Solutions to Practice Midterm for Stat 371, Fall 2013

1. (a) The P-value is:

\[ P(U \geq 6) = 0.03 + 0.03 + 0.02 + 0.01 + 0.01 = 0.10. \]

(b) The P-value is:

\[ P(U \leq -5) + P(U \geq 5) = 0.01 + 0.01 + 0.03 + 0.03 + 0.04 + \\
0.04 + 0.03 + 0.03 + 0.02 + 0.01 + 0.01 = 0.28. \]

(c) I need to solve the equation \( P(U \geq c) = 0.07 \) for the value of \( c \). By trial-and-error, I obtain:

\[ P(U \geq 7) = 0.03 + 0.02 + 0.01 + 0.01 = 0.07. \]

(d) I need to solve the equation \( P(U \leq c) = 0.04 \) for the value of \( c \). By trial-and-error, I obtain:

\[ P(U \leq -8) = 0.01 + 0.01 + 0.02 = 0.04. \]

2. First, I combine the data and sort:

<table>
<thead>
<tr>
<th>Response</th>
<th>12 13 14 14 16 16 16 28 29 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Rank</td>
<td>1 2 3.5 3.5 6 6 6 8 9 10</td>
</tr>
</tbody>
</table>

(a) Next, \( r_1 = 3.5 + 6 + 6 + 10 = 25.5 \) and \( r_2 = 1 + 2 + 3.5 + 6 + 8 + 9 = 29.5 \).

(b) We need to compute mean ranks:

\[ \bar{r}_1 = 25.5/4 > 6 \text{ is larger than } \bar{r}_2 = 29.5/6 < 5. \]

Thus, (i) is correct:

Descriptively based on ranks, the data from treatment 1 are larger than the data from treatment 2.

3. (a) The unadjusted variance is:

\[ \sigma^2 = \frac{n_1 n_2 (n + 1)}{12} = \frac{4(6)(11)}{12} = 22; \text{ and } \sigma = \sqrt{22} = 4.690. \]

(b) First, note that the values of \( t_i > 1 \) are: 3 and 2. Next, \[ \sum (t_i^3 - t_i) = (27 - 3) + (8 - 2) = 30. \]

Thus, the adjusted variance is:

\[ 22 - \frac{n_1 n_2 \sum (t_i^3 - t_i)}{12 n (n - 1)} = 22 - \frac{4(6)(30)}{12(10)(9)} = 21.33; \text{ and } \sigma = \sqrt{21.33} = 4.618. \]
4. The data are denoted by $x$’s and the sample size is $m = 100$.

(a) According to the Empirical Rule, about 68% of the data fall between:

$$c = \bar{x} - s = 600.00 - 25.38 = 574.62$$

and

$$d = \bar{x} + s = 625.38.$$ 

(b) By counting, 12 observations are smaller than $c$ and 14 observations are larger than $d$.

Thus, 

$$100 - (12 + 14) = 74,$$

or $74\%$ of the observations actually fall between the values of $c$ and $d$.

(c) Its deviation is $563 - \bar{x} = 563 - 600 = -37$.

(d) The sample size is even. The two center positions are 50 and 51. By inspection, $x_{(50)} = 600$ and $x_{(51)} = 601$. Thus, the median is $(600 + 601)/2 = 600.5$.

(e) The mean of data set A is 600.00. Thus, the sum of its 100 numbers is

$$m(\bar{x}) = 100(600.00) = 60,000.$$ 

Thus, the sum of the 97 numbers in data set B is:

$$60,000 - (530 + 656 + 670) = 58,144.$$ 

Thus, the mean of data set B is $58,144/97 = 599.42$.

(f) Data set B consists of 97 numbers, an odd sample size. Thus, position $(97 + 1)/2 = 49$ is the unique center position. The number in position 49 (Remember: don’t count 530; it has been deleted) is 600, the median.

5. There are four possible assignments, corresponding to the four options of the response (rank) to put on treatment 2.

(a) The computations are shown below:

<table>
<thead>
<tr>
<th>Tr. 1</th>
<th>$\bar{x}$</th>
<th>Tr. 2</th>
<th>$\bar{y}$</th>
<th>$u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 6, 12</td>
<td>6</td>
<td>12</td>
<td>12</td>
<td>$-6$</td>
</tr>
<tr>
<td>0, 6, 12</td>
<td>6</td>
<td>12</td>
<td>12</td>
<td>$-6$</td>
</tr>
<tr>
<td>0, 12, 12</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>6, 12, 12</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

Thus, the sampling distribution for $U$ is:

<table>
<thead>
<tr>
<th>$u$ :</th>
<th>-6</th>
<th>2</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(U = u)$ :</td>
<td>0.50</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>
(b) The ranks are 1, 2, 3.5 and 3.5. The computations are shown below:

<table>
<thead>
<tr>
<th>Tr. 1</th>
<th>$r_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3.5</td>
<td>6.5</td>
</tr>
<tr>
<td>1, 2, 3.5</td>
<td>6.5</td>
</tr>
<tr>
<td>1, 3.5, 3.5</td>
<td>8</td>
</tr>
<tr>
<td>2, 3.5, 3.5</td>
<td>9</td>
</tr>
</tbody>
</table>

Thus, the sampling distribution for $R_1$ is:

<table>
<thead>
<tr>
<th>$r_1$</th>
<th>$P(R_1 = r_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5</td>
<td>0.50</td>
</tr>
<tr>
<td>8</td>
<td>0.25</td>
</tr>
<tr>
<td>9</td>
<td>0.25</td>
</tr>
</tbody>
</table>

6. (a) The P-value for $\neq$ is smaller than 1. It is not twice the P-value for $<$; thus, it must be twice the P-value for $>$. Therefore, the P-value for $>$ is

$$0.5(0.2596) = 0.1298.$$

(b) To obtain the P-value for $\neq$, we double each of the other two P-values. The P-value for $\neq$ is the minimum of three numbers: the two we just calculated and 1. For the current situation, the minimum is

$$2(0.2422) = 0.4844.$$

7. (a) The nearly certain interval for the exact significance level is:

$$0.0523 \pm 3\sqrt{\frac{0.0523(0.9477)}{10,000}} = 0.0523 \pm 0.0067.$$

(b) The nearly certain interval for the exact significance level is:

$$0.0520 \pm 3\sqrt{\frac{0.0520(0.9480)}{10,000}} = 0.0520 \pm 0.0067.$$

(c) First, I note that

$$b = 7, c = 10, (b + c) = 7 + 10 = 17 \text{ and } (c - b) = 10 - 7 = 3.$$

The nearly certain interval for the difference between the exact significance levels is:

$$0.0003 \pm 0.0003\sqrt{\frac{10000(17) - 3^2}{9999}} = 0.0003 \pm 0.0012.$$

I have two comments:

i. I am nearly certain that the exact significance levels are equal or very close to being equal.

ii. The precision of this interval (half-width equals 0.0012) is much better than the precision of either of the earlier intervals.
8. (a) On quiz 4 the class had a great deal of difficulty with assigning ranks to ordered categorical responses. Therefore, I will provide quite a bit of detail. The combined data set consists of 100 numbers: 18 1’s; 27 2’s; 39 3’s; and 16 4’s. Thus, the sorted list consists of:

18 1’s, followed by 27 2’s, followed by 39 3’s, followed by 16 4’s.

Therefore:

- The 18 responses of ‘1’ are assigned the rank \((1 + 18)/2 = 9.5\);
- The 27 responses of ‘2’ are assigned the rank \((19 + 45)/2 = 32\);
- The 39 responses of ‘3’ are assigned the rank \((46 + 84)/2 = 65\); and
- The 16 responses of ‘4’ are assigned the rank \((85 + 100)/2 = 92.5\).

(b) The answer is:

\[ r_1 = 7(9.5) + 11(32) + 22(65) + 10(92.5) = 2773.5. \]

(c) We have:

\[ r_1 + r_2 = 100(101)/2; \text{ or } 2773.5 + r_2 = 5050; \text{ or } r_2 = 2276.5. \]

(d) Treatment 1. The mean rank on treatment 1, \(2773.5/50 = 55.47\), is larger than the mean rank on treatment 2, \(2276.5/50 = 45.53\). Note: Because the study is balanced, we don’t need to compute mean ranks to determine the correct answer.

(e) The mean is:

\[ \mu = n_1(n + 1)/2 = 50(101)/2 = 2,525. \]

(f) First, I note that the \(t_i’s\) are: 18, 27, 39 and 16. Thus:

\[ \sum(t_i^3 - t_i) = (18^3 - 18) + (27^3 - 27) + (39^3 - 39) + (16^3 - 16) = 5,814 + 19,656 + 59,280 + 4,080 = 88,830. \]

The variance is:

\[ \sigma^2 = \frac{50(50)(101)}{12} - \frac{50(50)(88,830)}{12(100)(99)} = 21,041.67 - 1,869.47 = 19,172.20. \]

(g) i. You would place 2,525 in the *Mean* box.

ii. You would place \(\sqrt{19,172.20} = 138.46\) in the *SD* box.

iii. You would select *Above*.

iv. You would place 2773.0 in its box. (Remember the continuity correction.)
9. (a) The distances between successive $x$ values are: 2, 3 and 4. Thus, $\delta$, which equals the minimum of these, is 2.

(b) The probability histogram is below. Note that the centers of the four rectangles are equal to our $x$'s: 0, 2, 5 and 9. Also, the base of each rectangle extends from $x - 1$ to $x + 1$, giving a width of $\delta = 2$.

10. Even a cursory glance at the data reveals that $u = \bar{x} - \bar{y}$ is a positive number. Thus, 0.2571 is the approximate P-value for $>$ and 0.5085 is the approximate P-value for $\neq$.

11. Even a cursory glance at the data reveals that $\bar{x} = \bar{y}$ and, hence, that $u = 0$. Thus, this is a trick question! The vassarstats site breaks down when $u = 0$ and all we know are the following facts:

- The exact P-value for $\neq$ equals 1.
- The exact P-value for $>$ equals the exact P-value for $<$, because of symmetry implied by balance, and both of these P-values are larger than 0.5000.