Case Study—Chi-Squared Test

- A city has a newly opened nuclear plant, and there are families staying dangerously close to the plant. A health safety officer wants to take this case up to provide relocation for the families that live in the surrounding area. To make a strong case, he wants to prove with numbers that an exposure to radiation levels is leading to an increase in diseased population. He formulates a contingency table of exposure and disease.
- Does the data suggest an association between the disease and exposure?

	Disease		**Total**
Exposure	Yes	No	
Yes	37	13	50
No	17	53	70
Total	54	66	120

© Copyright 2015, Simplilearn. All rights reserved.

Case Study—Chi-Squared Test (contd.)

Steps:

- Calculate the number of individuals of exposed and unexposed groups expected in each disease category (yes and no) if the probabilities were the same.
  - If there were no effect of exposure, the probabilities should be same and the chi-squared statistic would have a very low value.

Proportion of population exposed = (50/120) = 0.42

Proportion of population not exposed = (70/120) = 0.58

Thus, expected values:

Population with disease = 54

Exposure Yes : 54 * 0.42 = 22.5

Exposure No : 54 * 0.58 = 31.5

Population without disease = 66

Exposure Yes : 66 * 0.42 = 27.5

Exposure No : 66 * 0.58 = 38.5

© Copyright 2015, Simplilearn. All rights reserved.

Case Study—Chi-Squared Test (contd.)

simplilearn

- Calculate the Chi-squared statistic

$$\chi 2 = \Sigma \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

$$= \frac{(37-22.5)^2}{22.5} + \frac{(13-27.5)^2}{27.5} + \frac{(17-31.5)^2}{31.5} + \frac{(53-38.5)^2}{38.5}$$

$$= 29.1$$

- Calculate the degrees of freedom :

  (Number of rows – 1) X (Number of columns – 1)

  df = (2 – 1) X (2 – 1) = 1

- Calculate the p-value from the chi-squared table

  For chi-squared value 29.1 and degrees of freedom = 1, from the table, p-value is < 0.001

- Interpretation: There is 0.001 chance of obtaining such discrepancies between expected and observed values if there is no association

- Conclusion : There is an association between the exposure and disease.

© Copyright 2015, Simplilearn. All rights reserved.

Case Study—One Way ANOVA

- Marks obtained in the same subject by 3 students belonging to three different schools are given below.
- Does the data suggest any association between schools and marks?

School	A	B	C
Marks	82	83	38
	83	78	59
	97	68	55

Basic Idea : Partition the total variation in the data into the variance between groups and variance within groups.

© Copyright 2015, Simplilearn. All rights reserved.

Case Study—One Way ANOVA (contd.)

Steps:

- Calculate the means
  - School A : mean(82,83,97) = 87.3
  - School B : mean(83,78,68) = 76.3
  - School C : mean(38,59,55) = 50.6

- Calculate the grand mean
  - Grand mean = mean(82,83,97,83,78,68,39,59,55) = 71.4

- Calculating the variations
  - Sum of Squared Deviations about the grand mean, across all observed values : SS_{Total} = 2630.2
  - Sum of Squared Deviations of group mean about the grand mean – three group means against the grand mean : $SS_{Between}$ = 2124.2
  - Sum of Squared Deviations of observations within a group about their group mean; added across all groups : SS_{Within} = 506

© Copyright 2015, Simplilearn. All rights reserved.

Case Study—One Way ANOVA (contd.)

simplilearn

- Calculate the degrees of freedom for every variance
  - df_{Total} = Number of observations – 1 = 9 -1 = 8
  - $df_{Between}$ = Number of groups -1 = 3-1 = 2
  - df_{Within} = Number of observations – number of groups = 9-3 = 6
- Calculate the Mean Squared Variances
  - Mean Squared variance between groups : $MS_{Between} = SS_{Between} / df_{Between} = 2124.2/2 = 1062.1$
  - Mean Squared variance within groups : $MS_{Within} = SS_{Within} / df_{Within} = 506/6 = 84.3$
- Calculate the f-statistic
  - F-value : $MS_{Between} / MS_{Within} = 1062.1/84.3 = 12.59$
- Calculate the p-value from the F-table
  - p-value for given f-value 12.59 and degrees of freedom 2 and 6 is 0.007
- Conclusion : Since the p-value is less than alpha, we can conclude by rejecting the null hypothesis, that there is a difference in the marks obtained by students belonging to different groups.

© Copyright 2015, Simplilearn. All rights reserved.