

All Assessment Questions

TeachStat Research Group

Version: November 1, 2019

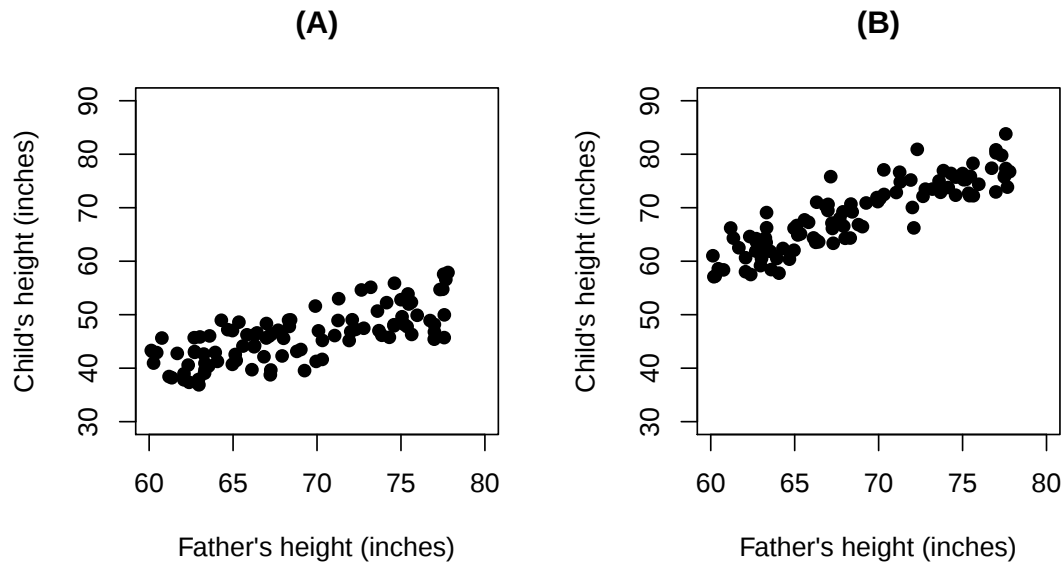
These are the **current** versions of the questions used in the assessment discussed in “Think-aloud interviews: A tool for exploring student statistical reasoning.” Note that the paper discusses some revisions, and these are the versions *after* revisions were made.

Some questions were adapted from other sources, such as textbooks or other concept inventories. These sources are indicated at the bottom of each page.

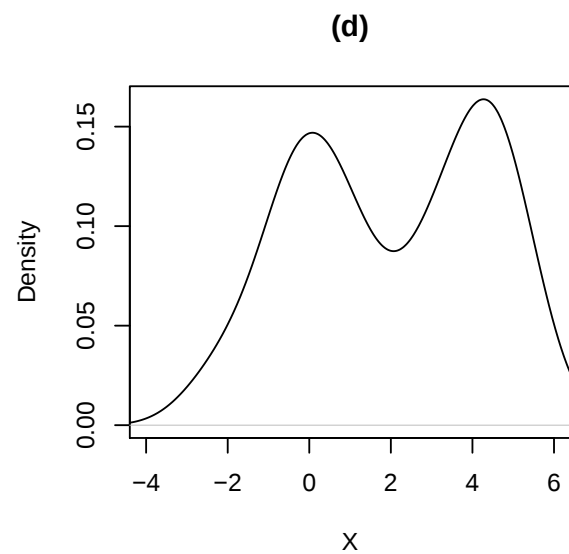
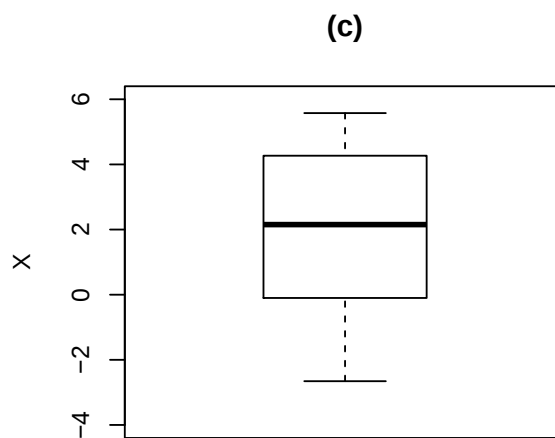
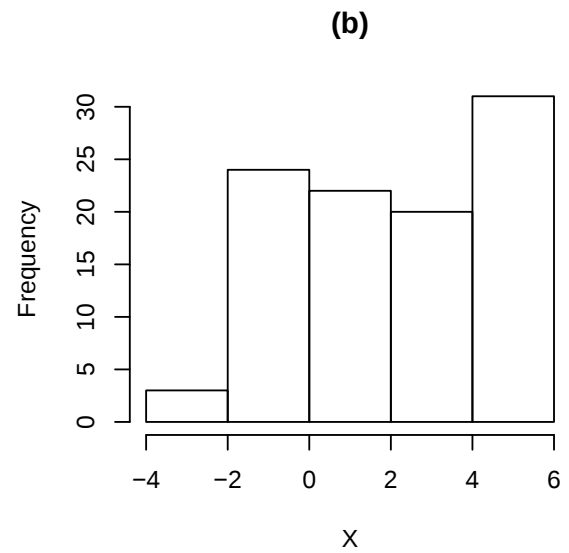
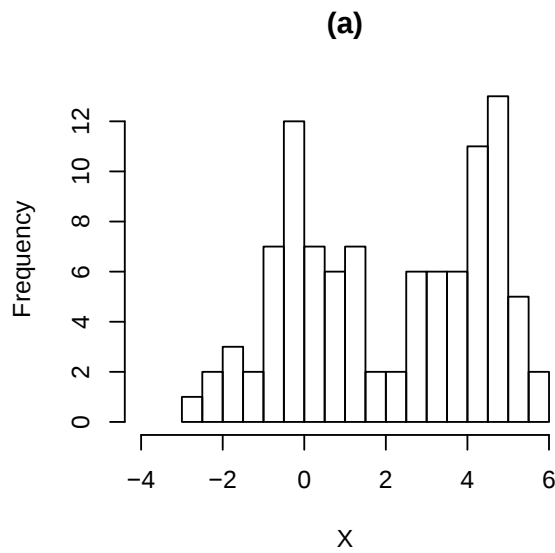
To protect the integrity of the questions, **correct answers are not indicated**. Answers are available on request from the authors.

1. Say we wanted to examine three groups—high school students, college students, and college graduates—and compare the relationship between sleep duration and caffeine consumption across the groups. Which is the most informative visualization?
 - A. For each group, make two histograms of the two variables.
 - B. For each group, make side-by-side boxplots of the two variables.
 - C. For each group, make a scatterplot of the two variables.
 - D. Make a mosaic plot of the three groups.

2. A study measured the heights of children at age 10 and again at age 25. Each height was plotted against the heights of the child's father. Which plot below shows the children's heights at age 10, and which shows their heights at age 25?



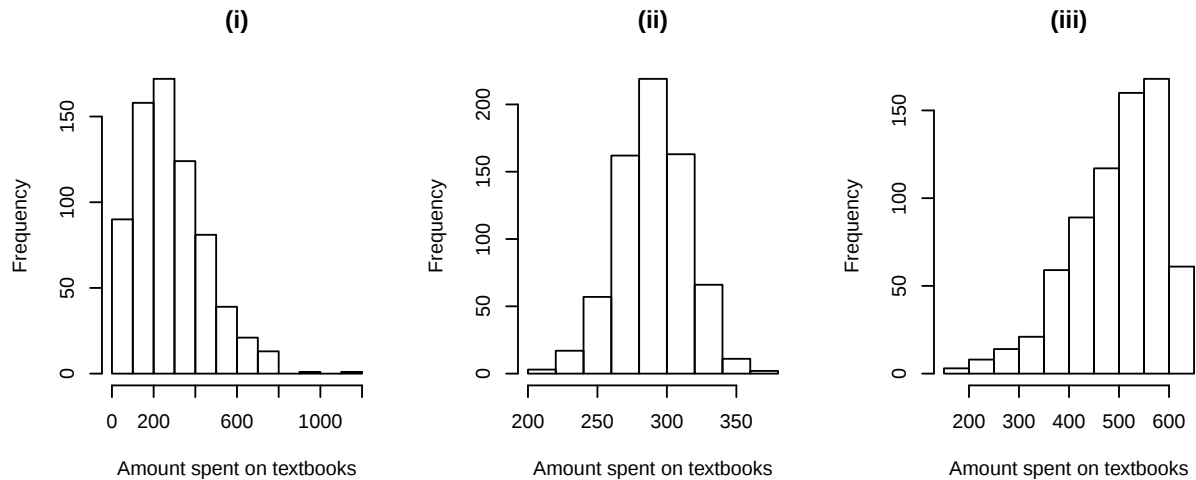
3. In which of the following graphs does the variable X have the greatest variation?



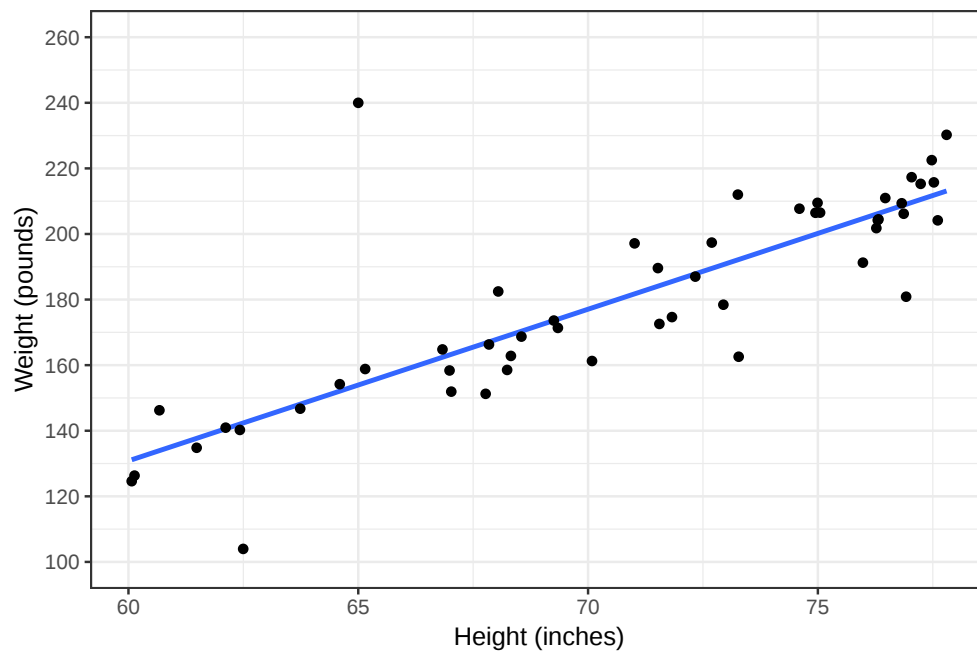
(e) There's not enough information to tell

-
4. (a) What graphic would be most appropriate to visualize the number of students in each major at Carnegie Mellon University?
- A. Scatterplot
 - B. Mosaic plot
 - C. Bar chart
 - D. Box plots
- (b) What graphic would be most appropriate to visualize the weights of burgers served at three different restaurants?
- A. Histogram
 - B. Scatterplot
 - C. Mosaic plot
 - D. Box plots
- (c) What graphic would be most appropriate to visualize the final exam grades for a large introductory course?
- A. Box plots
 - B. Mosaic plot
 - C. Scatterplot
 - D. Histogram
- (d) What graphic would be most appropriate to visualize the SAT scores and GPAs of freshman students?
- A. Scatterplot
 - B. Histogram
 - C. Box plots
 - D. Mosaic plot

5. For a sample of 700 college students, the average cost of textbooks for the fall semester was about \$290, with a standard deviation of about \$170. Would the histogram for the data look like (i), (ii), or (iii)?



6. The following scatter plot shows the heights and weights of a sample of 50 men, and the line of best fit.

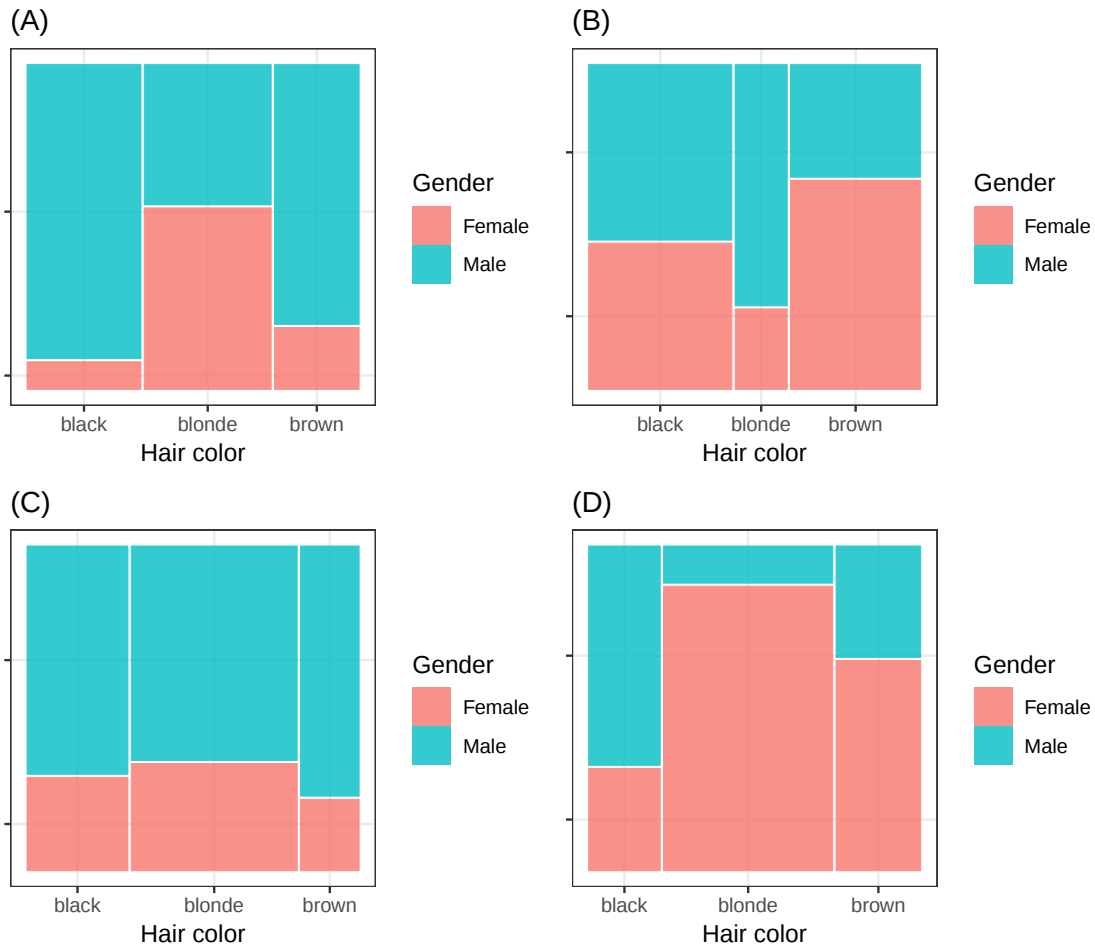


The average height is 70.7 inches; the average weight is 180 pounds.

What is approximately the average weight for men **in the population** that are 5'5" (65 inches) tall?

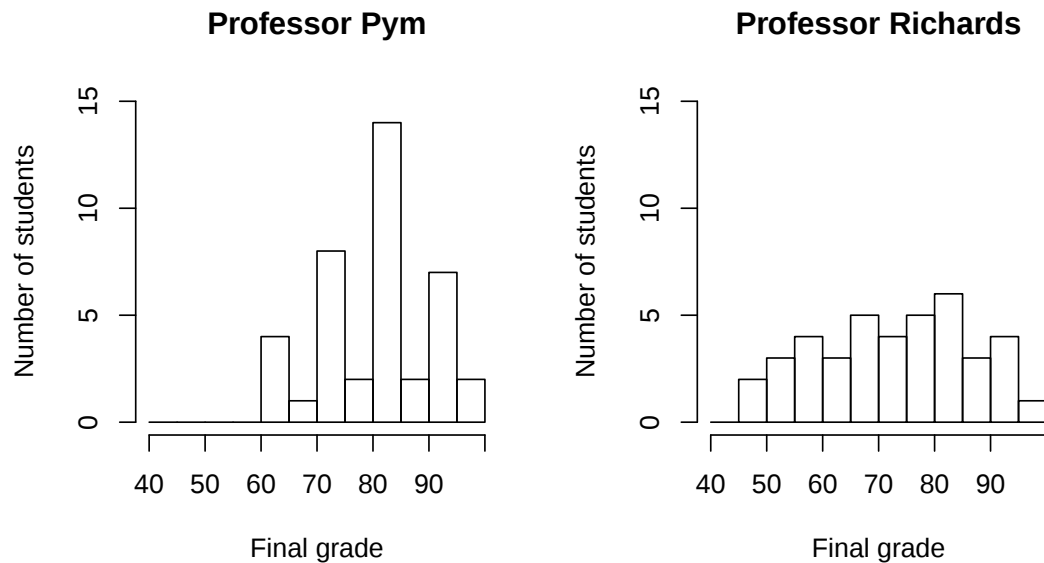
- A. Approximately 240 pounds
- B. Approximately 200 pounds
- C. Approximately 180 pounds
- D. Approximately 155 pounds
- E. We cannot approximate this because it requires extrapolation

7. Use the mosaic plots below to answer the following questions.



- (a) Which of the mosaic plots provides evidence that the variables 'Hair Color' and 'Gender' are independent? A. B. C. D.
- (b) Which of the mosaic plots provides evidence of the most uniform hair color distribution, ignoring gender? A. B. C. D.

8. An introductory statistics course was taught in two separate sections by two different instructors, professors Pym and Richards. Both sections had 40 students. At the end of the semester, the professors compared their students' final grades. A histogram of each class's grades is given below. Each histogram is on the same scale.



Which course has a greater variation in student final grades?

- A. Professor Pym's
- B. Professor Richards's
- C. The variation is equal.
- D. I can't tell from these graphs.

Question adapted from: Cooper and Shore (2008), "Students' Misconceptions in Interpreting Center and Variability of Data Represented via Histograms and Stem-and-leaf Plots", Journal of Statistics Education vol 16 no 2.

9. All three of the following lists of numbers have the same mean of 50. Without doing any calculations, say which one has the smallest standard deviation and which has the largest standard deviation.
- A. 25, 25, 25, 50, 75, 75, 75.
 - B. 25, 30, 40, 50, 60, 70, 75.
 - C. 25, 30, 40, 50, 50, 50, 50, 60, 70, 75.

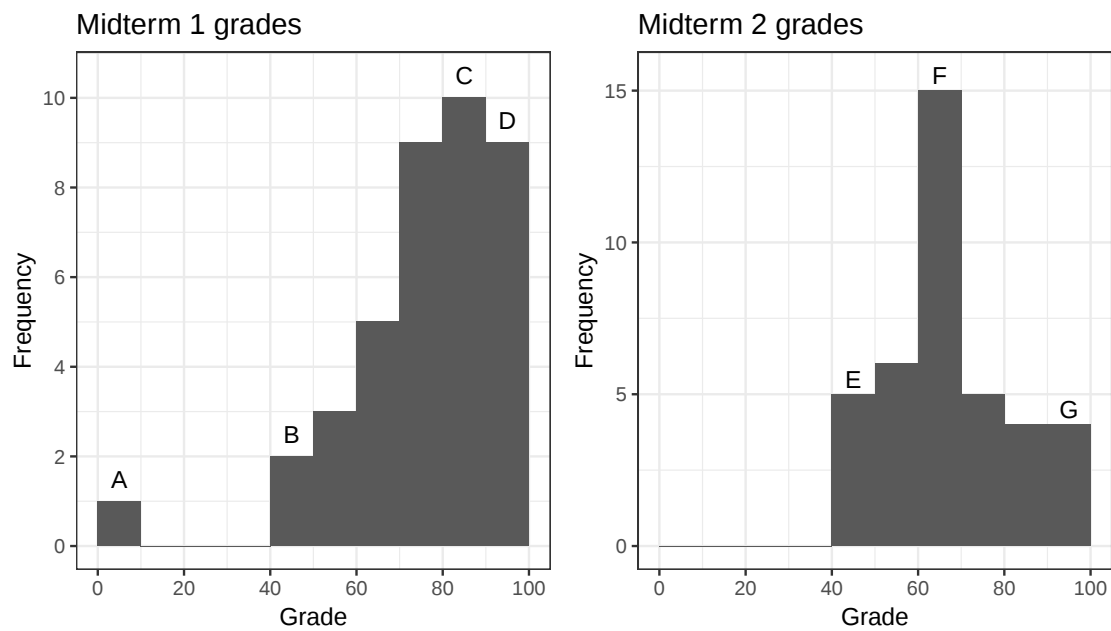
10. A teaching assistant gives a quiz to her students. There are ten questions on the quiz and no partial credit is given. She records the number right and the number wrong, for each student.

The average number of correct answers is 6.4 with a standard deviation of 2.0; the average number of wrong answers is 3.6 with the same standard deviation of 2.0.

The correlation coefficient between the number of correct answers and the number of wrong answers is

- A. 0
- B. -0.50
- C. $+0.50$
- D. -1
- E. $+1$
- F. can't tell without the data.

11. A statistics class had two different midterm exams. After each exam, the professor made a histogram of the midterm exam grades.



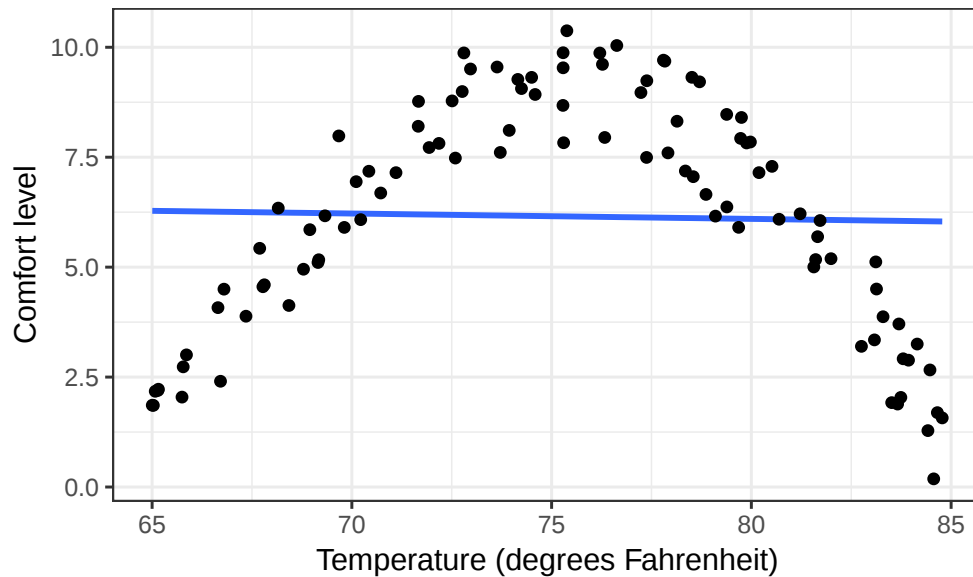
Circle any of the bars that appear to be outliers.

A. B. C. D. E. F. G.

12. A researcher conducted a survey of students to examine the relationship between two variables: sleep duration (in minutes) and quantity of caffeine consumed (in milligrams). Which type of plot would be the most informative?
- A. Two separate histograms of the two variables.
 - B. A scatterplot of the two variables.
 - C. Two side-by-side boxplots of the two variables.
 - D. Overlaid histograms of the two variables in a single plot.

13. A random selection of students on campus were asked to rate the comfort of their classroom on a scale of 0 to 10, where 0 is least comfortable and 10 is the most comfortable. The temperature of the classroom was also recorded.

Below is a scatterplot of the data, along with the line of best fit.



From the plotted data and the line of best fit, what can you conclude about the relationship between temperature and comfort?

- A. As temperature changes, comfort does not change; the average comfort is similar for any temperature.
- B. As temperature increases, comfort increases.
- C. As temperature increases, comfort decreases.
- D. Comfort is best at about 74 degrees, and decreases the farther the temperature is from 74.

14. A candy manufacturer produces 50% brown candies, and each bag they sell is a random assortment of candies. Sam plans to buy a large family-sized bag, and Kerry plans to buy a small fun-size bag. Which bag is more likely to have more than 70% brown candies?
- A. The large family-sized bag
 - B. The small fun-sized bag
 - C. Both bags are equally likely to have more than 70% brown candies

15. A number of draws will be made at random from the box

-3	-1	0	1	3
----	----	---	---	---

After each draw, you write down the number drawn and then put it back in the box.

You win \$100 if the average number of the draws is between -0.1 and 0.1 . Which is better, 100 draws or 1,000 draws? Or does it not matter?

- A. 100 draws
- B. 1,000 draws
- C. The number of draws doesn't matter

16. Farmer Brown has an iPod with 50 songs on it: 25 country songs and 25 jazz songs.

Farmer Brown listens to her music while relaxing after work. Her iPod plays the songs in random order.

On Monday, Farmer Brown listened to 4 songs, but on Tuesday she listened to 10 songs. On which day was she more likely to listen to *only* country music?

- A. Monday
- B. Tuesday
- C. Both are equally likely

17. You have two options:

- (i) You toss a fair coin 100 times: on each toss, if it lands heads you win \$1, if it lands tails you lose \$1.
- (ii) You draw 100 times at random from the box

-1	1	1
----	---	---

On each draw, you are paid (in dollars) the number on the ticket, so

-1

 costs you a dollar. You put the number back in the box after each draw.

Which option is better?

- A. Option (i)
- B. Option (ii)
- C. They are the same

18. Suppose we record whether college students took AP Computer Science in high school and whether or not they major in computer science.

200 students are surveyed. The results are:

	Majors in CS	Doesn't major in CS	
Took AP CS	10	40	50
Didn't take AP CS	10	140	150
	20	180	200

To determine whether students who take AP Computer Science are more likely to major in computer science than those who did not take AP Computer Science, which pair of values should we compare? Choose the best answer.

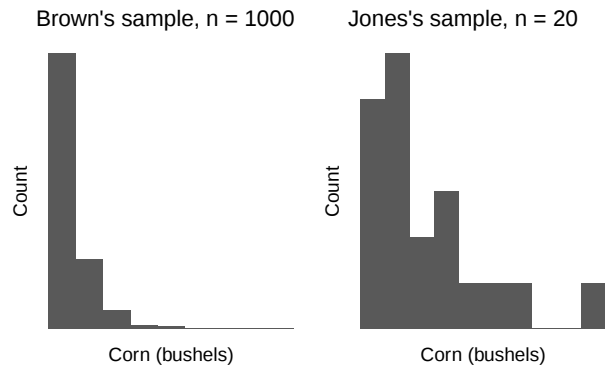
- A. 10 versus 10
- B. $20/200$ versus $180/200$
- C. $10/20$ versus $40/180$
- D. $10/50$ versus $10/150$
- E. $20/200$ versus $10/20$

19. Your friend flips two coins and looks at how they land, without showing you. One coin is gold and one is silver. Which of the following events has a higher probability?
- A. Both coins are heads, given that at least one is heads.
 - B. Both coins are heads, given that the gold coin is heads.
 - C. The events are equally likely.

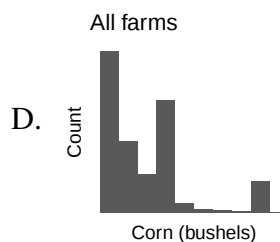
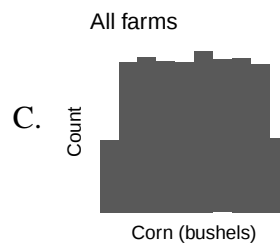
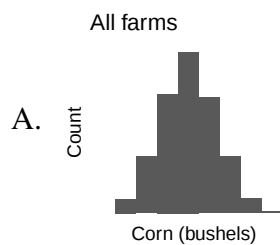
Question adapted from: Bertrand's Box paradox, adapted from Wikipedia

20. A coin will be tossed some number of times, and you win \$97 if the number of heads is exactly equal to half the number of tosses. Which of the following is best for you?
- A. 97 tosses
 - B. 100 tosses
 - C. 1,000 tosses
 - D. The number of tosses doesn't matter

21. Farmer Brown surveyed $n = 1000$ randomly selected farms and recorded their annual corn production (in bushels). Farmer Jones was also interested in corn production, and surveyed $n = 20$ other randomly selected farms and recorded their annual corn production. These are histograms of their results:



Based on these samples, which of the following would be the most likely result if the Department of Agriculture plotted the corn production of *every* corn farm in the country?



22. Suppose we flip a coin twice and record the outcome (Heads or Tails), **in order**.

Recall that the sample space is the set of all possible elementary outcomes.

(a) Which of the following is the sample space?

- A. $\{H\}$
- B. $\{H, T\}$
- C. $\{HH, TT, HT\}$
- D. $\{HH, TT, HT, TH\}$

(b) Does your answer change if the coin is unfair? Why or why not?

23. Farmer Brown is planning a trip to Las Vegas in December, but she doesn't know how much money to budget. She wants to know about the probability of the price of flights increasing over the next few months and the probability that she will win money gambling on the trip.

Suppose the probability of an increase in flight prices is 0.7 and the probability of winning money is 0.2. What is the probability of both events happening? Assume the events are independent.

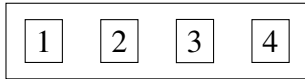
- A. $0.2 + 0.7$
- B. 0.2×0.7
- C. $1 - (0.2 \times 0.7)$
- D. $1 - (0.2 + 0.7)$
- E. $0.2^{0.7}$

24. Gambler Gary is rolling two dice, hoping for snake-eyes: a roll where both dice show a 1. On each roll, the probability of this result is $1/36$. Gary keeps rolling the dice until he's successful, and on average, it would take 36 rolls to win.

After 8 unsuccessful rolls, Gary's friend Fiona joins him to watch him roll the dice. How many rolls should Fiona expect to watch before Gary is successful?

- A. 36
- B. 28
- C. 44
- D. She totally jinxed him and he'll never roll snake-eyes

25. Two draws are made at random from the box containing



After taking out the first draw, a duck eats it, and nobody knows what was written on it. You draw a second time.

In this case, are the two draws independent? Explain.

- A. The draws are independent
- B. The draws are dependent
- C. Not enough information to tell

26. You have 25 ducks. You want to know which duck is the lightest and which is the heaviest, so you begin to weigh the ducks in a random order. However, after you have weighed five ducks, all the ducks fly away.

Which of the following events is most likely?

- A. The heaviest duck you weighed is also the heaviest of the 25
- B. You weighed the lightest duck of the 25
- C. The heaviest duck you weighed is the heaviest of the 25 **and** you weighed the lightest of the 25
- D. Either the heaviest duck you weighed is the heaviest of the 25 **or** you weighed the lightest of the 25

27. I'm about to flip a fair coin 6 times. After each flip, I'll record whether it landed heads (H) or tails (T).

Consider the following outcomes:

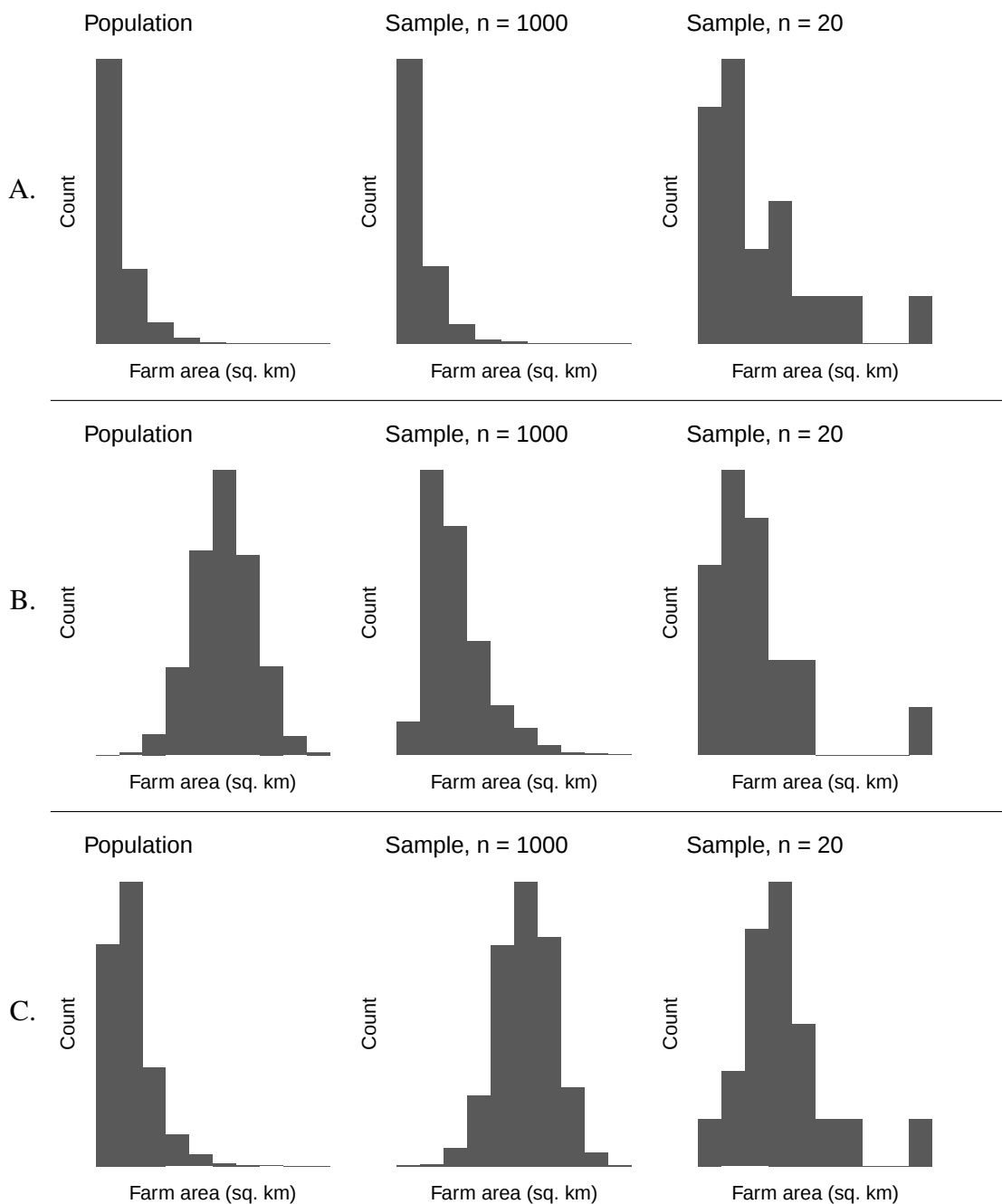
- (i) H T T H T H, in that order
- (ii) H H H H H H, in that order
- (iii) Three Hs and three Ts, in *any* order
- (iv) Six Ts, in *any* order

Which pair of statements is correct?

- A. (i) is more likely than (ii), **and**
(iii) is more likely than (iv).
- B. (i) and (ii) are equally likely, **and**
(iii) is more likely than (iv).
- C. (i) is more likely than (ii), **and**
(iii) and (iv) are equally likely,
- D. (i) and (ii) are equally likely, **and**
(iii) and (iv) are equally likely.

28. Farmer Brown collects data on the land area of farms in the US (in square kilometers). By surveying her farming friends, she collects the area of every farm in the US, and she makes a histogram of the population distribution of US farm areas. She then takes two random samples from the population, of sizes $n = 1000$ and $n = 20$, and plots histograms of the values in each sample.

One of the rows below shows her three histograms. Using the **shape** of the histograms, choose the correct row.



29. A bag has 9 pieces of fruit: 3 apples, 3 pears, and 3 oranges. Four times in a row, you pick out a piece of fruit, write down what it is, and put it back in the bag.

If the first 3 pieces of fruit were apples, what is the fourth piece **most** likely to be?

- A. A pear
- B. An apple
- C. An orange
- D. An orange or a pear are both equally likely and more likely than an apple.
- E. An apple, orange, or pear are all equally likely.

Question adapted from: Hirsch and O'Donnell (2001), "Representativeness in Statistical Reasoning: Identifying and Assessing Misconceptions", Journal of Statistics Education vol 9 no 2.

30. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days.

The forecast of 70% chance of rain can be considered most accurate if it rained on:

- A. 95% – 100% of those days.
- B. 85% – 94% of those days.
- C. 75% – 84% of those days.
- D. 65% – 74% of those days.
- E. 55% – 64% of those days.

Question adapted from: Konold (1995), "Issues in Assessing Conceptual Understanding in Probability and Statistics", Journal of Statistics Education vol 3 no 1 (adapted from Falk 1993, problem 5.1.1, p. 111).

31. Some potentially lucrative, but very uncertain, investments can be made independently. Each has the probability of 0.1 of being a success. As an investment program, a firm invests in 8 of these. Find the probability that the firm gets **at least** one success.
- A. 8×0.1
 - B. 0.1^8
 - C. 0.9^8
 - D. $1 - 0.9^8$
 - E. $1 - 0.1^8$

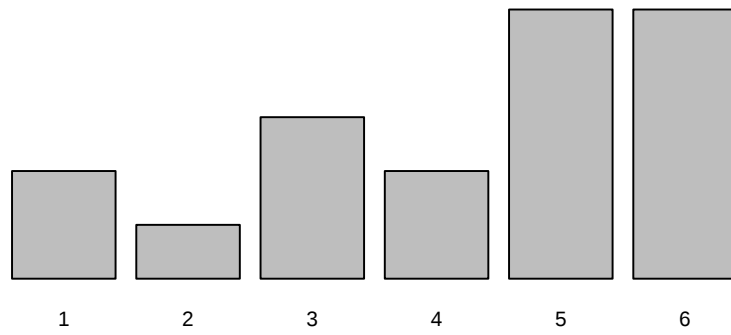
Question adapted from: Beginning Statistics with Data Analysis, by Mosteller, Fienberg, and Rourke, page 231, question 5

32. You roll a **fair** die six times. The rolls turn up 6, 5, 6, 6, 5, and 5.

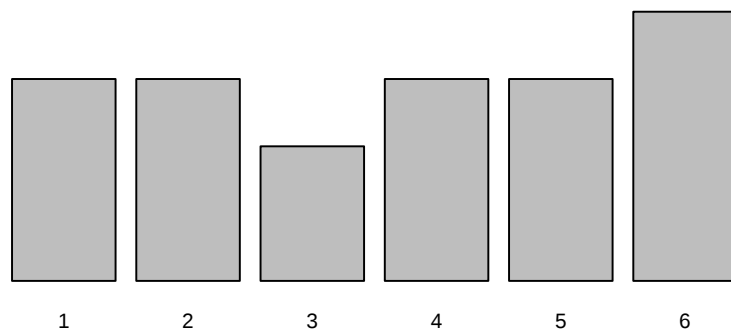
You roll the die twelve more times, then count up the results of all eighteen rolls.

Which of the plots below is the most likely plot of **the eighteen rolls**?

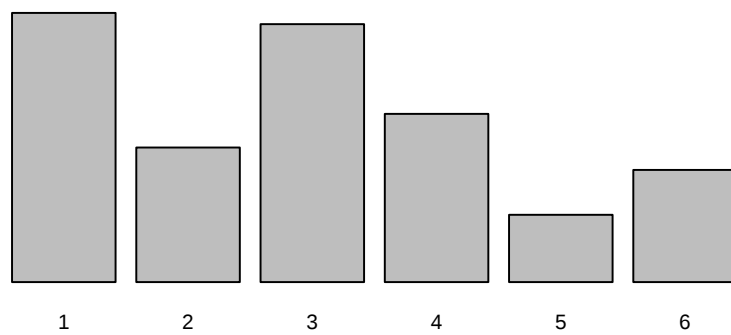
(A)



(B)



(C)



33. As a reward for getting the highest exam grade in your Statistics class, your professor lets you have a piece of fruit chosen from one of two bags: Bag 1 has three apples and one orange, and Bag 2 has six oranges and two apples.

However, instead of telling you which bag is which, the professor picks a bag at random, then picks one random fruit from that bag. What's the probability that the randomly chosen fruit is an apple?

- A. $5/12$
- B. $3/4$
- C. $1/2$
- D. $1/4$

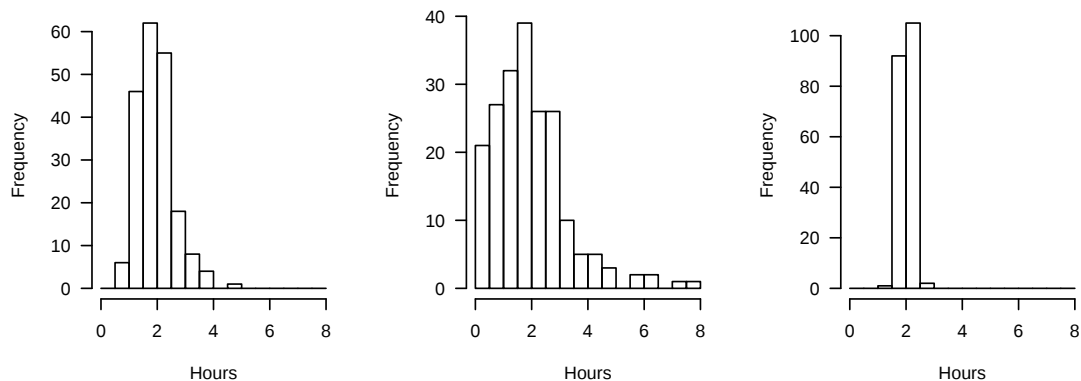
34. Jeri, Steve, and Cosma are conducting surveys of how many hours students study per day at a large public university.

Jeri talks to two hundred students, **one at a time**, and adds each student's answer to her histogram.

Steve talks to two hundred **groups of 5 students**. After asking each group of 5 students how much they study, Steve takes the **group's average** and adds it to his histogram.

Cosma talks to two hundred **groups of 50 students**. After asking each group of 50 students how much they study, Cosma takes the **group's average** and adds it to his histogram.

The three final histograms are shown below, in scrambled order.



Which of the following corresponds to the ordering of the histograms above, from left to right?

- A. Steve, Cosma, Jeri
- B. Steve, Jeri, Cosma
- C. Jeri, Cosma, Steve
- D. Cosma, Steve, Jeri

35. A box contains three black tickets numbered 1, 2, 3 and three white tickets numbered 2, 2, 3.



One ticket is drawn at random: you have to guess the number on the ticket.

You catch a glimpse of the ticket as it is drawn out of the box. You cannot make out the number, but you do see that the ticket is black.

What is the chance that the number on it will be 2?

- A. $1/6$
- B. $1/3$
- C. $1/2$
- D. $2/3$

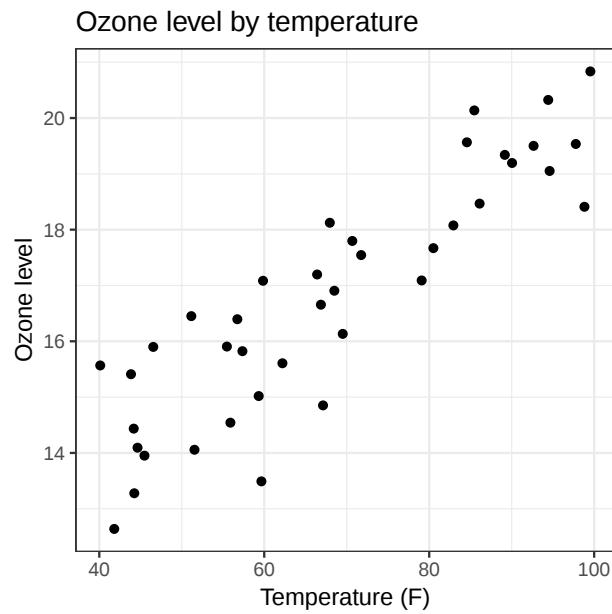
36. You are rolling a fair six-sided die. On its sides are the numbers -3 , -2 , -1 , 1 , 2 , and 3 , rather than the usual numbers 1 through 6 .

You will win \$100 if the **sum** of the rolls is zero.

Would you rather play the game by rolling the die exactly 10 times, or 100 times?

- A. 10 times
- B. 100 times
- C. 10 or 100 times would be equally good
- D. Neither—it would be impossible to win with 10 or 100 rolls

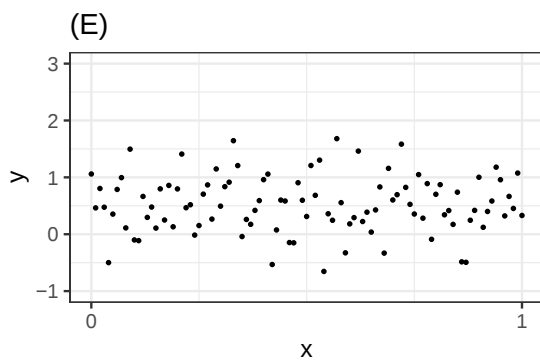
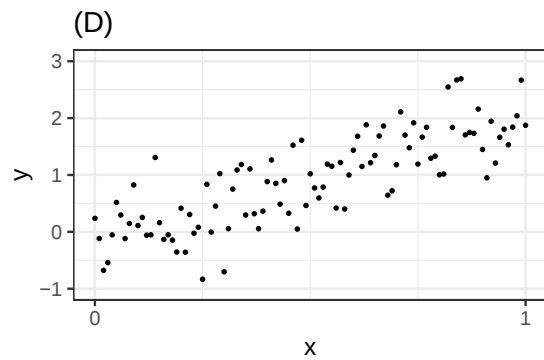
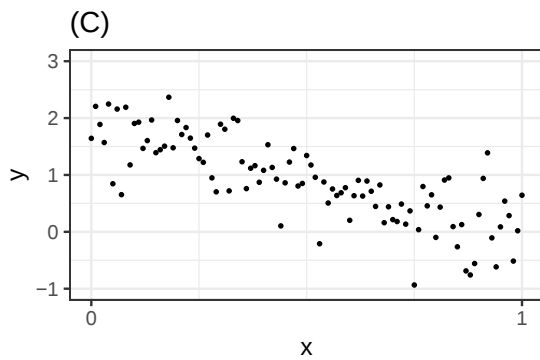
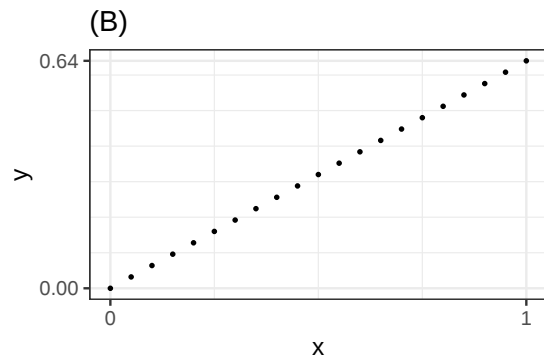
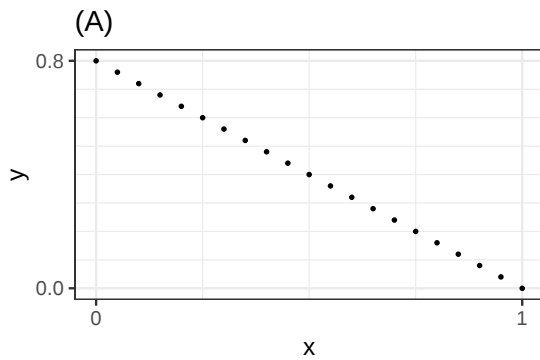
37. Below is a plot made by researchers who are interested in the relationship between the ozone level and the average temperature (in degrees Fahrenheit) of the atmosphere over the Arctic circle.



The formal inference procedure for which this plot would be most appropriate is:

- A. ANOVA
- B. Chi-squared test for dependence
- C. Hypothesis test for correlation coefficient
- D. Kolmogorov–Smirnov test for normality

38. To study the relationship between two variables, you make a scatterplot and calculate the correlation coefficient. You see a linear relationship, and a correlation coefficient of -0.8 . Which of the plots below did you observe?



39. Farmer Brown is about to conduct a study and will use the p -value of a hypothesis test to reach her conclusion.

- If $p < 0.01$, she will reject the null hypothesis at the 1% significance level.
- If $p < 0.05$, she will reject the null hypothesis at the 5% significance level.
- If $p < 0.10$, she will reject the null hypothesis at the 10% significance level.

What is the probability that Farmer Brown will falsely reject a true null hypothesis?

- A. 1%
- B. 5%
- C. 10%
- D. 1%, 5% or 10% depending on the p -value.

40. A survey of Californians found a statistically significant positive correlation between number of books read and nearsightedness.

Which of the following can we conclude about Californians?

- A. Reading books causes an increased risk of being nearsighted.
- B. Being nearsighted causes people to read more books.
- C. We cannot determine which factor causes the other, because correlation does not imply causation.
- D. We cannot draw any conclusions because Californians aren't a random sample of people.

41. A student conducted a survey of random samples of 100 CMU students and 100 University of Pittsburgh students. The survey asked two questions: how many hours you sleep per night, and how many hours you study each day.

The student found:

School	Mean sleep hours	Std. dev.	Mean study hours	Std. dev.
CMU	6.5	1	8.5	2
Pitt	7.5	1	7.5	2

Unfortunately, the student's computer crashed and lost the data. The student only remembers that for one variable the difference between schools was statistically significant, and for the other it was *not* statistically significant.

The student does not remember which was which.

Which variable was more likely the one with the statistically significant difference?

- A. Hours of sleep per night, because the standard deviation was smaller at each school.
- B. Hours of sleep per night, because students spend less time sleeping than studying on average.
- C. Hours spent studying, because the standard deviation was larger at each school.
- D. Hours spent studying, because students spend more time studying than sleeping on average.
- E. We can't tell without the data.

42. A bored student flips a quarter 250 times and gets 140 heads—that's 56%. That makes the 90% confidence interval for the probability of getting heads [51%, 61%].

What does this mean?

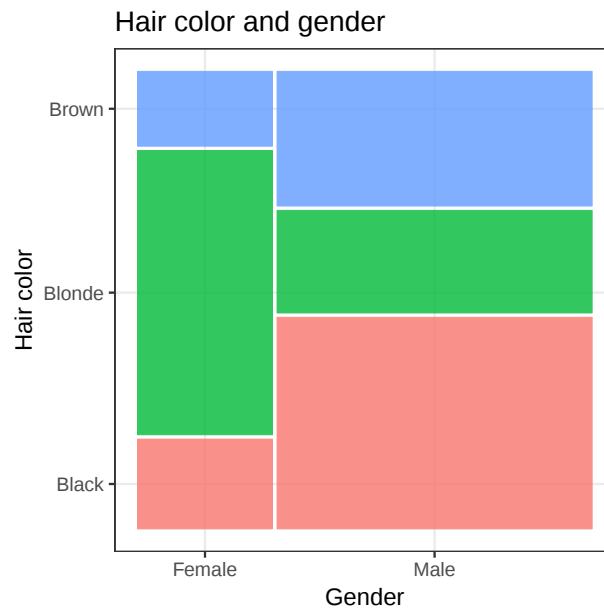
- A. We are 90% sure that in this experiment the quarter landed heads on between 51% and 61% of the flips.
- B. We are 90% sure that any flipped quarter will land on heads between 51% and 61% of the time.
- C. If you flip a quarter many times, you can be 90% sure of getting between 51% and 61% heads.
- D. 90% of all flipped quarters will land heads between 51% and 61% of the time.
- E. We are 100% sure that the true percentage of times this quarter landed on heads in this experiment was some percentage between 51% and 61%.

43. A student participates in a blind Coke versus Pepsi taste test. She correctly identifies which soda is which four times out of six tries. She claims that this proves that she can reliably tell the difference between the two soft drinks.

You want to determine the probability of anyone getting at least four right out of six tries just by random guessing. Which of the following would provide an accurate estimate of that probability?

- A. Have the student repeat this experiment many times and calculate the percent of the time she correctly distinguishes the brands.
- B. Simulate this on a computer, with a 50% chance of guessing the correct soft drink on each try, and calculate the percentage of times there are four or more correct guesses in six trials.
- C. Repeat this experiment with a large sample of people and calculate the percentage of people who make four correct guesses out of six tries.
- D. All of the above methods would provide an accurate estimate of the probability.

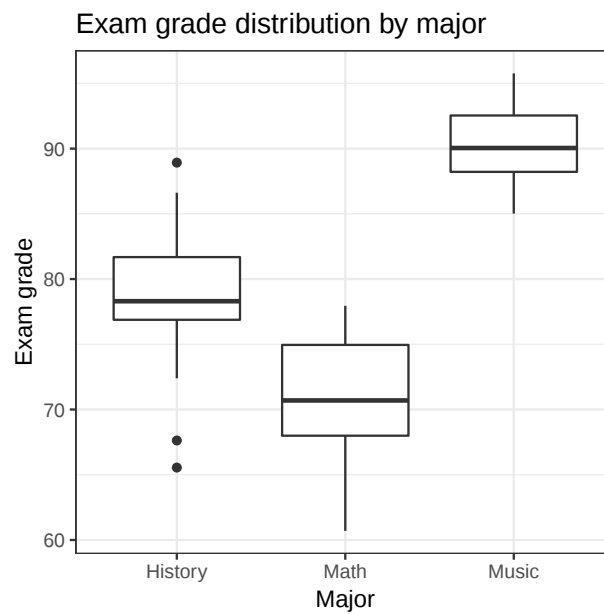
44. Below is a plot made by researchers who are interested in the relationship between hair color and gender.



The formal inference procedure for which this plot would be most appropriate is:

- A. ANOVA
- B. Chi-squared test for dependence
- C. Hypothesis test for correlation coefficient
- D. Kolmogorov-Smirnoff test for normality

45. Below is a plot made by researchers who are interested in the relationship between students' exam grades and their majors.



The formal inference procedure for which this plot would be most appropriate is:

- A. ANOVA
- B. Chi-squared test for dependence
- C. Hypothesis test for correlation coefficient
- D. Kolmogorov-Smirnoff test for normality

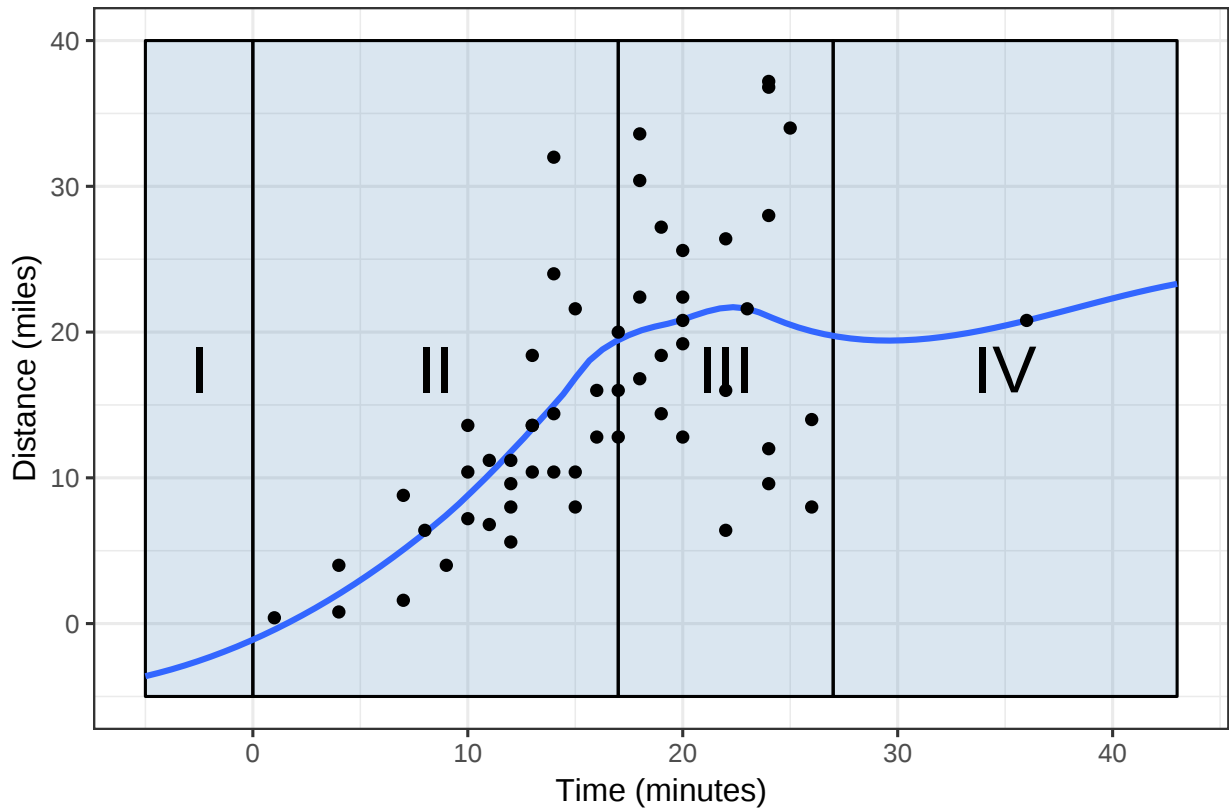
46. A professor experiments with a new way of teaching a class. The class is split into two sections of 20 students, with the students randomly assigned to each section, and the professor teaches one section normally and the other section with a new curriculum.

At the end of the semester, the professor compares the final exam grades between the two sections and finds no statistically significant difference.

What can the professor conclude?

- A. The new curriculum must not have any effect on final exam grades at all.
- B. The professor's analysis must be wrong—there should be a statistically significant difference.
- C. The curriculum may cause a difference in final exam grades, but systematic differences between the students in each section might have canceled it out.
- D. The curriculum may cause a difference in final exam grades, but the sample size could be too small for it to be statistically significant.

47. Displayed are times and distances of daily commutes. Overlaid on the scatterplot is a curve of best fit. When predicting the distance for a new observation, for which time ranges (represented by the shaded rectangles) would you be most confident in your prediction?



- A. I
- B. II
- C. III
- D. IV
- E. I and IV
- F. II and III
- G. I, II, III, and IV
- H. None

48. Farmer Brown obtains a random sample of many cows and chickens from many farms, and administers an IQ test to each. In this large sample, the average cow IQ score was 93.1, and the average chicken IQ score was 102.3.

Farmer Brown tests if chickens have a higher average IQ than cows, and gets a statistically significant p -value of 0.027. All test assumptions were met, and the sample size was quite large.

Can she conclude that, if she picks a cow and a chicken at random, the chicken is almost certainly smarter than the cow?

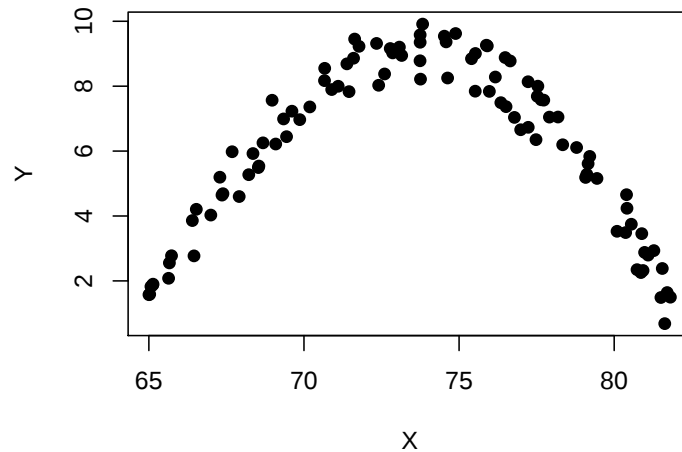
- A. No, because while the difference is statistically significant, it is only about the average of all chickens and cows, not individual chickens and cows.
- B. Yes, because there is only a 2.7% chance of any individual cow being smarter than any individual chicken.
- C. No, because while the difference is statistically significant, the small p -value means the difference between chickens and cows is very small.
- D. Yes, because the difference is statistically significant, which is strong evidence that chickens are smarter than cows.

49. Doctor Brown wants to test two different drugs, Fixitol and Solvix, and see if they reduce blood pressure. Doctor Brown gets 1,000 volunteers with high blood pressure, 500 men and 500 women, to participate in the study. Doctor Brown randomly assigns the men to receive Fixitol and the women to receive Solvix.

Because Doctor Brown used random assignment, if she observes a major difference between blood pressure between the two types of drugs, can she conclude the drugs were the cause?

- A. Yes, because random assignment eliminates the possibility of confounding.
- B. Yes, because the sample size (hundreds of patients) is very large, so the results should be quite accurate.
- C. No, because Doctor Brown used volunteers, not a random sample of all patients with high blood pressure.
- D. No, because if there is any gender difference in the medications' effects, that could account for the difference.

50. You create a scatterplot of two continuous variables X and Y :



Which of the following values is the best correlation r between X and Y to report?

- A. -0.5
- B. 0.5
- C. -0.9
- D. 0.9
- E. Correlation is not an appropriate measure for these two variables
- F. Need more information

51. Professor Smith wants to know if typing her introductory statistics exams in Comic Sans will improve their exam performance. To answer this question, she randomly gives half of the 200 students in her class an exam with all of the questions typed in Comic Sans, while the other students get the same exam with questions typed in Times New Roman.

After comparing the exam scores across both groups of students, Professor Smith finds that the students who were given the exam typed in Comic Sans had a higher average grade on the exam, compared to the average grade for students who did not receive the exam typed in Comic Sans. Professor Smith repeats this experiment across multiple semesters of her course and always sees the same result.

Which of the following is true?

- A. The result is statistical evidence that giving students exams typed in Comic Sans will lead to higher exam scores across the class.
- B. All teachers in every subject should print their exams in Comic Sans to improve their students' performance.
- C. Professor Smith can't draw any conclusions from these tests because other factors, such as the amount of hours students spent studying, might also affect their exam results.
- D. Professor Smith can't draw any conclusions from these tests because she randomly decided which students would receive the exam typed in Comic Sans instead of choosing students systematically, such as giving only the female students the exam typed in Comic Sans.

52. Carnegie Mellon University reports no significant difference between the mean GPAs of a sample of its male and female students ($p = 0.287$). The University of Pittsburgh also reports no significant difference between the mean GPAs of a sample of its male and female students ($p = 0.353$).

What can you conclude about students at each university?

- A. There is stronger evidence that male and female students have different GPAs at Carnegie Mellon University
- B. There is stronger evidence that male and female students have different GPAs at the University of Pittsburgh
- C. Neither university has strong evidence that male and female students have different GPAs
- D. There is not enough information to answer this question

53. Researchers for the city are examining local public swimming pools, and have observed a positive relationship between the number of lifeguards and the number of deaths from drowning: on days with more lifeguards, more swimmers drown. Which of the following is an appropriate conclusion that the researchers can draw?
- A. Increasing the number of lifeguards will result in more drowning deaths.
 - B. Pools should decrease the number of lifeguards to better protect swimmers.
 - C. Pools should increase the number of lifeguards to better protect swimmers.
 - D. We expect more drowning deaths on days with more lifeguards.

54. A clinical trial randomly assigned subjects to either practice mindfulness meditation or a placebo relaxation exercise as a treatment for a cold. The trial found that subjects who practiced mindfulness meditation had a shorter time to recovery than students assigned to the relaxation exercise, and the result was statistically significant.

Which conclusion does this support?

- A. Recovering faster from a cold causes subjects to meditate.
- B. Mindfulness meditation causes subjects to recover faster from a cold.
- C. We cannot draw any conclusions because correlation does not imply causation.
- D. We cannot draw any conclusions because assignment was random instead of systematic.

55. A research study randomly divided participants into two groups. One group received Vitamin E to take daily; the other group received only a placebo pill. The research study followed the participants for eight years to see how many developed cancer during that time.

Which of the following gives the *best* explanation of the purpose of randomization in this study?

- A. To reduce sampling variation in the estimated benefit of Vitamin E.
- B. To ensure that all potential cancer patients had an equal chance of being selected for the study.
- C. To ensure participants formed a representative sample of the population.
- D. To produce treatment groups with similar characteristics.
- E. To prevent skewness in the results.

56. A large clinical trial randomly assigned overweight patients to take either a new diet pill or a placebo. At the end of the trial, patients taking the diet pill weighed much less than those taking the placebo, and the difference was statistically significant.

Later, a survey asked a random sample of *different* people their weight and whether they were taking the diet pill. The survey found that people taking the diet pill weighed **more**, on average, than those not taking the pill, and the difference was statistically significant.

What is the most likely explanation of the contradictory results?

- A. The patients in the clinical trial who took the diet pill must have started the trial weighing less than those taking the placebo.
- B. The survey was not a controlled experiment, so its results do not prove that diet pills *cause* weight gain.
- C. The survey used a random sample of patients, so its results can't be generalized to a broader population.
- D. The clinical trial didn't use a random sample of patients, so its results do not prove that diet pills *cause* weight loss.

57. A scientist conducted a randomized trial to test if a new skin cream cures a rash. The results were:

	Improved	Got worse	Total
Treatment	223	75	298
No treatment	107	21	128
Total	330	96	426

Does the skin cream seem to work?

- A. Yes, because 223 patients improved under the treatment, many more than the 107 who improved without treatment.
- B. No, because 75 patients got worse under the treatment, many more than the 21 who got worse without treatment.
- C. Yes, because more than three times more patients improved (a total of 330) than got worse (a total of 96).
- D. No, because three times more patients improved than got worse under the treatment, but about five times more patients improved than got worse with no treatment.

Question adapted from: Kahan, Peters, Dawson, and Slovic (2017), "Motivated Numeracy and Enlightened Self-Government", Behavioral Public Policy 1 pp 54–86.

58. Two of your friends just emailed you two very different news articles. The first article discusses a recent survey that found that people who drink more coffee make more money annually. The headline says, “Drink More Coffee and Make Millions!” The second article conducted a similar survey and found that people who drink less coffee make more money annually. The headline says, “Drink Less Coffee and Make Millions!”

Which of the following statements is true?

- A. It’s more believable that people who drink more coffee make more money, so the first article’s headline is correct.
- B. It’s more believable that people who drink less coffee make more money, so the second article’s headline is correct.
- C. Because the studies’ conclusions conflict with each other, no conclusions can be drawn about the relationship between drinking coffee and income.
- D. Neither article’s conclusion should be trusted because other factors, such as a person’s age, might influence both the amount of coffee a person drinks and the amount of money they make.