

## 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

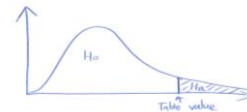
- **Chi-Square test** is a non-parametric test.
- Help to understand the relationship between **categorical variables**, for examples gender, income class and age group etc.
- Dealing with proportions
- Frequency
  - Observed data vs Expected data
- Each of the Expected frequency must be greater than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

### Test for Univariate data

#### ➤ Steps

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

3. Sketch a curve



### Test for Univariate data

4. Find the critical value for  $\chi^2 \rightarrow$  Table value.  
Need to know **df (n-1)** and  $\alpha$

5. Label step 4. result on the curve

6. Find the test statistic (value)

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

7. Label step 6. result on the curve

### Test for Univariate data

8. Decide to reject  $H_0$  or failed to reject  $H_0$

**Reject  $H_0$ .** There is **enough evidence** to conclude the data do not meet the expected distribution.

**Fail to Reject  $H_0$ .** 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  $\alpha = 0.05$

	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
<b>Total</b>	<b>50</b>	<b>50</b>

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

$H_0$ : 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  $\chi^2$  value

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

So, critical value is 9.48

TABLE 8 Types-Tail Areas for Chi-Square Distribution

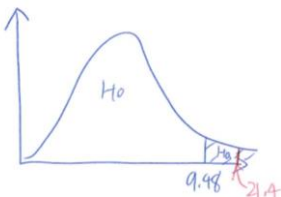
Right-tail area	df = 1	df = 2	df = 3	df = 4	df = 5
>0.100	<2.70	<4.00	<5.25	<6.25	<7.23
0.100	2.70	4.00	5.25	6.25	7.23
0.050	2.70	4.70	6.36	7.37	8.37
0.025	2.87	4.81	6.49	7.50	8.50
0.010	2.96	4.91	6.62	7.62	8.62
0.005	3.06	5.02	6.75	7.75	8.75
0.001	3.17	5.18	6.92	7.92	8.92
0.0005	3.28	5.31	7.08	8.08	9.08
0.0001	3.40	5.46	7.22	8.22	9.22
0.00005	3.53	5.62	7.40	8.40	9.40
0.00001	3.68	5.80	7.60	8.60	9.60
0.000005	4.01	6.39	8.04	9.34	10.34
0.000001	4.21	6.63	8.31	9.62	10.64
0.0000005	4.44	6.79	8.60	10.34	11.98
0.0000001	4.70	7.01	8.94	10.71	12.27
0.00000005	5.02	7.37	9.34	11.34	12.83
0.00000001	5.41	7.82	9.81	11.96	13.39
0.000000005	5.91	8.39	10.46	12.83	14.08
0.000000001	6.63	9.21	11.34	13.27	15.08
0.0000000005	7.87	10.59	12.83	14.86	16.74
0.0000000001	10.82	13.81	16.26	18.46	20.51
<0.00000000005	>10.82	>13.81	>16.26	>18.46	>20.51

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

O	E	O-E	(O-E) <sup>2</sup>	$\frac{(O-E)^2}{E}$
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
<b>50</b>	<b>50</b>			<b>21.4</b>

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

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



**Conclusion:**  
 Reject  $H_0$ . 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  $H_0$ . There is enough evidence to conclude that the observed trauma cases for the four seasons do not meet the expected distribution.

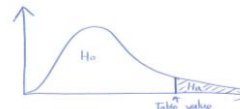
### Test for Homogeneity

➤Steps

1. Calculate the Expected data →  $\frac{(\text{Row Total}) (\text{Column Total})}{\text{Grand Total}}$

2. Define or State  
 $H_0$ : Proportions in each a category are the **same** for all b groups.  
 $H_a$ : 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  $\chi^2 \rightarrow$  Table value.  
 Need to know **Row** df (n-1) x **Column** df (n-1) and  $\alpha$

5. Label step 4. result on the curve

6. Find the test statistic (value)

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

7. Label step 6. result on the curve

### Test for Homogeneity

8. Decide to reject  $H_0$  or failed to reject  $H_0$

**Reject  $H_0$ .** 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  $H_0$ .** 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  $\alpha = 0.05$

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
<b>Column Total</b>	<b>298.00</b>	<b>7,596.00</b>	<b>7,894.00</b>

### 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
<b>Column Total</b>	<b>298.00</b>	<b>7,596.00</b>	<b>7,894.00</b>

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
<b>Column Total</b>	<b>298.00</b>	<b>7,596.00</b>	<b>7,894.00</b>

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

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

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

### 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  $\chi^2$  value

$\alpha = 0.05$ ,

Row df  $(n-1) \times$  Column df  $(n-1)$   
 $= (3 - 1)(2 - 1)$   
 $= 2 \times 1$   
 $= 2$

Right-tail area	df = 1	df = 2	df = 3	df = 4	df = 5
> 0.100	< 2.71	< 1.68	0.58	< 0.71	< 0.82
0.100	2.71	1.68	0.58	0.71	0.82
0.050	2.71	1.68	0.58	0.71	0.82
0.025	2.71	1.68	0.58	0.71	0.82
0.010	2.71	1.68	0.58	0.71	0.82
0.005	2.71	1.68	0.58	0.71	0.82
0.001	2.71	1.68	0.58	0.71	0.82
0.0005	2.71	1.68	0.58	0.71	0.82

So, critical value is **5.99**

### Example 2 – Heart Attacks Homogeneity Test – Step 3

**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	48.00	1,650.00	1,698.00
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

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

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

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

**Cont'**

**Conclusion:**  
**Reject  $H_0$ .** 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  $H_0$ . 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

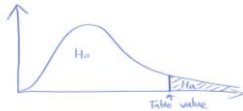
➤Steps

1. Calculate the Expected data →  $\frac{(\text{Row Total}) (\text{Column Total})}{\text{Grand Total}}$

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  $\chi^2 \rightarrow$  Table value.  
 Need to know Row df (n-1) x Column df (n-1) and  $\alpha$

5. Label step 4. result on the curve

6. Find the test statistic (value)

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

7. Label step 6. result on the curve

### Test for Independence

8. Decide to reject  $H_0$  or failed to reject  $H_0$   
**Reject  $H_0$ .** There is **enough evidence** to conclude that variable x and variable y are **dependent**.

**Fail to Reject  $H_0$ .** 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  $\alpha = 0.05$

	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 Independence Test – Step 1 Cont'

Observed (O)	Self-Report		Row Total
	No Pain	Pain	
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		Row Total
	No Pain	Pain	
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

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

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

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

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

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

**Example 3 – Pained Expression Independence Test – Step 4** **Cont'**

Critical  $\chi^2$  value

$\alpha = 0.05,$

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

TABLE F: Upper-Tail Areas for Chi-Square Distribution

Right tail area	df = 1	df = 2	df = 3	df = 4	df = 5
>0.100	0.004	1.68	0.878	0.711	0.625
0.100	0.004	1.68	0.878	0.711	0.625
0.050	0.004	1.68	0.878	0.711	0.625
0.025	0.004	1.68	0.878	0.711	0.625
0.010	0.004	1.68	0.878	0.711	0.625
0.005	0.004	1.68	0.878	0.711	0.625
0.001	0.004	1.68	0.878	0.711	0.625
0.0005	0.004	1.68	0.878	0.711	0.625
0.0001	0.004	1.68	0.878	0.711	0.625
0.00005	0.004	1.68	0.878	0.711	0.625
0.00001	0.004	1.68	0.878	0.711	0.625

*Note: In the original image, a red box highlights the value 3.84 at the intersection of the 0.050 row and the 1 df column.*

So, critical value is **3.84**

**Example 3 – Pained Expression Independence Test – Step 6** **Cont'**

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

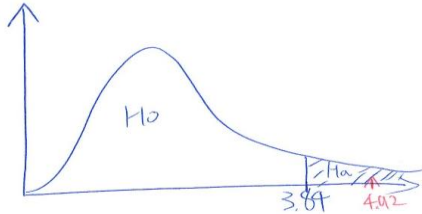
  

Expected (E)	Self-Report		Row Total
	No Pain	Pain	
Facial Expression No Pain	12.81	44.19	57.00
Facial Expression Pain	7.19	24.81	32.00
<b>Column Total</b>	<b>20.00</b>	<b>69.00</b>	<b>89.00</b>

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

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

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



**Conclusion:**  
 Reject  $H_0$ . 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  $H_0$ . 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