**AP Statistics** **Name:** ________________________________ **Block:** ____

Review Unit VIII – Inference with Means

**Directions:** Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

____ 1. Which of the following is *not* a property of the t-distribution?

A) The *t* curve is centered at 0 and is bell-shaped.
B) The *t* curve is more spread out than the z curve.
C) The *t* curve tends to spread out as the degrees of freedom increase
D) The formula for the t-interval is $\bar{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right)$

____ 2. A group of students at Podunk High School conduct a study to compare the mean number of chocolate chips per cookie between the "Chips Ahoy" brand and the "Chips Galore" brand. Based on the data they collect, they conduct a test on the hypotheses: $H_0$: $\mu_{ahoy} = \mu_{galore}$ $H_A$: $\mu_{ahoy} \neq \mu_{galore}$
When testing at $\alpha = 0.05$, they obtain a p-value of 0.037. Which of the following conclusions is justified?

A) The mean number of chocolate chips is equal between the two brands.
B) There is insufficient evidence of a difference between the mean number of chips between the brands.
C) On average, "Chips Ahoy" cookies have more chocolate chips than "Chips Galore".
D) On average, "Chips Ahoy" cookies have fewer chocolate chips than "Chips Galore".
E) There is evidence of a difference between the mean number of chips between the two brands.

3. The following studies involve comparing pH levels of soil samples to investigate the effects of acid rain on the environment. For which scenario(s) would a matched-pairs *t*-test be appropriate? *(Choose all that apply)*

I. The surface soil pH is measured for 8 randomly selected spots on the north side of a large lake and also for 8 randomly selected spots on the south side of a large lake. Each of the eight selected spots from the north side is randomly paired with one of the spots on the south side, and the difference in surface soil pH within each pair of spots is analyzed.

II. The surface soil pH and subsoil pH are measured for 8 randomly selected locations around a large lake. For each of the 8 locations, the difference between surface pH and subsoil pH is calculated, and these differences are analyzed.

III. The difference between surface soil pH and subsoil pH is calculated for 8 randomly selected locations on the north side of a lake as well as for 8 randomly selected locations on the south side of a lake. An analysis is conducted on whether there is a larger mean difference between surface and subsoil pH on the north side of the lake than on the south side.

4. Remember that the power of a test is defined as the probability of rejecting a false null hypothesis. Suppose that the null hypothesis of a test is $H_O$: $p = 0.50$.

i) Suppose $H_A$: $p > 0.50$, and that $H_A$ is true. For a fixed sample size and significance level $\alpha$, the power of the test will be greatest if the actual proportion is which of the following?

a) 0.40 b) 0.48 c) 0.52 d) 0.60 e) 0.64

ii) Suppose $H_A$: $p < 0.50$, and that $H_A$ is true. For a fixed sample size and significance level $\alpha$, the power of the test will be greatest if the actual proportion is which of the following?

a) 0.40 b) 0.48 c) 0.52 d) 0.60 e) 0.64

iii) Suppose $H_A$: $p \neq 0.50$, and that $H_A$ is true. For a fixed sample size and significance level $\alpha$, the power of the test will be greatest if the actual proportion is which of the following?

a) 0.40 b) 0.48 c) 0.52 d) 0.60 e) 0.64

5. A 98% confidence interval for the average adult male pulse rate is (68.76, 75.48).

   a) Calculate the point estimate for this interval.

   b) Calculate the margin of error for this interval.

6. The owners of a local footwear store wishes to know the mean amount of money spent on running shoes by runners at Lady Bird Lake, to see if consumer spending habits have changed since the recent recession. Previous corporate records indicate that the standard deviation for amount of money spent on running shoes is $21. If the store owners wish to estimate the mean amount of money spent with 98% confidence and a margin of error of no more than $10, what is the minimum number of customers that they should survey?

7. A large city newspaper periodically reports the mean cost of dinner for two people at restaurants in the city. The newspaper staff will collect data from a random sample of restaurants in the city and estimate the mean price using a 96 percent confidence interval. In past years, the standard deviation has always been very close to $35. Assuming that the population standard deviation is $35, which of the following is the minimum sample size needed to obtain a margin of error of no more than $5 ?

**NOTE: For these next 2 problems, if you need to refresh yourself on how confidence levels relate to alpha (significance levels), <u>please read pages 480-481 in your textbook</u> ("Confidence Intervals and Hypothesis Tests").**

8. School officials wish to determine if there has been an <u>increase</u> in the mean SAT math score at Podunk High School since 2008, when the mean score was 672. A 98% 1-sample *t*-interval is constructed from sample data collected from 50 randomly selected current students, and if the entire interval is <u>above</u> 672, then the officials will conclude that there has been an increase in the true mean SAT math score at Podunk. This is equivalent to performing a test on the following hypotheses:

   $H_0: \mu = 672 \qquad H_A: \mu > 672$

   Since a 98% confidence interval was used, what is the equivalent $\alpha$-level of the related 1-sample *t*-test?

9. Another school official wishes to determine if there has been a <u>change</u> in the mean SAT math score at Podunk High School since 2008. The same 98% 1-sample *t*-interval is used, and if 672 is outside of the interval in <u>either</u> direction, then the officials will conclude that there has been a change in the true mean SAT math score at Podunk. This is equivalent to performing a test on the following hypotheses:

   $H_0: \mu = 672 \qquad H_A: \mu \neq 672$

   Since a 98% confidence interval was used, what is the equivalent $\alpha$-level of the related 1-sample *t*-test?

10. The manager at Air Express feels that the weights of packages shipped recently are less than in the past, when packages had a mean weight of 37.5 lbs. A random sample of 18 recent shipments yielded a mean weight of 32.1 lb and standard deviation of 14.2 lb. The weights of these packages are normally distributed. Let $\mu$ represent the mean weight of packages shipped by this company. Conduct a test of significance to see if the mean weight of packages shipped recently has decreased from the past.
**For this problem, please practice finding the p-value BOTH by using the *t*-table AND by using the calculator.**

11. **Shut-ins** are adults who are too ill to leave their homes on a normal basis. Researchers asked 12 randomly selected shut-ins in the Dallas area about the number of hours of television they watched per week. The sample mean number of hours of TV watched per week was 81.3, with a sample standard deviation of 10.26.

a) Calculate and interpret a 90% confidence interval estimate for the mean number of hours of television watched per week by Dallas area shut-ins. **You may assume that the conditions for inference have been checked and verified.**

b) Explain the meaning of 90% confidence level in the context of the shut-in data.

12. **STEREOTYPE THREAT** Back in the old days, one common stereotype was that boys are better at math than girls*. But as a result of this "stereotype", could asking a girl to specify her gender before taking a math test negatively impact her performance on that test? A number of studies in the late 1990's sought to address this question.

Twenty female students that were taking the AP Calculus AB exam at Podunk High School were randomly selected for this study. All 20 took the same test, but half of the girls were randomly assigned to identify their gender before the exam, while the other half were asked to identify their gender after taking the test. The tables below show the raw AP Calculus Exam scores for these 20 students.
*(Note: Although AP scores are reported on a 1 – 5 scale, the raw score for an AP Calculus AB Exam can range from 0 – 108)*

**Group A (were asked to identify gender before the test)**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Raw AP Exam Score | 74 | 59 | 101 | 77 | 63 | 85 | 54 | 40 | 83 | 76 |

Mean Score: 71.2    Standard Deviation of Scores: 17.536

**Group B (were asked to identify gender after the test)**

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Raw AP Exam Score | 63 | 101 | 82 | 69 | 56 | 92 | 83 | 100 | 75 | 86 |

Mean Score: 80.7    Standard Deviation of Scores: 15.056

Do the data provide convincing evidence, at the $\alpha$ = 0.05 level, that the mean exam score for girls who are asked to identify their gender before the exam are lower than girls who are asked to identify their gender after the exam?

*For the record, your math teacher thinks that this is **absolutely** untrue.

13. **AP TATs** AP Statistics students at PHS (Podunk High School) are required to take the TAT (Thinking Assessment Test) before the end of the year to prove their aptitude in logical thinking (better study for those TATs!!!). A fellow teacher at PHS designs a 5-week after-school TAT preparatory course which claims to help students "beat" the TAT. A random sample of 9 students are selected to participate in a study, in which each of the students takes the TAT twice – once before and once again after participating in the prep course. Their data is shown below (possible scores range from 0 to 800; higher numbers are better).

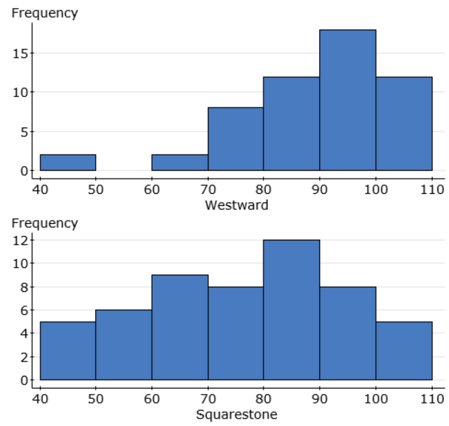
| Student | TI | VS | AS | SP | BC | JS | JE | JY | SS |
|---|---|---|---|---|---|---|---|---|---|
| Score Before TAT Prep | 725 | 620 | 600 | 677 | 750 | 595 | 630 | 640 | 616 |
| Score After TAT Prep | 760 | 615 | 624 | 685 | 740 | 600 | 648 | 650 | 615 |

a) Based on the data, is there evidence at the 5% significance level that the mean TAT score for students at PHS is higher after participating in the TAT preparatory course than before participating in the prep course?

b) Estimate the mean difference in TAT scores before and after taking the prep course for students at PHS by calculating a confidence interval. Use a confidence level that is appropriate based on the level of significance and hypotheses used in part (a).

14. A data analyst working for the Podunk Independent School District is in charge of determining how much of a difference there is in the mathematical abilities of students at two of its local high schools. Among all students who have completed at least Algebra II, separate random samples of 54 students at Westward High and 53 students from Squarestone High were selected to take an upper-level mathematics exam. The same test was given to both groups of students. The summary statistics and graphs of the sample data for the scores are shown below.

| | Westward | Squarestone |
|---|---|---|
| Number of students | 54 | 53 |
| Mean score | 87.9 | 75.1 |
| Standard Deviation | 13.11 | 17.35 |




a) Calculate the standard error of the difference between the mean scores for the two schools.

b) Use a 2-sample t-interval to estimate the difference in the mean exam scores between the two schools. Use a 90% confidence level. Be sure to verify the appropriate conditions.

c) Interpret the meaning of the 90% confidence level in context.

15. A company manufactures and markets a type of disposable soup bowl that is used in many ramen noodle restaurants. The company has created a new bowl that it believes insulates better than the current bowl. A random sample of soup bowls for each of the two types will be selected. In each sample, each bowl will be filled with the same amount of soup that has been heated to 130 degrees Fahrenheit (˚F). The amount of time (in minutes) it takes for the soup to cool to 100 ˚F will be measured for each bowl.

The hypotheses that the company will test are shown below, where $\mu_N$ is the true mean time it takes soup to cool from 130 ˚F to 105 ˚F in the new bowl and $\mu_C$ is the true mean time it takes soup to cool from 130 ˚F to 105 ˚F in the current bowl.

$H_O$: $\mu_N = \mu_C$
$H_A$: $\mu_N > \mu_C$

a) Describe a Type II error in the context of the study.

b) The company is concerned about the probability of a Type II error and the power of the test. Initially, the company was considering using a test procedure that uses a significance level of $\alpha = 0.01$, but is considering using a significance level of $\alpha = 0.10$ instead.
How would increasing the significance level from $\alpha = 0.01$ to $\alpha = 0.10$ affect the probability of a Type II error? How would it affect the power of the test?

c) The marketing department in the company has suggested that a 2-minute increase in the time it takes the soup to cool from 130 ˚F to 105 ˚F would be a noticeable improvement to customers. Suppose the company statistician estimates that the probability of committing a Type II error is 0.19 when the true mean cooling time for the new bowls is 2 minutes greater than the true mean cooling time for the current bowls, and when the test is conducted at a level of significance of $\alpha = 0.10$. Calculate and interpret the power of the test in the context of this problem.

**AP Statistics**
**Unit VIII Review – Inference with Means**

****ANSWERS ONLY**** *(for explanations, please come in for tutorials)*

1. C
2. E
3. II only

4. Since you want the largest effect size in the direction of the alternative hypothesis (Ha)....
   i) e
   ii) a
   iii) e

5. a) 72.12 (the point estimate or sample statistic is the <u>midpoint</u> of the interval!)
   b) 3.36 (the margin of error is the distance from the midpoint to either endpoint)
6. 24
7. 207
8. 0.01 (since you are only using 1 tail)
9. 0.02 (since you are using both tails)

10. 1-sample t-test (for means), df = 17. t = -1.613, *p*-value = 0.0625... (using the t-table: 0.05 < p-value < 0.10)
11. a) 1-sample t-interval, with df = 11: (75.98, 86.62)...
    b) In a large number of repeated random samples using the same sampling method, about 90% of the resulting intervals would contain the true mean number of TV watched per week by Dallas-area shut-ins.

12. 2-sample t-test (for a difference between two means)
    Assuming group 1 is "before" and group 2 is "after":
    t = -1.2998, df = 17.597, p-value = 0.1052
    Conclusion: Since p-value > alpha, we fail to reject H0. We lack evidence that the <u>mean</u> exam score is lower for girls who identify their gender before the exam than girls who identify their gender after the exam.

13. a) Matched pairs t-test (or "Paired t-test")
    Assuming group 1 is "after" and group 2 is "before":
    t = 1.946, df = 8, p-value = 0.0438
    b) Paired t- interval, using 90% confidence: (0.415, 18.251)

14. a) 2.977
    b) 2-sample t-interval (for "W" minus "S"): (7.856, 17.744)
    Note: In spite of the skewed histogram for "W", the sample size is still large enough.........
    c) If this sampling method were repeated a large number of times, about 90% of the resulting intervals would contain the true mean difference of test scores between the two schools.

15. a) In truth, the new bowls have a longer mean cooling time than the current bowls...
    ...but the company fails to detect this.
    b) *[Hint: alpha and beta are inversely related... and beta + power = 1 ... ... ... so............]*
    c) Power = 0.81
    Given that $\alpha$ = 0.10 and that the new bowls have a cooling time that is longer by 2 minutes:
    This is the probability of correctly detecting (or concluding/deciding) that the new bowls have a longer mean cooling time than the current bowls.