There are totally 42 points in the exam. The students with score higher than or equal to 40 points will receive 40 points. Please write down your name, student ID number, and mark your division below.

NAME: 

ID: 


1. (20 points). (1 point each) Fill in the blanks. Answers must be filled in the blanks.

(a) (3 points). Suppose 10 observations were independently observed from $N(\mu, 12.4)$ with sample mean 3.721. Then, the 90% confidence interval for $\mu$ is ____________ and the 95% confidence interval for $\mu$ is ____________. If we want the 90% confidence interval to be shorter than 0.1, we need the sample size $n$ to be at least ____________.

(b) (3 points). Suppose 8 observations were independently observed from $N(\mu, \sigma^2)$ with sample mean 17.86 and sample standard error 4.71, where $\mu$ and $\sigma^2$ are unknown. Then, the 95% confidence interval for $\mu$ is ____________. The 95% confidence interval for $\sigma^2$ is ____________ and the 95% confidence interval for $\sigma$ is ____________.

(c) (3 points). Suppose $X \sim \chi^2_n$, where $n$ is the degrees of freedom. Assume $c_1$ and $c_2$ satisfy $P(X > c_1) = 0.005$ and $P(X < c_2) = 0.005$. If $n = 3$, then $c_1 = __________$, $c_2 = __________$, and $P(X < -c_1) = __________$.

(d) (3 points). Let $T \sim t_n$, where $n$ is the degrees of freedom. Assume $c$ satisfies $P(|T| \leq c) = 0.99$. If $n = 3$, $P(T > c) = __________$, $P(T < -c) = __________$, and $c = __________$.

(e) (3 points). Let $F \sim F_{3,6}$. Assume $c_1$ satisfies $P(F > c_1) = 0.05$ and $P(F < c_2) = 0.05$. Then, $c_1 = __________$ and $c_2 = __________$, and $P(F < -c_2) = __________$.

(f) (3 points). Let $X \sim Bin(280, p)$ and we observed $x = 177$. Then, the 95% confidence interval for $p$ is ____________. To make the 95% confidence interval for $p$ to be less than 0.01, we need the sample size $n$ to be at least __________. Based on this data, if we want to test $H_0 : p \leq 0.5$ against $H_a : p > 0.5$, we conclude ____________ at significance level $\alpha = 0.05$.

(g) (2 point). Let $X_1, \cdots, X_n \sim N(\mu_1, 12)$ and $Y_1, \cdots, Y_n \sim N(\mu_2, 18)$ be random samples from two different populations. Suppose $\bar{x} = 21.44$ and $\bar{y} = 26.53$. If $n = 20$, then the 99% confidence interval for $\mu_1 - \mu_2$ is ____________. If we want the 99% confidence interval less than or equal to 0.2, we need $n$ at least __________.

Answer: (a) [1.889, 5.553], [1.538, 5.904], 13422; (b) [13.92, 21.80], [9.697, 91.893], [3.114, 9.586]; (c) 12.837, 0.072, 0; (d) 0.005, 0.005, 5.841; (e) 4.76, 0.112, 0; (f) [0.5757, 0.6886], 38416, $p > 0.5$; (g) $[-8.250, -1.930]$, 19970.
2. (6 points). Assume \( X \sim Bin(10, p) \). Consider the test

\[ H_0 : p \geq 0.6 \leftrightarrow H_a : p < 0.6. \]

Let the rejection region be \( \{0, 1, 2\} \).

(a) (2 points). Compute the type I error probability when \( p = 0.7 \).

The type I error probability when \( p = 0.7 \) is

\[ P(\bar{X} \leq 3|p = 0.7) = B(2; 10, 0.7) = 0.00159. \]

(b) (2 points). Compute the type II error probability when \( p = 0.3 \).

The type II error probability when \( p = 0.3 \) is

\[ P(X > 2|p = 0.3) = 1 - B(2; 10, 0.3) = 1 - 0.383 = 0.6172. \]

(c) (2 points). Find the significance level of such test.

The significance level is

\[ P(\bar{X} \leq 3|p = 0.6) = B(2; 10, 0.6) = 0.0123. \]

3. (6 points). The following summary data on the strength of joints is taken from an article.

<table>
<thead>
<tr>
<th>Side Coating</th>
<th>Sample Size</th>
<th>Sample Mean</th>
<th>Sample SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without</td>
<td>20</td>
<td>79.53</td>
<td>9.51</td>
</tr>
<tr>
<td>With</td>
<td>20</td>
<td>61.76</td>
<td>5.59</td>
</tr>
</tbody>
</table>

(a) (2 points). Compute the 95% two-sample \( t \) confidence interval for the difference of the true mean strength between the two groups (without side coating and with side coating).

The degree of freedom in the two-sample \( t \) procedure is

\[ \nu = \frac{(9.51^2/20 + 5.59^2/20)^2}{(9.51^2/20)^2/19 + (5.59^2/20)^2/19} = 30.73. \]

Thus, we need to choose \( \nu = 30 \). Then, the 95% two sample \( t \)-confidence interval for the difference is

\[ 79.53 - 61.76 \pm t_{0.025,30} \sqrt{\frac{9.51^2}{20} + \frac{5.59^2}{20}} = [12.73, 22.81]. \]

(b) (2 points). Compute the 95% pooled \( t \) confidence interval for the difference of the true mean strength between the two groups.

In the pooled \( t \) procedure, we have

\[ s_p^2 = \frac{19 \times 9.51^2 + 19 \times 5.59^2}{38} = 60.84 = 7.80^2. \]

Thus, the 95% pooled \( t \) confidence interval for the difference is

\[ 79.53 - 61.76 \pm t_{0.025,38} \times 7.8 \times \sqrt{\frac{1}{20} + \frac{1}{20}} = 17.77 \pm \frac{2.024 \times 7.8}{\sqrt{10}} = [12.77, 22.76]. \]
(c) (2 points). Test whether the variances of the two groups are equal and evaluate whether the pooled t procedure is a reasonable method at $\alpha = 0.1$.

We need $F_{0.05,19,19} \approx F_{0.05,20,19} = 2.23$, $F_{0.95,19,19} = F_{0.05,19,19}^{-1} = 0.448$ and $F^* = 9.51^2/5.59^2 = 2.89$. Since $2.89 \not\in [0.448, 2.23]$, we conclude $\sigma_1^2 \neq \sigma_2^2$. Thus, we say that the pooled t procedure is not a reasonable method at $\alpha = 0.1$.

4. (4 points). In order to study whether the stream water pH-value in the Great Smoky Mountains vary in seasons, a research study measured the pH-value in summer and winter in 10 locations. The data are given below:

<table>
<thead>
<tr>
<th></th>
<th>Location</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pH-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>summer</td>
<td>7.62</td>
<td>6.08</td>
<td>6.46</td>
<td>7.00</td>
<td>7.15</td>
<td>6.53</td>
<td>6.74</td>
<td>5.90</td>
<td>6.61</td>
</tr>
<tr>
<td>winter</td>
<td>7.97</td>
<td>6.50</td>
<td>6.84</td>
<td>7.43</td>
<td>7.16</td>
<td>6.86</td>
<td>7.11</td>
<td>6.09</td>
<td>6.95</td>
</tr>
<tr>
<td>difference</td>
<td>-0.35</td>
<td>-0.42</td>
<td>-0.38</td>
<td>-0.43</td>
<td>-0.01</td>
<td>-0.33</td>
<td>-0.37</td>
<td>-0.19</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

(a) (2 points). Test whether there is a significant difference of the water pH-value between summer and winter at significance level $\alpha = 0.05$.

$$\bar{d} = -0.327$$

and

$$s_d^2 = 0.01767.$$ 

Test $H_0 : \mu_d = 0$ versus $H_a : \mu_d \neq 0$. Since

$$\left| \frac{\bar{d}}{s_d/\sqrt{n}} \right| = \left| \frac{-0.327}{\sqrt{0.01767/10}} \right| = 7.77 > t_{0.025,9} = 2.262,$$

we reject $H_0$ and conclude they are significantly different.

(b) (2 points). Give a 95% confidence interval of the water pH-value difference in Great Smoky Mountains area.

The 95% confidence interval for the difference is

$$\bar{d} \pm t_{0.025,9} \times \frac{s_d}{\sqrt{10}} = -0.327 \pm 2.262 \times \frac{0.1329}{\sqrt{10}} = [-0.4221, -0.2319].$$

5. (6 points). The following table refers to a 1992 survey by the Wright State University School of Medicine and the United Health Services in Dayton, Ohio. The survey asked 2276 students in their final year of high school in a nonurban area near Dayton, Ohio whether they had even used alcohol, cigarettes, or marijuana.

<table>
<thead>
<tr>
<th>Alcohol Used</th>
<th>Cigarette Used</th>
<th>Marijuana Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
(a) (2 points). For those used alcohol, compute the 99% confidence interval of the difference of marijuana used rates between groups of cigarette used or not used.

Here we have \( m = 911 + 538 = 1449 \), \( \hat{p}_1 = 911/1449 = 0.6287 \), \( n = 44 + 456 = 500 \), and \( \hat{p}_2 = 44/500 = 0.088 \). The 99% confidence interval is

\[
\hat{p}_1 - \hat{p}_2 \pm z_{0.005} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_2)}{1449} + \frac{\hat{p}_2(1 - \hat{p}_2)}{500}} = [0.4944, 0.5869].
\]

(b) (2 points). For those cigarette used, compute the 99% confidence interval for the difference of marijuana used rate between groups of alcohol used or not used.

Here, we have \( m = 1449 \), \( n = 46 \), \( \hat{p}_1 = 911/1449 = 0.6287 \), and \( \hat{p}_2 = 3/46 = 0.0652 \). Then, the 99% confidence interval is

\[
\hat{p}_1 - \hat{p}_2 \pm z_{0.005} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_2)}{1449} + \frac{\hat{p}_2(1 - \hat{p}_2)}{46}} = [0.5201, 0.6069].
\]

(c) (2 points). Test whether the rate of marijuana used are the same between groups of cigarette used or not used at significance level \( \alpha = 0.01 \).

They are not the same since none of the confidence interval in (a) and (b) contain 0.