Test of Proportion and Association

Adejumo Ridwan Suleiman

Test of Proportion

Test of Proportion

- Involves categorical values
- Two possible outcomes **Success** or **Failure**
- Population proportion of success is denoted by p.

What is Proportion?

- Fraction of the total population that possesses a certain attribute.
- For example, what is the proportion of men in a given sample of 10 that have systolic high blood pressure above 120.
- I 4 men have systolic high blood pressure above 120, then $p = \frac{4}{10}$.
- Proportion can also be expressed in terms of percentage $p = \frac{4}{10} \cdot 100 = 40$

Sampling distribution of a proportion

- As sample size increases, sampling distribution of $\hat{p}$ becomes approximately normal
- The mean of the sampling distribution is p
- The standard deviation of the sampling disribution is $\sqrt{\frac{p(1-p)}{n}}$

One sample Z- test for population proportion - Assumptions

- Sample is randomly selected
- Observations are dichotomous
- Observations are independent of each other
- Sample size is large

Two sample Z-test for populaion proportion - Assumptions

- Same as that of one sample
- The two samples are independent of each other

Test of Association

What is a Contingency table?

- Display association between two catgorical variables
- One or both of which has more than two possible values

Examples

- Smoking and Lung Cancer

	Have Lung Cancer	**Does not have Lung Cancer**	**Total**
Smoke	20	35	55
Does not smoke	15	18	33
Total	35	53	88

Chi-square test statistic χ^2 - Assumptions

- Samples are randomly selected
- Observations are categorical
- Independence of Observations
- Each cell in the contingency table should have at least an expeced frequency of **5** in at most **20%** of the cells
- No cells should have expected frequency of 0
- Large sample size

Chi-square test statistic limitations

- Does not describe the strength of relationships
- Sensitive to sample size
- Sensitive to low expected frequencies

Fisher's Exact Test

- Used in place of chi-square test when
  - Cell counts are sparse
  - More than **20%** of the cells have expected frequencies of less than **5**
  - Sample size is small $n < 20$