Instructions:

1. Print your name clearly.

2. Do not open the exam until I say “go.”

3. Put away everything except a pencil, a calculator, and your two pages (two sides each) of notes (formula sheets).

4. Attempt all questions.

5. Show your work clearly. Correct answers without enough work may receive no credit.

6. Find the needed tables at the end of the packet. You may tear the tables sheet(s) free.

7. Don’t worry if you don’t finish the test. Just try to score as many points as you can.

8. If a question is ambiguous, resolve the ambiguity in writing. We will consider grading accordingly.

9. The exam will end when I call time. If you continue writing after I call time, you risk a penalty. (The alternative, that you get more time than your peers, is unfair.)

10. You are welcome to turn your exam in to me before I call time. However, if you are still here in the last five minutes, please remain seated until I’ve picked up all the exams.

11. Good luck!
<table>
<thead>
<tr>
<th>Question</th>
<th>Points</th>
<th>Your Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
1. In a summer survey of Madison second grade children, 15 out of 48 had played outside in the past 24 hours.

(a) Find a 90% confidence interval for the true proportion of students who played outside in the past 24 hours.

ANSWER:
The observed numbers of successes and failures, 15 and 33 = 48 − 15, are both > 5. \( \hat{\pi} = \frac{15}{48} = .3125; \) margin = \( z_{.05} \sqrt{\frac{.3125(1-.3125)}{48}} = 1.645(0.0669) \approx .1101; \) interval = .3125 ± .1101

(b) Are the data strong evidence that the population proportion of children who played outside in the past 24 hours is less than 1/2?

ANSWER:
• Hypotheses: \( H_0 : \pi = 1/2 \) vs. \( H_A : \pi < 1/2 \)
• Assumptions: We have a SRS. \( n = 48 \) and \( \pi = 1/2 \implies \) the expected numbers of successes and failures are 24 = 48(1/2) and 24 = 48(1 − 1/2), which are both > 5. I’ll use the test for one proportion.
• Test statistic: \( p = \hat{\pi} = \frac{15}{48} = .3125 \implies z = \frac{.3125-1/2}{\sqrt{\frac{.5(1-.5)}{48}}} \approx -2.60. \)
• P-value: \( P(Z < -2.60) = .0047 \)
• Conclusion: Reject \( H_0. \) The data are strong evidence that the population proportion of children played outside in the past 24 hours is less than 1/2.
2. Consider a test of $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ based on one sample from a normal population with known $\sigma$ and the power of the test to reject $H_0$ when $H_A$ is true because $\mu = \mu_A$. Mark each statement as TRUE or FALSE (circle one).

(a) TRUE / FALSE Increasing the separation $|\mu_0 - \mu_A|$ increases the power.

(b) TRUE / FALSE Increasing the probability of a type I error increases the power.

(c) TRUE / FALSE Increasing the probability of a type II error increases the power.

(d) TRUE / FALSE Increasing the sample size increases the power.

(e) TRUE / FALSE Increasing $\sigma$ increases the power.

ANSWER:

(a) TRUE Increasing the separation $|\mu_0 - \mu_A|$ increases the power.

(b) TRUE Increasing the probability of a type I error increases the power.

(c) FALSE Increasing the probability of a type II error increases the power.

(d) TRUE Increasing the sample size increases the power.

(e) FALSE Increasing $\sigma$ increases the power.
3. Each morning a hospital quality-control staff person checks two pulse oximeters by mounting them on a testing rig that simulates a stable patient. Experience shows that the oximeters give normally distributed readings with variances that match closely and change little, while the mean readings drift from one another over time. In one morning’s check, simple random samples of 5 readings were taken from the first oximeter and of 7 readings from the second. The average oxygen level from the first oximeter was 98.3 with standard deviation 0.31, while the average from the second was 96.1 with standard deviation 0.29.

- Find a 98% confidence interval for the true difference in mean readings from the two oximeters.

**ANSWER:**

We have normal populations and equal variances and so will use the 2-sample \( t \) confidence interval for a difference of two means.

\[
\begin{align*}
\hat{\sigma}_p^2 &= \frac{(5-1)0.31^2 + (7-1)0.29^2}{5+7-2} = 0.0889 \\
1 - \alpha &= 0.98 \implies \alpha = 0.02 \implies \frac{\alpha}{2} = 0.01 \text{ and we need } t_{5+7-2, 0.01} = t_{10, 0.01} = 2.764. \text{ The interval is } (\bar{X} - \bar{Y}) \pm t_{n_X + n_Y - 2} \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} = (98.3 - 96.1) \pm 2.764 \sqrt{0.0889/7 + 0.0889/5} = 2.10 \pm 0.48.
\end{align*}
\]

- Test whether the true mean readings from the two oximeters are different.

**ANSWER:**

We test \( H_0: \mu_X - \mu_Y = 0 \) vs. \( H_0: \mu_X - \mu_Y \neq 0 \).

Our (2-sample \( t \)) test statistic is 
\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(98.3 - 96.1) - 0}{\sqrt{0.0889/5 + 0.0889/7}} = 12.6,
\]

so our \( p \)-value is \( P(|T_{10}| > |12.6|) = 2P(T_{10} > 12.6) < 2(0.001) = 0.002 \), and we reject \( H_0 \). The data are strong evidence that the true mean readings are different.
4. A simple random sample of four young professionals is recruited to test Valerian tea, which is advertised as a sleeping aid. Two people are randomly chosen to take Valerian tea while the other two take Earl Grey tea. Two QQ plots are made, each containing two points that fit a straight line. However, every pair of points fits a straight line, so these plots were a waste of time. Here are the times to fall asleep (in minutes) for the four patients:

<table>
<thead>
<tr>
<th>tea</th>
<th>time to sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valerian</td>
<td>27</td>
</tr>
<tr>
<td>Valerian</td>
<td>41</td>
</tr>
<tr>
<td>Earl Grey</td>
<td>33</td>
</tr>
<tr>
<td>Earl Grey</td>
<td>48</td>
</tr>
</tbody>
</table>

Assuming that the two populations of sleep times have the same shape, use the Wilcoxon rank sum test to find the $p$-value for testing

$H_0$ : the two populations are identical vs.

$H_A$ : the new population is shifted left of the Earl Grey population

ANSWER:

**Rank data sample**

<table>
<thead>
<tr>
<th>rank</th>
<th>data</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>Valerian</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>Earl Grey</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>Valerian</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>Earl Grey</td>
</tr>
</tbody>
</table>

$R_{\text{Valerian}} = 1 + 3 = 4$

$R_{\text{min}} = \frac{2(2+1)}{2} = 3$

$U_{\text{obs}} = 4 - 3 = 1$

For the $p$-value, we need to know $P(U \leq U_{\text{obs}})$, so we write down all possible ranks for Valerian (doing it for Earl Grey, instead, would work too).

<table>
<thead>
<tr>
<th>Sample A ranks</th>
<th>1, 2</th>
<th>1, 3</th>
<th>1, 4</th>
<th>2, 3</th>
<th>2, 4</th>
<th>3, 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>$R_{\text{min}} = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$U = |0, 1, 2, 2, 3, 4|$

$p$-value $= P(U \leq 1) = \frac{2}{6} = \frac{1}{3}$, which is not small (compared to $\alpha = .05$), so we do not reject $H_0$. The data are not strong evidence of a shift.
5. A study on pavement deflection in an airport runway sent a Boeing 777 aircraft and a Boeing 747 aircraft taxiing across four sections of pavement, each with embedded sensors, with these deflections measured (in mm):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boeing 777</td>
<td>4.57</td>
<td>4.48</td>
<td>4.36</td>
<td>4.43</td>
</tr>
<tr>
<td>Boeing 747</td>
<td>4.01</td>
<td>3.87</td>
<td>3.72</td>
<td>3.76</td>
</tr>
</tbody>
</table>

Can you conclude that the (population) mean deflection is greater for the Boeing 777?

- Hypotheses:
  
- Assumptions you’re making:
  
- Test statistic:
  
- P-value:
  
- Conclusion:

**ANSWER:** The samples are not independent—they are paired by pavement section. Let $X =$ Boeing 777 deflection, $Y =$ Boeing 747 deflection, and $D = X - Y$.

- Hypotheses: $H_0 : \mu_D = 0$, $H_1 : \mu_D > 0$
- Assumptions: differences are a random sample from a normal population
- Test statistic: $\{D_i\} = .56,.61,.64,.67, n = 4, \bar{D} = .62$, and $s_D = .0469 \implies t = \frac{\bar{D} - \mu_0}{s_D/\sqrt{n}} = \frac{.62 - 0}{.0469/\sqrt{4}} = 26.44$
- P-value: $P(t_{4-1} > 26.44) \approx 0$ (The table says it’s < .001.)
- Conclusion: Reject $H_0$. The data are strong evidence that Boeing 777 mean deflection is greater.
6. A simple random sample of 8 pills was taken and the time (in minutes) to dissolve in water was measured for each pill. Here are the data:

24 72 294 18 54 66 120 48

A few summary statistics include \( n = 8 \), \( \bar{x} = 87 \), and \( s = 89.4 \). Plots suggest that the population is not normal, so a bootstrap is used. 10000 resamples with replacement are taken, with a \( \hat{t} \) calculated for each resample. Here are some quantiles of the resampling distribution of \( \hat{t} \):

<table>
<thead>
<tr>
<th>Probability</th>
<th>.001</th>
<th>.005</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.50</th>
<th>.90</th>
<th>.95</th>
<th>.99</th>
<th>.995</th>
<th>.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantile</td>
<td>-10.2</td>
<td>-8.4</td>
<td>-7.6</td>
<td>-5.7</td>
<td>-3.7</td>
<td>-0.1</td>
<td>1.0</td>
<td>1.3</td>
<td>1.9</td>
<td>2.1</td>
<td>2.7</td>
</tr>
</tbody>
</table>

(a) Find a 90% confidence interval for the population mean dissolve time for the pills.

**Answer:**

For a 90% interval, we have \( 1 - \alpha = .90 \implies \alpha = .10 \implies \alpha/2 = .05 \) and we need the .05 and .95 upper critical values, \( \hat{t}_{.05} = 1.3 \) and \( \hat{t}_{.95} = -5.7 \). Here is the interval:

\[
\left( \bar{x} - \hat{t}_{(\alpha/2)} \frac{s}{\sqrt{n}}, \bar{x} - \hat{t}_{(1-\alpha/2)} \frac{s}{\sqrt{n}} \right) = \left( 87 - (1.3) \frac{89.4}{\sqrt{8}}, 87 - (-5.7) \frac{89.4}{\sqrt{8}} \right) = (45.9, 267.2)
\]
(b) Estimate the $p$-value for a bootstrap test of $H_0 : \mu = 100$ vs. $H_A : \mu < 100$, where $\mu$ is the population mean dissolve time, at level $\alpha = .05$. (Hint: You cannot use R to estimate the $p$-value as we did in class. You can, nevertheless, estimate the $p$-value and draw a conclusion.)

ANSWER:

$$p\text{-value} = P(T < t_{obs})$$
$$= P(T < \frac{87 - 100}{89.4/\sqrt{8}})$$
$$= P(T < -.41)$$
$$= \text{between .10 and .50,}$$
$$\text{because the .10 quantile is } -3.7 \text{ and the .50 quantile is } -0.1$$
$$> (\alpha = .05)$$

Or, for the original $H_0 : \mu = 60$, which was a typo,

$$p\text{-value} = P(T < t_{obs})$$
$$= P(T < \frac{87 - 60}{89.4/\sqrt{8}})$$
$$= P(T < .85)$$
$$= \text{between .50 and .90,}$$
$$\text{because the .50 quantile is } -0.1 \text{ and the .90 quantile is 1.0}$$
$$> (\alpha = .05)$$

We do not reject $H_0$. The data are not strong evidence that the population mean dissolve time is less than 100 seconds.
(Alternately, since $\mu_0 = 100$ is in the confidence interval, above, we do not reject $H_0$.)