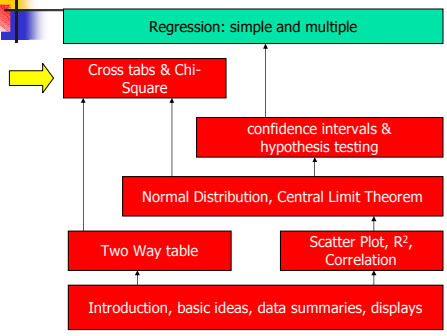
PUB – POS 316
Week 13

Analysis of Two-Way Tables: Chi-Square

Navid Ghaffarzadegan
navidg@gmail.com
Last updated – Feb, 10

Course Road Map



Introduction

- Data in categories
- Systematic Investigation of two-way tables.
- Chi-square, Chi-test
- Working with Excel.

Data in categories

- What is "data in categories?"
  - Two Variables are categorical:
    - X: Men or Women Y: Yes or No

	Gender	
Frequent of binge drinker	Men	Women
Yes	1630	1684
No	5550	8232

  - How should we analyze this data?
  - Joint Distribution: dist. of the whole data
  - Conditional distribution, Marginal distribution

Data in categories

	Gender		
Frequent of binge drinker	Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Conditional Distribution

	Gender	
Frequent of binge drinker	Men	Women
Yes	0.227019	0.1698265
No	0.772981	0.8301735

Systematic Investigation of two-way tables.

- How can we systematically compare two groups?
- Systematically:
  - With reporting the level of confidence.
  - Are we sure that the difference in two group is not just a matter of error in our study? (remember the issue of sampling vs. population?

Systematic Investigation of two-way tables.

- What do we expect to happen, if there is no systematic difference between male and female? (H_0)

Frequent of binge drinker	Gender Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Frequent of binge drinker	Gender Men	Women	total
Yes			3314
No			13782
	7180	9916	17096

Systematic Investigation of two-way tables.

- If there were no difference between male and female the conditional distribution would have shown that.
- In another word, numbers should show:
  - proportion of male (out of total male) that are binge drinkers = proportion of female (out of total female) that are binge drinkers.

Systematic Investigation of two-way tables.

- What do we expect to happen, if there is no systematic difference between male and female? (H_0)

Frequent of binge drinker	Gender Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Frequent of binge drinker	Gender Men	Women	total
Yes			3314
No			13782
	7180	9916	17096

Systematic Investigation of two-way tables.

- What do we expect to happen, if there is no systematic difference between male and female? (H_0)

Frequent of binge drinker	Gender Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Frequent of binge drinker	Gender Men	Women	total
Yes	1391.818	1922.182	3314
No	5788.182	7993.818	13782
	7180	9916	17096

Systematic Investigation of two-way tables.

- What do we expect to happen, if there is no systematic difference between male and female? (H_0)

Frequent of binge drinker	Gender Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

What we expect under the null hypothesis (No difference between male and female)

Frequent of binge drinker	Gender Men	Women	total
Yes	1391.818	1922.182	3314
No	5788.182	7993.818	13782
	7180	9916	17096

Systematic Investigation of two-way tables.

- What do we expect to happen, if there is no systematic difference between male and female?

Frequent of binge drinker	Gender Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
		9916	17096

Compare to see if we can reject the null hypothesis? Are we far enough from the null hypothesis?

Frequent of binge drinker	Gender Men	Women	total
Yes	1391.818	1922.182	3314
No	5788.182	7993.818	13782
	7180	9916	17096

Chi-square, Chi-test

Frequent of binge drinker	Gender: Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Frequent of binge drinker	Gender: Men	Women	total
Yes	1391.818	1922.182	3314
No	5788.182	7993.818	13782
	7180	9916	17096

- We can look at the difference between these numbers. Something like:
- (1630-1391)+(1684-1922)+(5550-5788)+(8232-7993)
- But again they cancel out! Can you guess what we should do?!
- This is what we look at:

$$X^2 = \frac{(1630-1391)^2}{1391} + \frac{(1684-1922)^2}{1922} + \frac{(5550-5788)^2}{5788} + \frac{(8232-7993)^2}{7933}$$

Chi-square, Chi-test

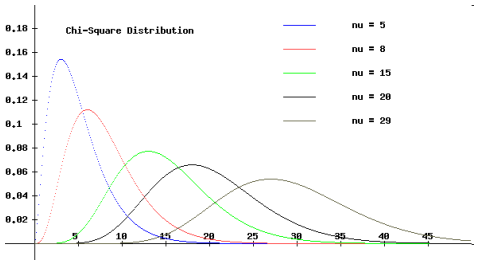
Frequent of binge drinker	Gender: Men	Women	total
Yes	1630	1684	3314
No	5550	8232	13782
Total	7180	9916	17096

Frequent of binge drinker	Gender: Men	Women	total
Yes	1391.818	1922.182	3314
No	5788.182	7993.818	13782
	7180	9916	17096

$$X^2 = \frac{(1630-1391)^2}{1391} + \frac{(1684-1922)^2}{1922} + \frac{(5550-5788)^2}{5788} + \frac{(8232-7993)^2}{7933}$$

- Now what should we do with this number?
- We know that as X^2 becomes larger, as we get far from the null hypothesis.
- In another word: P-value should decline.
- BUT X^2 does not follow z or t distributions!.. It follows X^2 (Chi-Square distribution)

Chi-square, Chi-test



df = (number of rows – 1)*(number of columns – 1)

Chi-square, Chi-test

- Work with the table.

Chi-square, Chi-test

- The procedure:
  - 1. State the null hypothesis
  - 2. Build the table under the null hypothesis.
  - 3. Calculate Chi-Square.
  - 4. Find the critical value in the table and see if you can REJECT the null hypothesis.
- (There is also a short cut!)

Working with Excel

- Another example in the Excel file.
- How excel can give additional help.
  - =CHITEST(actual_range,expected_range)
  - = CHIINV(probability,degrees_freedom)
    - This is our table!
  - = CHIDIST(x,degrees_freedom)
    - Gives the inverse of our table.. Asks for x and gives probability (p-value) to reject the null hypothesis