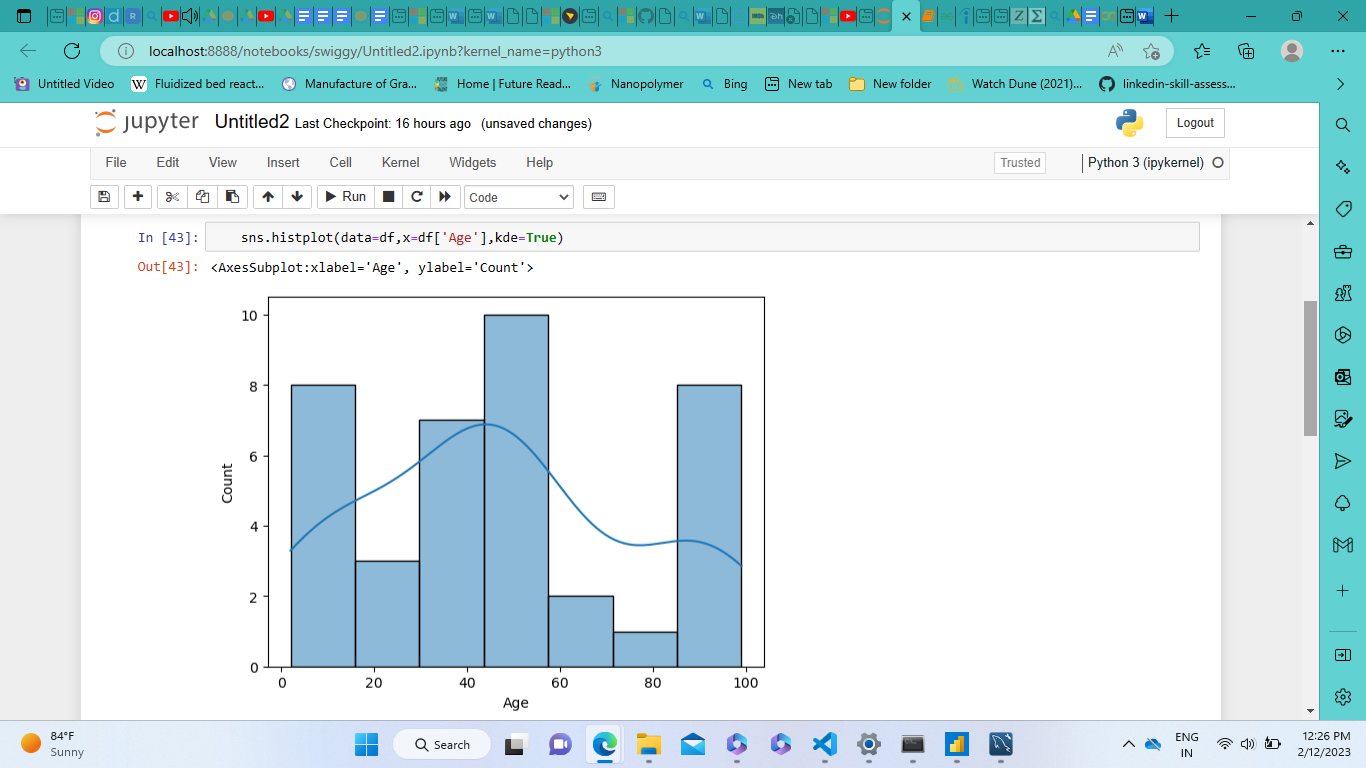
**Statistics Assignment 1**

1. **Plot a histogram,**

**10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99**



1. **In a quant test of the CAT Exam, the population standard deviation is known to be 100. A sample of 25 tests taken has a mean of 520. Construct an 80% CI about the mean.**

To construct an 80% Confidence Interval (CI) about the mean, we can use the formula for the Confidence Interval for a population mean when the population standard deviation is known:

**CI = X̄ ± Z \* (σ/√n)**

Where: X̄ = Sample mean (520 in this case)

Z = Z-score corresponding to the desired confidence level (80% confidence level corresponds to a Z-score of 1.28)

σ = Population standard deviation (known to be 100)

n = Sample size (25 in this case)

Now, let's calculate the Confidence Interval:

CI = 520 ± 1.28 \* (100/√25)

CI = 520 ± 1.28 \* (100/5)

CI = 520 ± 1.28 \* 20

CI = 520 ± 25.6

Lower bound of CI = 520 - 25.6 = 494.4

Upper bound of CI = 520 + 25.6 = 545.6

The 80% Confidence Interval about the mean is approximately 494.4 to 545.6. This means that we are 80% confident that the true population mean falls within this range based on the sample data.

1. **A car believes that the percentage of citizens in city ABC that owns a vehicle is 60% or less. A sales manager disagrees with this. He conducted a hypothesis testing surveying 250 residents & found that 170 residents responded yes to owning a vehicle.**
2. **State the null & alternate hypothesis.**
3. **At a 10% significance level, is there enough evidence to support the idea that vehicle owner in ABC city is 60% or less.**

**Null Hypothesis (H0)**: The percentage of citizens in city ABC that owns a vehicle is 60% or less.

**Alternate Hypothesis (H1)**: The percentage of citizens in city ABC that owns a vehicle is greater than 60%.

To test this hypothesis, we can use a one-tailed z-test for proportions, as we are testing if the proportion of vehicle owners is greater than 60%.

Here's the calculation to determine if there is enough evidence to support the idea that vehicle owners in ABC city is 60% or less:

* Calculate the sample proportion of vehicle owners:

Sample proportion (p̂) = Number of residents who own a vehicle / Total number of respondents

= 170 / 250

= 0.68 (or 68%)

* Calculate the standard error (SE) of the sample proportion:

SE = √(p̂ \* (1 - p̂) / n)

= √(0.68 \* (1 - 0.68) / 250)

= √(0.68 \* 0.32 / 250)

= √(0.2176 / 250)

= √0.0008704

≈ 0.0295

* Calculate the test statistic (z):

z = (p̂ - P) / SE

= (0.68 - 0.60) / 0.0295

= 0.08 / 0.0295

≈ 2.71

* Find the critical value at the 10% significance level (α = 0.10): The critical value for a one-tailed test at 10% significance level is approximately 1.28.
* Compare the test statistic with the critical value: Since the test statistic (2.71) is greater than the critical value (1.28), we reject the null hypothesis.

**Conclusion**: At a 10% significance level, there is enough evidence to support the idea that the percentage of citizens in city ABC who own a vehicle is greater than 60%. The sales manager's claim is supported by the survey data.

1. **What is the value of the 99 percentile?**

**2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,11,12**

Given data set: 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12

* Sort the data in ascending order: 2, 2, 3, 4, 5, 5, 5, 6, 7, 8, 8, 8, 8, 8, 9, 9, 10, 11, 11, 12
* Calculate the position of the 99th percentile:

99th percentile position = (99/100) \* (total number of data points)

= (99/100) \* 20

= 0.99 \* 20

= 19.8

* Since the position (19.8) is not a whole number, we need to interpolate the values at positions 19 and 20.

Value at position 19 = 11

Value at position 20 = 12

* Interpolation:

Value at the 99th percentile = Value at position 19 + (position fraction \* difference between values at positions 20 and 19)

= 11 + (0.8 \* (12 - 11))

= 11 + (0.8 \* 1)

= 11 + 0.8

= 11.8

So, the value at the 99th percentile in the given data set is approximately **11.8.**

1. **In left & right-skewed data, what is the relationship between mean, median & mode?**

**Draw the graph to represent the same.**

* In left-skewed (negatively skewed) data:

Mean < Median < Mode

* In right-skewed (positively skewed) data:

Mean > Median > Mode

Explanation:

**Mean**:

The mean is the arithmetic average of the data points. In left-skewed data, the mean is typically less than the median, as the long tail on the left side of the distribution pulls the mean towards lower values. In right-skewed data, the mean is usually greater than the median, as the long tail on the right side pulls the mean towards higher values.

**Median**:

The median is the middle value of the data when arranged in ascending order. In left-skewed data, the median is generally greater than the mode, as the majority of data points are concentrated towards the higher end, pulling the median towards higher values. In right-skewed data, the median is generally less than the mode, as the majority of data points are concentrated towards the lower end, pulling the median towards lower values.

**Mode**:

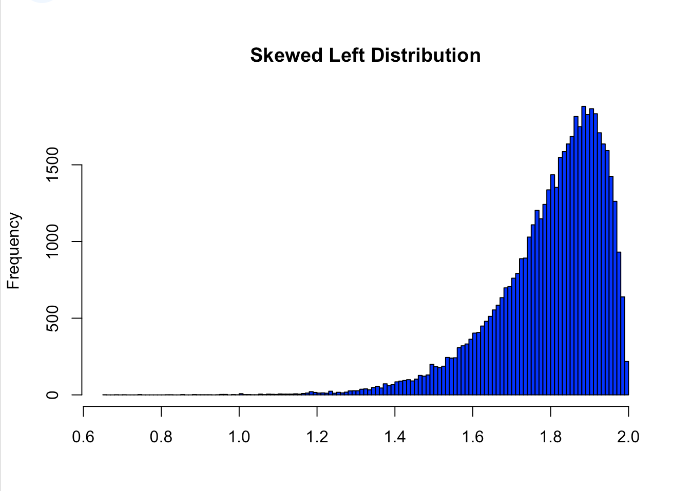
The mode is the value that appears most frequently in the data. In left-skewed data, the mode is typically greater than the median, as it represents the peak of the distribution towards the higher values. In right-skewed data, the mode is usually less than the median, as it represents the peak of the distribution towards the lower values.

Graphical Representation:

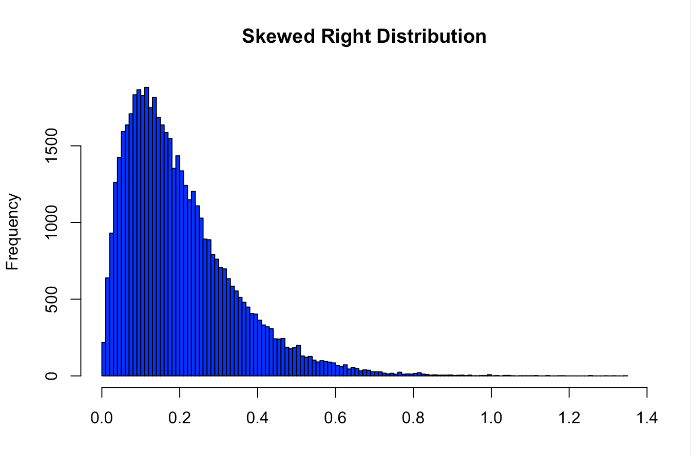
* In a left-skewed distribution, the tail extends towards the left, and the mean is on the left side of the median, which is on the left side of the mode.
* In a right-skewed distribution, the tail extends towards the right, and the mean is on the right side of the median, which is on the right side of the mode.

Below are the rough graphs for a left-skewed and a right-skewed distribution:

**Left-skewed distribution:**



**Right-skewed distribution:**

****