Serie 01: Beschreibende Statistik

Aufgabe 1

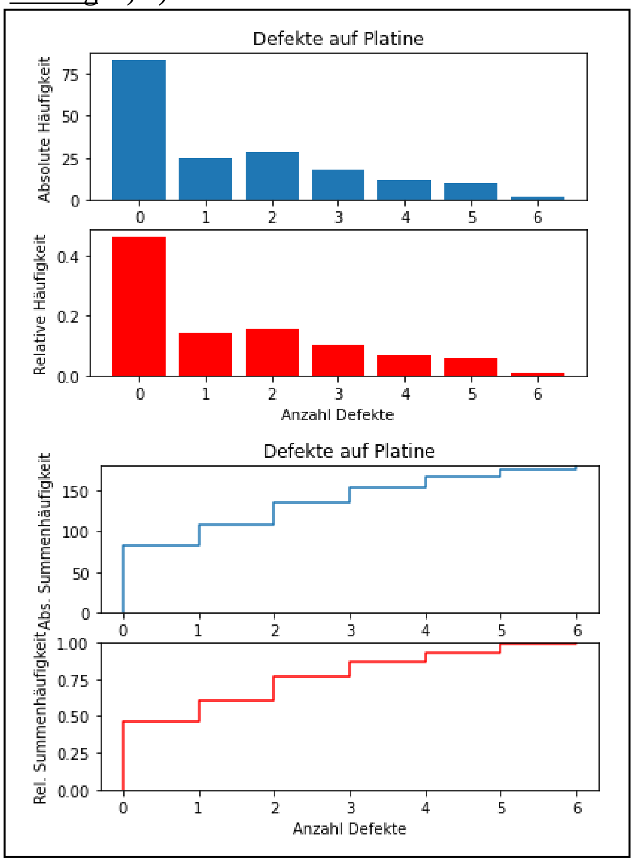
Gegeben ist die folgende Auswertung einer Qualitätskontrolle zur Anzahl der Defekte auf Platinen in einer Leiterplattenproduktion.

a_i	0	1	2	3	4	5	6
h_i	83	25	28	18	12	10	2

a) Zeichne die absolute bzw. relative Häufigkeit (PMF)

b) Zeichne die absolute bzw. relative Verteilungsfunktion (CDF)

Lösung: a) b)




```
import numpy as np
import matplotlib.pyplot as plt

## Aufgabe 1.1: Defekte auf Platine
# Defekte
X=np.arange(7.)
# absolute Häufigkeiten
hi=np.array([83., 25., 28., 18., 12., 10., 2.])
# relative Haeufigkeiten PDF
fi=hi/sum(hi)
# absolute Summenhäufigkeit
Hi=hi.cumsum(axis=0)
# kumulative Verteilung CDF
Fi=fi.cumsum(axis=0)

#Plot Häufigkeiten
plt.figure(1)
plt.subplot(211)
fig11 = plt.bar(X,hi)
plt.title('Defekte auf Platine')
plt.xlabel('Anzahl Defekte')
plt.ylabel('Absolute Häufigkeit')
plt.subplot(212)
fig12=plt.bar(X,fi,color='r')
plt.xlabel('Anzahl Defekte')
plt.ylabel('Relative Häufigkeit')

#Plot Summenfunktionen
plt.figure(2)
fig21=plt.subplot(211)
plt.step(X,Hi, where='post')
plt.title('Defekte auf Platine')
plt.xlabel('Anzahl Defekte')
plt.ylabel('Abs. Summenhäufigkeit')
fig22=plt.subplot(212)
plt.step(X,Fi, where='post',color='r')
plt.xlabel('Anzahl Defekte')
plt.ylabel('Rel. Summenhäufigkeit')
```

Aufgabe 2

Auf die Frage nach der Anzahl Geschwister geben 25 zufällig ausgewählte Studierende folgende Antworten

1 0 1 1 2 1 0 0 2 1 1 1 3 0 2 2 1 1 0 2 3 1 1 2 1

a) Fülle folgende Tabelle aus:

Anzahl Geschwister a_i	0	1	2	3
absolute Häufigkeit h_i				
relative Häufigkeit f_i				
kumulative absolute Häufigkeit H_i				
kumulative relative Häufigkeit F_i				

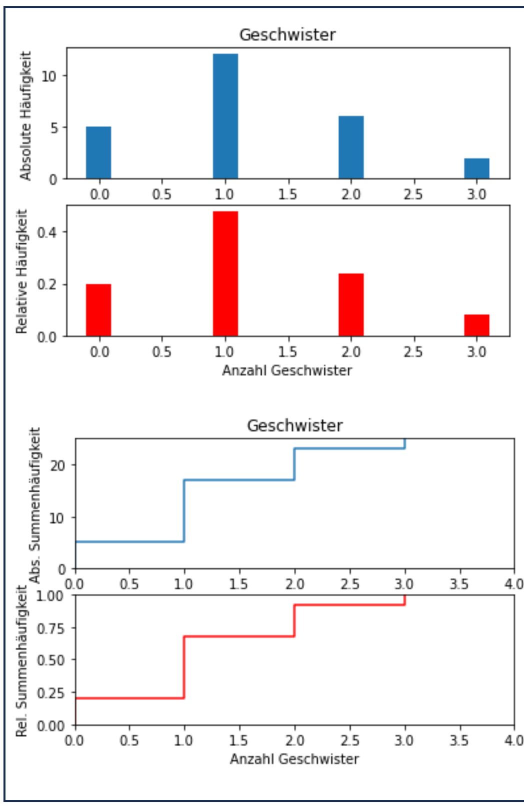
b) Zeichne die relative Häufigkeitsfunktion (PMF) und die kumulative Verteilungsfunktion (CDF).

<u>Lösung</u>:

a) Fülle folgende Tabelle aus:

Anzahl Geschwister a_i	0	1	2	3
absolute Häufigkeit h_i	5	12	6	2
relative Häufigkeit f_i	0.20	0.48	0.24	0.08
kumulative absolute Häufigkeit H_i	5	17	23	25
kumulative relative Häufigkeit F_i	0.20	0.68	0.92	1.00

b)

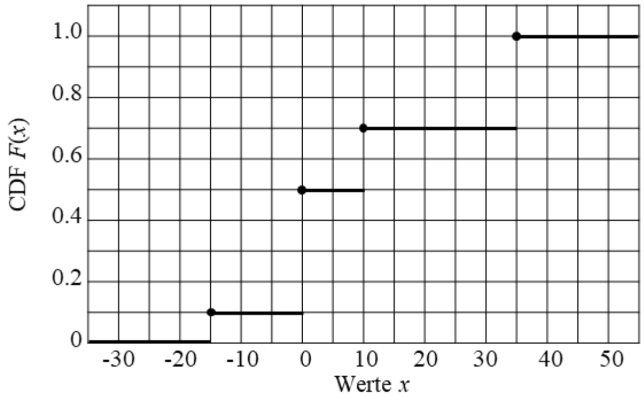



```
import numpy as np
import
matplotlib.pyplot as
plt

## Aufgabe 1.2:
Geschwister

#Daten
d=np.array([1,0,1,1,2,1
,0,0,2,1,1,1,3,0,2,2,1,
1,0,2,3,1,1,2,1])
#Geschwister
X, hi = np.unique(d,
return_counts=True)
print('X=',X)
# absolute Häufigkeiten
print('hi=',hi)
# relative
Haeufigkeiten PDF
fi=hi/sum(hi)
print('fi=',fi)
# absolute
Summenhäufigkeit
Hi=hi.cumsum(axis=0)
print('Hi=',Hi)
# kumulative Verteilung
CDF
Fi=fi.cumsum(axis=0)
print('Fi=',Fi)
```

```
#Plot Häufigkeiten
plt.figure(1)
plt.subplot(211)
fig11 = plt.bar(X,hi,  width=0.2)
plt.title('Geschwister')
plt.xlabel('Anzahl Geschwister')
plt.ylabel('Absolute Häufigkeit')
plt.subplot(212)
fig12=plt.bar(X,fi,color='r',
width=0.2)
plt.xlabel('Anzahl Geschwister')
plt.ylabel('Relative Häufigkeit')

#Plot Summenfunktionen
plt.figure(2)
fig21=plt.subplot(211)
plt.step(np.append([0],X),np.append
([0],Hi),where='post')
plt.title('Geschwister')
plt.xlabel('Anzahl Geschwister')
plt.ylabel('Abs. Summenhäufigkeit')
plt.xlim(0,4)
plt.ylim(0,25)
fig22=plt.subplot(212)
plt.step(np.append([0],X),np.append
([0],Fi),where='post',color='r')
plt.xlabel('Anzahl Geschwister')
plt.ylabel('Rel. Summenhäufigkeit')
plt.xlim(0,4)
plt.ylim(0,1)
```

Aufgabe 3

Ein nicht in Klassen eingeteilter Datensatz hat folgende kumulative Verteilungsfunktion (CDF):



a) Welche Werte kommen unter den Daten vor?

b) Welche relativen Häufigkeiten haben die vorkommenden Datenwerte?

c) Angenommen der Datensatz besteht aus insgesamt 40 Daten. Wie viele Daten haben dann genau den Wert 10? Wie viele höchstens den Wert 10? Und wie viele mindestens den Wert 10?

<u>Lösung:</u>

a) Werte: -15, 0, 10, 35

b) Relative Häufigkeiten: 0.1, 0.4, 0.2, 0.3

c) Anzahl Daten mit genau Wert 10: $0.2 \cdot 40 = 8$

Anzahl Daten mit höchstens Wert 10: $0.7 \cdot 40 = 28$

Anzahl Daten mit mindestens Wert 10: $0.5 \cdot 40 = 20$

Klassierte Daten

Aufgabe 4

Auf die Frage «Wie viele Minuten brauchst du von zu Hause bis an die ZHAW?» geben 18 zufällig ausgewählte Studierende folgende Antworten:

30 20 10 45 15 60 20 20 30 75 10 35 20 50 10 90 45 20

Wir teilen die Wegzeiten in folgende drei Klassen ein:

0 bis unter 15 Minuten, 15 bis unter 45 Minuten, 45 bis und mit 90 Minuten

a) Fülle nachfolgende Tabelle aus. Gib die nicht ganzzahligen Ergebnisse als Brüche und als auf 0.001 gerundete Dezimalzahlen an.

Wegzeiten in Minuten	[0, 15 [	[15, 45 [	[45, 90]
absolute Häufigkeit			
relative Häufigkeit			
relative Häufigkeitsdichte			
kumulative relative Häufigkeit			

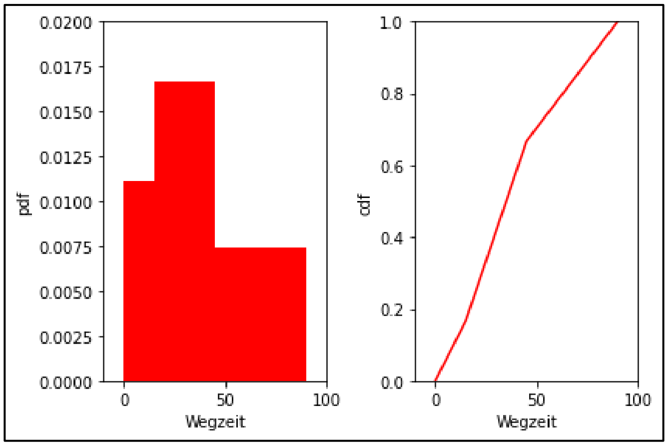
b) Zeichne die relative Häufigkeitsdichtefunktion (PDF) und die kumulative Verteilungsfunktion (CDF).

Lösung:

a) Fülle nachfolgende Tabelle aus.

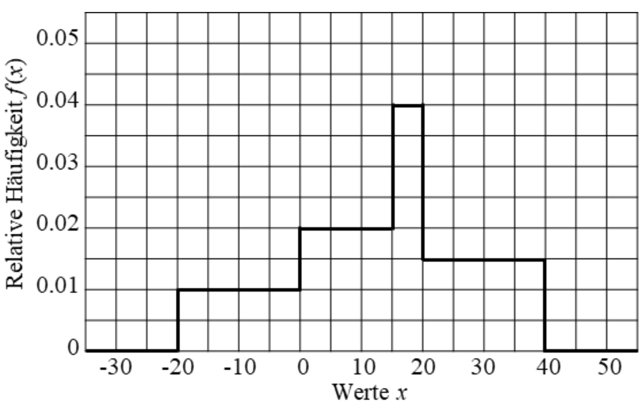
Wegzeiten in Minuten	[0, 15 [	[15, 45 [	[45, 90]
absolute Häufigkeit	3	9	6
relative Häufigkeit	1/6 ≈ 0.167	1/2 ≈ 0.500	1/3 ≈ 0.333
relative Häufigkeitsdichte	1/90 ≈ 0.011	1/60 ≈ 0.017	1/135 ≈ 0.007
kumulative relative Häufigkeit	1/6 ≈ 0.167	2/3 ≈ 0.667	1.000

b)



Aufgabe 5

Ein in vier Klassen eingeteilter Datensatz besitzt folgende relative Häufigkeitsdichtefunktion (PDF) $f(x)$.



a) Wie lauten die vier Klassen? Und welche relativen Häufigkeiten besitzen die Klassen?

b) Zeichne die zugehörige kumulative Verteilungsfunktion (CDF) $F(x)$.

c) Berechne mit linearer Interpolation $F(12)$ und $F(35)$.

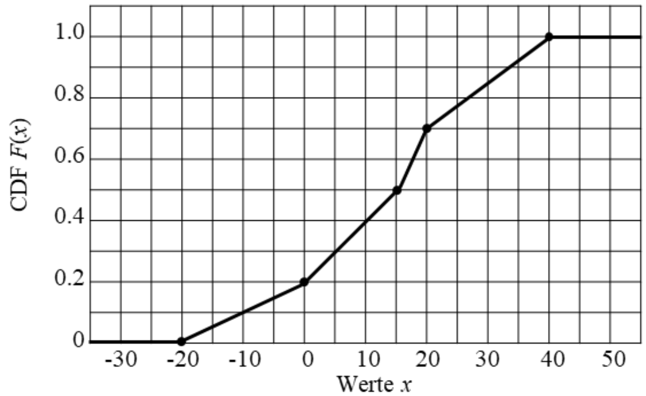
Lösung:

a) Klassen und relative Häufigkeiten der Klassen

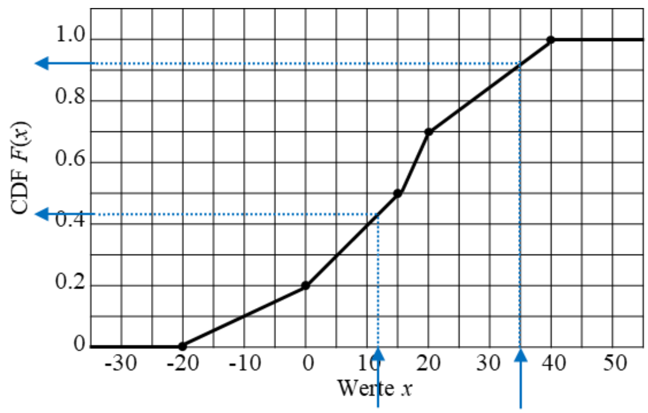
[-20,0[: 0.01· 20 = 0.2 [0,15[: 0.02 · 15 = 0.3

[15,20[: 0.04 · 5 = 0.2 [20,40[: 0.015 · 20 = 0.3

b)



c)



$$F(12) = 0.2 + \frac{0.5 - 0.2}{15 - 0} \cdot (12 - 0) = 0.44$$

$$F(35) = 0.7 + \frac{1 - 0.7}{40 - 20} \cdot (35 - 20) = 0.925$$

Aufgabe 6

Bei einer Firma werden in einem Monat 400 Lebensversicherungsverträge abgeschlossen. Nachstehend ist die klassifizierte Häufigkeitsverteilung für die Versicherungssummen angegeben.

Vers.sum. von .. bis unter …(in kCHF)	[4,10[	[10,20[	[20,30[	[30,40[	[40,80[	[80,120[
Anzahl Verträge	20	160	80	40	88	12

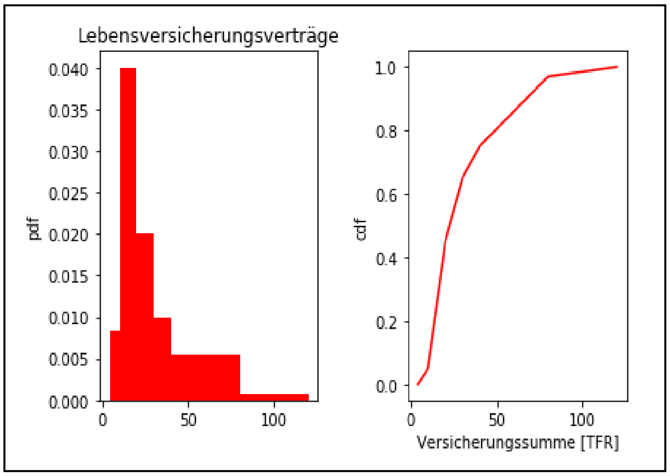
a) Zeichne das Histogramm der relativen Häufigkeiten (PDF).

b) Zeichne die kumulative Verteilungsfunktion für klassierte Daten (CDF).

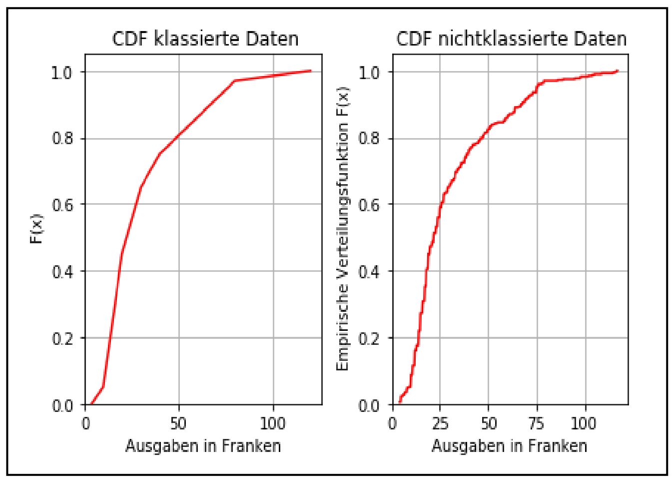
c) Simuliere die Stichprobe (rand… Befehl) und vergleiche die CDF der klassierten Daten mit der simulierten empirischen CDF der Stichprobe (optional).

Lösung: a) b) c)

```
import numpy as np
import matplotlib.pyplot as plt

## Aufgabe 1.6: Klassierte Daten
#Klassengrenzen
KlGr=np.array([4, 10, 20, 30, 40, 80, 120])
#absolute Klassenhaeufigkeiten
hi=np.array([20, 160, 80, 40, 88, 12])
n=np.sum(hi)
#Saeulenhoehen im Histogram der PDF
Kldiff=np.diff(KlGr)
Sumhi=np.sum(hi)
Saeulenhoehe=hi/np.diff(KlGr)/np.sum(hi)
Ser=np.append([Saeulenhoehe],[0])

#figure
plt.figure(1)
plt.subplot(121)
#Klassenmitten generieren
KlMitten=np.empty(KlGr.size-1)
for k in range (0,KlGr.size-1):
     KlMitten[k]=(KlGr[k+1]+KlGr[k])/2
plt.bar(KlMitten,Saeulenhoehe,width=Kldiff,color='r
ed')
plt.ylabel('pdf')
plt.title('Lebensversicherungsverträge')
plt.subplot(122)
plt.plot(KlGr,np.append([0],[np.cumsum(hi)/Sumhi]),
color='red')
plt.xlabel('Versicherungssumme [TFR]')
plt.ylabel('cdf')
plt.tight_layout()
```




```
#Stichprobe generieren für die klassierten Daten
#Klassierte Daten
St_s=np.empty(shape=[0, n])
for k in range (0,KlGr.size-1):
        St_s=np.concatenate((St_s,np.random.rand
int
        (KlGr[k],KlGr[k+1],size=hi[k])),axis=Non
e)

#Histogramm erstellen für die simulierten Daten
h_s,K_s=np.histogram(St_s,KlGr) #abs.
Klassenhäufigk.
Flaechen=h_s/n #Saeulenflächen im Histogramm
CDF=np.append([0],np.cumsum(Flaechen)) #Werte
der CDF

#Simulierte Daten
h_s_m,K_m = np.histogram (St_s, bins=n,
density=True)
CDF_m = np.cumsum (h_s_m)

plt.figure(2)
plt.subplot(121)
plt.plot(K_s,CDF,color='red') #
Verteilungsfunkt. CDF
plt.xlabel('Ausgaben in Franken')
plt.ylabel(' F(x)')
plt.grid()
plt.title(' CDF klassierte Daten')
plt.xlim(left=0)
plt.ylim(ymin=0)

plt.subplot(122)
plt.plot (K_m[1:], CDF_m/CDF_m[-1],color='red')
# Empirische Verteilungsfunktion CDF
plt.xlabel('Ausgaben in Franken')
```



Kenngrössen

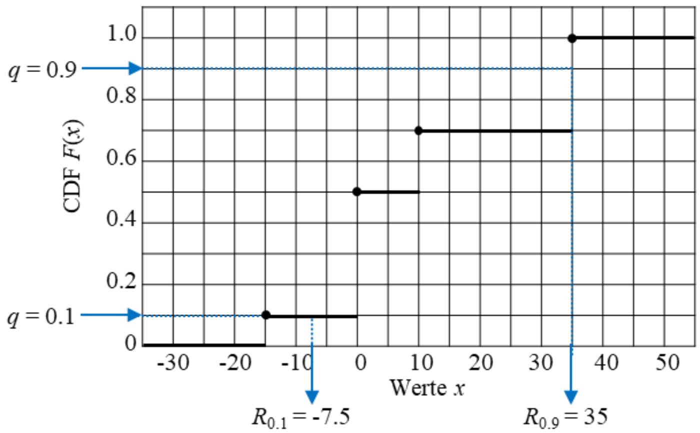
Aufgabe 7

a) Bestimme für Aufgabe 3 aus der gegebenen CDF graphisch das 0.1-Quantil und das 0.9-Quantil.

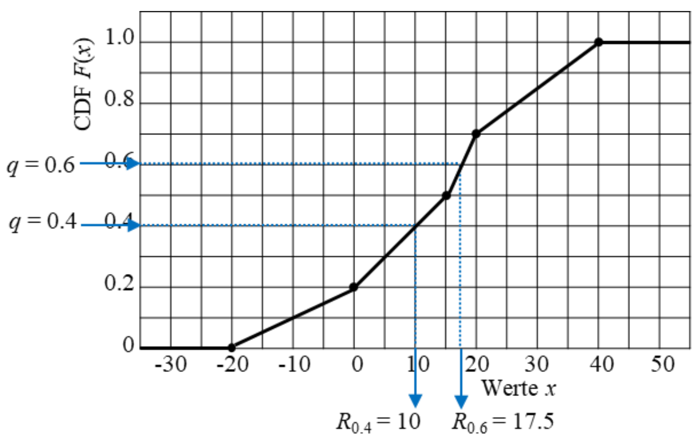
b) Bestimme für Aufgabe 5 aus der selbst erstellten CDF graphisch das 0.4-Quantil und das 0.6-Quantil.

Lösung:

a)



b)



Aufgabe 8

Zwei Weitspringer führen Statistik über die im Training erzielten Leistungen:

	[700,720[	[720,740[	[740,760[	[760,780[	[780,800[	[800,820[
A	19	24	26	27	10	5
B	4	8	52	40	32	24

Bei einem Wettkampf springt A 7.3m und B 7.5m weit.

a) Welcher Weitspringer ist bezogen auf seine Trainingsleistung am weitesten ‚unter Form'

b) Überlege, welches Mass geeignet ist, für diesen Vergleich.

Lösung:

Das Mass für die beiden Springer, um den Wettkampfsprung mit den Trainingssprängen zu vergleichen, sind die Quantile, d.h. es wird untersucht in welchem Bereich der Wettkampfsprung bezüglich der Trainingssprünge liegt.

a) Springer A: 7.3 m ist im Bereich 720-740. Wir berechnen wie viele Trainingssprünge darunter lagen, d.h. wir bestimmen $F_A(730)$ mit der Verteilungsfunktion $F_A(x)$ der klassierten Daten von Springer A.

$$\frac{F_A(740)-F_A(720)}{20}=\frac{q_a-F_A(720)}{730-720} \Leftrightarrow q_a=\frac{19}{111}+\frac{10\cdot 24/111}{20}=\frac{31}{111}\approx 0.2793$$

Genauso bestimmen wir $F_B(750)$ mit der Verteilungsfunktion $F_B(x)$ der klassierten Daten von Springer B.

$$\frac{F_B(760)-F_B(740)}{20}=\frac{q_B-F_B(740)}{750-740} \Leftrightarrow q_B=\frac{12}{160}+\frac{10\cdot 52/160}{20}=\frac{38}{160}=\frac{19}{80}=0.2375$$

Das heisst bei Springer B sind im Verhältnis weniger Trainingssprünge unter seinem Wettkampsprung. B ist damit im Vergleich schlechter als A.

b) Quantile sind das geeignete Mass. Bei A ist 730 das 0.2793-Quantil und bei B ist 750 das 0.2375-Quantil.

```
## Aufgabe 1.8
#Wir vergleichen die CDF WErte der beiden Springer für die beiden Weiten
RKLGr=np.array([720, 740, 760, 780, 800, 820])
print('Rechte Klassengrenzen = ',RKLGr)  #Rechte Klassengrenzenn
ha= np.array([19, 24, 26, 27, 10, 5])
print('abs. Häufigkeit Springer A = ',ha) #absolute Häufigkeiten Springer A
hb= np.array([4, 8, 52, 40, 32, 24])
print('abs. Häufigkeiten Springer B = ',hb) # absolute Häufigkeiten Springer B
CDFA= np.cumsum(ha)/np.sum(ha)
print('CDF Werte Springer A = ', CDFA) # CDF WErte von A
CDFB=np.cumsum(hb)/np.sum(hb)
print('CDF Werte Springer B = ', CDFB)# CDF WErte von B
#Bewertung der Ergebnisse durch Interpolation
WeiteA=730
WeiteB=750
#Obere und untere Klassen festlegen und dann interpolieren für A
Ua=np.where(RKLGr<WeiteA)[0]
Oa=np.where(RKLGr>=WeiteA)[0]
qa=(WeiteA-RKLGr[Ua[-1]])/(RKLGr[Oa[0]]-RKLGr[Ua[-1]])*(CDFA[Oa[0]]-CDFA[Ua[-1]])+CDFA[Ua[-1]]
print('qa = ',qa)
#Obere und untere Klassen festlegen und dann interpolieren für B
Ub=np.where(RKLGr<WeiteB)[0]
Ob=np.where(RKLGr>=WeiteB)[0]
print(CDFB[Ub[-1]])
qb=(WeiteB-RKLGr[Ub[-1]])/(RKLGr[Ob[0]]-RKLGr[Ub[-1]])*(CDFB[Ob[0]]-CDFB[Ub[-1]])+CDFB[Ub[-1]]
print('qb = ',qb)
```

Aufgabe 9

a) Bestimme für Aufgabe 1 den Mittelwert, den Modus und den Median

b) Bestimme für Aufgabe 6 durch Interpolation an der Summenkurve:
   - Wie viel % der Versicherten sind mit höchstens 18'000.- versichert?
   - Mit welchem Mindestbetrag sind die 20% der Personen versichert, die am höchsten versichert sind?

c) Berechne für Aufgabe 6 näherungsweise den Median und den Mittelwert.

Lösung:

a) Bei Aufgabe 1: $x_{med} = 1$, $x_{mod} = 0$, $\bar{x} \approx 1.38$

b) Bei Aufgabe 6:

Mit der Verteilungsfunktion $F(x)$ der klassierten Daten suchen wir $F(18)$:

$$\frac{F(20)-F(10)}{10} = \frac{q-F(10)}{18-10} \Leftrightarrow q = \frac{20}{400} + \frac{8\cdot 160/400}{10} = \frac{20+128}{400} = \frac{74}{200} = 0.37$$

37% sind mit höchstens 18.000 versichert.

Mit der Verteilungsfunktion $F(x)$ suchen wir x mit $F(x) = 0.80$, d.h. das 0.8-Quantil:

$0.8 \cdot 400 = 320$:

$$\frac{F(80)-F(40)}{40} = \frac{0.8-F(40)}{x-40}$$

$$\Leftrightarrow x = 40 + \frac{40\cdot(0.8-F(40))}{F(80)-F(40)} = 40 + \frac{40\cdot(0.8-300/400)}{88/400} = \frac{540}{11} \approx 49.09$$

Mindestbetrag ist demnach 49090.

c) Median:

$$\frac{F(30)-F(20)}{10} = \frac{0.5-F(20)}{x-20} \Leftrightarrow x_{med} = 20 + \frac{10\cdot(0.5-180/400)}{80/400} = 22.5$$

Mittelwert bei klassierten Daten (aus dem Mittelwert der Klassen):

4-9.999..:	7	
10-19.999…:	15	
20-29.999…:	25	etc.

$$\bar{x} = \frac{1}{400}[7\cdot 20 + 15\cdot 160 + 25\cdot 80 + 35\cdot 40 + 60\cdot 88 + 100\cdot 12] \approx 31.05$$

Aufgabe 10

Bei der Ermittlung der wöchentlichen Fahrzeiten verschiedener Personen im ÖV ergaben sich die folgenden Werte (in h):

2.1 2.4 2.8 3.1 4.2 4.9 5.1 6.0 6.4 7.3
10.8 12.5 13.0 13.7 14.8 17.6 19.6 23.0 25.0 35.2 39.6

a) Bestimme den Median.

b) Berechne den Mittelwert und die Standardabweichung, sowie die korrigierte Standardabweichung.

Lösung: $x_{med} = 10.8$, $\bar{x} \approx 12.8$, $\sigma \approx 10.42$, $k\sigma \approx 10.68$.

```
## Aufgabe 1.10
X=[ 2.1  2.4  2.8  3.1  4.2  4.9  5.1  6.0  6.4  7.3 ...
    10.8  12.5  13.0  13.7  14.8  17.6  19.6  23.0  25.0  35.2  39.6];
Med=median(X) # Median 10.8
Mittel=1/length(X)*sum(X) # Mittelwert 12.814285714285715
Var=1/length(X)*sum(X.^2)-Mittel^2 # Varianz 108.616
ST=sqrt(Var) # Standardabweichung 10.42
kVar=length(X)/(length(X)-1)*Var  # korrigierte Varianz 114.04
kST=sqrt(kVar) # korrigierte Standardabweichung 10.6792
```

Aufgabe 11

Gemäss dem Statistischen Amt des Kantons Zürich versteuerten im Kanton Zürich im Jahr 2023 rund 693'000 natürliche Personen ein Einkommen zwischen CHF 20'000 und CHF 200'000, wobei die Einkommen wie folgt verteilt waren:

Steuerbares Einkommen (in kCHF)	[20,40[	[40,60[	[60,80[	[80,100[	[100,150[	[150,200[
Anzahl Personen (in 1000)	148	186	141	87	98	33

a) Berechne die kumulativen relativen Häufigkeiten. Runde die Ergebnisse auf 0.01.
b) Berechne, welcher Anteil der Personen höchstens CHF 120'000 versteuern.
c) Berechne, welches maximale Einkommen die 30% der Personen haben, welche am wenigsten versteuern.
d) Welche Schiefe hat die Verteilung der steuerbaren Einkommen?
Gib ohne Rechnung an, welche Grössenbeziehung zwischen Median und Mittelwert der steuerbaren Einkommen besteht.
e) Berechne näherungsweise den Median und den Mittelwert der steuerbaren Einkommen.

<u>Lösung:</u>

a)

Steuerbares Einkommen (in kCHF)	[20,40[	[40,60[	[60,80[	[80,100[	[100,150[	[150,200[
abs. Häufigkeit (in 1000)	148	186	141	87	98	33
rel. Häufigkeit (in 1000)	0.21	0.27	0.20	0.13	0.14	0.05
kum. rel. Häufigkeit (in 1000)	0.21	0.48	0.68	0.81	0.95	1.0

b) Gegeben $x = 120$, gesucht $F(120)$:
$F(120) = 0.81 + \frac{120-100}{150-100} \cdot (0.95 - 0.81) \approx 0.866$

c) Gegeben $q = F(x) = 0.3$, gesucht $R_{0.3} = F^{-1}(0.3)$
$R_{0.3} = 40 + \frac{0.3-0.21}{0.48-0.21} \cdot (60 - 40) \approx 46.67$

d) Verteilung ist rechtsschief
Median ist kleiner als Mittelwert

e) Median:
$$x_{med} = 60 + \frac{0.5 - 0.48}{0.68 - 0.48} \cdot (80 - 60)$$
Mittelwert (mit Klassenmitten):
$\overline{x} \approx 0.21 \cdot 30 + 0.27 \cdot 50 + 0.2 \cdot 70 + 0.13 \cdot 90 + 0.14 \cdot 125 + 0.05 \cdot 175 = 71.75$

Boxplot

Aufgabe 12

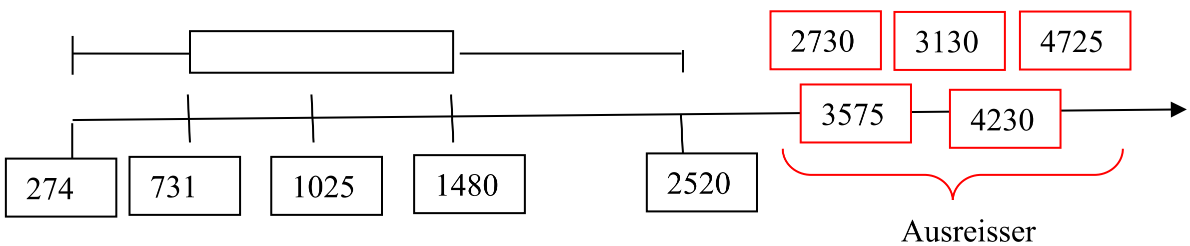
Zeichne für das **Beispiel** Nettomieten im Skript einen Boxplot.

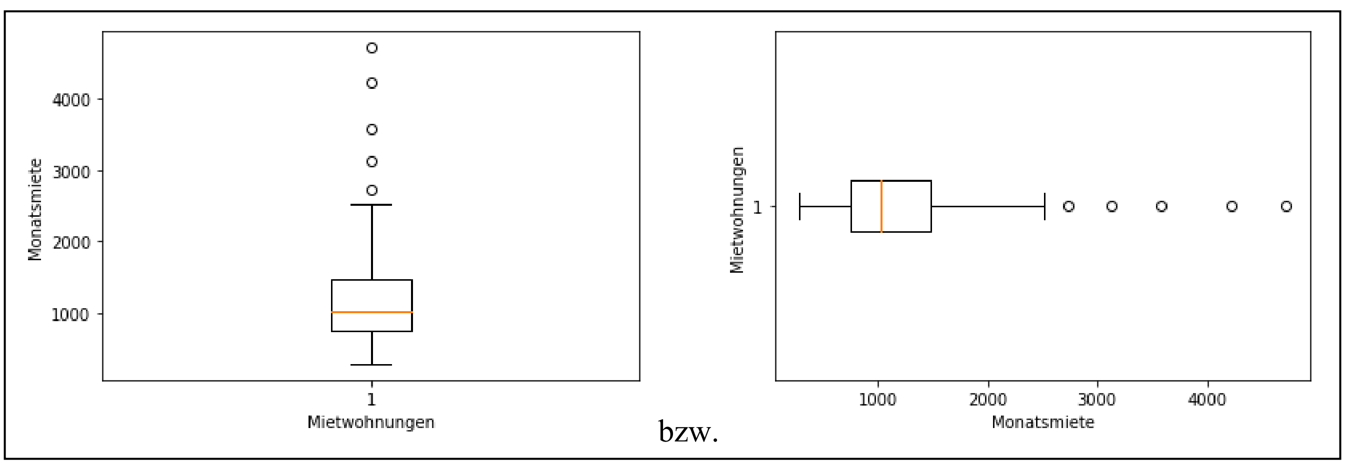
<u>Lösung</u>: $Q_1 = 735.5$, $Q_2 = 1025$, $Q_3 = 1480$,

Interquartilsabstand: $Q_3 - Q_1 = 1480 - 731 = 749$,

Maximale Antenne: $Q_3 + 1.5 \cdot (Q_3 - Q_1) = Q_3 + 1.5 \cdot 749 \approx 1480 + 1123.5 = 2603.5$

Minimale Antenne: $Q_1 - 1.5 \cdot (Q_3 - Q_1) = Q_1 - 1.5 \cdot 749 \approx 731 - 1123.5 < 0$





Aufgabe 13

Zu 36 zufällig ausgewählten Zeitpunkten wurde in einer petrochemischen Anlage die durchschnittliche Partikelkonzentration (in Mikrogramm pro Kubikmeter) der Luft gemessen.
Dabei ergaben sich folgende Messwerte:

5	18	15	7	23	220	130	85	103	25	80	7	24	6	13	65	37	25
24	65	82	95	77	15	70	110	44	28	33	81	29	14	45	92	17	53

Bestimme mit Python Histogramm, Verteilungsfunktion, Boxplot, Kennwerte (arith. Mittelwert, Q_1, Q_2, Q_3, (korrigierte) Standardabweichung).

Lösung:

```
import numpy as np
import matplotlib.pyplot as plt

## Aufgabe 1.13
X = np.array([5, 18, 15, 7, 23, 220, 130, 85, 103, 25, 80, 7, 24, 6,
13, 65, 37, 25,
     24, 65, 82, 95, 77, 15, 70, 110, 44, 28, 33, 81, 29, 14, 45, 92,
17, 53])
b=np.max(X)-np.min(X)

plt.figure(1)
plt.subplot(211)
plt.hist(X,density=1,bins=20)
plt.axis([0,250,0,0.03])
plt.ylabel('absolute Häufigkeit')
plt.title('Histogramm: Partikelkonzentration')
plt.subplot(212)
plt.hist(X,density=1,bins=5)
plt.axis([0,250,0,0.03])
plt.xlabel('Mikrogramm pro Kubikmeter')
plt.ylabel('absolute Häufigkeit')

plt.figure(2)
plt.hist(X,density=1,cumulative=True, bins=b)
plt.title('Empirische Verteilungsfunktion CDF')
plt.xlabel('Mikrogramm pro Kubikmeter')
plt.ylabel('absolute Häufigkeit')
```

```
plt.figure(3)
plt.hist(X,density=1,histtype='step',cumulati
ve=True, bins=b)
plt.title('Empirische Verteilungsfunktion
CDF')
plt.xlabel('Mikrogramm pro Kubikmeter')
plt.ylabel('absolute Häufigkeit')

plt.figure(4)
plt.boxplot(X)
plt.title('Partikelkonzentration :Boxplot')
plt.ylabel('Mikrogramm pro Kubikmeter')

#Kennwerte
print(np.percentile(X,[0.25,0.5,0.75]))
#[5.0875 5.175  5.2625]
print(np.mean(X)) #51.72
print(np.std(X)) #44.332, Standardabw.
print(np.std(X,ddof=1)) #44.96, korr. St.abw.
```