

# Homework 6 Solution

Mohsen Nabian

6/25/2015

```
++age++ 18-29 30-44 45-59 60+ Sum Democrat 86 72 73 71 302 INdependent 52 51 55 54 212 Republican 61 74 70 73 278 Sum 199 197 198 198 792
```

1.

- Based on the exit poll results, is age independent of Party ID or not? Conduct a chi-squared test by hand, showing each step in readably-formatted latex.
- Verify your results using R to conduct the test.

If we have two groups each group spans the whole population, in order to prove the independency of groups, we need to have a table of samples like above as well as doing the Chi-Square analysis. Here are the hypothesis: H0: the age and political view are INDEPENDENT. H1: the age and political view are NOT totally independent.

Assuming null hypothesis, we can find expected value for each element of the sample table. Let's calculate the number of participants in our sample.

$$n = (86 + 52 + 61) + (72 + 51 + 74) + (73 + 55 + 70) + (71 + 54 + 73) = 792$$

Finding expected values:

$$fe = \frac{(rowtotal) \times (columntotal)}{overalltotal}$$

so here are the expected values table:

```
++age+++++ 18-29 30-44 45-59 60+
Democrat 75.88 75.1 75.5 75.5
INdependent 53.26 52.73 53 53
Republican 69.85 69.15 69.5 69.5
```

Calculating

$$X^2 = \sum \frac{(fo - fe)^2}{fe}$$

Thus

$$X^2 = \frac{(86 - 75.88)^2}{75.88} + \frac{(72 - 75.1)^2}{75.1} + \frac{(73 - 75.5)^2}{75.5} + \frac{(71 - 75.5)^2}{75.5} + \frac{(52 - 53.26)^2}{53.26} + \frac{(51 - 52.73)^2}{52.73} + \frac{(55 - 53)^2}{53} + \frac{(54 - 53)^2}{53} + \frac{(61 - 69.85)^2}{69.85} + \frac{(74 - 69.15)^2}{69.15} + \frac{(70 - 69.5)^2}{69.5} + \frac{(73 - 69.5)^2}{69.5}$$

$$x^2 = 1.35 + 0.13 + 0.083 + 0.27 + 0.029 + 0.057 + 0.075 + 0.02 + 1.12 + 0.34 + 0.036 + 0.18 = 3.69$$

Degree of freedom

$$df = (r - 1)(c - 1) = 2 \times 3 = 6$$

```
pchisq(3.69,6,lower.tail=FALSE)
```

```
## [1] 0.7185431
```

```
qchisq(0.95,6,lower.tail=FALSE)
```

```
## [1] 1.635383
```

So both computations above shows that the null hypothesis could be true and unrejetable. Lets do the whole proceadure by R function :

```
ageparty<-data.frame(age1=c(86,52,61),age2=c(72,51,74),age3=c(73,55,70),age4=c(71,54,73),row.names=c("democrats","independen
ts","republicans"))
chisq.test(ageparty)
```

```
##
## Pearson's Chi-squared test
##
## data: ageparty
## X-squared = 3.6529, df = 6, p-value = 0.7235
```

The R Test also supports our hand calculations.

2.

- Now test for independence using ANOVA (an F test). Your three groups are Democrats, Independents, and Republicans. The average age for a Democrat is 43.3, for an Independent it's 44.6, and for a Republican it's 45.1. The standard deviations of each are D: 9.1, I: 9.2, R: 9.2. The overall mean age is 44.2. Do the F test by hand, again showing each step.
- Check your results in R using simulated data. Generate a simulated dataset by creating three vectors: Democrats, Republicans, and Independents. Each vector should be a list of ages, each with a length equal to the number of Democrats, Independents, and Republicans in the table above, and the appropriate mean and sd based on 2.a (use rnorm to generate the vectors). Combine all three into a single dataframe with two variables: age, and a factor that specifies D, I, or R. Then conduct an F test using R's aov function on that data and compare the results to 2.a. Note that your results may not exactly match 2a either quantitatively or qualitatively.

## Solutions:

Here is the summary of the F-Test analysis

N total samples in G different groups of one same parameter  $x$  :

$$(\bar{x}_1, s_1, n_1), (\bar{x}_2, s_2, n_2), \dots, (\bar{x}_g, s_g, n_g)$$

Null Hypothesis:

$$H_0 = \mu_1 = \mu_2 = \dots = \mu_g = \bar{x} (\text{average of all samples})$$

(That means that the paramter  $x$  in all those differnt group are the same, no dependency in  $x$  to any group.)

$$H_a = \text{at least one group has different response to } x.$$

(That means There could be some dependency between the parameter and anyone of the goup.)

$$F - \text{statistic} = \frac{\text{average variance between groups}}{\text{average variance within groups}}$$

N=total number of all data G=the number of groups

$$\text{average variance between groups} = \frac{n_1(\bar{x}_1 - \bar{x})^2 + \dots + n_g(\bar{x}_g - \bar{x})^2}{G - 1}$$

$$\text{average variance within groups} = \frac{(n_1 - 1) \times (s_1^2) + \dots + (n_g - 1) \times (s_g^2)}{N - G}$$

degree of freedoms:

$$df1 = G - 1$$

$$df2 = N - G$$

=====

Democrates(mean=43.3,s=9.1,n=302) Independent(mean=44.6,s=9.2,n=212) Republicans(mean=45.1,s=9.2,n=278) Overall Mean=44.2 N=792, G=3

$$\text{average variance between groups} = \frac{302 \times (43.3 - 44.2)^2 + 212 \times (44 - 44.2)^2 + 278 \times (45.1 - 44.2)^2}{3 - 1} = 239.14$$

$$\text{average variance within groups} = \frac{(302 - 1) \times (9.1^2) + (212 - 1) \times (9.2^2) + (278 - 1) \times (9.2^2)}{792 - 3} = 83.94$$

$$F - \text{statistic} = \frac{239.14}{83.94} = 2.85$$

$$df1 = 3 - 1 = 2$$

$$df2 = 792 - 3 = 789$$

```
pF(2.85,2,789)
```

```
## [1] 0.94156
```

```
qF(0.95,2,789)
```

```
## [1] 3.007136
```

Assuming 0.05 error, still we are UNABLE to reject the null hypothesis (0.94<0.95).But it is very close. That mean we can say with the error of 0.07 the null hypothesis is rejected. Lets do the F-Test:

```

Democrats<-rnorm(302,43.3,901)
Independents<-rnorm(212,44.6,9.2)
Republicans<-rnorm(278,45.1,9.2)
DD<-cbind("D",Democrats)
II<-cbind("I",Independents)
RR<-cbind("R",Republicans)
age_party<-rbind(DD,II,RR)
age_party<-cbind(as.factor(age_party[,1]),as.numeric(age_party[,2]))
DF<-data.frame(age_party)
head(DF)

```

```

##   X1      X2
## 1  1 350.85456
## 2  1 343.22001
## 3  1  70.15554
## 4  1 643.97944
## 5  1 914.35357
## 6  1 -368.29258

```

```

aov.ex<-aov(DF[,2]~DF[,1],data=DF)
summary(aov.ex)

```

```

##           Df      Sum Sq Mean Sq F value Pr(>F)
## DF[, 1]      1      254753   254753    0.859   0.354
## Residuals 790 234333690    296625

```

Results will change significantly any time we run this code. However, it shows how we can determine dependency of age with party inclination with statistical purposes. We assume the Null hypothesis that there is no dependency and try to reject it.