2012
multi-choice

STATISTICS
SECTION I
Time—1 hour and 30 minutes
Number of questions—40
Percent of total score—50

ections: Solve each of the following problems, using the available space for scratch work. Decide which is the : of the choices given and fill in the corresponding circle on the answer sheet. No credit will be given for ything written in the test book. Do not spend too much time on any one problem.

. Of the following dotplots, which represents the set of data that has the greatest standard deviation?

(A) 0 to 10

(B) 0 to 10

(C) 0 to 10

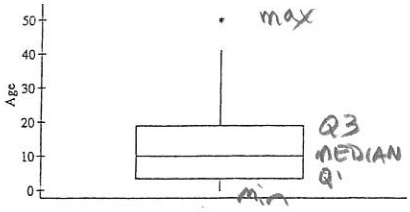
(D) 0 to 10

(E) 0 to 10

Std dev = avg distance data is from the mean

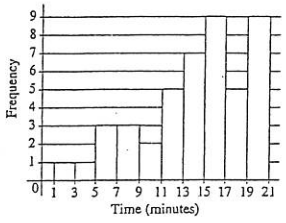
(B has more data far from mean, and less data close to mean)

2. A random sample of 374 United States pennies was collected, and the age of each penny was determined. According to the boxplot below, what is the approximate interquartile range (IQR) of the ages?



(A) 8
(B) 10
(C) 16
(D) 40
(E) 50

Q3 ≈ 20
Q1 ≈ 5
IQR = Q3 − Q1
≈ 20 − 5
≈ 15



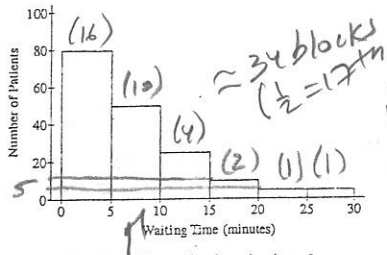
3. The histogram above shows the number of minutes needed by 45 students to finish playing a computer game. Which of the following statements is correct?

(A) The distribution is skewed to the right.
(B) The distribution is skewed to the left.
(C) The distribution appears to be normal.
(D) The distribution appears to be chi-square.
(E) The distribution appears to be uniform.

4. A bank surveyed all of its 60 employees to determine the proportion who participate in volunteer activities. Which of the following statements is true?

(A) The bank should not use the data from this survey because this is an observational study.
(B) The bank can use the result of this survey to prove that working for the bank causes employees to participate in volunteer activities.
(C) The bank did not select a random sample of employees, so the survey will not provide the bank with useful information.
(D) The bank would have to use the survey data to construct a confidence interval in order to estimate the proportion of employees who participate in volunteer activities.
(E) The bank does not need to use an inference procedure to determine the proportion of employees who participate in volunteer activities because the survey was a census of all employees.

5. The histogram below displays the frequencies of waiting times, in minutes, for 175 patients in a dentist's office.



Which of the following could be the median of the waiting times, in minutes?

(A) 2.50
(B) 7.25
(C) 12.25
(D) 15.00
(E) 17.50

≈ 34 blocks ($\frac{1}{2}$ = 17th block)
count 'blocks' to find where middle 'block' would be
17th block somewhere in here between 5 and 10
block ≈ 5

GO ON TO THE NEXT PAGE.

6. Data were collected on the amount, in dollars, that individual customers spent on dinner in an Italian restaurant. The quartiles for these data are given below.

(median)

Q1	Q2	Q3
$36.27	$44.27	$58.97

Which of the following statements must be true for these customers?

(A) At least half of the customers spent less than or equal to $44.27 and at least half spent greater than or equal to $44.27.
(B) Seventy-five percent of the customers spent between $36.27 and $58.97. (50%)
(C) Twenty-five percent of the customers spent less than or equal to $58.97 and the remaining 75 percent spent greater than or equal to $58.97. (25% spent more than or equal to $58.97)
(D) The mean amount spent by customers is $44.27. can't determine mean from median
(E) A majority of customers spent $44.27.

50% spent more, 50% spent less

GO ON TO THE NEXT PAGE.

7. The weight of adult male grizzly bears living in the wild in the continental United States is approximately normally distributed with a mean of 500 pounds and a standard deviation of 50 pounds. The weight of adult female grizzly bears is approximately normally distributed with a mean of 300 pounds and a standard deviation of 40 pounds. Approximately, what would be the weight of a female grizzly bear with the same standardized score (z-score) as a male grizzly bear with a weight of 530 pounds?

(A) 276 pounds
(B) 324 pounds
(C) 330 pounds
(D) 340 pounds
(E) 530 pounds

$$z_{male} = \frac{530 - 500}{50} = .6$$

$$z_{female} = \frac{x - 300}{40} = .6$$

$$x - 300 = .6(40)$$

$$x = .6(40) + 300$$

$$= 324$$

GO ON TO THE NEXT PAGE.

8. A company sells concrete in batches of 5 cubic yards. The probability distribution of X, the number of cubic yards sold in a single order for concrete from this company, is shown in the table below.

X = the number of cubic yards	10	15	20	25	30
Probability	0.15	0.25	0.25	0.30	0.05

The expected value of the probability distribution of X is 19.25 and the standard deviation is 5.76. There is a fixed cost to deliver the concrete. The profit Y, in dollars, for a particular order can be described by $Y = 75X - 100$. What is the standard deviation of Y?

(A) $332.00
(B) $432.00
(C) $532.00
(D) $1,343.75
(E) $1,400.00

Standard deviation is affected by multipliers, but not by add/subtract
here X is multiplied by 75 and then 100 is subtracted
only the multiply by 75 affects measures of spread
$\sigma_{cost} = 5.76(75) = \$432$

GO ON TO THE NEXT PAGE.

9. Based on a survey of a random sample of 900 adults in the United States, a journalist reports that 60 percent of adults in the United States are in favor of increasing the minimum hourly wage. If the reported percent has a margin of error of 2.7 percentage points, which of the following is closest to the level of confidence?

(A) 80.0%
(B) 90.0%
(C) 95.0%
(D) 95.5%
(E) 99.0%

proportions:

$CI = \hat{p} \pm z^* \sqrt{\frac{pq}{n}}$

margin of error

so $z^* \sqrt{\frac{pq}{n}} = .027$

$z^* \sqrt{\frac{(.6)(.4)}{900}} = .027$

$z^* = \frac{.027}{\sqrt{\frac{(.6)(.4)}{900}}} = 1.653$

z scores

normalcdf(−1.653, 1.653, 0, 1)

= .9017

≈ 90%

−1.653 0 +1.653

GO ON TO THE NEXT PAGE.

10. A compact disc (CD) manufacturer wanted to determine which of two different cover designs for a newly released CD will generate more sales. The manufacturer chose 70 stores to sell the CD. Thirty-five of these stores were randomly assigned to sell CDs with one of the cover designs and the other 35 were assigned to sell the CDs with the other cover design. The manufacturer recorded the number of CDs sold at each of the stores and found a significant difference between the mean number of CDs sold for the two cover designs. Which of the following gives the conclusion that should be made based on the results and provides the best explanation for the conclusion?

(A) It is not reasonable to conclude that the difference in sales was caused by the different cover designs because this was not an experiment.
(B) It is not reasonable to conclude that the difference in sales was caused by the different cover designs because there was no control group for comparison.
(C) It is not reasonable to conclude that the difference in sales was caused by the different cover designs because the 70 stores were not randomly chosen.
(D) It is reasonable to conclude that the difference in sales was caused by the different cover designs because the cover designs were randomly assigned to stores.
(E) It is reasonable to conclude that the difference in sales was caused by the different cover designs because the sample size was large.

random assignment is required for an experiment

GO ON TO THE NEXT PAGE.

11. The manager of a public swimming pool wants to compare the effectiveness of two laundry detergents, Detergent A and Detergent B, in cleaning the towels that are used daily. As each dirty towel is turned in, it is placed into the only washing machine on the premises. When the washing machine contains 20 towels, the manager flips a coin to determine whether Detergent A or Detergent B will be used for that load. The cleanliness of the load of towels is rated on a scale of 1 to 10 by a person who does not know which detergent was used. The manager continues this experiment for many days. Which of the following best describes the manager's study?

(A) A completely randomized design
(B) A randomized block design with Detergent A and Detergent B as blocks
(C) A randomized block design with the washing machine as the block
(D) A matched-pairs design with Detergent A and Detergent B as the pair
(E) An observational study

← blocking can only be done on differences in the subject

pairing is done with subjects (not treatment)

Book answer is A

(but this experiment doesn't involve random assignment)

Better would have been to number the towels as they came in and then randomly assign half to detergent A

(this design — what if later towels are dirtier than earlier towels?)

12. In the design of a survey, which of the following best explains how to minimize response bias?

(A) Increase the sample size.
(B) Decrease the sample size.
(C) Randomly select the sample.
(D) Increase the number of questions in the survey.
(E) Carefully word and field-test survey questions.

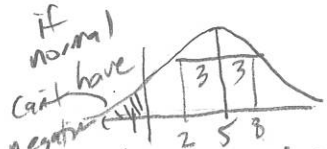
↓

when something about the survey influences people to change their responses.

13. For a sample of 42 rabbits, the mean weight is 5 pounds and the standard deviation of weights is 3 pounds. Which of the following is most likely true about the weights for the rabbits in this sample?

(A) The distribution of weights is approximately normal because the sample size is 42, and therefore the central limit theorem applies.
(B) The distribution of weights is approximately normal because the standard deviation is less than the mean.
(C) The distribution of weights is skewed to the right because the least possible weight is within 2 standard deviations of the mean.
(D) The distribution of weights is skewed to the left because the least possible weight is within 2 standard deviations of the mean.
(E) The distribution of weights has a median that is greater than the mean.

if normal can't have negative weight



↳ this would be redistributed

GO ON TO THE NEXT PAGE.

GO ON TO THE NEXT PAGE.

skip

14. When conducting a large sample test of $H_0 : p = p_0$ for a single proportion, the test statistic is $z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, where $\hat{p}$ is the sample proportion. Which of the following best explains the justification for the denominator of this test statistic?

(A) The standard deviation of $\hat{p}$ is known when the null hypothesis is true.

(B) The standard deviation of $\hat{p}$ is known when the alternative hypothesis is true.

(C) The sample size is large and therefore the standard deviation of p_0 is approximated well.

(D) The standard deviation of p_0 is known when the null hypothesis is true.

(E) The standard deviation of p_0 is known when the alternative hypothesis is true.

this question is worded strangely and they ended up dropping the question.

GO ON TO THE NEXT PAGE.

15. A polling firm is interested in surveying a representative sample of registered voters in the United States. The firm has automated its sampling so that random phone numbers within the United States are called. Each time a number is called, the procedure below is followed.

- If there is no response or if an answering machine is reached, another number is automatically called.
- If a person answers, a survey worker verifies that the person is at least 18 years of age.
- If the person is not at least 18 years of age, no response is recorded, and another number is called.
- If the person is at least 18 years of age, that person is surveyed.

Some people claim the procedure being used does not permit the results to be extended to all registered v
Which of the following is NOT a legitimate concern about the procedure being used?

(A) Registered voters with children under the age of 18 years may be underrepresented in the sample.

(B) Registered voters with unlisted telephone numbers may be underrepresented in the sample.

(C) Registered voters who have more than one telephone number may be overrepresented in the sample.

(D) Registered voters who live in households consisting of more than one voter may be underrepresented.

(E) People who are not registered to vote may bias the sample results.

They aren't using a list of phone numbers for registered voters, they are just calling randomly chosen phone numbers, so unlisted numbers could be called.

GO ON TO THE NEXT PAGE.

16. A complex electronic device contains three components, A, B, and C. The probabilities of failure for each component in any one year are 0.01, 0.03, and 0.04, respectively. If any one component fails, the device will fail. If the components fail independently of one another, what is the probability that the device will not fail in one year?

(A) Less than 0.01

(B) 0.078

(C) 0.080

(D) 0.922

(E) Greater than 0.99

$P(A \text{ not fail}) = 1 - .01 = .99$

$P(B \text{ not fail}) = 1 - .03 = .97$

$P(C \text{ not fail}) = 1 - .04 = .96$

$P(\text{no fail}) = P(A \text{ not fail}) \text{ and } P(B \text{ not fail}) \cdot P(C \text{ not fail})$

AND case

$= (.99)(.97)(.96)$

$= .92189$

18. When using a one-sample t-procedure to construct a confidence interval for the mean of a finite population, a condition is that the population size be at least 10 times the sample size. The reason for the condition is to ensure that

(A) the sample size is large enough

(B) the central limit theorem is applicable for the sample mean

(C) the sample standard deviation is a good approximation of the population standard deviation

(D) the degree of dependence among observations is negligible

(E) the sampling method is not biased

if you sample too many items, the probability of result on next trial is affected.

17. A large-sample 98 percent confidence interval for the proportion of hotel reservations that are canceled on the intended arrival day is (0.048, 0.112). What is the point estimate for the proportion of hotel reservations that are canceled on the intended arrival day from which this interval was constructed?

(A) 0.032

(B) 0.064

(C) 0.080

(D) 0.160

(E) It cannot be determined from the information given.

the statistic for this experiment, $\hat{p}$ (always in the middle of the CI)

$\hat{p} = \frac{.112 + .048}{2}$

$= .08$

GO ON TO THE NEXT PAGE.

GO ON TO THE NEXT PAGE.

19. A nonprofit organization plans to hold a raffle to raise funds for its operations. A total of 1,000 raffle tickets will be sold for $1.00 each. After all the tickets are sold, one ticket will be selected at random and its owner will receive $50.00. The expected value for the net gain for each ticket is −$0.95. What is the meaning of the expected value in this context?

(A) The ticket owners lose an average of $0.05 per raffle ticket.

(B) The ticket owners lose an average of $0.95 per raffle ticket.

(C) Each ticket owner will lose $0.95 per raffle ticket.

(D) A ticket owner would have to purchase 19 more tickets for the expected value of his or her net gain to increase to $0.00.

(E) A ticket owner has a 95 percent chance of having a ticket that is not selected.

20. Suppose that on a hypothesis test for a single population mean, H_a: $\mu < 10$. Assume that H_a is true. For a fixed sample size and significance level α, the power of the test will be greatest if the actual mean is which of the following?

(A) 8

(B) 9

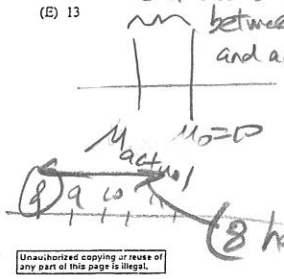
(C) 10

(D) 11

(E) 13

power of a test = % chance it will detect a difference as significant, if there is a difference to detect.

effect size = diff between μ_0 and actual μ

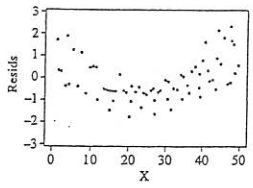


↑ the bigger the effect size, the easier it is to detect, P(detecting) ↑, so power ↑

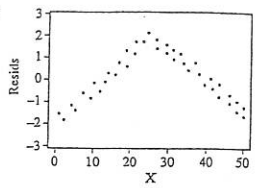
(8 has largest effect size)

21. The residual plots from five different least squares regression lines are shown below. Which of the plots provides the strongest evidence that its regression line is an appropriate model for the data and is consistent with the assumptions required for inference for regression?

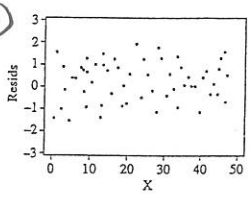
(A)



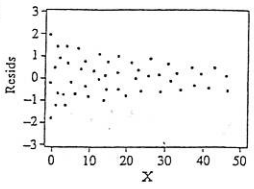
(B)



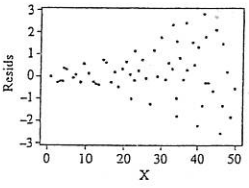
(C)



(D)



(E)



(least pattern in residuals = best fit)

22. A random sample of 50 students at a large high school resulted in a 95 percent confidence interval for the mean number of hours of sleep per day of (6.73, 7.67). Which of the following statements best summarizes the meaning of this confidence interval?

(A) About 95% of all random samples of 50 students from this population would result in a 95% confidence interval of (6.73, 7.67).

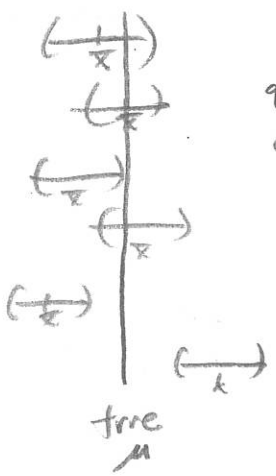
(B) About 95% of all random samples of 50 students from this population would result in a 95% confidence interval that covered the population mean number of hours of sleep per day.

(C) 95% of the students in the survey reported sleeping between 6.73 and 7.67 hours per day.

(D) 95% of the students in this high school sleep between 6.73 and 7.67 hours per day.

(E) A student selected at random from this population sleeps between 6.73 and 7.67 hours per day for 95% of the time.

confidence level (95%) is the probability that a confidence interval will "capture" the true population value.



95% of the CIs, centered at each sample $\bar{x}$, capture the true population μ

23. A local company is interested in supporting environmentally friendly initiatives such as carpooling among employees. The company surveyed all of the 200 employees at the downtown offices. Employees responded as to whether or not they own a car and to the location of the home where they live. The results are shown in the table below.

		Location of Home			
		Downtown Area In the City	Elsewhere In the City	Outside the City	Total
Car Ownership	Yes	10	15	35	60
	No	60	55	25	140
	Total	70	70	60	200

Which of the following statements about a randomly chosen person from these 200 employees is true?

(A) If the person owns a car, he or she is more likely to live elsewhere in the city than to live in the downtown area in the city. ($15 > 10$)

(B) If the person does not own a car, he or she is more likely to live outside the city than to live in the city (downtown area or elsewhere).

(C) The person is more likely to own a car if he or she lives in the city (downtown area or elsewhere) than if he or she lives outside the city.

(D) The person is more likely to live in the downtown area in the city than elsewhere in the city.

(E) The person is more likely to own a car than not to own a car.

24. A random sample of 432 voters revealed that 100 are in favor of a certain bond issue. A 95 percent confidence interval for the proportion of the population of voters who are in favor of the bond issue is

(A) $100 = 1.96\sqrt{\frac{0.5(0.5)}{432}}$

(B) $100 \pm 1.645\sqrt{\frac{0.5(0.5)}{432}}$

(C) $100 \pm 1.96\sqrt{\frac{0.231(0.769)}{432}}$

(D) $0.231 \pm 1.96\sqrt{\frac{0.231(0.769)}{432}}$

(E) $0.231 \pm 1.645\sqrt{\frac{0.231(0.769)}{432}}$

$\hat{p} = \frac{100}{432} = .231$

Confidence intervals (because they are centered at the statistic $\hat{p}$) should use $\hat{p}$, not p for the standard error calculation

$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$= .231 \pm 1.96\sqrt{\frac{(.231)(.769)}{432}}$

for 95%:

=invNorm(.025, 0, 1)

=(-)1.96

GO ON TO THE NEXT PAGE.

25. The commuting time for a student to travel from home to a college campus is normally distributed with a mean of 30 minutes and a standard deviation of 5 minutes. If the student leaves home at 8:25 A.M., what is the probability that the student will arrive at the college campus later than 9 A.M.?

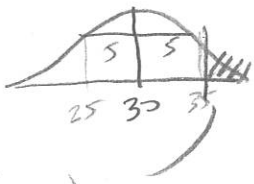
(A) 0.16

(B) 0.32

(C) 0.50

(D) 0.84

(E) 1.00



9:00AM
−8:25AM
35min

P(>35min):

=normalcdf(35, 999, 30, 5)

= .1587

GO ON TO THE NEXT PAGE.

26. In 2009 a survey of Internet usage found that 79 percent of adults age 18 years and older in the United States use the Internet. A broadband company believes that the percent is greater now than it was in 2009 and will conduct a survey. The company plans to construct a 98 percent confidence interval to estimate the current percent and wants the margin of error to be no more than 2.5 percentage points. Assuming that at least 79 percent of adults use the Internet, which of the following should be used to find the sample size (n) needed?

(A) $1.96\sqrt{\frac{(0.5)}{n}} \le 0.025$

(B) $1.96\sqrt{\frac{(0.5)(0.5)}{n}} \le 0.025$

(C) $2.33\sqrt{\frac{(0.5)(0.5)}{n}} \le 0.05$

(D) $2.33\sqrt{\frac{(0.79)(0.21)}{n}} \le 0.025$

(E) $2.33\sqrt{\frac{(0.79)(0.21)}{n}} \le 0.05$

for proportions

$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

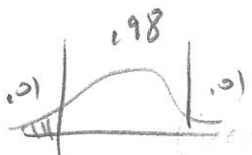
margin of error

$z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} \le .025$

$2.33\sqrt{\frac{(.5)(.5)}{n}} \le .025$

↑ really should be this. Use .5 for $\hat{p}$ for worst case

best choice here is d (using p_0 for $\hat{p}$)



=invNorm(.01, 0, 1)

= (-)2.326

GO ON TO THE NEXT PAGE.

27. A manufacturer claims its Brand A battery lasts longer than its competitor's Brand B battery. Nine batteries of each brand are tested independently, and the hours of battery life are shown in the table below.

Brand A	88	85	80	81	72	90	85	85	84
Brand B	80	79	77	82	75	81	77	73	78

Provided that the assumptions for inference are met, which of the following tests should be conducted to determine if Brand A batteries do, in fact, last longer than Brand B batteries?

(A) A one-sided, paired t-test

(B) A one-sided, two-sample t-test

(C) A two-sided, two-sample t-test

(D) A one-sided, two-sample z-test

(E) A two-sided, two-sample z-test

$H_0: \mu_A = \mu_B$

$H_A: \mu_A > \mu_B$ ↑ 1-sided

- 2 sample because these are not matched pairs
- this is numerical data (mean) so ~~t-test~~ (t tests are for categorical/proportions)

GO ON TO THE NEXT PAGE.

28. An experimenter conducted a two-tailed hypothesis test on a set of data and obtained a p-value of 0.44. If the experimenter had conducted a one-tailed test on the same set of data, which of the following is true about the possible p-value(s) that the experimenter could have obtained?

(A) The only possible p-value is 0.22.
(B) The only possible p-value is 0.44.
(C) The only possible p-value is 0.88.
(D) The possible p-values are 0.22 and 0.78.
(E) The possible p-values are 0.22 and 0.88.

p would be $\frac{.44}{2} = .22$

but...

$H_A: \mu < \mu_0$ $H_A: \mu > \mu_0$

1-sided tests could be upper or lower tailed so could be p-val = .22 or 1-.22

GO ON TO THE NEXT PAGE.

29. A randomized experiment was performed to determine whether two fertilizers, A and B, give different yields of tomatoes. A total of 33 tomato plants were grown; 16 using fertilizer A, and 17 using fertilizer B. The distributions of the data did not show marked skewness and there were no outliers in either data set. The results of the experiment are shown below.

	Fertilizer A	Fertilizer B
Average number of tomatoes per plant	19.54	23.39
Standard deviation	3.68	4.93
Number of plants	16	17

Which of the following statements best describes the conclusion that can be drawn from this experiment?

(A) There is no statistical evidence of difference in the yields between fertilizer A and fertilizer B ($p > 0.15$).
(B) There is a borderline statistically significant difference in the yields between fertilizer A and fertilizer B ($0.10 < p < 0.15$).
(C) There is evidence of a statistically significant difference in the yields between fertilizer A and fertilizer B ($0.05 < p < 0.10$).
(D) There is evidence of a statistically significant difference in the yields between fertilizer A and fertilizer B ($0.01 < p < 0.05$).
(E) There is evidence of a statistically significant difference in the yields between fertilizer A and fertilizer B ($p < 0.01$).

means, 2 indp. samples $H_0: \mu_A = \mu_B$

do a 2SampTTest using $H_A: \mu_A \neq \mu_B$

$\bar{x}_1 = 19.54$ $\bar{x}_2 = 23.39$ (no pooling)

$s_1 = 3.68$ $s_2 = 4.93$

$n_1 = 16$ $n_2 = 17$

p-val = .0161

GO ON TO THE NEXT PAGE.

30. In order to plan its next advertising campaign, the Trendy Motor Vehicle Company is investigating whether the type of vehicle and the color of vehicle are related. Each person in a random sample of size 275 selected from the company's mailing list was classified according to the type (car or truck) and the color of vehicle he or she drove. The data are shown in the table below.

		Vehicle Color				
		Red	Black	White	Tan	Green
Vehicle Type	Car	35	23	41	21	12
	Truck	27	55	39	12	10

Which of the following procedures would be most appropriate to use for investigating whether there is a relationship between vehicle type and color?

(A) A two-sample t-test
(B) A two-sample z-test
(C) A matched pairs t-test
(D) A chi-square goodness-of-fit test
(E) A chi-square test of independence

GO ON TO THE NEXT PAGE.

31. A large number of randomized experiments were conducted to determine whether taking a particular drug regularly would decrease the chance of getting a certain disease. For each of the experiments, the drug effect is the difference between the proportion of people taking the drug who got the disease and the proportion of people taking a placebo who got the disease. If the drug had no effect whatsoever, which of the following experimental results would be anticipated?

I. p-values will be greater than 0.05 for about 95 percent of the experiments.
II. There will be about an equal number of experiments showing positive and negative values of drug effect.
III. When 95 percent confidence intervals for the population drug effect are constructed, those confidence intervals include 0 about 95 percent of the time.

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

I)

If drug is really not effective H_0 is true, so we should be having $p > .05$ most of the time. $p < .05$ would be a type I error which occurs with prob $= \alpha = .05$

II) If there is no difference between drug and placebo the "best" treatment should be drug about half the time.

III) CI including 0 means results not significant, which should occur 95% of the time.

GO ON TO THE NEXT PAGE.

32. As part of a class project at a large university, Amber selected a random sample of 12 students in her major field of study. All students in the sample were asked to report their number of hours spent studying for the final exam and their score on the final exam. A regression analysis on the data produced the following partial computer output.

Predictor	Coef	SE Coef	T	P
Constant	62.328	4.570	13.64	0.000
Stdy Hours	2.697 = b	0.745 = s_b	3.62	0.005
S = 5.505	R-sq = 56.7%			

Amber wants to compute a 95 percent confidence interval for the slope of the least squares regression line in the population of all students in her major field of study. Assuming that conditions for inference are satisfied, which of the following gives the margin of error for the confidence interval?

(A) $(2.228)(0.74)$

(B) $(2.228)\left(\frac{0.74}{\sqrt{12}}\right)$

(C) $(2.228)(5.56)$

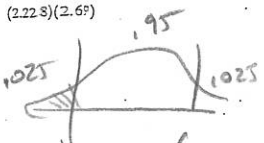
(D) $(2.228)\left(\frac{5.56}{\sqrt{12}}\right)$

(E) $(2.228)(2.69)$

$CI = b \pm t^* s_b$

margin of error

$= 2.228\ (.745)$



$t^* = invT(.025, 10)$ df $n-2$

$= (-)\ 2.228$

GO ON TO THE NEXT PAGE.

33. A mathematics competition uses the following scoring procedure to discourage students from guessing (choosing an answer randomly) on the multiple-choice questions. For each correct response, the score is 7. For each question left unanswered, the score is 2. For each incorrect response, the score is 0. If there are 5 choices for each question, what is the minimum number of choices that the student must eliminate before it is advantageous to guess among the rest?

(A) 0

(B) 1

(C) 2

(D) 3

(E) 4

Expected value > 2

if you guess w/ 4 choices

X	7	0
P	$\frac{1}{4}$	$\frac{3}{4}$

$EV = (7)(\frac{1}{4}) + (0)(\frac{3}{4}) = 1.75$

3 choices

	$\frac{1}{3}$	$\frac{2}{3}$

$(7)(\frac{1}{3}) + (0)(\frac{2}{3}) = 2.33$

now expected value from guessing is higher than 2

must eliminate at least 2

34. The probability that a new microwave oven will stop working in less than 2 years is 0.05. The probability that a new microwave oven is damaged during delivery and stops working in less than 2 years is 0.04. The probability that a new microwave oven is damaged during delivery is 0.10. Given that a new microwave oven is damaged during delivery, what is the probability that it stops working in less than 2 years?

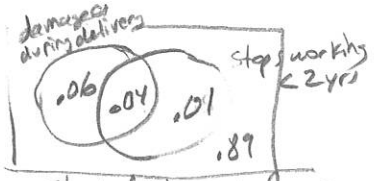
(A) 0.05

(B) 0.06

(C) 0.10

(D) 0.40

(E) 0.50



Start filling out from most overlapped region

$P(\text{stops} < 2\text{yrs} \mid \text{damaged during deliv}) = \frac{P(<2\text{yrs} \cap \text{DDD})}{P(\text{DDD})} = \frac{.04}{.06+.04} = \frac{.04}{.10} = .40$

GO ON TO THE NEXT PAGE.

35. Perchlorate is a chemical used in rocket fuel. People who live near a former rocket-testing site are concerned that perchlorate is present in unsafe amounts in their drinking water. Drinking water is considered safe when the average level of perchlorate is 24.5 parts per billion (ppb) or less. A random sample of 28 water sources in this area produces a mean perchlorate measure of 25.3 ppb. Which of the following is an appropriate alternative hypothesis that addresses their concern?

(A) $H_a: \mu < 25.3$

(B) $H_a: \mu > 25.3$

(C) $H_a: \mu < 24.5$

(D) $H_a: \mu > 24.5$

(E) $H_a: \mu \neq 24.5$

$H_0: \mu = 24.5$ ppb

$H_a: \mu > 24.5$ ppb

always same number

36. The computer output below shows the result of a linear regression analysis for predicting the concentration of zinc, in parts per million (ppm), from the concentration of lead, in ppm, found in fish from a certain river.

Response variable is Zinc (ppm)

Variable	Coefficient	Std Dev	T	P
Constant	16.3	4.90	3.32	0.003
Lead (ppm)	19.0	1.89	10.01	0.000
S = 16.17	R-Sq = 82.0%			

$\hat{y} = \mu$

Which of the following statements is a correct interpretation of the value 19.0 in the output?

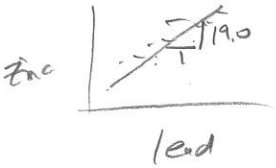
(A) On average there is a predicted increase of 19.0 ppm in concentration of lead for every increase of 1 ppm in concentration of zinc found in the fish.

(B) On average there is a predicted increase of 19.0 ppm in concentration of zinc for every increase of 1 ppm in concentration of lead found in the fish.

(C) The predicted concentration of zinc is 19.0 ppm in fish with no concentration of lead.

(D) The predicted concentration of lead is 19.0 ppm in fish with no concentration of zinc.

(E) Approximately 19% of the variability in zinc concentration is predicted by its linear relationship with lead concentration.



$\hat{Zinc} = 16.3 + 19.0(\text{lead})$

GO ON TO THE NEXT PAGE.

GO ON TO THE NEXT PAGE.

37. There were 5,317 previously owned homes sold in a western city in the year 2000. The distribution of the sales prices of these homes was strongly right-skewed, with a mean of \$206,274 and a standard deviation of \$37,881. If all possible simple random samples of size 100 are drawn from this population and the mean is computed for each of these samples, which of the following describes the sampling distribution of the sample mean?

(A) Approximately normal with mean \$206,274 and standard deviation \$3,788

(B) Approximately normal with mean \$206,274 and standard deviation \$37,881

(C) Approximately normal with mean \$206,274 and standard deviation \$520

(D) Strongly right-skewed with mean \$206,274 and standard deviation \$3,788

(E) Strongly right-skewed with mean \$206,274 and standard deviation \$37,881

n = 100 ⇒ approx normal

$\mu_{\bar{x}} = \mu = \$206,274$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$37881}{\sqrt{100}}$ \$3788

GO ON TO THE NEXT PAGE.

38. A physician believes that the exercise habits of East Coast adults are different from the exercise habits of West Coast adults. To study this, she gathers information on the number of hours of exercise per week from a random sample of East Coast adults and a random sample of West Coast adults. Which of the following might be an appropriate null hypothesis for this study?

(A) The average number of hours of exercise per week for East Coast adults is different from the average number of hours of exercise per week for West Coast adults.

(B) The average number of hours of exercise per week for East Coast adults is the same as the average number of hours of exercise per week for West Coast adults.

(C) The average number of hours of exercise per week for East Coast adults is greater than the average number of hours of exercise per week for West Coast adults.

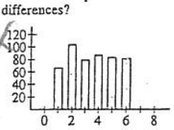
(D) The average number of hours of exercise per week for East Coast adults is less than the average number of hours of exercise per week for West Coast adults.

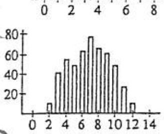
(E) The probability is 0.5 that an East Coast adult and a West Coast adult exercise an equal number of hours per week.

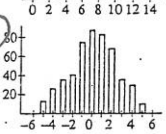
$H_0: \mu_{EC} = \mu_{WC}$

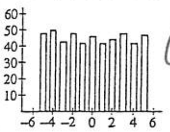
GO ON TO THE NEXT PAGE.

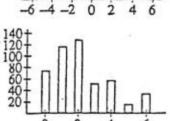
39. For a roll of a fair die, each of the outcomes 1, 2, 3, 4, 5, or 6 is equally likely. A red die and a green die are rolled simultaneously, and the difference of the outcomes (red – green) is computed. This is repeated for a total of 500 rolls of the pair of dice. Which of the following graphs best represents the most reasonable distribution of the differences?

(A)


(B)


(C)


(D)


(E)


largest diff = 6 − 1 = 5

smallest diff = −5

X diff | −5 −4 −3 −2 −1 0 1 2 3 4 5

(not uniform)

more detail: w/ n = 500 each die's value ≈ Normal

R − G = D

1 die | 1 2 3 4 5 6 ; 1/6 1/6 1/6 1/6 1/6 1/6

using L1, L2 1VarStats

$\mu_R = \mu_G = 3.5$

$\sigma_R = \sigma_G = 1.7078$

$\mu_D = \mu_R - \mu_G = 3.5 - 3.5 = 0$

$\sigma_D^2 = \sigma_R^2 + \sigma_G^2 = (1.7078)^2 + (1.7078)^2$

$\sigma_D = 2.415$

GO ON TO THE NEXT PAGE.

40. The probability of winning a certain game is 0.5. If at least 70 percent of the games in a series of n games are won, the player wins a prize. If the possible choices for n are

n=10, n=20, and n=100,

which value of n should the player choose in order to maximize the probability of winning a prize?

(A) n=10 only

(B) n=20 only

(C) n=100 only

(D) n=10 or n=20 only; the probabilities are the same.

(E) n=10 or n=20 or n=100; the probabilities are the same.

binomial model

n=10 X 0 1 2 3 ... 7 ... 10

1 − binomcdf(10, .5, 6) = .1719

n=20 X 0 ... 13, 14 ... 20

1 − binomcdf(20, .5, 13) = .0576

n=100 X 0 ... 69, 70 ... 100

1 − binomcdf(100, .5, 69) = .00039

END OF SECTION I

IF YOU FINISH BEFORE TIME IS CALLED, YOU MAY CHECK YOUR WORK ON THIS SECTION.

DO NOT GO ON TO SECTION II UNTIL YOU ARE TOLD TO DO SO.

MAKE SURE YOU HAVE DONE THE FOLLOWING.

- PLACED YOUR AP NUMBER LABEL ON YOUR ANSWER SHEET
- WRITTEN AND GRIDDED YOUR AP NUMBER CORRECTLY ON YOUR ANSWER SHEET
- TAKEN THE AP EXAM LABEL FROM THE FRONT OF THIS BOOKLET AND PLACED IT ON YOUR ANSWER SHEET

Answer Key for AP Statistics
Practice Exam, Section I

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

*Item 14 was not used in scoring.