

A graph with a curve

Description automatically generated

The normal curve overlaid on the histogram **does not** track the shape of the underlying data closely, and this is expected given that there are clear skew and tail deviations from normality in these distributions. **Spearman’s Rho or Kendall’s Tau** in view of the fact that these tests do not assume normality and therefore are more appropriate for data that is other than normal

## **Full marks if all the axes are clearly labelled, with units of measure, and there is a trendline running through the scatterplot (for correlation-type questions). IMPORTANT: The axes must use proper ENGLISH, not the wording from the dataset column if abbreviated, underscored, etc. The dependent variable is on the y axis.**

**X and Y Axes:**

The x-axis is titled **Star Rating**. The y-axis reads "Frequency," which is correct, Since the task emphasizes direct counts (how many restaurants fall within a specific star rating), labeling the y-axis as "Frequency" makes the chart more intuitive for this analysis.

**Title:**

It has a title that describes the plot: **Histogram of Star Ratings with Normal Curve Overlay**.

**Normal Bell Curve:**

The smooth normal bell curve on top of the histogram

## **Data Visualisation: For Comparison of Means/Medians and Correlation questions only. Is the Histogram correct?**

Yes

## Correctness

### Dependent variable:

The histogram describes the dependent variable (**StarRating**) appropriately, as it presents a frequency distribution.

### Normal Curve Overlay:

Having a normal curve overlay thus helps you to visually check for the data distribution if its normal or not, which is important in choosing the right statistical test to perform.

### Axes and Labels:

Axes are correctly labelled, and the plot has a sensible title fulfilling our third requirement for clear and informative visualization.

Suitability for Analysis: The histogram accurately enables

### Comparison of Means/Medians:

This depends on the shape of the distribution, to choose a part between parametric and non-parametric tests.

### Correlation assessment:

The difference is the main part, but the distribution of dependent variable cannot be neglected as well.

## **Data Visualisation: Does response state whether dep variable appears normally distributed or not? (N/A for comp of proportions)?**

The distribution of our dependent variable, **StarRating**, shows a small right skew away from the normal curve. It does not quite fit the bell-shaped curve of a normal distribution. The data does not follow perfect normality. Although the histogram displays near-symmetry, deviations suggest that this distribution does not meet the parametric assumption of normality.