

# Worksheet 5 - Analysis of variance (ANOVA)

## Single factor ANOVA

- Quinn & Keough (2002) - Chpt 8

## Question 1 - ANOVA and Tukey's test

Here is a modified example from Quinn and Keough (2002). Day and Quinn (1989) described an experiment that examined how rock surface type affected the recruitment of barnacles to a rocky shore. The experiment had a single factor, surface type, with 4 treatments or levels: algal species 1 (ALG1), algal species 2 (ALG2), naturally bare surfaces (NB) and artificially scraped bare surfaces (S). There were 5 replicate plots for each surface type and the response (dependent) variable was the number of newly recruited barnacles on each plot after 4 weeks.

### Format of day.csv data files

TREAT	BARNACLE
ALG1	27
..	..
ALG2	24
..	..
NB	9
..	..
S	12
..	..

**TREAT** Categorical listing of surface types. ALG1 = algal species 1, ALG2 = algal species 2, NB = naturally bare surface, S = scraped bare surface.

**BARNACLE** The number of newly recruited barnacles on each plot after 4 weeks.



[Open](#) the day data file.

Note that as with independent t-tests, variables are in columns with levels of the categorical variable listed repeatedly. Day and Quinn (1989) were interested in whether substrate type influenced barnacle recruitment. This is a biological question. To address this question statistically, it is first necessary to re-express the question from a statistical perspective.

**Q1-1.** From a classical hypothesis testing point of view, what is the statistical question they are investigating? That is, what is their statistical  $H_0$ ?

**Q1-2.** The appropriate statistical test for comparing the means of more than two groups, is an ANOVA. In the table below, list the assumptions of **ANOVA** along with how violations of each assumption are diagnosed and/or the risks of violations are minimized.

Assumption	Diagnostic/Risk Minimization
I.	
II.	
III.	

Using **boxplots** (HINT), examine the assumptions of normality and homogeneity of variance. Note that when sample sizes are small (as is the case with this data set), these ANOVA assumptions cannot reliably be checked using boxplots since boxplots require at least 5 replicates (and preferably more), from which to calculate the median and quartiles. As with regression analysis, it is the assumption of homogeneity of variances (and in particular, whether there is a relationship between the mean and variance) that is of most concern for ANOVA.

**Q1-3.** Check the assumption of homogeneity of variances by **plotting the sample (group) means against sample variances**. A strong relationship (positive or negative) between mean and variance suggests that the assumption of homogeneity of variances is likely to be violated.

a. Any evidence of non-homogeneity? (Y or N)

**Q1-4.** Test the null hypothesis that the population group means are equal by fitting a linear model (ANOVA) **single factor ANOVA** HINT. As with regression analysis, it is also a good habit to **examine the resulting diagnostics** (note that Leverage and thus Cook's D have no useful meaning for categorical X variables) and residual plot. If there are no obvious problems, then the analysis is likely to be reliable. **Examine the ANOVA table** HINT.

**Q1-5.** **Identify** the important items from the ANOVA output and fill out the following ANOVA table

Source of Variation	SS	df	MS	F-ratio
Between groups	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Residual (within groups)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

**Q1-6.** What is the probability that the observed samples (and the degree of differences in barnacle recruitment between them) or ones more extreme, could be collected from populations in which there are no differences in barnacle recruitment. That is, what is the probability of having the above F-ratio or one more extreme when the null hypothesis is true?

**Q1-7.** What statistical conclusion would you draw?

**Q1-8.** Write the results out as though you were writing a research paper/thesis. For example (select the phrase that applies and fill in gaps with your results):

The mean number of barnacles recruiting was (choose the correct option)

(choose correct option  (F = , df = , , P = )

different from the mean metabolic rate of female fulmars. To see how these results could be incorporated into a report, **copy and paste the ANOVA table from the R Console into Word and add an appropriate table caption** .

**Q1-9.**Such a result would normally be accompanied by a graph to illustrate the mean (and variability or precision thereof) barnacle recruitment on each substrate type. Construct such a **bar graph** showing the mean barnacle recruitment on each substrate type and an indication of the precision of the means with error bars. To see how these results could be incorporated into a report, **copy and paste the graph (as a metafile) into Word**.

**Q1-10.** Although we have now established that there is a statistical difference between the group means, we do not yet know which group(s) are different from which other(s). For this data a **Tukey's multiple comparison test** (to determine which surface type groups are different from each other, in terms of number of barnacles) is appropriate. **Complete the following table for Tukey's pairwise comparison of group means:** (HINT) include differences between group means and Tukey's adjusted P-values (in brackets) for each pairwise comparison.

	ALG1	ALG2	NB	S
ALG1	0.000 (1.00)			
ALG2	6.000 (0.165)	0.000 (1.000)		
NB	<input type="text"/> ( <input type="text"/> )	<input type="text"/> ( <input type="text"/> )	0.000 (1.000)	
S	<input type="text"/> ( <input type="text"/> )	<input type="text"/> ( <input type="text"/> )	<input type="text"/> ( <input type="text"/> )	0.000 (1.000)

**Q1-11.** What are your conclusions from the Tukey's tests?



**Q1-12.**A way of representing/summarizing the results of multiple comparison tests is to **incorporate symbols** into the bargraph such that similarities and differences in the symbols associated with bars reflect statistical outcomes. Modify the bargraph (HINT) so as to incorporate the Tukey's test results. Finally, generate an appropriate caption to accompany this figure.

## Question 2 - ANOVA and Tukey's test

Here is a modified example from Quinn and Keough (2002). Medley & Clements (1998) studied the response of diatom communities to heavy metals, especially zinc, in streams in the Rocky Mountain region of Colorado, U.S.A.. As part of their study, they sampled a number of stations (between four and seven) on six streams known to be polluted by heavy metals. At each station, they recorded a range of physiochemical variables (pH, dissolved oxygen etc.), zinc concentration, and variables describing the diatom community (species richness, species diversity H and proportion of diatom cells that were the early-successional species, *Achanthes minutissima*). One of their analyses was to ignore streams and partition the 34 stations into four zinc-level categories: background ( $< 20 \mu\text{g.l}^{-1}$ , 8 stations), low ( $21-50 \mu\text{g.l}^{-1}$ , 8 stations), medium ( $51-200 \mu\text{g.l}^{-1}$ , 9 stations), and high ( $> 200 \mu\text{g.l}^{-1}$ , 9 stations) and test null hypotheses that there we no differences in diatom species diversity between zinc-level groups, using stations as replicates. We will also use these data to test the null hypotheses that there are no differences in diatom species diversity between streams, again using stations as replicates.

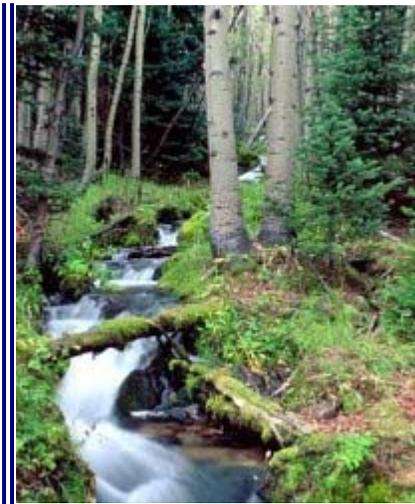
Format of medley.csv data files		
STATION	ZINC	DIVERSITY
ER1	BACK	2.27

...	...	...
ER2	HIGH	1.25
...	...	...
EF1	LOW	1.4
...	...	...
ER4	MEDIUM	1.62
...	...	...

**STATION** Uniquely identifies the sampling station from which the data were collected.

**ZINC** Zinc level concentration categories.

**DIVERSITY** Shannon-Weiner species diversity of diatoms



**Open** the medley data file.

Most statistical packages automatically order the levels of categorical variables alphabetically. Therefore, the levels of the ZINC categorical variable will automatically be ordered as (BACK, HIGH, LOW, MEDIUM). For some data sets the ordering of factors is not important. However, in the medley data set, it would make more sense if the factors were in the following order (BACK, LOW, MEDIUM, HIGH) as this would more correctly represent the relationships between the levels. Note that the ordering of a factor has no bearing on any analyses, it just makes the arrangement of data summaries within some graphs and tables more logical. It is therefore recommended that whenever a data set includes categorical variables, **reorder the levels of these variables** into a logical order. HINT

**Q2-1.** Check the ANOVA assumptions using a **factorial boxplot** HINT. Any evidence of skewness? Outliers? Does the spread of data look homogeneous between the different Zinc levels?



If the assumptions seem reasonable, **fit the linear model** (HINT), check the **residuals** (HINT) and if still there is no clear indication of problems, **examine the ANOVA table** (HINT).

**Q2-2.** Write the results out as though you were writing a research paper/thesis. For example (select the phrase that applies and fill in gaps with your results):

The mean diatom diversity was (choose the correct option)

(choose correct option  (F = , df = , , P = )

different between the four zinc level groups.

This can be abbreviated to  $F_{dfGroups, dfResidual} = \text{fratio}, P = \text{pvalue}$ . To see how the full anova table might be included in a report/thesis, **Copy and paste the ANOVA table into Word and add an appropriate table caption**.

**Q2-3.** Such a result would normally be accompanied by a graph to illustrate the means as well as an indication of the precision of the means. **Copy and paste the graph (as a metafile) into Word**.

**Q2-4.** Now determine which zinc level groups were different from each other, in terms of diatom species diversity, using a **Tukey's multiple comparison test** (HINT). **Incorporate symbols** these findings onto the graph.

## Question 3 - ANOVA and planned comparisons

Here is a modified example from Quinn and Keough (2002). Partridge and Farquhar (1981) set up an experiment to

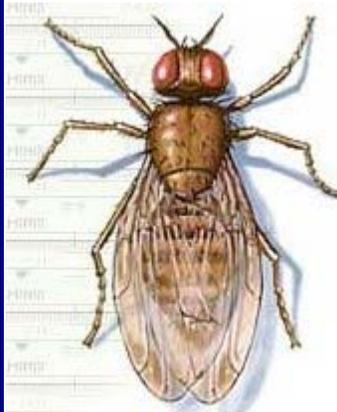
examine the effect of reproductive activity on longevity (response variable) of male fruitflies (*Drosophila* sp.). A total of 125 male fruitflies were individually caged and randomly assigned to one of five treatment groups. Two of the groups were used to investigate the effects of the number of partners (potential matings) on male longevity, and thus differed in the number of female partners in the cages (8 vs 1). There were two corresponding control groups containing eight and one newly pregnant female partners (with which the male flies cannot mate), which served as controls for any potential effects of competition between males and females for food or space on male longevity. The final group had no partners, and was used as an overall control to determine the longevity of un-partnered male fruitflies.

## Format of partridge.csv data files

GROUP	LONGEVITY
PREG8	35
..	..
NON0	40
..	..
PREG1	46
..	..
VIRGIN1	21
..	..
VIRGIN8	16
..	..

**GROUP** Categorical listing of female partner type.  
 PREG1 = 1 pregnant partner, NONE0 = no female partners, PREG8 = 8 pregnant partners, VIRGIN1 = 1 virgin partner, VIRGIN8 = 8 virgin partners.  
 Groups 1,2,3 - Control groups  
 Groups 4,5 - Experimental groups.

**LONGEVITY** Longevity of male fruitflies (days)



[Open](#) the partridge data file.

**Q3-1.** When comparing the mean male longevity of each group, what is the null hypothesis?


Note, normally we might like to adjust the ordering of the levels of the categorical variable (GROUP), however, in this case, the alphabetical ordering also results in the most logical ordering of levels.

**Q3-2.** Before performing the ANOVA, check the assumptions (Boxplots, scatterplot of Mean vs Variance) using the variable GROUP as the grouping (IV) variable for the X-axes. Is there any evidence that the ANOVA assumptions have been violated (Y or N)?

In addition to the global ANOVA in which the overall effect of the factor on male fruit fly longevity is examined, a number of other comparisons can be performed to identify differences between specific groups. As with the previous question, we could perform Tukey's post-hoc pairwise comparisons to examine the differences between each group. Technically, it is only statistically legal to perform  $n-1$  pairwise comparisons, where  $n$  is the number of groups. This is because if each individual comparison excepts a 5% ( $\alpha=0.05$ ) probability of falsely rejecting the  $H_0$ , then the greater the number of tests performed the greater the risk eventually making a statistical error. Post-hoc tests protect against such an outcome by adjusting the  $\alpha$  values for each individual comparison down. Consequently, the power of each comparison is greatly reduced.

This particular study was designed with particular comparisons in mind, while other pairwise comparisons would have very little biological meaning or importance. For example, in the context of the project aims, comparing group 1 with group 2 would not yield anything meaningful. As we have five groups ( $df=4$ ), we can do four **planned comparisons**.

**Q3-3.** In addition to the global ANOVA, we will address two specific questions by planned comparisons.

- a. "Is longevity affected by the presence of a large number of potential mates (8 virgin females compared to 1 virgin females)?" (contrast coefficients: 0, 0, 0, -1, 1)
- b. "Is longevity affected by the presence of any number of potential mates compared with either no partners or pregnant partners?" (contrast coefficients: -2, -2, -2, 3, 3)

**Q3-4.** Before we fit the linear model (perform the ANOVA), we need to define the contrast coefficients (and thus comparisons) that we wish to perform in addition to the global ANOVA. **Define the contrasts for the GROUP variable.**

**Q3-5.** If there is no evidence that the assumptions have been violated and the contrasts were successfully defined, **run the linear model, check residuals** and **examine the ANOVA table.**

**Q3-6.** Present the results of the planned comparisons as part of the following ANOVA table:

Source of Variation	SS	df	MS	F-ratio	Pvalue
Between groups	<input type="checkbox"/>				
8 virgin vs 1 virgin	<input type="checkbox"/>				
Partners vs Controls	<input type="checkbox"/>				
Residual (within groups)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		

Note that the Residual (within groups) term is common to each planned comparison as well as the original global ANOVA. **Copy and paste the ANOVA table from the R Console into Word and add an appropriate table caption**

**Q3-5.** Summarize the conclusions (statistical and biological) from the analyses.



**Q3-7.** Summarize the conclusions (statistical and biological) from the analyses.

- a. Global null hypothesis ( $H_0$ : population group means all equal)
- b. Planned comparison 1 ( $H_0$ : population mean of 8PREG group is equal to that of 1PREG)
- c. Planned comparison 2 ( $H_0$ : population mean of average of 1PREG and 8PREG groups are



equal to the population mean of average of CONTROL, 1VIRGIN and 8VIRGIN groups)

**Q3-8.** List any other specific comparisons that may have been of interest to this study. Remember that the total number of comparisons should not exceed the global degrees of freedom (4 in this case) and each outcome of each comparison should be independent of all other comparisons.

**Q3-9.** Attempt to redefine the set of contrasts including any additional ones you thought of from Q3-8 above. Assess whether or not orthogonality is still met. If it is, refit the linear model.

**Q3-10.** Finally, construct an appropriate **graph** to accompany the above analyses. **Save the graph as a jpeg image** and **import the graph into word**.

## Question 4 - ANOVA and planned comparisons

Snodgrass et al. (2000) were interested in how the assemblages of larval amphibians varied between depression wetlands in South Carolina, USA, with different hydrological regimes. A secondary question was whether the presence of fish, which only occurred in wetlands with long hydroperiods, also affected the assemblages of larval amphibians. They sampled 22 wetlands in 1997 (they originally had 25 but three dried out early in the study) and recorded the species richness and total abundance of larval amphibians as well as the abundance of individual taxa. Wetlands were also classified into three hydroperiods: short (6 wetlands), medium (5) and long (11) - the latter being split into those with fish (5) and those without (6). The short and medium hydroperiod wetlands did not have fish.

The overall question of interest is whether species richness differed between the four groups of wetlands. However, there are also specific questions related separately to hydroperiod and fish. Is there a difference in species richness between long hydroperiod wetlands with fish and those without? Is there a difference between the hydroperiods for wetlands without fish? We can address these questions with a single factor fixed effects ANOVA and planned contrasts using species richness of larval amphibians as the response variable and hydroperiod/fish category as the predictor (grouping variable).

### Format of snodgrass.csv data files

HYDROPERIOD	RICHNESS
Short	3
..	..
Medium	9
..	..
Longnofish	7
..	..
Longfish	12
..	..

**HYDROPERIOD** Categorical listing of the four hydroperiod/fish wetlands (short, medium and longnofish represent the hydroperiods of wetlands without fish; longfish represents wetlands with long hydroperiods that contain fish).



**RICHNESS** Species richness of larval amphibians

Open the snodgrass data file.

Reorder the factor levels of HYDROPERIOD into a more logical order (e.g. Short, Medium, Longnofish, Longfish)  
HINT

Q4-1. Examine the **group means and variances** and **boxplots** for species richness across the wetland categories HINT. Is there any evidence that any of the assumptions have been violated? ('Y' or 'N')

Q4-2. Now **fit a single factor ANOVA model** (HINT) and **examine the residuals** (HINT). Any evidence of skewness or unequal variances? Any outliers? Any evidence of violations? ('Y' or 'N')

Q4-3. As well as the overall analysis, Snodgrass et al. (2000) were particularly interested in two specific comparisons a) whether there was a difference in species richness between the long hydroperiod wetlands with and without fish, and b) whether there was a difference in species richness between permanent wetlands (long hydroperiods) and temporary wetlands (short and medium hydroperiods). What specific null hypotheses are being tested;

a.

---


b.

---


Q4-4. Define the appropriate contrast coefficients (and thus comparisons) (HINT) and check orthogonality (HINT) and define the contrast labels (HINT).

Q4-5. Perform the ANOVA with specific comparisons using the appropriate contrasts (HINT). Fill in the following table (HINT):

Source of Variation	SS	df	MS	F-ratio	Pvalue
Between groups	<input type="text"/>				
Long with vs nofish	<input type="text"/>				
Permanent vs Temporary	<input type="text"/>				
Residual (within groups)	<input type="text"/>	<input type="text"/>	<input type="text"/>		

Q4-6. What statistical conclusions would you draw from the overall ANOVA and the two specific contrasts and what do they mean biologically?


**Q4-7.** Finally, construct an appropriate **graph** to accompany the above analyses. **Save the graph as a jpeg image** and **import the graph into word**.

## Question 5 - ANOVA and power analysis

Laughing kookaburras (*Dacelo novaguineae*) live in cooperatively breeding groups of between two and eight individuals in which a dominant pair is assisted by their offspring from previous seasons. Their helpers are involved in all aspects of clutch, nestling and fledgling care. An ornithologist was interested in investigating whether there was an effect of group size on fledgling success. Previous research on semi captive pairs (without helpers) yielded a mean fledgling success rate of 1.15 (fledged young per year) with a standard deviation of 1.01. The ornithologist is planning to measure the success of breeding groups containing either two, four, six or eight individuals.

### Cooperative breeding in kookaburras



**Q5-1.** When designing an experiment or sampling regime around an ANOVA, it is necessary to consider the nature in which groups could differ from each other and the overall mean. Typically, following on from a significant ANOVA, the patterns are further explored via either multiple pairwise comparisons, or if possible, more preferably via planned contrasts (see Questions 1-4 above). The ability (power) to detect an effect of a factor depends on sample size as well as levels of within an between group variation. While ANOVA assumes that all groups are equally varied, levels of between group variability are dependent on the nature of the trends that we wish to detect. Do we wish to detect the situation where just one group differs from all the others, or two of the groups differ from others, or groups differ according to some polynomial trend, or some other combination?

For the purpose of example, we are going to consider the design implications of a number of potential scenarios. In each case we will try to accommodate an effect size of 50% (we wish to be able to detect a change of at least 50% in response variable - mean fledgling success rate), a power of 0.8 and a significance criteria of 0.05. Calculate the expected sample sizes necessary to detect the following:

- The mean fledgling success rate of pairs in groups of two individuals is 50% less than the mean fledgling success rate of individuals in groups with larger numbers of individuals - that is,  $\mu_{TWO} \neq \mu_{FOUR} = \mu_{SIX} = \mu_{EIGHT}$ . **HINT**
- The mean fledgling success rate of pairs in small groups (two and four) are equal and 50% less than the mean fledgling success rate of individuals in larger groups (six and eight) -

that is,  $\mu_{\text{TWO}} = \mu_{\text{FOUR}} \neq \mu_{\text{SIX}} = \mu_{\text{EIGHT}}$ . **HINT**



- c. The mean fledgling success rate increases linearly by 50% with increasing group size.

**HINT**



Note that we would not normally be contemplating accommodating both polynomial as well as non-polynomial contrasts or pairwise contrasts. We did so in Question 5-1 above purely for illustrative purposes!

**Q5-2.** Often when designing an experiment of survey it is useful to be able to estimate the relationship between power and sample size for a range of sample sizes and a range of outcomes so as to make more informed choices. This can be done by **plotting a power envelope**. Plot such an power envelope.

## Welcome to the end of Worksheet 5!