Practice Midterm for Stat 371, Fall 2013

Below is your practice midterm. Please note the following:

- I will send you the solutions to this practice exam as soon as they are available.
- For the actual midterm you will be given a few pages of notes. I will send them to you as soon as they are available. In the meantime, assume you have access to any formulas you need for the sample questions below.
- This practice midterm has a total of 55 points; the actual midterm will have 50 points.
- In order to prepare for the midterm, I recommend that you: complete this practice midterm; and review the quizzes, homework and practice problems. I recommend paying primary attention to this practice midterm.

1. **(8 points.)** Consider the following hypothetical sampling distribution for the test statistic $U$.

<table>
<thead>
<tr>
<th>$u$ :</th>
<th>−10</th>
<th>−9</th>
<th>−8</th>
<th>−7</th>
<th>−6</th>
<th>−5</th>
<th>−4</th>
<th>−3</th>
<th>−2</th>
<th>−1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(U = u)$ :</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.08</td>
<td>0.10</td>
<td>0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$u$ :</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(U = u)$ :</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Use the sampling distribution above to answer the following questions.

(a) Find the P-value for the alternative $> u = 6$.

(b) Find the P-value for the alternative $\neq u = −5$.

(c) Find the critical region for the alternative $>\alpha = 0.07$.

(d) Find the critical region for the alternative $<\alpha = 0.04$.

2. **(4 points.)** An unbalanced CRD yields the following data, sorted by treatment:

| Treatment 1: | 14 | 16 | 16 | 30 |
| Treatment 2: | 12 | 13 | 14 | 16 | 28 | 29 |

(a) Calculate the observed values $r_1$ and $r_2$ or the random variables $R_1$ and $R_2$. Show your work.

(b) Which of the following statements is correct?

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

ii. Descriptively based on ranks, the data from treatment 1 are smaller than the data from treatment 1.

iii. Descriptively based on ranks, the data from treatment 1 are neither smaller nor larger than the data from treatment 2.
3. *(2 points.)* Refer to the data in the previous problem. Compute the standard deviation of the sampling distribution of \( R_1 \) in each of the following ways:

(a) Without adjusting for the fact that there are ties in the data; i.e., an incorrect way.
(b) Adjusting for the fact that there are ties in the data.

4. *(7 points.)* Below are 100 sorted integer responses. Note that I am not telling you the values of \( x_{(4)} \) and \( x_{(97)} \); I do this to discourage you from spending your time summing the 100 numbers. In order to answer the questions below, all you need to know about these numbers are the facts obvious from this listing; namely,

\[
542 \leq x_{(4)} \leq 548 \text{ and } 642 \leq x_{(97)} \leq 645.
\]

Call this data set A.

\[
530 \quad 536 \quad 542 \quad x_{(4)} \quad 548 \quad 559 \quad 563 \quad 564 \quad 564 \quad 567
\]
\[
574 \quad 574 \quad 576 \quad 576 \quad 577 \quad 578 \quad 579 \quad 579 \quad 579 \quad 580
\]
\[
580 \quad 581 \quad 582 \quad 582 \quad 584 \quad 585 \quad 585 \quad 586 \quad 587 \quad 588
\]
\[
589 \quad 589 \quad 590 \quad 591 \quad 591 \quad 592 \quad 592 \quad 593 \quad 593 \quad 594
\]
\[
595 \quad 595 \quad 596 \quad 598 \quad 598 \quad 598 \quad 599 \quad 599 \quad 600 \quad 600
\]
\[
601 \quad 602 \quad 602 \quad 602 \quad 602 \quad 605 \quad 605 \quad 605 \quad 605 \quad 606
\]
\[
607 \quad 607 \quad 607 \quad 607 \quad 608 \quad 609 \quad 609 \quad 609 \quad 610 \quad 611
\]
\[
614 \quad 615 \quad 616 \quad 616 \quad 617 \quad 618 \quad 618 \quad 618 \quad 618 \quad 619
\]
\[
619 \quad 619 \quad 622 \quad 622 \quad 623 \quad 625 \quad 629 \quad 629 \quad 630 \quad 631
\]
\[
632 \quad 635 \quad 637 \quad 639 \quad 639 \quad 642 \quad x_{(97)} \quad 645 \quad 656 \quad 670
\]

The mean of data set A equals 600.00 and the standard deviation equals 25.38.

(a) According to the Empirical Rule, about 68% of the data fall between \( c \) and \( d \). Determine the numerical values of \( c \) and \( d \).

(b) Refer to part (a) and the values of \( c \) and \( d \) that you determined. What percentage of the observations actually fall between the values of \( c \) and \( d \)?

(c) Refer to the observation equal to 563. Determine its deviation (from the mean).

(d) Calculate the median of data set A.

(e) Delete observations 530, 656 and 670 from data set A; call the resultant set of 97 numbers data set B. Calculate the mean of data set B. **Hint:** Do not try to obtain your answer by summing the 97 numbers!

(f) Refer to the part(e). Calculate the median of data set B.
5. (5 points.) A CRD with \( n_1 = 3 \) and \( n_2 = 1 \) yields the following four sorted observations: 0, 6, 12 and 12.

(a) Calculate the sampling distribution of the random variable \( U \) which, as always, has observed value \( u = \bar{x} - \bar{y} \).

(b) Calculate the sampling distribution of \( R_1 \), the test statistic for the sum of ranks test.

6. (2 points.) I used Fisher’s tests on several balanced CRDs. Recall that balance implies symmetry around zero for the test statistic \( X \); i.e.,

\[
\text{for any number } b > 0, P(X \geq b) = P(X \leq -b) \text{ and } P(X \geq 0) = P(X \leq 0).
\]

For each of the situations below, determine the missing P-value. If it is impossible to determine the missing P-value, write impossible.

(a) The P-value for \(< \) is 0.9159 and the P-value for \( \neq \) is 0.2596. Determine the P-value for \( > \).

(b) The P-value for \(< \) is 0.2422 and the P-value for \( > \) is 0.8244. Determine the P-value for \( \neq \).

7. (9 points) Below is the table I used in the homework to find the critical regions for \( U \) and \( R_1 \) for Dawn’s study of her cat.

<table>
<thead>
<tr>
<th>Skeptic Correct</th>
<th>( U )</th>
<th>( R_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail to reject ( H_0 )</td>
<td>9,470</td>
<td>7</td>
</tr>
<tr>
<td>Reject ( H_0 )</td>
<td>10</td>
<td>513</td>
</tr>
<tr>
<td>Total</td>
<td>9,480</td>
<td>520</td>
</tr>
</tbody>
</table>

\( m = 10,000 \)

(a) As stated in the Course Notes the approximate significance level for the critical region \((U \geq 1.80)\) is 0.0523. Calculate the nearly certain interval for the exact significance level of this critical region.

(b) As stated in the Course Notes the approximate significance level for the critical region \((R_1 \geq 127.0)\) is 0.0520. Calculate the nearly certain interval for the exact significance level of this critical region.

(c) Calculate the nearly certain interval for the exact significance level from part (a) minus the exact significance level from part (b). Comment on your answer.
8. **(10 points)** I performed a balanced CRD with an ordinal categorical response. My data are below.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Low (1)</th>
<th>Low Medium (2)</th>
<th>High Medium (3)</th>
<th>High (4)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>11</td>
<td>22</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>16</td>
<td>17</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>27</td>
<td>39</td>
<td>16</td>
<td>100</td>
</tr>
</tbody>
</table>

(a) Assign ranks to the four categories.

(b) Using your ranks from part (a), determine the observed value of $R_1$.

(c) Using the fact that the sum of the $n$ ranks always equals $n(n + 1)/2$, determine the observed value of $R_2$.

(d) Based on ranks, which treatment gives larger responses? Explain your answer.

(e) Calculate the mean of the sampling distribution of $R_1$.

(f) Calculate the variance of the sampling distribution of $R_1$.

(g) Explain how you would use the Normal curve website to obtain an approximate P-value for these data and the alternative $> \text{ for the sum of ranks test.}$ In particular, answer the questions below:

i. Sticking with the default option, *Area from a value*, what number—be specific—should you place in the *Mean* box?

ii. What number—be specific—should you place in the *SD* box?

iii. Which option—*Above* or *Below* should you select?

iv. For your choice of option in (iii), what number—be specific—should you place in its box in order to have the approximate P-value appear in Results?

9. **(3 points)** The sampling distribution of the random variable $X$ is given in the table below.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td>0.24</td>
</tr>
<tr>
<td>9</td>
<td>0.48</td>
</tr>
</tbody>
</table>

(a) What is the value of $\delta$?

(b) Draw the probability histogram of this distribution. Remember to label the endpoints and heights of each rectangle.
10. (3 points) I entered the following artificial data into vassarstats:

<table>
<thead>
<tr>
<th>Treatment 1:</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 2:</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

I used vassarstats to obtain a simulation experiment with 10,000 runs for the test statistic $U$. My results are below:

- one-tailed: 0.2571.
- two-tailed: 0.5085.

Match each approximate P-value to its alternative.

11. (2 points) I entered the following artificial data into vassarstats:

<table>
<thead>
<tr>
<th>Treatment 1:</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 2:</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

I used vassarstats to obtain a simulation experiment with 10,000 runs for the test statistic $U$. My results are below:

- one-tailed: 0.0447.

Match this approximate P-value to its alternative.