Practice Problems for ST 372 Exam 2

1. For each of the following scenarios, specify which type of confidence interval or testing procedure you will use (e.g., one sample $z$-test, one sample $t$-test, large sample $z$-test, two sample $z$-test, two-sample $t$-test, paired $t$-test, $z$-test for proportion, $z$-test for two proportions etc.).

(a) A random sample of 57 adults from a particular community reveals 17 smokers. Researchers want to test whether the prevalence of smoking in this particular community is significantly different than that of the entire United States. The test we use is $2$-test for proportion (one-sample).

(b) Data on both indoor and outdoor concentration of Hexavalent chromium were collected for each house in a random sample of 33 houses. We want to test whether the mean indoor concentration is different from the mean outdoor concentration. Assume that the differences between the concentration measures are normally distributed. The test we use is paired $t$-test.

(c) Data were collected on yield of tomatoes for two different salinity levels of the soil (25 tomatoes in each group). We want to test whether the mean yields differ for different salinity levels. Assume the yields follow normal distribution for each group. The test we use is two-sample $t$-test.

(d) Coating weights for large pipes resulting from a galvanized coating process was measured for 20 randomly chosen pipes. Assume the weights follow a normal distribution. The production standards call for a true average weight of 200lb per pipe. We want to test whether the process satisfies the production standard. The test we use is one-sample $t$-test.

(e) The desired percentage of $\text{SiO}_2$ in a certain type of aluminous cement is 5.5%. We analyze 36 independently obtained samples in a particular production facility. We want to test whether the true average differs from the desired level. The test we use is Can't say ($n < 30$, pop is not normal).

(f) An article reported that 35 of 80 randomly selected broilers from brand A tested positively for salmonella while 66 of 80 broilers for brand B tested positive. We want to test whether the true proportion of non-contaminated broilers from brand A differs from that of brand B. The test we use is $z$-test for two proportions.
2. Suppose an experiment and a sample size are fixed and a test statistic is chosen. Then, decreasing the size of the rejection region to obtain a smaller \( \alpha \) value results in a larger \( \beta \) value for any parameter value consistent with \( H_0 \).

   (a) True

   (b) False

   Will everything else fixed, smaller type I error rate (\( \alpha \)) will result in larger type II error rate (\( \beta \)).

3. For hypothesis tests, the alternative hypothesis always includes the statement of equality between the parameter and the null value.

   (a) True

   (b) False

   Example: \( H_0: \mu = 0 \) vs \( H_a: \mu \neq 0 \)

4. Suppose you compute The Environmental Protection Agency (EPA) wants to determine whether the mean level \( \mu \) of a certain pollutant released into the atmosphere by a chemical company exceeds the EPA guidelines. The upper limit allowed by the EPA is 3 parts per million (ppm). The EPA wants to test \( H_0 : \mu = 3 \) versus \( H_a : \mu > 3 \). Suppose the EPA does not find the company in violation of the guidelines when the mean pollutant level actually exceeds 3 ppm. Then the EPA has made a

   (a) Type I error

   (b) Type II error

   (c) Correct decision

   (d) Cannot be determined

5. For a two-sample \( t \)-test, suppose that one cannot assume that the population variances are equal. It is determined that \( \nu = 9.85 \). Then the number of degrees of freedom used should be is \( 9 \) (round down)

6. Suppose that you are asked to calculate a p-value. You obtain a value of 3.24. Then you should

   (a) Reject \( H_0 \)

   (b) Fail to reject \( H_0 \)

   (c) Recalculate the p-value because you have made an error

   (d) What is a p-value? Never heard of it.
7. To understand the workings of the placebo effect, six patients were measured both with and without the placebo. Assume that the difference of measurements (treatment - control) is normally distributed. Suppose that \( \bar{d} = -0.326 \) and \( s_d = 0.181 \).

(a) Find a 95% confidence interval for the mean of the differences.

\[
\bar{d} \pm t_{0.025, 5} \cdot \frac{s_d}{\sqrt{n}}
\]

95% CI: \( \bar{d} \pm 0.025, 5 \cdot \frac{s_d}{\sqrt{n}} \)

\[
= -0.326 \pm \left( 2.571 \right) \left( \frac{0.181}{\sqrt{6}} \right)
\]

(b) Perform a hypothesis test for zero difference versus non-zero difference using \( \alpha = 0.05 \).

\[H_0: \mu_0 = 0 \quad \text{versus} \quad H_a: \mu_0 \neq 0,\]

where \( \mu_0 \) is the true mean difference.

\[
t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{-0.326}{0.181/\sqrt{6}}
\]

\[
= -4.412
\]

Since \( H_a: \mu_0 \neq 0 \), this is a two-sided test. So rejection region is \( t \leq -t_{0.025, 5} \) or \( t \geq t_{0.025, 5} \)

\[
= -2.571
\]

Conclusion: Since the test statistic value falls inside the rejection region \((-4.412 \text{ is } < -2.571 \text{ })\), we reject \( H_0 \).
9. Two catalysts are being analyzed to determine how they affect the main yield of a chemical process. Specifically, catalyst 1 is currently in use, but catalyst 2 is acceptable. Since catalyst 2 is cheaper, it should be adopted if it does not change the process yield. The plant data are as follows: \( n_1 = 8, \bar{x} = 91.73, s_1^2 = 3.89 \), and \( n_2 = 8, \bar{y} = 93.75, s_2^2 = 4.02 \). Assume that the process yields to which the catalysts were applied are normally distributed with equal variances. Find a 95% confidence interval for the difference in means.

\[
\mu_1 - \mu_2 = \text{mean yield using Catalyst 1 - mean yield using Catalyst 2}
\]

\[
95\% \text{ CI}
\]

Here \( \alpha = 1 - 0.95 = 0.05 \)

Also, this is a two sample \( t \) interval assuming equal variance.

The pooled \( \text{std dev} \) \( \bar{s} = \sqrt{\frac{(8-1)(4.02) + (8-1)(3.89)}{8 + 8 - 2}} \)

\[= 1.989 \]

\[
\text{CI} : \left( \frac{91.73 - 93.75}{\sqrt{\frac{1}{8} + \frac{1}{8}}} \right) \pm 2.145 \left( \frac{1.989}{\sqrt{\frac{1}{8} + \frac{1}{8}}} \right)
\]

\[= -2.02 \pm (2.145)(0.999) \]

\[= (-4.152, 0.112) \]
10. It is known that the amount of weight lost by males on a particular diet for a month follows a normal distribution $N(8, 5^2)$. Nutritionists developed a weight-reducing agent designed to increase the effectiveness of the diet. Let $X$ denote the amount of weight lost by males on the diet who also take the weight-reducing agent. It is believed that $X$ follows a $N(\mu, 5^2)$ distribution. To determine if the weight-reducing agent is effective, the nutritionists want to test the following hypotheses: $H_0 : \mu = 8$ versus $H_a : \mu > 8$.

(a) Suppose $X_1, \ldots, X_n$ denote the weight loss of $n = 16$ men on the diet and weight-reducing agent. If $H_0$ is true, what distribution does the sample mean $\bar{X}$ follow? (Be specific)

If $H_0$ is true ($\mu = 8$), then each $X_i$ has $N(8, 5^2)$ distribution.

So $\bar{X}$ has $N\left(8, \frac{5^2}{16}\right)$ distribution.

(b) Suppose we have collected the data and computed $\bar{x} = 10$. Perform a hypothesis test at $\alpha = 0.05$.

Use one sample $z$-test with known variance.

Test Statistic Value

$z = \frac{\bar{x} - \mu}{\text{Sd}(\bar{x})} = \frac{10 - 8}{\frac{5}{\sqrt{16}}} = 1.6$

Rejection Region

$z \geq z_{0.05}$ (Since $H_a$ is Right Sided)

<table>
<thead>
<tr>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail to reject $H_0$</td>
</tr>
</tbody>
</table>
11. To detect the presence of harmful insects in farm fields, one researcher put up boards covered with a sticky material and examined the insects trapped on the boards. The researcher wants to know which colors best attract insects. She placed 46 boards of each of the 2 colors (blue, green) at random locations in a field of oats and measured the number of beetles trapped.

Summary statistics:

<table>
<thead>
<tr>
<th>color</th>
<th>n</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>46</td>
<td>14.83</td>
<td>2.18</td>
</tr>
<tr>
<td>Green</td>
<td>46</td>
<td>31.16</td>
<td>2.57</td>
</tr>
</tbody>
</table>

Perform an appropriate hypothesis test to answer the researcher’s question. Make sure to clearly specify the null and alternative hypotheses, test statistic, rejection region (or p-value) and your conclusion. Use $\alpha = 0.05$.

Sample sizes for both groups are larger than 40, so use two sample $z$-test (large samples)

Hypotheses

$H_0: \mu_{\text{Blue}} - \mu_{\text{Green}} = 0$ vs $H_1: \mu_{\text{Blue}} - \mu_{\text{Green}} \neq 0$

Test statistic value

$$z = \frac{14.83 - 31.16}{\sqrt{\frac{2.18^2}{46} + \frac{2.57^2}{46}}} = -32.86$$

Rejection region

$$z \leq -2.58$$ or $$z \geq 2.58$$

Conclusion

$z = -32.86$ falls inside rejection region. So, reject $H_0$. 

7