

Extra Credit Assignment of Math 1680.009 (SP012):-

Instructor: Koshal Dahal

A Perfect assignment is worth 5, else is 0. A hard copy of completely-answered-Scranton (*Write your name (Last/First), UNT-ID #, with Math-class #, clearly at the top of your answered-Scranton*) of this assignment must be turned in by **May 1 (strictly before class starts, NO excuse!)**, to get the FULL credit. If you don't have all the solutions correct, then the instructor, **only if** deems necessary..., may provide you a partial credit, which may be at most 40% but not the higher. Also, let me know earlier, if any question is wrong or there is any mistake in the given answer-choices.

Warning!! You are FREE to use any "help/life-lines" from any sources: Book, friends, math lab, **excluding me!** to do this assignment. I encourage you all for the group discussion.

Good Luck in advance!!

Chap.2

Q.1: What is/are statistic(s)?

- A. Statistics (the discipline) is a way of reasoning, a collection of tools and methods, designed to help us understand the world.
- B. Statistics (plural) are particular calculations made from data.
- C. Data are values with a context. Data has a story to tell and our objective is to uncover that story
- D. all of the above

Q.2: Variables that are numbers are always quantitative.

- A. True
- B. False

Q.3: A(n) _____ holds information about the same characteristic for many cases.

- A. Element
- B. Population
- C. Variable
- D. Sample

Q.4: What are data?

- A. A list of numbers
- B. Numbers or labels together with context
- C. Context only
- D. A list of labels

Q.5: The context ideally tells:

- A. Who, What, and How
- B. When and Where
- C. Why
- D. All of the above

Q.6: A(n) _____ is an individual about whom or which we have data.

- A. Sample
- B. Population
- C. Variable
- D. Case

Q.7: The cases we actually examine to understand a larger group is the:

- A. Element
- B. Population
- C. Sample
- D. Variable

Q.8: All the cases we wish we knew about is the:

- A. Element
- B. Population
- C. Sample
- D. Variable

Q.9: A quantity or amount adopted as a standard of measurement is a(n):

- A. Element
- B. Sample
- C. Variable
- D. Unit

Q.10: The SPCA collects data about the dogs they house. Which is categorical?

- A. Breed
- B. Age
- C. Weight
- D. Veterinary costs

Q.11: School administrators collect data on the students attending the school. Which of the following is quantitative?

- A. Class (freshman, sophomore, etc.)
- B. Grade point average
- C. Whether the student is in AP class
- D. Whether the student has taken the SAT

Q.12: We collect these data from 50 students. Which variable is categorical?

- A. Eye color
- B. Head circumference
- C. Hours of homework last week
- D. Number of TV sets in at home

Q.13: Without which of the following two W's should we consider that we do not have data or useful information?

- A. Who and When
- B. Who and Why
- C. Why and How
- D. Who and What

Chap 3

Q.14: Which of the following displays percentages rather than counts?

- A. Frequency table
- B. Bar chart
- C. Relative frequency table
- D. Contingency table

Q.15: It is permissible to violate the area principle if

- A. the percentages do not add up to 100%.
- B. we need a flashy display to make a point.
- C. we do not care about being true to the data.
- D. None of the above.

Q.16: Which of the following best gives a quick impression of how a whole group is partitioned into smaller groups?

- A. Bar chart
- B. Frequency distribution
- C. Pie chart
- D. Contingency table

Q.17: TV viewers during the Super Bowl in 2007.

		Sex		
		Male	Female	Total
Response	Game	279	200	479
	Commercials	81	156	237
	Won't watch	132	160	292
	Total	492	516	1008

What percentage of the Super Bowl viewers were male?

- A. 19.8%
- B. 47.5%
- C. 58.2%
- D. 27.7%

Q.18: TV viewers during the Super Bowl in 2007.

		Sex		
		Male	Female	Total
Response	Game	279	200	479
	Commercials	81	156	237
	Won't watch	132	160	292
	Total	492	516	1008

What is the marginal distribution of those who watched the commercials only?

- A. 8.0%
- B. 23.5%
- C. 58.2%
- D. 27.7%

Q.19: Refer to the table of TV viewers during the Super Bowl in 2007 of Question 17.

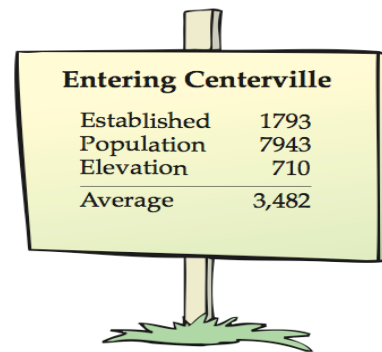
Given that a viewer did not watch the Super Bowl, what percentage were male?

- A. 45.2%
- B. 48.8%
- C. 26.8%
- D. 27.7%

Q.20: In a contingency table, when the distribution of one variable is the same for all categories of another, we say the variables are:

- A. Separate.
- B. Independent.
- C. Distinct.
- D. Dependent

Q.21: The following is an example of:



The image shows a yellow signpost on a wooden post with a small patch of grass at the base. The sign is titled "Entering Centerville" and contains a table with four rows of statistics.

Entering Centerville	
Established	1793
Population	7943
Elevation	710
Average	3,482

- A. Simpson's Paradox.
- B. Convolution.
- C. Aristotle's Paradox.
- D. Homer's Paradox.

Chap. 4

Q.22: Which if the following variables would most likely follow a Normal model?

- A. Family income
- B. Heights of singers in a co-ed choir
- C. Weights of male elephants
- D. Scores on an easy test

Q.23: You should use a histogram to display categorical data:

- A. True
- B. False

Q.24: Two sections of a class took the same quiz. Section A had 15 students who had a mean score of 80, and Section B had 20 students who had a mean score of 90. Overall, what was the mean score for all students on the quiz?

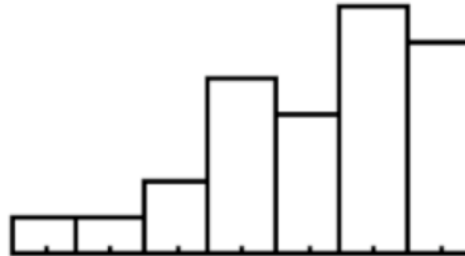
- A. 84.3
- B. 85.7
- C. None of these
- D. It cannot be determined.

Q.25: The SPCA has kept data records for the past 20 years. If they want to show the trend in the number of dogs they have housed, what kind of plot should they make?

- A. Boxplot
- B. Timeplot
- C. Bar graph
- D. Histogram

Q.26: Which is true of the data shown in the histogram?

- I. The distribution is skewed to the right.
- II. The mean is smaller than the median.
- III. We should use the median and IQR to summarize these data.



- A. I only
- B. II only
- C. III only
- D. II and III only
- E. I, II, and III

Q.27: If we want to discuss any gaps and clusters in the data set, which of the following should not be chosen to display the data set?

- A. Histogram
- B. Stem-and-leaf plot
- C. Boxplot
- D. Bar graph
- E. Dotplot

Q.28: We might choose to display data with a stemplot rather than a boxplot because a stemplot ...

- I. reveals the shape of a distribution.
- II. is better for large data sets.
- III. displays the actual data.

- A. I only
- B. II only
- C. III only
- D. I and III
- E. I, II, and III

Q.29: Which of the following are measures of the center of a distribution (circle all that apply):

- I. Mean
- II. Variance
- III. Standard deviation
- IV. Median

- A. I only
- B. III only
- C. I and IV
- D. IV only

Q.30: Which of the following are measures of the spread of a distribution (circle all that apply):

- I. Mean
- II. Variance
- III. Standard deviation
- IV. Median

- A. II only
- B. III only
- C. I and IV
- D. II and III

Chap. 5

Q.31: Which of the following is not part of the 5-number summary?

- A. The minimum and maximum
- B. The mean
- C. The median
- D. The quartiles: Q1 and Q3

Q.32: The five-number summary of credit hours for 24 students in an introductory statistics class is:

Min	Q1	Median	Q3	Max
13.0	15.0	16.5	18.0	22.0

From this we know that

- A. there are no outliers in the data.
- B. there is at least one low outlier in the data.
- C. there is at least one high outlier in the data.
- D. None of the above.

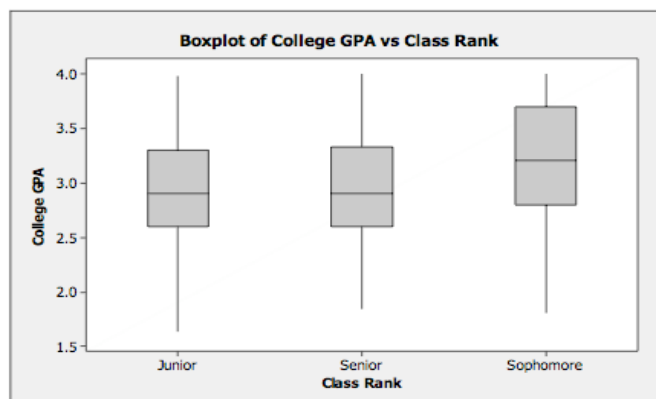
Q.33: Boxplots are particularly effective for displaying/comparing (circle all that apply):

- A. Categorical data
- B. Means
- C. Outliers & Groups

Q.34: The side-by-side boxplots below show cumulative GPAs for sophomores, juniors and seniors taking intro stats course in autumn 2003.

Which class had the lowest cumulative GPA?

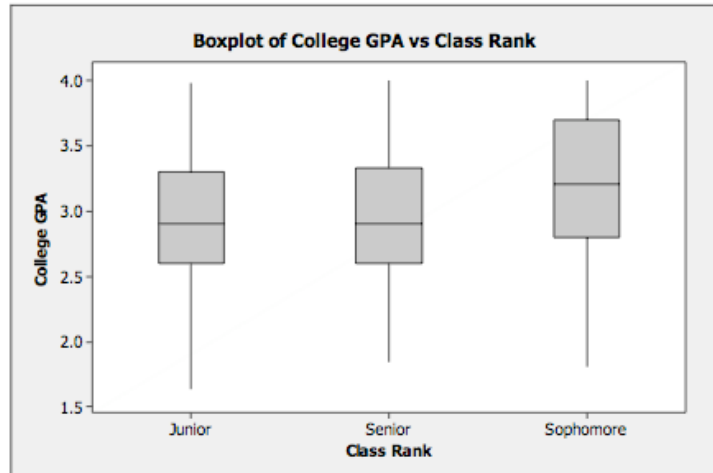
- A. Sophomore
- B. Junior
- C. Senior



Q.35: The side-by-side boxplots below show cumulative GPAs for sophomores, juniors and seniors taking intro stats course in autumn 2003.

Which class had the highest median GPA?

- A. Sophomore
- B. Junior
- C. Senior



Q.36: If you analyze data with outliers you can:

- A. Leave the outlier in place and proceed.
- B. Simply exclude the outlier.
- C. Exclude the outlier but discuss why.
- D. Never analyze data with outliers.

Q.37: An outlier is a point more than 1.0 IQR from either end of the box in a boxplot.

- A. True
- B. False

Q.38: When comparing two groups, use different scales, if necessary, for clarity and sizing.

- A. True
- B. False

Q.39: A far outlier is a point more than _____ IQR from either end of the box in a boxplot.

- A. 0.75
- B. 1.5
- C. 2.0
- D. 3.0

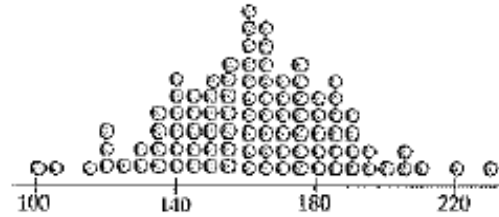
Chap.6

Q.40: Which of the following data summaries are changed by adding a constant to each data value?

- I. the mean
- II. the median
- III. the standard deviation

- A. I only
- B. III only
- C. I and II
- D. I and III
- E. I, II and III

Q.41: The best estimate of the standard deviation of the men's weights displayed in this dotplot is



- A. 10 pounds
- B. 15 pounds
- C. 25 pounds
- D. 40 pounds

Q.42: Suppose that a Normal model describes fuel economy (miles per gallon) for automobiles and that a Saturn has a standardized (z-score) of +2.2. This means that Saturns . . .

- A. get 2.2 miles per gallon.
- B. get 2.2 miles per gallon more than the average car.
- C. have a standard deviation of 2.2.
- D. achieve fuel economy that is 2.2 standard deviations better than the average car.

Q.43: Environmental researchers have collected data on rain acidity for years. Suppose that a Normal model describes the acidity (pH) of rainwater, and that water tested after last week's storm had a z-score of 1.8. This means that the acidity of that rain . . .

- A. had a pH of 1.8.
- B. had a pH 1.8 times that of average rainwater.
- C. had a pH 1.8 standard deviations higher than that of average rainwater.
- D. had a pH 1.8 higher than average rainfall.

Q.44: Suppose a Normal model describes the number of pages printer ink cartridges last for. If we keep track of printed page counts for the 47 printers in the company's office, which *must* be true?

- I. The page counts of those ink cartridges will be normally distributed.
 - II. The histogram for those page counts will be symmetric.
 - III. 95% of those page counts will be within 2 standard deviations of the mean.
- A. None
 - B. I only
 - C. II only
 - D. II and III
 - E. I, II, and III

Q.45: Students taking an intro stats class reported the number of credit hours that they were taking that quarter. Summary statistics are shown in the table.

Mean	Std Dev	Min	Q1	Median	Q3	Max
16.65	2.96	5	15	16	19	28

Suppose the college charges \$73 per credit hour plus a flat fee of \$35 per quarter. For example, a student taking 12 credit hours would pay $\$35 + 12(\$73) = \$911$ for that quarter.

What is the mean fee paid?

- A. \$216.05
- B. \$1250.45
- C. \$251.05
- D. \$1205

Q.46: Students taking an intro stats class reported the number of credit hours that they were taking that quarter. Summary statistics are shown in the table.

Mean	Std Dev	Min	Q1	Median	Q3	Max
16.65	2.96	5	15	16	19	28

Suppose the college charges \$73 per credit hour plus a flat fee of \$35 per quarter. For example, a student taking 12 credit hours would pay $\$35 + 12(\$73) = \$911$ for that quarter.

What is the standard deviation for the fees paid?

- A. \$216.08
- B. \$1250.48
- C. \$251.08
- D. \$1203

Q.47: Students taking an intro stats class reported the number of credit hours that they were taking that quarter. Summary statistics are shown in the table.

Mean	Std Dev	Min	Q1	Median	Q3	Max
16.65	2.96	5	15	16	19	28

Suppose the college charges \$73 per credit hour plus a flat fee of \$35 per quarter. For example, a student taking 12 credit hours would pay $\$35 + 12(\$73) = \$911$ for that quarter.

What is the median fee paid?

- A. \$216.08
- B. \$1250.48
- C. \$251.08
- D. \$1203

Q.48: In a Normal model, about 68% of the data fall within 2 standard deviations of the mean.

- A. True
- B. False

Chap. 7

Q.49: All but one of the following statements contains a mistake. Which one could be true?

- A. The correlation between height and weight is 0.568 inches per pound.
- B. The correlation between height and weight is 0.568.
- C. The correlation between the breed of a dog and its weight is 0.435.
- D. The correlation between gender and age is -0.171 .

Q.50: A correlation of zero between two quantitative variables means that

- A. we have done something wrong in our calculation of r .
- B. there is no association between the two variables.
- C. there is no linear association between the two variables.
- D. re-expressing the data will guarantee a linear association between the two variables.

Q.51: It takes a while for new employees to master a complex process. During the first month new employees work, a company tracks the number of days they have been on the job and the length of time it takes them to complete an assembly. The correlation is most likely to be

- A. exactly -1.0
- B. near -0.6
- C. near 0
- D. near $+0.6$

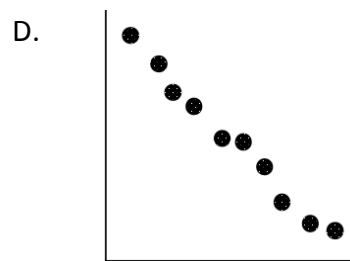
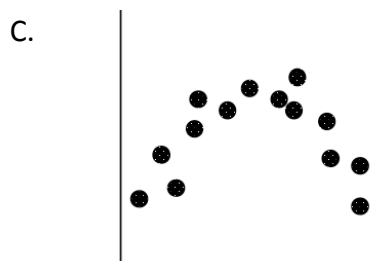
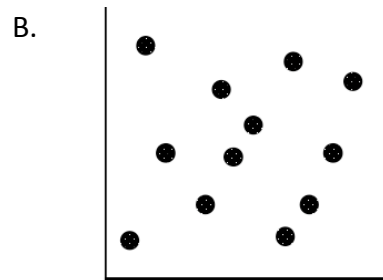
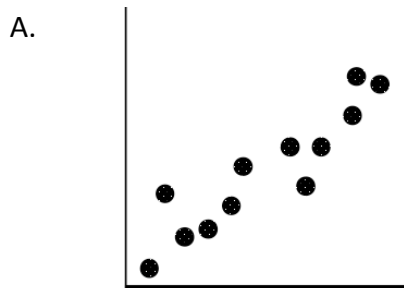
Q.52: A lakeside restaurant found the correlation between the daily temperature and the number of meals they served to be 0.40 . On a day when the temperature is two standard deviations above the mean, the number of meals they should plan on serving is _____ the mean.

- A. equal to
- B. 0.4 SD above
- C. 0.8 SD above
- D. 2.0 SD above

Q.53: For families who live in apartments the correlation between the family's income and the amount of rent they pay is $r = 0.60$. Which is true?

- I. In general, families with higher incomes pay more rent.
 - II. On average, families spend 60% of their income on rent.
 - III. The regression line passes through 60% of the (income\$, rent\$) data points.
- A. I only
 - B. II only
 - C. I and III only
 - D. I, II, and III

Q.54: Which scatterplot shows a strong association between two variables even though the correlation is probably near zero?



Q.55: The correlation between X and Y is $r = 0.35$. If we double each X value, decrease each Y by 0.20, and interchange the variables (put X on the Y -axis and vice versa), the new correlation is

- A. 0.35
- B. 0.50
- C. 0.90
- D. cannot be determined.

Q.56: A hidden variable that stands behind a relationship and determines it by simultaneously affecting the other two variables is called a _____ variable.

- A. lurking
- B. causal
- C. response
- D. Predictor

Chap. 8

Q.57: A residuals plot is useful because:

- I. It will help us to see whether a linear model makes sense.
 - II. It might show a pattern in the data that was hard to see in the original scatterplot.
- A. I only
 - B. II only
 - C. I and II
 - D. Neither I nor II

Q.58: To model a relationship with a regression line, a number of conditions need to be checked and met. Which of the following need not be checked?

- A. Quantitative variables condition
- B. Nearly normal condition
- C. Straight enough condition
- D. Outlier condition

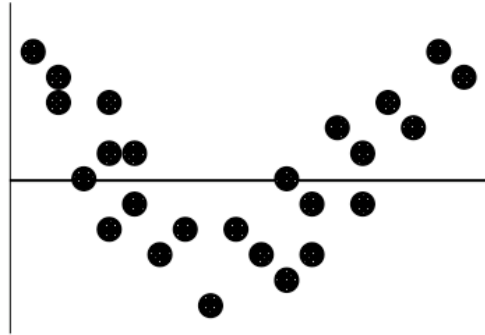
Q.59: A regression analysis of students' college grade point averages (GPA) and their high school GPAs found $r^2 = 0.311$. Which of the following is true?

- I. High school GPA accounts for 31.1% of college GPA.
 - II. 31.1 % of college GPAs can be correctly predicted with this model.
 - III. 31.1% of the variance in college GPA can be accounted for by the model.
- A. I only
 - B. II only
 - C. III only
 - D. None

Q.60: A regression analysis of students' AP Statistics test scores and the number of hours they spent doing homework found $r^2 = 0.32$. Which of the following is true?

- I. 32% of student test scores can be correctly predicted with this model.
 - II. Homework accounts for 32% of your grade in AP stats.
 - III. There is a 32% chance that you will get the score this model predicts for you.
- A. I only
 - B. II only
 - C. III only
 - D. None

Q.61: The residuals plot for a linear model is shown. Which is true?



- A. The linear model is okay because approximately the same number of points are above the line as below it.
- B. The linear model is okay because the association between the two variables is fairly good.
- C. The linear model is no good because the correlation is near 0.
- D. The linear model is no good because of the curve in the residuals.

Q.62: All but one of the statements below contain a mistake. Which one could be true?

- A. There is a high correlation between cigarette smoking and gender.
- B. The correlation between age and weight of a newborn baby is $r = 0.83$ ounces per day.
- C. The correlation between blood alcohol level and reaction time is $r = 0.73$.
- D. The correlation between a person's age and vision (20/20?) is $r = -1.04$

Q.63: A regression analysis of a company profits and the amount of money the company spent on advertising found $r^2 = 0.72$. Which of these is true?

- I. This model can correctly predict the profit for 72% of the companies.
- II. On average, about 72% of a company's profit results from advertising.
- III. On average, companies spend about 72% of their profits on advertising.

- A. None
- B. I only
- C. II only
- D. III only

Q.64: A consumer group collected data on standard color TV sets (no HD or large screen sets). They created a linear model to estimate the cost of the TV (in \$) based on the screen size (in inches). Which is the most likely slope of the line of best fit?

- A. 0.25
- B. 2.5
- C. 25.0
- D. 250.0

Chap.9

Q.65: By checking the Straight Enough Condition, we will definitely know whether the Linearity Assumption is true.

- A. True
- B. False

Q.66: If you find subsets among the population data, you should:

- A. Group all together and then perform the regression analysis.
- B. Choose data from each of the groups and perform the regression analysis.
- C. Perform separate regression analyses on each group.
- D. Do not perform regression analysis using any of the data.

Q.67: Extrapolation is

- A. okay to do as long as we are making predictions into the future.
- B. okay to do if we tell people that we are assuming the linear relationship will hold outside the range of the data.
- C. okay to do as long as there were no outliers in the original data.
- D. not okay to do.

Q.68: Which of the following is not a source of caution in regression analysis between two variables?

- A. Extrapolation
- B. A lurking variable
- C. Re-expressing data
- D. An outlier

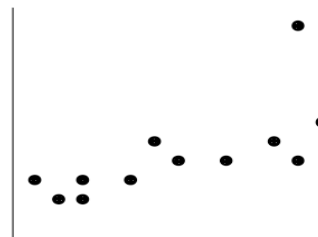
Q.69: Which statement about influential points is true?

- I. Removal of an influential point changes the regression line.
 - II. Data points that are outliers in the horizontal direction are more likely to be influential than points that are outliers in the vertical direction.
 - III. Influential points have large residuals.
- A. I only
 - B. I and II
 - C. II and III
 - D. I, II and III

Q.70: Two variables that are actually not related to each other may nonetheless have a very high correlation because they both result from some other, possibly hidden, factor. This is an example of

- A. Extrapolation
- B. A lurking variable
- C. Re-expressing data
- D. An outlier

Q.71: If the point in the upper right hand corner of the scatterplot is removed from the data set, then what will happen to the slope of the line of best fit (b) and the correlation (r)?



- A. Both will increase.
- B. Both will decrease.
- C. b will increase, r will decrease.
- D. b will decrease, and r will increase.

Q.72: Which statement about correlation is true?

- I. Regression based on data that are summary statistics tends to result in higher correlation.
 - II. If $r^2 = 0.95$, the response variable increases as the explanatory variable increases.
 - III. An outlier always decreases the correlation.
- A. None
 - B. I only
 - C. II only
 - D. III only

Chap. 11

Q.73: What is it about chance outcomes being random that makes random selection seem fair?

- I. Nobody can guess the outcome before it happens.
 - II. When we want things to be fair, usually some underlying set of outcomes will be equally likely.
 - III. Random outcomes display personal stakes in a particular outcome.
- A. I and II
 - B. II and III
 - C. I and III
 - D. All are not true.

Q.74: Examples of random selection are:

- A. Flipping a fair coin
- B. Rolling a fair die
- C. Using a spinner
- D. All of the above

Q.75: Numbers generated from a computer are generally:

- A. Only integers
- B. Random
- C. Pseudorandom
- D. Biased

Q.76: A simulation always models imaginary situations.

- A. True
- B. False

Q.77: Which of the following is not a step in building a simulation?

- A. Explain how you will model the component's outcome.
- B. State clearly what the response variable is.
- C. Run only one trial.
- D. State the conclusion.

Q.78: Which of the following is the first step in building a simulation?

- A. Explain how you will model the component's outcome.
- B. State clearly what the response variable is.
- C. State the conclusion.
- D. Identify the component to be repeated.

Q.79: Among a dozen eggs, three are rotten. A cookie recipe calls for two eggs; they will be selected randomly from that dozen. Which plan could be used to simulate the number of rotten eggs that might be chosen?

- I. Let 0, 1, and 2 represent the rotten eggs, and 3 – 11 the good eggs. Generate two random numbers 0 – 11, ignoring repeats.
 - II. Randomly generate a 0, 1, or 2 to represent the number of rotten eggs you get.
 - III. Since 25% of the eggs are rotten, let 0 = rotten and 1, 2, 3 = good. Generate two random numbers 0 – 3 and see how many 0's you get.
- A. I only
 - B. II only
 - C. III only
 - D. I or III
 - E. I, II, and III

Q.80: We call each time we obtain a simulated answer to our question a _____.

- A. random variable
- B. event
- C. trial
- D. Response

Chap.12

Q.81: A sample that consists of the entire population is called a:

- A. Simple sample
- B. Census
- C. Parameter
- D. Survey

Q.82: The natural tendency of randomly drawn samples to differ one from another is sometimes, unfortunately, called *sampling error*.

- A. True
- B. False

Q.83: Suppose your local school district decides to randomly test high school students for attention deficit disorder (ADD). There are three high schools in the district, each with grades 9-12. The school board pools all of the students together and randomly samples 250 students. Is this a simple random sample?

- A. Yes, because the students were chosen at random.
- B. Yes, because each student is equally likely to be chosen.
- C. Yes, because they could have chosen any 250 students from throughout the district.
- D. No, because we can't guarantee that there are students from each school in the sample.
- E. No, because we can't guarantee that there are students from each grade in the sample.

Q.84: Suppose the state decides to randomly test high school wrestlers for steroid use. There are 16 teams in the league, and each team has 20 wrestlers. State investigators plan to test 32 of these athletes by randomly choosing two wrestlers from each team. Is this a simple random sample?

- A. Yes, because the wrestlers were chosen at random.
- B. Yes, because each wrestler is equally likely to be chosen.
- C. Yes, because stratified samples are a type of simple random sample.
- D. No, because not all possible groups of 32 wrestlers could have been the sample.
- E. No, because a random sample of teams was not first chosen.

Q.85: The January 2005 Gallup Youth Survey telephoned a random sample of 1,028 U.S. teens aged 13-17 and asked these teens to name their favorite movie from 2004. *Napoleon Dynamite* had the highest percentage with 8% of teens ranking it as their favorite movie. Which is true?

- I. The population of interest is U.S. teens aged 13-17.
- II. 8% is a statistic and not the actual percentage of all U.S. teens who would rank this movie as their favorite.
- III. This sampling design should provide a reasonably accurate estimate of the actual percentage of all U.S. teens who would rank this movie as their favorite.

- A. I only
- B. II only
- C. III only
- D. I II, and III

Q.86: A chemistry professor who teaches a large lecture class surveys the students who attend his class on how he can make the class more interesting to get more students to attend. This survey method suffers from

- A. voluntary response bias
- B. nonresponse bias
- C. response bias
- D. undercoverage
- E. none of the above

Q.87: Which statement about bias is true?

- I. Bias results from random variation and will always be present.
 - II. Bias results from samples that do not represent the population.
 - III. Bias is usually reduced when sample size is larger.
- A. I only
 - B. II only
 - C. III only
 - D. I and III only, E. I, II, and III

Chap 13

Q.88: A company sponsoring a new Internet search engine wants to collect data on the ease of use of the new search engine. Which is the best way to collect the data?

- A. Census
- B. Sample survey
- C. Observational study
- D. Experiment

Q.89: We wish to compare the average ages of the math and science teachers at your high school. Which is the best way to collect the data?

- A. Census
- B. Sample survey
- C. Observational study
- D. Experiment

Q.90: More dogs are being diagnosed with thyroid problems than have been diagnosed in the past. A researcher identified 50 puppies not already diagnosed with a thyroid problem and followed the dogs for several years to see if any developed thyroid problems. This is a(n)

- A. Randomized experiment
- B. Survey
- C. Prospective study
- D. Retrospective study

Q.91: Which group(s) are affected by the placebo effect?

- A. The control group
- B. The treatment group
- C. Both the control and the treatment groups
- D. Neither the control nor the treatment group

Q.92: Double-blinding in experiments is important so that

- I. The evaluators do not know which treatment group the participants are in.
- II. The participants do not know which treatment group they are in.
- III. No one sees which treatment any of the participants are getting.

- A. I only
- B. II only
- C. III only
- D. I and II
- E. I, II, and III

Q.93: Which of the following is not required in an experimental design?

- A. Blocking
- B. Control
- C. Randomization
- D. Replication

Q.94: In an experiment the primary purpose of blinding is to reduce . . .

- A. bias.
- B. confounding.
- C. randomness.
- D. variation.

Q.95: In an experiment the primary purpose of blocking is to reduce . . .

- A. bias.
- B. confounding.
- C. randomness.
- D. variation.

Q.96: Does donating blood lower cholesterol levels? 50 volunteers have a cholesterol test, then donate blood, and then have another cholesterol test. Which aspect of experimental design is present?

- A. Randomization
- B. A control group
- C. A placebo
- D. Blinding
- E. None of these

Chap. 14

Q.97: Each occasion upon which we observe a random phenomenon is call a(n)

- A. trial
- B. event
- C. probability
- D. experiment

Q.98: For independent trials, the Law of Averages states that as the number of trials increases, the long run relative frequency of repeated events gets closer and closer to a single value.

- A. True
- B. False

Q.99: When the probability comes from the long run relative frequency of the event's occurrence, it is call a(n)...

- A. theoretical probability.
- B. random probability.
- C. personal probability.
- D. empirical probability.

Q.100: The probability of drawing a face card (JQK) from a deck of cards is

- A. $1/2$.
- B. $1/4$.
- C. $3/13$.
- D. $4/13$.

Q.101: A fair coin has come up "heads" 10 times in a row. The probability that the coin will come up heads on the next flip is

- A. less than 50%, since “tails” are due to come up.
- B. 50%.
- C. greater than 50%, since it appears that we are in a streak of “heads.”
- D. It cannot be determined.

Q.102: If $P(A)$ is 0.35, then $P(A^C)$ is

- A. 0
- B. 1
- C. 0.65
- D. 0.50

Q.103: Disjoint (mutually exclusive) sets are always independent.

- A. True
- B. False

Q.104: When we get to a stop light, it has to be either red, green or yellow. The $P(\text{red}) = 0.61$, $P(\text{green}) = 0.35$, and $P(\text{yellow}) = 0.04$. You travel this intersection every day. What is the probability the light will be yellow two days in a row when you arrive?

- A. 0.08
- B. 0.0016
- C. 0.96
- D. Cannot be determined

Chap. 15

Q.105: In an intro stats class, 57% of students eat breakfast in the morning and 80% of students floss their teeth. Forty-six percent of students eat breakfast and floss their teeth. What is the probability that a student from this class eats breakfast but does not floss their teeth?

- A. 9%
- B. 11%
- C. 34%
- D. 57%

Q.106: In an intro stats class, 57% of students eat breakfast in the morning and 80% of students floss their teeth. Forty-six percent of students eat breakfast and floss their teeth. What is the probability that a student from this class eats breakfast or flosses their teeth?

- A. 9%
- B. 11%
- C. 34%
- D. 91%

Q.107: Five juniors and four seniors have applied for two open student council positions. School administrators have decided to pick the two new members randomly. What is the probability that one junior and one senior are chosen for the two positions?

- A. 16.7%
- B. 27.8%
- C. 44.4%
- D. 55.6%

Q.108: Insurance company records indicate that 12% of all teenage drivers have been ticketed for speeding and 9% for going through a red light. If 4% have been ticketed for both, what is the probability that a teenage driver has been issued a ticket for speeding but not for running a red light?

- A. 3%
- B. 8%
- C. 12%
- D. 13%

Q.109: Which two events are most likely to be independent?

- A. being a senior; going to homeroom
- B. registering to vote; being left-handed
- C. having a car accident; having a junior license
- D. doing the Statistics homework; getting an A on the test
- E. having 3 inches of snow in the morning; being on time for school

Q.110: A poll of 120 Ithacans found that 30 had visited the new PRI museum, and that 80 had been to Home Depot. If it appeared that going to Home Depot and going to the museum were independent events, how many of those polled had been to both?

		Museum		Total
		Yes	No	
Home Depot	Yes	??		80
	No			40
Total		30	90	120

- A) 10
- B) 15
- C) 20
- D) It cannot be determined.

Q.111: A recent survey said that 56% of college students live on campus, 62% have a campus meal program, and 42% do both. Living on campus and having a meal plan are:

- A. Independent
- B. Disjoint
- C. Both independent and disjoint
- D. Neither independent nor disjoint

Q.112: Given a standard deck of 52 cards, what is the probability of drawing two face cards (JQK) in a row if the cards are drawn without replacement?

- A. 9/169
- B. 3/13
- C. 11/221
- D. 1/26

Chap.16

Q.113: A random variable where we can list all the outcomes is called a discrete random variable.

- A. True
- B. False

Q.114: Find the expected value using the chart below:

x	10	20	30
$P(X = x)$	0.3	0.5	0.2

- A. 1.0
- B. 30
- C. 19
- D. 0.5

Q.115: A wheel comes up green 50% of the time and red 50% of the time. If it comes up green, you win \$100, if it comes up red you win nothing. Intuitively, how much do you expect to win on one spin, on average?

- A. \$0
- B. \$50
- C. \$100
- D. Intuition does not play a part in statistics.

Q.116: BatCo, a company that sells batteries, claims that 99.5% of their batteries are in working order. How many batteries would you expect to buy, on average, to find one that does not work?

- A. 5
- B. 199
- C. 200
- D. 994
- E. 995

Q.117: A friend of yours plans to toss a fair coin 200 times. You watch the first 40 tosses, noticing that she got only 16 heads. But then you get bored and leave. If the coin is fair, how many heads do you expect her to have when she has finished the 200 tosses?

- A. 80
- B. 92
- C. 96
- D. 100

Q.118: Some marathons allow two runners to “split” the marathon by each running a half marathon. Alice and Sharon plan to “split” a marathon. Alice’s half marathon times average 92 minutes with a standard deviation of 4 minutes, and Sharon’s half-marathon times average 96 minutes with a standard deviation of 2 minutes. Assume that the women’s half marathon times are independent. The expected time for Alice and Sharon to complete a full marathon is $92 + 96 = 188$ minutes. What is the standard deviation of the total?

- A. 2 minutes
- B. 4.5 minutes
- C. 6 minutes
- D. 20 minutes

Q.119: Assume the heights of high school basketball players are normally distributed. For boys the mean is 74 inches with a standard deviation of 4.5 inches, while girl players have a mean height of 70 inches and standard deviation 3 inches. At a mixed 2-on-2 tournament teams are formed by randomly pairing boys with girls as teammates.

On average, how much taller do you expect the boy to be?

- A. 4"
- B. 2"
- C. 7.5"
- D. Cannot be determined

Q.120: Assume the heights of high school basketball players are normally distributed. For boys the mean is 74 inches with a standard deviation of 4.5 inches, while girl players have a mean height of 70 inches and standard deviation 3 inches. At a mixed 2-on-2 tournament teams are formed by randomly pairing boys with girls as teammates.

What will the variance be?

- A. 5.4
- B. 7.5
- C. 29.25
- D. Cannot be determined

Chap.17

Q.121: Bernoulli trials must be independent. If that assumption is violated, it is still okay to proceed as long as the sample is smaller than _____ % of the population

- A. 50
- B. 25
- C. 15
- D. 10

Q.122: Which of these has a Binomial model?

- A. The number of people we survey until we find someone who has taken Statistics
- B. The number of people in a class who have taken Statistics
- C. The number of aces in a five-card Poker hand
- D. The number of sodas students drink per day

Q.123: A Binomial model is approximately Normal if we expect at least

- A. 5 successes and 5 failures.
- B. 15 successes and 15 failures.
- C. 10 successes and 10 failures.
- D. 20 successes and 20 failures.

Q.124: The National Association of Retailers reports that 62% of all purchases are now made by credit card; you think this is true at your store as well. On a typical day you make 20 sales. What is the probability that your fourth customer is the first one who uses a credit card?

- A. 0.50

- B. 0.148
- C. 0.25
- D. 0.034

Q.125: According to Infoplease, 18.8% of the luxury cars manufactured in 2003 were silver. A large car dealership typically sells 50 luxury cars a month. If X represents the number of silver luxury cars sold in a typical month and the model is Binomial, what is the mean?

- A. 7.63
- B. 9.4
- C. 40.6
- D. 2.76

Q.126: According to Infoplease, 18.8% of the luxury cars manufactured in 2003 were silver. A large car dealership typically sells 50 luxury cars a month. If X represents the number of silver luxury cars sold in a typical month and the model is Binomial, what is the standard deviation?

- A. 7.63
- B. 9.4
- C. 40.6
- D. 2.76

Q.127: A Poisson model can be used to approximate a Binomial model when

- A. p is large and n is small.
- B. p is equal to n .
- C. p is small and n is large.
- D. The Poisson model cannot be used to approximate a Binomial model.

Chap.18

Q.128: The sampling distribution and distribution of the sample are synonymous.

- A. True
- B. False

Q.129: The variability we expect to see from one random sample to another is called _____ .

- A. the variance
- B. the standard deviation
- C. the sampling error
- D. the distribution variance

Q.130: The Central Limit Theorem (CLT) states that the sampling distribution model of the sample mean (and proportion) from a random sample is approximately Normal for any n , regardless of the distribution of the population, as long as the observations are independent.

- A. True
- B. False

Q.131: A certain population is approximately normal. We want to estimate its mean, so we will collect a sample. Which should be true if we use a large sample rather than a small one?

- I. The distribution of our sample data will be approximately normal.
 - II. The sampling distribution of the sample means will be approximately normal.
 - III. The variability of the sample means will be smaller.
- A. I only
 - B. II only
 - C. III only
 - D. II and III
 - E. I, II, and III

Q.132: Which of the following is not an assumption or condition that needs to be checked for the one-proportion z-test?

- A. Independence
- B. Randomization
- C. 10% Condition
- D. Success/Failure Condition
- E. Nearly Normal Condition

Q.133: A certain population is strongly skewed to the right. We want to estimate its mean, so we will collect a sample. Which should be true if we use a large sample rather than a small one?

- I. The distribution of our sample data will be closer to normal.
 - II. The sampling model of the sample means will be closer to normal.
 - III. The variability of the sample means will be greater.
- A. I only
 - B. II only
 - C. III only
 - D. I and III only
 - E. II and III only

Q.134: It is generally believed that electrical problems affect about 14% of new cars. An automobile mechanic conducts diagnostic tests on 128 new cars on the lot. How many successes (electrical failures) do you expect?

- A. 1.28
- B. 17.92
- C. 110.08
- D. Cannot determine from information given

Q.135: It is generally believed that electrical problems affect about 14% of new cars. An automobile mechanic conducts diagnostic tests on 128 new cars on the lot. What is the standard deviation?

- A. 0.347
- B. 0.001
- C. 0.031
- D. Cannot determine from information given.

Chap. 19

Q.136: The higher the level of confidence we want, the narrower our confidence interval becomes.

- A. True
- B. False

Q.137: The number of standard errors to move away from the mean of the sampling distribution to correspond to the specified level of confidence is called the _____.

- A. margin of error
- B. critical value
- C. confidence interval
- D. z-interval

Q.138: The Success/Failure Condition states that we must expect at least ____ “successes” and ____ “failures.”

- A. 5
- B. 10
- C. 15
- D. 25

Q.139: We have calculated a 95% confidence interval and would prefer for our next confidence interval to have a smaller margin of error without losing any confidence. In order to do this, we can

- I. change the z^* value to a smaller number.
 - II. take a larger sample.
 - III. take a smaller sample.
- A. I only
 - B. II only
 - C. III only
 - D. I and II
 - E. I and III

Q.140: Which is true about a 98% confidence interval for a population proportion based on a given sample?

- I. We are 98% confident that the sample proportion is in our interval.
 - II. There is a 98% chance that our interval contains the population proportion.
 - III. For the same sample, the interval is wider than a 95% confidence interval would be.
- A. None
 - B. I only
 - C. II only
 - D. III only

E. I and II

Q.141: We have calculated a confidence interval based on a sample of size $n = 100$. Now we want to get a better estimate with a margin of error that is only one-fourth as large. How large does our new sample need to be?

- A) 25
- B) 50
- C) 200
- D) 400
- E) 1600

Q.142: A newspaper article reported that a poll based on a sample of 1150 residents of a state showed that the state's Governor's job approval rating stood at 58%. They claimed a margin of error of $\pm 3\%$. What level of confidence were the pollsters using?

- A. 90%
- B. 99%
- C. 96%
- D. Cannot determine from information given.

Q.143: According to a September 2004 Gallup poll, about 73% of 18- to 29-year-olds said that they were registered to vote. A statistics professor asked her students whether or not they were registered to vote. In a sample of 50 of her students (randomly sampled from her 700 students), 35 said they were registered to vote. The 95% confidence interval for the true proportion of the professor's students who were registered to vote is given as

$$\hat{p} \pm z * SE(\hat{p}) = 0.70 \pm 1.96(0.065) = 0.70 \pm 0.127 = (0.573, 0.827)$$

i.e. We are 95% confident that between 57.3% and 82.7% of the professor's students are registered to vote. Now, what is the probability that the true proportion of the professor's students who were registered to vote is in your confidence interval?

- A. 95%
- B. 50%
- C. 100%
- D. There is no probability involved.

Q.144: A researcher wants to estimate the proportion of people who report the side effect of nausea when taking a drug to reduce anxiety. Of 25 people who take the drug, 8 report nausea. What can the researcher say with 90% confidence?

- A. Between 16% & 46% of people who take the drug will get the nauseas.
- B. Between 20% & 50% of people who take the drug will get the nauseas.
- C. Between 17% & 47% of people who take the drug will get the nauseas

Q.145: A News blog reported that a poll based on a sample of 800 voters the president's job approval rating stood at 62%. They claimed a ME of $\pm 3\%$. What levels of confidence were the pollsters using?

- A. 90%

- B. 92%
- C. 95%
- D. 97%

Chap.20

Q.146: A P -value indicates

- A. the probability that the null hypothesis is true.
- B. the probability that the alternative hypothesis is true.
- C. the probability of the observed statistic given that the null hypothesis is true.
- D. the probability of the observed statistic given that the alternative hypothesis is true.

Q.147: A small P -value indicates either that the observation is improbable or that the probability calculation was based on incorrect assumptions.

- A. True
- B. False

Q.148: In a hypothesis test, we check the Success/Failure Condition with the observed proportions.

- A. True
- B. False

Q.149: In a hypothesis test, the null hypothesis represents the status quo.

- A. True
- B. False

Q.150: According to a June 2004 Gallup poll, 28% of Americans “said there have been times in the last year when they haven’t been able to afford medical care.” Is this proportion higher for black Americans than for all Americans? In a random sample of 801 black Americans, 38% reported that there had been times in the last year when they had not been able to afford medical care. Which type of hypothesis test would you use?

- A. One-tail upper tail
- B. One-tail lower tail
- C. Two-tail
- D. Both A and B

Q.151: An online catalog company wants on-time delivery for 90% of the orders they ship. They have been shipping orders via UPS and FedEx but will switch to a new, cheaper delivery service (ShipFast) unless there is evidence that this service cannot meet the 90% on-time goal. As a test the company sends a random sample of orders via ShipFast, and then makes follow-up phone calls to see if these orders arrived on time. Which hypotheses should they test?

- A. $H_0: p < 0.90$ $H_A: p > 0.90$
- B. $H_0: p = 0.90$ $H_A: p > 0.90$
- C. $H_0: p > 0.90$ $H_A: p = 0.90$
- D. $H_0: p = 0.90$ $H_A: p < 0.90$

Q.152-155: Since many people have trouble programming their DVD players, an electronic company has developed what it hopes will be easier instructions. The goal is to have at least 95% of customers

succeed at being able to program their DVD players. The company tests the new system on 100 people, 92 of whom were successful. Is this strong evidence that the new system fails to meet the company's goal? A student's test of this hypothesis is provided. How many mistakes can you find?

Student's Test of Hypothesis

The student's test of the hypothesis is shown here.

$$H_0: p = 0.95 \quad \& \quad H_A: p > 0.95$$

SRS (Simple Random, Sample), $0.95(100) > 10$, $0.05(100) > 10$

$$\frac{92}{100} = 0.92; \quad SD(\hat{P}) = \sqrt{\frac{(0.95)(0.05)}{100}} = 0.022$$

$$z = \frac{0.92 - 0.95}{0.022} = -1.36$$

$$p = p(Z < -1.36) = 0.09$$

There is weak evidence that the new system does not work.

- Identify any mistakes in the student's statements of the null and alternative hypotheses. If there are any mistakes, show what the correct statements for the null and alternative hypotheses should be. Choose the correct answer below.
 - $H_0: p = 0.92$; $H_A: p > 0.92$
 - $H_0: p = 0.92$; $H_A: p < 0.92$
 - The student's statements for the null and alternative hypotheses are correct
 - $H_0: p = 0.95$; $H_A: p < 0.95$
- Identify any mistakes in the student's verification of assumptions and conditions. If there are any mistakes, state what the correct verification of assumptions and conditions should be. Choose the correct answer below.
 - The student's verification of assumptions and conditions is correct.
 - Unnecessary information is given; SRS (Simple Random Sample)
 - Unnecessary information is given; SRS (Simple Random Sample), 100 people < 10% of all people that program DVD players
 - The calculation of nq is incorrect and the 10% condition is missing; SRS (Simple Random Sample), $0.95(100) > 10$, $0.05(100) < 10$, 100 people < 10% of all people that program DVD players
- Identify any mistakes in the student's calculation of the P-value. If there are any mistakes, show what the correct calculations for the P-value should be. Choose the correct answer below.
 - The calculation of the probability is incorrect; $P = P(z > -1.36) = 0.91$
 - The student's calculation of $SD(\hat{p})$ is incorrect and carries through the calculation of the P-value; $SD(\hat{p}) = \sqrt{\frac{(0.92)(0.08)}{100}} = 0.0271$; $z = \frac{0.92 - 0.95}{0.0271} = -1.11$; $p = p(z < -1.11) = 0.09$
 - The student's calculation of the P-value is correct.
 - The calculation of the probability is incorrect; $P = P(z > -1.36) = 0.09$
- Identify any mistakes in the student's conclusion. If there are any mistakes, state what the correct conclusion should be assumed all requirements for the test are satisfied. Choose the correct answer below.
 - There is strong evidence that the new system does not work.
 - There is no evidence that the new system works.
 - There is strong evidence that the new system works.
 - The student's conclusion is correct.

Chap. 21

Q.156: The threshold P -value that determines when we reject a null hypothesis is called the

- A. Alpha value. B. power. C. beta value. D. level of significance.

Q.157: We commit a Type II error when the null hypothesis is true, but we mistakenly reject it.

- A. True B. False

Q.158: Suppose that a manufacturer is testing one of its machines to make sure that the machine is producing more than 97% good parts ($H_0: p = 0.97$ and $H_A: p > 0.97$). The test results in a P -value of 0.102. In reality, the machine is producing 99% good parts. What probably happens as a result of our testing?

- A. We correctly fail to reject H_0 .
B. We correctly reject H_0 .
C. We reject H_0 , making a Type I error.
D. We fail to reject H_0 , making a Type I error.
E. We fail to reject H_0 , making a Type II error.

Q.159: Which of the following is true about Type I and Type II errors?

- I. Type I errors are always worse than Type II errors.
II. Type II errors are always worse than Type I errors.
III. The severity of Type I and Type II errors depends on the situation being tested.
A. I only B. II only C. III only D. I and II

Q.160: We are about to test a hypothesis using data from a well-designed study. Which is true?

- I. A large P -value would be strong evidence against the null hypothesis.
II. We can set a higher standard of proof by choosing $\alpha = 10\%$ instead of 5%.
III. If we reduce the risk of committing a Type I error, then the risk of a Type II error will also decrease.
A) None
B) I only
C) II only
D) III only
E) I and II only

Q.161: Suppose that a device advertised to increase a car's gas mileage really does not work. We test it on a small fleet of cars (with H_0 : not effective), and our data results in a P -value of 0.004. What probably happens as a result of our experiment?

- A. We correctly fail to reject H_0 .
B. We correctly reject H_0 .
C. We reject H_0 , making a Type I error
D. We reject H_0 , making a Type II error.
E. We fail to reject H_0 , committing Type II error.

Q.162: We will test the hypothesis that $p = 60\%$ versus $p > 60\%$. We don't know it, but actually p is 70% . With which sample size and significance level will our test have the greatest power?

- A. $\alpha = 0.01, n = 200$
- B. $\alpha = 0.01, n = 500$
- C. $\alpha = 0.05, n = 200$
- D. $\alpha = 0.05, n = 500$
- E. The power will be the same as long as the true proportion P remains 70% .

Q.163: We are about to test a hypothesis using data from a well designed study. Which is true?

- I. A small P-value would be strong evidence against H_0
 - II. We can set a higher standard of proof by choosing $\alpha = 10\%$ instead of 5%
 - III. If we reduce the α level, we reduce the power of the test
- A. None
 - B. I only
 - C. II only
 - D. III only
 - E. I & III

Hurray!! You're done.