

## **QMB 6305 Analysis of Variance Project**

## Problem 8 from Midterm – TV watching habits

**Objective:** To understand the variance in TV watching habits across Detroit, San Diego, Miami, Buffalo. Conduct appropriate test for variances across these cities and understand the results.

**ANOVA calculates – Among groups variation/ Within groups variation.**

If this value is high, then there is a high among groups variation which implies there is a difference in the TV on time between these cities.

If the value is low, then there is more unexplained variance within the column than among groups variation, which means there is a lot more unexplained variation within each city than among cities. There is no difference in the TV on time between these cities. We cannot conclude, one or more city has a different TV on time when compared to others.

**About data:** 144 rows, 4 columns – Each column representing a city.

Software used: Minitab

To understand the variation in the “TV ON TIME” across different levels (4 cities)

Analysis of Variance					
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	1458643	486214	103.34	0.000
Error	572	2691240	4705		
Total	575	4149883			

### Analysis of Variance Table:

Source of variation – How much variation is due to the factors?, how much variation is due to errors and randomness?

DF for factors – Total factors – 1 (4-1)

Total - Total observations -1 (144\*4 -1) = 575

Adjusted SS – Total variation due to factors (1458643), Variance due to errors (2691240)

Adjusted MS – Adjusted SS/DF

F value – 486214/4705

P value – Probability for null Hypothesis. **The P value is close to 0, indicating we can REJECT THE NULL HYPOTHESIS**

**Null Hypothesis – All means are equal**

**Alternate Hypothesis – At least one mean is different**

### Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$

*Equal variances were assumed for the analysis.*

Equal variance assumption – Variance in Detroit, Miami, San Diego and Buffalo are equal

### Factor Information

Factor	Levels	Values
Factor	4	Detroit, SanDiego, Miami, Buffalo

### Means

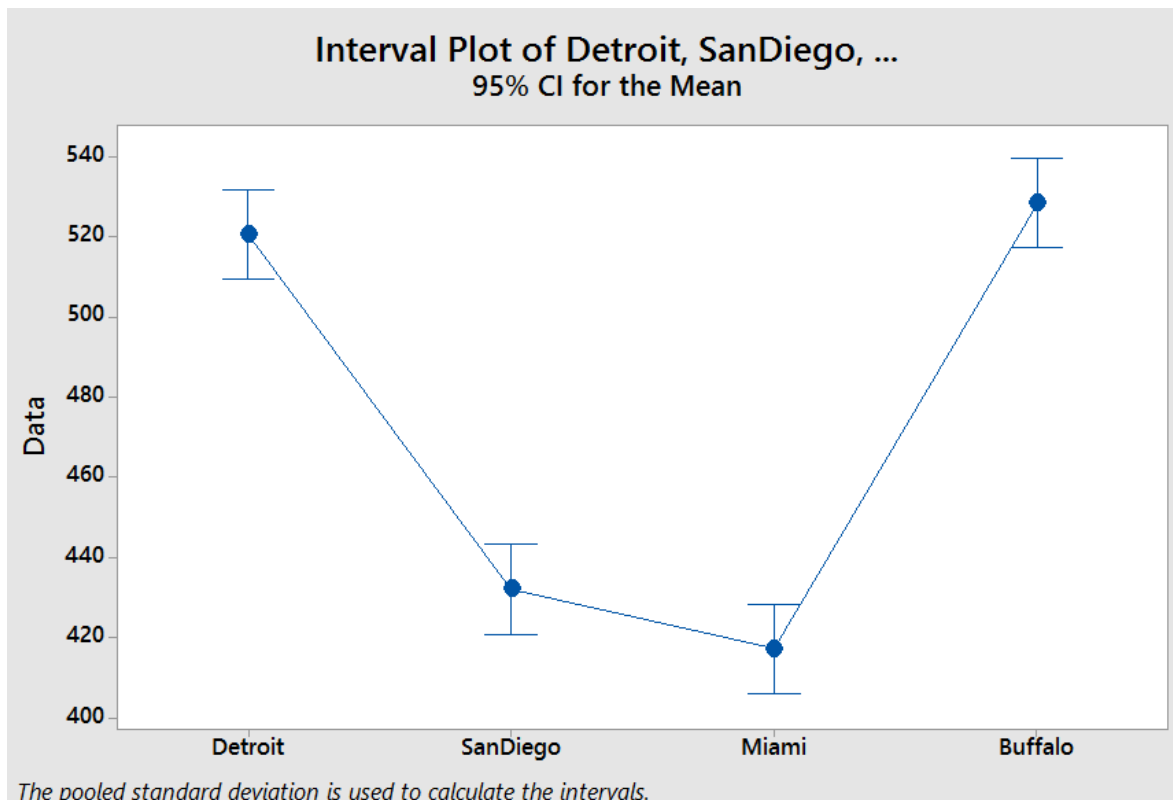
Factor	N	Mean	StDev	95% CI
Detroit	144	520.76	75.94	(509.53, 531.98)
SanDiego	144	432.13	68.14	(420.90, 443.36)
Miami	144	417.28	70.58	(406.06, 428.51)
Buffalo	144	528.55	58.55	(517.32, 539.78)

Pooled StDev = 68.5927

The above table gives data about the average amount of “TV on time” for each city in minutes. The population mean with 95 % Confidence interval is in the final column, for each city.

For ANOVA, the differences in the population mean has to be analyzed.

From the confidence intervals data, its seen that “Some of the confidence intervals overlap”, and “Some do not”. If they overlap, we cannot conclude that there is likely a difference in their population means. If they do not overlap, then there is a difference in the population means



San Diego and Miami overlap. Detroit and Buffalo overlap. But the two combinations do not overlap.

- San Diego is different from Detroit and Buffalo. San Diego and Miami do not seem to have difference.
- Miami is different from Detroit and Buffalo.
- Detroit is different from San Diego and Miami. Detroit and Buffalo do not seem to have difference.
- Buffalo is different from San Diego and Miami.

The confidence interval also from the table above follows the similar pattern like the graph.

## **COMPARISON OF MEANS**

Software: Minitab

### **Tukey Pairwise Comparisons**

#### **Grouping Information Using the Tukey Method and 95% Confidence**

Factor	N	Mean	Grouping
Buffalo	144	528.55	A
Detroit	144	520.76	A
SanDiego	144	432.13	B
Miami	144	417.28	B

*Means that do not share a letter are significantly different.*

The grouping infers that Buffalo and Detroit are in *Group A* and there is not statistical difference between them.

San Diego and Miami are in *Group B* and there is no statistical difference between them.

#### **Tukey Simultaneous Tests for Differences of Means**

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
SanDiego - Detroit	-88.63	8.08	(-109.37, -67.88)	-10.96	0.000
Miami - Detroit	-103.47	8.08	(-124.22, -82.72)	-12.80	0.000
Buffalo - Detroit	7.79	8.08	(-12.96, 28.54)	0.96	0.770
Miami - SanDiego	-14.85	8.08	(-35.60, 5.90)	-1.84	0.256
Buffalo - SanDiego	96.42	8.08	(75.67, 117.17)	11.93	0.000
Buffalo - Miami	111.26	8.08	(90.51, 132.01)	13.76	0.000

*Individual confidence level = 98.95%*

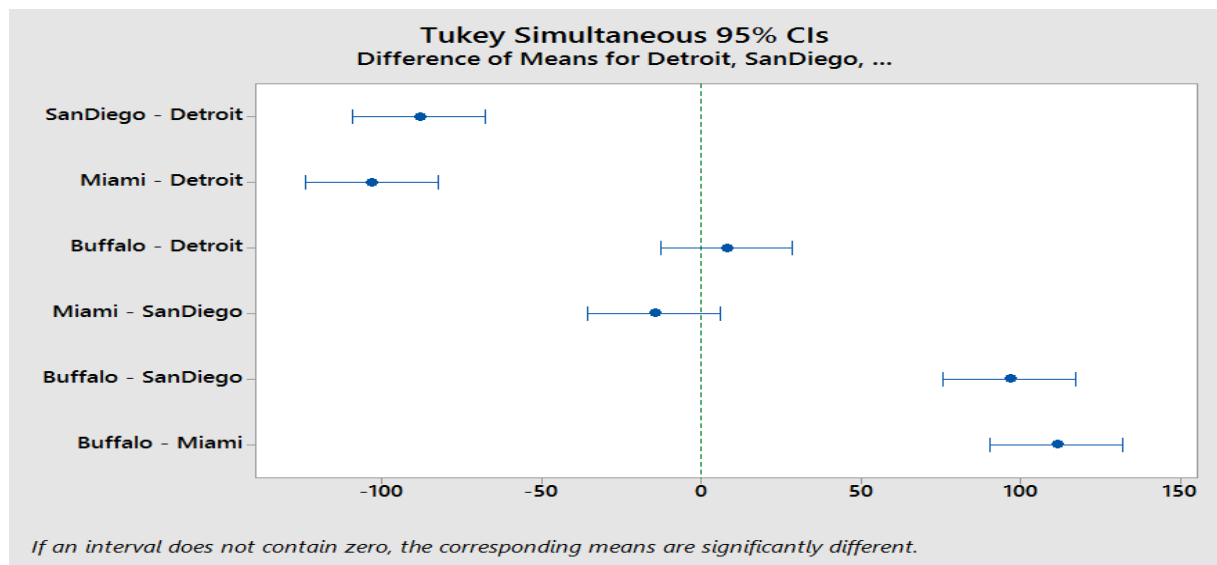
#### **Tukey Simultaneous 95% CIs**

The above table has the differences of means. If the 95% confidence interval contains "0", then there is not enough evidence to conclude there is a difference in the mean.

In the above table, Miami – San Diego and Buffalo – Detroit have a confidence interval that contains "0". The p value for those are too high to reject null hypothesis.

Null hypothesis – There is no difference between the two population means.

Alternate hypothesis – There is a difference between the two population means.



This graph is the confidence interval on the difference between means.

For Buffalo – Detroit and Miami – San Diego, the interval crosses “0” indicating, we cannot say there is a difference between the population means for these pairs. For others, there is evidence to conclude a difference.

### ASSUMPTIONS

- 1) Randomness – It is assumed that the observations are random. There is no specific picking of samples. All other settings remain the same across these cities. All observations are independent of each other
- 2) Normality - Samples for each group are from normally distributed populations.
- 3) Equality of variances – Variances of the populations represented by groups are equal.

### Tests

Method	Test Statistic	P-Value
Multiple comparisons	—	0.010
Levene	2.84	0.037

### Test for Equal Variances: Detroit, SanDiego, Miami, Buffalo

Levene test – for testing equality of variance.

P – value is less than 0.05, meaning we can reject the null hypothesis.

Null hypothesis – All variances are equal.

## Test for Equal Variances: Detroit, SanDiego, Miami, Buffalo

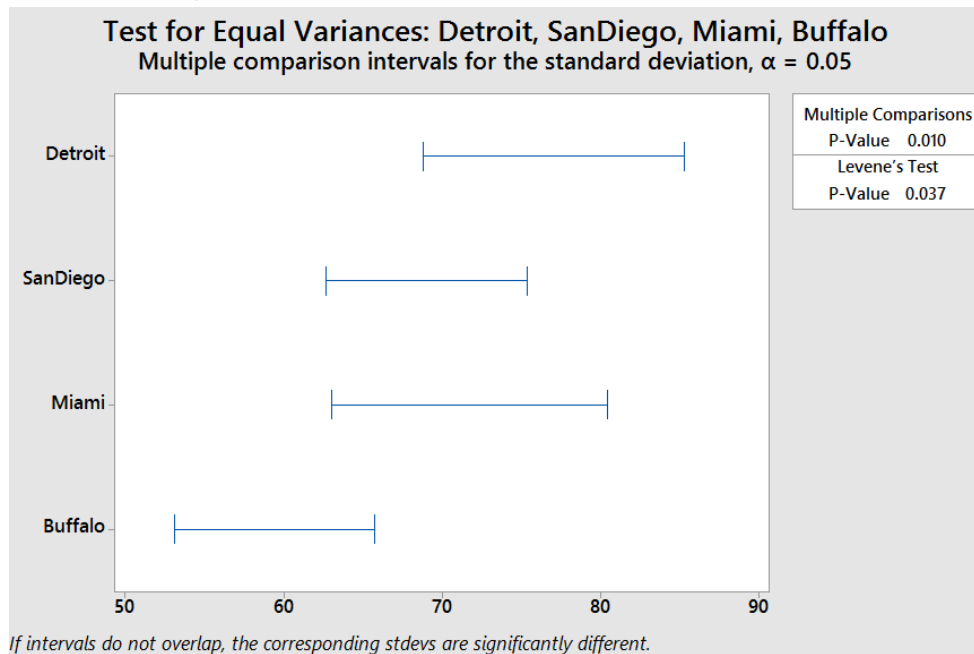
### Method

Null hypothesis All variances are equal  
Alternative hypothesis At least one variance is different  
Significance level  $\alpha = 0.05$

### 95% Bonferroni Confidence Intervals for Standard Deviations

Sample	N	StDev	CI
Detroit	144	75.9384	(66.1297, 88.7413)
SanDiego	144	68.1435	(60.3489, 78.3031)
Miami	144	70.5841	(60.2495, 84.1511)
Buffalo	144	58.5453	(51.4838, 67.7505)

Individual confidence level = 98.75%



The graph indicates that there is no overlap between Buffalo and Detroit. This implies that “Not all variances are equal” and we have evidence to reject the null hypothesis (“All variances are equal”)

This data fails in the assumption that “Variances are equal”.

### Problem 3 from Midterm – Battery life in Sonoran and Minnesotan

**Objective:** To understand the variance in Battery life between Minnesotan and Sonoran. Conduct appropriate test for variances across these conditions.

#### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	1	383.5	383.5	2.59	0.112
Error	72	10664.3	148.1		
Total	73	11047.8			

Null hypothesis – All means are equal.

P value is greater than 0.05, which means we do not have enough evidence to reject the NULL hypothesis.

#### Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$

*Equal variances were assumed for the analysis.*

#### Factor Information

Factor	Levels	Values
Factor	2	Sonoran, Minnesotan

#### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
12.1702	3.47%	2.13%	0.00%

#### Means

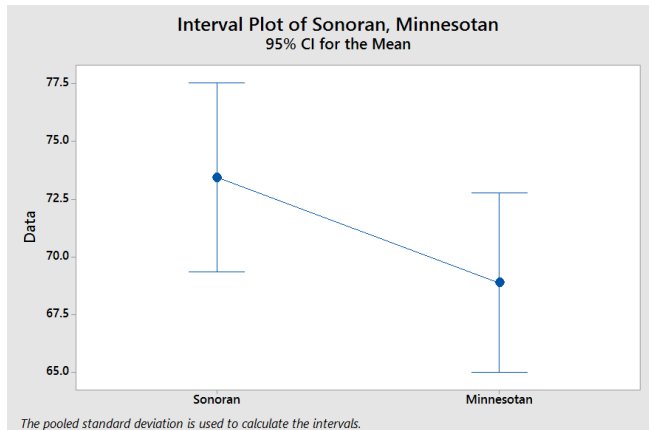
Factor	N	Mean	StDev	95% CI
Sonoran	35	73.46	12.42	(69.36, 77.56)
Minnesotan	39	68.90	11.94	(65.01, 72.78)

*Pooled StDev = 12.1702*

#### Interval Plot of Sonoran, Minnesotan

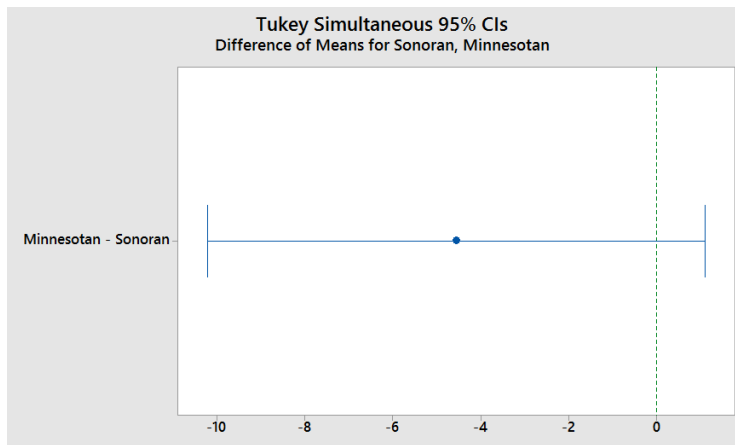
The confidence interval for the population mean for the Battery life in these two city conditions.

Since the confidence interval overlaps, we cannot conclude statistically that there is a difference in the means for these two cities.



There is an overlapping region in the graph, indicating there is not enough evidence to conclude statistical difference between these two cities.

### COMPARISON OF MEANS:



### Tukey Pairwise Comparisons

#### Grouping Information Using the Tukey Method and 95% Confidence

Factor	N	Mean	Grouping
Sonoran	35	73.46	A
Minnesotan	39	68.90	A

Means that do not share a letter are significantly different.

#### Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
Minnesotan - Sonoran	-4.56	2.83	(-10.21, 1.09)	-1.61	0.112

Individual confidence level = 95.00%

#### Tukey Simultaneous 95% CIs

The two cities are grouped in same group “A” indicating there is not a separable difference. The test for difference of means contains “0” in the confidence interval which is also indicated in the graph. The p value is greater than 0.05 and so we cannot reject the null hypothesis.

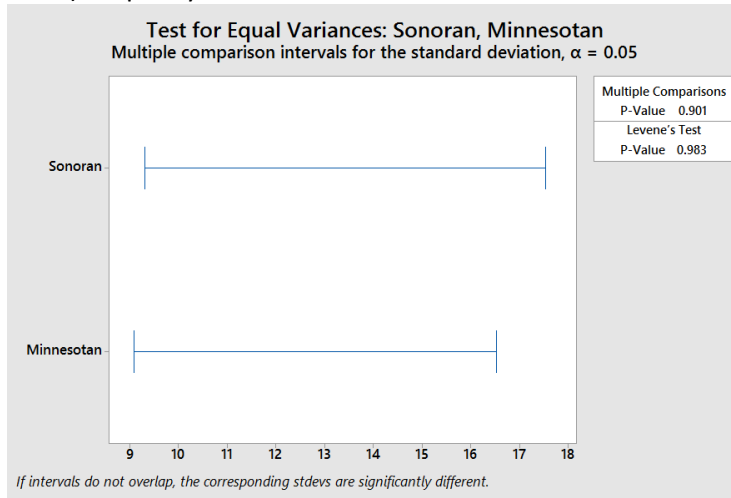
Null hypothesis – The two means are equal.



From the above results, it is evident that there is no enough evidence to say that these cities have difference.

### ASSUMPTIONS:

- 1) Randomness- Observations are random, independent and all other setting is same.
- 2) Normality – Samples are taken from normal distribution
- 3) Equality of variance



### Test for Equal Variances: Sonoran, Minnesotan

#### Method

Null hypothesis	All variances are equal
Alternative hypothesis	At least one variance is different
Significance level	$\alpha = 0.05$

#### 95% Bonferroni Confidence Intervals for Standard Deviations

Sample	N	StDev	CI
Sonoran	35	12.4176	(7.10220, 23.1966)
Minnesotan	39	11.9446	(7.73372, 19.5731)

Individual confidence level = 97.5%

#### Tests

Method	Test Statistic	P-Value
Multiple comparisons	—	0.901
Levene	0.00	0.983

#### Test for Equal Variances: Sonoran, Minnesotan

The test for equality of variance has a p value of 0.983 in which case we cannot reject the Null hypothesis. Null hypothesis - The variances are equal.

Also, the graph has an overlapping region indicating that the variances are equal.

The above data between two cities passes the assumptions for Homogeneity of Variance.

But, ANOVA does not give enough evidence to claim a statistical difference between these two cities.

## Problem 5 from Midterm – Time taken for different photographers

**Objective:** To understand the variance in setting times for photographic sessions among Rick Blaine, CharlieAllNut, FredCDobbs, HarrySteveMorgan. Conduct appropriate test for variances among these photographers and understand the results.

### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	1444.4	481.464	2811.08	0.000
Error	590	101.1	0.171		
Total	593	1545.4			

### Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$

*Equal variances were assumed for the analysis.*

Null hypothesis – All means are equal.

The F value for the model is very high indicating there is a variance among these factors. The p value is less than 0.05 indicating we can reject the null hypothesis. So, not all means are equal.

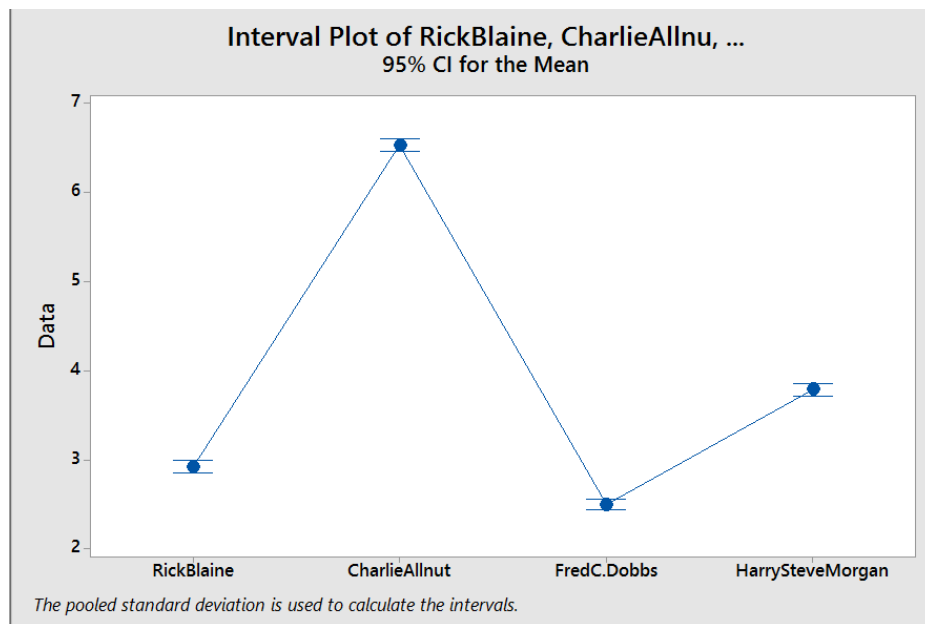
### Means

Factor	N	Mean	StDev	95% CI
RickBlaine	120	2.9233	0.4535	(2.8491, 2.9975)
CharlieAllnut	137	6.5328	0.5381	(6.4634, 6.6023)
FredC.Dobbs	195	2.5005	0.2846	(2.4423, 2.5587)
HarrySteveMorgan	142	3.7866	0.3903	(3.7184, 3.8548)

*Pooled StDev = 0.413852*

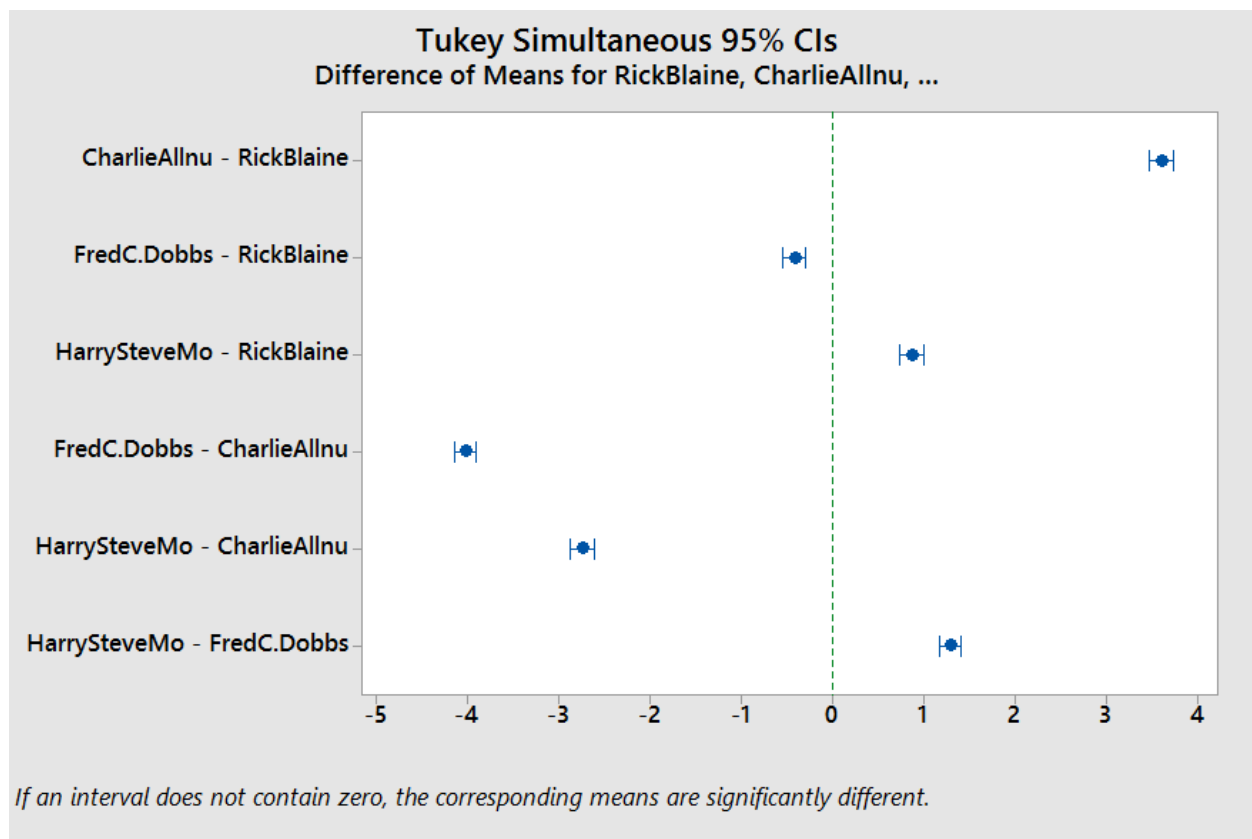
### Interval Plot of RickBlaine, CharlieAllnu, ...

The above table gives information about 95 % confidence interval for the population means for these 4 photographers. From the 95% CI, its evident that there is no overlapping region between the population means indicating that there is a difference among these 4 photographers.



The interval plot for the confidence interval for the population means of these 4 photographers do not have an overlapping region. Fred and Rick are close on their “time” but still there is no overlap. This graph shows that there is statistical difference among these 4 photographers

### **COMPARISON OF MEANS:**



If the interval between any two photographers contains "0" then they are not significantly different. However, no pairwise comparison has the "0" in the region indicating that each of them are statistically different.

## Tukey Pairwise Comparisons

### Grouping Information Using the Tukey Method and 95% Confidence

Factor	N	Mean	Grouping
CharlieAllnut	137	6.5328	A
HarrySteveMorgan	142	3.7866	B
RickBlaine	120	2.9233	C
FredC.Dobbs	195	2.5005	D

*Means that do not share a letter are significantly different.*

### Tukey Simultaneous 95% CIs

Each of the photographer is grouped into a different group namely A, B, C, D indicating that none of them are similar to each other in terms of their means.

#### Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
CharlieAllnu - RickBlaine	3.6095	0.0517	(3.4767, 3.7423)	69.76	0.000
FredC.Dobbs - RickBlaine	-0.4228	0.0480	(-0.5461, -0.2996)	-8.81	0.000
HarrySteveMo - RickBlaine	0.8633	0.0513	(0.7316, 0.9950)	16.82	0.000
FredC.Dobbs - CharlieAllnu	-4.0323	0.0461	(-4.1508, -3.9139)	-87.40	0.000
HarrySteveMo - CharlieAllnu	-2.7462	0.0496	(-2.8734, -2.6190)	-55.41	0.000
HarrySteveMo - FredC.Dobbs	1.2861	0.0457	(1.1689, 1.4033)	28.17	0.000

*Individual confidence level = 98.95%*

#### Tukey Simultaneous 95% CIs

#### Interval Plot of RickBlaine, CharlieAllnu, ...

The test for difference in mean also shows that each of the photographers are statistically different. The difference in the means between two photographers does not contain "0" in any of the cases. They are all positive or negative.

Null Hypothesis – All means are equal

The p value for each observation is less than 0.05 in which case, we can reject the null hypothesis. This test statistically proves that there is a difference in the setting time among these 4 photographers.

### ASSUMPTIONS:

- 1) Randomness- Each observation is independent, random and all other factors remain equal
- 2) Normality - The Samples are taken for normal distribution
- 3) Homogeneity of Variance – Variances represented by the populations are equal

## Method

Null hypothesis	All variances are equal
Alternative hypothesis	At least one variance is different
Significance level	$\alpha = 0.05$

## 95% Bonferroni Confidence Intervals for Standard Deviations

Sample	N	StDev	CI
RickBlaine	120	0.453508	(0.387385, 0.542203)
CharlieAllnut	137	0.538122	(0.462146, 0.638223)
FredC.Dobbs	195	0.284568	(0.248205, 0.330493)
HarrySteveMorgan	142	0.390350	(0.344557, 0.450146)

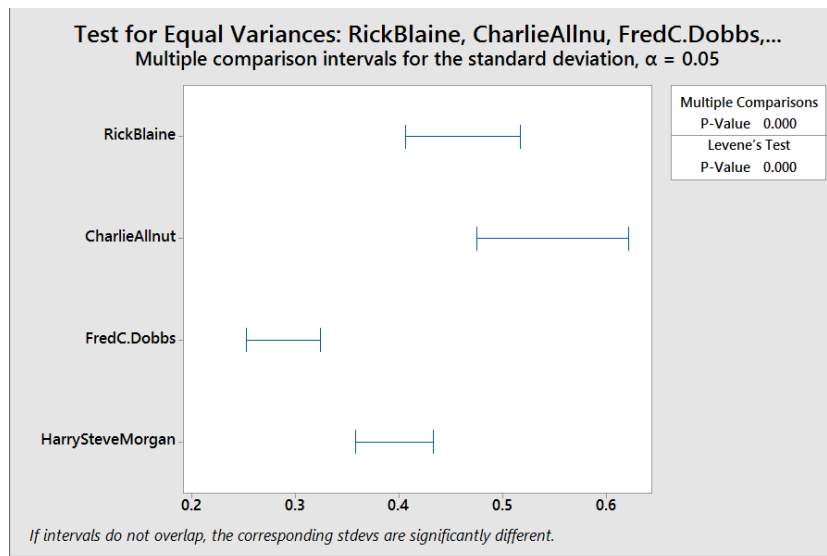
Individual confidence level = 98.75%

## Tests

Method	Test	
	Statistic	P-Value
Multiple comparisons	—	0.000
Levene	18.34	0.000

## Test for Equal Variances: RickBlaine, CharlieAllnu, FredC.Dobbs,...

The p value for the Levene test is close to 0, in which case we can reject the null hypothesis. So, we cannot say that All Variances are equal.



If all intervals overlap, then the variances are equal. Here, there is no overlap between Fred – Rick, Fred – Charlie, Fred – Harry, and Harry – Charlie. So, this proves that Not all Variances are equal.

The ANOVA results show that there is statistical difference in the setting time among these 4 photographers. But the assumptions for ANOVA test failed at the Equality of variances assumption.