

NON-PARAMETRIC TESTS

1. Introduction

- Inferential statistics where population parameters are not a requirement to calculate its value
- A process which is carried out in order to find out whether or not a particular statistical hypothesis is likely to be true
- A statistical test in which no assumption are made about any statistical parameter. This is similar to a test in which we do not assume that the data have any particular distribution

2. The X^2 test for goodness of fit

- Test hypothesis about the proportions of a population distribution
- Test how well the sample proportions fit the population proportion specified by the null hypothesis
- Example:
 - H_0 : there is no difference in proportion of people in different categories
 - Observed data (f_o)

Category 1	Category 2	Category 3
7	26	27

$$n = 60$$

- Expected population based on hypothesis (f_e)

Category 1	Category 2	Category 3
20	20	20

$$1/3 \text{ of } 60 \text{ in each category} = 1/3 * 60 = 20$$

- Difference between observed data and expected data ($f_o - f_e$)

Category 1	Category 2	Category 3
$7 - 20 = -13$	$26 - 20 = 6$	$27 - 20 = 7$

- Square the differences $(f_o - f_e)^2$

Category 1	Category 2	Category 3
$(-13)^2 = 169$	$6^2 = 36$	$7^2 = 49$

- $X^2 = \Sigma [(\text{square the difference}) / (\text{expected data})]$
- $X^2 = \Sigma [(f_o - f_e)^2 / (f_e)] \rightarrow$ chi square formula
- $X^2 = 169/20 + 36/20 + 49/20$
- $X^2 = 8.45 + 1.8 + 2.45$
- $X^2 = 12.7$
- Degree of freedom = number of column - 1
- Degree of freedom = $3 - 1 = 2$

3. Chi-square test for independence

- Test a relationship between two variables
- Each individual in the sample is measured or classified on two separate variables
- Example:
 - H_o : there is no relationship between preference of teaching method and gender of the students
 - Variable 1: teaching method
 - Variable 2: gender of students

- Observed data (f_o)

Gender	Teaching Methods		Total
	Lecture	Tutorial	
Male	25	25	50
Female	35	65	100
Total	60	90	150

- 40% students refer lecture
- 60% students prefer tutorial
- Expected data (f_e)

Gender	Teaching Methods		Total
	Lecture	Tutorial	
Male	*20	30	**50
Female	40	60	100
Total	**60	90	***150

- $f_e = (**\text{Row total} \times **\text{Column total}) / (***\text{whole total})$
 - *Example of calculation
 - $**50 \times **60 / ***150 = *20$
- Difference between expected and observed value ($f_o - f_e$)

Gender	Teaching Methods		Total
	Lecture	Tutorial	
Male	25 - 20 = 5	25 - 30 = -5	50
Female	35 - 40 = -5	65 - 60 = 5	100
Total	60	90	150

- Square of difference $(f_o - f_e)^2$
 - $5^2 = 25, (-5)^2 = 25, (-5)^2 = 25, 5^2 = 25$
- Formula for $X^2 = \Sigma [(f_o - f_e)^2 / (f_e)]$
 - So $X^2 = 25/20 + 25/30 + 25/40 + 25/60$
 - $X^2 = 1.25 + 0.83 + 0.63 + 0.42 = 3.13$
 - Degree of freedom = (no. Of column - 1) x (no. of row - 1)
 - Degree of freedom = $(2-1) \times (2-1) = 1$
 - Critical value = 3.84 when $\alpha = 0.05 \rightarrow$ refer to X^2 statistic table
- Interpretation
 - Calculated values is less than critical value
 - Therefore hypothesis is accepted (do not reject the null hypothesis), there is no difference
- Conclusion
 - There is no relationship between preference of teaching method and gender of the students

4. Chi-squared test for variance

- This is a test of the null hypothesis that the population variance is σ^2 .
- We have a sample of size n and we compute an unbiased estimate of the population variance s^2 using divisor $n-1$.
- The distribution used is X^2 statistics is $(n-1)^2 s^2 / \sigma^2$.
- We assume that the population is normally distributed
- For 95% level of confidence probability level are within 97.5% and 2.5%
- Find out the critical region in MS Excel CHINV (p, df)