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.
- Moore and Notz's Table B is provided on the last page of this exam.
- This exam contains 54 multiple choice questions.
  - Each question with worth 2 points.

  This exam is worth **100 points** (but it is possible to get up to 108 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 **2.5 hours** 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.*

HELPFUL FORMULAS

$$\textbf{margin of error} = \frac{1}{\sqrt{n}}$$

Measured value = True value + Bias + Random error.

$$\text{percentage change} = \frac{\text{amount of the change}}{\text{starting value}} \times 100\%.$$

$$\overline{x} = \frac{1}{n}\sum x \qquad s = \sqrt{\frac{1}{n-1}\sum (x-\overline{x})^2}$$

$$z = \frac{\text{observation} - \text{mean}}{\text{standard deviation}}$$

$$r = \frac{1}{n-1}\sum \left(\frac{x-\overline{x}}{s_x}\right)\left(\frac{y-\overline{y}}{s_y}\right) \qquad y = a + bx$$

$\widehat{p}$ is (approximately) normal with mean p and standard deviation $\sqrt{\dfrac{p(1-p)}{n}}$

$$\widehat{p} \pm z^* \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

C	80%	90%	95%	99%
z^*	1.28	1.64	1.96	2.58

$\overline{x}$ is (approximately) normal with mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$

$$\overline{x} \pm z^* \left(\frac{s}{\sqrt{n}}\right)$$

MULTIPLE CHOICE. Circle the best answer. Make sure your answer is clearly marked. Ambiguous responses will be marked wrong.

1. An ornithologist (someone who studies birds) observes a sample of 30 birds and measures the following variables on each bird:

$$\begin{aligned} x &= \text{ambient temperature (deg F)} \\ y &= \text{number of calories (Kcal).} \end{aligned}$$

She calculated the correlation to be $r = -0.70$ and the least-squares regression line to be

$$y = 38.2 - 0.36x.$$

What percentage of the variability in the number of calories measurements is explained by straight-line relationship with ambient temperature?

(a) 36%
(b) 38.2%
(c) 49%
(d) 84%

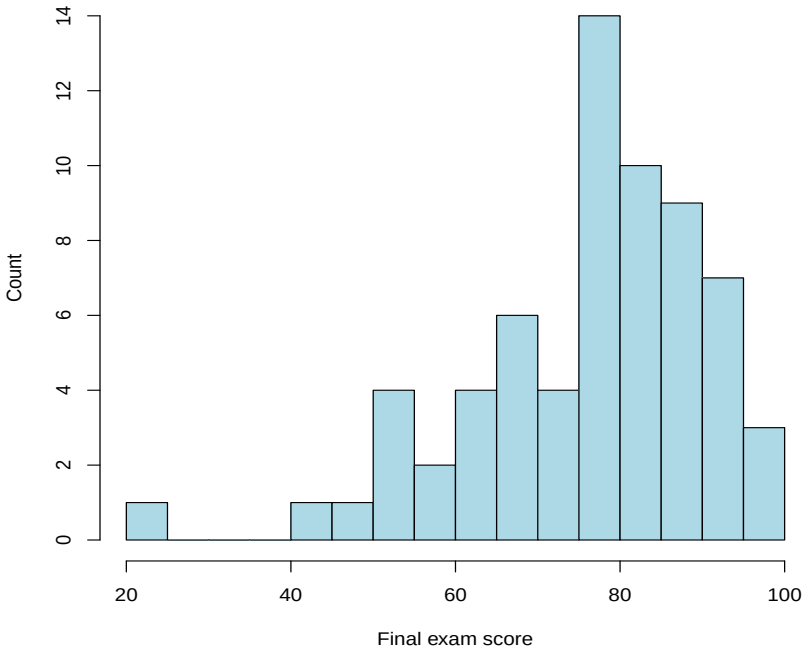
2. Moore and Notz's Table A, the "Table of Random Digits," consists of digits 0-9 which appear in random order. In which scenario would this table be useful?

(a) deciding which subjects will be assigned to different treatment groups in a comparative experiment
(b) assessing the level of bias and variability in responses from a convenience sample
(c) determining whether there is a causal relationship between two categorical variables
(d) assessing whether a quantitative variable is measured with bias and random error

3. In the language of clinical trials, what is meant by the term **equipoise**?

(a) The institutional review board ensures that all patients are treated in the same way and are protected from being harmed.
(b) Patients must be treated equally throughout the course of the trial.
(c) There is uncertainty about which treatment may benefit the patient the most.
(d) Equal randomization should be used in any setting comparing drugs or medical interventions.

4. In class, we looked at the distribution of final exam scores for 66 students I had in an introductory statistics course when I was a professor at Oklahoma State University in 2002:



A **density curve** which approximates this histogram would refer to what?

(a) the final exam score if I gave the same exam to another class of 66 students
(b) the final exam score distribution adjusted after accounting for midterm grades
(c) the final exam score distribution for a larger population of students from which my 66 students were sampled
(d) the final exam score distribution for those students who passed the final exam

5. In comparative drug studies, double-blind experiments are typically used. The purpose of keeping the diagnosing physicians ignorant of the treatment status of the subjects is to

(a) ensure subjects are randomly assigned to treatments.
(b) reduce sampling error.
(c) eliminate a possible source of bias.
(d) remove grounds for malpractice suits if statistically significant results are not determined.

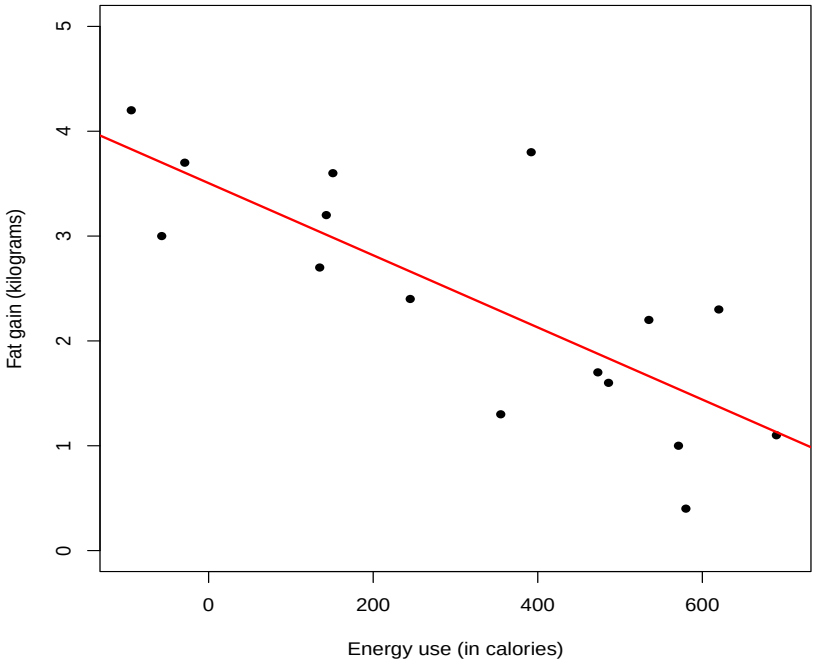
6. In an obesity study, researchers deliberately overfed 16 young healthy adults for 8 weeks. Afterwards, they measured the following variables on each adult:

$$\begin{aligned} x &= \text{amount of energy used in non-exercise activities (in calories)} \\ y &= \text{amount of fat gain (in kilograms).} \end{aligned}$$

I used R to calculate the equation of the least-squares regression line:

$$y = 3.5 - 0.0035x.$$

Below is a scatterplot of the data with the least-squares regression line superimposed.



What would you predict the fat gain to be for a young healthy adult who burns 400 calories during non-exercise activities?

(a) 3.8 kilograms
(b) 3.5 kilograms
(c) 2.1 kilograms
(d) 1.4 kilograms

7. Refer to the previous question (see previous page). Add a new observation to the scatterplot corresponding to a 17th adult:

$$\begin{aligned} x &= 0 \text{ calories} \\ y &= 0.5 \text{ kilograms.} \end{aligned}$$

If you added this observation and recalculated the least-squares regression line,

(a) the intercept would decrease and the slope would increase.
(b) the intercept would increase and the slope would increase.
(c) the intercept would decrease and the slope would decrease.
(d) the intercept would increase and the slope would decrease.

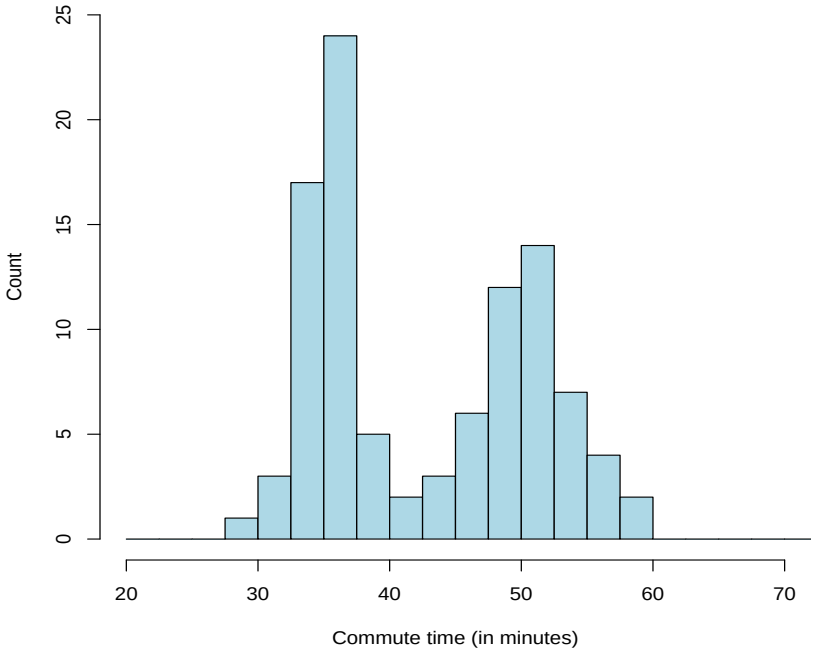
8. You are designing an experiment to see whether adding calcium to one's diet will reduce the blood pressure of middle-aged men. You plan to give added calcium to a group of men and see if their blood pressure is lower after six weeks. A control group with placebo will also be used. Preliminary information suggests the effect of adding calcium may be different for black men than for white men. Knowing this, you should use a

(a) randomized block design, with black men and white men as the blocks.
(b) a stratified design, where black men receive calcium and white men receive placebo.
(c) completely randomized design to ensure the treatment groups are similar on average in terms of race.
(d) randomized block design, with calcium and placebo as the blocks.

9. Researchers have long recognized that questions about sexuality are sensitive and participants may not be willing to reveal truthful responses to these questions. Which type of survey method would be useful if you wanted to conduct a sample survey with a question about one's sexuality? You would want your method to maintain anonymity of individuals' answers to the question.

(a) retrospective design with random matching
(b) randomized response technique
(c) randomized complete block design with sex as a block
(d) completely randomized design

10. I drive in to campus every day from Blythewood. Here is a histogram of my driving commute times (in minutes) for 100 days during 2019.



The shape of the histogram is identifying two different distributions—one that corresponds to shorter commute times and one that corresponds to longer commute times. What are the mean times for each distribution?

(a) 30 and 60 minutes
(b) 35 and 50 minutes
(c) 40 and 45 minutes
(d) 45 and 55 minutes

11. During October 1-23, 2023, Gallup conducted a telephone survey using a sample of 1,009 American adults. Each participant was asked,

Is there more crime in your area than there was a year ago, or less?

The survey found that 55% of the adults in the sample said there was more crime than a year ago. The number 55% is a

(a) statistic
(b) parameter

12. Government officials in the United Kingdom collected data on the counts and rates of maternal deaths for vaginal delivery and Cesarean delivery over a recent 3-year period. Here is a table of the data:

Delivery type	# of deliveries	# of deaths	Death rate per 100,000 deliveries
Vaginal	1,571,000	75	4.8
Cesarean	425,000	73	17.2

A government official says vaginal deliveries are more dangerous because a larger number of deaths occurred over this 3-year span. She is using a measurement which is

(a) biased.
(b) unreplicated.
(c) unreliable.
(d) invalid.

13. When we used the sample proportion $\widehat{p}$ to estimate the population proportion p, we learned $\widehat{p}$ was unbiased if the sampling design was a simple random sample. What does it mean for a statistic to be **unbiased**?

(a) Its value will be the same in every possible simple random sample.
(b) The sampling distribution of the statistic is symmetric.
(c) Its margin of error is 0.
(d) It neither systematically overestimates nor underestimates the population parameter on average.

14. Researchers randomly assigned 60 severe-stroke patients to receive either

1. tPA (a standard treatment for stroke)
2. tPA + blood chilling.

The researchers concluded, "the difference in recovery times between the two treatments was **not** statistically significant." What does this conclusion mean?

(a) The sample sizes (30 per treatment) were too small.
(b) More treatment groups should be used to obtain a statistically significant difference.
(c) The difference between the results for each group could be explained by random chance.
(d) The experiment was biased because researchers used severe-stroke patients only.

15. People with diabetes must control their blood glucose level. In a recent study, the fasting glucose level (mg/dl) was measured on a sample of $n = 15$ diabetes patients. Here are the data, already ordered from low to high:

```
78  95 103 112 134 141 145 147 153 158 172 172 200 271 359
```

Here is the five-number summary for these data:

```
quantile(glucose,type=2)
 0%  25%  50%  75% 100%
 78  112  147  172  359
```

How many outliers are in the data set? Use the 1.5(IQR) rule of thumb we talked about in class.

(a) 0
(b) 1
(c) 2
(d) 3 or more

16. A cardiologist observes a simple random sample of $n = 30$ adult white males and measures the systolic blood pressure (SBP, measured in mm Hg) on each male. The average of the 30 measurements was reported to be $\overline{x} = 114.9$ and the standard deviation was $s = 9.3$. A 95% confidence interval calculated from these data using

$$\overline{x} \pm z^* \left(\frac{s}{\sqrt{n}} \right)$$

gives $(111.6, 118.2)$. How do we interpret this interval?

(a) Ninety-five percent (95%) of the males in the sample had a SBP between 111.6 and 118.2 mm Hg.
(b) When considering the population of all adult white males, we are 95% confident the mean SBP of this population is between 111.6 and 118.2 mm Hg.
(c) The probability one randomly selected adult white male will have a SBP between 111.6 and 118.2 mm Hg is 0.95.
(d) Ninety-five percent (95%) of all adult white males will have a SBP between 111.6 and 118.2 mm Hg.

17. Researchers summarized the results of a randomized comparative experiment with two doses of fremanezumab and a placebo. Patients with chronic migraine were randomly assigned to one of three treatment groups:

1. Group 1: Receive fremanezumab quarterly
2. Group 2: Receive fremanezumab monthly
3. Group 3: Receive placebo.

Why is it so critical to use **randomization** when assigning subjects to one of the three treatment groups?

(a) We want to ensure all subjects are treated equally during the course of the experiment.
(b) We want to create three groups of subjects which are similar on average at the beginning of the experiment.
(c) We want to remove all lurking variables.
(d) We want to guarantee the results from the experiment will be statistically significant.

18. An editorial in the *New York Times* claims,

> *"An American woman is beaten by her husband or boyfriend every 15 seconds."*

This claim would correspond to 4 women per minute, 240 women per hour, 5,760 women per day, and **2,102,400 women per year**. However, this yearly count sharply disagrees with the results from the National Crime Victimization Survey, which reliably puts the yearly count of domestic violence at around 250,000 women per year.

Dr. Best, the author of *Damned Lies and Statistics*, would attribute this large discrepancy to some type of

(a) innumeracy
(b) random sampling error
(c) misclassification
(d) blinding

19. An orthopedic health researcher wants to compare the following medical interventions for individuals with severe knee pain due to inflammation:

- Group 1: Aggressive (epidural steroid injection into the knee)
- Group 2: Conservative (physical therapy on the knee)
- Group 3: Control (ibuprofin only).

Which could **not** be used in this experiment?

(a) randomization
(b) blocking
(c) replication
(d) blinding

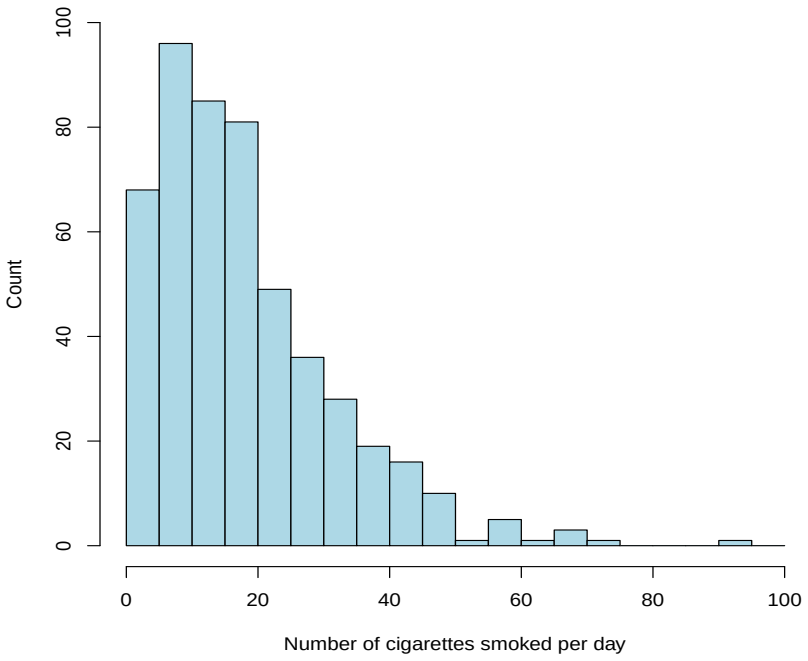
20. An advertisement for a new heartburn treatment says that it "reduces heartburn by 300%." What does this mean?

(a) The new treatment will offer more protection against a competing treatment that only reduces heartburn by 200%.
(b) Those taking the treatment are 300 times less likely to experience heartburn when compared to those who do not take the treatment.
(c) Those who do not take the treatment will experience 300% more heartburn.
(d) None of the above.

21. Sample surveys often involve contacting individuals by phone or over email. What is usually the largest source of nonsampling error in these types of surveys?

(a) probability weighting
(b) random sampling error
(c) nonresponse
(d) confounding

22. Researchers took a simple random sample (SRS) of $n = 500$ adult smokers in South Carolina. For each smoker in the sample, researchers recorded the number of cigarettes smoked per day. A histogram of the 500 observations is shown below:



Why is the mean number of cigarettes smoked per day larger than the median?

(a) The mean is resistant to outliers. The median's value is highly sensitive to them.
(b) The few large observations on the high side will inflate the value of the mean. The median will be largely unaffected by these values.
(c) The distribution is skewed to the left.
(d) The mean will increase if the standard deviation increases. The median doesn't depend on the standard deviation.

23. A brand of storage batteries has lifetimes which follow a normal distribution with mean $\mu = 3.2$ years and standard deviation $\sigma = 0.4$ years. Approximately what percentage of the batteries will have lifetimes between 2.8 and 3.6 years?

(a) 32%
(b) 68%
(c) 81.5%
(d) 95%

24. A designed experiment used 864 subjects who were at risk for colon cancer. Each subject was randomly assigned to one of four antioxidant groups:

- Group 1: Daily beta-carotene
- Group 2: Daily vitamins C and E
- Group 3: Daily beta-carotene + vitamins C and E
- Group 4: Control.

Each subject was followed for four years. Researchers documented whether each subject had been diagnosed with colon cancer. In this experiment, what is the **treatment**?

(a) whether a subject had been diagnosed with colon cancer
(b) the four-year follow-up period under observation
(c) the antioxidants
(d) the 864 subjects

25. Which statement about the correlation r is **false**?

(a) The correlation r is a number between 0 and 1.
(b) The correlation r describes linear relationships only.
(c) The correlation r is the same between x and y as it is between y and x.
(d) The correlation r does not have any units attached to it.

26. The Pew Research Center conducted a poll during October 25-30, 2017, where they asked participants,

Do you think the use of marijuana should be made legal?

Pew reported their poll consisted of telephone interviews with 1504 randomly selected American adults, and 752 of the 1504 adults (50%) answered "Yes" to the question above. What is the **sample**?

(a) all American adults who think marijuana should be made legal
(b) the 752 adults who answered "Yes" to the question above
(c) the 1504 adults who completed the telephone interview
(d) all American adults

27. What was the main point of the FATHERLY FRIENDLY FARMS example we did in class?

(a) The Hawthorne effect can make the results observed in controlled experiments differ from real-world implementation.
(b) Ethical issues can arise even when experimenting on animals.
(c) Distributions of data can reveal visual patterns the raw data cannot convey.
(d) Measurements of quantitative variables can suffer from bias and random error.

28. Hurricane Hugo was a Category 4 hurricane with sustained winds of 140 mph when it came inland during the early morning hours of September 22, 1989. A local meteorologist, who is predicting whether a Category 4 or 5 hurricane will hit South Carolina, recently said on the nightly news,

> *"It has been almost 35 years since Hurricane Hugo hit South Carolina in 1989. Based on our weather models, the law of averages says that we are probably due for another Category 4 or 5 hurricane in the next year or so."*

How is the meteorologist making an error in this statement?

(a) A repeat of Hurricane Hugo has probability 0. The law of averages requires probabilities larger than 0.
(b) The meteorologist is making an incorrect statement about what happens in the short term. The law of averages gives an assessment of long-term behavior.
(c) The law of averages only applies for quantitative variables−not whether or not something will happen (categorical).
(d) The law of averages can only be used to make statements about hurricane activity in the past. It can not be used to calculate a future probability.

29. The University of South Carolina has about 28,000 undergraduate students and 7,000 graduate students on the Columbia campus. A financial aid officer wants to select a sample of 300 undergraduate students and 100 graduate students. Which sampling design should she use?

(a) a cluster sample
(b) a stratified random sample
(c) a convenience sample
(d) a simple random sample

30. True or False. For any normal distribution, the mean and median are equal.

(a) True
(b) False

31. In the other class I teach this semester (STAT 512), we recently calculated the probability of an outcome to be 0.25. Which statement is true?

(a) The outcome will occur more often than not.
(b) If the outcome doesn't occur on the next three attempts, then it will on the fourth attempt.
(c) It would be highly unusual if this outcome occurred.
(d) None of the above.

32. University officials at the Russell House are trying to determine whether students like Five Guys or In-N-Out hamburgers better. They have a budget to recruit 40 students. Which design would allow for the most accurate and precise comparison?

(a) Have male students eat Five Guys. Have female students eat In-N-Out. Compare the groups with sex as a block.
(b) Randomly assign 20 students to eat Five Guys. The remaining 20 students eat In-N-Out. Compare the groups.
(c) Use a matched pairs design where all 40 students will eat both burgers. The order of which burger is eaten first is randomized for each student. Compare the groups.

33. Which of the following would be **unethical** in an experiment involving human subjects?

(a) not obtaining informed consent from the subjects
(b) releasing data which identifies names and addresses of the subjects
(c) placing the interests of science over the interests of the subjects
(d) all of the above.

34. A March 2018 Gallup survey asked a sample of 1,041 American adults if they wanted stricter laws covering the sale of firearms. A total of 697 of the survey respondents said "Yes" to this question. Assuming this was a simple random sample (SRS), I calculated a 95% confidence interval for the population proportion of all American adults who want stricter laws to be $(0.64, 0.70)$. How would a 90% confidence interval compare to a 95% confidence interval?

(a) It would be 10% longer than the 95% confidence interval.
(b) It would be the same length because the sample size is large.
(c) It would be 90% longer than the 95% confidence interval.
(d) It would be shorter than the 95% confidence interval.

35. An observational study examined iris (eye) coloration in a population of North American hawks. Here is a probability model for the iris color of female hawks in this population:

Outcome	Yellow	Light orange	Medium orange	Dark orange	Red
Probability	0.05	0.29	0.47	0.18	0.01

According to this model, what is the probability a randomly selected female hawk has an iris color that is **not** some shade of orange?

(a) 0.0005
(b) 0.06
(c) 0.94
(d) None of the above

36. The average score on an exam in organic chemistry was 60. The professor gave a retest and reported the average score increased by 25%. This means the average score on the retest was

(a) 45
(b) 65
(c) 75
(d) 85

37. Rasmussen Reports recently conducted a national telephone and online survey using a simple random sample (SRS) of $n = 2100$ American adults. Each participant was asked:

Do you think movies have gotten better in the past 20 years?

Suppose the population proportion of all American adults who think movies have gotten better in the past 20 years is $p = 0.30$. For samples of this size ($n = 2100$), 95% of all simple random samples will produce a sample proportion $\widehat{p}$ that falls between which two values?

(a) 0.295 and 0.305
(b) 0.28 and 0.32
(c) 0.25 and 0.35
(d) 0.20 and 0.40

38. For a sample of 100 healthy dogs, a veterinarian measured the glucose concentration in the right eye and also in the blood serum. The data below are the eye measurements as a percentage of the blood measurements.

```
> stem(glucose)

   6 | 69
   7 | 04
   7 | 77789
   8 | 00000122222333444
   8 | 55555556777788889999
   9 | 000011111111222233333444
   9 | 55555666677888889
  10 | 001111123
  10 | 59
  11 | 0
```

The stems above denote the tens digit and the leaves denote the units digit.

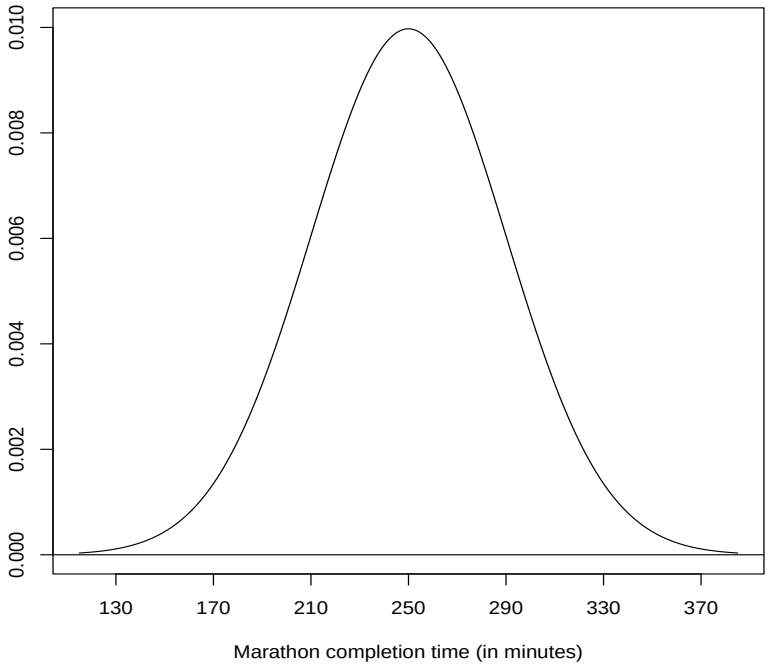
Glucose percentages at 100 or more identify dogs who are likely to become blind in the future. What percentage of dogs in the sample have glucose percentages that are larger than or equal to 100?

(a) 3 percent
(b) 5 percent
(c) 12 percent
(d) 30 percent

39. We talked about confidence statements for a population proportion p in Chapter 3, and we wrote many of them this semester. What are the two parts of a confidence statement?

(a) proportion and mean
(b) confidence level and margin of error
(c) center and spread
(d) bias and random sampling error

40. For a population of runners, the time it takes to complete the New York Marathon (in minutes) is normally distributed with mean $\mu = 250$ and standard deviation $\sigma = 40$. This population density curve is shown below:



What percentage of runners in this population will finish the marathon in longer than (more than) 310 minutes?

(a) 98.5%
(b) 93.32%
(c) 6.68%
(d) 1.5%

41. Before I wrote this question, I looked at my Rate My Professors web site profile. I observed that 52 students had given me ratings during the last 18 years for different courses I have taught at USC (over which time I have taught about 1,800 students). The collection of 52 students is best regarded as a

(a) simple random sample
(d) voluntary response sample
(c) probability sample
(b) stratified random sample

42. Suppose I administered this exam in two different rooms at the same time. I randomly assign each of my 133 students to one of the two rooms:

- Room 1: Classical music playing in the background
- Room 2: Hard rock music playing in the background.

Each student takes the exam in the room assigned to him/her. What type of experimental design is this?

(a) cluster sampling design
(b) completely randomized design
(c) matched pairs design
(d) randomized complete block design

43. I want to examine my credit card balance for each month between January 2000 through December 2023. Which graphical display would show me how my monthly credit card balance changes over time?

(a) line graph
(b) boxplot
(c) histogram
(d) stemplot

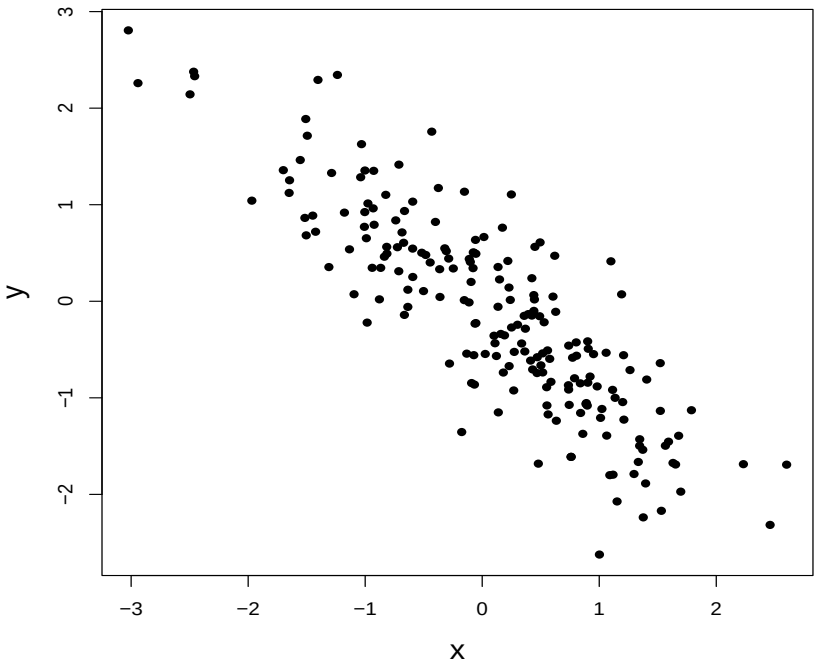
44. A large study used past records from Canada's national health care system to compare the effectiveness of two approaches to treat prostate cancer:

- Approach 1: Traditional surgery
- Approach 2: Radiation without surgery.

Investigators obtained records for thousands of past patients whose doctors had selected one approach or the other. The study found patients treated using Approach 2 were more likely to die within 10 years. Is this an experiment or an observational study?

(a) Observational study. The study aimed to determine a causal link between the approach to treat patients and 10-year survival with a large number of patients.
(b) Observational study. Investigators merely observed records from the past. Patients were not randomized to the two different approaches.
(c) Experiment. The "more likely" conclusion references a statistically significant result, which is only possible to determine in an experiment.
(d) Experiment. The study included two groups, and the investigators were interested in comparing the groups.

45. A scatterplot of two quantitative variables is shown below:



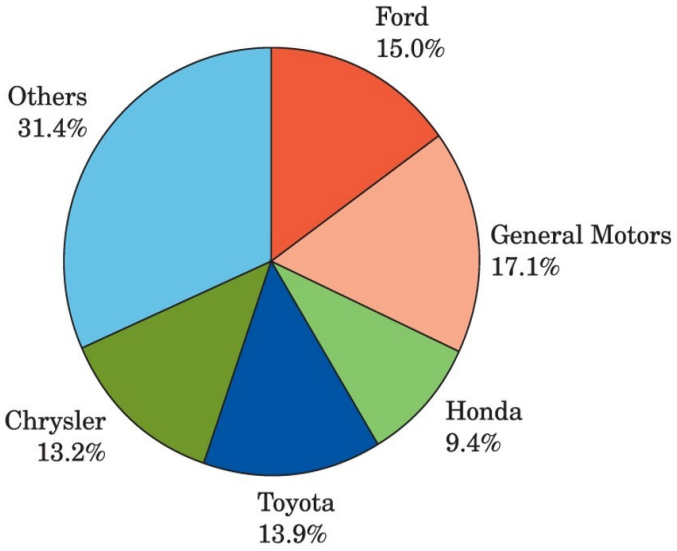
What is the correlation r closest to?

(a) -1.45
(b) -0.87
(c) -0.14
(d) -0.02

46. Next semester, your friend is taking a statistics class, and his instructor is making him calculate the standard deviation. For one data set, his answer is 14.2. He asks you what this number means in words. What would your answer be?

(a) It is the average distance from the mean if you averaged this distance over all the observations.
(b) It is the difference between the largest and smallest observations.
(c) It is the range over which the middle 50% of the data fall.
(d) It is the location of where a histogram of the data would balance if it were made of solid material.

47. Moore and Notz show the following pie chart of the percentage of passenger car sales in 2016 by various manufacturers:



What other graphical display could we use to display the distribution of this variable?

(a) line graph
(b) bar graph
(c) boxplot
(d) scatterplot

48. The Environmental Protection Agency sends four identical samples of polluted water to a laboratory for analysis. The lab measures the amount of dioxin in each sample. The lab reports dioxin concentrations of 141, 323, 74, and 923 parts per billion. Dioxin measurements made by the lab are

(a) misclassified.
(b) confounded.
(c) not reliable.
(d) unbiased.

49. The College of Arts and Sciences (CAS) consists of departments in the natural sciences, the social sciences, and the humanities/arts. The Dean wants to form a committee by selecting the chairs of three ($n = 3$) departments using a **simple random sample**. There are 21 departments in CAS:

Department	Code	Geography	Code	Department	Code
AA Studies	01	Geography	08	Psychology	15
Anthropology	02	History	09	Religious Studies	16
Biology	03	LLC	10	Sociology	17
Chem/Biochem	04	Mathematics	11	Statistics	18
Criminal Justice	05	Philosophy	12	Theatre/Dance	19
EOE	06	Physics	13	Vis Art/Design	20
English	07	Political Sci	14	Women's/Gen Studies	21

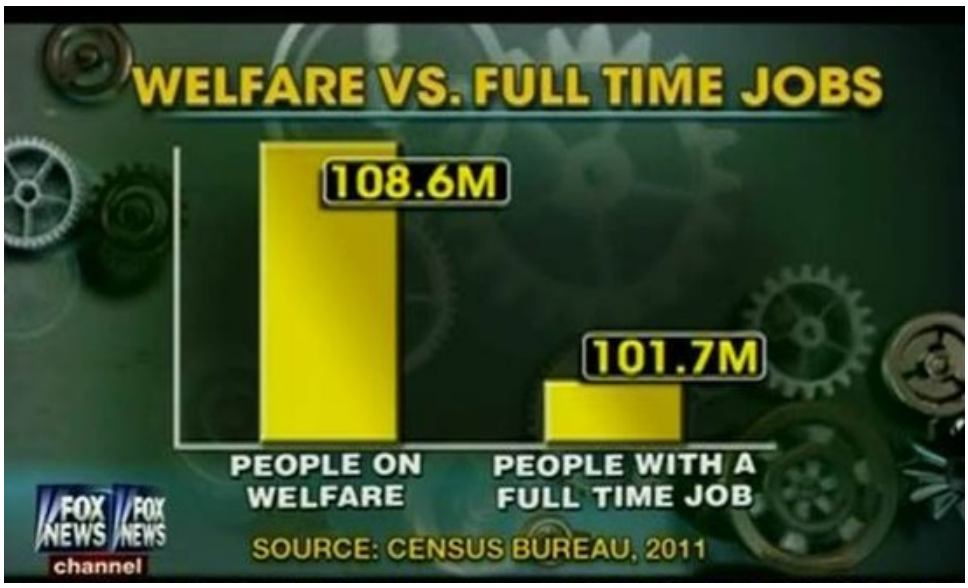
What is the defining characteristic of a simple random sampling design?

(a) This design will ensure each possible sample of 3 chairs has the same chance of being selected.
(b) This design will produce 3 chairs who are guaranteed to be representative of the population of chairs.
(c) This design will favor the selection of department chairs from certain departments over others.
(d) This design will give a sample of one chair from the natural sciences, one from the social sciences, and one from the humanities/arts.

50. A sociologist in Atlanta observes that ice cream sales and violent crime rates have a moderate to strong correlation ($r = 0.62$). Which of the following statements is **true**?

(a) The positive correlation is probably purely coincidental.
(b) As the values of one variable increases, the values of the other variable tend to decrease.
(c) Selling more ice cream causes more violent crime.
(d) There is likely a third variable, like temperature, which influences the values of both variables.

51. In August 2012, Fox News displayed this graph on one of its nightly programs:



Why is this graph misleading?

(a) Because there are two groups being compared, the viewer cannot tell if the difference between the groups is statistically significant.
(b) The bar height for welfare is about 4 times larger than the height for the full time job group. The difference in counts between the two groups is nowhere near that.
(c) A scatterplot between welfare recipients and people with a full time job would show a better description of the data.
(d) Counts of people are quantitative in nature. Fox News should have used a stemplot or a boxplot to display these two figures.

52. The World Health Organization uses a normal distribution with mean $\mu = 125$ and standard deviation $\sigma = 15$ as probability model for the systolic blood pressure (SBP) of American males. Last week when I visited the cardiologist, my SBP was 131, which corresponds to a standard score of $z = 0.4$. What does this mean?

(a) The probability my SBP is above the mean is 0.4.
(b) 40 percent of all American males have a SBP higher than mine.
(c) 40 percent of all American males have a SBP lower than mine.
(d) My SBP is 0.4 standard deviations above the mean SBP.

53. In class, we talked about the necrotizing enterocolitis study which involved 615 premature infants at Richland Hospital. Two variables recorded on each infant were the **birth weight** (in grams) and the **race** of the mother (AA, Hispanic, White, Other). Which types of variables are these?

(a) Birth weight is categorical and race is quantitative.
(b) Both variables are categorical.
(c) Both variables are quantitative.
(d) Birth weight is quantitative and race is categorical.

54. What was the unethical behavior in the **Duke-Potti example** we discussed in class? Recall we watched a 60 Minutes clip on this example.

(a) Dr. Potti lobbied the US government to fund his medical research, but he instead illegally funneled tax-payer money to his colleagues in North Dakota.
(b) Dr. Potti's medical experiments did not use blinding. This amplified the Hawthorne effect, making his experiments lack realism.
(c) Dr. Potti administered cancer treatments to poor people in Africa who had not given informed consent.
(d) Dr. Potti's primary discovery in research was a fraud. He manipulated data on patients to make his theory work.

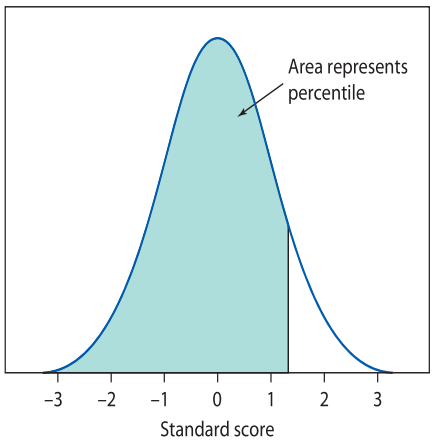


Table B Percentiles of the Normal distributions

Standard score → Percentile		Standard score → Percentile		Standard score → Percentile	
−3.4	0.03	−1.1	13.57	1.2	88.49
−3.3	0.05	−1.0	15.87	1.3	90.32
−3.2	0.07	−0.9	18.41	1.4	91.92
−3.1	0.10	−0.8	21.19	1.5	93.32
−3.0	0.13	−0.7	24.20	1.6	94.52
−2.9	0.19	−0.6	27.42	1.7	95.54
−2.8	0.26	−0.5	30.85	1.8	96.41
−2.7	0.35	−0.4	34.46	1.9	97.13
−2.6	0.47	−0.3	38.21	2.0	97.73
−2.5	0.62	−0.2	42.07	2.1	98.21
−2.4	0.82	−0.1	46.02	2.2	98.61
−2.3	1.07	0.0	50.00	2.3	98.93
−2.2	1.39	0.1	53.98	2.4	99.18
−2.1	1.79	0.2	57.93	2.5	99.38
−2.0	2.27	0.3	61.79	2.6	99.53
−1.9	2.87	0.4	65.54	2.7	99.65
−1.8	3.59	0.5	69.15	2.8	99.74
−1.7	4.46	0.6	72.58	2.9	99.81
−1.6	5.48	0.7	75.80	3.0	99.87
−1.5	6.68	0.8	78.81	3.1	99.90
−1.4	8.08	0.9	81.59	3.2	99.93
−1.3	9.68	1.0	84.13	3.3	99.95
−1.2	11.51	1.1	86.43	3.4	99.97