Session 6

Designing Experiments

Key Terms for This Session

Previously Introduced

- bias
- box plot
- distribution
- Five-Number Summary
- interquartile range
- mean
- median

New in This Session

- comparative experimental study
- comparative observational study
- comparative study
- design of a comparative study
- random assignment
- treatment

Introduction

The statistics problems in previous sessions primarily seek to answer questions about a single group. For example, our analysis and interpretation of "How many raisins are in a box of raisins?" typically focused on describing the number of raisins for one brand of raisins. Until now, our primary focus has been on describing characteristics of a single distribution by organizing data in graphs and tables and by determining numerical summary measures for data.

In this session, we will investigate the idea of comparative studies—studies that focus on comparing several different distributions. The analysis and interpretation of data from comparative studies depend on the method of data collection, which can be either experimental or observational. **[See Note 1]**

Learning Objectives

The goal of this session is to understand the concepts involved in comparing two or more groups. In this session, you will do the following:

- Investigate comparative studies, which can be experimental or observational
- Learn how to analyze and interpret results from comparative studies
- Learn how to design a comparative experiment
- Explore paired and unpaired comparative studies

Note 1. In this session, we are analyzing experimental results informally. More standard statistical methods use formal inference procedures to make generalizations from the results of experiments.

Part A: Comparative Studies (15 min.)

Many statistics problems require you to make comparisons. You may be interested in comparing the salaries for different professions, for example, or you might want to compare test scores of children in different reading programs. Comparative studies like this fall into two categories—experimental studies and observational studies. **[See Note 2]**

The purpose of a comparative *experimental* study is to determine the "cause and effect" of an action. In this situation, an experiment deliberately imposes a particular treatment on a group of individuals in order to gauge their responses. This allows the investigator to determine if the treatment caused a change in the individuals' responses.

For example, to determine the effect of taking aspirin on heart disease, an experiment was conducted using a large group of doctors as subjects. Half the doctors took aspirin every other day; the other half took a pill that looked and tasted like aspirin, but was not. The experimenters imposed the treatment (aspirin) on the individuals by deciding through random selection which half would take aspirin and which half would not. After several years, the incidence of heart attacks for the two groups was compared. The analysis of the results suggested that the aspirin dosage had a positive effect—that is, there was a lower rate of heart attacks among the doctors who received aspirin.

The purpose of a comparative *observational* study is to determine if there is a difference in measurements between groups of individual subjects. An observational study does not impose a treatment on the subjects—it observes them as they are. These types of studies do not allow the investigator to determine whether or not a treatment caused a change in the subjects' responses—since there is no treatment!

For example, studies of the effects of smoking on people are necessarily observational. For many reasons, we cannot select a group of subjects and force them to smoke for a long period of their lives and then select another group of subjects and not allow them to smoke. Instead, we would select a group of people who are smokers, select another group who are not smokers, and compare the health characteristics of the two groups.

The data for comparative studies like these (either observational or experimental) consist of at least two sets of measurements. Although there may be more than one variable in a study, we will restrict our attention to the analysis of data collected on one variable for now. We will use Five-Number Summaries and comparative box plots to analyze and interpret data from several different comparative studies.

Problem A1. Suppose a professor wanted to decide whether courses taught online are as effective as more traditional methods of instruction. The professor divides a class into two groups. One group receives traditional instruction, while the other takes the course online. At the end of the course, each group takes the same comprehensive exam.

a. Is this an experimental study or an observational study? Explain.

b. What treatment is under study?

Problem A2. Suppose you were designing a study to answer the question "Do people who choose to take a two-week vacation in the winter feel more positive upon returning to work than people who choose to take a two-week vacation in the summer?"

a. Describe how such a study might be designed.

b. Would this be an experimental study or an observational study? Why?

c. What treatment is under study?

Note 2. You will investigate the distinction between an observational study and an experimental study. Several examples are presented to illustrate the difference between these two types of studies. Group discussion should focus on how the nature of *experimental* studies allows us to make inferences about cause and effect. In contrast, an *observational* study can determine if there is a difference in measurements between two groups of individual objects but cannot conclusively attribute this difference to any particular factor.

Part A, cont'd.

Problem A3. Suppose you were designing a study to answer the question "Do people with diets high in fiber, fruits, and vegetables have a lower risk of colon cancer?"

a. Describe how such an experiment might be designed.
b. Could this be an experimental study? What problems might you encounter if you wanted to perform an experimental study to answer this question?
c. What treatment is under study?

Part B: Comparative Observational Studies (35 min.)

A New Raisin Question

Let's begin our investigation of comparative observational studies by returning to the raisin problem from Session 2. In that session, your investigation focused on a particular brand of raisins. In this session, you will return to the issue of comparing two brands of raisins: When the weights of the boxes are the same, how do the numbers of raisins in each box compare between the two brands?

Ask a question:
How do the numbers of raisins in boxes of Brand C and Brand D compare?

Collect appropriate data:
We counted 28 boxes of Brand C raisins and 36 boxes of Brand D raisins. Here are the ordered raisin counts for boxes of Brand C raisins:

25	25	25	26	26	26	26	26	27	27	27
28	28	28	28	28	28	28	28	28	29	29
29	30	30	31	32	32					

Here are the ordered raisin counts for boxes of Brand D raisins:

23	24	25	25	25	27	27	27	27	27	27	27
27	28	28	29	29	29	29	29	29	30	31	32
32	33	33	33	34	34	35	35	35	36	36	38

Part B, cont'd.

Analyze the data:

Here are the mean and median counts for each brand:

	Mean	Median
Brand C	27.9	28
Brand D	29.9	29

According to these data, Brand D typically has a few more raisins than Brand C. On average, Brand D has two more raisins than Brand C, and the median number of Brand D raisins is one more than the median number of Brand C raisins.

Based on the means and medians, you might conclude that the number of raisins in a box is about the same for both brands. Although it is useful to look at the means and medians, there are other aspects of the distribution you might want to consider.

Problem B1. Why is this raisin study observational as opposed to experimental?

Using Five-Number Summaries and Box Plots

Comparing two sets of measurements is not quite as simple as comparing two numbers. Because we are comparing a set of 28 measurements for Brand C with a set of 36 measurements for Brand D, any comparison must be based on percentages and not absolute frequencies. A comparison of the Five-Number Summaries is useful, since these quantities divide the ordered data into four groups, with approximately 25% of the data in each group. Here are the Five-Number Summaries for these data: **[See Note 3]**

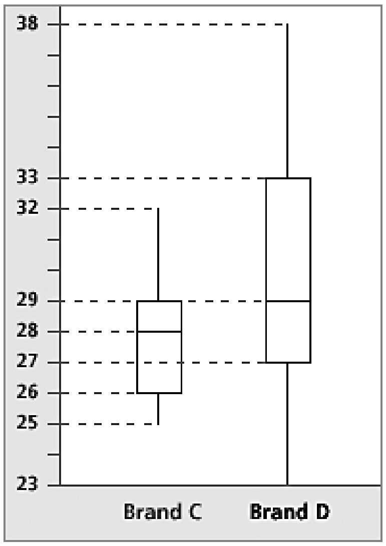
	Min	Q1	Med	Q3	Max
Brand C	25	26	28	29	32
Brand D	23	27	29	33	38

Note 3. Most people do not have difficulty comparing a single number with a single number, for example, noting that the median of one set of counts is greater than the median of another, or comparing one upper quartile with another. Some people, though, may have difficulty in comparing the distribution of one set of counts with the distribution of another.

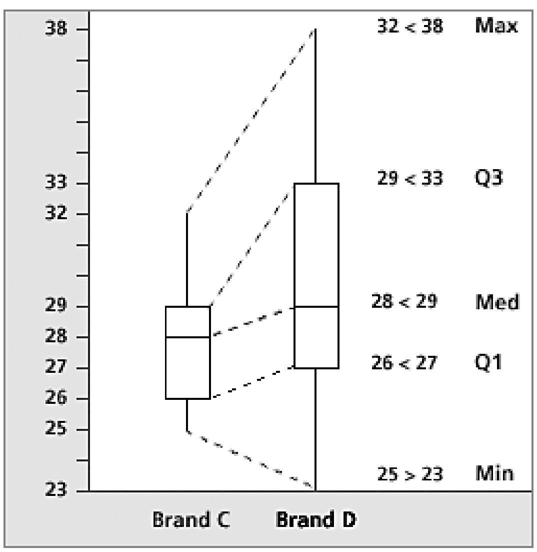
To compare one Five-Number Summary with another in the proper way requires a composite comparison of five numbers to five numbers; you must think beyond single-number comparisons. The box plots help to clarify this comparison, especially the variation within a group as indicated by the range and the interquartile range.

Part B, cont'd.

Here are the comparative box plots for these data:



You might start by comparing the actual values in the Five-Number Summaries. This will tell you where one set of measurements is located relative to the other set:



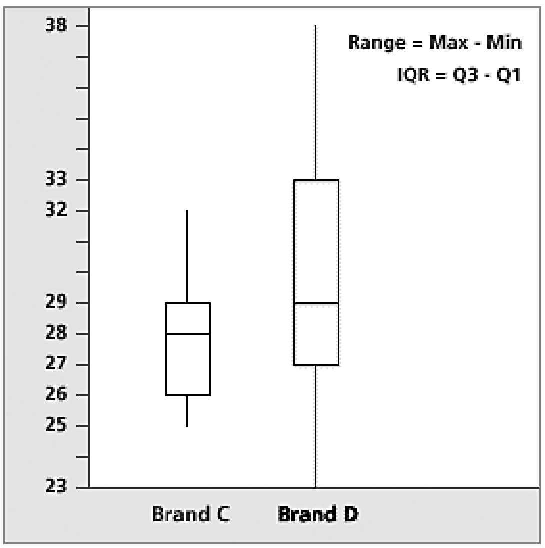
Note that with the exception of the minimum values, all summary measures for Brand D are higher than for Brand C. This suggests that boxes of Brand D tend to have more raisins than boxes of Brand C. In fact, since the third quartile for Brand D is greater than the maximum for Brand C, more than 25% of the boxes of Brand D have more raisins than *any* boxes of Brand C.

Part B, cont'd.

The Interquartile Range

Your comparison of the two sets of measurements should also consider the degree of variation within each set. This comparison can be based on the *range* of all the data (Max - Min) as well as on the range of the middle half of the data (which is called the Q-Spread or Interquartile Range, or simply the IQR), which is found by subtracting Q1 from Q3.

Let's look at the box plots again, and then calculate the range and the IQR:



	Range (Max - Min)	IQR (Q3 - Q1)
Brand C	32 - 25 = 7	29 - 26 = 3
Brand D	38 - 23 = 15	33 - 27 = 6

Based on the comparative box plots, there is more variation in the raisin counts for Brand D raisins than for Brand C raisins. The values for the ranges and IQR confirm this (Range C = 7, Range D = 15; IQR C = 3, IQR D = 6). Both the range and the IQR for Brand D are at least twice the range and the IQR for Brand C.

Consequently, although Brand D tends to have more raisins per box than Brand C, the smaller range and IQR for Brand C tell us that Brand C is more consistent than Brand D. Since the weights of boxes are the same, this would also suggest that the sizes of the raisins vary less for Brand C.

Part B, cont'd.

Problem B2. Brand A raisins come in boxes of the same weight as Brands C and D. Here are the ordered counts for 30 boxes of Brand A raisins:

23	25	25	26	26	26	26	27	27	27	27
28	29	29	29	30	30	31	31	31	32	32
32	33	34	34	35	35	36	39			

Compare the counts for Brands A and D. Make sure you consider where the data are located and the degree of variation. (You may have already determined the Five-Number Summary in Session 4.)

Problem B3. Brand B raisins come in boxes of the same weight as Brands A, C, and D. Here are the ordered counts for 27 boxes of Brand B raisins:

17	22	24	24	25	25	25	25	26	26	26	26
26	26	27	27	27	27	28	29	29	29	29	29
29	30	30									

Compare the counts for three brands: A, B, and C. Make sure you consider where the data are located and the degree of variation. (You may have already determined the Five-Number Summary in Session 4.)

Part C: Comparative Experimental Studies (65 min.)

Measuring Short-Term Recall

Now we'll move on to an investigation of comparative experimental studies. We'll begin with a problem related to human memory and, in particular, short-term recall. **[See Note 4]**

Ask a question:

Is it easier to remember strings of characters that spell familiar words than to remember arbitrary strings of characters?

Which list do you think would be easier for most people to remember—words like those in List A, or character strings (non-words) like those in List B?

List A: Words	List B: Non-Words
BOSTON	MZAPDR
EAR	CTG
CART	OXCS
BUG	AEA
PAPER	SKEOC

Collect appropriate data:

In order to explore this question, you'll need appropriate data—and to get this data, you'll need a method of measurement. **[See Note 5]**

Note 4. The activity in Part C provides an opportunity to consider the full statistical problem-solving process, with the primary focus on collecting the data. Useful data depend on an appropriate measurement/collection design, and you will have the opportunity to devise your own method of measurement. The activity asks you to consider several design options, and, in doing so, you will investigate such important ideas as randomization, blocking (pairing), and the effects of order.

After completing the memory experiment in Part C, consider conducting your own memory experiment to determine whether your results confirm the results in our example. One option is to use the Interactive Activity in Session 1, Part C, and reexamine the distance perception phenomenon. A paired design similar to the one used for the memory experiment could be used to compare subjects' length perception of line segments with arrows to their length perception of line segments without arrows.

Fathom Software, used by the onscreen participants, is helpful in creating graphical representations of data. You can use Fathom Software to complete Problems C7-C8, as well as Problem H1. For more information, go to the Key Curriculum Press Web site at http://www.keypress.com/fathom/.

Note 5. Measurement is the most important part of the statistics problem-solving process. Poor measurement will certainly produce poor conclusions! Most introductory statistics books or courses do not put a major emphasis on measurement. In this course, we encourage you to take some time to focus on measurement; this activity is good place to do so.

Part C, cont'd.

Problem C1. Describe the methods you might use to measure a person's ability to recall words and to recall non-words. Your description should be as specific as possible and include enough detail so that other people could follow your instructions well enough to perform the measurement themselves. Keep in mind that you want to use the results to make a fair comparison of the two sets of data you collect.

Problem C2. Consider your answer to Problem C1. Are there any ways in which the measurement process you described is biased? If so, think about ways you might try to remove this bias. **[See Tip C2, page 183]**

An Experiment

Here is one way to do the measurement.

Two lists are used:

1. List A contains 20 words of three characters each.
2. List B contains 20 non-words of three characters each.

To measure a person's ability to remember words (or non-words), each subject follows these steps:

1. Study the list for two minutes.
2. Pause for 20 seconds.
3. In one minute, list all the words or non-words that the subject can recall.

In this version of the experiment, we will show the list of non-words (List B) before the list of words. The subject's final score is the number of words (or non-words) he or she was able to correctly recall.

To perform this experiment, you'll need a subject and a timekeeper.

- Show the subject the list of non-words (List B). He or she has two minutes to study the list.
- When two minutes have elapsed, take the list of non-words away from the subject and wait for 20 seconds.
- The subject then has one minute to recall and write down non-words from the list.
- Finally, count the number of correct non-words the subject recalled.

Repeat the same sequence for words (List A).

List B: Non-Words		**List A: Words**	
MZA	LAZ	GOT	MOP
CTG	OSL	EAR	EAT
OXC	TNM	SAT	CUB
VLO	UMN	NOW	LOG
AVE	HIR	BUG	JOT
RPK	NYB	HOT	LOT
GJQ	FKP	NAP	NOD
DSH	ERI	CAR	TOP
IHS	PWD	DOT	SAD
QLJ	BUF	FOG	RUN

Try It Online! **www.learner.org**

This problem can be explored online as an Interactive Activity. Go to the *Data Analysis, Statistics, and Probability* Web site at www.learner.org/learningmath and find Session 6, Part C.

Part C, cont'd.

Problem C3. Think about the experiment design you described in Problem C1. Does the process you just went through (studying non-words, then writing non-words; studying words, then writing words) present any additional biases? How might you resolve any biases in your experiment design? **[See Tip C3, page 183]**

Experimental Design

You would not want to make conclusions about memory by examining only one person. Therefore, you should use more than one subject in this experiment.

Let's assume that you will use 16 subjects in your experiment. You will need to make some decisions about how to measure short-term recall in your 16 subjects.

Recall the original question: "Is it easier to remember strings of characters that spell familiar words than to remember arbitrary strings of characters?"

As stated, the question is perhaps not as specific as it should be. For example, we have not clarified the population of people we are studying. Age may have an impact on a person's ability to memorize. Are we interested in adults only?

If you believe that age makes a difference in a person's recall ability, then perhaps you need to refine the question.

Below are five different ways that you might assign the 16 subjects to groups and collect your measurements. These are referred to as *designs*. You should consider only one design at a time; do not move on to Design 2 until you have answered all four questions for Design 1, and so on. Your focus should be on the methods of data collection for each design.

Problem C4. For each design, answer the following questions:

a. What are the strengths of the design?

b. What are the weaknesses of the design?

c. Do you have concerns about bias that are not addressed by the design?

d. Do you have suggestions for improving the design?

Design 1

- Divide the 16 subjects into two groups of eight. Ask for volunteers to form the first group.
- When the two groups have been formed, measure each of the eight subjects in Group 1 using List A; measure each of the subjects in Group 2 using List B.
- Compare the results (the eight measurements) for Group 1 with the results for Group 2.

Design 2

- Measure each of the 16 subjects using List A. Then measure each of the 16 subjects using List B.
- Compare the results (all 16 measurements) for List A with the results for List B.

Design 3

- Divide the 16 subjects into two groups, and assign eight subjects to each group.
- When the two groups have been formed, measure each of the eight subjects in Group 1 using List A; measure each of the subjects in Group 2 using List B.
- Compare the results (the eight measurements) for Group 1 with the results for Group 2.

Part C, cont'd.

Problem C4, cont'd.

Design 4

- Divide the 16 subjects into two groups of eight. Assign eight subjects to each group.
- When the two groups have been formed, measure each of the eight subjects in Group 1 using List A. Then measure each of the subjects in Group 1 using List B.
- Measure each of the subjects in Group 2 using List B. Then measure each of the subjects in Group 2 using List A.
- Compare the results (all 16 measurements) for Group 1 with the results (all 16 measurements) for Group 2.

Design 5

- Divide the 16 subjects into two groups of eight. Randomly assign eight subjects to each group.
- When the two groups have been formed, measure each of the eight subjects in Group 1 using List A. Then measure each of the subjects in Group 1 using List B.
- Measure each of the subjects in Group 2 using List B. Then measure each of the subjects in Group 2 using List A.
- Compare the results (all 16 measurements) for Group 1 with the results (all 16 measurements) for Group 2.

Problem C5. Of the five designs in Problem C4, which do you think does the best job of eliminating potential sources of bias? Does this design eliminate *all* potential sources of bias, or are there further possible improvements?

Video Segment (approximate times: 12:38-13:59): You can find this segment on the session video approximately 12 minutes and 38 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing.

In this video segment, participants compare Designs 4 and 5. Watch this segment after completing Problem C5.

Why do participants decide that Design 5 is better than Design 4?

You may say that of the five designs, the best uses random assignment to divide the subjects into two groups: Half of the subjects see List A first and half see List B first, and each subject ultimately sees both lists. When measurements are paired in this way (i.e., each person reads both lists), we are able to compare two different measurements for each subject. Without pairing, we must compare measurements for different people, and the differences in the people themselves may affect the difference in the measurements. Pairing will prove especially useful when we analyze our data. **[See Note 6]**

Note 6. Groups or individuals working alone should discuss or reflect on the important ideas and issues of experimental design:

- Using volunteers for subjects or personally selecting subjects might bias results.
- The order in which tasks are performed may affect the outcomes of the tasks.
- Pairing in data collection can reduce variation in measurements. In this case, each person is paired with him- or herself. This enables us to eliminate the differences in memory recall caused by the individual differences among subjects from any observed difference.
- Random selection or assignment is intended to remove bias. Also, randomly assigning subjects to two groups is intended to average out their differences so that the two groups are more likely to be similar.

Part C, cont'd.

Problem C6.

a. Describe a method you might use to randomly assign the 16 subjects to two groups of eight subjects each.

b. Why would it be preferable to have each subject read both lists, instead of having eight subjects read List A and eight subjects read List B?

[See Tip C6, page 183]

Video Segment (approximate times: 20:47-25:22): You can find this segment on the session video approximately 20 minutes and 47 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing.

In this video segment, researchers at the Brigham and Women's Hospital discuss the design of a study on the effects of aspirin. Watch this segment after completing Problem C6.

How was the study on aspirin designed? What characteristics of the study's design are most important in eliminating bias?

Analysis of the Experiment

Analyze the data:

Sixteen subjects participated in the memory experiment using Design 5. Eight subjects were randomly assigned to each group. Each subject in Group 1 was first measured using List A. Then each subject in Group 1 was measured using List B. The order was reversed for Group 2; the subjects were first measured using List B, then List A. Here are the measurements from the experiment:

Number Recalled Correctly

Person	List A: Words	List B: Non-Words	Difference
1	12	9	3
2	15	6	9
3	12	5	7
4	12	4	8
5	10	5	5
6	3	5	-2
7	7	5	2
8	11	2	9
9	9	1	8
10	14	8	6
11	9	1	8
12	10	5	5
13	9	6	3
14	5	4	1
15	13	9	4
16	10	4	6

Because every subject was measured using both List A and List B, it is of interest to look at the differences in the subjects' scores (Score A - Score B), which are shown in the last column.

Part C, cont'd.

Problem C7.

a. Determine the Five-Number Summary for the 16 scores using List A.

b. Determine the Five-Number Summary for the 16 scores using List B.

c. Using the same scale, sketch the two box plots for List A and List B side by side.

d. Based on the summaries and box plots, how do scores using List A (Words) compare with scores using List B (Non-Words)?

The comparison of the two box plots indicates that the scores of subjects using List A tend to be higher than the scores of subjects using List B; therefore, subjects recalled words more readily than they recalled non-words. The difference in the two medians (10 - 5 = 5) indicates that people can typically recall five more words than non-words. However, there is somewhat more variation in the scores from List A (words) than from List B (non-words); the range for List A is 12 compared to 8 for List B, and the IQR for List A is 3 compared to 2 for List B.

Video Segment (approximate times: 16:15-17:04): You can find this segment on the session video approximately 16 minutes and 15 seconds after the Annenberg/CPB logo. Use the video image to locate where to begin viewing.

In this video segment, Professor Kader and participants display the results of the study on short-term recall on box plots and discuss what they see. Watch this segment after completing Problem C7.

Why are box plots helpful in studying the results of comparative experiments?

This comparative analysis does not take into account the advantage you gain from pairing each subject's scores from both lists and then examining the difference between the two scores (Score A - Score B). A positive difference occurs when the List A score is higher than the List B score; that is, the difference is positive when memory recall is better for words than for non-words. **[See Note 7]**

Problem C8. Examine the column of differences in the Number Recalled Correctly table:

a. How many of the differences are positive? How many are negative?

b. What does this suggest about memory recall of words versus non-words?

c. Determine the Five-Number Summary of the 16 differences.

d. Sketch the box plot of the differences.

e. Based on the Five-Number Summary and box plot of the differences in scores between the two lists, how do scores using List A compare with scores using List B?

Note 7. This is an informal analysis of results. Formal probability-based inferences were not considered. However, the issues of generalizing any results to a larger population need to be raised. What is the population? How general are the results? A more advanced analysis is required to make an inference about any larger population that our results might represent.

Homework

Problem H1. Conduct your own comparative study. It can be an observational study or an experimental study. The questions below should serve as guidelines as you proceed.

Ask a question:

a. What question would you like to answer with your study?

b. What population are you seeking to answer this question for? For example, the population might be teachers at your school, or boxes of Brand A raisins.

Collect appropriate data:

c. Come up with a design for your study, one that seeks to remove sources of potential bias. (See Design 5 of Problem C4 for a possible design strategy.) Remember that to answer your question, you will need two sets of data to compare. Make sure that you collect enough data to analyze; you should use at least 12 subjects (16, if possible).

d. If you've chosen an experimental study, what treatment will you be investigating? Remember that the goal of an experimental study is to judge the effectiveness of a particular treatment.

Analyze the data:

e. Determine Five-Number Summaries and box plots for each set of data you collect, and compare the box plots side by side.

Interpret the results:

f. Does your data answer the original question? Do your results suggest any new questions? If the data you collected did not answer the question, how might you modify the study to answer the question more thoroughly?

Suggested Readings

These readings are available as downloadable PDF files on the *Data Analysis, Statistics, and Probability* Web site. Go to:

www.learner.org/learningmath

Noether, Gottfried (1994). Mental Random Numbers: Perceived and Real Randomness. In *Teaching Statistics at Its Best* (pp. 40-41). University of Sheffield, England: Teaching Statistics Trust.

Tips

Part C: Comparative Experimental Studies

Tip C2. Remember, bias can come in many forms, including biases of the people conducting a measurement and biases of the people whose ability to recall words is being measured.

Tip C3. If you gave someone the same task twice, would you expect them to perform better, worse, or equally well the second time? If you did not answer "equally well," then that presents a source of bias for this experiment.

Tip C6. You need to make sure that the selection process is random and that the person(s) conducting the experiment does not have any influence on dividing the subjects into groups. Regarding the second question, what would happen if the eight subjects randomly chosen for List B were unusually smart?

Solutions

Part A: Comparative Studies

Problem A1.

a. This is an experimental study, since the online course is deliberately imposed on one group of participants.

b. The treatment here is the online course. The treatment is successful if students perform better on the exam after taking the online course.

Problem A2.

a. You might pick two groups of working adults, of roughly the same age and background, half of whom voluntarily take vacations during the winter and half of whom voluntarily take vacations during the summer. Then you'd give each a survey of their feelings upon returning to work and compare the results.

b. This is an observational study, since choosing when to take a vacation is not imposed on anyone.

c. If everyone takes his or her vacation voluntarily, then there is no treatment. If you were to change this to an experimental study and select when people go on vacation, then the treatment would be the choice of vacation time.

Problem A3.

a. You might select a number of people at random and tell half of them to eat a diet high in fiber, fruits, and vegetables. The other half would not receive any specific dietary instructions. Over a long period of time, measure the incidence of colon cancer in each group.

b. This could be an experimental study (as described in the answer to part [a]). You might encounter difficulty enforcing the imposed dietary changes, and there may be unforeseen health risks imposed on either group as a result of their diets.

c. The treatment under study here is a diet high in fiber, fruits, and vegetables.

Part B: Comparative Observational Studies

Problem B1. The raisin studies are observational because they observe the objects (raisins) as they are. There is no treatment deliberately imposed on any group of raisins, so there is no "cause and effect" to study.

Problem B2. Here are the Five-Number Summaries and box plots for each brand:

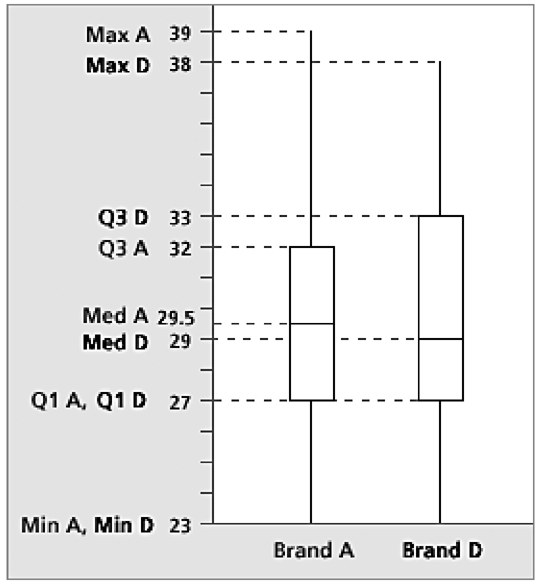
	Min	Q1	Med	Q3	Max
Brand A	23	27	29.5	32	39
Brand D	23	27	29	33	38

Solutions, cont'd.

Problem B2, cont'd.

The box plots indicate that the two sets of counts are very similar. The location indicators are all about the same: The Mins and Q1s are exactly the same, and the Meds, Q3s, and Maxes differ by 0.5, 1, and 1, respectively, which are not large differences relative to the magnitudes of the numbers we are comparing.

The degree of variation is similar for the two brands. The ranges for Brands A and D are 16 and 15, respectively, and the IQRs are 5 and 6.



Problem B3. Here are the Five-Number Summaries for each brand:

	Min	Q1	Med	Q3	Max
Brand A	23	27	29.5	32	39
Brand B	17	25	26	29	30
Brand C	25	26	28	29	32

The Five-Number Summaries for Brands A, B, and C suggest that Brand B has the fewest raisins in general. It has the smallest median (26), the smallest minimum (17), and the smallest maximum (30). Brand C has the least total variation and the highest minimum (25). Brand A has the most raisins in general, having the largest median (29.5) and by far the largest maximum (39); it also has the greatest variation.

Part C: Comparative Experimental Studies

Problem C1. Answers will vary, but here is one possible method of collection:

1. Each subject is given a specific amount of time to study List A—the list of words.
2. After a timed delay, each subject is given a specific amount of time to write down all the words the subject can recall from the list.
3. The subject is scored based on the number of words correctly recalled.

Repeat this process using List B—the non-words.

Because you want to use the results to make a comparison, you want the lists to be comparable. The two lists should contain the same number of character strings, and the character strings should be the same size. One way to do this is for both lists to contain "words" that are all the same length (say, four letters each); another is for both lists to contain "words" that are respectively the same lengths (say, the first "word" on each list is six letters, the next "words" are three letters, and so on).

Solutions, cont'd.

Problem C2. Yes, the process may be biased in the way the "non-words" are created—the person creating non-words may deliberately choose letter combinations that are exceptionally difficult to remember. Also, someone asked to perform the task twice, with List A and then List B, may perform better with List B simply because he or she has practiced the task by using List A first.

Problem C3. Answers will vary. One source of bias is that if people are all asked to try List A first, then List B, there may be bias introduced since the participants are performing the same task a second time.

Problem C4.

Design 1

a. One strength is that an equal number of participants form each group.

b. A major weakness is the selection of the groups; volunteers may be particularly eager to test their memories. Another weakness is that no participant takes both tests, so there is no possibility of directly comparing the results of the two lists for the same participant.

c. The selection of the participants is a large source of bias and is not addressed by the design.

d. A better design would randomly assign participants to each group and would take measurements for each list for all 16 participants.

Design 2

a. An equal number of participants uses each list, and all participants take both tests, allowing for more direct comparison.

b. A major weakness is that all participants use List A first. By using List B second, they may perform better (or worse) simply due to their prior experience from the first test, and not due to actual differences in the tests themselves.

c. A potential source of bias is the possibility that List B becomes easier or harder as a result of List A being used first.

d. A better design would randomly select half the participants to use List A first and half to use List B first.

Design 3

a. Each group has an equal number of participants, and participants do not determine their own groups.

b. A major weakness is the selection of the groups. The person conducting the study may deliberately place certain types of people in a group, either to deliberately skew the results or due to unconscious bias. Another weakness is that no participant takes both tests, so there is no possibility of directly comparing the results of the two lists for the same participant.

c. The fact that the person conducting the study selects the groups is an unaddressed source of bias.

d. A better design would randomly assign participants to each group and would take measurements for each list for all 16 participants.

Design 4

a. All participants use both lists, and an equal number of participants uses each list.

b. The major weakness is the selection of the groups. The person conducting the study may deliberately place certain types of people in a group, either to deliberately skew the results or due to unconscious bias.

c. The fact that the person conducting the study selects the groups is not addressed as a source of bias.

d. A better design would randomly assign participants to each group.

Solutions, cont'd.

Problem C4, cont'd.

Design 5

a. Groups are randomly assigned; all participants use both lists; and an equal number of participants use each list first.

b. There do not appear to be any major weaknesses.

c. There are no major sources of bias.

d. A better design might include more participants to increase the relevance and confidence of the findings.

Problem C5. Design 5 does the best job of removing bias. There are still small possible sources of bias, the most apparent of which is the method of creating the two lists. A specific way of randomly generating List B would be useful, as might a specific method of finding the words used for List A.

Problem C6.

a. There are many ways to randomly assign the 16 subjects to groups. One way is to place the 16 names on equal-size slips of paper, then draw eight of these slips from a hat. Another is to start with a list of last names of the 16 participants arranged in alphabetical order, then flip a coin for each individual. If the coin lands heads, the participant is assigned to Group 1, and if the coin lands tails, the participant is assigned to Group 2. This continues until eight subjects are assigned to a group; the remaining subjects are then assigned to the other group.

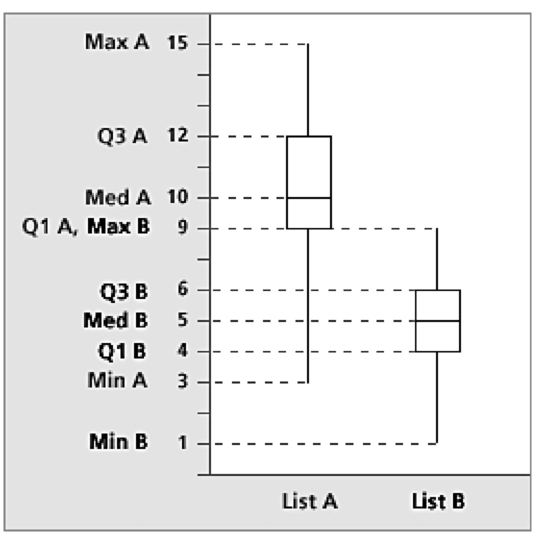
b. If different groups read each list, variation in the data may come from randomly picking only the best people to read one list. If each person reads both lists, this potential source of variation and bias is removed completely.

Problem C7.

a. and b. Here are the Five-Number Summaries for the two lists:

	Min	Q1	Med	Q3	Max
List A	3	9	10	12	15
List B	1	4	5	6	9

c. Here are the box plots:



d. Note that all measures of location (Min, Q1, Med, Q3, and Max) are higher for List A than for List B. The median for List A is twice the median for List B. In other words, people typically remembered twice as many "words" from List A as from List B. However, there is more variation in the number recalled correctly for List A than for List B. The range for List A is 12 (from 3 to 15), while the range for List B is 8 (from 1 to 9). The interquartile ranges for the two lists (3 and 2) are roughly equal.

One telling statistic is that the median of List A (10) is higher than the maximum of List B (9). This means that more than 50% of people scored higher on List A than anyone scored on List B.

Solutions, cont'd.

Problem C8.

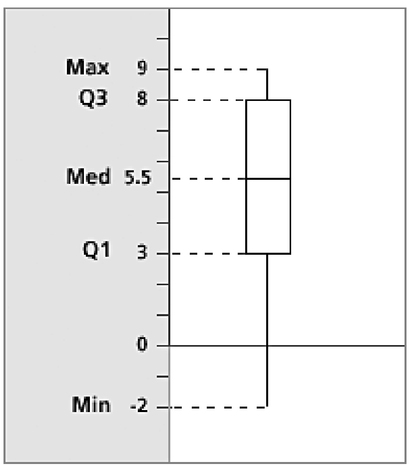
a. Fifteen of the 16 differences are positive. Only one is negative.

b. Since only one person had better recall using the list of non-words, and 15 others had better recall using the list of words, this suggests that words are significantly easier to recall than non-words.

c. Here is the Five-Number Summary of the differences:

Min	Q1	Med	Q3	Max
-2	3	5.5	8	9

d. Here is the box plot of the differences:



e. These results indicate that people are better at recalling words than non-words. Note that the entire interquartile range (the "box") is above the axis, which indicates that all of the center 50% of participants performed better with the list of words.

Homework

Problem H1. Answers will vary.