Univariate Statistical Analysis

Lecture 9

Chi-Square Test
Chapter 12

Today

- Test for univariate data (Goodness-Of-Fit)
- Two–way table
  - Test for Homogeneity
  - Test for Independence

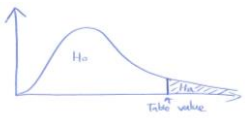
Chi-Square Test

- **Chi-Square test** is a non-parametric test.
- Help to understand the relationship between **categorical variables**, for examples gender, income class and age group etc.
- Dealing with proportions
- Frequency
  - Observed data vs Expected data
- Each of the Expected frequency must be greater than 5.

Test for Univariate data

- Steps

1. Calculate the Expected data → Average = $\frac{Total\ Counts}{n}$
2. Define or State
   $\boldsymbol{H_0}$: The (observed) data meet the expected distribution
   $\boldsymbol{H_a}$: The data do not meet the expected distribution
3. Sketch a curve



Test for Univariate data

4. Find the critical value for χ^2 → Table value.
Need to know df (n-1) and $\propto$

5. Label step 4. result on the curve

6. Find the test statistic (value)

$$\chi^2 = \sum_{all\ cells} \frac{(Observed\ count\ - Expected\ counts)^2}{Expected\ counts}$$

7. Label step 6. result on the curve

Test for Univariate data

8. Decide to reject $\boldsymbol{H_0}$ or failed to reject $\boldsymbol{H_0}$

Reject $\boldsymbol{H_o}$. There is **enough evidence** to conclude the data do not meet the expected distribution.

Fail to Reject $\boldsymbol{H_o}$. There is **insufficient evidence** to conclude that the data do not meet the expected distribution.

Example 1 – Tasty Dog Food (p.661, example 12.3)

Given the following set of data and $\propto$ =0.05

	Observed
Duck Liver pate	3
Spam	11
Dog Food	8
Pork liver pate	6
Liverwurst	22

Example 1 – Tasty Dog Food Cont'
Univariate data test – Step 1 and 2

	Observed (O)	Expected (E)
Duck Liver pate	3	10
Spam	11	10
Dog Food	8	10
Pork liver pate	6	10
Liverwurst	22	10
Total	**50**	**50**

Expected data = $\frac{Total\ Counts}{n} = \frac{50}{5} = 10$

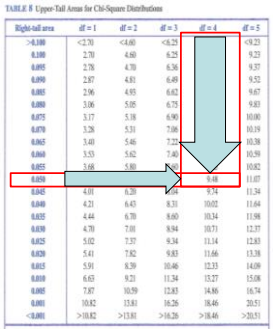
H_0: The data meet the expected distribution
H_a: The data do not meet the expected distribution

Example 1 – Tasty Dog Food Cont'
Univariate data test – Step 4

Critical χ^2 value

$\alpha = 0.05,$
$df = n - 1 = 5 - 1 = 4$

So, critical value is 9.48

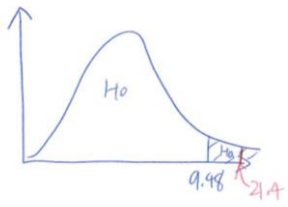


Example 1 – Tasty Dog Food Cont'
Univariate data test – Step 6

O	E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
3	10	-7	49	4.9
11	10	1	1	0.1
8	10	-2	4	0.4
6	10	-4	16	1.6
22	10	12	144	14.4
50	**50**			**21.4**

$$\chi^2 = \sum_{all\ cells} \frac{(Observed\ count\ - Expected\ counts)^2}{Expected\ counts} = \mathbf{21.4}$$

Example 1 – Tasty Dog Food Cont'
Univariate data test – Step 3,5,7 and 8



Conclusion:
Reject H_o. There is **enough evidence** to conclude that the observed five spreads do not meet the expected distribution.

Practice Question 1 – Trauma Seasons (p.663)

Practice Question 1 – Trauma Seasons (p.663)

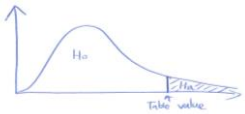
Solution:
Critical Value = 7.81
$\chi^2 = 30.19$

Reject H_o. There is enough evidence to conclude that the observed trauma cases for the four seasons do not meet the expected distribution.

Test for Homogeneity

➢Steps

1. Calculate the Expected data → $\frac{(Row\ Total)\ (Column\ Total)}{Grand\ Total}$
2. Define or State
   $\boldsymbol{H_0}$: Proportions in each a category are the same for all b groups.
   $\boldsymbol{H_a}$: Proportions in each a category are not all the same for all b groups.
3. Sketch a curve



Test for Homogeneity

4. Find the critical value for χ^2 → Table value.
Need to know Row df (n-1) x Column df (n-1) and $\propto$

5. Label step 4. result on the curve

6. Find the test statistic (value)

$$\chi^2 = \sum_{all\ cells} \frac{(Observed\ count\ - Expected\ counts)^2}{Expected\ counts}$$

7. Label step 6. result on the curve

Test for Homogeneity

8. Decide to reject $\boldsymbol{H_0}$ or failed to reject $\boldsymbol{H_0}$

Reject $\boldsymbol{H_o}$. There is **enough evidence** to conclude that the proportions in each *a* category are not all the same for all *b* groups.

Fail to Reject $\boldsymbol{H_o}$. There is **insufficient evidence** to conclude that the proportions in each *a* category are not all the same for all *b* groups.

Example 2 – Heart Attacks (p.667, example 12.4)

Given the following set of data and $\propto = 0.05$

Observed (O)			
	Survived	Did not Survive	**Row Total**
House or Townhouse	217.00	5,314.00	**5,531.00**
Apartment First or Second	35.00	632.00	**667.00**
Apartment Third or Higher	46.00	1,650.00	**1,696.00**
Column Total	**298.00**	**7,596.00**	**7,894.00**

Example 2 – Heart Attacks Homogeneity Test – Step 1 Cont'

Observed (O)			
	Survived	Did not Survive	**Row Total**
House or Townhouse	217.00	5,314.00	**5,531.00**
Apartment First or Second	35.00	632.00	**667.00**
Apartment Third or Higher	46.00	1,650.00	**1,696.00**
Column Total	**298.00**	**7,596.00**	**7,894.00**

Expected (E)			
	Survived	Did not Survive	**Row Total**
House or Townhouse	208.80	5,322.20	**5,531.00**
Apartment First or Second	25.18	641.82	**667.00**
Apartment Third or Higher	64.02	1,631.98	**1,696.00**
Column Total	**298.00**	**7,596.00**	**7,894.00**

$\frac{(298)(5,531)}{7,894} = 208.8$ $\quad$ $\frac{(7,596)(5,531)}{7,894} = 5,322.2$

$\frac{(298)(667)}{7,894} = 25.18$ $\quad$ $\frac{(7,596)(667)}{7,894} = 641.82$

$\frac{(298)(1,696)}{7,894} = 64.02$ $\quad$ $\frac{(7,596)(1,696)}{7,894} = 1,631.98$

Example 2 – Heart Attacks Homogeneity Test – Step 2 Cont'

H_0: Proportions in each survival category are the same for all three groups. (Does not matter which population group, the (not) survived proportions are the same)

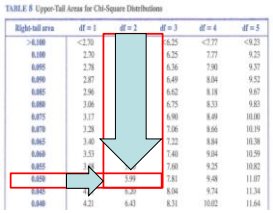
H_a: Proportions in each survival category are not all the same for all three groups.

Example 2 – Heart Attacks Homogeneity Test – Step 4 Cont'

Critical χ^2 value

$\alpha = 0.05$,

Row df (n–1) x Column df (n–1)
$= (3-1)(2-1)$
$= 2\,x1$
$= 2$



So, critical value is **5.99**

Example 2 – Heart Attacks Homogeneity Test – Step 6 Cont'

Observed (O)

	Survived	Did not Survive	Row Total
House or Townhouse	217.00	5,314.00	**5,531.00**
Apartment First or Second	35.00	632.00	**667.00**
Apartment Third or Higher	46.00	1,650.00	**1,696.00**
Column Total	**298.00**	**7,596.00**	**7,894.00**

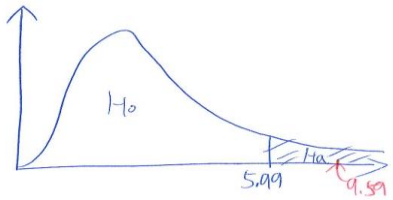
Expected (E)

	Survived	Did not Survive	Row Total
House or Townhouse	208.80	5,322.20	**5,531.00**
Apartment First or Second	25.18	641.82	**667.00**
Apartment Third or Higher	64.02	1,631.98	**1,696.00**
Column Total	**298.00**	**7,596.00**	**7,894.00**

$$\chi^2 = \frac{(217-208.8)^2}{208.8} + \frac{(35-25.18)^2}{25.18} + \cdots + \frac{(632-641.82)^2}{641.82} + \frac{(1650-1631.98)^2}{1631.98}$$

$$\chi^2 = 0.32 + 3.83 + \cdots + 0.18 + 0.2 = 9.59$$

Example 2 – Heart Attacks Homogeneity Test – Step 3,5,7 and 8 Cont'



Conclusion:
Reject H_o. There is **enough evidence** to conclude that the proportions in each survival category are NOT all the SAME for all three groups.

Practice Question 2 – Weight Off (p.670, Example 12.6)

Practice Question 2 – Weight Off (p.670, Example 12.6)

Solution:
Critical Value = 9.21
$\chi^2 = 13.77$

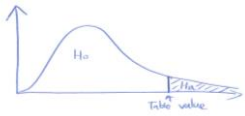
Reject H_o. There is enough evidence to conclude that the proportions in each weight gained category are different for all the three weight loss methods.

Test for Independence

➢Steps

1. Calculate the Expected data → $\frac{(Row\ Total)\ (Column\ Total)}{Grand\ Total}$
2. Define or State
   H_0: variable x and variable y are independent. (no association)

   H_a: variable x and variable y are dependent. (is an association)
3. Sketch a curve



Test for Independence

4. Find the critical value for χ^2 → Table value.
   Need to know Row df (n-1) x Column df (n-1) and $\propto$
5. Label step 4. result on the curve
6. Find the test statistic (value)

$$\chi^2 = \sum_{all\ cells} \frac{(Observed\ count\ - Expected\ counts)^2}{Expected\ counts}$$

7. Label step 6. result on the curve

Test for Independence

8. Decide to reject H_0 or failed to reject H_0

   Reject H_o. There is **enough evidence** to conclude that variable x and variable y are dependent.

   Fail to Reject H_o. There is **insufficient evidence** to conclude that variable x and variable y are dependent.

Example 3 – Pained Expression (p.673, example 12.7)

Given the following set of data and $\propto = 0.05$

	No Pain	Pain	Row Total
Facial Expression No Pain	17.00	40.00	**57.00**
Facial Expression Pain	3.00	29.00	**32.00**
Column Total	**20.00**	**69.00**	**89.00**

Example 3 – Pained Expression Cont' Independence Test – Step 1

Observed (O)	Self-Report		
	No Pain	Pain	**Row Total**
Facial Expression No Pain	17.00	40.00	**57.00**
Facial Expression Pain	3.00	29.00	**32.00**
Column Total	**20.00**	**69.00**	**89.00**

Expected (E)	Self-Report		
	No Pain	Pain	**Row Total**
Facial Expression No Pain	12.81	44.19	**57.00**
Facial Expression Pain	7.19	24.81	**32.00**
Column Total	**20.00**	**69.00**	**89.00**

$\frac{(20)(57)}{89} = 12.81$ $\frac{(69)(57)}{89} = 44.19$

$\frac{(20)(32)}{89} = 7.19$ $\frac{(69)(32)}{89} = 24.81$

Example 3 – Pained Expression Cont' Independence Test – Step 2

H_0: Facial expression and self-report pain are independent.

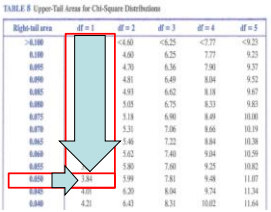
H_a: Facial expression and self-report pain are dependent.

Example 3 – Paired Expression Independence Test – Step 4 Cont'

Critical χ^2 value

$\alpha = 0.05,$

Row df (n–1) x Column df (n–1)
$= (2-1)(2-1)$
$= 1\ x 1$
$= 1$



So, critical value is **3.84**

Example 3 – Paired Expression Independence Test – Step 6 Cont'

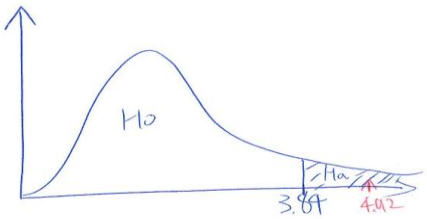
Observed (O)	Self-Report		
	No Pain	Pain	Row Total
Facial Expression No Pain	17.00	40.00	**57.00**
Facial Expression Pain	3.00	29.00	**32.00**
Column Total	**20.00**	**69.00**	**89.00**

Expected (E)	Self-Report		
	No Pain	Pain	Row Total
Facial Expression No Pain	12.81	44.19	**57.00**
Facial Expression Pain	7.19	24.81	**32.00**
Column Total	**20.00**	**69.00**	**89.00**

$$\chi^2 = \frac{(17-12.81)^2}{12.81} + \frac{(3-7.19)^2}{7.19} + \frac{(40-44.19)^2}{44.19} + \frac{(29-24.81)^2}{24.81}$$

$$\chi^2 = 1.37 + 2.44 + .04 + 0.71 = \mathbf{4.92}$$

Example 3 – Paired Expression Independence Test – Step 3,5,7 and 8 Cont'



Conclusion:
Reject H_o. There is enough evidence to conclude that facial expression and self-report pain are dependent.

Practice Question 3 – Sleep Quality (p.674, Example 12.8)

Practice Question 3 – Sleep Quality (p.674, Example 12.8)

Solution:
Critical Value = 11.34
$\chi^2 = 24.99$

Reject H_o. There is enough evidence to conclude that sleep quality and activity level are dependent.

Conclusion

- Test for univariate data (Goodness-Of-Fit)
- Two–way table
  - Test for Homogeneity
  - Test for Independence