

---

---

# MODULE 3

---

## TWO-WAY ANOVA

**Module Objectives:**

1. Identify the advantages of a two-factor design compared to separate one-factor designs.
2. Understand the hypotheses that can be tested with a two-way ANOVA.
3. Understand how the full and simple models are incorporated into the two-way ANOVA.
4. Understand how the row-factor, column-factor, and interaction df and SS are computed.
5. Understand how a significant row-factor, column-factor, and interaction effect is determined.
6. Understand how to read an ANOVA table and identify relationships within the table (e.g., partitioning).
7. Understand why main effects cannot be interpreted if there is an interaction effect present.
8. Present the results of a two-way ANOVA in an efficient, comprehensive, readable format.

**I**N CONTRAST TO A ONE-WAY ANOVA (see Module 2), a two-way ANOVA allows the simultaneous determination of whether the mean of a quantitative variable differs according to the levels of two factor variables or an interaction between the two. For example, consider the following situations where the effect of two factor explanatory variables on a quantitative response variable is examined:

- Determining the effect of UV-B light intensity and tadpole density on mean growth of plains leopard frog (*Rana blairi*) tadpoles (Smith *et al.* 2000).
- Whether mean monoterpene levels in Douglas firs (*Pseudotsuga menziesii*) differ depending on exposure to ambient or elevated levels of CO<sub>2</sub> and ambient or elevated levels of temperature (Snow *et al.* 2003).
- Whether mean microcystin-LR concentration (a common cyanotoxin) differed by time (days) and exposure type (control (no exposure), direct exposure, or indirect exposure) to duckweed (*Lemna gibba*) (LeBlanc *et al.* 2005).
- Determining if the mean Na-K-ATPase activity level differs among locations of the kidney and between normal and hypertensive rats (Garg *et al.* 1985).

The theory and application of two-way ANOVAs are discussed in this module. The presentation here depends heavily on the foundational material discussed in Modules 1 and 2.

### 3.1 Manipulating Two Factors

It is both possible and advantageous to manipulate more than one factor at a time to determine the effect of those factors on the response variable. For example, an experimenter may manipulate both the density of tadpoles and the level of exposure to UV-B light to determine the effect of both of these explanatory variables on the growth of tadpoles. The design of such an experiment and why it is beneficial to simultaneously manipulate two factors, as compared to varying each of those factors alone in two separate experiments, is examined in this section.

#### Design

The key principle to designing two-factor experiments is that each level of each factor must be combined with each level of the other factor to form treatments. For example, suppose that interest is in the impact of two different UV-B light intensity levels (simply called “High” and “Low”) and the density of tadpoles (1, 2, and 4 individuals per tank) on the growth of tadpoles. In this example, the levels of density must be completely “crossed” with the levels of UV-B light intensity to form six treatments for the tadpoles. The specific treatments would be “High-1”, “High-2”, “High-4”, “Low-1”, “Low-2”, and “Low-4” (Table 3.1).

Table 3.1. Schematic table depicting the crossing of two factors - Density of tadpoles and UV-B Light Intensity - resulting in a total of six treatments. The numbers in each cell represent individuals exposed to that treatment.

UVB Light	Density		
	1	2	4
High	68,49,65,5,58,	71,8,10,80,78,	26,46,6,62,4,
	34,18,30,27,25,	74,38,36,50,19,	72,39,67,45,57,
	1,7,13,31,51	55,73,16,29,83	37,24,21,84,86
Low	40,90,89,22,79,	43,54,82,20,12,	17,60,47,70,11,
	53,32,77,56,3,	85,61,76,15,69,	59,75,81,64,88,
	42,87,63,48,9	44,33,35,41,28	23,66,2,52,14

Before discussing the advantages of a two-factor experiment as compared to single-factor experiments, let's examine how the two one-factor-at-a-time (OFAT) experiments would be constructed. First, the 90 individuals available for the experiment would be split into two groups. One group would be used to determine the effect of density on tadpole growth and the other group would be used to determine the effect of UV-B light intensity on tadpole growth. The individuals within each group would then be randomly allocated to the different levels of each of the factors (Table 3.2).

Table 3.2. Schematic table of two OFAT experiments using UV-B light intensity and density of tadpoles. The numbers in each cell correspond to the random allocation of individuals to each treatment.

UV-B Light		Density		
High	Low	1	2	4
9,24,38,34,20,14,18, 37,15,13,41,39,8,17, 21,43,36,15,16,29,3	26,6,32,27,19,10,28, 40,33,22,2,11,42,4,23, 30,7,12,25,44,35,31	52,54,75,84,53, 66,80,79,71,47, 88,83,61,87,49	65,86,76,77,46, 63,55,59,60,72, 70,81,56,68,89	58,85,90,69,74, 51,67,62,50,78, 48,73,57,82,64

## Advantages

There are two major advantages of a two-factor experiment as compared to two separate OFAT experiments. First, the two-factor experiment allocates the individuals more efficiently than the two OFAT experiments. For example, in the two-factor design of the tadpole experiment, all 90 individuals are used to identify differences among the light intensities **AND** all 90 individuals are used to identify differences among the densities of tadpoles. However, in the OFAT experiments, only 44 individuals<sup>1</sup> are used to identify differences among the light intensities and only 45 individuals are used to identify differences among the tadpole densities. This efficiency in allocation of individuals effectively leads to an increased sample size for determining the effect of any one factor. An increased sample size means more statistical power<sup>2</sup> and, relatedly, the ability to correctly identify smaller effect sizes.<sup>3</sup> Alternatively, the increased efficiency means that the same power and detectable effect size can be obtained with a smaller sample size and, thus, lower costs.

◊ **One advantage of two-factor experiments is that individuals are allocated to treatments more efficiently. The more efficient use of individuals results in increased power and increased ability to detect small effect sizes, or lower costs to obtain the same power and effect size detectability.**

Two-factor experiments are also advantageous in that they allow the researcher to identify if an interaction exists between the two factors. An interaction among factors is said to exist if the effect of one factor on the response variable depends on the level of the other factor. For example, if the growth rate of the tadpoles increases from the 1 to the 2 to the 4 tadpole densities in the low light intensity treatments but decreases from the 1 to the 2 to the 4 tadpole densities in the high light intensity treatments (Figure 3.1-Left), then an interaction between the density of tadpoles and UV-B light level intensities is said to exist. However, if the pattern among tadpole densities is the same in both the low and high density levels (Figure 3.1-Right), then no interaction exists between the factors.

<sup>1</sup>One individual was not used in order to have equal numbers of individuals in each treatment.

<sup>2</sup>Recall that statistical power is the probability of correctly rejecting a false null hypothesis or, alternatively, correctly identifying a difference in means when one really exists.

<sup>3</sup>In this context an effect size is the difference in two means. Thus, the ability to detect a smaller effect size means that there is a higher chance of identifying a difference among means that are relatively close together.

△ **Interaction Effect:** when the effect of one factor on the response variable is significantly different among the levels of the other factor.

◇ **Another advantage of two-factor experiments is that an interaction between factors can be detected.**

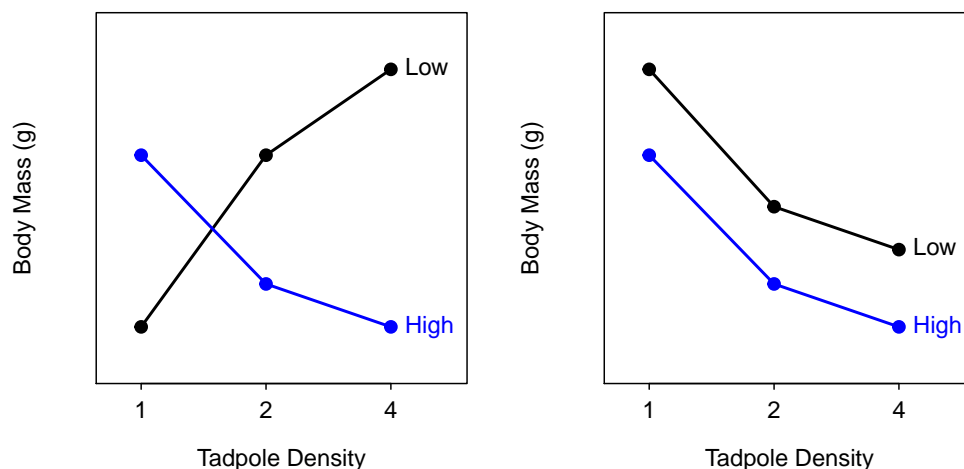


Figure 3.1. Interaction plot (mean growth rate for each treatment connected within the UV-B light intensity levels) for the tadpole experiment illustrating a hypothetical interaction effect (Left) and a lack of an interaction effect (Right).

## 3.2 Analytical Foundation

### 3.2.1 Terminology

Some terminology must be developed before discussing the objective criteria for determining the significance of effects. The tadpole growth experiment (Table 3.1) was an experiment with two factors: UV-B light intensity with two levels (High and Low) and the density of tadpoles with three levels (1, 2, and 4 tadpoles). The combination of these two factors created six treatments with 15 individuals or replicates per treatment.<sup>4</sup>

With the schematic arrangement of this experiment shown in Table 3.1, one of the factors can be generically labeled as the row factor (i.e., UV-B light intensity in this example) and the other factor can be generically labeled as the column factor (i.e., tadpole density). With this,  $r$  is defined as the number of levels of the row factor (i.e.,  $r = 2$  in this example) and  $i$  is an index for the rows (i.e.,  $i = 1, \dots, r$ ). Similarly,  $c$  is defined as the number of levels of the column factor (i.e.,  $c = 3$ ) and  $j$  is an index for the columns (i.e.,  $j = 1, \dots, c$ ). As was the case in this example, and for the sake of simplicity, let's only consider the situation where the same number of replicates are used per treatment.<sup>5</sup> With this consideration,  $n$  is defined as the number of replicates per treatment and  $k$  is an index for the individuals within a treatment (i.e.,  $k = 1, \dots, n$ ).

<sup>4</sup>These terms are not defined specifically here as it is assumed that you were introduced to basic experimental design terms in your introductory statistics course.

<sup>5</sup>An experiment where each treatment has the same number of replicates is called a balanced design.

With these definitions, the response variable recorded on the  $k$ th individual in the treatment defined by the  $i$ th level of the row factor and the  $j$ th level of the column factor is given by  $Y_{ijk}$ . Similarly, the mean response for the  $i$ th level of the row factor and the  $j$ th level of the column factor is given by  $\bar{Y}_{ij.}$ , the mean response for the  $i$ th level of the row factor is given by  $\bar{Y}_{i..}$ , the mean response for the  $j$ th level of the column factor is given by  $\bar{Y}_{.j.}$ , and the mean response regardless of level of any factor is given by  $\bar{Y}_{...}$ . The  $\bar{Y}_{ij.}$  are called *treatment means* or *group means*, the  $\bar{Y}_{i..}$  and  $\bar{Y}_{.j.}$  are called *level means*, and  $\bar{Y}_{...}$  is called the *grand mean*.

**Δ Treatment Mean:** The mean for a particular treatment,  $\bar{Y}_{ij.}$

**Δ Level Mean:** The mean for a particular row level,  $\bar{Y}_{i..}$ , or column level,  $\bar{Y}_{.j.}$

**Δ Grand Mean:** The mean for all individuals,  $\bar{Y}_{...}$

### 3.2.2 Models & SS

As with the one-way ANOVA, the two-way ANOVA begins as an analysis of two models. As usual, the full model uses a separate mean for each treatment group and the simple model uses a single common mean for all treatment groups. In essence, the full model says that each treatment mean should be modeled by a separate mean (Table 3.3), whereas the simple model says that each treatment mean should be modeled by a common mean (Table 3.4). As in the one-way ANOVA, the goal is to determine which of these models best fits the data, while accounting for differences in complexity.

Table 3.3. The means used to describe each treatment under the full model.

UVB Light	Density		
	1	2	4
High	$\bar{Y}_{11.}$	$\bar{Y}_{12.}$	$\bar{Y}_{13.}$
Low	$\bar{Y}_{21.}$	$\bar{Y}_{22.}$	$\bar{Y}_{23.}$

Table 3.4. The means used to describe each treatment under the simple model.

UVB Light	Density		
	1	2	4
High	$\bar{Y}_{...}$	$\bar{Y}_{...}$	$\bar{Y}_{...}$
Low	$\bar{Y}_{...}$	$\bar{Y}_{...}$	$\bar{Y}_{...}$

The first step in this process is the calculation of sums-of-squares. The  $SS_{Total}$  eventually leads (when converted to  $MS_{Total}$ ) to a measure of the variability not explained by the simple model and, thus, is

computed by summing the squared residuals about the simple model, or<sup>6</sup>

$$SS_{Total} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (Y_{ijk} - \bar{Y} \dots)^2$$

The  $SS_{Within}$  eventually leads (as  $MS_{Within}$ ) to a measure of the variability not explained by the full model and, thus, is computed by summing the squared residuals about the full model, or

$$SS_{Within} = \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij.})^2$$

The  $SS_{among}$  is usually found by subtraction (i.e.,  $SS_{Total} - SS_{Within}$ ), but can be shown to be

$$SS_{Among} = n \sum_{i=1}^r \sum_{j=1}^c (\bar{Y}_{ij.} - \bar{Y} \dots)^2$$

Thus,  $SS_{among}$  is a measure of the variability that was not explained by the simple model that was explained by the full model, or the variability among the treatment means.

In a two-way ANOVA, the  $SS_{Among}$  could be “large” (indicating a significant difference among the treatment means) because of large differences in means among the levels of the row factor, the column factor, the interaction of the row and column factors, or sampling variability. Thus, just as  $SS_{Total}$  was partitioned into parts (i.e., into  $SS_{Among}$  and  $SS_{Within}$ ), the  $SS_{Among}$  can be partitioned into parts due to differences in the levels of the row factor, differences in the levels of the column factor, and differences in the interaction among the row and column factors. In other words,

$$SS_{Among} = SS_{Row} + SS_{Column} + SS_{Interaction} \quad (3.2.1)$$

or, mathematically,

$$n \sum_{i=1}^r \sum_{j=1}^c (\bar{Y}_{ij.} - \bar{Y} \dots)^2 = cn \sum_{i=1}^r (\bar{Y}_{i..} - \bar{Y} \dots)^2 + rn \sum_{j=1}^c (\bar{Y}_{.j.} - \bar{Y} \dots)^2 + n \sum_{i=1}^r \sum_{j=1}^c (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y} \dots)^2 \quad (3.2.2)$$

When these sum-of-squares are converted into mean-squares, they represent the amount of variability that is explained by the row, column, or interaction effect, respectively.

◇ A large  $SS_{among}$  indicates that at least one treatment mean differs from another treatment mean.

◇ A large  $SS_{among}$  can be caused by differences among row levels, column levels, or the interaction between row and column levels.

<sup>6</sup>It is easiest to read a series of sums in a formula such as this from right-to-left. In this formula for example, one can see that the residuals are summed across the individuals within each treatment first (i.e., the  $k$  index is summed first), then across all columns (i.e., the  $j$  index is summed second), and then across all rows. In other words, the residuals are summed within each cell of the table representing the experiment, then these sums are summed for each column, and then these column sums are summed across the rows to get one final sum.

### 3.2.3 Understanding the Main Effect SS

Differences among treatment means can be attributable to either of the two factor variables or the interaction between these two variables. If the differences are due to the interaction between the two factor variables then a so-called “interaction effect” on the response variable exists. However, *if an interaction effect does not exist* and the differences is due to one of the factor variables, then a “main effect” on the response variable exists. Thus, a main effect suggests that a factor variable has an effect on the response variable regardless of the value of the other factor variable.

The main effect sums-of-squares identified with Equations (3.2.1) and (3.2.2) can best be understood by comparing the formulas to a symbolic table of means (Table 3.5). For example,  $SS_{Row}$  is computed by summing the residuals of the row marginal means (i.e., right most column;  $\bar{Y}_{i..}$ ) from the grand mean (i.e.,  $\bar{Y}_{...}$ ) and then, to ensure that all individuals were accounted for in the calculation, multiplying by the number of individuals used to compute the row marginal means (i.e.,  $cn$ ). If the means in the marginal row column are very different, then  $SS_{Row}$  will become large. Because large differences in row marginal means indicate a main effect of the row factor, then large values of  $SS_{Row}$  will indicate a main effect of the row factor. A similar argument can be made for the  $SS_{Column}$  calculation.

Table 3.5. Means (treatment, marginal row, marginal column, and grand) for the tadpole experiment.

UVB Light	Density			
	1	2	4	
High	$\bar{Y}_{11.}$	$\bar{Y}_{12.}$	$\bar{Y}_{13.}$	$\bar{Y}_{1..}$
Low	$\bar{Y}_{21.}$	$\bar{Y}_{22.}$	$\bar{Y}_{23.}$	$\bar{Y}_{2..}$
	$\bar{Y}_{.1.}$	$\bar{Y}_{.2.}$	$\bar{Y}_{.3.}$	$\bar{Y}_{...}$

The mean table shown in Table 3.6 are the actual results for the tadpole experiment. Differences among the two marginal means for the UV-B light intensity (e.g., 0.858 and 1.045) are small and, thus, are indicative of only a slight UV-B light intensity main effect. The relatively larger differences among the three marginal means for the tadpole densities (1.270, 0.816, and 0.768) are indicative of a stronger density main effect. The marginal means can also be plotted (Figure 3.2) to see the main effects.<sup>7</sup>

Table 3.6. Sample means (treatment, marginal row, marginal column, and grand) for the tadpole experiment.

UVB Light	Density			
	1	2	4	
High	1.065	0.759	0.748	0.858
Low	1.476	0.872	0.786	1.045
	1.270	0.816	0.767	0.951

Main effects can only be interpreted if the interaction effect does not exist. This is because if an interaction effect exists, then the effect of one factor on the response variable depends on the level of the other factor. Thus, a “main effect” differs depending on the level of the other “main effect” factor. Therefore, the interpretations discussed in this section should only be made when it has already been determined that an interaction effect does not exist.

<sup>7</sup>In later sections, objective measures of the size of main and interaction effects are developed.

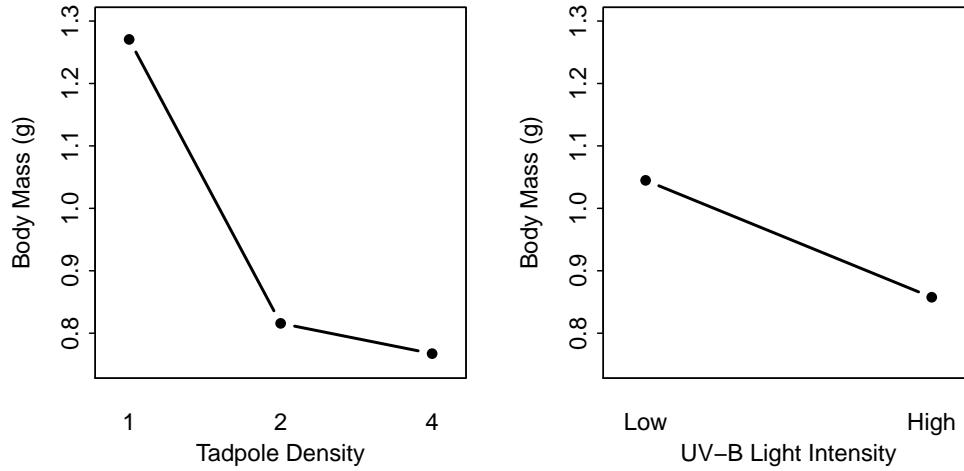


Figure 3.2. Main effects plots for the UV-B light intensity (Left) and tadpole densities (Right) factors.

◊ Main effects should only be interpreted if an interaction effect does NOT exist.

### 3.2.4 ANOVA Table

As is known from the one-way ANOVA, the  $SS$  are not sufficient for making conclusions and they must be converted to  $MS$  with the use of the  $df$ . The  $MS$  are then used to compute appropriate F test statistics and p-values. First, with the exception of  $df_{Interaction}$ ,  $df$  are computed rather intuitively. The total, among, and within  $df$  are the same as with the one way ANOVA, namely

- $df_{Total}$  = total number of individuals - 1 =  $rcn - 1$
- $df_{Within}$  = total number of individuals - number of treatment groups =  $rcn - rc = rc(n - 1)$
- $df_{Among}$  = total number of treatment groups - 1 =  $rc - 1$

As  $SS_{Among}$  was partitioned,  $df_{Among} = df_{Row} + df_{Column} + df_{Interaction}$ , where

- $df_{Row}$  = number of levels of the row factor - 1 =  $r - 1$
- $df_{Column}$  = number of levels of the column factor - 1 =  $c - 1$
- $df_{Interaction} = (r - 1)(c - 1)$

The  $MS$  are computed exactly as before, i.e.,  $MS = \frac{SS}{df}$ . Thus,  $MS_{Row} = \frac{SS_{Row}}{df_{Row}}$  and represents the variability in the response variable that **IS** explained by the different levels of the row factor. Similarly,  $MS_{Column}$  represents the variability in the response variable that **IS** explained by the different levels of the column factor.

Finally, the significance of a particular effect is defined by computing the ratio of the variability explained by that effect to the variability among individuals within each treatment group. In other words, an F test statistic is created by dividing the corresponding  $MS$  by  $MS_{Within}$  to determine the significance of a particular main or interaction effect. For example, the F test statistic for determining the significance of the row factor is  $\frac{MS_{Row}}{MS_{Within}}$ . This F test statistic will have  $df_{Row}$  numerator df and  $df_{Within}$  denominator df.

All results just discussed are succinctly summarized in an ANOVA table (Table 3.7).



Table 3.7. Symbolic ANOVA table for a generic two-way ANOVA.

Source	df	SS	MS	F	p-value
Row	$r - 1$	$SS_{Row}$	$\frac{SS_{Row}}{r-1}$	$\frac{MS_{Row}}{MS_{Within}}$	
Column	$c - 1$	$SS_{Column}$	$\frac{SS_{Column}}{c-1}$	$\frac{MS_{Column}}{MS_{Within}}$	
Interaction	$(r - 1)(c - 1)$	$SS_{Interaction}$	$\frac{SS_{Interaction}}{(r-1)(c-1)}$	$\frac{MS_{Interaction}}{MS_{Within}}$	
Within	$rc(n - 1)$	$SS_{Within}$	$\frac{SS_{Within}}{rc(n-1)}$		
Total	$rcn - 1$	$SS_{Total}$			

### 3.2.5 Interpreting the ANOVA Table

“Small” p-values in the ANOVA table are indicative of significant effects. For example, a small p-value in the “row” row would indicate a significant “row” factor main effect. Similarly, a small p-value in the “interaction” row would indicate a significant “interaction” between the “row” and “column” factors.

The main effects can NOT and should NOT be determined if there is a significant interaction effect. By definition, a significant interaction effect indicates that the effect of one factor depends on the level of the other factor. Thus, an ANOVA table should be read from the bottom to the top; i.e., first determine if there is a significant interaction effect. Example interpretations are illustrated in Section 3.4.

◊ Main effects should not be discussed if a significant interaction is present.

If there is a significant main effect in the absence of a significant interaction effect, then it is logical to want to identify exactly which levels of the factor are different. This can be accomplished with post-hoc multiple comparison tests as was done in Section 2.4 (e.g., Tukey’s HSD method).

### 3.2.6 Two-way ANOVA in R

#### Model & ANOVA

The only difference between fitting a two-way rather than a one-way ANOVA is that the formula entered as the first argument in `lm()` is of the form `response~factor1*factor2`. In other words, the entire model can be formed by simply identifying the interaction term because R will automatically include the main factors as terms in the model. The same model could also be declared with a formula of the form `response~factor1+factor2+factor1:factor2`.<sup>8</sup>

```
> tad <- read.csv("data/Tadpoles.csv")
> tad$uvb <- factor(tad$uvb, levels=c("Low", "High")) # re-order levels
> tad$density <- factor(tad$density) # force numeric values to be levels
> tad.lm1 <- lm(mass~uvb*density, data=tad)
```

The ANOVA table is extracted from this object with `anova()`. These results indicate that the interaction effect is insignificant ( $p = 0.3417$ ). Because the interaction term is insignificant, the main effects can be

<sup>8</sup>Note that the actual interaction term is indicated by separating the main effect names by a colon.

interpreted. The density main effect is strongly significant ( $p = 0.0004$ ) and the UV-B main effect is only weakly significant ( $p = 0.0912$ ).

```
> anova(tad.lm1)
Analysis of Variance Table

Response: mass
      Df Sum Sq Mean Sq F value    Pr(>F)
uvb     1  0.7889  0.78886    2.9204 0.0911577
density 2  4.6251  2.31253    8.5611 0.0004132
uvb:density 2  0.5876  0.29381    1.0877 0.3416820
Residuals 84 22.6900  0.27012
```

### Multiple Comparisons

The Tukey HSD procedure can be applied to either of the two main effects with `glht()`. As described for a one-way ANOVA, this function requires a saved `lm()` object as the first argument and a `mcp()` function as the second argument. The factor for which to make the comparisons should be used in `mcp()`. For example, the Tukey HSD results for the highly significant density main effect in the tadpole example are extracted below. These results indicate that mean mass at a density of 1 tadpole is significantly higher than the mean mass at densities of either 2 or 4 tadpoles. There is no difference in mean mass between the 2 and 4 tadpole densities. Furthermore, mean mass is between 0.152 and 1.057 g higher at a density of 1 then at a density of 2 tadpoles and between 0.238 and 1.143 g higher at a density of 1 then at a density of 4 tadpoles.

```
> tad.mc1 <- glht(tad.lm1,mcp(density="Tukey"))
> summary(tad.mc1)

      Estimate Std. Error    t value    p value
2 - 1 = 0 -0.60440000  0.1897785 -3.1847646 0.005653761
4 - 1 = 0 -0.69033333  0.1897785 -3.6375731 0.001382049
4 - 2 = 0 -0.08593333  0.1897785 -0.4528085 0.893299375

> confint(tad.mc1)

      Estimate      lwr      upr
2 - 1 -0.60440000 -1.0570910 -0.1517090
4 - 1 -0.69033333 -1.1430243 -0.2376423
4 - 2 -0.08593333 -0.5386243  0.3667577
```

Tukey's procedure should NOT be applied to the main effects if a significant interaction term has been found. In this example, multiple comparisons among the levels of each main factor are appropriate. However, if an interaction term was significant, then multiple comparisons must be made between each group defined by the combined levels of the two factors. This is most easily accomplished by creating a new variable that represents the levels of groups, fitting a one-way ANOVA using this new variable as the single factor, and then using the multiple comparison technique. It is not needed in this example, but for illustrative purposes, the Tukey's procedure is applied to the interaction term in the tadpole data as follows.

```
> tad$grps <- tad$uvb:tad$density      # create single-factor interaction variable
> tad.lm2 <- lm(mass~grps,data=tad)    # fit new one-way ANOVA model
> tad.mc2 <- glht(tad.lm2,mcp(grps="Tukey"))
> summary(tad.mc2)
```

	Estimate	Std. Error	t value	p value
Low:2 - Low:1 = 0	-0.60440000	0.1897785	-3.18476462	0.024132456
Low:4 - Low:1 = 0	-0.69033333	0.1897785	-3.63757309	0.006148247
High:1 - Low:1 = 0	-0.41166667	0.1897785	-2.16919496	0.263193140
High:2 - Low:1 = 0	-0.71686667	0.1897785	-3.77738517	0.003858549
High:4 - Low:1 = 0	-0.72793333	0.1897785	-3.83569875	0.003192655
Low:4 - Low:2 = 0	-0.08593333	0.1897785	-0.45280847	0.997528999
High:1 - Low:2 = 0	0.19273333	0.1897785	1.01556966	0.911498699
High:2 - Low:2 = 0	-0.11246667	0.1897785	-0.59262055	0.991275774
High:4 - Low:2 = 0	-0.12353333	0.1897785	-0.65093413	0.986613313
High:1 - Low:4 = 0	0.27866667	0.1897785	1.46837813	0.685176136
High:2 - Low:4 = 0	-0.02653333	0.1897785	-0.13981208	0.999992274
High:4 - Low:4 = 0	-0.03760000	0.1897785	-0.19812566	0.999956343
High:2 - High:1 = 0	-0.30520000	0.1897785	-1.60819021	0.595310416
High:4 - High:1 = 0	-0.31626667	0.1897785	-1.66650379	0.557590483
High:4 - High:2 = 0	-0.01106667	0.1897785	-0.05831358	0.999999902

### Interaction Plot

An interaction plot plots the means for each combination of the levels of the two factors (Figure 3.3). The means at the same level of one of the factor variables (i.e., the one NOT shown on the x-axis) are connected with lines. The interaction plot is an efficient method for visualizing the presence or absence of interactions in a two-way ANOVA. In the absence of an interaction effect, the interaction plot is still a good method for visualizing main effects.

An interaction plot is created with `fitPlot()`. This function requires a saved two-way ANOVA `lm()` object as its first argument. Which factor variable is plotted on the x-axis can be changed by including a `change.order=TRUE` argument. I prefer the factor with more levels to be the factor plotted on the x-axis. For large number of treatments, the confidence intervals shown in the graph may need to be turned off with the `interval=FALSE` argument to create a plot that is easier to interpret.

```
> fitPlot(tad.lm1,change.order=TRUE,xlab="Tadpole Density",ylab="Body Mass (g)")
```

Significance letters may be added to interaction plots with `addSigLetters()` as shown for a one-way ANOVA. When entering the significance letters, note that the means are plotted by the levels of the connected factor within the levels of the x-axis factor (in this example the means would be plotted as “Low-1”, “Low-2”, “Low-4”, “High-”, “High-2”, “High-4”). You will likely need to use trial-and-error to get the significance letters placed properly.

### Main Effect Plot

In situations where the interaction effect is not significant it may be more appropriate to construct a *main effects plot* of group means for each significant main factor (Figure 3.2). These plots are constructed by including the name of the main effect variable within quotes in `which=` in `fitPlot()`.

```
> fitPlot(tad.lm1,which="density",xlab="Tadpole Density",ylab="Body Mass (g)")
> fitPlot(tad.lm1,which="uvb",xlab="UV-B Light Intensity",ylab="Body Mass (g)")
```

Significance letters can also be added to these plots with `addSigLetters()`, also using `which=`.

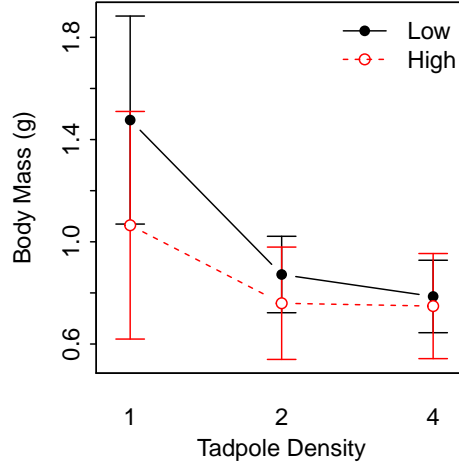


Figure 3.3. Interaction plot of treatment means from raw tadpole body mass data.

### 3.3 Assumptions

The assumptions for a two-way ANOVA are the same as for a one-way ANOVA. Thus, assumptions are checked in R in the same manner as described in Section 2.2 (i.e., `transChooser()`).

### 3.4 Example Analyses

#### 3.4.1 Tadpoles: UV-B Light Intensity & Density

In organisms with complex life cycles, each stage - embryo, larva, and adult - often has unique ecological requirements, and the performance of each stage can influence the performance of other stages. Changes in UV-B light intensity levels caused by changes in the ozone layer may affect some amphibian species. Little work has been done to examine the effect of UV-B levels on the different stages of amphibians.

Smith *et al.* (2000) designed an experiment to determine the effect of different intensities of UV-B light and different tadpole densities on the body mass (i.e., an indicator of growth) of plains leopard frogs (*Rana blairi*). Primarily, the scientists hypothesized that the body mass of the tadpoles would be lower at higher UV-B light intensities. However, they were also concerned that the effect of UV-B light on tadpole growth might be affected by the density of tadpoles (i.e., an interaction effect).<sup>9</sup> Secondly, they hypothesized that tadpole growth would be lower at higher densities. Thus, the statistical hypotheses to be examined were

$$H_0 : \text{"no UV-B light intensity effect"} : \mu_{Low} = \mu_{High}$$

$$H_A : \text{"UV-B light intensity effect"} : \mu_{Low} \neq \mu_{High}$$

$$H_0 : \text{"no tadpole density effect"} : \mu_1 = \mu_2 = \mu_4$$

$$H_A : \text{"tadpole density effect"} : \text{"At least one pair of means is different"}$$

$$H_0 : \text{"no interaction effect"}$$

$$H_A : \text{"interaction effect"}$$

where the subscripts identify the levels of each factor (specifically defined below).

<sup>9</sup>The scientists did not necessarily hypothesize that an interaction effect would exist but they wanted to make sure to detect one if it did because that would alter their conclusions about the effect of UV-B light intensity.

### Data Collection

A single clutch of *R. blairi* eggs were collected from a Liberty, MO pond. The eggs were divided into 36 groups of 20-30 eggs each. These groups of eggs were randomly allocated to either high or low UV-B treatments. The eggs were examined after eight days. Hatchlings from this portion of the experiment were then used in a further experiment to determine the effect of density on body mass. Thus, 15 “containers” were constructed that held 1, 2, or 4 hatchlings for each UV-B exposure. Thus, there were six treatments in this experiment from the two levels of UV-B exposure and three levels of density. Each tadpole was randomly selected for placement in one of the treatment “containers.” Experimental containers were filled with 450 ml of deionized water, the water was changed every three days, and the tadpoles were fed a constant ration of crushed algal flakes. After a total of 95 days, the tadpoles were removed from the experimental containers, blotted dry, and weighed to the nearest 0.001 g with an electronic balance.

### Exploratory Data Analysis and Assumption Checking

There appears to be within- and among-group independence as there would be no relationship between or effect of one container on any other container. Variances among the treatments appear to be non-constant (Levene’s  $p = 0.006$ ; Figure 3.4-Right), the residuals appear to be non-normal (Anderson-Darling  $p = 0.012$ ) though not strongly skewed (Figure 3.4-Left), and there appears to be at least one significant outlier (outlier test  $p = 0.017$ ; Figure 3.4-Right). The assumptions are not met and a transformation should be considered.

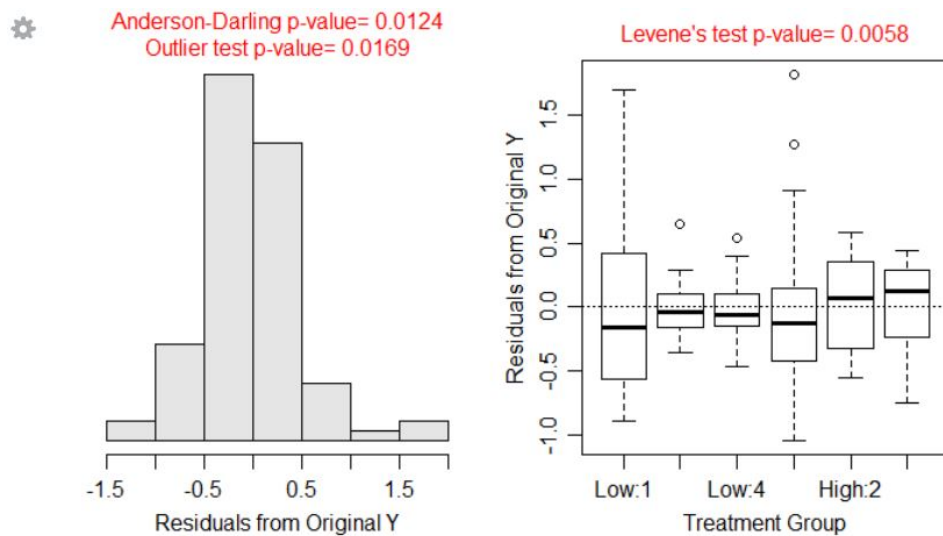


Figure 3.4. Residual plot (Left) and histogram of residuals (Right) from two-way ANOVA fit on raw tadpole data.

No transformation was found that simultaneously resulted in all assumptions being met. Therefore, the data will not be transformed. In addition, the removal of two potential outliers did not result in equal variances. Thus, the original raw data without any individuals removed will be analyzed but the reader is cautioned that the results may be misleading because of the assumption violations.

### Results

The complete raw data were examined with a two-way ANOVA. There was no evidence for an interaction effect ( $p = 0.342$ ; Table 3.8); thus, the main effects can be interpreted directly. There appears to be a significant difference in mean tadpole body mass among the different densities ( $p < 0.0005$ ; Table 3.8).

There is weak, but not significant, evidence of a difference between the two levels of UV-B light intensity ( $p = 0.091$ ; Table 3.8).

Table 3.8. Two-way ANOVA results for tadpole body mass.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
uvb	1	0.7889	0.78886	2.9204	0.0911577
density	2	4.6251	2.31253	8.5611	0.0004132
uvb:density	2	0.5876	0.29381	1.0877	0.3416820
Residuals	84	22.6900	0.27012		

Effect sizes for all pairs of levels are shown in Table 3.9. The mean body mass of tadpoles at a density of 1 tadpole was significantly higher than the mean body mass of tadpoles at densities of 2 and 4 tadpoles (Table 3.9; Figure 3.5). Mean tadpole body mass was not significantly different at densities of 2 and 4 tadpoles (Table 3.9; Figure 3.5).

Table 3.9. Tukey confidence intervals (95%) for differences in pairs of means between levels of UV-B light intensity and tadpole density.

	Estimate	lwr	upr
High - Low	-0.41166667	-0.7890621	-0.03427121
2 - 1	-0.60440000	-1.0571845	-0.15161550
4 - 1	-0.69033333	-1.1431178	-0.23754883
4 - 2	-0.08593333	-0.5387178	0.36685117

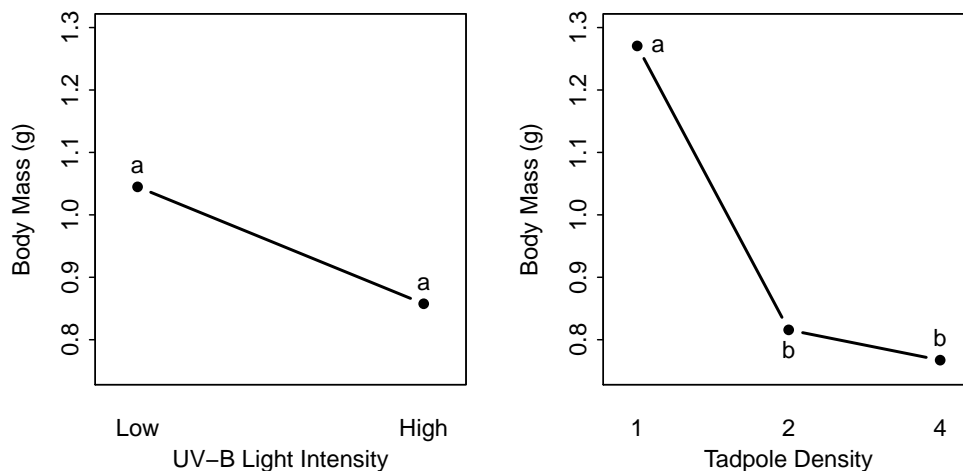


Figure 3.5. Main effects plots for the UV-B light intensity (Left) and tadpole densities (Right) factors. Group means with the same letters within a plot are not significantly different at the 5% level.

### Conclusion

There is very little support for the researcher's hypothesis that the body mass of tadpoles (i.e., growth rate) would be lower if the embryos were exposed to higher UV-B light intensities. While the tadpole body mass declined significantly when density increased from 1 to 2 tadpoles, but not from 2 to 4 tadpoles, there was no indication that the effect of UV-B light intensity differed depending on the density of tadpoles. In other words, the mean body mass of tadpoles was slightly lower at high UV-B light intensities for all densities examined.

It can be concluded that the different UV-B light intensities and tadpole densities caused the differences in tadpole body mass because the individual containers of tadpoles were randomly allocated to treatment groups (or, alternatively, the tadpoles were randomly allocated to containers) and all other variables were controlled. However, the inferences cannot be extended to a general population of containers of tadpoles because the 90 containers of tadpoles used in the experiment were not randomly chosen from the population of containers of tadpoles (or, alternatively, the tadpoles all came from one cluster of eggs).

### Appendix – R commands

```
Tadpoles <- read.csv("data/Tadpoles.csv")
Tadpoles$uvb <- factor(Tadpoles$uvb,levels=c("Low","High"))
Tadpoles$density <- factor(Tadpoles$density)
tad.lm1 <- lm(mass~uvb*density,data=Tadpoles)
transChooser(tad.lm1)
anova(tad.lm1)
tad.mci1u <- glht(tad.lm1,mcp(uvb="Tukey"))
summary(tad.mci1u)
confint(tad.mci1u)
tad.mci1d <- glht(tad.lm1,mcp(density="Tukey"))
summary(tad.mci1d)
confint(tad.mci1d)
fitPlot(tad.lm1,which="density",ylim=c(0.75,1.3),interval=FALSE,
        xlab="Tadpole Density",ylab="Body Mass (g)")
addSigLetters(tad.lm1,which="density",lets=c("a","b","b"),pos=c(4,1,3))
fitPlot(tad.lm1,which="uvb",ylim=c(0.75,1.3),interval=FALSE,
        xlab="UV-B Light Intensity",ylab="Body Mass (g)")
addSigLetters(tad.lm1,which="uvb",lets=c("a","a"),pos=c(3,3))
```

### 3.4.2 Blood Pressure & Na-K-ATPase

Sodium (Na) plays an important role in the genesis of high blood pressure, and the kidney is the principal organ that regulates the amount of sodium in the body. The kidney contains an enzyme, Na-K-ATPase, which is essential for maintaining proper sodium levels. If the enzyme does not function properly, then high blood pressure may result. The activity of this enzyme has been studied in whole kidneys, even though the kidney is known to contain many functionally distinct sites. To see whether any particular site of Na-K-ATPase activity was abnormal with hypertension, or whether this enzyme was uniformly abnormal throughout the kidney, Garg *et al.* (1985) studied Na-K-ATPase activity at different sites along the nephrons of normal rats and specially bred rats which spontaneously develop hypertension. The sites in the kidney examined by Garg *et al.* were the distal collecting tubule (DCT), cortical collecting duct (CCD), and outer medullary collecting duct (OMCD).

The researchers hypothesized that the level of Na-K-ATPase would be depressed at all sites in the rats with hypertension. This translates to the authors expecting a rat type effect but not an interaction effect. The authors were also interested in determining if there was a significant difference in the level of Na-K-ATPase activity at the different sites. Thus, the statistical hypotheses to be examined were

$$H_0 : \text{"no rat effect"} : \mu_{Normal} = \mu_{Hyper}$$

$$H_A : \text{"rat effect"} : \mu_{Normal} \neq \mu_{Hyper}$$

$$H_0 : \text{"no location effect"} : \mu_{DCT} = \mu_{CCD} = \mu_{OMCD}$$

$$H_A : \text{"location effect"} : \text{"At least one pair of means is different"}$$

$$H_0 : \text{"no interaction effect"}$$

$$H_A : \text{"interaction effect"}$$

where the subscripts identify the levels of each factor (specifically defined above).

### Data Collection

Twelve rats were randomly selected for each type of rat (i.e., normal rats and rats with hypertension) from all rats available in the author's laboratory. The site of the kidney (DCT, CCD, or OMCD) where the Na-K-ATPase activity ( $\text{pmol} \cdot (\text{min} \cdot \text{mm})^{-1}$ ) was recorded on a rat was also randomly selected so that measurements at each location were recorded from four rats.<sup>10</sup>

### Exploratory Data Analysis and Assumption Checking

There does not appear to be any among-groups dependence because only one site was studied for each rat. There also does not appear to be any within-group dependence because the rats were randomly selected for inclusion in each treatment (although there may be some question about independence in the original pool of available rats). Variances among the treatments appear to be constant (Levene's  $p = 0.403$ ), the residuals appear to be approximately normally distributed (Anderson-Darling  $p = 0.520$ ; Figure 3.6-Left), and there are no significant outliers (outlier test  $p = 0.153$ ). The raw data will be examined with a two-way ANOVA without transformation because the assumptions have been adequately met.

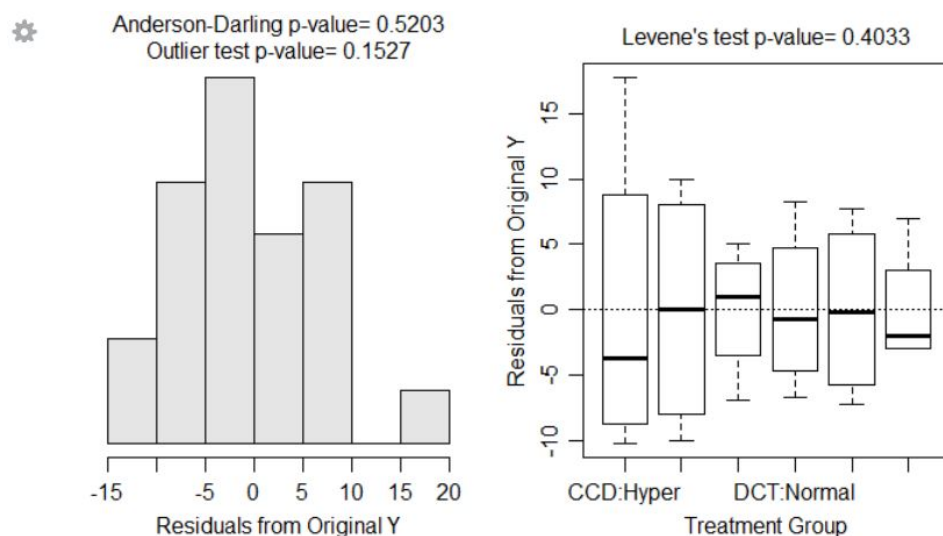


Figure 3.6. Residual plot (Left) and histogram of residuals (Right) from two-way ANOVA fit of Na-K-ATPase activity for each treatment.

### Results

There appears to be a significant interaction effect ( $p = 0.040$ ; Table 3.10); thus, the main effects cannot be interpreted directly from these results. Tukey's multiple comparison method was performed on the

<sup>10</sup>A better design would have measured the Na-K-ATPase activity at all three sites from the same rats. However, this would have violated the independence assumption and required other methods (repeated-measures or mixed-models) to analyze the data.



interaction effect to more specifically describe the differences among group means (Table 3.11; Figure 3.7). From these results it was concluded that the level of Na-K-ATPase activity was significantly lower for the hypertensive rats than for the normal rats only at the DCT site. In addition, the level of Na-K-ATPase activity was significantly greater at the DCT site than at the CCD and OMCD sites, which were not significantly different.

Table 3.10. Two-way ANOVA results for Na-K-ATPase activity levels.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
site	2	8287.0	4143.5	64.3928	6.27e-09
strain	1	459.4	459.4	7.1390	0.01555
site:strain	2	496.0	248.0	3.8541	0.04044
Residuals	18	1158.2	64.3		

Table 3.11. Results of Tukey's multiple comparisons among all pairs of treatments. [Note: only significant differences are shown to save space.]

	Estimate	Std. Error	t value	p value
DCT:Hyper - CCD:Hyper = 0	25.75	5.6722	4.5397	0.0030
DCT:Normal - CCD:Hyper = 0	46.50	5.6722	8.1979	0.0000
DCT:Hyper - CCD:Normal = 0	19.00	5.6722	3.3497	0.0355
DCT:Normal - CCD:Normal = 0	39.75	5.6722	7.0079	0.0000
DCT:Normal - DCT:Hyper = 0	20.75	5.6722	3.6582	0.0187
OMCD:Hyper - DCT:Hyper = 0	-32.75	5.6722	-5.7738	0.0002
OMCD:Normal - DCT:Hyper = 0	-34.00	5.6722	-5.9942	0.0001
OMCD:Hyper - DCT:Normal = 0	-53.50	5.6722	-9.4320	0.0000
OMCD:Normal - DCT:Normal = 0	-54.75	5.6722	-9.6524	0.0000

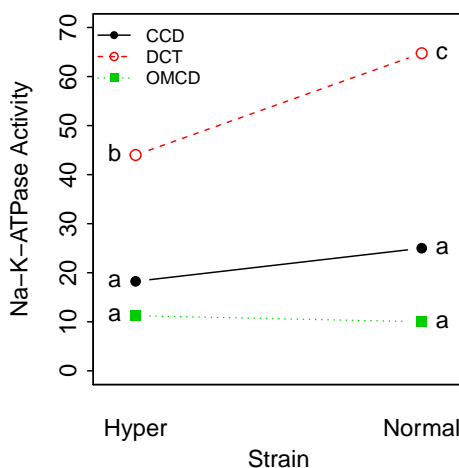


Figure 3.7. Plot of treatment means from raw tadpole body mass data. Means with different letters near them are significantly different.

## Conclusion

There is some support for the researcher's hypothesis that Na-K-ATPase activity levels would be lower in rats with hypertension. However, this support was found at only one site in the kidney, the distal collecting

tubule (DCT). At all other sites, no difference between the hypertensive and normal rats was observed. Furthermore, Na-K-ATPase activity was higher at the DCT site than either of the other two sites for both normal and hypertensive rats.

Strictly speaking, causal inferences cannot be made from this study because the rats were not randomly allocated to the hypertension and normal rat treatments. Recall that they either had this characteristic or not – they were randomly selected from their individual populations but they were not randomly selected from a population of rats and then allocated to the hypertension or normal treatments. In addition, there should be a concern that the special breeding of the rats for hypertension may have an effect on some other aspect of the rat's physiology. Finally, inferences to the entire population of rats in each group in Garg's laboratory can be made because these rats were randomly selected.

### Appendix – R commands

```
NAKATPase <- read.csv("NAKATPase.csv")
nak.lm1 <- lm(activity~site*strain,data=NAKATPase)
transChooser(nak.lm1)
anova(nak.lm1)
NAKATPase$group <- NAKATPase$site:NAKATPase$strain
nak.lm2 <- lm(activity~group,data=NAKATPase)
nak.mc2 <- glht(nak.lm2,mcp(group="Tukey"))
summary(nak.mc2)
confint(nak.mc2)
fitPlot(nak.lm1,change.order=TRUE,interval=FALSE,ylab="Na-K-ATPase Activity",
        xlab="Strain",ylim=c(0,70),legend="topleft",cex.leg=.75)
addSigLetters(nak.lm1,change.order=TRUE,lets=c("a","b","a","a","c","a"),
              pos=c(2,2,2,4,4,4))
```

[Turn the Page]

## 3.5 Summary Process

The following is a template for a process of fitting a two-way ANOVA model. Consider this process as you learn to fit one-way ANOVA models, but don't consider this to be a concrete process for all models.

1. Perform a thorough EDA of the quantitative response variable.
  - Pay close attention to the distributional shape, center, dispersion, and outliers within each level of the factor variable.
2. Fit the untransformed ultimate full (i.e., both main effects and the interaction effect) model [`lm()`].
3. Check the assumptions of the fit of the model [`transChooser()`].
  - Check equality of variances with a Levene's test and residual plot.
  - Check normality of residuals with a Anderson-Darling test and histogram of residuals.
  - Check for outliers with an outlier test, residual plot, and histogram of residuals.
4. If an assumption or assumptions are violated, then attempt to find a transformation where the assumptions are met.
  - Use the trial-and-error method [`transChooser()`], theory, or experience to identify a possible transformation.
  - If only an outlier exists (i.e., equal variances and normal residuals) and no transformation corrects the "problem," then consider removing the outlier from the data set.
  - Fit the ultimate full model with the transformed response or reduced data set.
5. Construct an ANOVA table for the full model [`anova()`] and interpret the interaction p-value.
6. If a significant interaction exists then ...
  - Do NOT interpret the main effects!!
  - Create a new variable that is the combination of the two main effect variables and fit a one-way ANOVA model with this variable as the only explanatory variable.
  - Use Tukey's HSD method on the new combined explanatory variable to determine which level means are different [`glht()`].
  - Summarize findings with significance letters [`addSigLetters()`] on an interaction plot [`fitPlot()` of the original model].
7. If a significant interaction does NOT exist then ...
  - Interpret the p-values for each main factor variable.
  - If differences among level means for a factor variable exist then use Tukey's HSD method to identify specific differences. Repeat for both factor variables if appropriate.
  - Summarize findings with significance letters [`addSigLetters()`] on a main effects plot [`fitPlot()` with the `which=` argument].
8. Give Professor Ogle a soft chocolate chip cookie.