ANOVA

Analysis Of Variance

Analysis of Variance

- Introduction
- Types of ANOVA

Introduction

Why ANOVA ?

- Using various tests for Hypothesis, we have been comparing two populations.
  - Independent Samples t-test (random)
  - Matched sample t-test (paired)

- However, this limit us to the comparison of two populations only.

- If you wish to compare the means of more than two populations each containing several levels or subgroups we use ANOVA

- **AN**alysis **O**f **VA**riance

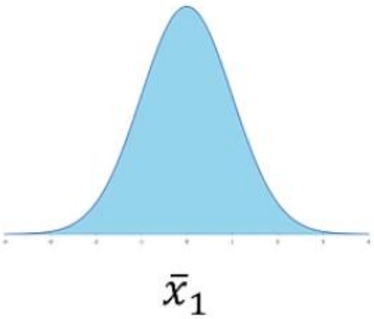
Concept of ANOVA

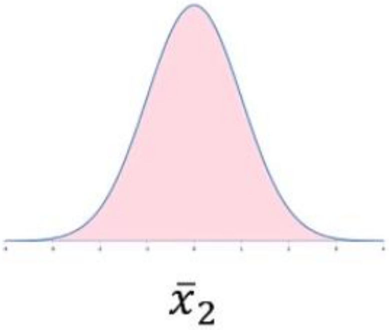
ANOVA is used when we wish to compare more than two populations/sample.

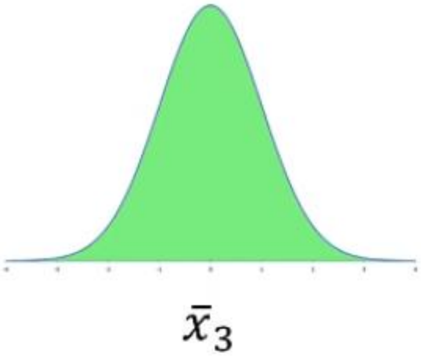
Suppose we want to compare THREE sample means to see if a difference exists among them or not.

Basically, What we are asking is:

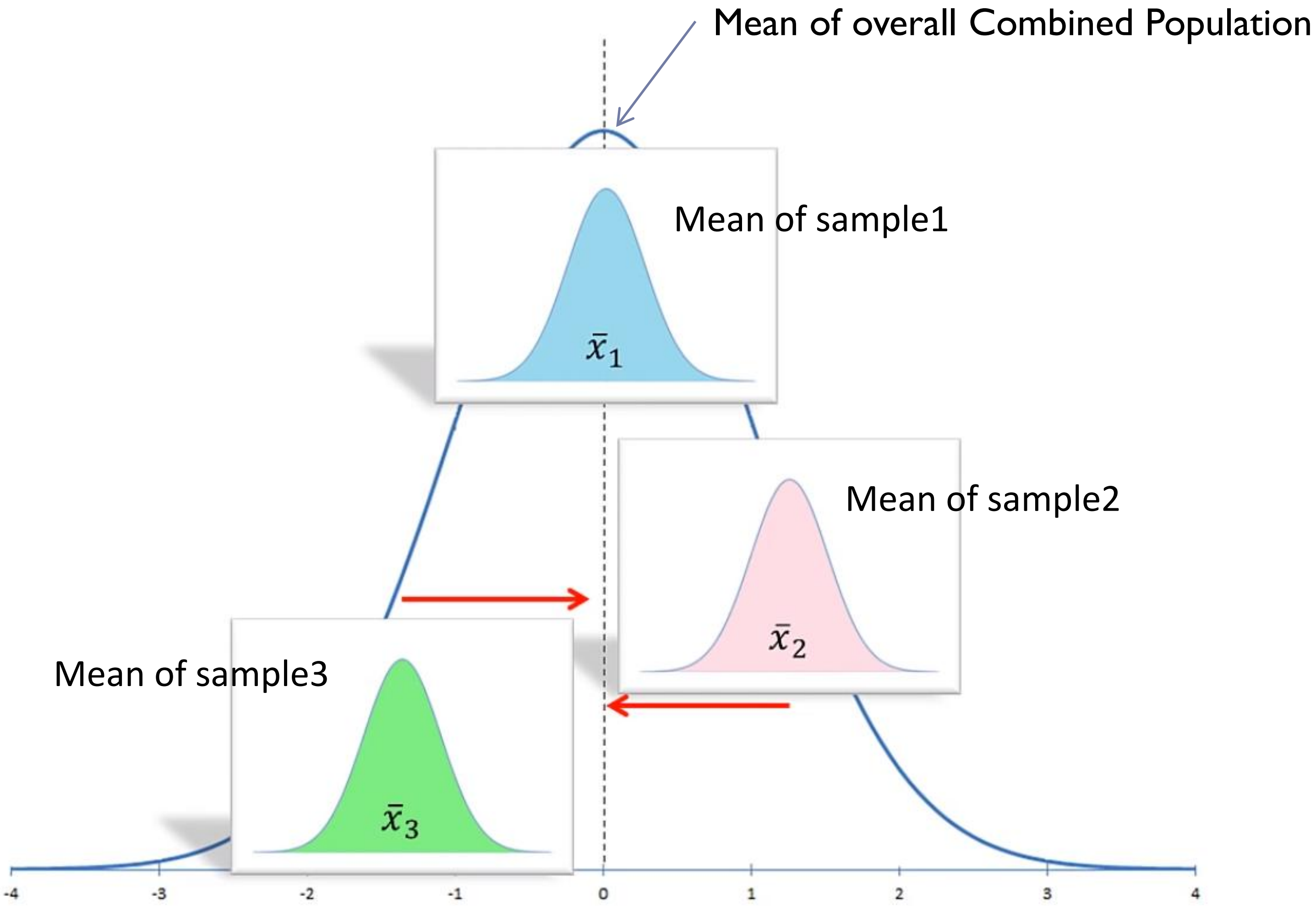
- Do all of these means comes from a common population ?
- Or they come from different/unique populations ?







Concept of ANOVA



Mean of overall Combined Population

Mean of sample1

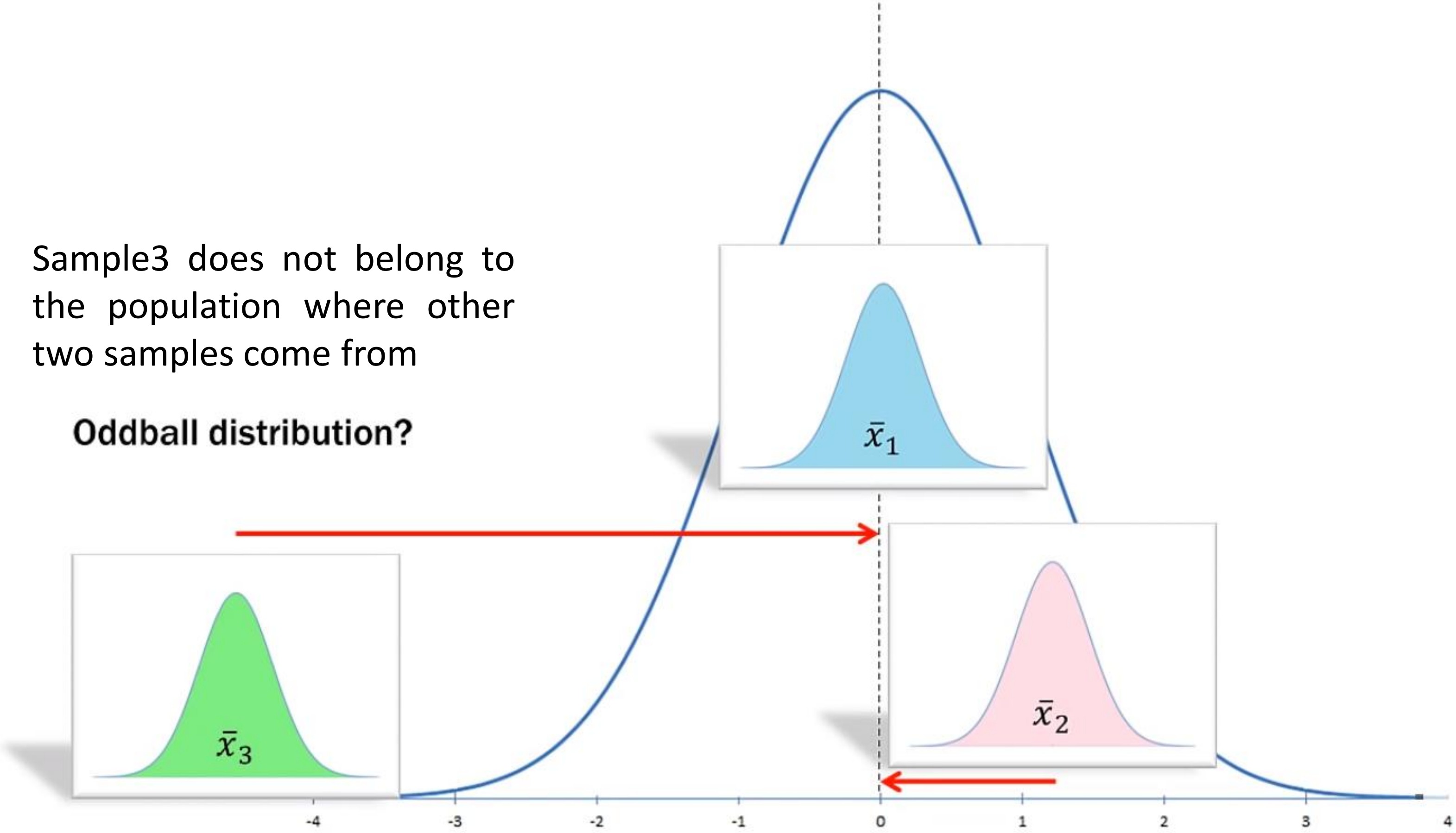
Mean of sample2

Mean of sample3

Concept of ANOVA

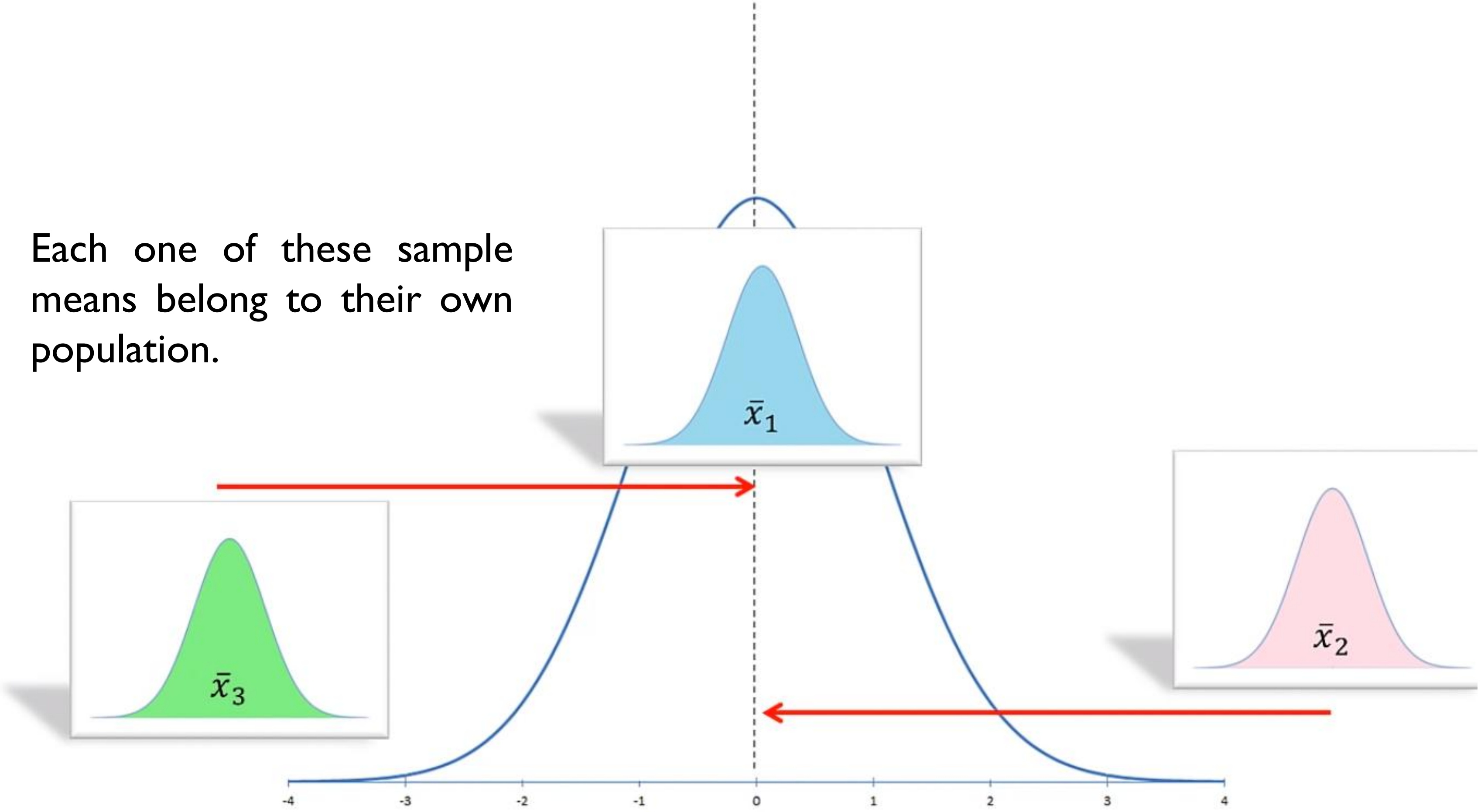
Sample3 does not belong to the population where other two samples come from

Oddball distribution?



Concept of ANOVA

Each one of these sample means belong to their own population.

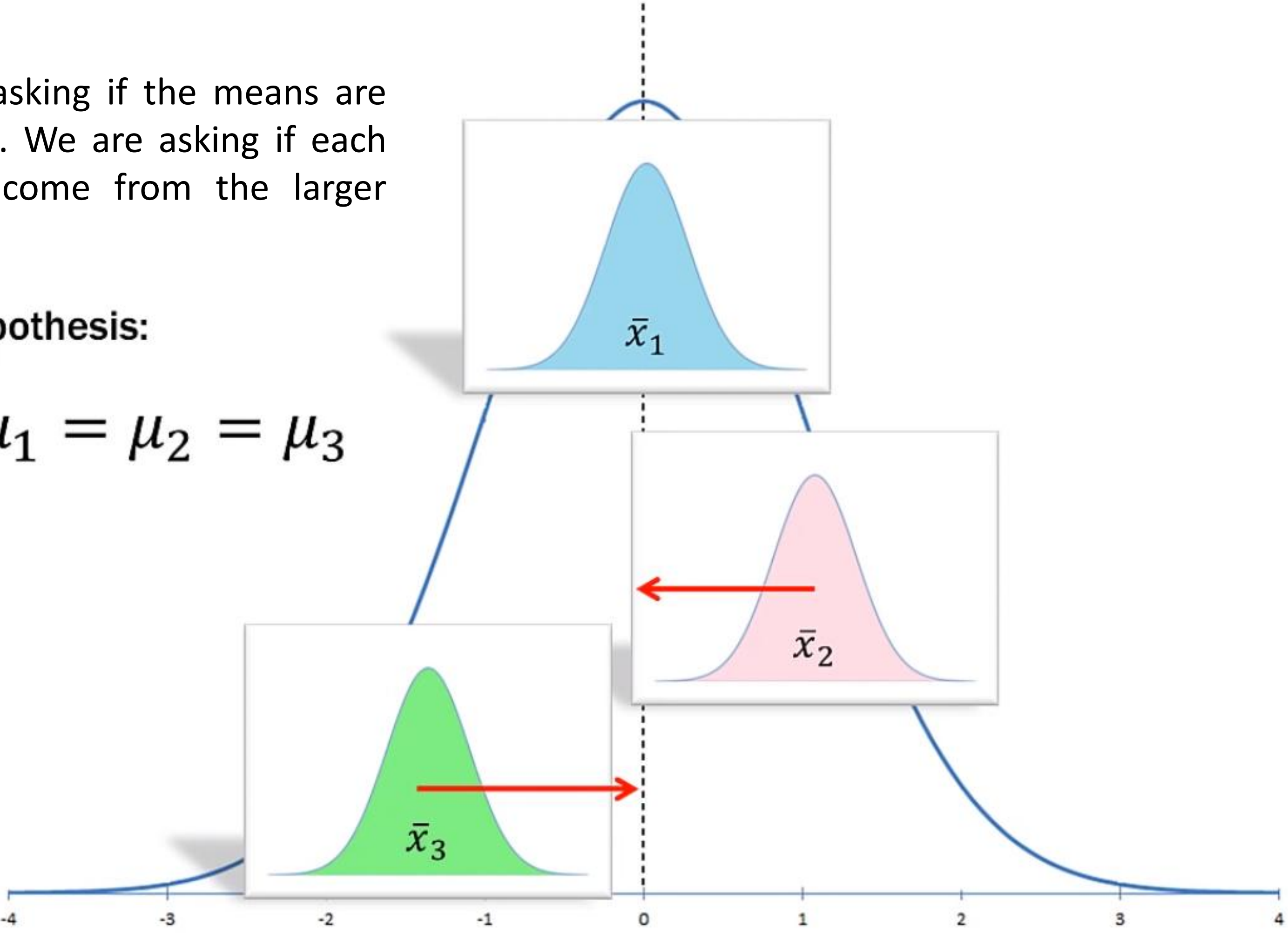


Concept of ANOVA

We are not asking if the means are exactly equal. We are asking if each mean likely come from the larger population

Null hypothesis:

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

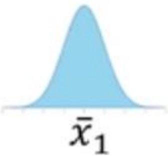


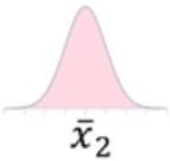
Concept of ANOVA

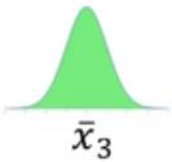
Why are we using ANOVA and why not multiple t-tests like we did in case of single sample and two sample ?

It is simple because the error rate compounds if we use t-test for all pairs. Let's see how ?

We have three samples:







The Pairs for t-test will be: $H_0: \bar{x}_1 = \bar{x}_2; \alpha = .05$ $H_0: \bar{x}_1 = \bar{x}_3; \alpha = .05$ $H_0: \bar{x}_2 = \bar{x}_3; \alpha = .05$

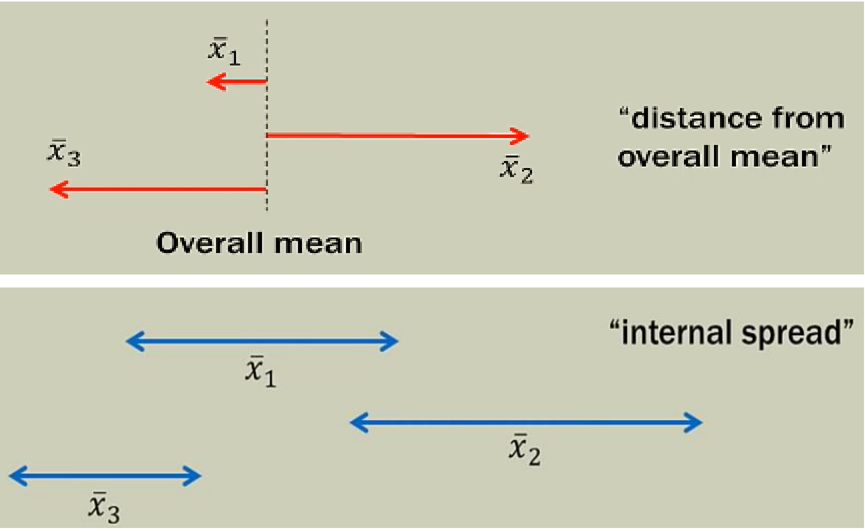
Pairwise Comparison means : **Three t-tests ALL with** α = 0.05
Alpha is Type I error rate (at 95% Confidence Interval)

The error compounds with each test: (0.95)*(0.95)*(0.95) = 0.857
α = 1 – 0.857 = 0.143

Concept of ANOVA

ANOVA is a variability ratio (F Ratio):

$$\frac{\textit{Variability}\ \text{AMONG/BETWEEN}\ \textit{the means}}{\textit{Variability}\ \text{AROUND/WITHIN}\ \textit{the distribution}} = \frac{\textit{Variance Between}}{\textit{Variance Within}}$$



$$\textit{Total Variance} = \textit{Variance Between} + \textit{Variance Within}$$

Partitioning – Separating total variance into its component parts

If the 'Variability BETWEEN the means' is greater, numerator will be relatively larger Hence ratio will be much greater than 1.
i.e. It means, the samples most likely do not come from the common population; **REJECT NULL HYPOTHESIS.**

Concept of ANOVA

$$\frac{Variance\ Between}{Variance\ Within} = \frac{Variance\ Among}{Variance\ Around}$$

$$\frac{LARGE}{small} = Reject\ H_0$$

At least one mean is an outlier

$$\frac{similar}{similar} = Fail\ to\ Reject\ H_0$$

Means are fairly close to overall mean and distributions overlap a bit

$$\frac{small}{LARGE} = Fail\ to\ Reject\ H_0$$

Means are very close to overall mean and the distributions melt together.

Types of ANOVA

Types of ANOVA

Types of ANOVA

- One Way ANOVA *(One Factor ANOVA)*
- Two way ANOVA without Replication *(Two Factor ANOVA)*
- Two way ANOVA with Replication *(Two Factor ANOVA)*

Why ANOVA ?

- Using various tests for Hypothesis, we have been comparing two populations.
  - Independent Samples t-test (random)
  - Matched sample t-test (paired)

- However, this limit us to the comparison of two populations only.

- If you wish to compare the means of more than two populations each containing several levels or subgroups we use ANOVA

- **AN**alysis **O**f **VA**riance

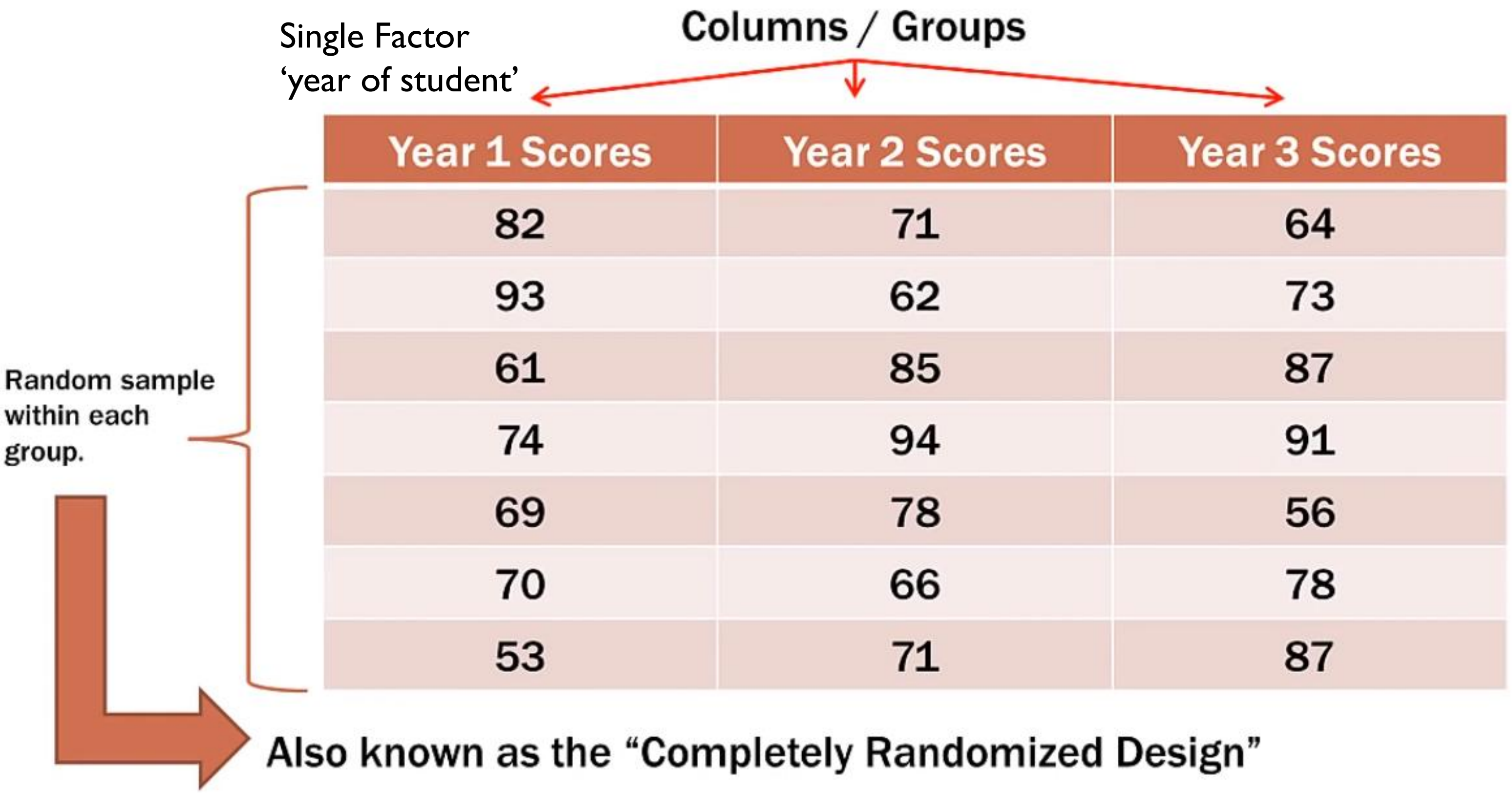
Example – One way ANOVA

Twenty One students at University of Madrid in Spain were selected for a test on few common topics combined.
7 first year students, 7 second year students, 7 third year students were randomly selected.

Students undertook assessment having maximum score of 100.
We are interested in whether or not a difference exists somewhere between the three different year levels ?

Example – One way ANOVA

Single Factor
‘year of student’



Columns / Groups

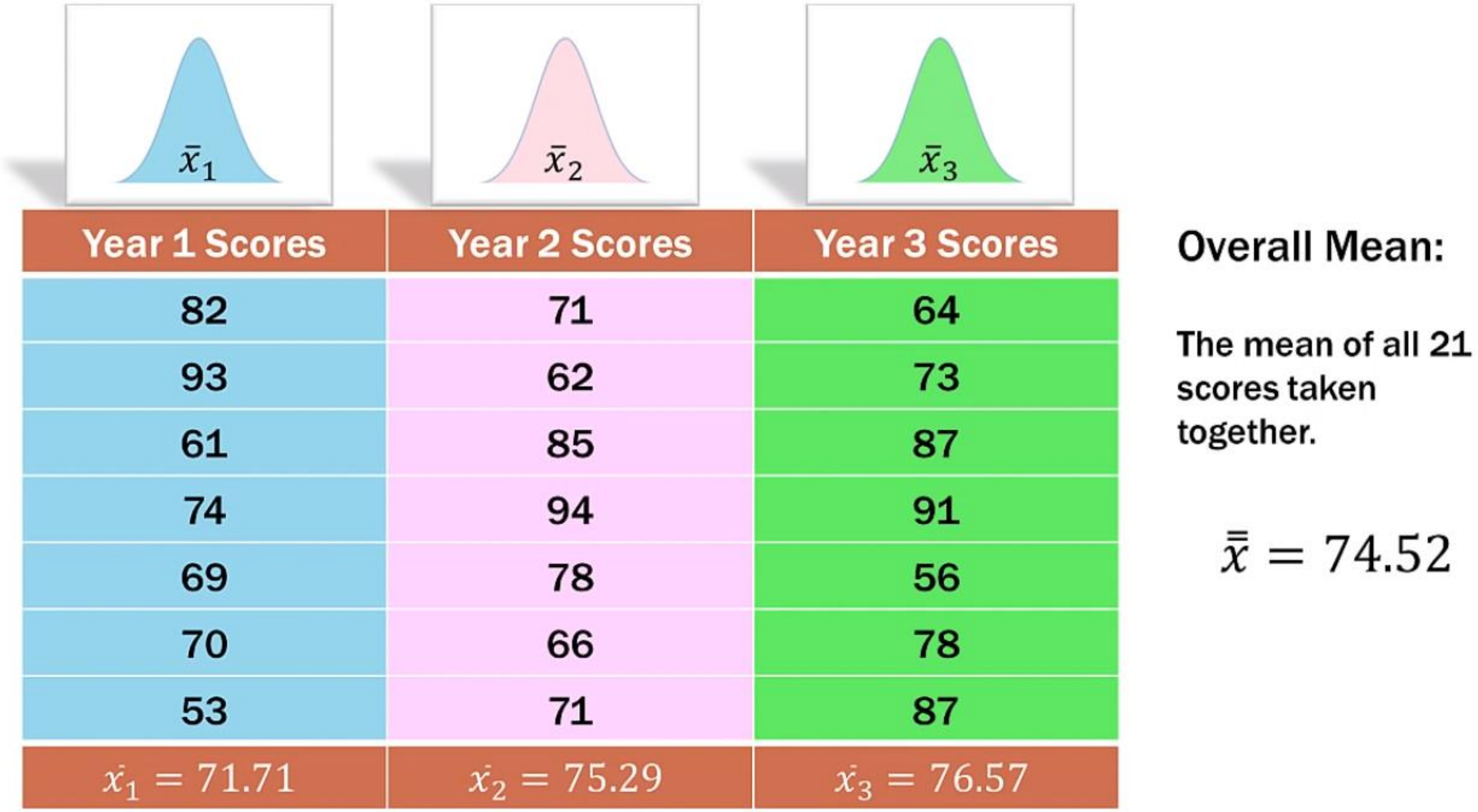
Year 1 Scores	Year 2 Scores	Year 3 Scores
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87

Random sample within each group.

Also known as the “Completely Randomized Design”

Example – One way ANOVA

Step 1: Calculate Mean of each column
Step 2: Calculate Overall Mean


$\bar{x}_1$
$\bar{x}_2$
$\bar{x}_3$

Year 1 Scores	Year 2 Scores	Year 3 Scores
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87
$\bar{x}_1 = 71.71$	$\bar{x}_2 = 75.29$	$\bar{x}_3 = 76.57$

Overall Mean:

The mean of all 21 scores taken together.

$$\bar{\bar{x}} = 74.52$$

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)

$$SS = \sum (x - \mu)^2$$

SST = SSC + SSE

Where

SST = Sum of square Totals or Total Sum of Squares, which is

Sum of square of (Each item in all samples – Overall Mean)

SSC = Sum of Square of Columns, which is

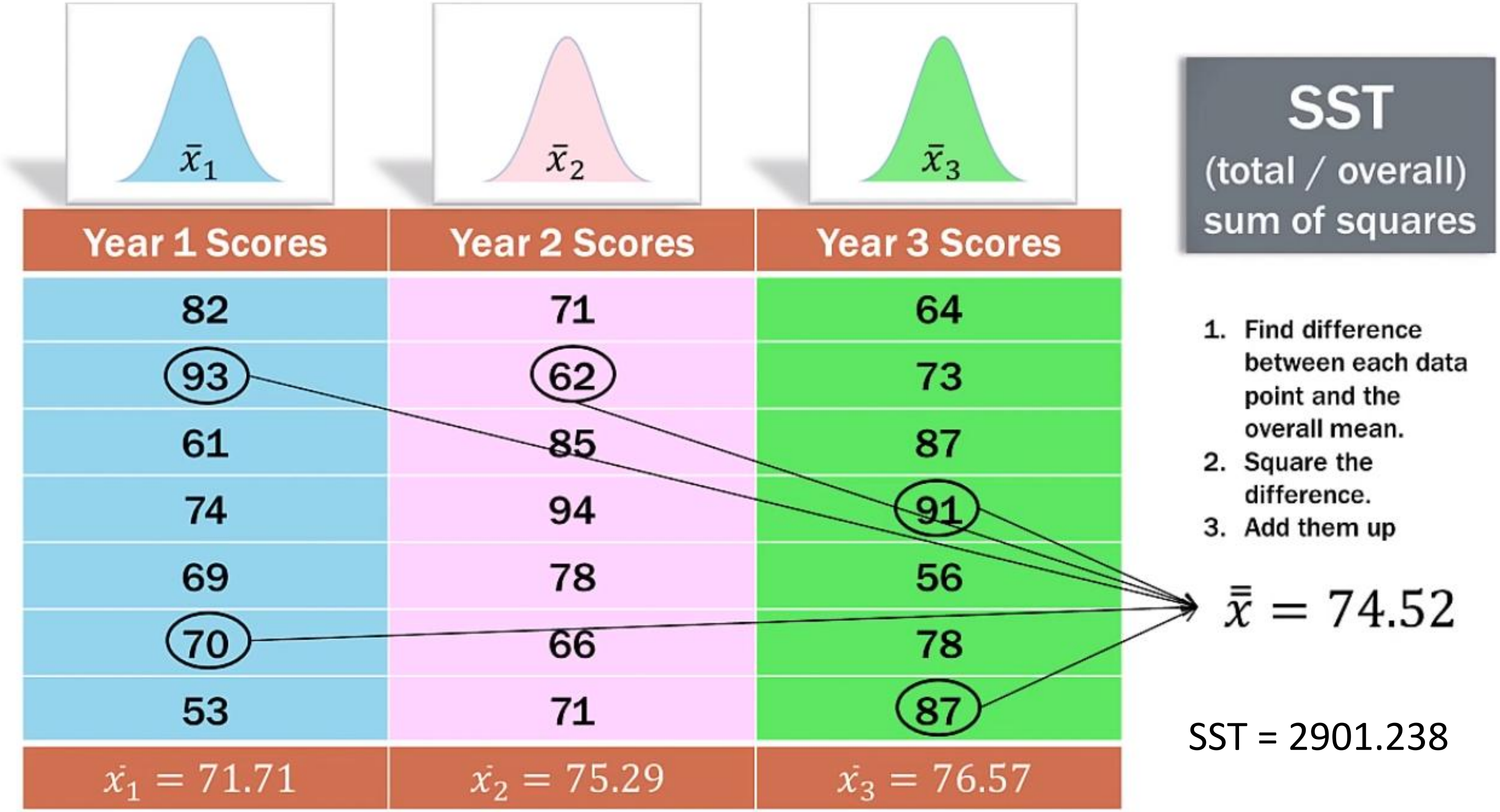
Sum of square of (Each Group Mean – Overall Mean)

SSE = Sum of Square or Sum of Square of Errors, which is

Sum of square if (Each item in a group – Mean of that group)

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



Year 1 Scores	Year 2 Scores	Year 3 Scores
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87
$\bar{x}_1 = 71.71$	$\bar{x}_2 = 75.29$	$\bar{x}_3 = 76.57$

SST

(total / overall) sum of squares

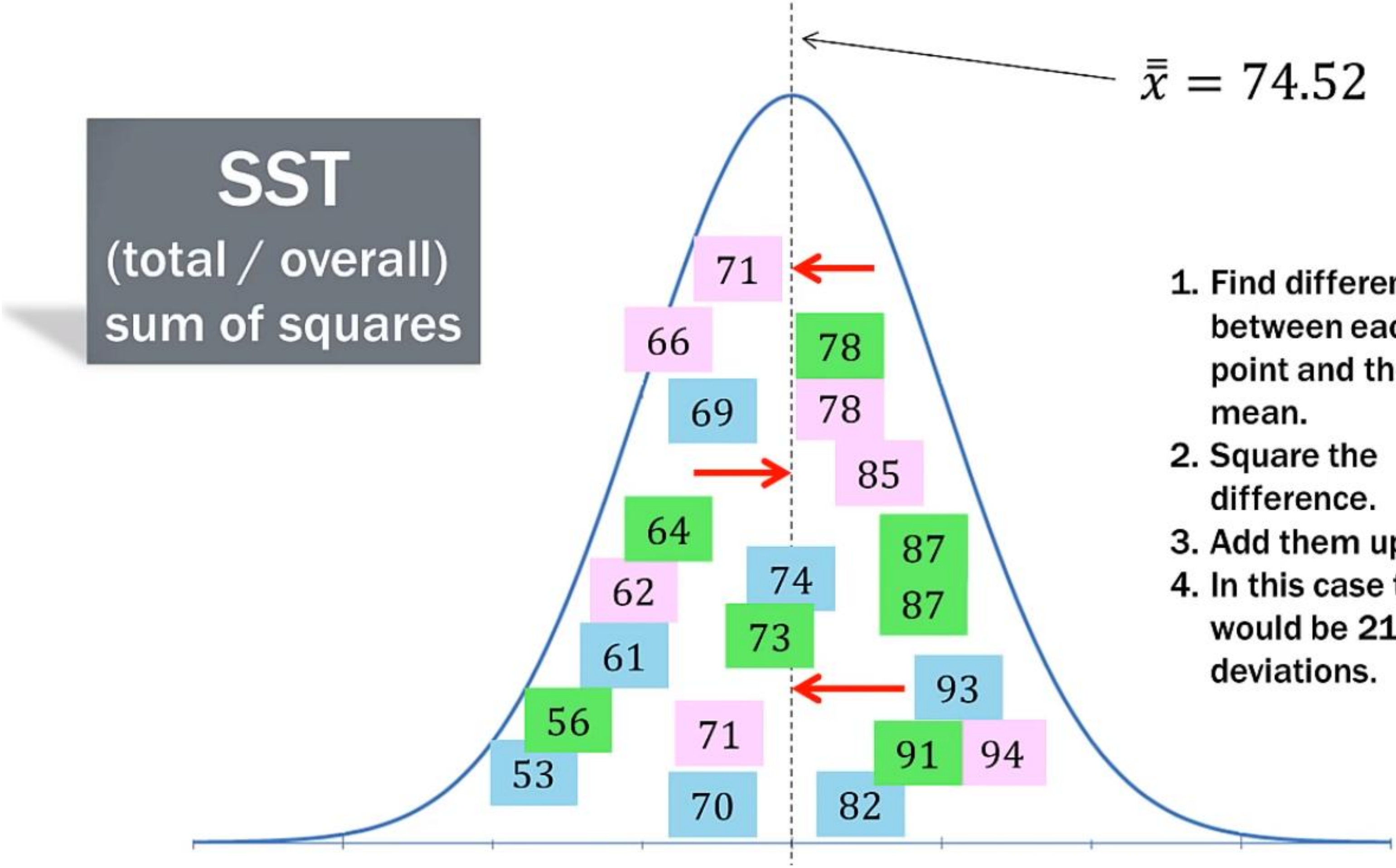
1. Find difference between each data point and the overall mean.
2. Square the difference.
3. Add them up

$$\bar{\bar{x}} = 74.52$$

SST = 2901.238

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



SST

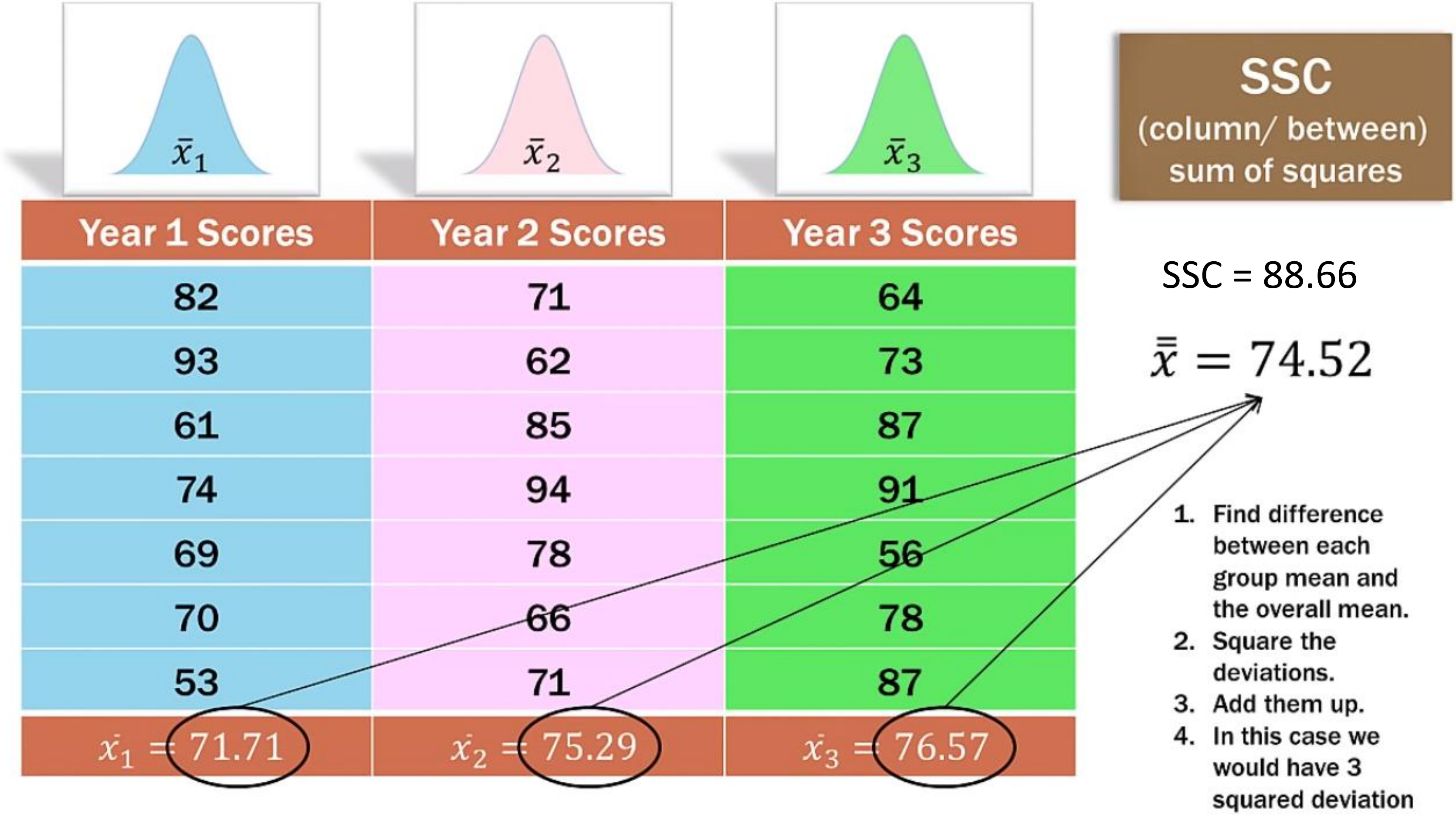
(total / overall) sum of squares

$\bar{\bar{x}} = 74.52$

1. Find difference between each data point and the overall mean.
2. Square the difference.
3. Add them up.
4. In this case there would be 21 squared deviations.

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



Year 1 Scores	Year 2 Scores	Year 3 Scores
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87
$\bar{x}_1 = 71.71$	$\bar{x}_2 = 75.29$	$\bar{x}_3 = 76.57$

SSC

(column/ between)
sum of squares

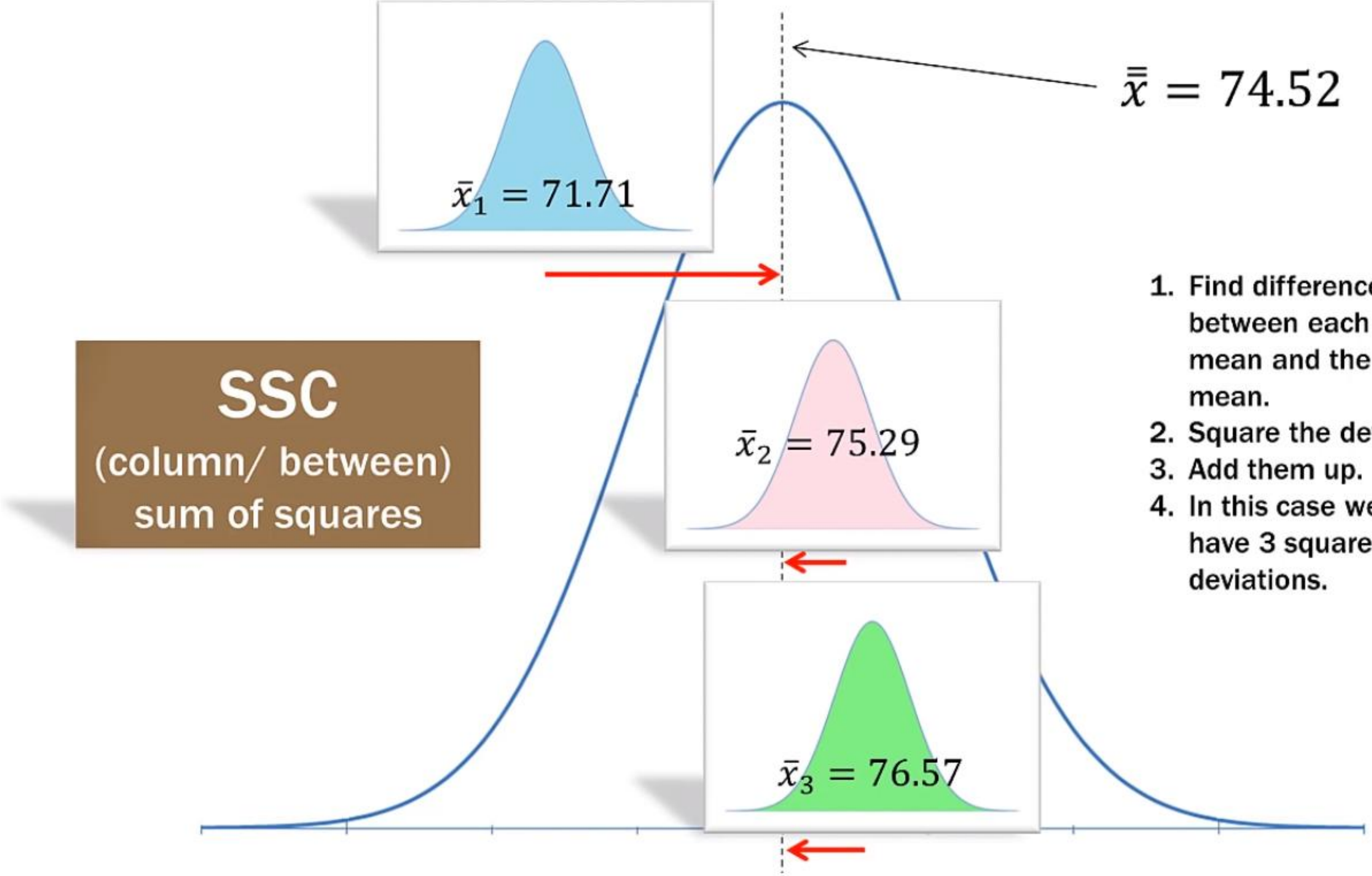
SSC = 88.66

$$\bar{\bar{x}} = 74.52$$

1. Find difference between each group mean and the overall mean.
2. Square the deviations.
3. Add them up.
4. In this case we would have 3 squared deviation

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



SSC

(column/ between) sum of squares

$\bar{\bar{x}} = 74.52$

$\bar{x}_1 = 71.71$

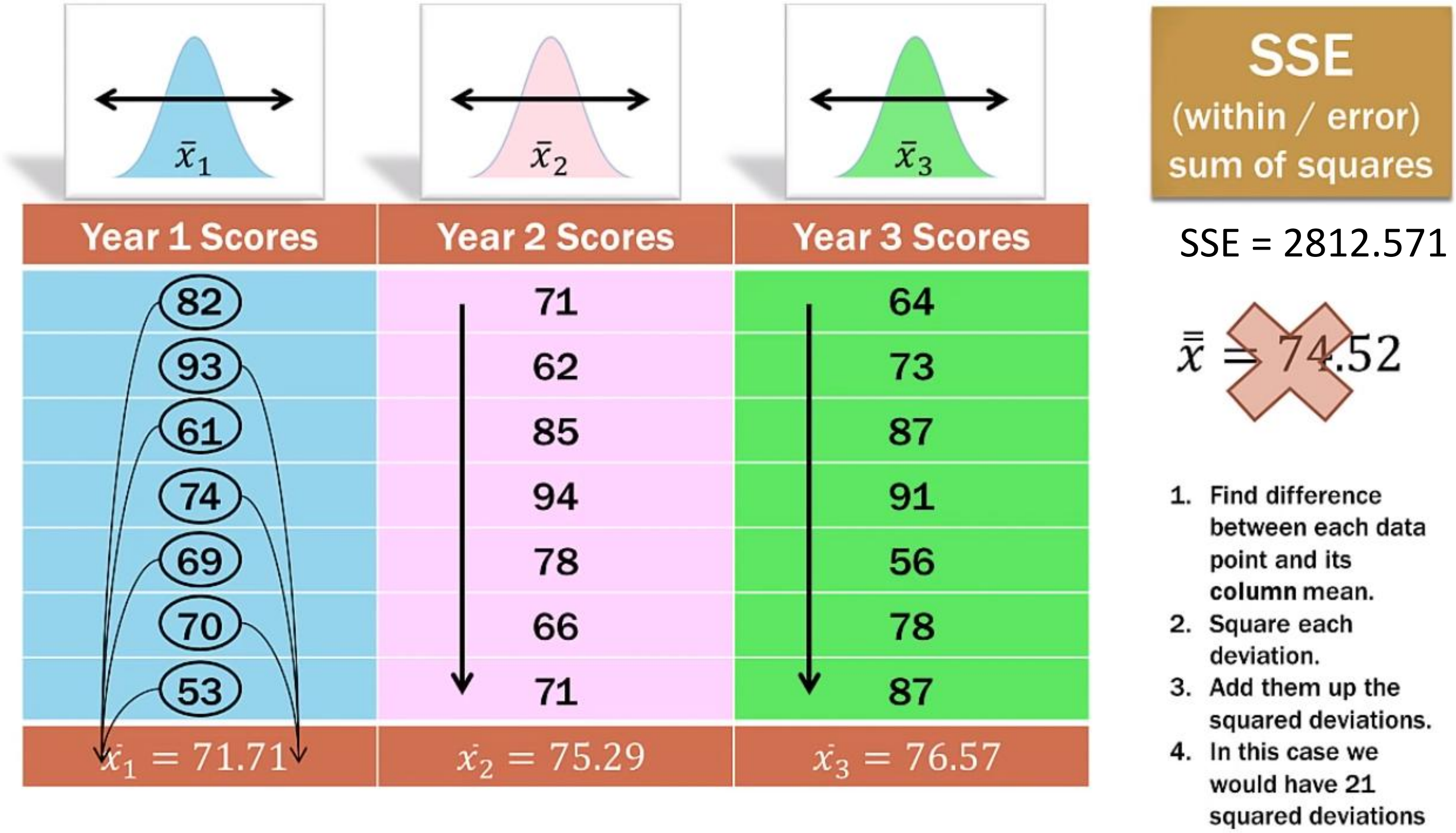
$\bar{x}_2 = 75.29$

$\bar{x}_3 = 76.57$

1. Find difference between each sample mean and the overall mean.
2. Square the deviations.
3. Add them up.
4. In this case we would have 3 squared deviations.

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



Year 1 Scores	Year 2 Scores	Year 3 Scores
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87
$\bar{x}_1 = 71.71$	$\bar{x}_2 = 75.29$	$\bar{x}_3 = 76.57$

SSE
(within / error)
sum of squares

SSE = 2812.571

$\bar{\bar{x}}\ \cancel{= 74.52}$

1. Find difference between each data point and its **column** mean.
2. Square each deviation.
3. Add them up the squared deviations.
4. In this case we would have 21 squared deviations

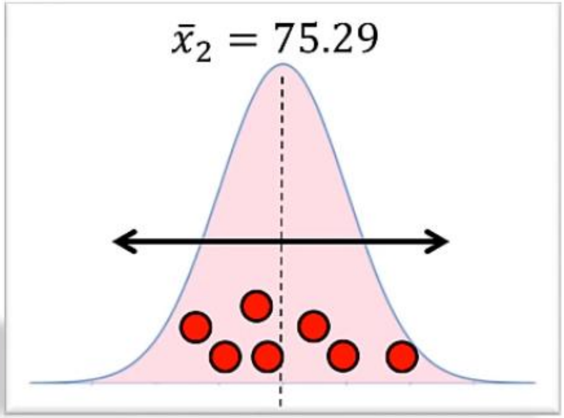
Example – One way ANOVA

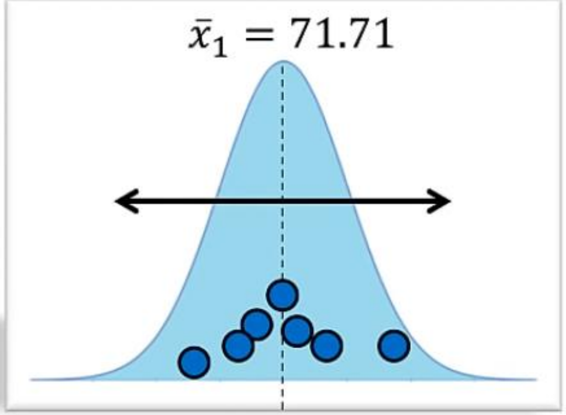
Step 3: Calculate Sum of Squares (SST, SSC, SSE)

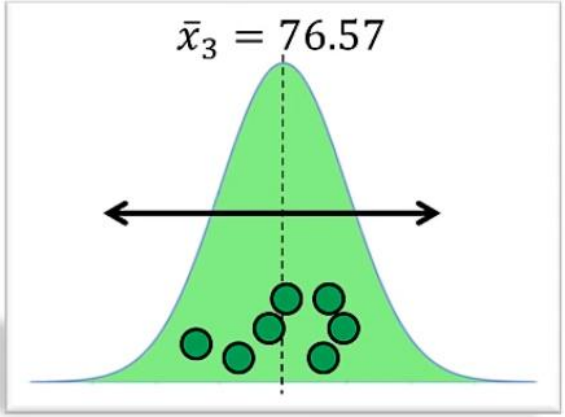
SSE

(within / error) sum of squares

1. Find difference between each data point and its **column** mean.
2. Square each deviation.
3. Add them up the squared deviations.
4. In this case we would have **21** squared deviations.

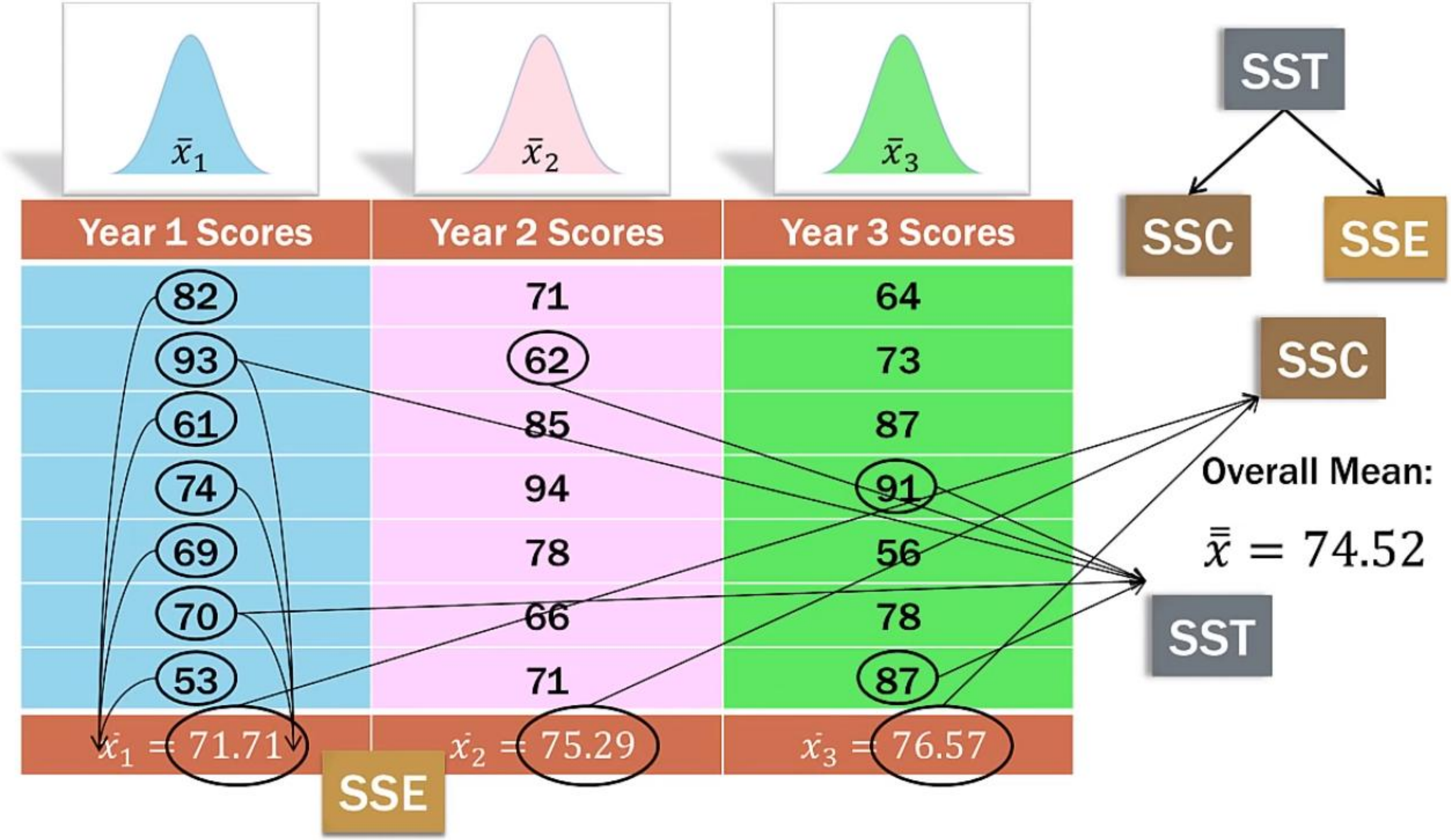






Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



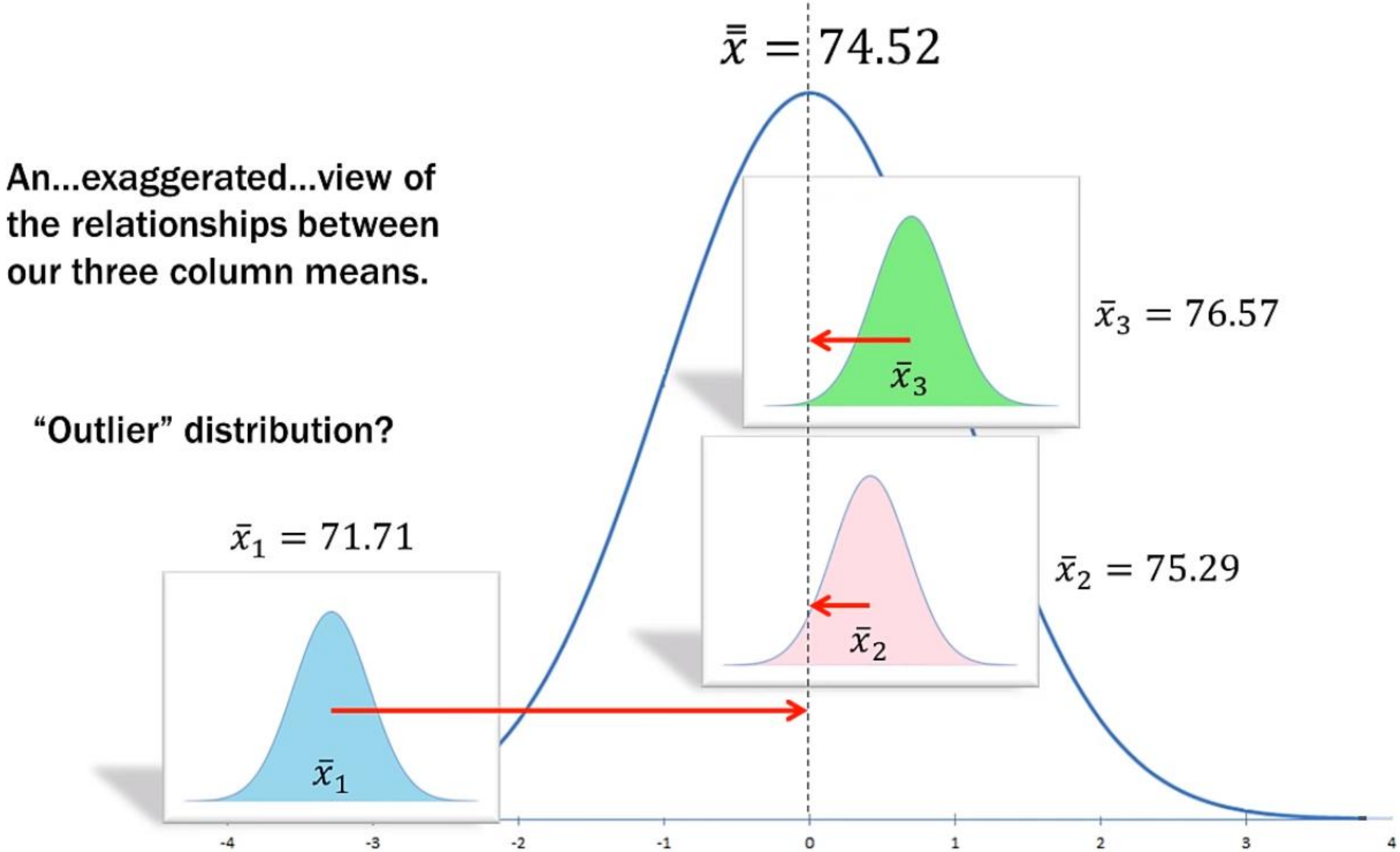
Year 1 Scores	Year 2 Scores	Year 3 Scores
82	71	64
93	62	73
61	85	87
74	94	91
69	78	56
70	66	78
53	71	87
$\bar{x}_1 = 71.71$	$\bar{x}_2 = 75.29$	$\bar{x}_3 = 76.57$

Overall Mean:

$$\bar{\bar{x}} = 74.52$$

Example – One way ANOVA

Step 3: Calculate Sum of Squares (SST, SSC, SSE)



An…exaggerated…view of the relationships between our three column means.

“Outlier” distribution?

Example – One way ANOVA

Step 4: Calculate Degree of Freedom (df), MSC and MSE

SSC $\qquad df columns = C - 1$ $\qquad MSC = \frac{SSC}{df_{columns}}$

SSE $\qquad df_{error} = N - C$ $\qquad MSE = \frac{SSE}{df_{error}}$

SST $\qquad df_{total} = N - 1$ $\qquad F = \frac{MSC}{MSE}$

N = total number of observations
C = Number of columns/treatments

Example – One way ANOVA

Step 4: Calculate Degree of Freedom (df), MSC and MSE

SSC $\qquad dfcolumns = 3 - 1 = 2 \qquad MSC = \frac{SSC}{df_{columns}} = \frac{88.66}{2} =$

44.33

SSE $\qquad dferror = 21 - 3 = 18 \qquad MSE = \frac{SSE}{df_{error}} = \frac{2812.571}{18} =$

156.254

SST $\qquad df_{total} = 21 - 1 = 20 \qquad F = \frac{MSC}{MSE}$

MSC = Mean Square Columns/treatments
MSE = Mean Square Error

Example – One way ANOVA

Step 5: Calculate F Ratio

SSC $\quad df columns = 3 - 1 = 2 \quad$ $MSC = \frac{SSC}{df_{columns}} = \frac{88.66}{2} = 44.33$

SSE $\quad df error = 21 - 3 = 18 \quad$ $MSE = \frac{SSE}{df_{error}} = \frac{2812.571}{18} = 156.254$

SST $\quad df_{total} = 21 - 1 = 20 \quad$ $F = \frac{MSC}{MSE} = \frac{44.33}{156.254} = 0.2837$

MSC = Mean Square Columns/treatments
MSE = Mean Square Error

Example – One way ANOVA

Step 6: Calculate F Critical Value

SSC $dfcolumns = 3 - 1 = 2$ $MSC = \frac{SSC}{df_{columns}} = \frac{88.66}{2} = 44.33$

SSE $dferror = 21 - 3 = 18$ $MSE = \frac{SSE}{df_{error}} = \frac{2812.571}{18} = 156.254$

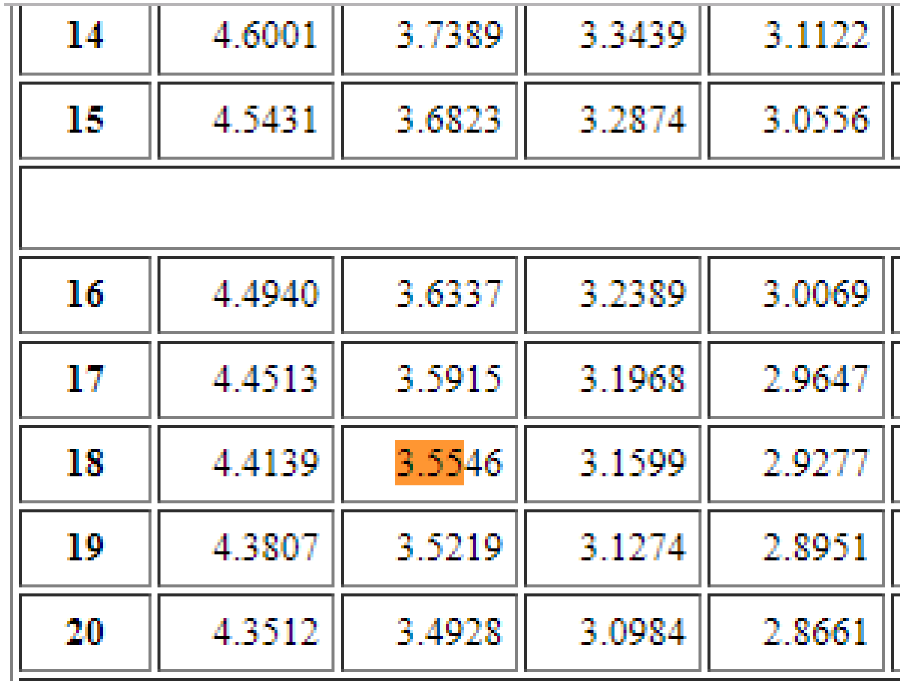
SST $df_{total} = 21 - 1 = 20$ $F = \frac{MSC}{MSE} = \frac{44.33}{156.254} = 0.2837$

Look up F statistic distribution table for alpha = 0.05 and degree of freedom of numerator (SSC) = 2 and degree of freedom for denominator (SSE) = 18

Example – One way ANOVA

Step 6: Calculate F Critical Value

Look up F statistic distribution table for alpha = 0.05 and degree of freedom of numerator (SSC) = 2 and degree of freedom for denominator (SSE) = 18. Refer F statistics table.



14	4.6001	3.7389	3.3439	3.1122
15	4.5431	3.6823	3.2874	3.0556
16	4.4940	3.6337	3.2389	3.0069
17	4.4513	3.5915	3.1968	2.9647
18	4.4139	3.5546	3.1599	2.9277
19	4.3807	3.5219	3.1274	2.8951
20	4.3512	3.4928	3.0984	2.8661

Example – One way ANOVA

Step 7: Inference

F Ratio = 0.2837
While the F critical value for alpha = 0.05 is
F critical = 3.5546

Is our F statistic (i.e. F Ratio) value larger or beyond F critical ?
No. Hence our NULL Hypothesis holds.
i.e. We fail to reject the H_0

Hence, there is no significant difference in mean test score by Year of Student.

Why ANOVA ?

- Using various tests for Hypothesis, we have been comparing two populations.
  - Independent Samples t-test (random)
  - Matched sample t-test (paired)

- However, this limit us to the comparison of two populations only.

- If you wish to compare the means of more than two populations each containing several levels or subgroups we use ANOVA

- **AN**alysis **O**f **VA**riance

Two-Way ANOVA ‘BLOCK’ Design

- In one-way ANOVA we selected random sample for each column/treatment group

- Two-way ANOVA allows us to ‘account for variation’ at the ROW level due to some other factor or grouping.
- i.e. in two-way ANOVA we add another dimension, the row dimension based on certain criteria.

- Here in two-way ANOVA, we attempt to minimize the ERROR variance by saying that some of the ERROR variance is actually due to the variance in the ROWS.

- So here, we now have 4 types of Sum of Squares (Sources of variance):
- **Total Variance = SSC + SSE + SSB (Sum of Square of Rows/Blocks)**

Example – Two way ANOVA – Without Rep

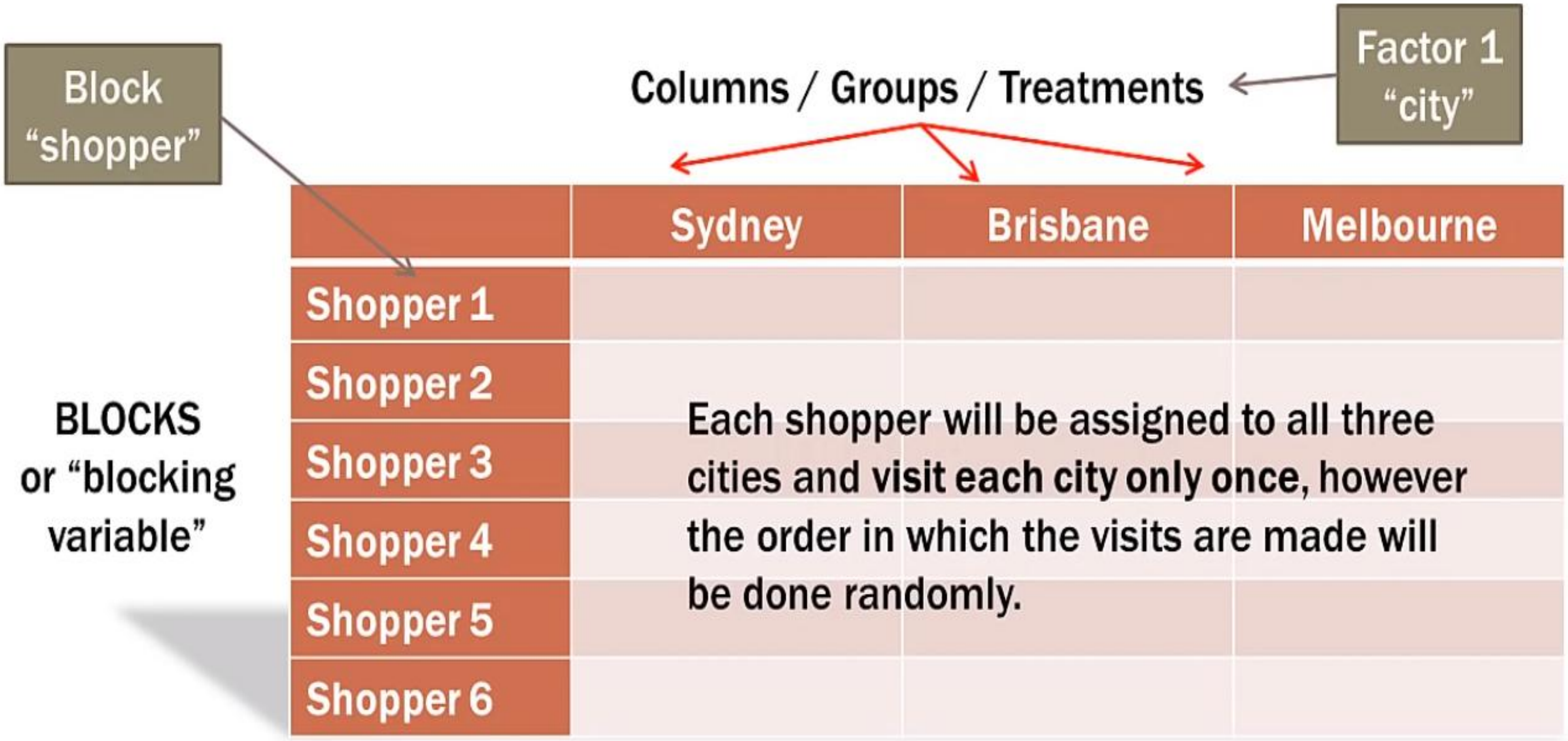
Starbucks under the pressure of Quality Control, sends out 6 Shopper inspectors as regular customers to the Australian cities of Sydney, Brisbane and Melbourne.

These 6 inspectors will visit the same stores in each 3 cities in a random manner. They will do the survey and check how well the store is managed, how good is the service and quality of products etc.

Here, Starbucks Management like to know If a difference in the Inspector ratings exists among the cities. Are they all same ? Is one significantly higher than the other two ?
Are all three different from each other ?

Note: What makes this problem good fit for Two-way ANOVA is that the Inspectors will have their own natural variation.

Example – Two way ANOVA – Without Rep



Block "shopper"

Columns / Groups / Treatments

Factor 1 "city"

	Sydney	Brisbane	Melbourne
Shopper 1			
Shopper 2			
Shopper 3			
Shopper 4			
Shopper 5			
Shopper 6			

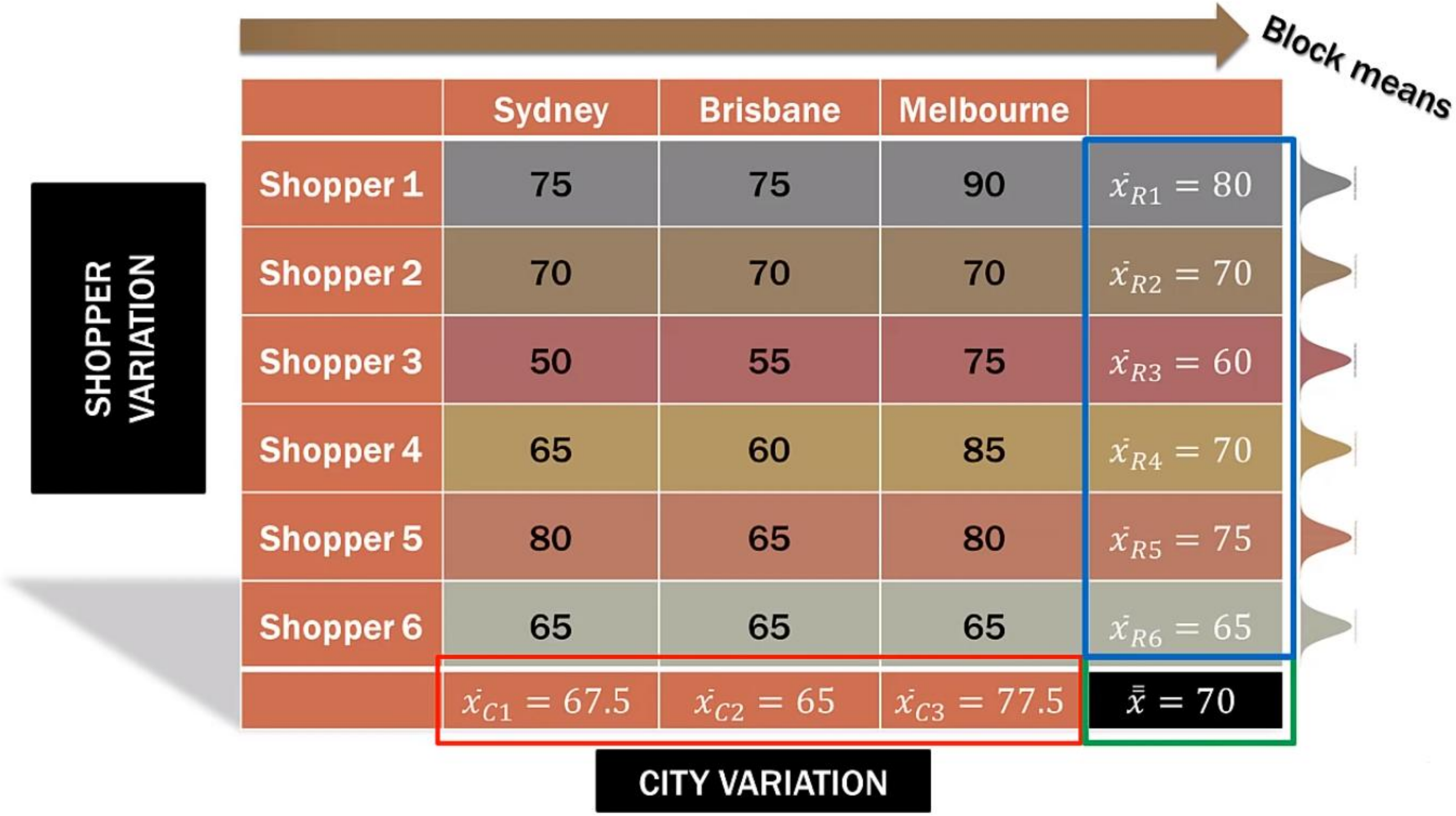
BLOCKS
or "blocking
variable"

Each shopper will be assigned to all three cities and **visit each city only once**, however the order in which the visits are made will be done randomly.

Example – Two way ANOVA – Without Rep

Step 1: Find Column means, Row Means

Step 2: Find the Overall Mean



Block means
SHOPPER VARIATION
CITY VARIATION

	Sydney	Brisbane	Melbourne	
Shopper 1	75	75	90	$\bar{x}_{R1} = 80$
Shopper 2	70	70	70	$\bar{x}_{R2} = 70$
Shopper 3	50	55	75	$\bar{x}_{R3} = 60$
Shopper 4	65	60	85	$\bar{x}_{R4} = 70$
Shopper 5	80	65	80	$\bar{x}_{R5} = 75$
Shopper 6	65	65	65	$\bar{x}_{R6} = 65$
	$\bar{x}_{C1} = 67.5$	$\bar{x}_{C2} = 65$	$\bar{x}_{C3} = 77.5$	$\bar{\bar{x}} = 70$

Example – Two way ANOVA – Without Rep

Step 3: Calculate Sum of Squares (SST, SSC, SSE, SSB)

$$SS = \sum (x - \mu)^2$$

SST = SSC + SSE + SSB

Where

SST = Sum of square Totals or Total Sum of Squares, which is
Sum of square of (Each item in all samples – Overall Mean)

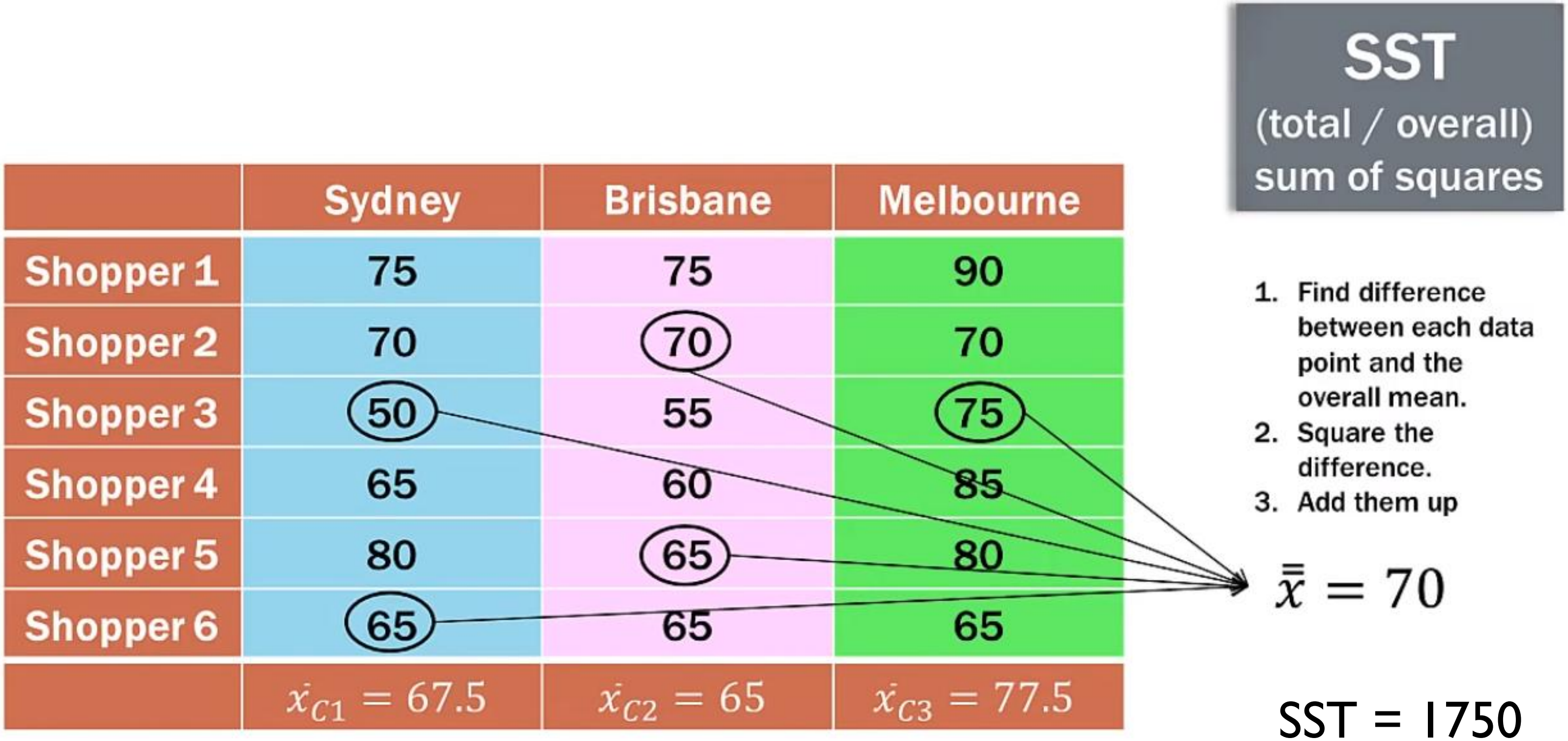
SSC = Sum of Square of Columns, which is
Sum of square of (Each Group Mean – Overall Mean)

SSE = Sum of Square or Sum of Square of Errors, which is
Sum of square if (Each item in a group – Mean of that group)

SSB = Sum of Square Errors of Blocks, which is
Sum of square of (Each average in block – Overall Mean)

Example – Two way ANOVA – Without Rep

Step 3: Calculate Sum of Squares (SST, SSC, SSE, SSB)



	Sydney	Brisbane	Melbourne
Shopper 1	75	75	90
Shopper 2	70	70	70
Shopper 3	50	55	75
Shopper 4	65	60	85
Shopper 5	80	65	80
Shopper 6	65	65	65
	$\bar{x}_{C1} = 67.5$	$\bar{x}_{C2} = 65$	$\bar{x}_{C3} = 77.5$

SST

(total / overall) sum of squares

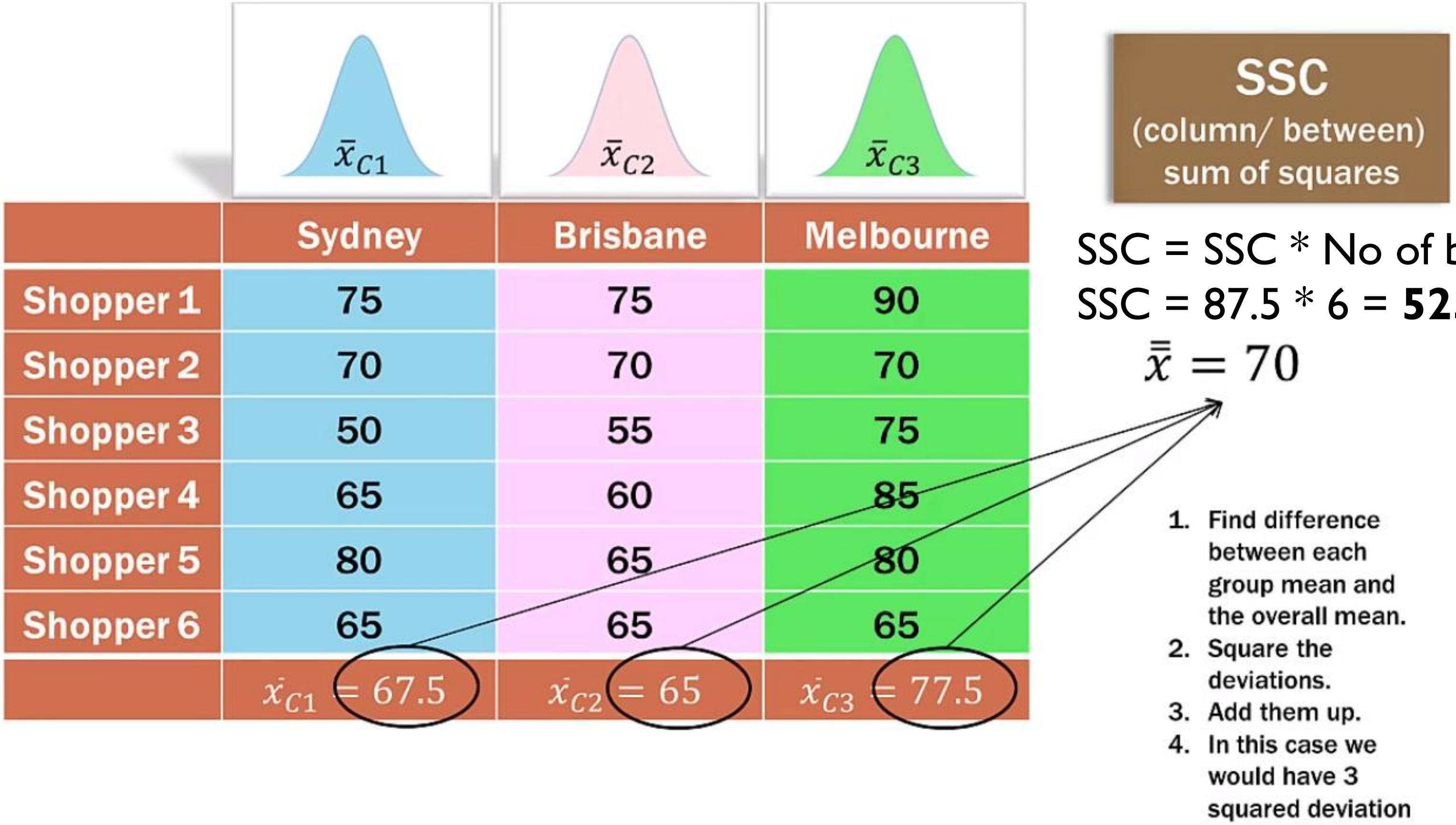
1. Find difference between each data point and the overall mean.
2. Square the difference.
3. Add them up

$\bar{\bar{x}} = 70$

SST = 1750

Example – Two way ANOVA – Without Rep

Step 3: Calculate Sum of Squares (SST, SSC, SSE, SSB)



	Sydney	Brisbane	Melbourne
Shopper 1	75	75	90
Shopper 2	70	70	70
Shopper 3	50	55	75
Shopper 4	65	60	85
Shopper 5	80	65	80
Shopper 6	65	65	65
	$\bar{x}_{C1} = 67.5$	$\bar{x}_{C2} = 65$	$\bar{x}_{C3} = 77.5$

SSC

(column/ between) sum of squares

SSC = SSC * No of blocks

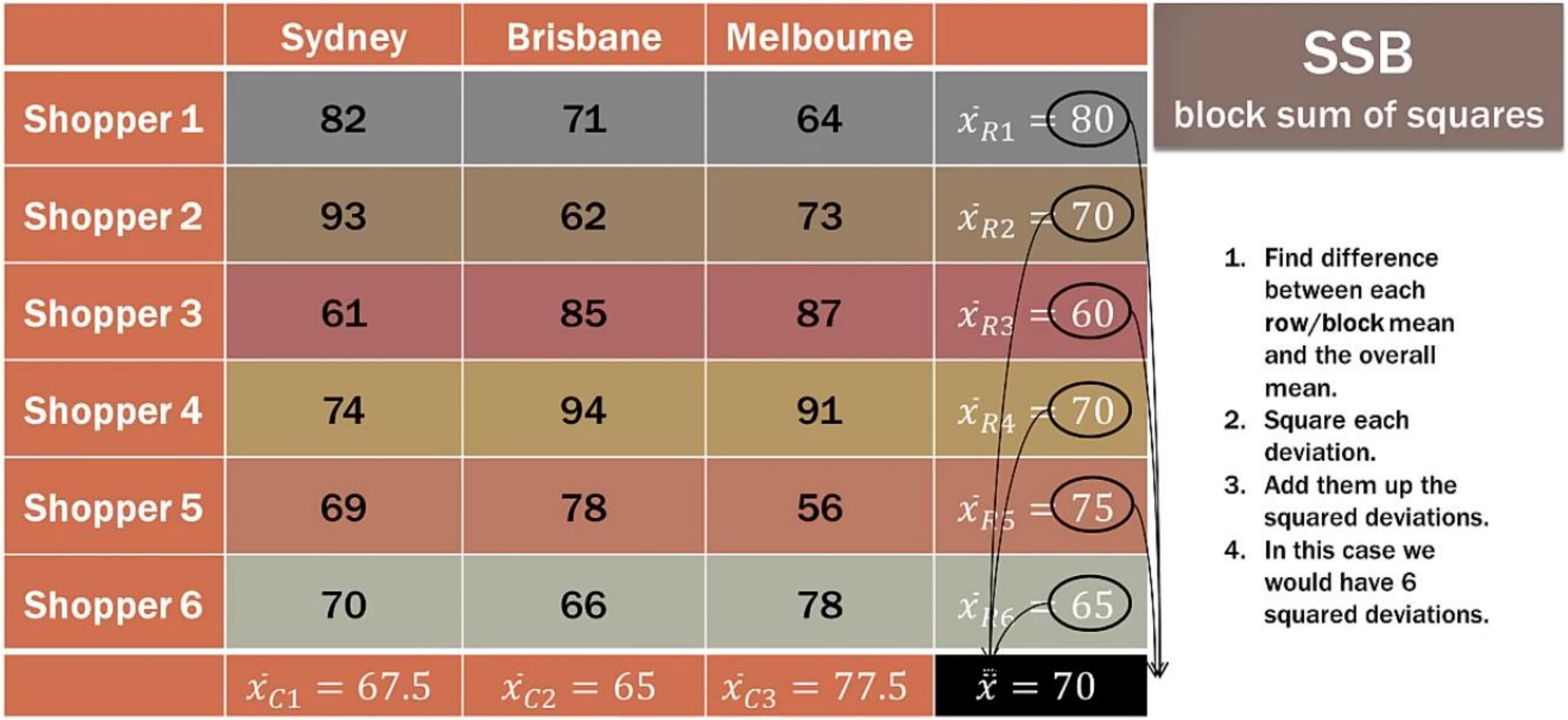
SSC = 87.5 * 6 = **525**

$$\bar{\bar{x}} = 70$$

1. Find difference between each group mean and the overall mean.
2. Square the deviations.
3. Add them up.
4. In this case we would have 3 squared deviation

Example – Two way ANOVA – Without Rep

Step 3: Calculate Sum of Squares (SST, SSC, SSE, SSB)



	Sydney	Brisbane	Melbourne	
Shopper 1	82	71	64	$\bar{x}_{R1} = 80$
Shopper 2	93	62	73	$\bar{x}_{R2} = 70$
Shopper 3	61	85	87	$\bar{x}_{R3} = 60$
Shopper 4	74	94	91	$\bar{x}_{R4} = 70$
Shopper 5	69	78	56	$\bar{x}_{R5} = 75$
Shopper 6	70	66	78	$\bar{x}_{R6} = 65$
	$\bar{x}_{C1} = 67.5$	$\bar{x}_{C2} = 65$	$\bar{x}_{C3} = 77.5$	$\bar{\bar{x}} = 70$

SSB

block sum of squares

1. Find difference between each row/block mean and the overall mean.
2. Square each deviation.
3. Add them up the squared deviations.
4. In this case we would have 6 squared deviations.

SSB = SSB * No of Columns/Groups = 250 * 3 = **750**

Example – Two way ANOVA – Without Rep

Step 3: Calculate Sum of Squares (SST, SSC, SSE, SSB)

$$SS = \sum (x - \mu)^2$$

SST = SSC + SSE + SSB

Hence

SSE = SST – SSC – SSB

SSE = 1750 – 525 – 750 = 475

SSE = 475

Example – Two way ANOVA – Without Rep

Step 4: Calculate Degree of Freedom (df), MSC, MSB and MSE

SSC $\quad dfcolumns = C - 1$ $\quad MSC = \frac{SSC}{df_{columns}}$

SSB $\quad df_{blocks} = B – 1$ $\quad MSB = \frac{SSB}{df_{blocks}}$

SSE $\quad dferror = (C - 1)(B - 1)$ $\quad MSE = \frac{SSE}{df_{error}}$

SST $\quad df_{total} = N - 1$ $\quad MST = \frac{SST}{df_{total}}$

N = total number of observations
C = Number of columns/treatments

Example – Two way ANOVA – Without Rep

Step 4: Calculate Degree of Freedom (df), MSC, MSB and MSE

SSC $\quad df columns = 3 - 1 = 2 \quad MSC =$
$\frac{525}{2}$=262.5

SSB $\quad df_{blocks} = 6 - 1=5 \quad MSB = \frac{750}{5}=150$

SSE $\quad df_{error} = (3-1)(6-1) = 10 \quad MSE = \frac{475}{10}$= 47.5

SST $\quad df_{total} = 18 - 1 = 17 \quad MST = \frac{1750}{17} = 102.941$

N = total number of observations

C = Number of columns/treatments

Example – Two way ANOVA – Without Rep

Step 5: Calculate F Ratios

We are interested in differences in the city ratings, hence the first ratio is important for us.

However, you can check for the variations in ratings of shoppers as well, for that we can make use of second F-ratio given below.

$$F = \frac{MSC}{MSE} = \frac{262.5}{47.5} = 5.526$$

$$F = \frac{MSB}{MSE} = \frac{150}{47.5} = 3.158$$

Example – Two way ANOVA – Without Rep

Step 6: Calculate F Critical Values

$$F = \frac{MSC}{MSE} = \frac{262.5}{47.5} = 5.526$$

$$F_{critical} = F_{\alpha=0.05,2,10} = 4.1028$$

Is our F statistic larger than $F_{critical}$?

Yes. Reject the H_0

Significant difference in Mean quality score by city.

$$F = \frac{MSB}{MSE} = \frac{150}{47.5} = 3.158$$

$$F_{critical} = F_{\alpha=0.05,5,10} = 3.3258$$

So there is some difference in the city scores even accounting for the variation in the shopper.

Two way ANOVA – With Replication

In a Two way ANOVA with replication, we have multiple measurements per cell.

This allows for a new type of measurement which is called the interaction between two factors.

Basically in Two Way ANOVA we look for significant interactions between the cells or columns.

Example - Two way ANOVA – With Replication

Example: The agricultural researchers are interested in the effectiveness of three different types of plant foods AA, BB and CC.

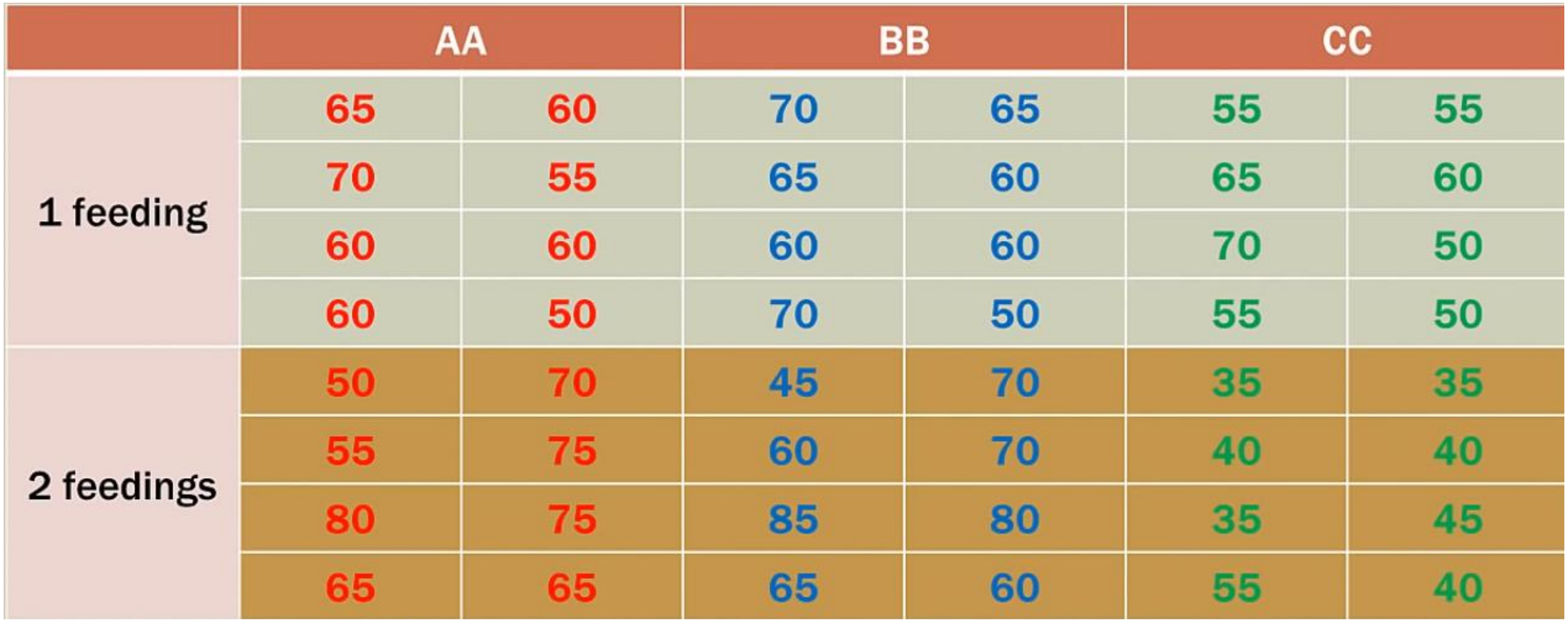
They designed an experiment in which they test each plant food AA, BB and CC with two feeding per day for 75 days after planting. Eight plants are tested for each combination.

The variable of interest is the plant height in centimeters (cm).
This same type of plant seed is used for the entire experiment.

Determine:

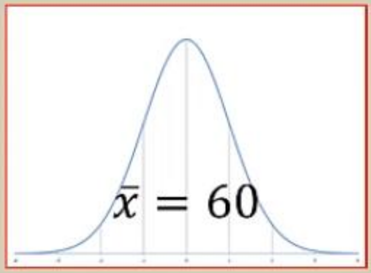
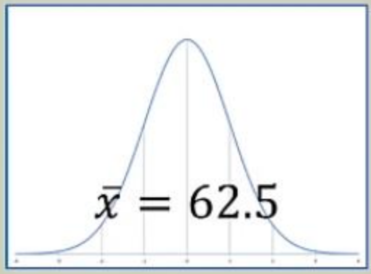
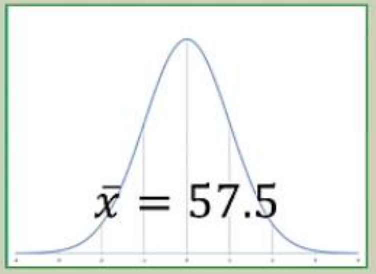
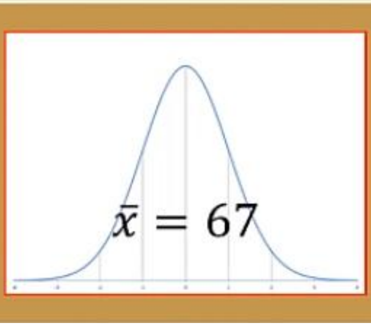
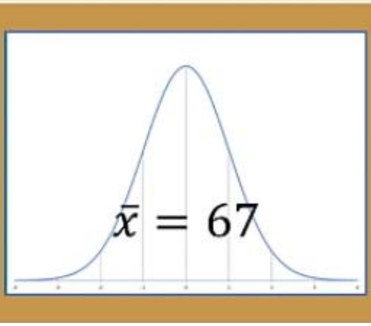
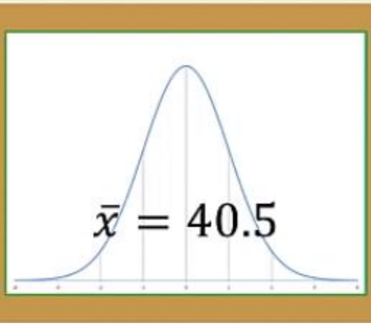
- Is there any difference in plant growth when comparing 1 feeding to 2 feedings per day ?
- Do some plants grow better once per day and others grow better at twice per day ?

Example - Two way ANOVA – With Replication



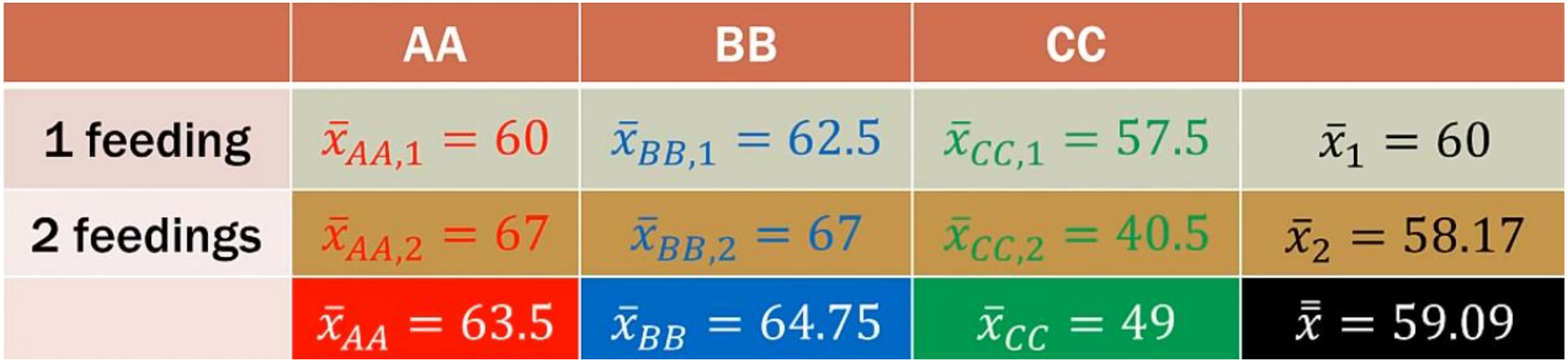
	AA		BB		CC	
1 feeding	65	60	70	65	55	55
	70	55	65	60	65	60
	60	60	60	60	70	50
	60	50	70	50	55	50
2 feedings	50	70	45	70	35	35
	55	75	60	70	40	40
	80	75	85	80	35	45
	65	65	65	60	55	40

Example - Two way ANOVA – With Replication

	AA	BB	CC
1 feeding	 $\bar{x} = 60$	 $\bar{x} = 62.5$	 $\bar{x} = 57.5$
2 feedings	 $\bar{x} = 67$	 $\bar{x} = 67$	 $\bar{x} = 40.5$

Each cell has its own mean and variance, which makes Two Factor with replication special.

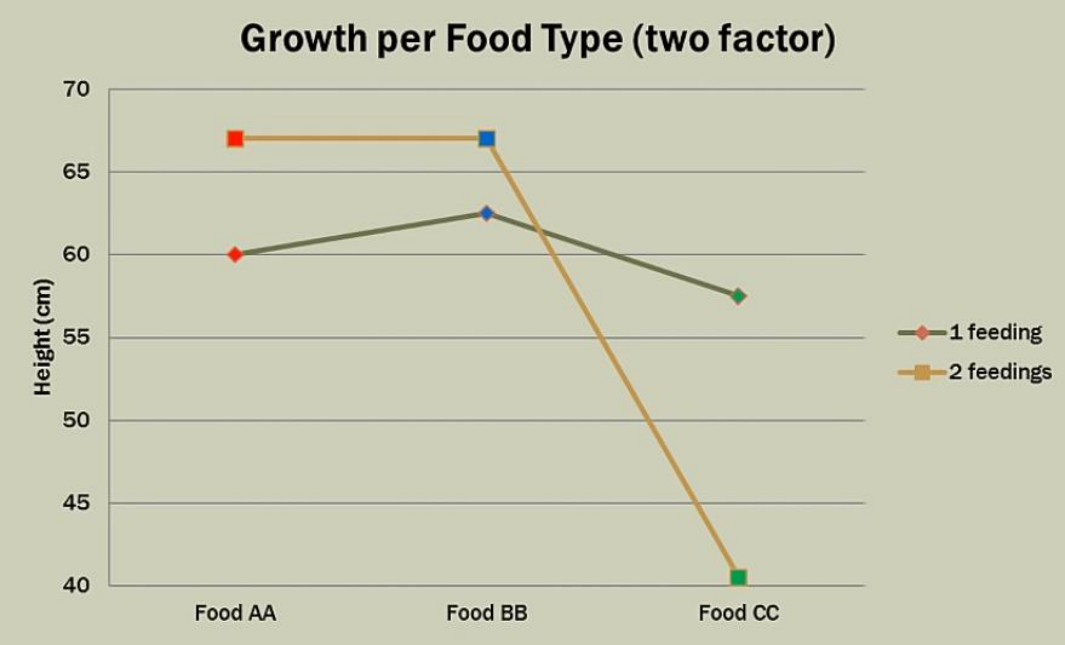
Example - Two way ANOVA – With Replication



	AA	BB	CC	
1 feeding	$\bar{x}_{AA,1} = 60$	$\bar{x}_{BB,1} = 62.5$	$\bar{x}_{CC,1} = 57.5$	$\bar{x}_1 = 60$
2 feedings	$\bar{x}_{AA,2} = 67$	$\bar{x}_{BB,2} = 67$	$\bar{x}_{CC,2} = 40.5$	$\bar{x}_2 = 58.17$
	$\bar{x}_{AA} = 63.5$	$\bar{x}_{BB} = 64.75$	$\bar{x}_{CC} = 49$	$\bar{\bar{x}} = 59.09$

Keep in mind that the points in the table are not the individual points but are the mean of their own distribution.

Example - Two way ANOVA – With Replication

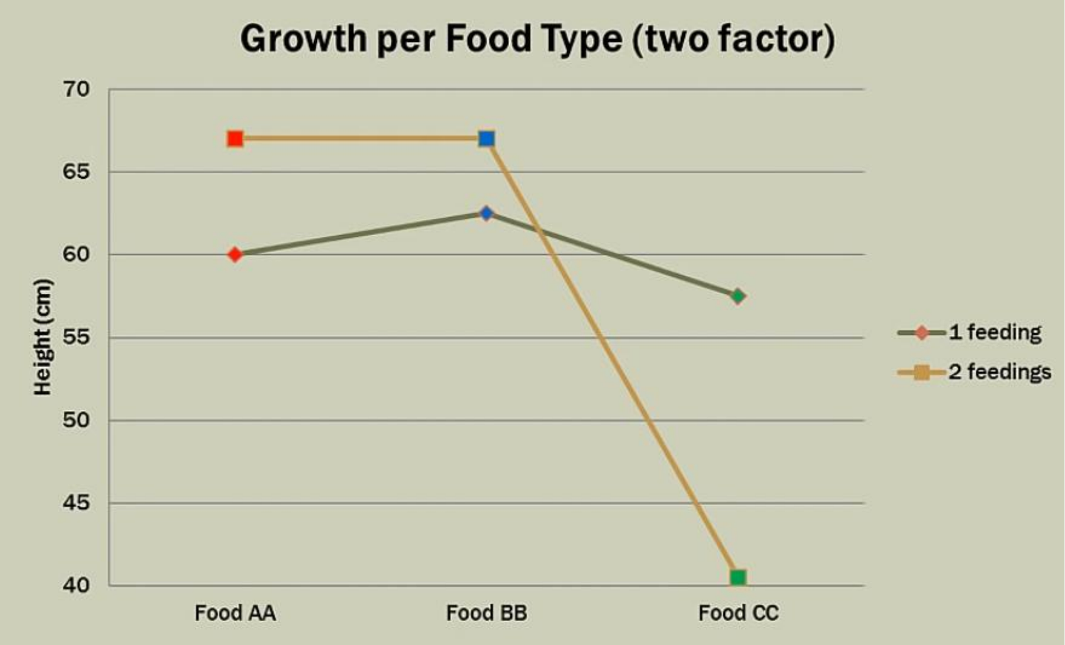


Questions:

- For AA, how many feedings produce more growth ?
- For BB, how many feedings produce more growth ?
- For CC, how many feedings produce more growth ?

Do two feedings produce more growth across ALL plant foods consistently ?

Example - Two way ANOVA – With Replication



Answers:
It appears that two feedings are better for AA and BB but not for CC.

This type of situation is called **Interaction.**

The **interaction** occurs when the effect of one factor changes for different levels of the other factor

In this case, the feeding frequency for food affects the plant growth.

Example - Two way ANOVA – With Replication

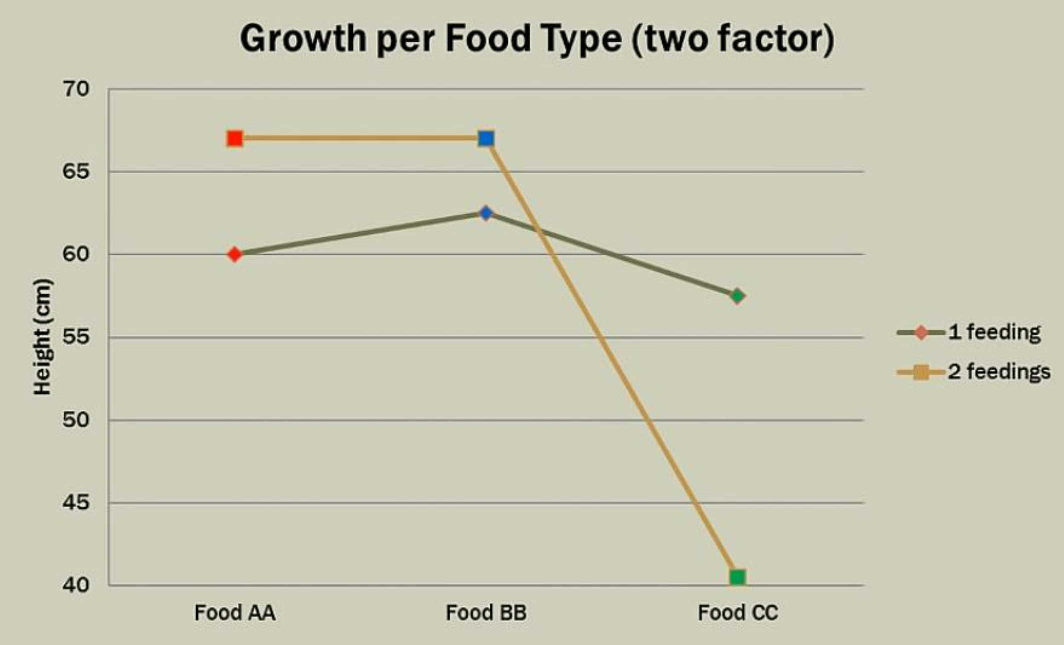
On a marginal means graph, usually:

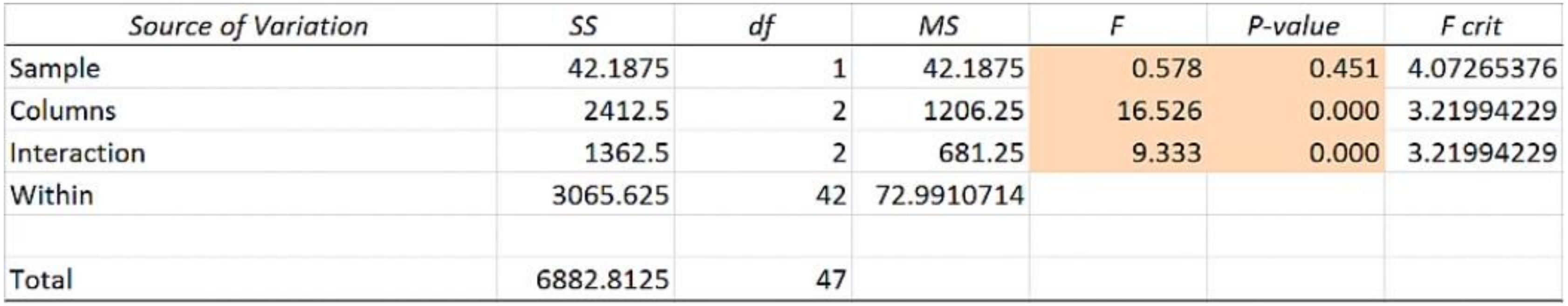
- The factor of interest or the factor with the most levels is on the horizontal axis.
- The dependent variable is on the vertical axis (what to measure)
- The second factor makes up the series lines by graphing the dependent variable for each level/category of the first factor

Look for the crossing or non-parallel lines on the graph i.e. interaction.

Always look for the interaction first.
If it is significant, then the relationship/significance of the factors **cannot be analyzed**. **The factors are too intertwined.**

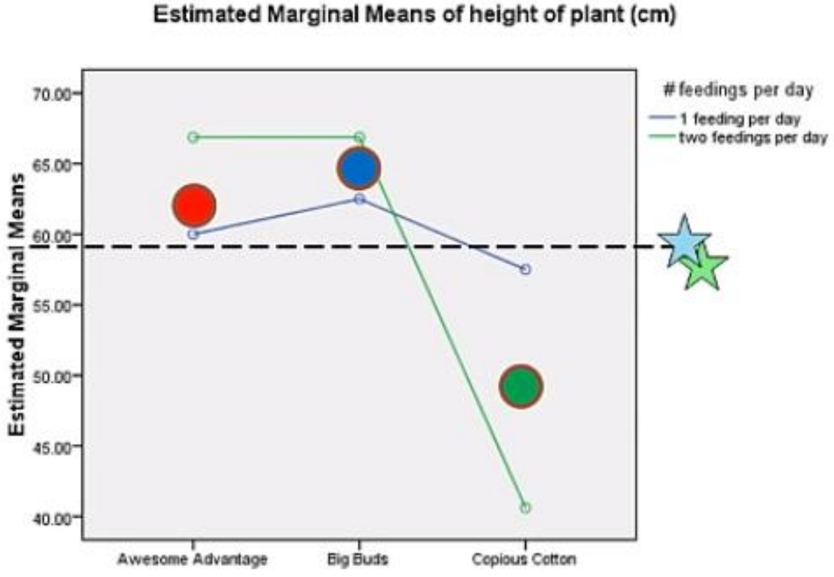
Example - Two way ANOVA – With Replication

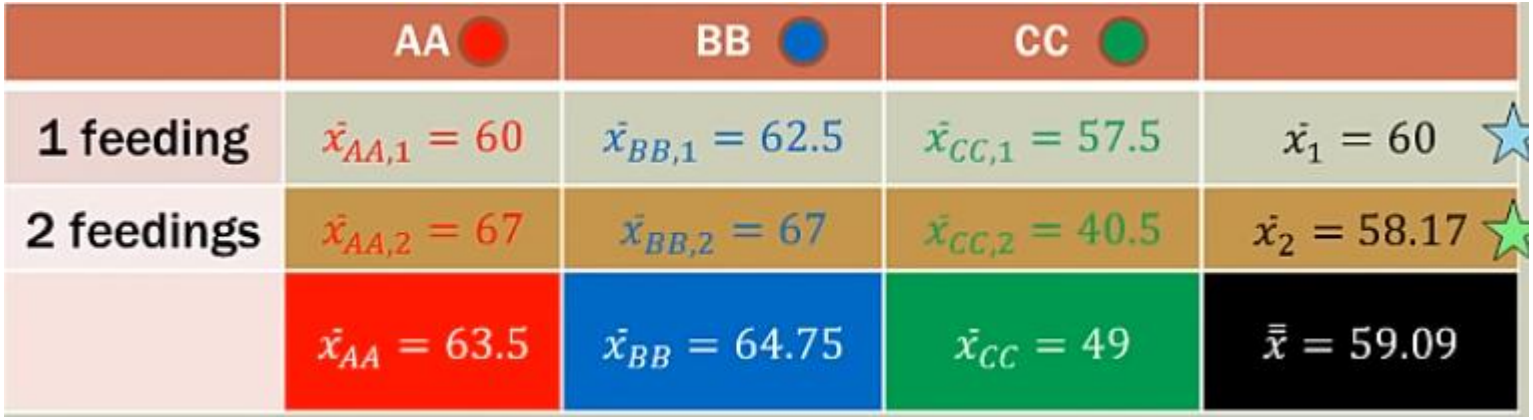




Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	42.1875	1	42.1875	0.578	0.451	4.07265376
Columns	2412.5	2	1206.25	16.526	0.000	3.21994229
Interaction	1362.5	2	681.25	9.333	0.000	3.21994229
Within	3065.625	42	72.9910714			
Total	6882.8125	47				

Example - Two way ANOVA – With Replication





	AA ●	BB ●	CC ●	
1 feeding	$\bar{x}_{AA,1} = 60$	$\bar{x}_{BB,1} = 62.5$	$\bar{x}_{CC,1} = 57.5$	$\bar{x}_1 = 60$ ☆
2 feedings	$\bar{x}_{AA,2} = 67$	$\bar{x}_{BB,2} = 67$	$\bar{x}_{CC,2} = 40.5$	$\bar{x}_2 = 58.17$ ☆
	$\bar{x}_{AA} = 63.5$	$\bar{x}_{BB} = 64.75$	$\bar{x}_{CC} = 49$	$\bar{\bar{x}} = 59.09$

Non-Parametric test

Mann-Whitney test:

It is a non-parametric test that is used to compare two population means that come from the same population, it is also used to test whether two population means are equal or not.

It does not assume any assumptions related to the distribution.

There are, however, some assumptions that are assumed:
1. The sample drawn from the population is random.
2. Independence within the samples and mutual independence is assumed.

Thank You