Tests for Qualitative Data Sets

Recall that qualitative data are categorical such as:

- Gender
- Favorite color
- Type of college degree earned
- Favorite football team
- Top five most popular movies

- It is impossible to calculate a population or a sample statistic like a mean or a standard deviation for qualitative data set.

- Therefore, hypothesis tests such as a z–test, or a t–test cannot be used with qualitative data sets.

We need another tool for dealing with qualitative data sets.

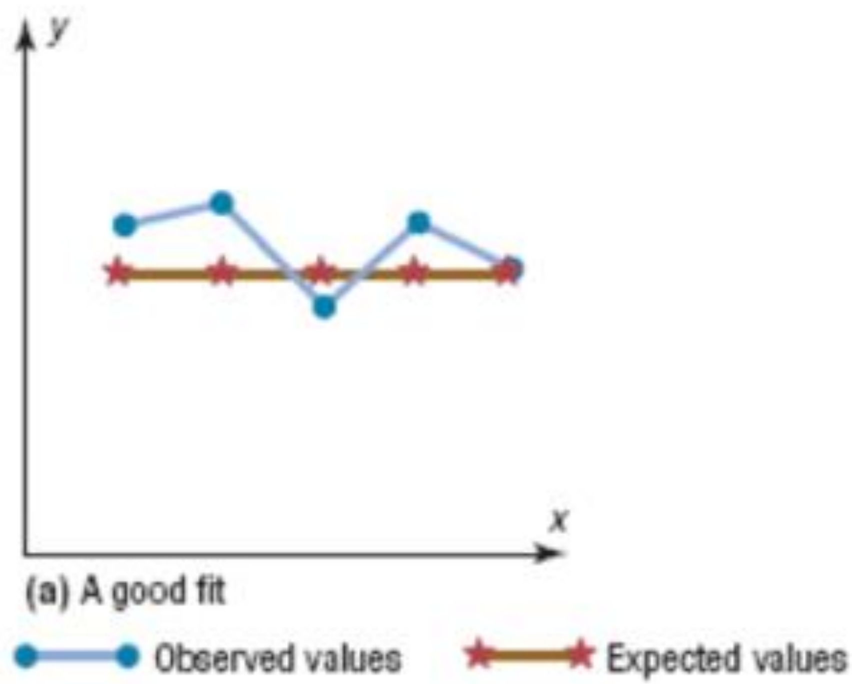
Chi Squared Tests

- One of the most common statistical tests for qualitative data is the chi-square test

χ^2 Test for Goodness of Fit

- The chi-squared statistic can be used to see whether a frequency distribution fits a specific pattern. This is referred to as the chi-square goodness-of-fit test.

What does a good fit versus a not-a-good fit?



When to choose this type of test?

1. **In order to meet customer demands, a manufacturer of running shoes may wish to see whether buyers show a preference for a specific style.**

2. **A traffic engineer may wish to see whether accidents occur more often on some days than on others, so that police patrols can increase accordingly.**

3. **An emergency service may want to see whether it receives more calls at certain times of the day than at others so that it can provide adequate staffing.**

The Hypothesis Test for χ^2 Goodness of Fit

4 Step Process

1) The hypotheses **<u>will be written as statements instead of using symbols</u>**

   In general,

   $H_0: Good\ fit$ (the sample data fits some pre–defined or theoretical pattern)
   $H_1: NOT\ a\ good\ fit$ (the sample data does not fit a pre–defined or theoretical pattern)

2) Set the critical range
3) The test value is $\chi^2 = \sum \frac{(O - E)^2}{E}$

   O = Observed frequencies

   E = Expected frequency
4) Make a decision

Example: Fruit Soda Flavors

A market analyst wished to see whether consumers have any preference among five flavors of a new fruit soda. A sample of 100 people provided the following data. Is there enough evidence to reject the claim **that there is no preference** in the selection of fruit soda flavors, using the data shown? Let α = 0.05.

	Cherry	Strawberry	Orange	Lime	Grape
Observed	32	28	16	14	10
Expected	20	20	20	20	20

Step 1: State the hypotheses and identify the claim.

H_0: Consumers show no preference in flavors (claim).

H_1: Consumers show a preference in soda flavors.

Step – 2: Find the critical value. [No. of attributes = 1]

Degrees of Freedom $= 5 - 1 = 4$ and $\alpha = 0.05$.

Step – 3: Compute the test value

	Cherry	Strawberry	Orange	Lime	Grape
Observed	32	28	16	14	10
Expected	20	20	20	20	20

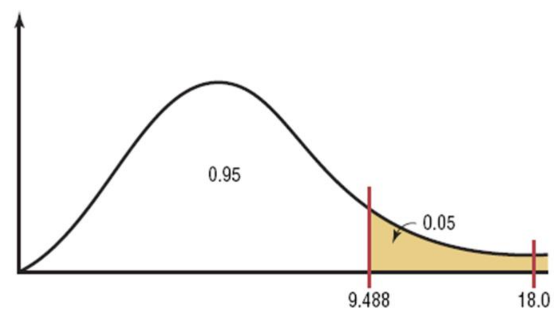
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(32-20)^2}{20} + \frac{(28-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(10-20)^2}{20} = 18.0$$

Step – 4: Make Decision

$\chi^2_{computed}$ **= 18.0** **and** $\chi^2_{(5-1,0.05)}$ **= 9.488**

$\chi^2_{computed} > \chi^2_{(5-1,0.05)}$. **H0 is rejected.**



There is enough evidence to reject the claim that consumers show no preference for the flavors

Exercise problem

Suppose we examine a sample of 300 children Left-Handed vs. Gender. Sample results organized in the following table

Gender	Hand Preference		Row Total
	Left	Right	
Female	12	108	120
Male	24	156	180
Column Total	36	264	300

Test whether the Proportion of females who are left handed is equal to the proportion of males who are left handed (α=0.05)

Step 1:

H0: Proportion of females who are left handed is equal to the proportion of males who are left handed

Step 2: Critical Range $\chi^2_{((row-1)x(column-1),0.05)}$

$\chi^2((2-1)x(2-1),0.05)=\chi^2(1,0.05) = 3.841$

Step 3: Compute χ^2 statistic. Before that prepare the table shown in the next slide.

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

O – Observed count

E – Expected count

Expected Count (E) = $\frac{row\ count\ x\ column\ count}{total\ count}$

Gender	Hand Preference		Row Total
	Left	Right	
Female	Observed = 12 Expected = 14.4	Observed = 108 Expected = 105.6	120
Male	Observed = 24 Expected = 21.6	Observed = 156 Expected = 158.4	180
Column Total	36	264	300

$$FL = \frac{36\ x\ 120}{300} = 14.4$$

$$FR = \frac{264\ x\ 120}{300} = 105.6$$

$$ML = \frac{36\ x\ 180}{300} = 21.6$$

$$MR = \frac{264\ x\ 180}{300} = 158.4$$

Now compute $\chi^2 = \sum \frac{(O-E)^2}{E}$

χ^2_{stat} =

$$= \frac{(12-14.4)^2}{14.4} + \frac{(108-105.6)^2}{105.6} + \frac{(24-21.6)^2}{21.6} + \frac{(156-158.4)^2}{158.4} = 0.7576$$

$\chi^2(1,0.05) = 3.841 > \chi^2_{stat} = 0.7576$

Therefore, Accept the null hypothesis H0.
i.e., Proportion of females who are left handed is equal to the proportion of males who are left handed

Exercise Problem – 1

The following data show the distribution of digits in the numbers chosen at random from a telephone directory:

Digit	0	1	2	3	4	5	6	7	8	9	Total
Count	1026	1107	997	966	1075	933	1107	972	964	853	10000

Test whether the digits may be taken to occur equally frequently in the directory

Solution

H0: The digits occur equally frequently in the directory.

Based on H0, the expected frequencies are 1000 in each of the digits.

Digit	0	1	2	3	4	5	6	7	8	9	Total
Count [O]	1026	1107	997	966	1075	933	1107	972	964	853	10000
Count [E]	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	10000
O-E	26	107	-3	-34	75	-67	107	-28	-36	-147	

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2(9,0.05) = 16.919$$

$$\chi^2(9,0.05) = 16.919 < \chi^2_{\,stat} = 58.542$$

H0 is rejected. i.e., the digits do not occur uniformly in the directory.

Exercise Problem – 2

The following data show the number of aircraft accidents that occurred during the various days of a week.

Day	MON	TUE	WED	THU	FRI	SAT
Count	15	19	13	12	16	15

Test whether the accidents are uniformly distributed over the week.

Solution

H0: Accidents occur uniformly distributed over the week.

Based on H0, the expected count on accidents are 15 in each of the days of week

Digit	MON	TUE	WED	THU	FRI	SAT
Count [O]	15	19	13	12	16	15
Count [E}	15	15	15	15	15	15
O-E	0	4	2	3	1	0

$\chi^2(5,0.05) = 11.07$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2$$

$\chi^2(5,0.05) = 11.07 > \chi^2_{stat} = 2$

H0 is accepted. i.e., the accidents may be regarded to occur uniformly over the week.

Exercise Problem – 3

The following table gives for a sample married women, the level of education and the marriage adjustment score

Level of Education	Marriage Adjustment Level				Total
	Very Low	Low	High	Very High	
PG Degree	24	97	62	58	241
UG Degree	22	28	30	41	121
School	32	10	11	20	73
Total	78	135	103	119	435

Can you conclude from this data that the higher the level of education greater is the degree of adjustment in marriage at 5% LoS?.

Solution

H0: There is no relation between the level of education and adjustment in marriage.

Level of Education	Marriage Adjustment Level								Total
	Very Low		Low		High		Very High		
PG Degree	24	43	97	75	62	57	58	66	241
UG Degree	22	22	28	37	30	29	41	33	121
School	32	13	10	23	11	17	20	20	73
Total	78		135		103		119		435

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 57.66 \quad \text{DoF} = (r-1)(c-1) = 2\text{x}3 = 6$$

$\chi^2(6,0.05) = 12.59 < \chi^2 = 57.66$. H0 is rejected.
It is concluded that there is relation between the level of Education and the adjustment in marriage.

Exercise Problem

1. A total number of 3759 individuals were interviewed in a public opinion survey on a political proposal. Of them, 1872 were men and the rest women. A total of 2257 individuals were in favour of the proposal and 917 were opposed to it. A total of 243 men were undecided and 442 women were opposed to the proposal. Do you justify or contradict the hypothesis that there is no association between gender and attitude?

Solution

Derived table from the given data

	Favoured	Opposed	Undecided	Total
Men	1154	475	243	1872
Women	1103	442	342	1887
Total	2257	917	585	3759

Calculate the Expected values [E]

	Favoured	Opposed	Undecided	Total
Men	1154 **1124**	475 **457**	243 **291**	1872
Women	1103 **1133**	442 **460**	342 **294**	1887
Total	2257	917	585	3759

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ = 18.76 and Dof = 1 x 2 = 2

$\chi^2(2,0.05) = 5.991 < \chi^2 = 18.76$. H0 is rejected.

There is no association between Gender and attitude on proposal