Notes:

Analysis of Variance (ANOVA) In Minitab

Key Learning Points

1. Describe the importance of using a one-way ANOVA.
2. Explain why to use a one-way ANOVA.
3. Utilize one-way ANOVA in improvement projects.

What is Analysis of Variance?

Analysis of Variance (ANOVA) tests the equality of two or more continuous population means. This is done by partitioning total variation into between group and within group components. For statistical significance, the variation between groups must be large relative to the variation within groups.

ANOVA tests if specific discrete inputs or factors have a significant effect on the continuous response. Classifications of each factor must have two or more levels. A one-way ANOVA is used when there is only one factor with at least two levels.

Highlight

A one-way ANOVA with just two levels is the same as a 2 Sample t-test.

Difference

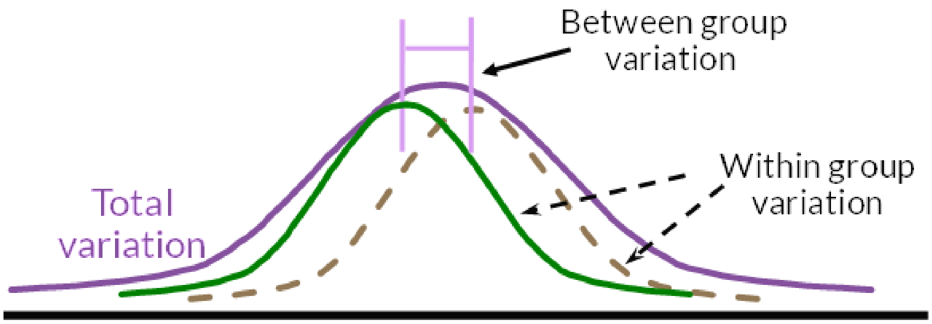
One-way ANOVA tests for significant differences between stratifications of a continuous population. The stratification variable is known as a factor and the individual classifications are levels or treatments.

Definitions

- Response Variable: The output of a process. This is often referred to as a dependent variable. It is a continuous measure.
- Factor: An input into a process. There are two more levels, or discrete intervals for each factor. Factors must be classified in discrete intervals.
- Treatment: A specific combination of factor levels whose effect is to be compared with other treatments.
- One-way ANOVA: When there is only one treatment but multiple levels.

Variation

Is the mean of the variation in these groups equal?



Reading ANOVA Results

There are multiple levels of analysis with an ANOVA test.

- P-value details the probability of being wrong if we conclude at least one difference. Must be >0.05 for 95% Significance.
- Confidence intervals of means identify significant differences between groups.
- R^2 is the coefficient of determination. It is the percentage of total variation in Y explained by the test variable.

Animal Hospital Example

Over the course of his prestigious career, Dr. D, a prominent veterinarian at the local animal hospital noticed that dogs blood coagulates in varying amounts of time. He posed the theory that this variation is due to each dogs diet.

Step 1: State the Practical Problem

The Practical Problem:

Does diet affect blood coagulation time in dogs?

Notes:

Step 2: Establish the Hypotheses

Notes:

H_o: $\mu_{\text{diet 1}} = \mu_{\text{diet 2}} = \mu_{\text{diet 3}} = \mu_{\text{diet 4}}$
or
H_o: τt's = 0

Interpretation of the null hypothesis: The average blood coagulation time of each diet is the same (or what the dog eats does NOT affect blood coagulation time).

H_a: at least one diet is different
or
H_a: τt's $\neq 0$

Interpretation of the alternative hypothesis: At least one diet affects the average blood coagulation time differently than another (or what type of diet the dog eats does affect blood coagulation time).

Step 3: Decide on Appropriate Statistical Test

ANOVA compares continuous population means between two or more stratifications.

Step 4: Set the Alpha Level

Dr. D wants to be 95% confident that he is making the correct conclusion.
$\alpha = 0.05$

Steps 5&6: Set the Power and Sample Size, and Collect the Data

As more factors and levels of factors are added to the test, the sample size needs to increase to support them.

When collecting the data, ensure that treatments and treatment levels are assigned randomly. Be sure to balance the design among treatments and levels.

The Data

- Twenty-four dogs receive one of four diets.
- The type of diet is the factor of interest.
- Blood coagulation time is the response (output).
- During the experiment, diets were assigned randomly to dogs.
- Blood samples were taken and tested in random order.

Diet A	Diet B	Diet C	Diet D
62	63	68	56
60	67	66	62
63	71	71	60
59	64	67	61
	65	68	63
	66	68	64
			63
			59

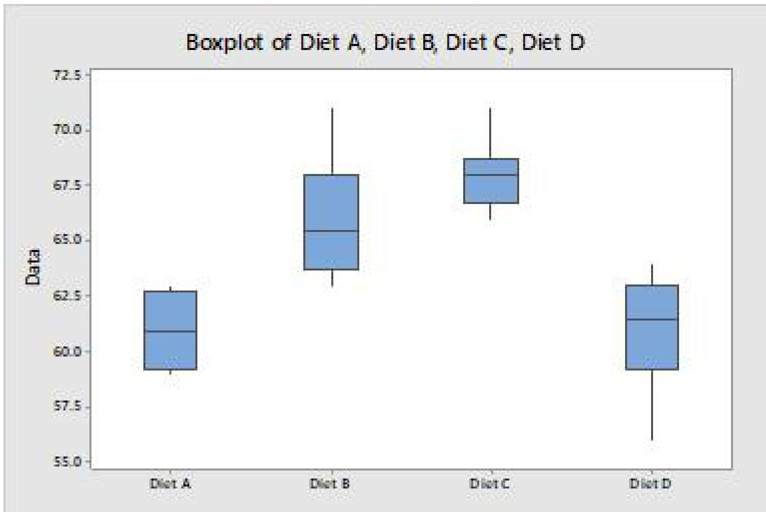
Notes:

Step 7: Use the Appropriate Graphical Tool To Explore the Data

A box plot is the perfect tool for exploring these data.

Box Plot

- **Minitab: Graph > Box Plot**
  - Multiple Ys's > Simple
  - Graph Variables Diet A – Diet D



Step 8: Check Data Assumptions

There are two assumptions that must be tested for ANOVA. First was the test for normality.

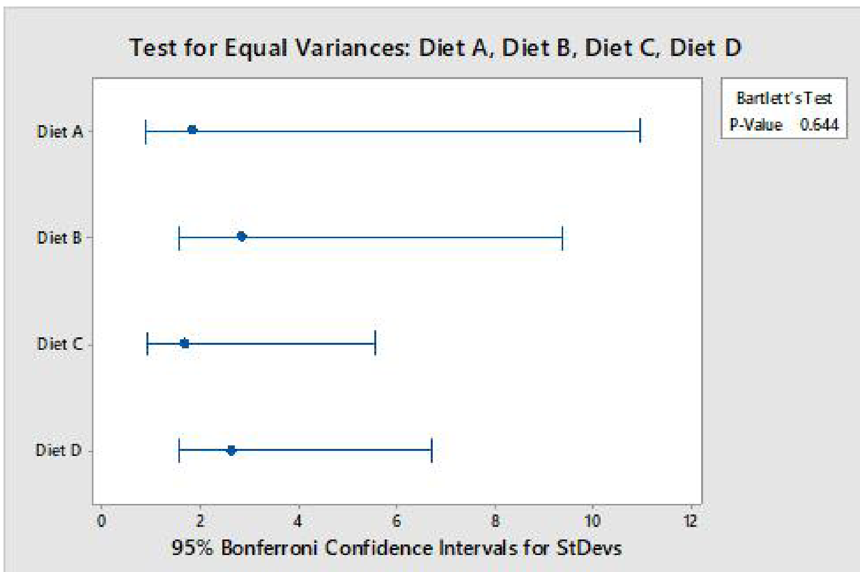
Then, we perform a test for equal variance to ensure that population variances were equal across all factor levels.

Test for Equal Variance

- **Minitab: Stat > ANOVA > Test for Equal Variances**
  - Response data are in separate column for each factor level
  - Responses Diet A – Diet D
  - Options > Use test based on normal distribution

Note: If variances are not equal, then the "between level" difference would be overwhelmed by the variation within one level. This assumption of equal variances generally holds, especially if your test is BALANCED (same # of observations in each level).

Results



Based on this test, Dr. D concluded that the variances are equal.

Step 9: Run the Statistical Test

Dr. D was now ready to create the ANOVA table in Minitab.

- **Minitab: Stat > ANOVA > One-Way**
  - Response data are in separate columns
  - Responses: Diet A-Diet D
  - Confidence Level: 95.0
  - Graph: Three in One

Notes:

One-Way ANOVA: Time vs. Diet

One-way ANOVA: Time versus Diet

Method

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

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Diet	4	Diet A, Diet B, Diet C, Diet D

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Diet	3	228.0	76.000	13.57	0.000
Error	20	112.0	5.600		
Total	23	340.0			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.36643	67.06%	62.12%	53.46%

Means

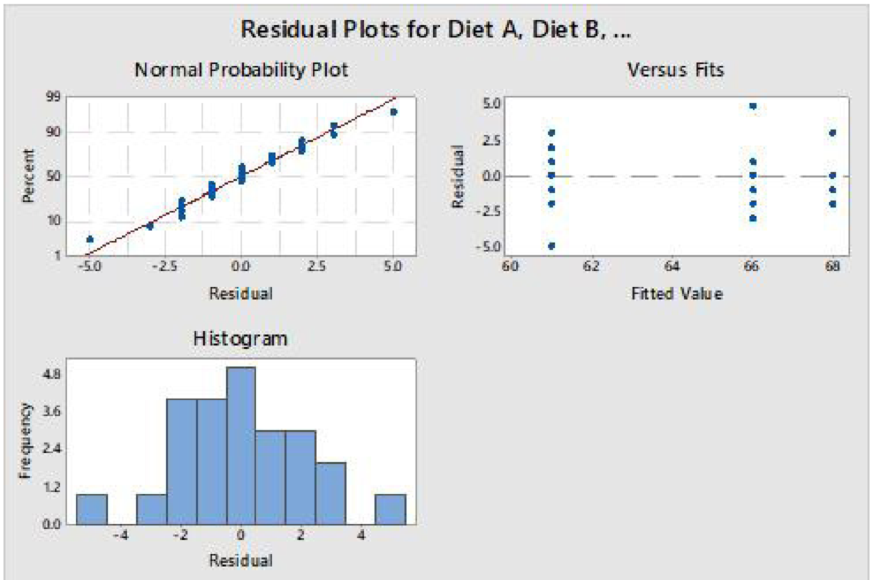
Diet	N	Mean	StDev	95% CI
Diet A	4	61.000	1.826	(58.532, 63.468)
Diet B	6	66.00	2.83	(63.98, 68.02)
Diet C	6	68.000	1.673	(65.985, 70.015)
Diet D	8	61.000	2.619	(59.255, 62.745)

Pooled StDev = 2.36643

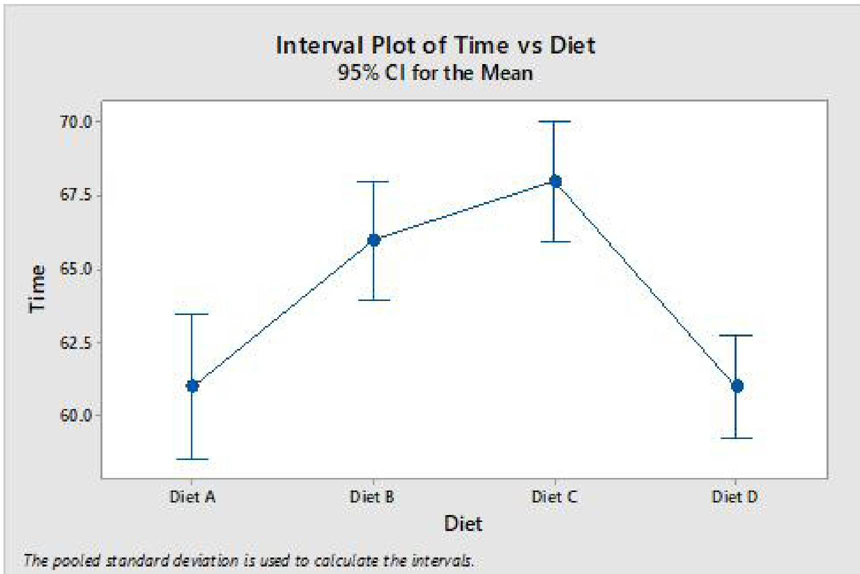
Interval Plot of Time vs Diet

Residual Plots for Time

Residual Plots for Time



Interval Plot of Time vs. Diet



Fits and Residuals

The Fit is just the mean of each Diet Type.
The Residual is (Observed Score – Mean).

Diet	Time	FITS	RESI
Diet A	62	61	1
Diet A	60	61	-1
Diet A	63	61	2
Diet A	59	61	-2
Diet B	63	66	-3
Diet B	67	66	1
Diet B	71	66	5
Diet B	64	66	-2
Diet B	65	66	-1
Diet B	66	66	0
Diet C	68	68	0
Diet C	66	68	-2
Diet C	71	68	3
Diet C	7	68	-1
Diet C	68	68	0
Diet C	68	68	0
Diet D	56	61	-5
Diet D	62	61	1
Diet D	60	61	-1
Diet D	61	61	0
Diet D	63	61	2
Diet D	64	61	3
Diet D	63	61	2
Diet D	59	61	-2

Notes:

Step 9a: Do the Assumptions for the Errors Hold?

Residuals are the difference between the expected blood coagulation times per animal and the actual blood coagulation. The expected time is equal to the sample average for each diet. If all model assumptions hold true, ANOVA residuals should be distributed normally.

Use the “three in one” or ”four in one” residual plots to check assumptions.

Statistical Conclusion

The P-Value (0.000) for the factor effect is less than the alpha level of 0.05, and therefore, the null hypothesis is rejected.

Diets A and D have significantly shorter coagulation times than Diets B and C.

Step 10: Translate the Statistical Conclusion Into a Practical Conclusion

Dr. D can conclude that diet type DOES affect blood coagulation time. In particular, if you would like to increase blood coagulation time, use diets B or C. If the goal is to decrease blood coagulation time, use diets A or D.

When Should ANOVA Be Used?

The ANOVA tests for differences between two or more continuous population means.

Pitfalls to Avoid

- Data must be normally distributed for the One-way ANOVA.
- The test will lose power if the variation between groups is not equal.
- Check residual plots to ensure model assumptions are valid.
- Be sure to test for power if you fail to reject your null hypothesis.

Notes: