SOLUTIONS - MIKE

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. yes!
(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?

so prop >.05

(A) H_0: $p \neq 0.05$, H_a: $p = 0.05$

B (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: $p = 0.05$

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?

1st exam $z = 1$ 2nd exam $z = 1$

find z-scores

(A) She scored much better on the first exam.
(B) She scored much better on the second exam.
C (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 → 3/10 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 (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. → 3 choices 7 choices
(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.

2002 AP STAT EXAM

40 MULTIPLE CHOICE

90 MINUTES

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

D

$\Sigma x \cdot p(x) = 1(.2) + 2(.3) + 3(.1) + 4(.08) + 5(.02) = 1.52$

$\times 25$

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

E

- stays same
- transformations don't change correlation

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

A

+4 each score so $\bar{x}$ ↑

S stays same... only a shift of data, no change in spread.

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 $\underset{\bar{x}}{25.5} \pm \underset{t^*}{2.228} \times \frac{3.01^{\,s}}{\sqrt{11}\ \sqrt{n}}$.

(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.

A

df = 10 use t-chart 95%

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

E

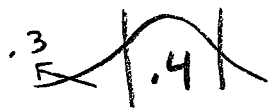
~~(A)~~ reduce nonresponse bias
~~(B)~~ reduce the effects of confounding variables
~~(C)~~ reduce bias due to the interviewer effect
~~(D)~~ decrease the variability in the population → you cannot change σ.
(E) decrease the standard deviation of the sampling distribution of the sample proportion

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?

4,000 = 40%

D

(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



$-.5244 = \frac{x-10}{.2}$

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").

B

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

$\frac{82 \cdot 92}{157}$

$\frac{\text{(column total)(row total)}}{\text{total}}$

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?

B

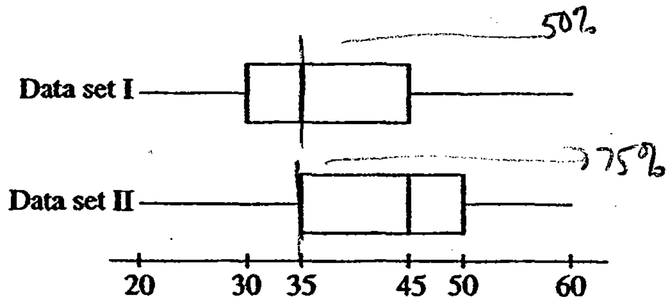
(A) One-sample z-test
(B) Two-sample t-test
(C) Paired t-test
(D) Chi-square goodness-of-fit test
(E) One-sample t-test

2 diff groups

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? more confid = more room for error

(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.

A



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. True
(B) The interquartile range of data set I is equal to the interquartile range of data set II. True
D
(C) The median of data set I is less than the median of data set II. True
(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. True

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?

split in groups, then sample

(A) Cluster
(B) Convenience
D
(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?

E

(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. he's considering 2 different variables

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?

B

(A) −7.85
(B) −4.60
(C) 4.60
(D) 5.00
(E) 7.85

$16.6 + .65(20) = 29.6$

$\text{Residual} = y - \hat{y}$
$25 - 29.6$

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

E

I. The t-distribution is symmetric. ✓
II. The t-distribution with k degrees of freedom has a ~~smaller~~ more 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

Brown Eyes	Green Eyes	Blue Eyes
O → 34	15	11
E → 30	15	15

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?

B

(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

$\sum \frac{(O-E)^2}{E}$

$\chi^2 = 1.6$

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?

B

(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

all of these show comparison easily of before and after raises

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
	slope	SEb		

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) $11.6559 \pm 1.725 \times 0.3153$

df = n − 2

95%

look up value in t-chart

A

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.

D

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.

= disjoint → cannot happen at same time

True

E

ex: P(A on test) / P(B on test)

disjoint ✓

P(A) depends on getting a B so they are NOT indep.

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?

D

~~(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. defn' of p-value
(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?

E

~~(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.~~ dangerous! cannot accurately analyze
(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.

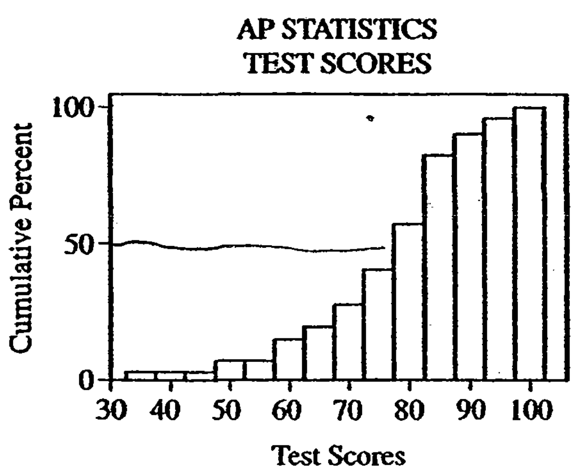
26. A quality control inspector must verify whether a machine that packages snack foods is working correctly. The inspector will randomly select a sample of packages and weigh the amount of snack food in each. Assume that the weights of food in packages filled by this machine have a standard deviation 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?

D

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

$$\left(\frac{z^*\sigma}{m}\right)^2 = n$$

$$\left(\frac{2.878(.30)}{.12}\right)^2 = 51.77 \to 52$$

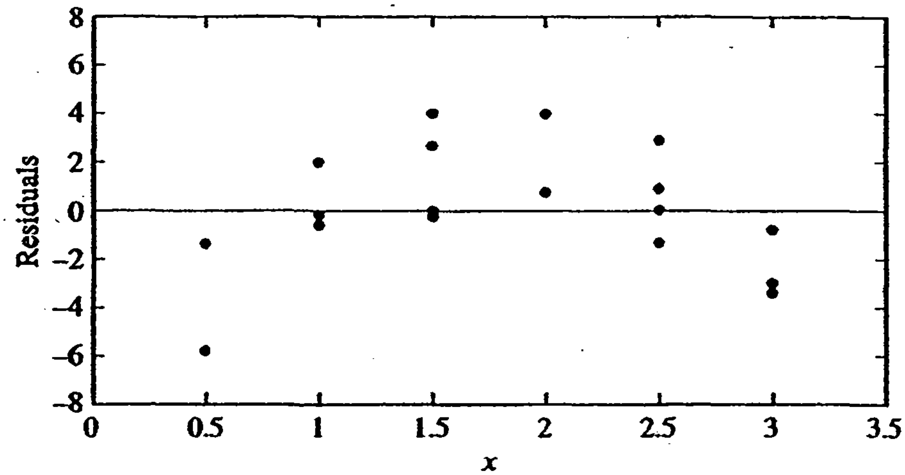


27. The figure above shows a cumulative relative frequency histogram of 40 scores on a test given in an AP Statistics class. Which of the following conclusions can be made from the graph?

(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~~ more than 50.~~
~~(C) Sixty percent of the students had test scores ~~above~~ below 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~~ least frequently.~~

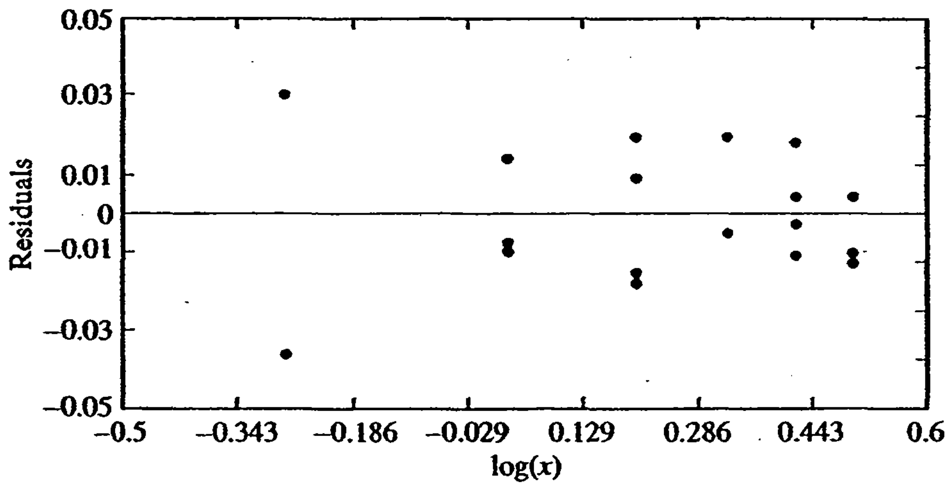
p.7

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.



bad model!

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



better model!

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.

E

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

(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.

D

P. 8

(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. only b/c of the ≠ test!

(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?

CLT

C

(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) $\mu_{\bar{x}} < 4.25$

$\mu_{\bar{x}} = \mu \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{2}}$

smaller

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?

B

(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. def'n of slope
(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?

E

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

combos success failure
$\binom{4}{2}(.2)^2(.8)^2$
binomial

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

B

(A) is estimating the population mean using the sample mean
(B) is using the sample variance as an estimate of the population variance the reason we use t & not z!
(C) is using data, rather than theory, to judge that the carbon content is normal know this!
(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?

$r^2 = .183$

C

(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. defn of r^2
(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 (C) $n = 20, \alpha = 0.05$

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

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

as $n \uparrow$, less variability

as $\alpha \uparrow$, reject more

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 X	2 ∝	2.5	3 •
Probability	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

B

(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.

	$\overline{X}$	
1 1	1	✓
1 2	1.5	X
1 3	2	∝
2 1	1.5	X
2 2	2	∝
2 3	2.5	
3 1	2	∝
3 2	2.5	
3 3	3	•

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. The pop is the pop → doesn't change.

(B) Ninety-five percent of the ~~sample~~ measurements fall between 12 and 18.
you are estimating μ, not individ measurements

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

(D) $P(12 \leq \bar{x} \leq 18) = 0.95$ This is only based off 1 sample.

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

If $\mu = 19$, it should be in the C.I., but it's not, so $\bar{x} = 15$ is unlikely

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?

$\mu_x = .35$

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

exactly 200

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

$p = \frac{200}{500} = .4$

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

exactly 200

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

$\sigma_x = \sqrt{\frac{p(1-p)}{n}}$

stdev based on what you believe. 35% against

E

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

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 Sally 2-sided significant at $\alpha = .05$ Betty 1-sided not signif

(B) −1.690

(C) 1.340

A (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 (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 You know the actual difference ⇒ No need to estimate

(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