SECTION 6: TWO-FACTOR ANOVA

6.1 Research Design and Assumptions of Two-Factor ANOVA

- Factorial design involves two or more factors affecting the response variable and each factor has two or more levels
- When there are two factors affecting the response variable, it is known as two-factor ANOVA
- Just like one-factor ANOVA, there is only one <u>response variable</u>
- There are two factors being tested at the same time to determine whether they have an effect on the variable being measured
- For each of the two factors, there are <u>two or more levels</u> or treatments being applied to the individuals or experimental units being test
- Often the two factors are categorical
- One or both of the factors may be quantitative, having several measurable levels or treatments
    - In this case, two-factor ANOVA is very similar to multiple linear regression
- Some of the <u>advantages setting up one experiment to test the effects of two-factors</u> on the response variable at the same time and analyzing the data with two-factor ANOVA, as opposed to doing two separate experiments for the effects of these two factors and analyzing them separately using one-factor ANOVA are as follows:
    1. Can test the effect of two factors at the same time, thus saving time and expenses
    2. Can compare the significance of the effects of the two factors
    3. Allows testing for interaction between the two factors
- **Main effects** = the effect of each factor considered separately
- **Interaction effect** = interaction between the two factors, that is, the effect of one factor on the response variable depends on the level of the other factor
    - The interaction effect may or may not occur

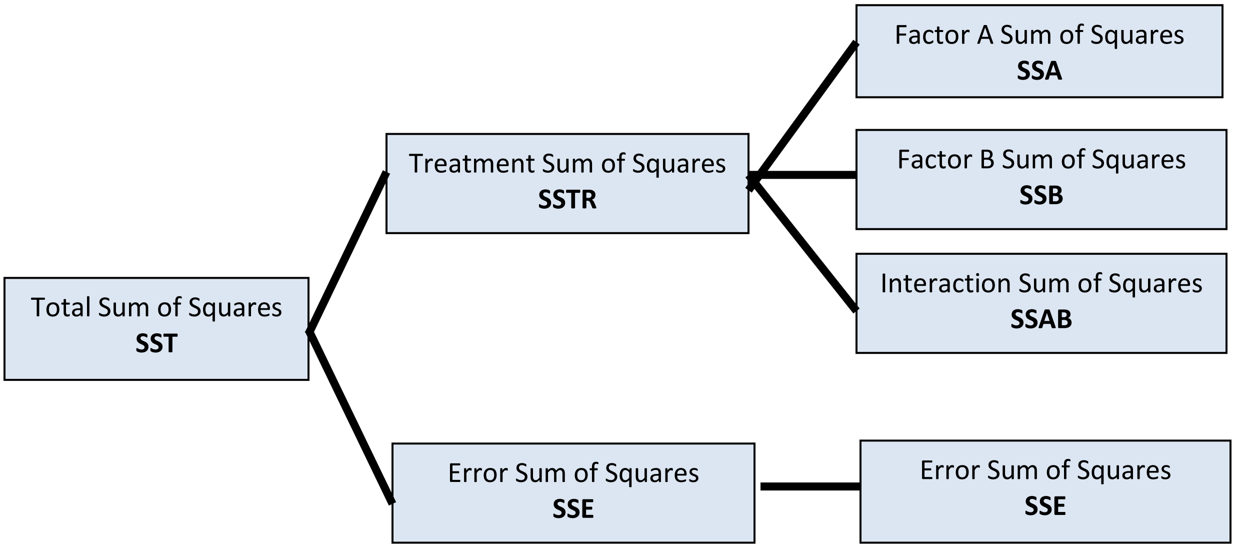
Fixed Effect Factors and Random Effect Factors

- **Fixed Effect Factors** are factors whereby the researcher deliberately fixes the levels of the factor because those are the levels of interest to him/her
- **Random Effect Factors** are factors for which the levels are selected or occur at random from a collection of possible levels

Assumptions for Two-Factor ANOVA

Assumptions (Conditions) for Two-Factor ANOVA

1. **Random sampling**
2. **Independent observations:** The observations of the response variable are independent of one another, though the levels of the factors do not need to be independent
3. **Normal distributions:** For <u>each combination</u> of treatments, the response variable must be normally distributed
4. **Equal standard deviations:** For <u>all combinations</u> of treatments, the standard deviations of the response variable are the equal.



Response = Overall mean + A Main Effect + B Main Effect + AB Interaction Effect + Error

6.2 Two-Factor ANOVA with Interaction (with Replication, Balanced Data)

- Sample size (i.e., number of observations) at least 2 for each combination of treatments
- **<u>Balanced data</u>** means that the sample size is the same for all combinations of treatments

Two-Factor ANOVA Identity for Sums of Squares (balanced data):

Total Sum of Squares = Factor A Sum of Squares + Factor B Sum of Squares
+ AB Interaction Sum of Squares + Error Sum of Squares

SST = SSA + SSB + SSAB + SSE

Two-Factor ANOVA Identity for Degrees of Freedom (balanced data):

df(SST) = df(SSA) + df(SSB) + df(SSAB) + df(SSE)

Or $n - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1) + (n - ab)$

Two-Factor ANOVA Hypothesis Test (With Interaction) (Non-Additive Model)

Purpose: To perform hypothesis tests for the main effects and interaction effects of two factors
Assumptions: Given above

Null and Alternative Hypotheses

Overall Model:	H_0: The overall model is not useful for making predictions
	H_a: The overall model is useful for making predictions
Factor A main effect:	H_0: There is no main effect due to Factor A
	Ha: There is a main effect due to Factor A
Factor B main effect:	H_0: There is no main effect due to Factor B
	Ha: There is a main effect due to Factor B
AB interaction effect:	H_0: The two factors do not interact
	Ha: The two factors interact

ANOVA table for Two-Factor Analysis of Variance (With Interaction) (Non-Additive Model)

Source of variation	SS	df	MS = SS/df	F-statistic
Overall model Corrected model	SSA +SSB +SSAB	df(A) +df(B) +dfAB) = ab – 1	$Corr\ MS = \frac{Corr\ SS}{Corr\ df}$	$F_{Overall} = \frac{Corr\ MS}{MSE}$
Factor *A*	*SSA*	*a* – 1	$MSA = \frac{SSA}{a-1}$	$F_A = \frac{MSA}{MSE}$
Factor *B*	*SSB*	*b* – 1	$MSB = \frac{SSB}{b-1}$	$F_B = \frac{MSB}{MSE}$
AB Interaction	*SSAB*	(*a* – 1)(*b* – 1)	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$F_{AB} = \frac{MSAB}{MSE}$
Error (within)	*SSE*	*n* – *ab*	$MSE = \frac{SSE}{n-ab}$	
Total	***SST***	***n* – 1**		

F (Overall model) = F (Corrected model) $= \frac{\text{Corrected } SS\,/\,(ab-1)}{\text{Error } SS\,/\,(n-ab)} = \frac{\text{Corrected } MS}{MSE}$

$$F_A = \frac{SSA\,/\,(a-1)}{SSE\,/\,(n-ab)} = \frac{MSA}{MSE} \qquad F_B = \frac{SSB\,/\,(b-1)}{SSE\,/\,(n-ab)} = \frac{MSB}{MSE} \qquad F_{AB} = \frac{SSAB\,/\,(a-1)(b-1)}{SSE\,/\,(n-ab)} = \frac{MSAB}{MSE}$$

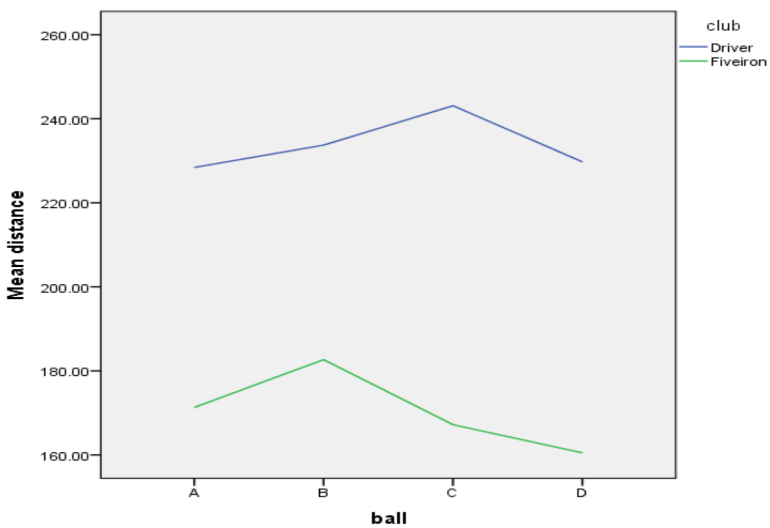
Where: a = number of levels of Factor A
b = number of levels of Factor B
n = total number of observations
= a x b x (no. of replicates per combination of treatments)

Research Problem Involving Two-Factor ANOVA, Followed by Multiple Comparisons Tests
Example on Golf Balls

A golfer investigated the distances that golf balls are hit and how this is affected by the brand of ball used and the club used. He set up an experiment using a completely randomized design. Moreover, it was a balanced design, with 4 replicates per combination of the two factors. He tested the following effects:

1. Effect of the four brands of balls (A, B, C and D)
2. Effect of two types of clubs (driver and five iron)
3. Effect of interaction between club and ball

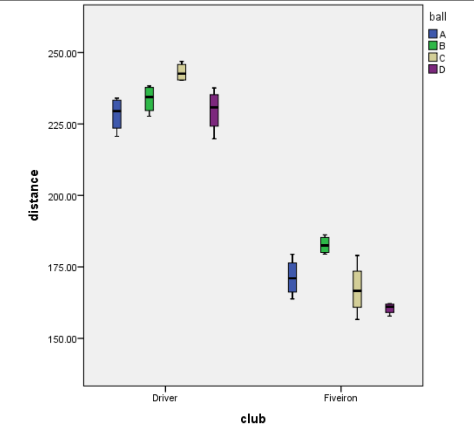
		Ball Brand			
		A	**B**	**C**	**D**
Club Type	**Driver**	226.4	238.3	240.5	219.8
		232.6	231.7	246.9	228.7
		234.0	227.7	240.3	232.9
		220.7	237.2	244.7	237.6
	Five Iron	163.8	184.4	179.0	157.8
		179.4	180.6	168.0	161.8
		168.6	179.5	165.2	162.1
		173.4	186.2	156.6	160.3



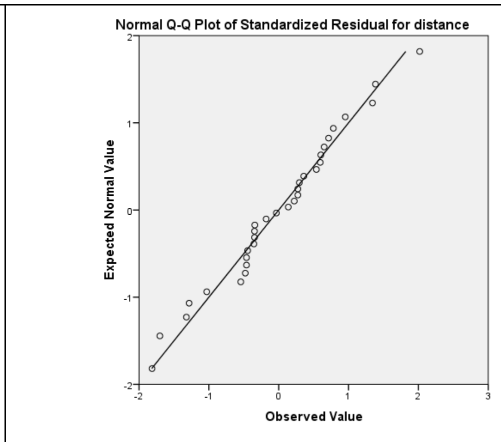
Observe the following in this Line Chart:

1. The huge separation of the two lines indicates a virtually certain effect of the club factor, that is, a difference between the average of the means for the driver and the average of the means for the five iron.
2. The fact that the two lines go considerably up and down from A to B to C to D indicates a likely effect of the brand of the balls.
3. The non-parallel lines indicate an interaction effect between club and ball.
   - The segments from B to C run in completely different directions
   - They don't have to cross; if they are not parallel that indicates interaction

Checking Assumptions



Side-by-side boxplots



Normal probability plot

Levene's Test of Equality of Error Variances[a]

Dependent Variable: distance

F	df1	df2	Sig.
1.269	7	24	.307

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + club + ball + club * ball

Checking Assumptions

>>>>>>>>>>>>

Randomness and Independence

"Completely randomized design" (stated in the question) covers both of these assumptions.

Normality

The above Q-Q Plot shows that the points are reasonably close to a straight line. Thus, the distance variable is approximately normally distributed.

Equal Standard Deviations

Levene's Test gives P = 0.307. Since this is a large P-value (at least > 0.05), H_0 is not rejected. Thus, the standard deviations for all the combinations of the two-factors are not significantly different.

>>>>>>>>>>>>

Note: Fiveiron*C: SD = 9.24, Fiveiron*D: SD = 1.96

The largest SD/smallest SD = 9.24/1.96 = 4.71 > 2

This rule of thumb should NOT be applied for k > 2

SPSS Output for Two-Factor Analysis of Variance (with Interaction)

Tests of Between-Subjects Effects

Dependent Variable: distance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33652.830[a]	7	4807.547	140.689	.000
Intercept	1306819.028	1	1306819.028	38243.115	.000
club	32086.778	1	32086.778	938.996	.000
ball	801.348	3	267.116	7.817	.001
club * ball	764.703	3	254.901	7.459	.001
Error	820.113	24	34.171		
Total	1341291.970	32			
Corrected Total	34472.942	31			

a. R Squared = .976 (Adjusted R Squared = .969)

- Suppose the numbers highlighted in yellow are not given

>>>>>>>>>>

(a) At the 5% significance level, perform the most appropriate test to determine whether the overall model is significant.

H_0: All treatment combinations (driver/ball) have equal means (the overall model is not significant).
Ha: At least two treatment combinations have different means (the overall model is significant).

Overall SS (Corrected SS) = SSA + SSB + SSAB = 32086.778 + 801.348 + 764.703 = 33652.829
(OR)
Overall SS (Corrected SS) = Corrected Total SS – Error SS = 34472.942 – 820.113 = 33652.829
[Note: Corrected Total SS = Total SS – Intercept SS = 1341291.970 – 1306819.028 = 34472.942]

Overall model df = df(A) + df(B) + df(AB) = (2-1) + (4-1) + [(2-1)(4-1)] = 1 + 3 + 3 = 7
(OR)
Overall model df = ab – 1 = (2)(4) – 1 = 7

Total sample size (n) = number of combinations x number of replicates per combination
= 2 club types x 4 ball types x 4 replicates per combination = 2 x 4 x 4 = 32
Error df = n – ab = 32 – (2)(4) = 24

F (Overall model) = F (Corrected model)

$$= \frac{\text{Corrected } SS/(ab-1)}{\text{Error } SS/(n-ab)} = \frac{\text{Corrected } MS}{MSE}$$

$$= \frac{33652.829/7}{820.113/24} = \frac{4807.54700}{34.17138} = 140.689$$

df (Overall model) = [(combinations – 1), (n-ab)]
= [(8 – 1), (32 – (2)(4)] = (7, 24)

P < 0.001 There is extremely strong evidence against Ho.
Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence that the treatment means for distance are not all the same for all club/ball combinations; therefore, the overall model is significant.

(b) At the 5% significance level, perform the most appropriate test to determine whether there is a main effect of club type on mean distance.

H_0: There is no main effect of club type. (The means for all club types are equal).
Ha: There is a main effect of club type. (At least two means for club types are different)

F (Main effect of club type) = $F_A = \dfrac{SSA/(a-1)}{SSE/(n-ab)} = \dfrac{MSA}{MSE}$

$$= \frac{32,086.778/(2-1)}{820.113/(32-(2)(4))} = \frac{32,086.778}{34.17138} = 938.996$$

df (club type) = [(a – 1), (n-ab)] = [(2 – 1), (32 – (2)(4)] = (1, 24)

P < 0.001   There is extremely strong evidence against Ho. Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence that there is a significant main effect of club type, that is, the mean distances are not all the same for the different club types, averaging over all brands of balls.

(c) At the 5% significance level, perform the most appropriate test to determine whether there is a main effect of ball brand on mean distance.

H_0: There is no main effect of ball brand. (The means for all brands of balls are equal.
Ha: There is a main effect of ball brand. (At least two means for ball brands are different)

F (Main effect of ball brand) = $F_B = \dfrac{SSB/(b-1)}{SSE/(n-ab)} = \dfrac{MSB}{MSE}$

$$= \frac{801.348/(4-1)}{820.113/(32-(2)(4))} = \frac{267.11600}{34.17138} = 7.817$$

df (ball brand) = [(b – 1), (n-ab)] = [(4 – 1), (32 – (2)(4)] = (3, 24)

P < 0.001 There is extremely strong evidence against Ho. Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence that there is a significant main effect of ball brand, that is, the mean distances are not all the same for the different ball brands, averaging over the two club types.

(d) At the 5% significance level, perform the most appropriate test to determine whether the effect of club type on mean distance depends on ball brand. (In other words, test whether there is an interaction effect between club type and ball brand.)

H_0: There is no interaction effect between club type and ball brand.(All interaction terms equal 0)
[Ball*Club: A*Driver = B*Driver = C*Driver = D*Driver = A*Five =B*Five = C*Five = D*Five = 0]
Ha: There is an interaction effect between club type and ball brand. (At least one interaction term is not 0)

F (Interaction effect of club type*ball brand)

$$F_{AB} = \frac{SSAB/(a-1)(b-1)}{SSE/(n-ab)} = \frac{MSAB}{MSE}$$

$$= \frac{764.703/(2-1)(4-1)}{820.113/(32-(2)(4))} = \frac{254.901}{34.17138} = 7.459$$

df (Interaction) = [(a – 1)(b-1), (n-ab)] = [(2 – 1)(4 – 1), (32 – (2)(4)] = (3, 24)

0.005 > P > 0.001   There is very strong evidence against Ho. Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence that there is a significant interaction effect of club type times ball brand, that is, at least one interaction term is not 0.

>>>>>>>>>>

Multiple Comparisons

Point Estimates + Confidence Intervals for each Combination of Club Type and Ball Brand

Estimates

Dependent Variable: distance

club	Ball	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Driver	A	228.425	2.923	222.393	234.457
	B	233.725	2.923	227.693	239.757
	C	243.100	2.923	237.068	249.132
	D	229.750	2.923	223.718	235.782
Fiveiron	A	171.300	2.923	165.268	177.332
	B	182.675	2.923	176.643	188.707
	C	167.200	2.923	161.168	173.232
	D	160.500	2.923	154.468	166.532

Pairwise Comparisons: For each ball brand, comparisons between club types

Pairwise Comparisons

Dependent Variable: distance

ball	(I) club	(J) club	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
A	Driver	Fiveiron	57.125*	4.133	.000	48.594	65.656
	Fiveiron	Driver	-57.125*	4.133	.000	-65.656	-48.594
B	Driver	Fiveiron	51.050*	4.133	.000	42.519	59.581
	Fiveiron	Driver	-51.050*	4.133	.000	-59.581	-42.519
C	Driver	Fiveiron	75.900*	4.133	.000	67.369	84.431
	Fiveiron	Driver	-75.900*	4.133	.000	-84.431	-67.369
D	Driver	Fiveiron	69.250*	4.133	.000	60.719	77.781
	Fiveiron	Driver	-69.250*	4.133	.000	-77.781	-60.719

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

- The point estimate for the multiple comparisons
  = Difference between each pair of means (SPSS calls this Mean Difference)
  = $(x_i - x_j)$
- For example, difference between the means for Ball A, Driver and Ball A, Fiveiron
  $= (x_i - x_j) = (228.425 - 171.300) = 57.125$ (Shown in the table above)

Pairwise Comparisons for Interactions (Club*Ball)
(Shown in this table for each club type, comparisons between ball brands)

- For example, the first comparison is: (Driver*Ball A compared with Driver*Ball B)

Pairwise Comparisons

Dependent Variable: distance

club	(I) ball	(J) ball	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Driver	A	B	-5.300	4.133	1.000	-17.184	6.584
		C	-14.675*	4.133	.010	-26.559	-2.791
		D	-1.325	4.133	1.000	-13.209	10.559
	B	A	5.300	4.133	1.000	-6.584	17.184
		C	-9.375	4.133	.196	-21.259	2.509
		D	3.975	4.133	1.000	-7.909	15.859
	C	A	14.675*	4.133	.010	2.791	26.559
		B	9.375	4.133	.196	-2.509	21.259
		D	13.350*	4.133	.021	1.466	25.234
	D	A	1.325	4.133	1.000	-10.559	13.209
		B	-3.975	4.133	1.000	-15.859	7.909
		C	-13.350*	4.133	.021	-25.234	-1.466
Fiveiron	A	B	-11.375	4.133	.067	-23.259	.509
		C	4.100	4.133	1.000	-7.784	15.984
		D	10.800	4.133	.092	-1.084	22.684
	B	A	11.375	4.133	.067	-.509	23.259
		C	15.475*	4.133	.006	3.591	27.359
		D	22.175*	4.133	.000	10.291	34.059
	C	A	-4.100	4.133	1.000	-15.984	7.784
		B	-15.475*	4.133	.006	-27.359	-3.591
		D	6.700	4.133	.709	-5.184	18.584
	D	A	-10.800	4.133	.092	-22.684	1.084
		B	-22.175*	4.133	.000	-34.059	-10.291
		C	-6.700	4.133	.709	-18.584	5.184

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

>>>>>>>>>>

Means Comparisons Diagrams (based on the Bonferroni Method)

Diagram for Driver

A	D	B	C
228.425	229.750	233.725	243.100

We can be 95% confident that, when using the Driver, the mean distance for Ball C is different than for Balls A and D. No other differences are significant.

Diagram for Five Iron

D	C	A	B
160.500	167.200	171.300	182.675

We can be 95% confident that, when using the Five Iron, the mean distance for Ball B is different than for Balls D and C. No other differences are significant.

>>>>>>>>>>

- The very different results of the multiple comparisons for Driver as opposed to Five Iron (since the ascending order of the means are different) also <u>indicates interaction</u> between club type and ball type
- These could be joined together to compare all 8 means in one diagram, but there would be complete separation between the means for Driver and means for Five Iron.

Means Comparisons for Interactions

- The multiple comparisons for interactions on the previous page do not show all of them because it just gives them separately for the two club types.
- Combinations = 4 x 2 = 8, so there are actually [8(8-1)]/2 = 28 multiple comparisons for interactions

6.3 Non-Additive (With Interaction) and Additive (Without Interaction) Models in Two-Factor ANOVA

- Non-additive Model (includes interaction) (Full Model)
- Additive Model (does not include interaction) (Reduced Model)
- Often compared using the Extra Sum-of-Squares F-test

Two-Factor ANOVA Hypothesis Test (Without Interaction) (Reduced Model) (Additive Model)

Purpose: To perform hypothesis tests for the main effects of two factors
Assumptions: Same as for Non-Additive Model

Null and Alternative Hypotheses

Overall Model: H_0: The overall model is not useful for making predictions
H_a: The overall model is useful for making predictions
Factor A main effect: H_0: There is no main effect due to Factor A
Ha: There is a main effect due to Factor A
Factor B main effect: H_0: There is no main effect due to Factor B
Ha: There is a main effect due to Factor B

ANOVA table for Two-Factor Analysis of Variance (Without Interaction) (Additive Model)

Source of variation	SS	df	MS = SS/df	F-statistic
Overall model Corrected model	SSA +SSB	df(A) +df(B)	$Corr\ MS = \dfrac{Corr\ SS}{Corr\ df}$	$F_{Overall} = \dfrac{Corr\ MS}{MSE}$
Factor *A*	*SSA*	*a* – 1	$MSA = \dfrac{SSA}{a-1}$	$F_A = \dfrac{MSA}{MSE}$
Factor *B*	*SSB*	*b* – 1	$MSB = \dfrac{SSB}{b-1}$	$F_B = \dfrac{MSB}{MSE}$
Error (within)	*SSE*	df(Total)-df(A) -df(B) =(n-1) -(a-1) -(b-1)	$MSE = \dfrac{SSE}{\text{Error } df}$	
Total	***SST***	***n* – 1**		

Note: For an Additive Model, Overall model df ≠ ab – 1 and Error df ≠ n – ab.

F (Overall model) = F (Corrected model) $= \dfrac{\text{Corrected } SS / \left[(a-1)+(b-1)\right]}{SSE / \text{Error } df} = \dfrac{\text{Corrected } MS}{MSE}$

$$F_A = \frac{SSA/(a-1)}{SSE/\text{Error } df} = \frac{MSA}{MSE} \qquad F_B = \frac{SSB/(b-1)}{SSE/\text{Error } df} = \frac{MSB}{MSE}$$

Where: $\text{Error } df = (n-1)-(a-1)-(b-1)$
a = number of levels of Factor A
b = number of levels of Factor B
n = total number of observations
= a x b x (no. of replicates per combination of treatments)

Previous Example on Golf Balls (Comparing Non-Additive and Additive Models)

A golfer investigated the distances that golf balls are hit and how this is affected by the brand of ball used and the club used and the interaction between ball brand and club type.

Consider the following two models

$$\text{Model 1}: \mu(\text{Distance} \mid \text{Club, Ball}) = \beta_0 + \text{Club} + \text{Ball}$$

[No interaction – Additive model >>> Reduced model]

$$\text{Model 2}: \mu(\text{Distance} \mid \text{Club, Ball}) = \beta_0 + \text{Club} + \text{Ball} + (\text{Club*Ball})$$

[With Interaction – Non-additive model >>> Full model]

Model 1: Additive Model (No Interaction – Reduced Model)

Tests of Between-Subjects Effects

Dependent Variable: distance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	32888.127	4	8222.03175	140.076	.000
club	32086.778	1	32086.778	546.652	.000
ball	801.348	3	267.116	4.551	.010
Error	1584.816	27	58.6969		
Corrected Total	34472.942	31			

a. R Squared = .976 (Adjusted R Squared = .969)

Full (Non-Additive) Model (with Interaction)	Reduced (Additive) Model (without Interaction)
n = 2 clubs x 4 balls x 4 replicates = 32	n = 2 clubs x 4 balls x 4 replicates = 32 (same)
Overall model df = df(A) + df(B) + df(AB) = (a – 1) + (b – 1) + (a -1)(b-1) = (2 – 1) + (4 – 1) + (2-1)(4-1) = 1 + 3 + 3 = 7 = ab – 1 = (2)(4) – 1 = 7	Overall model df = df(A) + df(B) = (a – 1) + (b – 1) = (2 – 1) + (4 – 1) = 1 + 3 = 4
Error df = n – ab = 32 – (2)(4) = 24	Error df = df(Total) – df(A) – df(B) = (n – 1) – (a – 1) – (b – 1) = (32 – 1) – (2 – 1) – (4 – 1) = 31 – 1 – 3 = 27

Note also:

SS_E (reduced) = SSAB (full) + SS_E from the full model (repeated below)
= 764.703 + 820.113 = 1584.816

And: df_E (reduced) = df_E (Full) + df(interaction) from the full model (repeated below)
= 24 + 3 = 27

And: SS Corrected (Reduced) = SS Corrected (Full) – SSAB
= 33652.830 – 764.703
= 32888.127

The Model from Page 5 is Repeated Here Below and can be called:

Model 2: Non-Additive Model (With Interaction – Full Model)

Tests of Between-Subjects Effects

Dependent Variable: distance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33652.830[a]	7	4807.547	140.689	.000
Intercept	1306819.028	1	1306819.028	38243.115	.000
club	32086.778	1	32086.778	938.996	.000
ball	801.348	3	267.116	7.817	.001
club * ball	764.703	3	254.901	7.459	.001
Error	820.113	24	34.171		
Total	1341291.970	32			
Corrected Total	34472.942	31			

a. R Squared = .976 (Adjusted R Squared = .969)

Extra Sum-of-Squares F-test

- This is repeated from previous sections of the notes.

Extra Sum-of-Squares F-Test

Null and alternative hypotheses:

H_0: Reduced model

H_a: Full model

Calculations for Extra Sum-of Squares F-test:

Extra Sum-of-Squares = SS_E (reduced) - SS_E (full)

Extra df = df_E (reduced) – df_E (full)

$$F = \frac{Extra \quad SS / Extra \quad df}{MS_E(\text{Full} \quad \text{model})}$$

$$= \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)}$$

Examine the distribution of the F-table at:

$$df = (Extra\_df, n-k)$$

Recall that, residual (error) = observed value – estimated value

Therefore, residual sum of squares or error sum of squares is:

$$SS_E = \sum (\text{observed value – estimated value})^2$$

Testing for An Interaction Effect by comparing an Additive and a Non-additive Model

(a) By comparing the additive model and the non-additive models presented in the tables above, at the 5% significance level, perform the most appropriate test to determine whether the effect of club type depends on ball brand (in other words, whether there is an interaction effect between club type and ball brand), after accounting for club type and ball brand.

>>>>>>>>>>

$H_0 : \mu(\text{Distance} \mid \text{Club, Ball}) = \beta_0 + \text{Club} + \text{Ball}$

[No interaction – Additive model >>> Reduced model (Model 1)]

$H_a : \mu(\text{Distance} \mid \text{Club, Ball}) = \beta_0 + \text{Club} + \text{Ball} + (\text{Club*Ball})$

[With Interaction – Non-additive model >>> Full model (Model 2)]

$$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)}$$

$$F = \frac{[1584.816 - 820.113]/[27-24]}{820.113/24} = \frac{764.703/3}{34.1714} = \frac{254.901}{34.1714} = 7.459$$

$$df = [Extra\ df, df_E(Full)] = [(27-24), 24] = (3, 24)$$

0.005 > P > 0.001; there is very strong evidence against Ho

Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence that there is a significant interaction effect of club type times ball brand, after accounting for club type and ball brand, that is, at least one interaction term is not 0.

Compare with the approach used on page 7: The F-statistic, df, and P-value are exactly the same.

(b) Now suppose the Interaction is ignored, use the additive model to determine whether either club type or ball brand have an effect on mean distance. Perform the test at the 5% significance level.

H_0: Neither factor (club type, ball brand) has an effect on mean distance.

Ha: At least one factor (club type or ball brand) has an effect on mean distance

$$F = \frac{[32086.778 + 801.348]/[(2-1)+(4-1)]}{1584.816/(27)}$$

$$F = \frac{[32{,}888.126]/[4]}{1584.816/(27)} = \frac{8222.0315}{58.6969} = 140.076$$

$df = (4, 27)$ P < 0.001 There is extremely strong evidence against Ho.

Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that either club type or ball brand or both have an effect on mean distance.

Note: Since the interaction term is omitted and the entire reduced model is just about the effect of club type and ball type, we get the same results if we test for the overall (reduced) model (corrected model).

>>>>>>>>>>

Compare the Results of the non-additive and additive models

	Full Model, OR Non-additive Model (Interaction)	**Reduced Model, OR Additive Model (No interaction)**
Overall Model	F = 140.689 Df = (7, 24) P-value = 0.00000000 $F_{0.001}$ = 5.23 (may be more significant than additive)	F = 140.076 Df = (4, 27) P-value = 0.00000000 $F_{0.001}$ = 6.33
Effect of Club type	F = 938.996 Df = (1, 24) P-value = 0.00000000	F = 546.652 Df = (1, 27) P-value = 0.00000000
Effect of Ball brand	F = 7.817 Df = (3, 24) P-value = 0.000824	F = 4.551 Df = (3, 27) P-value = 0.010478
Effect of interaction	F = 7.459 Df = (3, 24) P-value = 0.001074	**None**

Note:

1. The overall model and the effect of club type and ball brand were all more significant with the non-additive model (with interaction) than the additive model.
2. This is because, when the interaction term is significant (such as in this example), the interaction model is more accurate and more effective in showing significant effects.
3. If we ignore the interaction term (when actually it is significant), it becomes a confounding variable.

Example on Extra Sum-of-Squares F-test in Single-Factor ANOVA and Two-Factor ANOVA

Extra Sum-of-Squares F-test in Single-Factor ANOVA (Nested Design)
In a certain locality, middle-aged people between the ages of 40-50 years old had been members of two different groups for several years, an exercise group and a group that watched movies together. Based on random samples taken from these groups, a study was conducted to determine the possible effect of the activity group (exercise group and movie group) and gender (males and females) on systolic blood pressure (SBP) readings. At the 5% significance level, perform the most appropriate test, showing all steps, to determine whether there is a difference in mean SBP between females and males, after accounting for activity group (the exercise/movie groups). You may consider using any of the SPSS output shown in Tables 1–4 below. Assume that all assumptions are met for the required analyses.

Parameters are defined as follows:

μ_{E-F} = Exercise group-females

μ_{E-M} = Exercise group-males

μ_{M-F} = Movie group-females

μ_{M-M} = Movie group-males

Table 1: Summary statistics of SBP for the four groups, based on activity group and gender.

Descriptives						
SBP						
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Exercise group-Females	9	121.33	5.766	1.922	116.90	125.77
Exercise group-Males	9	117.44	3.779	1.260	114.54	120.35
Movie group-Females	9	128.11	6.623	2.208	123.02	133.20
Movie group-Males	9	135.44	7.732	2.577	129.50	141.39
Total	36	125.58	9.104	1.517	122.50	128.66

Table 2: The overall ANOVA table for comparison of mean SBP between the four groups.

ANOVA					
SBP					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1691.417	3	563.8056	14.91878	.000
Within Groups	1209.333	32	37.79167		
Total	2900.75	35			

Table 3: ANOVA table for comparison of SBP between the exercise and movie groups (ignoring gender).

ANOVA					
SBP					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1381.361	1	1381.361	30.91129	0.00000322
Within Groups	1519.389	34	44.68791		
Total	2900.750	35			

Table 4: ANOVA table for comparison of mean SBP between females and males (ignoring the exercise/movie groups).

ANOVA					
SBP					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	26.69444	1	26.69444	0.315795	0.5778
Within Groups	2874.056	34	84.53105		
Total	2900.75	35			

>>>>>>>>>>

If there is no difference in mean SBP between males and females in the exercise group,

then: $\mu_{E-F} = \mu_{E-M}$

If there is no difference in mean SBP between males and females in the movie group,

then: $\mu_{M-F} = \mu_{M-M}$

$H_0 : \mu_{E-F} = \mu_{E-M}$ and $\mu_{M-F} = \mu_{M-M}$

[Reduced model: Two-mean model (Table 3)]

$H_a : \mu_{E-F}, \mu_{E-M}, \mu_{M-F}, \mu_{M-M}$

[Full model: Four-mean model (Table 2)]

$$df_E(full) = n - k = 36 - 4 = 32$$

$$df_E(reduced) = n - k = 36 - 2 = 34$$

$$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)}$$

$$= \frac{[1519.389 - 1209.333]/[34-32]}{1209.333/32} = \frac{310.056/2}{37.79166} = \frac{155.02800}{37.79166} = 4.102$$

F-distribution, with $df = [Extra\ df, df_E(Full)] = [df_E(reduced) - df_E(full), n-k]$

$$df = [(n-2)-(n-4), n-4] = [(36-2)-(36-4), 36-4] = [2,32] \approx [2,30]$$

P-value: 0.025 < P < 0.05, which provides strong evidence against the null hypothesis.
Since P < α (0.05), reject H_0

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a difference in mean SBP between females and males, after accounting for activity group (the exercise/movie groups).

>>>>>>>>>>

Two-Factor ANOVA: Non-Additive (With Interaction) and Additive (Without Interaction) Models

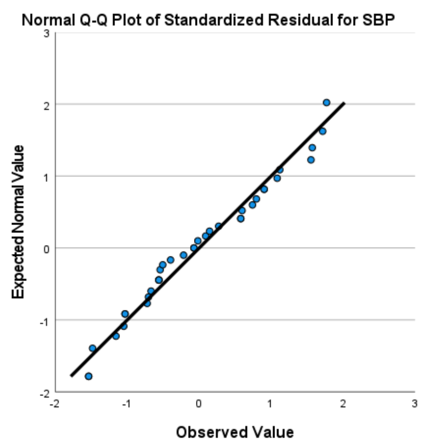
Activity group (exercise versus movie) and gender can also be considered as **two factors** possibly affecting the response variable (SBP). In this case, we can also test whether the interaction between activity group and gender has an effect on SBP.

Checking Assumptions

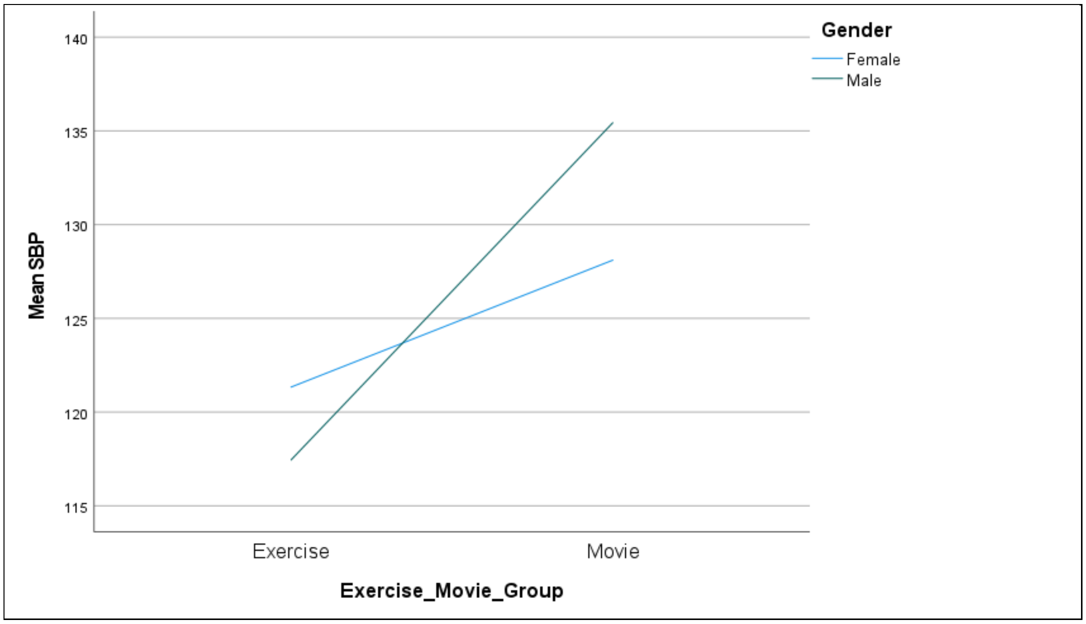
Levene's Test of Equality of Error Variances[a,b]

	Levene Stat	df1	df2	Sig.
Based on Mean	1.916	3	32	.147
Based on Median	1.686	3	32	.190
Based on Median and with adjusted df	1.686	3	27.8	.193
Based on trimmed mean	1.887	3	32	.152

Levene's Test



Normal probability plot



Line Plot: Non-parallel lines indicate interaction.

Two-Factor ANOVA: Non-Additive Model (With Interaction)

Tests of Between-Subjects Effects					
Dependent Variable: SBP					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1691.417[a]	3	563.806	14.919	0.000003
Intercept	567762.250	1	567762.250	15023.477	<.001
Activity Group	1381.361	1	1381.361	36.552	<.001
Gender	26.694	1	26.694	.706	.407
Activity Group * Gender	283.361	1	283.361	7.498	.010
Error	1209.333	32	37.792		
Total	570663.000	36			
Corrected Total	2900.750	35			
a. R Squared = .583 (Adjusted R Squared = .544)					

Note: Interaction Effect

F = 7.498, df = (1,32), P = 0.010

Two-Factor ANOVA: Additive Model (Without Interaction)

Tests of Between-Subjects Effects					
Dependent Variable: SBP					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1408.056[a]	2	704.028	15.564	0.000017
Intercept	567762.250	1	567762.250	12551.902	<.001
Activity Group	1381.361	1	1381.361	30.539	<.001
Gender	26.694	1	26.694	.590	.448
Error	1492.694	33	45.233		
Total	570663.000	36			
Corrected Total	2900.750	35			
a. R Squared = .485 (Adjusted R Squared = .454)					

Testing for An Interaction Effect by comparing the Additive and Non-additive Models
By comparing the additive and the non-additive models presented in the tables above, at the 5% significance level, perform the most appropriate test to determine whether the interaction between activity group and gender has an effect on SBP, after accounting for activity group and gender.

$$H_0 : \mu(\text{SBP} \mid \text{Group, Gender}) = \beta_0 + \text{Group} + \text{Gender}$$

[No interaction – Additive model >>> Reduced model]

$$H_a : \mu(\text{SBP} \mid \text{Group, Gender}) = \beta_0 + \text{Group} + \text{Gender} + (\text{Group*Gender})$$

[With Interaction – Non-additive model >>> Full model]

$$F = \frac{[SS_E(reduced) - SS_E(full)]/[df_E(reduced) - df_E(full)]}{SS_E(full)/df_E(full)}$$

$$F = \frac{[1492.694 - 1209.333]/[33-32]}{1209.333/32} = \frac{283.361/1}{37.79166} = 7.498$$

$$df = [Extra\ df, df_E(Full)] = [(33-32), 32] = (1, 32)$$

P-value: 0.01 < P < 0.025. There is strong evidence against H_0.
Since P < α (0.05), reject H_0.
Conclusion: At the 5% significance level, the data provide sufficient evidence that the interaction between activity group and gender has an effect on SBP, after accounting for activity group and gender.

Note again: The interaction term in the non-additive model and the ESS F-test comparing the non-additive and additive models give the exactly the same results: F-statistic, df, and P-value.

Comparison of Extra Sum-of-Squares F-test in Single-Factor ANOVA and Two-Factor ANOVA

- The Extra Sum-of-Squares F-test based on single-factor ANOVA was able to show that gender has an effect on SPB after accounting for the activity group (exercise versus movie). However, two-way ANOVA could not do this.
- Two-way ANOVA could be used to test for the interaction between activity group and gender. However, ESS F-test based on single-factor ANOVA may suggest a possible interaction effect, but it would not be possible to use it to objectively evaluate if there is an interaction effect.

6.4 Randomized Block Design

- Analyzed with Randomized Block Analysis of Variance (ANOVA)
- Considered as an extension of the paired design
- A special type of Two-Factor ANOVA where the “block” factor is not of interest to the researcher.
- One advantage is that it allows the researcher to eliminate the effect of the block factor so that it does not affect the real factor of interest.
- Thus, it is much more powerful in testing the effect of the factor of interest than ordinary ANOVA

Blocks in space

- Suppose an experiment was conducted to determine whether there is a difference in the effectiveness of four new types of fertilizers (A, B, C, and D)
- However, there is a gradient of conditions in the test area due to a gradual slope towards a river
- Experimental area is divided into blocks such that it can be assumed that the conditions are homogenous within each block, even though conditions vary among blocks.
    - Thus, it eliminates the effect of extraneous variables in space.
- In each block, each treatment is represented once.

Gradient in moisture &nutrients

Blocks (in space)				
1	C	A	D	B
2	B	D	A	C
3	B	C	D	A
4	D	A	B	C
5	A	C	D	B
	RIVER			

Blocks in time

- Eliminate the effect of time on the observations
    - E.g., If a researcher wants to compare 4 sites and he cannot take several measurements in all sites at the same time, he can take 1 measurement in each site every month
- This will eliminate the effect of temporal variation (e.g., seasonal variation or day-to-day variation) on the results.

Gradient in Time

	Abundance of birds			
Blocks (in time)	**Site A**	**Site B**	**Site C**	**Site D**
Oct	11	8	9	15
Nov	13	10	12	16
Jan	4	2	3	7
Feb	9	6	7	14
March	20	16	17	25

Applications in Medical Sciences, Education, Psychology, Etc.

- The before-and-after “treatment” can be extended to monitoring patients every few hours, once a month, etc., or subjects during some intervention program in education or psychology
- In this case, the blocks are the subjects or patients
- Greatly increases the power of the test in detecting responses of subjects/patients to treatments or programs

Randomized Block ANOVA

Extra Assumption (in addition to the other assumptions of two-way ANOVA:
There is no significant interaction between the main factor of interest (treatment) and the blocks factor (which is not of interest). This can be checked graphically with a line graph.

Null and alternative hypotheses for the factor of interest:
H_0: There is no significant difference between population means of the main factor of interest (treatment)
H_a: There is a significant difference between population means of the main factor of interest (treatment)

Null and alternative hypotheses for blocks (not always tested because not main interest):
H_0: There is no effect of the block factor.
Ha: There is an effect of the block factor

Calculations for Randomized Block ANOVA:

Sum of Squares	Defining formula
Treatment (SSTR)	$SSTR = \sum_{i=1}^{k} b(\bar{x}_{Ti} - \bar{x})^2$
Block (SSBL)	$SSBL = \sum_{i=1}^{b} k(\bar{x}_{Bj} - \bar{x})^2$
Total (SST)	$SST = \sum_{i,j} (x_{ij} - \bar{x})^2$
Error (SSE)	$SSE = SST - SSTR - SSBL$

Source of variation	SS	df	MS = SS/df	F-statistic
Treatment	SSTR	k – 1	MSTR = SSTR / (k – 1)	F = MSTR/MSE
Block	SSBL	b - 1	MSBL = SSBL / (b – 1)	F = MSBL/MSE
Error	SSE	(k – 1)(b – 1)	MSE = SSE / (k – 1)(b – 1)	
Total	SST	n - 1		

Note: Error df = (k – 1)(b – 1) = n – k – b + 1 = Total df – (Treatment df + Blocks df)

$$F_{Treatment} = \frac{SSTR / (k-1)}{SSE / (k-1)(b-1)} = \frac{MSTR}{MSE}$$

$$df = [(k-1), (k-1)(b-1)]$$

$$F_{Blocks} = \frac{SSBL / (b-1)}{SSE / (k-1)(b-1)} = \frac{MSBL}{MSE}$$

$$df = [(b-1), (k-1)(b-1)]$$

Where k = number of treatments, b = number of blocks, n = kb = total number of observations

Research Problem: An experiment was conducted to test the effectiveness of four types of fertilizers on eggplants (*Solanum melongena*). The test area was a low-lying area near a river. Eggplants are particularly sensitive to soil fertility and structure as well as soil moisture. Therefore, the gradients in these factors towards the river would have affected the experiment. Thus, a randomized block design was used to eliminate the effect of these gradients.

Table of raw data, including treatment means, block means and grand mean (Units = kg/m^2):

Block	Fertilizer Treatment				Block Means
	A	B	C	D	
1	2.7	2.8	2.9	2.8	$\bar{x}_{B1} = 2.8$
2	2.9	3.0	3.2	2.9	$\bar{x}_{B2} = 3.0$
3	3.0	3.1	3.3	3.0	$\bar{x}_{B3} = 3.1$
4	3.3	3.3	3.6	3.4	$\bar{x}_{B4} = 3.4$
5	3.3	3.4	3.5	3.4	$\bar{x}_{B5} = 3.4$
Treatment Means	$\bar{x}_{T1} = 3.04$	$\bar{x}_{T2} = 3.12$	$\bar{x}_{T3} = 3.30$	$\bar{x}_{T4} = 3.10$	$\bar{x} = 3.14$

Calculate Four Sums of Squares

Treatment SS

$$SSTR = \sum_{i=1}^{k} b(\bar{x}_{Ti} - \bar{x})^2 = 5(3.04-3.14)^2 + 5(3.12-3.14)^2 + 5(3.30-3.14)^2 + 5(3.10-3.14)^2 = 0.188$$

Block SS

$$SSBL = \sum_{i=1}^{b} k(\bar{x}_{Bj} - \bar{x})^2$$

$$= 4(2.8-3.14)^2 + 4(3.0-3.14)^2 + 4(3.1-3.14)^2 + 4(3.4-3.14)^2 + 4(3.4-3.14)^2 = 1.088$$

Total SS

$$SST = \sum_{i,j}(x_{ij} - \bar{x})^2 = (2.7-3.14)^2 + (2.8-3.14)^2 + (2.9-3.14)^2 + \ldots + (3.4-3.14)^2 = 1.308$$

Error SS

$$SSE = SST - SSTR - SSBL = 1.308 - 0.188 - 1.088 = 0.032$$

>>>>>>>>>>

Source of variation	SS	df	MS	F-statistic
Treatment	0.188	k – 1 = 4 – 1 = 3	0.06267	23.50
Block	1.088	b – 1 = 5 – 1 = 4	0.27200	101.87
Error	0.032	(k – 1)(b – 1) = 12	0.00267	
Total	1.308	n – 1 = kb – 1 = 19		

(a) At the 5% significance level, test whether there is a difference in mean eggplant yield between the four fertilizer treatments (the factor of interest).

H_0: There is no difference in mean eggplant yield between the four fertilizer treatments
Ha: There is a difference in mean eggplant yield between the four fertilizer treatments (at least two means are different).

$$F_{Treatment} = \frac{SSTR/(k-1)}{SSE/(k-1)(b-1)} = \frac{0.188/(4-1)}{0.032/(4-1)(5-1)} = \frac{0.06267}{0.00267} = 23.50$$

$$df = [(k-1),(k-1)(b-1)] = [(4-1),(4-1)(5-1)] = (3,12)$$

P < 0.001 There is extremely strong evidence against Ho.
Since P < α (0.05), reject Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a difference in mean eggplant yield between the four fertilizer treatments (at least two means are different).

(b) At the 5% significance level, test whether there is a difference in mean eggplant yield between blocks (in space), which is not the factor of interest.

H_0: There is no effect of the block factor on mean eggplant yield.
Ha: There is an effect of the block factor on mean eggplant yield.

F (Blocks)

$$F_{Blocks} = \frac{SSBL/(b-1)}{SSE/(k-1)(b-1)} = \frac{1.088/(5-1)}{0.032/(4-1)(5-1)} = \frac{0.27200}{0.00267} = 101.87$$

$$df = [(b-1),(k-1)(b-1)] = [(5-1),(4-1)(5-1)] = (4,12)$$

P < 0.001 There is extremely strong evidence against Ho.
Since P < α (0.05), reject Ho. [Exact P-value = 3.71 x 10^{-9}]

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a significant effect of blocks (in space) on mean eggplant yield (at least two means are different).

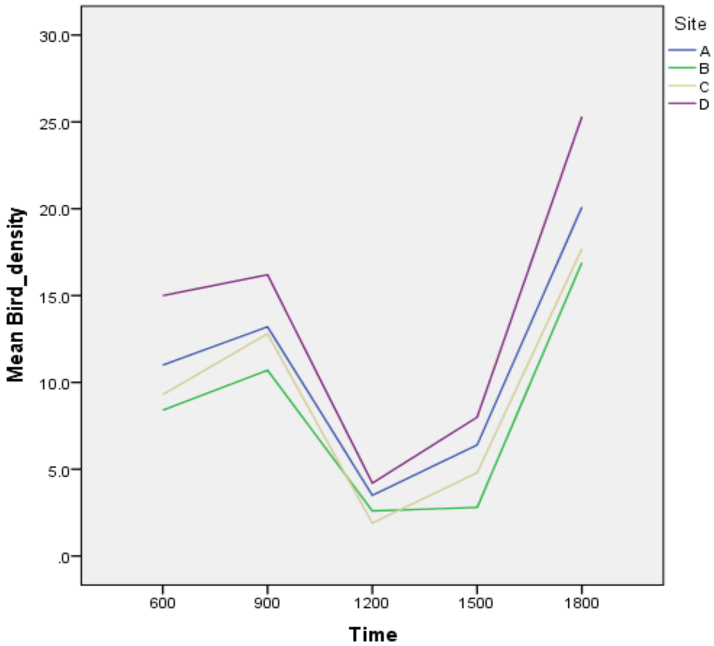
>>>>>>>>>>

Demonstrate what would happen if we IGNORE the Blocks Factor and analyze this data set with Single-Factor ANOVA (Use Excel)

Result: F = 0.895, df = (3, 16), P = 0.4650

Test for correlation as a measure of significant blocking effect

Example of Applying Randomized Block ANOVA to Blocks in Time
As part of planning wildlife conservation strategies, a group of ecologists wanted to determine whether there was a significant difference in bird density between four sites (A, B, C, and D). They chose a randomized block design where blocks were times of the day and they recorded bird density simultaneously at the four sites at 5 times of the day (600 hrs, 900 hrs, 1200 hrs, 1500 hrs, and 1800 hrs). Make use of the line graph and ANOVA table with missing values to answer the questions below.



Note:

1. If a block design was not used, the great variation in time of day would hide any differences between sites.
2. The lines are almost parallel, indicating little or no interaction between subject and test.

Tests of Between-Subjects Effects

Dependent Variable: Bird_density

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	798.279[a]	7	114.040	80.267	.000
Intercept	2221.832	1	2221.832	1563.844	.000
Site	84.876	3	28.292	19.913	.000
Time	713.403	4	178.351	125.533	.000
Error	17.049	12	1.421		
Total	3037.160	20			
Corrected Total	815.328	19			

a. R Squared = .979 (Adjusted R Squared = .967)

(a) Does the line graph above indicate interaction? Explain your answer.

The lines are more or less parallel, which means that there is no interaction between the factor of interest (site) and blocks (time of the day), which is not of interest, in their effect on bird density.

(b) At the 5% significance level, test whether there is a difference in mean bird density between the four sites (the factor of interest).

H_0: There is no difference in mean bird density between the four sites
Ha: There is a difference in mean bird density between sites (means of at least two sites are different).

$$F_{Treatment} = \frac{SSTR/(k-1)}{SSE/(k-1)(b-1)} = \frac{84.876/4-1}{17.049/(4-1)(5-1)} = \frac{28.292}{1.421} = 19.913$$

$$df = [(k-1),(k-1)(b-1)] = [(4-1),(4-1)(5-1)] = (3,12)$$

P < 0.001 Since P < α (0.05), reject Ho since there is extremely strong evidence.

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a difference in mean bird density between sites (the means of at least two sites are different).

(c) At the 5% significance level, test whether there is a difference in mean bird density between blocks (time of day), which is not the factor of interest.

H_0: There is no effect of the block factor (time of day).
Ha: There is an effect of the block factor

$$F_{Blocks} = \frac{SSBL/(b-1)}{SSE/(k-1)(b-1)} = \frac{MSBL}{MSE} = \frac{713.403/5-1}{17.049/(4-1)(5-1)} = \frac{178.351}{1.421} = 125.533$$

$$df = [(b-1),(k-1)(b-1)] = [(5-1),(4-1)(5-1)] = (4,12)$$

P < 0.001, Since P < α (0.05), reject Ho.
There is extremely strong evidence against Ho.

Conclusion: At the 5% significance level, the data provide sufficient evidence to conclude that there is a significant effect of blocks (time of day) on bird density.

Note: The block effect is probably much greater than the effect of site which can be seen in the Line Graph.

(d) Judging by the line graph and the ANOVA output do you think the same conclusion would have been reached if a completely randomized design (analyzed with one-way ANOVA) had been applied by the researchers instead of the randomized block design. Explain the logic of your answer.

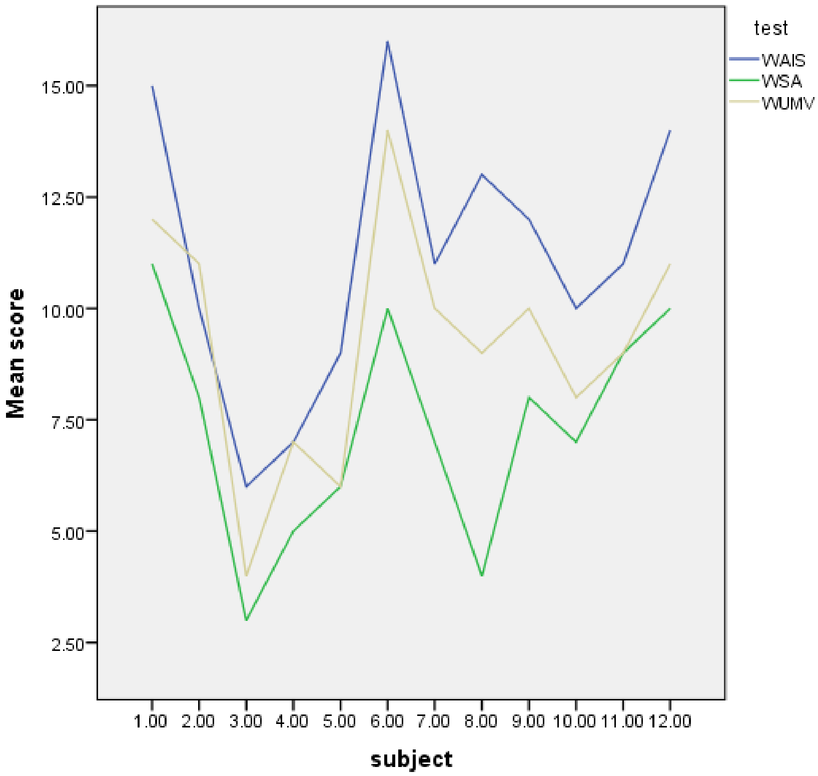
The line graph shows a huge variation in bird density from one time of the day to another. In fact, this variation is much greater than the variation in bird density between sites. Therefore, if the data were analyzed with completely randomized one-way ANOVA it would likely have shown no difference between sites because this difference would have been overshadowed by the difference over time. With such a research design, randomized block ANOVA is much more powerful than One-Way ANOVA.

Example on Comparing Language Test Scores
A random sample of 12 subjects were given three different types of language tests as follows:

1. WAIS vocabulary (linguistic)
2. Willner Unusual Meanings Vocabulary (WUMV) pragmatic
3. Willner-Sheerer Analyogy Test (WSA) pragmatic

Since people have very different linguistic ability, in order to account for, or eliminate, the subjects' abilities, the randomized block design was selected. Therefore the same 12 subjects took all three language tests. At the 5% significance level, test whether there is any difference in mean scores attained on the three language tests.

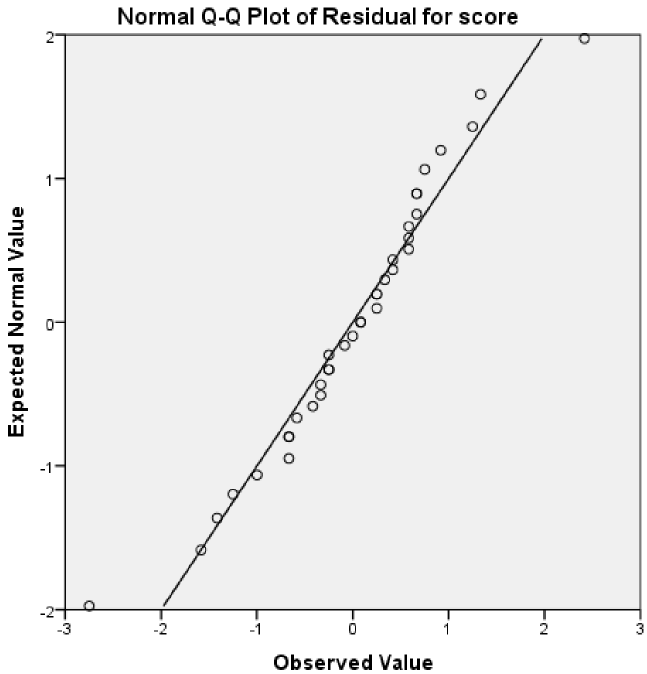
Line Graph



Note:

1. If a block design was not used, the great variation from one subject to the other would hide any differences between language tests.
2. The lines are almost parallel, indicating little or no interaction between subject and test.

Q-Q Plot to examine the assumption of normality



Note: the data points fall roughly along a straight line, indicating that the data are approximately normally distributed.

Randomized Block ANOVA (SPSS Output)

Tests of Between-Subjects Effects

Dependent Variable: score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	310.250[a]	13	23.865	17.214	.000
Intercept	3080.250	1	3080.250	2221.820	.000
test	88.167	2	44.083	31.798	.000
subject	222.083	11	20.189	14.563	.000
Error	30.500	22	1.386		
Total	3421.000	36			
Corrected Total	340.750	35			

a. R Squared = .910 (Adjusted R Squared = .858)

Test for a Difference in Mean Score Between the Three Language Tests
H_0: There is no difference in mean score between the three language tests
Ha: There is a difference in mean score between the three language tests

F (Treatment)

$$F_{Treatment} = \frac{SSTR/(k-1)}{SSE/(k-1)(b-1)} = \frac{MSTR}{MSE} = \frac{88.167/(3-1)}{30.500/(3-1)(12-1)} = \frac{44.0835}{1.3864} = 31.798$$

$$df = [(k-1),(k-1)(b-1)] = [(3-1),(3-1)(12-1)] = (2,22)$$

P < 0.001   Since P < α (0.05), reject Ho with extremely strong evidence

Conclusion: The data provide sufficient evidence that there is a significant difference in mean score between the three language tests

Test for Effect of Blocks
H_0: There is no effect of the block factor.
Ha: There is an effect of the block factor

F (Blocks)

$$F_{Blocks} = \frac{SSBL/(b-1)}{SSE/(k-1)(b-1)} = \frac{MSBL}{MSE} = \frac{222.083/(12-1)}{30.500/(3-1)(12-1)} = \frac{20.189}{1.3864} = 14.563$$

$$df = [(b-1),(k-1)(b-1)] = [(12-1),(3-1)(12-1)] = (11,22)$$

P < 0.001   Since P < α (0.05), reject Ho with extremely strong evidence

Conclusion: The data provides sufficient evidence that there is a significant effect of blocks.

Note: The block effect is probably much greater than the effect of language score (even though the F-statistic is lower) due to the higher numerator df. Also that can be seen in the Line Graph.

Multiple Comparisons

Dependent Variable: score
Tukey HSD

(I) test	(J) test	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
WAIS	WSA	3.8333*	.48069	.000	2.6258	5.0409
	WUMV	1.9167*	.48069	.002	.7091	3.1242
WSA	WAIS	-3.8333*	.48069	.000	-5.0409	-2.6258
	WUMV	-1.9167*	.48069	.002	-3.1242	-.7091
WUMV	WAIS	-1.9167*	.48069	.002	-3.1242	-.7091
	WSA	1.9167*	.48069	.002	.7091	3.1242

Based on observed means.
The error term is Mean Square(Error) = 1.386.
*. The mean difference is significant at the 0.05 level.

Conclusion: All pairwise comparisons show significant differences.