Do Not Write On This Worksheet

Sample Examination One

Section I

Time–1 hour and 30 minutes
Questions 1–40
Percent of total grade–50

Directions: The questions or incomplete statements that follow are each followed by five suggested answers or completions. Choose the response that best answers the question or completes the statement.

---

1. For the years 1950–1980, the number of heart disease deaths per 100,000 people in the United States were recorded. The regression line below was computed using a statistical software package. Which statement is the correct interpretation of the slope?

```
The regression equation is
Number of deaths=7387 - 3.63 year
```

(A) The number of heart disease deaths per 100,000 people has been dropping by an estimated 3.63 deaths per year on the average.
(B) There is an increase of approximately 7387 deaths per year.
(C) For every 3.63 years there is a decrease on the average of 1 death due to heart disease per 100,000 people.
(D) The regression line estimates that for every 3.63 years there is an average increase of 1 death due to heart disease per 100,000 people.
(E) Heart disease will be cured in the year 2036.

Answer

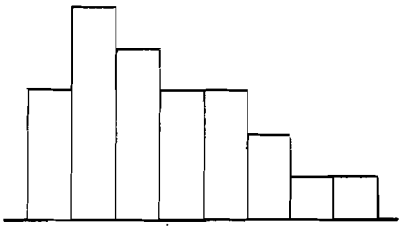
2. Which of the following statements about a linear regression model is true?

I. The sum of the residuals is always zero.
II. If $r^2 = 0$, the regression line is a horizontal line.
III. No pattern in the residual plot is an indication that a nonlinear model will show a better fit to the data than a straight-line regression model.

(A) I only
(B) I and II only
(C) I and III only
(D) II and III only
(E) I, II, and III

Answer

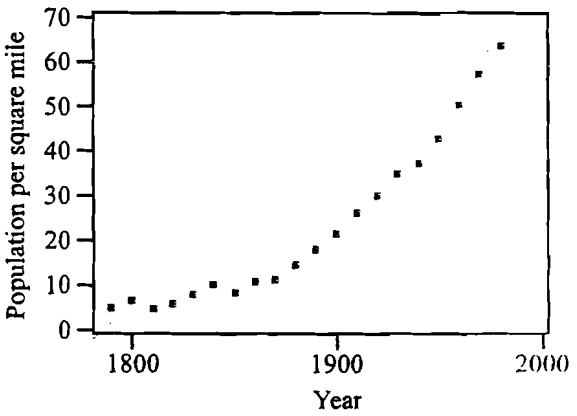
3. Which of the following is more likely to be true of this distribution?



(A) Mean = 3 Median = 3 Mode = 3
(B) Mean = 3.5 Median = 4 Mode = 3
(C) Mean = 4 Median = 3.5 Mode = 3
(D) Mean = 3.5 Median = 3.5 Mode = 5
(E) Mean = 3 Median = 2 Mode = 5

Answer

4. For the past two-hundred years population per square feet in a northwest suburb can be modeled using the exponential equation: $y = ab^x$ The scatter plot of the data is shown below. The R^2 value of the exponential model is 0.98.



Which of the following statements is true?

(A) If an attempt is made at fitting a straight line model to the original data, the corresponding residual plot would be approximately linear.
(B) If an attempt is made at fitting a straight line model to the original data, the corresponding residual plot would be scattered, and show no pattern.
(C) If an attempt is made at fitting a straight line model to the original data, the corresponding residual plot would be a straight line.
(D) Plotting the logarithm of the population per square mile against year should be approximately linear.
(E) Plotting the logarithm of the population per square mile against the logarithm of year should be approximately linear.

Answer

Use the following information for question 5 and 6

Those that study child development found a linear regression model for infants that uses age in months to predict height. A sample of 12 babies was randomly selected and the information shown below was generated.

S = 0.2560 R-Sq = 68.9% R-Sq(adj)= 69.5%

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Age	12	23.50	23.50	23.50	3.61	1.04
Height	12	79.850	79.800	79.860	2.302	0.665

5. The approximate slope of the least squares line is

(A) 0.44
(B) 0.53
(C) 1.08
(D) 1.30
(E) The slope cannot be determined from the given information

Answer

6. About what percent of the observed variation in the height can be explained by least-squares regression of height on age?

(A) 26%
(B) 29%
(C) 69%
(D) 83%
(E) 95%

Answer

7. In a statistics course, a linear regression equation was computed to predict the final exam score based on the score on the first test of the term. The equation was: $\hat{y} = 25 + 0.7x$ where y is the final exam score and x is the score on the first test. George scored 80 on the first test. On the final exam George scored 85. What is the value of his residual?

(A) -4
(B) 4
(C) 4.5
(D) 5
(E) 81

Answer

8. Suppose the regression line for a set of data, $\hat{y} = 3x + b$, passes through point (2, 5). If $\bar{x}$ and $\bar{y}$ are the sample means of the x and y values respectively, then $y =$

(A) $\bar{x}$
(B) $\bar{x} - 2$
(C) $\bar{x} + 5$
(D) $3\bar{x}$
(E) $3\bar{x} - 1$

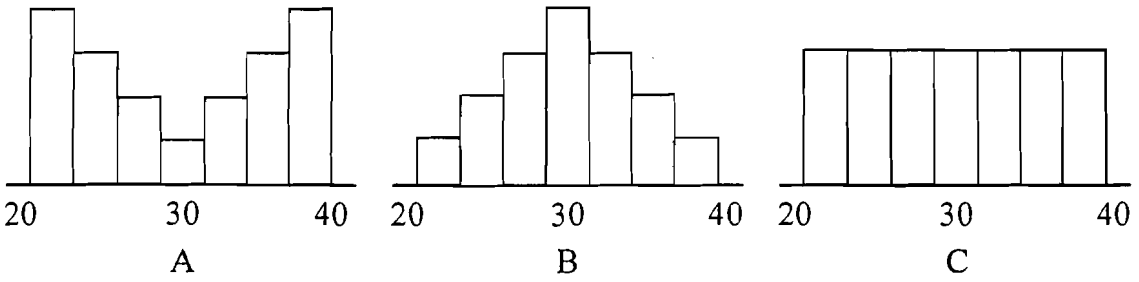
Answer

9. You have sampled 25 students to find the mean SAT scores at Morris Knolls High School. A 95% confidence interval for the mean SAT score is 900 to 1100. Which of the following statement gives a valid interpretation of this interval?

(A) 95% of the 25 students have a mean score between 900 and 1100.
(B) 95% of the population of all students at Morris Knolls have a score between 900 and 1,100.
(C) If this procedure were repeated many times, 95% of the resulting confidence intervals would contain the true mean SAT score at Morris Knolls.
(D) If this procedure were repeated many times, 95% of the sample means would be between 900 and 1100.
(E) If 100 samples were taken and a 95% confidence interval was computed, 5 of them would be in the interval from 900 to 1100.

Answer

Use the following graphs for questions 10–11. Assume the heights (frequencies) of each picture use the same scale.



10. Which distribution above has the smallest standard deviation?

(A) A
(B) B
(C) C
(D) It cannot be determined from the graphs
(E) All three distributions have the same standard deviation

Answer

11. In which distribution(s) would you be more likely to find the mean and median the same?

(A) A only
(B) B only
(C) C only
(D) A and B only
(E) A, B, and C

Answer

12. A 90% confidence interval for a population mean is determined to be 800 to 900. If the confidence is increased to 95% confidence while the sample statistics and sample size remain the same, the confidence interval for μ

(A) becomes narrower
(B) becomes 0.05
(C) does not change
(D) becomes wider
(E) becomes 0.025

Answer

13. A large sample hypothesis test with σ known of a null hypothesis $\mu = 15$ against the alternative hypothesis $\mu \neq 15$ results in the test statistic value of $z = 1.37$. Assuming σ is known, the corresponding p-value is approximately

(A) 0.0853
(B) 0.1707
(C) 0.4147
(D) 0.8293
(E) 0.9147

Answer

14. Multiplying each element of the data set by an integer n will have which affect on the mean and variance?

(A) There is no change in the mean and the variance is multiplied by n.
(B) The variance is divided by $\sqrt{n}$ and the mean is multiplied by n.
(C) The mean is multiplied by n and the variance is multiplied by n^2.
(D) The mean is multiplied by n and the variance is multiplied by $\sqrt{n}$.
(E) Since mean and variance are dependent on sample size, one cannot determine the affect of multiplying by n.

Answer

15. A candidate for mayor hires a statistician to determine the amount of support he has for the upcoming election. The statistician tests the null hypothesis that the population proportion equals 50% against the alternative hypothesis that the population proportion is greater than 50%. The results of a simple random sample of 91 registered voters can be found in the printout below.

```
Test of p = 0.5 vs p > 0.5
Success = for Candidate

Variable           X     N  Sample P   Exact P value
for Candidate     58    91  0.637363          0.0044
```

The conclusion which can be reached is:

I. The candidate can be quite confident that he has more than 50% of the vote.
II. The *p*-value of 0.0044 tells us that we cannot reject the null hypothesis and that the candidate has 50% or less of the vote.
III. The *p*-value of 0.0044 indicates that it is not very likely to get an observed value of 0.637 if the null hypothesis is true.

(A) I only
(B) II only
(C) III only
(D) I and II
(E) I and III

Answer

16. Which of the following are true?

I. The larger the sample, the smaller the spread in the sampling distribution.
II. Provided that the population size is significantly greater than the sample size, the spread of a sampling distribution is about the same no matter what the sample size.
III. Sampling distributions from non-normal populations are approximately normal provided *n* is large.

(A) II only
(B) III only
(C) I and II only
(D) I and III only
(E) I, II, and III

Answer

17. An inspection procedure at a manufacturing plant involves picking three items at random and then accepting the whole lot if at least two of the three items are in perfect condition. If in reality 80% of the whole lot is perfect, what is the probability that the lot will be accepted?

(A) 0.512
(B) 0.560
(C) 0.640
(D) 0.896
(E) 0.992

Answer

18. Given the data below, in conducting a test of association between gender and grade, what is the expected count for the number of males who earned a grade of B?

	A	B	C	D
Male	10	32	25	6
Female	5	41	14	12

(A) 32.3
(B) 35.5
(C) 36.8
(D) 41.0
(E) It cannot be determined

Answer

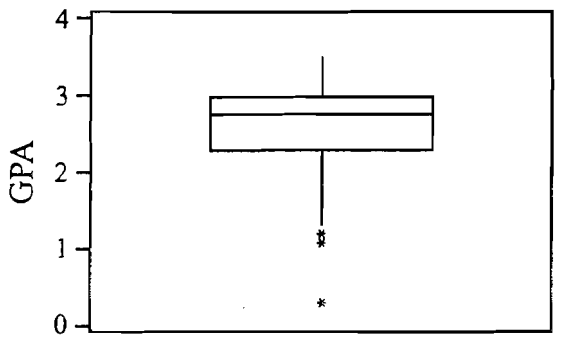
19. Which of the following is true?

I. A simple random sample is any sampling technique where each element of the population has the same chance of being selected.
II. A simple random sample is a sample where every set of n elements in the population has the same chance of being selected.
III. From a population of N=10 there are 90 equally likely possible samples of size 2 if we are sampling without replacement.

(A) I only
(B) II only
(C) III only
(D) I and II only
(E) I, II, and III

Answer

20. Which statement is true about the boxplot below?



I. It is a left skewed distribution which has outliers.
II. It is a symmetrical distribution which has outliers.
III. The interquartile range is less than 1.
IV. Approximately 75% of the observations have a GPA less than 3.

(A) I only
(B) II only
(C) II and III
(D) III and IV only
(E) I, III, and IV

Answer

21. The level of significance is always

(A) the maximum allowable probability of Type II error
(B) the maximum allowable probability of Type I error
(C) the same as the confidence
(D) the same as the p-value
(E) $1 - P$ (type II error)

Answer

22. A statistics student wishes to test the strength of various brands of paper towel. He chooses 5 brands and selects 6 towels from each brand. He numbers them 1–30. He randomly selects a towel and places it in an embroidery hoop. Exactly 10 ml of water and a large weight are placed in the center of the towel. The time it takes for the towel to break is recorded. In this case, the explanatory variable is the

(A) amount of time it takes for the towel to break
(B) 10ml of water and the large weight
(C) brand of paper towel
(D) large weight
(E) number of paper towels used in the experiment

Answer

23. A study was done to determine if the method of instruction (either lecture or discussion) depended on the type of class which was being taught. Twenty art classes, seventeen math classes and twenty-five science classes were observed. The method of instruction, discussion or lecture, was recorded. Which of the following best describes the relationship between method of teaching and type of class?

	discuss	lecture
arts	5	15
math	12	5
science	15	10

(A) There appears to be no relationship since the number of discussion class and the number of lecture classes was exactly the same.
(B) No association can be determined since the number of art, math and science classes were not exactly the same.
(C) There appears to be an association since the art class was less likely to use discussion then either math or science.
(D) There appears to be an association since the number of math and science classes is greater than the number of arts classes.
(E) A measure of association cannot be determined from these data.

Answer

24. A researcher interested in the age at which women are having their first child surveyed a simple random sample of 250 women having at least one child and found a approximately normal distribution with a mean age of 27.3 and a standard deviation of 5.4. According to the empirical rule, also known as the 68-95-99.7 rule, approximately 95% of the women had their first child between the ages of

(A) 11.1 years and 43.5 years
(B) 16.5 years and 38.1 years
(C) 21.9 years and 32.7 years
(D) 21.9 years and 38.1 years
(E) 25.0 years and 29.6 years

Answer

25. You want to estimate the mean SAT score for a population of students with a 90% confidence interval. Assume that the population standard deviation is $\sigma = 100$. If you want the margin of error to be approximately 10, which of the following would be the desired minimal sample size?

(A) 16
(B) 38
(C) 100
(D) 271
(E) 1476

Answer

26. At a certain high school, all students who take AP Psychology also take AP Statistics. From past records, the probability that a student gets a 5 in AP Psychology is one-fourth. The probability a student gets a 5 on the AP Statistics exam is one-seventh. What is the probability that a student will get at least one 5 when taking both exams

I. if the two events are independent?
II. if the students getting a 5 on the AP Statistics exam are a subset of those students getting a 5 on the AP Psychology exam?

(A) $\frac{11}{28}, \frac{1}{4}$
(B) $\frac{11}{28}, \frac{1}{7}$
(C) $\frac{9}{28}, \frac{1}{4}$
(D) $\frac{9}{28}, \frac{1}{7}$
(E) $\frac{5}{14}, \frac{1}{4}$

Answer

27. The scores of a standardized test designed to measure math anxiety are normally distributed with a mean of 100 and a standard deviation of 10 for a population of first year college students. Which of the following observations would you suspect is an outlier?

(A) 90
(B) 100
(C) 150
(D) 90, 100, and 150 are all outliers
(E) None of 90, 100, and 150 are outliers

Answer

28. Early studies of probability were conducted by the Italian mathematician Girolamo Cardano (1501–1576). One of the many dice games that Cardano studied was played with six 6-sided dice. Each of these six dice had five blank faces and one face with a number. The numbers 1 through 6 each appeared on one of the six dice. All 6 dice were rolled at once, and the payoff to the gambler was based on the sum of the numbers showing on the up faces. What is the expected value of the sum obtained by rolling all 6 dice?

(A) 3
(B) 3.5
(C) 6
(D) 36
(E) It cannot be determined from the given information

Answer

29. Which of the following is true?

I. If the sample size is constant, then reducing the probability of type I error will reduce the probability of type II error.
II. Increased power can be achieved by reducing the type II error.
III. If the p-value of a test is 0.015, the probability that the null hypothesis is true is 0.015.

(A) I only
(B) II only
(C) I and III only
(D) II and III only
(E) I, II, and III

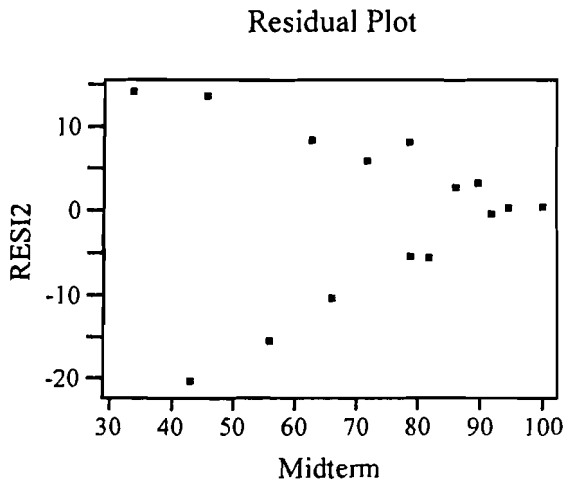
Answer

30. When two fair dice are rolled, what is the probability of getting a sum of 7 given that the first die rolled is an odd number?

(A) $\frac{1}{12}$
(B) $\frac{1}{9}$
(C) $\frac{1}{6}$
(D) $\frac{1}{4}$
(E) $\frac{1}{2}$

Answer

31. The residual plot below came from data which plotted grade at midterm against grade on final exam. A linear regression line was calculated. Which conclusion could be reached by analyzing the residual plot?



(A) An exponential curve could be used to predict final grade given midterm grade.
(B) Students did better on the final exam then they did on the midterm.
(C) There is evidence that a linear model is appropriate.
(D) There exists unequal variance throughout the model.
(E) There is no pattern evident in the residual plot.

Answer

32. The heights of American men aged 18 to 24 are approximately normal with a mean of 68 inches and a standard deviation of 2.5 inches. About 20% of these men are taller than

(A) 66 inches
(B) 68 inches
(C) 70 inches
(D) 72 inches
(E) 74 inches

Answer

33. Which of the following will most likely approximate a uniform distribution?

(A) heights of students at a particular high school
(B) weights of students at a particular high school
(C) SAT scores of seniors at a particular high school
(D) IQ scores of students at a particular high school
(E) ages of students at a particular high school

Answer

34. Polly takes three standardized tests. She scores 600 on all three. Using standard scores, or z-scores, rank her performance on the three tests from best to worst if the means and standard deviations for the tests are as follows:

	Mean	Standard Deviation
Test I	500	80
Test II	470	120
Test III	560	30

(A) I, II, and III
(B) III, II, and I
(C) I, III, and II
(D) III, I, and II
(E) II, I, and III

Answer

35. At a certain high school a simple random sample was taken asking fifty-two 11th and 12th graders their political affiliation. The following two-way table was established. If a χ^2 test of independence were performed on these data, what would be the corresponding degrees of freedom?

	11th Grade	12th Grade
Republican	11	5
Democrat	10	15
Independent	5	6

(A) 1
(B) 2
(C) 3
(D) 6
(E) 25

Answer

36. A student is interested in the effects of different walking styles on heart rate. He decides to use 30 volunteers from his school for his experiment. All 30 participants find their at-rest pulse rates. Each participant will walk twice for 10 minutes, once using a fast pace but with no arm movement and again using a fast pace, but with an exaggerated arm movement style. The experimenter throws a coin to determine which style each participant will walk first. All participants get sufficient rest between walks to let their pulse rates return to normal. The student then compares increased pulse rate based on the walk with no arm movement to increased pulse rate based on the walk with exaggerated arm movement for each student.

Which of the following statements is true?

(A) This is an observational study, and not an experiment, therefore no conclusion can be reached regarding walking style and pulse rate.
(B) Observations in this study are independent.
(C) Blocking is used in this study to reduce difference in increased pulse rates among individual students.
(D) Because subjects were not assigned randomly to a control or an experimental group, the design of the experiment was flawed.
(E) This is an example of a completely randomized comparative experiment without blocking.

Answer

37. A cup of coffee at a Mini Mart is usually 50 cents. The manager of the store decides to let the customer roll a die. If the die is a 6, the customer gets the cup of coffee for free. If the die is a 1, 2, 3, 4 or 5 the coffee is $1. A customer buys a cup of coffee at the mart once a day, five days a week. A statistics class wishes to know the number of cups of coffee which were free to the customer using a simulation. There were 25 students in the class. Each student did 4 simulations. Which frequency table below is most likely to contain the results from these 100 trials?

(A)

Number of free cups of coffee	Frequency
0	14
1	18
2	17
3	21
4	14
5	16

(B)

Number of free cups of coffee	Frequency
0	0
1	3
2	5
3	14
4	41
5	37

(C)

Number of free cups of coffee	Frequency
0	0
1	0
2	3
3	17
4	39
5	41

(D)

Number of free cups of coffee	Frequency
0	39
1	42
2	14
3	4
4	1
5	0

(E)

Number of free cups of coffee	Frequency
0	3
1	17
2	36
3	30
4	13
5	1

Answer

38. For the owner of the Mini Mart in question 37, assume it costs 35 cents to make one cup of coffee. Using the process of charging customers \$1.00 if they get a 1, 2, 3, 4 or 5 on a die and no charge if the die is a 6, what is the approximate expected profit per cup to the owner?

(A) 0.33
(B) 0.48
(C) 0.50
(D) 0.65
(E) 0.83

Answer

39. An airline claims that its planes are, on average, less than 8 minutes late in landing. The appropriate hypotheses for examining the claim are

(A) $H_o: \mu < 8$
$H_a: \mu \geq 8$

(B) $H_o: \mu = 8$
$H_a: \mu \neq 8$

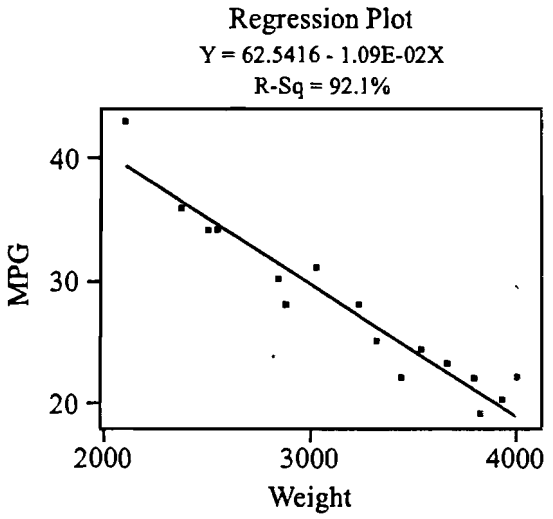
(C) $H_o: \mu = 8$
$H_a: \mu < 8$

(D) $H_o: \mu = 8$
$H_a: \mu > 8$

(E) $H_o: \mu = 7$
$H_a: \mu < 7$

Answer

40. A student was interested in the relationship between weight of a car and gas consumption measured in mpg. He selected sixteen different automobiles and recorded their weights along with their advertised mpg. The regression equation and regression plot are shown below.



What affect would the addition of the point (4,300 lbs., 15.63mpg) have on the value of r^2?

(A) It will have no effect on r^2 because it lies on the line.
(B) It will have no effect on r^2 because its value is the same as $(\overline{x}, \overline{y})$.
(C) It will decrease r^2 because it is an outlier.
(D) It will increase r^2 because it is an influential point which lies on the least squares line.
(E) It will increase r^2 because every additional point will increase the percent variation in y due to the relationship with the least squares line.

Answer

Section II
Part A

Time–65 minutes
Questions 1–5
Percent of Section II grade–75

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

1. Johannes Kepler spent years trying to discover a relationship between the time (in days) for a planet to revolve around the sun and the distance (in millions of kilometers) of the planet from the sun. A computer printout of a regression analysis is shown below for all nine planets in our universe.

Predictor	Coef	StDev	T	P
Constant	-5826	3303	-1.76	0.121
Distance	15.027	1.218	12.34	0.000

S = 7471 R-Sq = 95.6% R-Sq(adj) = 95.0%

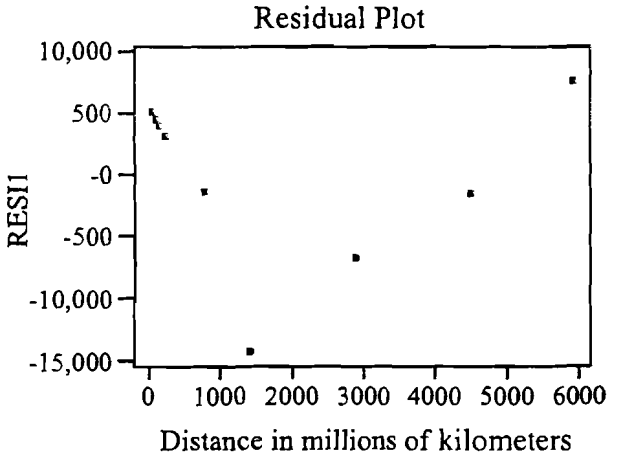
(a) What is the least squares line which could be used to predict the time it takes for a planet to revolve around the sun given the distance a planet is from the sun?

(b) Interpret the slope in the context of the problem.

(c) For a distance of 15000 million kilometers, what is the predicted time for one revolution around the sun?

(d) Explain clearly what the p-value in the "DISTANCE" row of the computer printout means.

(e) The residual plot is given below. Does the plot indicate that a line is an appropriate model to use for these data? Explain why or why not.



2. A study was conducted to investigate effects of regular exercise on weight loss of adults. Nine subjects selected randomly from a population were weighed before a program of regular exercise and then again three weeks after. Data are listed below as weights of subjects in kilograms before and after the exercise program. The claim is that the weights after the program are less than the weights before the program

Before Exercise Program	After Exercise Program
99	97
57	55
62	58
69	66
74	75
77	76
59	56
92	88
70	65

Do these data support the claim that the exercise program is effective in reducing weight? Use statistical evidence to justify your response. State clearly your assumptions.

3. A student wishes to study the effects of sugar on pulse rate. He knows that males and females pulse rates react differently to the ingestion of sugar. There are 36 volunteers, 18 males and 18 females who will participate in the study.

   (a) Design an experiment which would study the effects of sugar on pulse rate. Carefully describe all steps and procedures.

   (b) What are the explanatory and response variables?

   (c) What is done to control differences between males and female?

   (d) What is your control group?

   (e) Draw a diagram to illustrate your design.

4. A real estate agent in Texas wished to determine if there was a relationship between where a house was located and the number of bedrooms the house has. The agent selected the past 518 home sales during the months of June through September. The data collected is shown below:

	CITIES		
	Dallas	**Fort Worth**	**Other**
2 bedrooms	40	31	11
3 bedrooms	157	91	61
4 bedrooms	46	44	37

(a) Calculate the relative percentages of 2, 3, and 4 bedroom homes for each of the 3 cities.

(b) Do these data provide convincing evidence of an association between the location of a house and number of bedrooms a house has? Give appropriate statistical evidence to support your conclusion.

5. An experimenter wanted to determine if weights differed between smokers and non-smokers. She randomly selected 92 people, all from the same age group and records if they smoke or they do not smoke. She then records their current weight. Shown below is a computer analysis of the variable weight for smokers and non-smokers.

Variable	Smokes	N	Mean	Median	TrMean	StDev
Weight	No	64	142.03	141.00	141.12	23.60
	Yes	28	152.29	151.50	152.54	22.89

Variable	Smokes	SE Mean	Minimum	Maximum	Q1	Q3
Weight	No	2.95	95.00	215.00	-123.50	155.00
	Yes	4.33	108.00	190.00	132.00	170.00

(a) Is this an experiment? Explain why or why not. How does this affect the conclusions we can draw from the study?

(b) Are there any outliers in either the smokers or nonsmokers group? Show the work to verify your answer.

Section II
Part B

Time–25 minutes
Question 6
Percent of Section II grade–25

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

6. Two swimmers have been training for a big race in the 100 meter backstroke. Their times are known to be normally distributed with the following distributions (in seconds). We can also assume that the times of the two swimmers are independent.

swimmer 1: mean = 76.76	standard deviation = 3.08
swimmer 2: mean = 77.9	standard deviation = 2.49

(a) The swimmers keep track of their progress over the season. What is the mean and standard deviation of the first swimmer's time (x) minus the second swimmer's (y) time?

$\mu_{x-y}=$
$\sigma_{x-y}=$

(b) Using your results in part (a), what is the probability the second swimmer will beat the first swimmer on any given race? Round your answer to the nearest hundredth. Please note that the lowest time wins a race.

(c) Using your results in part (b), what is the probability that the second swimmer will beat the first swimmer in 6 or more of 10 independent races?

(d) If a sample of 10 races were taken from each swimmer, what is the probability that the average time of the ten races for the first swimmer would be better than the average time for the ten races of the second swimmer?