PUBH620 Biostatistics

Life in Statistics

Week 4

One-way ANOVA and Post-Hoc Testing

Lecture notes by Dr Brandon Cheong and Dr Mike Steele

ELECTRONIC WARNING NOTICE

In this lecture…

(clicking the links below will direct you to the topic page)

Topic Learning Objectives (TLOs)

- Understand the concepts of ANOVA and appreciate the impact of these concepts in public health practice
- To develop the ability to identify public health problems where ANOVA and extensions of ANOVA can be used

4.1 One-Way ANOVA

The Analysis of Variance is widely used and is an extremely powerful statistical test in public health research.

It is often used to compare the means of three or more groups however, it has other uses as well:

- ANOVA may be used as an **exploratory technique** to observe data within and between groups.
- ANOVA can be used for **estimating parameters** of a research design model.
- ANOVA can also be used for **building models** to account for data and explain underlying processes

ANOVA is deemed more powerful than running multiple t-tests because it doesn't increase the chances of type I errors.

ANOVA does this by performing a single test that tests the null hypothesis for 'k' groups. 'k' being the number of experimental groups or samples.

4.1.1 Introduction to one-way ANOVA

The One-Way ANOVA is a statistical test that compares the means of three or more groups.

It is used to test for a statistical difference between sample means of three or more groups.

If there is a statistical difference, then post-hoc testing is required.

Summary of definitions

- Response: The variable of interest to be measured in the experiment is referred to as the response or dependent variable. (Note: this term is commonly used in regression.
- Factor: Those variable(s), whose effect on the response is of interest, are called factors or, when used in a regression, independent variables.
- Level: The values of a factor to be used in an experiment are called levels.
- Experimental Unit: is the object on which the response variable and factors are observed. This can be the person or the animal or whatever you are measuring.

4.1.1 Introduction to one-way ANOVA

Eg. Do different animal diets result in different body weights?

Since we only have one factor (diet) on the variable (bodyweight of animal), we call this a “One-Way ANOVA”. The levels here are the type of diet, let’s say we have 4 diets: Diet A, Diet B, Diet C and Diet D and 19 animals. The design of the experiment should have each experimental unit (“observation” or what SPSS calls a “case”) assigned at random to receive one of the four diets.

Ideally we would try to have similar numbers of animals assigned to each diet. However, having unbalanced sample sizes is OK as we can just use statistical software to do the analysis.

4.1.1 Introduction to one-way ANOVA

Observed values:

	Diet A (kg)	Diet B (kg)	Diet C (kg)	Diet D (kg)	
	60.80	68.70	102.60	87.90	
	57.00	67.70	102.10	84.20	
	65.00	74.00	100.20	83.10	
	58.60	66.30	96.50	85.70	
	61.70	69.80		90.30	
Mean	**60.62**	**69.30**	**100.35**	**86.24**	**78.01**

78.01 is known as the Grand Mean, which is just the mean of all 4 groups together.

4.1.1 Introduction to one-way ANOVA

When we perform a one-way ANOVA, we usually address it in two parts. This involves us looking at the relative difference between:

- The variability *between* the treatment/group means (of all four diet groups).
- The variability *within* each treatment/group (known as the residual).

What this means is that we will be more likely to reject the null hypothesis if there is a large variability between the treatment means, and a small variability within each treatment.

Keeping in mind that our H_0 for a one-way ANOVA is that there is no difference in the means of the four diet groups.

So, the two components that we look at are:

1. $Total\ sum\ of\ squares = Treatment\ sum\ of\ squares + Residual\ sum\ of\ squares$
2. $Residual\ values = Observed\ values\ - Expected\ values$

4.1.1 Introduction to one-way ANOVA

Expected values:

Are the calculated means from observed values.

Diet A (kg)	Diet B (kg)	Diet C (kg)	Diet D (kg)
60.62	69.30	100.35	86.24
60.62	69.30	100.35	86.24
60.62	69.30	100.35	86.24
60.62	69.30	100.35	86.24
60.62	69.30		86.24

Residual values:

Are observed values minus the expected values.

Diet A (kg)	Diet B (kg)	Diet C (kg)	Diet D (kg)
0.18	-0.60	2.25	1.66
-3.62	-1.60	1.75	-2.04
4.38	4.70	-0.15	-3.14
-2.02	-3.00	-3.85	-0.54
1.08	0.50		4.06

4.1.2 Assumptions of one-way ANOVA

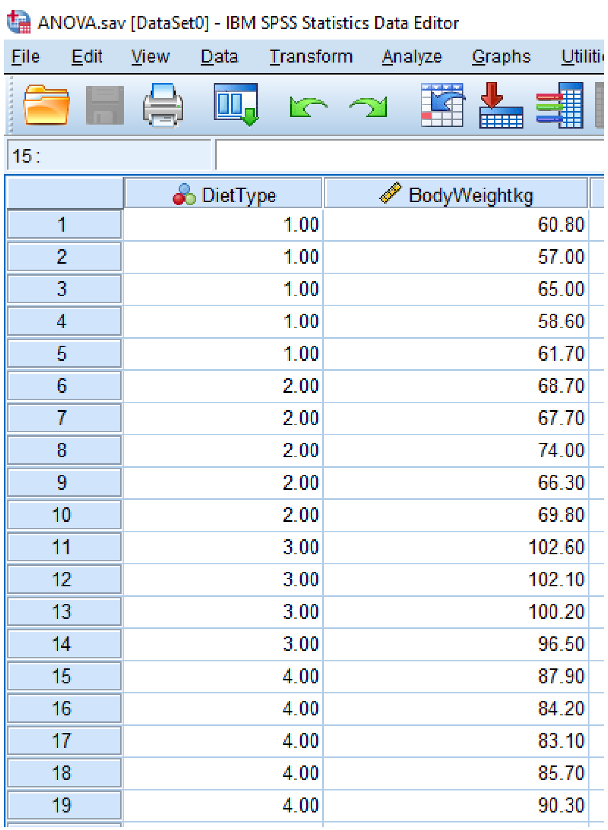
Like all statistical tests, the one-way ANOVA also has assumptions that need to be checked:

- Scale of Measurement. The dependent variable (response variable) data must be measured on a ratio or interval scale.
- Independence. Each participant should only participate only once in the research, and should not influence the participation of others.
- Normality. Each group should be approximately normally distributed however, ANOVA is adaptable to violations of this assumption.
- Homogeneity of Variances. Each group should have an approximately equal amount of variability.

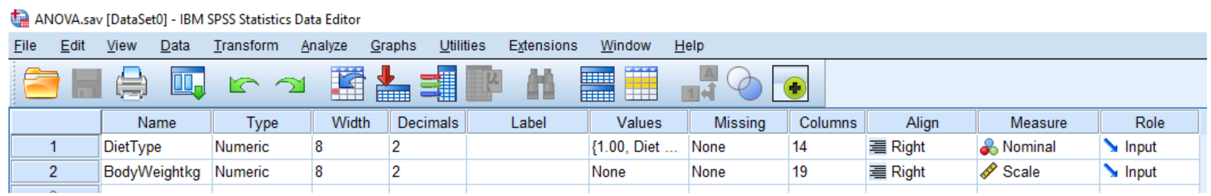
Check these assumptions using SPSS!

4.1.2 Assumptions of one-way ANOVA

Checking ANOVA assumptions in SPSS



	DietType	BodyWeightkg
1	1.00	60.80
2	1.00	57.00
3	1.00	65.00
4	1.00	58.60
5	1.00	61.70
6	2.00	68.70
7	2.00	67.70
8	2.00	74.00
9	2.00	66.30
10	2.00	69.80
11	3.00	102.60
12	3.00	102.10
13	3.00	100.20
14	3.00	96.50
15	4.00	87.90
16	4.00	84.20
17	4.00	83.10
18	4.00	85.70
19	4.00	90.30

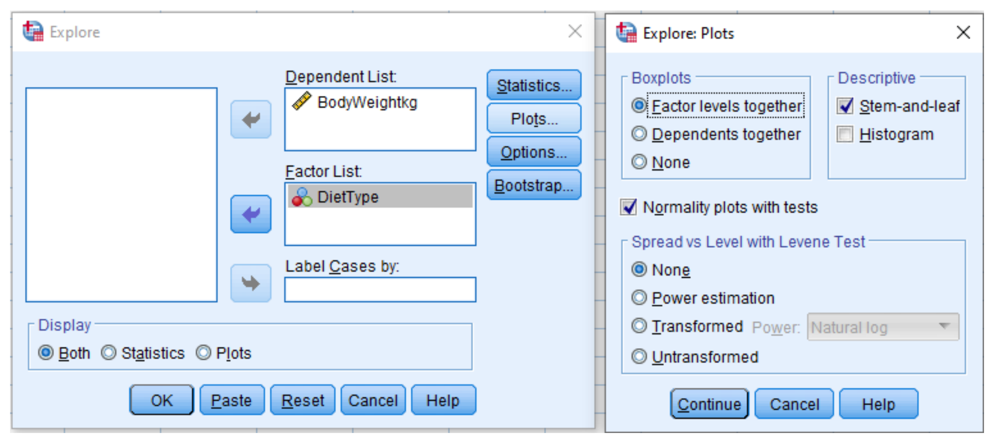


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	DietType	Numeric	8	2		{1.00, Diet ...	None	14	Right	Nominal	Input
2	BodyWeightkg	Numeric	8	2		None	None	19	Right	Scale	Input

<Analyze>

<Descriptive Statistics>

<Explore> etc. etc…



4.1.2 Assumptions of one-way ANOVA

Checking ANOVA assumptions in SPSS

Data are measured in kg – Assumption 1 satisfied

Data are independent – Assumption 2 satisfied

Data are Normal – Assumption 3 satisfied

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	DietType	Statistic	df	Sig.	Statistic	df	Sig.
BodyWeightkg	Diet A	.162	5	.200*	.979	5	.931
	Diet B	.232	5	.200*	.923	5	.552
	Diet C	.236	4	.	.885	4	.360
	Diet D	.174	5	.200*	.961	5	.815

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Data have equal variances – Need to assess this using Levene’s test option when running a one-way ANOVA in SPSS.

4.1.3 Calculating the test statistic F for one-way ANOVA

The Total Sum of Squares is equal to the sum of the squared deviations between the observed values and the grand mean. The grand mean is the average of all the 19 animals body weights, which is 78.01 kg.

$$
\begin{aligned}
&Total\ sum\ of\ squares\\
&= (60.80 - 78.01)^2 + (65 - 78.01)^2 + (58.6 - 78.01)^2 + \cdots + (90.30 - 78.01)^2\\
&= 4354.698
\end{aligned}
$$

This formula is very similar to how we would calculate variance. The only difference is that here we do not divide by *n-1*.

You can see how this might be a pain to calculate by hand! If you do decide to do it ‘by hand’…use Excel!

4.1.3 Calculating the test statistic F for one-way ANOVA

The Treatment sum of squares (Between Groups Variance) is calculated by squaring the difference between each of the respective group means and grand means then summing them all up together:

$$Treatment\ sum\ of\ squares$$
$$= 5 \times (60.62 - 78.01)^2 + 5 \times (69.30 - 78.01)^2 + 4$$
$$\times (100.35 - 78.01)^2 + 5 \times (86.24 - 78.01)^2 = 4226.348$$

Again, use Excel to do these calculations!

4.1.3 Calculating the test statistic F for one-way ANOVA

The Residual sum of squares (Within Groups Variance) is calculated by squaring the difference between observed and expected values for each group and then summing them all up together:

Residual sum of squares = $(60.8 - 60.62)^2 + (57.0 - 60.62)^2 + (65.0 - 60.62)^2 + (58.6 - 60.62)^2 + (61.7 - 60.62)^2 + (68.7 - 69.30)^2 + (67.7 - 69.30)^2 + (74.0 - 69.30)^2 + (66.3 - 69.30)^2 + (69.8 - 69.30)^2 + (102.6 - 100.35)^2 + (102.1 - 100.35)^2 + (100.2 - 100.35)^2 + (96.5 - 100.35)^2 +$

$(87.9 - 86.24)^2 + (84.2 - 86.24)^2 + (83.1 - 86.24)^2 + (85.7 - 86.24)^2 + (90.3 - 86.24)^2$

=128.350

Or just use SPSS!

4.1.3 Calculating the test statistic F for one-way ANOVA

Now we have calculated three of the more involved summary statistics required for the obtaining the test statistic F:

Source of Variation	Sum of squares
Treatment sum of squares (between groups)	4226.348
Residual sum of squares (within groups)	128.350
Total sum of squares	4354.698

In order to get the test statistic F, we need to calculate the degrees of freedom for each variance component:

df (between) = k – 1, where k = number of treatment groups

df (within) = N - k, where N = total number of participants (animals)

Therefore df(between) and df(within) = 3 and 15 respectively.

4.1.3 Calculating the test statistic F for one-way ANOVA

Final ANOVA table to calculate test statistic F:

The variance is also known as the 'mean square'. Variance is calculated by taking the sum of squares and dividing it by the degrees of freedom.

F is then the ratio of the mean square treatment over the mean square residual:

$$F = \frac{MST}{MSR}$$

Source of Variation	SS	df	Mean Square	F
Treatment sum of squares (between groups)	4226.348	3	4226.348/3 = 1408.783	MST/MSR = 164.635
Residual sum of squares (within groups)	128.350	15	128.350/15 = 8.557	
Total sum of squares	4354.698	18		

In practice, just use SPSS!

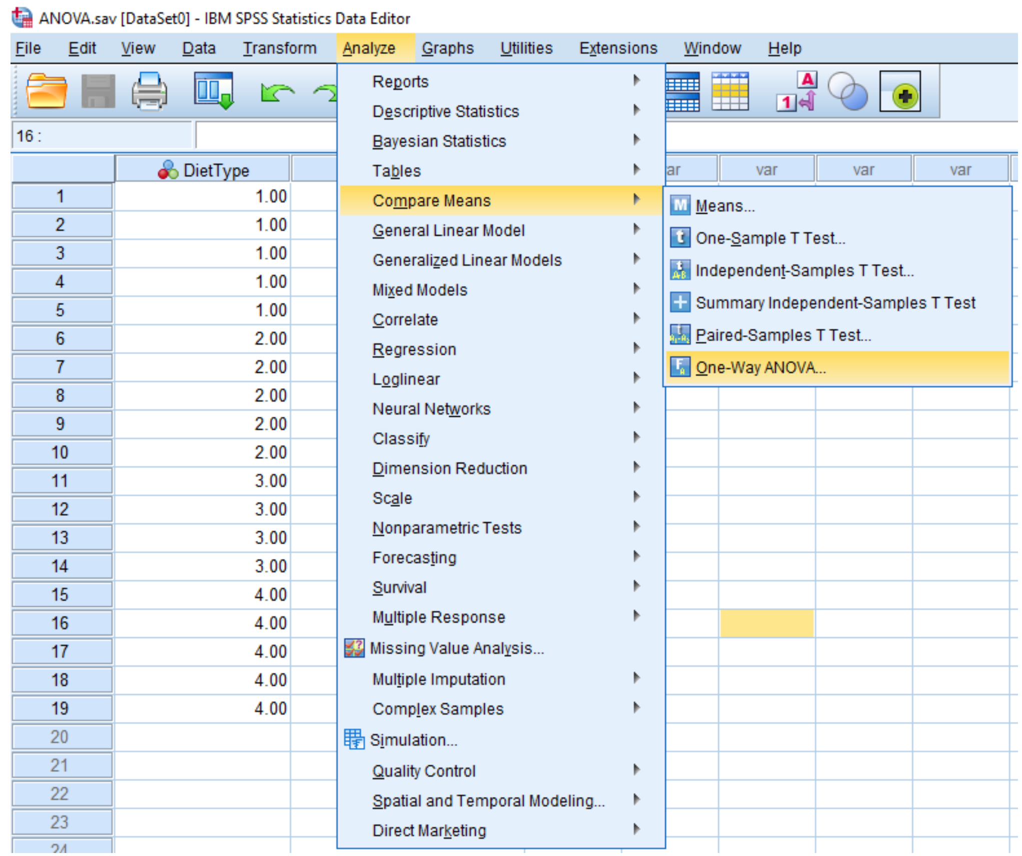
4.1.4 One-way ANOVA in SPSS

Click

<Analyze>

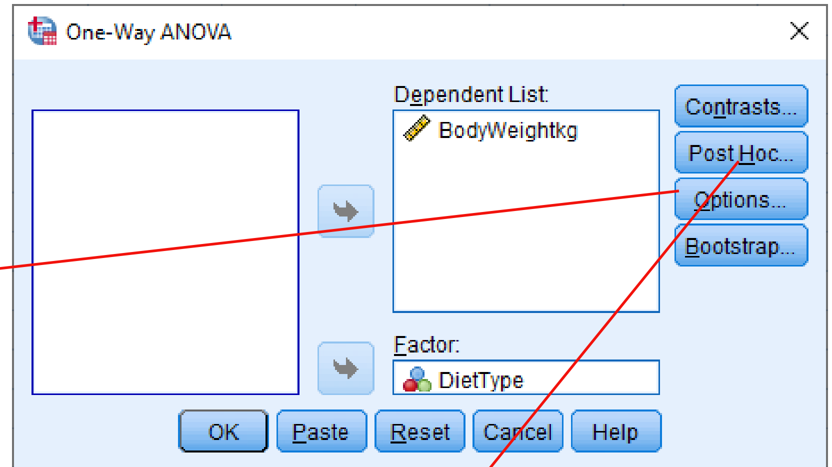
<Compare Means>

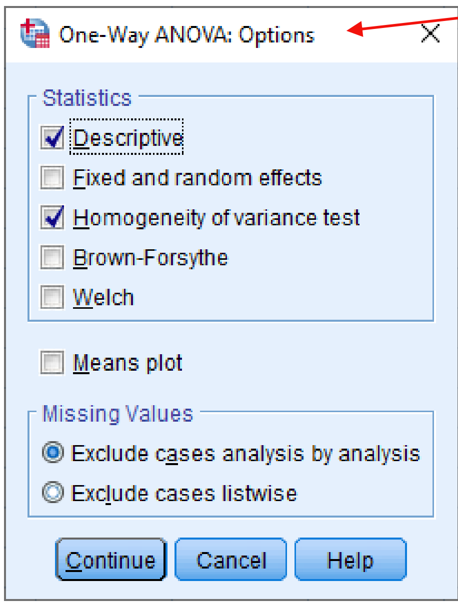
<One-Way ANOVA>

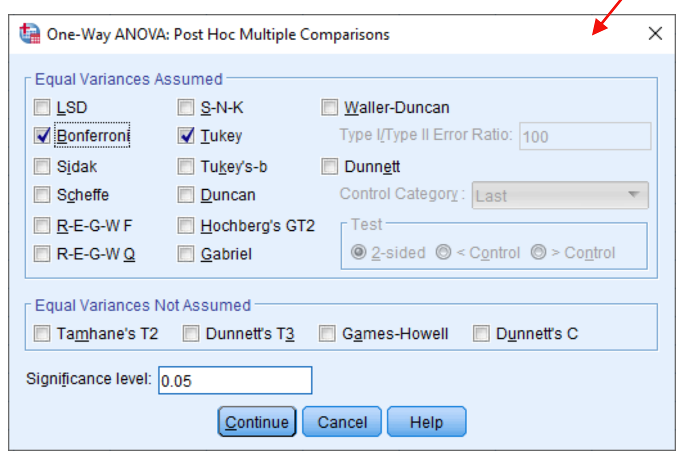


4.1.4 One-way ANOVA in SPSS

ANOVA in SPSS







Post-Hoc testing only required if ANOVA gives statistically significant result. Only need to choose **one** of Bonferroni or Tukey!

4.1.4 One-way ANOVA in SPSS

ANOVA output in SPSS

Descriptives

BodyWeightkg

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Diet A	5	60.6200	3.06464	1.37055	56.8148	64.4252	57.00	65.00
Diet B	5	69.3000	2.92660	1.30882	65.6661	72.9339	66.30	74.00
Diet C	4	100.3500	2.76707	1.38353	95.9470	104.7530	96.50	102.60
Diet D	5	86.2400	2.89620	1.29522	82.6439	89.8361	83.10	90.30
Total	19	78.0105	15.55402	3.56834	70.5137	85.5073	57.00	102.60

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
BodyWeightkg	Based on Mean	.034	3	15	.991
	Based on Median	.024	3	15	.995
	Based on Median and with adjusted df	.024	3	14.638	.995
	Based on trimmed mean	.035	3	15	.991

P-value of 0.991 shows non-significance for Levene's test. Hence, we have equal variances.

4.1.4 One-way ANOVA in SPSS

ANOVA output in SPSS

ANOVA

BodyWeightkg

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4226.348	3	1408.783	164.642	.000
Within Groups	128.350	15	8.557		
Total	4354.698	18			

Just like we calculated! We have statistical significance ($p < 0.001$) so need to use post-hoc testing to do pairwise comparison.

4.1.5 Post-Hoc Testing

Post-hoc is latin for “after the fact”, which gives an idea of its purpose in statistical testing. It is only used after completing a one-way ANOVA test that gave a statistically significant result. **If the ANOVA test was not significant then post-hoc testing is not required**.

There is much debate over which procedure is the best to routinely employ. The two most common are the Bonferroni and Tukey’s test. If you reject the null hypothesis in a one-way ANOVA, either test is fine.

These tests involve performing multiple tests for each pair of means. Although a series of tests are performed, the significance level of the test is adjusted, so that in total the probability of committing a type I error remains equal to α.

These tests are easily done through SPSS by clicking the “POST-HOC” button in the one-way ANOVA menu and selecting the desired test (Bonferroni OR Tukey).

4.1.5 Post-Hoc Testing

The difference between Bonferroni's and Tukey's test

The Bonferroni procedure uses t-tests to perform pairwise comparisons between group means but controls overall error rate for each test by taking the significance (α) level and dividing it by the number of comparisons made ie. α becomes α/n. Hence, the observed significance level is adjusted for the multiple comparisons being made.

The Tukey's procedure uses the "Honest Significant Difference (HSD)", which is a value that represents the differences between groups ie. To compare every mean with every other mean. Overall error rate α is set for the collection of all pairwise comparisons.

In practice, the Bonferroni procedure tends to be more conservative and often the Tukey's HSD test is preferred however, in this unit, choosing either will be fine.

4.1.5 Post-Hoc Testing

A multiple comparisons table from a Tukey's post-hoc test

It is important to pay attention to the Mean Difference (I-J) and Sig. columns. This will give you details about where the differences in means are.

Multiple Comparisons

Dependent Variable: BodyWeightkg

Tukey HSD

(I) DietType	(J) DietType	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Diet A	Diet B	-8.68000*	1.85005	.001	-14.0121	-3.3479
	Diet C	-39.73000*	1.96227	.000	-45.3856	-34.0744
	Diet D	-25.62000*	1.85005	.000	-30.9521	-20.2879
Diet B	Diet A	8.68000*	1.85005	.001	3.3479	14.0121
	Diet C	-31.05000*	1.96227	.000	-36.7056	-25.3944
	Diet D	-16.94000*	1.85005	.000	-22.2721	-11.6079
Diet C	Diet A	39.73000*	1.96227	.000	34.0744	45.3856
	Diet B	31.05000*	1.96227	.000	25.3944	36.7056
	Diet D	14.11000*	1.96227	.000	8.4544	19.7656
Diet D	Diet A	25.62000*	1.85005	.000	20.2879	30.9521
	Diet B	16.94000*	1.85005	.000	11.6079	22.2721
	Diet C	-14.11000*	1.96227	.000	-19.7656	-8.4544

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

4.1.5 Post-Hoc Testing

Homogenous subset table from Tukey's test

As each diet has its own subset (1, 2, 3 and 4), this shows that each Diet type is significantly different from each other.

If Diet type mean values were found in the same column then there is no statistical difference between those particular diet types eg. If the value 69.30 was in column 1 under 60.62 then there would be no significant difference in their means according to SPSS.

BodyWeightkg

Tukey HSD[a,b]

		Subset for alpha = 0.05			
DietType	N	1	2	3	4
Diet A	5	60.6200			
Diet B	5		69.3000		
Diet D	5			86.2400	
Diet C	4				100.3500
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 4.706.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

4.1.6 Calculating Effect Size for one-way ANOVA

The Effect Size is an important parameter to calculate because it gives us an indication of the magnitude and importance of the study. It tells us the proportion of variability in the data that can be attributed to the independent variable (type of diet).

Eta-squared (η^2) is used to denote the effect size for a one-way ANOVA and is calculated using this formula:

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}}$$

Using the ANOVA table, you can get the values above:

$$\eta^2 = \frac{4226.348}{4354.698} = 0.971$$

ANOVA

BodyWeightkg

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4226.348	3	1408.783	164.642	.000
Within Groups	128.350	15	8.557		
Total	4354.698	18			

This tells us that we can attribute 97.1% of the variability in body weight to the type of diet they are on.

Cohen's conventions state that $\eta^2 = 0.01$ is small, $\eta^2 = 0.059$ is medium and $\eta^2 = 0.138$ is large.

4.1.7 Detailing the results of the one-way ANOVA

Complete statistical summary:

To investigate the impact of four different diets on the body weights of 19 animals, a one-way ANOVA was used.

Kolmogorov-Smirnov and Shapiro-Wilk tests were used to confirm normality of each diet group. A Levene's test showed a non-significant result, F (3, 15) = .03, p = .99, and thus the assumption of homogeneity of variance was not violated.

The ANOVA was statistically significant, indicating that there was a significant difference in the mean body weights of all four diet groups, $F(3, 15) = 164.64$, $p < .001$, $\eta^2 = .97$, which indicates a large effect.

Post hoc analyses with Tukey's HSD evidenced that Diet C (M = 100.35, SD = 2.77) had a significantly higher mean body weight than Diet A (M = 60.62, SD = 3.06), Diet B (M = 69.30, SD = 2.93) and Diet D (M = 86.24, SD = 2.90). All diet types were significantly different from each other.

Adapted from (Allen et al., 2019)