Homework Set #3 Problem 2

Costa Huang

March 22, 2018

0.1 Input the data

First, let's input the data into a observed table:

```
In [12]: import pandas as pd

         obs = pd.DataFrame(
             index=[0, 1, 2],
             data={
                 0: [42, 64, 20],
                 1: [309, 246, 115],
                 2: [31, 46, 27]
             }
         )
         obs

         # Row 0 = Democrat
         # Row 1 = Republican
         # Row 3 = Other

         # Col 1 = We Know All Relevant Facts
         # Col 2 = Some Relevant Facts Withheld
         # Col 3 = Not sure
```

Out[12]:

	0	1	2
0	42	309	31
1	64	246	46
2	20	115	27

Let X denote the partisanship of a given registered voter and Y denote the opinion (s)he may have on the issue of Kennedy's assassination. We want to test the hypothesis

$$H_0 : X \text{ is independent from } Y$$
$$H_1 : X \text{ is not independent from } Y$$

Let n_{ij} to denote observed at i-th row and j-th column. Then we build the probability table where the MLE is used to estimate each $p_{ij} = \frac{n_{i+}n_{j+}}{n^2}$:

```
In [19]: prob = pd.DataFrame(index=[0,1,2], columns=[0,1,2])

         for i, row in obs.iterrows():
             for j, column in obs.iteritems():
                 prob.iloc[i,j] = row.sum() * column.sum() / 900 ** 2

         prob
```

```
Out[19]:            0         1          2
         0  0.0594222  0.315975  0.0490469
         1  0.0553778  0.294469  0.0457086
         2     0.0252     0.134     0.0208
```

Now we are ready to build the expectancy table where each $e_{ij} = np_{ij}$

```
In [23]: exp = pd.DataFrame(index=[0,1,2], columns=[0,1,2])

         for i, row in obs.iterrows():
             for j, column in obs.iteritems():
                 exp.iloc[i,j] = row.sum() * column.sum() / 900

         exp
```

```
Out[23]:        0        1        2
         0  53.48  284.378  44.1422
         1  49.84  265.022  41.1378
         2  22.68    120.6    18.72
```

Finally, by Theorem 10.5.1, we use calculate

$$d_2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim X^2_{(r-1)(c-1)} = X^2_4$$

```
In [21]: d2 = 0

         for i, row in obs.iterrows():
             for j, column in obs.iteritems():
                 d2 += (obs.iloc[i,j] - exp.iloc[i,j]) ** 2 / exp.iloc[i,j]

         d2
```

```
Out[21]: 18.710948240704745
```

Since $d_2 = 18.71 \geq 9.48 = X^2_{0.95,4}$, we reject the null hypothesis that X and Y are independent.