Homework 4

1.Go to the website and see how many tumors showed deletions in these regions for each type of cancer. Then, use the chi-square test to evaluate the null ($\alpha = 0.05$) that the population proportions of deletions of the genome region are the same for each of these types of cancers. What do you conclude?

Solution. From the data set on the web, we know 12 tumors out of 66 showed deletions for RCC and 23 out of 95 tumors showed deletions for bladder carcinoma. We can build the following 2 by 2 table:

	RCC	Bladder carcinoma	Total
Deletions	12	23	35
No Deletions	54	72	126
Total	66	95	161

$$\begin{aligned} \chi^2 &= \frac{(12 - 66 \cdot 35/161)^2}{66 \cdot 35/161} + \cdots + \frac{(72 - 95 \cdot 126/161)^2}{95 \cdot 126/161} \\ &= 0.832 \end{aligned}$$

The degree of freedom for our test statistic is (2-1)(2-1)=1. From the chi-square table, we know ${\chi^2}_{1-0.05}$=3.84. Since $0.832 < 3.84$, so we don't reject the null hypothesis that the population proportions of deletions of the genome region are the same for each of these types of cancers at the significance level 0.05.

2. Construct a 95% confidence interval for the true underlying proportion of CWD in Wisconsin (we've been calling this p) using the fact that 15 cases were found in 500 deer. What do you conclude about CWD?

Solution.

$$\begin{aligned} \hat{p} &= 15/500 = 0.03 \\ z_{1-\alpha/2} &= z_{0.975} = 1.96 \\ 1.96\sqrt{\hat{p}\hat{q}/500} &= 0.015 \end{aligned}$$

So the 95% confidence interval is [0.03-0.015,0.03+0.015]=[0.015,0.045]. Comparing this with the proportion in 2001, which is 0.0087, we conclude that the CWD is spreading over these years from 2001.

9.12 To compare the white-blood counts of patients on teh medical and surgical services in Table 2.11 when normality is not assumed, what test can be used? Perform the test in Problem 9.11, and report a p-value.

Solution. We can use the Wilcoxon Rank Sum Test here, because the two groups surgical and medical group have different sample sizes. The following table can help us to find the test statistic:

WBC	Service	Rank
8	1	15.5
5	1	5.5
12	2	22.5
4	2	2.5
11	2	20.0
6	2	9.5
8	1	15.5
7	1	13.0
7	1	13.0
12	2	22.5
7	1	13.0
3	2	1.0
11	2	20.0
14	2	24.5
11	2	20.0
9	2	17.0
6	2	9.5
6	2	9.5
5	1	5.5
6	2	9.5
10	2	18.0
14	2	24.0
4	1	2.5
5	2	5.5
5	1	5.5

From the above table, we know the sum of ranks for the medical group is

89. Then I go back to Table 12 look at $n_1 = 9$, $n_2 = 16$ with different α. We can start from $\alpha = .10$, then lower limit is 87, upper limit is 147. So we don't reject. Then move right to $\alpha = 0.05$, lower limit is 82 and upper limit is 152. So we don't reject. Actually if we don't reject at the beginning we won't reject at the end, because the accetance region is bigger and bigger. Then our conclusion is p-value> 0.10.

9.45 What nonparametric test can be used to compare the median pod weight for inoculated versus noninoculated plants(Table 8.35)? Implement the test and report a two-tailed p-value.

Solution. Since the two groups are independent, we can use the Wilcoxon Rank Sum Test. The following table can help us to get the test statistic:

pod weight(I)	1.76,	1.45,	1.03,	1.53,	2.34,	1.96,	1.79,	1.21
rank	12,	9,	7,	10,	16,	14,	13,	8

pod weight(U)	0.49,	0.85,	1.00,	1.54,	1.01,	0.75,	2.11,	0.92
rank	1,	3,	5,	11,	6,	2,	15,	4

The sum of ranks for the inoculated group is 89. Using $n_1 = 8$, $n_2 = 8$, from table 12 we know if we choose $\alpha = 0.05$, then we reject H_0. If we choose $\alpha = 0.02$, then we don't reject. So the p-value is between 0.02 and 0.05.

10.99 What test can be used to compare the percentage of obese and normal individuals with at least a high-school education?

Solution. We can use chi-square test.

10.100 Implement the test in Problem 10.99, and report a two-tailed p-value.

Solution. We can build the following table:

	Normal	Obese	Total
High-school education	70.07	53.12	123.19
No High-school education	6.93	10.88	17.81
Total	77	64	141

So the chi-square test statistic is 2.027. With degree of freedom 1, we know the p-value is between 0.10 and 0.25. (the method to get the range is similar to 9.45)

10.101 What test can be used to compare the percentage of individuals with at least a high-school education between the three BMI groups in Table 10.40?

Solution. We can still use the chi-square test here.

10.102 Implement the test in Problem 10.101, and report a two-tailed p-value.

Solution. We can get the following table:

	Normal	overweight	Obese	Total
High-school education	70.07	105	53.12	228.19
No High-school education	6.93	15	10.88	32.81
Total	77	120	64	261.00

The chi-square test statistic is 2.036 with degree of freedom 2. From Table 6 and d=2, we know the p-value is between 0.25 and 0.5.