Chapter II The 2002 AP Statistics Examination

Exam Content and Format

The 2002 AP Statistics Exam contained questions from all four major content areas — Exploring Data, Planning a Study, Probability, and Statistical Inference — in both the 40-question multiple-choice section and the 6-question free-response section. Each part contributed 50 percent to the total exam score.

2002 AP Statistics Exam Format	
Multiple-Choice (Section I)	
40 questions	90 minutes
Free-Response (Section II)	
Part A	
5 questions	65 minutes
Part B	
1 question	25 minutes

The six free-response questions covered the content areas of data exploration, sampling and experimental design, probability, statistical inference, and fitting models to data. The first five were short-answer questions; the sixth was a longer investigative task worth 25 percent of the maximum possible score for the free-response section.

Question 1: The purpose of question 1 was to assess the student's ability to interpret a graphical display and to reason using interval estimates. To receive full credit for this question, the student needed to comment on the increasing precision of the estimates over time and to use the intervals in the graphical display to discuss the degree of support provided for each of the two competing theories. Good communication was particularly important on this question.

Question 2: The purpose of question 2 was to assess understanding of some basic principles of experimental design, including pairing/blocking, randomization, and blinding. A response describing a matched-pairs design with appropriate randomization and a correct discussion of double blinding received full credit for this question.

Question 3: The purpose of question 3 was to evaluate the student's ability to compute probabilities based on the normal distribution and to evaluate the student's knowledge of properties of the distribution of a sum of independent random variables. To receive full credit for this question, the student was required to compute and interpret a probability, correctly determine the mean and standard deviation of a sum of independent random variables, and then use the computed mean and standard deviation to correctly compute a second probability.

Question 4: The purpose of question 4 was to determine if the student could read standard statistics computer output and to assess understanding of correlation and influential points in a regression analysis. To receive full credit for this question, the response had to include the correct equation for the least squares regression line, a correct interpretation of the correlation coefficient, and a discussion of whether it was reasonable to use the given line over a restricted range of airplane sizes.

Question 5: The purpose of question 5 was to evaluate whether the student could carry out a test of hypotheses and state conclusions in context. To receive full credit for this question, the student needed to identify two distinct pairs of hypotheses in part (a) and then in part (b) to identify an appropriate test procedure, check (not just state) any necessary conditions for the test, and then, based on the result of the test, give an appropriate conclusion in context.

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

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

1. Which of the following is a key distinction between well designed experiments and observational studies?

(A) More subjects are available for experiments than for observational studies.
(B) Ethical constraints prevent large-scale observational studies.
(C) Experiments are less costly to conduct than observational studies.
(D) An experiment can show a direct cause-and-effect relationship, whereas an observational study cannot.
(E) Tests of significance cannot be used on data collected from an observational study.

---

2. A manufacturer of balloons claims that p, the proportion of its balloons that burst when inflated to a diameter of up to 12 inches, is no more than 0.05. Some customers have complained that the balloons are bursting more frequently. If the customers want to conduct an experiment to test the manufacturer's claim, which of the following hypotheses would be appropriate?

(A) $H_0: p \neq 0.05, H_a: p = 0.05$
(B) $H_0: p = 0.05, H_a: p > 0.05$
(C) $H_0: p = 0.05, H_a: p \neq 0.05$
(D) $H_0: p = 0.05, H_a: p < 0.05$
(E) $H_0: p < 0.05, H_a: p = 0.05$

GO ON TO THE NEXT PAGE

3. Lauren is enrolled in a very large college calculus class. On the first exam, the class mean was 75 and the standard deviation was 10. On the second exam, the class mean was 70 and the standard deviation was 15. Lauren scored 85 on both exams. Assuming the scores on each exam were approximately normally distributed, on which exam did Lauren score better relative to the rest of the class?

(A) She scored much better on the first exam.
(B) She scored much better on the second exam.
(C) She scored about equally well on both exams.
(D) It is impossible to tell because the class size is not given.
(E) It is impossible to tell because the correlation between the two sets of exam scores is not given.

---

4. Suppose that 30 percent of the subscribers to a cable television service watch the shopping channel at least once a week. You are to design a simulation to estimate the probability that none of five randomly selected subscribers watches the shopping channel at least once a week. Which of the following assignments of the digits 0 through 9 would be appropriate for modeling an individual subscriber's behavior in this simulation?

(A) Assign "0, 1, 2" as watching the shopping channel at least once a week and "3, 4, 5, 6, 7, 8, and 9" as not watching.
(B) Assign "0, 1, 2, 3" as watching the shopping channel at least once a week and "4, 5, 6, 7, 8, and 9" as not watching.
(C) Assign "1, 2, 3, 4, 5" as watching the shopping channel at least once a week and "6, 7, 8, 9, and 0" as not watching.
(D) Assign "0" as watching the shopping channel at least once a week and "1, 2, 3, 4, and 5" as not watching; ignore digits "6, 7, 8, and 9."
(E) Assign "3" as watching the shopping channel at least once a week and "0, 1, 2, 4, 5, 6, 7, 8, and 9" as not watching.

5. The number of sweatshirts a vendor sells daily has the following probability distribution.

Number of Sweatshirts x	0	1	2	3	4	5
$P(x)$	0.3	0.2	0.3	0.1	0.08	0.02

If each sweatshirt sells for $25, what is the expected daily total dollar amount taken in by the vendor from the sale of sweatshirts?

(A) $5.00

(B) $7.60

(C) $35.50

(D) $38.00

(E) $75.00

---

6. The correlation between two scores X and Y equals 0.8. If both the X scores and the Y scores are converted to z-scores, then the correlation between the z-scores for X and the z-scores for Y would be

(A) −0.8

(B) −0.2

(C) 0.0

(D) 0.2

(E) 0.8

7. Suppose that the distribution of a set of scores has a mean of 47 and a standard deviation of 14. If 4 is added to each score, what will be the mean and the standard deviation of the distribution of new scores?

	Mean	Standard Deviation
(A)	51	14
(B)	51	18
(C)	47	14
(D)	47	16
(E)	47	18

---

8. A test engineer wants to estimate the mean gas mileage μ (in miles per gallon) for a particular model of automobile. Eleven of these cars are subjected to a road test, and the gas mileage is computed for each car.

A dotplot of the 11 gas-mileage values is roughly symmetrical and has no outliers. The mean and standard deviation of these values are 25.5 and 3.01, respectively. Assuming that these 11 automobiles can be considered a simple random sample of cars of this model, which of the following is a correct statement?

(A) A 95% confidence interval for μ is $25.5 \pm 2.228 \times \frac{3.01}{\sqrt{11}}$.

(B) A 95% confidence interval for μ is $25.5 \pm 2.201 \times \frac{3.01}{\sqrt{11}}$.

(C) A 95% confidence interval for μ is $25.5 \pm 2.228 \times \frac{3.01}{\sqrt{10}}$.

(D) A 95% confidence interval for μ is $25.5 \pm 2.201 \times \frac{3.01}{\sqrt{10}}$.

(E) The results cannot be trusted; the sample is too small.

9. A volunteer for a mayoral candidate's campaign periodically conducts polls to estimate the proportion of people in the city who are planning to vote for this candidate in the upcoming election. Two weeks before the election, the volunteer plans to double the sample size in the polls. The main purpose of this is to

(A) reduce nonresponse bias
(B) reduce the effects of confounding variables
(C) reduce bias due to the interviewer effect
(D) decrease the variability in the population
(E) decrease the standard deviation of the sampling distribution of the sample proportion

---

10. The lengths of individual shellfish in a population of 10,000 shellfish are approximately normally distributed with mean 10 centimeters and standard deviation 0.2 centimeter. Which of the following is the shortest interval that contains approximately 4,000 shellfish lengths?

(A) 0 cm to 9.949 cm

(B) 9.744 cm to 10 cm

(C) 9.744 cm to 10.256 cm

(D) 9.895 cm to 10.105 cm

(E) 9.9280 cm to 10.080 cm

GO ON TO THE NEXT PAGE

11. The following two-way table resulted from classifying each individual in a random sample of residents of a small city according to level of education (with categories "earned at least a high school diploma" and "did not earn a high school diploma") and employment status (with categories "employed full time" and "not employed full time").

	Employed full time	Not employed full time	Total
Earned at least a high school diploma	52	40	92
Did not earn a high school diploma	30	35	65
Total	82	75	157

If the null hypothesis of no association between level of education and employment status is true, which of the following expressions gives the expected number who earned at least a high school diploma and who are employed full time?

(A) $\frac{92 \cdot 52}{157}$

(B) $\frac{92 \cdot 82}{157}$

(C) $\frac{82 \cdot 52}{92}$

(D) $\frac{65 \cdot 52}{92}$

(E) $\frac{92 \cdot 52}{82}$

GO ON TO THE NEXT PAGE

12. The manager of a factory wants to compare the mean number of units assembled per employee in a week for two new assembly techniques. Two hundred employees from the factory are randomly selected and each is randomly assigned to one of the two techniques. After teaching 100 employees one technique and 100 employees the other technique, the manager records the number of units each of the employees assembles in one week. Which of the following would be the most appropriate inferential statistical test in this situation?

(A) One-sample z-test

(B) Two-sample t-test

(C) Paired t-test

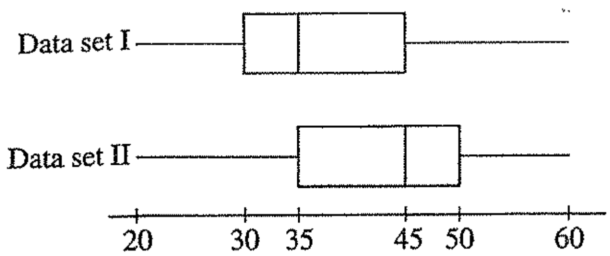
(D) Chi-square goodness-of-fit test

(E) One-sample t-test

13. A random sample has been taken from a population. A statistician, using this sample, needs to decide whether to construct a 90 percent confidence interval for the population mean or a 95 percent confidence interval for the population mean. How will these intervals differ?

(A) The 90 percent confidence interval will not be as wide as the 95 percent confidence interval.
(B) The 90 percent confidence interval will be wider than the 95 percent confidence interval.
(C) Which interval is wider will depend on how large the sample is.
(D) Which interval is wider will depend on whether the sample is unbiased.
(E) Which interval is wider will depend on whether a z-statistic or a t-statistic is used.

GO ON TO THE NEXT PAGE



14. The boxplots shown above summarize two data sets, I and II. Based on the boxplots, which of the following statements about these two data sets CANNOT be justified?

(A) The range of data set I is equal to the range of data set II.

(B) The interquartile range of data set I is equal to the interquartile range of data set II.

(C) The median of data set I is less than the median of data set II.

(D) Data set I and data set II have the same number of data points.

(E) About 75% of the values in data set II are greater than or equal to about 50% of the values in data set I.

GO ON TO THE NEXT PAGE

15. A high school statistics class wants to conduct a survey to determine what percentage of students in the school would be willing to pay a fee for participating in after-school activities. Twenty students are randomly selected from each of the freshman, sophomore, junior, and senior classes to complete the survey. This plan is an example of which type of sampling?

(A) Cluster

(B) Convenience

(C) Simple random

(D) Stratified random

(E) Systematic

---

16. Jason wants to determine how age and gender are related to political party preference in his town. Voter registration lists are stratified by gender and age-group. Jason selects a simple random sample of 50 men from the 20 to 29 age-group and records their age, gender, and party registration (Democratic, Republican, neither). He also selects an independent simple random sample of 60 women from the 40 to 49 age-group and records the same information. Of the following, which is the most important observation about Jason's plan?

(A) The plan is well conceived and should serve the intended purpose.
(B) His samples are too small.
(C) He should have used equal sample sizes.
(D) He should have randomly selected the two age groups instead of choosing them nonrandomly.
(E) He will be unable to tell whether a difference in party affiliation is related to differences in age or to the difference in gender.

GO ON TO THE NEXT PAGE

17. A least squares regression line was fitted to the weights (in pounds) *versus* age (in months) of a group of many young children. The equation of the line is

$$\hat{y} = 16.6 + 0.65t,$$

where $\hat{y}$ is the predicted weight and t is the age of the child. A 20-month-old child in this group has an actual weight of 25 pounds. Which of the following is the residual weight, in pounds, for this child?

(A) −7.85

(B) −4.60

(C) 4.60

(D) 5.00

(E) 7.85

---

18. Which of the following statements is (are) true about the t-distribution with k degrees of freedom?

I. The t-distribution is symmetric.
II. The t-distribution with k degrees of freedom has a smaller variance than the t-distribution with $k + 1$ degrees of freedom.
III. The t-distribution has a larger variance than the standard normal (z) distribution.

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

GO ON TO THE NEXT PAGE

Brown Eyes	Green Eyes	Blue Eyes
34	15	11

19. A geneticist hypothesizes that half of a given population will have brown eyes and the remaining half will be split evenly between blue- and green-eyed people. In a random sample of 60 people from this population, the individuals are distributed as shown in the table above. What is the value of the χ^2 statistic for the goodness of fit test on these data?

(A) Less than 1

(B) At least 1, but less than 10

(C) At least 10, but less than 20

(D) At least 20, but less than 50

(E) At least 50

---

20. A small town employs 34 salaried, nonunion employees. Each employee receives an annual salary increase of between \$500 and \$2,000 based on a performance review by the mayor's staff. Some employees are members of the mayor's political party, and the rest are not.

Students at the local high school form two lists, A and B, one for the raises granted to employees who are in the mayor's party, and the other for raises granted to employees who are not. They want to display a graph (or graphs) of the salary increases in the student newspaper that readers can use to judge whether the two groups of employees have been treated in a reasonably equitable manner.

Which of the following displays is <u>least</u> likely to be useful to readers for this purpose?

(A) Back-to-back stemplots of A and B

(B) Scatterplot of B *versus* A

(C) Parallel boxplots of A and B

(D) Histograms of A and B that are drawn to the same scale

(E) Dotplots of A and B that are drawn to the same scale

21. In a study of the performance of a computer printer, the size (in kilobytes) and the printing time (in seconds) for each of 22 small text files were recorded. A regression line was a satisfactory description of the relationship between size and printing time. The results of the regression analysis are shown below.

Dependent variable: Printing Time

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	53.3315	1	53.3315	140
Residual	7.62381	20	0.38115	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	11.6559	0.3153	37	≤ 0.0001
Size	3.47812	0.294	11.8	≤ 0.0001

R squared = 87.5%   R squared (adjusted) = 86.9%

s = 0.6174 with 22 − 2 = 20 degrees of freedom

Which of the following should be used to compute a 95 percent confidence interval for the slope of the regression line?

(A) $3.47812 \pm 2.086 \times 0.294$

(B) $3.47812 \pm 1.96 \times 0.6174$

(C) $3.47812 \pm 1.725 \times 0.294$

(D) $11.6559 \pm 2.086 \times 0.3153$

(E) $11.6559 \pm 1.725 \times 0.3153$

GO ON TO THE NEXT PAGE

22. A study of existing records of 27,000 automobile accidents involving children in Michigan found that about 10 percent of children who were wearing a seatbelt (group SB) were injured and that about 15 percent of children who were not wearing a seatbelt (group NSB) were injured. Which of the following statements should NOT be included in a summary report about this study?

(A) Driver behavior may be a potential confounding factor.

(B) The child's location in the car may be a potential confounding factor.

(C) This study was not an experiment, and cause-and-effect inferences are not warranted.

(D) This study demonstrates clearly that seat belts save children from injury.

(E) Concluding that seatbelts save children from injury is risky, at least until the study is independently replicated.

---

23. Which of the following statements is true for two events, each with probability greater than 0 ?

(A) If the events are mutually exclusive, they must be independent.

(B) If the events are independent, they must be mutually exclusive.

(C) If the events are not mutually exclusive, they must be independent.

(D) If the events are not independent, they must be mutually exclusive.

(E) If the events are mutually exclusive, they cannot be independent.

GO ON TO THE NEXT PAGE

24. A consulting statistician reported the results from a learning experiment to a psychologist. The report stated that on one particular phase of the experiment a statistical test result yielded a p-value of 0.24. Based on this p-value, which of the following conclusions should the psychologist make?

(A) The test was statistically significant because a p-value of 0.24 is greater than a significance level of 0.05.
(B) The test was statistically significant because $p = 1 - 0.24 = 0.76$ and this is greater than a significance level of 0.05.
(C) The test was not statistically significant because 2 times $0.24 = 0.48$ and that is less than 0.5.
(D) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 24% of the time.
(E) The test was not statistically significant because, if the null hypothesis is true, one could expect to get a test statistic at least as extreme as that observed 76% of the time.

---

25. A new medication has been developed to treat sleep-onset insomnia (difficulty in falling asleep). Researchers want to compare this drug to a drug that has been used in the past by comparing the length of time it takes subjects to fall asleep. Of the following, which is the best method for obtaining this information?

(A) Have subjects choose which drug they are willing to use, then compare the results.

(B) Assign the two drugs to the subjects on the basis of their past sleep history without randomization, then compare the results.

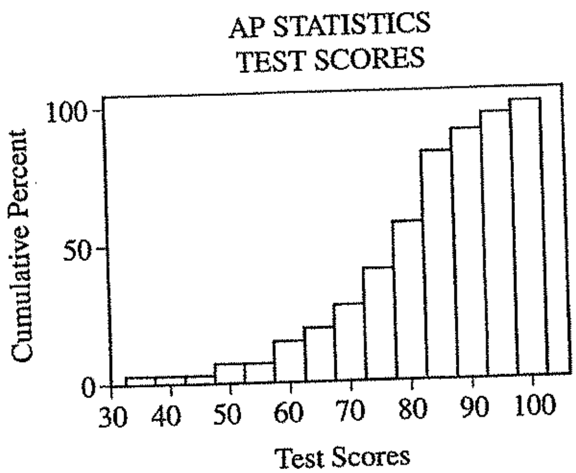
(C) Give the new drug to all subjects on the first night. Give the old drug to all subjects on the second night. Compare the results.

(D) Randomly assign the subjects to two groups, giving the new drug to one group and no drug to the other group, then compare the results.

(E) Randomly assign the subjects to two groups, giving the new drug to one group and the old drug to the other group, then compare the results.

26. A quality control inspector must verify whether a machine that packages snack foods is working correctly. The inspector will randomly select a sample of packages and weigh the amount of snack food in each. Assume that the weights of food in packages filled by this machine have a standard deviation of 0.30 ounce. An estimate of the mean amount of snack food in each package must be reported with 99.6 percent confidence and a margin of error of no more than 0.12 ounce. What would be the minimum sample size for the number of packages the inspector must select?

(A) 8
(B) 15
(C) 25
(D) 52
(E) 60



27. The figure above shows a cumulative relative frequency histogram of 40 scores on a test given in an AP Statistics class. Which of the following conclusions can be made from the graph?

(A) There is greater variability in the lower 20 test scores than in the higher 20 test scores.

(B) The median test score is less than 50.

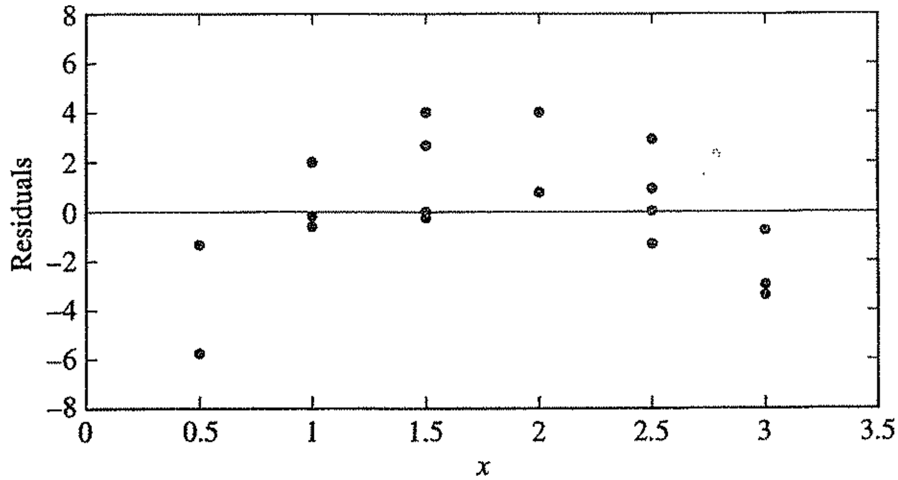
(C) Sixty percent of the students had test scores above 80.

(D) If the passing score is 70, most students did not pass the test.

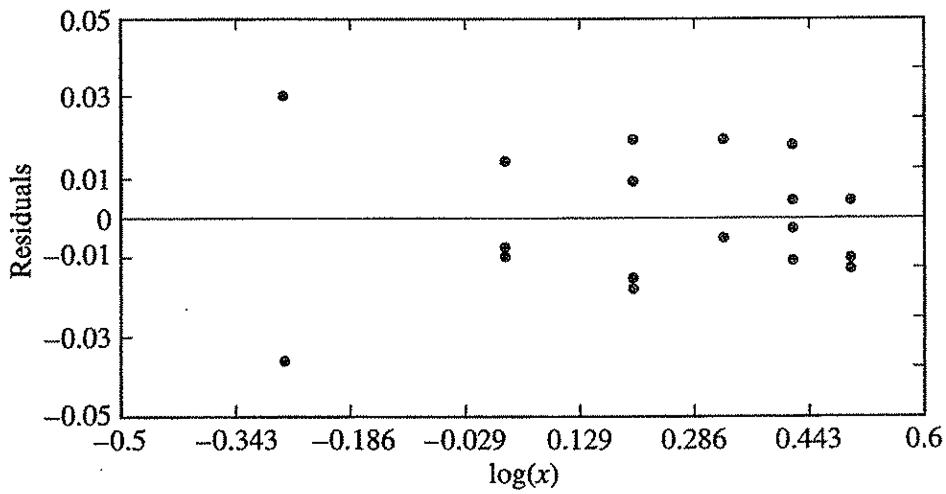
(E) The horizontal nature of the graph for test scores of 60 and below indicates that those scores occurred most frequently.

GO ON TO THE NEXT PAGE

28. Two measures x and y were taken on 18 subjects. The first of two regressions, Regression I, yielded $\hat{y} = 24.5 + 16.1x$ and had the following residual plot.



The second regression, Regression II, yielded $\widehat{\log(y)} = 1.6 + 0.51 \log(x)$ and had the following residual plot.



Which of the following conclusions is best supported by the evidence above?

(A) There is a linear relationship between x and y, and Regression I yields a better fit.

(B) There is a linear relationship between x and y, and Regression II yields a better fit.

(C) There is a negative correlation between x and y.

(D) There is a nonlinear relationship between x and y, and Regression I yields a better fit.

(E) There is a nonlinear relationship between x and y, and Regression II yields a better fit.

29. The analysis of a random sample of 500 households in a suburb of a large city indicates that a 98 percent confidence interval for the mean family income is ($41,300, $58,630). Could this information be used to conduct a test of the null hypothesis H_0: $\mu = 40{,}000$ against the alternative hypothesis H_a: $\mu \neq 40{,}000$ at the $\alpha = 0.02$ level of significance?

(A) No, because the value of σ is not known.

(B) No, because it is not known whether the data are normally distributed.

(C) No, because the entire data set is needed to do this test.

(D) Yes, since $40,000 is not contained in the 98 percent confidence interval, the null hypothesis would be rejected in favor of the alternative, and it could be concluded that the mean family income is significantly different from $40,000 at the $\alpha = 0.02$ level.

(E) Yes, since $40,000 is not contained in the 98 percent confidence interval, the null hypothesis would not be rejected, and it could be concluded that the mean family income is not significantly different from $40,000 at the $\alpha = 0.02$ level.

---

30. The population {2, 3, 5, 7} has mean $\mu = 4.25$ and standard deviation $\sigma = 1.92$. When sampling with replacement, there are 16 different possible ordered samples of size 2 that can be selected from this population. The mean of each of these 16 samples is computed. For example, 1 of the 16 samples is (2, 5), which has a mean of 3.5. The distribution of the 16 sample means has its own mean $\mu_{\bar{x}}$ and its own standard deviation $\sigma_{\bar{x}}$. Which of the following statements is true?

(A) $\mu_{\bar{x}} = 4.25$ and $\sigma_{\bar{x}} = 1.92$
(B) $\mu_{\bar{x}} = 4.25$ and $\sigma_{\bar{x}} > 1.92$
(C) $\mu_{\bar{x}} = 4.25$ and $\sigma_{\bar{x}} < 1.92$
(D) $\mu_{\bar{x}} > 4.25$
(E) $\mu_{\bar{x}} < 4.25$

GO ON TO THE NEXT PAGE

31. A wildlife biologist is interested in the relationship between the number of chirps per minute for crickets (y) and temperature. Based on the collected data, the least squares regression line is $\hat{y} = 10.53 + 3.41x$, where x is the number of degrees Fahrenheit by which the temperature exceeds 50°. Which of the following best describes the meaning of the slope of the least squares regression line?

(A) For each increase in temperature of 1° F, the estimated number of chirps per minute increases by 10.53.

(B) For each increase in temperature of 1° F, the estimated number of chirps per minute increases by 3.41.

(C) For each increase of one chirp per minute, there is an estimated increase in temperature of 10.53° F.

(D) For each increase of one chirp per minute, there is an estimated increase in temperature of 3.41° F.

(E) The slope has no meaning because the units of measure for x and y are not the same.

---

32. In a carnival game, a person can win a prize by guessing which one of 5 identical boxes contains the prize. After each guess, if the prize has been won, a new prize is randomly placed in one of the 5 boxes. If the prize has not been won, then the prize is again randomly placed in one of the 5 boxes. If a person makes 4 guesses, what is the probability that the person wins a prize exactly 2 times?

(A) $\frac{2!}{5!}$

(B) $\frac{(0.2)^2}{(0.8)^2}$

(C) $2(0.2)(0.8)$

(D) $(0.2)^2(0.8)^2$

(E) $\binom{4}{2}(0.2)^2(0.8)^2$

GO ON TO THE NEXT PAGE

33. An engineer for the Allied Steel Company has the responsibility of estimating the mean carbon content of a particular day's steel output, using a random sample of 15 rods from that day's output. The actual population distribution of carbon content is not known to be normal, but graphic displays of the engineer's sample results indicate that the assumption of normality is not unreasonable. The process is newly developed, and there are no historical data on the variability of the process. In estimating this day's mean carbon content, the primary reason the engineer should use a t-confidence interval rather than a z-confidence interval is because the engineer

(A) is estimating the population mean using the sample mean
(B) is using the sample variance as an estimate of the population variance
(C) is using data, rather than theory, to judge that the carbon content is normal
(D) is using data from a specific day only
(E) has a small sample, and a z-confidence interval should never be used with a small sample

---

34. Each of 100 laboratory rats has available both plain water and a mixture of water and caffeine in their cages. After 24 hours, two measures were recorded for each rat: the amount of caffeine the rat consumed, X, and the rat's blood pressure, Y. The correlation between X and Y was 0.428. Which of the following conclusions is justified on the basis of this study?

(A) The correlation between X and Y in the population of rats is also 0.428.
(B) If the rats stop drinking the water/caffeine mixture, this would cause a reduction in their blood pressure.
(C) About 18 percent of the variation in blood pressure can be explained by a linear relationship between blood pressure and caffeine consumed.
(D) Rats with lower blood pressure do not like the water/caffeine mixture as much as do rats with higher blood pressure.
(E) Since the correlation is not very high, the relationship between the amount of caffeine consumed and blood pressure is not linear.

---

35. In a test of the hypothesis $H_0: \mu = 100$ *versus* $H_a: \mu > 100$, the power of the test when $\mu = 101$ would be greatest for which of the following choices of sample size n and significance level α?

(A) $n = 10$, $\alpha = 0.05$

(B) $n = 10$, $\alpha = 0.01$

(C) $n = 20$, $\alpha = 0.05$

(D) $n = 20$, $\alpha = 0.01$

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

GO ON TO THE NEXT PAGE

36. An urn contains exactly three balls numbered 1, 2, and 3, respectively. Random samples of two balls are drawn from the urn with replacement. The average, $\overline{X} = \frac{X_1 + X_2}{2}$, where X_1 and X_2 are the numbers on the selected balls, is recorded after each drawing. Which of the following describes the sampling distribution of $\overline{X}$?

(A)

$\overline{X}$	1	1.5	2	2.5	3
Probability	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

(B)

$\overline{X}$	1	1.5	2	2.5	3
Probability	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

(C)

$\overline{X}$	1	1.5	2	2.5	3
Probability	0	0	1	0	0

(D)

$\overline{X}$	1	1.5	2	2.5	3
Probability	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{3}{5}$	$\frac{1}{10}$	$\frac{1}{10}$

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

---

37. A simple random sample produces a sample mean, $\bar{x}$, of 15. A 95 percent confidence interval for the corresponding population mean is 15 ± 3. Which of the following statements must be true?

(A) Ninety-five percent of the population measurements fall between 12 and 18.

(B) Ninety-five percent of the sample measurements fall between 12 and 18.

(C) If 100 samples were taken, 95 of the sample means would fall between 12 and 18.

(D) $P(12 \le \bar{x} \le 18) = 0.95$

(E) If $\mu = 19$, this $\bar{x}$ of 15 would be unlikely to occur.

GO ON TO THE NEXT PAGE

38. Suppose that public opinion in a large city is 65 percent in favor of increasing taxes to support the public school system and 35 percent against such an increase. If a random sample of 500 people from this city are interviewed, what is the approximate probability that more than 200 of these people will be against increasing taxes?

(A) $\binom{500}{200}(0.65)^{200}(0.35)^{300}$

(B) $\binom{500}{200}(0.35)^{200}(0.65)^{300}$

(C) $P\left(z > \dfrac{0.40-0.65}{\sqrt{\dfrac{(0.65)(0.35)}{500}}}\right)$

(D) $P\left(z > \dfrac{0.40-0.35}{\sqrt{\dfrac{(0.4)(0.6)}{500}}}\right)$

(E) $P\left(z > \dfrac{0.40-0.35}{\sqrt{\dfrac{(0.35)(0.65)}{500}}}\right)$

GO ON TO THE NEXT PAGE

39. As lab partners, Sally and Betty collected data for a significance test. Both calculated the same z-test statistic, but Sally found the results were significant at the $\alpha = 0.05$ level while Betty found that the results were not. When checking their results, the women found that the only difference in their work was that Sally had used a two-sided test, while Betty used a one-sided test. Which of the following could have been their test statistic?

(A) −1.980

(B) −1.690

(C) 1.340

(D) 1.690

(E) 1.780

---

40. A student working on a history project decided to find a 95 percent confidence interval for the difference in mean age at the time of election to office for former American Presidents versus former British Prime Ministers. The student found the ages at the time of election to office for the members of both groups, which included all of the American Presidents and all of the British Prime Ministers, and used a calculator to find the 95 percent confidence interval based on the t-distribution. This procedure is not appropriate in this context because

(A) the sample sizes for the two groups are not equal

(B) the entire population was measured in both cases, so the actual difference in means can be computed and a confidence interval should not be used

(C) elections to office take place at different intervals in the two countries, so the distribution of ages cannot be the same

(D) ages at the time of election to office are likely to be skewed rather than bell-shaped, so the assumptions for using this confidence interval formula are not valid

(E) ages at the time of election to office are likely to have a few large outliers, so the assumptions for using this confidence interval formula are not valid

END OF SECTION I

Table 4.2 — AP Statistics Scoring Worksheet

Section I: Multiple Choice

[________ − (1/4 × ________)] × 1.2500 = ________ = ________

Number correct (out of 40) — Number wrong — Multiple-Choice Score (If less than zero, enter zero.) — Weighted Section I Score (Do not round)

Section II: Free Response

Question 1 ________ (out of 4) × 1.8750 = ________ (Do not round)

Question 2 ________ (out of 4) × 1.8750 = ________ (Do not round)

Question 3 ________ (out of 4) × 1.8750 = ________ (Do not round)

Question 4 ________ (out of 4) × 1.8750 = ________ (Do not round)

Question 5 ________ (out of 4) × 1.8750 = ________ (Do not round)

Question 6 ________ (out of 4) × 3.1250 = ________ (Do not round)

Sum = ________ Weighted Section II Score (Do not round)

AP Grade Conversion Chart Statistics

Composite Score Range*	AP Grade
68-100	5
53-67	4
40-52	3
29-39	2
0-28	1

*The students' scores are weighted according to formulas determined in advance each year by the Development Committee to yield raw composite scores; the Chief Reader is responsible for converting composite scores to the 5-point AP scale.

Composite Score

________ + ________ = ________

Weighted Section I Score + Weighted Section II Score = Composite Score (Round to nearest whole number.)

Chapter III Answers to the 2002 AP Statistics Exam

Section I: Multiple Choice

Listed below are the correct answers to the multiple-choice questions, the percentage of AP students who answered each question correctly by AP grade, and the total percentage answering correctly.

Section I Answer Key and Percent Answering Correctly

Item No.	Correct Answer	Percent Correct by Grade 5	4	3	2	1	Total Percent Correct
1	D	97	93	90	86	78	88
2	B	97	92	86	79	62	81
3	C	97	91	85	76	56	79
4	A	99	97	92	83	59	84
5	D	99	97	90	79	53	81
6	E	89	76	66	55	37	62
7	A	99	96	91	83	62	84
8	A	88	73	59	46	33	56
9	E	77	54	37	27	20	39
10	D	76	51	35	26	19	37
11	B	95	86	73	59	41	68
12	B	85	73	67	63	58	67
13	A	94	85	75	64	46	70
14	D	99	96	89	79	51	80
15	D	97	94	88	79	56	81
16	E	98	95	91	86	65	85
17	B	88	78	64	50	28	58
18	E	66	40	26	19	15	29
19	B	91	78	63	48	31	59
20	B	91	76	59	46	30	57

Item No.	Correct Answer	Percent Correct by Grade 5	4	3	2	1	Total Percent Correct
21	A	79	51	32	20	14	34
22	D	94	88	80	69	46	73
23	E	43	26	19	16	14	21
24	D	95	82	65	49	27	60
25	E	93	92	91	90	78	88
26	D	83	65	49	37	23	48
27	A	65	46	33	26	18	34
28	E	93	76	53	37	22	52
29	D	91	79	64	51	34	61
30	C	72	48	33	23	18	35
31	B	98	94	86	78	53	80
32	E	87	71	58	45	29	55
33	B	78	54	37	25	18	38
34	C	82	54	28	14	10	33
35	C	43	22	16	15	15	20
36	B	88	72	57	43	29	54
37	E	34	12	6	3	2	9
38	E	71	48	34	26	18	36
39	A	24	16	13	11	9	13
40	B	88	73	60	47	28	56

2002 AP Statistics Exam Multiple Choice Solutions

1. Answer: (D)

(A) is FALSE → There are not necessarily more subjects available for either experiments or observational studies. There are probably more subjects available for observational studies if anything because people are more apt to be alright with someone observing them than to take place in an experiment.
(B) is FALSE → In anything, ethical constraints might prevent certain large scale experiments from taking place, not observational studies.
(C) is FALSE → On average, experiments tend to be more costly than observational studies.
(D) is TRUE → Only through a well-designed experiment can we determine a causal relationship between our variables whereby in an observational study, we can only determine if there is an association between our variables.
(E) is FALSE → You can do significance tests on any data, whether it comes from an observational study or an experiment.

2. Answer: (B)

Choices (A) and (E) are out from the get-go because the null hypothesis must have $p = 0.05$ in it. Since we are checking to see if the manufacturer is wrong (and the manufacturer says that it is "no more than 0.05"), we must be testing for the idea that the proportion is more than 5% or H_a: $p > 0.05$.

Note: Our parameter of interest in this example is p, the true proportion of balloons that burst when inflated to a diameter above 12 inches. This is because this is categorical data…the balloons either burst or don't.

3. Answer: (C)

If you want to see how these Lauren performed in a relative sense and the scores follow a normal distribution, then you may use z scores:

1st exam: $z = \dfrac{x-\mu}{\sigma} = \dfrac{85-75}{10} = 1$ | 2nd exam: $z = \dfrac{x-\mu}{\sigma} = \dfrac{85-70}{15} = 1$

Since the z scores are the same [both times she scored exactly one standard deviation above the rest of the class], she scored about equally as well on both exams. The class size and correlation between the two scores is irrelevant for this question.

4. Answer: (A)

Recall that when you are setting up a simulation you must **disregard the values of the numbers** and simply treat them as non-numerical objects. Since 30% of all subscribers watch the shopping channel at least once a week, make 30% of the 10 numbers represent those people and the other 70% of the 10 numbers represent the other people. The numbers do not represent quantities, they only represent outcomes.

5. Answer: (D)

There are a couple different ways you can do this:

Method 1: Figure out the expected number of sweatshirts that are sold:

$$E(x) = \mu = \sum x_i p(x_i) = (0*0.3)+(1*0.2)+(2*0.3)+(3*0.1)+(4*0.08)+(5*0.02) = 1.52$$

Now since each sweatshirt is sold for $25, they should expect to make:

$$\$25 \times 1.52 = \$38$$

Method 2: Convert the original data from number of sweatshirts to amount of money earned:

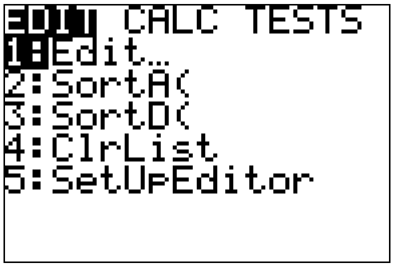
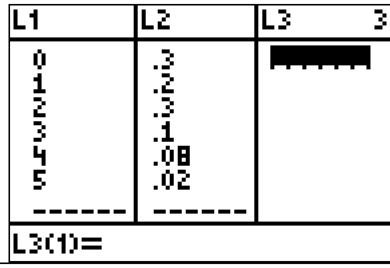
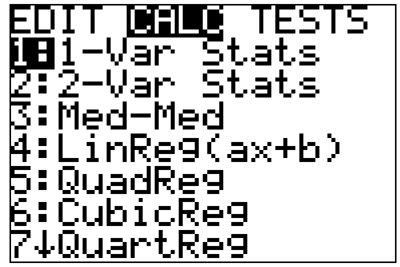
Money per sweatshirt *x*	$0	$25	$50	$75	$100	$125
P(*x*)	0.3	0.2	0.3	0.1	0.08	0.02

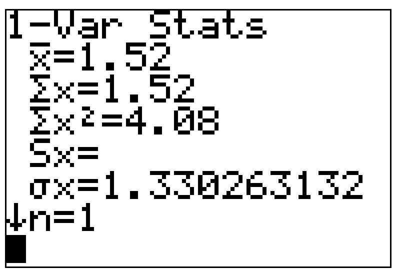
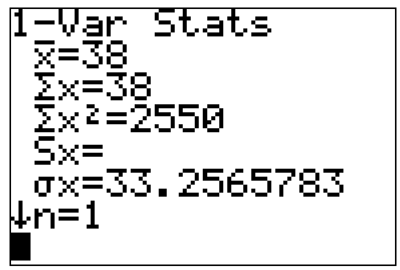
Now do the same thing to find the expected revenue:

$$E(x) = (0*0.3)+(25*0.2)+(50*0.3)+(75*0.1)+(100*0.08)+(125*0.02) = \$38.00$$

Keep in mind, though, that you don't have to do these by hand unless it's a free response question, and even then you need only show the formulas for the purpose of presentation.

If it's a multiple choice question like this, just do it in the calculator:

<table>
<tr>
<td>1. Press STAT and press ENTER.

</td>
<td>2. Enter all the x values into L_1 and all the probabilities into L_2

</td>
<td>3. Press STAT → CALC → 1-Var-Stats

</td>
</tr>
<tr>
<td>3. Now tell it to do 1-Var Stats on L_1 with a frequency of L_2.

</td>
<td>4. Press ENTER and the $\bar{x}$ is your mean (expected value) and σ_x is your standard deviation.

</td>
<td>4. You can also do this from the perspective of money… same exact process.

</td>
</tr>
</table>

6. Answer: (E)

Recall that the correlation coefficient (*r*) is not affected by multiplying, adding, subtracting, or dividing (any linear transformation does not affect the correlation coefficient). The only thing that could affect the correlation coefficient was if you were to do some sort of non-linear transformation such as taking the log, taking the square root, or squaring every data value. So it remains the same.

7. Answer: (A)

Since a constant is being added to every data value, the mean will be affected but the standard deviation will not…

$$\mu = 47 + 4 = 51.$$

$$\sigma = 14$$

Now, had the question said that we, say, multiply every number by 2, then both the mean and the standard deviation would be affected:

$$\mu = 47 \times 2 = 94$$

$$\sigma = 14 \times 2 = 28$$

Also, had the question given you the variance and doubled every data value, that would have changed the answer as well. Pretend that the variance is 14 and we multiply every value in the data set by 2. You have to square the constant because the variance is in square units:

$$\text{Variance} = \sigma^2 = 2^2 \times 14 = 56$$

8. Answer: (A)

Immediately (E) is not the correct answer because even though the sample size is small (n = 11), the question states that the data appears 'unimodal and symmetric' in which case the small sample size is no big deal. Recall that for a one-sample t-interval the degrees of freedom are n – 1. On the t-table, look up the critical value for 95% confidence and n – 1 = 11 – 1 = 10 degrees of freedom and you get 2.228:

df					
10	1.372	1.812	2.228	2.359	2.764
:	:	:	:	:	:
	80%	90%	95%	96%	98%

Confidence Level *C*

The formula one sample t-interval is:

$$\bar{x} \pm t^*_{n-1} \times \frac{s}{\sqrt{n}} \rightarrow 25.5 \pm 2.228 \times \frac{3.01}{\sqrt{11}}$$

9. Answer: (E)

This guy is going to increase the sample size in hopes of fixing one of the following problems:

(A) is FALSE → Non-response bias is when you ask the question to people and they blatantly refuse to respond. The more people you ask, the more who will refuse to respond.
(B) is FALSE → Confounding variables are variables that you did not consider and that affect the response variable, making it difficult to ascertain the true effect. Increasing the sample size does not get rid of these. Only a well-designed experiment can help minimize the effects of confounding variables.
(C) is FALSE → The interviewer effect will not diminish as more people are surveyed. The interviewer effect is when the person conducting the interview (their demeanor, appearance, etc) makes it to where it is difficult to respond honestly. It is a form of response bias (i.e. anything that biases your response)
(D) is FALSE → You can not alter the variability in the population. There is no variability in the population. Variability only refers to sampling.
(E) is TRUE → As you increase the sample size, sampling variability (the degree to which statistics will differ from sample to sample) will decrease always. So, the standard deviation of the sampling distribution of the sample proportion will also decrease. Its formula is as follows:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$$

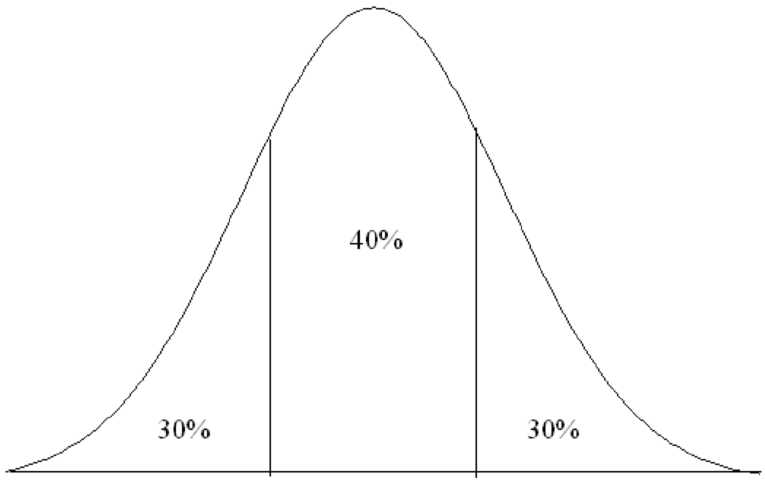
Notice that if we were to increase the sample size, this quantity would invariably decrease.

10. Answer: (D)

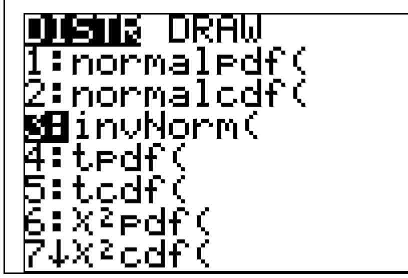
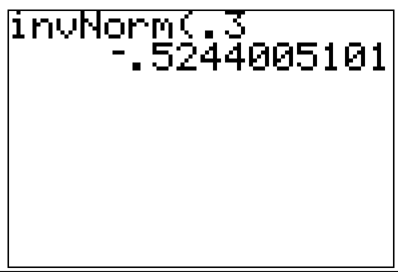
Since this question asks for the ‘shortest distance’ that will still give you 4,000 shellfish, it must be centered about the mean. Also, since there are 10,000 shellfish in the population, the question is asking for the shortest distance that will contain:

$$\frac{4,000}{10,000} = 40\% \text{ of the shellfish}$$

Using the ghetto-looking picture below, you realize that you have the area under the curve and you need the values on the *x*-axis:



In this case, use the TI83 to find both values (although you will notice that you need only one):

1. Press [2nd] → [VARS] → invNorm(	2. The command is invNorm(area to left) = *z* 

Now, just solve for the missing value *x*:

$$z = \frac{x - \mu}{\sigma}$$

$$-0.5244 = \frac{x - 10}{0.2}$$

x = 9.895 cm.

11. Answer: (B)

This is asking for a component of the χ^2 test for independence. Look at wording in the question: “If the null hypothesis of no association between level of education and employment status is true…”. This is asking for the expected value under the assumption of independence. The formula for this value is:

$$\text{Expected} = \frac{\text{Row} \times \text{Column}}{\text{Grand}} = \frac{92 \cdot 82}{157}$$

12. Answer: (B)

1. Since the data is quantitative (we are measuring the <u>mean</u> number of units assembled per employee), the goodness of fit test is irrelevant because that deals with categorical data.

2. Since there are 2 groups, we have to look for pairing. There is no evidence to believe that the experimental units are being paired in any way (no matching in the question).

3. The sampling procedure seems to indicate that the experimental units were sampled independent from one another. So he should use a two-sample *t*-test.

<u>Note</u>: Had there only been one group of employees and you had to distinguish between whether to use a *z*-test or a *t*-test, you have to see if the population standard deviation is known or not. If it is, use a *z*-test. If not, use a *t*-test.

13. Answer: (A)

This question harps at the different qualities of confidence intervals that you have to know for this exam:

(A) is TRUE → As your confidence level goes up, so too does the width of the confidence interval [a 99% CI is wider than a 95% CI which is wider than a90% CI, etc]
(B) is FALSE → It contradicts what I just said for choice (A).
(C) is FALSE → Sample size is not the only thing that affects the width of a confidence interval. Also, the question makes no mention of the sample size. However, if the question had not changed the confidence level, but did change the sample size, a <u>larger</u> sample size will yield a <u>narrower</u> confidence interval and thus more precision.
(D) is FALSE → If your sample is biased, there is no way you should be performing any type of statistical inference. Period.
(E) is FALSE → Whether you use a *z*-interval or a *t*-interval depends on whether you know the population standard deviation.

14. Answer: (D)

Let's go through these one at a time:

(A) is TRUE → Range is measured as max – min and in this case, the maximum and minimum values for both boxplots look about the same.

(B) is TRUE → The IQR is measured as Q3 – Q1:

For Data set I: Q3 – Q1 = 45 – 30 = 15
For Data set II: Q3 – Q1 = 50 – 35 = 15

(C) is TRUE → The median is measured by the center line in the boxplot, which for data set I is at approximately 35 and for data set II is approximately 45. Keep in mind that on a boxplot, this is a measure of the <u>median</u>, not the mean. The only time that this could be considered a measure of the mean would be if the data set looked approximately symmetrical at which time the mean and median would be approximately equal. Also, you can tell from this picture what the shape of each data set would be. Data set I looks slightly skewed right (in which case the mean would be above the median) and data set II looks slightly skewed right (in which case the mean would be less than the median).

(D) is FALSE → We have no idea what the sample sizes are for each one of these data sets. You cannot tell the sample sizes solely from box-and-whisker plots.

(E) is TRUE → The first quartile for data set II is approximately 35. At the first quartile, 75% of all data points are above that value (and 25% are below it). For data set I, the median is at about 35 as well. The median is the place where 50% of the data points are below it and 50% of the data points are above it. So, the statement "about 75% of the values in data set II are greater than about 50% of the data values in data set I" must be true.

15. Answer: (D)

What is the difference between each of the sampling procedures?

(A) **Cluster** sampling is when you take a random group that is already relatively heterogeneous (mixed) on the variable of interest. For example, we just grab my 1st period class as a representative sample of all my stats classes because we felt that they were a good mix on the variable of interest.

(B) **Convenience** sampling is a very bad thing. It is when you solely grab those people that are conveniently located to you. Imagine if I took a survey of all teachers during our weekly meeting to see how all teachers felt about the administration. This would not work too well because I don't know if their opinions are representative of all opinions. It's just <u>convenient</u> for me because they're right there.

(C) **Simple random** sampling is when we simply randomly sample from our population of interest. If we used a random number generator, or put names in a hat, we would be conducting a simple random sample

(D) **Stratified random** sampling is when we break people off into similar groups and then within each group (or strata), we sample people randomly. We want to divide the subjects into strata such that the people are as alike as possible on the variable of interest. This question is an example of stratified random sampling because they broke the class off into freshmen, sophomores, juniors, and seniors and then selected students from each stratum.

(E) **Systematic** sampling is when you take every n^{th} person. If I were to stand in the hall area of a school and sample every, say, 5^{th} person, this would be systematic sampling.

16. Answer: (E)

Jason wants to determine how age and gender are related to political party preference:

(A) is FALSE → This is for bunch of reasons we are about to discuss.
(B) is FALSE → Sample size is not important. It's about quality, not quantity. A good small sample is better than a big crappy sample.
(C) is FALSE → Equal sample sizes are not required for your sample to be representative.
(D) is FALSE → The question says that he did choose the samples randomly. Liars.
(E) is TRUE → If he's trying to see the effect of political party preference on gender and age group, he has done a stupid thing. He has selected two independent samples that differ on both variables. Now, since they are different genders and different ages, he cannot tell if the difference in political preference is due to the difference in age groups or the difference in genders. In other words, they are confounded

17. Answer: (B)

17. A residual is always measured as A – P = Actual – Predicted. Since you are not given a predicted value, you have to plug the *x* value into the equation in order to get the predicted value:

$$\hat{y} = 16.6 + 0.65 \times 20 = 29.6$$

Actual – Predicted = 25 – 29.6 = –4.60.

Also, let's look at the slope and intercept (just for fun):

Slope: We predict, on average, that for each additional month of age, the child gains an additional 0.65 pounds of weight.

Intercept: We predict, on average, that when a baby is zero months old, he will weigh 16.6 pounds. This is actually describing the weight of the baby at birth, so it makes sense contextually. It is possible, though, that the data was not gathered from values that stretched down to zero so it is probably extrapolation to interpret this value.

18. Answer: (E)

Recall that the *t*-distribution is used when the standard deviation for the population is unknown. Here are the properties of the *t*-distribution that you may have to know:

1. It is symmetric and centered at zero
2. It is indexed by degrees of freedom
3. It is shorter and has more area in the tails than the z-distribution
4. A *z*-distribution is a *t*-distribution with infinite degrees of freedom
5. As degrees of freedom increase, the *t*-distribution approaches a *z*-distribution and thus has less variability as these degrees of freedom go up.

So…

I. The *t*-distribution is symmetric → TRUE [see above]

II. The *t*-distribution with *k* degrees of freedom has smaller variance than the *t*-distribution with *k* + 1 degrees of freedom → FALSE [A *t*-distribution with *k* degrees of freedom has <u>more</u> variance than the *t*-distribution with *k* + 1 degrees of freedom]

III. The *t*-distribution has a larger variance than the standard normal (*z*) distribution → TRUE [A *t*-distribution by definition has more variability than a *z* distribution.]

19. Answer: (B)

19. The values given in the question are observed values for the 60 people who were sampled:

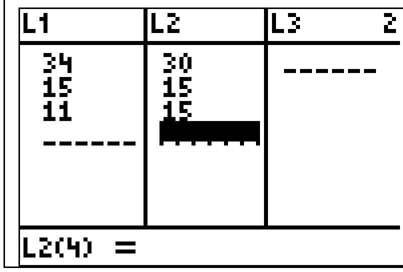
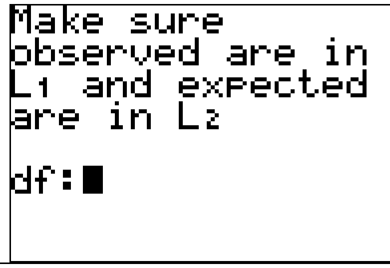
Brown Eyes	Green Eyes	Blue Eyes
34	15	11

It is hypothesized, though, that half of all people have brown eyes and the other half are evenly divided between green eyes and blue eyes. By that accord, the expected counts for these *n* = 60 people are as follows:

Brown Eyes	Green Eyes	Blue Eyes
$60\times 0.5=30$	$60\times 0.25=15$	$60\times 0.25=15$

$$\chi^2=\sum\frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}}=\frac{(34-30)^2}{30}+\frac{(15-15)^2}{15}+\frac{(11-15)^2}{15}=1.6$$

Again, this is one of those questions you can do in the calculator. Keep in mind, though, that the goodness of fit test will <u>not</u> work if you try to put the data in a matrix and perform it like you would the tests for independence and homogeneity. You have to do this using the program I gave you:

1. Make sure all observed values are in L_1 and all expected values in L_2	2. Press PRGM and run the CHIGOF program	3. Let it know that there are 2 degrees of freedom (#categories – 1) and you're done!
		

20. Answer: (B)

Since we are attempting to <u>compare</u> two distributions, you have to be careful what you decide is/isn't an "appropriate" display:

(A) **Back-to-back stem and leaf plots** would serve nicely since we are trying to compare the distributions of these data values. This way, the distributions are side-by-side and they can be ascertained accordingly.
(B) **Scatterplot of B *versus* A** would not serve this purpose in the least. As a matter of fact, it would be impossible to create because in order to do that you require <u>bivariate</u> data (two variables taken from each experimental unit). We haven't done that here because the only variable we collected from each person is their salary.
(C) **Parallel boxplots of A and B** would serve the same purpose as the back to back stem and leaf plots because it would show the distributions of the salaries side by side.
(D) **Histograms of A and B that are drawn to the same scale** should also serve this purpose because you can see the shape, center, and spread of each group side by side. Also, it's very important that you use the same scale for each because if they're not on the same scale, it can be misleading.
(E) **Dotplots of A and B that are drawn on the same scale** could also be useful because they can show the distributions of each variable side by side as well.

I think this is an important question because you may very well be asked to draw the 'appropriate' display for some data and you have to know what to do. If you want to <u>compare</u> distributions, everything but a scatterplot will get that done. A scatterplot can show a <u>relationship</u> between to variables (each garnered from the same experimental unit).

21. Answer: (A)

When reading these outputs, you have to know how to acquire the least squares regression line. Ignore all the stuff at the top (Regression, Residual, etc). Also, ignore R squared (adjusted) as this does not pertain to you. Here are the only parts of the output you need for this question:

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	11.6559	0.3153	37	≤ 0.0001
Size	3.47812	0.294	11.8	≤ 0.0001

The only confidence interval you will ever be asked to create for regression is one for the slope of the least squares regression line. Recall, though, that degrees of freedom for regression are $n - 2 = 22 - 2 = 20$ (which is conveniently given to you in this example). You must tehn look up the critical value for 95% confidence and 20 degrees of freedom:

df					
20	1.325	1.725	2.086	2.197	2.528
:	:	:	:	:	:
	80%	90%	95%	96%	98%

Confidence Level *C*

The formula for the CI is as follows:

$$b_1 \pm t^*_{n-2} \times SE_{b_1}$$

$$3.47812 \pm 2.086 \times 0.294$$

$$(2.864836,\ 4.091404)$$

Even though the actual confidence interval is not in the question, it's good to know what it means. We are 95% confident that for each additional kilobyte of size, the document requires somewhere between an additional 2.864836 minutes and 4.091404 minutes to print.

22. Answer: (D)

This question harps at the distinction between an observational study and a well-designed experiment. Since there was no random assignment of subjects to treatment groups, this is not an experiment. It is just an observational study. It's actually a retrospective study because past data was collected to acquire this information. Recall that the main difference between an observational study and an experiment is that an experiment can determine causality (provided that it's well-designed) and an observational study can only determine that a relationship exists between the variables. So…

(A) is TRUE → This should be included in the report because even though seatbelts were probably a factor, they probably weren't the only factor. This study could be confounded by driver behavior. That is, it may be hard to tell if the injury (or lack thereof) was because of the seatbelt or because of the driver's behavior.

(B) is TRUE → This should be included in the report because, again, you cannot tell if the injury was due to the seatbelt or the child's location in the car. Studies have shown that children who sit in the front of a car are more apt to be injured than children who sit in the back seat. So this, too, is a confounding factor.

(C) is TRUE → This should be included in the report because this was not an experiment. We didn't take kids in throw them into a car (some with seatbelts on, some without), crash the car into a pole, and record how things panned out. Since we cannot do this (for ethical reasons… hello?) a cause and effect relationship cannot be determined even if it does not exist.

(D) is FALSE → This should not be included in the report because in an observational study, we can only determine that in accidents where seatbelts were worn, less children were injured.

(E) is TRUE → This should be included in the report because one of the main caveats of experimental design is replication. If something is true, it should always be true, not just in this example.

23. Answer: (E)

In probability theory, we have three main types of events:

1. **Mutually Exclusive (a.k.a. disjoint) Events**: Outcomes can not happen at the same time.
2. **Independent Events**: Outcomes can happen at the same time, they just don't affect each other.
3. **Neither Independent nor Disjoint (a.k.a. dependent) Events**: Outcomes can happen at the same time and they do affect each other.

(A) is FALSE → If the events are mutually exclusive, they cannot be independent. They cannot exist in both categories at once.

(B) is FALSE → **I**f the events are independent, they cannot be mutually exclusive. They cannot exist in both categories at once.

(C) is FALSE → If the events are not mutually exclusive, they could be either independent or dependent.

(D) is FALSE → If the events are not independent, they could still be either mutually exclusive or dependent.

(E) is TRUE → These events cannot exist in both categories at once. If the events are categorized as mutually exclusive, they cannot be independent.

24. Answer: (D)

Here, the results of a statistical test and subsequent p-value are given and you are asked to make assertions based on this information:

Remember that in order for a test to be "statistically significant", the p-value must be less than the alpha level. Since the p-value is 0.24, you do not reject Ho and thus the results are not statistically significant at any reasonable alpha level [0.01, 0.05, or 0.10].

(A) is FALSE → In order for these data to be statistically significant, the p-value must be less than the alpha level.
(B) is FALSE → You need not do anything to the p-value. It is what it is. Leave it alone.
(C) is FALSE → Once again, you don't need to double the p-value unless you're given a one-tailed test and you need to convert it to a two-tailed test. In no way is this indicated here.
(D) is TRUE → This is the definition of a p-value. It is the probability of viewing data as extreme or more extreme than that which you saw, supposing the null hypothesis were correct.
(E) is FALSE → Even though it is the correct definition of what a p-value is, there is no reason to subtract that value from one.

25. Answer: (E)

This question deals with the appropriate design of a study. Make sure that you understand the different components that deal with the design and the implementation of a study as well as the different biases that could arise from your study.

(A) is FALSE → You cannot simply let subjects pick whichever drug they do/don't take.
(B) is FALSE → When you design a study, you must always make it to where the subjects are randomly assigned to treatments.
(C) is FALSE → Even though it is a paired design, you still need to utilize randomization. That is, you still need to randomly assign which group gets the old drug first and which gets the new drug first.
(D) is FALSE → The question states that they want to compare the old drug to the new drug. If you don't use the old drug, all you have evidence for is the idea that the drug works better than nothing. Bang-up job, genius.
(E) is TRUE → The subjects are randomly assigned to two groups and one group gets the old drug, the other gets the new drug. Perfect.

26. Answer: (D)

26. This is a question about sample size determination. It's likely that this question (or one like it) will come up on the exam because this is one of the formulas that you have to memorize. The formulas for sample size determination depend upon the data:

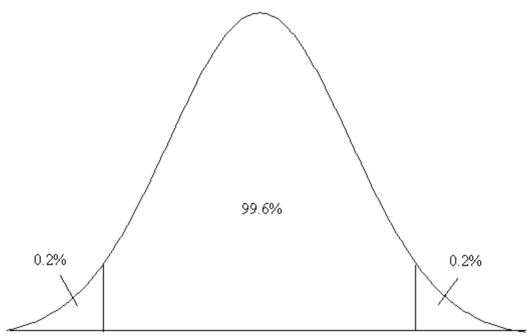
$$\text{Quantitative} \rightarrow \qquad n=\left(\frac{z\times\sigma}{m}\right)^2$$

$$\text{Categorical} \rightarrow \qquad n=\left(\frac{z}{m}\right)^2 p\left(1-p\right)$$

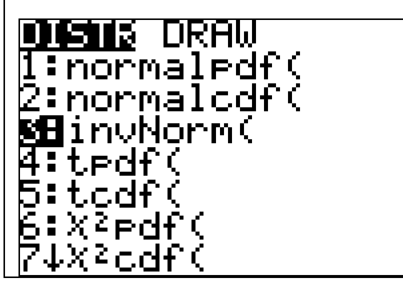
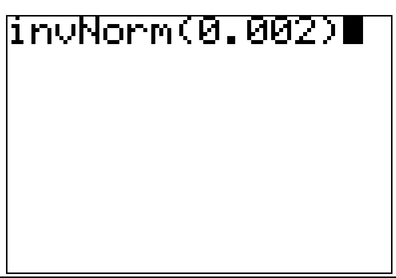
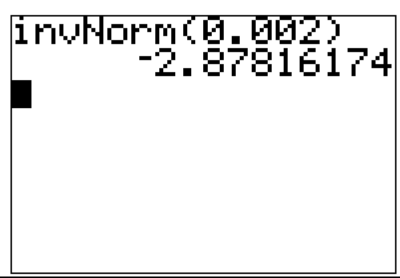
In the formula, the critical *z* comes from the confidence level and the margin of error (*m*) will be in the question noted as either (a) "margin of error" or (b) it will follow the word "within".

Also, for the formula that deals with categorical data, if there is no estimate for *p* (or if the question asks for the "most conservative estimate"), use 0.5.

Because the question asks for a critical *z* value that is not on the t-table (99.6% confidence), you have to get it yourself using the calculator:



Now use the calculator to get the critical value:

1. Press 2nd → VARS → 3:invNorm(	2. Type invNorm(area to left)	3. The value will inevitably be negative, so make it positive.
		

Now just use the quantitative formula because this is quantitative data (it's a machine measuring the weights of snack foods):

$$n = \left(\frac{z \times \sigma}{m}\right)^2 = \left(\frac{2.878 \times 0.30}{0.12}\right)^2 = 51.76$$

Always round up! → $n = 52$

<u>Note</u>: The critical z value for a confidence level <u>very</u> close to 99.6% is on the t table (99.5% confidence) as $z = 2.807$. If you use this number in the formula, you get a very similar answer:

$$n = \left(\frac{z \times \sigma}{m}\right)^2 = \left(\frac{2.807 \times 0.30}{0.12}\right)^2 = 49.24$$

Among the choices, $n = 52$ is the 'minimum' sample size that would achieve his goals.

27. Answer: (A)

Be careful with these questions. This is a <u>cumulative</u> <u>frequency</u> histogram, not just a histogram. For example, look at the test score of 70. The cumulative frequency is about 25% which means that about 25% of the students scored a 70 on the test <u>or less</u>. Knowing this…

(A) is TRUE → The variability of the data can be determined by looking at the horizontal distance covered before achieving that percentile. I would guess that the 20th percentile [lower 20%] is at a test score around 65. So the lower 20% of the test scores ranged from about 35 to about 65. The 80th percentile [upper 20%] is at a test score of about an 80. So the upper 20% of test scores are spread out from about 80 to 100. So, the range of the lower 20% test scores is approximately 65 – 35 = 30 and the range of the upper 20% of test scores is approximately 100 – 80 = 20. Since the lower 20% of test scores are more spread out, they have more variability.

(B) is FALSE → The median test score should be at the 50% mark on the cumulative percent. Look at 50 on the y-axis and trace the graph to the right and you will see the median score is somewhere between 70 and 80. Not less than 50.

(C) is FALSE → If you look at the test score of 80, the cumulative frequency for that value is 60%. This means that 60% of the students scored an 80 <u>or less</u>.

(D) is FALSE → If you look at the cumulative frequency for a test score of 70, the cumulative frequency is about 25%. This means that 25% of students scored a 70 <u>or less</u> and logically, 75% of the students scored a 70 <u>or greater</u>.

(E) is FALSE → The values which occurred the most frequently are those for which there were the biggest "jumps" in the data. There seems to be the biggest "jump" in cumulative frequency from the score of 80 to the score of 85. This means that these values occurred the most frequently.

28. Answer: (E)

Remember that in order for a linear model to be appropriate for a set of bivariate data, the residual plot must have random scatter above and below zero with no discernable patterns. The first residual plot (of the original variables, *x* and *y*) shows a clear curved pattern in the residuals. The second regression is then run on the transformed variables (the log of *x* and the log of *y* were both taken) and the residual plot for the transformed data looks much better. So…

(A) is FALSE → There is definitely a nonlinear relationship between *x* and *y* as seen by the curved residual plot and Regression I definitely does not yield a better fit.
(B) is FALSE → For the same reason as (A) … there is not a linear relationship between *x* and *y*.
(C) is FALSE → You cannot tell if the correlation is negative or positive based upon a residual plot.
(D) is FALSE → Even though there is a nonlinear relationship between *x* and *y*, Regression I is still not the better fit.
(E) is TRUE → There is a nonlinear relationship between *x* and *y* and Regression II yields the better fit.

29. Answer: (D)

Recall that a two-sided hypothesis test and a confidence interval with a complimentary alpha level will always yield consistent results. In this case, the confidence level is 98% and the alpha level is 0.02, so these results will be consistent.

(A) is FALSE → Even though the question doesn't indicate that we know the value of σ, the sample size of 500 renders that fact almost irrelevant. At a sample size of 500, *s* is an excellent estimate of σ.
(B) is FALSE → Again, the population need not necessarily be normally distributed if the sample size is 500. The sampling distribution is approximately normal regardless of that fact.
(C) is FALSE → You do not need the entire data set to get this question. All you need is the confidence interval and the hypothesis test and you're good to go.
(D) is TRUE → The value for the null hypothesis is 40,000. The 98% confidence interval is entirely above 40,000 so the null hypothesis would be rejected. We are 98% confident that the true mean income is not $40,000. Also, though not asked in this question, you can also conclude that the p-value is less than 0.02 since you reject H_o.
(E) is FALSE → The question says that the null hypothesis is not rejected. This is not correct. If the value from your hypothesis test is not in your confidence interval, then you must reject the null.

30. Answer: (C)

On your formula sheet, you are given the following information regarding the sampling distribution of the sample mean:

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \sigma/\sqrt{n}$$

Since we are taking samples of size 2 from this (very small) population, we know the following about the sampling distribution of $\bar{x}$:

$$\mu_{\bar{x}} = \mu = 4.25$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.92/\sqrt{2} = 1.358$$

So… $\mu_{\bar{x}} = 4.25$ and $\sigma_{\bar{x}} < 1.92$

31. Answer: (B)

Since I've gone over this one a bunch of times in previous examples, we'll just go over the **right** answer:

Recall that the slope is defined as follows:

"We predict, on average, that for each additional increase in *x*, *y* will increase by [<u>enter slope here</u>]"

Since *x* is the number of degrees Fahrenheit by which the temperature exceeds 50^{o} and *y* is the number of chirps per minute, the slope must mean that for each additional 1^{o} increase in temperature, the number of chirps per minute increases by 3.41. [Remember that 10.53 is <u>not</u> the slope. It's the *y*-intercept].

32. Answer: (E)

This example follows a binomial distribution. Recall the 4 qualifications for a binomial distribution:

1) There are two outcomes (success and failure)
2) The probability of success is constant
3) The trials are independent
4) There are a fixed number of trials

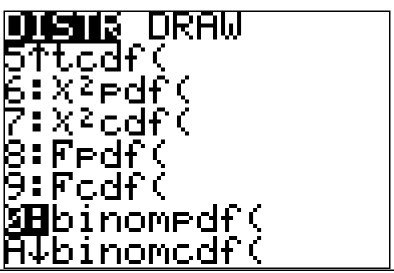
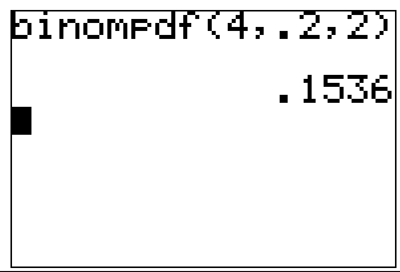
This example works nicely:

1) There are 2 outcomes [you either win the prize or you don't]
2) The probability of success is constant [It's always 0.20 because there are 5 boxes and one prize]
3) The trials are independent [If you win this time, it has no bearing on you winning the next time]
4) There are a fixed number of trials [You are guessing 4 times]

Since the question is asking for the probability of getting <u>exactly</u> *k* successes within *n* trials (in this case, exactly 2 correct out of 4 guesses), you use the binomial formula given to you on your formula sheet:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}=\binom{4}{2}(0.2)^2(0.8)^2$$

This question just wants you to use the formula with the numbers in the appropriate places. If you had to evaluate this, though, you can do it in the calculator:

1. Press [2nd] → [VARS] → 0:binompdf(	2. Type binompdf (*n*, *p*, *k*)	3. This gives you the probability of getting <u>exactly</u> 2 successes out of 4 trials if the probability of success is 0.20.
		

(A) $\frac{2!}{5!}$ → This gives the number of possible ways in which you can get 2 successes out of 5 trials. It is called a permutation. This is <u>not</u> covered in this class

(B) $\frac{(0.2)^2}{(0.8)^2}$ → This is a crap answer. Did you fall for this? Don't fall for this.

(C) 2(0.2)(0.8) → This would follow a binomial distribution if there were 2 ways the event could occur.

(D) $(0.2)^2(0.8)^2$ → This would have been the solution if the question had asked: "What is the probability of guessing correctly on the first two attempts, but then guessing incorrectly on the last two".

33. Answer: (B)

In this question, you are kind of tricked because the sample size is small, so you think that inference cannot be done. However, the question states that dotplots of the data indicate that the assumption of normality is not unreasonable, so the small sample size is not important. Remember that you use a *t*-distribution to estimate the *z*-distribution when the population standard deviation is unknown. So…

(A) is FALSE → We <u>always</u> construct confidence intervals using our statistics in hopes of capturing parameters. In this case, we are using the sample mean carbon content to give us insight into what the population mean carbon content could be.
(B) is TRUE → The question states that "there are no historical data on the variability of the process". This means that the population standard deviation is unknown so we use the sample standard deviation to estimate it. I know the answer says that the sample variance is used as an estimate of the population variance, but remember:

$$\text{Variance} = \text{Standard Deviation}^2$$

If the population standard deviation is unknown, then so is the variance.

(C) is FALSE → We should always use data rather than theory to judge normality.
(D) is FALSE → You can use data from however many days you think it takes to get a representative sample. In no way does this mean that you have to use a *t* distribution.
(E) is FALSE → It is good to use a *t*-distribution when the sample size is small, but the main reason is because the standard deviation for the population is unknown. You can use a *z*-interval with a small sample size if the population standard deviation is known and the population from which the data comes is approximately normal. So the statement, "A *z*-interval should never be used with a small sample" is false.

34. Answer: (C)

This example uses bivariate data because two measures were recorded for each rat: the amount of caffeine consumed (*x*) and the rat's blood pressure (*y*). The correlation is calculated as 0.428. This is *r*. This means is there is a moderately weak positive correlation between the amount of caffeine consumed by a rat and that rat's blood pressure. So…

(A) is FALSE → The correlation for our sample is 0.428. This is not the correlation for the entire population. Although, if these 100 rats were randomly sampled from the population, the population correlation should be close to that.
(B) is FALSE → Correlation does not imply causation. In no way will not drinking the caffeine cause a reduction in blood pressure. There could be a multitude of confounding variables that we do not know about.
(C) is TRUE → It is the coefficient of determination that describes the percent of variability in y explained by *x*. In this example:

$$r^2 = 0.428^2 = 0.183$$

This means that about 18% of the variability in a rat's blood pressure can be explained by the amount of caffeine the rat consumed. Another way of saying this (you don't have to know this for free response questions, but it could be on the multiple choice section). About 18% of the variation in blood pressure can be explained by the linear association between blood pressure and caffeine consumed. This is equally as viable an answer.

(D) is FALSE → We don't know what the rats like. Rats don't talk. This is stupid.

(E) is FALSE → A weak correlation does not necessarily mean that the relationship is nonlinear. This could be the case because nonlinear relationships weaken the correlation, but it could just be a sort of weak linear association.

35. Answer: (C)

Remember that the power of a test [known as $(1-\beta)$] is not something that you have to calculate; just something you have to know. The power of a test is the probability that you correctly reject a false null hypothesis (it's your BS detector). Recall that to increase the power, the two ways to do that are to increase the sample size and/or increase the effect size. Also, remember:

$$\alpha \uparrow \Rightarrow \beta \downarrow \Rightarrow (1-\beta) \uparrow$$

So, to achieve the highest power, you just have to find the highest alpha level (Type I Error Rate) and the highest sample size will give you the most power.

36. Answer: (B)

Since the balls are being replaced, you can easily know all possible combinations of balls that can be drawn and thus find the average of each:

Balls Drawn	$\overline{X}$
1 and 1	1
1 and 2	1.5
1 and 3	2
2 and 1	1.5
2 and 2	2
2 and 3	2.5
3 and 1	2

3 and 2	2.5
3 and 3	3

Now there are 9 possibilities, you know what the sampling distribution of $\overline{X}$ is just by looking at how often each $\overline{X}$ occurs:

$\overline{X}$	1	1.5	2	2.5	3
Probability	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}=\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

37. Answer: (E)

Since the simple random sample yields an $\bar{x}$ of 15 $\pm$ 3, this gives a resulting confidence interval of (12, 18). This can be viewed from two perspectives:

(1) We are 95% confident that μ lies between 12 and 18.

(2) In all intervals constructed the same way (of the same sample size from the same population), we expect 95% of them to capture μ.

So…

(A) is FALSE → To discuss "95% of population measurements" is a stupid statement. We have never covered "population measurements".
(B) is FALSE → This is because "95 of sample measurements" will look like whatever the hell they want to look like. However, if we create intervals around these 'sample measurements', 95% of these intervals should capture μ. However, our confidence interval is not the benchmark for all confidence intervals.
(C) is FALSE → This is for the same reason as choice (B). Again, our confidence interval is not the benchmark for all confidence intervals. It's close to the right answer, but it's actually incorrect and that's why most students chose it. Had it said "In 100 intervals constructed the same way, we expect 95 of them to capture μ", that would have been correct.
(D) is FALSE → This is because $P(12 \le \bar{x} \le 18)$ = 1. The sample average is <u>always</u> directly in the center of any confidence interval.
(E) is TRUE → This is because in all intervals constructed the same way, we expect 95% of them to capture μ. So if μ=19, we would be kind of surprised. We are 95% confident that μ is between 12 and 18. Only 5% of all 95% confidence intervals will not contain μ, and it looks like we're in that 5% of intervals. So if μ = 19, this $\bar{x}$ of 15 would be unlikely.

38. Answer: (E)

Since this question is asking for the answer in terms of those who are <u>against</u> the tax increase, do it entirely from that perspective ($p = 0.35$):

This is probably the way that these questions will be asked in that they will use a rather large sample size ($n = 500$) as sort of a hint that you should use a normal model. Since there are at least 10 successes and at least 10 failures:

$$np = (500)(0.35) = 175 \geq 10 \text{ and } n(1-p) = (500)(0.65) = 325 \geq 10$$

…a normal model is appropriate. This makes these two choices incorrect:

(A) $\binom{500}{200}(0.65)^{200}(0.35)^{300} \rightarrow$ This is the probability of getting <u>exactly</u> 200 people who are <u>in favor of</u> the taxes (because it uses 0.65 first) in the sample of 500 people.

(B) $\binom{500}{200}(0.35)^{200}(0.65)^{300} \rightarrow$ This is the probability of getting <u>exactly</u> 200 people who are <u>against</u> the taxes (because it uses 0.35 first) in the sample of 500 people.

Now that you have a normal model, you need a mean and a standard deviation. These are given to you on your formula sheet:

$$\mu_{\hat{p}} = p = 0.65$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.35)(0.65)}{500}}$$

The question asks for the probability that more than 200 people are against the taxes. This is the same as asking for the probability that more than $200/500 = 0.40$ are against the taxes. Now, just convert all the information into a z-score and find the probability of getting a z-score <u>bigger</u> than that:

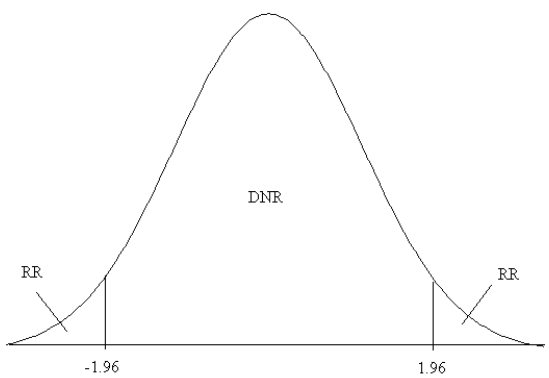
$$z = \frac{x - \mu}{\sigma} = \frac{0.40 - 0.35}{\sqrt{\frac{(0.35)(0.65)}{500}}}$$

39. Answer: (A)

The only way that a two-sided test can find significance (a.k.a. reject H_0) and a one-sided test does not is if the researcher chose the wrong direction for the one-sided alternative. Let's discuss this from the perspective or the rejection region(s):

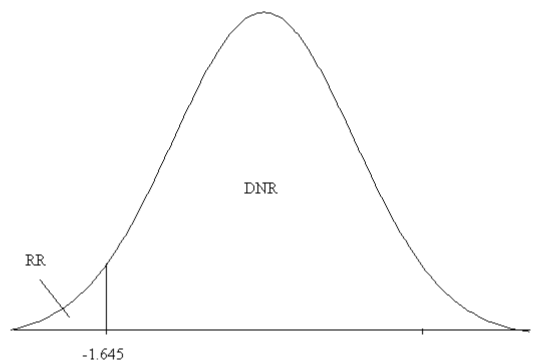
<u>Sally</u>

Results were significant (rejected H_0)
Used two-sided test

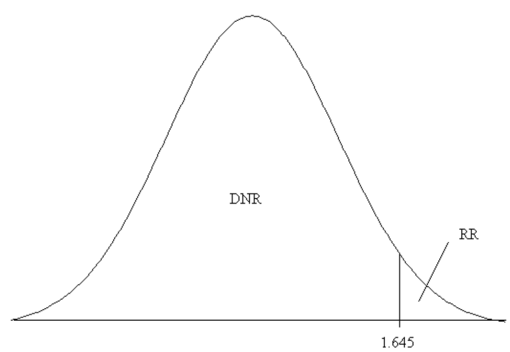


<u>Betty</u>

Results were <u>not</u> significant (did <u>not</u> reject H_0)
Used one-sided test (direction not known)



OR



The toughest part of this question is its ambiguity. Betty could have chosen the 'wrong' direction in her hypothesis test. Therefore, just check to see if any of the choices for a test statistic fall in Sally's rejection region and doesn't fall into <u>either</u> of Betty's (since you don't know which one it was). The only answer choice that satisfies this is (A) –1.980.

40. Answer: (B)

Here, the person is attempting to do a two sample *t*-interval:

(A) is FALSE → The two sample sizes do not need to be equal in order to perform any inference on independent means.
(B) is TRUE → The student gathered the ages of every prime minister and every president, so a confidence interval is unnecessary. He has the parameters (the true mean ages of both populations). He's a winner. Game over.
(C) is FALSE → The distribution of ages could very well not be similar and the test can still be performed. They just both have to be large samples or from normal populations in order to be valid.
(D) is FALSE → We don't really know what the shapes of the ages would be. They're probably all centered around late 50's with a fairly normal distribution, though you cannot really speculate to this effect without the data.
(E) is FALSE → This is for the same reasons as choice (D). You have no idea what the distributions looks like and cannot make assumptions to that effect.