M 358K TERM PROJECT DESCRIPTION

The purpose of your project is to experience the entire statistical process of studying a question, including the following aspects:

1. Choosing the question and formulating it carefully.

2. Designing the data collection procedure.

3. Collecting the data.

4. Analyzing the data

5. Writing up your results, including the method of data collection, the data obtained, any difficulties encountered in data collection, your analysis, and any interpretations or conclusions that can be drawn from your data and analysis.

You should do your project in a group of three or four students. (Individual projects will be permitted only in unusual circumstances.)

Your question should meet the following three criteria:

1. It falls into one of the classifications I - VII below.

2. It can be studied by inference from a sample to the entire population. In particular, the population you are studying must be large enough so that you can choose a sample meeting the criteria listed below for the type of question involved. Also, you should be willing and able to choose a sample that is reasonably random.

Types of questions you may choose to study:

Caution: The examples given here are meant only to illustrate the type of question; they are not necessarily as precisely formulated as your project question will ultimately need to be.

(The chapter given in parentheses is the chapter of the book where analysis of the type of question is discussed.)

I. What is the mean of some quantitative variable? (Chapter 7)

Examples:
    a. What is the average amount of sleep a UT student gets on Sunday night?
    b. What has been the average increase in U. S. college and university tuition and fees in the past year?

II. How do the means of a quantitative variable in tow populations compare? (Chapter 7)

 Examples:
    a. Do male or female students study more on average?
    b. On which of your two favorite running courses is your average speed higher?

III. What is the proportion of a population that has a certain characteristic? (Chapter 8)

   Example:
    a. What proportion of UT students recycle their Daily Texan?
    b. Is the person answering a random phone call more likely to be male or female?

IV. How do the proportions of two populations having a certain characteristic compare? (Chapter 8)

Examples:
    a. Are you better at making three-point jump shots from in front of the basket or from the left corner?
    b. If you asked people about their opinions on banning X-rated movies, would more favor banning if you asked, "Do you think the government should forbid the showing of X-rated movies?' or if you asked, "Do you think the government should allow the showing of X-rated movies?"

V. (Chapter 9) A. How do the proportions of more than two populations having a certain characteristic compare?
     or
  B. How do the proportions of two or more populations falling into certain categories (three or more) compare?

   Examples:
    a. Do a person's smoking habits depend on their parents' smoking habits? In other words, is the proportion of smokers the same or different depending on whether both parents smoke, only one parent smokes, or neither parent smokes? (Variant A -- three populations)

b. Do the proportions of male and female students earning A's, B's, C's, D's, and F's differ significantly? (Variant B -- two populations and five categories)

VI. What is the relationship, if any, between two quantitative variables? (Chapter 10)

   Examples:
   a. What is the relationship between steps per minute and running speed?
   b. What is the relationship between the amount of nitrogen fertilizer applied and height of a plant?

   (This type of question is about regression, but you will need to use the techniques in Chapter 10 as well as the techniques we've already studied. Depending on what your data show, you might need to go beyond what we do in class for this type of question to transform your data so the regression methods fit.)

   All of the above types of questions can be studied using techniques we will cover in this course. Another type of question that can be studied using techniques in the book that we won't study in class is:

VII. How do means of a quantitative variable for more than two populations compare? (Chapter 12)

   Examples:
   a. Do compact, midsize, or large cars get better city gas mileage?
   b. What colors work best for insect traps?

You may choose a question for your project of this type, *but if you do, you will need to learn the material in Chapter 12 on your own and use it in your analysis*. Also, you might need to transform your data so that the appropriate assumptions for the test are satisfied.

Criteria for population and sample size

   Your population and sample sizes will need to meet the following guidelines, depending on the type of question you are studying. *These are only rough guidelines; there may be more considerations depending on your specific question.* The guidelines below are labeled by the same Roman numeral used for the corresponding question type above.

I. (What is the mean of some quantitative variable?) The sample size should be at least 15, unless the data are strongly skewed or have outliers, in which case the sample size should be at least 40.

II. (How do the means of a quantitative variable in two populations compare?) Each samples should be at least 8, unless the data are strongly skewed or have outliers, in which case the sample size should be at least 20. It is best to have sample sizes approximately equal for the two groups.

III. (What is the proportion of a population that has a certain characteristic?) The population should be at least 10 times the size of the sample. The sample size should be at least 30. Each subgroup in the sample (e.g., students in the sample who recycle their Texans, and students who don't) should be at least 10.

IV. (How do the proportions of two populations having a certain characteristic compare?) Both populations should be at least 10 times as large as the sample taken from them. Both samples should be at least 30. Each subgroup in the sample (e.g., number of shots made from in front of the basket, number missed from the corner, etc.) should be at least 5.

V. (Comparing proportions with more than two populations or more than two categories.) The sample should be at least five times as large as the total number of subgroups considered (number of populations times number of categories).

VI. (What is the relationship, if any, between two quantitative variables?) The sample size should be at least 15. However, if the residuals are skewed, your sample should be at least 40; fifty would be better.

VII. (How do several means compare?) Unless you have good reasons to believe that the data for each group are normally distributed with the same standard deviation, you should try to get sample sizes of at least 30 for each group.

Start thinking about questions you might want to study for your project. Ask yourself if they fit one of the categories above.
- Identify clearly what your population and what your variables are.
- Think about whether you need to do an experiment or observational study (see Sections 3.1 - 3.3), and how you can choose your sample (see Sections 3.1 - 3.3).
- Discuss your questions with classmates. Start forming groups to work on projects. (Note: You can send email to the whole class from Blackboard.)

Things to consider when deciding on your project topic:

Choosing a project group:
- Make sure everyone in your group is interested in the topic and has common goals. For example, if you want to try to do an A project, you're going to have conflicts if other group members only care about just passing.
- Be sure there are times you can meet together to work on your project.
- Be sure to exchange email addresses and phone numbers.

Ethical guidelines:
- If your project involves human subjects, the participants must be adults (18 or over) who are in a position to freely give informed consent.
- The study should not subject the participants (or the researchers) to any foreseeable physical or psychological risk.
- If your project involves collecting survey data, you do not need to obtain written informed consent, but you should ask potential participants if they are willing to participate in a survey about [topic] which will take about [time], and the participants should be told that their responses will be confidential (or anonymous if that is possible). You must respect this confidentiality or anonymity. (*Confidential* means you will not tell anyone else what a subject said or did; *anonymous* means you will not know which subject said or did what. Be sure not to tell subjects that their responses will be anonymous if in fact they will not be.)