For each category in the dataset:

1. Count: the number of occurrences of each category in the sample (you can find from Dataset)

2. Proportion (p): The relative frequency of each category calculated as:

Where the estimated proportion for category *k;*

is the count of observations in category

*N* is the total number of observations in the dataset.

3. Standard Error (SE): The measure of variability of the proportion, calculated as:

where - is a prior count (set to 2) to adjust for small sample sizes and prevent extreme estimates.

4. The formula for the 95% confidence interval (CI) using t-distribution

where is the t-critical value for a 95% confidence level,

df (degrees of freedom) is calculated as N - 1

5. Entropy Measure for Categorical Diversity

The Shannon Entropy Measure is computed to quantify the diversity of income levels in the dataset:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Income Level** | **n** | **Estimated Proportion (up to 2 digits)** | **Standard Error (up to 3 digits)** | **95% CI (t-distribution)**  **(up to 3 digits) lower and upper** | **Entropy Measure (up to 3 digits)** |
| Middle |  |  |  |  |  |
| Low |  |  |  |  |
| High |  |  |  |  |