**DISPLAYING DATA UNIT PLAN**

***I. Unit Introduction***

Measure – A single number that represents some characteristic of an entire data set, or the relation of a datum to some such characteristic. Measures are a kind of statistic (or parameter.)

Some data displays require measures and some do not. In this unit we will look at displays that do not require measures. Sometimes this is called constructing from “raw” data. Later, once we have learned to calculate measures in the next few units, we will construct an additional display that relies on some of them.

We will be looking at two main types of display in this section: The **table** and the **graph**. (Some of the graphs are based on the tables.)

**Common symbols in this unit:**

“Big Sigma” ∑ – means “add up all the \_\_\_\_\_\_” or “sum of all the \_\_\_\_\_\_\_\_\_”

n – sample size

**Key Idea:** Each data display we will look at in this section is only appropriate for **one type of data: Qualitative or Quantitative.**

**Key Idea:** Every good data display has a title and appropriate labels!

**II. Displaying Qualitative Data**

**A.) Section II Introduction**

A simple way to display qualitative data is a basic **qualitative data table**. This table consists of three columns:

Type (of datum), for example “Eagles,” “Steelers,” “Ravens,” “Other,” “None.”

Frequency how many of each type was collected in the study. Its symbol is a “script f”: . The sum of this column should equal the sample size. That is:

Relative frequency - The proportion of the total frequency in each category. That is:

Give example by collecting sports data from the class. **Give a badly-worded survey question and ask students to improve it!**

**After the table is constructed, discuss how *rf* relates to Probability.**

We can use this table to construct the graph known as the **bar chart**, and, by adding another column, we can make the **pie chart** graph.

**B.) Bar charts** –

construct a bar chart from the sports data. **We can choose to make our y-axis either frequency or relative frequency.**

**Note:** A bar chart with a break in the y-axis is considered misleading. Try to avoid its use.

A special kind of bar chart is the Pareto chart, where the bars are arranged in descending order.

Arrange the sports data in a Pareto chart.

p. 53 Try it yourself #5. **These numbers look big – how do we construct an appropriate vertical axis**?

**C.) Pie charts** –

In a pie chart, we add a fourth row to the data table, the angle. That is:

for each type.

Construct a pie chart from the sports data.

p. 52 Try it yourself #4.

**Handout:** Pie and Bar Charts Answer Key (to the HW below)

**HW:** p. 58 #20-23; p.59 #29-30; p.108 #11-12; p. 111 #3

**III.) Displaying Quantitative Data**

**A.) Section III Introduction**

**- Quantitative data sets can be classified as either univariate or multivariate**.

**-** A data set is Univariate if each datum consists of a single measurement. (cf. multivariate)

- A special kind of multivariate set is the bivariate set – where each datum consists of a pair of numbers. This is a kind of paired data set. An example would be recording the temperature and the rainfall each day. Each datum would be a pair: (*temp., rainfall*)

**B.) Displaying Univariate Data Sets**

**1.) Basic Univariate Displays**

- **a.) Stem-and-leaf plots**:

[Collect some data from the class. Good questions include number of cousins, age of mother.]

**b.) Dot plots:**

p. 51 Ex 3, TIY 3

**HW:** p. 56 #1, #3-10; p. 57-58 #15-19; p. 107 #7-8

**2.) Frequency Distributions**

**-** We were able to construct stem-and-leaf and dot plots without any prior table, but **for univariate data sets**, the basis for most displays is the frequency distribution.

Class: An interval of fixed with used to organize a data set.

Frequency: The number of data that fell within a given class. Each class has a frequency.

Span: A set of classes in the same data set such that:

a.) All classes are the same width

b.) The classes do not overlap

c.) Each datum falls within exactly one class.

Frequency Distribution: A span of classes, their associated frequencies, and other relevant statistics. A frequency distribution is usually displayed as a table (called a frequency table), and includes the following columns:

a.) Class

b.) Midpoint\* (optional but useful later on)

c.) Tally\* (only included on the “clipboard table” of the person collecting the data)

d.) Frequency

e.) Relative Frequency

f.) Cumulative Frequency\* (optional)

To construct a frequency table, you must first find the **class width (w)**. This is done by dividing the **range** by the desired number of classes, then rounding up to the **NEXT** whole number. That is:

This is true EVEN WHEN *range/(desired # of classes) is already an integer. So if you get 4, you have to round all the way up to 5!!*

In case you have forgotten, range is found by taking the highest-valued datum and subtracting from it the lowest-valued datum. That is:

The start point (or **lower class limit**) of the first class is the minimum datum. The start point of each subsequent class is the start point of its previous class’ start point plus the class width.

p. 35 Example 1 (but we’re going to do it a little differently than in the book!)

Now let’s look at some of the “optional” columns on the frequency table. From the previous section we’re already familiar with how to find **relative frequency**. In case you’ve forgotten:

To find the **midpoint** of the first (lowermost) class, we take the lower class limit, add to it the lower class limit of the subsequent class, and divide by two.

Find each subsequent midpoint by adding the previous midpoint to the class width.

The symbol for midpoint is **x**. This is because it can serve as a “synthetic” datum as we will see.

To summarize in equation form:

To find the **cumulative frequency** of a class, take the frequency of that class and add to it the frequency of all previous (lower) classes. Think of it like a “vacuum cleaner” sweeping up frequencies as it goes along. That is:

p. 37 Example 2 (but we’re going to do it a little differently than in the book!)

**HW:** Frequency Tables Worksheet #1, questions #1-10, #14-23

**3.) Displays based on Frequency Distributions**

**a.) Histogram**

A Histogram is a lot like a bar graph, except:

1.) The x-axis is the range of the data set.

2.) The bars must touch, to emphasize the **continuity of quantitative data**.

You may put a break on the horizontal axis, but not the vertical axis.

The y-axis can be either *f* or *rf*.

Let’s make a Histogram from the Examples 1 & 2 data from yesterday.

**b.) Frequency Polygons**

A frequency polygon is a line graph connecting all the points of the form

Plus two additional points:

and

Where “F” is taken to denote the final (uppermost) class.

This ensures that the line graph is connected to the x-axis at both ends. This shape is a polygon and can be useful in calculating **probabilities**.

Again, let’s use our data from Examples 1 & 2 to make a frequency polygon.

**Discuss how this can be used to find probabilities**.

**c.) Ogives**

An ogive (pronounced: “O-jive”) is a line graph connecting all the points of the form

Plus one additional point: . This ensures the ogive is anchored to the horizontal axis. This is also useful to calculate probabilities

**HW:** Remaining questions on Frequency Tables Worksheet #1; All of F.T. Worksheet #2

**C.) Displaying Bivariate (Paired Data) Sets**

Concept of **correlation** (rho) Direct, inverse and independent relationships

* independent vs. dependent variables
* Scatter plots
* definition of a time series (top of p. 55)
* Time series charts

**Handout:** Scatter Plots and Time Series Answer Key (to the HW below)

**HW:** p. 59 #25-28, p.107 #9, p. 108 #10

**IV. Unit Review**

**Handout:** Displaying Data Unit Outline

**HW:** Displaying Data Worksheets #1 & #2

Unit Test