# Advanced Statistics

Analysis of Variance(ANOVA)

Dr. P.K.Viswanathan(Professor Analytics)

# ANOVA Basics

- This technique is part of the domain called “Experimental Designs”.
- This helps in establishing in a precise fashion the Cause - Effect relation amongst variables.
- From the Statistical Inference Point of View, ANOVA is an extension of independent t test for testing the equality of two population means.
- When we have to compare more than two population means, we use ANOVA
- Typically, the null hypothesis( $H_0$ ) is as under:
- $H_0$ : $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots\ldots = \mu_k$ for testing the equality of Population Means for k populations

# ANOVA-One Way Classification

## *Assumptions involved in using ANOVA*

- The samples drawn from different populations are independent and random.
- The response variables of all the populations are normally distributed.
- The variances of all the populations are equal.

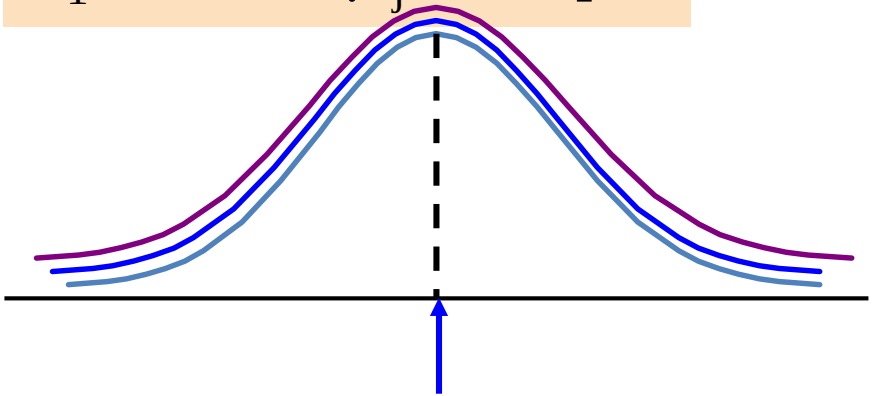
# Hypotheses of One-Way ANOVA

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$
  - All population means are equal
- $H_1 :$ Not all of the population means are equal
  - For at least one pair, the population means are unequal

# One-Way ANOVA
# Null Hypothesis($H_0$=True)

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

$$H_1 : \text{Not all } \mu_j \text{ are equal}$$

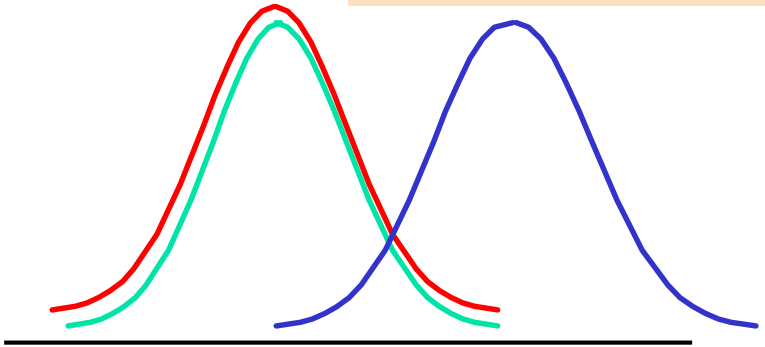


$$\mu_1 = \mu_2 = \mu_3$$

# One-Way ANOVA
# Alternative Hypothesis($H_1$=True)

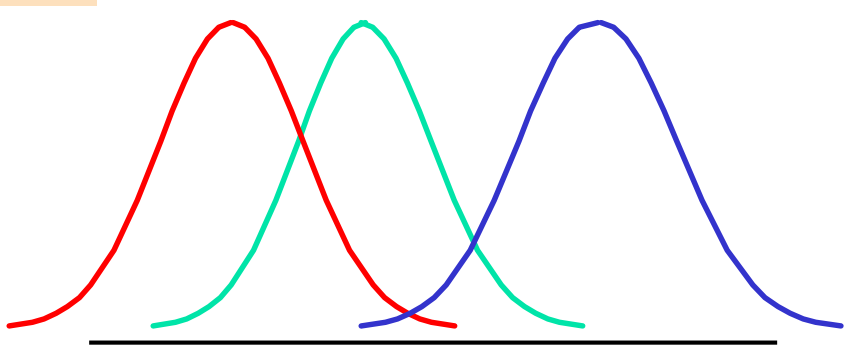
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

$$H_1 : \text{Not all } \mu_j \text{ are equal}$$



$\mu_1 = \mu_2 \neq \mu_3$

or

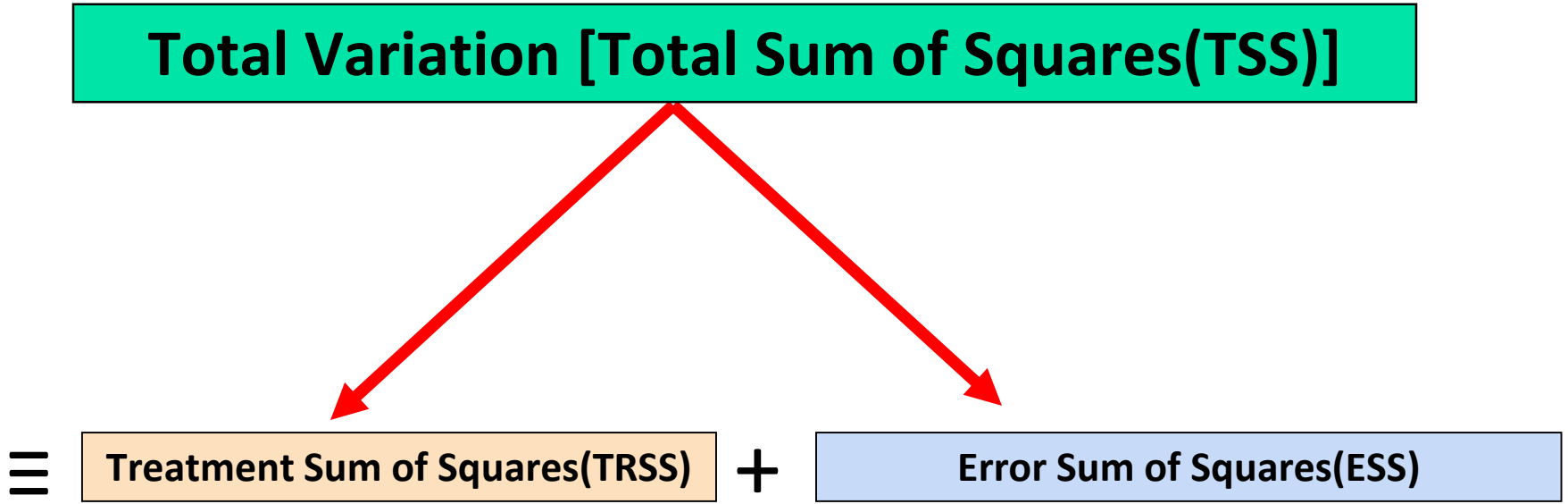


$\mu_1 \neq \mu_2 \neq \mu_3$

# ANOVA Basics

- The beauty of ANOVA is that it performs the test of equality of more than two population means by actually analyzing the variance.
- In simple terms, ANOVA decomposes the total variation into components of variation. That is, explaining the changes in the response variable caused by these components.
- To put it succinctly, the total sum of squares is equal to the sum of squares due to causes.

# Partition of Total Variation(Information Content



# ANOVA-One Way Classification-Example

- A supermarket is interested in knowing whether it should go for a quarter-page, half-page, or a full-page advertisement for a Product.
- In order to choose the size of the advertisement that will bring in the most store traffic, the supermarket can use ANOVA technique.
- Here, you are trying to establish a cause-effect relationship between store traffic and the various sizes of advertisement.

# ANOVA-One Way Classification

## *How One-Way Classification Works in Practice?*

- **Total Sum of Squares ≡ Treatment Sum of Squares + Error Sum of Squares**.

- The word treatment is generic and as such may denote different methods, machines, different advertisement copy platforms, different strategies, different brands and the like.

- The variation in sum of squares of the response variable (dependent variable) is caused only by treatment and any thing unexplained by the treatment is attributed to error term.

# One Way ANOVA- Application

Sporting goods manufacturing company wanted to compare the distance traveled by golf balls produced using four different designs. Ten balls were manufactured with each design and were brought to the local golf course for the club professional to test. The order in which the balls were hit with the same club from the first tee was randomized so that the pro did not know which type of ball was being hit. All 40 balls were hit in a short period of time, during which the environmental conditions were essentially the same. The results (distance traveled in yards) for the four Designs are stored in **Golfball.csv**

At the 0.05 level of significance, is there evidence of a difference in the mean distances traveled by the golf balls with different designs?

Problem 10.64, Chapter 10, Page 381 of the Textbook
Business Statistics- A First Course 7th Edition Pearson Education Indian Edition

# Anova Output

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Design | 3 | 2990.99 | 997.00 | 53.03 | 2.73E-13 |
| Residuals | 36 | 676.82 | 18.80 | | |

# TukeyHSD Test

| | diff | lwr | upr | p adj |
|---|---|---|---|---|
| Design2-Design1 | 11.9020 | 6.6795 | 17.1245 | 2.65E-06 |
| Design3-Design1 | 19.9740 | 14.7515 | 25.1965 | 1.64E-11 |
| Design4-Design1 | 22.0080 | 16.7855 | 27.2305 | 8.89E-13 |
| Design3-Design2 | 8.0720 | 2.8495 | 13.2945 | 0.00103 |
| Design4-Design2 | 10.1060 | 4.8835 | 15.3285 | 4.51E-05 |
| Design4-Design3 | 2.0340 | -3.1885 | 7.2565 | 0.72211 |

# Interesting Application of Two Factor Anova/Ancova
# Testing The Effects of Price and Advertising

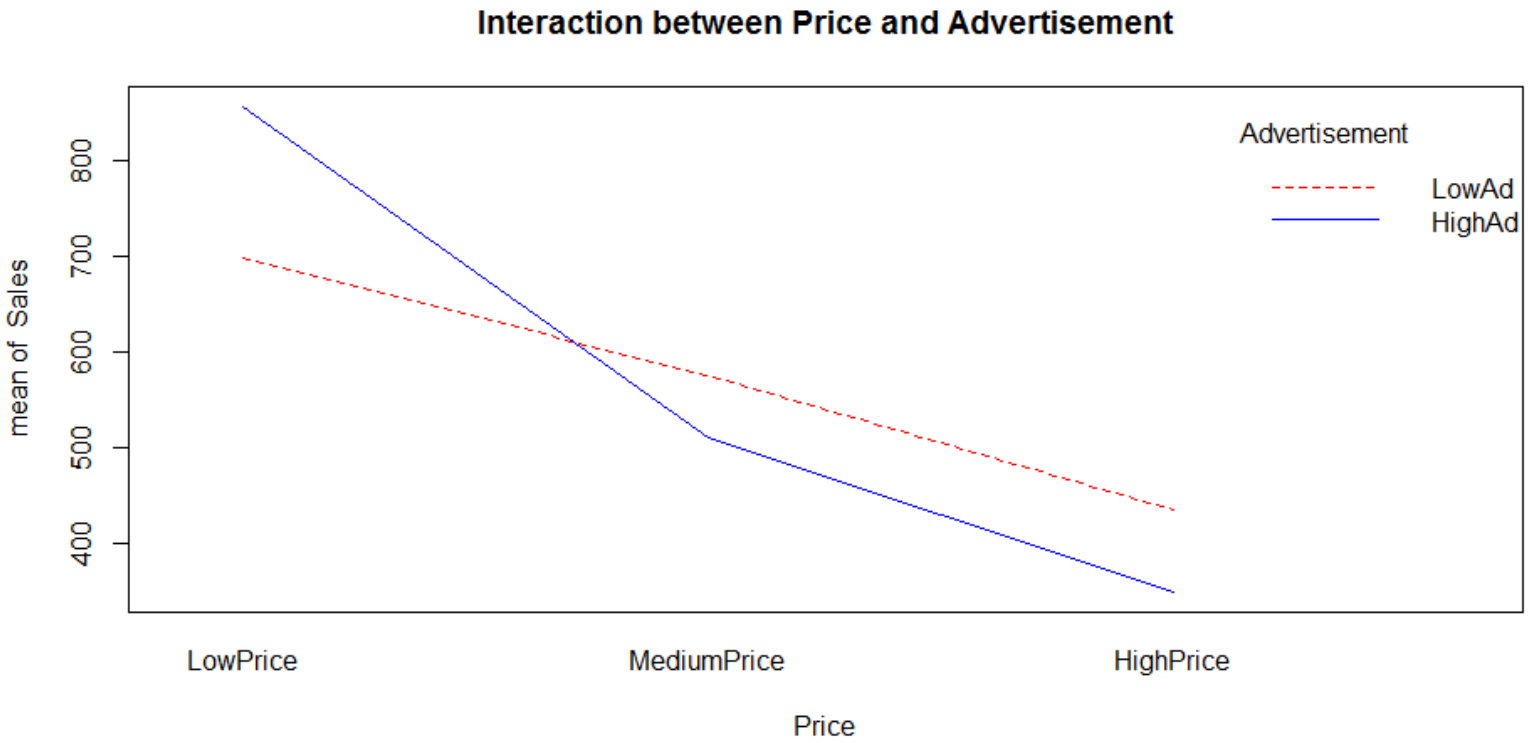
Newfood Product Management opted to conduct  a market test

Experiment using a balanced two factor design with three levels of price(low, medium, high) and two levels of advertising (low and high). Each combination of price and advertising was used in four different stores resulting in a total of 24 observations. The data are given in paul-newfood.csv. Perform anova/ancova and interpret the results.

# Newfood-Anova-Main Effect

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Price | 2 | 600412.5833 | 300206.2917 | 13.6640 | 0.0002 |
| Advertisement | 1 | 32.6667 | 32.6667 | 0.0015 | 0.9696 |
| Residuals | 20 | 439412.5833 | 21970.6292 | | |

# Newfood-Interaction Effect



# Newfood-Anova-Interaction Effect

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Price | 2 | 600412.5833 | 300206.2917 | 14.7819 | 0.0002 |
| Advertisement | 1 | 32.6667 | 32.6667 | 0.0016 | 0.9685 |
| Price:Advertisement | 2 | 73850.0833 | 36925.0417 | 1.8182 | 0.1909 |
| Residuals | 18 | 365562.5000 | 20309.0278 | | |

# Newfood-Ancova-Interaction Effect

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| StoreSize | 1 | 191540.9780 | 191540.9780 | 18.8113 | 0.0004 |
| Price | 2 | 501543.2182 | 250771.6091 | 24.6284 | 0.0000 |
| Advertisement | 1 | 128371.9891 | 128371.9891 | 12.6075 | 0.0025 |
| Price:Advertisement | 2 | 45304.1585 | 22652.0793 | 2.2247 | 0.1386 |
| Residuals | 17 | 173097.4895 | 10182.2053 | | |

# Newfood-Ancova-Interaction Effect-Adjusted for store Size

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Price | 2 | 386945.4676 | 193472.7338 | 19.0011 | 0.0000 |
| Advertisement | 1 | 128371.9891 | 128371.9891 | 12.6075 | 0.0025 |
| Price:Advertisement | 2 | 45304.1585 | 22652.0793 | 2.2247 | 0.1386 |
| Residuals | 17 | 173097.4895 | 10182.2053 | | |