Exercise 2 - DV1597

Marius Stokkedal, Matk20@student.bth.se

1. **P(A) = 0.3 and P(B) = 0.7**
    a. No, because the result is determined if the events are independent and/or mutually exclusive. If they for example are independant, $P(A \cap B) = P(A) \cdot P(B)$ and if they're mutually exclusive $P(A \cap B) = 0$. If the events instead are dependent the result is $P(A|B) = \frac{P(A \cap B)}{P(B)}$ or $P(B|A) = \frac{P(A \cap B)}{P(A)}$ depending on which event has happened. $P(A|B)$ is read as "probability of A happening given that B has happened".
    b. Independent random events
        i. $P(A \cap B) = P(A) \cdot P(B) = 0.3 \cdot 0.7 = 0.21$
        ii. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.7 - 0.21 = 0.79$
        iii. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.21}{0.7} = 0.3$ (Same as $P(A)$ because the events are independent)
2. **Compare datasets**

Formula for standard deviation:

$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ where n is the size of the dataset and $\bar{x}$ is the sample-mean

a. 1,7,7,7,9,12,12,1,14
    i. Mean: $\frac{1+7+7+7+9+12+12+1+21}{9} = 7.77\ldots$
    ii. Standard deviation: 4.6
b. 1,7,7,7,9,12,12,12,21
    i. Mean: $\frac{1+7+7+7+9+12+12+12+21}{9} = 9$
    ii. Standard deviation: 4.0

Dataset *a* has a lower mean and a bigger standard deviation

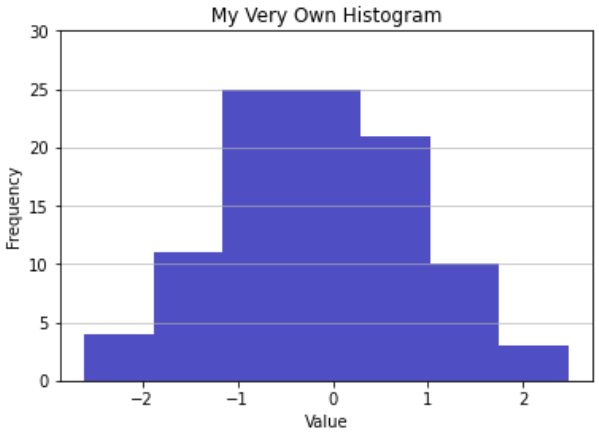
c. -10,0,0,0,17,27,40,40
    i. Mean: $\frac{-10+17+27+40+40}{8} = 14.25$
    ii. Standard deviation: 19.6
d. -30,0,0,0,17,27,40,40
    i. Mean: $\frac{-30+17+27+40+40}{8} = 11.25$
    ii. Standard deviation: 24.0

In this case dataset *b* has a lower mean but *a* has a lower standard deviation.

3. According to the *68-95-99.7* rule, also known as the empirical rule, ≈68%, ≈95% and ≈99.7% of the values must be located within one, two and three standard deviations from the mean, respectively.[1] By using numpy and spicy.stats in python I calculated the following: 67.92%, 95.29% and 99.71% of the values lie within one, two and three standard deviations of the mean, respectively. In conclusion, the data seem to have

[1] "10 Normality Tests-Python (2020) | Towards Data Science." 18 sep.. 2020, https://towardsdatascience.com/normality-tests-in-python-31e04aa4f411. Öppnades 20 apr.. 2022.

normality as it follows the *68-95-99.7* rule. This can also be shown if the data is plotted [figure 1], it does not form a perfect bell curve, but it still has the general shape and the bar count is quite low and therefore sensible for differences in value distribution.



[Figure 1]