Topic 03: Review of Statistics and Probability

C. Canoza | I.J. Timbungco | N.K. Vitales | S.C. Zarraga | 2022 | Fudamentals of Machine Learning

```
import pandas as pd
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
```

Find the median for the data 8, 5, 7, 10, 15, 21.

```
# code

data = [8, 5, 7, 10, 15, 21]
data = sorted(data)

print(f"median: {np.median(data)}")
```

```
median: 9.0
```

The following data represents the survey regarding the heights (in cm) of 51 girls of Class x. Find the median height.

Height (in cm)	Number of Girls
Less than 140	4
Less than 145	11
Less than 150	29
Less than 155	40
Less than 160	46
Less than 165	51

Get the mean and median of the height

```
data_height = sorted([140, 145, 150, 155, 160, 165])
print(f"mean height: {np.mean(data_height)} \nmedian height: {np.median(data_he
```

```
mean height: 152.5
median height: 152.5
```

In the given set of data: 2, 4, 5, 5, 6, 7, the mode of the data set is ___ since it has appeared in the set twice.

```
## What is the mode

data = sorted([2, 4, 5, 5, 6, 7])
print(f"mode: {stats.mode(data)}")
```

```
mode: ModeResult(mode=array([5]), count=array([2]))
```

Find the variance and standard deviation of the following scores on an exam: 92, 95, 85, 80, 75, 50

```
exam_data = [92, 95, 85, 80, 75, 50]
print(f"exam variance: {np.var(exam_data)} \nexam standard deviation: {np.std(e
```

```
exam variance: 219.58333333333334
exam standard deviation: 14.818344486930156
```

Baye's Theorem

```
def bayes_theorem(p_b, p_g_given_b, p_g_given_not_b):
  # calculate P(not B)
  not_b = 1 - p_b
  # calculate P(G)
  p_g = p_g_given_b * p_b + p_g_given_not_b * not_b
  # calculate P(B|G)
  p_b_given_g = (p_g_given_b * p_b) / p_g
  return p_b_given_g
#P(B)
p_b = 1/4
# P(G|B)
p_g_given_b = 1
# P(G|notB)
p_g_given_not_b = 1/3
# calculate P(B|G)
result = bayes_theorem(p_b, p_g_given_b, p_g_given_not_b)
# print result
print('P(B|G) = %.2f%%' % (result * 100))
```

```
P(B|G) = 50.00%
```