MODULE 4

Two main chi-squared tests:

- **Chi-squared X^2, goodness of fit**
  Determines whether an observed categorical variable follows an expected distribution
  - **Null Hypothesis (H_0)** – The variable follows the expected distribution
  - **Alternative Hypothesis (H_1)** – The variable does not follow the expected distribution
  - $X^2 = \sum \frac{(Observed - Expected)^2}{Expected}$
  - Main steps:
    1. Identify the Null and Alternative Hypotheses
    2. Calculate the chi-square test statistic
    3. Calculate p-value
       Find the area under the X^2 curve
       Degrees of freedom = number of categorical levels - 1
    4. Make a conclusion

- **Chi-squared X^2, test of independence** (aka test of homogeneity)
  Determines whether two categorical variables are associated with each other or not
  - **Null Hypothesis (H_0)** – The variables are independent
  - **Alternative Hypothesis (H_1)** – The variables are not independent
  - $E_{ij} = \frac{R_i C_j}{T}$
  - $expected\ value = \frac{column\ total * row\ total}{overall\ total}$
  - Main steps:
    1. Identify the Null and Alternative Hypotheses
    2. Calculate the chi-square test statistic
    3. Calculate p-value
       Find the area under the X^2 curve
       Degrees of freedom = (no. of rows – 1) (no. of columns – 1)
    4. Make a conclusion

Analysis of Variance (ANOVA)
A group of statistical techniques that test the difference of means between three or more groups

ANOVA assumptions:

1. The dependent values within each group should be normally distributed
2. The variances across groups are equal
3. Observations are independent of each other

One-way ANOVA
Compares the means of one continuous dependent variable based on three or more groups of one categorical variable

- **Null Hypothesis (H_0)** – The means of each group are equal
- **Alternative Hypothesis (H_1)** – The means of each group are not equal

Steps:

1. Calculate group means and grand mean (overall mean)
2. Calculate the sum of squares between groups (SSB) and the sum of squares within groups (SSW)

$$SSB = \sum_{g=1} n_g(M_g - M_G)^2$$

Where: n_g = number of sample in g^{th} group, M_g = mean of the g^{th} group, M_G = grand mean

$$SSW = \sum_{g=1} \sum_{i=1} (x_{gi} - M_g)^2$$

Where: x_{gi} = sample I of the g^{th} group

3. Calculate mean squares for both SSB and SSW

$$Mean\ Squares\ Between\ Groups\ (MSSB) = \frac{SSB}{k-1}$$

Where: k = number of groups

$$Mean\ Squares\ Within\ Groups (MSSW) = \frac{SSW}{n - k}$$

Where: $\mathrm{n} = \text{total number of samples in all groups}$

4. Compute the F-statistic
   F-statistic is the ratio of MSSB over MSSW
5. Use the F-distribution and F-statistic to get the p-value
6. Make a conclusion

Two-way ANOVA
Compares the means of one continuous dependent variable based on three or more groups of two categorical variables

Post hoc test
Performs a pairwise comparison between all available groups while controlling for the error rate

Tukey's HSD (honestly significantly different)
Compare all the pairs of groups and determine which pairs are different from one another while controlling the fact that running multiple hypothesis tests all at once

ANCOVA (Analysis of covariance)
A statistical technique that tests the difference of means between three or more groups while controlling for the effects of covariates (variables irrelevant to the test)

Covariate – an independent variable that can influence the outcome of a statistical test but it is not direct interest

MANOVA (Multivariate analysis of variance)
Compares how two or more continuous outcome variables vary according to categorical independent variables

- One-way MANOVA
  One categorical independent variable
- Two-way MANOVA
  Two categorical independent variables

MANCOVA (Multivariate analysis of covariance)
Compares how two or more continuous outcome variables vary according to categorical independent variables while controlling for covariates