$\qquad$ Per:
Review Chapters 13, 3, 4
MULTIPLE CHOICE. Write the letter corresponding to the best answer.
E 1.* The Physicians' Health Study, a large medical experiment involving 22,000 male physicians, attempted to determine whether aspirin could help prevent heart attacks. In this study, one group of about l1,000 physicians took an aspirin every other day, while a control group took a placebo. After several years, it was determined that the physicians in the group that took aspirin had significantly fewer heart attacks than the physicians in the control group. Which of the following statements explains why it would not be appropriate to say that everyone should take an aspirin every other day?
I. The study included only physicians, and different results may occur in individuals in other occupations.
II. The study included only males and there may be different results for females.
III. Although taking aspirin may be helpful in preventing heart attacks, it may be harmful to some other aspects of health.
A) I only
B) II only
C) III only
D) II and III only
E) I, II, and III

E 2. When considering graphical displays for data, which of the following statements is/are true?
A) Bar graphs are appropriate for categorical data (but not numerical data).
B) Histograms and dotplots are appropriate for numerical data (but not categorical data).
C) Dotplots are most appropriate for distributions of numerical data when there are a relatively small number of data points (generally fewer than 50).
D) Histograms are most appropriate for distributions of numerical data when the number of values in the distribution is large (generally more than 50 or 100).
E) All of the above statements are true.

B 3. A researcher plans a study to examine the attitudes of residents of California towards a proposal in Congress to declare English to be the official language of the state. He obtains a random sample of 50 residents of one community in California and all agree to participate. Which of the following statements is true?
$\times$ A) This is a poorly designed survey because it is a voluntary response sample.
$\checkmark$ B) The design of the study may be biased because the sample may not represent the population of interest.
$\times$ C) It is a well-designed survey because of the $100 \%$ response rate.
$\times$ D) As long as the respondents were randomly selected, there is no bias.
$x$ E) A more accurately designed study would have included opinions on this issue from residents in other states.

A 4. Which of the following is not required in an experimental design?
A) blocking
B) control
C) randomization
D) replication
E) All are required in an experimental design.

D 5. An experimenter believes that two new exercise programs are more effective than any current exercise routines and wishes to compare the effectiveness of these two new exercise programs on physical fitness. The experimenter is trying to determine whether or not a control group, which follows neither of these new programs but continues with current exercise routines, would be beneficial. Which of the following can be said about the addition of a control group?
A) A control group would eliminate the placebo effect.
B) A control group would eliminate the need for blinding in the study.
C) A control group would allow the experimenter to determine which of the two exercise programs improves physical fitness the most.
(D) A control group would allow the experimenter to determine if either of the exercise programs is more effective than current programs for physical fitness.
E) There would be no added benefit to having a control group.

丷ㅡ 6. Look back at question \#5. If that question shows up on the test and it is NOT MULTIPLE CHOICE, would you still know how to answer the question correctly?
A) Yup!
B) No.
C) Maybe?
D) Okay, I can take a hint...
E) Let me look back at \#5 one more time...

E 7. Obtaining a sample of students in a high school by randomly sampling from each grade level is an example of
A) convenience sampling
C) simple random sampling
E) stratified random sampling
B) cluster sampling
D) voluntary response sampling
H) systematic random sampling
8. Describe each of the following as either a statistic or a parameter. Then write the appropriate symbol for each.

9. An article on peanut butter in Consumer Reports reported the following scores for various brands:

| Creamy |  |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 56 | 44 | 62 | 36 | 39 | 53 | 50 | 65 | 45 | 22 |
| 40 | 56 | 68 | 41 | 30 | 40 | 50 | 56 | 30 |  |

Construct a comparative stem-and-leaf display (the use of split stems is NOT necessary for this problem). Then write a few sentences to compare the two types of peanut butter.

In general, the distribution of scores for crunchy is higher than the scores for creamy.
Center: Creamy has a slight higher mode (in the 50's) than crunchy (in the 40's).
Shape: Both distributions are unimodal. The distribution for creamy is perhaps skewed slightly to the lower numbers, while the distribution for crunchy is slightly skewed to the higher scores (but one could make the case for "approximately symmetric" for both of these).
Spread: Both distributions have similar spreads, although it appears from the plot that crunchy has a larger spread. Crunchy has a spread of $34-80$, and creamy has a spread of 22-68.
10. Madeline and Judy are conducting an experiment to see if students who eat animal crackers before a short memory test will perform better on the test versus students who do not eat any animal crackers. Their subjects will include students in all 4 grade levels in high school, and will also include both males and females. Both of the girls think that performance on this memory test will vary between grade levels (specifically, that the memory scores of $9^{\text {th }}$ and $10^{\text {th }}$ graders will be significantly different than the memory scores of $11^{\text {th }}$ and $12^{\text {th }}$ graders). They also believe that performance on the test will not vary significantly between males and females.

The girls wish to conduct a randomized block experimental study, however they cannot agree on an appropriate blocking variable. Judy thinks that they should block by gender, while Madeline thinks they should block by grade level. Which girl is correct? Explain.

Generally, you should block by whichever variable has a LARGER association with the response variable. Since the girls believe that there is MORE of an association between GRADE LEVEL and MEMORY PERFORMANCE, they should choose to block by grade level. So Madeline is correct.

## EXPLANATION:

This means that once they separate the subjects by grade level, students in each group will be split up - half will get to eat animal crackers, and half will not.

This way, Madeline and Judy will be able to better compare the difference in scores between students who ate animal crackers versus those who didn't.

The goal in blocking is to create groups that are HOMOGENOUS in some way that associates with the response variable (which is memory performance scores).

Since there is not believed to be a significant difference between males and females (in terms of memory performance), blocking by gender would not help minimize the variability, since both blocks could still contain students of all 4 grades in each block.
11. Sixty people in a mall were interviewed. They were asked about the highest level of education they had completed and whether or not they smoked cigarettes.
a) Approximately what proportion of smokers completed high school only?

$$
\text { about } 63 \%
$$

b) Did a greater number of nonsmokers or smokers complete 4 or more years of college?

$$
\begin{aligned}
& \text { We dou't know. } \\
& \text { (we only have proportions, not counts) }
\end{aligned}
$$

c) Did a greater proportion of nonsmokers or

|  | Smo | tion |
| :---: | :---: | :---: |
| $90 \%$ $80 \%$ $70 \%$ | 4+-year college <br> 2-yr | 4+-year college |
| 50\% |  | 2-yr |
| $30 \%$ <br> 20\% <br> 10\% | School | $\begin{aligned} & \text { High } \\ & \text { School } \end{aligned}$ |

Smoker
Nonsmoker

$$
\begin{aligned}
& \text { A greater proportion of NON-smokers } \\
& \text { completed } 4+\text { years of college. }
\end{aligned}
$$

d) For this group of people is there an association between level of education and smoking? Explain.

Yes.
A significantly larger PROPORTION (don't say "number!!!") of smokers only completed up through high school compared to non-smokers.

Also, non-smokers were more likely (proportionally) to complete a 4+ year college than smokers.
12. In a survey of airline travelers, subjects were observed in the coach section of airplanes to determine if men or women are bothered by a seatmate of the opposite gender using the common armrest.

Discontentment Felt When Seat-mate Used Common Armrest: Males and Females

|  | Bothered (B) | Not Bothered (NB) | Total |
| :--- | :---: | :---: | :---: |
| Females (F) | 19 | 26 | 45 |
| Males (M) | 38 | 17 | 55 |
| Total | 57 | 43 | 100 |

a) What is the percent of males who are bothered by a seatmate using the common armrest?

$$
38 / 55 \approx 69.1 \%
$$

b) What is the percent of people surveyed that are bothered?

$$
57 / 100=57 \%
$$

13. A nutritionist wants to study the effect of storage time ( 6,12 , and 18 months) on the amount of vitamin $C$ present in dried fruit when stored for these lengths of time. Vitamin C is measured in milligrams per 100 milligrams of fruit. The researcher has 60 total packages of dried fruit: 30 of them are dried apple, and the other 30 are dried peach.
a) Identify the experimental units: Packages of dried fruit
the explanatory variable(s) or treatment(s): Storage time of 6 mouths, 12 months, or 18 mouths
the response variable(s): amount of vitamic $C$
b) Explain how you would carry out a completely randomized experiment for this study.


Randomization Procedure: Assign each package a number from 1-60. Write each number on a slip of paper and put all 60 slips in a hat (make sure each slip of paper is the same size, and be sure to stir the slips in the hat). Without looking, draw 20 slips of paper. The packages of fruit with those numbers are assigned to group 1 ( 6 months of storage). Draw 20 more slips from the hat and place those packages of fruit into group 2 ( 12 months of storage). The remaining 20 packages of fruit are placed into group 3 (18 months of storage).
c) Describe the changes that would be made to your experiment in part (B) by incorporating blocking.

Since the different types of fruit may differ in amounts of Vitamin C, we will block by type of fruit.

d) Can the nutritionist generalize his/her findings to all types of dried fruit? Explain.

No. Because this experiment was conducted with dried apple and dried peach, we can only generalize the findings to dried apple and dried peach (and not to other types of dried fruit).
14. Temperatures from random samples of cities on a hot summer day from two different regions of the European continent. Write a few sentences to describe and compare the distributions of temperatures.

[Student must COMPARE center, shape and spread IN CONTEXT]
The distribution of temperatures for the Southwest has a lower mode (in the mid80's) than for the Northwest (where most of the temperatures are spread out in the 80's and 90's).

The distribution for the Northwest is skewed to the left, while the distribution of temperatures in the Southwest is roughly symmetric and bell-shaped (approximately normal).

The distribution of temperatures in the Northwest has greater variability (going from the mid-60's up to around 101), while the temperatures in the Southwest range from 72 to 99.
15. The dentists in a dental clinic would like to determine if there is a difference between the number of new cavities in people who eat an apple a day and in people who eat less than one apple a week. They are going to conduct a study with 50 people in each group.

Fifty clinic patients who report that they routinely eat an apple a day and 50 clinic patients who report that they eat less than one apple a week will be identified. The dentists will examine the patients and their records to determine the number of new cavities the patients have had over the past two years. They will then compare the number of new cavities in the two groups.

If the mean number of new cavities for those who ate an apple a day was statistically significantly smaller than the mean number of new cavities for those who ate less than one apple a week, could one conclude that the lower number of new cavities can be attributed to eating an apple a day? Explain.

No, we cannot conclude that eating an apple a day leads to a reduction in the number of new cavities. The patients were not randomly assigned to treatments, therefore this was an observational study (not an experiment). Causal (aka,
"cause-and-effect") relationships can only be drawn from well-designed randomized experiments.

## OR

The student may state that a possible confounding/lurking variable is responsible for the reduction of new cavities; however this variable MUST be linked to eating an apple a day.

FOR INSTANCE, perhaps patients who eat an apple a day also eat well and/or are more conscious about their general health, and perhaps this confounding variable has more to do with a reduction in cavities than just eating an apple.

