Practice Problems and Exams
SECTION I: MULTIPLE-CHOICE

1. The universe or "totality of items or things" under consideration is called
   a) a sample.
   b) a population.
   c) a parameter.
   d) a statistic.

2. A summary measure that is computed to describe a characteristic of an entire
   population is called
   a) a parameter.
   b) a census.
   c) a statistic.
   d) the scientific method.

3. Which of the following is a discrete quantitative variable?
   a) The Dow Jones Industrial average
   b) The volume of water released from a dam
   c) The distance you drove yesterday.
   d) The number of employees of an insurance company

4. The classification of student class designation (freshman, sophomore, junior, senior)
   is an example of
   a) a categorical random variable.
   b) a discrete random variable.
   c) a continuous random variable.
   d) a parameter.
SECTION II: TRUE OR FALSE

1. A sample is the portion of the universe that is selected for analysis.
   True

2. The possible responses to the question “How long have you been living at your current residence?” are values from a continuous variable.
   True

3. A continuous variable may take on any value within its relevant range even though the measurement device may not be precise enough to record it.
   True

4. Student grades (A to F) are an example of continuous numerical data.
   False

5. A statistic is usually unobservable while a parameter is usually observable.
   False

6. The answer to the question “How do you rate the quality of your business statistics course” is an example of an ordinal scaled variable.
   True

7. The type of TV one owns is an example of a numerical variable.
   False

8. Whether the university is private or public is an example of a categorical variable.
   True

9. The grade level (K-12) of a student is an example of a nominal scaled variable.
   False

10. The amount of water consumed by a person per week is an example of a continuous variable.
    True

11. To learn to use statistical programs, you need to understand the underlying statistical concepts.
    True
CHAPTER 3 - NUMERICAL DESCRIPTIVE MEASURES

Practice Exam - Solution

SECTION I: MULTIPLE-CHOICE

1. Which of the following statistics is not a measure of central tendency?
   a) Arithmetic mean.
   b) Median.
   c) Mode.
   d) Q₃.

2. Which of the arithmetic mean, median, and mode are resistant measures of central tendency?
   a) The arithmetic mean and median only.
   b) The median and mode only.
   c) The mode and arithmetic mean only.
   d) All the three are resistant measures.

3. Which of the following statements about the median is not true?
   a) It is more affected by extreme values than the arithmetic mean.
   b) It is a measure of central tendency.
   c) It is equal to Q₂.
   d) It is equal to the mode in bell-shaped "normal" distributions.

4. In a perfectly symmetrical distribution
   a) the range equals the interquartile range.
   b) the interquartile range equals the arithmetic mean.
   c) the median equals the arithmetic mean.
   d) the variance equals the standard deviation.

5. In general, which of the following descriptive summary measures cannot be easily approximated from a boxplot?
   a) The variance.
   b) The range.
   c) The interquartile range.
   d) The median.

6. Which descriptive summary measures are considered to be resistant statistics?
   a) The arithmetic mean and standard deviation.
   b) The interquartile range and range.
   c) The mode and variance.
   d) The median and interquartile range.
7. In perfectly symmetrical distributions, which of the following is NOT a correct statement?
   a) The distance from Q1 to Q2 equals to the distance from Q2 to Q3.
   b) The distance from the smallest observation to Q1 is the same as the distance from Q3 to the largest observation.
   c) The distance from the smallest observation to Q2 is the same as the distance from Q2 to the largest observation.
   d) **The distance from Q1 to Q3 is half of the distance from the smallest to the largest observation.**

8. Which of the following is NOT sensitive to extreme values?
   a) The range.
   b) The standard deviation.
   c) **The interquartile range.**
   d) The coefficient of variation.

9. According to the Chebyshev rule, at least 75% of all observations in any data set are contained within a distance of how many standard deviations around the mean?
   a) 1
   b) 2
   c) 3
   d) 4

10. According to the Chebyshev rule, at least what percentage of the observations in any data set are contained within a distance of 2 standard deviations around the mean?
    a) 67%
    b) 75%
    c) 88.89%
    d) 95%

**SECTION II: TRUE OR FALSE**

1. The median of the values 3.4, 4.7, 1.9, 7.6, and 6.5 is 1.9.
   **False**

2. In a set of numerical data, the value for Q3 can never be smaller than the value for Q1.
   **True**

3. In right-skewed distributions, the distance from Q3 to the largest observation exceeds the distance from the smallest observation to Q1.
   **True**

4. A boxplot is a graphical representation of a 5-number summary.
   **True**

5. The line drawn within the box of the boxplot always represents the median.
   **True**

6. A population with 200 elements has an arithmetic mean of 10. From this information, it can be shown that the population standard deviation is 15.
   **False**
7. An economics professor bases his final grade on homework, two midterm examinations, and a final examination. The homework counts 10% toward the final grade, while each midterm examination counts 25%. The remaining portion consists of the final examination. If a student scored 95% in homework, 70% on the first midterm examination, 96% on the second midterm examination, and 72% on the final, his final average is 79.8%.

True

8. The coefficient of variation measures variability in a data set relative to the size of the arithmetic mean.

True

9. The coefficient of variation is a measure of central tendency in the data.

False

10. If a set of data is perfectly symmetrical, the arithmetic mean must be identical to the median.

True

11. If the data set is approximately bell-shaped, the empirical rule will more accurately reflect the greater concentration of data close to the mean as compared to the Chebyshev rule.

True

12. The Z scores can be used to identify outliers.

True

13. As a general rule, an observation is considered an extreme value if its Z score is greater than −3.

False

14. As a general rule, an observation is considered an extreme value if its Z score is less than −3.

True

15. The Z score of an observation measures how many standard deviations is the value from the mean.

True

TABLE (A)
Health care issues are receiving much attention in both academic and political arenas. A sociologist recently conducted a survey of citizens over 60 years of age whose net worth is too high to qualify for Medicaid and have no private health insurance. The ages of 25 uninsured senior citizens were as follows:

<table>
<thead>
<tr>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
</tr>
<tr>
<td>61</td>
</tr>
<tr>
<td>62</td>
</tr>
<tr>
<td>63</td>
</tr>
<tr>
<td>64</td>
</tr>
<tr>
<td>65</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>68</td>
</tr>
<tr>
<td>68</td>
</tr>
<tr>
<td>69</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>73</td>
</tr>
<tr>
<td>73</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>76</td>
</tr>
<tr>
<td>76</td>
</tr>
<tr>
<td>81</td>
</tr>
<tr>
<td>81</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>86</td>
</tr>
<tr>
<td>87</td>
</tr>
<tr>
<td>89</td>
</tr>
<tr>
<td>90</td>
</tr>
<tr>
<td>92</td>
</tr>
</tbody>
</table>

1. Referring to TABLE (A), calculate the arithmetic mean age of the uninsured senior citizens to the nearest hundredth of a year.

74.04 years

2. Referring to TABLE (A), identify the median age of the uninsured senior citizens.
73 years

3. Referring to TABLE (A), identify the first quartile of the ages of the uninsured senior citizens.
65.5 years

4. Referring to TABLE (A), identify the third quartile of the ages of the uninsured senior citizens.
81.5 years

5. Referring to TABLE (A), identify the interquartile range of the ages of the uninsured senior citizens.
16 years

6. Referring to TABLE (A), identify which of the following is the correct statement.
   a) **One fourth of the senior citizens sampled are below 65.5 years of age.**
   b) The middle 50% of the senior citizens sampled are between 65.5 and 73.0 years of age.
   c) The average age of senior citizens sampled is 73.5 years of age.
   d) All of the above are correct.

7. Referring to TABLE (A), identify which of the following is the correct statement.
   a) One fourth of the senior citizens sampled are below 64 years of age.
   b) The middle 50% of the senior citizens sampled are between 65.5 and 73.0 years of age.
   c) **25% of the senior citizens sampled are older than 81.5 years of age.**
   d) All of the above are correct.

8. Referring to TABLE (A), what type of shape does the distribution of the sample appear to have?
   Slightly positive or right-skewed.

9. Referring to TABLE (A), calculate the variance of the ages of the uninsured senior citizens correct to the nearest hundredth of a year squared.
   94.96 years$^2$

10. Referring to TABLE (A), calculate the standard deviation of the ages of the uninsured senior citizens correct to the nearest hundredth of a year.
    9.74 years

11. Referring to TABLE (A), calculate the coefficient of variation of the ages of the uninsured senior citizens.
    13.16%
**An Introduction to Statistics Course (ECOE 1302)**
Spring Semester 2009-2010

**Chapter 6 - The Normal Distribution**
**Practice Exam - Solution**

Instructors: Dr. Samir Safi     Mr. Ibrahim Abed

**SECTION I: MULTIPLE-CHOICE**

1. In its standardized form, the normal distribution
   a. has a mean of 0 and a standard deviation of 1.
   b. has a mean of 1 and a variance of 0.
   c. has an area equal to 0.5.
   d. cannot be used to approximate discrete probability distributions.

2. If a particular batch of data is approximately normally distributed, we would find that approximately
   a. 2 of every 3 observations would fall between \( \pm 1 \) standard deviation around the mean.
   b. 4 of every 5 observations would fall between \( \pm 1.28 \) standard deviations around the mean.
   c. 19 of every 20 observations would fall between \( \pm 2 \) standard deviations around the mean.
   d. All the above.

3. For some value of \( Z \), the probability that a standard normal variable is below \( Z \) is 0.2090. The value of \( Z \) is
   a. \(-0.81\)
   b. \(-0.31\)
   c. 0.31
   d. 1.96

4. For some positive value of \( X \), the probability that a standard normal variable is between 0 and \(+2X\) is 0.1255. The value of \( X \) is
   a. 0.99
   b. 0.40
   c. 0.32
   d. 0.16

5. If we know that the length of time it takes a college student to find a parking spot in the library parking lot follows a normal distribution with a mean of 3.5 minutes and a standard deviation of 1 minute, find the probability that a randomly selected college student will take between 2 and 4.5 minutes to find a parking spot in the library parking lot.
   i. 0.0919
   ii. 0.2255
   iii. 0.4938
6. The owner of a fish market determined that the average weight for a catfish is 3.2 pounds with a standard deviation of 0.8 pound. A citation catfish should be one of the top 2% in weight. Assuming the weights of catfish are normally distributed, at what weight (in pounds) should the citation designation be established?
   a. 1.56 pounds
   b. **4.84 pounds**
   c. 5.2 pounds
   d. 7.36 pounds

**Section II: TRUE Or FALSE**

1. The probability that a standard normal random variable, \( Z \), falls between –1.50 and 0.81 is 0.7242.
   **True**

2. The probability that a standard normal random variable, \( Z \), is below 1.96 is 0.4750.
   **False**

3. The probability that a standard normal random variable, \( Z \), falls between –2.00 and –0.44 is 0.6472.
   **False**

4. A worker earns $15 per hour at a plant and is told that only 2.5% of all workers make a higher wage. If the wage is assumed to be normally distributed and the standard deviation of wage rates is $5 per hour, the average wage for the plant is $7.50 per hour.
   **False**

5. Any set of normally distributed data can be transformed to its standardized form.
   **True**

6. The "middle spread," that is the middle 50% of the normal distribution, is equal to one standard deviation.
   **False**

7. If a data batch is approximately normally distributed, its normal probability plot would be S-shaped.
   **False**

**SECTION III: FREE RESPONSE PROBLEMS**

1. A company that sells annuities must base the annual payout on the probability distribution of the length of life of the participants in the plan. Suppose the probability distribution of the lifetimes of the participants is approximately a normal distribution with a mean of 68 years and a standard deviation of 3.5 years.
   a. What proportion of the plan recipients would receive payments beyond age 75?
      
      **0.0228**

   b. Find the age at which payments have ceased for approximately 86% of the plan participants.
2. A food processor packages orange juice in small jars. The weights of the filled jars are approximately normally distributed with a mean of 10.5 ounces and a standard deviation of 0.3 ounce. Find the proportion of all jars packaged by this process that have weights that fall below 10.875 ounces.

\[ \text{Proportion} = 0.8944 \]

3. The owner of a fish market determined that the average weight for a catfish is 3.2 pounds with a standard deviation of 0.8 pound. Assuming the weights of catfish are normally distributed, the probability that a randomly selected catfish will weigh less than 2.2 pounds is _______?

\[ \text{Probability} = 0.1056 \]

4. The amount of pyridoxine (in grams) in a multiple vitamin is normally distributed with \( \mu = 110 \) grams and \( \sigma = 25 \) grams.

a. What is the probability that a randomly selected vitamin will contain between 100 and 110 grams of pyridoxine? 0.1554

b. What is the probability that a randomly selected vitamin will contain between 82 and 100 grams of pyridoxine? 0.2132

c. What is the probability that a randomly selected vitamin will contain at least 100 grams of pyridoxine? 0.6554

d. What is the probability that a randomly selected vitamin will contain between 100 and 120 grams of pyridoxine? 0.3108

e. What is the probability that a randomly selected vitamin will contain less than 100 grams of pyridoxine? 0.3446

f. What is the probability that a randomly selected vitamin will contain less than 100 grams or more than 120 grams of pyridoxine? 0.6892

g. Approximately 83% of the vitamins will have at least how many grams of pyridoxine? 86.25 using Table E.2

5. You were told that the amount of time lapsed between consecutive trades on the New York Stock Exchange followed a normal distribution with a mean of 15 seconds. You were also told that the probability that the time lapsed between two consecutive trades to fall between 16 to 17 seconds was 13%. The probability that the time lapsed between two consecutive trades would fall below 13 seconds was 7%.

a. What is the probability that the time lapsed between two consecutive trades will be longer than 17 seconds? 7% or 0.07

b. What is the probability that the time lapsed between two consecutive trades will be longer than 17 seconds?7% or 0.07

c. What is the probability that the time lapsed between two consecutive trades will be between 13 and 14 seconds?13% or 0.13

d. What is the probability that the time lapsed between two consecutive trades will be between 15 and 16 seconds? 30% or 0.30

e. What is the probability that the time lapsed between two consecutive trades will be between 14 and 15 seconds?30% or 0.30

f. What is the probability that the time lapsed between two consecutive trades will be between 13 and 16 seconds?73% or 0.73
g. What is the probability that the time lapsed between two consecutive trades will be between 14 and 17 seconds? 73% or 0.73

h. The probability is 20% that the time lapsed will be shorter how many seconds? 14 seconds

i. The probability is 80% that the time lapsed will be longer than how many seconds? 14 seconds

j. The middle 60% of the time lapsed will fall between which two numbers? 14 seconds and 16 seconds

k. The middle 86% of the time lapsed will fall between which two numbers? 13 seconds and 17 seconds

6. Suppose Z has a standard normal distribution with a mean of 0 and standard deviation of 1.
   a. The probability that Z is less than 1.15 is _________. **0.8749**
   b. The probability that Z is less than 1.15 is _________. **0.8749**
   c. The probability that Z is more than 0.77 is _________. **0.2206**
   d. The probability that Z is less than -2.20 is _________. **0.0139**
   e. The probability that Z is more than -0.98 is _________. **0.8365**
   f. The probability that Z is between -2.33 and 2.33 is _________. **0.9802**
   g. The probability that Z is between -2.89 and -1.03 is _________. **0.1496**
   h. The probability that Z is between -0.88 and 2.29 is _________. **0.7996**
   i. The probability that Z values are larger than _________ is 0.3485. **0.39**
   j. The probability that Z values are larger than _________ is 0.6985. **-0.52**
   k. So 27% of the possible Z values are smaller than _________ **-0.61**
   l. So 85% of the possible Z values are smaller than _________ **1.04**
   m. So 96% of the possible Z values are between _________ and _________ (symmetrically distributed about the mean). **-2.05 and 2.05 or -2.06 and 2.06**
   n. So 50% of the possible Z values are between _________ and _________ (symmetrically distributed about the mean). **-0.67 and 0.67 or -0.68 and 0.68**

7. The owner of a fish market determined that the average weight for a catfish is 3.2 pounds. He also knew that the probability of a randomly selected catfish that would weigh more than 3.8 pounds is 20% and the probability that a randomly selected catfish that would weigh less than 2.8 pounds is 30%.
   a. The probability that a randomly selected catfish will weigh less than 3.6 pounds is _______. **70% or 0.7**
   b. The probability that a randomly selected catfish will weigh between 2.6 and 3.6 pounds is _______. **50% or 0.5**
   c. The middle 40% of the catfish will weigh between _____ pounds and _____ pounds. **2.8 and 3.6**

**TABLE (A)**
The manager of a surveying company believes that the average number of phone surveys completed per hour by her employees has a normal distribution. She takes a sample of 15 days output from her employees and determines the average number of surveys per hour on these days. The ordered array for this data is: 10.0, 10.1, 10.3, 10.5, 10.7, 11.2, 11.4, 11.5, 11.7, 11.8, 11.8, 12.0, 12.2, 12.2, 12.5.
a. Referring to Table (A), the first standard normal quantile is ________.-1.5341
b. Referring to Table (A), the fourth standard normal quantile is ________.-0.6745
c. Referring to Table (A), the ninth standard normal quantile is ________.+0.1573
d. Referring to Table (A), the fourteenth standard normal quantile is ________.+1.1503
e. Referring to Table (A), the last standard normal quantile is ________.+1.5341

Table (B)

The number of column inches of classified advertisements appearing on Mondays in a certain daily newspaper is normally distributed with population mean 320 and population standard deviation 20 inches.

a. Referring to Table (B), for a randomly chosen Monday, what is the probability there will be less than 340 column inches of classified advertisement? 0.8413
b. Referring to Table (B), for a randomly chosen Monday, what is the probability there will be between 280 and 360 column inches of classified advertisement? 0.9544 using Table E.2
c. Referring to Table (B), for a randomly chosen Monday the probability is 0.1 that there will be less than how many column inches of classified advertisements? 294.4
d. Referring to Table (B), a single Monday is chosen at random. State in which of the following ranges the number of column inches of classified advertisement is most likely to be:
   i.  300 --320
   ii. 310 --330
   iii. 320 -- 340
   iv. 330 -- 350
SECTION I: MULTIPLE-CHOICE

1. Sampling distributions describe the distribution of
   a) parameters.
   b) statistics.
   c) both parameters and statistics.
   d) neither parameters nor statistics.

2. The Central Limit Theorem is important in statistics because
   a) for a large $n$, it says the population is approximately normal.
   b) for any population, it says the sampling distribution of the sample mean is approximately normal, regardless of the sample size.
   c) for a large $n$, it says the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.
   d) for any sized sample, it says the sampling distribution of the sample mean is approximately normal.

3. Which of the following statements about the sampling distribution of the sample mean is incorrect?
   a) The sampling distribution of the sample mean is approximately normal whenever the sample size is sufficiently large ($n \geq 30$).
   b) The sampling distribution of the sample mean is generated by repeatedly taking samples of size $n$ and computing the sample means.
   c) The mean of the sampling distribution of the sample mean is equal to $\mu$.
   d) The standard deviation of the sampling distribution is equal to $\sigma$.

4. Which of the following is true about the sampling distribution of the sample mean?
   a) The mean of the sampling distribution is always $\mu$.
   b) The standard deviation of the sampling distribution is always $\sigma$.
   c) The shape of the sampling distribution is always approximately normal.
   d) All of the above are true.

5. Sales prices of baseball cards from the 1960s are known to possess a skewed-right distribution with a mean sale price of $5.25 and a standard deviation of $2.80. Suppose a random sample of 100 cards from the 1960s is selected. Describe the sampling distribution for the sample mean sale price of the selected cards.
   a) Skewed-right with a mean of $5.25 and a standard error of $2.80
   b) Normal with a mean of $5.25 and a standard error of $0.28
   c) Skewed-right with a mean of $5.25 and a standard error of $0.28
   d) Normal with a mean of $5.25 and a standard error of $2.80
6. For sample size 16, the sampling distribution of the mean will be approximately normally distributed  
   a) regardless of the shape of the population.  
   b) **if the shape of the population is symmetrical.**  
   c) if the sample standard deviation is known.  
   d) if the sample is normally distributed.

7. For sample sizes greater than 30, the sampling distribution of the mean will be approximately normally distributed  
   a) **regardless of the shape of the population.**  
   b) only if the shape of the population is symmetrical.  
   c) only if the standard deviation of the samples are known.  
   d) only if the population is normally distributed.

8. Which of the following is NOT a reason for the need for sampling?  
   e) It is usually too costly to study the whole population.  
   f) It is usually too time consuming to look at the whole population.  
   g) It is sometimes destructive to observe the entire population.  
   h) **It is always more informative by investigating a sample than the entire population.**

9. A sample of 300 subscribers to a particular magazine is selected from a population frame of 9,000 subscribers. If, upon examining the data, it is determined that no subscriber had been selected in the sample more than once,  
   a) the sample could not have been random.  
   b) the sample may have been selected without replacement or with replacement.  
   c) the sample had to have been selected with replacement.  
   d) the sample had to have been selected without replacement.

10. A telemarketer set the company’s computerized dialing system to contact every 25th person listed in the local telephone directory. What sampling method was used?  
    a) Simple random sample  
    b) **Systematic sample**  
    c) Stratified sample  
    d) Cluster sample

11. Which of the following sampling methods will more likely be susceptible to ethical violation when used to form conclusions about the entire population?  
    a) Simple random sample  
    b) Cluster sample  
    c) Systematic sample  
    d) **Convenient sample**

12. At a computer manufacturing company, the actual size of computer chips is normally distributed with a mean of 1 centimeter and a standard deviation of 0.1 centimeter. A random sample of 12 computer chips is taken. What is the standard error for the sample mean?  
    a) 0.029  
    b) 0.050  
    c) 0.091  
    d) 0.120
SECTION II: TRUE OR FALSE

1. The amount of time it takes to complete an examination has a skewed-left distribution with a mean of 65 minutes and a standard deviation of 8 minutes. If 64 students were randomly sampled, the probability that the sample mean of the sampled students exceeds 71 minutes is approximately 0.
   True

2. The Central Limit Theorem is considered powerful in statistics because it works for any population distribution provided the sample size is sufficiently large and the population mean and standard deviation are known.
   True

3. As the sample size increases, the standard error of the mean increases.
   False

4. If the population distribution is unknown, in most cases the sampling distribution of the mean can be approximated by the normal distribution if the samples contain at least 30 observations.
   True

5. If the amount of gasoline purchased per car at a large service station has a population mean of $15 and a population standard deviation of $4 and it is assumed that the amount of gasoline purchased per car is symmetric, there is approximately a 68.26% chance that a random sample of 16 cars will have a sample mean between $14 and $16.
   True

6. If the amount of gasoline purchased per car at a large service station has a population mean of $15 and a population standard deviation of $4 and a random sample of 64 cars is selected, there is approximately a 95.44% chance that the sample mean will be between $14 and $16.
   True

7. Suppose \( \mu = 50 \) and \( \sigma^2 = 100 \) for a population. In a sample where \( n = 100 \) is randomly taken, 95% of all possible sample means will fall between 48.04 and 51.96.
   True

8. The Central Limit Theorem ensures that the sampling distribution of the sample mean approaches normal as the sample size increases.
   True

9. For distributions such as the normal distribution, the arithmetic mean is considered more stable from sample to sample than other measures of central tendency.
   True

10. The amount of bleach a machine pours into bottles has a mean of 36 oz. with a standard deviation of 0.15 oz. Suppose we take a random sample of 36 bottles filled by this machine. The sampling distribution of the sample mean has a standard error of 0.15.
    False
11. The mean of the sampling distribution of a sample proportion is the population proportion, $\pi$.
   True

12. The standard error of the sampling distribution of a sample proportion is $\sqrt{\frac{p(1-p)}{n}}$
   where $p$ is the sample proportion.
   False

13. A sample of size 25 provides a sample variance of 400. The standard error, in this case equal to 4, is best described as the estimate of the standard deviation of means calculated from samples of size 25.
   True

**SECTION III: FREE RESPONSE QUESTIONS**

1. The average score of all pro golfers for a particular course has a mean of 70 and a standard deviation of 3.0. Suppose 36 golfers played the course today. Find the probability that the average score of the 36 golfers exceeded 71. 0.0228

2. At a computer manufacturing company, the actual size of computer chips is normally distributed with a mean of 1 centimeter and a standard deviation of 0.1 centimeter. A random sample of 12 computer chips is taken.
   a. What is the probability that the sample mean will be between 0.99 and 1.01 centimeters? 0.2736 using Table E.2
   b. What is the probability that the sample mean will be below 0.95 centimeters? 0.0418 using Table E.2
   c. Above what value do 2.5% of the sample means fall? 1.057

3. The amount of pyridoxine (in grams) per multiple vitamin is normally distributed with $\mu = 110$ grams and $\sigma = 25$ grams. A sample of 25 vitamins is to be selected.
   a. What is the probability that the sample mean will be between 100 and 120 grams? 0.9544 using Table E.2
   b. What is the probability that the sample mean will be less than 100 grams? 0.0228
   c. What is the probability that the sample mean will be greater than 100 grams? 0.9772
   d. So, 95% of all sample means will be greater than how many grams? 101.7757
   e. So, the middle 70% of all sample means will fall between what two values? 104.8 and 115.2

4. The amount of time required for an oil and filter change on an automobile is normally distributed with a mean of 45 minutes and a standard deviation of 10 minutes. A random sample of 16 cars is selected.
   a. What would you expect the standard error of the mean to be? 2.5 minutes
   b. What is the probability that the sample mean is between 45 and 52 minutes? 0.4974
   c. What is the probability that the sample mean will be between 39 and 48 minutes? 0.8767
5. The amount of bleach a machine pours into bottles has a mean of 36 oz. with a standard deviation of 0.15 oz. Suppose we take a random sample of 36 bottles filled by this machine.
   a. The probability that the mean of the sample exceeds 36.01 oz. is __________. 0.3446
   b. The probability that the mean of the sample is less than 36.03 is __________. 0.8849
   c. The probability that the mean of the sample is between 35.94 and 36.06 oz. is __________. 0.9836
   d. The probability that the mean of the sample is between 35.95 and 35.98 oz. is __________. 0.1891
   e. So, 95% of the sample means based on samples of size 36 will be between __________ and __________. 35.951 and 36.049 ounces

6. To use the normal distribution to approximate the binomial distribution, we need ______ and ______ to be at least 5. $n\pi$ and $n(\pi - 1)$

7. Assume that house prices in a neighborhood are normally distributed with standard deviation $20,000. A random sample of 16 observations is taken. What is the probability that the sample mean differs from the population mean by more than $5,000? 0.3174 using Table E.2

**TABLE (A)**

Times spent studying by students in the week before final exams follow a normal distribution with standard deviation 8 hours. A random sample of 4 students was taken in order to estimate the mean study time for the population of all students.

a. Referring to Table (A), what is the probability that the sample mean exceeds the population mean by more than 2 hours? 0.3085
b. Referring to Table (A), what is the probability that the sample mean is more than 3 hours below the population mean? 0.2266
c. Referring to Table (A), what is the probability that the sample mean differs from the population mean by less than 2 hours? 0.3830 using Table E.2
d. Referring to Table (A), what is the probability that the sample mean differs from the population mean by more than 3 hours? 0.4532 using Table E.2

**TABLE (B)**

According to a survey, only 15% of customers who visited the web site of a major retail store made a purchase. Random samples of size 50 are selected.

a. Referring to Table (B), the average of all the sample proportions of customers who will make a purchase after visiting the web site is ________. 0.15 or 15%
b. Referring to Table (B), the standard deviation of all the sample proportions of customers who will make a purchase after visiting the web site is ________. 0.0505

c. Referring to Table (B), what proportion of the samples will have between 20% and 30% of customers who will make a purchase after visiting the web site? 0.1596
d. Referring to Table (B), what proportion of the samples will have less than 15% of customers who will make a purchase after visiting the web site? 0.5

e. Referring to Table (B), what is the probability that a random sample of 50 will have at least 30% of customers who will make a purchase after visiting the web site? 0.0015

f. Referring to Table (B), 90% of the samples will have less than what percentage of customers who will make a purchase after visiting the web site? 21.47%

g. Referring to Table (B), 90% of the samples will have more than what percentage of customers who will make a purchase after visiting the web site? 8.536 using Table E.2

TABLE (C)

According to an article, 19% of the entire U.S. population have high-speed access to the Internet. Random samples of size 200 are selected from the U.S. population.

a. Referring to Table (C), the population mean of all the sample proportions is 19% or 0.19

b. Referring to Table (C), the standard error of all the sample proportions is 0.0277

c. Referring to Table (C), among all the random samples of size 200, _____ % will have between 14% and 24% who have high-speed access to the Internet. 92.82 using Table E.2

d. Referring to Table (C), among all the random samples of size 200, _____ % will have between 9% and 29% who have high-speed access to the Internet. 99.97

e. Referring to Table (C), among all the random samples of size 200, _____ % will have more than 30% who have high-speed access to the Internet. 0.0000 or virtually zero

f. Referring to Table (C), among all the random samples of size 200, _____ % will have less than 20% who have high-speed access to the Internet. 64.06 using Table E.2

g. Referring to Table (C), among all the random samples of size 200, 90 % will have less than _____% who have high-speed access to the Internet. 22.55 using Table E.2

h. Referring to Table (C), among all the random samples of size 200, 90 % will have more than _____% who have high-speed access to the Internet. 15.45
**An Introduction to Statistics Course (ECOE 1302)**  
Spring Semester 2012-2013  
**Chapter 8 - CONFIDENCE INTERVAL ESTIMATION**  
**Practice Exam - Solution**  
Instructors: Dr. Samir Safi     Mr. Ibrahim Abed

**SECTION I: MULTIPLE-CHOICE**

1. The width of a confidence interval estimate for a proportion will be  
   a) narrower for 99% confidence than for 95% confidence.  
   b) wider for a sample size of 100 than for a sample size of 50.  
   c) **narrower for 90% confidence than for 95% confidence.**  
   d) narrower when the sample proportion is 0.50 than when the sample  
      proportion is 0.20.

2. A 99% confidence interval estimate can be interpreted to mean that  
   a) if all possible samples are taken and confidence interval estimates are  
      developed, 99% of them would include the true population mean somewhere  
      within their interval.  
   b) we have 99% confidence that we have selected a sample whose interval does  
      include the population mean.  
   c) **Both of the above.**  
   d) None of the above.

3. Which of the following is not true about the Student’s \( t \) distribution?  
   a) It has more area in the tails and less in the center than does the normal  
      distribution.  
   b) **It is used to construct confidence intervals for the population mean when  
      the population standard deviation is known.**  
   c) It is bell shaped and symmetrical.  
   d) As the number of degrees of freedom increases, the \( t \) distribution approaches  
      the normal distribution.

4. It is desired to estimate the average total compensation of CEOs in the Service  
   industry. Data were randomly collected from 18 CEOs and the 97% confidence  
   interval was calculated to be ($2,181,260, $5,836,180). Which of the following  
   interpretations is correct?  
   a) **97% of the sampled total compensation values fell between $2,181,260 and  
      $5,836,180.**  
   b) We are 97% confident that the mean of the sampled CEOs falls in the interval  
      $2,181,260 to $5,836,180.  
   c) In the population of Service industry CEOs, 97% of them will have total  
      compensations that fall in the interval $2,181,260 to $5,836,180.  
   d) **We are 97% confident that the average total compensation of all CEOs  
      in the Service industry falls in the interval $2,181,260 to $5,836,180.**
5. A confidence interval was used to estimate the proportion of statistics students that are females. A random sample of 72 statistics students generated the following 90% confidence interval: (0.438, 0.642). Based on the interval above, is the population proportion of females equal to 0.60?
   a) No, and we are 90% sure of it.
   b) No. The proportion is 54.17%.
   c) Maybe. 0.60 is a believable value of the population proportion based on the information above.
   d) Yes, and we are 90% sure of it.

6. A confidence interval was used to estimate the proportion of statistics students that are female. A random sample of 72 statistics students generated the following 90% confidence interval: (0.438, 0.642). Using the information above, what total size sample would be necessary if we wanted to estimate the true proportion to within ±0.08 using 95% confidence?
   a) 105
   b) 150
   c) 420
   d) 597

7. When determining the sample size necessary for estimating the true population mean, which factor is not considered when sampling with replacement?
   a) The population size.
   b) The population standard deviation.
   c) The level of confidence desired in the estimate.
   d) The allowable or tolerable sampling error.

8. Suppose a 95% confidence interval for $\mu$ turns out to be (1,000, 2,100). To make more useful inferences from the data, it is desired to reduce the width of the confidence interval. Which of the following will result in a reduced interval width?
   a) Increase the sample size.
   b) Increase the confidence level.
   c) Increase the population mean.
   d) Increase the sample mean.

9. A major department store chain is interested in estimating the average amount its credit card customers spent on their first visit to the chain’s new store in the mall. Fifteen credit card accounts were randomly sampled and analyzed with the following results: $\bar{X} = \$50.50$ and $s^2 = 400$. Construct a 95% confidence interval for the average amount its credit card customers spent on their first visit to the chain’s new store in the mall assuming that the amount spent follows a normal distribution.
   a) $\$50.50 \pm \$9.09$
   b) $\$50.50 \pm \$10.12$
   c) $\$50.50 \pm \$11.00$
   d) $\$50.50 \pm \$11.08$
10. Private colleges and universities rely on money contributed by individuals and corporations for their operating expenses. Much of this money is put into a fund called an endowment, and the college spends only the interest earned by the fund. A recent survey of 8 private colleges in the United States revealed the following endowments (in millions of dollars): 60.2, 47.0, 235.1, 490.0, 122.6, 177.5, 95.4, and 220.0. Summary statistics yield $X = 180.975$ and $s = 143.042$. Calculate a 95% confidence interval for the mean endowment of all the private colleges in the United States assuming a normal distribution for the endowments.

\[ a) \ 180.975 \pm 94.066 \]
\[ b) \ 180.975 \pm 99.123 \]
\[ c) \ 180.975 \pm 116.621 \]
\[ d) \ 180.975 \pm 119.586 \]

11. A university dean is interested in determining the proportion of students who receive some sort of financial aid. Rather than examine the records for all students, the dean randomly selects 200 students and finds that 118 of them are receiving financial aid. If the dean wanted to estimate the proportion of all students receiving financial aid to within 3% with 99% reliability, how many students would need to be sampled?

\[ a) \ n = 1,844 \]
\[ b) \ n = 1,784 \]
\[ c) \ n = 1,503 \]
\[ d) \ n = 1,435 \]

12. An economist is interested in studying the incomes of consumers in a particular region. The population standard deviation is known to be $1,000. A random sample of 50 individuals resulted in an average income of $15,000. What is the width of the 90% confidence interval?

\[ a) \ 232.60 \]
\[ b) \ 364.30 \]
\[ c) \ 465.23 \]
\[ d) \ 728.60 \]

13. The head librarian at the Library of Congress has asked her assistant for an interval estimate of the mean number of books checked out each day. The assistant provides the following interval estimate: from 740 to 920 books per day. If the head librarian knows that the population standard deviation is 150 books checked out per day, approximately how large a sample did her assistant use to determine the interval estimate?

\[ a) \ 2 \]
\[ b) \ 3 \]
\[ c) \ 12 \]
\[ d) \ It \ cannot \ be \ determined \ from \ the \ information \ given. \]

14. The head librarian at the Library of Congress has asked her assistant for an interval estimate of the mean number of books checked out each day. The assistant provides the following interval estimate: from 740 to 920 books per day. If the head librarian knows that the population standard deviation is 150 books checked out per day, and she asked her assistant to use 25 days of data to construct the interval estimate, what confidence level can she attach to the interval estimate?

\[ a) \ 99.7\% \]
\[ b) \ 99.0\% \]
\[ c) \ 98.0\% \]
\[ d) \ 95.4\% \]
SECTION II: TRUE OR FALSE

1. A race car driver tested his car for time from 0 to 60 mph, and in 20 tests obtained an average of 4.85 seconds with a standard deviation of 1.47 seconds. A 95% confidence interval for the 0 to 60 time is 4.52 seconds to 5.18 seconds.
   False

2. Given a sample mean of 2.1 and a population standard deviation of 0.7 from a sample of 10 data points, a 90% confidence interval will have a width of 2.36.
   False

3. A random sample of 50 provides a sample mean of 31 with a standard deviation of \(s=14\). The upper bound of a 90% confidence interval estimate of the population mean is 34.32.
   True

4. Other things being equal, as the confidence level for a confidence interval increases, the width of the interval increases.
   True

5. For a \(t\) distribution with 12 degrees of freedom, the area between –2.6810 and 2.1788 is 0.980.
   False

6. A sample of 100 fuses from a very large shipment is found to have 10 that are defective. The 95% confidence interval would indicate that, for this shipment, the proportion of defective fuses is between 0 and 0.28.
   False

7. A point estimate consists of a single sample statistic that is used to estimate the true population parameter.
   True

8. The \(t\) distribution is used to develop a confidence interval estimate of the population mean when the population standard deviation is unknown.
   True

9. The standardized normal distribution is used to develop a confidence interval estimate of the population proportion when the sample size is sufficiently large.
   True

10. The width of a confidence interval equals twice the sampling error.
    True

11. A population parameter is used to estimate a confidence interval.
    False

12. Holding the sample size fixed, increasing the level of confidence in a confidence interval will necessarily lead to a wider confidence interval.
    True

13. Holding the width of a confidence interval fixed, increasing the level of confidence can be achieved with a lower sample size.
    False
SECTION III: FREE RESPONSE QUESTIONS

Table (A)

A quality control engineer is interested in the mean length of sheet insulation being cut automatically by machine. The desired length of the insulation is 12 feet. It is known that the standard deviation in the cutting length is 0.15 feet. A sample of 70 cut sheets yields a mean length of 12.14 feet. This sample will be used to obtain a 99% confidence interval for the mean length cut by machine.

1. the critical value to use in obtaining the confidence interval is ________.  
2.58

2. the confidence interval goes from ________ to ________.  
12.09 to 12.19

3. the confidence interval indicates that the machine is not working properly.  
True

4. the confidence interval is valid only if the lengths cut are normally distributed.  
False

5. suppose the engineer had decided to estimate the mean length to within 0.03 with 99% confidence. Then the sample size would be ________.  
165.8724 rounds up to 166

Table (B)

The actual voltages of power packs labeled as 12 volts are as follows: 11.77, 11.90, 11.64, 11.84, 12.13, 11.99, and 11.77.

1. a confidence interval for this sample would be based on the t distribution with ________ degrees of freedom.  
6

2. the critical value for a 99% confidence interval for this sample is __________.  
3.7074

3. a 99% confidence interval for the mean voltage of the power packs is from ________ to _________.  
11.6367 to 12.0891

4. a 95% confidence interval for the mean voltage of the power pack is wider than a 99% confidence interval.  
False

5. a 99% confidence interval will contain 99% of the voltages for all such power packs.  
False

6. a confidence interval estimate of the population mean would only be valid if the distribution of voltages is normal.  
True
7. a 90% confidence interval calculated from the same data would be narrower than a
99% confidence interval.
True

8. it is possible that the 99% confidence interval calculated from the data will not
contain the mean voltage for the sample.
False

9. it is possible that the 99% confidence interval calculated from the data will not
contain the mean voltage for the entire population.
True

TABLE (C)

A hotel chain wants to estimate the average number of rooms rented daily in each month.
The population of rooms rented daily is assumed to be normally distributed for each
month with a standard deviation of 24 rooms.

1. during January, a sample of 16 days has a sample mean of 48 rooms. This
information is used to calculate an interval estimate for the population mean to be
from 40 to 56 rooms. What is the level of confidence of this interval?
81.76%

2. during February, a sample of 25 days has a sample mean of 37 rooms. Use this
information to calculate a 92% confidence interval for the population mean.
28.60 to 45.40

3. The county clerk wants to estimate the proportion of retired voters who will need
special election facilities. The clerk wants to find a 95% confidence interval for the
population proportion which extends at most 0.07 to either side of the sample
proportion. How large a sample must be taken to assure these conditions are met?
196

4. The county clerk wants to estimate the proportion of retired voters who will need
special election facilities. Suppose a sample of 400 retired voters was taken. If 150
need special election facilities, calculate an 80% confidence interval for the
population proportion.
0.344 to 0.406

TABLE (D)

The president of a university is concerned that illicit drug use on campus is higher than
the 5% acceptable level. A random sample of 250 students from a population of 2000
revealed that 7 of them had used illicit drug during the last 12 months.

1. what is the critical value for the 90% one-sided confidence interval for the proportion
of students who had used illicit drug during the last 12 months?
1.28 (using Table E.2)

2. what is the upper bound of the 90% one-sided confidence interval for the proportion
of students who had used illicit drug during the last 12 months?
0.0405
3. the president can be 90% confident that no more than 5% of the students at the university had used illicit drug during the last 12 months.
   **True**

4. using the 90% one-sided confidence interval, the president can be 95% confident that no more than 5% of the students at the university had used illicit drug during the last 12 months.
   **False**

5. using the 90% one-sided confidence interval, the president can be 85% confident that no more than 5% of the students at the university had used illicit drug during the last 12 months.
   **True**
SECTION I: MULTIPLE-CHOICE

1. Which of the following would be an appropriate null hypothesis?
   a. The mean of a population is equal to 55.
   b. The mean of a sample is equal to 55.
   c. The mean of a population is greater than 55.
   d. Only (a) and (c) are true.

2. Which of the following would be an appropriate alternative hypothesis?
   a. The mean of a population is equal to 55.
   b. The mean of a sample is equal to 55.
   c. The mean of a population is greater than 55.
   d. The mean of a sample is greater than 55.

3. A Type II error is committed when
   a. we reject a null hypothesis that is true.
   b. we don't reject a null hypothesis that is true.
   c. we reject a null hypothesis that is false.
   d. we don't reject a null hypothesis that is false.

4. The power of a test is measured by its capability of
   a. rejecting a null hypothesis that is true.
   b. not rejecting a null hypothesis that is true.
   c. rejecting a null hypothesis that is false.
   d. not rejecting a null hypothesis that is false.

5. If an economist wishes to determine whether there is evidence that average family
   income in a community exceeds $25,000
   a. either a one-tailed or two-tailed test could be used with equivalent results.
   b. a one-tailed test should be utilized.
   c. a two-tailed test should be utilized.
   d. None of the above.

6. If the p-value is less than $\alpha$ in a two-tailed test,
   a. the null hypothesis should not be rejected.
   b. the null hypothesis should be rejected.
   c. a one-tailed test should be used.
   d. no conclusion should be reached.
7. It is possible to directly compare the results of a confidence interval estimate to the results obtained by testing a null hypothesis if
   a. a two-tailed test for $\mu$ is used.
   b. a one-tailed test for $\mu$ is used.
   c. Both of the previous statements are true.
   d. None of the previous statements is true.

8. The symbol for the power of a statistical test is
   a. $\alpha$.
   b. $1 - \alpha$.
   c. $\beta$.
   d. $1 - \beta$.

9. How many Kleenex should the Kimberly Clark Corporation package of tissues contain? Researchers determined that 60 tissues is the average number of tissues used during a cold. Suppose a random sample of 100 Kleenex users yielded the following data on the number of tissues used during a cold: $\overline{X} = 52, s = 22$. Give the null and alternative hypotheses to determine if the number of tissues used during a cold is less than 60.
   a. $H_0: \mu \leq 60$ and $H_1: \mu > 60$.
   b. $H_0: \mu \geq 60$ and $H_1: \mu < 60$.
   c. $H_0: \overline{X} \geq 60$ and $H_1: \overline{X} < 60$.
   d. $H_0: \overline{X} = 52$ and $H_1: \overline{X} \neq 52$.

10. How many Kleenex should the Kimberly Clark Corporation package of tissues contain? Researchers determined that 60 tissues is the average number of tissues used during a cold. Suppose a random sample of 100 Kleenex users yielded the following data on the number of tissues used during a cold: $\overline{X} = 52, s = 22$. Suppose the alternative we wanted to test was $H_1: \mu < 60$. State the correct rejection region for $\alpha = 0.05$.
    a. Reject $H_0$ if $t > 1.6604$.
    b. Reject $H_0$ if $t < -1.6604$.
    c. Reject $H_0$ if $t > 1.9842$ or $Z < -1.9842$.
    d. Reject $H_0$ if $t < -1.9842$.

11. How many Kleenex should the Kimberly Clark Corporation package of tissues contain? Researchers determined that 60 tissues is the average number of tissues used during a cold. Suppose a random sample of 100 Kleenex users yielded the following data on the number of tissues used during a cold: $\overline{X} = 52, s = 22$. Suppose the test statistic does fall in the rejection region at $\alpha = 0.05$. Which of the following conclusion is correct?
    a. At $\alpha = 0.05$, there is not sufficient evidence to conclude that the average number of tissues used during a cold is 60 tissues.
    b. At $\alpha = 0.05$, there is sufficient evidence to conclude that the average number of tissues used during a cold is 60 tissues.
    c. At $\alpha = 0.05$, there is not sufficient evidence to conclude that the average number of tissues used during a cold is not 60 tissues.
    d. At $\alpha = 0.10$, there is sufficient evidence to conclude that the average number of tissues used during a cold is not 60 tissues.
12. We have created a 95% confidence interval for $\mu$ with the result (10, 15). What decision will we make if we test $H_0 : \mu = 16$ versus $H_1 : \mu \neq 16$ at $\alpha = 0.10$?
   a. Reject $H_0$ in favor of $H_1$.
   b. Accept $H_0$ in favor of $H_1$.
   c. Fail to reject $H_0$ in favor of $H_1$.
   d. We cannot tell what our decision will be from the information given.

13. A ________________ is a numerical quantity computed from the data of a sample and is used in reaching a decision on whether or not to reject the null hypothesis.
   a. significance level
   b. critical value
   c. test statistic
   d. parameter

14. The owner of a local nightclub has recently surveyed a random sample of $n = 250$ customers of the club. She would now like to determine whether or not the mean age of her customers is over 30. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. The appropriate hypotheses to test are:
   a. $H_0 : \mu \geq 30$ versus $H_1 : \mu < 30$.
   b. $H_0 : \mu \leq 30$ versus $H_1 : \mu > 30$.
   c. $H_0 : \bar{X} \geq 30$ versus $H_1 : \bar{X} < 30$.
   d. $H_0 : \bar{X} \leq 30$ versus $H_1 : \bar{X} > 30$.

15. The owner of a local nightclub has recently surveyed a random sample of $n = 250$ customers of the club. She would now like to determine whether or not the mean age of her customers is over 30. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. If she wants to be 99% confident in her decision, what rejection region should she use?
   a. Reject $H_0$ if $t < -2.34$.
   b. Reject $H_0$ if $t < -2.55$.
   c. Reject $H_0$ if $t > 2.34$.
   d. Reject $H_0$ if $t > 2.58$.

16. The owner of a local nightclub has recently surveyed a random sample of $n = 250$ customers of the club. She would now like to determine whether or not the mean age of her customers is over 30. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. Suppose she found that the sample mean was 30.45 years and the sample standard deviation was 5 years. If she wants to be 99% confident in her decision, what decision should she make?
   a. Reject $H_0$.
   b. Accept $H_0$.
   c. Fail to reject $H_0$.
   d. We cannot tell what her decision should be from the information given.
17. The owner of a local nightclub has recently surveyed a random sample of $n = 250$ customers of the club. She would now like to determine whether or not the mean age of her customers is over 30. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. Suppose she found that the sample mean was 30.45 years and the sample standard deviation was 5 years. If she wants to be 99% confident in her decision, what conclusion can she make?
   a. There is not sufficient evidence that the mean age of her customers is over 30.
   b. There is sufficient evidence that the mean age of her customers is over 30.
   c. There is not sufficient evidence that the mean age of her customers is not over 30.
   d. There is sufficient evidence that the mean age of her customers is not over 30.

18. The owner of a local nightclub has recently surveyed a random sample of $n = 250$ customers of the club. She would now like to determine whether or not the mean age of her customers is over 30. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. Suppose she found that the sample mean was 30.45 years and the sample standard deviation was 5 years. What is the $p$-value associated with the test statistic?
   a. 0.3577
   b. 0.1423
   c. 0.0780
   d. 0.02

19. The marketing manager for an automobile manufacturer is interested in determining the proportion of new compact-car owners who would have purchased a passenger-side inflatable air bag if it had been available for an additional cost of $300. The manager believes from previous information that the proportion is 0.30. Suppose that a survey of 200 new compact-car owners is selected and 79 indicate that they would have purchased the inflatable air bags. If you were to conduct a test to determine whether there is evidence that the proportion is different from 0.30, which test would you use?
   a. $Z$-test of a population mean
   b. $Z$-test of a population proportion
   c. $t$-test of population mean
   d. $t$-test of a population proportion

20. The marketing manager for an automobile manufacturer is interested in determining the proportion of new compact-car owners who would have purchased a passenger-side inflatable air bag if it had been available for an additional cost of $300. The manager believes from previous information that the proportion is 0.30. Suppose that a survey of 200 new compact-car owners is selected and 79 indicate that they would have purchased the inflatable air bags. If you were to conduct a test to determine whether there is evidence that the proportion is different from 0.30 and decided not to reject the null hypothesis, what conclusion could you draw?
   a. There is sufficient evidence that the proportion is 0.30.
   b. There is not sufficient evidence that the proportion is 0.30.
   c. There is sufficient evidence that the proportion is 0.30.
   d. There is not sufficient evidence that the proportion is not 0.30.
SECTION II: TRUE OR FALSE

1. For a given level of significance, if the sample size is increased, the power of the test will increase.
   True

2. Suppose, in testing a hypothesis about a proportion, the p-value is computed to be 0.043. The null hypothesis should be rejected if the chosen level of significance is 0.05.
   True

3. Suppose, in testing a hypothesis about a proportion, the Z test statistic is computed to be 2.04. The null hypothesis should be rejected if the chosen level of significance is 0.01 and a two-tailed test is used.
   False

4. The smaller is the p-value, the stronger is the evidence against the null hypothesis.
   True

5. A sample is used to obtain a 95% confidence interval for the mean of a population. The confidence interval goes from 15 to 19. If the same sample had been used to test the null hypothesis that the mean of the population is equal to 20 versus the alternative hypothesis that the mean of the population differs from 20, the null hypothesis could be rejected at a level of significance of 0.05.
   True

6. A sample is used to obtain a 95% confidence interval for the mean of a population. The confidence interval goes from 15 to 19. If the same sample had been used to test the null hypothesis that the mean of the population is equal to 18 versus the alternative hypothesis that the mean of the population differs from 18, the null hypothesis could be rejected at a level of significance of 0.05.
   False

7. A sample is used to obtain a 95% confidence interval for the mean of a population. The confidence interval goes from 15 to 19. If the same sample had been used to test the null hypothesis that the mean of the population is equal to 20 versus the alternative hypothesis that the mean of the population differs from 20, the null hypothesis could be rejected at a level of significance of 0.10.
   True

8. A sample is used to obtain a 95% confidence interval for the mean of a population. The confidence interval goes from 15 to 19. If the same sample had been used to test the null hypothesis that the mean of the population is equal to 20 versus the alternative hypothesis that the mean of the population differs from 20, the null hypothesis could be rejected at a level of significance of 0.02.
   False

9. A sample is used to obtain a 95% confidence interval for the mean of a population. The confidence interval goes from 15 to 19. If the same sample had been used to test the null hypothesis that the mean of the population is equal to 20 versus the alternative hypothesis that the mean of the population differs from 20, the null hypothesis could be accepted at a level of significance of 0.02.
   False
Microsoft Excel was used on a set of data involving the number of parasites found on 46 Monarch butterflies captured in Pismo Beach State Park. A biologist wants to know if the mean number of parasites per butterfly is over 20. She will make her decision using a test with a level of significance of 0.10. The following information was extracted from the Microsoft Excel output for the sample of 46 Monarch butterflies:

<table>
<thead>
<tr>
<th>$n$</th>
<th>46</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic Mean</td>
<td>28.00</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>25.92</td>
</tr>
<tr>
<td>Standard Error</td>
<td>3.82</td>
</tr>
</tbody>
</table>

Null Hypothesis: $H_0: \mu \leq 20.000$; $\alpha = 0.10$; $df = 45$; $T$ Test Statistic = 2.09; One-Tailed Test Upper Critical Value = 1.3006; $p$-value = 0.021; Decision = Reject.

1. the parameter the biologist is interested in is:
   a. the mean number of butterflies in Pismo Beach State Park.
   b. the mean number of parasites on these 46 butterflies.
   c. the mean number of parasites on Monarch butterflies in Pismo Beach State Park.
   d. the proportion of butterflies with parasites.

2. state the alternative hypothesis for this study.
   $H_1: \mu > 20.000$

3. what critical value should the biologist use to determine the rejection region?
   a. 1.6794
   b. 1.3011
   c. 1.3006
   d. 0.6800

4. True or False: the null hypothesis would be rejected.
   True

5. True or False: the null hypothesis would be rejected if a 4% probability of committing a Type I error is allowed.
   True

6. True or False: the null hypothesis would be rejected if a 1% probability of committing a Type I error is allowed.
   False

7. the lowest level of significance at which the null hypothesis can be rejected is ______.
   0.021

8. True or False: the evidence proves beyond a doubt that the mean number of parasites on butterflies in Pismo Beach State Park is over 20.
   False

9. True of False: the biologist can conclude that there is sufficient evidence to show that the average number of parasites per Monarch butterfly in Pismo Beach State Park is over 20 using a level of significance of 0.10.
   True
10. True or False: the biologist can conclude that there is sufficient evidence to show that the average number of parasites per Monarch butterfly in Pismo Beach State Park is over 20 with no more than a 5% probability of incorrectly rejecting the true null hypothesis.

True

11. True or False: the biologist can conclude that there is sufficient evidence to show that the average number of parasites per Monarch butterfly in Pismo Beach State Park is over 20 with no more than a 1% probability of incorrectly rejecting the true null hypothesis.

False

12. True or False: the value of $\beta$ is 0.90.

False

13. True or False: if these data were used to perform a two-tailed test, the $p$-value would be 0.042.

True

### TABLE (B)

A bank tests the null hypothesis that the mean age of the bank's mortgage holders is less than or equal to 45, versus an alternative that the mean age is greater than 45. They take a sample and calculate a $p$-value of 0.0202.

1. True or False: the null hypothesis would be rejected at a significance level of $\alpha = 0.05$.

True

2. True or False: the null hypothesis would be rejected at a significance level of $\alpha = 0.01$.

False

3. True or False: the bank can conclude that the average age is greater than 45 at a significance level of $\alpha = 0.01$.

False

4. if the same sample was used to test the opposite one-tailed test, what would be that test's $p$-value?
   a. 0.0202
   b. 0.0404
   c. 0.9596
   d. **0.9798**
A major home improvement store conducted its biggest brand recognition campaign in the company's history. A series of new television advertisements featuring well-known entertainers and sports figures were launched. A key metric for the success of television advertisements is the proportion of viewers who “like the ads a lot”. A study of 1,189 adults who viewed the ads reported that 230 indicated that they “like the ads a lot.” The percentage of a typical television advertisement receiving the “like the ads a lot” score is believed to be 22%. Company officials wanted to know if there is evidence that the series of television advertisements are less successful than the typical ad (i.e. if there is evidence that the population proportion of “like the ads a lot” for the company’s ads is less than 0.22) at a 0.01 level of significance.

1. The parameter the company officials is interested in is:
   a. the mean number of viewers who “like the ads a lot”.
   b. the total number of viewers who “like the ads a lot”.
   c. the mean number of company officials who “like the ads a lot”.
   d. the proportion of viewers who “like the ads a lot”.

2. State the null hypothesis for this study.
   \[ H_0 : \pi \geq 0.22 \]

3. State the alternative hypothesis for this study.
   \[ H_1 : \pi < 0.22 \]

4. What critical value should the company officials use to determine the rejection region?
   \[ -2.3263 \]

5. The null hypothesis will be rejected if the test statistics is
   a. greater than 2.3263
   b. less than 2.3263
   c. greater than \(-2.3263\)
   d. less than \(-2.3263\)

6. True or False: the null hypothesis would be rejected. False

7. The lowest level of significance at which the null hypothesis can be rejected is ______\(0.0135\)

8. The largest level of significance at which the null hypothesis will not be rejected is ______\(0.0135\)

9. True or False: the company officials can conclude that there is sufficient evidence to show that the series of television advertisements are less successful than the typical ad using a level of significance of 0.01. False

10. True or False: the company officials can conclude that there is sufficient evidence to show that the series of television advertisements are less successful than the typical ad using a level of significance of 0.05. True

11. True or False: the value of \(\beta\) is 0.90. False

12. What will be the \(p\)-value if these data were used to perform a two-tailed test? \(0.027\)
SECTION I: MULTIPLE-CHOICE

1. The $t$ test for the difference between the means of 2 independent populations assumes that the respective
   a. sample sizes are equal.
   b. sample variances are equal.
   c. populations are approximately normal.
   d. All of the above.

2. If we are testing for the difference between the means of 2 related populations with samples of $n_1 = 20$ and $n_2 = 20$, the number of degrees of freedom is equal to
   a. 39.
   b. 38.
   c. 19.
   d. 18.

3. In testing for differences between the means of two related populations, the null hypothesis is
   a. $H_0 : \mu_D = 2$.
   b. $H_0 : \mu_D = 0$.
   c. $H_0 : \mu_D < 0$.
   d. $H_0 : \mu_D > 0$.

4. In testing for differences between the means of two independent populations, the null hypothesis is:
   a. $H_0 : \mu_1 - \mu_2 = 2$.
   b. $H_0 : \mu_1 - \mu_2 = 0$.
   c. $H_0 : \mu_1 - \mu_2 > 0$.
   d. $H_0 : \mu_1 - \mu_2 < 2$.

5. When testing $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_1 : \mu_1 - \mu_2 > 0$, the observed value of the $Z$-score was found to be $-2.13$. The $p$-value for this test would be
   a. 0.0166.
   b. 0.0332.
   c. 0.9668.
   d. 0.9834.
6. If we wish to determine whether there is evidence that the proportion of items of interest is higher in group 1 than in group 2, the appropriate test to use is
   a) the \( Z \) test for the difference between two proportions.
   b) the \( F \) test for the difference between two variances.
   c) the pooled-variance \( t \) test for the difference between two proportions.
   d) the \( F \) test for the difference between two proportions.

**SECTION II: TRUE OR FALSE**

1. The sample size in each independent sample must be the same if we are to test for differences between the means of 2 independent populations.
   **False**

2. When testing for differences between the means of 2 related populations, we can use either a one-tailed or two-tailed test.
   **True**

3. The test for the equality of 2 population variances assumes that each of the 2 populations is normally distributed.
   **True**

4. When the sample sizes are equal, the pooled variance of the 2 groups is the average of the 2 sample variances.
   **True**

5. A researcher is curious about the effect of sleep on students’ test performances. He chooses 60 students and gives each 2 tests: one given after 2 hours’ sleep and one after 8 hours’ sleep. The test the researcher should use would be a related samples test.
   **True**

6. A statistics professor wanted to test whether the grades on a statistics test were the same for upper and lower classmen. The professor took a random sample of size 10 from each, conducted a test and found out that the variances were equal. For this situation, the professor should use a \( t \) test with related samples.
   **False**

7. In testing the difference between two proportions using the normal distribution, we may use a two-tailed \( Z \) test.
   **True**
**SECTION III: FREE RESPONSE QUESTIONS**

**TABLE (A)**

A researcher randomly sampled 30 graduates of an MBA program and recorded data concerning their starting salaries. Of primary interest to the researcher was the effect of gender on starting salaries. Analysis of the mean salaries of the females and males in the sample is given below.

<table>
<thead>
<tr>
<th></th>
<th>Population 1 Sample</th>
<th>Population 2 Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesized Difference</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>Level of Significance</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Sample Size</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Sample Mean</td>
<td>48266.7</td>
<td>55000</td>
</tr>
<tr>
<td>Sample Standard Deviation</td>
<td>13577.63</td>
<td>11741.29</td>
</tr>
<tr>
<td>Difference in Sample Means</td>
<td>-6733.3</td>
<td>-6733.3</td>
</tr>
<tr>
<td>t-Test Statistic</td>
<td>-1.40193</td>
<td>-1.40193</td>
</tr>
<tr>
<td>Lower-Critical Value</td>
<td>-1.70113</td>
<td>-1.70113</td>
</tr>
<tr>
<td>p-Value</td>
<td>0.085962</td>
<td>0.085962</td>
</tr>
</tbody>
</table>

1. The researcher was attempting to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. According to the test run, which of the following is an appropriate alternative hypothesis?
   a. $H_1: \mu_{\text{females}} > \mu_{\text{males}}$
   b. $H_1: \mu_{\text{females}} < \mu_{\text{males}}$
   c. $H_1: \mu_{\text{females}} \neq \mu_{\text{males}}$
   d. $H_1: \mu_{\text{females}} = \mu_{\text{males}}$

2. The researcher was attempting to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. From the analysis in Table 10-2, the correct test statistic is:
   a. 0.0860
   b. 1.4019
   c. 1.7011
   d. 6,733.33
3. the researcher was attempting to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. The proper conclusion for this test is:
   a. At the $\alpha = 0.10$ level, there is sufficient evidence to indicate a difference in the mean starting salaries of male and female MBA graduates.
   b. **At the $\alpha = 0.10$ level, there is sufficient evidence to indicate that females have a lower mean starting salary than male MBA graduates.**
   c. At the $\alpha = 0.10$ level, there is sufficient evidence to indicate that females have a higher mean starting salary than male MBA graduates.
   d. At the $\alpha = 0.10$ level, there is insufficient evidence to indicate any difference in the mean starting salaries of male and female MBA graduates.

4. the researcher was attempting to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. What assumptions were necessary to conduct this hypothesis test?
   a. Both populations of salaries (male and female) must have approximate normal distributions.
   b. The population variances are approximately equal.
   c. The samples were randomly and independently selected.
   d. **All of the above assumptions were necessary.**

5. what is the 99% confidence interval estimate for the difference between two means?
   $-20,004.90$ to $6,538.30$

6. what is the 95% confidence interval estimate for the difference between two means?
   $-16571.55$ to $3,104.95$

7. what is the 90% confidence interval estimate for the difference between two means?
   $-14,903.61$ to $1,437.01$

**TABLE (B)**

To test the effectiveness of a business school preparation course, 8 students took a general business test before and after the course. The results are given below.

<table>
<thead>
<tr>
<th>Student</th>
<th>Exam Score Before Course (1)</th>
<th>Exam Score After Course (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>530</td>
<td>670</td>
</tr>
<tr>
<td>2</td>
<td>690</td>
<td>770</td>
</tr>
<tr>
<td>3</td>
<td>910</td>
<td>1,000</td>
</tr>
<tr>
<td>4</td>
<td>700</td>
<td>710</td>
</tr>
<tr>
<td>5</td>
<td>450</td>
<td>550</td>
</tr>
<tr>
<td>6</td>
<td>820</td>
<td>870</td>
</tr>
<tr>
<td>7</td>
<td>820</td>
<td>770</td>
</tr>
<tr>
<td>8</td>
<td>630</td>
<td>610</td>
</tr>
</tbody>
</table>

8. the number of degrees of freedom is
   a. 14.
   b. 13.
   c. 8.
   d. **7.**
9. The value of the sample mean difference is _______ if the difference scores reflect the results of the exam after the course minus the results of the exam before the course.
   a. 0  
   b. 50  
   c. 68  
   d. 400

10. The value of the standard error of the difference scores is
   a. 65.027  
   b. 60.828  
   c. **22.991**  
   d. 14.696

11. What is the critical value for testing at the 5% level of significance whether the business school preparation course is effective in improving exam scores?
   a. 2.365  
   b. 2.145  
   c. 1.761  
   d. **1.895**

12. At the 0.05 level of significance, the decision for this hypothesis test would be:
   e. **reject the null hypothesis.**
   f. do not reject the null hypothesis.
   g. reject the alternative hypothesis.
   h. It cannot be determined from the information given.

13. At the 0.05 level of significance, the conclusion for this hypothesis test would be:
   i. **the business school preparation course does improve exam score.**
   j. the business school preparation course does not improve exam score.
   k. the business school preparation course has no impact on exam score.
   l. It cannot be drawn from the information given.

14. The calculated value of the test statistic is _______.
   **2.175**

15. The p-value of the test statistic is _______.
   **0.0331 (using Excel) or ‘between 0.025 and 0.05’ (using Table E.3 with 7 degrees of freedom)**

16. True or False: in examining the differences between related samples we are essentially sampling from an underlying population of difference "scores."
   **True**
TABLE (C)

To investigate the efficacy of a diet, a random sample of 16 male patients is drawn from a population of adult males using the diet. The weight of each individual in the sample is taken at the start of the diet and at a medical follow-up 4 weeks later. Assuming that the population of differences in weight before versus after the diet follow a normal distribution, the $t$-test for related samples can be used to determine if there was a significant decrease in the mean weight during this period. Suppose the mean decrease in weights over all 16 subjects in the study is 3.0 pounds with the standard deviation of differences computed as 6.0 pounds.

1. the $t$ test should be ______-tailed.
   one

2. the computed $t$ statistic is ______.
   2.00

3. there are ______ degrees of freedom for this test.
   15

4. the critical value for a one-tailed test of the null hypothesis of no difference at the $\alpha = 0.05$ level of significance is ______.
   1.7531

5. a one-tailed test of the null hypothesis of no difference would ______ (be rejected/not be rejected) at the $\alpha = 0.05$ level of significance.
   be rejected

6. the $p$-value for a one-tailed test whose computed $t$ statistic is 2.00 is between ______.
   0.025 and 0.05

7. if we were interested in testing against the two-tailed alternative that $\mu_d$ is not equal to zero at the $\alpha = 0.05$ level of significance, the null hypothesis would ______ (be rejected/not be rejected).
   not be rejected

8. the $p$-value for a two-tailed test whose computed statistic is 2.00 is between ______.
   0.05 and 0.10

9. what is the 95% confidence interval estimate for the mean difference in weight before and after the diet?
   −0.20 to 6.20

10. what is the 99% confidence interval estimate for the mean difference in weight before and after the diet?
    −1.42 to 7.42

11. what is the 90% confidence interval estimate for the mean difference in weight before and after the diet?
    0.37 to 5.63
TABLE (D)

A few years ago, Pepsi invited consumers to take the “Pepsi Challenge.” Consumers were asked to decide which of two sodas, Coke or Pepsi, they preferred in a blind taste test. Pepsi was interested in determining what factors played a role in people’s taste preferences. One of the factors studied was the gender of the consumer. Below are the results of analyses comparing the taste preferences of men and women with the proportions depicting preference for Pepsi.

Males: $n = 109, p_M = 0.422018$ 
Females: $n = 52, p_F = 0.25$
$p_M - p_F = 0.172018$  
$Z = 2.11825$

1. to determine if a difference exists in the taste preferences of men and women, give the correct alternative hypothesis that Pepsi would test.
   a) $H_1: \mu_M - \mu_F \neq 0$
   b) $H_1: \mu_M - \mu_F > 0$
   c) $H_1: \pi_M - \pi_F \neq 0$
   d) $H_1: \pi_M - \pi_F = 0$

2. suppose Pepsi wanted to test to determine if the males preferred Pepsi more than the females. Using the test statistic given, compute the appropriate $p$-value for the test.
   a) 0.0171
   b) 0.0340
   c) 0.2119
   d) 0.4681

3. suppose Pepsi wanted to test to determine if the males preferred Pepsi less than the females. Using the test statistic given, compute the appropriate $p$-value for the test.
   a) 0.0170
   b) 0.0340
   c) 0.9660
   d) 0.9830

4. suppose that the two-tailed $p$-value was really 0.0734. State the proper conclusion.
   a) At $\alpha = 0.05$, there is sufficient evidence to indicate the proportion of males preferring Pepsi differs from the proportion of females preferring Pepsi.
   b) At $\alpha = 0.10$, there is sufficient evidence to indicate the proportion of males preferring Pepsi differs from the proportion of females preferring Pepsi.
   c) At $\alpha = 0.05$, there is sufficient evidence to indicate the proportion of males preferring Pepsi equals the proportion of females preferring Pepsi.
   d) At $\alpha = 0.08$, there is insufficient evidence to indicate the proportion of males preferring Pepsi differs from the proportion of females preferring Pepsi.

5. construct a 90% confidence interval estimate of the difference between the proportion of males and females who prefer Pepsi.
   0.05 to 0.30

6. construct a 95% confidence interval estimate of the difference between the proportion of males and females who prefer Pepsi.
   0.02 to 0.32
7. construct a 99% confidence interval estimate of the difference between the proportion of males and females who prefer Pepsi.

\[-0.03 \text{ to } 0.37\]

**TABLE (E)**

A corporation randomly selects 150 salespeople and finds that 66% who have never taken a self-improvement course would like such a course. The firm did a similar study 10 years ago in which 60% of a random sample of 160 salespeople wanted a self-improvement course. The groups are assumed to be independent random samples. Let \( \pi_1 \) and \( \pi_2 \) represent the true proportion of workers who would like to attend a self-improvement course in the recent study and the past study, respectively.

1. if the firm wanted to test whether this proportion has changed from the previous study, which represents the relevant hypotheses?
   a) \( H_0: \pi_1 - \pi_2 = 0 \) versus \( H_1: \pi_1 - \pi_2 \neq 0 \)
   b) \( H_0: \pi_1 - \pi_2 \neq 0 \) versus \( H_1: \pi_1 - \pi_2 = 0 \)
   c) \( H_0: \pi_1 - \pi_2 \leq 0 \) versus \( H_1: \pi_1 - \pi_2 > 0 \)
   d) \( H_0: \pi_1 - \pi_2 \geq 0 \) versus \( H_1: \pi_1 - \pi_2 < 0 \)

2. if the firm wanted to test whether a greater proportion of workers would currently like to attend a self-improvement course than in the past, which represents the relevant hypotheses?
   a) \( H_0: \pi_1 - \pi_2 = 0 \) versus \( H_1: \pi_1 - \pi_2 \neq 0 \)
   b) \( H_0: \pi_1 - \pi_2 \neq 0 \) versus \( H_1: \pi_1 - \pi_2 = 0 \)
   c) \( H_0: \pi_1 - \pi_2 \leq 0 \) versus \( H_1: \pi_1 - \pi_2 > 0 \)
   d) \( H_0: \pi_1 - \pi_2 \geq 0 \) versus \( H_1: \pi_1 - \pi_2 < 0 \)

3. what is the unbiased point estimate for the difference between the two population proportions?
   a) 0.06
   b) 0.10
   c) 0.15
   d) 0.22

4. what is/are the critical value(s) when performing a Z test on whether population proportions are different if \( \alpha = 0.05 \)?
   a) \( \pm 1.645 \)
   b) \( \pm 1.96 \)
   c) \(-1.96\)
   d) \( \pm 2.08 \)

5. what is/are the critical value(s) when testing whether population proportions are different if \( \alpha = 0.10 \)?
   a) \( \pm 1.645 \)
   b) \( \pm 1.96 \)
   c) \(-1.96\)
   d) \( \pm 2.08 \)
6. what is/are the critical value(s) when testing whether the current population proportion is higher than before if $\alpha = 0.05$?
   a) $\pm 1.645$
   b) $+ 1.645$
   c) $\pm 1.96$
   d) $+ 1.96$

7. what is the estimated standard error of the difference between the two sample proportions?
   a) 0.629
   b) 0.500
   c) 0.055
   d) 0

8. what is the value of the test statistic to use in evaluating the alternative hypothesis that there is a difference in the two population proportions?
   a) 4.335
   b) 1.96
   c) 1.093
   d) 0

9. the company tests to determine at the 0.05 level whether the population proportion has changed from the previous study. Which of the following is most correct?
   a) Reject the null hypothesis and conclude that the proportion of employees who are interested in a self-improvement course has changed over the intervening 10 years.
   b) **Do not reject the null hypothesis and conclude that the proportion of employees who are interested in a self-improvement course has not changed over the intervening 10 years.**
   c) Reject the null hypothesis and conclude that the proportion of employees who are interested in a self-improvement course has increased over the intervening 10 years.
   d) Do not reject the null hypothesis and conclude that the proportion of employees who are interested in a self-improvement course has increased over the intervening 10 years.

10. construct a 99% confidence interval estimate of the difference in proportion of workers who would like to attend a self-improvement course in the recent study and the past study.
   $-0.08$ to $0.20$

11. construct a 95% confidence interval estimate of the difference in proportion of workers who would like to attend a self-improvement course in the recent study and the past study.
   $-0.05$ to $0.17$

12. construct a 90% confidence interval estimate of the difference in proportion of workers who would like to attend a self-improvement course in the recent study and the past study.
   $-0.03$ to $0.15$
SECTION I: MULTIPLE-CHOICE

1. The \( Y \)-intercept \( (b_0) \) represents the
   a) predicted value of \( Y \) when \( X = 0 \).
   b) change in estimated average \( Y \) per unit change in \( X \).
   c) predicted value of \( Y \).
   d) variation around the sample regression line.

2. The slope \( (b_1) \) represents
   a) predicted value of \( Y \) when \( X = 0 \).
   b) the estimated average change in \( Y \) per unit change in \( X \).
   c) the predicted value of \( Y \).
   d) variation around the line of regression.

3. If the correlation coefficient \( (r) = 1.00 \), then
   a) the \( Y \)-intercept \( (b_0) \) must equal 0.
   b) the explained variation equals the unexplained variation.
   c) there is no unexplained variation.
   d) there is no explained variation.

4. If the correlation coefficient \( (r) = 1.00 \), then
   a) all the data points must fall exactly on a straight line with a slope that equals 1.00.
   b) all the data points must fall exactly on a straight line with a negative slope.
   c) all the data points must fall exactly on a straight line with a positive slope.
   d) all the data points must fall exactly on a horizontal straight line with a zero slope.

5. Assuming a linear relationship between \( X \) and \( Y \), if the coefficient of correlation \( (r) \) equals \(-0.30\),
   a) there is no correlation.
   b) the slope \( (b_1) \) is negative.
   c) variable \( X \) is larger than variable \( Y \).
   d) the variance of \( X \) is negative.
6. The strength of the linear relationship between two numerical variables may be measured by the
   a) scatter diagram.
   b) coefficient of correlation.
   c) slope.
   d) Y-intercept.

7. In a simple linear regression problem, \( r \) and \( b_1 \)
   a) may have opposite signs.
   b) must have the same sign.
   c) must have opposite signs.
   d) are equal.

SECTION II: FREE RESPONSE QUESTIONS

TABLE (A)

A candy bar manufacturer is interested in trying to estimate how sales are influenced by the price of their product. To do this, the company randomly chooses 6 small cities and offers the candy bar at different prices. Using candy bar sales as the dependent variable, the company will conduct a simple linear regression on the data below:

<table>
<thead>
<tr>
<th>City</th>
<th>Price ($)</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>River Falls</td>
<td>1.30</td>
<td>100</td>
</tr>
<tr>
<td>Hudson</td>
<td>1.60</td>
<td>90</td>
</tr>
<tr>
<td>Ellsworth</td>
<td>1.80</td>
<td>90</td>
</tr>
<tr>
<td>Prescott</td>
<td>2.00</td>
<td>40</td>
</tr>
<tr>
<td>Rock Elm</td>
<td>2.40</td>
<td>38</td>
</tr>
<tr>
<td>Stillwater</td>
<td>2.90</td>
<td>32</td>
</tr>
</tbody>
</table>

1. what is the estimated slope parameter for the candy bar price and sales data?
   a. 161.386
   b. 0.784
   c. 3.810
   d. 48.193

2. what is the estimated average change in the sales of the candy bar if price goes up by $1.00?
   a. 161.386
   b. 0.784
   c. 3.810
   d. 48.193

3. what is the coefficient of correlation for these data?
   a. 0.8854
   b. 0.7839
   c. 0.7839
   d. 0.8854

4. what is the percentage of the total variation in candy bar sales explained by the regression model?
   a. 100%
   b. 88.54%
   c. 78.39%
   d. 48.19%
TABLE (B)

The management of a chain electronic store would like to develop a model for predicting the weekly sales (in thousand of dollars) for individual stores based on the number of customers who made purchases. A random sample of 12 stores yields the following results:

<table>
<thead>
<tr>
<th>Sales (Thousands of Dollars)</th>
<th>Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.20</td>
<td>907</td>
</tr>
<tr>
<td>11.05</td>
<td>926</td>
</tr>
<tr>
<td>8.21</td>
<td>713</td>
</tr>
<tr>
<td>9.21</td>
<td>741</td>
</tr>
<tr>
<td>9.42</td>
<td>780</td>
</tr>
<tr>
<td>10.08</td>
<td>898</td>
</tr>
<tr>
<td>6.73</td>
<td>510</td>
</tr>
<tr>
<td>7.02</td>
<td>529</td>
</tr>
<tr>
<td>6.12</td>
<td>460</td>
</tr>
<tr>
<td>9.52</td>
<td>872</td>
</tr>
<tr>
<td>7.53</td>
<td>650</td>
</tr>
<tr>
<td>7.25</td>
<td>603</td>
</tr>
</tbody>
</table>

1. what are the values of the estimated intercept and slope? 
   \[1.4464, 0.0100\]

2. what is the value of the coefficient of determination? 
   \[0.9453\]

3. what is the value of the coefficient of correlation? 
   \[0.9723\]
SECTION I: MULTIPLE-CHOICE

For each question in this section, circle the correct answer. Each problem is worth 1 point.

5. Which of the following is a continuous quantitative variable?
   a) The amount of milk produced by a cow in one 24-hour period
   b) The color of a student’s eyes
   c) The number of employees of an insurance company
   d) The number of gallons of milk sold at the local grocery store yesterday

6. The classification of student major (accounting, economics, management, marketing, other) is an example of
   a) a discrete random variable.
   b) a continuous random variable.
   c) a categorical random variable.
   d) a parameter.

7. In a right-skewed distribution
   a) the median equals the arithmetic mean.
   b) the median is larger than the arithmetic mean.
   c) the median is less than the arithmetic mean.
   d) none of the above.

8. When extreme values are present in a set of data, which of the following descriptive summary measures are most appropriate:
   a) interquartile range and median.
   b) CV and range.
   c) arithmetic mean and standard deviation.
   d) variance and interquartile range.

9. According to the Chebyshev rule, at least 93.75% of all observations in any data set are contained within a distance of how many standard deviations around the mean?
   a) 1
   b) 4
   c) 2
   d) 3

5. Which of the following about the normal distribution is NOT true?
   a. It is a discrete probability distribution.
6. Theoretically, the mean, median, and mode are the same.
c. About $2/3$ of the observations fall within $\pm 1$ standard deviation from the mean.
d. Its parameters are the mean, $\mu$, and standard deviation, $\sigma$.

6. For some positive value of $A$, the probability that a standard normal variable is between 0 and $A$ is 0.4332. The value of $A$ is
   a. 0.10
   b. 1.50
   c. 0.50
   d. 1.00

12. The standard error of the population proportion will become larger
   a) as population proportion approaches 0.
   b) as population proportion approaches 1.00.
   c) as the sample size increases.
   d) as population proportion approaches 0.50.

9. Why is the Central Limit Theorem so important to the study of sampling distributions?
   a) It allows us to disregard the size of the sample selected when the population is not normal.
   b) It allows us to disregard the shape of the sampling distribution when the size of the population is large.
   c) It allows us to disregard the size of the population we are sampling from.
   d) It allows us to disregard the shape of the population when the sample size is large.

10. The evening host of a dinner reached into a bowl, mixed all the tickets around, and selected the ticket to award the grand door prize. What sampling method was used?
    a) Simple random sample
    b) Systematic sample
    c) Stratified sample
    d) Cluster sample

**SECTION II: TRUE OR FALSE**

For each question in this section, indicate whether the sentence is TRUE or False. Each problem is worth 1 point.

1. (   ) A sample is the totality of items or things under consideration.
   False

2. (   ) The quality (“terrible”, “poor”, “fair”, “acceptable”, “very good” and “excellent”) of a day care center is an example of a numerical variable.
   False

3. (   ) In a sample of size 40, the sample mean is 15. In this case, the sum of all observations in the sample is $\sum X_i = 600$.
   True
4. The probability that a standard normal random variable, \( Z \), is between 1.50 and 2.10 is the same as the probability \( Z \) is between – 2.10 and – 1.50.
   True

5. As a general rule, an observation is considered an extreme value if its \( Z \) score is greater than 1.
   False

6. The mean of the sampling distribution of a sample proportion is the population proportion, \( \pi \).
   True

**SECTION III: FREE RESPONSE QUESTIONS**

(i) (4 Points) The fill amount of bottles of a soft drink is normally distributed, with a mean of 2.0 liters and a standard deviation of 0.06 liter. If you select a random sample of 36 bottles, what is the probability that the sample mean will be
1. (2 Points) between 1.99 and 2.0 liters? **0.3413**

2. (2 Points) The probability is 99% that the sample mean amount of soft drink will be at least how much? **1.9767**

(ii) (4 Points) You were told that the mean score on a statistics exam is 75 with the scores normally distributed. In addition, you know the probability of a score between 55 and 60 is 4.41% and that the probability of a score greater than 90 is 6.68%.
   a. (2 Points) What is the probability of a score greater than 95? **2.27% or 0.0227**

   b. (2 Points) The middle 86.64% of the students will score between which two scores? **60 and 90**
(iii) **(6 Points)** A study at a college in the west coast reveals that, historically, 45% of their students are minority students. Suppose that a random sample of size 75 are selected.

a. **(2 Points)** What is the standard error of the proportions of students in the samples who are minority students? \(0.05745\)

b. **(2 Points)** What is the probability that between 30% and 50% of the students in the sample will be minority students? \(0.8033\)

c. **(2 Points)** 80% of the samples will have less than what sample proportion of minority students? \(49.83\)
The Islamic University of Gaza
Faculty of Commerce
Department of Economics and Political Sciences

An Introduction to Statistics Course (ECOE 1302)
Spring Semester 2009-2010

Final Examination
17/6/2010

Question #1: [22 ½ Points]
For each question in this section, circle the correct answer. Each problem is worth 1 ½ points.

7. The t test for the mean difference between 2 related populations assumes that the
   a. population sizes are equal.
   b. sample variances are equal.
   c. population of differences is approximately normal or sample sizes are large enough.
   d. All of the above.

8. The null and alternative hypotheses to determine if the number of tissues used during a cold is less than 60.
   a) \( H_0 : \mu \leq 60 \) and \( H_1 : \mu > 60 \).
   b) \( H_0 : \mu \geq 60 \) and \( H_1 : \mu < 60 \).
   c) \( H_0 : \bar{X} \geq 60 \) and \( H_1 : \bar{X} < 60 \).
   d) \( H_0 : \bar{X} = 52 \) and \( H_1 : \bar{X} \neq 52 \).

3. When extreme values are present in a set of data, which of the following descriptive summary measures are most appropriate:
   a. CV and range.
   b. arithmetic mean and standard deviation.
   c. interquartile range and median.
   d. variance and interquartile range.

4. The Y-intercept \((b_0)\) represents the
   a. estimated average \(Y\) when \(X = 0\).
   b. change in estimated average \(Y\) per unit change in \(X\).
   c. predicted value of \(Y\).
   d. variation around the sample regression line.

5. The t distribution
   e) assumes the population is normally distributed.
   f) approaches the normal distribution as the sample size increases.
   g) has more area in the tails than does the normal distribution.
   h) All of the above.
6. An economist is interested in studying the incomes of consumers in a particular region. The population standard deviation is known to be $1,000. A random sample of 50 individuals resulted in an average income of $15,000. What total sample size would the economist need to use for a 95% confidence interval if the width of the interval should not be more than $100?
   a) $n = 1537$
   b) $n = 385$
   c) $n = 40$
   d) $n = 20$

7. The width of a confidence interval estimate for a proportion will be
   a) narrower for 99% confidence than for 95% confidence.
   b) wider for a sample size of 100 than for a sample size of 50.
   c) **narrower for 90% confidence than for 95% confidence.**
   d) narrower when the sample proportion is 0.50 than when the sample proportion is 0.20.

8. In testing for differences between the means of two related populations, the null hypothesis is
   3. $H_0 : \mu_D = 2$.
   4. $H_0 : \mu_D = 0$.
   5. $H_0 : \mu_D < 0$.
   6. $H_0 : \mu_D > 0$.

9. Which of the following would be an appropriate alternative hypothesis?
   e. The mean of a population is equal to 55.
   f. The mean of a sample is equal to 55.
   g. **The mean of a population is greater than 55.**
   h. The mean of a sample is greater than 55.

10. A Type I error is committed when
    a) **we reject a null hypothesis that is true.**
    b) we don't reject a null hypothesis that is true.
    c) we reject a null hypothesis that is false.
    d) we don't reject a null hypothesis that is false.

11. If an economist wishes to determine whether there is evidence that average family income in a community exceeds $25,000
    a) either a one-tailed or two-tailed test could be used with equivalent results.
    b) **a one-tailed test should be utilized.**
    c) a two-tailed test should be utilized.
    d) None of the above.

12. If the $p$-value is less than $\alpha$ in a two-tailed test,
    a) the null hypothesis should not be rejected.
    b) **the null hypothesis should be rejected.**
    c) a one-tailed test should be used.
    d) no conclusion should be reached.
13. It is possible to directly compare the results of a confidence interval estimate to the results obtained by testing a null hypothesis if
   a) a two-tailed test for $\mu$ is used.
   b) a one-tailed test for $\mu$ is used.
   c) Both of the previous statements are true.
   d) None of the previous statements is true.

14. We have created a 95% confidence interval for $\mu$ with the result (10, 15). What decision will we make if we test $H_0 : \mu = 16$ versus $H_1 : \mu \neq 16$ at $\alpha = 0.05$?
   a) Reject $H_0$ in favor of $H_1$.
   b) Accept $H_0$ in favor of $H_1$.
   c) Fail to reject $H_0$ in favor of $H_1$.
   d) We cannot tell what our decision will be from the information given.

15. Suppose a 95% confidence interval for $\mu$ turns out to be (1,000, 2,100). Give a definition of what it means to be “95% confident” as an inference.
   a) In repeated sampling, the population parameter would fall in the given interval 95% of the time.
   b) In repeated sampling, 95% of the intervals constructed would contain the population mean.
   c) 95% of the observations in the entire population fall in the given interval.
   d) 95% of the observations in the sample fall in the given interval.

Question #2: [22½ Points]
For each question in this section, indicate whether the sentence is TRUE or False. Each problem is worth 1½ points.

1. (     ) When we test for differences between the means of 2 independent populations, we can only use a two-tailed test.
   False
2. (     ) Repeated measurements from the same individuals is an example of data collected from 2 related populations.
   True
3. (     ) The statement of the null hypothesis always contains an equality.
   True
4. (     ) The smaller is the $p$-value, the stronger is the evidence against the null hypothesis.
   True
5. (     ) The $t$ distribution is used to construct confidence intervals for the population mean when the population standard deviation is unknown.
   True
6. (     ) Given a sample mean of 2.1 and a population standard deviation of 0.7 from a sample of 10 data points, a 90% confidence interval will have a width of 2.36.
   False
7. (     ) The sample mean is a point estimate of the population mean.
   True
8. (     ) A point estimate consists of a single sample statistic that is used to estimate the true population parameter.
   True
9. (     ) In a set of numerical data, the value for Q2 is always halfway between Q1 and Q3.
   False
10. (     ) The coefficient of variation is a measure of relative variation.
    True
11. (True) The t-distribution approaches the standardized normal distribution when the number of degrees of freedom increases.

12. (True) For a given data set, the confidence interval will be wider for 95% confidence than for 90% confidence.

13. (True) A sampling distribution is a distribution for a statistic.

14. (False) The type of TV one owns is an example of an ordinal scaled variable.

15. The probability that a standard normal random variable, Z, falls between –1.50 and 0.81 is 0.7242.

**Question #3: [16 Points]**

To test the effectiveness of a business school preparation course, 8 students took a general business test before and after the course. Denote \( X_1 \): Before Course (1), \( X_2 \): After Course (2), Difference \( D = X_2 - X_1 \), Suppose \( \overline{D} = -50 \), \( S_D = 65.03 \)

a. (2 Points) Sate the null and alternative hypotheses to determine the effectiveness of a business school preparation course.

b. (6 Points) Using the sample information provided, calculate the value of the test statistic.

c. (4 Points) Compute the P-value.

d. (2 Points) State your decision at level of significance \( \alpha = .05 \)

e. (2 Points) State your conclusion

**Question #4: [24 Points]**

The dean of a college is interested in the proportion of graduates from his college who have a job offer on graduation day. He is particularly interested in seeing if there is a difference in this proportion for accounting and economics majors. In a random sample of 100 of each type of major at graduation, he found that 65 accounting majors and 52 economics majors had job offers. If the accounting majors are designated as “Group 1” and the economics majors are designated as “Group 2,” perform the appropriate hypothesis test using a level of significance of 0.05.

a. (2 Points) Sate the hypotheses the dean should use

\[ H_0: \pi_1 - \pi_2 = 0 \text{ versus } H_1: \pi_1 - \pi_2 \neq 0 \]

b. (6 Points) Using the sample information provided, calculate the value of the test statistic.

\[ Z = 1.866 \]

c. (4 Points) construct a 95% confidence interval estimate of the difference in proportion between accounting majors and economic majors who have a job offer on graduation day. Interpret.

\(-0.01\text{ to } 0.27\)
d. (4 Points) Compute the P-value.

0.0621

d. (2 Points) State your decision based on your result in Part (c)

e. (2 Points) State your decision based on your result in Part (d)

f. (2 Points) Are the two decisions in (d) and (e) consistent? Why or why not?

g. (2 Points) State your conclusion

**Question #5: [15 Points]**

The managers of a brokerage firm are interested in finding out if the number of new clients a broker brings into the firm affects the sales generated by the broker. They sample 12 brokers and determine the number of new clients (X) they have enrolled in the last year and their sales amounts in thousands of dollars (Y).

\[
\begin{align*}
\sum_{i=1}^{12} X_i &= 301, \quad \sum_{i=1}^{12} Y_i = 549, \quad \sum_{i=1}^{12} X_i Y_i = 14868, \\
\sum_{i=1}^{12} X_i^2 &= 8531, \quad \sum_{i=1}^{12} Y_i^2 = 26681, S_X^2 = 89.17, S_Y^2 = 142.20
\end{align*}
\]

a) (6 Points) Compute the values of the estimated intercept and slope. Explain

**Slope = 1.12, Y-intercept = 17.7**

b) (5 Points) Compute the value of the coefficient of correlation. Interpret

c) (2 Points) Compute the value of the coefficient of determination. Interpret

**0.785:** 78.5% of the total variation in sales generated can be explained by the number of new clients brought in.

d) (2 Points) Compute the prediction for the amount of sales (in $1,000s) for a person who brings 25 new clients into the firm.

**45.66**

**Bonus:**

(a) (3 Points) For a normally distributed variable, what is the probability between \(\mu - 0.67\sigma\) and \(\mu + 0.67\sigma\) equals .50?
(b) (3 Points) A fast food chain sells hamburger that they claim has sodium content of 650 milligrams. A simple random sample of 35 hamburgers was analyzed for sodium content. A 99% confidence interval for the population mean sodium content, \( \mu \), of such hamburgers is (652, 672). If we were to use the preceding data to test the hypotheses \( H_0: \mu = 650 \) versus \( H_a: \mu \neq 650 \). At a 1% significance level, would we reject the null hypothesis? Explain.
SECTION I: MULTIPLE-CHOICE
For each question in this section, circle the correct answer. Each problem is worth 1 point.

10. The process of using sample statistics to draw conclusions about true population parameters is called
   a) statistical inference.
   b) the scientific method.
   c) sampling.
   d) descriptive statistics.

11. A summary measure that is computed to describe a characteristic from only a sample of the population is called
   a) a parameter.
   b) a census.
   c) a statistic.
   d) the scientific method.

12. In a right-skewed distribution
   a) the median equals the arithmetic mean.
   b) the median is less than the arithmetic mean.
   c) the median is larger than the arithmetic mean.
   d) none of the above.

11. When extreme values are present in a set of data, which of the following descriptive summary measures are most appropriate:
   a) CV and range.
   b) arithmetic mean and standard deviation.
   c) interquartile range and median.
   d) variance and interquartile range.

12. In its standardized form, the normal distribution
   a) has a mean of 0 and a standard deviation of 1.
   b) has a mean of 1 and a variance of 0.
   c) has an area equal to 0.5.
   d) cannot be used to approximate discrete probability distributions.
7. If a particular batch of data is approximately normally distributed, we would find that approximately
   a. 2 of every 3 observations would fall between ±1 standard deviation around the mean.
   b. 4 of every 5 observations would fall between ±1.28 standard deviations around the mean.
   c. 19 of every 20 observations would fall between ±2 standard deviations around the mean.
   d. All the above.

8. The Central Limit Theorem is important in statistics because
   e) for a large \( n \), it says the population is approximately normal.
   f) for any population, it says the sampling distribution of the sample mean is approximately normal, regardless of the sample size.
   g) for a large \( n \), it says the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.
   h) for any sized sample, it says the sampling distribution of the sample mean is approximately normal.

13. Which of the following statements about the sampling distribution of the sample mean is incorrect?
   a) The sampling distribution of the sample mean is approximately normal whenever the sample size is sufficiently large (\( n \geq 30 \)).
   b) The sampling distribution of the sample mean is generated by repeatedly taking samples of size \( n \) and computing the sample means.
   c) The mean of the sampling distribution of the sample mean is equal to \( \mu \).
   d) The standard deviation of the sampling distribution of the sample mean is equal to \( \sigma \).

14. The width of a confidence interval estimate for a proportion will be
   i) narrower for 99% confidence than for 95% confidence.
   j) wider for a sample size of 100 than for a sample size of 50.
   k) narrower for 90% confidence than for 95% confidence.
   l) narrower when the sample proportion is 0.50 than when the sample proportion is 0.20.

15. A 99% confidence interval estimate can be interpreted to mean that
   a) if all possible samples are taken and confidence interval estimates are developed, 99% of them would include the true population mean somewhere within their interval.
   b) we have 99% confidence that we have selected a sample whose interval does include the population mean.
   c) Both of the above.
   d) None of the above.
SECTION II: TRUE OR FALSE
For each question in this section, indicate whether the sentence is TRUE or False. Each problem is worth 1 point.

1. (       ) The possible responses to the question “How many times in the past three months have you visited a city park?” are values from a discrete variable.

15. (       ) Other things being equal, as the confidence level for a confidence interval increases, the width of the interval increases.

16. (       ) The $t$ distribution is used to develop a confidence interval estimate of the population proportion when the population standard deviation is unknown.

17. (       ) The confidence interval obtained will always correctly estimate the population parameter.

18. (       ) In estimating the population mean with the population standard deviation unknown, if the sample size is 12, there will be 6 degrees of freedom.

19. (       ) As the sample size increases, the effect of an extreme value on the sample mean becomes smaller.

20. (       ) A sampling distribution is a distribution for a statistic.

21. (       ) The probability that a standard normal random variable, $Z$, is less than 5.0 is approximately 0.

22. (       ) The "middle spread," that is the middle 50% of the normal distribution, is equal to one standard deviation.

23. (       ) In a set of numerical data, the value for Q2 is always halfway between Q1 and Q3.

SECTION III: FREE RESPONSE QUESTIONS

(i) (2 Points) The assets in billions of dollars of the five largest bond funds are 19.5, 16.8, 13.7, 12.8, and 10.9. Compute the standard deviation for this population of the five largest bond funds.
(ii) (4 Points) You were told that the mean score on a statistics exam is 75 with the scores normally distributed. In addition, you know the probability of a score between 55 and 60 is 4.41% and that the probability of a score greater than 90 is 6.68%.

a. (2 Points) What is the probability of a score greater than 95?

b. (2 Points) The middle 86.64% of the students will score between which two scores?

(iii) (2 Points) The head of a computer science department is interested in estimating the proportion of students entering the department who will choose the new computer engineering option. Suppose there is no information about the proportion of students who might choose the option. What size sample should the department head take if she wants to be 95% confident that the estimate is within 0.10 of the true proportion?

(iv) (2 Points) A hotel chain wants to estimate the average number of rooms rented daily in each month. The population of rooms rented daily is assumed to be normally distributed for each month with a standard deviation of 24 rooms. During February, a sample of 25 days has a sample mean of 37 rooms. Use this information to calculate a 92% confidence interval for the population mean.
Question #1 (10 Points):
A corporation randomly selects 150 salespeople and finds that 66% who have never taken a self-improvement course would like such a course. The firm did a similar study 10 years ago in which 60% of a random sample of 160 salespeople wanted a self-improvement course. The groups are assumed to be independent random samples. Let $\pi_1$ and $\pi_2$ represent the true proportion of workers who would like to attend a self-improvement course in the recent study and the past study, respectively. At 5% significance level, test whether this proportion has changed from the previous study by using critical value and p-value approaches.
Question #2 (10 Points):
The Excellent Drug Company claims its aspirin tablets will relieve headaches faster than any other aspirin on the market. To determine whether the company claim is valid, random samples of size 15 are chosen from aspirins made by Excellent and the Simple Drug Company. An aspirin is given to each of the 30 randomly selected persons suffering from headaches and the number of minutes required for each to recover from the headache is recorded. The sample results are:

<table>
<thead>
<tr>
<th>Simple</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.9</td>
<td>8.4</td>
</tr>
<tr>
<td>2.14</td>
<td>2.05</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

Sample mean
Sample standard deviation
Sample Size

At 5% significance level, test whether Excellent aspirin cures headaches significantly faster than Simple aspirin.
Question #3 (10 Points):
(a) (7 Points) To test the effectiveness of a business school preparation course, 8 students took a general business test before and after the course. Denote $X_1$: Before Course(1), $X_2$: After Course(2), Difference $D= X_2 - X_1$. Suppose $\bar{D} = 50$, $S_0 = 65.03$. Test at 5% significance level to determine the effectiveness of a business school preparation course.
A statistician selected a sample of 16 receivable accounts. He reported that the sample information indicated the mean of the population ranges from $4,739.80 and $5,260.20 and the sample standard deviation is $400. He neglected to report what confidence level he had used. Based on the above information, what is confidence level?
The Islamic University of Gaza
Faculty of Commerce
Department of Economics and Political Sciences

An Introduction to Statistics Course (ECOE 1302)
Spring Semester 2011

Final Examination
6/6/2011

Question #1: [30 Points]
For each question in this section, circle the correct answer. Each problem is worth 2 points.

Write your choices in the following table

<table>
<thead>
<tr>
<th>Q8</th>
<th>Q7</th>
<th>Q6</th>
<th>Q5</th>
<th>Q4</th>
<th>Q3</th>
<th>Q2</th>
<th>Q1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15</td>
<td>Q14</td>
<td>Q13</td>
<td>Q12</td>
<td>Q11</td>
<td>Q10</td>
<td>Q9</td>
<td></td>
</tr>
</tbody>
</table>

1- If we wish to determine whether there is evidence that the proportion of items of interest is the same in group 1 as in group 2, the appropriate test to use is

a) the \( Z \) test  
\( b) \) the \( \chi^2 \) test  
\( c) \) Both a) and b)  
\( d) \) Neither of a) nor b)

2- When testing for independence in a contingency table with 3 rows and 2 columns, there are ______ degrees of freedom.

\( a) \) 2  
\( b) \) 5  
\( c) \) 6  
\( d) \) 12

Questions 3-4 refer to the following information:

3- A copy machine dealer has on the number \( x \) of copy machines at each of 89 customer locations and the number \( y \) of service calls in a month at each location. Summary calculations give

\[
\bar{x} = 8.4, \quad s_x = 2.1, \quad \bar{y} = 14.2, \quad s_y = 3.8, \quad \text{and} \quad r = 0.86.
\]

What is the slop of the least squares regression line of number of service calls on number of copiers?

\( a) \) None of these  
\( b) \) 0.48  
\( c) \) 1.56  
\( d) \) Can't tell from the information given

4- About what percent of the variation in number of service calls is explained by the linear relation between number of service calls and the number of machines?

\( a) \) 86%  
\( b) \) 86%  
\( c) \) 93%  
\( d) \) 74%  
\( e) \) Can't tell from the information given
Questions 5-6 refer to the following information:
5- A 95% confidence interval for the mean reading achievement score for a population of third grades is (40, 50). The margin of error of this interval is

2.5  d) 5  c) 10  b) 95%  a)
e) The answer cannot be determined from the information given

6- The sample mean is

47.5  d) 42.5  c) 45  b) 0.95  a)
e) The answer cannot be determined from the information given

7- Using the same set of data, you compute a 95% confidence interval and a 99% confidence interval. Which of the following statement is correct?
The 99% interval is wider  b) The intervals have the same width  a) You cannot be determined which interval is wider unless you know n and s
e) The 95% interval is wider
c) The 95% interval is wider
d) You cannot be determined which interval is wider unless you know n and s

8- The P-value for a z test of \( H_0 : \mu = .5 \) vs. \( H_a : \mu < .5 \), where \( z = -2.36 \) is:

a) \( P(z > -2.36) \)  b)\( P(z < -2.36) \)  c) \( 2P(z > -2.36) \)  d) \( 2P(z > -2.36) \)

9 – Suppose you take a simple random sample from a population known to be normally distributed, but the value of \( \sigma \) is unknown. Your sample size is \( n = 10 \). Which formula below should be used to find the 90% confidence interval for the mean?

\[
\bar{x} \pm 1.645 \frac{s}{\sqrt{10}} \quad \text{b) } \bar{x} \pm 1.645 \frac{\sigma}{\sqrt{10}} \quad \text{c) } \bar{x} \pm 1.833 \frac{s}{\sqrt{10}} \quad \text{d) } \bar{x} \pm 1.833 \frac{\sigma}{\sqrt{10}}
\]

10- A type I error is made by
a) failing to reject \( H_0 \) when it is true.  b) rejecting \( H_0 \) when it is false.
c) rejecting \( H_0 \) when it is true.  d) failing to reject \( H_0 \) when it is false.

11- The degrees of freedom of a paired \( t \) test based on \( n = 20 \) pairs is

a) 38  b) 20  c) 19  d) 10  e) None of these

12- Which of the following is not a statistical hypothesis?
a) \( \mu < 100 \)  b) \( \bar{X} > 100 \)  c) \( \mu > 100 \)  d) \( \mu \neq 100 \)

Questions 13-14 refer to the following information:
13- The height of Palestinian men aged 18 to 24 are approximately normally distribution with mean 170 cm and standard deviation 6 cm. Half of all young men are shorter than

a) 164 cm  
(b) 170 cm  
(c) 176 cm  
(d) Can't tell, because the median height is not given.

14- Only about 5% of young men have heights outside the range

a) 158 cm to 176 cm  
b) 164 cm to 176 cm  
c) 152 cm to 188 cm  
d) 146 cm to 188 cm

15- An airplane is only allowed a gross passenger weight of 6,885 kg. If the weights of passengers traveling by air between two cities have a mean of 80 kg and a standard deviation of 18 kg, the approximate probability that the combined weight of 81 passengers will exceed 6,885 kg is:

0.0000  
d) 0.0062  
c) 0.9938  
b) 0.3906  
a) 0.9999

**Question #2: [20 Points]**

Weekly sales (in thousand of dollars) for individual stores based on the number of customers who made purchases. A random sample of 12 stores yields the following results:

\[
\sum_{i=1}^{12} x_i = 8589, \sum_{i=1}^{12} y_i = 103.34, \sum_{i=1}^{12} x_i^2 = 6450413, \sum_{i=1}^{12} y_i^2 = 922.035, \sum_{i=1}^{12} x_i y_i = 76997.25
\]

e) (6 Points) Compute the value of the coefficient of correlation. Interpret

f) (8 Points) Compute the values of the estimated intercept and slope. Interpret
g) (2 Points) Compute the prediction for the amount of sales for 1000 customers who made purchases.

h) (4 Points) Compute the value of the coefficient of determination. Interpret
Question #3: [10 Points]
A market researcher investigated consumer preferences for Coca-Cola and Pepsi before a taste test and after a taste test. The following table summarizes the results from a sample of 200 consumers:

<table>
<thead>
<tr>
<th>SOFT DRINK</th>
<th>PREFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepsi</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>82</td>
</tr>
<tr>
<td>Coca-Cola</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>118</td>
</tr>
</tbody>
</table>

Using a chi-square test to see if there is evidence of a difference in preference for Coca-Cola before and after the taste test by allowing a 5% probability of incorrectly concluding that there is a difference when there is in fact no difference.

- \( H_0: \)
- \( H_a: \)
- Test Statistic:
  - Compute the “P-value” or the rejection region:
  - Decision:
  - Conclusion:
Question #4: [10 Points]
A company wanted to know if attending a course on "how to be a successful salesperson" can increase the average sales of its employees. The company sent six of its salespersons to attend this course. The following table gives the week sales of these salespersons before and after they attended this course.

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>14</td>
<td>9</td>
<td>25</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>19</td>
<td>14</td>
<td>24</td>
<td>24</td>
<td>18</td>
</tr>
</tbody>
</table>

Using the 5% significance level, can you conclude that the mean weekly sales for all salespersons increase as a result of attending this course? Assume that the population of paired difference has a normal distribution.

- **H₀:**
  - H₀:
  - Hₐ:

- **Test Statistic:**
  - Compute the “P-value” or the rejection region:

- **Decision:**

- **Conclusion:**
Question #5: [10 Points]
In 2010, 5% of job applicants who were tested for smoking failed the test. At the 0.05 level, test the claim that the failure rate is now lower if a random sample of 1500 current job applicants results in 60 failures.

- **H₀:**
  - $H_0: \quad \text{Failure rate is 5%}$

- **Hₐ:**
  - $H_1: \quad \text{Failure rate is lower than 5%}$

- **Test Statistic:**

  - Compute the “P-value” or the rejection region:

- **Decision:**

  - Conclusion:
Question #6: [20 Points]
(a) [10 Points] An article reports that (4.0, 5.6) is a 95% confidence interval for the mean length of stay, in days, of patients in hospital for a particular operation. Suppose the sample size is 50, find the sample mean and the standard deviation.

(b) [10 Points] How large a sample size is needed to estimate the mean annual income of CCC company to be within $2000 with probability 0.99? Suppose there is no prior information about the standard deviation of annual income of the CCC company, but we guess that about 68% of their incomes are between $10000 and $40,000 and that this distribution of incomes is approximately bell shaped.