

**AP STATISTICS**  
**Summer Assignment 2009**  
**Mr. José Colon**

Carefully read and answer the questions in this document.

On page 6, you will see a TI SKILL assignment. Use of a calculator will be an integral part of this course. Instructions are given for the TI83 – TI84 series. You need to become familiar with entering and editing variable lists.

If you have any other calculator, I will be glad to help you if you bring the reference book that comes with your calculator. All class instruction is given on the TI 84-plus. The TI 84 has the same screen and operations as the TI-83.

There is a graded homework assignment on the last page of this document.

This assignment is DUE the first day of class in August 2009.

Late work will not be accepted.

## DATA

Data *are* what we manipulate in Statistics. Data *are* the information in the Information Age.

What are data? First of all, data are plural. One piece of information is a datum. Two or more are data.

When you look at data, you might see something like this:

NAME	PHONE	GR	English	Math	GPA	ZIPCODE
Smith, John	699-2800	11	Eng III CP	Algebra II	3.92	29223

This list of data about John Smith tells us some information.

I can see that the name is John Smith, his phone number, his grade, two classes he is in, his GPA and his zip code.

This list of variables about one person is called a **record**.

A record of just one person isn't particularly interesting unless you know that person. You might be more interested in a whole list of records.

NAME	PHONE	GR	English	Math	GPA	ZIPCODE
Smith, John	699-2800	11	Eng III CP	Alg II	3.92	29223
Jones, Maria	699-2801	12	Eng IV H	PreCalc	4.05	29223
Anders, Andy	699-2802	10	Eng II CP	Alg II	2.23	29223
Bald, Beth	699-2803	10	Eng I	Geo CP	2.01	29223
Cold, Connie	699-2804	12	Eng IVH	AP STAT	7.24	29069

Now that we have a lot of data, we need to know something about that data. We ask the question, "WHAT ARE THE DATA?"

We can answer this using the 5Ws: **Who, What, When, Where, Why**. We also implicitly include an "H" (**How**) with this group.

When we examine data, we want to be able to answer the questions about the data:

**Who** are the data about?

**What** was the data that was collected?

**When** was it collected?

**Where** was it collected?

**Why** was it collected?

**How** was it collected?

Answering these questions can provide a **CONTEXT** for the data. If we can't answer ALL of these questions, then we don't have good data. We might have useless information.

Identify the WHO, WHAT, WHEN, WHERE, WHY and HOW for the following situations.

In a study appearing in the journal SCIENCE, a research team reports that plants in southern England are flowering earlier in the spring. Records of the first flowering dates for 385 species over a period of 47 years indicate that flowering has advanced an average of 15 days per decade, an indication of climate warming according to the authors.

WHO:

WHAT:

WHEN:

WHERE:

WHY:

HOW:

Scientists at a major pharmaceutical firm conducted an experiment to study the effectiveness of an herbal compound to treat the common cold. They exposed each patient to a cold virus, then gave them either the herbal compound or a sugar solution known to have no effect on colds. Several days later, they assessed each patient's condition using a cold severity scale ranging 0–5. They found no evidence of the benefits of the compound.

WHO:

WHAT:

WHEN:

WHERE:

WHY:

HOW:

Answers:

Flowers:     Who   385 species of flowers.  
                  What   date of first flowering  
                  When   not specified  
                  Where  Southern England  
                  Why   To find indications of a warming of the overall climate  
                  How   not specified.

Cold remedy   Who   experiment volunteers  
                  What   herbal cold remedy or sugar solution and cold severity  
                  When   not specified  
                  Where  Major pharmaceutical firm  
                  Why   To test efficacy of herbal remedy on common cold  
                  How   The scientists set up an experiment

Now that you've got a general idea of the 5W's, lets look at them a bit more closely.

## WHO

Who is not who is doing the study. Who is the person or thing being examined. Sometimes they are called respondents in a survey or subjects or participants in a study. A generic term for the *who* is **experimental unit**. Be sure you know who the *who* are or you may not know what the data are telling you.

## WHAT

The characteristics of the *who* that you observe are called variables. Each variable has a name identifies what has been observed or measured.

There are two types of variables.

- **CATEGORICAL:** characteristics or names. Examples include hair color, gender, breed of a pet dog.
- **QUANTITATIVE:** Variables recorded in numbers that are actually used as numbers (it makes sense to add, subtract, multiply and divide them). Examples include height, weight, yearly income, age.
  - **UNITS:** Quantitative variables have UNITS. Height is in inches, weight is in kilograms, yearly income is in euros and age is in minutes. Units need to be specified with quantitative variables.

## **CAUTIONS:**

All numbers do not indicate quantitative variables: it makes no sense to do mathematical operations on zip codes. Zip Codes are numbers, but they are CATEGORICAL in nature.

**Age:** sometimes variables can be deceptive. If I ask your age in years or days, then it would be a quantitative variable. I could get an average age of all the students in a class. This would make sense. However, I could also take AGE as a categorical variable if I separated groups into categories “Child”, “Teenager”, “Young Adult,” and “Adult.”

**Scales:** Have you ever taken a survey where the answers to a statement are STRONGLY AGREE , AGREE, DISAGREE, and STRONGLY DISAGREE ? You could assign values of 4, 3, 2 and 1 to these responses. They do measure some sort of order, but are also categorical in nature. These are called **ORDINAL** variables. It depends on what type of analysis you want to do with them as to whether you want to consider them quantitative or categorical.

## **Identify the following variables about YOU as categorical or quantitative.**

last name	hair color	height	phone number	cell phone number
shirt size	siblings	gender	birth order	grade point average

Reconsider the two problems you had earlier.

Read the problems again and identify the variables. Indicate whether the variables are quantitative or categorical. If the variables are quantitative, include the units.

In a study appearing in the journal SCIENCE, a research team reports that plants in southern England are flowering earlier in the spring. Records of the first flowering dates for 385 species over a period of 47 years indicate that flowering has advanced an average of 15 days per decade, an indication of climate warming according to the authors.

Scientists at a major pharmaceutical firm conducted an experiment to study the effectiveness of an herbal compound to treat the common cold. They exposed each patient to a cold virus, then gave them either the herbal compound or a sugar solution known to have no effect on colds. Several days later, they assessed each patient’s condition using a cold severity scale ranging 0-5. They found no evidence of the benefits of the compound.

## ANSWERS:

Identify the following variables about YOU as categorical or quantitative.

Categorical: last name, hair color, both phone numbers, gender.

Quantitative: height (inches), siblings (units), birth order, grade point average.

Shirt size? Depends on how you interpret this one. SMALL, MEDIUM, LARGE would be categorical. For men's shirts, a "15" would mean a 15 inch neck and would be quantitative.

Birth order. ORDINAL. Depending on how you are going to use it, this variable can either be quantitative or categorical.

Flowers:

Variables – Date of first flowering Quantitative Units: Days

Cold Treatment

Variables – Type of treatment – categorical

Severity of cold rating – quantitative (scale of 0–5) possibly ordinal.

Consider the following problem. Identify the 5W's and H and identify all variables and variable types.

**A study compared the lifetimes of actors who had won Academy Awards (Oscars) with those of actors who had been nominated but had not won and with actors who had been in the movies that led to the awards but had not been nominated. The study found that actors who had won tended to live longer than the other actors considered.**

## WHEN

The WHEN of the data is the time it was gathered. Knowing the date of the survey can make a great deal of difference. An opinion survey about oil prices in 1999 would have a significant difference from an opinion survey about oil prices in 2006. If this information is not given, indicate as such.

## WHERE

As the WHEN is significant, so is the where. Opinions gathered in South Carolina might be different than opinions gathered in North Dakota.

## WHY

Often the most revealing of these questions is WHY was the data gathered. If you're trying to see if a new medication is effective and you're the FDA gathering the data, you have a different perspective than if you're the drug company making the medication. Be skeptical. Knowing the motivation for the data collection has much to do with interpreting the statistics we will compute later in the course.

## How

How was the data gathered? You might get different answers on an anonymous questionnaire than by asking the same questions face-to-face. Are you gathering your own data, or are you relying on records that someone else made a decade ago. Did I ask my friends their opinion? Or did I ask the opinion of randomly selected people?

## WHAT CAN GO WRONG?

- Just because your variable has numbers doesn't mean it's quantitative.
- Always be skeptical. Make sure you know what the research is about.

**TI Skills:** Enter values of variables into Lists in the calculator.

Consider the height (in inches) of a basketball team's players:

63	75	71	71
73	74	63	61

**TO ENTER DATA:** Press the STAT button. Choose EDIT from the horizontal menu. You'll see a row of column names L1, L2, L3, and so on. Move the cursor to the first position under the L1.

Begin to type in the numbers, one datum at a time, followed by the ENTER key (or the down arrow). Enter all 8 data.

**TO CHANGE A DATUM:** The Second 63 should really be a 62. To change it, move the cursor onto the second 63 and re enter the new value, 62, followed by an ENTER.

**TO DELETE A DATUM:** Highlight the datum you wish to delete and press the DEL key.

**TO CLEAR THE DATALIST:** Move to cursor atop the L1. Hit CLEAR, ENTER. Everything should disappear.

**TO RECLAIM THE DATALIST:** STAT – EDIT – SetUpEditor.

### **Graded Homework Assignment:**

Through your own research (Internet, library, etc.) find an ABSTRACT for **two** different studies. Try to find studies in an area that has interest to you. A copy of the Abstract should be followed by an identification of the 5W's and H, the variables and the variable types with units.

This will be your first grade in AP STATISTICS and is due on the first day of class. Work to be submitted should be neat and legible and should include a cover page with your name, date and title of your work.

You will be asked to write the Richland School District Honor Code on all submitted work for this course:

“On my honor, I have neither given nor received unauthorized assistance on this assignment. I understand that any violation of the RSD2 honor code will result in academic and disciplinary action.”