**Chapter 2.4 Creating Distributions From Data**

**Excel Tutorial**

Distributions help summarize many characteristics of a data set by describing how often certain values for a variable appear in the data set. Distribution can be created for both categorical and quantitative data, and they assist the analyst in gauging variation.

It is often useful to create a frequency distribution for a data set. A **frequency distribution** is a summary of data that shows the number (frequency) of observations in each of several nonoverlapping classes, AKA bins.

The data in the **SoftDrinks** file is a sample of 50 soft drink purchases. Each purchase is for one of five popular soft drink, which define the five bins: Coca-Cola, Diet Coke, Dr. Pepper, Pepsi, and Sprite.

1. **Create a Frequency Distribution in Excel**
   1. In D1 type “Bins”
   2. In D2 through D6 Type the name of the soft Drink Companies
   3. In E2 enter the formula: =COUNTIF($A$2:$B$26, D2)
   4. Copy the formula into cells E3-E6

**Observations: Coca-Cola is leader, then Pepsi, Diet Coke etc.**

**Why when we copied the formula the original data reference did not shift down the spreadsheet?**

Because of the absolute reference: $$ that kept the same reference. But the D column cells changed with the formula copies.

1. **Create Relative Frequency and Percent Frequency Columns**

**Relative Frequency:** the fraction or proportion of items belonging to a bin relative to the whole data set

**Percent Frequency:** Relative frequency as a percentage.

1. Title column F “Relative Frequency” and Column G “Percent Frequency”
2. In Cell E7 Type Total and Auto-sum the F column
3. In Cell C2 type formula =$E2/$E$7
4. Copy Formula down column F
5. In Column G copy cells over
6. Then make the column a percentage

**What are some possible uses of this distribution?**

Concession stand has an upcoming game or concert. If this data is representative of their customer population they can use it to purchase soft drinks for the game or concert. If they buy 1000 drinks total, then 380 of them should be coke. Etc.

1. **Frequency Distributions for Quantitative Data**

Steps: 1. Determine the number of nonoverlapping bins

2. Determine the width of each bin

3. Determine the bin limits

**Audit Time Data File – Records the number of days it took to complete an Audit**

1. (n = 20) so use about 5 bins
2. (Largest data value – Smallest data value ) / Number of bins

= (33 – 12) / 5 = 4.2 ~ 5 Bins

1. Bin Limits:
   1. 10-14
   2. 15-19
   3. 20-24
   4. 25-29
   5. 30-34
2. Type Bins in C1 and Frequency in D1
3. Type the upper limits in C2:C6: 14,19,24,29,34
4. Select Cells D2:D6
5. Type Formula =FREQUENCY(A2:A21, C2:C6)
6. Press CTRL+SHIFT+ENTER for an array.
7. **Histograms using Audit Time Data File**

Histograms are common graphical presentation of quantitative data. It can be used to summarize frequency, relative frequency, or percent frequency distributions.

The frequency measure of each class is shown by drawing a rectangle whose base is the class limits on the horizontal axis and the height is the corresponding frequency measure.

**Version 1**

* 1. Click the **Data** Ribbon
  2. Click Data Analysis in the Analyze group.
  3. Choose **Histogram** in the box
     1. **Input Range:** A2:A21
     2. **Bin Range:** C2:C6
     3. **Output Options:** New Worksheet Ply
     4. Select the Chart Box for **Chart Output**
     5. **Click Ok**

If you don’t see the **Data Analysis** button:

* 1. Click File
  2. Options
  3. Add-ins
  4. Manage: “Excel Add-ins -> Go
  5. Check off the Analysis Tool Pak or ToolPak – VBA

**Version 2 – From Frequency Table**

1. Select Data A2:B6
2. Click Insert Tab
3. Click Historgram Symbol
4. Click More Statistical Charts
5. Click Column and Click OK

**Version 3 – From Original Data**

1. Highlight data Cells A2:A21
2. **Clink Insert Tab**
3. Blue Histogram Symbol
4. **Click OK**
5. Right Click on Horizontal Axis Numbers
6. **Select** format Axis
7. **You have limited options for changing the format in this menu**

**How to reduce Spacing in-between bars of the histogram**

1. Right Click one of the histogram columns
2. Select Format Data Series
3. Click Series Option
4. Set Gap Width to 0%

**Discuss Skewness – Histogram Slide**

1. **Cumulative Distributions**

Uses the classes, class widths, and class limits from before but instead of the frequency we show the cumulative frequency – which shows the number of data items with value less than or equal to the upper class limit of each class.

* 1. Make another column in your frequency table called “Cumulative Frequency”
  2. In Row 2 of that column – type Formula: =SUM($D$2:D2)
  3. Then Drag down the square to populate the rest of the column
  4. Make another column and title it “Cumulative Relative Frequency”
  5. In Row 2 type formula = E2/20 Which is the cumulative frequency divided by total
  6. Drag down to populate column
  7. Create another column titled “Cumulative Percent Frequency”
  8. In Row 2 type formula: F2 \*100
  9. Drag down to populate Column

**These cumulative frequency columns help tell you information like 85% of the audits were completed in 24 days or less. Etc.**