

Serie1

September 30, 2021

Wahrscheinlichkeitsrechnen (wer)
BSc Data Science @ FHNW Brugg
Lukas Reber
Serie 1

1 Aufgabe 1

quantitativ stetig
quantitativ kategorisch
qualitativ
quantitativ stetig
quantitativ stetig
qualitativ

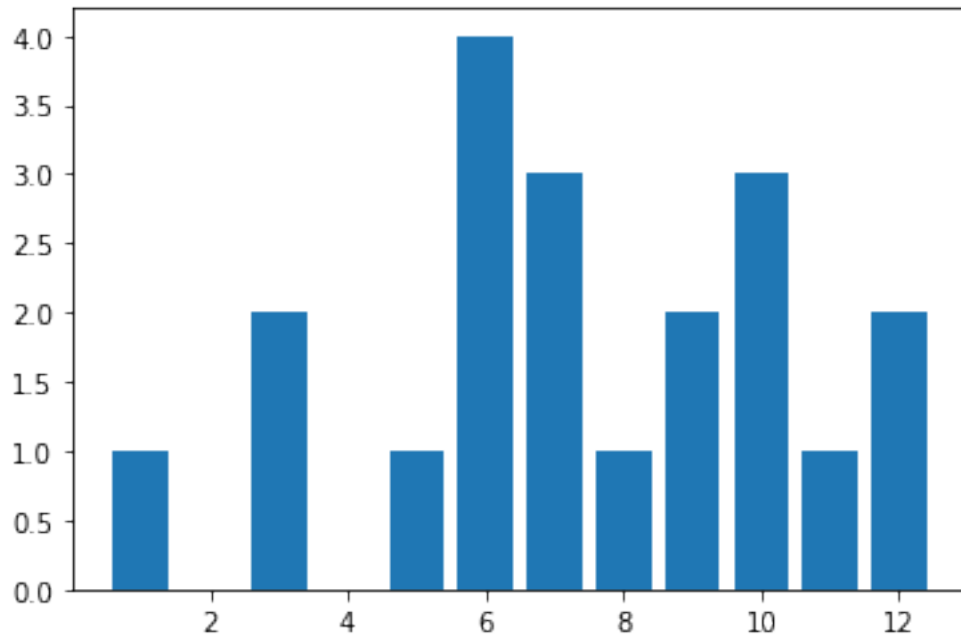
2 Aufgabe 2

```
[2]: import matplotlib.pyplot as plt
import numpy as np

l = [6,8,3,10,1,12,3,12,7,5,10,9,6,6,11,7,6,10,9,7]

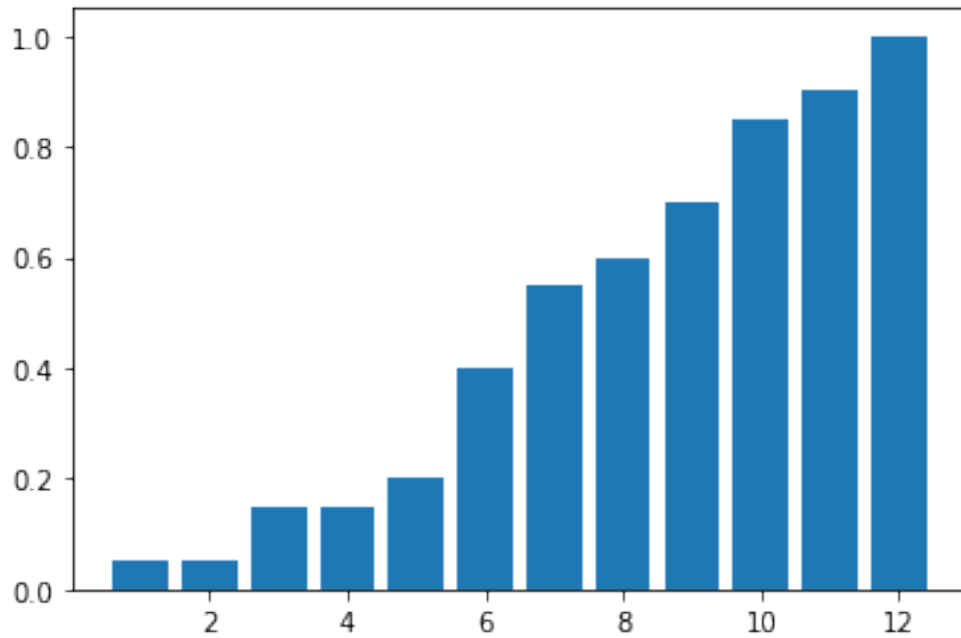
dict = {}
[dict.update({x: l.count(x)}) for x in list(set(l))]

plt.bar(dict.keys(),dict.values())
plt.show()
```



```
[3]: # relative Summenhäufigkeit berechnen
s = 0
sh = {}
t = sum(dict.values())
for i in range(1,13):
    if dict.get(i):
        s = s + round((dict.get(i) / t),2)
        sh.update({i: s})

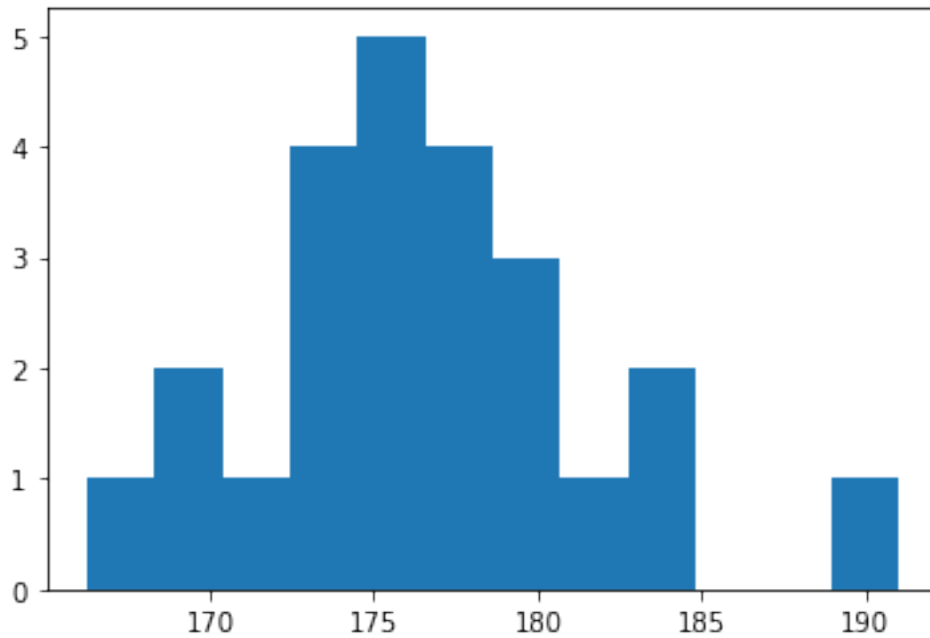
plt.bar(sh.keys(),sh.values())
plt.show()
```



3 Aufgabe 3

```
[4]: x = [182.0, 176.0, 191.0, 173.5, 183.0, 178.0, 179.0, 174.0, 184.5, 174.0, 180.  
↪5, 174.5, 178.5, 169.0, 178.5, 168.5, 179.0, 177.0, 172.5, 171.0, 176.5, 167.  
↪0, 175.0, 175.0]
```

```
[5]: # a)  
plt.hist(x, bins=12, range=(166.25,191))  
plt.show()
```

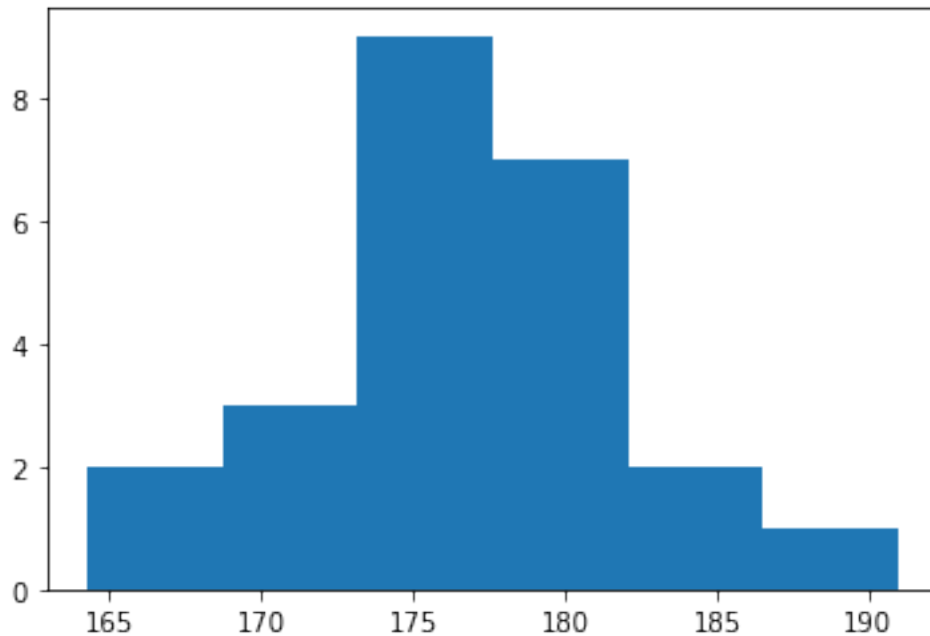


Spannweite ist $191 - 167 = 24$, mit binsize 12 ist jeder Bin 2 cm.

Wird wohl irgendwie auch gehen, in diesem Fall aber von Hand sehr viel einfacher :-)

Frage: Was ist die Ordinale?

```
[6]: # c)
plt.hist(x, bins=6, range=(164.25,191))
plt.show()
```



4 Aufgabe 4

Grundsätzlich sollten alle Bins dieselbe Grösse haben Zu wenig Klassen: Ungewöhnliche Verteilung wird ggf. nicht erkannt Zu viele Klassen: ggf. sehr viel “Noise”, welches die wichtigen Faktoren nicht mehr erkennen lassen

Verschiedene Formeln für die Berechnung der Anzahl Bins und Bin Size:
<https://answerminer.com/static/e9ebb1ee2256c796bb1f9ea05af152d1/9e74b/histogram.png>

5 Aufgabe 5

```
[7]: x = np.array([12.1, 12.2, 12.0, 12.3, 12.3, 12.4, 12.1, 12.4, 12.0, 12.1])
      x.mean()

      print(f'Fehlender Wert: {x[-1]}')
```

Fehlender Wert: 12.1

```
[8]: x.std()
```

```
[8]: 0.14456832294800986
```

```
[9]: np.median(x)
```

```
[9]: 12.149999999999999
```

6 Aufgabe 6

```
[10]: x = np.array([2.5, 1.8, 3.0, 2.2, 2.9, 1.9, 2.3, 2.6])
      print(f'arithmetisches Mitte: {x.mean()}, Median: {np.median(x)}')
```

arithmetisches Mitte: 2.4, Median: 2.4

7 Aufgabe 7

```
[11]: x = np.array([9,9,9,9,9,10,11,12,13])
      print(f'arithmetisches Mitte: {x.mean()}, Median: {np.median(x)}, Std: {x.
      ↪std()}')
```

arithmetisches Mitte: 10.111111111111111, Median: 9.0, Std: 1.4487116456005886

8 Aufgabe 8

```
[12]: x = np.array(1)
      print(f'Durchschnitt: {x.mean()}')
      print(f'Median: {np.median(x)}')
      print(f'Modus: {x[np.argmax(np.unique(x, return_counts=1)[1])]}')
      p25,p75 = np.percentile(x,[25,75])
      print(f'IQR: {p75-p25}')
      print(f'Empirische Varianz: {x.var()}')
      print(f'Empirische Standardabweichung: {x.std()}')
      print(f'Spannweite: {x.max()-x.min()}')
```

Durchschnitt: 7.4

Median: 7.0

Modus: 10

IQR: 4.0

Empirische Varianz: 8.74

Empirische Standardabweichung: 2.9563490998188966

Spannweite: 11

9 Aufgabe 9

Antwort: 23 oder 25

10 Aufgabe 10

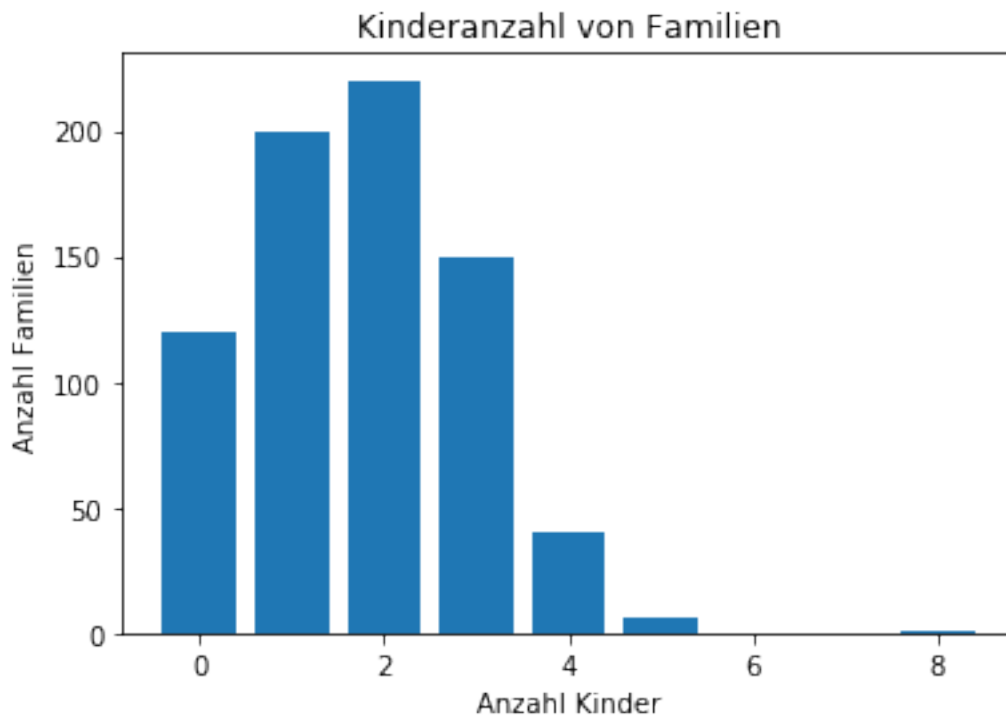
```
[13]: fam = {
      0: 120,
      1: 200,
      2: 220,
      3: 150,
```

```

4: 40,
5: 6,
8: 1
}

plt.bar(fam.keys(),fam.values())
plt.xlabel("Anzahl Kinder")
plt.ylabel("Anzahl Familien")
plt.title("Kinderanzahl von Familien")
plt.show()

```



```

[14]: fam_arr = np.array(list(fam.values()))

print(f'Durchschnitt: {fam_arr.mean()}')
print(f'Median: {np.median(fam_arr)}')
print(f'Varianz: {np.var(fam_arr)}')

```

Durchschnitt: 105.28571428571429

Median: 120.0

Varianz: 7048.775510204082

11 Aufgabe 11

Arithmetisches Mittel:

$$\frac{n}{m+n}$$

Standardabweichung:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - (\frac{n}{m+n}))^2}$$