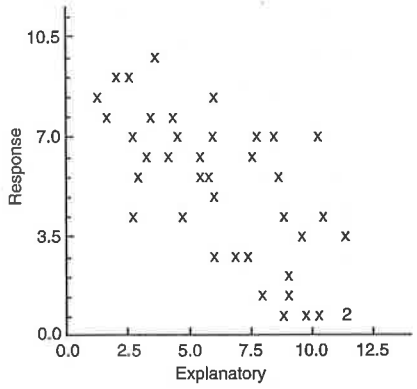
AP Statistics Practice Examination 2

Multiple Choice Statistics Section I

Time: 90 minutes

1. For the given scatterplot, what is the correct regression output?



(A)

Predictor	Coef	SE Coef	T	P
Constant	9.1104	0.7745	11.76	0.000
Explanatory	0.6181	0.1085	−5.70	0.000
S = 2.02551	R-Sq = 46.1%	R-Sq(adj) = 44.7%		

(B)

Predictor	Coef	SE Coef	T	P
Constant	9.1104	0.7745	11.76	0.000
Explanatory	−0.6181	0.1085	−5.70	0.000
S = 2.02551	R-Sq = 67.9%	R-Sq(adj) = 66.3%		

(C)

Predictor	Coef	SE Coef	T	P
Constant	9.1104	0.7745	11.76	0.000
Explanatory	0.6181	0.1085	−5.70	0.000
S = 2.02551	R-Sq = 67.9%	R-Sq(adj) = 66.3%		

(D)

Predictor	Coef	SE Coef	T	P
Constant	9.1104	0.7745	11.76	0.000
Explanatory	−0.6181	0.1085	−5.70	0.000
S = 2.02551	R-Sq = 46.1%	R-Sq(adj) = 44.7%		

(E)

Predictor	Coef	SE Coef	T	P
Constant	−0.6181	0.7745	11.76	0.000
Explanatory	9.1104	0.1085	−5.70	0.000
S = 2.02551	R-Sq = 46.1%	R-Sq(adj) = 44.7%		

2. A magazine claims that 25.1% of all women enjoy gardening. A researcher believes the percentage is higher and performs a test of H_0: $p = 0.251$ versus H_a: $p > 0.251$. A random sample of 100 women yields significant results at the $\alpha = 0.05$ level. Which of the following statements about the confidence interval used to estimate the true population proportion of women who enjoy gardening must be true?
   (A) A 90% confidence interval contains the proportion 0.251.
   (B) A 90% confidence interval does not contain the proportion 0.251, because the value of 0.251 is above the upper limit of the interval.
   (C) A 90% confidence interval does not contain the proportion 0.251, because the value of 0.251 is below the lower limit of the interval.
   (D) A 95% confidence interval does not contain the proportion 0.251, because the value of 0.251 is above the upper limit of the interval.
   (E) A 95% confidence interval does not contain the proportion 0.251, because the value of 0.251 is below the lower limit of the interval.

3. Executives from a music label believe they are losing money from the number of CDs that individuals burn illegally. They would like to determine the number of CDs the average computer owner burns in a month. What size sample would they need to take to estimate this number with a margin of error of 0.5 with 95% confidence, if they assume the standard deviation is 3?
   (A) 10
   (B) 17
   (C) 24
   (D) 98
   (E) 139

4. As part of a statistics project, a student decides to find a 95% confidence interval for the difference in average ages of students and faculty. The student, through diligent research, is able to record the age of all faculty members and all students and then calculates the 95% confidence interval using the t-distribution. Which of the following is a consideration the student failed to take into account?

   (A) The group of teachers and students are not independent. Therefore, the assumptions for using the two-sample t-interval are not valid.

   (B) The ages of teachers and students are not likely to be normally distributed. Therefore, the assumptions for using the two-sample t-interval are not valid.

   (C) The distribution of student ages is likely to have a few large outliers. Therefore, the assumptions for using the two-sample t-interval are not valid.

   (D) In both cases, the student had data for the entire population. Therefore, the actual difference in average ages can be computed, and a confidence interval is not valid.

   (E) Because there are most likely 40 or more students and 40 or more faculty members, there is nothing the student failed to take into account, and the confidence interval is valid.

5. Randomly selected individuals were asked about their physical activity. Of 75 randomly selected men, 30 had walked for exercise in the preceding two weeks. Of 75 randomly selected women, 36 had walked for exercise in the preceding two weeks. Assume independence between the samples. Is there evidence to show a significant difference in the proportion of men and the proportion of women who walk for exercise?

   (A) Because the proportions are different, there is evidence to show a significant difference in the proportions of men and women who walk for exercise.

   (B) With $p = 0.162$, there is insufficient evidence to show a significant difference in the proportions of men and women who walk for exercise.

   (C) With $p = 0.324$, there is insufficient evidence to show a significant difference in the proportions of men and women who walk for exercise.

   (D) With $p = 0.838$, there is insufficient evidence to show a significant difference in the proportions of men and women who walk for exercise.

   (E) The conditions necessary to perform a significance test have not been met; therefore, a conclusion cannot be drawn.

6. Which of the following is a true statement about experimental design?
   (A) Replication is a key component in experimental design. Thus, an experiment needs to be conducted on repeated *samples* before generalizing results.
   (B) Control is a key component in experimental design. Thus, a control group that receives a placebo is a *requirement* for experimentation.
   (C) Randomization is a key component in experimental design. Randomization is used to *reduce* bias.
   (D) Blocking eliminates the effects of *all* lurking variables.
   (E) The placebo effect is a concern for *all* experiments.

7. An experimenter believes that two new exercise programs are more effective than any current exercise routines and wishes to compare the effectiveness of these two new exercise programs on physical fitness. The experimenter is trying to determine whether or not a control group, which follows neither of these new programs but continues with current exercise routines, would be beneficial. Which of the following can be said about the addition of a control group?
   (A) A control group would eliminate the placebo effect.
   (B) A control group would eliminate the need for blinding in the study.
   (C) A control group would allow the experimenter to determine which of the two exercise programs improves physical fitness the most.
   (D) A control group would allow the experimenter to determine if either of the exercise programs is more effective than current programs for physical fitness.
   (E) There would be no added benefit to having a control group.

8. A large fast-food chain is changing vendors for its children's meal toys. The vendor claims that equal quantities of the four types of toys have been manufactured and will be distributed randomly among the restaurants. One restaurant received 89 of one type of toy, 95 of a second type, 106 of a third type, and 110 of the fourth type of toy in a shipment of 400 toys. If we consider this shipment to be a random sample of toys, does this shipment provide sufficient evidence to contradict the vendor's claim?
   (A) Yes, since the store did not receive 100 of each type of toy.
   (B) Yes, since a test of significance yields significant results at the 0.10 but not at the 0.05 level.
   (C) Yes, since a test of significance yields significant results at the 0.05 but not at the 0.01 level.
   (D) Yes, since a test of significance yields significant results at the 0.01 but not at the 0.001 level.
   (E) No, since a test of significance yielded results that were not significant at even the 0.10 level.

9. Concert attendance for a stadium is normally distributed with a standard deviation of 7641. If a concert with 41,293 people in attendance is in the top 2% of all concert attendance records, what is the mean concert attendance?
   (A) 25,601
   (B) 27,917
   (C) 54,670
   (D) 56,986
   (E) 77,163

10. A congressman mails a questionnaire to his constituents asking if the United States should use military force to overthrow violent dictators in controversial areas of the world. Of the 500 people who respond, 35% believe the United States should use military force in this situation. On a talk show, the politician claims that only 35% of his constituents (with a 4% margin of error) believe in using military force. Which assumption for constructing a confidence interval is violated?
   (A) The population is ten times as large as the sample.
   (B) The data constitute a simple random sample from the population of interest.
   (C) The count of successes, $n\hat{p}$, is 10 or more.
   (D) The count of failures, $n(1 - \hat{p})$, is 10 or more.
   (E) There are no violations for constructing a confidence interval.

11. Owners of a day-care chain wish to determine the proportion of families in need of day care for the town of Bockville. Bockville is estimated to have 1000 families. The owners of the day-care chain randomly sample 50 families and find that 60% of them have a need for day-care services. Which of the following is a condition necessary for constructing a confidence interval for a **proportion** that has *not* been met?
   (A) The data constitute a representative random sample from the population of interest.
   (B) The sample size is less than 10% of the population size.
   (C) The counts of those who need day care and those who don't need day care are 10 or more.
   (D) The distribution of sample values is approximately normally distributed.
   (E) All conditions necessary for constructing a confidence interval for the proportion seem to be met.

12. Which of the following sample designs does *not* contain a source of bias?

(A) A politician would like to know how her constituents feel about a particular issue. As a result, her office mails questionnaires about the issue to a random sample of adults in her political district.

(B) A company uses the telephone directory to randomly select adults for a telephone survey to gauge their feelings toward items manufactured by the company.

(C) An interviewer selects a random sample of individuals to question about a particular issue. Since some of the individuals are not informed about the issue, the interviewer gives background and his personal view on the issue before recording their responses.

(D) A news show asks viewers to call a toll-free number to express their opinions about a recent high-publicity trial.

(E) One thousand numbered tickets are sold as a fund-raiser. Five numbers are chosen randomly, and the individuals with the winning ticket numbers each win $10.

13. The cause of death and the age of the deceased are recorded for 440 patients from a hospital.

	15–24	25–34	35–44	45–54	55–64	Total
Accident	14	12	15	12	7	60
Homicide	5	4	3	0	0	12
Heart disease	1	3	14	34	63	115
HIV	0	3	6	4	0	13
Cancer	2	4	17	47	89	159
Other	3	7	16	26	43	95
Total	25	33	71	123	202	440

Use these values to estimate the probability that a person at this hospital died as a result of an accident if it is known the person was between the ages of 45 and 54.

(A) 0.0273

(B) 0.0976

(C) 0.1364

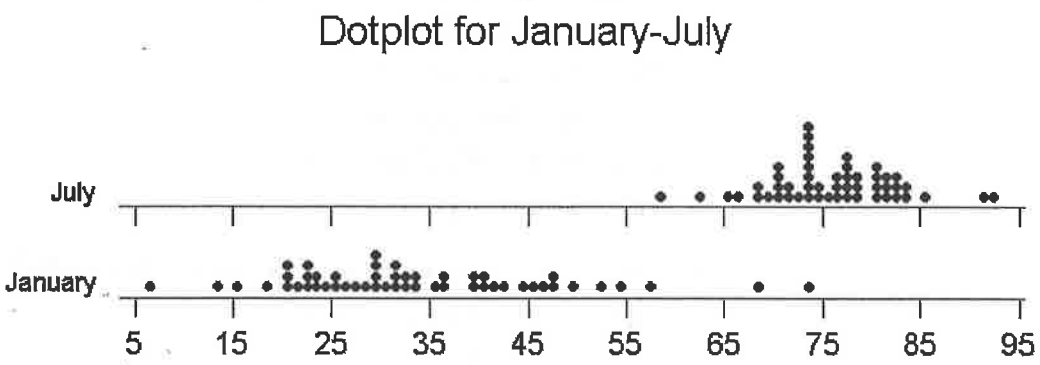
(D) 0.2000

(E) 0.4878

14. Two different Internet sites claim to offer the Web's lowest hotel rates for major U. S. cities. To test the claim, a consumer group randomly selects 50 hotels, and checks the rate charged for these hotels on both sites. To determine if there is a significant difference in rates between these two sites, which significance test is appropriate?
    (A) Two-sample z-test
    (B) Two-sample t-test
    (C) Matched-pairs test
    (D) χ^2 test of independence
    (E) Linear regression t-test

15. Which of the following does *not* represent a probability density function?
    (A) $f(x) = 1$ for $0 \leq x \leq 1$
    (B) $f(x) = \frac{1}{3}x$ for $0 \leq x \leq \sqrt{6}$
    (C) $f(x) = \frac{1}{4}x + \frac{1}{4}$ for $0 \leq x \leq 2$
    (D) $f(x) = \begin{cases} 0.2, & 0 \leq x \leq 3 \\ 0.4, & 3 < x \leq 4 \\ 0, & \text{elsewhere} \end{cases}$
    (E) $f(x) = \begin{cases} 0.1, & 0 \leq x \leq 8 \\ 0.2, & 8 < x \leq 10 \\ 0, & \text{elsewhere} \end{cases}$

16. The following dotplots show the mean temperature (in degrees Fahrenheit) for a sample of cities in North America. Both January and July temperatures are shown. What is one statement that can be made with certainty from an analysis of the dotplots?



(A) Every city has a higher average temperature in July than in January.

(B) The distribution of temperatures in July is skewed right, while the distribution of temperatures in January is skewed left.

(C) There is more variability in average temperatures in January than in July.

(D) The median average temperature for January is higher than the median average temperature for July.

(E) There are no outliers in the average temperatures for January or July.

17. The following table shows the preferred exercise for a random sample of 223 men of various ages.

Physical Activity/Age	18–31	32–45	46–59	60–73	Over 74
Jogging	23	14	9	1	0
Cycling	19	19	14	11	8
Swimming	10	8	5	3	1
Weight Lifting	34	21	12	6	5

If the type of exercise is independent of age, how many men over the age of 74 would we expect to prefer cycling?

(A) 3

(B) 4

(C) 8

(D) 11

(E) 14

18. A drug company wishes to test a new drug. A researcher assembles a group of volunteers and randomly assigns them to one of two groups—one to take the drug and one to take a placebo. In addition, the company wants the experiment to be double-blind. What is the meaning of double-blind in this situation?

    (A) The volunteers in both groups are blindfolded when they take the drug or placebo.

    (B) The volunteers in both groups do not know whether they are taking the drug or the placebo.

    (C) Neither the volunteers nor the drug company executives know which volunteers are taking the drug and which are taking the placebo.

    (D) Neither the volunteers nor the evaluator know which volunteers are taking the drug and which are taking the placebo.

    (E) As long as the subjects are randomly assigned to the two groups, there is no need to make the experiment double-blind.

19. A baseball recruiter visits a high school where a player has a batting average of 0.450. (This means that he gets a hit in 45% of his at-bats.) What is the probability that the recruiter won't see the player get a hit until his third at-bat?

    (A) $(0.450)^2(0.550)$

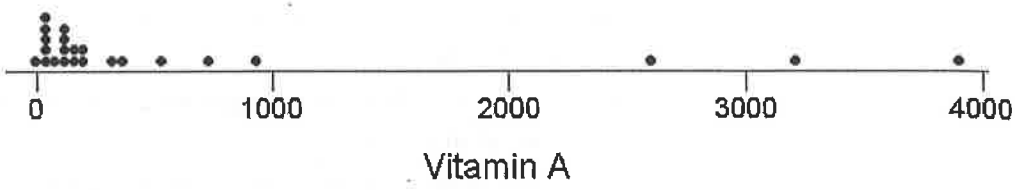
    (B) $(0.550)^2(0.450)$

    (C) $\binom{3}{1}(0.450)(0.550)^2$

    (D) $\binom{3}{1}(0.550)(0.450)^2$

    (E) $\binom{3}{2}(0.450)(0.550)^2$

20. The following graph shows the vitamin A content (in IUs, International Units) for 23 types of fruit.



Which of the following would be the best measure to describe the center of this distribution?

(A) Mean
(B) Median
(C) Standard deviation
(D) Interquartile range
(E) Range

21. Pearson High School students have cumulative grade point averages as shown in the table.

GPA / Class	≥ 4.0	3.0–4.0	2.0–3.0	1.0–2.0	< 1.0	Total
Sophomores	43	121	114	22	10	310
Juniors	26	102	84	16	5	233
Seniors	15	87	100	10	7	219
Total	84	310	298	48	22	762

Which of the following statements is *not* true?

(A) About 39% of sophomores have *at least* a 3.0 GPA.
(B) Sophomores represent 39% of GPAs from 3.0 to 4.0.
(C) Seniors represent about 29% of the reported GPAs at Pearson High School.
(D) Only about 3% of seniors have GPAs *less than* 1.0.
(E) About 11% of the reported GPAs are juniors with GPAs from 2.0 to 3.0.

22. The coefficient of determination, r^2, between two variables is computed to be 81%. Which of the following statements *must* be true?

(A) Large values of the explanatory variable correspond with large values of the response variable.

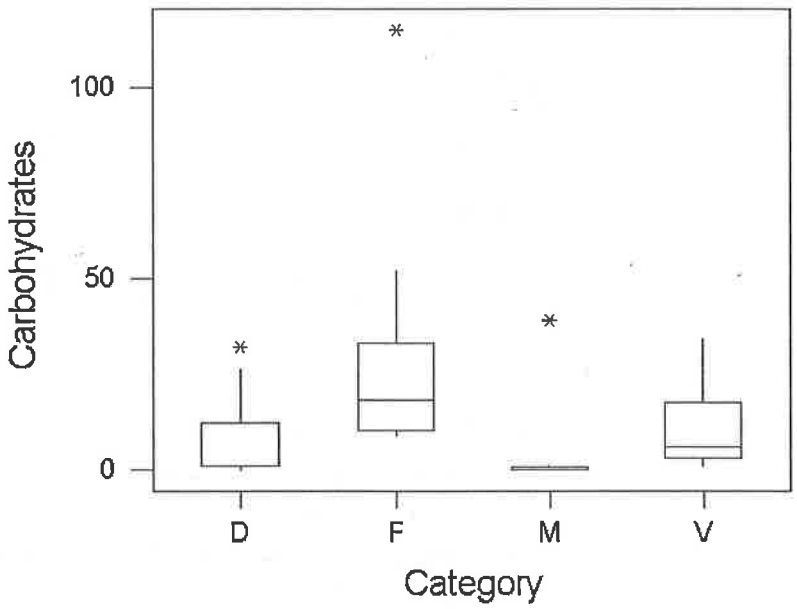
(B) Large values of the explanatory variable correspond with small values of the response variable.

(C) A cause-and-effect relationship exists between the explanatory and response variables.

(D) There is a strong, positive, linear relationship between the explanatory and response variables.

(E) Approximately 81% of the variability in the response variable is explained by linear regression on the explanatory variable.

23. The carbohydrate content (in grams) for serving sizes of select dairy (D), fruit (F), meat (M), and vegetable (V) items is recorded, yielding the following graphical information.



Which of the four food categories would have the smallest value for *all* measures of spread?

(A) Dairy

(B) Fruit

(C) Meats

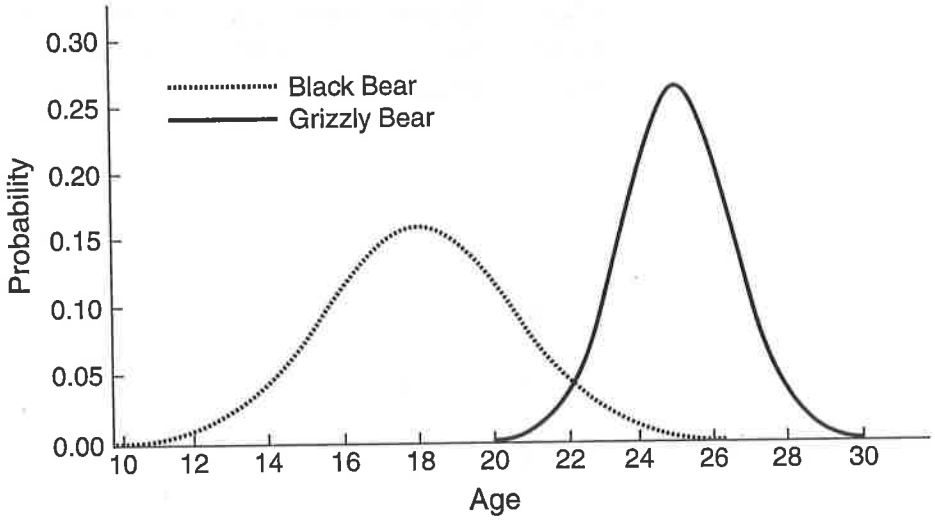
(D) Vegetables

(E) Meats with the outlier removed

24. The correlation between the depth (in feet) and length (in feet) for a sample of caves is found to be -0.346. If the measurement of depth is converted to meters, what will be the resulting correlation? (1 ft = 0.3048 m)

(A) -0.627
(B) -0.346
(C) -0.105
(D) 0.105
(E) 0.346

25. Data are collected from zoos for the age at death (in years) for black bears and grizzly bears. The graphs below model the data collected from these zoos.



Which conclusion can be drawn from this graph?

(A) Grizzly bears tend to live longer than black bears, but their variability is smaller.
(B) Black bears tend to live longer than grizzly bears, but their variability is smaller.
(C) Grizzly bears tend to live longer than black bears, and their variability is larger.
(D) Black bears tend to live longer than grizzly bears, and their variability is larger.
(E) Grizzly bears tend to live longer than black bears, and the two variabilities are approximately equal.

26. In a very large school district, the food services administrator wishes to determine the proportion of students who will buy a school lunch to within ± 0.03. Using the most conservative estimate for p, how many students should this administrator survey to have 90% confidence?

(A) 164
(B) 271
(C) 457
(D) 752
(E) 1844

27. If two events, A and B, are mutually exclusive, then the probability that both A and B occur simultaneously is

(A) 0.
(B) 1.
(C) $P(A) + P(B)$.
(D) $P(A) + P(B) - P(A \cap B)$.
(E) $P(A)\,P(B)$.

28. Which of the following is *not* a characteristic for t-distributions?

(A) The t-distributions are mound-shaped.
(B) The t-distributions are centered at 0.
(C) The t-distributions have more area in the tails than a normal distribution.
(D) The t-distributions use s as an estimate of σ.
(E) As the number of degrees of freedom decreases, the t-distribution approaches the normal distribution.

29. Which of the following is a valid discrete probability distribution?

(A)

x	−10	−9	−8	−7	−6	−5	−4	−3	−2	−1
$P(x)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

(B)

x	−2	1	2	4
$P(x)$	0.2	0.6	0.2	0.1

(C)

x	1	2	3
$P(x)$	0.3	0.2	0.1

(D)

x	1	2	3	4
$P(x)$	0.1	0.2	0.3	−0.1

(E)

x	−2	−1	1	2
$P(x)$	−0.3	−0.2	0.2	0.3

30. For a given school year, a reporter has been told that the average teacher's salary was \$59,500 with a standard deviation of \$17,200. The reporter also knows that teachers will be receiving raises of 3.25% for the next school year. What would the reporter write for the new average teacher's salary and standard deviation?

    (A) mean \$1934; standard deviation \$559
    (B) mean \$59,500; standard deviation \$17,200
    (C) mean \$59,500; standard deviation \$17,759
    (D) mean \$61,434; standard deviation \$17,200
    (E) mean \$61,434; standard deviation \$17,759

31. The average life expectancy for a male in eastern Africa is 45 years. Ten years ago, a major health organization opened a health clinic in a large village located in eastern Africa. The organizers believe the life expectancy for this village has increased as a result of the health care. What are the appropriate hypotheses for a significance test?

    (A) $H_0: \mu = 45; H_a: \mu \neq 45$
    (B) $H_0: \mu = 45; H_a: \mu > 45$
    (C) $H_0: \mu = 45; H_a: \mu < 45$
    (D) $H_0: \mu \neq 45; H_a: \mu = 45$
    (E) $H_0: \mu > 45; H_a: \mu = 45$

32. Since many individuals walk around their homes in their socks, a manufacturer has created a material for socks that is believed to be more resistant to wear than cotton. The manufacturer wishes to test this belief over a period of a month. Given a group of volunteers, which of the following designs will *best* test this new material's resistance to wear?

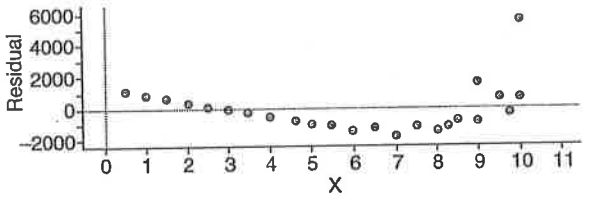
    (A) Have the volunteers wear the socks made from the new material for a month, and check the wear on the socks at the end of the month.
    (B) Allow half of the volunteers to wear cotton socks, while the other half wear socks made of the new material. Compare the wear on the socks at the end of the month.
    (C) Randomly assign half of the volunteers to wear cotton socks, while the other half wear socks made of the new material. Compare the wear on the socks at the end of the month.
    (D) Randomly assign half of the volunteers to wear cotton socks, while the other half wear socks made of the new material. At the end of two weeks, the volunteers should change sock types. Compare the wear on the socks at the end of the month.
    (E) For each volunteer, randomly choose which foot wears a cotton sock, while the other foot wears a sock made of the new material. Compare the wear on the socks at the end of the month.

33. Given the information below, which of the statements is true?

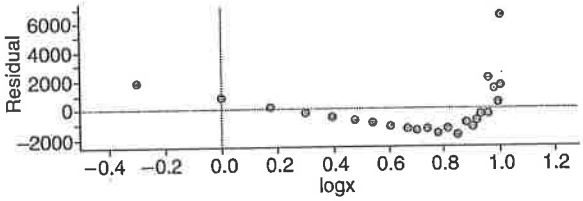
x	2	4	6	8	10
$P(x)$	0.3	0.2	0	0.4	0.1

(A) The expected value of the random variable is 6.
(B) The expected value of the random variable is 0.
(C) The variance of the random variable is 1.
(D) The expected value of the random variable is 11.6.
(E) The variance of the random variable is 8.64.

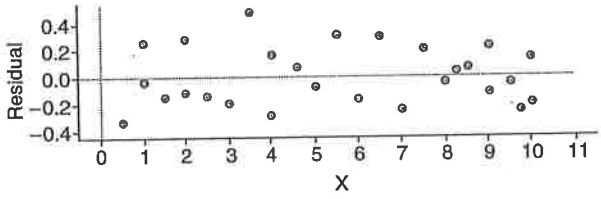
34. Residual plots and their corresponding linear regression equations are shown for a set of data. Using the equation with the best fit, find the predicted value of y when $x = 7$.



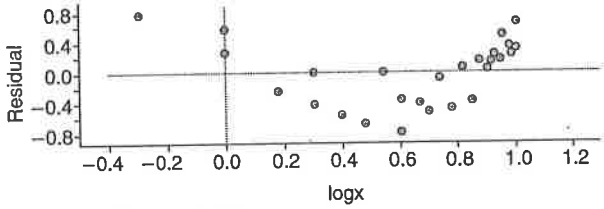
$\hat{Y} = 480X - 1400;$



$\hat{Y} = 3310\log x - 850;$



$\log\hat{y} = 0.387X - 0.051;$



$\log\hat{y} = 3.27\log x + 0.011;$

(A) 454.988
(B) 594.936
(C) 1947.275
(D) 1960.000
(E) Not enough information is given to determine the predicted value for y.

35. A university is proposing a new procedure for professors to gain tenure. To gauge sentiment about the proposal, the university intends to randomly sample five professors, five assistant professors, five associate professors, five adjunct professors, and five visiting professors. This is an example of what type of sampling design?
    (A) Simple random sample
    (B) Stratified random sample
    (C) Systematic random sample
    (D) Cluster sample
    (E) Convenience sample

36. The weights of women are approximately normally distributed. This week, the z-score of weight for a member of a weight-watching group is 1.25. Which of the following is a correct interpretation of this z-score?
    (A) This week the member weighs 1.25 lb more than last week.
    (B) This week the member weighs 1.25 lb less than last week.
    (C) This week the member weighs 1.25 lb more than the average woman.
    (D) This week the member weighs 1.25 standard deviations more than she did last week.
    (E) This week the member weighs 1.25 standard deviations more than the average woman.

37. In a city, 13.5% of the labor force are members of a union. If a random sample of 75 adults is taken from this city, what is the probability that between 15% and 20% of them are union members?

(A) $$P\left(\frac{0.15 - 0.135}{\sqrt{\frac{0.135(1 - 0.135)}{75}}} < z < \frac{0.20 - 0.135}{\sqrt{\frac{0.135(1 - 0.135)}{75}}}\right)$$

(B) $$P\left(\frac{0.15 - 0.135}{\sqrt{\frac{0.15(1 - 0.15)}{75}}} < z < \frac{0.20 - 0.135}{\sqrt{\frac{0.15(1 - 0.15)}{75}}}\right)$$

(C) $$P\left(\frac{0.15 - 0.135}{\sqrt{\frac{0.15(1 - 0.15)}{75}}} < z < \frac{0.20 - 0.135}{\sqrt{\frac{0.2(1 - 0.2)}{75}}}\right)$$

(D) $$P\left(\frac{0.15 - 0.135}{\sqrt{\frac{0.2(1 - 0.2)}{75}}} < z < \frac{0.20 - 0.135}{\sqrt{\frac{0.2(1 - 0.2)}{75}}}\right)$$

(E) $$P\left(\frac{0.135 - 0.15}{\sqrt{\frac{0.15(1 - 0.15)}{75}}} < z < \frac{0.135 - 0.2}{\sqrt{\frac{0.2(1 - 0.2)}{75}}}\right)$$

38. Having read about the positive effects of gingko biloba on memory, some precocious statistics students decide to conduct their own experiment to test the herb's effects. Close to 50 of their classmates, all in good health and representing a variety of ethnic groups, volunteer to take part in the experiment, and the students randomly assign half of the volunteers to take gingko. The other half take a placebo. The students perform a memory test on the volunteers at the beginning of the experiment and a second test eight weeks later. After analyzing their results, they find no memory improvement in the gingko group versus the placebo group. Assuming the students followed all aspects of good experimental design, which of the following can be concluded?

    (A) Gingko biloba does not improve memory, and no one should take it to improve memory.

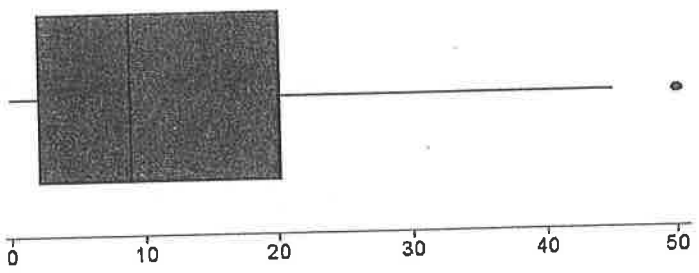
    (B) Gingko biloba does not improve memory in healthy individuals and should only be taken by individuals exhibiting signs of dementia.

    (C) Gingko biloba does not improve memory in healthy teenagers and should only be taken by adults.

    (D) Gingko biloba does not improve memory in healthy teenagers and should only be taken by adults in poor health.

    (E) Gingko biloba does not improve memory in healthy teenagers, and further studies should be conducted to determine its effectiveness in other groups.

39. For the given boxplot, which are the correct summary statistics?



(A)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Data Val	47	12.23	9.00	11.16	12.16	1.77

Variable	Minimum	Maximum	Q1	Q3
Data Val	0.00	50.00	2.00	20.00

(B)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Data Val	47	12.23	9.00	11.16	12.16	1.77

Variable	Minimum	Maximum	Q1	Q3
Data Val	0.00	45.00	2.00	20.00

(C)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Data Val	47	9.00	9.00	11.16	12.16	1.77

Variable	Minimum	Maximum	Q1	Q3
Data Val	0.00	50.00	2.00	20.00

(D)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Data Val	47	9.00	9.00	11.16	12.16	1.77

Variable	Minimum	Maximum	Q1	Q3
Data Val	0.00	45.00	2.00	20.00

(E)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Data Val	47	12.23	20.00	11.16	12.16	1.77

Variable	Minimum	Maximum	Q1	Q3
Data Val	0.00	50.00	9.00	45.00

40. Data have been collected, and a statistician conducts a test of significance using the data. The statistician is considering the effects associated with Type I and Type II errors under these circumstances. Which of the following is true?

I. Reducing the probability of a Type I error increases power.
II. Reducing power increases the probability of a Type II error.
III. Reducing the probability of a Type I error increases the probability of a Type II error.

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

SECTION II: FREE RESPONSE
Part A

90 minutes
Suggested time: 65 minutes

1. A shipment of computer chips consists of 10,000 units. The manufacturer claims that the probability of selecting a defective unit is 0.003.
   a. How many defective units should the receiver of the shipment expect?

   b. What is the probability that the receiver of the shipment will get more defective units than expected?

   c. What is the probability that fewer than 18 units will be defective in this shipment?

2. A manufacturer has created a pole from a new material and believes that pole-vaulters can improve the heights of their vaults by using this new type of pole.
   a. In order to test this hypothesis, the manufacturer visits Kennett High School. He selects a student who is new to vaulting. He records a sample of 25 vaults where the student uses a pole made from the customary material. Then he records a sample of 25 vaults where the student uses a pole constructed from the new material. He analyzes the data and finds a significant improvement in heights with the new pole. Comment on the design of this experiment.

b. Could you improve upon the design of the experiment in part (a)? If you answer yes, describe *your* design.

c. As pole-vaulters improve their performances, longer poles become a necessity. Describe an experiment to test the new material that takes different pole lengths into account.

3. Suppose that the weights of a name-brand cereal vary normally with mean $\mu = 11.13$ oz and standard deviation $\sigma = 0.08$ oz. The advertised weight is 11 oz. For the equivalent generic brand with an advertised weight of 11 oz, the weights vary normally with mean $\mu = 11.15$ oz and standard deviation $\sigma = 0.16$ oz.

   a. For each of the brands, find the probability that the weight of a box of cereal will be less than the advertised weight.

   b. For each of the brands, find the probability that the average weight for a purchase of four boxes of cereal will exceed 11.25 oz.

   c. If you believe strongly in getting what you pay for, that is, you want to make sure you get at least the weight advertised on the box, which brand of cereal are you more likely to buy and why?

   d. If you wanted to get the most for your money, that is, you would like to get much more than the advertised weight, which brand of cereal are you more likely to buy and why?

4. Based on results from a random sample of 35 individuals, a company advertises that individuals who use its diet supplement lose an average of 4.6 lb with a standard deviation of 1.2 lb during the first week of dieting.

   a. A graph of the data reveals no strong skew or outliers. Does this provide evidence of a significant weight loss?

   b. The advertisement fails to mention that these results were part of a double-blind experiment to compare the supplement with a placebo. In the control group, the weight loss was 3.7 lb with a standard deviation of 2.3 lb in the first week of dieting. The control group consisted of a random sample of 32 individuals. A graph of the data for the control group reveals no strong skew or outliers. Is there evidence to show that the group taking the diet supplement lost more weight than the control group in the first week?

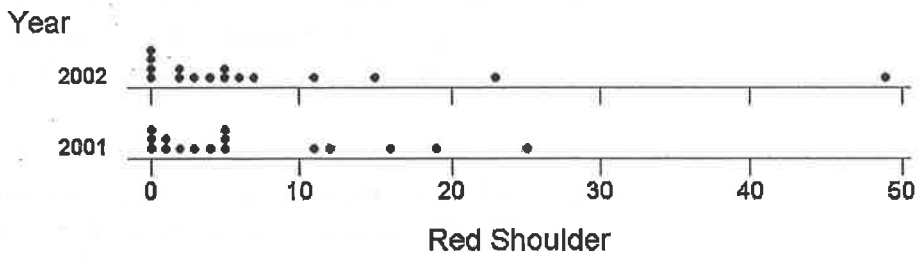
5. In eastern Pennsylvania, the peak migration for red-shouldered hawks is October 15–31. During 2001 and 2002, the owner of a mountaintop home recorded the number of these hawks that migrated past her home each day during the peak migration season. The following summary statistics resulted:

Variable	Year	N	Mean	Median	TrMean	StDev
Red Shoulder	2001	16	6.81	4.50	6.00	7.62
	2002	16	8.25	4.50	5.93	12.53

Variable	Year	SE Mean	Minimum	Maximum	Q1	Q3
Red Shoulder	2001	1.90	0.00	25.00	1.00	11.75
	2002	3.13	0.00	49.00	0.50	10.00

   a. Are there any outliers in the distribution of migration data for red-shouldered hawks for 2002? Describe the procedure you used to determine outliers, and justify your answer based on that procedure.

b. Dotplots for the distributions of migration data for both years are given.



If you said there were outliers in part (a), would the value(s) still be outliers if we combined the results from 2001 and 2002? What is the total number of outliers for the combined data?

c. Write a few sentences comparing the migration of red-shouldered hawks in 2001 and 2002.

d. Using the combined results from 2001 and 2002, if a friend visits during the peak migration season, what is the probability the friend will see on a given day 20 or more red-shouldered hawks?

e. Using the combined results from 2001 and 2002, what is the most probable number of red-shouldered hawks the same friend would see on a given day during the peak migration season?

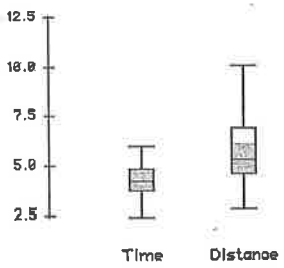
Part B

Suggested time: 25 minutes

6. A hiker records distances (in miles) and times (in hours) for a random sample of 25 of her hikes, resulting in the following information.

Summary of	Distance
No Selector	
Count	25
Mean	5.952
Median	5.7
StdDev	2.71356
IntQRange	2.975

Summary of	Time
No Selector	
Count	25
Mean	3.7
Median	3.5
StdDev	1.54448
IntQRange	2.0625



Dependent variable is: **Time**
No Selector
R squared = 77.3% R squared (adjusted) = 76.3%
s = 0.7515 with 25 - 2 = 23 degrees of freedom

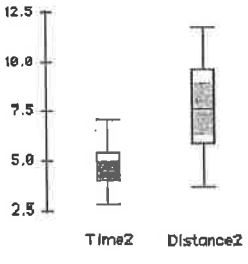
Source	Sum of Squares	df	Mean Square	F-ratio
Regression	44.2594	1	44.2594	78.4
Residual	12.9906	23	0.564807	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	0.721346	0.3685	1.96	0.0625
Distance	0.500446	0.05653	8.85	≤ 0.0001

A friend also records distances (in miles) and times (in hours) for a random sample of 20 of his hikes, resulting in the following information.

Summary of	Distance2
No Selector	
Count	20
Mean	7.585
Median	7.65
StdDev	2.4577
IntQRange	3.7

Summary of	Time2
No Selector	
Count	20
Mean	4.5875
Median	4.5
StdDev	1.2577
IntQRange	1.375



Dependent variable is: **Time2**
No Selector
R squared = 85.5% R squared (adjusted) = 84.7%
s = 0.4952 with 20 - 2 = 18 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	26.1202	1	26.1202	107
Residual	4.41416	18	0.245231	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	0.968919	0.3677	2.64	0.0168
Distance2	0.477071	0.04623	10.3	≤ 0.0001

These friends are trying to determine whether or not they would be well suited to hike together.

a. One of the friends believes that hiking distances are important to determining compatibility. Is there any evidence of a significant difference in hiking distances between the two hikers?

b. The other friend believes it is the length of the hike in time that best determines compatibility. Is there any evidence of a significant difference in hiking times between the two hikers?

c. Should the two friends hike together? Use your results from parts (a) and (b) to support your answer.

d. Rather than look at the choice of hikes, which would be reflected in the distances and times of the hikes reported, a mutual friend suggests that the two should instead look at their hiking rate. This friend conjectures that hikers with rates of speed within 0.1 mph of one another make compatible hikers. Record the speed at which these friends hike. Would your answer to part (c) change if their mutual friend is correct in his conjecture?