1. The plot below shows the value of the Mann-Whitney U statistic calculated from two samples, one of which is shifted relative to the other by various amounts. The .025 quantile for the Mann–Whitney statistic with sample 10 and 5 is 9.

   ![Plot showing the Mann-Whitney U statistic](image)

   a. Sketch the construction of the Hodges–Lehman estimator above. Use a dotted line.

   The Hodges-Lehmann estimator is where the above curve crosses the null expectation of the MWW statistic, which, since at least one of the group sizes is even, happens for a range of shift values. The Hodges–Lehmann estimator is taken to be the midpoint between the two vertical dotted lines.

   b. Sketch the 95% confidence interval for the difference in location between these samples on the graph above.

   Draw horizontal lines at 9, and 50+1-9=42. These lines lie along horizontal parts of the curve of the test statistic. End points of the CI are the most extreme parts here.

   Total for this question: 8.
2. Nine children collect candy on Halloween. Three children fell into one of three age groups, 5-8, 9-12, and 13-15. Additionally, they follow one of three collection strategies: They visit their friends houses, they visit rich people’s houses, and they visit houses with fake graveyards out front. Below is their candy yield, in pounds:

<table>
<thead>
<tr>
<th>Age</th>
<th>Friends</th>
<th>Rich</th>
<th>Fake graveyard</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-8</td>
<td>0.75</td>
<td>0.41</td>
<td>0.92</td>
</tr>
<tr>
<td>9-12</td>
<td>0.88</td>
<td>1.45</td>
<td>1.36</td>
</tr>
<tr>
<td>13-15</td>
<td>5.88</td>
<td>0.25</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Test the null hypothesis that the age groups are uniform in their candy yield, in a sense does not take ordering of age group into account, and which allows for different results based on collection strategy. Use a test that does not require assumptions on the response variable distribution, except for continuity. Here and below, test at a .05 level. A full answer includes a $p$-value, and a conclusion to reject or not reject the null hypothesis.

*Here are the ranks, within strategy:*

<table>
<thead>
<tr>
<th>Age</th>
<th>Friends</th>
<th>Rich</th>
<th>Fake graveyard</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-8</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9-12</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13-15</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Average per group are $5/3 = 1.66$, $8/3 = 2.66$, and $5/3 = 1.66$. Overall average ranks are all 2. Differences from averages are $-1/3$, $2/3$, $-1/3$. Sum of squared differences is $6/9 = 2/3$. Friedmantestis $T = 12L \sum_{k=1}^{K} (\bar{R}_{k.} - \frac{1}{2}(NK + 1))^2 / [K(KN + 1)]$ for $L = 3$, $K = 3$, $N = 1$. So $T = 12 \times 3 \times (2/3) / (3 \times 4) = 2$. Compare with $\chi^2_2$ table, you see $p > .1$. Do not reject the null hypothesis that distribution of candy yield is independent of age.

*Total for this question: 10.*

3. Nine children collecting candy on Halloween. Three children fell into one of three age groups, 5-8, 9-12, and 13-15. Below is their candy yield, in pounds:

<table>
<thead>
<tr>
<th>Age</th>
<th>Friends</th>
<th>Rich</th>
<th>Fake graveyard</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-8</td>
<td>0.75</td>
<td>0.41</td>
<td>0.92</td>
</tr>
<tr>
<td>9-12</td>
<td>0.88</td>
<td>1.45</td>
<td>1.36</td>
</tr>
<tr>
<td>13-15</td>
<td>5.88</td>
<td>0.25</td>
<td>0.36</td>
</tr>
</tbody>
</table>

a. Test the null hypothesis that the age groups are uniform in their candy yield, in a sense does not take ordering of age group into account. Use a test that does not require assumptions on the response variable distribution, except for continuity. Here and below, test at a .05 level. A full answer includes a $p$-value, and a conclusion to reject or not reject the null hypothesis.
Ranks are given below:

<table>
<thead>
<tr>
<th>5-8</th>
<th>4</th>
<th>3</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-12</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>13-15</td>
<td>9</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Test statistic is \( W = \frac{12}{[(N + 1)N]} \sum_{k=1}^{K} (R_k - n_k(N + 1)/2)/n_k \) for \( N = 9 \), \( K = 3 \), \( n_1 = n_2 = n_3 = 3 \). The sums of ranks are 13, 20, and 12, and so the difference from the averages are −2, 5, and −3. The sum of squared differences is 38, and the statistic is \( (12/(10 \times 9))(38/3) = 1.69 \). Compare to a \( \chi^2 \) distribution. \( p > .1 \). Do not reject the null hypothesis of equal yield in each age group.

b. Repeat this test in a way that reflects the expectation that yield should either increase or decrease with age.

Use the Jonckheere-Terpstra Test. Separate Mann-Whitney tests are: Middle vs. Young: 8, Old vs. Young 3, Old vs. Middle: 3. Test statistic is the sum of these, 14. Expectation is \( 3 \times 3 \times 3/2 = 13.5 \). Variance is \( \sum_{i=2}^{K} n_i m_{i-1} (m_i + 1)/12 \) for \( n_i = 3 \), \( K = 3 \), \( m_i = \sum_{j=1}^{i} n_j \), \( m_1 = 3 \), \( m_2 = 6 \), \( m_3 = 9 \). So variance is \( 3 \times 3 \times 7 + 3 \times 6 \times 10 = 63 + 180 = 263/12 \). Z score is \( .5/\sqrt{263/12} = .107 \). \( p \)-value is .914. Do not reject the null hypothesis that distribution of yield does not depend on age.

c. Consider doing a traditional parametric test of the hypothesis in part a. (Do not actually do this test.) Would you expect similar results? Why or why not?

No. The large value for the first yield for the older students is an outlier which would make a normal approximation inappropriate.

Total for this question: 20.

4. The following data reflect the effectiveness of a certain medication.

\[-1.1 \ -0.2 \ -0.1 \ 0.9 \ 1.0 \ 1.9 \]
Values of 0 are indicative of no effect. Positive and negative values are indicative of positive and negative effect respectively. These values have been rounded to one decimal place to ease your calculation.

a. Making no assumption about the distribution of these data, other than that they are independent and identically distributed draws from a continuous distribution, test the null hypothesis of zero population effect, vs. the alternative of either a positive or negative effect. Here and below, test at a .05 level. A full answer includes a \( p \)-value, and a conclusion to reject or not reject the null hypothesis.

Use a sign test. Use as the test statistic the number of positive values, 3, out of six trials. The \( p \)-value is 1. Do not reject the null hypothesis.

b. Suggest an assumption about the distribution of these data that would allow for more powerful inference.

Under the assumption of symmetry, the signed rank test is more powerful.

c. Under this stronger assumption, repeat the test of the same null and alternative hypotheses as above.

These data, ranked by absolute value, are

\[-0.1 \ -0.2 \ 0.9 \ 1.0 \ -1.1 \ 1.9\]

The signed rank statistic is the sum of the ranks of the positive entries, \( 3 + 4 + 6 = 13 \).

The expected value under the null hypothesis is \( (6 \times (6 + 1))/4 = 10.5 \), and the variance is \( 6 \times (2 \times 6 + 1) \times (6 + 1)/24 = 6 \times 13 \times 7/24 = 91/4 = 22.75 \). The \( z \)-score is \( (13-10.5)/\sqrt{22.75} = 0.524 \) (or, if I do a correction for continuity, \( (12.5-10.5)/\sqrt{22.75} = .4193 \)). The \( p \)-value is 0.600. Do not reject the null hypothesis.

d. Under this stronger assumption, estimate median of the populaton median.

Walsh averages are -1.10, -0.65, -0.60, -0.10, -0.05, 0.40, -0.20, -0.15, 0.35, 0.40, 0.45, 0.90, 0.90, 1.40, 1.00, 1.45, 1.90. Sorted, these are -1.10, -0.65, -0.60, -0.20, -0.15, -0.10, -0.05, 0.35, 0.40, 0.40, 0.45, 0.85, 0.90, 0.90, 1.00, 1.40, 1.45, 1.90. The median is the 11th of these, 0.40. The Hodges-Lehmann estimator of the median is 0.40.

Total for this question: 28.
5. Parents who buy candy for Halloween follow one of two strategies: Leaving their porch lights on, to attract trick or treaters, or leaving the light off, so they get all of the candy to themselves (at the risk of vandalism from disappointed children). Empirical cumulative distribution functions for the amount of weight gain by parents following each of these strategies are plotted below.

Researchers note that, although the empirical CDFs look remarkably different, the Wilcoxon two-sample test p-value is near 1.

a. Comment on why such obviously different distributions are not distinguished by the Wilcoxon test.

The Wilcoxon statistic is designed to distinguish distributions that differ by a shift. These distributions have different shapes and spreads, but not noticeably different locations, and so the Wilcoxon statistic cannot distinguish them.

b. Suggest a test that might better distinguish these groups.

One might use an Ansari-Bradley test to test for difference in spread, or a Kolmogorov Smirnov test to distinguish between shapes. Siegel-Tukey and Cramer-von Mises respectively are also acceptable.
6. Investigators are considering using either a two-sample T test, or a Wilcoxon rank sum test. The Wilcoxon test has the advantage of not requiring assumptions about the underlying distribution of the random observations, but has the disadvantage of having lower power for some distributions. Describe a criterion for measuring potential loss of power when choosing one statistical test instead of another.

One could calculate asymptotic relative efficiency, which is the limit of the inverse ratio of sample sizes needed to maintain the same power against the same alternative, as one of the sample sizes increases. In the case when the test statistic has an approximately normal distribution, with a mean a differentiable function \( \mu(\theta) \) of the parameter \( \theta \) under the alternative, and a variance of the form \( \sigma^2(\theta)/n \), this is the ratio of the efficacies, defined to be the \( \frac{\mu'(0)}{\sigma(0)} \).

Total for this question: 4. Total for exam: 80.