Lösung

Lösung zu Teilaufgabe 1 (10 BE)

Die Angaben im Aufgabentext können in die folgende Mehrfeldertafel übersetzt werden:

	m	w	
GS	1 533	15 504	17 037
GYM	5 452	6 663	12 115
andere	11 858	19 347	31 205
	18 843	41 514	60 357

Dabei steht GS für Grundschule und GYM für Gymnasium.

Mit Hilfe der Angaben aus der Tabelle ergibt sich für den Anteil der weiblichen Lehrkräfte an den beruflichen und allgemeinbildenden Schulen wegen $\frac{41\,514}{60\,357} \approx 0{,}688$ ein Wert von etwa 69 %.

Die Wahrscheinlichkeit, dass eine zufällig ausgewählte weibliche Lehrkraft an einer Grundschule angestellt war, beträgt

$$P_{\mathrm{w}}(\mathrm{GS}) = \frac{P(\mathrm{GS} \cap \mathrm{w})}{P(\mathrm{w})} = \frac{|\mathrm{GS} \cap \mathrm{w}|}{|\mathrm{w}|} = \frac{15\,504}{41\,514} \approx 0{,}37.$$

Lösung zu Teilaufgabe 2 (9 BE)

In dieser Aufgabe soll die Wahrscheinlichkeit dafür berechnet werden, dass unter 8 Lehrkräften aus dem Werra-Meißner-Kreis, die zufällig ausgewählt wurden, genau 5 Frauen und 3 Männer sind. Es sind drei Lösungsvorschläge gegeben, die bewertet werden sollen. Dabei ist außerdem bekannt, dass ein Vorschlag eine exakte Lösung liefert, ein anderer eine Näherungslösung. Der dritte Vorschlag ist falsch.

Ansatz A
Der erste Ansatz

$$P(\mathrm{E}) = \frac{\binom{577}{5} \cdot \binom{346}{3}}{\binom{923}{8}}$$

liefert das richtige Ergebnis.

Aus den 923 im Werra-Meißner-Kreis beschäftigen Lehrkräften werden zufällig 8 Personen ausgewählt. Die Auswahl geschieht ohne Wiederholung und ohne Beachtung der Reihenfolge, sodass $\binom{923}{8}$ Kombinationen möglich sind. Analog wird die Anzahl der möglichen Teilmengen zur Auswahl der 5 weiblichen und 3 männlichen Lehrkräfte mit Hilfe des Binomialkoeffizienten berechnet, für die Wahrscheinlichkeit ergibt sich der oben genannte Ansatz.

Ansatz B
Ansatz B liefert das falsche Ergebnis. Mit

$$P(\mathrm{E}) = \frac{577}{923} \cdot \frac{576}{922} \cdot \ldots \cdot \frac{573}{919} \cdot \frac{346}{918} \cdot \frac{345}{918} \cdot \frac{344}{916}$$

wird lediglich die Wahrscheinlichkeit dafür berechnet, dass erst 5 weibliche und anschließend noch 3 männliche Lehrkräfte ausgesucht werden, die Anzahl der möglichen Kombinationen fehlt. Wird Ansatz B um den Faktor $\binom{8}{5}$ ergänzt, stimmt er mit Ansatz A überein und liefert das richtige Ergebnis.

Ansatz C
Mit

$$P(\mathrm{E}) = \binom{8}{5} \cdot \left(\frac{577}{923}\right)^5 \cdot \left(\frac{346}{923}\right)^3$$

wird ein Näherungswert für die gesuchte Wahrscheinlichkeit ermittelt. Hier wird im Gegensatz zu Ansatz B mit $\binom{8}{5}$ die Anzahl der möglichen Kombinationen berücksichtigt, die Wahrscheinlichkeiten für einen weiblichen/ männlichen Lehrer werden hier jedoch als konstant angesehen, obwohl sie sich in Wirklichkeit nach jeder Auswahl ändern, weil sich die Anzahl der insgesamt Möglichen verringert.

Für einen großen Stichprobenumfang ist es grundsätzlich legitim, eine Bernoulli-Kette anzusetzen und die -leichte- Änderung der zu Grunde liegenden Wahrscheinlichkeit nicht zu berücksichtigen. Ein exaktes Ergebnis wird jedoch nur durch den Ansatz der hypergeometrischen Verteilung aus Ansatz A erreicht.

Lösung zu Teilaufgabe 3.1 (4 BE)

In Teilaufgabe 3.1 soll untersucht werden, ob eine großangelegte Werbekampagne zur Erhöhung des Männeranteils unter den Studienanfängern für das Grundschullehramt Erfolg zeigt oder nicht. Hierfür werden 50 Abiturientinnen und Abiturienten nach ihren Berufszielen befragt, es wird ein einseitiger Signifikanztest zur Bewertung der Ergebnisse durchgeführt:

In der Nullhypothese H_0 wird immer davon ausgegangen, dass sich nichts verändert hat, die Werbeagentur untersucht daher die Alternativhypothese H_1 mit $p_1 > 0{,}1$. Finden sich unter den 50 befragten Schülern 8 oder mehr männliche Studienanfänger des Grundschullehramtes, soll die Nullhypothese verworfen und der Erfolg der Werbekampagne anerkannt werden. Finden sich maximal 7 Schüler, so wird weiter von einem zu Grunde liegenden Anteil männlicher Studienanfänger von $p_0 = 0{,}1$ ausgegangen.

Ein Fehler 2. Art liegt immer dann vor, wenn die Nullhypothese fälschlicherweise angenommen wird. Im konkreten Fall bedeutet das, dass die Werbekampagne zwar erfolgreich war, sich aber dennoch unter den 50 befragten Schülern maximal 7 männliche Studienanfänger finden.

Der Fehler 2. Art kann nur berechnet werden, wenn eine Wahrscheinlichkeit für die Alternativhypothese bekannt ist. In der Aufgabe war $p_1 = \frac{1}{6}$ vorgegeben, damit ergibt sich mit X der Anzahl der männlichen Studienanfänger:

$$\begin{aligned}
P\left(\text{Fehler 2. Art}\right) &= P_{\frac{1}{6}}\left(X \leq 7\right) \\
&= \mathrm{F}\left(50; \frac{1}{6}; 7\right) = \sum_{k=0}^{7} \binom{50}{k} \cdot \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{50-k} \\
&= 0{,}3911
\end{aligned}$$

Der Fehler 2. Art beträgt bei diesem Signifikanztest also etwa 39 %.

Lösung zu Teilaufgabe 3.2 (7 BE)

In dieser Teilaufgabe soll ein weiterer Hypothesentest entwickelt werden, der die Behauptung der Werbeagentur auf dem 5%-Niveau bestätigen soll. Es werden 225 Studienanfänger befragt, die kritische Zahl k ist gesucht, oberhalb derer die Nullhypothese verworfen werden muss.

Das Signifikanzniveau eines Hypothesentests gibt die Größe des Fehlers 1. Art an. Im betrachteten Fall liegt ein Fehler 1. Art dann vor, wenn die Werbekampagne zwar keinen Einfluss auf die Studienwahl hatte und die Anzahl der männlichen Studienanfänger unverändert bei $p_0 = 0{,}1$ geblieben ist, sich unter den 225 Studienanfängern jedoch trotzdem k oder mehr männliche Studienanfänger finden.

k muss jetzt so bestimmt werden, dass

Unterscheide:

$X \leq k$

$$\Rightarrow z = \frac{k - \mu + 0{,}5}{\sigma}$$

$X < k$

$$\Rightarrow z = \frac{k - \mu - 0{,}5}{\sigma}$$

$$P_{0,1}(X \geq k) = 1 - \mathrm{F}(225; 0{,}1; (k-1)) \leq 0{,}05$$

gilt. Sofern eine entsprechende Tabelle vorhanden ist, lässt sich $k \geq 31$ direkt ablesen. Finden sich unter den 225 Studienanfängern 31 oder mehr männliche Studenten, muss die Nullhypothese auf dem 5%-Niveau abgelehnt werden.

Alternativ kann die Bedingung mit Hilfe der Normalverteilung angenähert werden, weil wegen $\sigma = \sqrt{225 \cdot 0{,}1 \cdot 0{,}9} = 4{,}5$ $\sigma > 3$ gilt. Es ist

$$\begin{aligned} z &= \frac{k - \mu - 0{,}5}{\sigma} \\ &= \frac{k - 22{,}5 - 0{,}5}{4{,}5}. \end{aligned}$$

Es muss $1 - \Phi(z) \leq 0{,}05$ gelten, daraus ergibt sich $z \geq 1{,}65$ und $k \geq 31$.

Erläuterung: *Binomial- und Normalverteilung I*

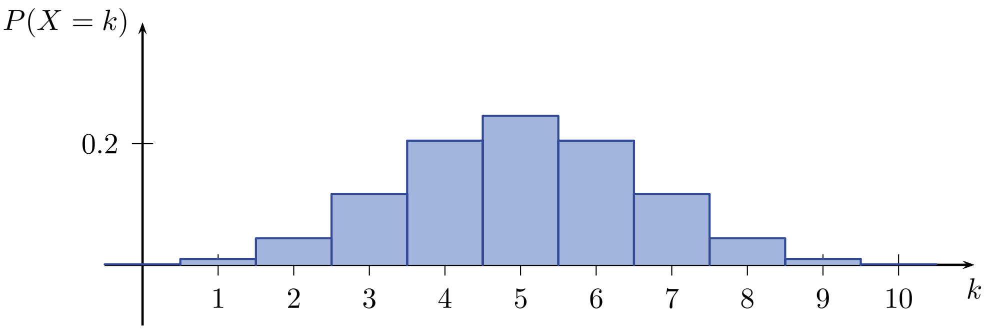
Die Binomialverteilung

Die Binomialverteilung ist die Wahrscheinlichkeitsverteilung eines Bernoulli-Experimentes, also eines Zufallsexperimentes mit genau zwei Ausgängen. Für ein n-stufiges Zufallsexperiment mit Trefferwahrscheinlichkeit p beträgt die Wahrscheinlichkeit für k Treffer gerade

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

der Erwartungswert der Verteilung wird über $\mu=n\cdot p$ erhalten, die Standardabweichung über $\sigma=\sqrt{n\cdot p\cdot(1-p)}$.

Die folgende Abbildung zeigt die Wahrscheinlichkeitsverteilung einer binomialverteilten Zufallsgröße X mit $n=10$ und $p=0{,}5$:



Die Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung, sie bezieht sich also immer auf eine endliche Menge n. Jedem $k<n$ wird eine feste Wahrscheinlichkeit $P(X=k)$ zugeordnet, eine Wahrscheinlichkeit $P\left(a\leq X\leq b\right)$ wird durch

$$P\left(a\leq X\leq b\right)=\sum_{x=a}^{b}\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}$$

berechnet, dabei werden die Flächeninhalte der einzelnen Streifen des Histogramms aufsummiert.

Erläuterung: *Binomial- und Normalverteilung II*

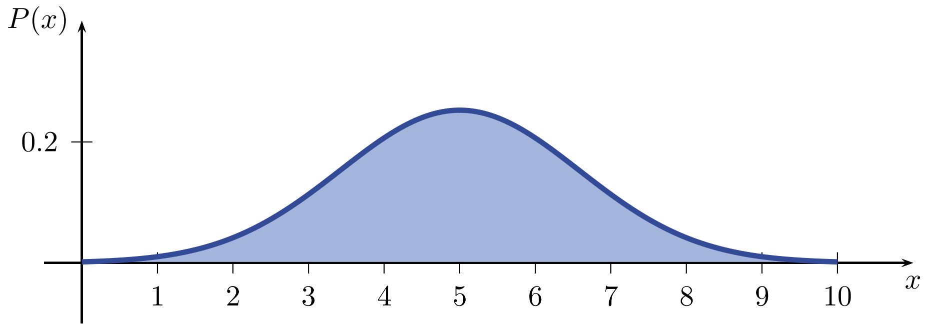
Die Normalverteilung

Im Gegensatz zur Binomialverteilung handelt es sich bei der Normalverteilung um eine kontinuierliche Verteilung. Sie erstreckt sich über eine unendliche Menge, jedem Punkt der Menge wird die Wahrscheinlichkeit 0 zugeordnet. Der Graph der Normalverteilung

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

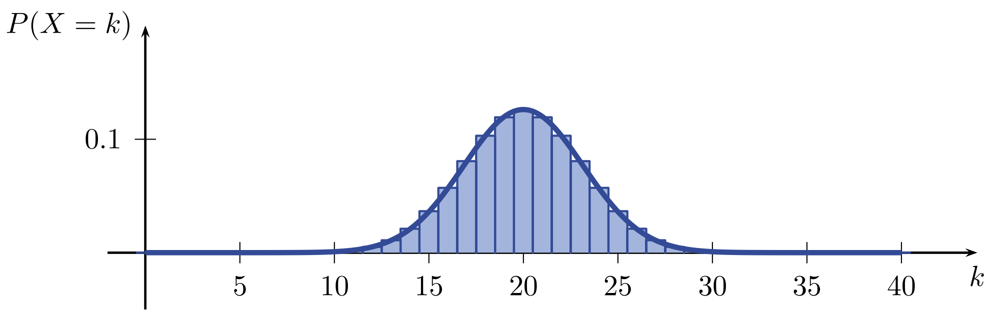
wird auch als Gauß'sche Glockenkurve bezeichnet.

Die folgende Abbildung zeigt die Wahrscheinlichkeitsverteilung einer normalverteilten Zufallsgröße mit $\mu = 5$ und $\sigma = 1{,}58$.



Für $n \to \infty$ konvertiert die Binomialverteilung gegen die Normalverteilung. Für einen genügend großen Stichprobenumfang kann daher die Binomialverteilung durch die Normalverteilung angenähert werden.
Als Kriterium für die Anwendbarkeit der Näherung gibt es den Satz von Moivre-Laplace: Gilt $\sqrt{np(1-p)} > 3$, so kann die Normalverteilung als Näherung verwendet werden.

Der folgende Graph zeigt die Wahrscheinlichkeitsverteilung einer binomialverteilten Zufallsgröße mit $n = 40$ und $p = 0{,}5$. Wegen $\sqrt{40 \cdot 0{,}5^2} = 10$ kann die Verteilung durch die entsprechende Normalverteilung angenähert werden:

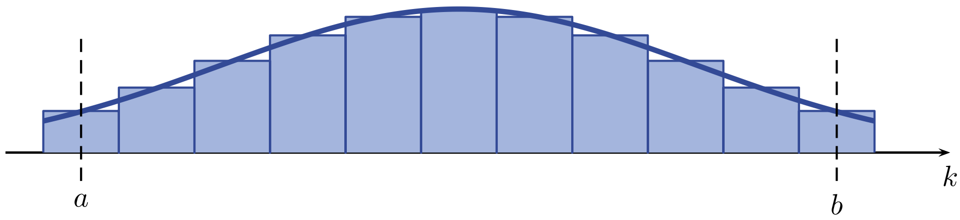


Erläuterung: *Binomial- und Normalverteilung III*

Die Wahrscheinlichkeit $P\,(a \leq X \leq b)$ kann jetzt durch das Integral der Gauß-Kurve angenähert werden:

$$P\,(a \leq X \leq b) \approx \int_a^b f(x)\,\mathrm{dx} = \mathrm{F(b)} - \mathrm{F(a)}$$

Dabei muss weiter berücksichtigt werden, dass die Verwendung der Grenzen a und b im Integral verglichen mit den Streifen im Histogramm der Binomialverteilung nicht die vollständige Fläche berücksichtigen:



In den Integralgrenzen muss also die Hälfte der Breite der Streifen, also 0,5, addiert bzw. subtrahiert werden, je nach dem ob es sich um die obere oder die untere Grenze handelt:

$$P\,(a \leq X \leq b) \approx \int_{a-0{,}5}^{b+0{,}5} f(x)\,\mathrm{dx} = \mathrm{F(b+0{,}5)} - \mathrm{F(a-0{,}5)}$$

Um das Integral nicht für jede Verteilung neu berechnen zu müssen, wird die Normalverteilung auf eine Verteilung mit $\mu = 0$ und $\sigma = 1$, die sogenannte Standardnormalverteilung, mit

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{z^2}{2}}$$

normiert. Die Verteilungsfunktion dieser Standardnormalverteilung

$$\Phi(z) = \int_{\infty}^{z} \frac{1}{\sqrt{2\pi}}\,e^{-\frac{t^2}{2}}\,\mathrm{dt}$$

ist tabellarisiert und muss daher nicht jedes Mal neu gebildet werden.

Um die Werte der Standardnormalverteilung nutzen zu können, muss lediglich die Zufallsgröße der Binomialverteilung normiert werden. Der Erwartungswert der normierten Zufallsgröße soll bei $\mu = 0$ liegen, dies wird durch die Subtraktion von $\mu = n \cdot p$ realisiert. Die Division durch $\sigma = \sqrt{np(1-p)}$ normiert die Standardabweichung auf 1:

$$P\,(a \leq X \leq b) \approx \Phi\left(\frac{b-\mu+0{,}5}{\sigma}\right) - \Phi\left(\frac{a-\mu-0{,}5}{\sigma}\right)$$