Stochastik

Hausaufgabenblatt 10

Patrick Gustav Blaneck

Letzte Änderung: 8. Januar 2022

1. In den 30 Museen der Stadt Artima gab es im letzen Monat jeweils X Neuerwerbungen pro Museum. Dabei sei folgende Urliste entstanden:

$$\begin{array}{cccccccccc} 2 & 4 & 3 & 5 & 5 & 2 & 3 & 1 & 5 & 6 \\ 4 & 7 & 8 & 3 & 2 & 8 & 3 & 6 & 4 & 6 \\ 5 & 7 & 3 & 3 & 2 & 5 & 4 & 4 & 3 & 11 \end{array}$$

   (a) Erstellen Sie eine Tabelle mit der absoluten und relativen Häufigkeit bzw. Summenhäufigkeit der Neuerwerbungen X pro Museum.

Lösung:

Es gilt:

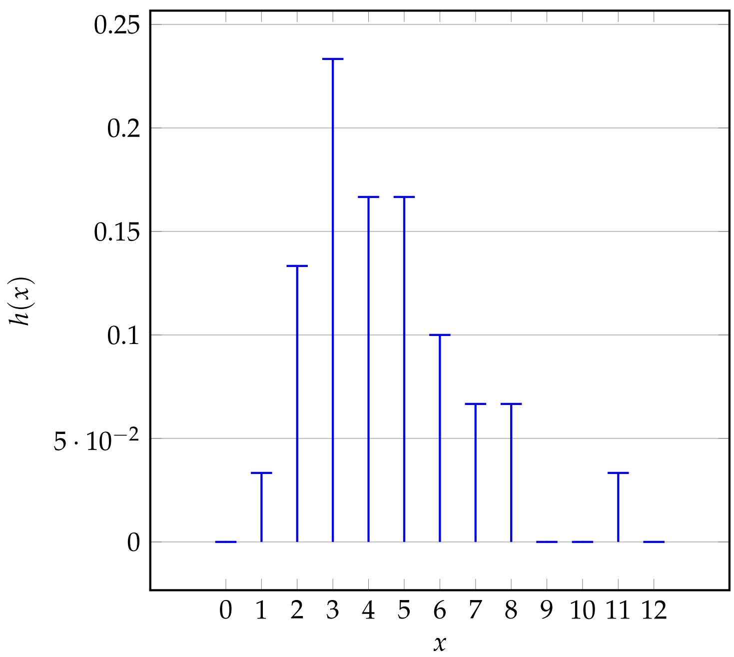
x_i	n_i	h_i	H_i
1	1	$1/30$	$1/30$
2	4	$2/15$	$1/6$
3	7	$7/30$	$2/5$
4	5	$1/6$	$17/30$
5	5	$1/6$	$11/15$
6	3	$1/10$	$5/6$
7	2	$1/15$	$9/10$
8	2	$1/15$	$29/30$
11	1	$1/30$	1

$\square$

(b) Zeichnen Sie im Anschluss

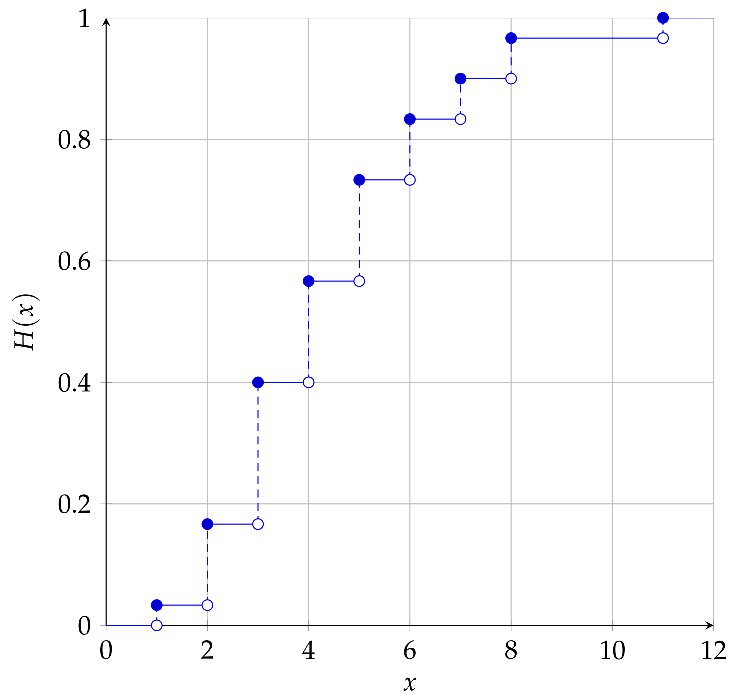
i. das zugehörige Stabdiagramm

Lösung:



ii. die empirische Verteilungsfunktion.

Lösung:



(c) Berechnen Sie

i. das arithmetische Mittel

Lösung:

Es gilt:

$$\bar{x} = \frac{1}{n} \cdot \sum_i x_i = \frac{1}{30} \cdot 134 = \frac{67}{15} \approx 4.467$$

$\square$

ii. den Median

Lösung:

Es gilt:

$$\tilde{x} = x_{1/2} = 4$$

$\square$

iii. das 10%-Quantil

Lösung:

Es gilt:

$$x_{1/10} = 2$$

□

iv. obere Quartil

Lösung:

Es gilt:

$$x_{3/4} = 6$$

□

(d) Berechnen Sie die empirische Varianz und die empirische Standardabweichung.

Lösung:

Es gilt:

$$\begin{aligned}
s^2 &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2 \right) \\
&= \frac{1}{29} \cdot \left(\sum_{i=1}^{30} x_i^2 - 30 \cdot \left(\frac{67}{15} \right)^2 \right) \\
&= \frac{1}{29} \cdot \left(\sum_{i=1}^{30} x_i^2 - \frac{8978}{15} \right) \\
&= \frac{1}{29} \cdot \sum_{i=1}^{30} x_i^2 - \frac{8978}{435} \\
&= \frac{1}{29} \cdot 740 - \frac{8978}{435} \\
&= \frac{2122}{435} \approx 4.878
\end{aligned}$$

Und damit auch:

$$s = \sqrt{s^2} = \sqrt{\frac{2122}{435}} \approx 2.209$$

□

2. Die nachfolgende Tabelle gibt eine Übersicht über die Anzahl der verkauften Bücher zu unterschiedlichen Preisen in einer Buchhandlung im Laufe eines Tages:

Buchpreis (in €)	Anzahl der verkauften Bücher
$[0;10)$	5
$[10;30)$	15
$[30;50)$	20
$[50;80)$	12
$[80;120)$	8

(a) Berechnen Sie die jeweiligen absoluten und relativen Klassenhäufigkeiten

Lösung:

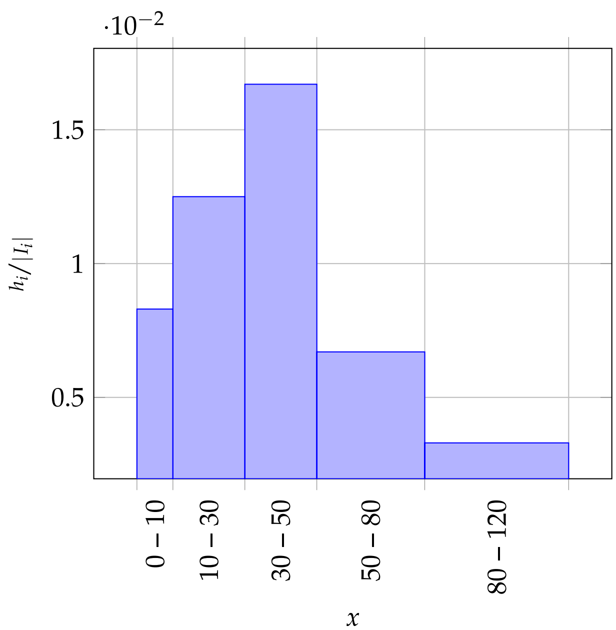
Es gilt:

I_i	n_i	h_i	H_i	$h_i/\lvert I_i\rvert$
$[0;10)$	1	$1/12$	$1/12$	$1/120 \approx 0.0083$
$[10;30)$	4	$1/4$	$1/3$	$1/80 = 0.0125$
$[30;50)$	7	$1/3$	$2/3$	$1/60 \approx 0.0167$
$[50;80)$	5	$1/5$	$13/15$	$1/150 \approx 0.0067$
$[80;120)$	5	$2/15$	1	$1/300 \approx 0.0033$

□

(b) Zeichnen Sie das zugehörige Histogramm.

Lösung:



□

(c) Bestimmen Sie

i. das arithmetische Mittel,

Lösung:

Es gilt:

$$\bar{x} \approx \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot \alpha_i = \sum_{i=1}^{k} h_i \cdot \alpha_i = \frac{5}{12} + 5 + \frac{40}{3} + 13 + \frac{40}{3} = \frac{541}{12} \approx 45.083$$

□

ii. den Median sowie

Lösung:

Offensichtlich ist die Einfallsklasse gegeben mit

$$I_3 = [30; 50) = [a_3; b_3)$$

Es gilt damit:

$$\tilde{x} = a_3 + \frac{1/2 - H_2}{h_3} \cdot (b_3 - a_3) = 30 + \frac{1/2 - 1/3}{1/3} \cdot (50 - 30) = 40$$

□

iii. das obere und untere Quartil.

Lösung:

Offensichtlich ist die Einfallsklasse für das untere Quantil gegeben mit

$$I_2 = [10; 30) = [a_2; b_2)$$

und die für das obere Quartil mit

$$I_4 = [50; 80) = [a_4; b_4)$$

Es gilt damit:

$$x_{1/4} = a_2 + \frac{1/4 - H_1}{h_2} \cdot (b_2 - a_2) = 10 + \frac{1/4 - 1/12}{1/4} \cdot (30 - 10) = \frac{70}{3}$$

und

$$x_{3/4} = a_4 + \frac{3/4 - H_3}{h_4} \cdot (b_4 - a_4) = 50 + \frac{3/4 - 2/3}{1/5} \cdot (80 - 50) = \frac{130}{2} = 62.5$$

□

3. In einer (kleinen) Bankfiliale werden die vergebenen Kredite untersucht:

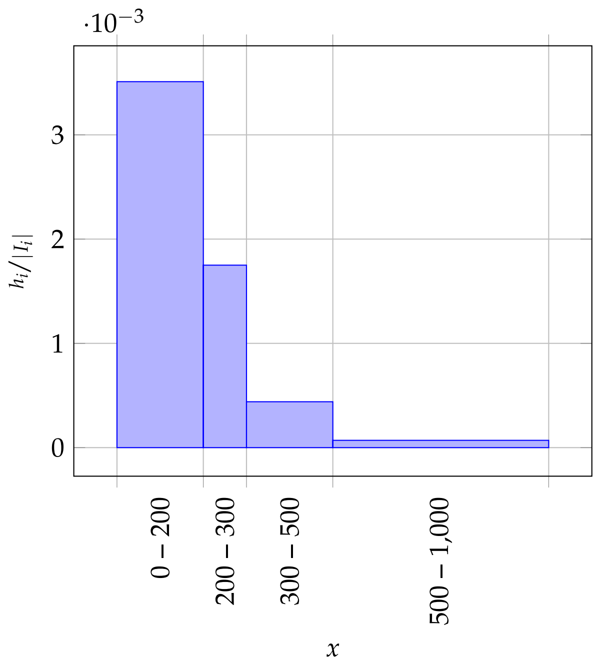
Kredithöhe (in Tausend €)	Anzahl der Kredite
$[0; 200)$	40
$[200; 300)$	10
$[300; 500)$	5
$[500; 1000)$	2

(a) Erstellen Sie ein Histogramm.

Lösung:

Es gilt:

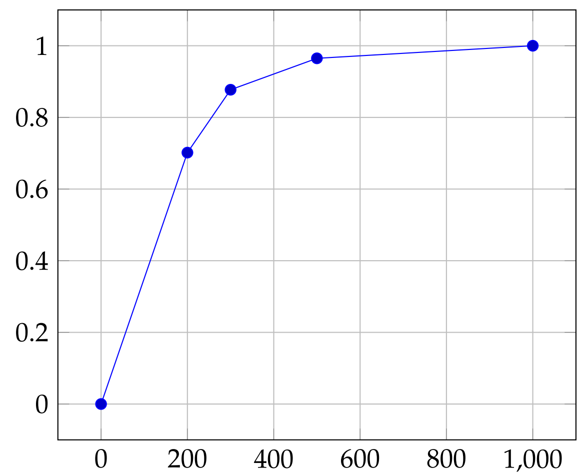
I_i	n_i	h_i	H_i	$h_i/\|I_i\|$
$[0; 200)$	40	$40/57$	$40/57$	$1/285 \approx 0.00351$
$[200; 300)$	10	$10/57$	$50/57$	$1/570 \approx 0.00175$
$[300; 500)$	5	$5/57$	$55/57$	$1/2280 \approx 0.00044$
$[500; 1000)$	2	$2/57$	1	$1/14250 \approx 0.00007$



□

(b) Zeichnen Sie die empirische Verteilungsfunktion.

Lösung:



$\square$

(c) Bestimmen Sie den Median und das obere Quartil.

Lösung:

Offensichtlich ist die Einfallsklasse für den Median gegeben mit

$$I_1 = [0; 200) = [a_1; b_1)$$

und die für das obere Quantil mit

$$I_2 = [200; 300) = [a_2; b_2)$$

Es gilt damit:

$$\tilde{x} = a_1 + \frac{1/2 - H_0}{h_1} \cdot (b_1 - a_1) = 0 + \frac{1/2}{40/57} \cdot (200 - 0) = \frac{285}{2} = 142.5$$

und

$$x_{3/4} = a_2 + \frac{3/4 - H_1}{h_2} \cdot (b_2 - a_2) = 200 + \frac{3/4 - 40/57}{10/57} \cdot (300 - 200) = \frac{455}{2} = 227.5$$

$\square$

4. Die Schülerinnen und Schüler wurden vor der Klausur anonym befragt, wie viele Stunden Schlaf sie vor der Klausur gehabt haben:

Note	2	4	3	5	6	6	1	2	5	4
Schlaf	9	5	6	5	1	2	9	8	4	7

(a) Berechnen Sie aus den Daten

i. die empirische Kovarianz

Lösung:

N und S seien wie offensichtlich definiert. TO DO: Umbenennen in X und Y

Es gilt:

$$\operatorname{Cov}(X,Y) = s_{XY} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i y_i - n \cdot \bar{x} \cdot \bar{y} \right)$$

Wir berechnen zuerst die arithmetischen Mittel von X und Y wie folgt:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{10} \cdot (2+4+3+5+6+6+1+2+5+4) = \frac{19}{5} = 3.8$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{10} \cdot (9+5+6+5+1+2+9+8+4+7) = \frac{28}{5} = 5.6$$

Damit gilt:

$$\begin{aligned}
\operatorname{Cov}(X,Y) = s_{xy} &= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \\
&= \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i y_i - n \cdot \bar{x} \cdot \bar{y} \right) \\
&= \frac{1}{9} \left(\sum_{i=1}^{n} x_i y_i - 10 \cdot \frac{19}{5} \cdot \frac{28}{5} \right) \\
&= \frac{1}{9} \left(\sum_{i=1}^{n} x_i y_i - \frac{1064}{5} \right) \\
&= \frac{1}{9} \sum_{i=1}^{n} x_i y_i - \frac{1064}{45} \\
&= \frac{1}{9} \cdot 172 - \frac{1064}{45} \\
&= -\frac{68}{15} \approx -4.53
\end{aligned}$$

□

ii. den empirischen Korrelationskoeffizient

Lösung:

Es gilt:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Wir berechnen zuerst die empirischen Standardabweichungen s_x und s_y von X und Y wie folgt:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{9}\sum_{i=1}^{n}\left(x_i - \frac{19}{5}\right)^2 = \frac{46}{15} \implies s_x = \sqrt{\frac{46}{15}} = \frac{\sqrt{690}}{15}$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{9}\sum_{i=1}^{n}\left(y_i - \frac{28}{5}\right)^2 = \frac{38}{5} \implies s_y = \sqrt{\frac{38}{5}} = \frac{\sqrt{190}}{5}$$

Damit gilt dann:

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-68/15}{\sqrt{690}/15 \cdot \sqrt{190}/5} = -\frac{34\sqrt{1311}}{1311} \approx -0.93903$$

□

iii. die lineare Regression

Lösung:

Die Regressionsgerade ist gegeben mit

$$\hat{y}_i = \hat{a} + \hat{b}x$$

Es gilt:

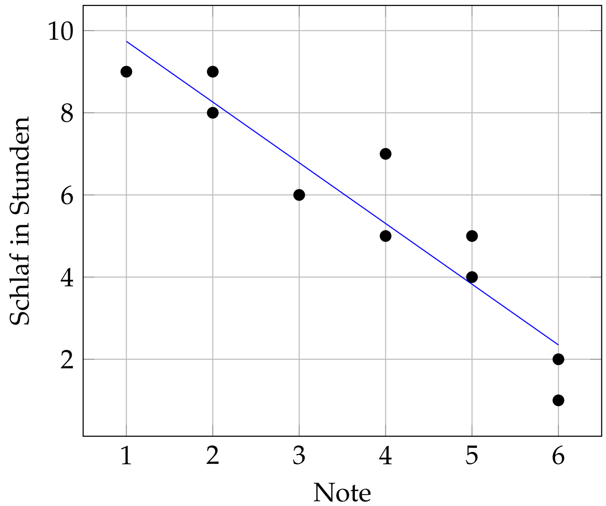
$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{-68/15}{46/15} = -\frac{34}{23} \approx -1.478$$

$$\hat{a} = \bar{y} - \hat{b} \cdot \bar{x} = \frac{28}{5} + \frac{34}{23} \cdot \frac{19}{5} = \frac{258}{23} \approx 11.217$$

□

(b) Erstellen Sie ein Streudiagramm und zeichnen Sie die Regressionsgerade ein.

Lösung:



□