

Statistical Problem Sets in WeBWork

STATISTICAL PROBLEM SETS IN WEBWORK

PETER STAAB AND RACHAEL
NORTON

ROTEL (Remixing Open Textbooks with an
Equity Lens) Project
Fitchburg, Massachusetts



Statistical Problem Sets in WeBWorK Copyright © 2023 by Rachael Norton and Peter Staab is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

CONTENTS

Land Acknowledgement	vii
Preface	1

Part I. Main Body

1.1 Types of Data and Variables	7
1.2 Data Sampling	9
1.3 Levels of Measurement	14
1.4 Experimental Design	16
2.1, 2.2 Graphs and Percentages	18
2.4 Box plots	22
2.5 Center of data	28
3 Probability	34
4.1, 4.2 Random Variables	40
6.1, 6.2 Normal Distribution	42
7 Central Limit Theorem	43
8.1, 8.2 Confidence Intervals	46

8.3-9.3 Confidence Intervals and Hypothesis Testing	49
9.4, 9.5 Hypothesis Testing	53
12.1, 12.2 Linear Equations and Scatter Plots	55
12.3, 12.4 Linear Regression	58
Grant Information	67

LAND ACKNOWLEDGEMENT

Land Acknowledgement Statement for the ROTEL Grant

As part of ROTEL Grant's mission to support the creation, management, and dissemination of culturally-relevant textbooks, we must acknowledge Indigenous Peoples as the traditional stewards of the land, and the enduring relationship that exists between them and their traditional territories. We acknowledge that the boundaries that created Massachusetts were arbitrary and a product of the settlers. We honor the land on which the Higher Education Institutions of the Commonwealth of Massachusetts are sited as the traditional territory of tribal nations. We acknowledge the painful history of genocide and forced removal from their territory, and other atrocities connected with colonization. We honor and respect the many diverse indigenous people connected to this land on which we gather, and our acknowledgement is one action we can take to correct the stories and practices that erase Indigenous People's history and culture.

Identified tribes and/or nations of Massachusetts

Historical nations:

- Mahican
- Mashpee
- Massachuset
- Nauset
- Nipmuc
- Pennacook
- Pocomtuc
- Stockbridge
- Wampanoag

Present day nations and tribes:

- Mashpee Wampanoag Tribe
- Wampanoag Tribe of Gay Head Aquinnah
- Herring Pond Wampanoag Tribe
- Assawompsett-Nemasket Band of Wampanoags
- Pocasset Wampanoag of the Pokanoket Nation
- Pacasset Wampanoag Tribe
- Seaconke Wampanoag Tribe
- Chappaquiddick Tribe of the Wampanoag Indian Nation
- Nipmuc Nation (Bands include the Hassanamisco, Natick)
- Nipmuck Tribal Council of Chaubunagungamaug

- Massachusetts Tribe at Ponkapoag

In the event that we have an incorrect link or are missing an existing band/nation, please let us know so that we may correct our error.

Suggested readings

Massachusetts Center for Native American Awareness

A guide to Indigenous land acknowledgment

‘We are all on Native Land: A conversation about Land Acknowledgements’ YouTube video

Native-Land.ca | Our home on native land (mapping of native lands)

Beyond territorial acknowledgments – âpihtawikosisân

Your Territorial Acknowledgment Is Not Enough

PREFACE

This book consists of problems which were adapted from an introductory-level statistics course. We aimed to improve accessibility and rewrite problems which reinforced gender binary, white supremacy, or ageism or were otherwise exclusionary or culturally irrelevant to students.

Our work was supported by the Remixing Open Textbooks through an Equity Lens (ROTEL) project, which is funded by a U.S. federal grant through The Open Textbook Pilot program. Our goal was to sift through the homework sets we currently use to teach Applied Statistics, an introductory course in statistics at Fitchburg State University, and rewrite any problems which were not accessible or did not promote equity and inclusion. The students at Fitchburg State University are racially diverse; approximately 10% of students are Hispanic or Latino, and 10% are Black or African American. About 45% of students are enrolled full-time, more than 40% of which are Pell Grant recipients and many of whom work at least part-time.

The problems in this project originated from WeBWorK (<https://openwebwork.org>), an open-source online homework platform in which students answer questions in an interactive manner. Problems are randomized for individual students and,

for nearly all problems, students receive immediate feedback about their answers. There are over 55,000 problems in the current library that have been written by mathematicians, statisticians and other scientists across the world.

We began with the homework sets on WeBWorK that we currently use when we teach Applied Statistics. We read every problem, keeping a watchful eye on the following aspects:

- **Accessibility.** WeBWorK is web-based, and extensive work has been done to promote accessibility online. However, individual problems need to be examined to make sure they comply with current accessibility standards.
- **Correct use of language promoting diversity, equity, and inclusivity.** We sought to improve problems that used antiquated notions of gender, propagated stereotypes, reinforced eurocentrism, or were otherwise outdated or exclusionary.

Out of 90 problems, we identified 39 problems that fell short in terms of accessibility or diversity, equity, and inclusivity. We revised these problems, and they have been included in this book. We normally use these problems as a companion to the OpenStax textbook *Introductory Statistics* by Barbara Illowsky and Susan Dean. As such, the numbering and titles of the problem sets align with the sections of the textbook. However, these problems may be used to supplement or

replace existing course materials for any introductory statistics course to make the content more inclusive and culturally relevant to students.

The most common way we improved accessibility was to provide alternate text for embedded graphics so that students with visual impairments can access the problems. There were several ways we changed the language in problems to promote diversity, equity, and inclusion. They ran the gamut from rewriting the entire problem to changing the names and pronouns of the people in the problem.

In fact, changing the names and pronouns of the people in a problem was an important aspect of this project. We updated an existing code, called a macro, that chooses names randomly to draw from a more diverse set of names. We also added pronoun information to each name, assuring that about 5-10% of persons in the macro had nonbinary pronouns. In the end, our code allows different versions of the same problem to appear for different users, and in each version the name, pronouns (subject, object, and possessive forms), and verb forms all agree and sound natural. For example, one version of a problem may read, “Amelia has a ribbon that is 22cm long. If she cuts off 6cm, how much does she have left?” We include this example here because problems for which our only change was names and pronouns are not included in this book. In May 2023, this macro was accepted into the WeBWorK codebase for others to use as well.

When a version of a problem is generated on WeBWorK,

the version is usually generated with random values for the variables and random names and pronouns, as seen at the end of the previous paragraph. Due to the nature of this publication, however, the problems in this document are a single version of those actually used in courses. If you use the problems on the WeBWorK platform, a randomized version will be generated for each student, adding to the diversity of identities represented in the problems. We invite you to contact one of us (Peter Staab, pstaab@fitchburgstate.edu) for help if you decide to use these problems in a course to ensure that you get the most out of them.

About the Authors

Peter Staab teaches at Fitchburg State University in Massachusetts. He has been using/developing OER for 15 years, long before he knew the term. He comes to this diversity and equity project with a desire for his students, who are increasingly coming to campus with diverse backgrounds. In addition, he looks toward the future as his children, both adopted and African-American, continue through middle and high school and their peer groups are quite diverse in economic, gender and racial background.

Rachael Norton taught mathematics at Fitchburg State University for three years. She greatly enjoyed belonging to a rare, majority female, mathematics department at FSU, but she

and her husband ultimately decided to relocate to be closer to family. She now teaches at St. Olaf College in southern Minnesota. She lives in the Twin Cities, a vibrant and diverse community with a high percentage of East Africans, two of whom are her husband and son. On Thursdays, you can find Rachael and her family eating delicious home-made Ethiopian food at her mother-in-law's house.

Acknowledgments

We want to thank a few people who have been instrumental in getting this project over the finish line. Rachel Graddy, who was the Director of Disability Services at Fitchburg State University, reviewed all problems for accessibility. Junior Peña, Director of Student Diversity, Equity, & Belonging Programs at Fitchburg State University, reviewed all of the problems for DEI language. Both Junior and Rachel gave us crucial feedback to improve this project. Alex Jordan helped tremendously with the random name macro, and Jackie Kremer has been an overall leader at Fitchburg State with ROTEL and a cheerleader for all of us. Lastly, we thank the statewide ROTEL staff for supporting us in the finishing touches, polishing the text, and keeping us organized to the end.

1.1 TYPES OF DATA AND VARIABLES

Problem 1

Which of the following show examples of collecting categorical data? CHECK ALL THAT APPLY.

- A. MacroSoft wants to measure the number of computers owned by households in Vancouver, BC, in order to sell more software.
- B. The professor is interested in the nationality of the students taking STAT 200. A WeBWorK question was made to ask the students their nationality.
- C. Counting the number of students in STAT 200 with black hair, brown hair, blonde hair, white hair, red hair, another hair colour, or bald.
- D. A doctor designates newborns as having low birth weight if they weigh less than 2500g (5.5 pounds), and as not having low birth weight otherwise.
- E. None of the Above

Correct Answers

B, C, D

Problem 2

Which of the following selected variables, associated with clinical trials of a drug, are quantitative variables? CHECK ALL THAT APPLY.

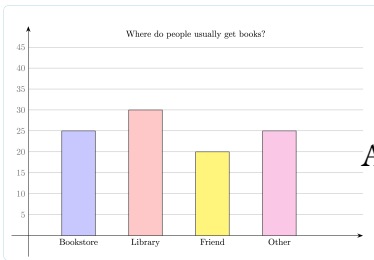
- A. Assigned Sex – Female, Male, Intersex
- B. Smoking history (cigarettes per year)
- C. Age (in years) < 20, 20-30, 30-40, 40-50, 50-60, 60 or above
- D. Body mass index (in kg/m^2)
- E. Dosage form – 1=tablet, 2=capsule, 3=liquid solution, 4=other

Correct Answers

B, D

1.2 DATA SAMPLING

Problem 1



Use the bar graph above to fill in the blanks.

- A. _ people said that they get books in a bookstore or a library.
- B. _ more people said that they get books from the library than said that they get books from a friend.
- C. _ people in the survey did not get their books from the library.

Solution

To tell how many people usually got books from the library, look at the bar above the word “Library”. Its height can be found from seeing where the line at the top of the bar would hit the y-axis (if we continued it leftward) since the heights are marked there. That height is the number of people who got

books from the library and in this case, the answer is 30.

To determine how many more people get books from a library than a friend, subtract the bar height of the friend bar from the bar height of the library bar or $30 - 20 = 10$.

The ways to find out how many people did not get their books from the library is to subtract the number who got their books from the library from 100. In this case, the answer is 70.

Correct Answers

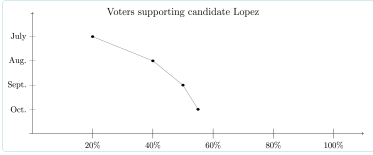
- A. 55
- B. 10
- C. 70

Problem 2

This table shows the results of polls taken during the four months preceding an election between two candidates – Lopez and Ye.

Percentage of Voters Choosing Lopez			
July	Aug.	Sept.	Oct.
20	40	50	55

Ye published this graph in a newspaper.



How could this graph be misleading about popularity?

- A. The graph gives the impression that popularity is changing at a constant rate.
- B. The graph does not show the number of voters polled each month.
- C. The graph does not show the percent of voters who chose Ye.
- D. The graph gives the impression that popularity is decreasing

Solution

While all of the choices state facts that are correct, the one unusual thing that would not be expected by the readers is that a line with negative slope would indicate increasing popularity. A person who did not look carefully would assume that a negative slope would indicate decreasing popularity. That is: the graph gives the impression that Lopez's popularity is decreasing.

Correct Answers

D

Problem 3

A simple random sample of 30 residents from Seattle is taken to estimate the median income of all Seattle residents.

Is this study

- A. REPRESENTATIVE?
- B. NON-REPRESENTATIVE?

A simple random sample of people over age 18 is taken to estimate the mean weight of all adults.

Is this study

- A. REPRESENTATIVE?
- B. NON-REPRESENTATIVE?

A telephone survey is conducted during the day in order to determine the chances of a certain candidate winning an election.

Is this study

- A. REPRESENTATIVE?
- B. NON-REPRESENTATIVE?

Using a sample of 40 patients from a local hospital, researchers

measured cholesterol level in an attempt to estimate the mean cholesterol level of U.S. citizens.

Is this study

- A. REPRESENTATIVE?
- B. NON-REPRESENTATIVE?

Correct Answers

A

A

B

B

1.3 LEVELS OF MEASUREMENT

Problem 1

Determine whether the following possible responses should be classified as interval, nominal or ordinal data.

1. Do you feel that the stay in New York was sufficiently long?
2. Which of the following features of the hotel in New York did you find most attractive: location, facilities, room size, or price?
3. What is the maximum number of hours per day that you would like to spend travelling?
4. What is the current temperature outside?

Correct Answers

1. NOMINAL
2. NOMINAL
3. INTERVAL

4. INTERVAL

Problem 2

Determine whether the following possible responses should be classified as ratio, interval, nominal or ordinal data.

1. Your hometown
2. The starting salaries of new Ph.D. graduates from a statistics program
3. The type of car a person drives
4. The size of fries (small, medium, large) ordered by a sample of Burger King customers

Correct Answers

1. NOMINAL
2. RATIO
3. NOMINAL
4. ORDINAL

1.4 EXPERIMENTAL DESIGN

Problem 1

A study of human development showed two types of movies to a group of children. Chips were available in a bowl, and the investigators compared the number of chips eaten by the children while watching the different kinds of movies. One kind was shown at 8 A.M. and another at 11 A.M. It was found that during the movie shown at 11 A.M., more chips were eaten than during the movie shown at 8 A.M. The investigators concluded that the different types of movies had an effect on appetite.

(a) The explanatory variable in this experiment is

- A. type of chip.
- B. type of movie.
- C. hunger level.
- D. None of the above.

(b) The response variable in this experiment is

- A. the time the movie was shown.
- B. the bowls.
- C. the different kinds of movies.
- D. the number of chips eaten.

(c) A lurking variable in this experiment is

- A. the different kinds of movies.
- B. the number of chips eaten.
- C. the bowls.
- D. the time the movie was shown.

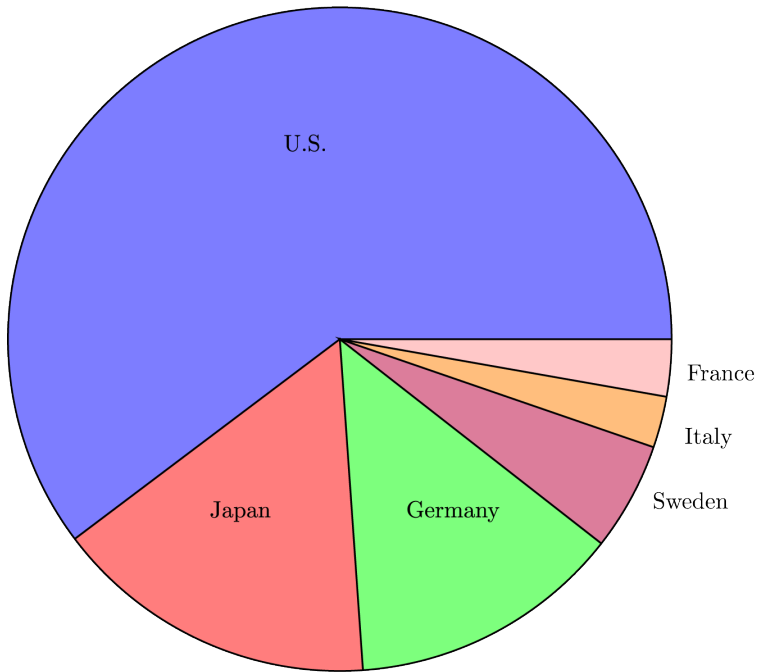
Correct Answers

- (a) B
- (b) D
- (c) D

2.1, 2.2 GRAPHS AND PERCENTAGES

Problem 1

Consumers Union measured the gas mileage in miles per gallon of 38 1978-79 model automobiles on a special test track. The pie chart below provides information about the country of manufacture of the model cars used by Consumers Union.



1. A pie chart is equivalent to a
 - A. timeplot
 - B. histogram.
 - C. bar chart.
 - D. scatter plot.
 - E. None of the above.

2. Based on this pie chart, we may conclude that
 - A. more than half of the cars in the study were from the

United States.

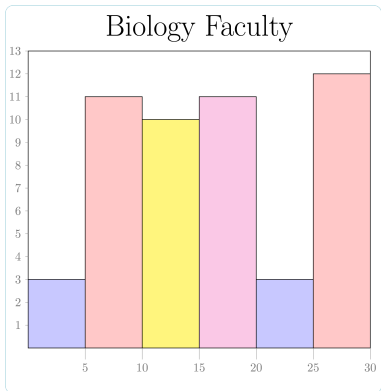
- B. Swedish cars get gas mileages that are between those of Japanese and American cars.
- C. German luxury cars made by Mercedes Benz, Audi, Porsche, and BMW represent approximately one quarter of the cars tested.
- D. Japanese cars get significantly lower gas mileage than cars of other countries. This is because their slice of the pie is at the bottom of the chart.
- E. None of the above.

Correct Answers

- 1. C
- 2. A

Problem 2

The histogram below gives the length of service of members of the Department of Biology at a particular university. The horizontal axis represents years of service in intervals 0 to 4.9, 5 to 9.9, etc., and the vertical axis represents the number of faculty.



What percent of the department faculty have less than 10 years of service?

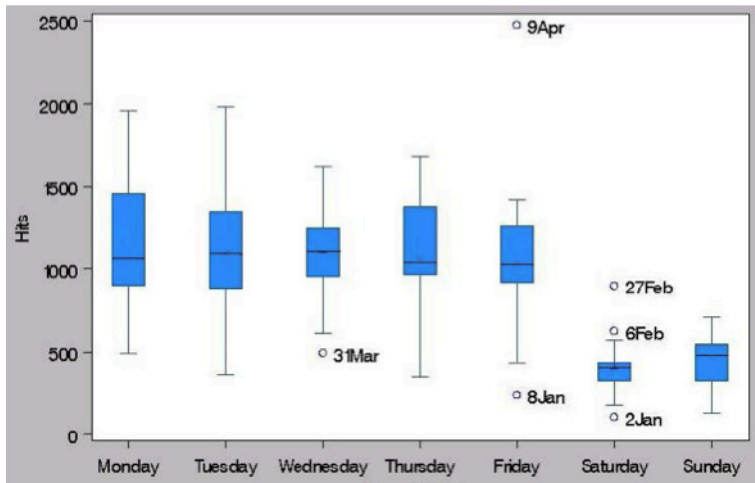
Correct Answers

28

2.4 BOX PLOTS

Problem 1

The following are boxplots of the numbers of hits at a certain website for the different days of the week.



1. Which days had outliers?

- A. Sunday
- B. Monday
- C. Tuesday
- D. Wednesday

E. Thursday

F. Friday

G. Saturday

2. Which day had the largest median?

3. Which day had the largest 3rd quartile?

4. Less than 25 percent of Thursdays had a higher number of hits than the busiest Saturday.

- True
- False

Correct Answers

1. DFG

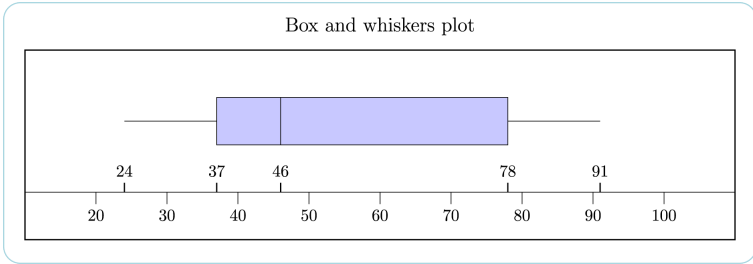
2. Wednesday

3. Monday

4. False

Problem 2

Consider the following box and whisker plot. Find the indicated values of the represented data.



1. Minimum:
2. First quartile:
3. Median:

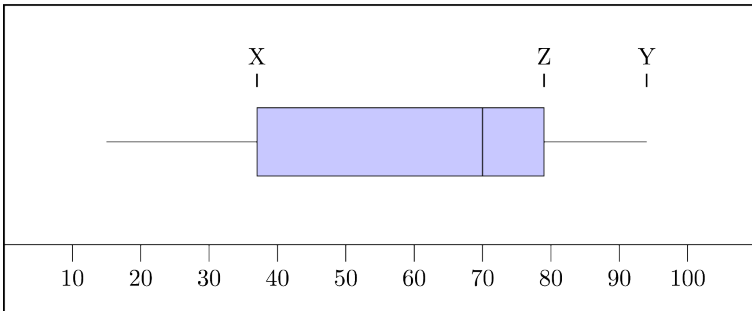
Correct Answers

1. 24
2. 37
3. 46

Problem 3

Consider the following box and whisker plot. Match the letters with the values.

Box and whiskers plot



1. X

2. Y

3. Z

A. Mean

B. Q_1

C. Max

D. Right Pivot

E. Q_3

Correct Answers

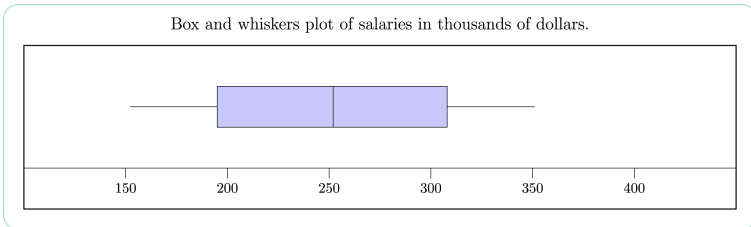
1. B

2. C

3. E

Problem 4

The boxplot below represents annual salaries of attorneys in thousands of dollars in Los Angeles.



About what percentage of the attorneys have salaries between \$152,000 and \$252,000?

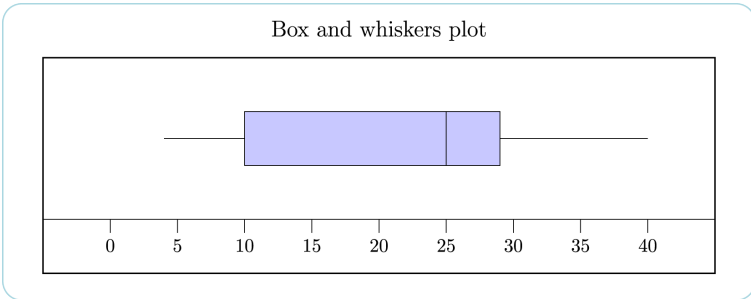
- 10
- 30
- 35
- 50
- None of these

Correct Answers

50

Problem 5

A boxplot for a set of data is given below. Find the five-number summary.



Find the minimum:

Find Q_1 :

Find the median:

Find Q_3 :

Find the maximum:

Correct Answers

Minimum: 4

Q_1 : 10

Median: 25

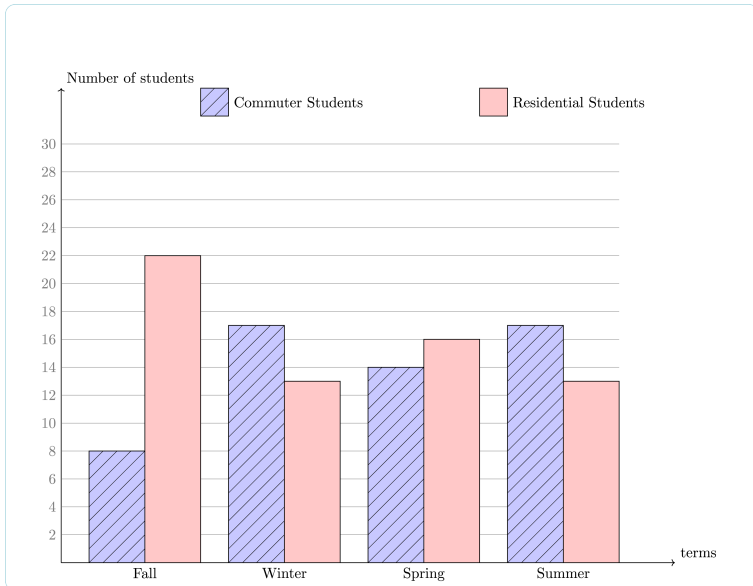
Q_3 : 29

Maximum: 40

2.5 CENTER OF DATA

Problem 1

This bar graph compares commuter and residential student enrollment in a certain class at a college in a certain school year.



Answer the following questions.

Question 1: In these 4 terms, the mean number of commuter students in this class was:

Question 2: In these 4 terms, the mean number of residential students in this class was:

Question 3: In these 4 terms, the median of commuter students in this class was:

Question 4: In these 4 terms, the median of residential students in this class was:

Question 5: If you want to know, in all 4 terms, there are more commuter or residential students, should you compare the mean or median?

Solution

Question 1

To find the mean number of commuter students in this class, we first add up the number of commuter students in all 4 terms:

$$8 + 17 + 14 + 17 = 56$$

Now we can find the mean by dividing the sum by the number of terms (4):

$$\text{mean} = \frac{56}{4} = 14$$

Solution: In these 4 terms, the mean number of commuter students in this class was 14.

Question 2

To find the mean number of residential students in this

class, we first add up the number of residential students in all 4 terms:

$$22 + 13 + 16 + 13 = 64$$

Now we can find the mean by dividing the sum by the number of terms (4):

$$\text{mean} = \frac{64}{4} = 16$$

Solution: In these 4 terms, the mean number of residential students in this class was 16.

Question 3

To find the median number of commuter students, we first need to order the number of commuter students in all 4 terms:

$$8, 14, 17, 17$$

Solution: In these 4 terms, the median of commuter students in this class was 15.5.

Question 4

To find the median number of residential students, we first need to order the number of residential students in all 4 terms:

$$13, 13, 16, 22$$

There are two numbers in the middle: '13' and '16'. The median is the mean of these two numbers: $\frac{13 + 16}{2} = 14.5$.

Solution: In these 4 terms, the median of residential students in this class was 14.5.

Question 5

To compare the total number of commuter and residential students in all 4 terms, we should compare the mean. The median is not accurate because some numbers were not considered; as a comparison, each number is used when the mean is calculated.

The mean of residential students is bigger than that of commuter students, so more residential students attended this class in all four terms. Notice that the median of commuter students is actually bigger than that of residential students.

Don't think that the median is not useful. When there are outliers (very big or small numbers compared to most numbers in the group), the median is better than mean. That's why in newspapers, you will more than likely see "the median house price" than "the mean house price," etc.

Correct Answers

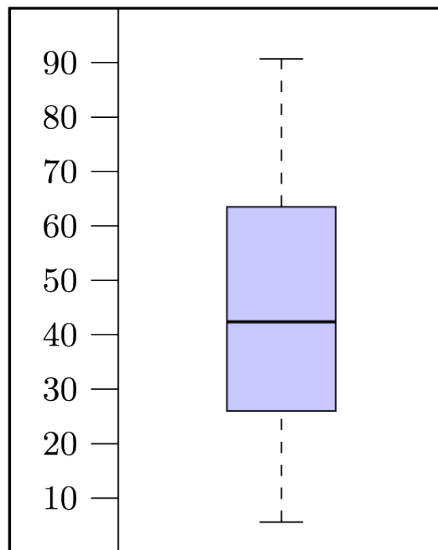
1. 14
2. 16
3. 15.5

4. 14.5
5. mean

Problem 2

The following diagram shows a boxplot for the lifetimes (in months) of a sample of 30 lightbulbs.

Box and whiskers plot for lifetime of lightbulbs



(a) What is the median lifetime?

- 5.6

- 26.0
- 42.4
- 63.5
- 90.7

(b) Which of the following statements is correct?

- The mean is larger in value than the median
- The mean is smaller in value than the median
- The mean is equal in value to the median
- There is insufficient information to make a comparison of the mean and median

Correct Answers

- a. 42.4
- b. The mean is larger than the median

3 PROBABILITY

Problem 1

Suppose that you have 12 cards. 7 are gray and 5 are yellow. The 7 gray cards are numbered 1, 2, 3, 4, 5, 6, and 7. The 5 yellow cards are numbered 1, 2, 3, 4, and 5. The cards are well shuffled. You randomly draw one card.

- G = the event the card drawn is gray
- E = the event the card drawn is even-numbered

$$P(G) =$$

$$P(E) =$$

$$P(G \text{ and } E) =$$

$$P(G \text{ or } E) =$$

$$P(G|E) =$$

$$P(E|G) =$$

Correct Answers

$$P(G) = 0.583333$$

$$P(E) = 0.416667$$

$$P(G \text{ and } E) = .25$$

$$P(G \text{ or } E) = .75$$

$$P(G|E) = .6$$

$$P(E|G) = .428571$$

Problem 2

Suppose that you have 15 cards. 9 are gray and 6 are yellow. The cards are well shuffled. You randomly draw two cards, **with replacement**.

- G_1 = the event the first card drawn is gray
- G_2 = the event the second card drawn is gray
- E = the event at least one card is gray

$$P(G_1 \text{ and } G_2) =$$

$$P(E) =$$

$$P(G_2|G_1) =$$

Correct Answers

$$P(G_1 \text{ and } G_2) = 0.36$$

$$P(E) = 0.84$$

$$P(G_2|G_1) = 0.6$$

Problem 3

In a box of assorted cookies, 43% contain chocolate, 12% contain nuts, and 3% contain both chocolate and nuts. Jada is allergic to both chocolate and nuts.

What is the probability that a cookie contains chocolate or nuts (they can't eat it)?

What is the probability that a cookie does not contain chocolate or nuts (they can eat it)?

Correct Answers

- 52%
- 48%

Problem 4

A soft-serve ice cream shop sold 226 ice creams in a day. The following table identifies the ice creams they sold by flavor and whether they were served in a cup or cone. Fill in the missing values.

Flavor	Chocolate	Vanilla	Swirl	Strawberry	Total
Cup	20		17	4	47
Cone	82	12		15	
Total		18			

- A. What is the probability that a randomly selected ice cream was served in a cup?
- B. What is the probability that a randomly selected ice cream was either chocolate or swirl?
- C. What is the probability that a randomly selected ice cream was chocolate served in a cup?
- D. What is the probability that a randomly selected ice cream was strawberry, given that it was served in a cone?
- E. What is the probability that a randomly selected ice cream was not chocolate?

Correct Answers

Flavor	Chocolate	Vanilla	Swirl	Strawberry	Total
Cup	20	6	17	4	47
Cone	82	12	70	15	179
Total	102	18	87	19	226

- A. 0.207965
- B. 0.836283
- C. 0.0884956
- D. 0.0837989
- E. 0.548673

Problem 5

One hundred nurses were surveyed about their living situation. The results of this survey are given in the following table.

	<i>Renter (R)</i>	<i>Homeowner (R')</i>	<i>Total</i>
Lives alone (A)	29	21	50
Does not live alone (A')	35	15	50
Total	64	36	100

If a nurse is selected at random from those surveyed, find the probability of each of the following events.

1. The nurse is a renter or lives alone. Probability =
2. The nurse is a homeowner or does not live alone.
Probability =
3. The nurse is a renter or does not live alone. Probability =

Correct Answers

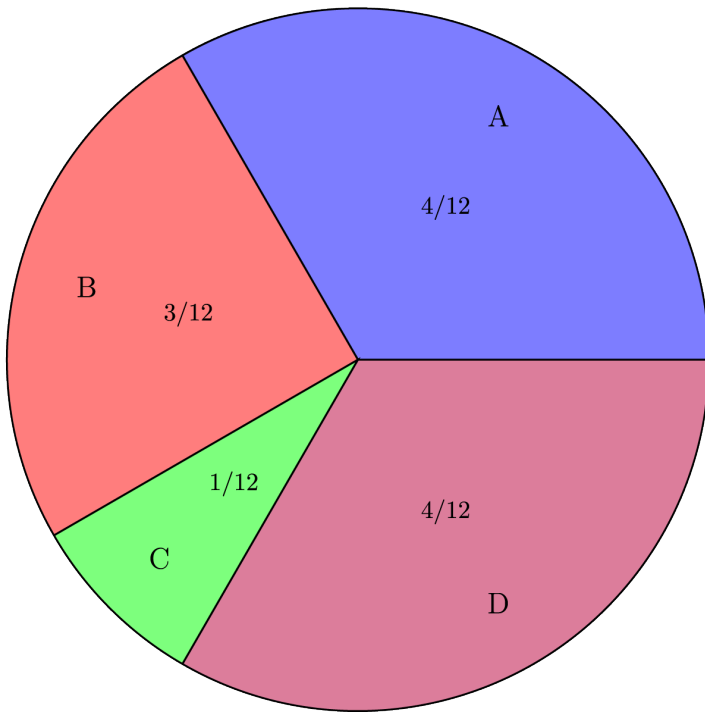
1. Probability = $1 - \frac{15}{100}$
2. Probability = $1 - \frac{29}{100}$
3. Probability = $1 - \frac{21}{100}$

4.1, 4.2 RANDOM VARIABLES

Problem 1

Consider the following game of chance based on the spinner below:

Each spin costs \$4. If the spinner lands on B the player wins \$10. If the spinner stops on D the player wins a dime. Otherwise, the player wins nothing.



Calculate the player's expected profit.

Note: Express your answer to at least three decimal places in dollar form.

Correct Answers

-1.46667

6.1, 6.2 NORMAL DISTRIBUTION

Problem 1

According to a study done by statistics students, the height for Chinese adult males is normally distributed with an average of 66 inches and a standard deviation of 2.5 inches. Suppose one Chinese adult male is randomly chosen.

1. Find the probability that the person is between 65 and 70 inches.
2. What percent of Chinese adult males are over 71.5 inches?

Correct Answers

1. 0.600623
2. 0.0139035

7 CENTRAL LIMIT THEOREM

Problem 1

A can of Ocean brand tuna is supposed to have a net weight of 6 ounces. The manufacturer tells you that the net weight is actually a Normal random variable with a mean of 5.95 ounces and a standard deviation of 0.2 ounces. Suppose that you draw a random sample of 42 cans.

Part i) Suppose the number of cans drawn is doubled. How will the standard deviation of sample mean weight change?

- It will increase by a factor of 2.
- It will decrease by a factor of 2.
- It will decrease by a factor of square root of 2.
- It will increase by a factor of square root of 2.
- It will remain unchanged.

Part ii) Suppose the number of cans drawn is doubled. How will the mean of the sample mean weight change?

- It will increase by a factor of square root of 2.

- It will increase by a factor of 2.
- It will decrease by a factor of square root of 2.
- It will decrease by a factor of 2.
- It will remain unchanged.

Part iii) Consider the statement: 'The distribution of the mean weight of the sampled cans of Ocean brand tuna is Normal.'

- It is a correct statement, but it is not a result of the Central Limit Theorem.
- It is a correct statement, and it is a result of the Central Limit Theorem.
- It is an incorrect statement. The distribution of the mean weight of the sample is not Normal.

Solution

The distribution of the weight of a single can of tuna is given to be Normal with mean 5.95 oz. and standard deviation 0.2 oz.

(i) The sample mean weight of a sample of 42 such cans is a random variable that follows the Normal distribution with mean 5.95 and standard deviation $\frac{0.2}{\sqrt{42}}$. If we were to double

the sample size to 84, the standard deviation of the sample mean would decrease by a factor of $\sqrt{2}$.

(ii) Changing the sample size does not affect the mean of the sampling distribution of the sample mean, which would remain at 5.95.

(iii) The sample mean here is normally distributed, though there is no need to apply the Central Limit Theorem for this result. Instead, the result follows exactly since the distribution of the weights of the individual cans is Normal.

Correct Answers

- i. C
- ii. E
- iii. A

8.1, 8.2 CONFIDENCE INTERVALS

Problem 1

A report in a research journal states that the average decrease in systolic blood pressure of people on a certain blood pressure medication is 24 mmHg with a margin of error of ± 4 mmHg with confidence level $C = 95\%$.

- a. According to this information, the mean decrease in systolic blood pressure of people on this medication, μ , could be as low as _mmHg.
- b. If the study is repeated, how large should the sample size be so that the margin of error would be less than 2 mmHg? (Assume $\sigma = 9$ mmHg.)

Correct Answers

- a. 20
- b. 78

Problem 2

An online used car company sells pre-owned cars. For 30 randomly selected transactions, the mean price is 2000 dollars.

- a. Assuming the population standard deviation for transaction prices is 300 dollars, obtain a 99% confidence interval for the mean price of all transactions. Please carry at least three decimal places in intermediate steps. Give your final answer to the nearest two decimal places.

Confidence interval: (,)

- b. Which of the following is a correct interpretation for your answer in part (a)? Select ALL the correct answers, there may be more than one.
1. 99% of the cars they sell have a price that lies inside this interval.
 2. There is a 99% chance that the mean price of all transactions lies in the interval.
 3. If we repeat the study many times, approximately 99% of the calculated confidence intervals will contain the mean price of all transactions.
 4. We can be 99% confident that the mean price for this sample of 30 transactions lies in the interval.
 5. We can be 99% confident that all of the cars they sell have a price inside this interval.
 6. None of the above.

Correct Answers

- a. 1858.91, 2141.09
- b. Choice 2, choice 3

Problem 3

How much money do winners go home with from the television quiz show *Jeopardy* ? To determine an answer, a random sample of winners was drawn and the amount of money each won was recorded and listed below. Estimate with 94% confidence the mean winning's for all the show's players.

43460, 38213, 42876, 38921, 36709, 26735, 35472, 31319, 42833, 27159, 32799, 39271, 44016, 28191, 37087

high end of confidence interval =

low end of confidence interval =

Correct Answers

- High end: 39480.1636458906
- Low end: 33194.6363541094

8.3-9.3 CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

Problem 1

Elaine wants to determine a 99 percent confidence interval for the true proportion p of high school students in the area who attend their home basketball games. Out of n randomly selected students, she finds that exactly half attend their home basketball games. About how large would n have to be to get a margin of error less than 0.01 for p ?

Note: Use the values for z^* from a z-table or t-table, and round to the smallest integer that works.

$$n \approx$$

Correct Answers

16590

Problem 2

For each situation, state the null and alternative hypotheses: (Type "mu" for the symbol μ , e.g. **mu** > **1** for the mean is greater than 1, **mu** < **1** for the mean is less than 1, **mu not = 1** for the mean is not equal to 1. Please do not include units such as "mm" or "\$" in your answer.)

(a) The diameter of a spindle in a small motor is supposed to be 6 millimeters (mm) with a standard deviation of 0.15mm. If the spindle is either too small or too large, the motor will not work properly. The manufacturer measures the diameter in a sample of 15 spindles to determine whether the mean diameter has moved away from the required measurement. Suppose the sample has an average diameter of 4.51 mm.

H_0 :

H_a :

(b) Destiny thinks that prices in Springfield are lower than the rest of the country. They read that the nationwide average price of a certain brand of laundry detergent is \$55 with standard deviation \$2.17. They take a sample from 3 local Springfield stores and find the average price for this same brand of detergent is \$20.81.

H_0 :

H_a :

Correct Answers

(a) $\mu = 6$; $\mu \neq 6$

(b) $\mu = 55$; $\mu < 55$

Problem 3

Consider the following hypothesis test. The null hypothesis is "The mean body temperature for humans is 98.6 degrees Fahrenheit," and the alternative hypothesis is "The mean body temperature for humans differs from 98.6 degrees Fahrenheit."

Answer the following questions.

a. "The mean body temperature for humans in fact is 98.6 degrees Fahrenheit, but the result of the sampling leads to the conclusion that the mean body temperature for humans differs from 98.6 degrees Fahrenheit" is a

- A. Type II error
- B. correct decision
- C. Type I error

b. "The mean body temperature for humans in fact differs from 98.6 degrees Fahrenheit, and the result of the sampling leads to that conclusion" is a

- A. Type II error
- B. correct decision
- C. Type I error

c. "The mean body temperature for humans in fact is 98.6 degrees Farenheit, and the result of the sampling does not lead to the rejection of the fact that the mean body temperature is 98.6 degrees Farenheit" is a

- A. Type II error
- B. correct decision
- C. Type I error

d. "The mean body temperature for humans in fact differs from 98.6 degrees Farenheit, but the result of the sampling fails to lead to that conclusion" is a

- A. Type II error
- B. correct decision
- C. Type I error

Correct Answers

- a. C
- b. B
- c. B
- d. A

9.4, 9.5 HYPOTHESIS TESTING

Problem 1

Jaidyn thinks that people living in a rural environment have a healthier lifestyle than other people. He thinks that the average lifespan in the USA is 77 years. A random sample of 14 obituaries from newspapers from rural towns in Massachusetts give $\bar{x} = 81.09$ and $s = 0.91$. Does this sample provide evidence that people living in rural communities live longer than 77 years?

a. State the null and alternative hypotheses: (Type “mu” for the symbol μ , e.g. **mu** > **1** for the mean is greater than 1, **mu** < **1** for the mean is less than 1, **mu not** = **1** for the mean is not equal to 1)

H_0 :

H_a :

b. Find the test statistic, $t =$

c. To a 10% level of significance, does this sample provide evidence that people living in rural communities live longer than 77 years. [Yes/No]

Correct Answers

a. $\mu = 77; \mu > 77$

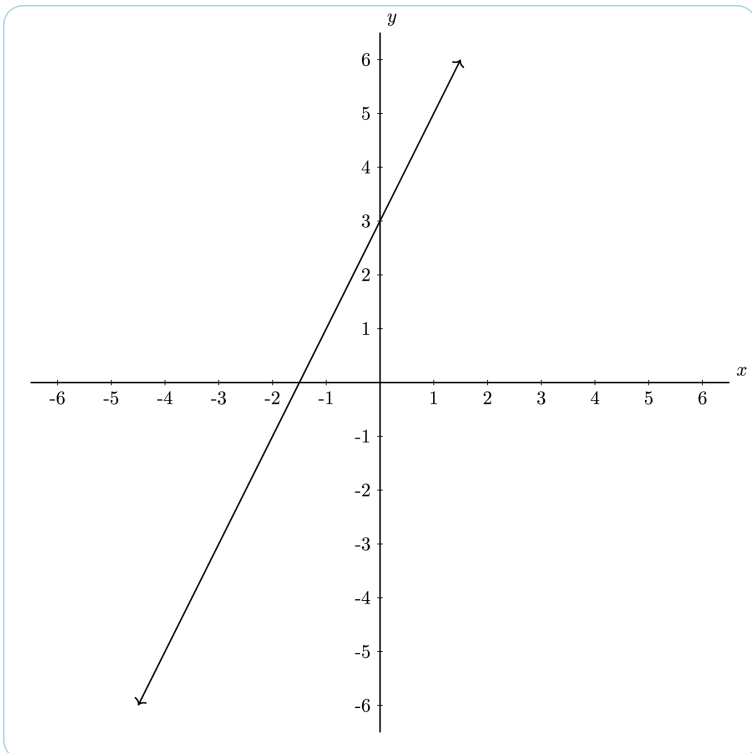
b.
$$\frac{81.09 - 77}{\frac{0.91}{\sqrt{14}}}$$

c. Yes

12.1, 12.2 LINEAR EQUATIONS AND SCATTER PLOTS

Problem 1

The following is a graph of a line:



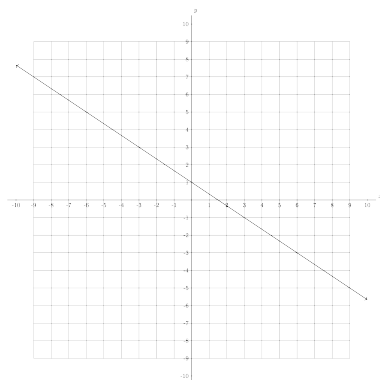
Write the equation of the line $y =$

Correct Answers

$$2x + 3$$

Problem 2

The following is a graph of a line:



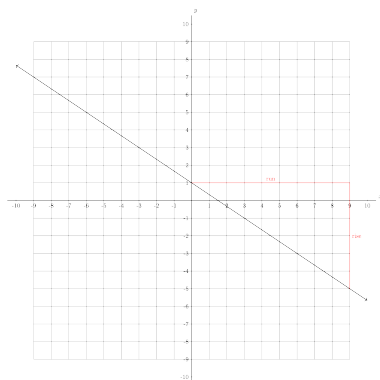
Write the equation of the line $y =$

Solution

A line's slope-intercept equation has the form $y = mx + b$, where m is the slope and b is the y -intercept. We first find the slope.

To find the slope of a line from its graph, we identify two points, and then draw a slope triangle. It's wise to choose points with integer coordinates. For this problem, we choose $(0, 1)$ and $(9, -5)$.

Next, we draw a slope



triangle and find the “rise” and “run”. In this problem, the rise is -6 and the run is 9 .

$$\begin{aligned} \text{slope} &= \frac{\text{rise}}{\text{run}} \\ &= \frac{-6}{9} \\ &= -\frac{2}{3} \end{aligned}$$

This line’s slope is $-\frac{2}{3}$.

It’s clear in the graph that this line’s y -intercept is $(0, 1)$.

So this line’s slope-intercept equation is $y = -\frac{2}{3}x + 1$

Correct Answers

$$-0.666667x + 1$$

12.3, 12.4 LINEAR REGRESSION

Problem 1

For each problem, select the best response.

(a) The owner of a chain of supermarkets notices that there is a positive correlation between the sales of beer and the sales of ice cream over the course of the previous Seasons when sales of beer were above average, sales of ice cream also tended to be above average. Likewise, during seasons when sales of beer were below average, sales of ice cream also tended to be below average. Which of the following would be a valid conclusion from these facts?

- a. A scatterplot of monthly ice cream sales versus monthly beer sales would show that a straight line describes the pattern in the plot, but it would have to be a horizontal line.
- b. Sales records must be in error. There should be no association between beer and ice cream sales.
- c. The sale of beer and ice cream may both be affected by another variable such as the outside temperature.

- d. Evidently, for a significant proportion of customers of these supermarkets, drinking beer causes a desire for ice cream or eating ice cream causes a thirst for beer.
- e. None of the above.

(b) A researcher observes that, on average, the number of children in cities with major league baseball teams is larger than in cities without major league baseball teams. The most plausible explanation for this observed association is

- a. the high number of children is responsible for the presence of more major league baseball teams (more children means potentially more fans at the ballpark since people like to bring their kids to baseball games, thus making it attractive for an owner to relocate to such cities).
- b. the presence of a major league baseball team causes the number of children to rise (perhaps people decide to have children so they can bring them to the ballpark).
- c. the observed association is purely coincidental. It is implausible to believe the observed association could be anything other than accidental.
- d. the association is due to the presence of a lurking variable (major league teams tend to be in large cities with more people, hence a greater number of children).
- e. None of the above.

(c) When possible, the best way to establish that an observed association is the result of a cause-and-effect relation is by means of

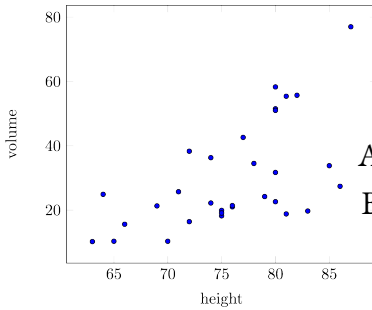
- a. the correlation coefficient.
- b. a well designed experiment.
- c. the square of the correlation coefficient.
- d. the least squares regression line.
- e. None of the above.

Correct Answers

- C
- D
- B

Problem 2

The height (in feet) and volume of usable lumber (in cubic feet) of 32 cherry trees are measured by a researcher. The goal is to determine if volume of usable lumber can be estimated from the height of a tree. The results are plotted below.



(a) In this study, the response variable is

- A. height.
- B. height or volume. It doesn't matter which is considered the response.
- C. neither height nor volume.

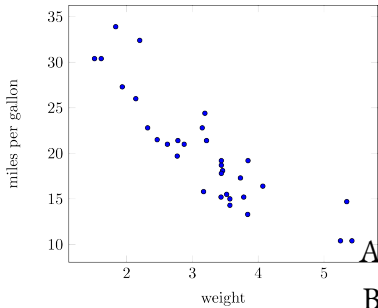
The measuring instrument used to measure height is the response variable.

- D. volume.

(b) The scatterplot suggests

- A. there is an outlier in the plot.
- B. there is a positive association between height and volume.
- C. both A and B.
- D. neither A nor B.

Now consider the following scatterplot of the weight of cars (in thousands of pounds) versus the miles they travel per gallon of gas consumed (mpg).



(a) A plausible value for the correlation between weight and mpg is

- A. -1.0
- B. +0.2
- C. -0.9
- D. +0.7

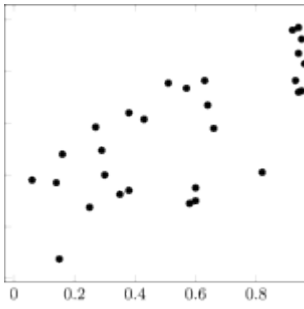
Correct Answers

- D
- C
- C

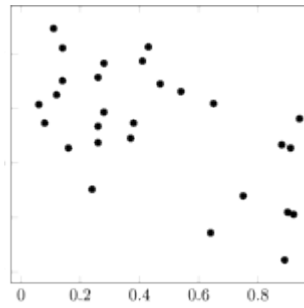
Problem 3

Match the correlation coefficients with their scatterplots. Select the letter of the scatterplot below which corresponds to the correlation coefficient.

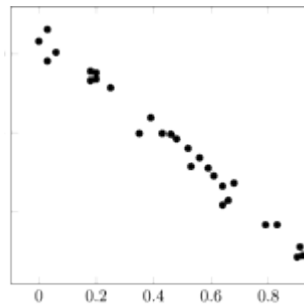
A.



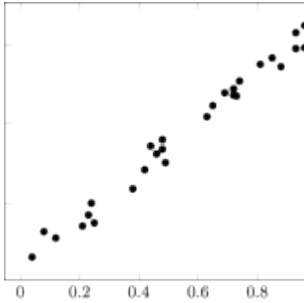
B.



C.



D.



- $r = -0.66$
- $r = -0.99$
- $r = 0.70$
- $r = 0.99$

Correct Answers

- A. $r = 0.70$
- B. $r = -0.66$
- C. $r = -0.99$
- D. $r = 0.99$

Problem 4

Data were collected from all 10 year olds in a soccer league. Based on this dataset, a least squares regression model was

fitted to predict weight Y (in kg) from height X (in cm). The model fitted was

$$Y = 0.99X - 101.24$$

The interpretation of the **slope** of the regression line would be

- A. For each 1 cm increase in height, you would expect the weight to increase by 0.99 kg
- B. For each 1 kg increase in weight, you would expect the height to decrease by 101.24 cm
- C. For each 1 kg increase in weight, you would expect the height to increase by 0.99 cm
- D. For each 1 cm increase in height, you would expect the weight to decrease by 101.24 kg

Correct Answers

A

Problem 5

Is the number of games won by a softball team in a season related to the team batting average? The table below shows the batting average (in thousandths) and the number of games won of 8 teams.

Team	Batting Average	Games Won
1	272	40
2	262	45
3	317	30
4	304	41
5	276	44
6	286	45
7	263	50
8	320	37

Using batting average as the explanatory variable x , do the following:

- The correlation coefficient is $r =$
- The equation of the least squares line is $\hat{y} =$

Correct Answers

- 0.815699
- $-0.214446x + 103.073$

GRANT INFORMATION

The U.S. Department of Education, the granting agency for the ROTEL project, requires information about the grant be included in the back matter. The text for this section is provided below.

The contents of this book were developed under a grant from the Fund for the Improvement of Postsecondary Education, (FIPSE), U.S. Department of Education. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

For more information about the ROTEL Project, please visit our project website.