Statistik I für Human- und Sozialwissenschaften

5. Übung

Lösungsvorschlag

Gruppenübung

G 12 Die Wahrscheinlichkeit dafür, dass ein Samenkorn einer bestimmten Blumenart unter geeigneten Versuchsbedingungen innerhalb von 14 Tagen keimt, betrage 0.95. Bei einem Versuch werden 100 solche Samenkörner beobachtet. Die Zufallsvariable X beschreibe die Anzahl derjenigen Samenkörner, die innerhalb von 14 Tagen keimen.

a) Geben Sie unter geeigneten Annahmen die Verteilung von X an.

b) Geben Sie den Erwartungswert und die Varianz von X an.

c) Berechnen Sie die Wahrscheinlichkeit $P(X \geq 98)$.

d) Bestimmen Sie mit Hilfe des Zentralen Grenzwertsatzes jeweils einen Näherungswert für die Wahrscheinlichkeit $P(X \geq 98)$. Überprüfen Sie dabei die Voraussetzungen des Satzes.

a) *Geht man davon aus, dass die Samenkörner unabhängig voneinander keimen, ist X binomialverteilt mit den Parametern $n = 100$ und $p = 0.95$.*

b) *Es gilt für eine binomialverteilte Zufallsvariable $E(X) = np$ und $Var(X) = np(1-p)$. Mit den Parametern aus a) folgt daher:*

$$E(X) = 100 \cdot 0.95 = 95 \qquad \text{und} \qquad Var(X) = 100 \cdot 0.95 \cdot 0.05 = 4.75.$$

c) *Wir beachten, dass X diskret ist und berechnen mit*

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

$$\begin{aligned} P(X \geq 98) &= P(X = 98) + P(X = 99) + P(X = 100) \\ &= \binom{100}{98} \cdot 0.95^{98} \cdot 0.05^2 + \binom{100}{99} \cdot 0.95^{99} \cdot 0.05 + 0.95^{100} \\ &\approx 0.0812 + 0.0312 + 0.0059 = 0.1183. \end{aligned}$$

d) *Beim Zentralen Grenzwertsatz geht man von einer Folge von unabhängigen Zufallsvariablen $X_1, X_2, \ldots$ aus mit $E(X_i) = \mu_i < \infty$ und $Var(X_i) = \sigma_i^2 < \infty$ für alle $i \in \mathbb{N}$. Der Satz besagt, dass die standardisierte Summe der ersten n Zufallsvariablen für großes n näherungsweise $N(0,1)$-verteilt ist.*

Es beschreibe X_i die Lebensdauer eines Kornes, dann gilt in unserer Situation für X:

$$X = \sum_{i=1}^{100} X_i,$$

also $n = 100$. Aus b) wissen wir ja schon, dass $E(X) = 95$ und $Var(X) = 4.75$ gilt, so *dass wir wie folgt standardisieren müssen:*

$$\begin{aligned} P(X \geq 98) &= 1 - P(X \leq 97) \\ &= 1 - P\left(\frac{X-95}{\sqrt{4.75}} \leq \frac{97-95}{\sqrt{4.75}}\right) \\ &\approx 1 - \Phi\Big(\frac{97-95}{\sqrt{4.75}}\Big) \approx 1 - \Phi(0.92) \\ &= 1 - 0.8212 = 0.1788 \ . \end{aligned}$$

Im Vergleich zum exakten Ergebnis (s. c)) ist die Näherung ziemlich schlecht. Das liegt vor allem an dem Wert für p, der sehr weit vom symmetrischen Fall ($p = 0.5$) entfernt ist, so dass die Verteilung von X hier sehr schief ist, was nicht so gut zur symmetrischen Normalverteilung passt.
Auf obiges Ergebnis kommt man auch, wenn man streng nach der Formel im Skript vorgeht und $\mu_1 + \ldots + \mu_{100} = 1 \cdot 0.95 + \ldots + 1 \cdot 0.95 = 100 \cdot 0.95 = 95$ bzw. $\sigma_1^2 + \ldots + \sigma_{100}^2 = 1 \cdot 0.95 \cdot 0.05 + \ldots + 1 \cdot 0.95 \cdot 0.05 = 100 \cdot 0.95 \cdot 0.05 = 4.75$ berechnet.

G 13 In einem Labor wird eine Flüssigkeit maschinell in 40 Reagenzgläser gefüllt. Die Maschine ist auf einen Abfüllwert von 10 [ml] eingestellt. Aus Erfahrung weiß man, dass dabei eine Standardabweichung von 0.5 [ml] auftritt. Nach einem weiteren Verarbeitungsschritt, welcher das Volumen der abgefüllten Menge nicht beeinflusst, werden die 40 Proben in ein Gefäß umgefüllt. Mit Y bezeichnen wir das Volumen der insgesamt abgefüllten Flüssigkeit.

a) Bestimmen Sie den Erwartungswert und die Varianz von Y.

b) Berechnen Sie mit Hilfe des Zentralen Grenzwertsatzes einen Näherungswert für die Wahrscheinlichkeit, dass Y um mehr als 2 ml von $E(Y)$ abweicht.

c) Wiederholen Sie a) und b) für 80, 120 und 160 Reagenzgläser. Was fällt Ihnen auf?

a) *Bezeichnet man mit X_i die Abfüllmenge im i-ten Reagenzglas ($i = 1, \ldots, 40$), ergibt sich Y als die Summe dieser Zufallsvariablen, also*

$$Y = X_1 + \ldots + X_{40}.$$

Dabei sind die X_i unabhängig und identisch verteilt mit Erwartungswert 10 [ml] und Varianz 0.5^2 [ml^2]. Daraus folgt

$$\begin{aligned} E(Y) &= E(X_1) + \ldots + E(X_{40}) = 40 \cdot E(X_1) = 400, \\ Var(Y) &= Var(X_1) + \ldots + Var(X_{40}) = 40 \cdot Var(X_1) = 40 \cdot 0.25 = 10. \end{aligned}$$

b) *Es geht um die Wahrscheinlichkeit $P(|Y - E(Y)| > 2)$. Dabei ist Y eine Summe unabhängiger identischer verteilter Zufallsvariablen (vgl. a)), weshalb der Zentrale Grenzwertsatz hier angewendet werden kann. Wir formen um:*

$$\begin{aligned} P(|Y - E(Y)| > 2) &= 1 - P(|Y - E(Y)| \leq 2) \\ &= 1 - P(-2 \leq Y - 400 \leq 2) \\ &= 1 - P\left(\frac{-2}{\sqrt{10}} \leq \frac{Y-400}{\sqrt{10}} \leq \frac{2}{\sqrt{10}}\right) \\ &\approx 1 - \left(\Phi\left(\frac{2}{\sqrt{10}}\right) - \Phi\left(\frac{-2}{\sqrt{10}}\right)\right) \\ &= 1 - (2\Phi(0.63) - 1) = 2 \cdot (1 - \Phi(0.63)) \\ &= 2 \cdot (1 - 0.7357) = 0.5286 \ . \end{aligned}$$

c) *Wir führen die Rechnungen komplett mit der Variablen n für die Anzahl der Reagenzgläser durch und setzen anschließend die drei vorgegebenen Werte ein. Zunächst erhält man in a) für $Y = X_1 + \ldots + X_n$*

$$E(Y) = n \cdot 10 \qquad \text{und} \qquad Var(Y) = n \cdot 0.25\ .$$

Daraus folgt für die Berechnung der Wahrscheinlichkeit eine andere Standardisierung (der Erwartungswert ist bei der Rechnung übrigens nicht von Belang)

$$\begin{aligned} P(|Y - E(Y)| > 2) &= 1 - P(|Y - E(Y)| \leq 2) \\ &= 1 - P(-2 \leq Y - E(Y) \leq 2) \\ &= 1 - P\left(\frac{-2}{\sqrt{0.25n}} \leq \frac{Y - E(Y)}{\sqrt{0.25n}} \leq \frac{2}{\sqrt{0.25n}}\right) \\ &= 1 - P\left(\frac{-4}{\sqrt{n}} \leq \frac{Y - E(Y)}{\sqrt{0.25n}} \leq \frac{4}{\sqrt{n}}\right) \\ &\approx 1 - \left(\Phi\left(\frac{4}{\sqrt{n}}\right) - \Phi\left(\frac{-4}{\sqrt{n}}\right)\right) \\ &= 1 - \left(2\Phi\left(\frac{4}{\sqrt{n}}\right) - 1\right) = 2 \cdot \left(1 - \Phi\left(\frac{4}{\sqrt{n}}\right)\right). \end{aligned}$$

Für $n = 80$ erhält man damit

$$\begin{aligned} P(|Y - E(Y)| > 2) &\approx 2 \cdot \left(1 - \Phi\left(\frac{4}{\sqrt{n}}\right)\right) \\ &= 2 \cdot (1 - \Phi(0.45)) = 2 \cdot (1 - 0.6736) = 0.6528. \end{aligned}$$

Analog ergibt sich

$$\begin{aligned} &\text{n=120:} \quad P(|Y - E(Y)| > 2) \approx 2 \cdot (1 - \Phi(0.37)) = 2 \cdot (1 - 0.6443) = 0.7114, \\ &\text{n=160:} \quad P(|Y - E(Y)| > 2) \approx 2 \cdot (1 - \Phi(0.32)) = 2 \cdot (1 - 0.6255) = 0.749\ . \end{aligned}$$

Die Wahrscheinlichkeiten wachsen mit n, was daran liegt, dass auch die Varianzen der Summe mit wachsendem n größer werden und somit eine Abweichung von mehr als 2 ml (dieser Wert bleibt ja konstant) vom Erwartungswert wahrscheinlicher wird.

G 14 Berechnen Sie folgende Integrale:

$$\text{a)} \int_1^2 (6x^2 - 6x + 4)\, dx, \quad \text{b)} \int_{\frac{5}{3}}^2 e^{3t-5}\, dt, \quad \text{c)} \int_{-1}^1 x e^x\, dx.$$

a) *Die Stammfunktion von x^n lautet $\frac{1}{n+1}x^{n+1}$ für $n \geq 0$. Damit erhalten wir:*

$$\begin{aligned} \int_1^2 (6x^2 - 6x + 4)\, dx &= (2x^3 - 3x^2 + 4x)\Big|_1^2 \\ &= (16 - 12 + 8) - (2 - 3 + 4) = 9. \end{aligned}$$

b) *Wir beachten, dass die Stammfunktion von e^x wieder e^x ist und berücksichtigen die innere Ableitung:*

$$\begin{aligned} \int_{\frac{5}{3}}^2 e^{3t-5}\, dt &= \frac{1}{3}\, e^{3t-5}\Big|_{5/3}^2 \\ &= \frac{1}{3}e^{3 \cdot 2 - 5} - \frac{1}{3}e^{3 \cdot 5/3 - 5} = \frac{e - 1}{3}. \end{aligned}$$

c)

$$\begin{aligned}
\int_{-1}^{1} \underbrace{x}_{g(x)} \underbrace{e^x}_{f'(x)}\, dx &= = \underbrace{e^x}_{f(x)} \underbrace{x}_{g(x)} \Big|_{-1}^{1} - \int_{-1}^{1} \underbrace{e^x}_{f(x)} \cdot \underbrace{1}_{g'(x)}\, dx \\
&= e^1 \cdot 1 - e^{-1} \cdot (-1) - (e^x)\Big|_{-1}^{1} \\
&= e + \frac{1}{e} - (e - e^{-1}) = \frac{2}{e}.
\end{aligned}$$

Hausübung

H 7 Nach der Einnahme eines bestimmten Medikaments treten bei einer zufällig ausgewählten Person mit Wahrscheinlichkeit $p = 0.02$ Nebenwirkungen auf. Das Medikament werde 200 Patienten verabreicht. Man kann annehmen, dass die Nebenwirkungen bei den einzelnen Personen unabhängig voneinander auftreten. Die Zufallsvariable X beschreibe die Anzahl der Personen, bei denen Nebenwirkungen beobachtet werden.

a) Wie ist die Zufallsvariable X verteilt?

b) Berechnen Sie die Wahrscheinlichkeit $P(X \leq 2)$.

c) Berechnen Sie die Wahrscheinlichkeit $P(X > 1)$.

d) Geben Sie mit Hilfe des Zentralen Grenzwertsatzes jeweils einen Näherungswert für die Wahrscheinlichkeiten $P(X \leq 2)$ und $P(4 \leq X \leq 9)$ an.

a) *Die Zufallsvariable X ist binomialverteilt mit Parametern $n = 200$ und $p = 0.02$.*

b)

$$\begin{aligned}
P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
&= 0.98^{200} + \binom{200}{1} \cdot 0.98^{199} \cdot 0.02 + \binom{200}{2} \cdot 0.98^{198} \cdot 0.02^2 \\
&\approx 0.0176 + 0.0718 + 0.1458 = 0.2352.
\end{aligned}$$

c) *Wir berechnen die gesuchte Wahrscheinlichkeit über die Gegenwahrscheinlichkeit:*

$$\begin{aligned}
P(X > 1) &= 1 - (P(X = 0) + P(X = 1)) \\
&= 1 - \left(0.98^{200} + \binom{200}{1} \cdot 0.98^{199} \cdot 0.02\right) \\
&\approx 1 - (0.0176 + 0.0718) = 0.9106.
\end{aligned}$$

d) *Es gilt $E(X) = np = 4$ sowie $Var(X) = np(1-p) = 3.92$. Die Standardisierung erfolgt demnach so:*

$$\begin{aligned}
P(X \leq 2) &= P\left(\frac{X-4}{\sqrt{3.92}} \leq \frac{2-4}{\sqrt{3.92}}\right) \\
&\approx \Phi\Big(\frac{-2}{\sqrt{3.92}}\Big) \approx \Phi(-1.01) = 1 - 0.8438 = 0.1562
\end{aligned}$$

$$\begin{aligned}
P(4 \leq X \leq 9) &= P(X \leq 9) - P(X \leq 3) \\
&= P\left(\frac{X-4}{\sqrt{3.92}} \leq \frac{9-4}{\sqrt{3.92}}\right) - P\left(\frac{X-4}{\sqrt{3.92}} \leq \frac{3-4}{\sqrt{3.92}}\right) \\
&\approx \Phi\Big(\frac{9-4}{\sqrt{3.92}}\Big) - \Phi\Big(\frac{3-4}{\sqrt{3.92}}\Big) \\
&\approx \Phi(2.53) - \Phi(-0.51) \\
&= 0.9943 - (1 - 0.6950) = 0.6893\ .
\end{aligned}$$

H 8 Bei der Beladung eines LKW mit Kisten muss darauf geachtet werden, dass das Gewicht der Ladung höchstens 7.8 Tonnen beträgt. Die Gewichte [in kg] der einzelnen Kisten sollen durch identisch stetig verteilte Zufallsvariablen $X_1, \ldots, X_n$ beschrieben werden, für die folgende Dichte angenommen wird :

$$f(x) = \begin{cases} \frac{1}{30} & \text{für } 105 \leq x \leq 135 \\ 0 & \text{sonst.} \end{cases}$$

a) Bestimmen Sie den Erwartungswert und die Varianz des Gewichts einer einzelnen Kiste.

b) Berechnen Sie unter der Unabhängigkeitsannahme einen Näherungswert für die Wahrscheinlichkeit, dass das zulässige Gewicht der Ladung eingehalten wird, wenn auf dem LKW $n = 64$ Kisten geladen werden.

a) *Das Gewicht einer einzelnen Kiste kann durch eine Zufallsvariable X beschrieben werden, welche die in der Aufgabenstellung genannte Dichte besitzt. Es gilt:*

$$\begin{aligned} E(X) &= \int_{105}^{135} x f(x) dx = \int_{105}^{135} \tfrac{1}{30} x dx \\ &= \tfrac{1}{60} x^2 \Big|_{105}^{135} = \tfrac{135^2}{60} - \tfrac{105^2}{60} = 120 \\ Var(X) &= E(X^2) - (E(X))^2 = \int_{105}^{135} x^2 f(x) dx - 120^2 \\ &= \int_{105}^{135} x^2 \tfrac{1}{30} dx - 120^2 = \tfrac{1}{90} x^3 \Big|_{105}^{135} - 120^2 = 75 \end{aligned}$$

b) *Unter Verwendung des Zentralen Grenzwertsatzes erhält man*

$$\begin{aligned} P(Y \leq 7800) &= P\left(\frac{Y - E(Y)}{\sqrt{Var(Y)}} \leq \frac{7800 - E(Y)}{\sqrt{Var(Y)}} \right) \\ &= P\left(\frac{Y - 64 \cdot 120}{\sqrt{64 \cdot 75}} \leq \frac{7800 - 64 \cdot 120}{\sqrt{64 \cdot 75}} \right) \\ &= P\left(\frac{Y - 7680}{\sqrt{4800}} \leq \frac{7800 - 7680}{\sqrt{4800}} \right) \\ &= P\left(\frac{Y - 7680}{\sqrt{4800}} \leq \sqrt{3} \right) \\ &\approx \Phi(\sqrt{3}) \approx \Phi(1.73) = 0.958. \end{aligned}$$