Problem 2: (Students)

Introduction:

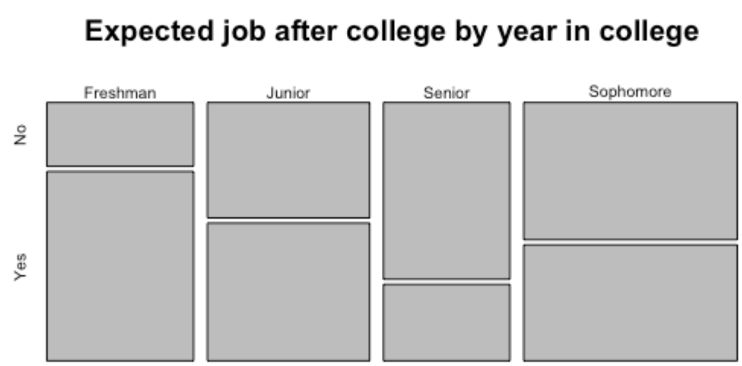
In this problem we are using the student dataset, and we will explore the relationship between students' years in college and if they believe it possible have a job after graduating college.

Our analysis is based on the following steps. Firstly, we calculate the sample estimate probability of undergraduate who believe they could have a job after graduation, for each grade. This will be done by taking the amount of students per year in school who said they will have a job after college and dividing that by the total amount of students in that year sampled. After this we will find if there is an overall dependence between year in college and their belief of having a job after college by using a *Pearson's chi squared test.*

Summary:

Below is the contingency table for each year

	Freshman	Sophomore	Junior	Senior	Totals
Yes	92	82	74	32	280
No	31	97	62	74	264
Totals	123	179	136	106	544



Above is the mosaic table of each year in college versus whether or not they believe they will have a job upon graduation. From the plot we can see that freshman are the most optimistic about having a job coming out of college, and that seniors are the least optimistic. Sophomores and Juniors seem to be very similar in how much they believe they will have a job upon graduation.

Below are the sample estimates for each year:
For Freshman the number that said yes is 92, and 31 said no. The sample estimate is therefore 92/(92+31) = .7480.
For Sophomores the number that said yes is 82, and 97 said no. The sample estimate is therefore 82/(82+97) = .4581
For Juniors the number that said yes is 74, and 62 said no. The sample estimate is therefore 74/(74+62) = .5441
For Seniors the number that said yes is 32, and 74 said no. The sample estimate is therefore 32/(32+74) = .3018

Analysis:

For this part, we will be comparing all of the years to see whether or not there is dependence, between students' year in college and whether or not they believe having a job after college. We will calculate the **Pearson's Chi squared statistic** to test for dependence between the years.

H(Null): Year in college is independent of whether they believe they will have a job after college or not
H(alternative): Year in college is dependent of whether they believe they will have a job after college

We will evaluate this with an alpha of .05.
We used R(***See Appendix for detailed code***) to find the Pearson's Chi squared value and here were the results:

X-squared = 48.781, df = 3, p-value = 1.452e-10

With these results we can **reject the null hypothesis** due to the p value being less than .05, and say that there is definitely dependence between the year in college and whether or not they believe they will have a job after college.

Now we want to test for independence between each of the years and we will do that by creating confidence intervals for the relative risk between each of the years. We will be making 95% confidence intervals, and since we will be making 6 confidence intervals we will need to adjust our alpha with relation to that, and for each confidence interval we will **use.05/6 instead of .05**, so our corresponding ***Z-value will be equal to 2.635.***

First we will find the CI between Freshman($\pi1$) and sophomores($\pi2$).
The relative risk is .7480/.4851 = 1.542
The confidence interval is:
$Ln(1.542) \pm Z(1-\alpha/2)\{[(1-\pi1)/y1 + (1-\pi2)/y2]\}^{(1/2)}$
$=0.433 \pm 2.635(.09497) = 0.433 \pm .2502$
$=(.1828,.6832)$
After exponentiating we get (1.201,1.980)

Next we will find the CI between Freshman($\pi1$) and Juniors($\pi3$).
The relative risk is .7480/.5441 = 1.3747
The confidence interval is:
$Ln(1.3747) \pm Z(1-\alpha/2)\{[(1-\pi1)/y1 + (1-\pi3)/y3]\}^{(1/2)}$
$=0.3182 \pm 2.635(.09456) = 0.3182 \pm .2492$
$=(.069,.5674)$
After exponentiating we get (1.071, 1.764)

The next CI we will find is between Freshman($\pi1$) and Seniors($\pi4$).
The relative risk is .7480/.3012 = 2.4834
The confidence interval is:
$Ln(2.4834) \pm Z(1-\alpha/2)\{[(1-\pi1)/y1 + (1-\pi4)/y4]\}^{(1/2)}$
$=0.9096 \pm 2.635(.1568) = 0.9096 \pm .413$
$=(.4966,1.3226)$
After exponentiating we get (1.643, 3.753)

The next CI we will find is between Sophomores($\pi2$) and Juniors($\pi3$).
The relative risk is .4851/.5441 = .8916
The confidence interval is:
$Ln(.8916) \pm Z(1-\alpha/2)\{[(1-\pi2)/y2 + (1-\pi3)/y3]\}^{(1/2)}$
$= -.1147 \pm 2.635(.1115) = -.1147 \pm .2938$
$=(-.4085,.1791)$
After exponentiating we get (.6646, 1.196)

The next CI we will find is between Sophomores($\pi2$) and Seniors($\pi4$).
The relative risk is .4851/.3012 = 1.6106
The confidence interval is:
$Ln(1.6106) \pm Z(1-\alpha/2)\{[(1-\pi2)/y2 + (1-\pi4)/y4]\}^{(1/2)}$

$= .4766 \pm 2.635(.1677) = .4766 \pm .4419$
$=(.0347,.9185)$
After exponentiating we get (1.035, 2.506)

Our last CI is between Junior($\pi 3$) and Seniors($\pi 4$).
The relative risk is .5441/.3012 = 1.8064
The confidence interval is:
$Ln(1.8064) \pm Z(1-\alpha/2)\{[(1-\pi 3)/y3 + (1-\pi 4)/y4]\}^{(1/2)}$
$= .5913 \pm 2.635(.1673) = .5913 \pm .4408$
$=(.1505,1.0321)$
After exponentiating we get (1.162, 2.807)
We created a function (***See Appendix for code***)to calculate the confidence interval in r to check our manual calculations.

Interpretation:

From the Pearson's Chi Squared test we were able to see with 95% confidence that there was **overall dependence** between the year in college and whether or not they think they will have a job coming out of college. The Chi Squared test tests for overall significance of between the variables, but to test whether or not between two individual years in college we **used confidence intervals** for the relative risk between each combination of two years.

For the 95% confidence intervals we found that for all relative risks involving freshman the probability that they believe they will have a job coming out of college is higher than for Sophomore, Junior, or Seniors. We see this because in each confidence interval, (1.201,1.980) for Freshman vs Sophomore, (1.071, 1.764) for Freshmen vs Juniors, (1.643, 3.753) for Freshmen vs Seniors, the interval is greater than 1, meaning that the **probability freshmen believe they will get a job is higher than the students from the other three grades**.

Next we will analyze the 95% confidence intervals involving Sophomores. We have already analyzed the relationship between freshmen and sophomores so we will go to junior first. The confidence interval between sophomores and juniors is (.6646, 1.196). This interval includes 1, which means that at a 5% significance level we can say that whether or not you believe you will have a job upon graduation is **independent whether you are a sophomore or a junior**. This means that being a sophomore or a junior does not affect one's probability of believing they will have a job upon graduation. The interval between sophomores and seniors is (1.035, 2.506), showing that we are 95% confident that **sophomores are more likely to believe they will have a job upon graduation than seniors** since the interval is greater than one.

Lastly, we analyze our final confidence interval which is between juniors and seniors. The 95% confidence interval is (1.162, 2.807), telling us that the belief of having a job upon graduation is **dependent on which year, junior or senior**, you are. Furthermore, juniors are more likely to have this belief than seniors since the interval is greater than 1.

Conclusion:

Overall we can see that the belief that one will have a job upon graduation is **<u>mostly dependent</u>** on which year in school that person is. Comparing all columns at the same time and all pairs of columns, except for between sophomores and juniors, we saw that whether one believes they will have a job upon graduation is dependent on year. We also found that the <u>one exception was between juniors and sophomores</u>, and that these two years are independent. We can also say that freshmen believe they will have a job at the highest rate, due to all their confidence intervals being greater than one, and that juniors believe they will have a job with the second highest probability, and then sophomores, and lastly seniors.