

Part 1: Multiple Choice. Circle the letter corresponding to the best answer.

- a 1. You wish to survey people who have brought in their cars for service during the past month. You decide to pick a random sample of gas stations in the city and then survey all the customers from those who had work done during the past month. This is an example of
- (a) Cluster Sampling
 - (b) Convenience Sampling
 - (c) Simple Random Sampling
 - (d) Stratified Random Sampling
 - (e) Systematic Sampling
- d 2. A bank wishes to survey its customers. The decision is made to randomly pick ten customers who just have banking accounts, ten who just have savings accounts, and ten who have both. This is an example of
- (a) Cluster Sampling
 - (b) Convenience Sampling
 - (c) Simple Random Sampling
 - (d) Stratified Random Sampling
 - (e) Systematic Sampling
- C 3. A sales representative wishes to survey her client base of 47 companies. She has 47 business cards, all identical size, from her contacts in the companies, and decides to drop them all in a small box, shake them up, and reach in to pick 5 cards for her sample. This is an example of
- (a) Cluster Sampling
 - (b) Convenience Sampling
 - (c) Simple Random Sampling
 - (d) Stratified Random Sampling
 - (e) Systematic Sampling
- C 4. Two studies are run to compare the experiences of low-income families receiving food stamps to those receiving cash subsidies. The first study interviews 50 families who have been in each government program for at least 2 years, while the second randomly assigns 50 families to each program and interviews them after two years. Which of the following is a true statement?
- (a) Both are observational studies because of the time involved.
 - (b) Both are observational studies because there are no control groups
 - (c) The first is an observational study, the second is an experiment.
 - (d) The first study is an experiment, the second is an observational study.
 - (e) Both studies are experiments, because in each, families are receiving treatments (food stamps or cash)
- a 5. Can changing diet reduce blood pressure? Vegetarian and low-salt diets are both promising. Men with high blood pressure are assigned at random to four diets: normal diet, vegetarian with unrestricted salt, normal with restricted salt, and vegetarian with restricted salt. The 240 subjects are labeled 001 to 240. Software randomly assigns 60 subjects to each of the four diets. This is an example of
- (a) Completely Randomized Design
 - (b) Randomized Block Design
 - (c) Matched Pairs Design
 - (d) An Observational Study
 - (e) A Simple Random Sample
- C 6. The Community Intervention Trial for Smoking Cessation asked whether a community wide advertising campaign would reduce smoking. The researchers located 11 pairs of communities, each pair similar in location, size, economic status, and so on. One community in each pair participated in the advertising campaign and the other did not. This is an example of
- (a) Completely Randomized Design
 - (b) Randomized Block Design, but not matched pairs
 - (c) Matched Pairs Design
 - (d) An Observational Study
 - (e) A Stratified Random Sample

C 7. Which of the following describes a situation in which it is reasonable to reach a cause and effect conclusion based on data from a statistical study?

- (a) The study is based on a random sample from a population of interest.
- (b) The study is observational, and the sample used is not a convenience sample.
- (c) The study is an experiment that uses random assignment to assign volunteers to treatments
- (d) The study is observational, and the two samples used are not a convenience samples.
- (e) It is always reasonable to reach a cause and effect conclusion based on data from a statistical study.

e 8. You want to take a simple random sample of 50 of the 816 students who live in a dormitory on campus. You label the students 001 to 816 in alphabetical order. In the table of random digits, you read the entries ~~95592~~ ~~94007~~ ~~69769~~ ~~33547~~ ~~72450~~ ~~16632~~ ~~81194~~ ~~14873~~
The first three students in your sample have labels

- ~~(a) 955, 929, 400~~
- ~~(b) 400, 769, 769~~
- (c) 559, 294, 007
- (d) 929, 400, 769
- (e) 400, 769, 335

ignore 000 and 817-999
and repeats

a 9. Consider an experiment to investigate the effectiveness of different insecticides in controlling pests and their impact on tomato plants. What is the best reason for randomly assigning treatment levels (spraying or not spraying) to the experimental units (farms)?

- (a) randomization is a good way to create two groups of farms that are as similar as possible, so that comparisons can be made between the two groups.
- (b) randomization eliminates the impact of any confounding variables.
- (c) randomization makes the analysis easier since the data can be collected and entered into the computer in any order.
- (d) randomization ensures that the study is double-blind.
- (e) randomization reduces the impact of outliers.

a 10. Control groups are used in experiments in order to

- (a) control the effects of outside variables on the outcome.
- (b) control the subjects of a study to ensure that all participate equally.
- (c) guarantee that someone other than the investigators, who have a vested interest in the outcome, controls how the experiment is conducted.
- (d) achieve a proper and uniform level of randomization.
- (e) reduce the variability in results.

d 11. A study of treatments for angina (pain due to low blood supply to the heart) compared bypass surgery, angioplasty, and use of drugs. The study looked at the medical records of thousands of angina patients whose doctors had chosen one of these treatments. It found that the average survival time of patients given drugs was the highest. What do you conclude?

- ~~(a) This study proves that drugs prolong life and should be the treatment of choice.~~
- ~~(b) We can conclude that drugs prolong life because the study was a comparative experiment.~~
- ~~(c) We can't conclude that drugs prolong life because the patients were volunteers.~~
- (d) We can't conclude that drugs prolong life because it was an observational study.
- (e) We can't conclude that drugs prolong life because no placebo was used.

d 12. You want to know the opinions of American high school teachers on the issue of establishing a national proficiency test as a prerequisite for graduation from high school. You obtain a list of all high school teachers belonging to the National Education Association and mail a survey to a random sample of 2500 teachers. In all, 1347 of the teachers return the survey. Of those that responded, 32% say they favor some sort of national proficiency test. Which of the following is true?

- (a) Since random sampling was used, we feel confident that 32% of all high school teachers favor the test
- (b) We cannot trust these results since the survey was mailed.
- (c) Because over half the sample responded, we feel confident that about 32% of all high school teachers favor the test.
- (d) The results of this survey may be affected by non-response bias.
- (e) The results of this survey cannot be trusted due to voluntary response bias.

a 13. Few people want to eat discolored French fries. Potatoes are kept refrigerated before being cut into French fries to prevent spoiling and preserve flavor. Unfortunately, immediate processing of cold potatoes causes discoloring due to complex chemical reactions. The potatoes must therefore be brought to room temperature before processing. Researchers want to design an experiment in which tasters will rate the color and flavor of French fries from several groups of potatoes. The potatoes will be freshly picked, stored for a month at room temperature, or stored for a month refrigerated. They will then be sliced and cooked. Identify the experimental units, the treatment, and the response respectively:

- (a) Potatoes, storage method, color and flavor
- (b) Potatoes, color and flavor, ~~storage method~~
- (c) Random assignment, storage method, color and flavor
- (d) French fries, cooking method, ~~taste~~
- (e) Taste, French fries, ~~cooking method~~

C 14. Do abandoned children placed in foster homes do better than similar children placed in an institution? The Bucharest Early Intervention Project found that the answer is a clear "Yes." The subjects were 136 young children abandoned at birth and living in orphanages in Bucharest, Romania. Half the children, chosen at random, were placed in foster homes (at the expense of the study since foster homes were not easily available in Romania). The remaining stayed in orphanages. What conclusion can we draw from the study?

- (a) That living in foster care is better than living in an orphanage. *population was in Romania*
- (b) That there is an association between living in foster care and doing better but we cannot establish cause and effect.
- (c) That living in foster care in Romania is better than living in an orphanage.
- (d) We can make no conclusion because the results were confounded by lurking variables.
- (e) We can make no conclusion because there were not enough participants.

C 15. A new headache remedy was given to a group of 25 subjects who had headaches. Four hours after taking the new remedy, 20 of the subjects reported that their headaches had disappeared. From this information you conclude

- (a) that the remedy is effective for the treatment of headaches.
- (b) nothing, because the sample size is too small.
- (c) nothing, because there is no control group for comparison.
- (d) that the new treatment is better than aspirin.
- (e) that the remedy is not effective for the treatment of headaches.

b 16. Find the variance, standard deviation, and range.
 {9, 7, 8, 6, 9, 12, 11, 5, 9, 10}

- (a) Variance: ~~5.0~~; standard deviation: 2.2, range: 7
- (b) Variance: 4.7, standard deviation: 2.2, range: 7
- (c) Variance: 4.7, standard deviation: 2.0, range: 7
- (d) Variance: ~~5.0~~; standard deviation: 2.2, range: 6
- (e) Variance: ~~5.0~~; standard deviation: 2.0, range: 7

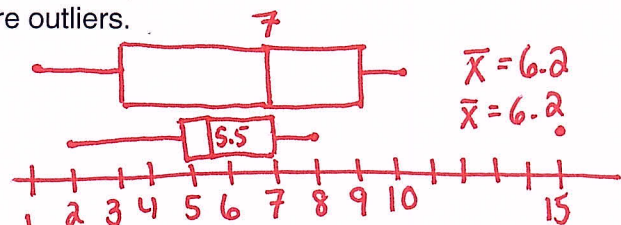
variance = (std dev)²
S_x = 2.170509413
range = 12 - 7
Var = (2.170509413)²
= 4.7

b 17. Find the mean and median from the data sets below. Which measure of central tendency would be appropriate measure to use for each? Why?

Data set 1: {7, 7, 2, 8, 5, 10, 9, 1, 10, 3}

Data set 2: {6, 2, 6, 5, 3, 8, 5, 7, 15, 5}

- (a) Datasets 1 and 2: mean: 6.2; because there are no outliers
- (b) Dataset 1: mean: 6.2; because there are no outliers. Dataset 2: median: 5.5; because 15 is an outlier.
- (c) Dataset 1: median: 7; because there are no outliers. Dataset 2: mean: 6.2; because 15 is an outlier.
- (d) Datasets 1 and 2: median: 6.2; because 10 and 15 are outliers.
- (e) Datasets 1 and 2: mean: 6.2; because 10 and 15 are outliers.



Part 2: Free Response

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

18. An ultra marathon, as you may guess, is a footrace longer than the 26.2 miles of a marathon. Runners commonly develop respiratory infections after an ultra marathon. Will taking 600mg of Vitamin C daily reduce those infections? Researchers randomly assigned ultra marathoners to receive either vitamin C or a placebo. Separately, they also randomly assigned the treatments to a group of non runners the same age as the runners. All subjects were watched for 14 days after the big race to see if infections developed.

(a) Explain why this is an experiment and not an observational study.

This is an experiment because researchers imposed a treatment (vitamin C or placebo) on the subjects.

(b) Identify the explanatory and response variables.

Explanatory: 600mg of vitamin C
Response: respiratory infections 14 days after a ultra marathon.

(c) Identify the experimental design used in this study. Justify your answer.

Block design. We performed the experiment on two groups (blocks). One group was runners and the other group was non-runners.

(d) The report of the study said "Sixty-eight percent of the runners in the placebo group reported the development of symptoms of upper respiratory tract infection after the race; this was significantly more than reported by the vitamin C supplemented running group (33%)." What is the statistical meaning of the word "significantly" in the context of this study?

That means we are unlikely to get a difference (68% vs 33%) that large or larger just by chance. Therefore, we can attribute the difference to the supplementing of vitamin C.

(e) This was a controlled experiment. Describe how it was controlled and explain the purpose of doing so.

For each block, there was a group with an inactive treatment to compare to. It allows us to attribute the response to the treatment instead of to lurking variables.

19. Many people start their day off with a jolt of caffeine from coffee or a soft drink. Most experts agree that people who take in large amounts of caffeine each day may suffer from physical withdrawal symptoms if they stop ingesting their normal amounts of caffeine.

(a) It is well known that all people react differently to withdrawal. What sort of experimental design would you choose for this study, and why?

Matched pairs design. or pair with themselves!

Since everyone reacts differently to withdrawal, we would want to pair up people that have similar withdrawal symptoms and similar caffeine intake so we can have accurate comparison data for the response variable.

(b) Explain why an experiment involving 100 people is preferable to one involving 10 people.

Replication: There more subjects used, the less variability in our results. It allows outside variables to become balanced between the two groups w/ random assignment.

(c) Assume that 100 people are available for the study. Describe a design for this experiment. Be sure to include a description of how you assign individuals to the treatment groups.

For each person: flip a coin (heads = caffeine, tails = no caffeine) to determine their caffeine consumption for 3 days. Monitor symptoms associated with withdrawal. Give a week recovery and switch treatments. Monitors symptoms. Compare data.
 use placebo!

20. Bias is present in each of the following sampling designs. In each case, identify the type of bias involved and state whether you think the sample result obtained is lower or higher than the actual value for the population.

(a) There is an announcement in the daily bulletin that student parking is going to change and come to the main office during lunch to fill out a survey whose results will be used to alter how things are arranged in the student parking lot. The survey results come out and say that 60% of all students feel that seniors should have the first two rows of parking in the parking lot.

- Voluntary response bias (no random selection, students choose who participates)
- It overestimates the amount of students who feel seniors should get the first 2 rows of parking.

(b) A simple random sample of households in a city is performed and the head of household is asked "Do you think our Mayor should be run out of office for all the ways he has mishandled the finances for the city?" An overwhelming 92% of the head of households responded "yes".

- Bias from the way the question is worded.
- It overestimates the percentage of people who feel that the Mayor should be removed from office

