Review+Practice

May 30 , 2012

- Final: Tuesday June 5 8:30-10:20
- Venue: Sections AA and AB (EEB 125), sections AC and AD (EEB 105) , sections AE and AF (SIG 134)
- Format: Short answer.
- Bring: calculator, BRAINS
- Will be provided formula sheet, normal tables.

1 Brian gathers data from his classmates about the computers they own. What are the observational units in this data set?

Name	O.S.	Amount of Memory	Year Purchased
Joe	Windows XP Pro	256 MB	2003
Max	Windows ME	128 MB	2000
Sue	Mac OS X	640 MB	2004
Jean	Windows 2000	256 MB	2002
Bill	Mac OS X	1 GB	2005

A Brian's classmates

B The type of operating system

C Joe, Max, Sue, Jean, Bill

D The amount of memory

2 In the study described in the article "Educating women saves millions of kids" American researchers analyzed 915 censuses and surveys from 175 countries tracking education and child deaths. What are the two variables under study and what is the observational unit?

3 In Brian's dataset, which of the variables is not considered to be numerical?

   A Names of Brian's classmates
   B Type of operating system
   C Amount of memory
   D Year purchased.

4 The Seattle Times conducts a poll on whether the people of Seattle believe red light cameras are effective. The variable measured on each individual is: (circle all that apply)

   A categorical
   B qualitative
   C quantitative
   D continuous
   E discrete

5 In the study described in the article "Educating women saves millions of kids" what is the scale of measurement for each variable?

6 Does drinking orange juice alleviate back pain? 50 volunteers (being treated for back pain) were asked to drink 3 glasses of orange juice every morning for two weeks. As a result, most volunteers said that pain was lessened enough for them to ask their doctors to reduce their medication. What is the explanatory variable?

    A Drinking orange juice

    B Amount of back pain

    C The fifty volunteers

    D The medication

7 Which of these is a valid response (or outcome) variable?

    A Average amount of back pain for the 50 volunteers at end of study.

    B Amount of back pain experienced by an individual at end of study.

8 Consider three situations:

I. You ask your doctor about the health risks involved with going out in the cold without a jacket.

II. On cold days, you record which of your classmates come to school without a jacket. Later, you record how many of those students become sick.

III. You find 30 adults and divide them into two groups. The first group is told not to wear jackets on cold days, the other group is told to wear jackets on cold days. You then compare the number from each group who get sick after a string of cold days.

Which of the above is an example of an observational study? II

9 An experiment is being done to test whether a new drug will reduce eye puffiness. Two groups of 50 are randomly chosen: one group is given the new drug treatment; the second is given a simple cream with no active ingredients. The group who was given the new drug treatment reported that 45% had reduced eye puffiness. In the second group with the simple cream, 20% had reduced puffiness. This second group is an example of: (choose the one that best fits)

A A placebo control group
B Randomization
C A control group
D Treatments

Where Does Data Come From?

10 Randomization is important in experimental design because it:

    A Reduces bias

    B Creates groups that are similar in all variables

    C Mitigates the effects of lurking variables

    D All of the above

11 The following feature of a designed experiment (when present) enables cause-effect conclusions to be drawn.

    A Placebo

    B Double blinding

    C Random assignment

    D Informed consent

12 The primary problem with observational studies is:

    A We cannot directly observe the results.

    B We cannot determine associations between variables.

    C We cannot determine cause and effect relationships between variables.

13 Epidemiologists find an association between high levels of cholesterol in the blood and heart disease. They conclude that cholesterol causes heart disease. However, a statistician argues that smoking confounds the association. This means one of the following. Which one?

A Smoking causes heart disease.

B Smoking is associated with heart disease, and smokers have high levels of cholesterol in their blood.

C Smokers tend to eat a less healthful diet than non-smokers. Thus, smokers have high levels of cholesterol in the blood, which in turn causes heart disease.

D The percentage of smokers is about the same among persons with high or low levels of cholesterol in the blood.

14 Sample surveys such as the Current Population Survey are an important kind of

   A Census
   B Observational study
   C Experiment
   D Observational unit

15 The Seattle Times conducts a poll on whether the people of Seattle believe red light cameras are effective. The newspaper contacts 1000 subscribers. The population of this poll is:

   A The 1000 people surveyed
   B Those who favor or disapprove of red light cameras
   C Those who are affected by red light cameras
   D People who live in Seattle

16 Ballard High School announces the results of a survey – 31% of the senior class has an MP3 player. This result was based on a random sample of 100 seniors. What is the parameter?

    A The random sample of 100 students

    B Ballard High School

    C The percentage of the senior class who has an MP3 player

    D 31%.

Where Does Data Come From?

17 American Idol asks viewers to call in to vote on their favorite performer. This is an example of:
   - A convenience sampling
   - B voluntary response sampling
   - C simple random sampling

18 You want to choose an S.R.S of eight people from a class of 24 seated in 8 rows of 3. How would you get your sample? Circle all that apply.
   - A Choose the first eight people who walk in the door.
   - B Choose the eight people that you know would answer the survey.
   - C Arrange the items alphabetically and choose the first eight people in the list.
   - D Assign them numbers from 00 to 23 and use random digits to pick 8 nos.
   - E Choose one person at random from each of the 8 rows.
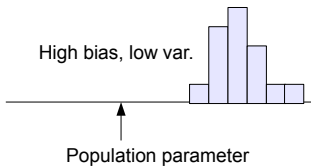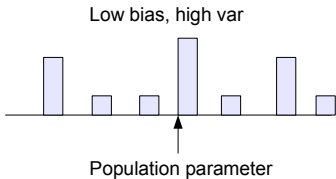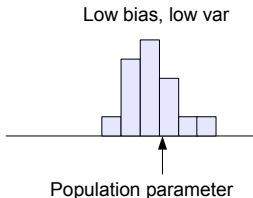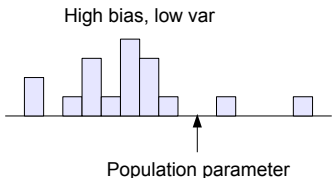
WHERE DOES DATA COME FROM?

19 Suppose a packaging inspector decides to inspect a sample from a crate of eggs for freshness. Each crate has 5 trays of 30 eggs each stacked on top of each other. The inspector decides to examine only the top row. Why is this sample biased?

    A Each egg did not have an equal chance of being chosen in the sample.

    B The top row may not be representative of the entire crate of eggs for freshness.

    C The inspector was convenience sampling which is typically biased.

    D All of the above.

20 The following figure shows the behavior of the sample statistic in many samples in four situations. Label each graph as as showing high or low bias and as showing high or low variability.

High bias, low var

Population parameter

Low bias, low var

Population parameter

Low bias, high var

Population parameter

High bias, low var.

Population parameter

WHERE DOES DATA COME FROM?

21 What kind of error is this an example of: During a phone survey, the person being called hangs up immediately?

   A Processing errors
   B Nonresponse
   C Response error
   D Wording

22 What kind of error is this an example of: During a phone survey, a person lies because they are embarassed?

   A Processing errors
   B Nonresponse
   C Response error
   D Wording

23 True or false: Confidence statements enable us to quantify random sampling errors. True

24 True or False and explain: The size of the population determines the variability of the statistic from a random sample. Which answer is the correct one?

A True. We must consider the size of the population when we determine the size of the sample.

B True. Larger populations reduce variability.

C False. Large populations determine bias not variability.

D False. Variability of a sample statistic depends on the size of the sample and not on the size of the population.

25 By choosing a sample with probability methods we can
- A Reduce bias
- B Increase bias
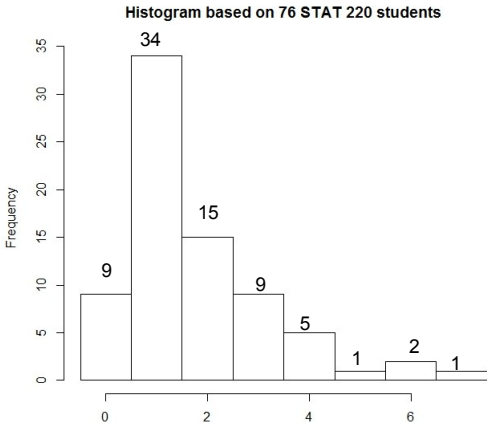- C Reduce variability
- D Increase variability

26 By increasing the size of a random sample, we can:
- A Reduce bias
- B Increase bias
- C Reduce variability
- D Increase variability

DESCRIBING DISTRIBUTIONS

27 Use the histogram to answer the following questions:

Histogram based on 76 STAT 220 students

Shape? Center? Spread? Median? Q1? Q3? Mean? S.D.?

DESCRIBING DISTRIBUTIONS

28 The summary measures that are most influenced by extreme observations are

   A Mean
   B Median
   C Mode
   D Standard deviation

29 For a right skewed histogram, the mean is _____ the median.

   A larger than
   B smaller than
   C equal to

Describing Distributions

30 Which of the following are properties of normal curves?

   A They are uniquely defined by giving its mean and standard deviation

   B The mean is at the center of symmetry of the curve.

   C They describe the distribution of statistics like sample proportions and sample means

   D All of the above

   E None of the above

DESCRIBING DISTRIBUTIONS

31 A local sub shop lists the carbohydrate content in each of
its healthy choice sandwiches. The distribution of
carbohydrate content is approximately normal with mean
40 carbohydrates and a standard deviation of 2
carbohydrates. What percentage of healthy choice
sandwiches are less than 38.5 grams of carbohydrates?

A 13.5%

B 16%

C 23%

D 36%

E 50%

32 Scores on the math part of the SAT test are normally distributed with a mean score of 500 and a standard deviation of 100. If a person scores 625 on the math part of the SAT, what is their standard score?

  A 1

  B 1.25

  C -2

  D -1.25

33 How high must a student score on the math SAT to fall within the top 18% of all scores?

  A 550

  B 590

  C 650

Describing Distributions

Relationships Between Quantitative Variables

Graph I   Graph II   Graph III
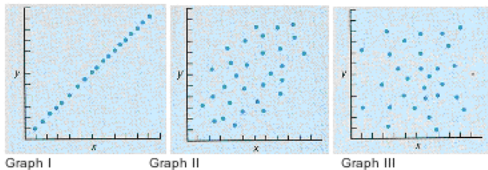
34 Which graph has the largest value of r?
- A   Graph 1
- B   Graph 2
- C   Graph 3
- D   None of the above

DESCRIBING DISTRIBUTIONS

35 When calculating *r*, what values do you use?

A Median
B Percentile
C Normal distribution
D Standard score

Describing Distributions

A regression analysis was made of the final scores from the midterm scores in a large course with the following results:

average midterm score ≈ 50, S.D. ≈ 25,
average final score ≈ 55, S.D. ≈ 15, r ≈ 0.60

36 Calculate the slope and intercept of the regression line fit to these data.

37 What is the regression estimate of the final score for a student whose midterm score was 80?

38 This prediction is likely to be off by _____ points or so.

39 Suppose a student with a midterm score of 80 ended up scoring 65 on the final. What is their residual?

40 As part of their training, air force pilots make two practice landings, with instructors, and are rated on performance. The instructors discuss the ratings with the pilots after each landing. Regression analysis shows that pilots who make poor landings the first time tend to do better the second time. Conversely, pilots who make good landings the first time tend to do worse the second time. The conclusion: criticism helps the pilots while praise makes them do worse. As a result, instructors were told to criticize all landings, good or bad. Was this justified by the facts? Yes or no and explain.

TESTS OF SIGNIFICANCE

41 What proportion of adults in the U.S. always vote in the Presidential elections? An article in *American Demographics* reports this percentage as 67%. To test this claim, a random sample of 800 adults was taken and 512 (64%) of them said that they always vote in the Presidential elections. Is this good evidence that the percentage of Americans who always vote in Presidential elections is different from 67%?