Übungen zur Wahrscheinlichkeitsrechnung und Statistik für Biologen

Blatt 3

1. Aufgabe (Berechnung von Erwartungswerten) Sei X eine Zufallsvariable mit möglichen Werten in $\{1, 2, 3\}$ und Verteilung

$$\Pr(X = 1) = \frac{1}{2} \quad \Pr(X = 2) = \frac{1}{3} \quad \Pr(X = 3) = \frac{1}{6}$$

Berechnen Sie per Hand

(a) den Erwartungswert $\mathbb{E}X$ von X (Ergebnis: $\frac{5}{3}$),

(b) $\mathbb{E}(X^2)$,

(c) die Varianz $\mathrm{Var}(X)$ von X,

(d) $\mathbb{E}(X^2 + X)$, und

(e) $\mathbb{E}\Big[(X + 1)^2\Big]$.

(Hinweis: Verwenden Sie dazu die Definition des Erwartungswerts und/oder Linearität.)

2. Aufgabe Sei X die Zufallsvariable aus Aufgabe 1. Desweiteren sei Y eine Zufallsvariable mit Werten in $\{0, 1\}$. Die gemeinsame Verteilung von (X, Y) sei

$$\begin{aligned}
&\Pr(X = 1, Y = 0) = \frac{1}{2} \qquad \Pr(X = 1, Y = 1) = 0 \\
&\Pr(X = 2, Y = 0) = \frac{1}{6} = \Pr(X = 2, Y = 1) \\
&\Pr(X = 3, Y = 0) = \frac{1}{12} = \Pr(X = 3, Y = 1).
\end{aligned}$$

Berechnen Sie per Hand

(a) $\Pr(Y = 0)$ und $\Pr(Y = 1)$,

(b) den Erwartungswert $\mathbb{E}Y$ von Y,

(c) $\mathbb{E}(Y^2)$,

(d) die Standardabweichung $\sqrt{\mathrm{Var}(Y)}$ von Y,

(e) die Kovarianz $\mathrm{Cov}(X, Y)$ von X und Y, und

(f) die Korrelation $\mathrm{Cor}(X, Y)$ zwischen X und Y.

3. Aufgabe Sei $n \in \mathbb{N}$ und seien $x_1, x_2, \ldots, x_n$ reelle Zahlen (z.B. $n = 5$ und $x_1 = 2.5$, $x_2 = 5$, $x_3 = 5$, $x_4 = -1$, $x_5 = -1$). Ziehe rein zufällig eine dieser Zahlen. Sei X die gezogene Zahl. (Zur besseren Vorstellung: Beschrifte n Kugeln gleicher Bauart mit den Zahlen $x_1, \ldots, x_n$. Durchmische die Kugeln in einem grünen Beutel und ziehe eine Kugel. Sei X die Zahl auf der gezogenen Kugel.) Zeigen Sie:

(a) Die Verteilung von X erfüllt

$$\Pr(X = a) = \frac{1}{n}\#\{i\colon x_i = a\} \qquad \text{für alle } a \in \{x_1, \ldots, x_n\},$$

wobei $\#\{i\colon x_i = a\}$ die Anzahl der Elemente der Menge ist. Im Zahlenbeispiel ist beispielsweise $\Pr(X = -1) = \frac{2}{5}$. (Diese Verteilung ist die sogenannte *empirische Verteilung* von $x_1, \ldots, x_n$.)

(b) Der Erwartungswert von X ist

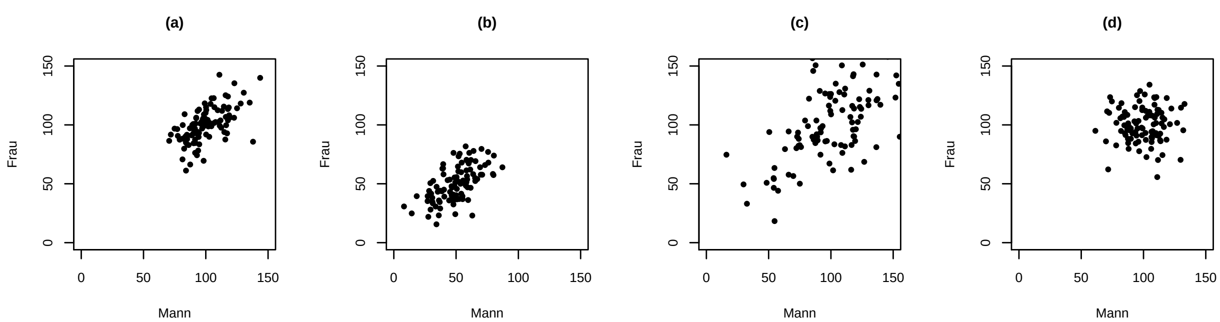
$$\mathbb{E}X = \frac{1}{n}\sum_{i=1}^{n} x_i =: \overline{x}$$

(c) Die Varianz von X ist

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

(d) Geben Sie $\mathbb{E}X$ und $\mathrm{Var}(X)$ für das Zahlenbeispiel $n = 6, x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 6$ an ($\rightsquigarrow$ Würfel).

4. Aufgabe (Qualitative Aussagen aufgrund der Korrelation) 100 (simulierte) Ehepaare nahmen an einem Intelligenztest teil. Der Mittelwert des Resultats war bei beiden Geschlechtern $\overline{x}_F = \overline{x}_M = 100$ mit Standardabweichung $s_F = s_M = 15$. Wenn man jedes Ehepaar als zweidimensionale Beobachtung (x_F, x_M) auffasst, ergab sich ein Korrelationskoeffizient $\rho = 0{,}6$.
Welcher der folgenden Scatterplots passt zu diesen Beobachtungen? Begründen Sie für Bilder (a)–(d), warum sie in Frage kommen oder nicht.



Gegeben seien gepaarte Beobachtungen (x, y). Stimmen Sie den folgenden Aussagen a)–d) zu? Begründen Sie jeweils Ihre Antwort.

(a) Wenn der Korrelationskoeffizient gleich $-0{,}75$ ist, gehören unterdurchschnittliche x-Werte eher zu überdurchschnittlichen y-Werten.

(b) Wenn die y-Werte stets kleiner als die x-Werte sind, ist der Korrelationskoeffizient negativ.

(c) Wenn der Korrelationskoeffizient 1 oder -1 ist, kann man mittels des x-Werts den y-Wert exakt vorhersagen.

(d) Wenn der Korrelationskoeffizient gleich 0 ist, dann sind die den Beobachtungen zu Grunde liegenden Verteilungen unabhängig voneinander.

5. Aufgabe Seien X, Y, Z drei reelle Zufallsvariablen und $c \in \mathbb{R}$ eine Konstante. Zeigen Sie

(a) $\text{Cov}(c \cdot X, Y) = c \cdot \text{Cov}(X, Y)$,

(b) $\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$,

(c) $\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$.

Es gelte für den Rest der Aufgabe $\mathbb{E}Z = 0$ und $\text{Var}(Z) = 1$. Seien $\mu, \sigma \in \mathbb{R}$ Konstanten. Berechnen Sie

(d) $\mathbb{E}(\sigma \cdot Z + \mu)$

(e) $\text{Var}(\sigma \cdot Z + \mu)$

(Hinweis: Nutzen Sie die Linearität des Erwartungswertes aus. Die Definition des Erwartungswertes ist hier nicht zu verwenden.)

6. Aufgabe Bei der Weihnachtsfeier des 30-köpfigen Frauenchors "Die Silberkehlen" stellt sich zu vorgerückter Stunde heraus, dass genau die Hälfte der Damen die Zunge rollen kann. Der Chorleiter, hauptberuflich Oberstudienrat für Musik und Mathematik, bemerkt daraufhin mit seiner ganz und gar unathletischen, aber etwas spitzen Zunge: "Na, dann sehen wir ja jetzt, dass Musikalität und Zungenrollen einander eher ausschließen, denn normalerweise können ja 2/3 aller Frauen die Zunge rollen. Unter der Nullhypothese, dass der Chor rein zufällig zusammengesetzt ist, ist die Wahrscheinlichkeit, dass genau die Hälfte die Zunge rollen kann, 0,024687, also unter dem allgemein anerkannten Signifikanzniveau von 0,05." Der zweite Sopran, Diplom-Psychologin und Zungenrollerin, wendet daraufhin etwas säuerlich ein, dass hier die Wahrscheinlichkeit, um 5 oder mehr vom Erwartungswert abzuweichen, bei 0,0789 liege, und dass daher von Signifikanz keine Rede sein könne.

(a) Überprüfen Sie, dass unter der genannten Nullhypothese gilt: Die Wahrscheinlichkeit, dass genau die Hälfte die Zunge rollen kann, ist 0,024687. Die Wahrscheinlichkeit, um 5 oder mehr vom Erwartungswert abzuweichen, ist 0,0789. Benutze zur Berechnung der Binomialverteilung den R-Befehl `dbinom`. Beispielsweise ist `dbinom(20,size=30,prob=2/3)` die Wahrscheinlichkeit, bei 30 Versuchen mit Erfolgswahrscheinlichkeit $\frac{2}{3}$ genau 20 Erfolge zu sehen.

(b) Welche Argumentation ist sinnvoller?

(c) Der Oberstudienrat schlägt implizit folgenden Test vor, um bei einer Gruppe von 30 Frauen zu testen, ob "eine Frau mit Wahrscheinlichkeit 2/3 die Zunge rollen kann": Sei k die Anzahl der 30 Frauen, die die Zunge rollen kann. Falls $\Pr(X = k) =$`dbinom(k,size=30,prob=2/3)`$\leq 5\%$, dann lehne die Nullhypothese "eine Frau kann mit Wahrscheinlichkeit 2/3 die Zunge rollen": ab. Mit welcher Wahrscheinlichkeit wird bei diesem Test fälschlicherweise Signifikanz angezeigt? Bestimmen Sie dazu die Menge $A = \{k \in \{0, 1, \ldots, 30\} : \text{ Nullhypothese wird abgelehnt } \}$ und berechnen Sie die Wahrscheinlichkeit von A unter bin(30,2/3).