



**AP Statistics
MidTerm Exam
STUDY GUIDE
2013-14**

The real exam:

** covers material from chapters 2-15

Multiple Choice – **50% of grade** (budget about 60 minutes for this part)

- 25 questions
- They will be from tests we've taken this year, or this Study Guide, or be of very similar nature to those on this Study Guide

Free Response (budget about 60 total minutes for parts A & B combined)

- Part A = **37.5% of grade** (5 shorter questions)
- Part B = **12.5% of grade** (1 longer question)
- They will be from tests we've taken this year, or this Study Guide, or be of very similar nature to those on this Study Guide

Name _____

Period _____

SECTION I

Number of Questions – 40

Percent of Total Grade – 50

Directions: Solve each of the following problems, using the available space (or extra paper) for scratchwork. Decide which is the best of the choices given and place that letter on the ScanTron sheet. No credit will be given for anything written on these pages for this part of the test. Do not spend too much time on any one problem.

- An outlier is an observation that
 - is seen more frequently than the other observations in the data set.
 - is seen less frequently than the other observations in the data set.
 - is always smaller than the other observations in the data set.
 - is always larger than the other observations in the data set.
 - is significantly different from the other observations in the data set.

- Suppose that a Normal model describes the acidity (pH) of rainwater, and that water tested after last week's storm had a z-score of +1.8. This means that the acidity of that rain
 - had a pH 1.8 standard deviations higher than that of average rainwater.
 - varied with a standard deviation of 1.8.
 - had a pH 1.8 higher than average rainfall.
 - had a pH 1.8 times that of average rainwater.
 - had a pH of 1.8.

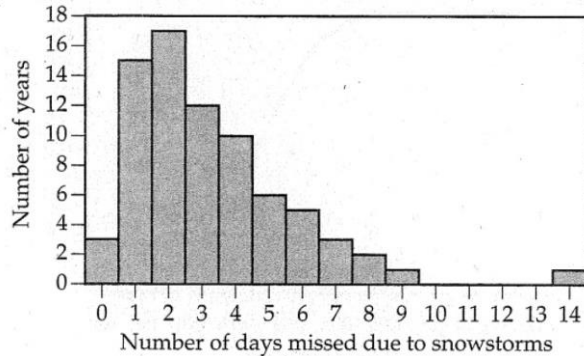
- A large company has offices in two locations, one in New Jersey and one in Utah. The mean salary of office assistants in the New Jersey office is \$28,500. The mean salary of office assistants in the Utah office is \$22,500. The New Jersey office has 128 office assistants and the Utah office has 32 office assistants. What is the mean salary paid to the office assistants in this company?

A. \$22,500	B. \$23,700	
C. \$25,500	D. \$27,300	E. \$28,500

- Which statement about bias is true?
 - Bias results from random variation and will always be present.
 - Bias results from a sampling method likely to produce samples that do not represent the population.
 - Bias is usually reduced when sample size is larger.

A. I only	B. III only	
C. II only	D. I and III only	E. II and III only

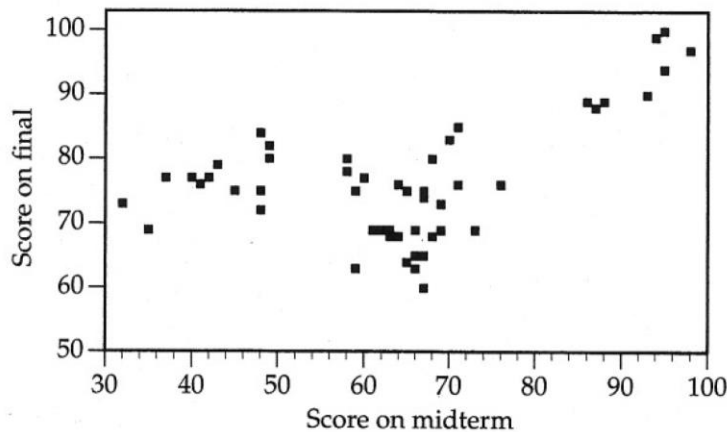
5. In the northern US, schools are sometimes closed during winter due to severe snowstorms. At the end of the school year, schools have to make up for the days missed. The following graph shows the frequency distribution of the number of days missed due to snowstorms per year, using data collected from the past 75 years.



Which of the following should be used to describe the center of this distribution?

- A. Mean, because it is an unbiased estimator.
 - B. Median, because the distribution is skewed.
 - C. IQR, because it excludes outliers and includes on the middle 50 percent of data.
 - D. First quartile, because the distribution is left skewed.
 - E. Standard deviation, because it is unaffected by the outliers.
6. Increasing the frequencies in the tails of a distribution will
- A. increase the standard deviation.
 - B. not affect the standard deviation.
 - C. not affect the standard deviation as long as the increases are balanced on each side of the mean.
 - D. reduce the standard deviation.
 - E. none of these
7. The probability that a car will skid on a bridge on a rainy day is 0.75. Today the weather station announced that there is a 20 percent chance of rain. What is the probability that it will rain today and that a car will skid on the bridge?
- A. 0.0300
 - B. 0.0375
 - C. 0.1500
 - D. 0.3000
 - E. 0.9500
8. Randomizing
- I. can protect you against factors you know are in the data.
 - II. can help protect you against factors that you aren't even aware of.
 - III. makes it impossible for us to draw inferences about the population when we see only a sample.
- A. none of these
 - B. III only
 - C. I, II, and III
 - D. II and III only
 - E. I and II only

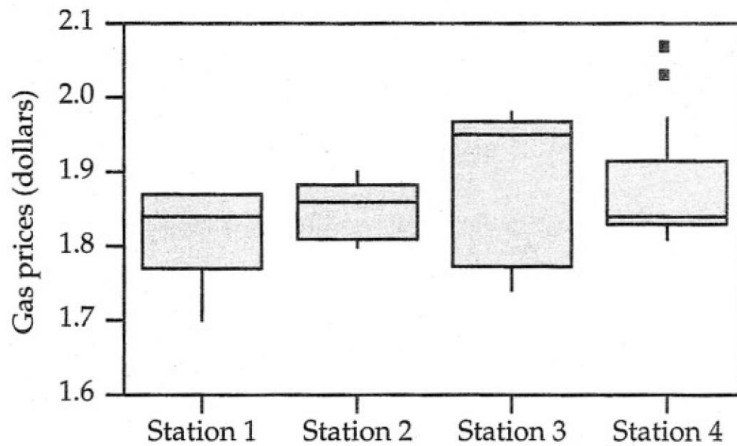
9. Extra study sessions were offered to students after the MidTerm to help improve their understanding of statistics. Student scores on the MidTerm and the Final Exam were recorded. The following scatterplot shows the Final test scores against the MidTerm test scores.



Which of the following statements correctly interprets the scatterplot?

- A. All students have shown significant improvement in the Final Exam scores as a result of the extra study sessions.
 - B. The extra study sessions were of no help. Each student's Final Exam score was about the same as his or her score on the MidTerm.
 - C. The extra study sessions further confused the students. All student scores decreased from MidTerm to Final Exam.
 - D. Students who scored below 55 on the MidTerm showed considerable improvement on the Final Exam; those who scored between 55 and 80 on the MidTerm showed minimal improvement on the Final Exam; and those who scored above 80 on the MidTerm showed almost no improvement on the Final Exam.
 - E. Students who scored below 55 on the MidTerm showed minimal improvement on the Final Exam; those who scored between 55 and 80 on the MidTerm showed moderate improvement on the Final Exam; and those who scored above 80 on the MidTerm showed considerable improvement on the Final Exam.
10. All but one of the statements below contain a mistake. Which one could be true?
- A. The correlation between height and weight is 0.568 inches per pound.
 - B. If the correlation between blood alcohol level and reaction time is 0.73, then the correlation between reaction time and blood alcohol level is -0.73.
 - C. The correlation between the breed of a dog and its weight is 0.435.
 - D. The correlation between gender and age is -0.171.
 - E. The correlation between weight and length of foot is 0.488.

11. A resident of Auto Town was interested in finding the cheapest gas prices at nearby gas stations. On randomly selected days over a period of one month, he recorded the gas prices (in dollars) at four gas stations near his house. The box plots of gas prices are as follows:



Which station has more consistent gas prices?

- A. Station 1 B. Station 2
 C. Station 3 D. Station 4 E. Station 4 without the outliers
12. A residuals plot is helpful because
- I. it will help us see whether our model is appropriate.
 II. it might show a pattern in the data that was hard to see in the original scatterplot.
 III. it will clearly identify influential points.
- A. I only B. II only
 C. I, II, and III D. I and III only E. I and II only
13. The observed time (in minutes) it takes a runner to complete a marathon are approximately normally distributed. The z-score for his running time this week was -2. Which one of the following statements is a correct interpretation of his z-score?
- A. This week his time was two minutes lower than his time last week.
 B. This week his time was two minutes lower than his best time ever.
 C. This week his time was two minutes lower than his average time.
 D. This week his time was two standard deviations lower than his average time.
 E. This week his time was two standard deviations lower than his time last week.
14. A numerical measure of the direction and strength of a linear association is its
- A. association. B. slope.
 C. predicted value. D. intercept. E. correlation.

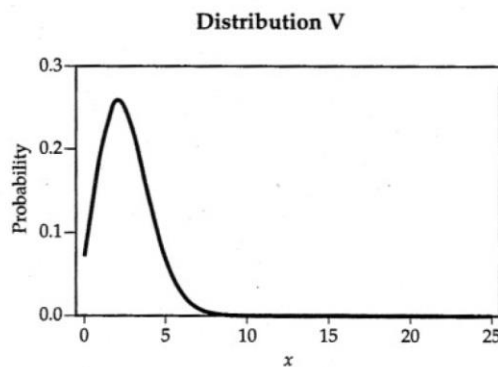
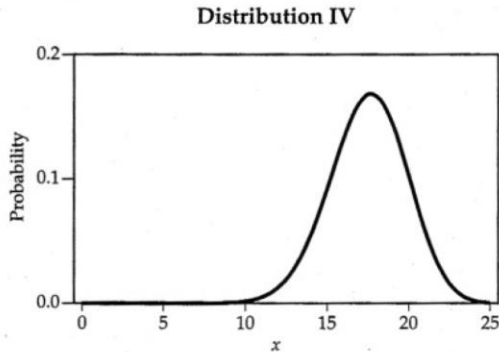
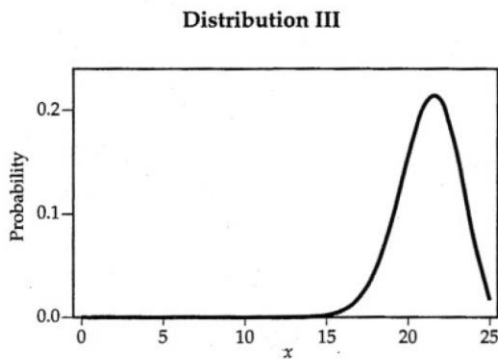
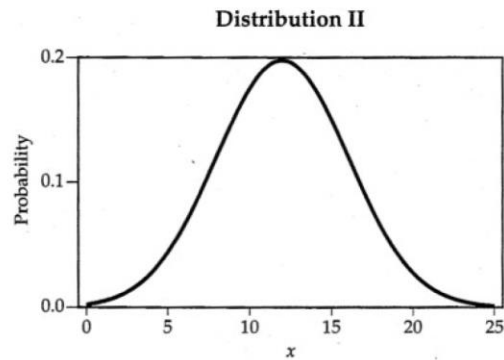
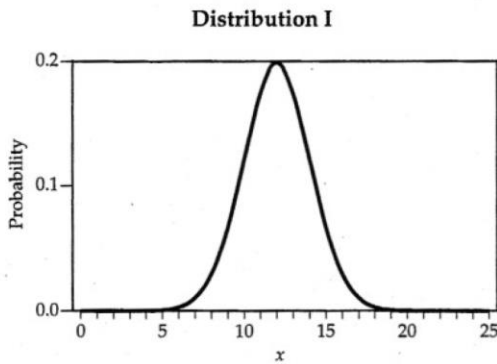
15. The probability that Ted enrolls in an English class is $\frac{1}{3}$. If he does enroll in an English class, the probability that he enrolls in a mathematics class is $\frac{1}{5}$. What is the probability that her enrolls in both classes?
- A. $\frac{1}{15}$ B. $\frac{2}{15}$ C. $\frac{7}{15}$ D. $\frac{3}{5}$ E. $\frac{13}{15}$
16. The linear regression model contains
- I. two easily-estimated parameters.
 - II. a meaningful measure of how well the model fits the data.
 - III. the ability to predict new values.
 - IV. a self-check to help us avoid silly mistakes.
- A. I and III only B. I, II, and III only
C. all of these D. none of these E. II only
17. A school committee member is lobbying for an increase in the gasoline tax to support the county school system. The local newspaper conducted a survey of county residents to assess their support for such an increase. What is the population of interest here?
- A. all school-aged children
B. all county residents
C. all county residents with school-aged children
D. all county residents with children in the county school system
E. all county school system teachers
18. With observational data, as opposed to data from a designed experiment, there is no way to be sure that a(n) _____ is not the cause of any apparent association.
- A. causal variable B. confounding variable
C. outlier D. lurking variable E. regression variable
19. A manufacturer of ready-bake cake mixes is interested in designing a study to test the effects of four different temperature levels ($300^{\circ}, 325^{\circ}, 350^{\circ}, 375^{\circ}\text{F}$), two different types of pans (glass and metal), and three different types of ovens (gas, electric, and microwave) on the texture of its cakes. How many different treatment combinations are to be used in this study?
- A. 3 B. 9 C. 18 D. 20 E. 24

20. A regression analysis of students' AP Statistics scores and the number of hours they spent doing homework found $R^2 = 0.32$. Which of these is true?

- I. 32% of student test scores can be correctly predicted with this model.
- II. Homework accounts for 32% of your grade in AP Stats.
- III. There's a 32% chance that you'll get the score this model predicts for you.

- A. all of these
- B. none of these
- C. II only
- D. III only
- E. I and II

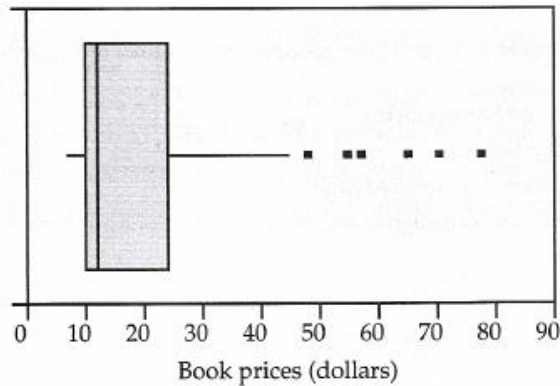
21. Which one of the following distributions could have a mean of approximately 12 and a standard deviation of approximately 2?



- A. Distribution I
- B. Distribution II
- C. Distribution III
- D. Distribution IV
- E. Distribution V

22. The relationship between the selling price of a car (in \$1,000) and its age (in years) is estimated from a random sample of cars of a specific model. The relation is given by the formula $\text{selling price} = 15.9 - 0.983(\text{age})$. Which of the following can we conclude from this equation?
- A. For every year the car gets older, the selling price goes down by approximately 9.83 percent.
 - B. A new car costs on the average \$9,830.
 - C. For every year the car gets older, the selling price drops by approximately \$983.
 - D. A new car costs \$16,883.
 - E. For every year the car gets older, the selling price drops by approximately \$1,590.
23. After a frost warning was issued, the owner of a large orange grove asked his workers to spray all his trees with water. The water was supposed to freeze and form a protective covering of ice around the orange blossom. Nevertheless, the owner suspected that some trees suffered considerable damage due to the frost. To estimate the proportion of trees that suffered more than 50 percent damage due to the frost, he took a random sample of 100 trees from his grove. What constitutes an observation (measurement) in this experiment?
- A. the proportion of trees that suffered more than 50 percent damage due to frost
 - B. the number of trees affected by the frost
 - C. for each sampled tree, whether it suffered more than 50 percent damage or at most 50 percent damage
 - D. for each sampled tree, whether it was sprayed with water or not sprayed with water
 - E. the number of trees sampled from the grove
24. For which of the following purposes would it be most unreasonable to use a census?
- A. to determine the proportion of students with a learning disability in a small rural area high school
 - B. to determine the proportion of red snappers with a high mercury level in the Gulf of Mexico
 - C. to determine the difference between the proportion of engineering professors and the proportion of business professors in favor of the new teaching initiative at a large university
 - D. to determine the mean wage earned by construction workers in a small town
 - E. to determine the mean selling price of houses in your neighborhood
25. A small kiosk at the Atlanta airport carries souvenirs in the price range of \$3.99 to \$29.99, with a mean price of \$14.75. The airport authorities decide to increase the rent charged for a kiosk by five percent. To make up for the increased rent, the kiosk owner decides to increase the prices of all items by 50 cents. As a result, which of the following will happen?
- A. The mean price and the range of prices will increase by 50 cents.
 - B. The mean price will remain the same, but the range of prices will increase by 50 cents.
 - C. The mean price and the standard deviation of prices will increase by 50 cents.
 - D. The mean price will increase by 50 cents, but the standard deviation of prices will remain the same.
 - E. The mean price and the standard deviation will remain the same.

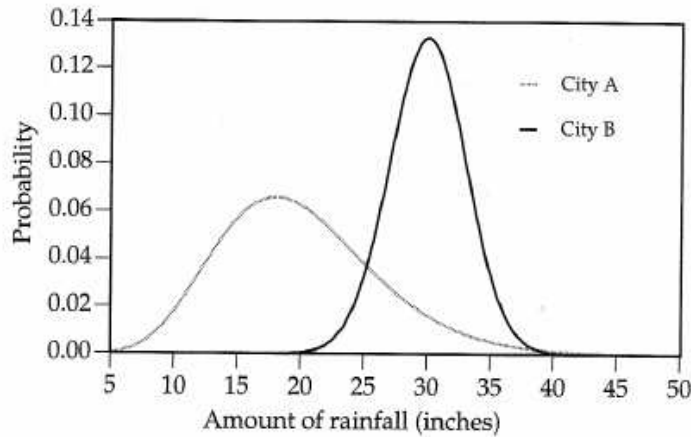
26. The following boxplot summarizes the prices of books at a bookstore.



Which of the following is true about the prices of the books at this store?

- A. This store carries more books priced above the mean than below the mean.
 - B. This store carries more books priced below the mean than above the mean.
 - C. This store carries about the same number of books at different prices in the entire price range.
 - D. The mean price of books at this store is the same as the median price of books.
 - E. The mean price of books at this store is lower than the median price of books.
27. An experiment was designed to test the effects of three different types of paint on the durability of wooden toys. Because boys and girls tend to play differently with toys, a randomly selected group of children was divided into two groups by sex. Which of the following statements about this experiment is true?
- A. There are three types of paint and two sex groups, giving a total number of six treatment combinations in this experiment.
 - B. This experiment should be blocked by paint type.
 - C. This experiment should be blocked by sex.
 - D. This is a completely randomized experimental design.
 - E. This is a matched-pairs experimental design in which one boy and one girl are matched by age to form a pair, then they are assigned to a group.
28. Scores on the take-home part of a test are approximately normally distributed, with a mean of 40 and a standard deviation of 3. Scores on the in-class part of the same test are also approximately normally distributed, with a mean of 34 and a standard deviation of 6. The final score on the test is the sum of the score on the take-home part and the score on the in-class part. What are the mean and standard deviation of the final test scores?
- A. Mean 74, standard deviation 4.50
 - B. Mean 37, standard deviation 6.71
 - C. Mean 37, standard deviation 4.50
 - D. Mean 74, standard deviation 6.71
 - E. Mean 74, standard deviation 3.00

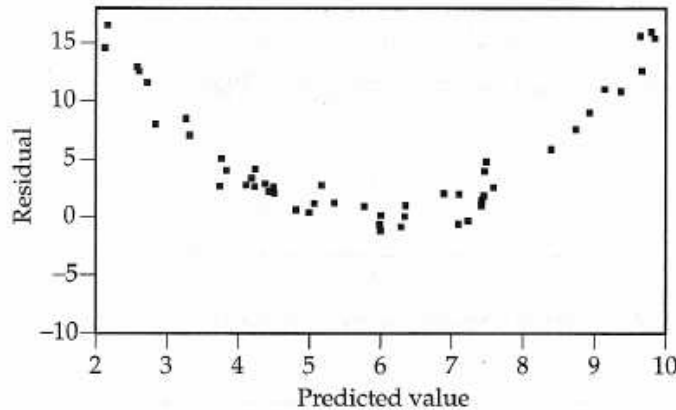
29. The following graph summarizes data collected on annual rainfall in two cities for the past 150 years.



Which of the following conclusions can be made from this graph?

- A. The cities have different mean annual rainfalls, but the range of their annual rainfalls is approximately the same.
 - B. On average, City B gets more rain than City A, but it has a smaller range of annual rainfall.
 - C. On average, City B gets more rain than City A, and it has a larger range of annual rainfall.
 - D. On average, City A gets more rain than City B, but it has a smaller range of annual rainfall.
 - E. On average, City A gets more rain than City B, and it has a larger range of average rainfall.
30. On a boxplot, an observation is classified as an outlier if
- A. it is smaller than $\bar{X} - 3S$ or larger than $\bar{X} + 3S$.
 - B. it is larger than $\bar{X} - 3S$ and smaller than $\bar{X} + 3S$.
 - C. it is smaller than Q_1 or larger than Q_3 .
 - D. it is smaller than $Q_1 - (1.5)(IQR)$ or larger than $Q_3 + (1.5)(IQR)$.
 - E. it is larger than $Q_1 - (1.5)(IQR)$ and smaller than $Q_3 + (1.5)(IQR)$.
31. Which of the following statements about any two events A and B is true?
- A. $P(A \cup B)$ implies events A and B are independent.
 - B. $P(A \cup B) = 1$ implies events A and B are mutually exclusive.
 - C. $P(A \cap B) = 0$ implies events A and B are independent.
 - D. $P(A \cap B) = 0$ implies events A and B are mutually exclusive.
 - E. $P(A \cap B) = P(A) - P(B)$ implies A and B are equally likely events.

32. Sixty pairs of measurements were taken at random to estimate the relation between variables X and Y . A least squares line was fitted to the collected data. The resulting residual plot is as follows:



Which of the following conclusions is appropriate?

- A. A line is an appropriate model to describe the relation between X and Y .
 - B. A line is not an appropriate model to describe the relation between X and Y .
 - C. The assumption of The Law of Averages has been violated.
 - D. The variables X and Y are not related at all.
 - E. Without knowing what variables X and Y are, there is no way to form an appropriate conclusion about a possible linear connection between them.
33. An insurance agent is successful in selling a life insurance policy to 20 percent of the customers he contacts. He decides to construct a simulation to estimate the mean number of customers he needs to contact before being able to sell a policy. Which of the following schemes should he use to do the simulation?
- A. Assign numbers “0, 1” to successfully selling a policy to a customer and numbers “2, 3, 4, 5, 6, 7, 8, 9” to failing to sell a policy to a customer.
 - B. Assign numbers “0, 1” to successfully selling a policy to a customer and numbers “2, 3, 4” to failing to sell a policy to a customer.
 - C. Assign number “0” to successfully selling a policy to a customer and number “1” to failing to sell a policy to a customer.
 - D. Assign numbers “0, 1, 2, 3, 4” to successfully selling a policy to a customer and numbers “5, 6, 7, 8, 9” to failing to sell a policy to a customer.
 - E. Assign number “20” to successfully selling a policy to a customer and numbers “1, 3, 5, 7, 9, 11, 13, 15, 17, 19” to failing to sell a policy to a customer.
34. A newspaper reporter examined police reports of accidents during the past 12 months to collect data about the speed of a car and its stopping distance. The reporter then constructed a scatterplot and computed a correlation coefficient to show the relation between a car’s speed and its stopping distance. This is an example of
- A. a double-blind study.
 - B. a single-blind study.
 - C. a study involving no blinding at all.
 - D. an observational study.
 - E. a well-designed experiment.

35. A company with offices in five different countries is interested in estimating the proportion of its employees in favor of banning smoking on the office premises. It is known that the views of people from different countries on issues like smoking tend to vary due to the influence of different local social structures. Which of the following is an appropriate sampling technique to use in this situation, and why?
- Stratified sample, because the population is divided into five strata – namely, the five offices in five different countries.
 - Stratified sample, because people in different countries use different languages and thus will not be able to communicate to each other the results of a simple random sample.
 - Simple random sample, because the study has not been attempted before.
 - Simple random sample, because this is the only method that gives unbiased results.
 - Simple random sample, because this is the only method for which inferential techniques are available.
36. In a clinic, 50 patients with sleep disorders are randomly assigned to one of two different groups. Patients in one group are given medication before bedtime. Patients in the other group are given blindfolds and played soft music at bedtime. Each patient is attached to a machine that records breathing patterns. From the patterns, it is possible to determine if the patient is awake or asleep. The data will be used to decide which method is more effective in helping patients with sleep disorders. Which of the following statements is correct in the context of this experiment?
- This is a single-blind experiment, because only one group uses blindfolds.
 - This is a single-blind experiment, because only patients and not doctors use blindfolds.
 - This is a double-blind experiment, because patients are blindfolded and the doctor does not know which patient receives which treatment.
 - This experiment cannot be a single-blind experiment, because many patients do not like to be blindfolded.
 - This experiment cannot be a double-blind experiment, because patients will know which treatment they are receiving, although the examining doctor might not.
37. Data were collected on two variables, X and Y , and a least squares regression line was fitted to the data. The estimated equation for this data is $y = -2.29 + 1.70x$. One point has $x = 5$, $y = 6$. What is the residual for this point?
- A. 7.91 B. 6.21 C. 0.21 D. -0.21 E. -2.91
38. The relation between studying time (in hours) and grade on a semester exam (0-100) in a random sample of students in math class was found to be $\text{grade} = 50.25 + 10.2(\text{hours})$.
- How will a student's grade be affected if she studies for two hours?
- Her grade will go up by approximately 40.05 points.
 - Her grade will go up by approximately 20.4 points.
 - Her grade will remain unchanged.
 - Her grade will go down by approximately 10.2 points.
 - It cannot be determined from the information given.

39. A group of scientists wanted to estimate the proportion of geese returning to the same site for the next breeding season. Suppose they decided to increase the sample size from 200 to 2,000. How will this affect the standard deviation of the sample?
- A. The standard deviation will increase.
 - B. The standard deviation will decrease.
 - C. The standard deviation will remain unaffected.
 - D. The standard deviation will more closely resemble a bimodal distribution.
 - E. The standard deviation will more closely resemble a skewed distribution.
40. A company has 400 employees. Their mean income is \$20,500, and the standard deviation of their incomes is \$3,750. The distribution of incomes is normally distributed. How many of the 400 employees do you expect to have an income of between \$13,000 and \$28,000?
- A. 50 B. 100 C. 200 D. 390 E. 400

SECTION II

Part A

Percent of Section II Grade – 75

Directions: Show all of your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanations.

Zip Drive Lifetimes.

41. The lifetimes of zip drives marketed by ZipZap, Inc. are normally distributed, with a mean lifetime of 18 months and a standard deviation of three months. The company plans to offer a new warranty guaranteeing the replacement of failed zip drives during the warranty period. The replacement of failed drives is costly, however, and the company can afford to replace up to only two percent of its drives. Rounded to the nearest whole number, how many months of warranty should the company offer with these zip drives?

Choosing a Printer.

42. A company executive is interested in purchasing printers for his division. He contacts a wholesaler, who shows him two printers comparable in quality and price made by two different manufacturers. Both printers are suitable for the kind of printing generally done in his division. To help him decide between these two printers, the wholesaler offers to loan him a printer of each kind for one day. Since the printers' printing time per page depends on the type of the document printed, the executive selects 25 different documents of varying size and complexity for printing. Help this executive design an experiment to compare the mean printing times of the two printers.

Cabins and Campsites.

A state park has campsites available for pitching tents. Visitors are allowed to pitch tents only at the designated campsites. In addition to campsites, the park also has some log cabins for rent. The demand for cabins and campsites changes from month to month. The renting records for year 2001 are as follows:

Month	Number of Cabins Rented	Number of Campsites Rented
Jan	21	3
Feb	16	5
Mar	12	8
Apr	10	15
May	8	18
Jun	10	25
Jul	12	28
Aug	10	30
Sep	10	25
Oct	24	15
Nov	22	4
Dec	10	4

43a. Using a sheet of graph paper, display the data graphically to compare the pattern of cabin rentals with that of campsite rentals.

43b. Examine your graphical display. Write a few sentences describing the rental history of cabins and campsites in 2001.

SECTION II

Part B

Question 44

Percent of Section II Grade – 25

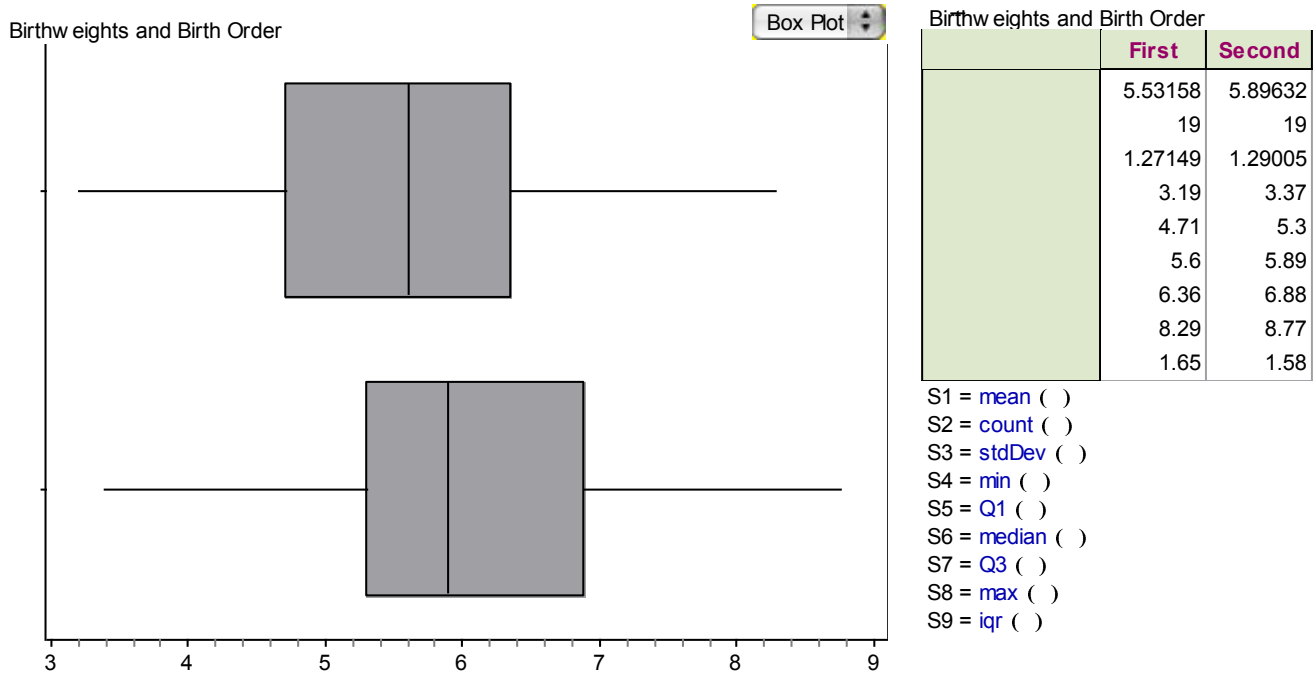
Directions: Show all of your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanations.

Birthweights and Birth Order.

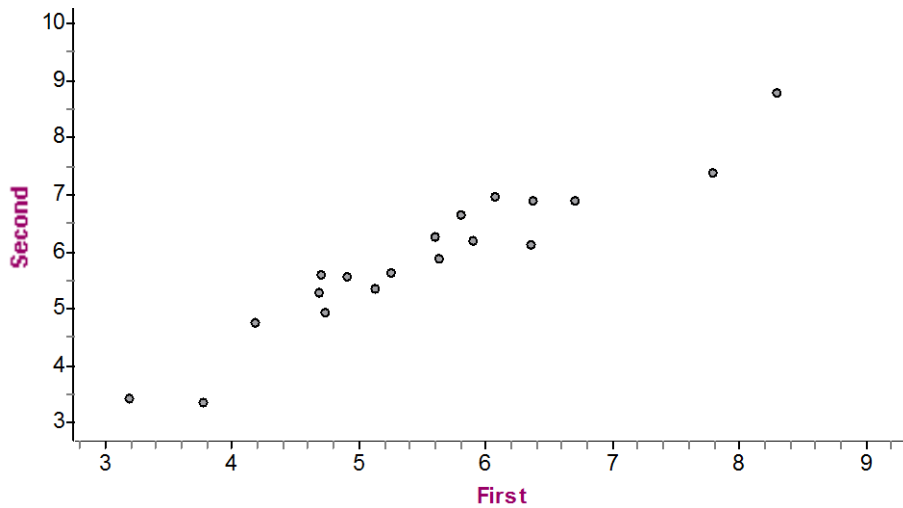
A physician who has delivered babies for over 20 years claims that the birthweights of babies tend to be higher than the birthweight of their next older sibling. In other words, a parent's second child tends to have a higher birthweight than the first child, the third child tends to have higher birthweight than the second child, and so on. To verify the doctor's statement, her office assistant gets a random sample of mothers with at least two children from the hospital records, and records the birthweights of the first two children. The birthweights for the children of 19 mothers are given in the table below.

Mother	Birthweight of First Child (in pounds)	Birthweight of Second Child (in pounds)
1	5.13	5.37
2	5.25	5.65
3	6.71	6.89
4	4.71	5.61
5	3.77	3.37
6	5.81	6.65
7	8.29	8.77
8	6.36	6.13
9	6.08	6.97
10	4.91	5.58
11	3.19	3.43
12	5.64	5.89
13	6.37	6.88
14	5.90	6.21
15	4.73	4.93
16	5.60	6.25
17	4.68	5.30
18	7.79	7.38
19	4.18	4.77

44a. Does this data provide any evidence to support the physician's theory? Justify your answer using the following statistical evidence.

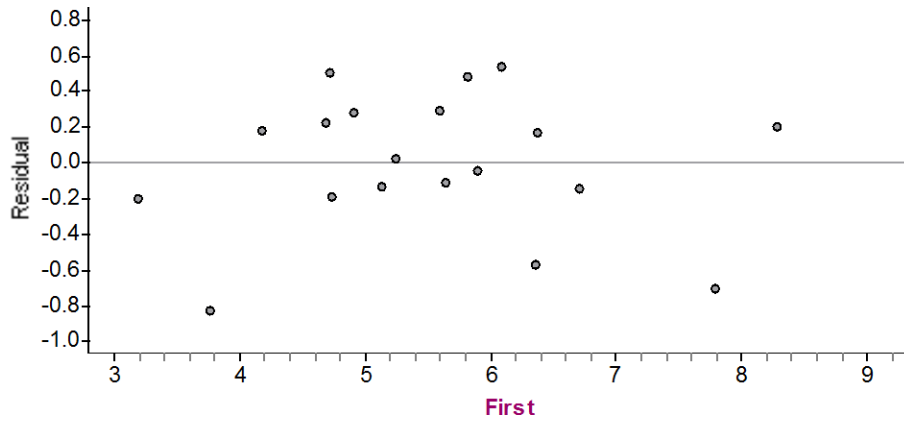


44b. The scatterplot of birthweights of first-born children and second-born children is given below.



Describe the relation between the birthweights of first-born children and the birthweights of second-born children.

44c. The regression analysis of the data resulted in the following outcome.



Is there a significant relation between the birthweights of the first-born children and the birthweights of the second-born children? Justify your answer using statistical evidence.

44d. Suppose the doctor is getting ready to deliver the second child of a mother whose first child weighed 7.2 pounds. Predict the birthweight of this second child.

44e. What is an appropriate range (to be at least 99.7% inclusive) of birthweights of second-born children, if the first-born child weighed 7.2 pounds?



AP Statistics
MidTerm Exam
STUDY GUIDE
2013-14

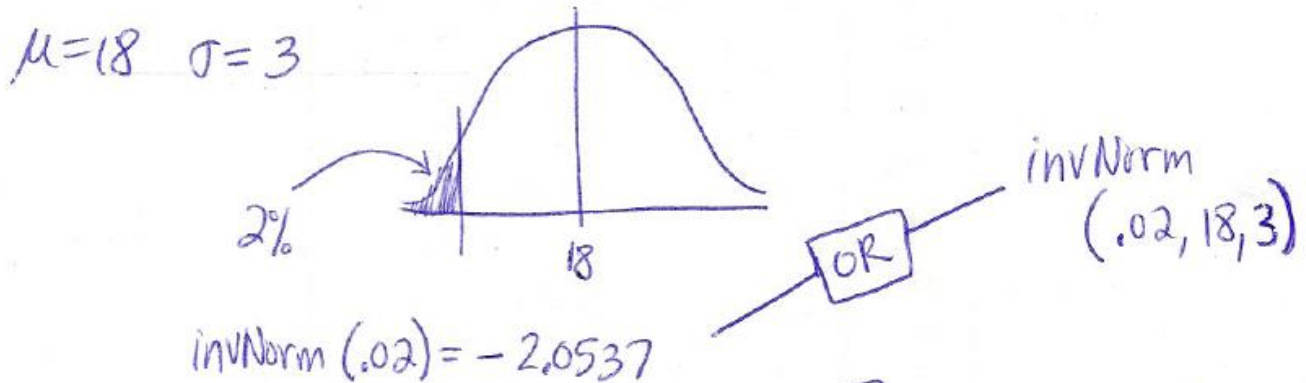
SOLUTIONS

1	E		21	A
2	A		22	C
3	D		23	C
4	C		24	B
5	B		25	D
6	A		26	B
7	C		27	C
8	E		28	D
9	D		29	B
10	E		30	D
11	B		31	D
12	E		32	B
13	D		33	A
14	E		34	D
15	A		35	A
16	C		36	E
17	B		37	D
18	D		38	B
19	E		39	B
20	B		40	D

SECTION II
Part A
SOLUTIONS

Zip Drive Lifetimes.

41. The lifetimes of zip drives marketed by ZipZap, Inc. are normally distributed, with a mean lifetime of 18 months and a standard deviation of three months. The company plans to offer a new warranty guaranteeing the replacement of failed zip drives during the warranty period. The replacement of failed drives is costly, however, and the company can afford to replace up to only two percent of its drives. Rounded to the nearest whole number, how many months of warranty should the company offer with these zip drives?



$$Z = \frac{X - \mu}{\sigma}$$

$$-2.0537 = \frac{X - 18}{3}$$

$$-6.1612 = X - 18$$

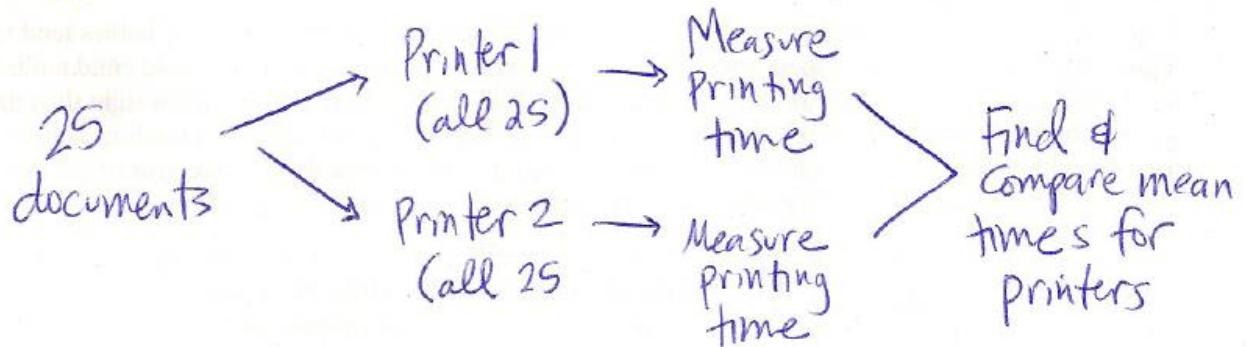
$$11.8 = X$$

The company should offer a 12-month warranty with these zip drives so as not to exceed replacing 2% of them.

Choosing a Printer.

42. A company executive is interested in purchasing printers for his division. He contacts a wholesaler, who shows him two printers comparable in quality and price made by two different manufacturers. Both printers are suitable for the kind of printing generally done in his division. To help him decide between these two printers, the wholesaler offers to loan him a printer of each kind for one day. Since the printers' printing time per page depends on the type of the document printed, the executive selects 25 different documents of varying size and complexity for printing. Help this executive design an experiment to compare the mean printing times of the two printers.

one approach:



no matter what approach you take (choose all 25 documents, or fewer than 25) the order the documents are selected to be printed must be random. Use a RNT to accomplish this.

After the experiment, the printer w/ the faster mean time will be chosen to be purchased.

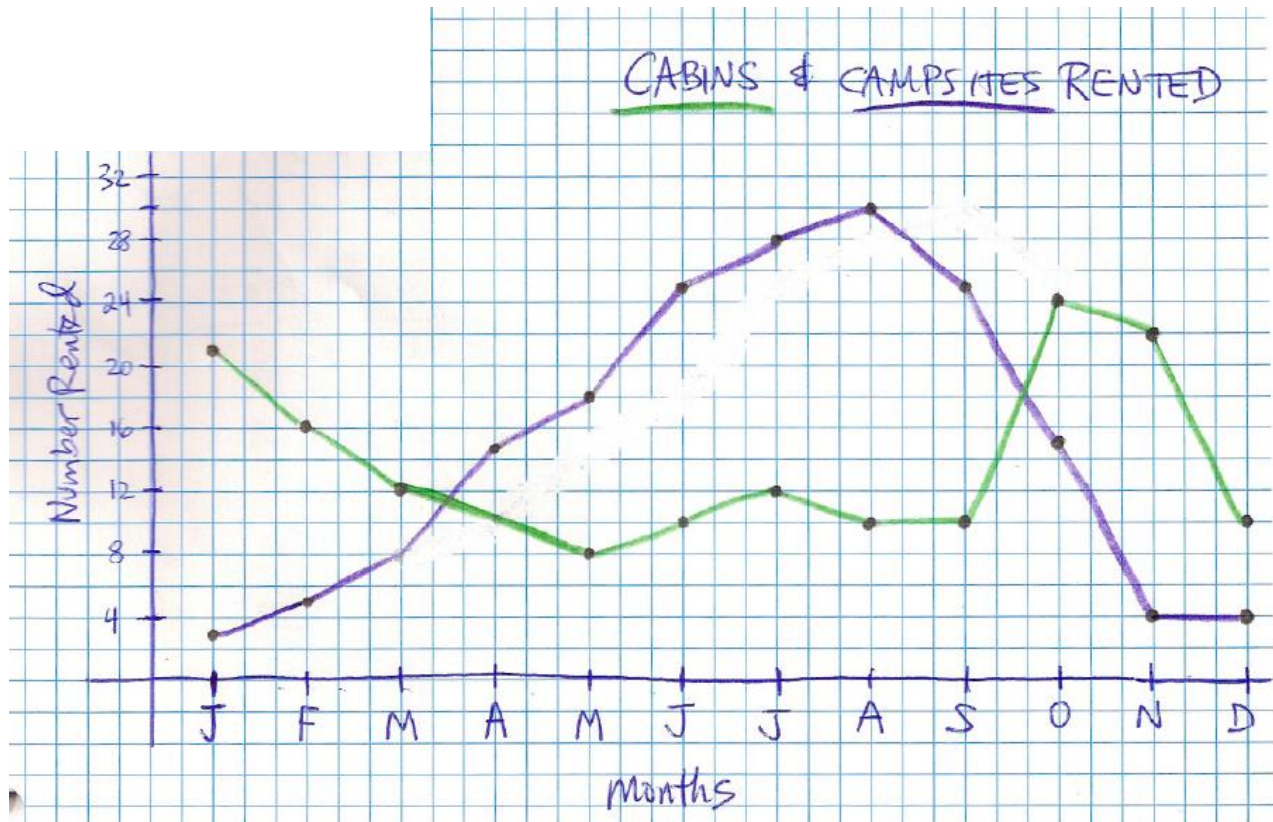
- blinding??
- control??

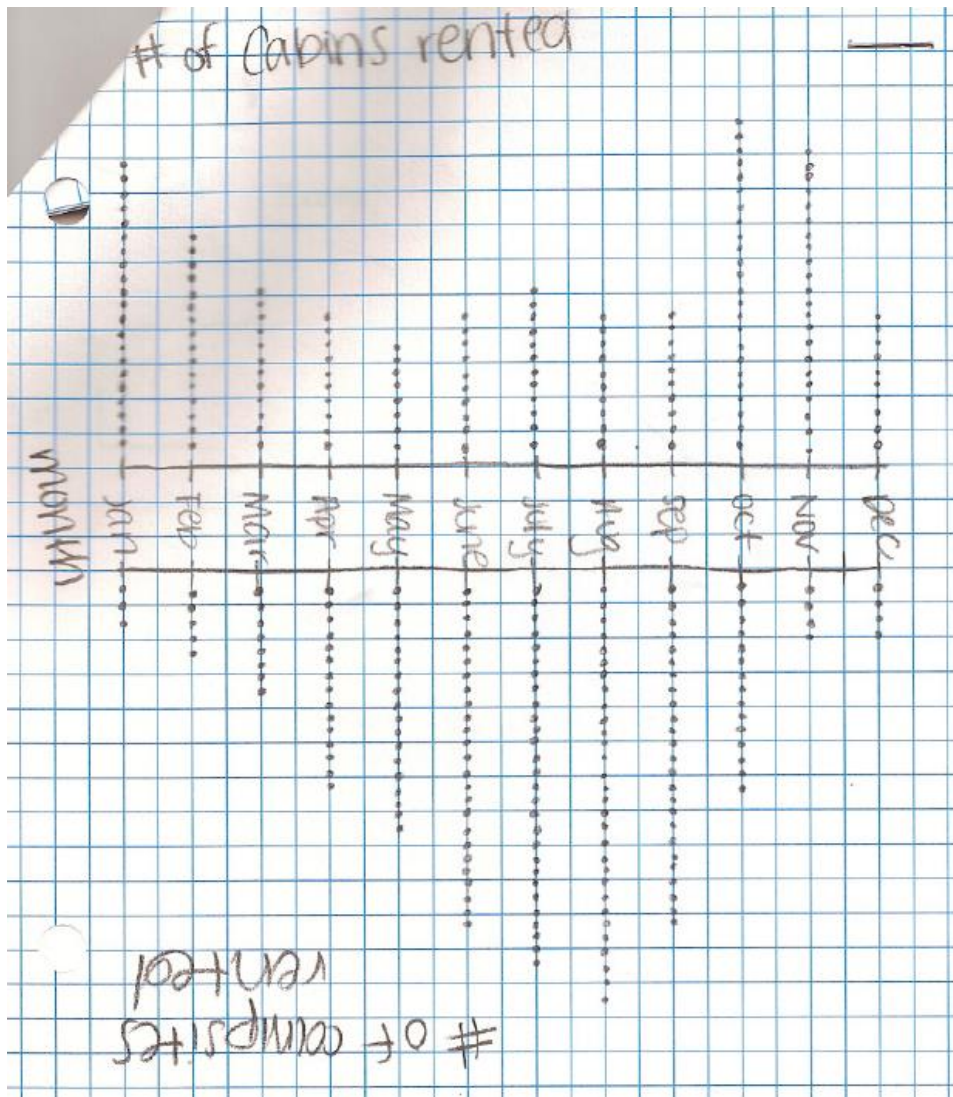
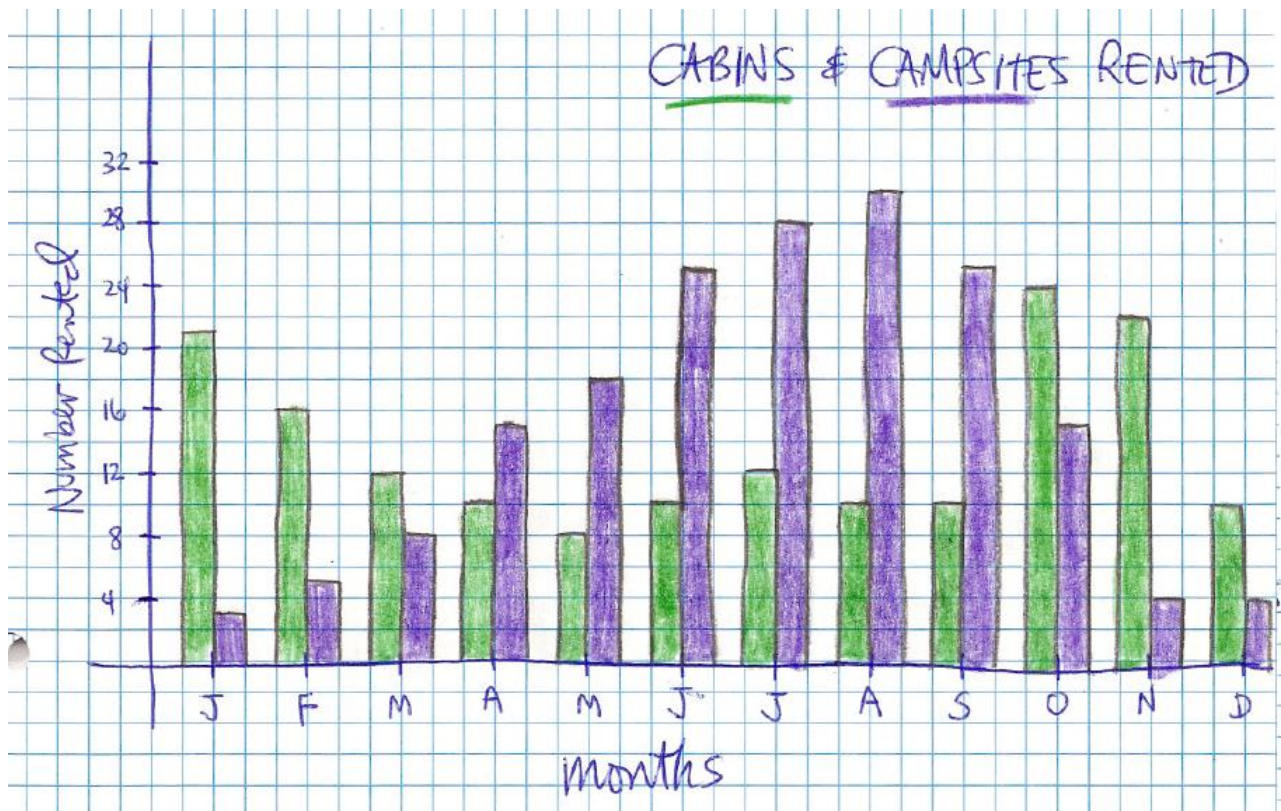
Cabins and Campsites.

A state park has campsites available for pitching tents. Visitors are allowed to pitch tents only at the designated campsites. In addition to campsites, the park also has some log cabins for rent. The demand for cabins and campsites changes from month to month. The renting records for year 2001 are as follows:

Month	Number of Cabins Rented	Number of Campsites Rented
Jan	21	3
Feb	16	5
Mar	12	8
Apr	10	15
May	8	18
Jun	10	25
Jul	12	28
Aug	10	30
Sep	10	25
Oct	24	15
Nov	22	4
Dec	10	4

43a. Using a sheet of graph paper, display the data graphically to compare the pattern of cabin rentals with that of campsite rentals.





43b. Examine your graphical display. Write a few sentences describing the rental history of cabins and campsites in 2001.

In general, more campsites are rented during the summer months, with rentals peaking in August. The cabin rentals are higher during the fall & winter months, except in December. Both cabin & campsite rentals are down in December.

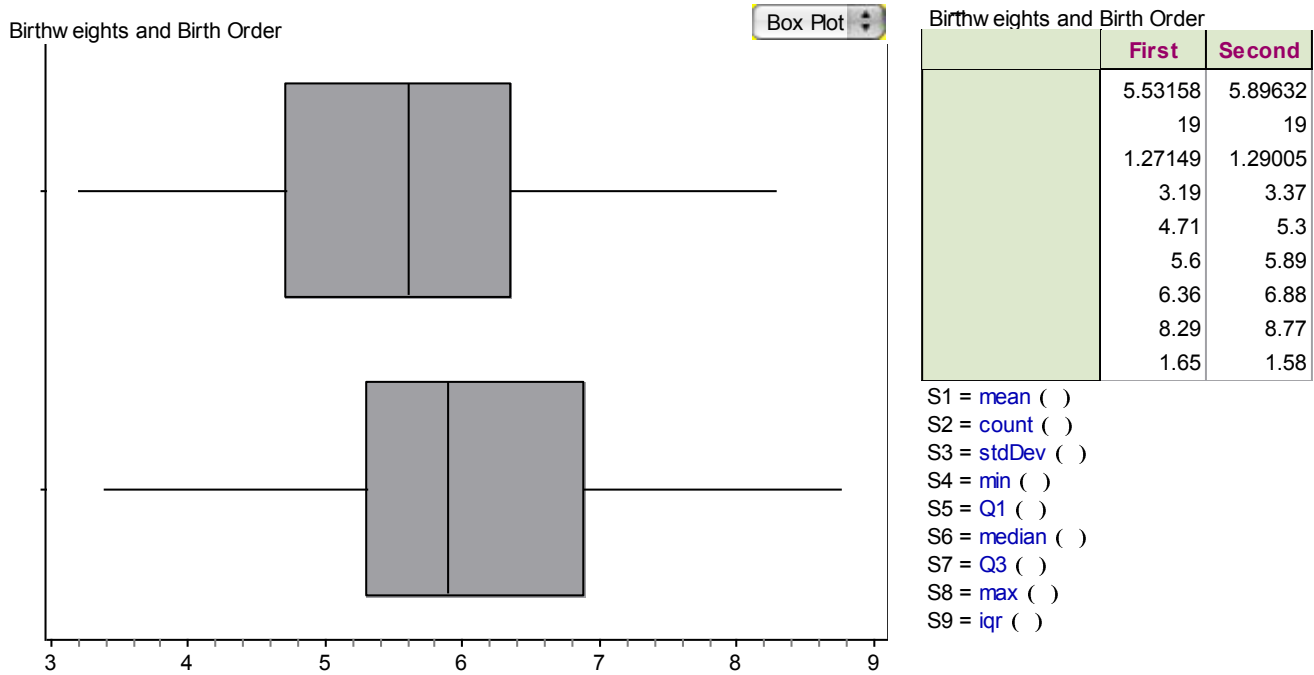
SECTION II
Part B
SOLUTIONS

Birthweights and Birth Order.

A physician who has delivered babies for over 20 years claims that the birthweights of babies tend to be higher than the birthweight of their next older sibling. In other words, a parent's second child tends to have a higher birthweight than the first child, the third child tends to have higher birthweight than the second child, and so on. To verify the doctor's statement, her office assistant gets a random sample of mothers with at least two children from the hospital records, and records the birthweights of the first two children. The birthweights for the children of 19 mothers are given in the table below.

Mother	Birthweight of First Child (in pounds)	Birthweight of Second Child (in pounds)
1	5.13	5.37
2	5.25	5.65
3	6.71	6.89
4	4.71	5.61
5	3.77	3.37
6	5.81	6.65
7	8.29	8.77
8	6.36	6.13
9	6.08	6.97
10	4.91	5.58
11	3.19	3.43
12	5.64	5.89
13	6.37	6.88
14	5.90	6.21
15	4.73	4.93
16	5.60	6.25
17	4.68	5.30
18	7.79	7.38
19	4.18	4.77

44a. Does this data provide any evidence to support the physician's theory? Justify your answer using the following statistical evidence.

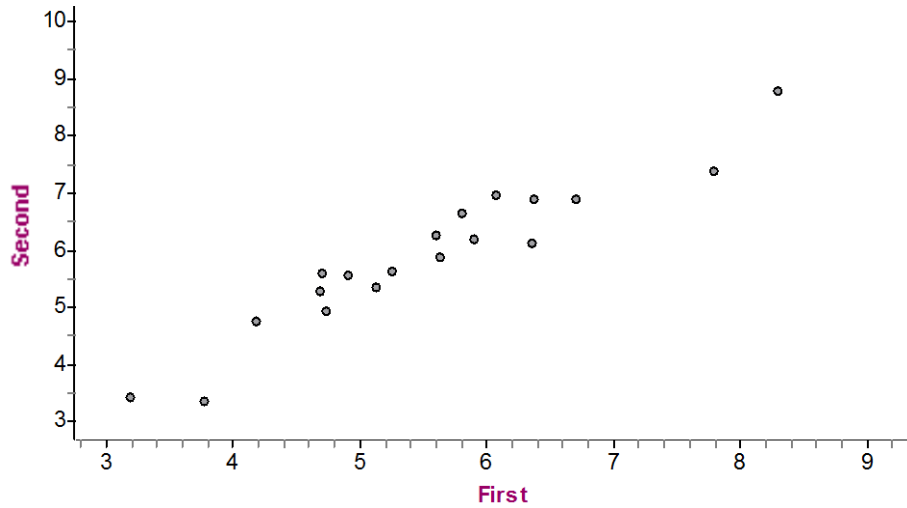


Each of the statistics in the Five-Number Summary is higher for the second child than the corresponding statistic for the first child.

The boxplots illustrate this.

This data provides significant evidence SUPPORTING the physician's statement.

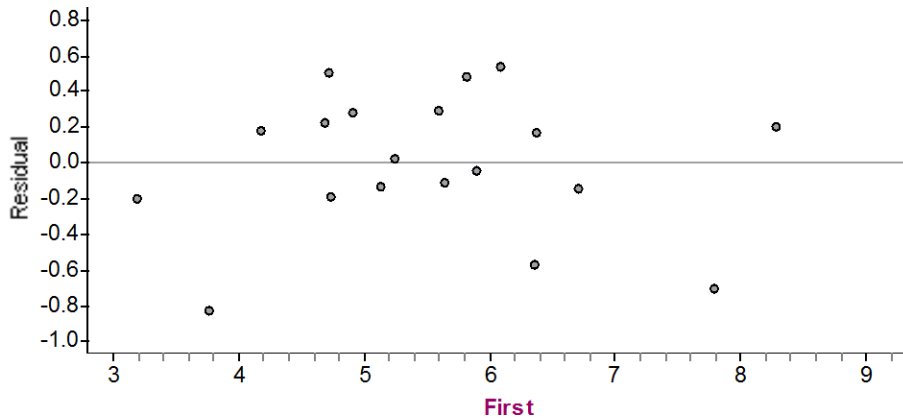
44b. The scatterplot of birthweights of first-born children and second-born children is given below.



Describe the relation between the birthweights of first-born children and the birthweights of second-born children.

There appears to be a strong positive linear relation between the birthweights of first & second children.

44c. The regression analysis of the data resulted in the following outcome.



$$\text{second} = 0.55 + 0.967\text{first}$$

$$R^2 = 0.91$$

$$SD = 0.3772$$

Is there a significant relation between the birthweights of the first-born children and the birthweights of the second-born children? Justify your answer using statistical evidence.

The residual plot shows little to no pattern, indicating that a linear model is appropriate for this data. We can continue to conclude that there is sufficient evidence to indicate a relation between the weights of the first & second children.

44d. Suppose the doctor is getting ready to deliver the second child of a mother whose first child weighed 7.2 pounds. Predict the birthweight of this second child.

$$\begin{aligned}\widehat{\text{Second}} &= 0.55 + 0.967 \text{ first} \\ &= 0.55 + 0.967(7.2)\end{aligned}$$

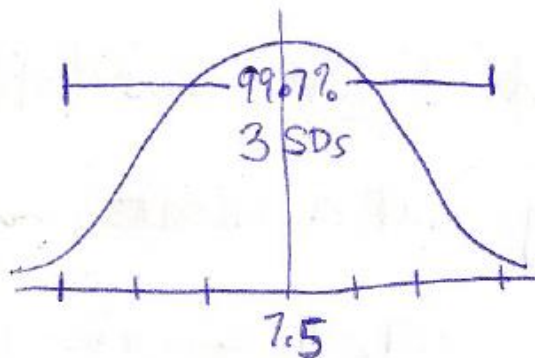
$$\widehat{\text{second}} = 7.5$$

The predicted birthweight of the second child is approximately 7.5 pounds.

44e. What is an appropriate range (to be at least 99.7% inclusive) of birthweights of second-born children, if the first-born child weighed 7.2 pounds?

$$\mu = 7.5$$

$$\sigma = 0.3772$$



$$7.5 - 3(0.3772) = 6.3684$$

$$7.5 + 3(0.3772) = 8.6316$$

To be 99.7% inclusive, the range of birthweights for the second-born child would be

6.37 to 8.63 pounds.