GROUND RULES:

- **Print** your full name clearly at the top of this page. Use the name that appears on university records.
- This is a closed-book and closed-notes exam. You can not use external notes of any kind.
- You may use a calculator. You may not use your phone as a calculator.
- This exam contains three parts:
  - Part 1. Multiple Choice. 22 questions, 2 points each (**44 points** total)
  - Part 2. Short Answer. 2 questions, 8 points each **(16 points** total)
  - Part 3. Extra Credit. 1 question (4 points total).

  This exam is worth **60 points** (but it is possible to get up to 64 points).
- Any discussion or inappropriate communication between you and another examinee, as well as the appearance of any unnecessary material, is not allowed. All violations will be reported to the Student Conduct and Academic Integrity Office immediately.
- You have **50 minutes** to complete this exam.

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 best answer. Make sure your answer is clearly marked. Ambiguous responses will be marked wrong.

1. A recent Quinnipiac survey asked a sample of American adults if they agreed with a statement about the severity of global warming. There were 1,001 Americans in the sample.

Quinnipiac reported a margin of error of $\pm 5\%$ points assuming a 95% confidence level. However, when I used our formula for margin of error, I obtained

$$\frac{1}{\sqrt{1001}} \approx 0.03 \text{ (or about 3\%)}.$$

What is the most likely reason for the discrepancy?

(a) Quinnipiac's estimate was probably determined from a question that was poorly worded.
(b) The survey design was probably biased.
(c) The sample of 1,001 Americans was probably not a simple random sample.
(d) The non-response rate was probably large.

2. In class, we talked about Dr. Judith Singer's article which described her reaction to a *New York Times* editorial on teaching creationism in Louisiana high schools. What statement most accurately summarizes Dr. Singer's analysis?

(a) The *New York Times* failed to acknowledge the study they cited included male students only.
(b) We need to be careful when performing research studies using subjects who are children.
(c) There was a statistically significant difference in exam performance between teachers who taught creationism and teachers who taught evolution.
(d) **Problems with non-response, interviewer bias, and question wording can greatly distort the results of a sample survey or census.**

3. A sample survey where the goal is to sample every individual in the population is called a

(a) population frame
(b) **census**
(c) meta-analysis
(d) control group

4. Does regular exercise reduce bone loss in post-menopausal women? A researcher finds 100 post-menopausal women (80 white; 20 non-white). She assigns white women to a regular exercise program. She assigns non-white women to a program that does not involve exercise. At the end of the study, she finds a statistically significant difference between the two groups in terms of bone loss reduction. However, her paper summarizing the experiment gets rejected from every journal. Why?

(a) A different number of white and non-white subjects was used.
(b) She did not use enough subjects.
(c) This is an observational study and it does not protect the confidentiality of the women.
(d) **She did not use randomization to assign women to the treatment groups.**

5. If properly performed, which type of research study can give the best evidence about **causality**?

(a) **a randomized comparative experiment**
(b) a randomized-response survey
(c) a retrospective study
(d) an observational study

6. I want to take a **simple random sample** (SRS) of 5 students from our class. How should I do this?

(a) **Assign each student a numerical code and obtain 5 randomly selected numbers. Those 5 students whose codes match the random numbers will constitute the sample.**
(b) Take the class roster I have on `my.sc.edu`, and select every 32nd name I see (there are 160 students in the class).
(c) Select 5 students that sit in the first two rows of the classroom.
(d) Send a mass email to the class asking for volunteers and choose the first 5 students who respond.

7. In class, we talked about the **replication crisis** in the social sciences and elsewhere. What does this refer to?

(a) Non-sampling errors prevent researchers from calculating a sample survey's margin of error.
(b) Researchers do not adhere to basic research ethics when performing experiments on human subjects.
(c) **Findings from published observational studies and experiments may be refuted in follow-up studies.**
(d) It is becoming more difficult to obtain accurate and precise estimates in observational studies.

8. The Registrar's Office at USC selects a simple random sample of 100 undergraduate students (from the 27,280 who attend USC) and records the number of hours each student is taking this semester. The average number of hours among the 100 students is 13.2. Is this number (13.2) a statistic or a parameter?

(a) **statistic**
(b) parameter

9. In an experiment, why is it important to use **randomization** when assigning subjects to treatments?

(a) **We would like treatment groups to contain subjects that are similar on average.**
(b) We would like our experiment to produce statistically significant results.
(c) We would like all subjects to be treated equally during the experiment.
(d) We would like to remove the placebo effect for the control group.

10. A researcher wants to estimate the proportion of all Hong Kong residents (aged 18-39) who have human papillomavirus infection. The sampling design will first divide the population into different groups based on age range:

- 18-24 years
- 25-29 years
- 30-34 years
- 35-39 years.

Simple random samples of 185 individuals will then be taken from each group. What type of **sampling design** is this?

(a) convenience sample
(b) cluster random sample
(c) **stratified random sample**
(d) voluntary response sample

11. We have talked throughout the semester about the necrotizing enterocolitis study where caffeine was given to premature infants. This was a carefully designed randomized comparative experiment.

(a) True
(b) **False**

12. Researchers want to study the well-being of college students, which is defined as being physically and mentally healthy and having a positive outlook on life. Here is how they will sample the students:

> "We will send mass emails to students at 6 institutes of higher learning (3 public universities and 3 private universities). Our research assistant will also approach students in public areas of the institutes at various parts of the day."

Which of the following is a **non-sampling error** in this example?

(a) Students don't respond to the email they receive.
(b) Students don't respond to the on-campus research assistant when approached.
(c) Students who do respond don't answer the well-being questions truthfully.
(d) **All of the above.**

13. I recently read research proposals on different topics. Here were the research questions posed in four different proposals:

(a) How does sleep quality affect academic performance in university students?
(b) **Does circumcision (removal of penis foreskin) lead to a decrease in sexually transmitted diseases?**
(c) Do fertilizers with extra nitrogen produce crops which are more resistant to drought?
(d) How do leadership styles impact employee retention?

In which of these proposals would it **not** be appropriate to do a randomized comparative experiment?

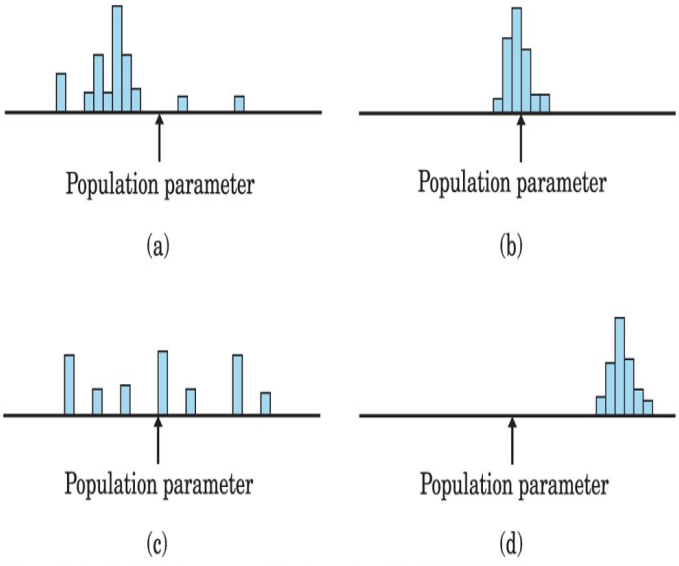
14. I read the following survey question recently online:

> *"Do you agree with the following statement: The federal government is too large and has gotten totally out of control and threatens our basic liberties. Should we or should we not clear house and commit to drastic change?"*

Using this question in a survey will likely lead to bias caused by a serious

(a) random sampling error
(b) margin of error
(c) **non-sampling error**
(d) frame error

15. Here are four distributions which contain many simulated values of a statistic (like the simulation we did in class). The population parameter in each distribution is identified.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Which distribution corresponds to a statistic which is biased but has small variability? Circle the correct answer above. **D**

16. A survey's **margin of error** accounts for uncertainty in the sample proportion. What specifically does it account for?

(a) **random sampling error**
(b) the bias caused by sampling from an inaccurate population frame
(c) the bias caused by participants who did not respond
(d) All of the above

17. Gallup News Service conducted a poll during October 5-11, 2017 with 1,028 American adults (aged 18 and over), asking,

Are you in favor of the death penalty for a person convicted of murder?

Fifty-five (55%) of the 1,028 adults responded "Yes" to this question. In this example, what is the **population**?

(a) the 55% who responded "Yes"
(b) the 1,028 adults who were asked the question above
(c) all American adults in favor of the death penalty
(d) **all American adults**

18. A researcher is performing a study to assess new parental leave policies and the incidence of postpartum depression among recent mothers. She records many variables. Two of them are the mother's age (in years) and the most recent birth type (natural birth or Cesarean section). Which types of variables are these?

(a) Both variables are quantitative.
(b) Both variables are categorical.
(c) **Mother's age is quantitative and birth type is categorical.**
(d) Mother's age is categorical and birth type is quantitative.

19. In class, we discussed the problems associated with the *Literary Digest* survey. What topic did this survey address?

(a) the damaging effects of excessive social media exposure
(b) **the 1936 presidential election between Landon and Roosevelt**
(c) if low covid-19 vaccination rates were related to socio-economic status
(d) whether celebrating Christmas should be allowed in public schools

20. The **randomized-response technique** would be best suited for which of the following survey questions?

(a) Did you attend a professional sporting event during the past year?
(b) Do you think crime has increased nationally since President Biden took office?
(c) Do you drive an American-made automobile?
(d) **Did you cheat on your taxes last year?**

21. A sixth-grade teacher wants to compare two methods of teaching fractions to his students. One of his classes will receive a more traditional method of instruction. His other class will learn fractions using a "new math" approach that the Department of Education now recommends. This study will last the entire semester.

In this example, what does "comparative design" mean?

(a) **Students in different classes will be treated and evaluated in the same way during the entire semester.**
(b) The teacher will use different examinations that are written by experts in each modality of instruction.
(c) Before the semester starts, students will be randomized to the different classes.
(d) The teacher will match each student in one class with a similar student in the other class.

22. Researchers recruited 100 amateur boxers to participate in a study that compared rest and massage before boxing. After a 10-minute workout where each boxer threw 400 punches, the boxers were randomized to one of two groups:

- **Group 1:** Massage (boxer receives a 20 minute massage)
- **Group 2:** Rest (boxer rests for 20 minutes).

Before they returned for a second workout, the heart rate (beats per minute) was measured.

Is this an experiment or an observational study?

(a) **experiment**
(b) observational study

PART 2: SHORT ANSWER. Give detailed responses. Please write clearly and legibly.

1. Rasmussen Reports recently conducted a national survey using a simple random sample (SRS) of 1,224 American adults. Each participant was asked:

How likely is it that your next automobile purchase will be an electric car?

The survey found that 355 of the 1224 adults in the sample said it was likely their next automobile purchase will be an electric car.

(a) Calculate the sample proportion of adults who say it is likely their next automobile purchase will be an electric car.

Answer:

$$\widehat{p} = \frac{355}{1224} \approx 0.29 \quad (\text{or } 29\%).$$

(b) Rasmussen reports "the margin of sampling error is ±3 percentage points with a 95% level of confidence." Do a calculation that shows where the "±3 percentage points" figure comes from.

Answer:

$$\textbf{margin of error} = \frac{1}{\sqrt{1224}} \approx \frac{1}{34.99} \approx 0.03 \quad (\text{or } 3\%).$$

(c) Write a 95% confidence statement. A confidence statement is a well-written sentence.

Answer: We are 95% confident the proportion of all American adults whose next automobile purchase will be an electric car is between 0.26 and 0.32 (or 26% and 32%).

2. The Bucharest Early Intervention Project was an experiment carried out in Bucharest, Romania. The goal was to compare children who were assigned to foster care (families with foster parents) to children who were assigned to live in institutional care facilities. There were 136 children in the experiment (aged 6-31 months). There were two groups:

- **Group 1:** Foster care
- **Group 2:** Institutional care.

(a) What is the "treatment" in this experiment? **Answer:** Type of care

(b) The children were randomly assigned to the different groups (68 children per group). Describe how this could be done. Be specific.

Answer: We could label the children with codes from 001-136. We could then use the Table of Random Digits to select the first 68 children whose codes match the random numbers identified. These children would be assigned to Group 1. The other 68 children would be assigned to Group 2.

You could also use R to do this. Select a SRS of 68 codes between 1 and 136. The children identified by these codes are assigned to Group 1. The remaining children are assigned to Group 2.

(c) One of the response variables was the child's IQ score at the age of 18. Give two lurking variables which may prevent investigators from establishing a causal link between care type and IQ score.

Answer: There are many possible answers here. For example, sex/gender, race/ethnicity, type of school attended, genetic traits, psychological factors or learning disabilities, characteristics of the biological parents, etc.

(d) When comparing the IQ score averages, the difference between the two care type groups was highly statistically significant (even after adjusting for lurking variables). Explain what this means.

Answer: The difference between the averages of the two groups was so large that it was likely not due to random chance.

PART 3: EXTRA CREDIT. Give a detailed response. Please write clearly and legibly.

On the first day of class, I showed you a quote by John Tukey,

> *"Statisticians get to play in everyone else's back yard."*

Think about all of the examples we have talked about so far this semester (and even on this exam). What do you think Dr. Tukey meant when he said this?

Answer: Statistics is used in all disciplines. For example,

- medicine like the NEC/caffeine example
- education like the sex education with at-risk kids and modality of instruction (online/in-person classes) examples
- agriculture like the pig/salmonella example
- political science and sociology to take political polls and polls on current events
- engineering like the board drying time example.

All areas use data so they benefit from statistical thinking. Statisticians get to interact with researchers from other areas and help them with their studies and experiments.