# Assignment III  Sample Answers

For Q1 and Q2 which ask you to prepare frequency tables and Pareto charts, a sample answer for the Covid-16 related deaths by race is as the followings:
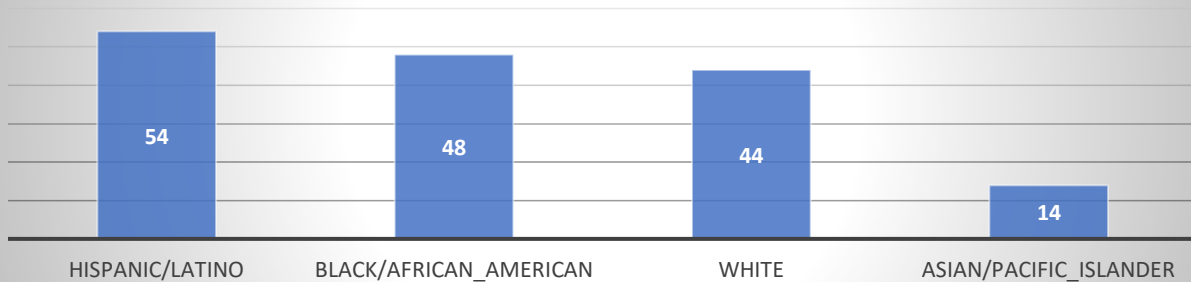
| Freqeuncy Table | | Relative Freqeuncy Table | |
|---|---|---|---|
| Row Labels | Frequencies | Row Labels | Relative Frequencies |
| Hispanic/Latino | 54 | Hispanic/Latino | 33.75% |
| Black/African_American | 48 | Black/African_American | 30.00% |
| White | 44 | White | 27.50% |
| Asian/Pacific_Islander | 14 | Asian/Pacific_Islander | 8.75% |
| Grand Total | 160 | Grand Total | 100.00% |

**Pareto Chart for Frequencies of Covid Death Cases by Race in NY**

| | |
|---|---|
| HISPANIC/LATINO | 54 |
| BLACK/AFRICAN_AMERICAN | 48 |
| WHITE | 44 |
| ASIAN/PACIFIC_ISLANDER | 14 |

**Pareto Chart for Relative Frequencies of Covid Death Cases by Race in NY**

| | |
|---|---|
| HISPANIC/LATINO | 33.75% |
| BLACK/AFRICAN_AMERICAN | 30.00% |
| WHITE | 27.50% |
| ASIAN/PACIFIC_ISLANDER | 8.75% |
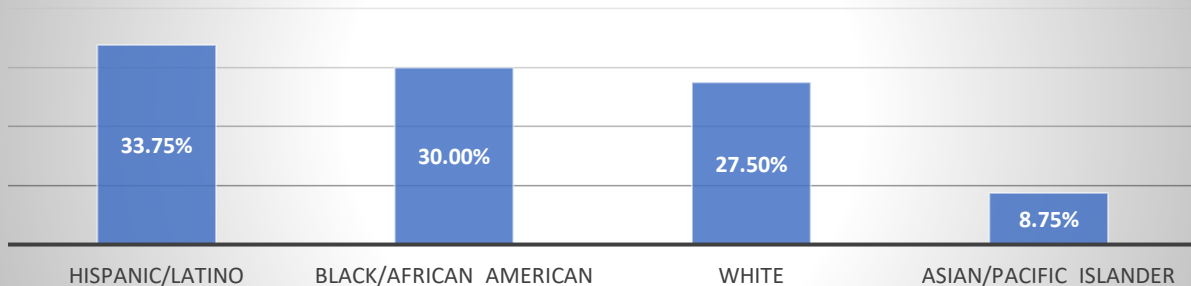
For Question Q3 a sample answer is as the following:

a)

| Summary Statistics | | |
| --- | --- | --- |
| | Covid Infection Rate | Average Income |
| Mean | 1 | 25 |
| Median | 2 | 24 |
| Min | 3 | 23 |
| Max | 4 | 22 |
| Range | 5 | 21 |
| First Quartile | 6 | 20 |
| Second Quartile | 7 | 19 |
| Third Quartile | 8 | 18 |
| Interquartile | 9 | 17 |
| 10th Percentile | 10 | 16 |
| 20th Percentile | 11 | 15 |
| 30th Percentile | 12 | 14 |
| 40th Percentile | 13 | 13 |
| 50th Percentile | 14 | 12 |
| 60th Percentile | 15 | 11 |
| 70th Percentile | 16 | 10 |
| 80th Percentile | 17 | 9 |
| 90th Percentile | 18 | 8 |
| 95th Percentile | 19 | 7 |
| 99th Percentile | 20 | 6 |
| Variance | 21 | 5 |
| Standard Deviation | 22 | 4 |
| Coefficient Variation | 23 | 3 |
| Skewness | 24 | 2 |
| Kurtosis | 25 | 1 |
| Covariance | 26 | |
| Correlation Coefficient | 27 | |

b)

**Covid Infection Rate & Average Income**

c)



**Histogram - Covid Infection Rate**
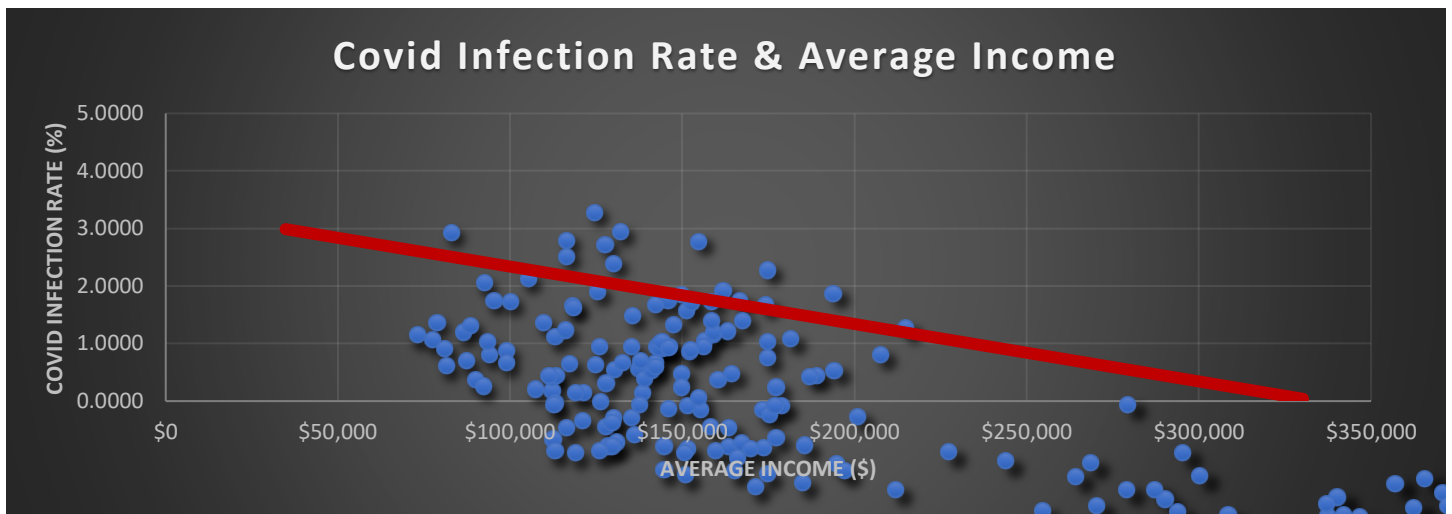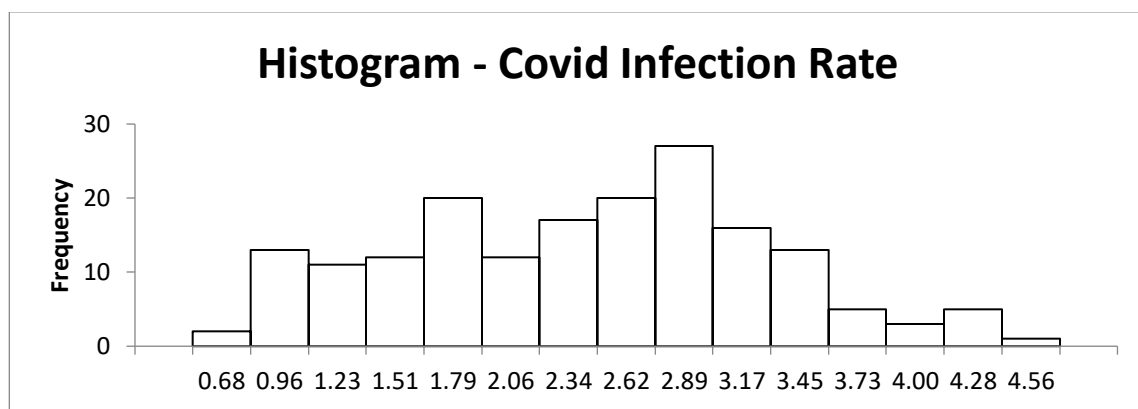
Do the same for Income

d)

i.   **Mean:**

$$\bar{x} = \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

**Mean of Covid-19 Infection Rate**

$$\overline{\text{Infection Rate}} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1000}{500} = 2$$

**Mean of Income**

$$\overline{\text{Income}} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \cdots.$$

ii.  **Variance:**

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n-1}$$

**Variance of Covid-19 Infection Rate**

$$s_{Covid}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \frac{1000}{500} = 2$$

**Variance of Income**

$$s_{income}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n-1} = \cdots$$

iii.  **Standard Deviation:**

$$s = \sqrt{Variance} = \sqrt{s^2}$$

**Standard Deviation of Covid-19 Infection Rate**

$$s_{Covid} = \sqrt{s_{Covid}^2} = \sqrt{2}$$

**Standard Deviation of Income**

$$s_{Income} = \sqrt{s_{Income}^2} = \cdots$$

iv.  **Coefficient of Variation:**

$$CV = \frac{s}{\bar{x}}$$

**Coefficient of Variation of Covid-19 Infection Rate**

$$CV_{Covid} = \frac{s_{Covid}}{\text{Infection Rate}} = \frac{\sqrt{2}}{2} = 0.70$$

**Coefficient of Variation of Income**

$$CV_{Income} = \frac{S_{Income}}{\overline{Income}} = \dots$$

## v. Skewness:

$$S = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{n}}{s^3} = \frac{\frac{(x_1-\bar{x})^3+(x_2-\bar{x})^3+(x_3-\bar{x})^3+\cdots+(x_n-\bar{x})^3}{n}}{s^3}$$

### Skewness of Covid-19 Infection Rate

$$S_{Covid} = \frac{\frac{(x_1-\bar{x})^3+(x_2-\bar{x})^3+(x_3-\bar{x})^3+\cdots+(x_n-\bar{x})^3}{n}}{s^3} = \frac{\frac{1000}{177}}{\sqrt{2}^3} = 1.99$$

### Skewness of Income

$$S_{Income} = \frac{\frac{(x_1-\bar{x})^3+(x_2-\bar{x})^3+(x_3-\bar{x})^3+\cdots+(x_n-\bar{x})^3}{n}}{s^3} = \dots$$

## vi. Kurtosis:

$$K = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{n}}{s^4} == \frac{\frac{(x_1-\bar{x})^4+(x_2-\bar{x})^4+(x_3-\bar{x})^4+\cdots+(x_n-\bar{x})^4}{n}}{s^4}$$

### Kurtosis of Covid-19 Infection Rate

$$K_{Covid} = \frac{\frac{(x_1-\bar{x})^4+(x_2-\bar{x})^4+(x_3-\bar{x})^4+\cdots+(x_n-\bar{x})^4}{n}}{s^4} = \frac{\frac{1000}{177}}{\sqrt{2}^4} = 1.41$$

### Kurtosis of Income

$$K_{Income} = \frac{\frac{(x_1-\bar{x})^4+(x_2-\bar{x})^4+(x_3-\bar{x})^4+\cdots+(x_n-\bar{x})^4}{n}}{s^4} = \dots$$

**vii.   Correlation Coefficients:**

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y} = \frac{\frac{(x_1-\bar{x})(y_1-\bar{y})+(x_2-\bar{x})(y_2-\bar{y})+(x_3-\bar{x})(y_3-\bar{y})+\cdots+(x_n-\bar{x})(y_n-\bar{y})}{n-1}}{s_x s_y}$$

**Correlation Coefficients between Covid-19 Infection Rate and Average Income**

$$r = \frac{\frac{-5883886.25}{176}}{50688.22} = -0.6595 \approx -65.95\%$$

**e)** The neighborhoods that are creating these extreme values in Average Income are:

| Zipcode | Neighborhood | Borough |
|---------|--------------|---------|
| 1 | A | X (Average Income: $......) |
| 2 | B | Y (Average Income: $......) |
| 3 | C | Z (Average Income: $......) |

The neighborhoods that are creating these extreme values in the infection rate:

| Zipcode | Neighborhood | Borough |
|---------|--------------|---------|
| 1 | A | X (Covid Infection rate: ......%) |
| 2 | B | Y (Covid Infection rate: ......%) |
| 3 | C | Z (Covid Infection rate: ......%) |