TECHNISCHE UNIVERSITÄT WIEN

Institut f. Stochastik u. Wirtschaftsmathematik

Schriftliche Prüfung

Statistik u. Wahrscheinlichkeitstheorie (f. Inf.)

LVA–Nr.: 107.254

VO: W 2015

Prüfer: W. Gurker

1–12–2016

Nachname	[redacted]		
Vorname	[redacted]		
Kenn	033 [redacted]	Matr	[redacted]

Dieser Teil wird nur vom Prüfer ausgefüllt!

1	4,5	2	4	3	3,5	4	1,5
5	3	6	4	7	–	8	5

Pkt: 25,5 (40)	Note: 3
	Gesamt:

Datum Unterschrift

Betrachten Sie die folgende – bereits geordnete – Stichprobe der Größe $n = 10$:

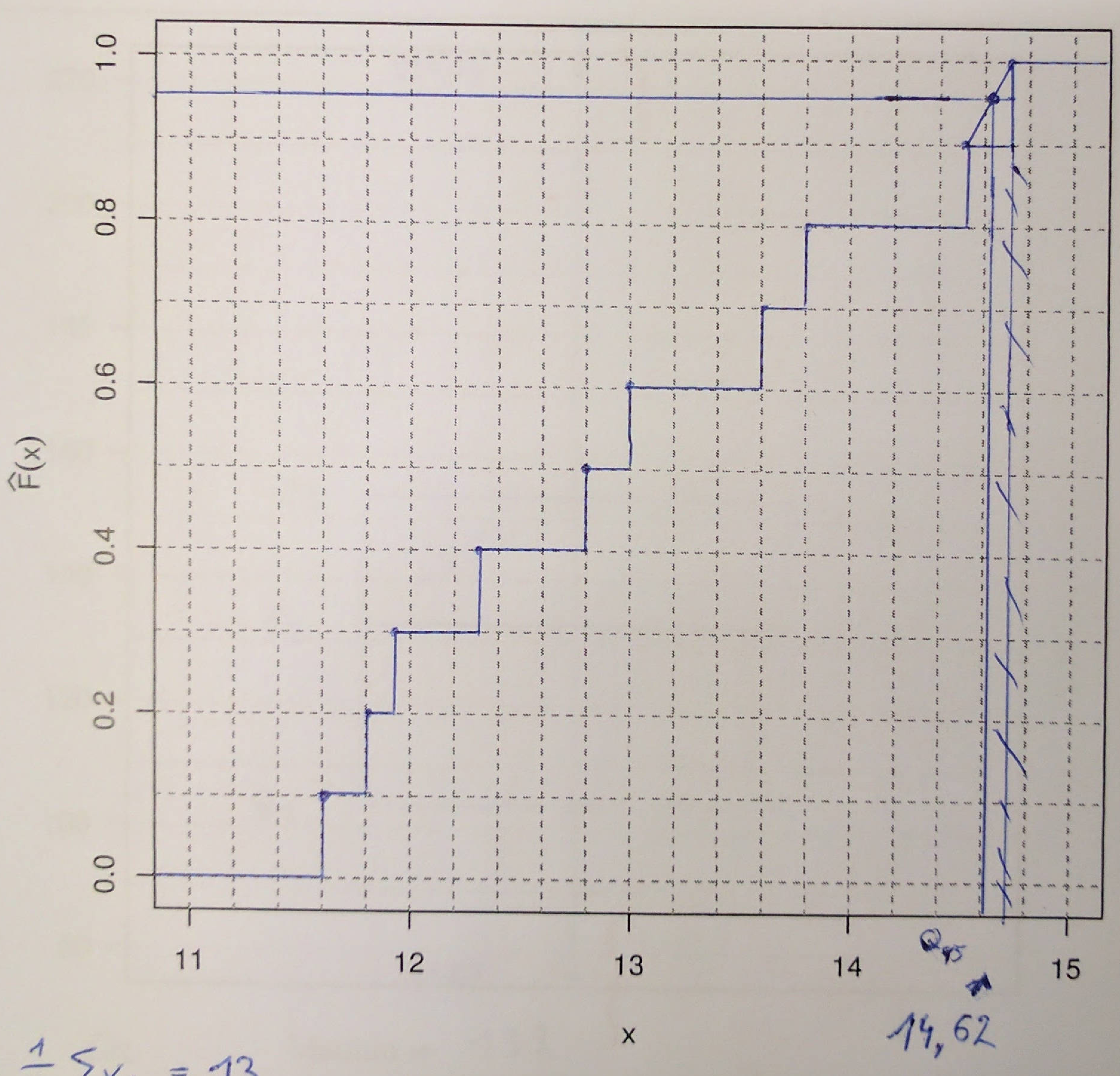
11.6 11.8 11.9 12.3 12.8 13.0 13.6 13.8 14.5 14.7

[2] Zeichnen Sie in die unten stehende Grafik die empirische Verteilungsfunktion.

[1] Bestimmen Sie grafisch das 95%–Quantil vom Typ 4. 14,62

[1] Bestimmen Sie den Stichprobenmittelwert. 13

[1] Bestimmen Sie die Stichprobenvarianz und -streuung.



$\bar{x} = \frac{1}{n} \sum x_i = 13$

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot (1,96 + 1,44 + 1,21 + 0,49 + 0,4 + 0 + 0,36 + 0,64 + 2,25 + 2,89)$

$= \frac{11,64}{9} = 1,29\dot{3}$ 1,25333~

$s = \sqrt{s^2} = 1,137$ 1,12

Aufgabe 2

Betrachten Sie die folgende – bereits geordnete – Stichprobe der Größe $n = 19$:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$x_{(i)}$	87	87	93	99	103	105	119	129	130	132	138	145	145	152	153	160	180	195	211

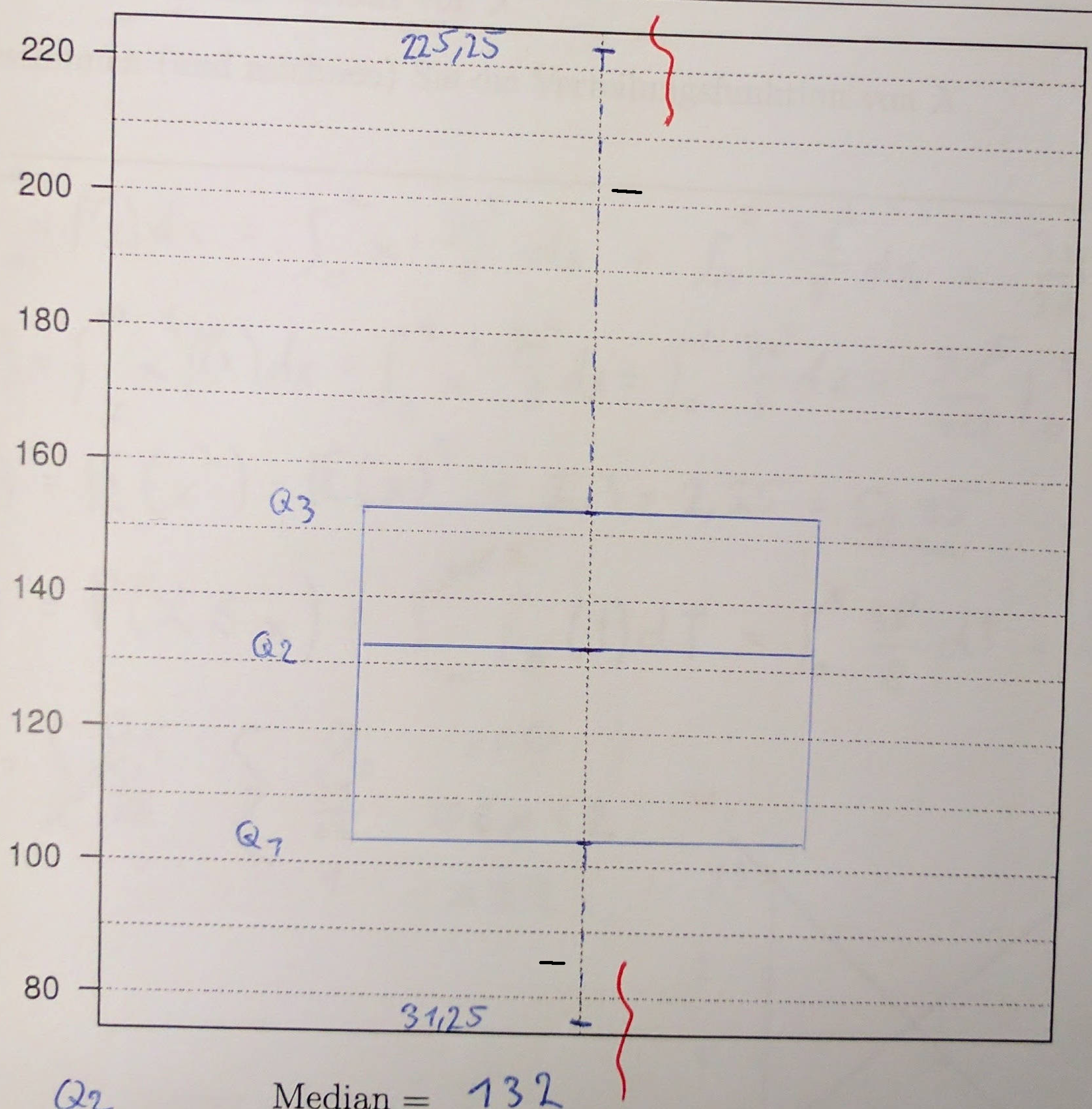
Q_1 Q_2 Q_3

[1] Bestimmen Sie den Median.

[1] Bestimmen Sie die Hinges.

[1] Bestimmen Sie auf Basis der Hinges die Fences.

[2] Zeichen Sie in die unten stehende Grafik den Boxplot der Daten.



Q_2 Median = 132

Q_1 lower Hinge = 104

Q_3 upper Hinge = 152,5

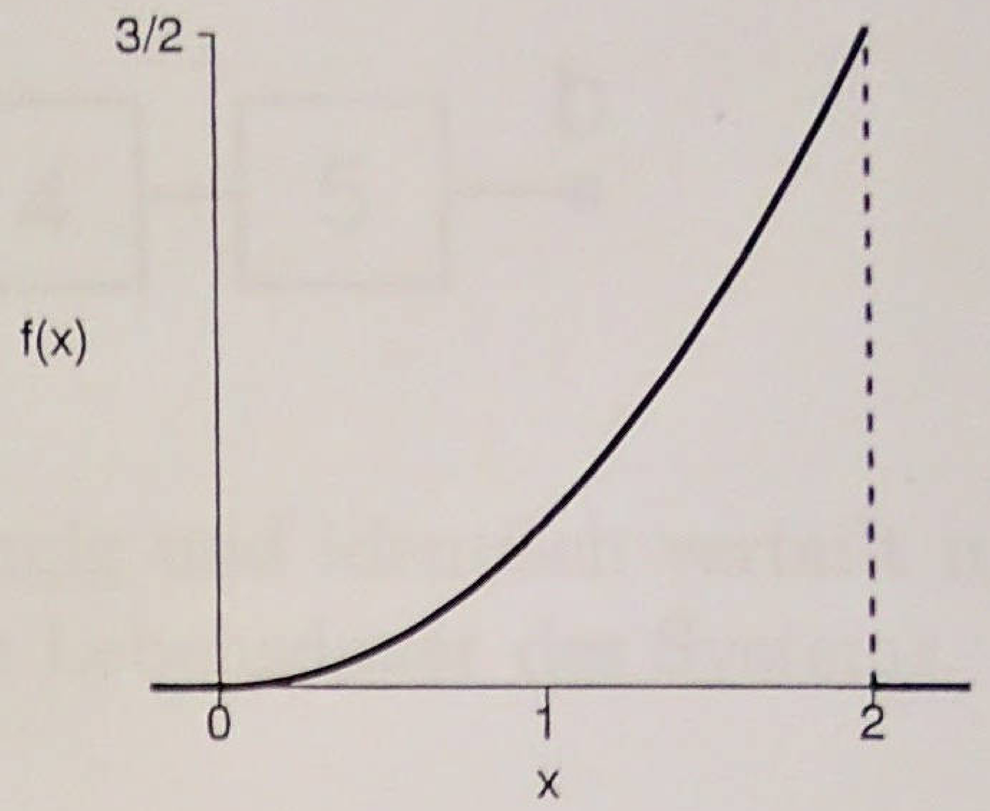
lower Fence = $Q_1 - 1{,}5 \cdot (Q_3 - Q_1) = 31{,}25$

upper Fence = $Q_3 + 1{,}5 \cdot (Q_3 - Q_1) = 225{,}25$

Liegen die Fences außerhalb des Minimums und Maximums, zeichnet man diese bei diesen statt außerhalb.

Die Dichte einer sG X lautet wie folgt:

$$f(x) = \begin{cases} \frac{3x^2}{8} & 0 < x < 2 \\ 0 & \text{sonst} \end{cases}$$



[1] Bestimmen Sie den Erwartungswert von X.

[2] Bestimmen Sie die Varianz von X.

[2] Bestimmen (und zeichnen) Sie die Verteilungsfunktion von X.

---

a) $\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^2 x \cdot \frac{3x^2}{8}\, dx = \int_0^2 \frac{3x^3}{8}\, dx = \frac{3x^4}{32}\Big|_0^2 = \underline{1{,}5}$

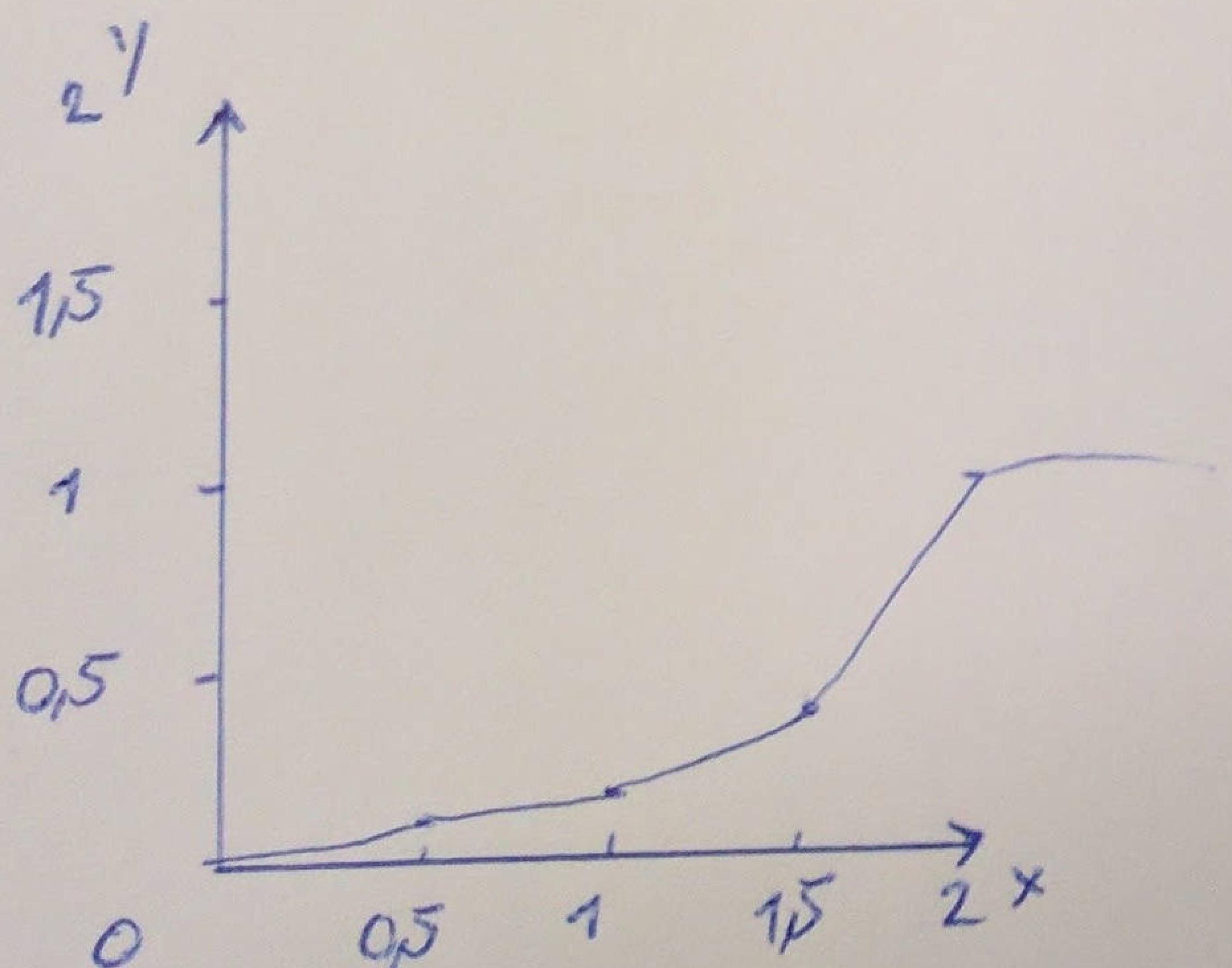
b) $\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^2 x^2 \cdot \frac{3x^2}{8}\, dx = \int_0^2 \frac{3x^4}{8}\, dx = \frac{3x^5}{40}\Big|_0^2 = \underline{2{,}4}$ ✓

$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 2{,}4 - 2{,}25 = \underline{0{,}15}$

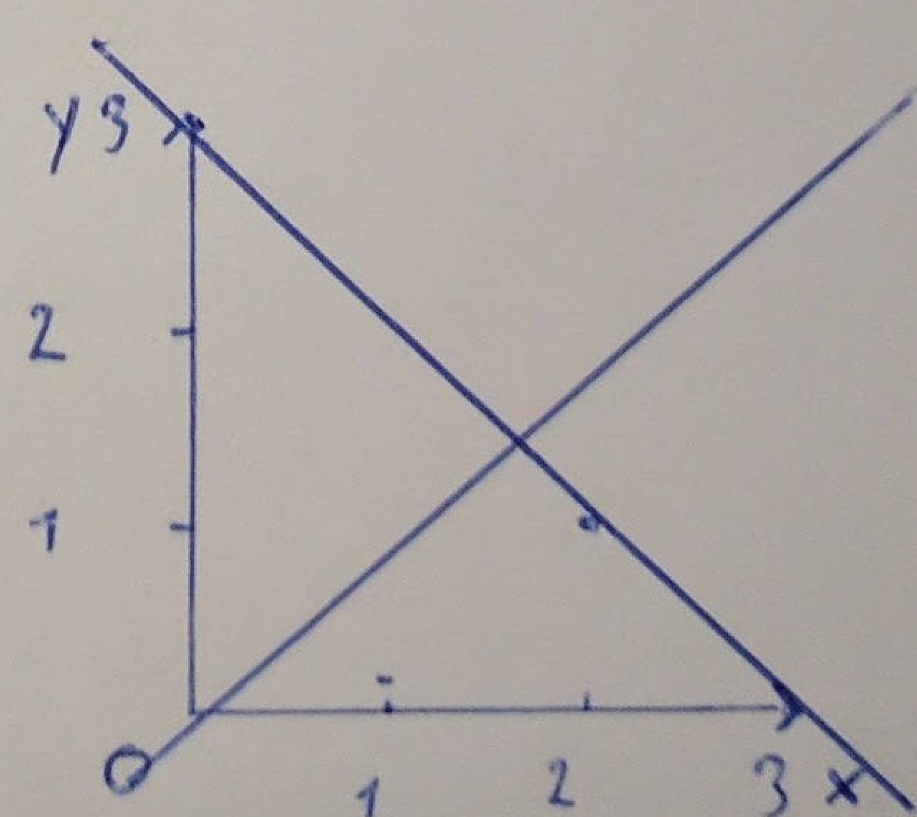
c) $F_X(x) = P(X \leq x) = \int_{-\infty}^{\cancel{\infty}\ x} f_X(t)\, dt = \int_0^{\cancel{2}\ x} \frac{3t^2}{8}\, dt = \frac{3t^3}{24}\Big|_0^2 = 1$ ✓

$\frac{t^3}{8}\Big|_0^x$?

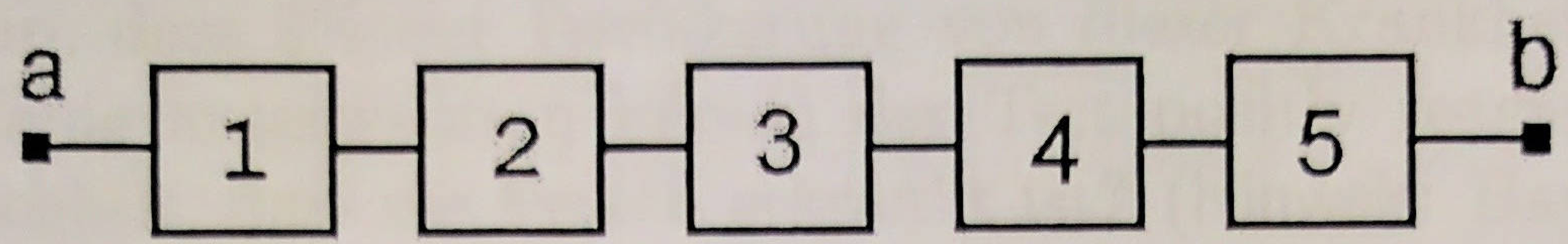
$F_X(x) = \cancel{\begin{cases} 0 \\ \frac{3t^3}{24} \end{cases}} \quad \begin{cases} 0 & x \leq 0 \\ \frac{3x^3}{24} & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$ ✓



schöner!



Die logische Struktur eines Systems sei gegeben wie folgt:



Die Lebensdauern der Komponenten seien unabhängig und identisch verteilt mit Dichte $f(x) = \frac{1}{10} e^{-x/10} I_{(0,\infty)}(x)$ ($\widehat{=}$ Exp($\tau = 10$)). X sei die Lebensdauer des Systems.

[2] Verteilungsfunktion von X ? (Mit Herleitung!)

[1] Dichte von X ? (Um welche Verteilung handelt es sich?)

[2] Erwartungswert, Varianz, Streuung von X ?

---

a) $F_{min}(x) = 1 - e^{-5x/10}$ ← wie bestimmt?

b) $f_{min}(x) = F'_{min}(x) = \frac{5}{10} e^{-5x/10} = \frac{1}{2} e^{-x/2}$ Vert.?

c) $IE(x) = \cancel{1} \frac{1}{2} = 0,5$

$Var(x) = \frac{1}{\lambda^2} = \frac{1}{4} = 0,25$

$s(x) = \sqrt{Var(x)} = IE(x) = 0,5$

$F_{min}(x) = 1 - \Pi_{i=1}^{n}[1 - (1 - e^{-\lambda x})] = 1 - e^{-n\lambda x}$ fuer $x > 0$

Exponentialverteilung

$$IE(x) = \frac{\tau}{n} = \frac{10}{5} = 2 = \tau x$$

$$Var(x) = \tau x^2 = 4$$

$$s(x) = \sqrt{Var(x)} = 2$$

2

[2] Angenommen, ein Test reagiert zu 95% positiv, sollte eine bestimmte Krankheit vorliegen, zeigt aber auch zu 10% ein falsch-positives Resultat. Wenn man davon ausgehen kann, dass 3% der Bevölkerung von dieser Krankheit betroffen ist, und bei einer zufällig ausgewählten Person der Test positiv reagiert, wie groß ist die Wahrscheinlichkeit, dass die Person erkrankt ist? (Hinweis: Bayes'sche Formel)

~~P~~ K... positiver Test
F... ~~false~~ ist falsch-positiv
B... Bevölkerungsanteil mit Krankheit

$P(K) = 0{,}95$
$P(\neg K) = 0{,}05$
$P(K/B) = 0{,}03$
~~P(K)~~ $P(\neg K | \neg B) = 0{,}1$
$P(K|\neg B) = 0{,}97$

$$\frac{P(K) \cdot P(K/B)}{P(K) \cdot P(K/B) + P(\neg K | \neg B) \cdot P(K | \neg B)}$$

$$\frac{0{,}95 \cdot 0{,}03}{0{,}95 \cdot 0{,}03 + 0{,}1 \cdot 0{,}97} = 0{,}227 \approx \underline{23\%}$$

[1] Ergänzung zu Aufgabe 3: Bestimmen Sie allgemein einen Ausdruck für das p–Quantil x_p $(0 < p < 1)$ der Verteilung. Welchen Wert hat speziell der Median?

$x_p = F^{-1}(p)$ fuer $p \in (0,1)$

$$F_x(x) = \quad x = \frac{y^3}{8}$$

$$8x = y^3$$

$$y = \sqrt[3]{8x}$$

$$x_{0,5} = \sqrt[3]{8 * 0,5} = \sqrt[3]{4}$$

1

[2] Die sG X sei nach P(λ) (Poisson) verteilt mit $\lambda = 9$. Berechnen Sie (approximativ) die Wahrscheinlichkeit $P(7 \le X \le 11)$. (Hinweis: Normalapproximation der P(λ)–Verteilung; rechnen Sie mit Stetigkeitskorrektur.)

$$P(7 \le X \le 11) = \Phi\left(\frac{b + 0{,}5 - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - 0{,}5 - \lambda}{\sqrt{\lambda}}\right)$$

$$= \Phi\left(\frac{11{,}5 - 9}{\sqrt{9}}\right) - \Phi\left(\frac{6{,}5 - 9}{\sqrt{9}}\right)$$

$$= \Phi(0{,}83) - \Phi(-0{,}83)$$

$$= 0{,}7967 - 0{,}7967 = 0$$?

- 0,2033 = 0,5934

Eine Stichprobe der Größe $n = 50$ von $X \sim \mathsf{P}(\lambda)$ (Poisson) war wie folgt:

x	0	1	2	3	4	5
Häufigkeit	9	17	15	5	3	1

2 [2] Bestimmen Sie allgemein für n Beobachtungen den ML–Schätzer von λ. (Mit genauer Herleitung!)

– [1] ML–Schätzwert von λ auf Basis der gegebenen Stichprobe?

b) $\hat{\lambda} = \frac{1}{n}\sum x_i = \cancel{8\ \frac{9+17+15+5+3+1}{6} = 8,\dot{3}}$

$$\frac{0*9+1*17+2*15+3*5+4*3+5*1}{50} = 1,58$$

a) $\cancel{L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)}$

$$L(\lambda) = \prod_{i=1}^{n} p(x_i;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}\cdot e^{-\lambda}}{x!}$$

$$\ln L(\lambda) = \sum_{i=1}^{n} \ln \frac{\lambda^{x_i}\cdot e^{-\lambda}}{x!} = \sum_{i=1}^{n} x_i \cdot \ln x - \lambda - \ln x!$$

~~Nullsetzen~~ Ableiten und Nullsetzen:

$$\frac{d \ln L(\lambda)}{d\lambda} = \sum \frac{x_i}{\lambda} - 1 = \frac{1}{\lambda}\left(\sum x_i\right) - n$$

$$0 = \frac{1}{\lambda}\left(\sum x_i\right) - n$$

$$\hat{\lambda} = \frac{1}{n}\sum x_i$$

2 [2] Bei einer Befragung von 200 wahlberechtigten Personen sagen 104, dass sie vorhaben, für einen bestimmten Kandidaten zu stimmen. Bestimmen Sie ein (approximatives) 90%–Konfidenzintervall für den Wähleranteil dieses Kandidaten.

$$\hat{p} = \frac{104}{200} = 0,52$$

$$KI = \hat{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$$

$$= 0,52 \pm 1,6449 \cdot 0,0353$$

$$KI = [0,462\ ; 0,578]$$

$\alpha = 0,1 \rightarrow \frac{\alpha}{2} = 0,05$

$z_{0,95} = 1,6449$

$$\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} = \sqrt{\frac{0,52\cdot 0,48}{200}} = \sqrt{\frac{0,2496}{200}} = 0,0353$$

Für eine Stichprobe x der Größe $m = 8$ von $X \sim \mathsf{N}(\mu_X, \sigma_X^2)$ ergab sich:

```
> data.frame(m=length(x), mean=mean(x), var=var(x), sd=sd(x))
  m  mean   var    sd
  8 11.79 7.687 2.773
```

[2] Testen Sie zum Niveau 5%: $\mathcal{H}_0 : \mu_X = 10$ gegen $\mathcal{H}_1 : \mu_X > 10$

$$T_0 = \frac{\bar{x}_n - \mu_0}{\frac{s_n}{\sqrt{n}}} = \frac{11,79 - 10}{\frac{2,773}{\sqrt{8}}} = 1,826$$

H_0 verwerfen falls: $T_0 > t_{n-1;1-\alpha}$

$$t_{7;0,95} = 1,895$$

$1,826 < 1,895 \rightarrow H_0$ wird nicht verworfen.

Für eine von der obigen Stichprobe unabhängige zweite Stichprobe y der Größe $n = 10$ von $Y \sim \mathsf{N}(\mu_Y, \sigma_Y^2)$ ergab sich:

```
> data.frame(n=length(y), mean=mean(y), var=var(y), sd=sd(y))
   n  mean   var    sd
  10 15.14 10.19 3.193
```

Wenn man davon ausgehen kann, dass $\sigma_Y^2 = \sigma_X^2$:

[1] Bestimmen Sie den gepoolten Varianzschätzer s_p^2?

[2] Bestimmen Sie ein 95% Konfidenzintervall für $\mu_Y - \mu_X$.

$$s_p^2 = \frac{(m-1)S^2x + (n-1)S^2y}{m-n-2} = \frac{(8-1) * 7,687 + (10-1) * 10,19}{8+10-2} = 9,094$$

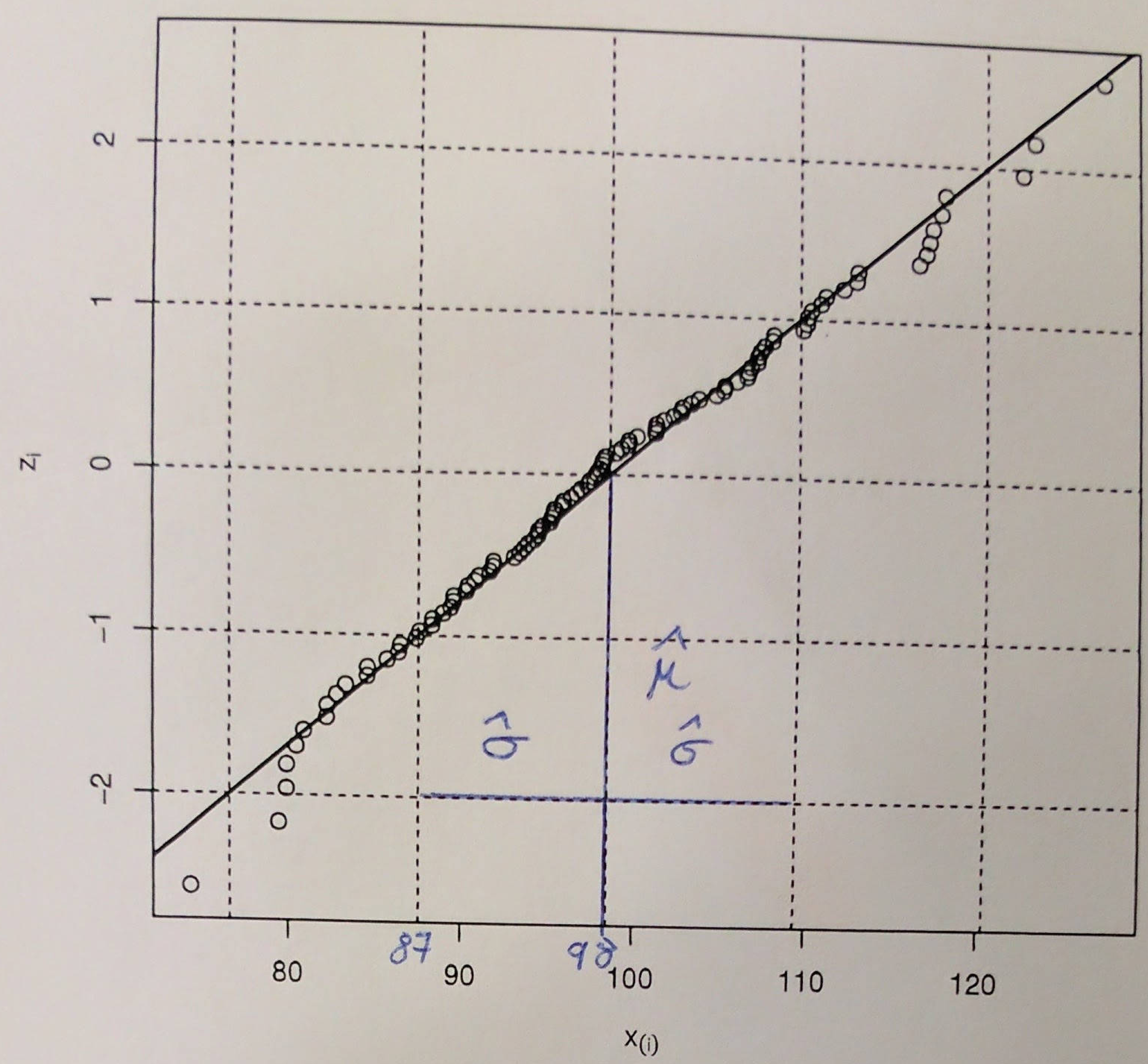
$$KI : \bar{x} - \bar{y} \pm t_{m+n-2;1-\frac{\alpha}{2}} * s_p\sqrt{\frac{1}{m} + \frac{1}{n}}$$

$$11,79 - 15,14 \pm t_{16;0,975} * \sqrt{9,094} * \sqrt{\frac{1}{8} + \frac{1}{10}}$$

$$KI : (-6,387; -0,313)$$

2

[2] Ein Datensatz der Größe $n = 100$ ergibt im Normal-QQ-Plot das folgende Bild:



Stammen die Daten aus einer Normalverteilung? ☒ ja ☐ nein ✓

Wie groß sind (etwa) Mittelwert $\widehat{\mu} \approx 98$ und Streuung $\widehat{\sigma} \approx 11$?

3

[3] Stammen die folgenden Beobachtungen aus einer stetigen uniformen Verteilung auf dem Intervall $(0, 1)$? (Testen Sie mit $\alpha = 10\%$.)

Klasse	(0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1]
Häufigkeit	13	22	14	23	28

Klasse	X_i	p_{i0}	np_{i0}	$(X_i - np_{i0})^2/np_{i0}$
(0;0,2]	13	1/5	20	2,45
(0,2;0,4]	22	1/5	20	0,2
(0,4;0,6]	14	1/5	20	1,8
(0,6;0,8]	23	1/5	20	0,45
(0,8;1]	28	1/5	20	3,2
	100	1	100	8,1

Q

H_0 wird verworfen, falls

$Q_{k-1} > \chi^2_{k-1;1-\alpha}$

$Q_4 > \chi^2_{4;0,9}$

$8,1 > 7,779$

H_0 wird verworfen

✓