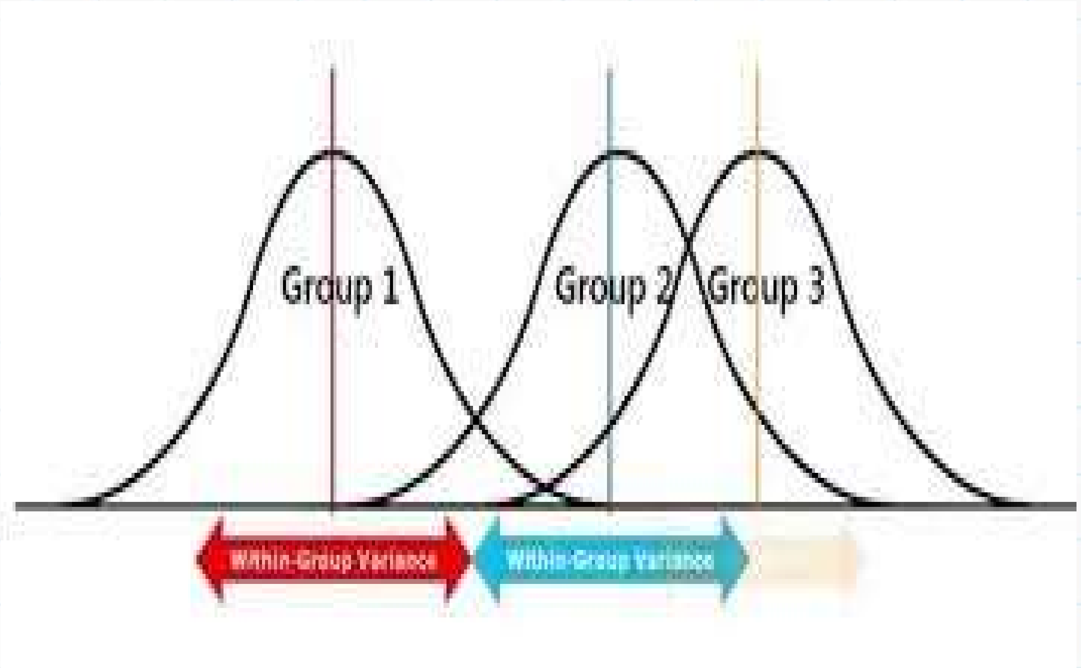
ANOVA

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent groups.



ANOVA Table

Source	Sum of squares	Degree of Freedom	Mean squares	F	F-test
Treatment	SS_T	k-1	$MS_T = \frac{SS_T}{k-1}$	$F = \frac{MS_T}{MS_E}$	$F > F_{\alpha,k-1,N-k}$?
Error	SS_E	N-k	$MS_E = \frac{SS_E}{N-k}$		
Total	TotalSS	N-1			

MANOVA is short

for

Multivariate ANalysis Of VAriance

MANOVA

- Multivariate analysis of variance (MANOVA) is an extension of analysis of variance (ANOVA).
- MANOVA is just an ANOVA with several dependent variables.

Analysis of Variance

$$Y_1 = X_1 + X_2 + X_3 + \cdots + X_n$$

(metric) (nonmetric)

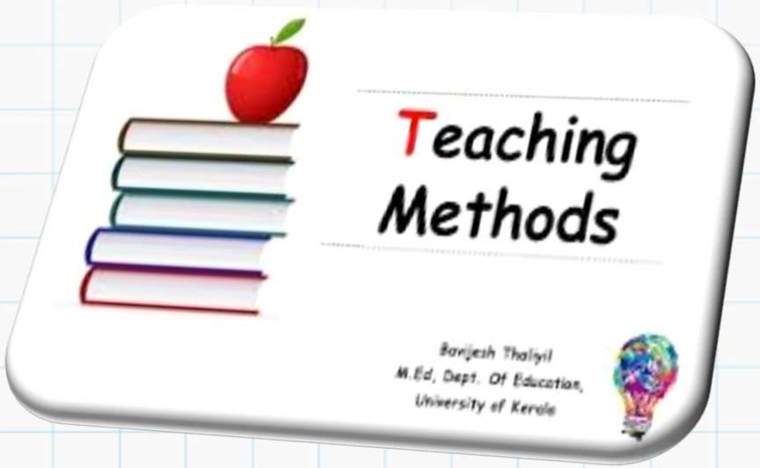
Multivariate Analysis of Variance

$$Y_1 + Y_2 + Y_3 + \cdots + Y_n = X_1 + X_2 + X_3 + \cdots + X_n$$

(metric) (nonmetric)

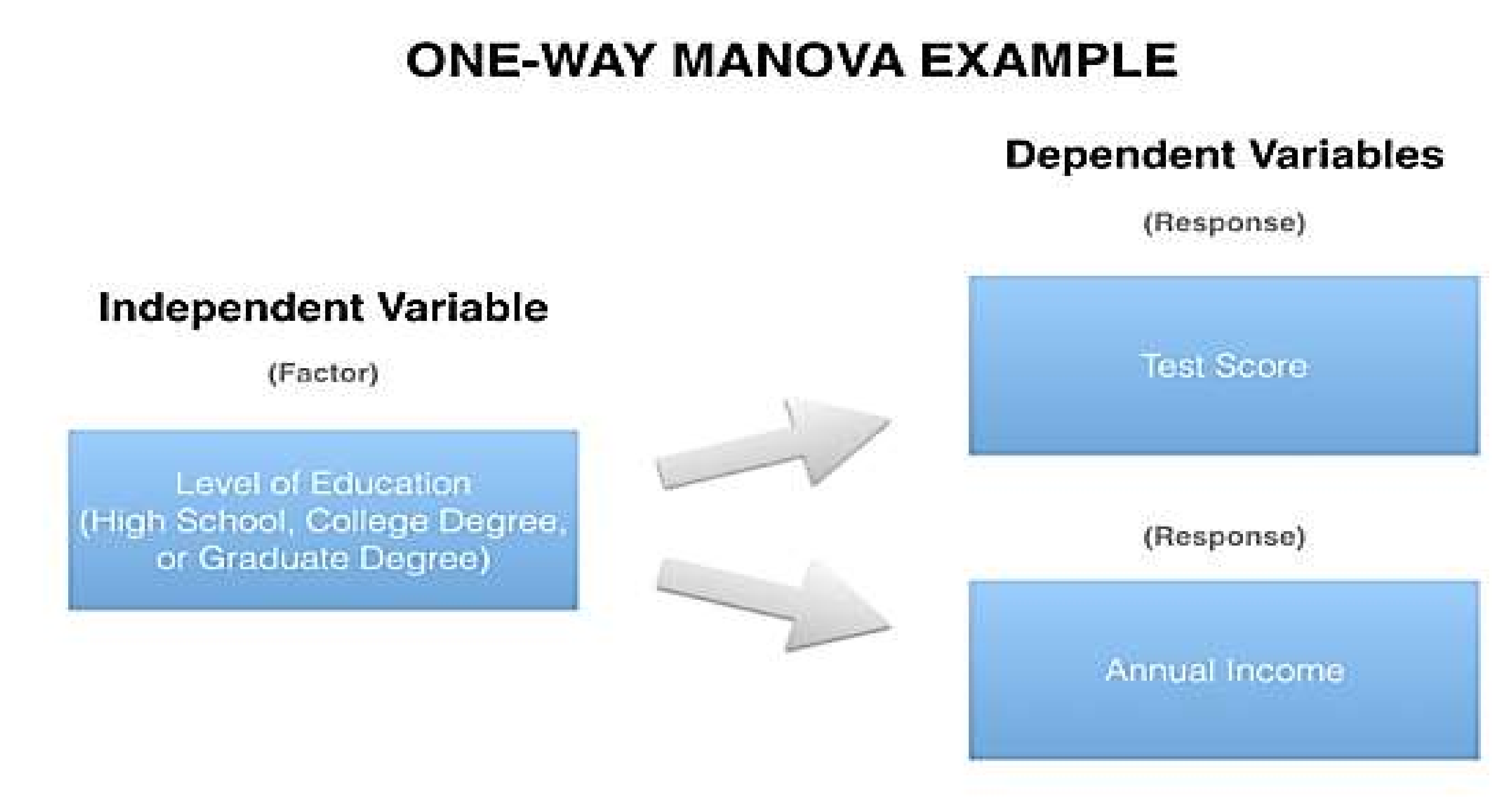
MANOVA

Suppose we are studying three different teaching methods for a course. This variable is our independent variable. We also have student satisfaction scores and test scores. These variables are our dependent variables. We want to determine whether the mean scores for satisfaction and tests differ between the three teaching methods.

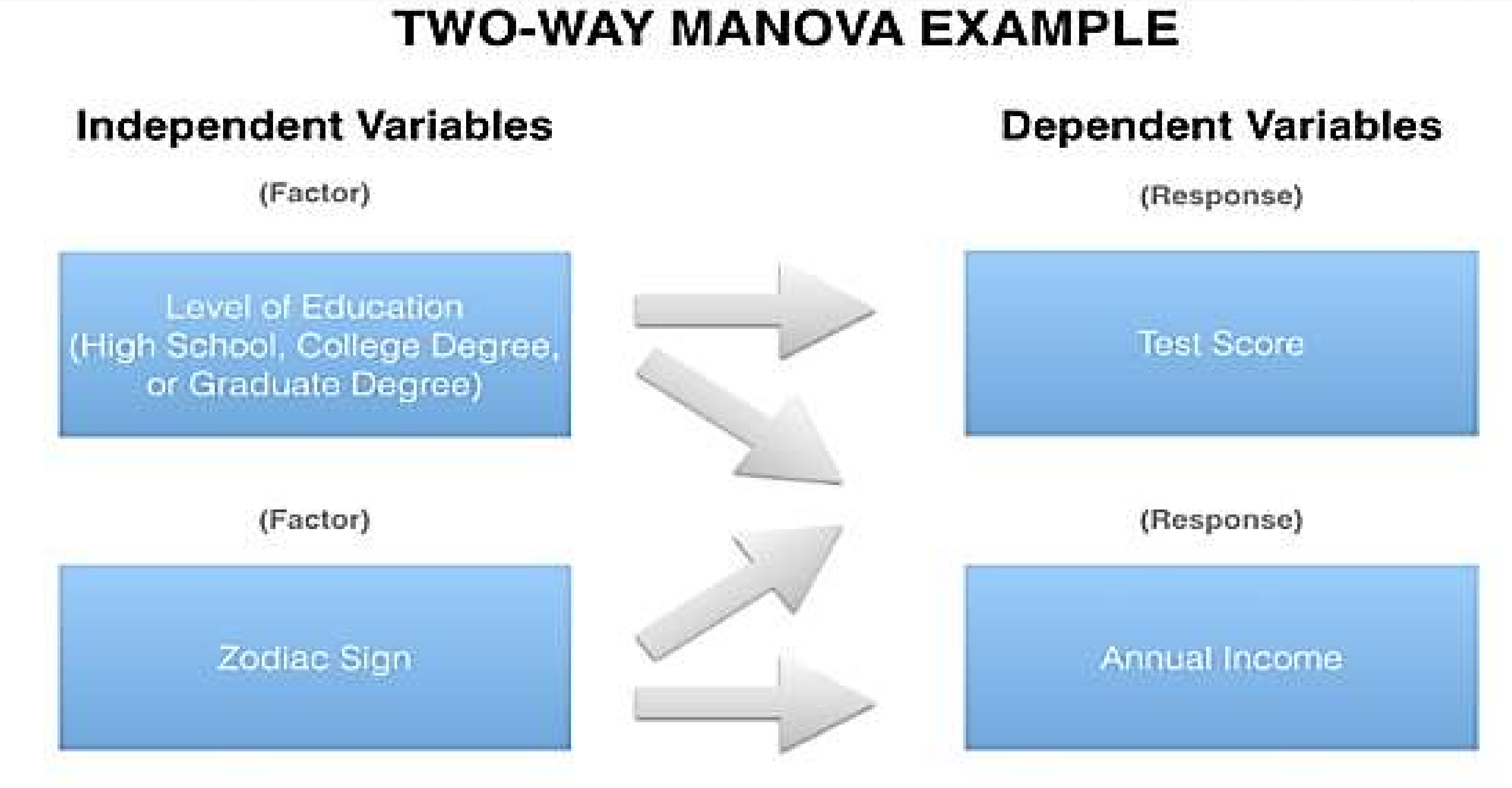


You might wish to test the hypothesis that gender and ethnicity interact to influence a set of job-related outcomes including attitudes toward co-workers, attitudes toward supervisors, feelings of belonging in the work environment, and identification with the corporate culture

MANOVA


ONE-WAY MANOVA EXAMPLE
Dependent Variables
(Response)
Test Score
Independent Variable
(Factor)
Level of Education
(High School, College Degree,
or Graduate Degree)
(Response)
Annual Income

MANOVA


TWO-WAY MANOVA EXAMPLE
Independent Variables
Dependent Variables
(Factor)
(Response)
Level of Education (High School, College Degree, or Graduate Degree)
Test Score
(Factor)
(Response)
Zodiac Sign
Annual Income

MANOVA

ANOVA

$H_0 : \mu_1 = \mu_2 = \ \ldots\ \mu_k$ vs $H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \cdots \neq \mu_k$

Null hypothesis (H_0) = all the group means are equal, that is, they come from the same population.

MANOVA

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{bmatrix} = \ldots\ldots\ldots = \begin{bmatrix} \mu_{1k} \\ \mu_{2k} \\ \vdots \\ \mu_{pk} \end{bmatrix}$$

Null hypothesis (H_0) = all the group mean vectors are equal, that is, they come from the same population.

μ_{pk} = means of variable p, group k

1. **Wilks Lambda**

$$\Lambda^* = \frac{|\mathbf{E}|}{|\mathbf{H}+\mathbf{E}|}$$

Here, the determinant of the error sums of squares and cross products matrix **E** is divided by the determinant of the total sum of squares and cross products matrix **T** = **H** + **E**. If **H** is large relative to **E**, then |**H** + **E**| will be large relative to |**E**|. Thus, we will reject the null hypothesis if Wilks lambda is small (close to zero).

2. **Hotelling-Lawley Trace**

$$T_0^2 = trace(\mathbf{HE}^{-1})$$

Here, we are multiplying **H** by the inverse of **E**; then we take the trace of the resulting matrix. If **H** is large relative to **E**, then the Hotelling-Lawley trace will take a large value. Thus, we will reject the null hypothesis if this test statistic is large.

3. **Pillai Trace**

$$V = trace(\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1})$$

Here, we are multiplying **H** by the inverse of the total sum of squares and cross products matrix **T** = **H** + **E**. If **H** is large relative to **E**, then the Pillai trace will take a large value. Thus, we will reject the null hypothesis if this test statistic is large.

4. **Roy's Maximum Root: Largest eigenvalue of $\mathbf{HE}^{-1}$**

Here, we multiply **H** by the inverse of **E**, and then compute the largest eigenvalue of the resulting matrix. If **H** is large relative to **E**, then the Roy's root will take a large value. Thus, we will reject the null hypothesis if this test statistic is large.

Working Procedure:

Step I: Arrange the observations of all groups in matrix form.

Step II: Set the Hypothesis.

Step III: Find the row wise average and grand average.

Step IV: For each observation of all groups, calculate:

$$\underbrace{x_{ni}}_{\text{Observation}} = \underbrace{\bar{x}}_{\text{Mean}} + \underbrace{[\bar{x}_n - \bar{x}]}_{\text{Treatment effect}} + \underbrace{[x_{ni} - \bar{x}_n]}_{\text{Residual}}$$

Step V: Write all observations of all groups in matrix form:

Observation=Mean + Treatment effect+ Residual

Working Procedure:

Step VI: Calculate sum of squares for observation, mean, treatment effect and residuals for all groups i.e.

SS_{obs}, SS_{mean}, SS_{te} and SS_{res}

[Note: $SS_{obs} = SS_{mean} + SS_{te} + SS_{res}$]

Step VII: Calculate Corrected sum of squares for all groups:

$SS_t = SS_{obs} - SS_{mean}$

Step VIII: Obtain cross product contributions, we proceed row by row in the arrays for the two variables for observation, mean, treatment and residuals.

Step IX: Calculate total corrected cross product=Observed cross product- Mean Cross product.

Working Procedure:

Step X: Find B stands for variances of observations between samples

$$B = \begin{bmatrix} SS_{tr}\, group1 & \text{Total Cross Product} \\ \text{Total Cross Product} & SS_{tr}\, group2 \end{bmatrix}$$

Step XI: Find W - stands for variances of observations within sample.

$$W = \begin{bmatrix} SS_{res}\, group1 & \text{Residual Cross Product} \\ \text{Residual Cross Product} & SS_{res}\, group2 \end{bmatrix}$$

Working Procedure:

Step XII: Manova Table:

<table>
<tr><th>Source of Variation</th><th>Matrix of SS and cross product</th><th>Degree of freedom</th><th>Wilk's lambda [equivalent form of the F-test]</th><th>Test Statistics</th></tr>
<tr><td>Treatment</td><td>B</td><td>g-1</td><td rowspan="3">$\wedge^* = \frac{|W|}{|W+B|}$
$=$</td><td rowspan="3">$\Rightarrow \left[\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right]\left[\frac{N-g-1}{g-1}\right]$</td></tr>
<tr><td>Residual</td><td>W</td><td>N-g</td></tr>
<tr><td>Total(Corrected)</td><td>B+W</td><td>N-1</td></tr>
</table>

Working Procedure:

Step XIII: Find the tabulated/Critical value for F test with degree of freedom $\overset{\nu_1}{2(g-1)}$ and $\overset{\nu_2}{2(N-g-1)}$ at LOS 5 %.

Step XIV: Comparison and Conclusion:

$F_{cal} > F_{tab}$

If Test Statistic > Critical Value then reject Null Hypothesis

If Test Statistic < Critical Value then do not reject Null Hypothesis

$F_{cal} < F_{tab}$

Problem 1:

Consider the following dataset where there are 2 dependent variables and 3 groups.

		y_1	y_2	
	Group	**For DV 1**	**For DV 2**	
x_1	Group 1	9, 6, 9	3,2,7	$n_1 = 3$
x_2	Group 2	0,2	4,0	$n_2 = 2$
x_3	Group 3	3,1,2	8,9,7	$n_3 = 3$

Whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

Step I: Arrange the observations of all groups in matrix form.

$$\begin{pmatrix} \begin{bmatrix} 9 \\ 3 \end{bmatrix} & \begin{bmatrix} 6 \\ 2 \end{bmatrix} & \begin{bmatrix} 9 \\ 7 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 4 \end{bmatrix} & \begin{bmatrix} 2 \\ 0 \end{bmatrix} & \\ \begin{bmatrix} 3 \\ 8 \end{bmatrix} & \begin{bmatrix} 1 \\ 9 \end{bmatrix} & \begin{bmatrix} 2 \\ 7 \end{bmatrix} \end{pmatrix} \begin{array}{l} \longrightarrow \text{Group } 1 \\ \longrightarrow \text{Group } 2 \\ \longrightarrow \text{Group } 3 \end{array}$$

$$\bar{x}_1 = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\bar{x}_3 = \begin{bmatrix} 2 \\ 8 \end{bmatrix} \quad \bar{x} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

Step III: Find the row wise average and grand average.

$$\bar{x}_1 = \begin{bmatrix} 9+6+9/3 \\ 3+2+7/3 \end{bmatrix} ; \quad \bar{x}_2 = \begin{bmatrix} 0+2/2 \\ 4+0/2 \end{bmatrix} ; \quad \bar{x}_3 = \begin{bmatrix} 3+1+2/3 \\ 8+9+7/3 \end{bmatrix}$$

$$x = \begin{bmatrix} \dfrac{9+6+9+0+2+3+1+2}{8} \\ 3+2+7+4+0+8+9+7/8 \end{bmatrix}$$

Step IV: For each observation of all groups, calculate:

$$x_{ni} = \overline{x} + [\overline{x}_n - \overline{x}] + [x_{ni} - \overline{x}_n]$$

Observation Mean Treatment effect Residual

for DV 1

G I → 9 6 9 ⇒ x_{11} x_{12} x_{13}

G II → 0 2 ⇒ x_{21} x_{22}

G III → 3 1 2 ⇒ x_{31} x_{32} x_{33}

$3 = x_{31}$, $\overline{x} = 4$, $\overline{x}_3 = 2$

$3 = x_{31} = 4 + [2 - 4] + [3 - 2]$

$= 4 + [-2] + [1] \quad = 3$

$\overline{x}_1 = 8$, $\overline{x}_2 = 1$, $\overline{x}_3 = 2$

Step V: Write all observations of all groups in matrix form:

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 & \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 & \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & \\ 1 & -1 & 0 \end{pmatrix}$$

obs²   mean   treatment effect   residual

Step VI: Calculate sum of squares for observation, mean, treatment effect and residuals for all groups i.e. SS_{obs}, SS_{mean}, SS_{te} and SS_{res}

$$SS_{obs} = 9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2 = 216$$

$$SS_{mean} = 8(4^2) = 128$$

$$SS_{te} = 3(4^2) + 2(-3)^2 + 3(-2)^2 = 78$$

$$\boxed{\begin{aligned} SS_{obs} &= SS_{mean} + SS_{te} + SS_{r} \\ &= 128 + 78 + 10 \\ &= 216 \end{aligned}}$$

$$SS_{res} = 1^2 + (-2)^2 + (1)^2 + (1)^2 + (-1)^2 = 10$$

Step VII: Calculate Corrected sum of squares for all groups:

$SS_t = SS_{obs} - SS_{mean}$

$= 216 - 128$

$SS_t = 88$