1. (a) What do we mean by power when we say a Generalized Likelihood Ratio Test has the most power of any valid test?

   Solution: When we talk about power in this course we are referring to the probability of rejecting the null hypothesis when the alternative hypothesis is true. Because all of our tests have only an probability of rejecting the null hypothesis when the null hypothesis is true, we want to simultaneously maximize the power which means rejecting a false null hypothesis as often as possible.

   (b) What is the advantage of a sign test over performing a paired t test?

   Solution: A sign test has the advantage that we need to make fewer assumptions. A t test requires the data to be at least approximately normal in order for the test to be valid, but a sign test does not need to have normal observations. The data can be skewed or discrete or contain outliers. A sign test can also be simpler to compute than a t test.

   (c) Why might you suggest that a researcher perform a Wilcoxon sign-rank test rather than a simple sign test?

   Solution: We might suggest that a researcher perform a Wilcoxon sign-rank test rather than a simple sign test because the ranks contain additional information about the difference between the two samples. The Wilcoxon sign-rank test has more power that the sign test. It is better able to detect a significant difference between the pairs. It will more often reject the null hypothesis when the alternative is true.

2. Suppose that we have 100 subjects in a medical experiment. The subjects all have their blood pressure taken before and after the study period. Let μ_B be the mean beforehand and μ_A be the mean after the study. We wish to test

$$H_0 : \mu_B = \mu_A \text{ versus } H_A : \mu_B \neq \mu_A$$

   We will assume that subjects are all independent, and the measurements have a normal distribution. Furthermore, assume that the variance of the difference between the before and after measurements is known to be $\sigma_d^2 = 125$.

   (a) If we propose a test statistic

$$Z = \frac{|\bar{x}_A - \bar{x}_B|}{\sigma_d/\sqrt{n}}$$

   and a critical value of 1.645, then what is the level of this test?

   Solution: If we propose a test statistic

$$Z = \frac{|\bar{x}_A - \bar{x}_B|}{\sigma_d/\sqrt{n}}$$

   we can use our central limit theorem to say that

$$\bar{x}_A - \bar{x}_B = \bar{d} \sim \mathcal{N}\left(\mu_A - \mu_B, \frac{\sigma_d^2}{n}\right)$$

Therefore, under the null hypothesis where the mean of the differences is 0,

$$P(Z > 1.645) = P\left(\frac{\bar{d}}{\sigma_d/\sqrt{n}} > 1.645\right) + P\left(\frac{\bar{d}}{\sigma_d/\sqrt{n}} < -1.645\right) = 0.05 + 0.05 = 0.1$$

This means that $\alpha = 0.1$ is the level of this test.

(b) What is the power of this test if the expected decrease in the blood pressure is actually 5.0mm?

Solution: The power of this test when the expected decrease in the blood pressure is actually 5.0mm is

$$\begin{aligned} &P\left(\frac{\bar{d}}{\sigma_d/\sqrt{n}} > 1.645\right) + P\left(\frac{\bar{d}}{\sigma_d/\sqrt{n}} < -1.645\right) = \\ &= P\left(\frac{\bar{d}+5}{\sigma_d/\sqrt{n}} > 1.645 + \frac{5}{\sqrt{125/100}}\right) + P\left(\frac{\bar{d}+5}{\sigma_d/\sqrt{n}} < -1.645 + \frac{5}{\sqrt{125/100}}\right) \\ &= P(Z > 6.117136) + P(Z < 2.827136) = 5 \times 10^{-10} + 0.9977 = 0.9977 \end{aligned}$$

(c) What is the power of this test if the expected value before the study was $\mu_B = 127mm$ while the expected value afterwards was $\mu_A = 124mm$

Solution: If the expected value before the study was $\mu_B = 127mm$ while the expected value afterwards was $\mu_A = 124mm$, then the expected value of the difference is going to be -3mm. The power is therefore going to be

$$\begin{aligned} &P\left(\frac{\bar{d}}{\sigma_d/\sqrt{n}} > 1.645\right) + P\left(\frac{\bar{d}}{\sigma_d/\sqrt{n}} < -1.645\right) = \\ &= P\left(\frac{\bar{d}+3}{\sigma_d/\sqrt{n}} > 1.645 + \frac{3}{\sqrt{125/100}}\right) + P\left(\frac{\bar{d}+3}{\sigma_d/\sqrt{n}} < -1.645 + \frac{3}{\sqrt{125/100}}\right) \\ &= P(Z > 4.328) + P(Z < 1.038) = 7.5 \times 10^{-6} + 0.8504 = 0.8504 \end{aligned}$$

3. A genomics experiment had two independent samples of bacteria culture. The sample A consists of 265 specimens with a genetic mutation, and the sample B consists of 543 specimens without the mutation. The mean levels of carbon in sample A was 222.965 grams and in sample B was 232.438. We also estimated the standard deviations in each sample as $s_A = 39.87$ and $s_B = 68.76$.

(a) Calculate the appropriate test statistic for the two-sided test. You can assume the variance for specimens from each group is the same.

Solution: The appropriate test statistic for the two-sided test, assuming the variance for specimens from each group is the same, is

$$t = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

where

$$\begin{aligned}
s_p^2 &= \frac{1}{265 + 543 - 2}(264 s_A^2 + 542 s_B^2) \\
&= \frac{264(1589.6169) + 542(4727.9376)}{806} = 3700.0013 \\
\bar{x} - \bar{y} &= 222/965 - 232.438 = -9.473 \\
\frac{1}{n} + \frac{1}{m} &= \frac{1}{265} + \frac{1}{543} = 0.005615
\end{aligned}$$

Thus,

$$t = \frac{-9.473}{\sqrt{3700.0013(0.005615)}} = -2.0783$$

(b) Is there a statistically significant difference between the grams in each sample? Assume an $\alpha = 0.05$ level.

Solution: There are 806 degrees of freedom in this t distribution so it is safe to say that it is nearly normal. We will use our two-sided critical value of 1.96 for an $\alpha = 0.05$ level test. Our test statistic is less than -1.96 so we reject the null hypothesis and conclude there is a statistically significant difference between the mean levels of carbon in the two samples.

4. Acme Tile Company wants to compare the performance between two kinds of acoustical tiles to see whether the different materials change the acoustic properties of rooms. They used experimental rooms where they could install Tile A, take a measurement, then install Tile B and measure again. Two reverberation times were recorded in each of the 8 rooms, once for each type of tile.

	Tile A	Tile B
Room 1	10	12
Room 2	9	10
Room 3	7	8
Room 4	15	18
Room 5	23	21
Room 6	11	15
Room 7	6	3
Room 8	17	17

(a) Perform a sign test to determine if there is a statistical difference between the two tiles at an $\alpha = 0.05$ level. Report your conclusion.

Solution:

Room	Difference	Sign
Room 1	-2	-
Room 2	-1	-
Room 3	-1	-
Room 4	-3	-
Room 5	2	+
Room 6	-4	-
Room 7	3	+
Room 8	0	

To perform a sign test, we calculate the differences. We have a tie in Room 8 so we will set this room aside in our analysis. There are 2 rooms out of 7 that have a positive sign. The exact P value for a binomial test where $H_0 : p = 0.5$ is

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{2^7} + \frac{7}{2^7} + \frac{21}{2^7} = 0.2266$$

$$\text{P-value} = 2P(X \leq 2) = \frac{58}{128} = 0.4531$$

This is not less than our level α so we would accept the null hypothesis that the two tiles have the same average reverberation times.

(b) Calculate the Wilcoxon sign-rank test statistic for the same test.

Solution: The Wilcoxon sign-rank test statistic depends on the ranks for the first 7 rooms (the room 8 is still a tie and so we leave it out)

Room	Difference	Rank
Room 1	-2	3.5
Room 2	-1	1.5
Room 3	-1	1.5
Room 4	-3	5.5
Room 5	2	3.5
Room 6	-4	7
Room 7	3	5.5

The ranks for ties are computed by averaging:

$$\frac{1+2}{2} = 1.5, \quad \frac{3+4}{2} = 3.5, \quad \frac{5+6}{2} = 5.5.$$

The sum of the ranks of the positive differences is $T^+ = 3.5 + 5.5 = 9$.

(c) Use a normal approximation to determine if the test statistic from part (b) is significant at an $\alpha = 0.05$ level. What would you conclude about the tiles?

Solution: The normal approximation may not be very good because $n < 25$, but we can still see what it would indicate.

$$E(T) = \frac{n(n+1)}{4} = 14, \qquad Var(T) = \frac{n(n+1)(2n+1)}{24} = \frac{14(15)}{6} = 35.$$

$$\begin{aligned}\text{P-value} &= 2P(T \leq 9) = 2P(T \leq 9.5) = 2P\left(Z \leq \frac{9.5 - 14}{\sqrt{35}}\right) \\ &= 2P(Z \leq -0.7606) \approx 2 \cdot 0.2234 = 0.4468\end{aligned}$$

Note: we need to use the continuity correction: we want to make sure that 9 is included So we accept the null hypothesis at an $\alpha = 0.05$ level. Again, we conclude that the two tiles have average reverberation times that are not statistically significantly different.

(d) Calculate the exact probability that the sign-rank test statistic would be $T \leq 1$ conditional on the ranks in this experiment.

Solution: To calculate the exact probability that the sign-rank test statistic would be $T \leq 1$ conditional on the ranks in this experiment, we consider the rankings

$$1.5, 1.5, 3.5, 3.5, 5.5, 5.5, 7.$$

Given that we had this pattern of tie scores, we would see that the only way that $T \leq 1$ is if $T = 0$ and all of the differences are negative.

$$P(T \leq 1) = P(T = 0) = \frac{1}{2^7} = 0.007813$$

5. For a comparison of the academic effectiveness of two junior high schools A and B, an experiment was designed using ten sets of identical twins, each twin having just completed the sixth grade. In each case, the twins in the same set had obtained their previous schooling in the same classrooms at each grade level. One child was selected at random from each set and assigned to school A. The other was sent to school B. Near the end of the ninth grade, an achievement test was given to each child in the experiment. The results are shown in the accompanying table.

Twin Pair	1	2	3	4	5	6	7	8	9	10
A	67	80	65	70	86	50	63	81	86	60
B	39	75	69	55	74	52	56	72	89	47

(a) Using the sign test, test the hypothesis that the two schools are the same in academic effectiveness, as measured by scores on the achievement test, against the alternative that the schools are not equally effective. Give the attained significance level. What would you conclude with $\alpha = 0.05$?

Solution: There are 7 out of 10 of the twins where student A did better than student

B. The probability of this occurring is

$$P(M \geq 7) = 2^{-10}\left(\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right) = \frac{176}{1024}$$

so that the P value is

$$2\frac{176}{1024} = 0.34375.$$

This is not significant at the $\alpha = 0.05$ because the P value is bigger than α.

(b) Suppose it is suspected that junior high school A has a superior faculty and better learning facilities. Test the hypothesis of equal academic effectiveness against the alternative that school A is superior. What is the p-value associated with this test?

Solution: For the one-sided alternative, the P -value is $\frac{176}{1024} = 0.171875$.

(c) What answers are obtained if Wilcoxons signed-rank test is used in analyzing the data? Compare these answers with the answers obtained above.

Solution: The ranks of the differences are

Pair	Absolute Difference	Rank	Sign
1	28	10	+
2	5	4	+
3	4	3	-
4	15	9	+
5	8	6	+
6	2	1	-
7	7	5	+
8	9	7	+
9	3	2	-
10	13	8	+

The sum of the ranks of the negative observations is $T^- = 3 + 1 + 2 = 6$. The critical value for $\alpha = 0.05$ from Table 9 is 8 for a two-sided test and 11 for a one-sided test. Thus, we get a significant difference between the two samples in either the one-sided or two-side tests.

Therefore, the test is more significant than the sign test. This is often true because we are able to use more information in the data (from the ranks) to get a more powerful test.

6. 239 subjects had their cholesterol measured, and then were put on high-fiber diets. After a month on the high-fiber diet, the cholesterol was measured again. The mean LDL cholesterol level before the experiment was 118.5 mg/DL. The mean level after the subjects were on a diet for a month was 113.7 mg/DL. The sample standard deviation of the amount of cholesterol each patient lost was 38.37 mg/DL.

   (a) Test whether or not the high-fiber data significantly reduced the LDL cholesterol level at an $\alpha = 0.05$ level.

Solution: $n = 239$, $\bar{x} = 118.5$ and $\bar{y} = 113.7$ with $s_d = 38.37$. The test statistic is

$$Z = \frac{118.5 - 113.7}{38.37/\sqrt{239}} = 1.934$$

This is greater than the critical value 1.645 and therefore it is significant.

(b) If the true reduction in cholesterol was 10 mg/DL, what is the power of this test?

Solution: The power of the test is

$$\begin{aligned} P(Z > 1.645) &= P\left(\frac{\bar{x} - \bar{y}}{38.37/\sqrt{239}} > 1.645\right) \\ &= P\left(\frac{\bar{x} - \bar{y} - 10}{38.37/\sqrt{239}} > 1.645 - \frac{10}{38.37/\sqrt{239}}\right) \\ &= P(Z > 1.645 - 4.049) \\ &= P(Z > -2.38) \\ &= 1 - 0.0087 \\ &= 0.9913 \end{aligned}$$

7. We have the outcomes from 437 experimental plots comparing two varieties of soy beans. We want to test whether there is difference between variety A and variety B in the amount of lysine in the beans produce. Each plot is planted with both varieties so that we can compare the yields under similar growing conditions.

   (a) In 186 of the 437 plots, variety A produced a greater yield of lysine than variety B. Use a normal approximation to calculate a P value for testing whether the yields were the same.

Solution: In 186 of the 437 plots, variety A produced a greater yield of lysine than variety B. Using continuity correction

$$\text{P-value} = P(M \leq 186) = P(M \leq 186.5)$$

The normal approximation takes

$$z = \frac{186.5 - 437/2}{\sqrt{437/4}} = -3.06$$

Thus, the probability of being smaller than this is

$$P(Z \leq -3.06) = 0.0011$$

so that we get a P value of 0.0022.

   (b) What assumptions do we need to make for the test in part (a) to be valid?

> **Solution:** We only need to assume that yield in the separate plots are independent and that all have the pairs all have the same distribution so that the probability that variety A has a higher yield is the same for each plot.

(c) Suppose we want to perform a Wilcoxon Signed-Rank test and calculated $T^- = 38,025$. Calculate a P-value using the appropriate normal approximation.

> **Solution:** Suppose we want to perform a Wilcoxon Signed-Rank test and calculated $T^- = 38,025$. The expected value of T^- is
>
> $$n(n+1)/4 = 437(438)/4 = 47,851.5,$$
>
> and the variance of the test statistic is
>
> $$n(n+1)(2n+1)/24 = (437)438(875)/24 = 6,978,343.75.$$
>
> Thus, using continuity correction
>
> $$Z = \frac{38,025.5 - 47,851.5}{\sqrt{6,978,343.75}} = -3.72$$
>
> which yields a P value
>
> $$2P(Z \leq -3.72) = 2(0.0001) = 0.0002.$$

8. A number of fish were collected from two lakes and the length of each fish was recorded. The X column refers to fish from Xavier Lake, and the Y column refers to Yearling Lake.

#	X	Y
1	8.5	7.3
2	9.2	7.6
3	10.4	10.0
4	9.1	7.4
5	8.6	8.7
6	9.4	

The lengths were measured in inches, and the first column refers to the order in which the fish were caught. We want to test whether the fish in the two lakes are the same size or not.

(a) Calculate the appropriate rank-sum statistic W .

> **Solution:**

#	X	Rank	Y	Rank
1	8.5	4	7.3	1
2	9.2	8	7.6	3
3	10.4	11	10.0	10
4	9.1	7	7.4	2
5	8.6	5	8.7	6
6	9.4	9		

The appropriate rank-sum statistic $W = 22$.

(b) Find a P-value for the MannWhitney test that the distributions of the fish sizes are the same in the two lakes.

Solution: The U statistic is $U = 22 - 5(6)/2 = 22 - 15 = 7$. This is less than the expected value of 15. From the table 8 on pages 862867 of the textbook the $P(U \leq 7) = 0.0887$ which makes the P-value 0.1774.

(c) What is the advantage of this test over the two-sample t test?

Solution: The advantage of this nonparametric test over the two-sample t test is that we do not need to assume that the data is normally distributed. It is likely that size of each fish drawn from the lake has a distribution that is not exactly normal, and this test is still appropriate under those conditions. Unfortunately, the nonparametric test has somewhat less power.

9. Suppose that we observe a single Poisson distributed random variable with pmf

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0.1.2\ldots$$

and we wish to test

$$H_0 : \lambda = 10 \text{ vs. } H_a : \lambda > 10.$$

(a) Calculate the appropriate likelihood ratio for the Likelihood Ratio Test. (Hint: you can use the fact that the MLE is $\hat{\lambda} = X$.)

Solution: Note to avoid confusion between two different λ's: I will use λ as a parameter of Poisson distribution, and Λ as a ratio of maximized likelihoods - an LRT test statistic.
The likelihood is

$$L(\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

Then LRT test statistic is

$$\Lambda = \frac{\max_{\lambda\in\Theta_0} L(\lambda)}{\max_{\lambda\in\Theta} L(\lambda)},$$

where $\Theta_0 = 10$ and $\Theta = \{\lambda : \lambda \geq 10\}$. Under H_0 and H_A the value of λ, which maximizes the likelihood is $\hat{\lambda} = \max\{x, 10\}$. Thus, if $x < 10$, then $\Lambda = 1$. If $x \geq 10$,

$$\Lambda = \frac{e^{-10}\frac{10^x}{x!}}{e^{-x}\frac{x^x}{x!}} = e^{x-10}\left(\frac{10}{x}\right)^x.$$

(b) Calculate the size of the test (α) that has critical region $\{X \leq 3\}$.

Solution: The size of this test is the probability of the set under the null hypothesis.

$$\begin{aligned} P(X \leq 3|\lambda = 10) &= \sum_{k=0}^{3} P(X = k|\lambda = 10) \\ &= e^{-10} + 10e^{-10} + e^{-10}\frac{100}{2} + e^{-10}\frac{1000}{6} \\ &= 0.01034 \end{aligned}$$

(c) Calculate the power of the test that has critical region $\{X \leq 3\}$ when $\lambda = 5$.

Solution: The power of the test when $\lambda = 5$ is

$$\begin{aligned} P(X \leq 3|\lambda = 5) &= \sum_{k=0}^{3} P(X = k|\lambda = 5) \\ &= e^{-5} + 5e^{-5} + e^{-5}\frac{25}{2} + e^{-5}\frac{125}{6} \\ &= 0.26503 \end{aligned}$$

(d) Is the test with critical region $\{X \leq 3\}$ a LRT for the above hypotheses? Justify your answer.

Solution: This is not the LRT for this test. The Likelihood ratio is (for $x > 10$)

$$\Lambda = e^{x-10}\left(\frac{10}{x}\right)^{x} = \exp\{x\ln 10 - x\ln x + x - 10\}.$$

The derivative of this function with respect to X is

$$\frac{d}{dx}\Lambda = (\ln 10 - \ln x)\exp\{x\ln 10 - x\ln x + x - 10\}.,$$

which is negative for any x¿10. Therefore Λ is a decreasing function in x, which implies that the LRT is of the form $\{X \geq k\}$ for some constant k.
Our set $\{X \leq 3\}$ is not of this form and therefore cannot be a LRT. In fact, for $x \leq 3$ the ratio $\Lambda = 1$, and $\Lambda < 1$ for $x > 10$(eg. $\Lambda(11) = 0.952741$). Therefore there is no k such that the set $\{\Lambda < k\} = \{X \leq 3\}$.

10. A survey asked people to rate their opinion on the economy in January 2016 and January 2017 on a scale from 1 to 10 where 10 meant they were extremely optimistic, 1 meant they were extremely pessimistic, and 5 indicated that they were neutral. Here are the results

Respondent	2016	2017
A	6	6
B	3	8
C	4	9
D	7	6
E	8	5
F	6	9
G	1	10
H	2	7
I	2	8

We wish to test $H_0 : \mu_{2016} = \mu_{2017}$ versus $H_A : \mu_{2016} < \mu_{2017}$.

(a) Calculate the signed-rank test statistic T .

Solution:

Respondent	2016	2017	Difference	Rank
A	6	6	0	
B	3	8	-5	5
C	4	9	-5	5
D	7	6	1	1
E	8	5	3	2.5
F	6	9	-3	2.5
G	1	10	-9	8
H	2	7	-5	5
I	2	8	-6	7

The signed-rank test statistic $T = T^+$ is

$$T^+ = 1 + 2.5 = 3.5$$

(b) For this set of ranks, calculate the probability under the null hypothesis that $T \leq 3$.

Solution: For this set of ranks, there are four ways that $T^+ \leq 3$

T /Rank	1	2.5	2.5	5	5	5	7	8
0	-	-	-	-	-	-	-	-
1	+	-	-	-	-	-	-	-
2.5	-	+	-	-	-	-	-	-
2.5	-	-	+	-	-	-	-	-

Each of these outcomes has a 2^{-8} under the null hypothesis. Therefore,

$$P(T^+ \leq 3) = 4(2^{-8}) = \frac{1}{64} = 0.015625.$$

The level of a two sided test that rejected when $T \leq 3$ would therefore be

$$\alpha = 2P(T^+ \leq 3) = \frac{1}{32} = 0.03125$$

because of the symmetry in the distribution.

11. There are two sections of 120B that were given the same homework assignments. A simple random sample of homework scores from the two sections are in the table:

Section A	Section B
50	50
50	47
48	43
	42

I want a nonparametric test to see if there is a significant difference between the two sections.

(a) Calculate the appropriate test statistic U.

Solution: The rank sum for Group A$= 6 + 6 + 4 = 16$ and the possible minimum $\frac{3\times 4}{2} = 6$.

$$U^+ = 16 - 6 = 10$$

$$U^- = 12 - 10 = 2$$

Therefore $U = \min\{U^+, U^-\} = 2$.

(b) Is the difference significant at the $\alpha = 0.1$ level? (You can use the table.)

Solution: According to the table, $P(U \leq 2) = 0.1143$ which means that the P -value is 0.228, and the test is not significant at the $\alpha = 0.1$ level.

(c) Calculate the exact $P(U \leq 1)$ for two samples drawn from these values. (Dont use the table in the textbook.)

Solution: There are $\binom{7}{3} = 35$ ways that these ranks can be distributed among the two samples. There are two ways that $U = 0$ is possible: either Sample A has

$$\{6,6,6\} \quad (W = 18,\ U^+ = 18 - 6 = 12,\ U^- = 12 - 12 = 0,\ U = \min\{0,12\} = 0)$$

or

$$\{1,2,3\} \quad (W = 6,\ U^+ = 6 - 6 = 0,\ U^- = 12 - 0 = 12,\ U = \min\{0,12\} = 0).$$

There is only one way (because of the ties) to get $U = 1$: Sample A has

$$\{1,2,4\} \quad (W = 7,\ U^+ = 7 - 6 = 1,\ U^- = 12 - 1 = 11,\ U = \min\{1,11\} = 1).$$

Therefore,

$$P(U \leq 1) = \frac{3}{35} = 0.0857.$$