SECTION 3: SEVERAL POPULATION MEANS

- The purpose of Analysis of Variance (ANOVA) is to compare several (more than two) groups or means – that is, to find the **difference** among **more than two groups**
- Measurements of **only one variable** are recorded, but they come from different populations, treatments or groups
- The one variable being measured can be considered as the **response variable**

3.1 F-Distribution

- All types of ANOVA utilize the F-distribution
- The **F-statistic** is arrived at by calculating two types of variations and dividing one by the other
- The F-statistic has two numbers for degrees of freedom:
    - The **numerator degrees of freedom**, which corresponds to the type of variation placed in the numerator when calculating the test statistic, and
    - The **denominator degrees of freedom**, which corresponds to the variation placed in the denominator of the F-statistic
- These degrees of freedom are denoted, for example

$$\text{df} = (\underset{\text{Numerator df}}{12}, \underset{\text{denominator df}}{35})$$

- There are an infinite number of F-distributions, each identified by the two degrees of freedom.

> **Basic Properties of F-Curves**
>
> **Property 1:** The total area under an F-curve equals 1.
>
> **Property 2:** An F-curve starts at 0 on the horizontal axis and extend indefinitely to the right, approaching, but never touching, the horizontal axis.
>
> **Property 3:** An F-curve is right skewed.
>
> **Property 4:** At df = (∞, ∞), F = 1.000 at all significance levels.

Sketch of Two Different F-curves, df = (3, 40) and df = (15, 100)

The F-Table

- The F-table gives the areas (or probabilities) under the curve to the right of given values of F
- **Numerator degrees of freedom** (indicated as ***dfn***) are shown along the top of each page
- **Denominator degrees of freedom** (indicated as ***dfd***) are shown along the sides of each page.
- For any given combination of dfn and dfd, the F-values are given in a cluster and their significance levels are indicated along the sides.
- The critical values of F are always ≥ 1, though the calculated (observed) values may be < 1

Examples

- Find $F_{0.05}$ at df = (5,23) = 2.64
- Find $F_{0.025}$ at df = (11,180) ≈ (10, 100) = 2.18

Guidelines for Using P-values as Criteria for Rejection of H_0 and Statistical Significance

P-value	Strength of Evidence Against H_0
P > 0.10	Weak
0.05 < P ≤ 0.10	Moderate
0.01 < P ≤ 0.05	Strong
0.001 < P ≤ 0.01	Very strong
P ≤ 0.001	Extremely strong

3.2 ANOVA: Assumptions and Logic

- While the pooled t-test is used to compare one variable measured in two populations, ANOVA is used to compare one variable measured in more than two populations
- **One-Way ANOVA** (also called **Single-Factor ANOVA**)
  - Used to compare the values of one variable between (among) several groups or populations that are affected by **one factor**
  - This one factor may also be considered as one explanatory variable
  - The different values of the factor are called levels of that factor or treatment
- **Two-Way ANOVA** (also called **Two-Factor ANOVA**)
  - Used to compare the values of one variable among populations that are classified or grouped according to **two factors**
  - So, we can consider these as two explanatory variables
- Factors may be categorical variables or quantitative variables
- **Multiway Factorial ANOVA** deals with comparisons where more than two factors affect the populations
- **Randomized-block ANOVA** is an extension of the Paired-sample t-test, where you have more than two samples "blocked" in time or space or by some relationship.
- The Meaning of Analysis of Variance is that **we analyze and compare variances among populations with variance within the populations**
- The following terms are used synonymously:
  Groups = Treatments = Samples (taken from Populations)

- The **F-statistic** is:

$$F = \frac{\text{Between Groups (Samples) Variability}}{\text{Within Groups (Samples) Variability}}$$

OR

$$F = \frac{\text{Treatment Mean Square (variation between samples)}}{\text{Error Mean Square (variation within samples)}}$$

- **"Error" = "Residual" = "Within Groups Variability"**

- Although the purpose of ANOVA is to compare several population means and the sample means are calculated during the analysis; in the end, the F-statistic only makes a comparison of variability (among and within), thus the term "Analysis of Variance"

One-Way ANOVA: Three Sources of Variation

Three Sources of Variation and Sums of Squares in One-Way ANOVA

For one-way ANOVA of ***k*** population means,

Total Sum of Squares (SS_{Total}) = total variation between and within samples or groups

Treatment Sum of Squares ($SS_{Treatment}$) = variation between treatments or groups

Error (or Residual) Sum of Squares (SS_{Error}) = variation within treatments or groups

One-Way ANOVA Identity:

$$SS_{Total} = SS_{Treatment} + SS_{Error}$$

Mean Squares and F-Statistic in One-Way ANOVA

Treatment mean square ($MS_{Treatment}$ **)**

= treatment sum of squares divided by treatment degrees of freedom

$$MS_{Treatment} = SS_{Treatment} / (k-1)$$

Where k = number of populations being compared

Error mean square (MS_{Error} **)**

= error sum of squares divided by error degrees of freedom

$$MS_{Error} = SS_{Error} / (n-k)$$

Where n = total number of observations

F-Statistic (F)

= the ratio of the variation between groups to the variation within groups

$$F = \frac{\text{Between Groups Variability}}{\text{Within Groups Variability}}$$

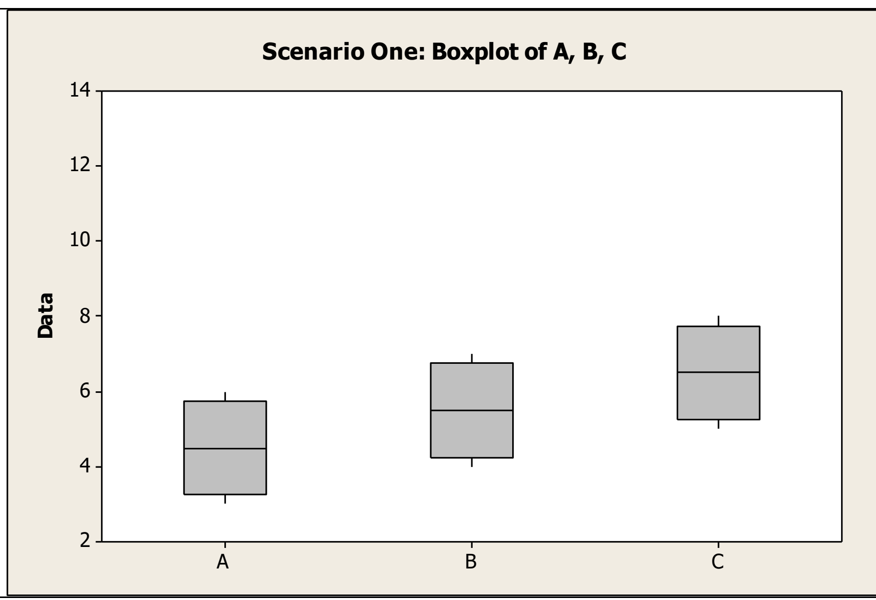
$$F = \frac{MS_{Treatment}}{MS_{Error}} = \frac{SS_{Treatment} / (k-1)}{SS_{Error} / (n-k)}$$

Three Scenarios to Explain the Logic of One-Way ANOVA

Scenario One

- Small variation among groups and small variation within groups
- No significant difference at α = 0.05

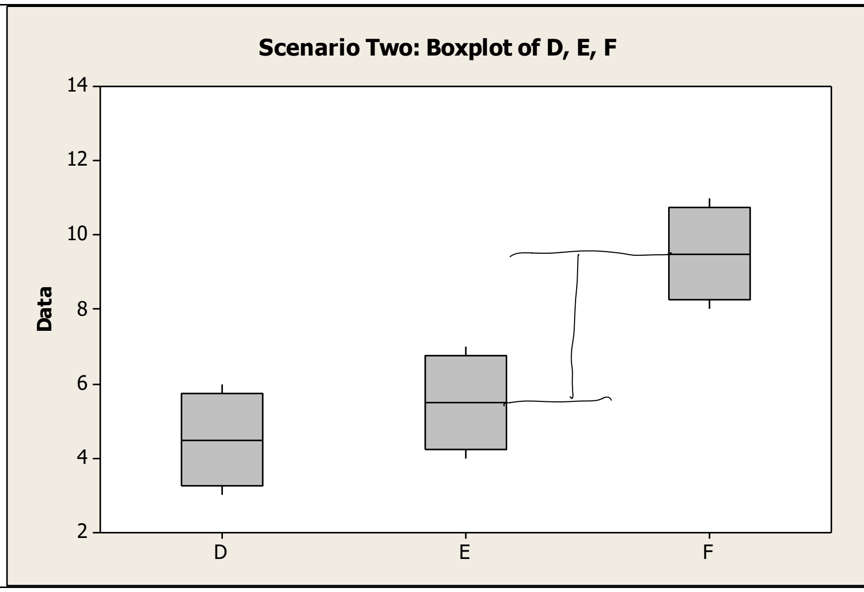
A	B	C
4	6	8
3	5	7
5	7	5
6	4	6



Scenario Two

- Larger variation among groups than within
  [This was done by adding 3 to each observation in Treatment C above]
- There is an extremely significant difference at α = 0.05

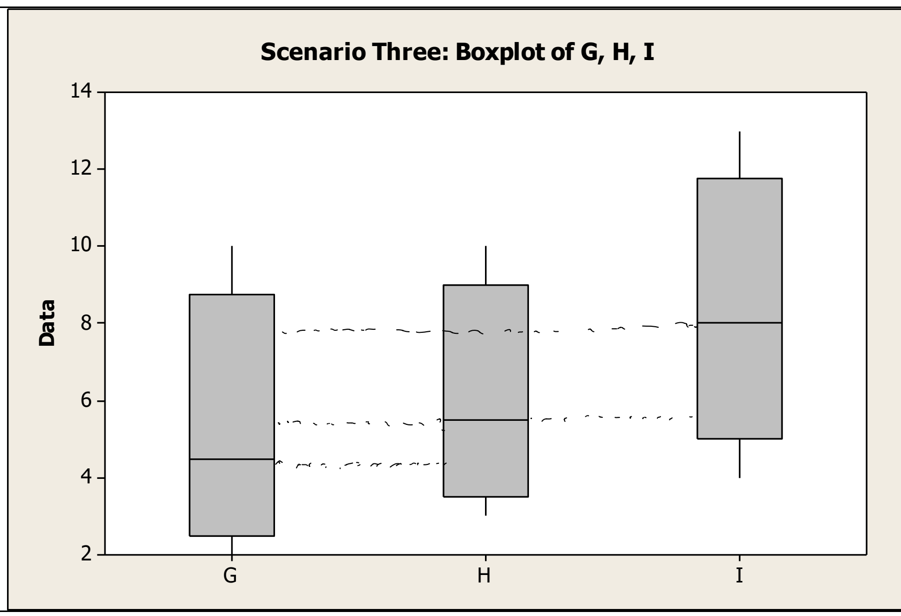
D	E	F
4	6	11
3	5	10
5	7	8
6	4	9



Scenario Three

- Large variation among groups (means), but even larger variation within groups
- No significant difference at α = 0.05

G	H	I
4	5	13
5	10	8
2	3	4
10	6	8



Computer Output: Scenario One

Summary Statistics

Groups	*Count*	*Sum*	*Average*	*Variance*
A	4	18	4.5	1.666667
B	4	22	5.5	1.666667
C	4	26	6.5	1.666667

ANOVA Table

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	8	2	4	2.4	0.146095	4.256495
Within Groups	15	9	1.666667			
Total	23	11				

Computer Output: Scenario Two

Summary Statistics

Groups	*Count*	*Sum*	*Average*	*Variance*
D	4	18	4.5	1.666667
E	4	22	5.5	1.666667
F	4	38	9.5	1.666667

ANOVA Table

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	56	2	28	16.8	0.000916	4.256495
Within Groups	15	9	1.666667			
Total	71	11				

Computer Output: Scenario Three

Summary Statistics

Groups	*Count*	*Sum*	*Average*	*Variance*
G	4	21	5.25	11.58333
H	4	24	6	8.666667
I	4	33	8.25	13.58333

ANOVA Table

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	19.5	2	9.75	0.864532	0.453485	4.256495
Within Groups	101.5	9	11.27778			
Total	121	11				

3.3 One-Way ANOVA Hypothesis Test

One-Way ANOVA Hypothesis Test

Purpose: To test for the difference between several (***k***) population means.

Assumptions:

1. Simple random samples from each population (implies independent sampling within populations)
2. Independent samples (All ***k*** samples are sampled independently of each other)
3. All populations being compared are normally distributed
4. Equal population standard deviations

Step 1: Check the purpose and assumptions

Step 2: State the null and alternative hypotheses:

H_0: $\mu_1 = \mu_2 = \ldots\mu_k$ (One-mean model)

H_a: Not all the means are equal. (***k***-mean model)

$\exists \mu_1, \mu_2 \in M$ s.t. $\mu_1 \neq \mu_2$

Step 3: Obtain the three sums of squares (SS_{Total}, SS_T and SS_E) and construct a **One-way ANOVA table** to obtain the calculated value of the F-statistic

$$SS_{Treatment} = \sum\sum(\bar{y}_j - \bar{\bar{y}})^2 = \sum n_j(\bar{y}_j - \bar{\bar{y}})^2$$

$$SS_{Error} = \sum\sum(y_{ij} - \bar{y}_j)^2$$

$$SS_{Total} = \sum\sum(y_{ij} - \bar{\bar{y}})^2$$

Nemerator

One-Way ANOVA Table

Source of variation	SS	df	MS = SS/df	F-statistic
Treatment (Between groups)	$SS_{Treatment}$	$k-1$	$MS_{Treatment} = \frac{SS_{Treatment}}{k-1}$	$F = \frac{MS_{Treatment}}{MS_{Error}}$
Error (Within groups)	SS_{Error}	$n-k$	$MS_{Error} = \frac{SS_{Error}}{n-k}$	
Total	SS_{Total}	$n-1$		

$$F = \frac{SS_{Treatment}/(k-1)}{SS_{Error}/(n-k)} = \frac{MS_{Treatment}}{MS_{Error}}$$

denominator

Step 4: Decide to reject or not reject Ho

***df* =** (numerator degrees of freedom, denominator degrees of freedom)

$df = (k-1, n-k)$ or F_{n-k}^{k-1}

If the P-value ≤ α, we reject H_0 (otherwise do not reject H_0)

Never double the p-value

Step 5: Conclusion in terms of the research problem

Note: ANOVA is a **one-tailed test** and the ANOVA **table is one-tailed**.

Example of One-Way ANOVA: Experiment on Yield of Different Varieties of Sorghum
An experiment was conducted to compare the yield of three varieties of sorghum by planting them in plots in a completely randomized design in a uniform field, obtaining data as shown below. The data are normally distributed and the three samples have equal variances. At the 5% significance level, test whether there is a difference in the mean yield of the three varieties.

Variety A	Variety B	Variety C
5	6	10
8	5	8
7	7	11
6	8	10
	9	8

Step 1: Check purpose and assumptions

- Purpose: To compare ***k*** population means
- The three populations are normally distributed, with equal variance
- The three samples are random and independent

Step 2: H_0: $\mu_1 = \mu_2 = \mu_3$ (There is no difference in mean yield among the three varieties) (One-mean model)

H_a: Not all the means are the same for the yield of the three varieties. (Three-mean model)

Step 3: Obtain the three sums of squares and construct a one-way ANOVA table to obtain the F-statistic

Quantity	Variety A	Variety B	Variety C	**Grand**
Total	26	35	47	108
Sample size	4	5	5	14
Mean	6.5	7	9.4	**7.7143**

mean of means

Calculate Treatment Sum of Squares (Measures variation between groups):

Quantity	Variety A	Variety B	Variety C	Totals
$(\bar{y}_j - \bar{\bar{y}})$	6.5 – 7.7143 = -1.2143	7 – 7.7143 = -0.7143	9.4 – 7.71428 = 1.6857	
$n_j(\bar{y}_j - \bar{\bar{y}})^2$	4 x (-1.214)^2 = 5.898	5 x (-0.7143)^2 = 2.551	5 x (1.6857)^2 = 14.208	$\sum n_j(\bar{y}_j - \bar{\bar{y}})^2 = 22.657$

Variation from the grand mean.

$$SS_{Treatment} = \sum n_j(\bar{y}_j - \bar{\bar{y}})^2 = 22.657$$

Calculate Error (or Residual) Sum of Squares (Measures variation within groups):

	Variety A	**Variety B**	**Variety C**	
	$(5-6.5)^2$ =2.25	$(6-7)^2$ =1	$(10-9.4)^2$ =0.36	
	$(8-6.5)^2$ =2.25	$(5-7)^2$ =4	$(8-9.4)^2$ =1.96	
	$(7-6.5)^2$ =0.25	$(7-7)^2$ =0	$(11-9.4)^2$ =2.56	
	$(6-6.5)^2$ =0.25	$(8-7)^2$ =1	$(10-9.4)^2$ =0.36	
		$(9-7)^2$ = 4	$(8-9.4)^2$ =1.96	
$\sum(y_{ij} - \bar{y}_j)^2$	5	10	7.2	$\sum\sum(y_{ij} - \bar{y}_j)^2 = 22.2$

$$SS_{Error} = \boxed{\sum}\boxed{\sum(y_{ij} - \bar{y}_j)^2} = 22.2$$

Sum of the variations

$$SS_{Total} = \sum\sum(y_{ij} - \bar{\bar{y}})^2 = SS_{Treatment} + SS_{Error} = 22.657 + 22.2 = 44.857$$

[The Total Sum of Squares is not actually required in order to calculate the F-statistic.]

The value of the Total Sum of Squares can be verified as follows:

	Variety A	**Variety B**	**Variety C**	
	$(5-7.7143)^2 = 7.367$	$(6-7.7143)^2 = 2.939$	$(10-7.7143)^2 = 5.224$	
	$(8-7.7143)^2 = 0.082$	$(5-7.7143)^2 = 7.367$	$(8-7.7143)^2 = 0.082$	
	$(7-7.7143)^2 = 0.510$	$(7-7.7143)^2 = 0.510$	$(11-7.7143)^2 = 10.796$	
	$(6-7.7143)^2 = 2.939$	$(8-7.7143)^2 = 0.082$	$(10-7.7143)^2 = 5.224$	
		$(9-7.7143)^2 = 1.653$	$(8-7.7143)^2 = 0.082$	
$\sum(y_{ij} - \bar{\bar{y}}_j)^2$	10.898	12.551	21.408	$\sum\sum(y_{ij} - \bar{\bar{y}}_j)^2$ =44.857

Excel Output

Summary Statistics

Groups	*Count*	*Sum*	*Average*	*Variance*
Var. A	4	26	6.5	1.666667
Var. B	5	35	7	2.5
Var. C	5	47	9.4	1.8

One-Way ANOVA Table

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	22.65714	2	11.32857	5.613256	0.020887	3.982298
Within Groups	22.2	11	2.018182			
Total	44.85714	13				

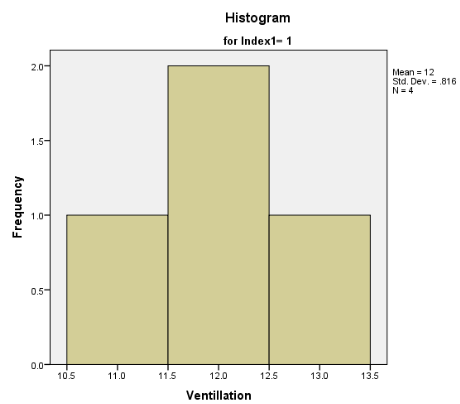
Step 4:
df = (k – 1, n - k) = (2, 11)   From the F-table: 0.025 > P > 0.01. The exact P-value = 0.020887.
So, there is strong evidence against H_0.
Since P-value < α (0.05), reject H_0.

Step 5:
At the 5% significance level, the data provide sufficient evidence to conclude that there is a difference in the mean yield of the three varieties (that is, at least two means are different).

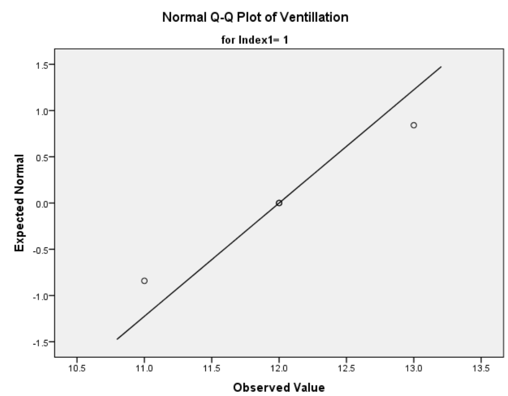
Example: Effect of Certain Diseases on Human Ventilation Rates
The normal resting ventilation rate is about 6 liters per minute (L/min) in healthy people, but is higher in people with a disease. The table below shows the ventilation rates of random samples of patients suffering from three different diseases. At the 1% significance level, determine whether there is a difference in the mean ventilation rates of people suffering from these three diseases.

Ventilation rate (L/min)		
Cancer	**Heart disease**	**Diabetes**
12	14	12
11	13	15
13	16	13
12	16	13
	13	14
	15	14
	15	

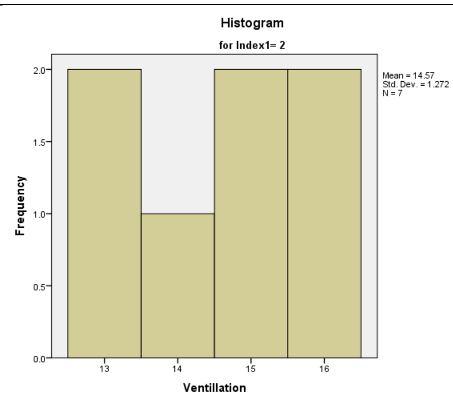
Checking Assumptions (SPSS Output)



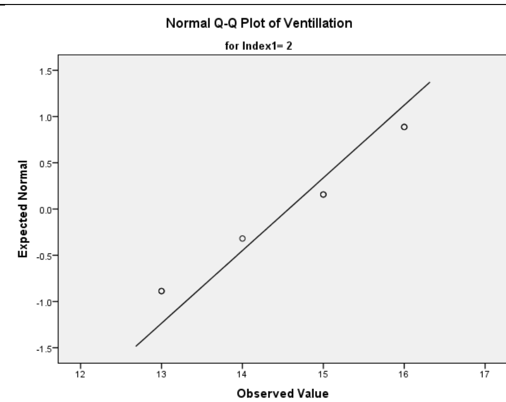
Histogram for Cancer



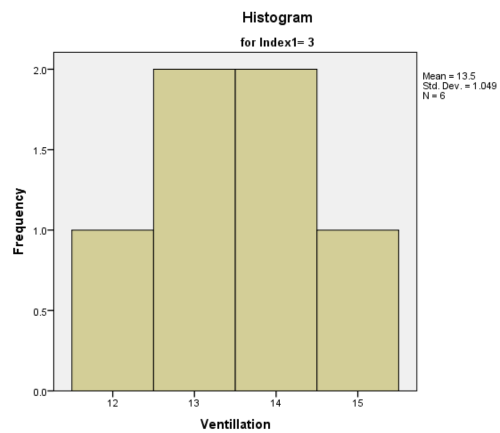
Normal Probability Plot for Cancer



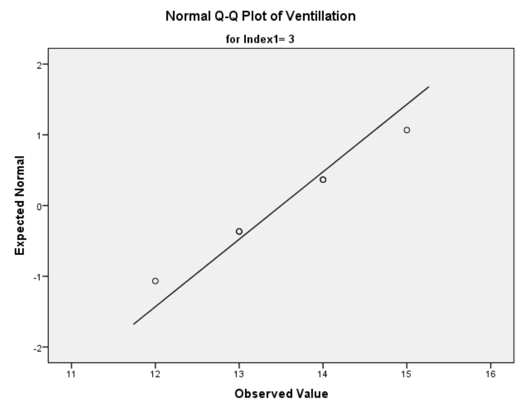
Histogram for Heart Disease



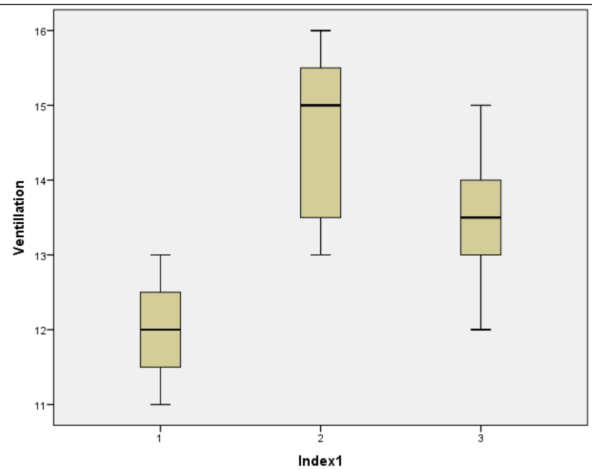
Normal Probability Plot for Heart Disease



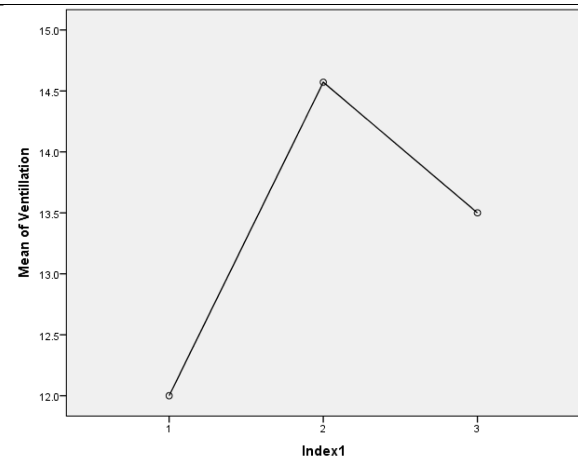
Histogram for Diabetes



Normal Probability Plot for Diabetes



Side-by-side Boxplots for Cancer, Heart Disease and Diabetes



Means Plot comparing mean ventilation rates for patients with cancer, heart disease and diabetes

Test of Homogeneity of Variances

Ventilation

Levene Statistic	df1	df2	Sig.
1.351	2	14	.291

SPSS Output

Descriptives

Ventilation

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	4	12.00	.816	.408	10.70	13.30	11	13
2	7	14.57	1.272	.481	13.39	15.75	13	16
3	6	13.50	1.049	.428	12.40	14.60	12	15
Total	17	13.59	1.460	.354	12.84	14.34	11	16

$SS_{Treatment}$

ANOVA

Ventilation

Source of variation	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16.903	2	8.452	6.874	.008
Within Groups	17.214	14	1.230		
Total	34.118	16			

Suppose that only partial ANOVA output is given, so the numbers highlighted in yellow are not given.

>>>>>>>>>> SS_{Error}

$H_0: \mu_1 = \mu_2 = \mu_3$ (no diff between means)

$H_A: \mu_1, \mu_2, \mu_3$ (at least 1 diff between means)

k = # of pop being compared

n : total num of observations $\sum_{i=1}^{k} n_i$

$$F = \frac{SS_T/(k-1)}{SS_E/(n-k)} = \frac{MS_T}{MS_E} = \frac{16.903/(3-1)}{17.214/(17-3)} = \frac{8.452}{1.230} = 6.874$$

$df = (2, 14)$: p-value = $0.005 < \text{p-value} < 0.01$

p-value $< \alpha$ of 0.05

Very strong evidence against H_0

$P < \alpha = 0.01$ ∴ we reject H_0

for medical experiment

>>>>>>>>>>

Experiment to test the ultimate strength of stainless steel, steel alloy and titanium alloy
An experiment was conducted to test the ultimate strength (in MPa's) of random samples of stainless steel, steel alloy and titanium alloy. Below is incomplete output of one-way ANOVA obtained from SPSS.

Summary Statistics				
Groups	*Count*	*Sum*	*Average*	*Variance*
Stainless Steel	5	4320	864	1930
Steel alloy	7	5740	820	1633.333
Titanium alloy	7	6240	891.4286	1347.619

Calculate the F-statistic by filling in missing values.
>>>>>>>>>>

ANOVA

Ultimate strength

Source of variation	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	18110.08	2	9055.0400	5.66	0.014
Within Groups	25605.17	16	1600.3564		
Total	43715.79				

df(2,16) = P < 0.001

>>>>>>>>>>
At the 5% significance level, what conclusion can you draw regarding the ultimate strength of the three materials?

(a) There is no significant difference in ultimate strength of the three materials.
(b) Ultimate strength of titanium alloy is greater than that of steel alloy, but is not greater than that of stainless steel.
(c) Ultimate strength of Titanium alloy is greater than that of both steel alloy and stainless steel.
(d) All the means for ultimate strength of the three materials are different.
(e) At least two of the means for ultimate strength of the three materials are different.

Answer: e

Filling in Missing Values in an ANOVA Table
>>>>>>>>>>

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12409.96		111.4	5.546	0.001 < P < 0.005
Within Groups	461.9	23	20.08		
Total	12871.86	28			

>>>>>>>>>>

3.4 Multiple Comparisons (= Unplanned comparisons)

- If, and only if, one-way ANOVA results in rejecting the null hypothesis, then it is often desirable to do multiple comparisons in order to determine which means are different the which other means
- Known as pairwise comparisons
- The number of pairwise comparisons that are possible for a given question is given by:
  k(k-1)/2, where k = number of means (groups) being compared

- There are several types of multiple comparisons, including:
  1. Tukey multiple-comparisons (also called HSD = honest significant difference)
     - Requires that the sample sizes for all groups be the same (or very similar)
     - When sample sizes are equal, the CIs are shorter than other methods and therefore more likely to show differences
  2. Bonferroni method
     - Can be used for a general case where sample sizes are different
     - Can control the overall error rate
  3. Fisher method
  4. Scheffe method - results in wider CIs than Tukey's test and therefore more conservative
  5. Least significant difference (LSD) - not suitable if the number of groups being compared is large
  6. Student-Newman-Keuls (SNK) test

3.4.1 Tukey Multiple Comparisons

Tukey Multiple Comparisons

Purpose: To determine pairwise differences between ***k*** population means when the null hypothesis has been rejected in one-way ANOVA.

Assumptions (same as for One-way ANOVA):

Step 1: At the given confidence level, 1 – α, find the critical value q_α at
df = (k, n – k) in the appropriate statistical table.

Step 2: Obtain the endpoints of the confidence interval for the difference, $\mu_i - \mu_j$

$$(\bar{y}_i - \bar{y}_j) \pm \frac{q_\alpha}{\sqrt{2}} \times \sqrt{MSE}\sqrt{(1/n_i)+(1/n_j)}$$

Where, ***MSE*** = Error mean square from one-way ANOVA table

Do so for all possible pairs of means with ***i < j*** and summarize the confidence intervals in a table.

[**Note:** There will be k(k – 1)/2 pairwise differences.]

Step 3: Compile the results in a matrix and declare two population means different if the confidence interval for the difference does not contain 0; otherwise, do not declare the two population means different.

Step 4: Conclusion

Summarize the results in a **means comparisons diagram** by ranking the sample means from smallest to largest and by connecting with lines those whose population means were not declared different.

And: Interpret the results of the multiple comparisons **in words**

Example for Tukey Multiple comparisons: Experiment on Yield of Different Varieties of Sorghum
An experiment was conducted to compare the yield of three varieties of sorghum by planting them in plots in a randomized design in a uniform field. One-way ANOVA resulted in rejecting the null hypothesis and thus drawing the conclusion that not all means are equal (or at least two means are different). Perform Tukey multiple comparisons to determine which pairs of means are different at the 95% confidence level.

Information already known based on ANOVA is as follows:

Groups	Mean	Sample size
Variety A	6.5	4
Variety B	7	5
Variety C	9.4	5

Error mean square (MSE) = 2.0182

At ***df* = (*k*, *n* – *k*)** = (3, 11) and **α** = 0.05, the critical value q_α = 3.82
>>>>>>>>>>

$$q_\alpha = 3.82$$

$$m = \binom{3}{2} = 3 = \frac{3(3-1)}{2}$$

$$(\bar{Y}_1 - \bar{Y}_2) \pm \frac{q_\alpha}{\sqrt{2}} \times \sqrt{MSE} \sqrt{\left(\frac{1}{n_i}\right) + \left(\frac{1}{n_j}\right)}$$

$$(-0.5) \pm \frac{3.82}{\sqrt{2}} \times \sqrt{2.0182} \sqrt{\frac{1}{4} + \frac{1}{5}}$$

Calculate this first

$$(-3.07, 2.07)$$

mat	Var A	Var B	Var C
Var A	—	—	—
Var B		—	—
Var C			—

Var A	Var B	Var C
6.5	7	9.4

>>>>>>>>>>

3.4.2 Bonferroni's Method of Multiple Comparisons

Bonferroni's Method of Multiple Comparisons

Purpose: To determine pairwise differences between ***k*** population means when the null hypothesis has been rejected in one-way ANOVA.

Step 1: Find the number of multiple comparisons (m) that are possible:

$m = \frac{k(k-1)}{2}$, where k = number of means (groups) being compared

Step 2: Calculate the individual comparison-wise error rate (α_I) based on the family-wise (experiment-wise) error rate (α_F) or confidence level (1 – α_F) given:

$$\alpha_I = \frac{\alpha_F}{m}$$

Step 3: Find the Critical value of ***t*** at ***df = n – k*** for α_I /2: $t_{df,\alpha_I/2}$

Step 4: Calculate the margin of error (ME) for each comparison (group ***i*** vs. Group ***j***):

$$ME_{ij} = Crit.value \times S.E.(\overline{y}_i - \overline{y}_j)$$

$$ME_{ij} = t_{n-k,\alpha_I/2} \times \sqrt{MSE}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Step 5: Declare two population means different if the absolute value of the difference between their sample means is greater than or equal to the corresponding margin of error

$\mu_i - \mu_j \neq 0$, **if** $\left|\overline{y}_i - \overline{y}_j\right| \geq ME_{ij}$

Present the results in a matrix

Step 6: Summarize the results in a **means comparisons diagram** by ranking the sample means from smallest to largest and by connecting with lines those whose population means were not declared different <u>and</u> **state the conclusion in words**.

ANOVA: Effect of Certain Diseases on Human Ventilation Rates
It was concluded with very strong evidence that there is a difference in the mean ventilation rates of people suffering from the three diseases tested (cancer, heart disease and diabetes) [One-way ANOVA: F = 6.87, df = (2, 14), P-value = 0.008324]. Perform Bonferroni's method of multiple comparisons to determine which pairs of means are different at the 95% confidence level.

Disease	**Cancer**	**Heart disease**	**Diabetes**
Sample mean	12	14.57	13.5
Sample size	4	7	6

ANOVA Table						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	16.90336	2	8.451681	6.873566	0.008325	6.514884
Within Groups	17.21429	14	1.229592			
Total	34.11765	16				

>>>>>>>>>>

$m = 3$

ind. error rate $= \alpha_I = \frac{\alpha_F}{m} = \frac{0.05}{3} = 0.0167$

Crit. Val $= t_{\alpha/2, 14} = t_{0.008, 14} = t_{0.005, 14} = 2.977$

always go for the largest crit. val.

$ME_{ij} = t_{\alpha_I} \times \sqrt{MS_E} \times \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

$= 2.977 \sqrt{\frac{17.21429}{14}} \times \sqrt{\frac{1}{4} + \frac{1}{7}}$

$M_{12} = 2.129$

$M_{23} = 1.836$

	Cancer	HD	Diabetes
Can	—	—	—
HD	$7.57 > 2.064$	—	—
Diabetes	$1.5 < 2.131$	$1.07 < 1.86$	—

Cancer 12   Diabets 13.5   HD

>>>>>>>>>> Conclusion is basically the same.

3.5 Linear Combinations (Contrasts) (=Planned Comparisons)

- The multiple comparisons discussed above are sometimes called “unplanned comparisons”
- Linear combinations, on the other hand, are planned comparisons
- Ideally, the means to be compared should be planned before collecting the data

Linear Combinations (Contrasts)

Step 1: Develop the linear combination by deciding which means or groups of means you want to compare.

$$\gamma_{D-E} = \frac{(\mu_{1,1} + \mu_{1,2} + \ldots + +\mu_{1,d})}{d} - \frac{(\mu_{2,1} + \mu_{2,2} + \ldots + +\mu_{2,e})}{e}$$

Where ***D*** and ***E*** are combinations of means to be compared and ***d*** and ***e*** are the number of means within those combinations, respectively

Then, define the parameter of the contrast, which will take the following general form (where γ is the Greek letter “gamma”):

$$\gamma = C_1\mu_1 + C_2\mu_2 + \ldots + C_k\mu_k$$

Check to be sure that the coefficients add up to 0 (This makes it a contrast)

$$\sum_{i-1}^{k} C_i = C_1 + C_2 + \ldots + C_k = 0$$

Step 2: State the hypothesis

Null hypothesis is $H_0 : \gamma = 0$

Alternative hypothesis may be:

$H_a : \gamma \neq 0$ or $H_a : \gamma < 0$ or $H_a : \gamma > 0$

Step 3: Calculate the estimate (sample contrast), standard error of the estimate and the t-statistic

Estimate: $\hat{\gamma} = C_1\bar{y}_1 + C_2\bar{y}_2 + \ldots + C_k\bar{y}_k$

$$SE(\hat{\gamma}) = s_p\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_k^2}{n_k}}$$

Where $s_p =$ pooled (common) standard deviation, and

$$s_p = \sqrt{MSE} = \sqrt{\frac{(n_1 - 1)s_1^2 + \ldots + (n_k - 1)s_k^2}{n-k}}$$

$$t = \frac{\hat{\gamma} - 0}{SE(\hat{\gamma})}$$

Step 4: Decide to reject or not reject Ho by comparing the P-value at $df = n - k$ with the significance level (α) and state the strength of the evidence against Ho.

Step 5: Write the conclusion in words in terms of the research problem.

Confidence Interval for a Linear Contrast

Confidence Interval for a Linear Contrast

Step 1: For a given confidence level (1 – α), find the Critical Value ($t_{\alpha/2}$) at **df = n - k**

Step 2: Calculate (or state):

Parameter: $\gamma = C_1\mu_1 + C_2\mu_2 + \ldots + C_k\mu_k$

Estimate: $\hat{\gamma} = C_1\overline{y}_1 + C_2\overline{y}_2 + \ldots + C_k\overline{y}_k$

Standard error of the estimate:

$$SE(\hat{\gamma}) = s_p\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_k^2}{n_k}}$$

Where $s_p = \sqrt{MS_E} = \sqrt{\frac{(n_1 - 1)s_1^2 + \ldots + (n_k - 1)s_k^2}{n - k}}$

Endpoints of the confidence interval:

$$\hat{\gamma} \pm Crit.Value \times SE(\hat{\gamma})$$

from the t-table

Step 3: Interpret the confidence interval in terms of the research problem

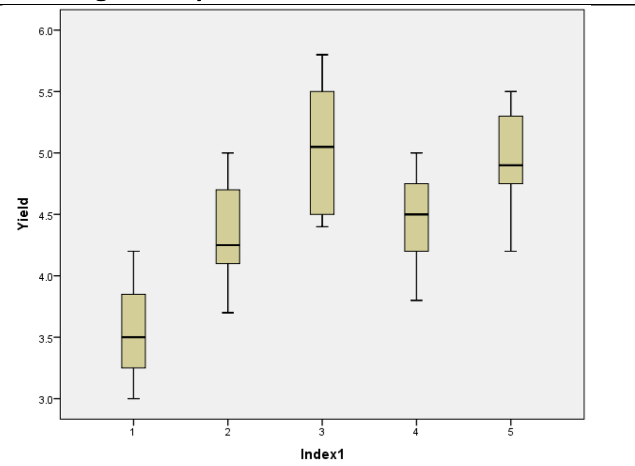
Applications of Linear Contrasts in a Rice Experiment

Azolla-Anabaena (below) is an endophytic association (symbiosis) between a water fern and a blue-green alga. *Utricularia*-Cyanophyta is an epiphytic association. Both have been used as biofertilizers ("living fertilizers") to increase rice crop yield, since both associations have been shown to fix nitrogen.

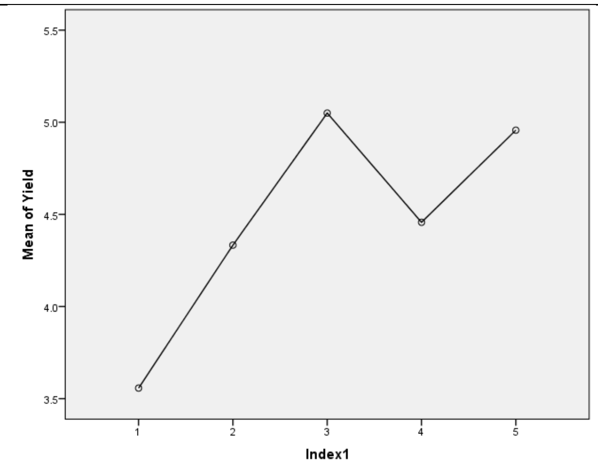
An experiment was conducted to test the effect of these biofertilizers on rice crop yield as compared to two different levels of chemical nitrogen fertilizer and a control, obtaining raw data as shown below.

Yield of rice (t ha^{-1})				
Control	***Utricularia***	***Azolla***	**N-Level 1**	**N-Level 2**
3	4.3	4.8	4.1	4.2
3.5	4.1	5.3	4.5	4.6
3.4	5	4.4	5	4.9
3.1	4.7	5.5	3.8	5.4
3.6	3.7	5.8	4.3	5.5
4.1	4.2	4.5	4.8	4.9
4.2			4.7	5.2

Checking Assumptions



Side-by-side boxplots: Standard deviations look quite similar



Means plots: Shows the relative yield of rice under the different treatments

Test of Homogeneity of Variances

Yield

Levene Statistic	df1	df2	Sig.
.410	4	28	.800

- P-value = 0.800, which is very large, so do not reject the null hypothesis. Therefore, there is insufficient evidence to conclude that there is a difference in the variances of the 5 treatments.

One-way ANOVA resulted in rejecting the null hypothesis with extremely strong evidence, as shown in the ANOVA output below.

Descriptives

Yield

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1 (Control)	7	3.557	.4577	.1730	3.134	3.980
2 (Utricularia)	6	4.333	.4590	.1874	3.852	4.815
3 (Azolla)	6	5.050	.5683	.2320	4.454	5.646
4 (N1)	7	4.457	.4198	.1587	4.069	4.845
5 (N2)	7	4.957	.4577	.1730	4.534	5.380
Total	33	4.458	.7040	.1226	4.208	4.707

One-Way ANOVA Hypothesis Test

ANOVA

Yield

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9.621	4	2.405	10.793	.000
Within Groups	6.240	28	.223		
Total	15.861	32			

The exact P-value is 2.02×10^{-5}

Research Question:
At the 5% significance level, test for differences in effectiveness between the following using linear contrasts (as two-tailed tests):

1. Control versus the biofertilizers
2. The biofertilizers versus the chemical nitrogen fertilizers
3. Level 1 of both types of fertilizers (use the epiphytic association as Level 1 of the biofertilizers) versus Level 2 of both types of fertilizers (use the endophytic association as Level 2 of the biofertilizers)

Linear Contrast 1: Control (C) versus the Biofertilizers (B)
>>>>>>>>>>

Step 1: $Y = C_1\mu_1 + \cdots + C_k\mu_k$

$$Y_{C-B} = \frac{\mu_C}{1} - \frac{(\mu_V + \mu_A)}{2}$$

Parameter $Y_{C-B} = \mu_C - \frac{1}{2}\mu_V - \frac{1}{2}\mu_A = 0$

Step 2: $H_0: Y = 0 \quad H_A: Y \neq 0$

Step 3: $\hat{Y}_{C-B} = \bar{Y}_C - \frac{1}{2}\bar{Y}_V - \frac{1}{2}\bar{Y}_A$

$= 3.557 - \frac{1}{2}(4.333) - \frac{1}{2}(5.050)$

$= -1.1345$

$SE = S_p = \sqrt{MS_E} = \sqrt{0.22285} = 0.4721$

$SE(\hat{Y}) = S_p\sqrt{\frac{C_1^2}{n_1} + \cdots + \frac{C_k^2}{n_k}}$

$= (0.4721)\sqrt{\frac{(1)^2}{2} + \frac{(-\frac{1}{2})^2}{6} + \frac{(-\frac{1}{2})^2}{6}}$

$= (0.4721)(0.4756)$

$SE(\hat{Y}) = 0.2245$

$$t = \frac{\hat{Y} - 0}{SE(\hat{Y})} = \frac{-1.1345}{0.2245} = -5.033$$

Linear Contrast 2: The biofertilizers (B) versus the chemical nitrogen (N) fertilizers

$$Y_{B-N} = \frac{C_1(\mu_V + \mu_A)}{2} - \frac{C_2(\mu_{N_1} + \mu_{N_2})}{2}$$

$$Y_{B-N} = \frac{1}{2}\mu_V + \frac{1}{2}\mu_A - \frac{1}{2}\mu_{N_1} - \frac{1}{2}\mu_{N_2}$$

$H_0: Y = 0 \quad H_A: Y \neq 0$

Estimate $\bar{Y}_{B-N} = \frac{1}{2}\bar{Y}_V + \frac{1}{2}\bar{Y}_A - \frac{1}{2}\bar{Y}_{N_1} - \frac{1}{2}\bar{Y}_{N_2}$

$$= \frac{1}{2}(4.333) + \frac{1}{2}(5.050) - \frac{1}{2}(4.457) - \frac{1}{2}(4.957)$$

$$= -0.0155$$

$$SE(\hat{Y}) = (0.4721)\sqrt{\frac{(\frac{1}{2})^2}{6} + \frac{(\frac{1}{2})^2}{6} + \frac{(\frac{-1}{2})^2}{7} + \frac{(\frac{-1}{2})^2}{7}}$$

$$\sqrt{0.1547}$$

$$0.4721 \times 0.3933 = 0.1857 = \frac{-0.0155}{0.1857}$$

$$= -0.083$$

Linear Contrast 3: Level 1 (L1) of the both types of fertilizers versus Level 2 (L2) of both types of fertilizers

>>>>>>>>>>

$$H_0 : \gamma = 0 \quad \text{versus } H_a : \gamma \neq 0$$

Estimate

$$\hat{\gamma}_{L1-L2} = \frac{1}{2}\overline{y}_U + \frac{1}{2}\overline{y}_{N1} - \frac{1}{2}\overline{y}_A - \frac{1}{2}\overline{y}_{N2}$$
$$= \frac{1}{2}(4.33) + \frac{1}{2}(4.46) - \frac{1}{2}(5.05) - \frac{1}{2}(4.96) = \frac{-1.22}{2} = -0.61$$

Standard error of the estimate

$$SE(\hat{\gamma}) = 0.4721\sqrt{\frac{(1/2)^2}{6} + \frac{(1/2)^2}{7} + \frac{(-1/2)^2}{6} + \frac{(-1/2)^2}{7}}$$
$$= (0.4721)(0.3934) = 0.1857$$

Observed value of the t-statistic:

$$t = \frac{\hat{\gamma} - 0}{SE(\hat{\gamma})} = \frac{-0.61 - 0}{0.1857} = -3.285$$

Combining the results of all 3 linear contrasts

Step 4: Decide to reject or not reject Ho:

$$df = n - k = 33 - 5 = 28 \qquad \alpha = 0.05$$

Linear Contrast	t-statistic	P-value	Decision	Strength of evidence
Control vs. Biofertilizers	– 5.053	P < 0.001	Reject Ho	Extremely strong
Biofertilizers vs. N Fert.	– 0.083	P > 0.50	Not reject Ho	Weak
Level 1 vs. Level 2	– 3.276	0.005 < P < 0.002	Reject Ho	Very strong

Step 5: There was extremely strong evidence that the biofertilizers (both combined) resulted in a difference in (greater) crop yield in comparison with the control. There was no difference in crop yield between the biofertilizers (both combined) and the nitrogen fertilizers (both levels combined). There was very strong evidence that Level 2 (combining both the biofertilizer and the nitrogen fertilizer) resulted in a difference in (greater) crop yield than Level 1.

Calculate a 95% confidence intervals for these Linear Contrasts

At $df = n - k = 33 - 5 = 28$, $t_{\alpha/2} = t_{0.05/2} = 2.048$

$\hat{\gamma}$ ± Critical Value of t x $SE(\hat{\gamma})$

Linear Contrast	Estimate	SE(Estimate)	Endpoints	Include 0
Control vs. Biofertilizers	– 1.1345	0.2245	(– 1.59, – 0.67)	No
Biofertilizers vs. N Fert.	– 0.0155	0.1857	(– 0.40, 0.36)	Yes
Level 1 vs. Level 2	– 0.6085	0.1857	(– 0.99, – 0.23)	No

Research Conclusion: Biofertilizers *Azolla-Anabaena* and *Utricularia*-Cyanophyta can be applied on rice to increase crop yield with effects comparable to the application of chemical nitrogen fertilizers. At the same time, these biofertilizers save costs and are an “environmentally-friendly” alternative. Also, the biofertilizers help to control weeds.

Post Hoc Tests (Tukey's and Bonferroni's Methods) on Rice Experiment (Using SPSS Output)

Since the one-way ANOVA table above shows that there is a difference in the mean yield of rice between the 5 treatments, it is appropriate to perform multiple comparisons tests and linear contrasts to determine which means are different.

Multiple Comparisons

Dependent Variable: Yield

	(I) Index1	(J) Index1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	C	U	-.7762*	.2626	.046	-1.541	-.011
		A	-1.4929*	.2626	.000	-2.258	-.728
		N1	-.9000*	.2523	.011	-1.635	-.165
		N2	-1.4000*	.2523	.000	-2.135	-.665
	U	C	.7762*	.2626	.046	.011	1.541
		A	-.7167	.2725	.092	-1.511	.077
		N1	-.1238	.2626	.989	-.889	.641
		N2	-.6238	.2626	.152	-1.389	.141
	A	C	1.4929*	.2626	.000	.728	2.258
		A	.7167	.2725	.092	-.077	1.511
		N1	.5929	.2626	.189	-.172	1.358
		N2	.0929	.2626	.996	-.672	.858
	N1	C	.9000*	.2523	.011	.165	1.635
		U	.1238	.2626	.989	-.641	.889
		A	-.5929	.2626	.189	-1.358	.172
		N2	-.5000	.2523	.301	-1.235	.235
	N2	C	1.4000*	.2523	.000	.665	2.135
		U	.6238	.2626	.152	-.141	1.389
		A	-.0929	.2626	.996	-.858	.672
		N1	.5000	.2523	.301	-.235	1.235
Bonferroni	C	U	-.7762	.2626	.063	-1.576	.024
		A	-1.4929*	.2626	.000	-2.293	-.693
		N1	-.9000*	.2523	.013	-1.669	-.131
		N2	-1.4000*	.2523	.000	-2.169	-.631
	U	C	.7762	.2626	.063	-.024	1.576
		A	-.7167	.2725	.137	-1.547	.114
		N1	-.1238	.2626	1.000	-.924	.676
		N2	-.6238	.2626	.246	-1.424	.176

	A	C	1.4929*	.2626	.000	.693	2.293
		U	.7167	.2725	.137	-.114	1.547
		N1	.5929	.2626	.320	-.207	1.393
		N2	.0929	.2626	1.000	-.707	.893
	N1	C	.9000*	.2523	.013	.131	1.669
		U	.1238	.2626	1.000	-.676	.924
		A	-.5929	.2626	.320	-1.393	.207
		N2	-.5000	.2523	.574	-1.269	.269
	N2	C	1.4000*	.2523	.000	.631	2.169
		U	.6238	.2626	.246	-.176	1.424
		A	-.0929	.2626	1.000	-.893	.707
		N1	.5000	.2523	.574	-.269	1.269

*. The mean difference is significant at the 0.05 level.

>>>>>>>>>>

Means tukey comp. diagram

Control	U	N1	N2	A
3.557	4.333	4.457	4.957	5.050

We can be 95% confident that control is different from others, but no other groups in others are different for each other.

Means Bonferroni comp. diagram

Control	U	N1	N2	A
3.557	4.333	4.457	4.957	5.050

We can be 95% confident that the control is different from N1, N2 and A but not different from U. The others are not different from each other.

>>>>>>>>>>

Linear Contrasts from SPSS Output (Previously done by hand calculations)

Contrast Coefficients

Contrast	Index1				
	1 (Control)	2 (Utricularia)	3 (Azolla)	4 (N1)	5 (N2)
1	1	-.5	-.5	0	0
2	0	.5	.5	-.5	-.5
3	0	.5	-.5	.5	-.5

try to use proportional coefficients with int coefficients

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Yield	Assume equal variances	1	-1.135	.2245	-5.053	28	.000
		2	-.015	.1857	-.083	28	.934
		3	-.608	.1857	-3.276	28	.003
	Does not assume equal variances	1	-1.135	.2284	-4.967	13.543	.000
		2	-.015	.1898	-.082	19.193	.936
		3	-.608	.1898	-3.206	19.193	.005

Perform Contrasts at α = 0.05 (Use ONLY Output for Equal Variances)

Linear Contrast 1: Control (C) versus the Biofertilizers (B)

$$\gamma_{C-B} = \mu_C - \frac{1}{2}\mu_U - \frac{1}{2}\mu_A$$

Estimate $\hat{\gamma}_{C-B}$ = -1.135 Standard error of the estimate $SE(\hat{\gamma})$ = 0.2245

t = -5.053 ➔ P = 0.000 ➔ Reject Ho ➔ With extremely strong evidence

Linear Contrast 2: The biofertilizers (B) versus the chemical nitrogen (N) fertilizers

$$\gamma_{B-N} = \frac{1}{2}\mu_U + \frac{1}{2}\mu_A - \frac{1}{2}\mu_{N1} - \frac{1}{2}\mu_{N2}$$

Estimate $\hat{\gamma}_{B-N}$ = -0.015 Standard error of the estimate $SE(\hat{\gamma})$ = 0.1857

t = -0.083 ➔ P = 0.934 ➔ Do not reject Ho ➔ Weak evidence

Contrast 3: Level 1 (L1) of the both types of fertilizers vs. Level 2 (L2) of both types of fertilizers

$$\gamma_{L1-L2} = \frac{1}{2}\mu_U + \frac{1}{2}\mu_{N1} - \frac{1}{2}\mu_A - \frac{1}{2}\mu_{N2}$$

Estimate $\hat{\gamma}_{L1-L2}$ = -0.608 Standard error of the estimate $SE(\hat{\gamma})$ = 0.1857

t = -3.276 ➔ P = 0.003 ➔ Reject Ho ➔ With very strong evidence

Comparison of Tukey's Multiple Comparisons, Bonferroni Multiple Comparisons and Linear Contrasts

1. In this study, sample sizes were nearly equal for all treatments, making Tukey's test suitable. Tukey's test showed that it is slightly more powerful than the Bonferroni Method since it showed a difference between Control and Utricularia whereas Bonferroni did not. This is mainly because the Bonferroni Method reduces <u>individual comparison-wise error rate</u> and makes the confidence intervals wider and less precise.

2. Linear Contrasts were more effective (and powerful) than the multiple comparisons tests in detecting differences between groups. Thus, these <u>planned</u> comparisons are very useful.

3.6 Reduced Models and the Extra Sum-of-Squares F-test in Single-Factor ANOVA

- Classifies two models: a reduced model and a full model
  - Null hypothesis is the reduced model, which is a special case of the full model obtained by imposing some restrictions
  - Alternative hypothesis is the full model, which is a general model that is found to adequately describe the data

Extra-Sum-of-Squares F-test

- Also called Partial F-test or Nested F-test

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)

$$F = \frac{(Extra\ SS)/(Extra\ df)}{SS_E(\text{Full})/df_E(Full)}$$

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

Examine the distribution of the F-table at:

$$df = [Extra\ df, df_E(Full)] = [Extra\ df, n-k]$$

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

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

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

Research problem: An educational researcher conducted a study to determine a possible effect of the initial interest of students (Low, Medium, Super) in a statistics course (as expressed at the start of the course) on their final grades. The study was based on a random sample of 72 students (24 in each interest group). For each level of interest, there were equal numbers of females and males.

(a) At the 5% significance level, perform the most appropriate test (showing all steps) to determine whether there is a difference in mean grades between students having different levels of interest, as expressed at the start of the course (that is, determine whether at least two means are different). For this test, use only the SPSS output shown in this part (a), that is Tables 1 and 2 (with missing values).

Table 1: Summary statistics of grades for the 3 treatment groups for level of interest.

Descriptives

Grade

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Low	24	68.1250	6.46269	1.31919	65.3960	70.8540
Medium	24	73.8750	6.34043	1.29424	71.1977	76.5523
Super	24	78.0000	7.10786	1.45089	74.9986	81.0014
Total	72	73.3333	7.71682	.90944	71.5200	75.1467

Table 2: ANOVA table for the comparison of grades for 3 treatment groups with respect to level of interest (ignoring gender).

ANOVA

Grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1180.750	2	590.375	13.368	.000012
Within Groups	3047.250	69	44.163		
Total	4228.000	71			

Suppose the numbers highlighted in yellow in the table above were not given

H_0: $\boldsymbol{\mu_1 = \mu_2 = \mu_3}$ (One-mean model)

There is no difference in the mean grades of students having different levels of interest.

H_a: $\boldsymbol{\mu_1 , \mu_2 , \mu_3}$ (Three-mean model)

Not all the mean grades of students having different levels of interest are equal. (Or, there is a difference in the mean grades between at least groups.)

k = number of populations being compare = 3

n = total sample size = 72

$$F = \frac{SSTR / k-1}{SSE / n-k} = \frac{MSTR}{MSE} = \frac{590.375}{3047.250 / 72-3} = \frac{590.375}{44.163} = 13.37$$

For df = (k – 1, n – k) = (2, 69) P < 0.001 There is extremely strong evidence against Ho.

Since P-value < α (0.05), reject Ho.

At the 1% significance level, the data provide sufficient evidence to conclude that there is a difference in the mean grades between students having different levels of interest, as expressed at the start of the course (that is, at least two means are different).

(b) The researcher then realized that he had been ignoring the possible effect of gender in the experiment. He did further analysis of the data and came up with the SPSS output in Tables 3 – 6 below. At the 5% significance level, perform the most appropriate test, showing all steps, to determine whether there is a difference in the mean grades between students having different levels of interest after accounting for the effect of gender. For this test, you may consider using any of the SPSS output shown below (Tables 3 – 6) or shown in part (a) (Tables 1 – 2).

Table 3: Summary statistics of grades for the 2 gender groups.

Descriptives

Grade

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Female	36	71.7500	7.32657	1.22109	69.2710	74.2290
Male	36	74.9167	7.87174	1.31196	72.2533	77.5801
Total	72	73.3333	7.71682	.90944	71.5200	75.1467

Table 4: ANOVA table for the comparison of grades for the two gender groups (ignoring level of interest).

ANOVA

Grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	180.500	1	180.500	3.122	.082
Within Groups	4047.500	70	57.821		
Total	4228.000	71			

Table 5: Summary statistics of grades for 6 groups (for all the combinations of levels of interest and gender).

Descriptives

Grade

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
Low-Female	12	65.4167	5.93079	1.71207	61.6484	69.1849
Medium-Female	12	75.5833	6.11196	1.76437	71.7000	79.4667
Super-Female	12	74.2500	5.62664	1.62427	70.6750	77.8250
Low-Male	12	70.8333	6.01261	1.73569	67.0131	74.6536
Medium-Male	12	72.1667	6.35085	1.83333	68.1315	76.2018
Super-Male	12	81.7500	6.57993	1.89946	77.5693	85.9307
Total	72	73.3333	7.71682	.90944	71.5200	75.1467

Table 6: ANOVA table for the comparison of grades for 6 groups (for all the combinations of levels of interest and gender).

ANOVA

Grade

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1764.333	5	352.867	9.453	.000001
Within Groups	2463.667	66	37.328		
Total	4228.000	71			

\>>>>>>>>>> If there is no effect for females

$$\mu_{L-F} = \mu_{M-F} = \mu_{S-F}$$

If there is not effect for males

$$\mu_{L-M} = \mu_{M-M} = \mu_{S-M}$$

$$H_0: \mu_{L-M} = \mu_{M-M} = \mu_{S-M}, \ \mu_{L-F} = \mu_{M-F} = \mu_{S-F}$$

Reduced model = 2-mean model (table 4)

$$H_A: \mu_{L-M}, \mu_{M-M}, \mu_{S-M}, \mu_{L-F}, \mu_{M-F}, \mu_{S-F}$$

Full model = 6-mean model (table 6)

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

$$= \frac{[4047.500 - 2463.667]/[70-66]}{2463.667/66} \qquad F = 10.604$$

$df = [4, 66]$ ← the diff between the full model and the reduced model.

$p < 0.001$, ∴ Therefore there is an extremely strong evidence against H0

$P = 0.00000107$

\>>>>>>>>>>

Example on Application of One-Way ANOVA and the Extra-Sum-of-Squares F-Test

In a certain university there are ten sections of Statistics 252 being taught in the same semester. There are four instructors (A, B, C, and D) and each instructor teaches the sections shown in the table below. Each section has 50 students enrolled. For parts (a), (b), and (c) below, clearly define the best procedure to be applied, but you do not need to actually perform the test since no data is given. In particular, choose the most appropriate test, state the null and alternative hypotheses in terms of the parameters defined in the table, and state the null distribution of the test statistic (that is, name the distribution of the test statistic and indicate the degrees of freedom). Assume all the required assumptions are satisfied.

Define: μ_i =mean mark of the i^{th} section, i = 1, 2, ..., 10

Instructor	Number of lecture sections	Parameters (subscript is the lecture section number)
A	3	μ_1, μ_2, μ_3
B	3	μ_4, μ_5, μ_6
C	2	μ_7, μ_8
D	2	μ_9, μ_{10}

>>>>>>>>>>

(a) Determine whether there is any difference in mean marks between the 10 sections.

$H_0: \mu_1 = \mu_2 = \cdots = \mu_{10}$ one-mean model

$H_A: \mu_1, \mu_2, \ldots, \mu_{10}$ 10-mean model

One factor ANOVA F-test for all 10 sections $k=10$, each with 50 students $\therefore$ $n=500$ df[9, 490]

F-distribution

(b) Determine if any instructor has different mean marks between their own sections.

$H_0: \mu_1 = \mu_2 = \mu_3$, $\mu_4 = \mu_5 = \mu_6$, $\mu_7 = \mu_8$, $\mu_9 = \mu_{10}$ 4-mean model

$H_A: \mu_1, \mu_2, \ldots, \mu_{10}$ 10-mean model

Null distribution

Extra SS F-test

df (6, 490)

(c) Suppose lecture sections 1, 4, 7, and 9 are evening classes and all the other sections are daytime classes, determine whether there is any difference in mean marks either between the evening classes or between the daytime classes.

$H_0: \mu_1 = \mu_4 = \mu_7 = \mu_8, \mu_2 = \mu_3 = \mu_5 = \mu_6 = \mu_8 = \mu_{10}$ 2-mean model

$H_A: \mu_1, \mu_2, \ldots, \mu_{10}$

ESS F-test

df(8, 490)

\>>>>>>>>>>

Example Combining Extra-Sum-of-Squares F-Test with a Review of Other Procedures Covered in This Section

Research Problem: A coral reef researcher measured the heights of randomly sampled *Acropora formosa* colonies along the reef crests of Mbudya Island and Fungu Yasin, on both the landward and seaward sides, making a total of four sites. At the four sites, the heights were normally distributed and the standard deviations were approximately equal. One-way ANOVA, performed at the 5% significance level, showed that there was a difference in the mean heights at the four sites. Use the output in Tables 1 – 5 to answer the questions below.

Table 1: Two-Sample t-test (assuming Equal Variances and independent samples) for the difference in mean height of *Acropora formosa* at Mbudya and Fungu Yasin (data from landward and seaward combined)

	Mbudya	*Fungu Yasin*
Mean	63.4375	60.10714286
Variance	131.47984	124.6917989
Observations	32	28
Pooled Variance	128.31989	
Hypothesized Mean Difference	0	
df	58	
t Stat	1.1361148	
P(T<=t) one-tail	0.1302906	
P(T<=t) two-tail	0.2605812	

Table 2: Summary Statistics for the Four Coral Reef Sites

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Mbudya-Landward	18	1214	67.44444	111.7908		
Mbudya-Seaward	14	816	58.28571	116.5275		
Fungu Yasin-Landward	16	1023	63.9375	107.7958		
Fungu Yasin-Seaward	12	660	55	109.2727		

Table 3: ANOVA table for comparison of all four means

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1373.944	3	457.9814	4.113888	0.010445	2.769431
Within Groups	6234.239	56	111.3257			
Total	7608.183	59				

Table 4: ANOVA table for comparison of Mbudya versus Fungu Yasin (landward and seaward combined)

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	165.6298	1	165.6298	1.290757	0.260581	4.006873
Within Groups	7442.554	58	128.3199			
Total	7608.183	59				

Table 5: ANOVA table for comparison of Landward sites versus Seaward sites (Mbudya and Fungu Yasin combined)

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1200.009	1	1200.009	10.86121	0.001679	4.006873
Within Groups	6408.174	58	110.4858			
Total	7608.183	59				

Define the parameters as follows:

μ_{LM} = mean Acropora height at Landward side of Mbudya

μ_{LF} = mean ... Landward side of Fungu Yasin

μ_{SM} = mean ... Seaward side of Mbudya

μ_{SF} = mean ... Seaward side of Fungu Yasin

(a) The coral reef researcher suspected that the difference between sites was mainly due to the effects of the landward environment (sheltered) versus the seaward environment (exposed to strong wave action). Perform the most appropriate test (a single **overall** test), at the 5% significance level, to determine whether there was a difference in *Acropora formosa* heights between the landward and seaward sides of these reefs, after accounting for the effects of different reefs.

Parameters: (Defined same as above)

If there is no Landward/Seaward effect at Mbudya, then: $\mu_{LM} = \mu_{SM}$

If there is no Landward/Seaward effect at Fungu Yasin, then: $\mu_{LF} = \mu_{SF}$

$H_0 : \mu_{LM} = \mu_{SM}$ and $\mu_{LF} = \mu_{SF}$

[Reduced model: Two means model for only Mbudya and Fungu Yasin]

$H_a : \mu_{LM}, \mu_{SM}, \mu_{LF}, \mu_{SF}$ [Not all four reef sites have the same mean height]

[Full model: Four means model]

From ANOVA table for comparison of Mbudya versus Fungu Yasin (reduced model) (Table 4)

SS_E (Two means model) = 7442.554 and df_E =58

From ANOVA table for comparison of all four means (full model)

SS_E (Four means model) = 6234.239 and df_E =56 (Table 3)

Hence,

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

Extra SS = 7442.554 – 6234.239 = 1208.315

Extra df = df_E (reduced) – df_E (full) = 58 – 56 = 2

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

$$= \frac{1208.315/2}{6234.239/56} = 5.427$$

For the Extra-Sum-of-Squares F-test, $df = (Extra\_df, n-k) = (2,56)$

Thus, 0.005 < P < 0.01, which provides very strong evidence against the null hypothesis

Since P < α (0.05), reject Ho

Conclusion: At the 5% significance level, there is sufficient evidence to conclude that there is a difference in mean height of the coral *Acropora formosa* between the landward and seaward sides of these coral reefs (Mbudya and Fungu Yasin combined).

(b) Suppose the researcher, then also wanted to check if there was any difference between Mbudya and Fungu Yasin, after accounting for the effect of landward versus seaward sides. Again, perform the most appropriate test (a single **overall** test) at the 5% significance level.

If there is no effect of reef (Mbudya/Fungu Yasin) on the Landward side, then: $\mu_{LM} = \mu_{LF}$

If there is no effect of reef (Mbudya/Fungu Yasin) on the Seaward side, then: $\mu_{SM} = \mu_{SF}$

$H_0 : \mu_{LM} = \mu_{LF}$ and $\mu_{SM} = \mu_{SF}$

[Reduced model: Two means model for only Landward and Seaward]

$H_a : \mu_{LM}, \mu_{LF}, \mu_{SM}, \mu_{SF}$

[Full model: Four means model]

Using the ANOVA table for Landward versus Seaward (reduced model) (=Two means model) (Table 5)
And the ANOVA table for comparison of all four means (full model) (=Four means model) (Table 3)

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

$$= \frac{[6408.174 - 6234.239]/[58-56]}{6234.239/56} = \frac{173.935/2}{6234.239/56} = 0.781$$

For the Extra-Sum-of-Squares F-test, $df = (Extra\_df, n-k) = (2,56)$

Thus, P > 0.25, which provides weak evidence against the null hypothesis

Since P > α (0.05), do not reject Ho

Conclusion: At the 5% significance level, there is insufficient evidence to conclude that there is a difference in mean height of the coral *Acropora formosa* between the Mbudya and Fungu Yasin (Landward and Seaward sides combined).

(c) Compare of the pooled two-mean t-test and the single-factor ANOVA F-test with respect to purpose, assumptions, hypotheses and statistical results.

>>>>>>>>>>

Compare the purpose: The purpose of both is to test for a difference between means

Compare assumptions:

Compare hypotheses:

Statistical results of the pooled two-mean t-test:

Statistical results of the single-factor ANOVA F-test:

>>>>>>>>>>

(d) Define a linear combination (using the 4 parameters (means) defined above) to compare the overall mean height of *Acropora formosa* at Mbudya and Fungu Yasin. In addition, determine the estimate of the contrast using the output in Table 2, but you don't have to perform a complete test. How does this estimate of the difference compare to your estimate in part (c)? Whether it is the same or different, explain the reason.

$$\gamma = \frac{(\mu_{LM} + \mu_{SM})}{2} - \frac{(\mu_{LF} + \mu_{SF})}{2}$$
$$\gamma = \frac{1}{2}\mu_{LM} + \frac{1}{2}\mu_{SM} - \frac{1}{2}\mu_{LF} - \frac{1}{2}\mu_{SF}$$
$$\hat{\gamma} = \frac{1}{2}\bar{y}_{LM} + \frac{1}{2}\bar{y}_{SM} - \frac{1}{2}\bar{y}_{LF} - \frac{1}{2}\bar{y}_{SF}$$
$$= \frac{1}{2}(67.44444) + \frac{1}{2}(58.28571) - \frac{1}{2}(63.9375) - \frac{1}{2}(55.0000) = 3.3963$$

This estimate (3.3963) is slightly different from the estimate in part (c) (3.3304). This difference is due to different sample sizes. If the same sizes had been the same, the estimates would have been exactly the same.

(e) Define a linear combination to compare the overall mean height of *Acropora formosa* between landward and seaward sides (regardless of the reef). Use this linear combination to carry out a test, at the 5% significance level, whether there is a difference in mean height between landward and seaward sides.

Contrast: $\gamma = \frac{(\mu_{LM} + \mu_{LF})}{2} - \frac{(\mu_{SM} + \mu_{SF})}{2}$

$$H_0 : \gamma = 0 \qquad\qquad H_a : \gamma \neq 0$$

Estimate: $\hat{\gamma} = \frac{1}{2}\bar{y}_{LM} + \frac{1}{2}\bar{y}_{LF} - \frac{1}{2}\bar{y}_{SM} - \frac{1}{2}\bar{y}_{SF}$

$$= \frac{1}{2}(67.44444) + \frac{1}{2}(63.9375) - \frac{1}{2}(58.28571) - \frac{1}{2}(55.0000) = 9.0481$$

Standard error of the estimate:

$$SE(\hat{\gamma}) = s_p\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_k^2}{n_k}} \qquad\qquad s_p = \sqrt{MSE} = \sqrt{111.3257} = 10.5511$$

$$SE(\hat{\gamma}) = (10.5511)\sqrt{\frac{(1/2)^2}{18} + \frac{(1/2)^2}{16} + \frac{(-1/2)^2}{14} + \frac{(-1/2)^2}{12}} = 2.75534$$

Observed value of the test-statistic

$$t = \frac{\hat{\gamma} - 0}{SE(\hat{\gamma})} = \frac{9.0481}{2.75534} = 3.284$$

$df = n - k = 60 - 4 = 56$ P-value: (0.0005 < P < 0.001) x 2 ➔ 0.001 < P < 0.002

Since P < α (0.05), reject H_0 with very strong evidence.

Conclusion: At the 5% significance level, there is a difference in mean height of *Acropora formosa* between landward and seaward sides.

(f) Does the effect of the side of the reef (landward or seaward) depend on the reef (Mbudya or Fungu Yasin)? Define a linear combination and carry out a test (at α = 0.05) to answer this question.

The effect of the side of the reef are:

For Mbudya: $\mu_{LM} - \mu_{SM}$

For Fungu Yasin: $\mu_{LF} - \mu_{SF}$

If the effect of the side of the reef does not depend upon which reef it is, then:

$\mu_{LM} - \mu_{SM} = \mu_{LF} - \mu_{SF}$ ➔which means that: $\mu_{LM} - \mu_{SM} - \mu_{LF} + \mu_{SF} = 0$

Thus, the linear combination is:

$$\gamma = \mu_{LM} - \mu_{SM} - \mu_{LF} + \mu_{SF}$$

The estimate for the linear combination is:

$$\hat{\gamma} = \bar{y}_{LM} - \bar{y}_{SM} - \bar{y}_{LF} + \bar{y}_{SF} = 67.44444 - 58.28571 - 63.9375 + 55 = 0.2212$$

Standard error of the estimate:

$$SE(\hat{\gamma}) = (10.5511)\sqrt{\frac{(1)^2}{18} + \frac{(-1)^2}{16} + \frac{(-1)^2}{14} + \frac{(1)^2}{12}} = 5.51107$$

Observed value of the t-statistic:

$$t = \frac{\hat{\gamma} - 0}{SE(\hat{\gamma})} = \frac{0.2212}{5.51107} = 0.04014$$

$df = n - k = 60 - 4 = 56$

P-value: (P > 0.25) x 2 ➔ P > 0.50

Since P > α (0.05), do not reject H_0 since there is weak evidence against it.

Conclusion: At the 5% significance level, the effect of the side of the reef (landward or seaward) does not depend on the reef (Mbudya or Fungu Yasin).

(g) Use the Bonferroni method to calculate two simultaneous 96% confidence intervals for the difference in mean height of Acropora Formosa between the landward side and seawards side for each reef separately.

So, we need to find 96% familywise confidence intervals for:

i) Effect of the side of the reef for Mbudya: $\gamma_M = \mu_{LM} - \mu_{SM}$

ii) Effect of the side of the reef for Fungu Yasin: $\gamma_F = \mu_{LF} - \mu_{SF}$

For 96% confidence, α = 0.04

$$\alpha_I = \frac{\alpha_F}{m} = \frac{0.04}{2} = 0.02$$

[**Note:** Here we do not use the formula $m = \frac{k(k-1)}{2}$ because this is not multiple comparisons where we want to compare all possible means pairwise; but rather, we are calculating 2 simultaneous confidence intervals, so m = 2.]

The critical value = $t_{n-k,\alpha_I/2} = t_{60-4,0.02/2} = t_{56,0.01} = 2.403$

Using the formula: $\hat{\gamma} \pm \text{Critical value} \times SE(\hat{\gamma})$

For the effect of the side of the reef at Mbudya:

$$\hat{\gamma}_M = \bar{y}_{LM} - \bar{y}_{SM} = 67.44444 - 58.28571 = 9.15873$$

$$s_p = \sqrt{MSE} = \sqrt{111.3257} = 10.5511$$

$$SE(\hat{\gamma}) = (10.5511)\sqrt{\frac{(1)^2}{18} + \frac{(-1)^2}{14}} = 3.75987$$

$$9.15873 \pm 2.403 \times 3.75987 \Rightarrow (9.15873 \pm 9.0350)$$

$$(0.124, 18.194)$$

For the effect of the side of the reef at Fungu Yasin:

$$\gamma_F = \bar{y}_{LF} - \bar{y}_{SF} = 63.9375 - 55.0000 = 8.9375$$

$$SE(\hat{\gamma}) = (10.5511)\sqrt{\frac{(1)^2}{16} + \frac{(-1)^2}{12}} = 4.02927$$

$$8.9375 \pm 2.403 \times 4.02927 \Rightarrow (8.9375 \pm 9.6823)$$

$$(-0.745, 18.620)$$

3.7 The Kruskal-Wallis test (a Nonparametric Equivalent of One-Way ANOVA)

- Can be used in all situations where there are k independent samples and k > 2
- If the data fit the assumptions of ANOVA, the Kruskal-Wallis Test will be $3/\pi$ =95.5% as powerful as ANOVA
- If the data do not fit the assumptions of ANOVA, the Kruskal-Wallis Test will be more powerful than ANOVA
- The data are ranked in order from lowest to highest (across all k groups) and calculations are performed on the ranks
- Where there are tied observations, assign the average rank to the tied observations

Importance of the Kruskal-Wallis test and other Nonparametric tests

<u>If</u>

- one or more of the data sets being compared are not normally distributed nor are they lognormal,

<u>And</u>

- when the one or more sample sizes are less than 30 (Central Limit Theorem),

<u>The Kruskal-Wallis test is the only valid option</u> (one-factor ANOVA cannot be performed)

- Also, since the Kruskal-Wallis test converts the raw data to ranks, it is <u>not affected by outliers</u> or unequal standard deviations, while these would affect one-way ANOVA.

Kruskal-Wallis Test

Purpose: To test for a difference between k population (where k > 2)

Assumptions:

1. Simple Random samples
2. Independent samples
3. Same-shape populations
4. All sample sizes are 5 or greater

The null and alternative hypotheses:

H_0: The population distributions of ***k*** populations are identical.
H_a: The population distributions of ***k*** populations are not all identical, that is, at least two are different.

Calculating the test statistic:

First, rank the data from all k samples combined, from lowest to highest
Assign average ranks where there are tied observations

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

Where n = total number of observations

$n_1, n_2, \ldots, n_k$ denote sample sizes of samples 1, 2,...k

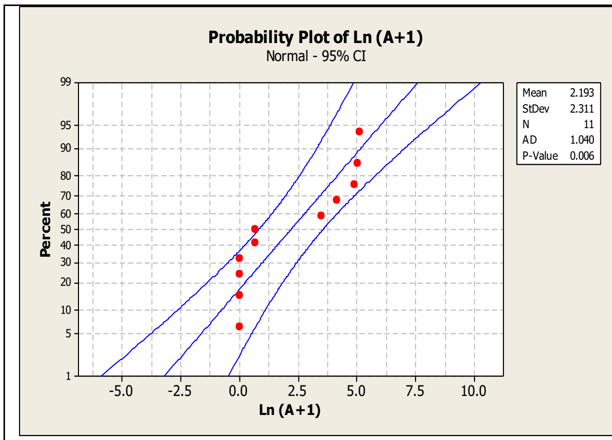
$R_1, R_2, \ldots, R_k$ denote the sums of the ranks for the sample data

Critical values of ***H*** follow the χ_α^2 distribution with **df = k – 1**

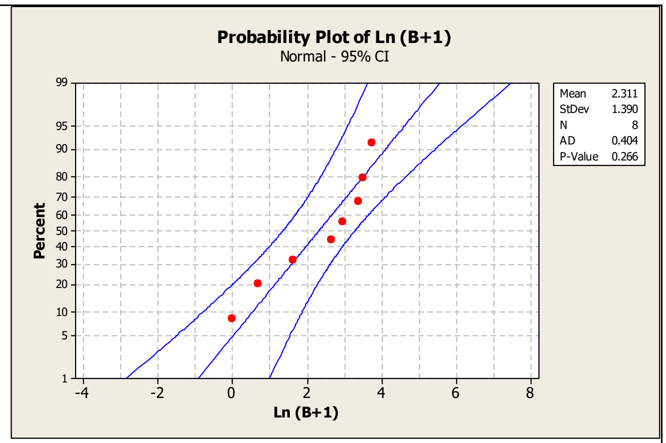
Research Problem:
Pitfall traps are inserted into the soil at ground level in three grasslands (A, B and C) in order to determine whether there is a difference in the abundance of ants in the three grasslands. Test this hypothesis at the 10% significance level.

	Number of ants per pitfall trap										
Grassland A	168	0	62	0	1	135	0	0	155	32	1
Grassland B	0	13	28	32	18	4	41	1			
Grassland C	144	1	3	135	45	3	122	4	110		

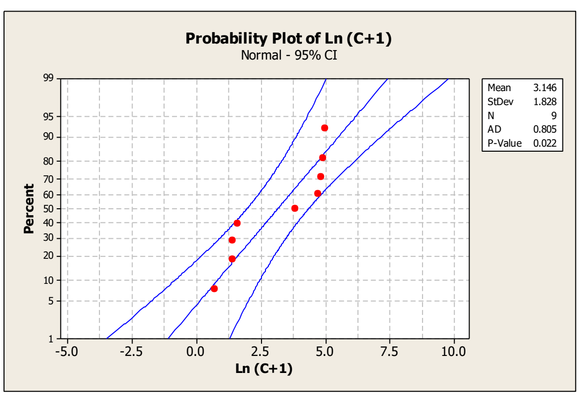
This shows an <u>aggregated</u> distribution.
(**Note** the difference between aggregated, random and regular (even) distributions in space or in time.)



Normal Probability Plot for logged transformed data from grassland A



Normal Probability Plot for logged transformed data from grassland B



Normal Probability Plot for logged transformed data from grassland C

Step 1: The purpose is to compare k populations

- 3 independent random samples
  However:
  - The data are neither normal nor <u>lognormal</u>
  - Sample size is < 30, therefore the Central Limit Theorem does not apply
  - Therefore, the Kruskal-Wallis Test must be performed
- Same shape distributions, as indicated in the NPPs above
- Sample size of all groups ≥ 5.

Step 2: H_0: There is no difference in the abundance of ants in the three grasslands.
H_a: There is a difference in the abundance of ants in the three grasslands (at least two are different).

Step 3: Calculate the test statistic H
Rank the data from lowest to highest, assigning average ranks where there are tied observations.
>>>>>>>>>>

	Grassland A		Grassland B		Grassland C	
	No. of ants	Rank	No. of ants	Rank	No. of ants	Rank
	168	28	0	3	144	26
	0	3	13	14	1	7.5
	62	21	28	16	3	10.5
	0	3	32	17.5	135	24.5
	1	7.5	18	15	45	20
	135	24	4	12.5	3	10.5
	0	3	41	19	122	33
	0	3	1	7.5	4	12.5
	155	27			110	22
	32	17.5				
	1	7.5				
Sum of ranks (R_j)		145		104.5		156.5
Sample size (n_j)		11		8		9

$n = 28$

$$H = \frac{12}{n(n+1)} \sum \ldots = \frac{12}{28(29)} \left[\frac{(145)^2}{11} + \frac{(104.5)^2}{8} + \frac{(156.5)^2}{9} \right] - 3(28+1)$$

$= 1.647$

$df = k - 1 = 2$

$p > 0.200$

>>>>>>>>>>