Ergebnis.

g) Berechnen Sie die Wahrscheinlichkeiten aus d) und e) mit der Tabelle der Normalverteilung und bestimmen Sie die prozentuale Abweichung der Werte bezogen auf die der Binomialverteilung.

Kumulierte Binomialverteilung für $n = 100$ und $p = 0{,}7$

k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$
50	0,000	56	0,002	62	0,053	68	0,367	74	0,837	80	0,991
51	0,000	57	0,004	63	0,080	69	0,451	75	0,886	81	0,995
52	0,000	58	0,007	64	0,116	70	0,538	76	0,924	82	0,998
53	0,000	59	0,012	65	0,163	71	0,623	77	0,952	83	0,999
54	0,001	60	0,021	66	0,221	72	0,704	78	0,971	84	1,000
55	0,001	61	0,034	67	0,289	73	0,776	79	0,984	85	1,000

E11	**Ergebnisse**		
	a)	Siehe ausführliche Lösungen.	
	b)	70,4% aller Schüler treiben gerne Sport.	
	c)	Siehe ausführliche Lösungen	
	d)	A:	$P(A)=P(M\cap S)=0,299$
		B:	$P(B)=P(S)=0,704$
		C:	$P(C)=P_M(S)=\frac{P(M\cap S)}{P(M)}=\frac{0,161}{0,46}=0,35$
		D:	$P(D)=P_S(W)=\frac{P(S\cap W)}{P(S)}=\frac{0,405}{0,704}\approx 0,575$
	e)	(1)	$P(X=70)=P(X\le 70)-P(X\le 69)=0,538-0,451=\underline{\underline{0,087}}$
		(2)	$P(X<75)=P(X\le 74)=\underline{\underline{0,837}}$
		(3)	$P(60\le X\le 71)=P(X\le 71)-P(X\le 59)=0,623-0,012=\underline{\underline{0,611}}$
		(4)	$P(X>75)=1-P(X\le 75)=1-0,886=\underline{\underline{0,114}}$
	f)	Siehe ausführliche Lösungen.	
	g)	Siehe ausführliche Lösungen.	

Ausführliche Lösungen

A11 Vierfeldtafel

a) 75% der 54% weiblicher Schüler lieben Sport $\Rightarrow W\cap S=0,75\cdot 0,64=0,405$

65% der 46% männlichen Schüler lieben Sport $\Rightarrow M\cap S=0,65\cdot 0,46=0,299$

	S	$\overline{S}$	
M	0,299	0,161	0,46
W	0,405	0,135	0,54
	0,704	0,296	1

M : Mann
W : Frau
S : sportbegeistert
$\overline{S}$: mag kein Sport

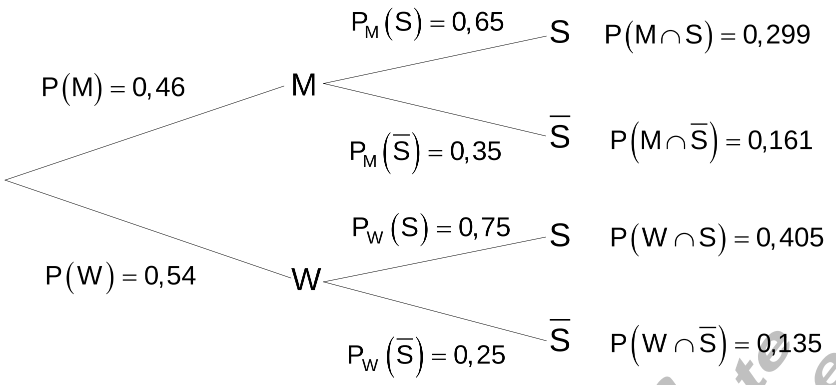
A11 Sporttreibende Schüler

b) $P(S)=0,704$

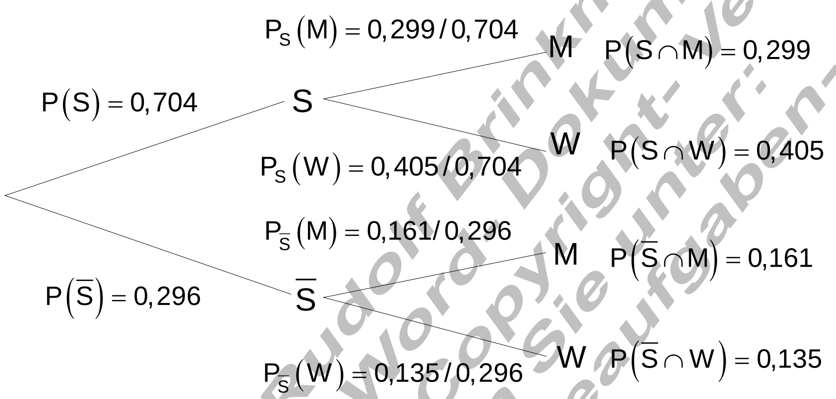
70,4% aller Schüler treiben gerne Sport.

A11 Baumdiagramme und Pfadwahrscheinlichkeiten

c) Das Baumdiagramm



Der inverse Baum



A11 Bedingte Wahrscheinlichkeit

d) $P(A)=P(M\cap S)=0,299$

Die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter Schüler männlich ist und gerne Sport treibt, beträgt 0,299.

$P(B)=P(S)=0,704$

Die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter Schüler gerne Sport treibt, ist 0,704.

$$P(C)=P_M(\overline{S})=\frac{P(M\cap \overline{S})}{P(M)}=\frac{0,161}{0,46}=0,35$$

Die Wahrscheinlichkeit dafür, dass ein Schüler, von dem man weiß, das er männlich ist, ungern Sport treibt, ist 0,35.

$$P(D)=P_S(W)=\frac{P(S\cap W)}{P(S)}=\frac{0,405}{0,704}\approx 0,575$$

Die Wahrscheinlichkeit dafür, dass ein Schüler, von dem man weiß, das er gerne Sport treibt, weiblich ist, beträgt etwa 0,575.

A11	Intervallwahrscheinlichkeiten		
	e)	(1)	$P(X=70)=P(X\le 70)-P(X\le 69)=0{,}538-0{,}451=\underline{\underline{0{,}087}}$ Die Wahrscheinlichkeit dafür, unter 100 ausgewählten Schülern genau 70 sportbegeisterte zu finden beträgt 0,087.
		(2)	$P(X<75)=P(X\le 74)=\underline{\underline{0{,}837}}$ Die Wahrscheinlichkeit dafür, unter 100 ausgewählten Schülern weniger als 75 sportbegeisterte zu finden beträgt 0,837.
		(3)	$P(60\le X\le 71)=P(X\le 71)-P(X\le 59)=0{,}623-0{,}012=\underline{\underline{0{,}611}}$ Die Wahrscheinlichkeit dafür, unter 100 ausgewählten Schülern mindestens 60 und höchstens 71 sportbegeisterte zu finden beträgt 0,611.
		(4)	$P(X>75)=1-P(X\le 75)=1-0{,}886=\underline{\underline{0{,}114}}$ Die Wahrscheinlichkeit dafür, unter 100 ausgewählten Schülern mehr als 75 sportbegeisterte zu finden beträgt 0,114.

A11	Aufgabenanalyse und aufstellen der Hypothesen.	
	f)	Es soll überprüft werden, ob der Anteil der Schüler die gerne Sport treiben 70% beträgt. Da weder eine eindeutige Abweichung nach oben oder nach unten vermutet wird, handelt es sich um einen zweiseitigen Hypothesentest. Die Hypothesen lauten: Nullhypothese: H_0: $p = 0{,}7$; Alternativhypothese H_1: $p \neq 0{,}7$. Der Ablehnungsbereich, bestimmt durch das Signifikanzniveau von 10%, verteilt sich gleichmäßig auf beide Seiten.

A11	Hypothesentest	
	f)	Nullhypothese : $H_0 : p = 0,7$ Signifikanzniveau : $\alpha \le 10\%$ Daten: $n = 100 ; p = 0,7 ; \mu = n \cdot p = 100 \cdot 0,7 = 70$ $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0,3} = \sqrt{21} \approx 4,583$ Es ist ein zweiseitiger Hypothesentest durchzuführen, denn eine geringe Anzahl, wie auch eine hohe Anzahl von Erfolgen spricht gegen H_0. Bei einem Signifikanzniveau von 10% sind folgende Intervalle zu betrachten: $\underbrace{\{\quad 5\% \quad\}}_{\text{Ablehnungsbereich für } H_0} \{\quad 90\% \quad\} \underbrace{\{\quad 5\% \quad\}}_{\text{Ablehnungsbereich für } H_0}$ Linksseitiger Ablehnungsbereich: $P(X \le k_1) \le 0,05 \Rightarrow k_1 \le 61$ denn $P(X \le 61) = 0,034$ Für den Annahmebereich gilt: $P(62 \le X \le k_2) \ge 0,9$ denn das Signifikanzniveau ist $\alpha \le 0,1$ Zu bestimmen ist die obere Grenze des Annahmebereichs. $P(X \le k_2) - P(X \le 61) \ge 0,9$ $\Leftrightarrow P(X \le k_2) - 0,034 \ge 0,9 \mid + 0,034$ $\Leftrightarrow P(X \le k_2) \ge 0,934$ Diese Bedingung ist für $k_2 = 77$ erfüllt, denn $P(X \le 77) = 0,952$ Damit wird der Annahmebereich: $A = \{62 \ldots 70 \ldots 77\}$ und der Ablehnungsbereich: $\overline{A} = \{0 \ldots 61\} \cup \{78 \ldots 100\}$ Kontrolle des Signifikanzniveaus: $P(\overline{A}) = P(X \le 61) + P(X \ge 78) = P(X \le 61) + [1 - P(X \le 77)]$ $= 0,034 + [1 - 0,952] = 0,034 + 0,048 = 0,082 < 0,1$

A11	Auswertung	
	f)	Der Fehler 1. Art beträgt 0,082. Mit einer Wahrscheinlichkeit von etwa 8,2% kann das Stichprobenergebnis im Ablehnungsbereich von H_0 liegen, so dass in diesem Fall die wahre Hypothese p = 0,7 zu Unrecht verworfen wird.

A11 Wahrscheinlichkeitsberechnung mit Hilfe der Tabelle für normalverteilte Zufallsvariablen.

g) Daten: $n = 100\,;\, p = 0{,}7\,;\, \mu = n \cdot p = 100 \cdot 0{,}7 = 70$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0{,}3} = \sqrt{21} \approx 4{,}583$$

$P(X = 70) = P(69{,}5 \le X \le 70{,}5)$

Umgebungsradius: $r = 0{,}5 \Rightarrow z = \frac{r}{\sigma} = \frac{0{,}5}{4{,}583} \approx 0{,}11 \Rightarrow P(X = 70) \approx 0{,}088$

bezogen auf 0,087 aus e) (1) ergibt das eine

Änderung auf $\frac{100\%}{0{,}087} \cdot 0{,}088 \approx \underline{\underline{101{,}1\%}}$ also 1,1% mehr

Der mit dieser Methode berechnete Wert ist etwa 1,1% zu groß.

A11 Wahrscheinlichkeitsberechnung mit Hilfe der Tabelle für normalverteilte Zufallsvariablen.

g) Daten: $n = 100\,;\, p = 0{,}7\,;\, \mu = n \cdot p = 100 \cdot 0{,}7 = 70$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0{,}3} = \sqrt{21} \approx 4{,}583$$

$P(X \le 74) \quad \{0...65\}\{66...70...74\}\{75...100\}$

$= \frac{1}{2}\left[1 - P(66 \le X \le 74)\right] + P(66 \le X \le 74) = \frac{1}{2} + \frac{1}{2} \cdot P(66 \le X \le 74)$

$P(66 \le X \le 74) = P(65{,}5 \le X \le 74{,}5)$

Umgebungsradius: $r = 4{,}5 \Rightarrow z = \frac{r}{\sigma} = \frac{4{,}5}{4{,}583} \approx 0{,}98 \Rightarrow P(66 \le X \le 74) \approx 0{,}673$

$P(X \le 74) = \frac{1}{2} + \frac{1}{2} \cdot 0{,}673 \approx 0{,}837$ bezogen auf 0,837 aus e) (2) ergibt das eine

Abweichung von $\underline{\underline{0\%}}$

Der mit dieser Methode berechnete Wert ist gleich groß.

A11 Wahrscheinlichkeitsberechnung mit Hilfe der Tabelle für normalverteilte Zufallsvariablen.

g) Daten: $n = 100\,;\ p = 0,7\,;\ \mu = n \cdot p = 100 \cdot 0,7 = 70$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0,3} = \sqrt{21} \approx 4,583$$

$P(60 \le X \le 71)$ $\quad \{60...68\}\{69...70...71\}\{72..80\}$ asymmetrische Umgebung

$$= \frac{1}{2}\left[P(60 \le X \le 80) - P(69 \le X \le 71)\right] + P(69 \le X \le 71)$$

$$= \frac{1}{2}\left[P(60 \le X \le 80) + P(69 \le X \le 71)\right]$$

$$P(60 \le X \le 80) = P(59,5 \le X \le 80,5)$$

Umgebungsradius: $r = 10,5 \Rightarrow z = \frac{r}{\sigma} = \frac{10,5}{4,583} \approx 2,29 \Rightarrow P(60 \le X \le 80) \approx 0,978$

$$P(69 \le X \le 71) = P(68,5 \le X \le 71,5)$$

Umgebungsradius: $r = 1,5 \Rightarrow z = \frac{r}{\sigma} = \frac{1,5}{4,583} \approx 0,33 \Rightarrow P(69 \le X \le 71) \approx 0,259$

$P(60 \le X \le 71) = \frac{1}{2}\left[0,978 + 0,259\right] \approx 0,619$ bezogen auf 0,611 aus e) (3)

ergibt das eine Änderung auf $\frac{100\%}{0,611} \cdot 0,619 \approx \underline{\underline{101,3\%}}$ also 1,3% mehr

Der mit dieser Methode berechnete Wert ist etwa 1,3% zu groß.

A11 Wahrscheinlichkeitsberechnung mit Hilfe der Tabelle für normalverteilte Zufallsvariablen.

g) Daten: $n = 100\,;\ p = 0,7\,;\ \mu = n \cdot p = 100 \cdot 0,7 = 70$

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0,3} = \sqrt{21} \approx 4,583$$

$P(X > 75)$ $\quad \{0...64\}\{65...70...75\}\{76...100\}$

$$= \frac{1}{2}\left[1 - P(65 \le X \le 75)\right]$$

$$P(65 \le X \le 75) = P(64,5 \le X \le 75,5)$$

Umgebungsradius: $r = 5,5 \Rightarrow z = \frac{r}{\sigma} = \frac{5,5}{4,583} \approx 1,2 \Rightarrow P(65 \le X \le 75) \approx 0,77$

$P(X > 75) = \frac{1}{2}\left[1 - 0,77\right] \approx 0,115$ bezogen auf 0,114 aus e) (4) ergibt das eine

Abweichung von $\frac{100\%}{0,114} \cdot 0,115 \approx \underline{\underline{100,9\%}}$ also 0,9% mehr

Der mit dieser Methode berechnete Wert ist 0,9% zu groß.

A11	Wahrscheinlichkeitsberechnung mit Hilfe der Tabelle für normalverteilte Zufallsvariablen.	
	g)	Daten: $n = 100$; $p = 0,7$; $\mu = n \cdot p = 100 \cdot 0,7 = 70$ $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{70 \cdot 0,3} = \sqrt{21} \approx 4,583$ Kontrolle des Signifikanzniveaus: $P(\overline{A}) = P(X \le 61) + P(X \ge 78)$ $P(X \le 61)$ $\{0...61\}\{62...70...78\}\{79...100\}$ $= \frac{1}{2}\left[1 - P(62 \le X \le 78)\right]$ $P(X \ge 78)$ $\{0...62\}\{63...70...77\}\{78...100\}$ $= \frac{1}{2}\left[1 - P(63 \le X \le 77)\right]$ $P(\overline{A}) = 1 - \frac{1}{2}\left[P(62 \le X \le 78) + P(63 \le X \le 77)\right]$ $P(62 \le X \le 78) = P(61,5 \le X \le 78,5)$ Umgebungsradius: $r = 8,5 \Rightarrow z = \frac{r}{\sigma} = \frac{8,5}{4,583} \approx 1,85 \Rightarrow P(62 \le X \le 78) \approx 0,936$ $P(63 \le X \le 77) = P(62,5 \le X \le 77,5)$ Umgebungsradius: $r = 7,5 \Rightarrow z = \frac{r}{\sigma} = \frac{7,5}{4,583} \approx 1,64 \Rightarrow P(63 \le X \le 77) \approx 0,899$ $P(\overline{A}) = 1 - \frac{1}{2}\left[0,936 + 0,899\right]$ $P(\overline{A}) = 1 - \frac{1}{2}\left[0,936 + 0,899\right] \approx 0,0825$ bezogen auf 0,082 aus f) ergibt das eine Abweichung von $\frac{100\%}{0,082} \cdot 0,0825 \approx \underline{\underline{100,6\%}}$ also $0,6\%$ mehr Der mit dieser Methode berechnete Wert ist 0,6% zu groß.