

# CS130/230 Lecture 12

## Introduction to StatView

Thursday, March 11, 2004

### **Types of Data Analysis**

When doing data analysis, we are interested in two types of summaries:

- 1) Statistical Summaries (e.g. descriptive, hypothesis testing)
- 2) Visual Summaries (e.g. tables, graphs)

Statistics is sometimes broken up into two different areas:

- 1) Descriptive Statistics - a situation is described by the statistics by the collection, summarization, organization and presentation of data.
- 2) Inferential Statistics - where inferences are made from samples of the population (e.g. smokers smoking a pack of cigarettes per day have a higher cholesterol). In this area we get into Hypothesis testing.

In the Descriptive Statistics world, we are concerned about each of the following. Just give a general description of the meaning of each of the following terms:

- Mean
- Median
- Mode

Here is an interesting problem that Descriptive Statistics can help us get a handle on.

### ***Problem 1***

A paint manufacturer tested two experimental brands of paint over a period of months to determine how long they would last without fading. Here are the results:

Brand A	Brand B
10	25
20	35
60	40
40	45
50	35
30	30

What do the descriptive statistics tell us about the paint with regard to fading?

## Histogram

Let's see how good the random number generator in Excel really is.

### **Problem 2**

Import the random number file we created at the beginning of class into StatView and let's create a histogram of the random data. Make sure that you shut Excel down before you open the file in StatView.

**Part I:** Import the data.

**Part II:** Create a histogram of numbers. (1) Analyze Menu -> New View (2) Click on the Frequency Distribution Triangle (3) Select Histogram (4) Select Create Analysis and click ok (you do not to change any of the options at the moment (5) Select the random number from the variables box on the right. If you can't see the random number variable make sure that you have the correct dataset selected in the drop down box.

**Question:** Based on what you see, how good is the random number generator?

## Scatterplots of One Variable

Another type of graph is a Cell Plot. Cell plots are use to show the means for a variable of your choice split by some nominal variable.

### **Problem 3**

There is a sample data file called "Lipid Data". I would like you to take this file and produce a bar chart in the cell plot option showing the mean weight of the people in the file split by Gender. Also make a plot of the mean Cholesterol split by Gender.

These two plots really allow us to examine one variable of interest. What if we want to examine the relationship between two variables?

## Using More Than One Variable

In statistics, we can define two types of variables:

(1) independent - "it is what it is" and nothing influences it (e.g. Gender)

(2) dependent - most likely dependent on another variable (e.g. Cholesterol may be dependent on age)

### **Problem 4**

Consider the following table which shows the number of bushels of wheat produced for the given rainfall amounts:

Rainfall	2.5	3	4.5	7.6	9.5	10.3
Bushels	37	43	42	46	48	51

The rainfall amount is given in inches.

We want to plot this data onto a scatterplot (scattergram) and find a trendline that best fits the data. This is similar to the regression exercises that we did in Excel.

**Part I:** Create a new dataset and add the rainfall and bushel information to this dataset.

**Part II:** Select New View from the Analyze menu and go to Regression Plot under the Regression option. Select the simple option for the moment. This will draw perform linear regression. Determine which variable is the dependent one and which is independent and plot the data.

**Part III:** How many bushels of wheat will be produced if the rainfall amount was 6.2 inches?

**Part IV:** How much rainfall would we need to have to produce 60 bushels of wheat?