

DIRECTIONS:

- This exam contains two parts:
 - Part 1. 44 Multiple Choice [88 points]
 - Part 2. 1 Short Answer [12 points]
- On Part 1, circle the correct response for each question. Make sure that your answer is clearly marked. You will not receive partial credit for any work done in Part 1. On Part 2, write clearly (in complete sentences) and explain your thoughts/answers carefully.
- This is a closed-book, closed notes examination. You may use a calculator if you wish.
- Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material, will be dealt with severely. Violations may result in an “F” for the class, suspension, or expulsion.
- This exam is worth a total of 100 points. **Print your name at the top of this page in the upper right hand corner.** *Good Luck!!*

HONOR PLEDGE FOR THIS EXAM:

After you have finished the exam, please read the following statement and sign your name below it.

I promise that I did not discuss any aspect of this exam with anyone other than the instructor, that I neither gave nor received any unauthorized assistance on this exam, and that the work presented herein is entirely my own.

PART 1: MULTIPLE CHOICE. Circle the correct response for each question. *Make sure that your answer is clearly marked.*

1. In class, what did we call a statement that is **always** true?
 - (a) an anecdotal statement
 - (b) a probabilistic statement
 - (c) a deterministic statement
 - (d) a random statement

2. What is the **key difference** between an observational study and an experiment?
 - (a) Observational studies are usually carried out using probability sampling designs; experiments do not use sampling.
 - (b) Experiments eliminate the chance for biased results; observational studies are often plagued with bias.
 - (c) In observational studies, we merely observe individuals' responses; in experiments, we deliberately impose a condition to influence the individuals' responses.
 - (d) None of the above.

3. What is **statistical inference**?
 - (a) The practice of using "cherry-picked" statistics to thrust one's opinion onto others.
 - (b) The process by which sample results are used to make statements about population parameters.
 - (c) The manner in which ethical issues are addressed in an experiment involving human subjects.
 - (d) None of the above.

4. In the language of surveys, what is a **frame**?
 - (a) a sampling design where each sample has the same chance of being chosen
 - (b) a device used to eliminate poorly-worded survey questions
 - (c) a non-treatment designed to look like the real treatment
 - (d) a complete list of every member of the population

5. In class, we talked about the HIV intravenous drug use (IDVU) data from Houston. In one instance, we separated the IDVU individuals by race (Hispanic, white, and black). What term best describes **race** in this example?
 - (a) response variable
 - (b) stratifying variable
 - (c) experimental variable
 - (d) binary variable

6. What does the **margin of error** quantify?
- (a) the chance of making a correct statement
 - (b) the uncertainty arising from using a statistic to estimate a population parameter
 - (c) the amount of information gained from using an experiment when compared to using an observational study
 - (d) None of the above
7. A **confidence statement** includes what two things?
- (a) margin of error and bias
 - (b) bias and variability
 - (c) bias and confidence level
 - (d) confidence level and margin of error
8. In an experiment, what did we call the group of individuals (subjects) that receives a **placebo**?
- (a) control group
 - (b) experimental group
 - (c) blocking group
 - (d) matched group
9. The last time I went to Atlanta, I decided to catch a Braves game. As I looked throughout the stands, I wondered what proportion of attendees were Braves fans. To answer this question, I asked 10 people in the row in front of me if, in fact, they were Braves fans. What type of **sampling design** did I use?
- (a) a probability sample
 - (b) simple random sample
 - (c) cluster sample
 - (d) convenience sample
10. On the way back from Atlanta, I stopped at the University of Georgia (UGA, in Athens) and met with a colleague to discuss some research in statistical genetics. While I was there, I stopped at a local coffee shop and asked 10 UGA undergraduates if they were Braves fans (4 said they were). Which statement do you think is **valid** based on these undergraduates' responses?
- (a) Forty percent of all Americans are Braves fans.
 - (b) Forty percent of all Georgia residents are Braves fans.
 - (c) Forty percent of all UGA undergraduates are Braves fans.
 - (d) None of the above.

11. *Does drinking coffee tend to increase a student's performance in school?* To answer this, I walked down to the Russell House and found 10 coffee-drinking undergraduates and 10 non-coffee drinking undergraduates. I asked them for their cumulative GPA and compared the GPAs for the two groups. Is this an experiment or an observational study?

- (a) experiment
- (b) observational study

12. Why is the **randomized response technique** helpful?

- (a) It eliminates all sources of variability in a survey.
- (b) It guarantees that researchers will not be unethical.
- (c) It provides a way for sensitive information to be collected while preserving an individual's anonymity.
- (d) It enables one to perform an experiment under conditions which can be generalized to a larger population of subjects.

13. Which one of the following statements is **false**?

- (a) When we attempt to make a statement about a population, based on the results of a sample, there will always be a degree of uncertainty.
- (b) If we take repeated random samples from a population, sample statistics computed from the samples will likely be different.
- (c) The margin of error present in the simple random sampling design will always increase when we take larger samples.
- (d) The margin of error for the randomized response technique is computed differently than it is for the simple random sampling design.

14. In class, we talked about the drug **laetrile** (a drug for cancer patients). The reason this drug gained so much popularity is attributed to

- (a) a carefully designed experiment.
- (b) a small margin of error.
- (c) a diligent institutional review board.
- (d) accounts of anecdotal information.

15. In an experiment, what does it mean when we say that two variables are **confounded**?

- (a) The effects that each variable has on the response variable can not be plausibly determined.
- (b) The variables were randomly chosen to represent the population.
- (c) We can be confident in the experimental results involving these two variables.
- (d) The experiment likely suffered from nonresponse because of the variables recorded.

16. A simple random sample of $n = 400$ Birmingham, AL residents was recently asked the question: *Have you ever been the victim of a hate crime?* From this sample, 100 residents responded “yes.” From this information, what is the value of the **sample proportion** of hate-crime victims?

- (a) 0.05
- (b) 0.10
- (c) 0.20
- (d) 0.25

17. In Question 16, what is the **margin of error** associated with the sample proportion?

- (a) 0.05
- (b) 0.10
- (c) 0.20
- (d) 0.25

18. I read an advertisement recently in which a credit card company promised that I could reduce my debt by 150 percent. Which of the following statements is (are) **true**?

- (a) This is possible if my debt is more than 150 dollars.
- (b) This is possible if my debt has recently increased by at least 150 percent.
- (c) The company’s claim makes no sense.
- (d) Both (a) and (b).

19. In clinical trials, government regulations require that institutional review boards (IRB) consist of at least 5 people, including at least one **non-scientist**. Why should review boards contain people who are not scientists?

- (a) The IRB needs to convince the government that pharmaceutical companies are getting the profits that they deserve.
- (b) There needs to be a balance of educational backgrounds among those who are reviewing the results.
- (c) The government requires that patients responses are confidential.
- (d) Without at least one non-scientist the margin of error in the results would be unacceptably high.

20. True or False. We use sample statistics to estimate population parameters.

- (a) True
- (b) False

A group of medical researchers studied the effects of two different diets: **LOW CARB** and **LOW FAT**. The researchers randomly assigned each of 132 severely obese patients to one of the two diets (66 patients to each diet) and monitored their results for 6 months. Use this information to answer Questions 21-25.

21. Which of the following would be a useful **response variable**?
- (a) weight loss after 6 months
 - (b) gender
 - (c) weight before starting the study
 - (d) diet assigned
22. Based on the description above, what type of **experimental design** is being used?
- (a) matched pairs design
 - (b) completely randomized design
 - (c) observational study with a dual-treatment
 - (d) randomized response design
23. The researchers themselves cautioned **against** interpreting these results as a general endorsement for a low-carb diet for those wanting to lose weight. Why was this caution likely made?
- (a) This is an observational study, and we know it is difficult to generalize the results.
 - (b) This experiment did not use blinding.
 - (c) This experiment did not preserve subject confidentiality because we know how much weight was lost.
 - (d) This experiment only used severely-obese subjects.
24. The researchers said that the results from this experiment were **statistically significant**. What does this mean?
- (a) The two diets performed similarly.
 - (b) The difference observed in the two diets was large.
 - (c) The difference observed in the two diets was likely not due to chance.
 - (d) The difference observed in the two diets was confounded with a lurking variable.
25. A follow-up study will be conducted next year using 132 new patients. However, the patients will first be put into groups as measured by their **initial weight** (mildly obese, moderately obese, severely obese). After this classification is performed, the two diets will be randomized to subjects within each group. In this new study, what type of variable is initial weight?
- (a) lurking variable
 - (b) randomized variable
 - (c) blocking variable
 - (d) binary variable

26. In a telephone survey, we called 2,000 Columbia residents but we only get a complete interview from 400 residents. What is the **response rate**, measured as a percentage?

- (a) 20 percent
- (b) 40 percent
- (c) 50 percent
- (d) 95 percent

27. In the language of experiments, what is a **treatment**?

- (a) a data frame that consists of all experimental subjects
- (b) a group of subjects that are similar in some way
- (c) a condition that is applied to the subjects
- (d) an analysis for replicated studies

28. True or False. **Confidence statements** are statements applicable only to the sample of individuals measured.

- (a) True
- (b) False

29. What was the main point of the FATHERLY FRIENDLY FARMS exercise?

- (a) It is very important to have an institutional review board (IRB) review the results of a clinical trial.
- (b) Hollywood stars use statistics misleadingly to drive home their political agendas.
- (c) Measurements that we make on individuals can be biased and may not be reliable.
- (d) None of the above.

30. Starting at Line 150 in Table A (**Table of Random Digits**), I read off the following digits:

07511 88915 41267 16853

Our population consists of 50 subjects, labeled 01, 02, 03, 04, and so on, up through the last subject labeled 50. We would like to choose a simple random sample (SRS) of size $n = 3$. Reading from **left to right**, which subjects would be chosen?

- (a) 07, 51, 18
- (b) 07, 51, 88
- (c) 07, 18, 89
- (d) 07, 18, 15

31. Recently, I found the following question in an online poll (at www.foxnews.com):

Should states fund tuition for illegal aliens?

What is **wrong** with this question?

- (a) Political questions should never be asked.
- (b) All online questions should be worded in the past tense.
- (c) A negative connotation is attached to the word “illegal.”
- (d) Everyone agrees that illegal aliens should be deported.

32. What does it mean for a measurement to have **predictive validity**?

- (a) The measurement must be easy to observe.
- (b) The measurement must always be correct.
- (c) The measurement must be able to predict success on tasks that are related to the property being measured.
- (d) None of the above.

33. The number of sports-related injuries treated in US hospital emergency rooms in 2001 was recorded; we discussed these data in class. For fishing, there were 84,115 injuries treated. For ice hockey, there were 54,601 injuries treated. Based on this information, why can we **not** deduce that hockey is actually less injury-prone than fishing?

- (a) We did not take into account the number of individuals who participate in both sports.
- (b) Data confidentiality was not preserved.
- (c) We did not conduct a controlled experiment to study this phenomenon.
- (d) The difference in the number of injuries is not that large.

34. On HW#1, I asked you to visit www.misterpoll.com to examine several of the current polls on that site. Which statement **best** describes the polls located on at this site?

- (a) The results are trustworthy because most of the sample sizes were larger than 1000.
- (b) The results are trustworthy only if the results can be supported by other online polls.
- (c) The results are likely generalizable only to those who use the internet.
- (d) None of the above.

35. True or False. The margin of error formula we used for the sample proportion is only valid if the confidence level used is **95 percent**.

- (a) True
- (b) False

36. Before the 2004 Presidential election, the confidence statement was given:

“We are 95 percent confident that between 45.6% and 54.4% of likely voters in Iowa will vote for John Kerry.”

Assuming that this information was based on a random sample, what is the **margin of error** associated with this statement?

- (a) 2.2 percent
- (b) 4.4 percent
- (c) 8.8 percent
- (d) 17.6 percent

37. An advertisement for a home security system says,

“According to the FBI statistics, over 26 percent of home burglaries take place between Memorial Day and Labor Day.”

Which statement is likely to be **most accurate**? Recall that there are 14 weeks between Memorial Day and Labor Day (roughly the end of May through the beginning of September) and that there are 52 weeks during the year.

- (a) Burglars wait for the summer months when homeowners are often on summer vacation.
- (b) The statistics are likely not accurate because people are not at home to report crimes.
- (c) Authorities are too busy during the summer to suitably address home surveillance systems.
- (d) None of the above.

38. True or False. A **voluntary response sample** is an example of a probability sample.

- (a) True
- (b) False

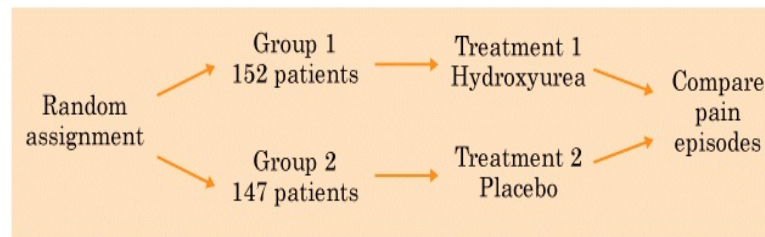
39. True or False. Nonsampling errors can not occur in a census.

- (a) True
- (b) False

40. True or False. Other things being equal, larger confidence levels provide a smaller margin of error.

- (a) True
- (b) False

41. Your textbook describes an experiment to compare the effectiveness of two treatments (drug/placebo) to subjects with sickle-cell anemia. The following design appears in your text:



Which experimental design best describes this experiment?

- (a) completely randomized design
- (b) randomized block design
- (c) matched pairs design
- (d) Latin-square design

42. In the language of clinical trials, to what does **equipoise** refer?

- (a) Randomization is used to assign subjects to drug treatments.
- (b) Patient records are kept confidential.
- (c) Patients are treated equally in every way possible (except for which treatment is assigned).
- (d) There is a genuine uncertainty about which treatment is best.

43. In carrying out a survey, which of the following would be best classified as a **sampling error**?

- (a) individual nonresponse
- (b) the use of poorly-worded questions
- (c) responses which are not truthful
- (d) undercoverage

44. Which one of the following characteristics is probably the **most difficult** to measure?

- (a) your weight
- (b) your shoe size
- (c) your desire to succeed in college
- (d) your systolic blood pressure

PART 2: SHORT ANSWER.

I asked you read the Judith Singer's *Chance* paper, "Afraid to Discuss Evolution?," and we discussed this paper in class. This paper summarized a *New York Times* article. Summarize the findings of Singer's paper and/or our classroom discussion. In particular,

- specifically state the topic discussed in the *New York Times* article
- identify the population of individuals targeted in the *New York Times* article
- state at least 3 things that led to Singer's attack of the article's questionable conclusion
- state at least 1 thing you learned from our discussion.

Successfully addressing each of these topics will earn you 12 points. Use clearly-written and complete sentences.