

Aufgabe 3.1: Quantile, emp. Vtlgsfkt. und Boxplot (teilw. R) (5 Punkte)



Gegeben ist der folgende Datenvektor:

```
x <- c(1, 2, 1, 2, 10, 10, 20, 1, 2, 1)
```

Lösen Sie folgende Aufgaben jeweils zuerst auf Papier und dann mit R:

- (i) Geben Sie die empirische Verteilungsfunktion $F_{10}(x)$ an und zeichnen Sie den Graph von F . Mit R brauchen Sie nur den Graphen zu zeichnen.
- (ii) Berechnen Sie jeweils alle der folgenden empirischen Quantile x_{p_i} mit

```
p <- c(0.2, 0.25, 0.4, 0.5, 0.6, 0.75, 0.99)
```

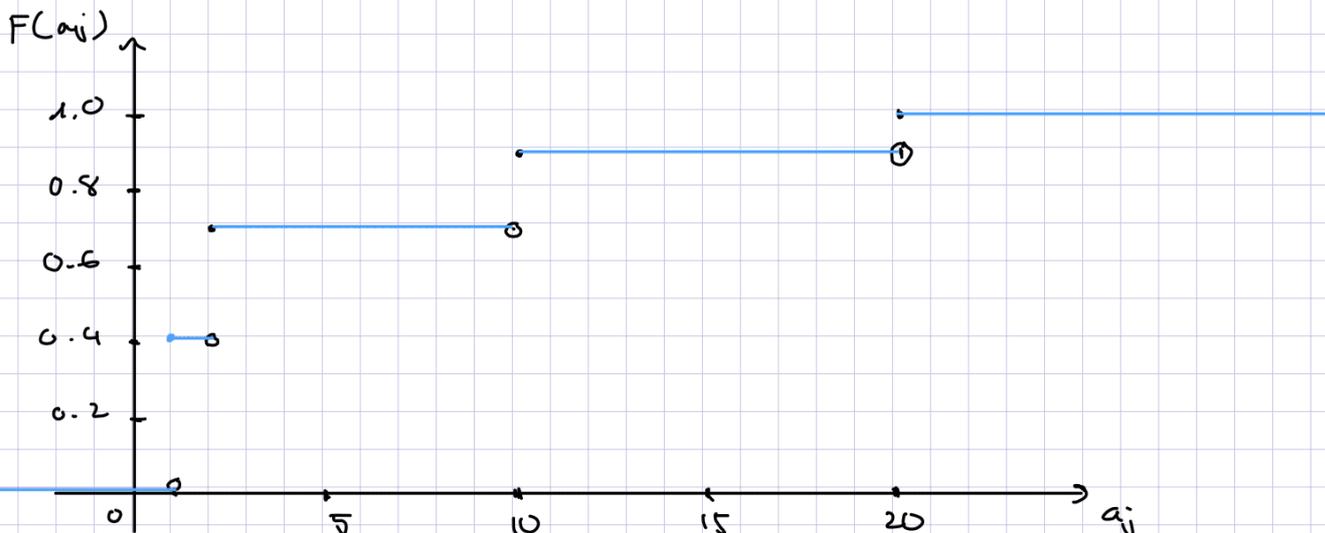
wobei p_i der i -te Vektoreintrag ist ($i = 1, \dots, 7$).

- (iii) Zeichnen Sie einen Boxplot zum Datenvektor x .

1 1 1 2 2 2 10 10 20

(i)	j	1	2	3	4	Σ
	a_j	1	2	10	20	
	$g(a_j)$	4	3	2	1	$\bar{\Sigma} = 10$
	$h(a_j)$	4	7	9	10	
	$g(a_j)$	0,4	0,3	0,2	0,1	$\bar{\Sigma} = 1$
	$F(a_j)$	0,4	0,7	0,9	1,0	

$$F_{10}(x) = \begin{cases} 0 & \text{für } x < 1 \\ 0,4 & \text{für } 1 \leq x < 2 \\ 0,7 & \text{für } 2 \leq x < 10 \\ 0,9 & \text{für } 10 \leq x < 20 \\ 1 & \text{für } x \geq 20 \end{cases}$$

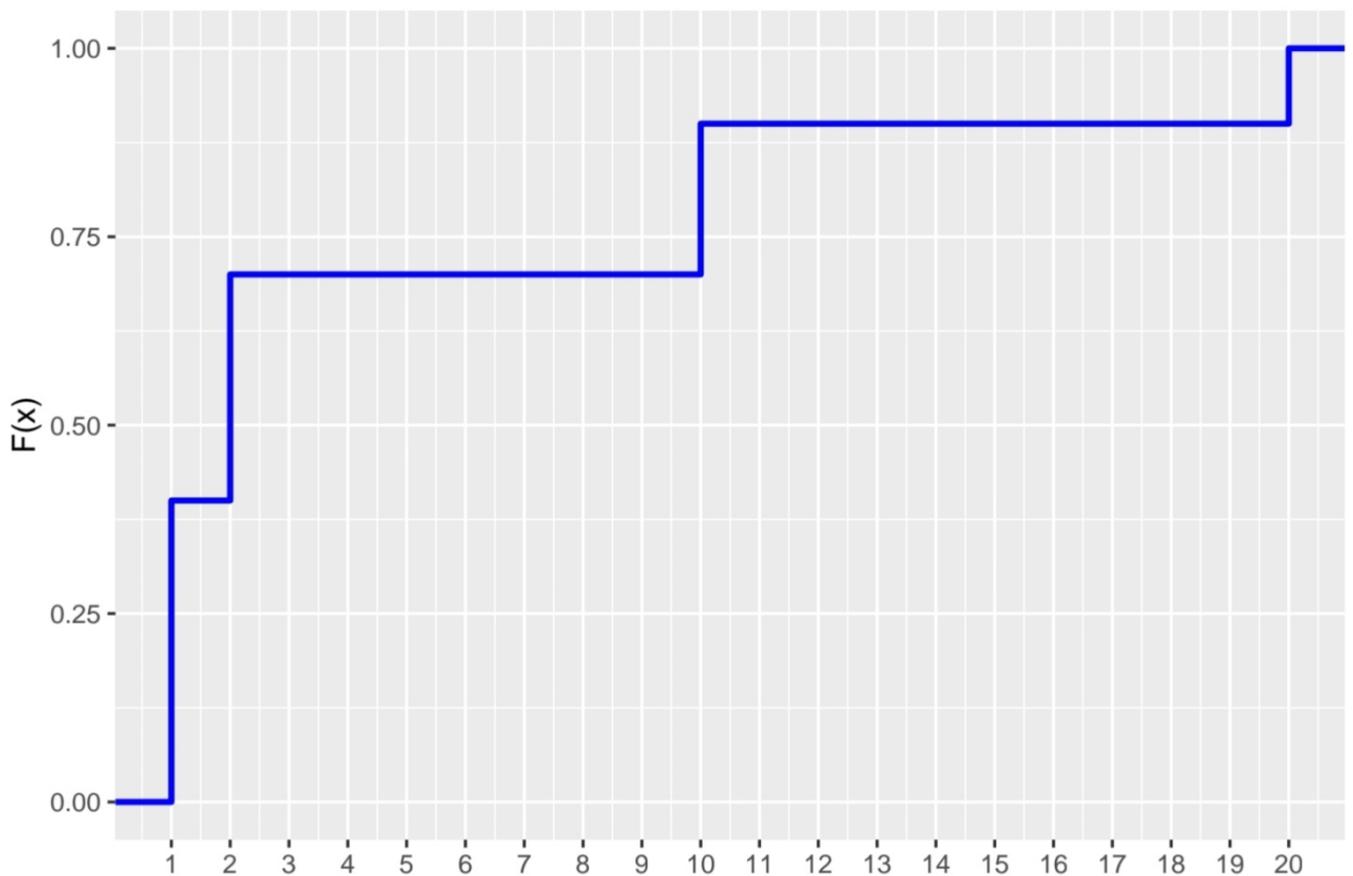


Aufgabe 3.1

```
##(i)
library(pacman)
p_load(ggplot2, tidyverse)
x <- c(1,2,1,2,10,10,20,1,2,1)
x_sorted <- sort(x)
x.tib <- tibble(x)

F_x <- cumsum(table(x_sorted))/length(x)

x.tib |> ggplot(aes(x = x)) +
  stat_ecdf(linewidth = 1, color = "blue") +
  labs(y = "F(x)") +
  scale_x_continuous(breaks = 0:20)
```



$$(ii) \quad x_{P_{0.2}} = x_{[2,3]} = 1$$

$$x_{P_{0.25}} = x_{[3]} = 1$$

$$x_{P_{0.4}} = x_{[4,5]} = 1.5$$

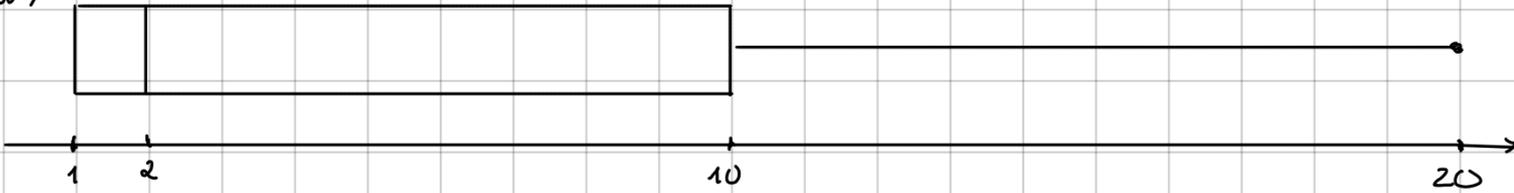
$$x_{P_{0.5}} = x_{[5,6]} = 2$$

$$x_{P_{0.6}} = x_{[6,7]} = 2$$

$$x_{P_{0.75}} = x_{[8]} = 10$$

$$x_{P_{0.99}} = x_{[10]} = 20$$

(iii)



50 Hochschulabsolventen wurden bezüglich ihrer Studiendauer (in Semestern) befragt.

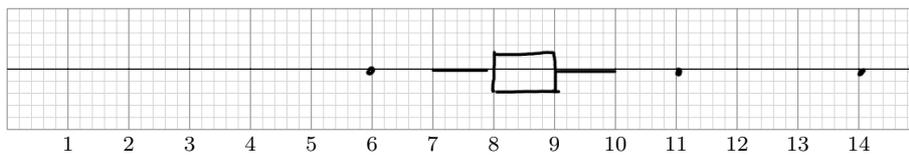
Anzahl Semester	a_i	6	7	8	9	10	11	14
Anzahl Absolventen	h_i	3	9	18	11	6	2	1
	H_i	3	12	30	41	47	49	50
	f_i	0,06	0,18	0,36	0,22	0,12	0,04	0,02
	$F_{50}(a_i)$	0,06	0,24	0,6	0,82	0,94	0,98	1

(i) Beim Aufschreiben und Auswerten sind einige Teile der Tabelle verloren gegangen. Vervollständigen Sie diese.

Hinweis: H_i bezeichnet die kumulative, absolute Häufigkeit bis zur Ausprägung a_i , also $H_i := \sum_{j:a_j \leq a_i} h_j$.

(ii) Bestimmen Sie händisch alle fünf Quartile.

(iii) Zeichnen Sie einen Boxplot der Daten:



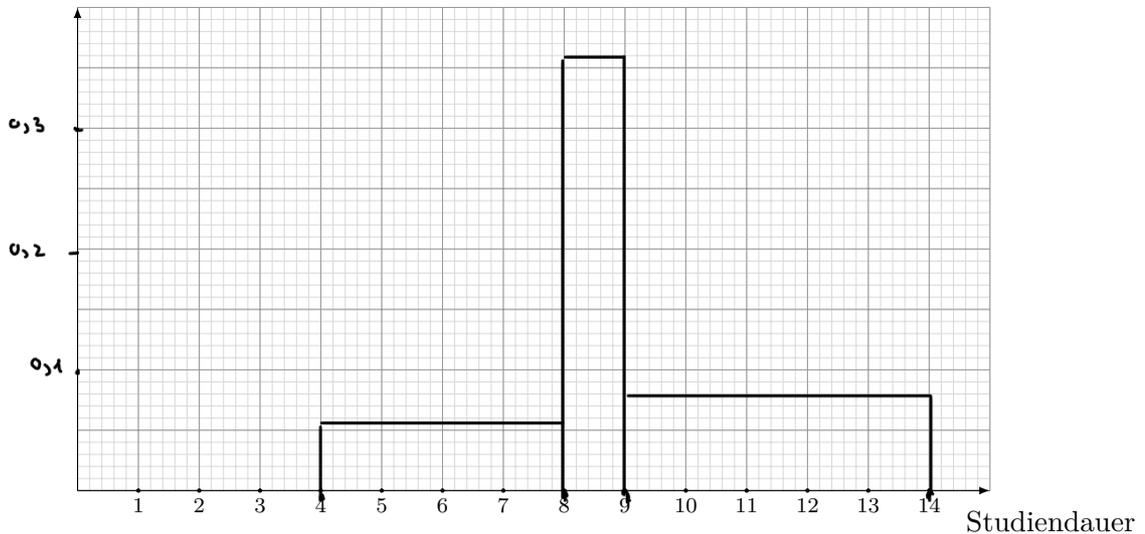
(iv) Geben Sie jeweils *einen* R-Befehl an mit dem Sie

- i.) die obigen Daten einem R-Objekt (Vektor) x zuweisen,
- ii.) alle Quartile ausgeben und
- iii.) das (mittlere) 25%, 50%, 75% und 95%-Quantil gleichzeitig aus dem Vektor x berechnen können.

(v) Die Daten sollen nun klassiert werden in *schnelle Absolventen* (mehr als vier, aber weniger als acht Semester), *normale Absolventen* und *langsame Absolventen* (neun bis maximal 14 Semester). Füllen Sie dazu die Tabelle auf der rechten Seite aus und zeichnen Sie das zugehörige Histogramm. Die Fläche des Histogramms soll auf 1 normiert sein.

Klassen	[4, 8)	[8, 9)	[9, 14]
Anzahl	12	18	20
Breite	4	1	5
Höhe	0,06	0,36	0,08

Häufigkeitsdichte



Bemerkung: Beschriften Sie die Achsen!

