Tutorial 9

Research Methods for Politcal Science A

Michele McArdle

8 & 9 December 2020

Outline

1. Two Sample T-Test
2. Chi-Square Test

Two Sample T-test

If we want to compare two independent samples we can use a two sample t-test.

The samples need to meet some criteria:
- samples need to have the same size - samples need to have the same variance

Hypotheses in a two sample t-test

In our H_0 we assume that the means of both populations is equal.

$H_0 : \ \mu_1 \ = \ \mu_2$

Hypotheses in a two sample t-test

In a two-tailed two sample t-test

$H_{alt}: \ \mu_1 \ \neq \mu_2$

Hypotheses in a one sample t-test

In a two-tailed two sample t-test

$H_{alt}:\ \mu_1\ <\mu_2$
or
$H_{alt}:\ \mu_1\ >\mu_2$

Sample t-score

We calculate the sample 1 t-score with the following variable:

$$t = \frac{\bar{X_1} - \bar{X_2}}{s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p = \sqrt{\frac{sd_1^2 + sd_2^2}{2}}$$

Sample t-score in R

```r
sample1 <- rnorm(1000, 100, 15)
sample2 <- rnorm(1000,  95, 15)
sample3 <- rnorm(1000,  75, 15)

#Two tailed test
sp <- sqrt((sd(sample2)^2 + sd(sample3)^2)/2)
numerator <- mean(sample1) - mean(sample2)
denominator <- sp*sqrt(1/1000 + 1/1000)
t <- numerator/denominator
t
```

```
## [1] 7.106174
```

Exercise

Calcualte the t-score for samples 2 and 3.

Sample t-score in R

```
t.crit <- qt(0.05/2, df = (1000-1))
t.crit
```

```
## [1] -1.962341
```

```
t <= t.crit
```

```
## [1] FALSE
```

```
t >= t.crit*-1
```

```
## [1] TRUE
```

Exercise

Do a two-tailed two sample t-test for samples 1 and 3.

Chi-Square Test

A chi-square test is based on chi-square distribution.

It is also one-sided.

$\chi^2 > \chi^2_{\alpha,\ df}$

Chi-Square Test

The formula to calculate the chi-square value:

$$\chi^2 = \sum_i \frac{n_{observed} - n_{expected}}{n_{expected}}$$

We also need to calculate the degrees of freedom:

$$df = (r - 1) * (c - 1)$$

Chi-Square

We ask people there favourite colour.

	Green	Red	Orange	Pink	Blue
Student	25	80	50	20	75
Expected	50	50	50	50	50

Chi-Square

$$\chi^2 = \frac{(25-50)^2}{50} + \frac{(80-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(20-50)^2}{50} + \frac{(75-50)^2}{50} = 61$$

$$df = (5-1) * (2-1) = 4 * 1 = 4$$

Chi-Square Test

We then compare this to the critical value which we can look up in a table or calculate with R.

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312

Chi-Square Test

```
chi2 = (25-50)^2/(50) + (80-50)^2/(50) +
       (50-50)^2/(50)+ (20-50)^2/(50) +
       (75-50)^2/(50)
df = (5-1)*(2-1)


chi2
```

```
## [1] 61
```

```
df
```

```
## [1] 4
```

Chi-Square Test

```r
chi2.crit <- qchisq(p = 0.05, df = 4)
chi2.crit
```

```
## [1] 0.710723
```

```r
chi2 >= chi2.crit
```

```
## [1] TRUE
```

Exercise

Test whether we can reject the Null Hypothesis:

	Green	Red	Orange	Pink	Blue	Violet
Student	60	50	70	45	52	58
Expected	55	55	55	55	55	55

$$\chi^2 = \sum_i \frac{n_{observed} - n_{expected}}{n_{expected}}$$
$$df = (r-1) * (c-1)$$

```
# qchisq(p = , df = )
```

Exercise

```r
chi2 = (60-55)^2/(55) + (50-55)^2/(55) +
       (70-55)^2/(55) + (45-55)^2/(55) +
       (52-55)^2/(55) + (58-55)^2/(55)
df = (6-1)*(2-1)


chi2
```

```
## [1] 7.145455
```

```r
df
```

```
## [1] 5
```

Exercise

```r
chi2.crit <- qchisq(p = 0.01, df = 5)

 chi2.crit
```

```
## [1] 0.5542981
```