Patrick Fitzgerald
ADS 534 Statistical Modeling
HW 6

1. The following table contains data from a study concerning the distribution of party affiliation in a neighborhood. The interest was whether there was an association between registered political party and the race.

Race					
Party	White	Black	Hispanic	Asian	**Total**
Democrat	431	190	180	95	896
Republican	390	248	159	106	903
Independent	406	177	164	82	829
Total	1227	615	503	283	2628

To determine whether there was an association between registered political party and the race, we would like to perform the Chi-square test.

. (a) We first calculate the table of expected counts. Table below lists the expected counts calculated in SAS. The table is incomplete as results for four cells in the lower-right corner of the table are removed and replaced by a, b, c, and d respectively. Complete the table by calculating the values of a, b, c, d. Remember to show your calculation process.

Race					
Party	White	Black	Hispanic	Asian	**Total**
Democrat	a	b	171.71	96.19	896
Republican	c	d	172.96	96.89	902
Independent	386.75	193.81	158.73	88.92	828
Total	1227	615	503	282	2627

```
> #### (row i total x column j total)/(total count in table)

> a <- 896*1227/2628

> a

[1] 418.3379

> b <- 896*615/2628

> b

[1] 209.6804

> c <- 903*1227/2628

> c

[1] 421.6062

> d <- 903*615/2628

> d

[1] 211.3185
```

. (b) What are the null hypothesis and alternative hypothesis of this test?

H_0: there is no association between political party and race vs H_a: there is an association between political party and race

. (c) Performing data analysis in SAS, we got the chi-square test statistic χ^2 = 16.50 (with p-value 0.011). What are the degrees of freedom for this χ^2 test statistic?

df = (r-1)(c-1) = (3-1)(4-1) = 2*3 = 6, where r = number of rows and c = the number of columns

2. A survey is conducted to see whether people have a certain symptom X. The following table breaks down these results by gender:

Symptom X by Gender

Having symptom X	Male	Female
Yes	9	36
No	39	63

. (a) Under the hypothesis of independence between row and column variables, calculate the expected cell count for the two cells in the column for women.

```
> FXyes <- ((36+9)*(36+63))/(9+36+39+63)

> FXyes

[1] 30.30612

> FXno <- ((63+39)*(36+63))/(9+36+39+63)

> FXno

[1] 68.69388
```

. (b) What is the chi-square χ^2 statistic for this contingency table? Is it significant at the $\alpha = 0.05$ level? Justify your answer. (Use Chi-square distribution table in textbook)

```
> symptom

   Male Female

Yes    9     36

No    39     63

> chisq.test(symptom)
```

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  symptom

X-squared = 3.9287, df = 1, p-value = 0.04747
```

We see our p-value (0.04747) is $< \alpha$ (0.05) and therefore is significant at the 0.05 level.

. (c) Calculate the relative odds (i.e., the OR) of having symptom X, for men to women.

```
> #### calculate odds for males and females to have symptom X

> malesone <- 9/48

> malestwo <- 1-malesone

> males <- malesone/malestwo

> males

[1] 0.2307692

> femalesone <- 36/99

> femalestwo <- 1-femalesone

> females <- femalesone/femalestwo
```

```
> females

[1] 0.5714286

> or <- males/females

> or

[1] 0.4038462
```

. (d) Calculate 95% confidence interval for the relative odds (i.e., the OR) in part (c).

```
> #### calculate 95% CI for the LOG of OR
> log_estimated_odds <- log(or)
> standard_error_log_of_or <- sqrt((1/9)+(1/36)+(1/39)+(1/63))
> ci_log_low <- log_estimated_odds-1.96*standard_error_log_of_or
> ci_log_high <- log_estimated_odds+1.96*standard_error_log_of_or
> paste(ci_log_low, ci_log_high)
[1] "-1.73920905700604 -0.0742335047099716"
>
> #### calculate 95% CI for OR
> ci_low <- exp(ci_log_low)
> ci_high <- exp(ci_log_high)
> paste(ci_low, ci_high)
[1] "0.175659282164616 0.928454869942444"
```

Thus, we can see that a 95% confidence interval for our relative odds is (0.176, 0.928)