

Assignment 1

(Descriptive Statistics, Probability & Probability Distributions)



Table of Content

1. [Descriptive Statistics](#)
2. [Probability](#)
3. [Probability Distributions](#)

Import the required libraries

```
In [1]: # type your code here
```

Let's begin with some hands-on practice exercises

1. Descriptive Statistics



1. The rainfall (in mm) in the city is recorded for 10 days. Find the rainfall value under which 60% of the rainfall would lie.

Given data:

```
rainfall (in mm) = [86, 74, 90, 108, 65, 84, 75, 92, 102, 93]
```

```
In [1]: from scipy import stats as st

rainfall = [86, 74, 90, 108, 65, 84, 75, 92, 102, 93]      # given rainfall data in mm

percentile = 60      # to calculate the 60th percentile of the data.

result = st.scoreatpercentile(rainfall, percentile)

print(f"The rainfall value under which {percentile}% of the rainfall would lie is {result:.1f} mm.")
```

The rainfall value under which 60% of the rainfall would lie is 90.8 mm.



2. The performance of John and Jack in the monthly exams is given below. Identify the most consistent student.

Given data:

```
John = [42, 38, 47, 35, 39, 46]
Jack = [32, 36, 46, 49, 28, 30]
```

```
In [6]: import statistics as st

John = [42, 38, 47, 35, 39, 46]
Jack = [32, 36, 46, 49, 28, 30]

x = st.stdev(John)
y = st.stdev(Jack)

print(f"Standard deviation of John's scores = {x:.2f}")
print(f"Standard deviation of Jack's scores = {y:.2f}")

if x < y:
    print('John's exam scores are more consistent than Jack's')
else:
    print('Jack's exam scores are more consistent than John's')
```

```
Standard deviation of John's scores = 4.71
Standard deviation of Jack's scores = 8.73
John's exam scores are more consistent than Jack's
```



3. The amount of Calcium, Potassium and Iron in the chocolate cookies of 5 different brands is collected. Find out which mineral can be neglected while comparing the different cookie brands?

Given data:

Calcium (in mg) = [132, 138.56, 147.2, 145.6, 139]
Potassium (in mg) = [122.2, 116, 106.6, 119.52, 128]
Iron (in mg) = [2.96, 3.02, 3.01, 2.99, 2.93]

```
In [7]: import numpy as np

calcium = [132, 138.56, 147.2, 145.6, 139]
potassium = [122.2, 116, 106.6, 119.52, 128]
iron = [2.96, 3.02, 3.01, 2.99, 2.93]

cv_calcium = np.std(calcium) / np.mean(calcium)
cv_potassium = np.std(potassium) / np.mean(potassium)
cv_iron = np.std(iron) / np.mean(iron)

print(f"CV of Calcium: {cv_calcium:.4f}")
print(f"CV of Potassium: {cv_potassium:.4f}")
print(f"CV of Iron: {cv_iron:.4f}")

if cv_calcium < cv_potassium < cv_iron:
    print("Calcium can be neglected as it shows the least variability among the three minerals.")
elif cv_potassium < cv_calcium < cv_iron:
    print("Potassium can be neglected as it shows the least variability among the three minerals.")
else:
    print("Iron can be neglected as it shows the least variability among the three minerals.")

CV of Calcium: 0.0389
CV of Potassium: 0.0600
CV of Iron: 0.0111
Iron can be neglected as it shows the least variability among the three minerals.
```

2. Probability



4. A magician wants you to select three cards from a pack of 52 cards. You have to choose a card without replacement. Find the probability that one card is a queen?

In [44]:

```
import math

# Number of ways to choose 1 queen out of 4
n_ways_choose_1_queen = comb(4, 1)

# Number of ways to choose 2 cards from the remaining 48 cards
n_ways_choose_2_cards = comb(48, 2)

# Total number of ways to choose 1 queen and 2 other cards
n_ways_choose_1_queen_and_2_cards = n_ways_choose_1_queen * n_ways_choose_2_cards

# Total number of ways to choose any 3 cards from a pack of 52
n_ways_choose_3_cards = comb(52, 3)

# Probability of choosing one queen when selecting three cards from a pack of 52 cards without replacement
prob = n_ways_choose_1_queen_and_2_cards / n_ways_choose_3_cards

print(f"The probability of selecting one queen card is {prob:.3f}.")
```

The probability of selecting one queen card is 0.204.



5. The odds that a Canadian picked at random will be left-handed are 3:14. What is the probability of selecting a left-handed person?

In [6]:

```
odds_left_handed = (3, 14)

# Calculate the total number of outcomes
total_outcomes = sum(odds_left_handed)

# Calculate the probability of selecting a Left-handed Canadian
prob_left_handed = odds_left_handed[0] / total_outcomes

print(f"The probability of selecting a left-handed Canadian at random is {prob_left_handed:.3f}.")
```

The probability of selecting a left-handed Canadian at random is 0.176.



6. Consider two dice are rolled simultaneously. Find the probability that the sum of the two numbers is at least 9.

```
In [7]: from itertools import product

# Define the sample space
dice = [1, 2, 3, 4, 5, 6]
sample_space = list(product(dice, repeat=2))

# Count the number of outcomes where the sum of the two numbers is at least 9
n_outcomes = len([i for i in sample_space if sum(i) >= 9])

# Calculate the probability
prob = n_outcomes / len(sample_space)

print(f"The probability that the sum of the two numbers is at least 9 when two dice are rolled is {prob:.3f}.")
```

The probability that the sum of the two numbers is at least 9 when two dice are rolled is 0.278.



7. The university wants to recruit a professor with a PhD in Mathematics given that he is older than 30 years. The university has got a total of 120 applications out of which 45 applicants, are older than 30 years. Out of these 45 applicants, 27 are Masters in Mathematics and others possess PhD. What is the probability of selecting a suitable candidate?

```
In [11]: total_applications = 120
applicants_over_30 = 45
masters_in_math = 27
phd_in_math = applicants_over_30 - masters_in_math

prob = phd_in_math / total_applications

print(f"The probability of selecting a suitable candidate is {prob:.2f}.")
```

The probability of selecting a suitable candidate is 0.15.



8. A school is planning to send a team of 6 students for the national science quiz. After taking three quiz rounds in the school, the principal has shortlisted two teams. The first team is consists of 4 boys and 2 girls. The other team is consists of 3 boys and 3 girls. The news reporter of school selects a team at random and a girl student is selected from this team randomly for the interview. Find the probability that the first team was chosen by the news reporter.

```
In [45]: # Probability that the first team is chosen and a girl is selected for the interview
prob_first_team_girl = (1/2) * (2/6)
```

```
# Probability that the second team is chosen and a girl is selected for the interview
prob_second_team_girl = (1/2) * (3/6)

# Total probability of selecting a girl student for the interview
prob_girl = prob_first_team_girl + prob_second_team_girl

# Probability that the first team was chosen by the news reporter,
# given that a girl student was selected for the interview
prob_first_team_given_girl = prob_first_team_girl / prob_girl

print(f"The probability that the first team was chosen by the news reporter is {prob_first_team_given_girl:.2f}.")
```

The probability that the first team was chosen by the news reporter is 0.40.

3. Probability Distributions



9. In a factory, the probability of producing a defective plastic box is 0.32. A sample of 50 boxes is collected. What is the probability that exactly 15 boxes are defective? (Note : Define the Random Variable)

```
In [19]: from scipy.stats import *

# Let X = defective box be the random variable

n = 50      # sample of 50 boxes is collected
p = 0.32    # probability of producing a defective plastic box is 0.32
k = 15      # probability that exactly 15 boxes are defective

probability=binom.pmf(k, n, p)

print(f"The probability that exactly {k} boxes are defective in a sample of {n} boxes is {prob:.4f}.")
```

The probability that exactly 15 boxes are defective in a sample of 50 boxes is 0.1168.

```
In [20]: from math import comb

n = 50
p = 0.32
k = 15

prob = comb(n, k) * (p**k) * ((1-p)**(n-k))

print(f"The probability that exactly {k} boxes are defective in a sample of {n} boxes is {prob:.4f}.")
```

The probability that exactly 15 boxes are defective in a sample of 50 boxes is 0.1168.



10. On a security gate, a truck passes at a rate of 4 per hour. What is the probability that at most 10 trucks will pass in 90 minutes?(Note : Define the Random Variable)

```
In [8]: from scipy.stats import *

# Let X = number of trucks that pass through the security gate in 90 minutes

mean = 4          # truck passes at a rate of 4 per hour
mu = 4/60 * 90    # trucks will pass in 90 minutes
k = 10            # probability that at most 10 trucks will pass in 90 minutes

prob=poisson.cdf(k, mu=6)

print(f"The probability that at most {k} trucks will pass through the security gate in 90 minutes is {prob:.4f}.")
```

The probability that at most 10 trucks will pass through the security gate in 90 minutes is 0.9574.



11. The dietitian has offered a meal plan for 30 patients. After two months of starting the diet, she has collected the data of weights for all the patients. The average weight is 80 kg with a standard deviation of 20 kg. What is the probability that the weight of a person is between 70 to 90 kg? Define the Random Variable

```
In [46]: from scipy.stats import *

# Given data
n = 30
avg = 80
std = 20
a = 70
b = 90
z1 = (b - avg) / std    # standardize the variable with x = 90
z2 = (a - avg) / std    # standardize the variable with x = 70

prob = norm.cdf(z1)-norm.cdf(z2)

print(f"The probability of a person following the dietitian's meal plan is between {a} to {b} kg is {prob:.4f}.")
```

The probability of a person following the dietitian's meal plan is between 70 to 90 kg is 0.3829.