Eco 221-02 Statistics: Descriptive Statistics-Tabular and Graphical
Chun-Pin Hsu

A frequency distribution is a tabular summary of data showing the frequency (or number) of items in each of several nonoverlapping classes.

Example:

Guests staying at a hotel were asked to rate the quality of their accommodations as being ***excellent***, ***above average***, ***average***, ***below average***, or ***poor***. The ratings provided by a sample of 20 guests are:

Show: Frequency Distribution

The relative frequency of a class is the fraction or proportion of the total number of data items belonging to the class.

A relative frequency distribution is a tabular summary of a set of data showing the relative frequency for each class. The percent frequency of a class is the relative frequency multiplied by 100. A percent frequency distribution is a tabular summary of a set of data showing the percent frequency for each class.

Show: Relative Frequency and Percent Frequency Distributions

Bar Graph

A bar graph is a graphical device for depicting qualitative data. On one axis (usually the horizontal axis), we specify the labels that are used for each of the classes.

A frequency, relative frequency, or percent frequency scale can be used for the other axis (usually the vertical axis).

Using a bar of fixed width drawn above each class label, we extend the height appropriately.

The bars are separated to emphasize the fact that each class is a separate category.

Show: Bar Graph

Pie Chart

The pie chart is a commonly used graphical device for presenting relative frequency distributions for qualitative data. First draw a circle; then use the relative frequencies to subdivide the circle into sectors that correspond to the relative frequency for each class. Since there are 360 degrees in a circle, a class with a relative frequency of .25 would consume .25(360) = 90 degrees of the circle.

Show: Pie Chart

Summarizing Quantitative Data

Frequency Distribution, Relative Frequency and Percent Frequency Distributions, Dot Plot, Histogram, Cumulative Distributions, and Ogive.

.Guidelines for Selecting Number of Classes

- Use between 5 and 20 classes. Data sets with a larger number of elements usually require a larger number of classes. Smaller data sets usually require fewer classes. Use classes of equal width.
- Approximate Class Width =

$$\frac{\text{Largest Data Value} - \text{Smallest Data Value}}{\text{Number of Classes}}$$

For Auto Repair case, if we choose six classes:
Approximate Class Width = (109 - 52)/6 = 9.5 ≅ 10

Show: Frequency Distribution, Relative Frequency and Percent Frequency Distributions

Dot Plot

One of the simplest graphical summaries of data is a dot plot.
A horizontal axis shows the range of data values. Then each data value is represented by a dot placed above the axis.

Show:Dot Plot

Histogram

Another common graphical presentation of quantitative data is a histogram.
The variable of interest is placed on the horizontal axis. A rectangle is drawn above each class interval with its height corresponding to the interval's frequency, relative frequency, or percent frequency. Unlike a bar graph, a histogram has no natural separation between rectangles of adjacent classes.

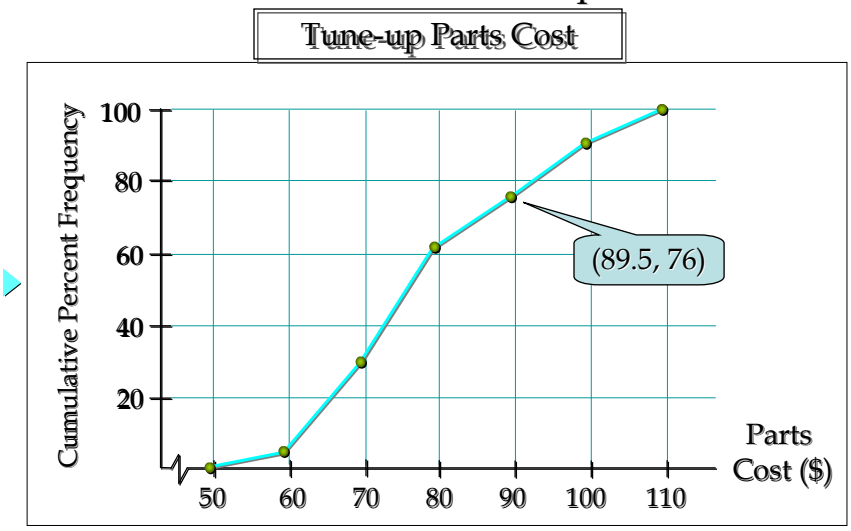
- Symmetric (Show)
  - Left tail is the mirror image of the right tail
  - Examples: heights and weights of people
- Moderately Skewed Left (Show)
  - A longer tail to the left
  - Example: exam scores
- Moderately Right Skewed (Show)
  - A Longer tail to the right
  - Example: housing values
- Highly Skewed Right (Show)
  - A very long tail to the right
  - Example: executive salaries

Cumulative frequency distribution - shows the number of items with values less than or equal to the upper limit of each class..
Cumulative relative frequency distribution – shows the proportion of items with values less than or equal to the upper limit of each class.
Cumulative percent frequency distribution – shows the percentage of items with values less than or equal to the upper limit of each class.
(Show)

Ogive

An ogive is a graph of a cumulative distribution. The data values are shown on the horizontal axis. Shown on the vertical axis are the cumulative frequencies, or cumulative relative frequencies, or cumulative percent frequencies. The frequency (one of the above) of each class is plotted as a point. The plotted points are connected by straight lines.

Ogive with
Cumulative Percent Frequencies



Exploratory Data Analysis
Cross tabulations and Scatter Diagrams

The techniques of exploratory data analysis consist of simple arithmetic and easy-to-draw pictures that can be used to summarize data quickly
One such technique is the stem-and-leaf display

Stem-and-Leaf Display

A stem-and-leaf display shows both the rank order and shape of the distribution of the data. It is similar to a histogram on its side, but it has the advantage of showing the actual data values. The first digits of each data item are arranged to the left of a vertical line. To the right of the vertical line we record the last digit for each item in rank order. Each line in the display is referred to as a stem. Each digit on a stem is a leaf.
(Show: Stem-and-Leaf Display)
If we believe the original stem-and-leaf display has condensed the data too much, we can stretch the display by using two stems for each leading digit(s). Whenever a stem value is stated twice, the first value corresponds to leaf values of 0 - 4, and the second value corresponds to leaf values of 5 - 9.
A single digit is used to define each leaf. In the preceding example, the leaf unit was 1. Leaf units may be 100, 10, 1, 0.1, and so on. Where the leaf unit is not shown, it is assumed to equal 1. **(Show: Examples)**

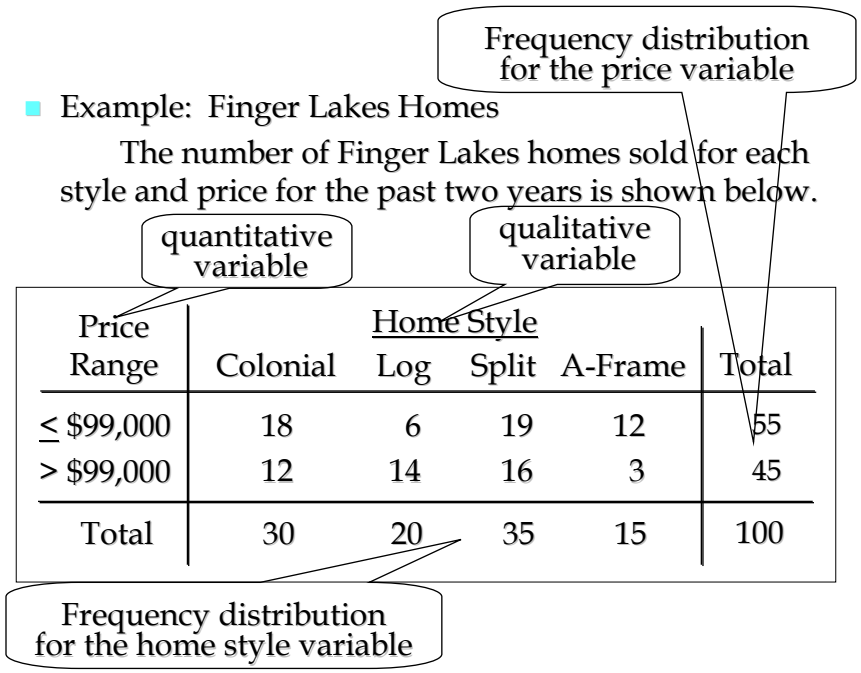
Crosstabulation

- A crosstabulation is a tabular summary of data for two variables.

ν Crosstabulation can be used when:

- one variable is qualitative and the other is quantitative,
- both variables are qualitative, or
- both variables are quantitative.

- The left and top margin labels define the classes for the two variables



- Example: Finger Lakes Homes

The number of Finger Lakes homes sold for each style and price for the past two years is shown below.

Price Range	Home Style: Colonial	Log	Split	A-Frame	Total
≤ $99,000	18	6	19	12	55
> $99,000	12	14	16	3	45
Total	30	20	35	15	100

Insights Gained from Preceding Crosstabulation

The greatest number of homes in the sample (19) are a split-level style and priced at less than or equal to $99,000.

Only three homes in the sample are an A-Frame style and priced at more than $99,000.

Crosstabulation: Row or Column Percentages

Converting the entries in the table into row percentages or column percentages can provide additional insight about the relationship between the two variables.

(Show Examples)

Crosstabulation: Simpson's Paradox

Data in two or more crosstabulations are often aggregated to produce a summary crosstabulation. We must be careful in drawing conclusions about the relationship between the two variables in the aggregated crosstabulation.

Simpson' Paradox: In some cases the conclusions based upon an aggregated crosstabulation can be completely reversed if we look at the unaggregated data suggests the overall relationship between the variables.