MODULE 5

ONE-WAY IVR

Module Objectives:

1. Understand how indicator variable are created and what they are used for.
2. Understand how interaction variables are created and what they are used for.
3. Describe the order that indicator and interaction variables are entered into models.
4. Understand how to perform hypothesis tests of equal y-intercepts or slopes among groups.
5. Understand how to interpret the results of an ANCOVA.
6. Describe the benefits of an ANCOVA versus the creation of ratios or the use of residuals.

S IMPLE LINEAR REGRESSION IS A POWERFUL tool used to evaluate the relationship between two quantitative variables (Module 4). It is common for a researcher to compare simple linear regression models fit to separate groups of individuals. *Indicator variable regression* (IVR),[1] is a method to make these comparisons. For example, IVR methods could be used in each of the following situations:

1. Determine if the relationship between foot length and tail length of possums differs between possums collected in Victoria and possums collected elsewhere (Woods and Hellgren 2003).
2. Determine if the relationship between propodus height and claw closing force differs among three types of crabs (Yamada and Boulding 1998).
3. Determine if the proportional area covered by an invasive plant differs between sites after adjusting for different resident times for the plant (i.e., how long it has been present at a site) (Mullerova 2005).
4. Determine if the snout-to-vent length (SVL) to body mass relationship differs between sexes of iguanid lizards (Vitt and Goldberg 1983).
5. Determine if the liver weight of yellow perch, adjusted for somatic weight, differed between four locations – two reference sites and two sites potentially impacted by pulp and paper mills (Karels and Oikari 2000).
6. Determine if the relationship between total body electrical conductivity and lean body mass (a method used to measure body fat) differs between live and dead birds of various species (Castro *et al.* 1990).
7. Determine if the relationship between the body mass of porcupines (*Erithizon dorsatum*) and days since October 15th differs among life stages AND sexes (Sweitzer and Berger 1993).
8. Determine if the relationship between clutch size and female length for spiders differs depending on whether the spider species is from web-building or cursorial genera (Simpson 1995).

In each of these cases, a single quantitative explanatory variable is considered along with one factor explanatory variable. Because there is only one factor variable, these specific situations require a *one-way IVR* model. Furthermore, it is common practice to call the quantitative explanatory variable in these models a *covariate*. This terminology will be used throughout the rest of this module.

Δ **One-way IVR**: A regression model with a single quantitative explanatory variable and a single factor variable.

Δ **Covariate**: The quantitative explanatory variable in an IVR model.

5.1 Indicator Variables

A *factor* variable is a categorical variable with two or more levels (Weisberg 2005b).[2] A factor variable must be converted to numeric codes in order to be included in a linear regression model. An indicator variable contains the numeric "codes" created from the factor variable. Specifically, an *indicator variable* consists of zeroes and ones, where a one indicates that the individual has a certain characteristic and a zero indicates that it does not have the characteristic.[3]

---

[1]This is not a standard name. Some authors call it *dummy variable regression* (e.g., Fox 1997), whereas others call it *analysis of covariance* (ANCOVA). I will not use either of these terms as I prefer the word "indicator" to "dummy" and I will reserve the ANCOVA method for the situation where the separate SLR models are known or assumed to have equal slopes. Thus, ANCOVA models are a subset of the IVR models discussed here. It should also be noted that some authors (e.g., Weisberg 2005) do not use a separate name for IVR but simply include it as a multiple linear regression model.

[2]Factor variables were considered extensively in the ANOVA modules (see Module 2 and Module 3).

[3]Other coding schemes exist. For example, "contrast coding" uses -1 and 1. However, the 0 and 1 coding scheme leads to simpler interpretations and calculations and, thus, is used throughout this course.

Δ **Indicator Variable**: A variable that is a numerical representation of a dichotomous (or binary or binomial) categorical variable.

$\diamond$ **Indicator variables allow the inclusion of categorical variables into regression analyses.**

$\diamond$ **Indicator variables will be coded as 0 for individuals that do NOT have the characteristic of interest and 1 for individuals that do have the characteristic of interest.**

An indicator variable should be named after the characteristic denoted by a one so that it will be easy to remember what the indicator variable represents and how the coding was made.[4] For example, if an indicator variable is called *CONTROL*, then a 0 means the individual was not in the control group and a 1 means that it was in the control group. If an indicator variable is called *FEMALES*, then a 0 means that the individual is not a female (i.e., a male) and a 1 means that the individual is a female.

$\diamond$ **Indicator variables should be named after the "1" group so that it is easy to remember the coding scheme being used.**

As an example, Makarewicz (2003) studied the relationship between the concentration of mirex in the tissue of salmon and the weight of the salmon.[5] They collected these data on two species of salmon – coho (*Oncorhynchus kisutch*) and chinook salmon (*Oncorhynchus tshawytscha*) – and were interested in determining if the relationship between mirex and weight differed between these two species. For this purpose, an indicator variable may be defined from the *species* factor variable as such,[6]

$$speciescoho = \left\{ \begin{array}{l} 1 \text{ if a coho salmon} \\ 0 \text{ if NOT a coho salmon} \end{array} \right.$$

Combinations of indicator variables can be used to code factor variables that consist of more than two levels. For example, Makarewicz (2003) also were interested in determining if the relationship between mirex concentrations and weight of the salmon differed among six collection years. For simplicity and space considerations, let's initially consider that they were interested in only three years – 1977, 1982, and 1986. In this simpler situation, two indicator variables would be derived from the *year* factor variable such as,

$$year1982 = \left\{ \begin{array}{l} 1 \text{ if collected in 1982} \\ 0 \text{ if NOT collected in 1982} \end{array} \right.$$

and

$$year1986 = \left\{ \begin{array}{l} 1 \text{ if collected in 1986} \\ 0 \text{ if NOT collected in 1986} \end{array} \right.$$

Note that it is possible to create a third dummy variable to represent salmon collected in 1977. However, this indicator variable would be redundant with *year1982* and *year1986*. In other words, if it is known that

---

[4]Indicator variables will be named automatically by R. Fortunately, it follows this naming rule.

[5]Mirex is a chlorinated hydrocarbon that was commercialized as an insecticide. Use of mirex was banned in 1976 because of its impact on the environment.

[6]This indicator variable could have been named *coho* for succinctness. However, the name chosen here will be consistent with names for indicator variables automatically created by R. In addition, the indicator variable could have been chosen to represent chinook salmon.

year1982=0 and that *year1986*=0, then the salmon must have been collected in 1977. This illustrates the rule that the number of indicator variables required is one less than the number of levels to be coded.[7] Thus, five indicator variables would be needed to include information from all six years that salmon were collected by Makarewicz (2003).

◇ **One less indicator variable is needed then the number of categories in the variable.**

The level denoted by the group with "0"s for all related indicator variables is called the *reference group*. In the last example above, the salmon collected in 1977 would be considered the reference group. It will be shown in a later section that all comparisons of intercepts and slopes in the analysis will "refer" to this group of individuals.

Δ **Reference group**: The category or group that is represented by zeroes for all indicator variables.

Suppose, for illustration purposes, that fish collected in 1982 were to be considered the reference group rather than fish collected in 1977. In this situation, the *year1982* indicator variable would be replaced with the *year1977* indicator variable in the model.

◇ **Changing the reference group in an analysis requires changing the indicator variables used in the analysis.**

5.2 Interaction Variables

An *interaction variable* is an explanatory variable that is the product of two or more other explanatory variables (see Module 3). In IVR models, interactions between the covariate and the indicator variables are used to determine if the effect of the covariate on the response variable depends on the levels represented by the indicator variables. In other words, these interaction variables allow one to model the influence of a factor variable on the covariate's effect on the response variable.[8]

Δ **Interaction Variable**: An explanatory variable that is the product of two or more explanatory variables.

◇ **Interaction variables in IVR are used to determine if the effect of the covariate on the response variable differs among groups defined by the indicator variable(s).**

The interaction between two variables in an IVR model may best be illustrated by looking at the examples in Figure 5.1. In the graph on the left, the effect of salmon weight (the covariate) on mirex concentration in the tissue (the response) is the same for coho and chinook salmon. This is illustrated by each group having the same slope.[9] However, in the graph on the right, different slopes indicate that the effect of salmon weight on mirex concentration in the tissue differs between the two species. In this particular example, the relationship is "flatter" for the coho salmon than for the chinook salmon. Thus, an interaction exists for the

---

[7]This is only true if an intercept term exists in the regression model, which will be the case for all models in this course.
[8]This is the same general interpretation of interaction variables as discussed for a two-way ANOVA.
[9]In regression the "effect" of the explanatory variable on the response variable is measured by the slope.

situation depicted in Figure 5.1-Right, because the effect of the covariate on the response depends on which group is being considered. Graphically, an interaction is illustrated by non-parallel lines.
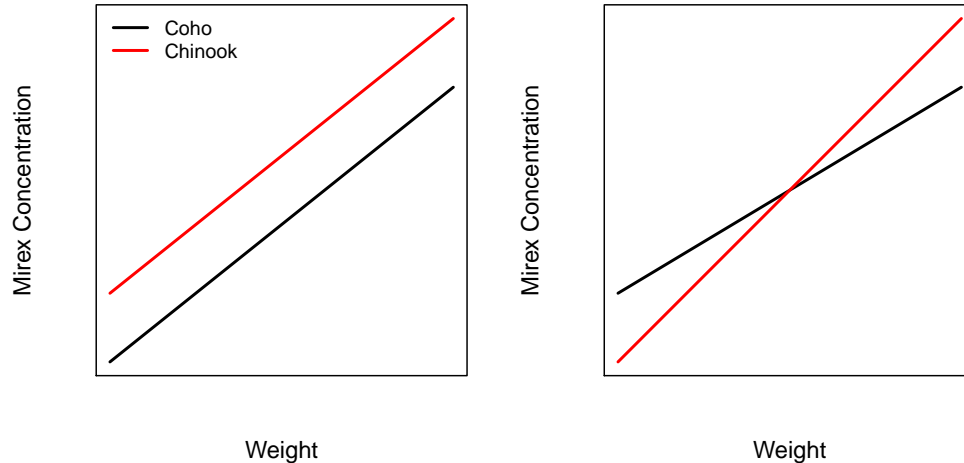

Figure 5.1. Idealized fitted-line plots representing SLR fit to two groups. The graph on the left indicates the absence of an interaction. The graph on the right indicates the presence of an interaction.

⬦ **Parallel lines indicate non-interaction. Intersecting lines (actual or eventual) indicate an interaction.**

5.3 Models & Sub-Models

Indicator and interaction variables can be used to determine if simple linear regression lines fit to different groups differ among those groups. To illustrate this point, consider a situation with one quantitative response variable (Y), one covariate (X), and one indicator variable (Z) representing a single dichotomous factor variable.[10] In addition, an interaction between the covariate and the indicator variable is constructed. Thus, the generic model fit for these data is

$$\mu_{Y|X,Z} = \alpha + \beta X + \delta_1 Z + \gamma_1 X * Z \tag{5.3.1}$$

The following symbols should be used for parameters, as was illustrated in this model, to make the interpretation of future models simpler:

- α: an overall intercept term.
- β: coefficient on the covariate.
- δ_i: coefficient on the ith indicator variable.
- γ_i: coefficient on the interaction between the covariate and the ith indicator variable.

---

[10]This situation is exactly the situation that exists for examining the effect of weight on mirex concentration between the two species of salmon.

The model in Equation (5.3.1) can be reduced to *submodels* that correspond to each group. The reduction of an overall regression model to appropriate submodels is accomplished by substituting appropriate values for the indicator variable into the overall regression model. In other words, the submodel for the reference group represented in Equation (5.3.1) is found by substituting a *0* for *Z* in Equation (5.3.1) and simplifying.[11] Similarly, the submodel for the non-reference group is found by substituting a *1* for *Z* in Equation (5.3.1) and simplifying.[12] These substitutions and simplifications are shown below.

Group	**Z=**	**Submodel ($\mu_{Y\mid\cdots} =$)**
Does NOT Have (*Reference*)	0	$= \alpha + \beta X + \delta_1 * 0 + \gamma_1 * 0 * X$ $= \alpha + \beta X$
Does Have	1	$= \alpha + \beta X + \delta_1 * 1 + \gamma_1 * 1 * X$ $= \alpha + \beta X + \delta_1 + \gamma_1 X$ $= (\alpha + \delta_1) + (\beta + \gamma_1)X$

> ◇ **Submodels are found by appropriately substituting "0"s and "1"s for the indicator variable(s) in the overall IVR model.**

Examination of the simplified submodels for each group shows that each submodel is a linear model. The submodel for the "Does NOT Have" group indicates that this group is modeled with a slope of β and an intercept of α. In contrast, the submodel for the "Does Have" group has a slope of $\beta + \gamma_1$ and an intercept of $\alpha + \delta_1$. Furthermore, from

$$\text{"Does Have" Intercept} - \text{"Does Not Have" Intercept} = (\alpha + \delta_1) - \alpha = \delta_1$$

it is seen that δ_1 is the difference in intercepts between the "Does Have" and the "Does Not Have" groups. A similar computation can be made to show that γ_1 is the difference in slope between the "Does Have" and the "Does Not Have" groups. Thus, interpretations of the four coefficients from the fit of Equation (5.3.1) are summarized in Figure 5.2 and as follows,

- α: intercept for the "Does NOT Have" group
- β: slope for the "Does NOT Have" group
- δ_1: *difference in intercept* between the "Does Have" and the "Does NOT Have"
- γ_1: *difference in slope* between the "Does Have" and the "Does NOT Have"

These interpretations show why the "Does NOT Have" group is called the reference group. The coefficients in the reference submodel represent the intercept and slope for that group. These coefficients also appear in the other submodels. The additional coefficients in the non-reference group submodels represent differences between the non-reference group and the reference group – every difference is computed relative to the value for the reference group. Thus, all other coefficients are only meaningful relative to the reference group.

> ◇ **The intercept coefficient and the coefficient on the quantitative explanatory variable are estimates of the intercept and slope, respectively, for the reference group.**

---

[11] All coefficients multiplied by a zero "drop out" of the model.
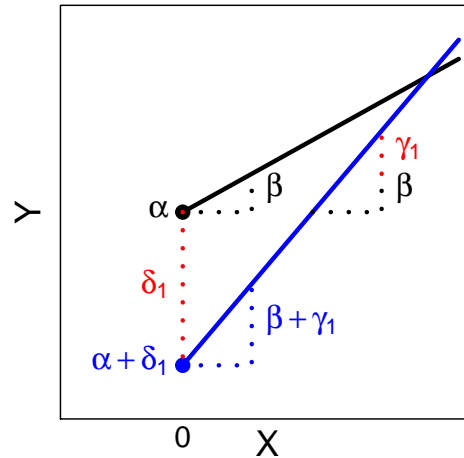[12] All coefficients multiplied by one can be simplified and combined with other coefficients.

Figure 5.2. Representation of two sub-models fit with Equation (5.3.1) and the meaning of the parameter estimates. Note that the black line is the "reference" sub-model.

⋄ **All coefficients on indicator variables are estimated differences in *intercepts* between a non-reference group and the reference group.**

⋄ **All coefficients on interaction variables are estimated differences in *slopes* between a non-reference group and the reference group.**

Further note that, with the suggested parameter symbols, the δ coefficients will represent differences in intercepts and the γ coefficients will represent differences in slopes.

⋄ **δ coefficients represent the difference in *intercept* between some non-reference group and the reference group.**

⋄ **γ coefficients represent the difference in *slope* between some non-reference group and the reference group.**

Thus, Equation (5.3.1) is used to determine which of four situations best represents the data (Figure 5.3):

- *"Coincident Lines"*: A single regression line represents the relationship between the response variable and the covariate for both groups (i.e., $\delta_1 = 0$ and $\gamma_1 = 0$).
- *"Parallel Lines"*: Separate parallel lines with different y-intercepts are needed for both groups (i.e., $\delta_1 \neq 0$ and $\gamma_1 = 0$),
- *"Same Intercept"*: Separate non-parallel lines with the same y-intercept are needed for both groups (i.e., $\delta_1 = 0$ and $\gamma_1 \neq 0$),
- *"Completely Separate Lines"*: Two completely separate lines are needed for both groups (i.e., $\delta_1 \neq 0$ and $\gamma_1 \neq 0$).

The covariate, indicator, and interaction variables should be entered into the model in that order to allow consistent and simple interpretations.
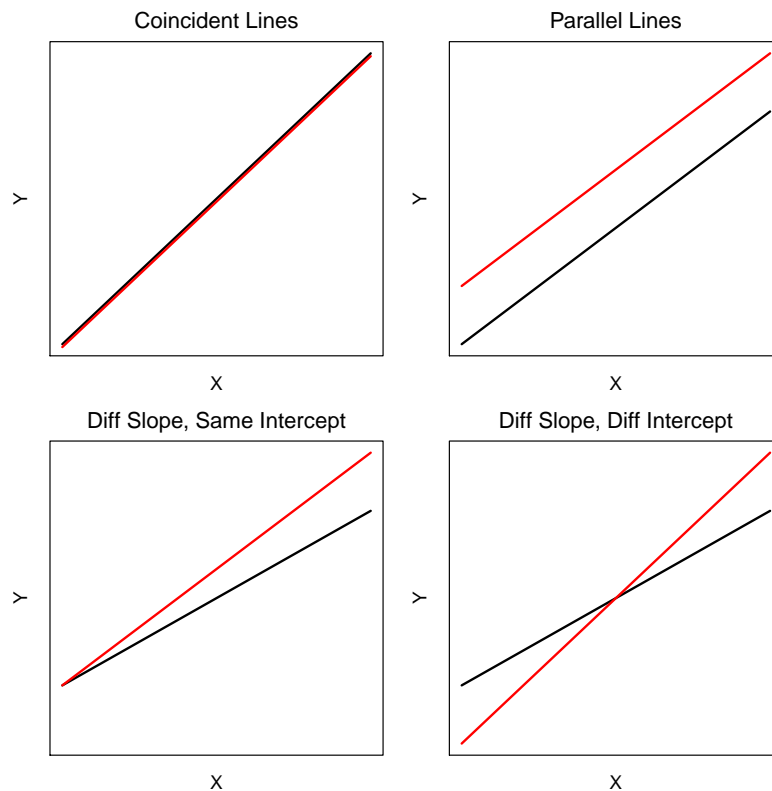
Figure 5.3. Hypothetical depictions of four situations that can occur for the relationship between a response variable (Y) and a covariate (X) for two groups of data.

◇ **The variables are entered into an IVR in the following order: (1) covariate(s), (2) indicator variables, and (3) interaction variables.**

As a specific example, the following model is used to determine if the relationship between mirex concentration (*mirex*) and weight (*weight*) differs between coho and chinook salmon,

$$\mu_{mirex|weight,speciescoho} = \alpha + \beta weight + \delta_1 speciescoho + \gamma_1 speciescoho * weight \qquad (5.3.2)$$

The submodels corresponding to each group are found by substituting appropriate values for the indicator variable, *speciescoho*, into Equation (5.3.2). These substitutions and reductions are shown below.

Group	*speciescoho=*	Submodel ($\mu_{mirex|\cdots} =$)
Chinook (*reference*)	0	$= \alpha + \beta weight + \delta_1 * 0 + \gamma_1 * 0 * weight$ $= \alpha + \beta weight$
Coho	1	$= \alpha + \beta weight + \delta_1 * 1 + \gamma_1 * 1 * weight$ $= \alpha + \beta weight + \delta_1 + \gamma_1 weight$ $= (\alpha + \delta_1) + (\beta + \gamma_1)weight$

Thus, the interpretation of the coefficients fit with Equation (5.3.2) are:

- α: intercept for chinook salmon
- β: slope for chinook salmon
- δ_1: *difference in intercept* between coho and chinook salmon
- γ_1: *difference in slope* between coho and chinook salmon

Fitting Models, Extracting Coefficients, & Making Predictions in R

Fitting the full model used to make comparisons among groups is exceptionally easy to do in R because R automatically creates the required indicator and interaction variables.[13] The full model can be fit to the data with a formula argument in `lm()` that is similar to what was used for a two-way ANOVA. Specifically, the full model is fit with a formula of the form `response~covariate*factor`, where `response` represents the quantitative response variable, `covariate` represents the quantitative explanatory variable, and `factor` represents the factor explanatory variable.[14] Estimates and confidence intervals for the parameters identified above (i.e., α, β, δ_1, and γ_1) are obtained by submitting the saved `lm()` object to `coef()` and `confint()`.

As a specific example, consider that the mirex concentration in salmon data consists of these variables:

1. *mirex*: the measured mirex concentration in the tissue ($mg \cdot kg^{-1}$).
2. *weight*: the measured weight of the fish (kg).
3. *species*: a factor variable indicating the species (*coho* or *chinook*).

```
> Mirex <- read.csv("Mirex.csv")
```

The model expressed in Equation (5.3.2) is then fit and the coefficient estimates and confidence intervals are extracted with `coef()` and `confint()`.[15] Thus, $\hat{\alpha}$=0.1355, $\hat{\beta}$=0.0096, $\hat{\delta}_1$=-0.0509, and $\hat{\gamma}_1$=0.0121. Thus, for example, the sample slope for chinook salmon is 0.0096 and the sample slope for coho salmon is -0.0121 smaller than the sample slope for chinook salmon. Furthermore, one is 95% confident that β is between 0.0039 and 0.0153 and γ_1 is between -0.0011 and 0.0253.

```
> mirex.lm1 <- lm(mirex~weight*species,data=Mirex)
> cbind(Ests=coef(mirex.lm1),confint(mirex.lm1))
                            Ests        2.5 %      97.5 %
(Intercept)           0.135451558   0.093132341 0.17777078
weight                0.009607284   0.003886599 0.01532797
speciescoho          -0.050914887  -0.113215438 0.01138566
weight:speciescoho    0.012110273  -0.001079885 0.02530043
```

Of course, inferential methods are needed to determine if these result suggest a real difference in parameters between the populations. For example, a statistical test of whether $\gamma_1 = 0$ would be used to determine if the population slope differed between coho and chinook salmon. These types of tests are the subject of a subsequent section.

---

[13]Unfortunately, this ease comes at the cost of the actual model to be fit being somewhat transparent to the user. The underlying structure of the model (e.g., indicator variables) can be viewed by sending the saved `lm()` object to `model.matrix()`.

[14]As with two-way ANOVA, this simple declaration, using an asterisk to combine the covariate and the factor variable, is an efficient way to tell R to include both main variables and the interaction term. In other words, the `response~covariate*factor` code is shorthand for the `response~covariate+factor+covariate:factor` code.

[15]The order of the variables should be as discussed previously – covariate followed by the indicator variable.

As with SLR, the model fit may be visualized with a fitted line plot as constructed with `fitPlot()`. The fitted line plot for an IVR will contain different symbols and lines for each group in the model. The symbols and lines will be identified with a legend which can be moved from the default location with a string in `legend=`. [*The plot produced by the code below is not shown here.*]

```
> fitPlot(mirex.lm1,legend="topleft",ylab="Mirex Concentration",xlab="Weight")
```

Predicted values for IVR models are constructed by submitting the `lm()` object and a data.frame of observed values of the explanatory variables to `predict()` or `predictionPlot()` as was described for SLR models. One must make sure, though, that the data.frame contains an observed value for each explanatory variable found in the `lm()` object. For example, the predicted value, with a 95% prediction interval, for a 3 kg coho salmon is found below.

```
> predict(mirex.lm1,data.frame(weight=3,species="coho"),interval="prediction")
        fit         lwr       upr
1 0.1496893 -0.03240232 0.331781
```

Similarly, the predicted values, with 95% prediction intervals, for 5 kg chinook AND a 3 kg coho salmon are found below

```
> predict(mirex.lm1,data.frame(weight=c(5,3),species=c("chinook","coho")),
         interval="prediction")
        fit          lwr       upr
1 0.1834880   0.001557795 0.3654182
2 0.1496893  -0.032402319 0.3317810
```

5.4 Hypothesis Testing

5.4.1 Overall F-Test

As with the other linear models considered thus far, IVR models can be cast as a comparison of two models. The *ultimate full model* is the model that contains all of the explanatory variables of interest (i.e., the covariate plus all indicator and interaction variables). The *ultimate simple model* is the same as it was with ANOVA and SLR; i.e., the model with no explanatory variables and just a constant at α or, equivalently, $\overline{Y}$. The overall F-test is a comparison of the ultimate full to the ultimate simple model.

Δ **Ultimate Full Model**: Model that contains ALL explanatory variables of interest.

Δ **Ultimate Simple Model**: Model that contains NO explanatory variables.

The overall ANOVA table for IVRs considers three sources of variability – regression, residual, and total. SS_{total} is, as always, found by summing the squared residuals from the ultimate simple model.[16] $SS_{residual}$ is the sum of squared residuals from the full model, with $df_{residual}$ equal to the number of individuals minus the number of parameters in the model. As with the SLR, $SS_{regression}$ is the difference between SS_{total} and $SS_{residual}$ and represensts the improvement in fit from using the ultimate full versus the ultimate simple model. Correspondingly, $df_{regression}$ are the difference between the number of parameters in the ultimate simple (always 1) and full models.

---

[16]With corresponding df_{total} of $n-1$, consistent with all other linear models considered thus far.

The MS and F test are computed as before. $MS_{regression}$ and $MS_{residual}$ have the same interpretation as before – $MS_{regression}$ is the amount of variability that the full model explained and $MS_{residual}$ is the amount of variability that the full model could not explain. However, in IVR, the $MS_{regression}$ is how much variability **ALL** of the explanatory variables combined could explain. In other words, the *COMBINED* effect of all explanatory variables is considered in the overall ANOVA F-test for IVR.

⬦ **The $MS_{Regression}$ in IVR is a measure of the combined variability in the response variable that IS explained by *ALL* of the explanatory variables being considered.**

Thus, a significant F test indicates that a substantial portion of the variability in the response variable is explained by all of the explanatory variables combined. Importantly, a significant F test indicates that there is a significant relationship between the response and some combination, including potentially only one, of the explanatory variables. The F test does not immediately indicate which, or which combination, of the explanatory variables forms the significant relationship. Methods for determining which variables are significant are considered in the next two sections. As a general rule, the overall F test should be consulted first to determine if some significant relationship exists. If no significant relationship exists, then one need not examine specific relationships further. If a significant relationship is found, then the methods of the next sections can be considered.

⬦ **A significant F test in an IVR indicates that at least one of the explanatory variables is important for explaining the variability of the response variable.**

⬦ **An overall F-test (and corresponding p-value) indicates whether or not there is a significant relationship between the response variable and all of the explanatory variables taken as a group. Post-hoc methods are required to further dissect the specific nature of the relationship.**

5.4.2   Partial F-Tests

Not every model comparison related to IVR needs to involve the ultimate simple and full models. For example, it may be of interest to determine if the two lines in the mirex concentration data corresponding to the coho and chinook salmon are *parallel*. In other words, is the relationship between mirex concentration and weight the same for coho and chinook salmon.

A *parallel lines test* (or equal slopes test) compares the ultimate full model to the model that contains no interactions between the covariate and the indicator variables. The hypotheses for this test may be written as

$$H_O : \text{``The coho and chinook lines are parallel''}$$
$$H_A : \text{``The coho and chinook lines are NOT parallel''}$$

or in terms of coefficients as (referring back to Equation (5.3.2))

$$H_O : \gamma_1 = 0$$
$$H_A : \gamma_1 \neq 0$$

However, most usefully, the hypotheses for the parallel lines test can be written in terms of models as

$$H_O : \mu_{mirex|\cdots} = \alpha + \beta weight + \delta_1 speciescoho$$
$$H_A : \mu_{mirex|\cdots} = \alpha + \beta weight + \delta_1 speciescoho + \gamma_1 speciescoho * weight$$

Models, and their corresponding hypotheses, can be compared by plugging $SS_{residuals}$ from the competing models into Equation (1.5.2) to generate an F test statistic. This F test statistic, similar to all other F test statistics based on model comparisons, is a ratio of the variability that WAS explained by the full model to the variability that was not explained by the full model. A "large" F indicates that the full model explains significantly more of the variability in the response variable then the simple model does.

This process can be illustrated with the mirex concentration data. First, the ultimate full model (which corresponds to H_A in this example) is fit (as was done previously) and the SS are extracted with `anova()`.

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

Response: mirex
               Df  Sum Sq  Mean Sq F value    Pr(>F)
weight          1 0.22298 0.222980 26.8586 9.155e-07
species         1 0.00050 0.000498  0.0600   0.80690
weight:species  1 0.02744 0.027444  3.3057   0.07158
Residuals     118 0.97964 0.008302
```

The df and SS of interest, at this point, are in the "Residuals" row of the output. Thus, RSS_{Full}=0.980, df_{Full}=118, and $RMS_{UltimateFull}$=0.0083. The simple model is, however, the model with just the covariate and the indicator variables, which is fit and the SS are extracted below.

```
> mirex.lm2 <- lm(mirex~weight+species,data=Mirex)
> anova(mirex.lm2)
Analysis of Variance Table

Response: mirex
           Df  Sum Sq  Mean Sq F value    Pr(>F)
weight      1 0.22298 0.222980 26.3481 1.124e-06
species     1 0.00050 0.000498  0.0589    0.8087
Residuals 119 1.00708 0.008463
```

Thus, RSS_{Simple}=1.007 and df_{Simple}=119 and the resultant F test statistic is

$$F = \frac{\frac{1.007-0.980}{119-118}}{0.0083} = 3.31$$

With 1 and 118 df the p-value is $p = 0.072$, which provides weak, but not significant, evidence that the full model is "better" than the simple model. Thus, one would conclude that the coho and chinook salmon have equal slopes and the relationship between mirex concentration and fish weight is the same for both species.

Δ **Parallel Lines Test**: An F test to determine if all groups in an IVR analysis can be described by regression lines with the exact same slope.

◇ **A parallel lines test is conducted by comparing the ultimate full model to the model that contains the covariate and indicator variables (i.e., no interaction variables with the quantitative variable are included).**

If the groups in an IVR are found to be parallel (i.e., have equal slopes) then the analyst should proceed to determine if the groups also have statistically equal-intercepts. *It must be noted here, though, that a test of equal-intercepts is relevant only if the lines are known or have been found to be parallel.* The reasoning for this warning can be illustrated in at least two ways. First, recall that the difference in intercepts come from examination of the coefficients on the indicator variables. The indicator variables stem from one original factor variable and, thus, can be thought of as "main effects." If a parallel lines test indicates non-parallel lines then it has been found that the interaction term is an important contributor to the model. As was learned with a two-way ANOVA, if an interaction exists in the model then the main effect terms should not be interpreted.

Δ **equal-intercepts Test**: An F test to determine if all groups in an IVR analysis can be described by regression lines with the exact same intercept *given that the lines have equal slopes.*

◇ **An equal-intercepts test is only appropriate if the lines are known or have been found to be parallel.**

Second, if the lines are not parallel then the significance of the intercept term depends on the relative magnitude of the slopes and "how far" the observed data is from $X = 0$. For example, the intercepts in Figure 5.4-Left are statistically different. However, if the center of the observed data was on $X = 0$ as in Figure 5.4-Right then the intercepts would not be statistically different. In summary, the test of equal-intercepts is not a useful test if the slopes differ between the groups.
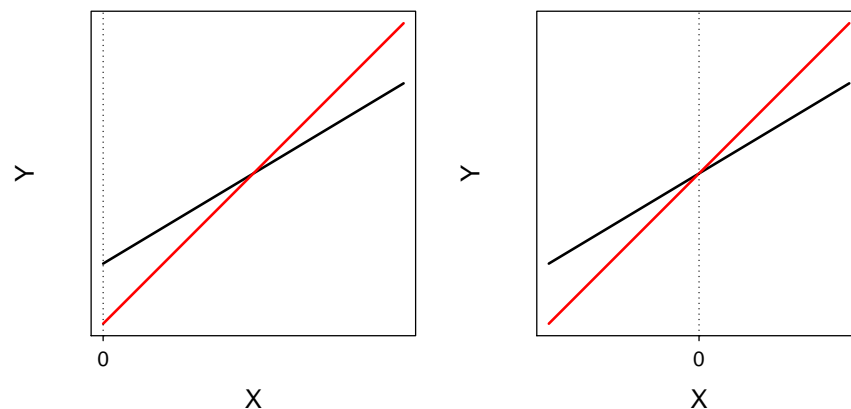

Figure 5.4. Representation of two sub-model fits with non-parallel lines. The left plot illustrates a situation where $X = 0$ is at the left margin of the observed data. The right plot illustrates a situation where $X = 0$ is at the center of the observed data. Overall, the two plots illustrate how the difference in intercepts depends on where $X = 0$ is relative to the two fitted lines.

On the other hand, if the lines are parallel then the equal-intercepts test is a very important test. With parallel lines the difference in intercepts is a measure of the vertical difference between the lines for every

value of the covariate. Thus, the difference in intercepts is a measure of the difference in the response variable between the groups when the covariate is held constant. In these situations, the equal-intercepts test tests whether this constant difference is significantly different than 0 or not.

> ◇ **With parallel lines the equal-intercepts test tests whether the the mean of the response variable when the quantitative variable is held constant differs between the groups.**

The model hypotheses for the equal-intercepts test, assuming parallel lines, are,

$$H_O : \mu_{mirex|\cdots} = \alpha + \beta weight$$
$$H_A : \mu_{mirex|\cdots} = \alpha + \beta weight + \delta_1 speciescoho$$

As the full model does not have an interaction term then it clearly states that this test of equal-intercepts follows knowing or testing that the slopes are parallel.

The full model in these hypotheses is not the ultimate full model, rather it is the simple model from the parallel lines test. Thus, RSS_{Full}=1.007 and df_{Full}=119. However, $RMS_{UltimateFull}$=0.0083 still holds.

The simple model is fit and the SS are extracted below.

```
> mirex.lm3 <- lm(mirex~weight,data=Mirex)
> anova(mirex.lm3)
Analysis of Variance Table

Response: mirex
           Df  Sum Sq  Mean Sq F value    Pr(>F)
weight      1 0.22298 0.222980  26.556 1.019e-06
Residuals 120 1.00758 0.008396
```

Thus, RSS_{Simple}=1.008 and df_{Simple}=120. The F test statistic,

$$F = \frac{\frac{1.008-1.007}{120-119}}{0.0083} = 0.06$$

With 1 and 118 df the p-value is $p = 0.807$. Thus, if the lines are considered to be parallel then there does not appear to be a significant difference between the intercepts for the chinook and coho salmon lines. Thus, it appears that the regression of mirex concentration on fish weight can be modeled by the same line (i.e., same slopes and same intercepts).

The default hypothesis test comparing the ultimate full and ultimate simple models is called the overall F-test. The hypothesis tests that compare a full and a simple model of which at least one is not an ultimate model is called a *partial F-test*. Partial F-tests are very useful for examining hypothesis tests that require combining groups of coefficients (e.g., "parallel lines" and "equal intercepts" tests).

> △ **Overall F Test**: The F-test that compares the ultimate full to the ultimate simple model.

> △ **Partial F Test**: An F-test where at lease one of the full and simple models is not the ultimate full or simple model.

Partial F Tests in R

Most of the F tests just described can all be computed efficiently by submitting the saved `lm()` object from the fit of the *ultimate full model* to `anova()`. For example, a close examination of the following results shows that the p-values for the parallel lines and equal-intercepts tests (assuming parallel lines) are found in the lines corresponding to the interaction and factor variables, respectively. This a general result; i.e., these two p-values will always be found in these two lines. As the equal interecepts test assumes parallel lines the equal-intercepts test p-value should never be used if the parallel lines p-value is less than α (i.e., indicating non-parallel lines).

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

Response: mirex
               Df  Sum Sq  Mean Sq F value    Pr(>F)
weight          1 0.22298 0.222980 26.8586 9.155e-07
species         1 0.00050 0.000498  0.0600   0.80690
weight:species  1 0.02744 0.027444  3.3057   0.07158
Residuals     118 0.97964 0.008302
```

> ◇ **The p-value for the parallel lines test is found in the row from `anova()` labelled with the interaction variable.**

> ◇ **The p-value for the equal-intercepts test (assuming parallel lines) is found in the row from `anova()` labelled with the factor variable.**

> ◇ **The equal-intercepts test p-value should NOT be interpreted if the parallel lines test p-value is less than α.**

5.4.3 Coefficients t-Tests

Most computer programs, including R, construct default hypothesis tests for the coefficients from a model fit. The hypothesis tests about individual model coefficients are constructed in the same manner as described for SLR (Module 4). Specifically, these default tests test that a coefficient is not equal to zero with the t test statistic found in Equation (4.2.1). The test statistics and p-values for the coefficients from fitting Equation (5.3.2) are found with `summary()`.

Despite the familiarity and simplicity of these calculations, one must be very careful when interpreting the results. These tests test the significance of a variable in describing the variability of the response variable *after all other explanatory variables have been included in the model.* In other words, these tests test the importance of a particular explanatory variable adjusted for every other variable in the model. For example, the test of the *speciescoho* coefficient in Equation (5.3.2) tests whether the intercepts differ between chinook and coho salmon assuming that *weight:speciescoho* is in the model. In other words, this test tests whether the intercepts differ assuming that the slopes are unequal. This is problematic because it is testing a "main effect" after an interaction effect is in the model. Thus, in general, **the test statistics and p-values corresponding to the model coefficients are not useful**.

> ◇ **The default t-tests for the model coefficients test the importance of an explanatory variable after all other explanatory variables are included in the model.**

> ◇ **The default t-tests for the model coefficients are generally inappropriate.**

The actual coefficient values are useful, however, because they are estimates of the intercept and slope for the reference group or for the difference in intercepts and slopes as described in previous sections. Thus, model coefficient values will be interpreted but their significance will not be determined from the default t-tests. Methods for testing which slopes are different or which intercepts are different will be described in the next section.

5.5 Multiple Comparisons

Comparing Slopes

To motivate the topics of this section, consider a change in the intent for the mirex concentration in salmon data. Suppose that interest is in determining if the relationship between mirex concentration in the tissue and the weight of the salmon differs among the six collection years, rather than between the two species. The following code was used to partially address this by performing the parallel lines test and, if appropriate, the equal-intercepts test.[17]

```
> Mirex$year <- factor(Mirex$year)
> mirex.lma <- lm(mirex~weight*year,data=Mirex)
> anova(mirex.lma)
Analysis of Variance Table

Response: mirex
             Df  Sum Sq  Mean Sq F value    Pr(>F)
weight        1 0.22298 0.222980 60.2061    4.7e-12
year          5 0.50667 0.101333 27.3607  < 2.2e-16
weight:year   5 0.09351 0.018703  5.0499 0.0003326
Residuals   110 0.40740 0.003704
```

The interaction term p-value ($p < 0.0005$) provides very strong evidence that the six lines are not parallel. Because the lines are not parallel, the equal-intercepts test (second line of output) was not interpreted. The next question after a result such as this is to ask which lines have different slopes – does just one line have a different slope, do all lines have different slopes, etc. An examination of the fitted-line plot for this model (Figure 5.5) suggests that the first four years have similar slopes, 1996 has a shallower slope, and 1999 has an even shallower slope. However, it is impossible to determine which slopes are significantly different from this graph.

One could look at the coefficient p-values from `summary()` to determine which groups differ from the reference group.[18] However, there is no efficient way to make comparisons among groups that don't include the reference group and there is the same problem with multiple comparisons discussed in Section 2.4. Unfortunately, a method such as Tukey's HSD method does not exist for comparing multiple slopes. Thus, a generic multiple comparison method must be used.

---

[17]The second line is used to convert *year* from a numeric to a factor variable.
[18]Examination of these tests is appropriate in this case because the interaction terms are added at the "end of the model" and their interpretation "after all other variables are in the model" is exactly as was intended.
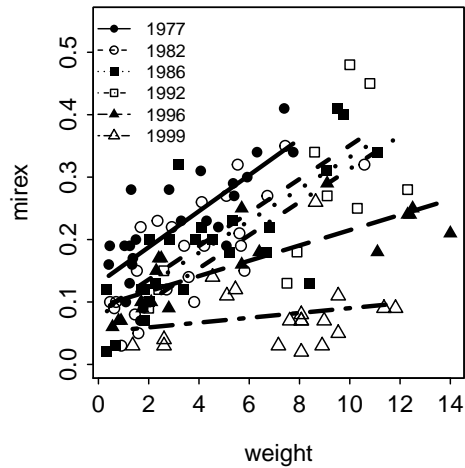
Figure 5.5. Fitted line plot for the salmon mirex concentration by year linear model results.

The "false discovery rate" (FDR), rather than the experimentwise error rate, has become popular, especially in the ecological literature (see Verhoeven *et al.* 2005). The FDR is the expected proportion of false discoveries (i.e., rejecting H_0 when it is in fact true) among the rejected null hypotheses. Controlling the FDR is a less stringent condition than controlling the experimentwise error rate. Therefore, methods that use the FDR are more powerful than other *ad hoc* methods that control the experimentwise error rate (e.g., Bonferroni methods). Methods based on the FDR have been described by Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001), and Storey (2002). The FDR methods work by multiplying the individual comparison p-values by a factor greater than 1. This "adjusted p-value" is then compared to the desired α for the desired false discovery rate. The details of these methods will not be discussed further here either but the interested reader is directed to the very readable summary by Verhoeven *et al.* (2005). Because of the increased power related to FDR-based methods, it is suggested that you use these methods when comparing slopes among a large number of groups.

> Δ **False Discovery Rate**: the expected proportion of false discoveries among the rejected null hypotheses.

For example, the following results summarize the comparisons of all slopes among all pairs of groups in the mirex concentration by year analysis using the FDR control method.

```
comparison     diff X95..LCI X95..UCI p.unadj   p.adj
 1982-1977 -0.00202 -0.01678  0.01273 0.78627 1.00000
 1986-1977 -0.00331 -0.01670  0.01008 0.62556 1.00000
 1992-1977 -0.00233 -0.01676  0.01211 0.75010 1.00000
 1996-1977 -0.01661 -0.02903 -0.00418 0.00929 0.09394
 1999-1977 -0.02500 -0.03964 -0.01036 0.00099 0.01386
 1986-1982 -0.00128 -0.01393  0.01137 0.84111 1.00000
 1992-1982 -0.00030 -0.01405  0.01345 0.96535 1.00000
 1996-1982 -0.01458 -0.02621 -0.00295 0.01445 0.11790
 1999-1982 -0.02297 -0.03694 -0.00901 0.00149 0.01788
 1992-1986  0.00098 -0.01130  0.01326 0.87451 1.00000
 1996-1986 -0.01330 -0.02314 -0.00346 0.00854 0.09394
```

```
1999-1986 -0.02169 -0.03421 -0.00917 0.00084 0.01260
1996-1992 -0.01428 -0.02550 -0.00306 0.01310 0.11790
1999-1992 -0.02267 -0.03630 -0.00904 0.00132 0.01716
1999-1996 -0.00839 -0.01988  0.00309 0.15049 1.00000
```

These results (in the *adj.p* column) indicate that the slopes for 1977, 1982, 1986, and 1992 are all equal, the slopes for 1996 and 1999 are significantly lower than the slopes for the other four years, and the slopes from 1996 and 1999 are equal. Thus, it appears that the slope between mirex concentration in the tissue and fish weight remained constant from 1977 to 1992, declined significantly in 1996, and remained lower in 1999.

Comparing Intercepts

If the lines are all parallel among groups but the intercepts are different, then the same multiple comparisons problem is encountered. However, the differences among all pairs of intercepts can be analyzed with a slight modification of Tukey's HSD method. The modification is that Tukey's procedure must be constructed with the *adjusted values*. The adjusted values in an IVR with parallel lines are the original observed values of the response variable adjusted as if all individuals had the exact same value of the covariate. Typical common values to be used for the covariate are 0 and the mean of the covariate.

A conceptual visualization of this "adjustment" process is to fit the parallel lines model to the observed data (Figure 5.6-Top Left), imagine tilting the fitted line plot from this model fit such that the fitted lines have a slope of zero (Figure 5.6-Top Right), and then "compressing" all of the points left-to-right so that they are all centered on $X = 0$ or $X = \overline{X}$ (Figure 5.6-Bottom Left). A one-way ANOVA model is then used to determine if the means of the adjusted values are different (Figure 5.6-Bottom Right).[19] More importantly, Tukey's HSD method can then be applied to the adjusted data to determine which adjusted means and, hence which intercepts, are significantly different.

As an example analysis of the intercepts, consider the mirex concentration data JUST for the four years from 1977 to 1992 (the years for which the lines were parallel). This reduced data set is obtained from the larger data set with,

```
> Mirex1 <- filterD(Mirex,!year %in% c("1996","1999"))
```

The parallel lines test and equal-intercepts tests for this reduced data set are obtained below. The interaction term p-value ($p = 0.976$) confirms the conclusion that the slopes for these four years are all equal. The indicator variable p-value ($p = 0.001$) indicates that at least one pair of the intercepts for these four years are different.

---

[19]The results of this one-way ANOVA model should indicate a significant difference among mean adjusted values if the equal-intercepts hypothesis was rejected.
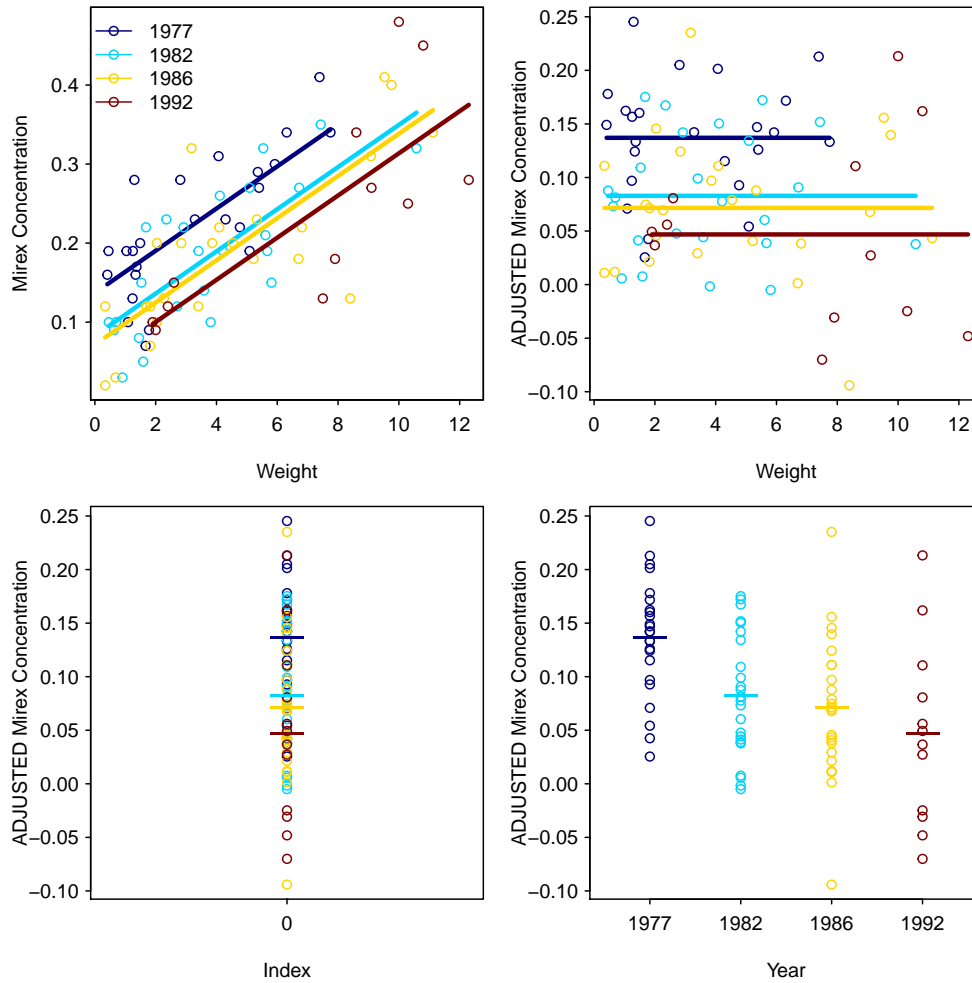
Figure 5.6. Plots illustrating the process of adjusting observed values to a common scale to allow the use of Tukey's HSD method for detecting differences in intercepts for parallel regression lines. The process is detailed in the text but proceeds left-to-right and then top-to-bottom on this graph.

```
> mirex1.lma <- lm(mirex~weight*year,data=Mirex1)
> anova(mirex1.lma)
Analysis of Variance Table

Response: mirex
            Df  Sum Sq Mean Sq  F value     Pr(>F)
weight       1 0.43886 0.43886 103.7272 7.431e-16
year         3 0.07602 0.02534   5.9895   0.001016
weight:year  3 0.00089 0.00030   0.0704   0.975568
Residuals   76 0.32155 0.00423
```

The results from the Tukey HSD method applied to the adjusted data are shown below. These results suggest that the intercept for 1977 is significantly higher than the intercepts for the other three years. The intercepts for the other three years are all statistically equal.

```
 comparison        diff     X95..LCI       X95..UCI         p.adj
  1982-1977 -0.05416588 -0.10225344 -0.006078329 0.0209412606
  1986-1977 -0.06545122 -0.11353877 -0.017363664 0.0033259143
  1992-1977 -0.09021635 -0.14911133 -0.031321360 0.0007448995
  1986-1982 -0.01128533 -0.05937289  0.036802220 0.9267699991
  1992-1982 -0.03605046 -0.09494545  0.022844523 0.3810617955
  1992-1986 -0.02476513 -0.08366011  0.034129857 0.6885397505
```

Thus, these results suggest that the regression of mirex concentration on fish weight is coincident for 1982, 1986, and 1992. In addition, the adjusted mean mirex concentration in 1977 is higher than the adjusted mean mirex concentration in 1982, 1986, and 1992 for all salmon weight values. Thus, the *relationship* between mirex concentration and salmon weight did not change from 1977 to the following three years but the mirex concentrations declined significantly no matter the weight of the fish.

Multiple Comparisons in R

The results from comparing all possible pairs of slopes is obtained by submitting the saved `lm()` object containing the ultimate full model to `compSlopes()`. This function defaults to using the preferred `"fdr"` method. The FDR control results for the mirex concentration by year shown above were obtained with

```
> compSlopes(mirex.lma)
```

If the lines are found to be parallel and there is some difference in the intercepts, then the specific differences in intercepts can be found by submitting the saved `lm()` object to `compIntercepts()`.[20] This function converts the observed data to adjusted values, fits the required one-way ANOVA model, and then uses Tukey's procedure to identify any differences in adjusted values. The Tukey HSD results shown above for the mirex concentration example for the four years between 1977 and 1992 were obtained as shown below.[21]

```
> mirex1.lmb <- lm(mirex~weight+year,data=Mirex1)
> compIntercepts(mirex1.lmb)
```

---

[20]The `lm()` object should not contain an interaction term as testing the intercepts is only appropriate on parallel lines. However, `compIntercepts()` will adjust, though a warning will be issued, to using a model without an interaction term if a model with an interaction term is supplied.

[21]The same results could be extracted from the ultimate full model as `compIntercepts()` will identify the interaction term and then fit the model without it before testing the intercepts.

5.6   Assumptions & Transformations

The assumptions for an IVR are the same as those for an SLR – independence, linearity, homoscedasticity, normality, and no outliers (Section 4.3). Methods for assessing the validity of these assumptions in an IVR are also essentially the same as those used for an SLR. In other words, fitted-line and residual plots will be used to examine linearity and assess homoscedasticity, an Anderson-Darling test and a histogram of the residuals will be used to assess the normality of residuals, and a residual plot and an outlier test will be used to diagnose outliers and influential points.

> ◇ **The assumptions of an IVR are the same as the assumptions for an SLR.**

> ◇ **Methods for assessing IVR assumptions are the same as those methods used for SLR.**

The only point of note for these analyses is that the assumptions should be assessed on the ultimate full model. If the assumptions are not met on the ultimate full model, then transformations for the response and the covariate[22] should be considered such that a transformed version of the ultimate full model meets the assumptions.[23] Once an ultimate full model is found where all of the assumptions are met, then it is assumed that all assumptions will be met for any model that is a nested subset of the ultimate full model. In other words, the assumptions do not need to be re-checked if non-significant explanatory variables are removed from the ultimate model.

> ◇ **The assumptions are assessed for the ultimate full model.**

> ◇ **Assumptions do not need to be assessed for simpler models that are nested within the ultimate full model if the assumptions were met for the ultimate full model.**

5.7   Example Analyses

5.7.1   Crab Claws

As part of a study of the effects of predatory intertidal crab species on snail populations, Yamada and Boulding (1998) measured the mean closing forces and the propodus heights of the claws on several individuals of three species of crabs – *Hemigrapsus nudus*, *Lophopanopeus bellus*, and *Cancer productus*. In one part of their research they hypothesized that the closing force of claws of species that were mollusc specialists (*Lophopanopeus bellus* and *Cancer productus*) would be greater than the closing force of claws of species that were mollusc generalists (*Hemigrapsus nudus*). The closing force would be adjusted for the size of the claw because of a suspected positive relationship between the size of the claw and the closing force.

Data Collection

Claw closing forces were measured with an *in vivo* "calibration device" that consisted of a piece of heavy wire bent into an Ʊ shape and a strain gauge glued to a piece of 0.76 mm steel shimstock and hot-melt glued to the straight edge of the wire. A linear relationship was found between the force exerted on the calibrating device by known weights and the output of the Wheatstone bridge containing the device's strain gauge. Claw closing forces were measured by positioning the mid-dactyl and mid-propodus fingers of a crab's claws into

---

[22]If the covariate is transformed then any interaction terms with the covariate should be recreated using the transformed covariate.
[23]The methods and rules for transforming these variables in an IVR are the same as those used for an SLR (see Section 4.4).

the upper and lower wire loops of the strain gauge and waiting for the crab to "attack" so that it closed its claw and pulled the loops together. Crabs were selected from the three species possessing propodus heights between 4.5 and 13 mm. Both right and left claws of 11-15 individuals per species were measured repeatedly and the highest reading per crab was recorded.

The data file to be analyzed consists of the following variables and their meanings:

- *spec*: a factor with levels *cp*, *hn*, and *lb* corresponding to abbreviations for the three crab species.
- *propodus*: measurement of claw propodus height (mm).
- *force*: measurement of claw closing force.

In the models, *spec* will be converted to indicator variables of the form *spechn* (=1 if *Hemigrapsus nudus*, =0 otherwise) and *speclb* (=1 if *Lophopanopeus bellus*, =0 otherwise). In addition, interaction variables between *propodus* and the indicator variables from *spec* will be constructed.

EDA & Assumption Checking

The initial ultimate full model fit to these data was

$$\mu_{force|propodus,spec} = \alpha + \beta propodus + \delta_1 spechn + \delta_2 speclb + \gamma_1 spechn * propodus + \gamma_2 speclb * propodus$$

The fit of this model exhibited linearity, a slight heteroscedasticity (Figure 5.7-Right), approximately normal residuals (Anderson-Darling $p = 0.584$), and no significant outliers (outlier test $p > 1$). Possible transformations for the response variable were considered through a trial-and-error method.
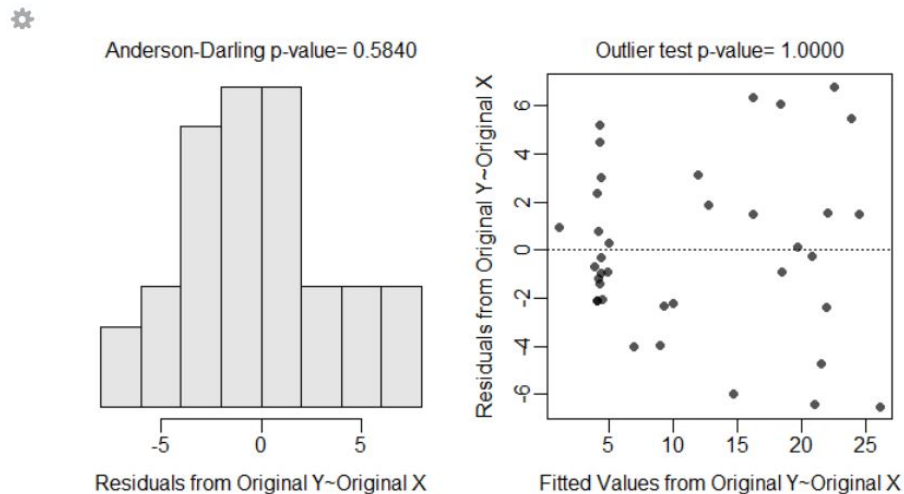

Figure 5.7. Histogram of residuals (Left) and residual plot (Right) for the regression of claw closing force on propodus height.

A square-root transformation of closing force resulted in linearity, homoscedasticity (Figure 5.8-Right), approximately normal residuals ($p = 0.303$), and no significant outliers (outlier test $p > 1$). With the assumptions largely met, this analysis will proceed by fitting the following ultimate full model to all indvidiuals in the data set

$$\mu_{sqrt(force)|...} = \alpha + \beta propodus + \delta_1 spechn + \delta_2 speclb + \gamma_1 spechn * propodus + \gamma_2 speclb * propodus$$
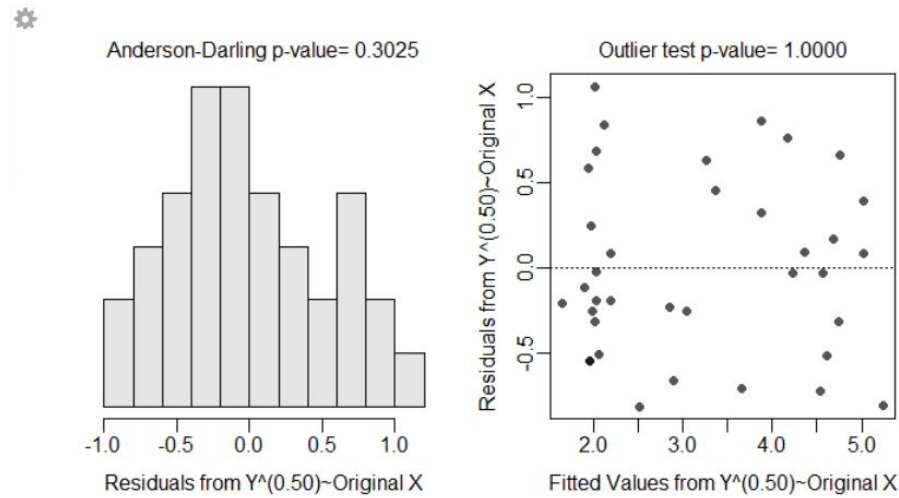
Figure 5.8. Histogram of residuals (Left) and residual plot (Right) for the regression of square-root of claw closing force on propodus height.

Results

The parallel lines test indicated that at least two of the three regression lines (by species) of the square root of closing force on propodus height had different slopes ($p < 0.0005$; Table 5.1). Post hoc multiple comparisons of all pairs of slopes, controlling the false discovery rate at 0.05, indicated that the slope for *Hemigrapsus nudus* was significantly lower than the slopes for the other two species, which were not significantly different (Table 5.2). Post hoc multiple comparisons for whether each slope is equal to zero or not indicated that the slopes for *Cancer productus* and *Lophopanopeus bellus* were both significantly greater than zero, whereas the slope for *Hemigrapsus nudus* was not significantly different from zero (Table 5.2).

Table 5.1. ANOVA table for ultimate full model using the transformed crab claw data.

```
              Df  Sum Sq Mean Sq F value     Pr(>F)
propodus       1 29.4494 29.4494  93.468 7.135e-11
spec           2 15.6884  7.8442  24.896 3.563e-07
propodus:spec  2  6.7015  3.3507  10.635 0.0003043
Residuals     31  9.7673  0.3151
```

The *Hemigrapsus nudus* crabs were removed from the data set to determine if the intercepts of the parallel relationships between the square root of closing force and propodus height differed between the *Cancer productus* and *Lophopanopeus bellus* crabs. These results confirmed that these two species have equal slopes ($p = 0.134$; Table 5.3) and that they have equal intercepts ($p = 0.494$; Table 5.3). A visual of the model fit is shown in Figure 5.9.

Conclusion

These results indicate that the relationship between square root closing force and propodus height is the same for *Cancer productus* and *Lophopanopeus bellus* crabs. In these two species of crabs, the square root closing force increases as the propodus height increases. In contrast, there is no observable statistical relationship between square root closing force and propodus height for *Hemigrapsus nudus* crabs.

Table 5.2. Post hoc multiple comparisons of all pairs of slopes (top) and for comparison of each slope to zero (bottom) for the transformed crab claw data.

```
                  Multiple Slope Comparisons (using the 'holm' adjustment)
  comparison     diff   95% LCI   95% UCI p.unadj    p.adj
1      hn-cp -0.32405 -0.55558 -0.09253 0.00762 0.01524
2      lb-cp  0.20622 -0.04844  0.46088 0.10872 0.10872
3      lb-hn  0.53027  0.28919  0.77135 0.00009 0.00027


                     Slope Information (using the 'holm' adjustment)
  level  slopes  95% LCI 95% UCI p.unadj    p.adj
2    hn 0.04118 -0.11190 0.19427 0.58715 0.58715
1    cp 0.36524  0.19155 0.53893 0.00016 0.00032
3    lb 0.57145  0.38522 0.75769 0.00000 0.00000
```

Table 5.3. ANOVA results from fitting the ultimate full model on the transformed crab claw data for only the *Cancer productus* and *Lophopanopeus bellus* crabs.

```
              Df  Sum Sq Mean Sq F value      Pr(>F)
propodus       1 20.8363 20.8363 59.4969 2.869e-07
spec           1  0.1699  0.1699  0.4853    0.4945
propodus:spec  1  0.8594  0.8594  2.4540    0.1337
Residuals     19  6.6540  0.3502
```
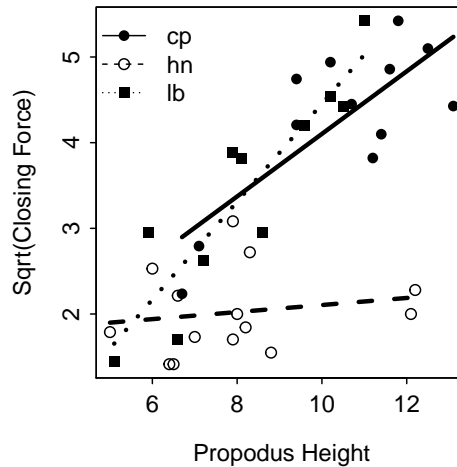

Figure 5.9. Fitted line plot from the fit of the ultimate model on the transformed data.

Appendix – R commands

```
> CrabClaw <- read.csv("CrabClaw.csv")
> crab.lm1 <- lm(force~propodus*spec,data=CrabClaw)
> transChooser(crab.lm1)
> CrabClaw$sqrt.force <- sqrt(CrabClaw$force)
> crab1.lm1 <- lm(sqrt.force~propodus*spec,data=CrabClaw)
> fitPlot(crab1.lm1,legend="topleft",xlab="Propodus Height",ylab="Sqrt(Closing Force)")
> anova(crab1.lm1)
```

```
> compSlopes(crab1.lm1)
> CrabClaw1 <- filterD(CrabClaw,spec!="hn")
> crab2.lm1 <- lm(sqrt.force~propodus*spec,data=CrabClaw1)
> anova(crab2.lm1)
```

5.7.2 Possum Morphometrics

Introduction

Many Australian mammals, from a wide range of taxonomic groups, exhibit variation in external morphology throughout their geographic distributions. A notable example is the common brushtail possum (*Trichosurus vulpecula*), which is characterized by marked changes in coat color and body size throughout its distribution in Australia. Variations in the closely related mountain brushtail possum (*Trichosurus caninus*) have been relatively unstudied. Lindenmayer *et al.* 1995 reported on the results of a study of the external morphology of the mountain brushtail possum from throughout its known geographic range. One aspect of their study was to determine if the relationship between the total length and foot length of a possum differed between possums captured near Victoria and those captured elsewhere.

Data Collection

Lindenmayer *et al.* (1995) examined morphometric measurements on each of 104 mountain brushtail possums (*Trichosurus caninus gilby*), trapped at seven sites from Southern Victoria to central Queensland in October and November 1993. Possums were trapped in large wire cage traps baited with apple. A total of 50 traps was used at each site. All animals captured were sedated with "Zoletil" to facilitate the collection of a range of morphometric measurements. Animals were released at the point of capture when recovery was complete. Several morphometric measurements were made but only the following variables will be used in this analysis:

- *pop*: a factor with levels *Vic* for capture near Victoria and *other* for capture near New South Wales or Queensland.
- *totlen*: measurement of the total length of the animal (cm).
- *footlen*: measurement of the length of the food (mm).

The *pop* variable will be "re-leveled" so that possums captured near Victoria will serve as the reference group. In addition, in the models, *pop* will be converted to an indicator variable of the form *popother* (=1 if not from Victoria, =0 if from Victoria) and an interaction variable between *footlen* and the indicator variable will be constructed. Individual 41 was removed from this analysis because the the foot length measurement was missing.

EDA & Assumption Checking

The initial ultimate full model fit to these data was

$$\mu_{totlen|footlen,pop} = \alpha + \beta footlen + \delta_1 popother + \gamma_1 popother * footlen$$

The fit of this model is largely linear and homoscedastic (Figure 5.10) with approximately normal residuals (Anderson-Darling $p = 0.207$) and no detectable outliers (outlier test $p = 0.113$). However, there are several points that "stand out" from the main cluster (Figure 5.10) – individual 31 had a total length much longer than would be expected given its foot length, individuals 39 and 43 each had total lengths much shorter

than would be expected given their foot lengths, and individual 41 had a foot length and total length shorter than most of the individuals from Victoria. It should be noted that all of these individuals are from Victoria and that three of the individuals are the shortest three possums recorded (in terms of total length). It is impossible to tell if any of these points are in error; however, it is possible that they have a heavy influence on the fitted line for Victoria. The combined influence of these points will be examined in the results.
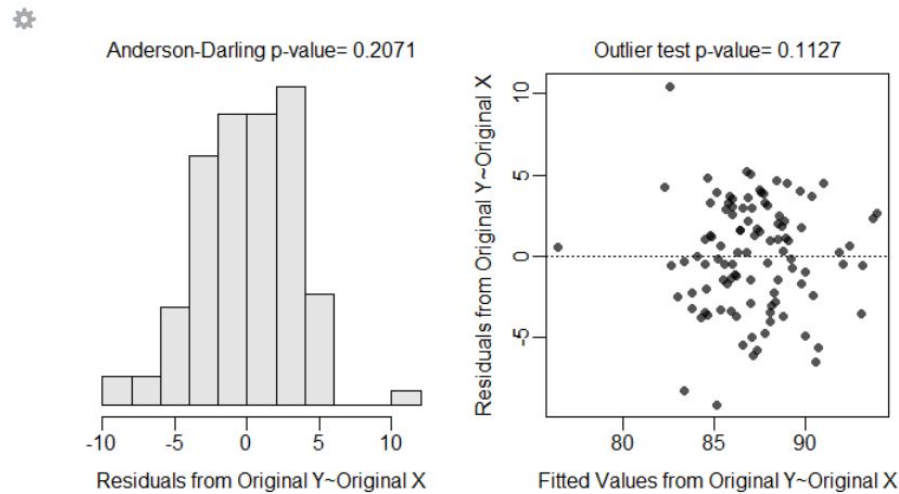

Figure 5.10. Histogram of residuals (Left) and residual plot (Right) for the regression of possum total length on foot length.

Results

Possums from Victoria and possums from other locations had similar slopes for the relationship between total length and foot length (parallel lines test $p = 0.323$; Table 5.4). The estimated common slope represents an average 0.74 to 1.24 cm increase in total length for a 1 mm increase in foot length. Assuming parallel lines, the relationship between total length and foot length for possums from Victoria and those from other areas had different intercepts (equal-intercepts test $p < 0.0005$; Table 5.4). Thus, the mean total length adjusted to a common foot length was between 3.92 and 8.34 cm less for possums from other locations than for possums from Victoria.

Table 5.4. ANOVA results from fitting the ultimate full model on the raw possum morphometric data.

```
            Df  Sum Sq Mean Sq F value     Pr(>F)
footlen      1  375.36  375.36  32.124 1.428e-07
pop          1  353.26  353.26  30.233 2.995e-07
footlen:pop  1   11.53   11.53   0.987    0.3229
Residuals   99 1156.79   11.68
```

The reduced model

$$\mu_{totlen|footlen,pop} = \alpha + \beta footlen + \delta_1 popother$$

can be used to predict the total length of an possum given a measurement of the possum's foot length. For example, predicted total lengths for possums from both locations with foot lengths of 65 and 70 mm are shown in Table 5.5.

Table 5.5. Predicted total lengths, with 95% prediction intervals, for the reduced model using all observations for possums from both locations with foot lengths of 65 and 70 mm.

```
footlen   pop  fit  lwr  upr
     65    Vic 80.3 73.2 87.4
     65  other 86.4 79.5 93.2
     70    Vic 85.2 78.3 92.1
     70  other 91.3 84.4 98.3
```

Removal of the four individuals identified in assumption checking did not change the overall result of a parallel relationship ($p = 0.407$) with unequal-intercepts ($p < 0.0005$; Figure 5.11). However, the common slope was estimated to be slightly (though likely not significantly) steeper at between 0.70 and 1.20 cm increase in total length for a 1 mm increase in foot length. In addition, the difference in adjusted mean total lengths decreased slightly (though likely not significantly) to between 4.36 and 6.82 cm. These results did not substantially alter the predictions for the 65 and 70 mm foot length individuals (Table 5.6).
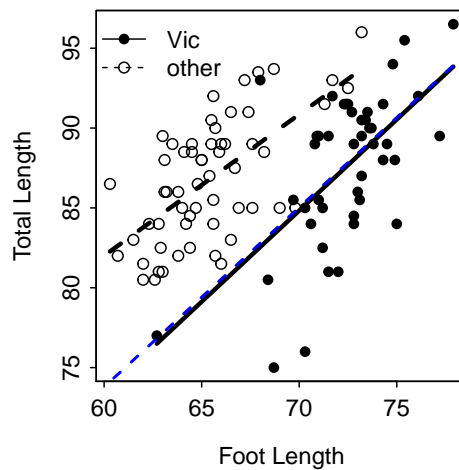

Figure 5.11. Fitted line plot from the fit of the ultimate model on the untransformed possum data with the four observation flagged in Figure ?? removed from the analysis. The dashed line represents the fitted line with the four individuals not removed.

Table 5.6. Predicted total lengths, with 95% prediction intervals, for the reduced model with four observations removed, for possums from both locations with foot lengths of 65 and 70 mm.

```
footlen   pop  fit  lwr  upr
     65    Vic 80.8 74.4 87.3
     65  other 86.4 80.3 92.5
     70    Vic 85.6 79.4 91.7
     70  other 91.1 84.9 97.4
```

Conclusion

A significant relationship between total length and foot length was observed for possums collected near Victoria and those collected elsewhere. The relationship was similar for possums collected at the two locations. However, the intercepts were significantly different indicating that possums from Victoria had a significantly

lower total length at every foot length. Thus, possums from Victoria seemed to be smaller (in total length) even after differences in foot length sizes between the two locations were considered.

The original data contained four possums from Victoria that appeared unusual relative to other observations. Three of these individuals had relatively small foot lengths and one individual had a much larger total length than would be expected given its foot length. Removal of these four individuals, however, did not substantially change the results of this study.

Appendix – R commands

```
> possum <- read.csv("possum1.csv")
> possum$pop <- factor(possum$pop,levels=c("Vic","other"))
> possum1 <- filterD(possum,!is.na(totlen),!is.na(footlen))
> op2.lm1 <- lm(totlen~footlen*pop,data=possum1)
> transChooser(op2.lm1)
> anova(op2.lm1)
> op2.lm2 <- lm(totlen~footlen+pop,data=possum1)
> compIntercepts(op2.lm2)
> pdf <- data.frame(footlen=c(65,65,70,70),pop=c("Vic","other","Vic","other"))
> predictionPlot(op2.lm2,pdf,interval="prediction")

> possum2 <- possum1[-c(31,39,41,43),]
> op3.lm1 <- lm(totlen~footlen*pop,data=possum2)
> anova(op3.lm1)
> op3.lm2 <- lm(totlen~footlen+pop)
> compIntercepts(op3.lm2)
> predictionPlot(op3.lm2,pdf,interval="prediction")
```

[Turn the Page]

5.8   Summary Process

The following is a template for a process of fitting a one-way IVR 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.
   - Pay close attention to the form, strength, and outliers on the scatterplot of the response and explanatory variables separated by each level of the factor [`coplot()`].[24]

2. Fit the untransformed ultimate full model [`lm()`].

3. Check the assumptions of the fit of the model [`transChooser()`].
   - Check the linearity of the relationship with residual plots.
   - Check homoscedasticity with residual plots.
   - Check normality of residuals with a Anderson-Darling test and histogram of residuals.
   - Check for outliers with the outlier test.

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 possible transformations for the response and explanatory variables.
   - If only an "unusual" or influential observation exists (i.e., linear, homoscedastic, and normal residuals) and no transformation corrects the "problem" then consider removing that observation from the data set.
   - Fit the ultimate full model with the transformed variables or reduced data set.

5. Construct an ANOVA table for the full model [`anova()`]. Interpret the F-test for the interaction between the covariate and the factor variable (i.e., the parallel lines test).
   - If the lines are NOT parallel then determine which slopes are different [`compSlopes()`]. SKIP to last step.
   - If the lines are parallel then continue with next step.

6. Interpret the F-test for the factor variable on the ANOVA table from the fit of the ultimate full model (i.e., equal-intercepts test).
   - If the intercepts are NOT equal then determine which mean adjusted values (i.e., intercepts) are different [`compIntercepts()`].
   - If the intercepts are equal (i.e., coincident lines) then continue with next step.

7. Interpret the F-test for the covariate on the ANOVA table from the fit of the ultimate full model to determine if there is a significant relationship between the response and the covariate.

8. Summarize findings with a fitted-line plot.
   - Use the ultimate full model to construct the fitted-line plot but comment in your conclusion about significance of differences in slopes and intercepts.

[**Turn the Page**]

---

[24]The use of the `coplot()` function was not discussed in the module. It's use for this purpose is rather simple and the output is fairly intuitive. The `coplot()` function requires a model formula of the form `response∼covariate | factor`. The | symbol means "conditioned on" which can be interpreted as "separated by."

5.9 ANCOVA

Analysis of covariance (ANCOVA) is a term that is used to describe linear models that contain both factor and quantitative explanatory variables (Fox 1997). Traditionally, the term ANCOVA has been restricted to the situation where it is assumed that the regression of the response on the explanatory variable has the same slope among levels of the factor variable. With this definition, the ANCOVA model is a subset of the more general IVR models constructed in the previous sections of this module. You should be aware when reading the literature because some authors use ANCOVA in the same way that IVR is used in this book. However, in this book, an IVR will refer to all models that have both factor and quantitative explanatory models and ANCOVA will refer to all IVR models where it is assumed that the slopes are equal.[25]

> Δ **ANCOVA model**: A subset of IVR models where the submodels have equal slopes.

The derivation of the name – analysis of covariance – seems to stem from the fact that the quantitative explanatory variable is often called a *covariate* and that the analysis is closely related to ANOVA. In the latter regard, it should be noted that an ANCOVA is used to determine if the mean of a quantitative response variable differs among two or more groups defined by a factor variable. However, an ANCOVA tests for the difference in means of the response variable among the groups AFTER adjusting for the relationship between the response variable and the covariate.

> Δ **Covariate**: The quantitative explanatory variable in an IVR model.

> ◇ **An ANCOVA tests for difference in means of the response variable among levels of the factor variable after adjusting for the relationship between the response variable and the covariate.**

The process of an ANCOVA consists of two steps already discussed in previous sections. First, it must be shown with a parallel lines test (see Section 5.4.2) that the slopes for all groups are equal. If not, a strictly defined ANCOVA cannot be completed. Second, if it is shown that the groups have parallel lines, then an equal-intercepts test (see Section 5.4.2) is used to determine if the level means of the response variable adjusted for the covariate are all equal or not. If not, then a Tukey's HSD test is conducted to determine which adjusted means are significantly different. The traditional ANCOVA uses means of the response variable that are adjusted to the mean of the covariate.[26] The possum example above is an example of an ANCOVA because it was shown that the lines had equal slopes, whereas the crab example above is not an example of an ANCOVA because the slopes were unequal.

> ◇ **A traditional ANCOVA requires that all groups in the data have equal slopes.**

> ◇ **A traditional ANCOVA adjusts the values of the response variable to the mean of the covariate.**

The ANCOVA works on the premise that if the relationship between the response variable and the covariate

---

[25]That is, the model does not contain an interaction term among indicator variables and the quantitative explanatory variable.

[26]The value that the response variables are adjusted to is not important when looking at the differences. However, it may be important when presenting summaries (e.g., the mean) of the adjusted values. To adjust the values to the mean of the covariate the `common.cov=` argument in the `compIntercepts()` function should be set to `mean(x)`, where x is replaced with the name of the covariate.

is represented by parallel lines for the groups, then the difference between the groups adjusted for this relationship is represented by the difference in intercepts between the groups. In other words, if the groups have parallel lines but different intercepts, then the difference in the intercepts is a constant difference at all values of the covariate; thus, the difference in intercepts is the difference in the value of the response variable adjusted to any value of the covariate.

5.9.1 ANCOVA in ecology

An ANCOVA may be appropriate in any situation where the relationship between a quantitative explanatory variable and the quantitative response variable is thought to increase the variance of the response variable and, thus, mask real differences among the groups or to cause a separation among the groups that is not due to the groups themselves. For example, researchers studying the effect of a treatment on the weight gain (i.e., response) of an animal may be worried that the initial size of the animals (i.e., covariate) results in an increase in variance of the weight gain. In another example, researchers examining the number of eggs produced by a species of fish (i.e., response) in two lakes may be concerned that the observed difference in number of eggs is a result of one lake having larger fish (i.e., size is the covariate) than the other lake, rather than a "real" difference in number of eggs between the lakes. Concerns such as these are so common in ecological and biological research that many researchers will compute a ratio of the response variable to the covariate[27] or construct residuals from a common line regression[28] before using a one-way ANOVA to test for differences among the treatments or groups. The fallacy of creating these ratios and the advantage of properly using an ANCOVA (or an IVR in general) is illustrated in this section with these two hypothetical situations.

Hypothetical Problem I

Suppose that a researcher was interested in the effect of two temperatures on the mean weight gain of a particular type of fish. Specifically, this researcher tested the following alternative hypothesis: $H_A : \mu_8 \neq \mu_{10}$, where the subscripts represent temperatures. The researcher measured the weight of each fish before the experiment began (hereafter, called initial body weight) so that fish of the same general size could be "paired." One of each "pair" of fish was then randomly allocated to the 8^oC treatment and the other was allocated to the 10^oC treatment. There were nine fish in each treatment for a total of 18 fish in the experiment. After a prescribed period of time the weight gain of each fish was recorded (i.e., difference in weight from the beginning to the end of the experimental period).

It may seem reasonable that the effect of initial weight was "handled" by pairing fish of comparable size at the beginning of the experiment. If this is believed, then the analyst would likely proceed with a one-way ANOVA.[29] A means plot (Figure 5.12-Left) and the one-way ANOVA (Table 5.7) both indicate that there is no significant difference in weight gain between the two treatments. However, if weight gain is plotted against initial weight (Figure 5.12-Right), then it is immediately seen that the fish in the 8^oC treatment always gained more weight then the paired fish at the 10^oC treatment. Thus, the conclusion of no effect of temperature from the initial ANOVA analysis does not fit the intuitive conclusion made from examining the data in Figure 5.12-Right.

---

[27]The use of ratios will be discussed in subsequent examples. However, the use of ratios has been criticized or the use of ANCOVA is superior to ratios by LeCren (1951), Atchley *et al.* (1976), Packard and Boardman (1988), Jackson *et al.* (1990), Raubenheimer and Simpson (1992), Albrecht *et al.* (1993), Garcia-Berthou *et al.* (1993), Raubenheimer (1995), and Beaupre and Dunham (1995).

[28]The use of residuals was popularized by Jakob (1996) but has been convincingly criticized by Maxwell *et al.* (1985), Hayes and Shonkwiler (1996), Smith (1999), and Garcia-Berthou (2001).

[29]A two-sample t-test would also work in this situation because there are only two treatments. Recall, though, that it was shown in the Module 2 that a one-way ANOVA for two groups is the same as a two-sample t-test.

Table 5.7. One-way ANOVA results for mean weight gain by temperature treatment.

```
          Df Sum Sq Mean Sq F value Pr(>F)
group      1  14.04  14.045  0.4304 0.5211
Residuals 16 522.12  32.632
```
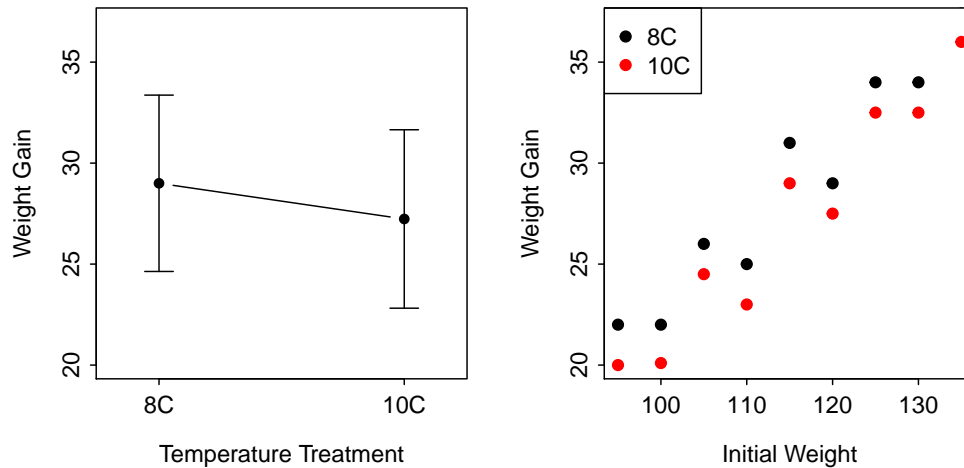

Figure 5.12. Means plot for the mean weight gain by temperature treatment (Left) and scatterplot of weight gain versus initial weight by temperature treatment (Right).

The results from fitting the model

$$\mu_{wt.gain|\cdots} = \alpha + \beta wt.init + \delta_1 group10C + \gamma_1 group10C * wt.init$$

showed that the regression lines were parallel (parallel lines test p=0.9235) but the intercepts were not equal (equal-intercepts test p=0.0333; Figure 5.13-Left). Tukey's HSD test estimated that the intercept for fish in the 10^oC treatment was 1.77 g lower than the intercept for fish in the 8^oC treatment with a 95% confidence interval of 0.28 to 3.25 g. Because the slopes are equal, one can also conclude that the adjusted means, i.e., the mean weight gain at the mean initial weight, are different by the same amount (Figure 5.13-Right).

This example illustrates how the inclusion of a "strong" covariate can mask the effects of a significant but "weak" factor variable. In the initial analysis using just a one-way ANOVA on weight gain, the increased variability due to the initial weight of the fish was not acknowledged and was thus left in the "unexplained" residual variability (i.e., $MS_{Residual} = 32.63$; Table 5.7). The variability explained by the temperature treatment ($MS_{group} = 14.04$; Table 5.7) was very small relative to the unexplained variability and, thus, the temperature treatment was not considered significant. However, when this large source of variability was explained by including the covariate in the ANCOVA, then the "unexplained" residual variability decreased dramatically (i.e., $MS_{Residual} = 2.35$; Table 5.8). The variability explained by the temperature treatment remains the same but it now "appears" very large because the amount of variability unexplained is relatively small. Thus, the temperature treatment becomes significant after the effect of initial weight on weight gain has been removed because the unexplained variability was significantly reduced. In other words, the use of the ANCOVA helped identify a significant treatment effect that was masked by the increased variability due to the relationship between the response variable and a (previously ignored) covariate. The reduction in residual variability can also be seen by the relatively "tight" confidence intervals for the adjusted means in Figure 5.13-Right relative to those for the unadjusted means in Figure 5.12-Left.
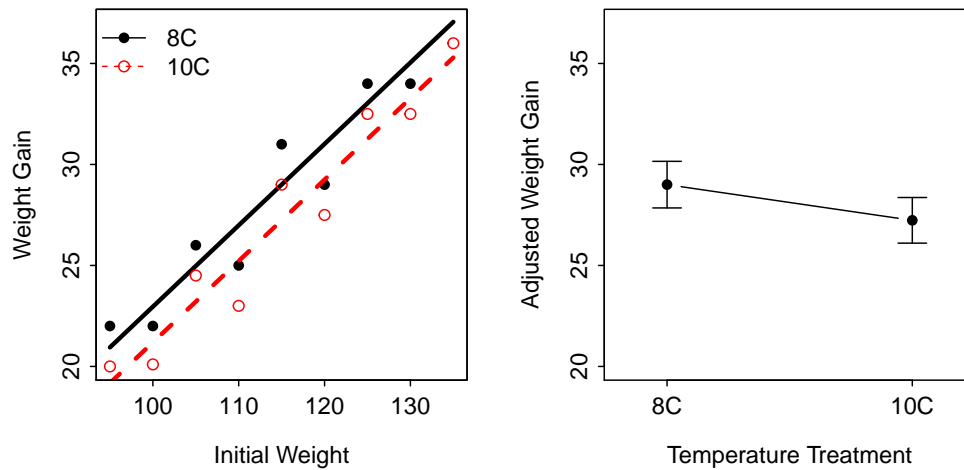
Figure 5.13. Fitted line plot for the regression of weight gain on initial weight by temperature treatment (Left) and means plot for the adjusted mean weight gain (at the mean initial weight) by temperature treatment (Right).

Table 5.8. One-way ANCOVA results for mean weight gain adjusted by initial weight by temperature treatment.

```
          Df Sum Sq Mean Sq  F value    Pr(>F)
wt.init    1 486.82  486.82 206.8897 3.505e-10
group      1  14.05   14.05   5.9688   0.02741
Residuals 15  35.30    2.35
```

> ◇ **An ANCOVA may increase the power to detect a difference among treatment means if a covariate is strongly related to the response variable. In these instances a significant difference in treatment adjusted means may be detected when a difference in unadjusted treatment means was not detected.**

As mentioned previously, some researchers will attempt to remove the effect of the initial weight of the fish by computing a ratio of weight gain to initial weight. This ratio can be thought of as a proportional growth measurement. Then, to determine if there is a difference in proportional growth between the two treatments, the researchers will continue, under the assumption that the effect of the initial weight of the fish has been removed, with a one-way ANOVA on the proportional growth measurement. However, a scatterplot of proportional growth versus initial weight shows that there is STILL a strong relationship between these two variables (Figure 5.14). Thus, the "effect" of initial weight has NOT been removed from the weight gain variable. In addition, a one-way ANOVA on the proportional weight variable did not detect a difference in the mean of this ratio between the two treatments (p=0.1677) despite the rather obvious differences indicated in the scatterplot (i.e., the values for all fish kept at 8^oC are higher than fish of the same initial size kept at 10^oC). It appears then, that this ratio did not completely remove the effect of the covariate and, thus, did not allow detection of the significant treatment effect present in the data.

> ◇ **Contrived response variables that consists of the ratio between an observed response variable and a covariate do not completely remove the effect of the covariate. An ANCOVA model should be used instead.**
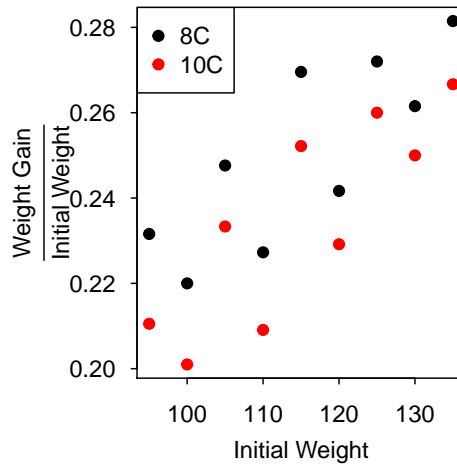
Figure 5.14. Scatterplot of the ratio of the gain in body weight to the initial body weight versus initial body weight.

Hypothetical Problem II

In this example, suppose that a researcher was interested in determining if the gonad weight (g) differed between the same species of fish from two different locations. In particular, the researcher tested the following statistical hypotheses: $H_A : \mu_A \neq \mu_B$ where the subscripts represent locations. The researcher measured the body weight with the gonads removed (called somatic weight) and the gonad weight of nine randomly selected fish from each location.

A one-way ANOVA provides strong evidence for a difference in gonad weight between the two lakes (Table 5.9; Figure 5.15-Left). However, a plot of gonad weight versus somatic weight (Figure 5.15-Right) indicates that there is also a very strong difference in somatic weight between the two locations; indeed every fish at location B is larger then every fish at location A. It is also clear from this plot that there is a very strong relationship between gonad weight and somatic weight. Thus, it may be possible that the observed difference in gonad weight is mostly due to the differences in the somatic weight covariate.

Table 5.9. One-way ANOVA results for mean gonad weight by location.

```
          Df Sum Sq Mean Sq F value    Pr(>F)
loc        1 1378.1 1378.12  45.511 4.696e-06
Residuals 16  484.5   30.28
```

The results from fitting the model

$$\mu_{wt.gonad|\cdots} = \alpha + \beta wt.somat + \delta_1 locB + \gamma_1 wt.somat * locB$$

showed that the submodel regression lines were parallel (parallel lines test p=0.8486) and the intercepts were equal (equal-intercepts test p=0.9309). Thus, in this example, the two locations are fit by the same regression line (Figure 5.16-Left). This result also indicates that the mean gonad weight adjusted to a common somatic weight does NOT differ between the two locations (Figure 5.16-Right). The difference observed in the observed gonad weights WAS due to the difference in somatic weights and the strong relationship between gonad weight and somatic weight.

This example illustrates how the inclusion of a "strong" covariate and a significant difference in the covariate among groups can make the "weak" effect of a factor variable appear significant. In the case of the one-
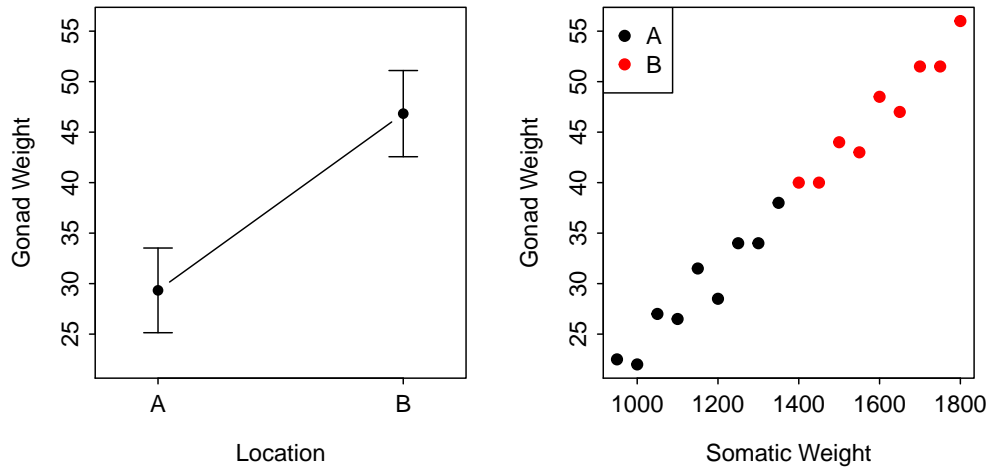
Figure 5.15. Means plot for the mean gonad weight by location (Left) and scatterplot of gonad weight versus somatic weight by location (Right).
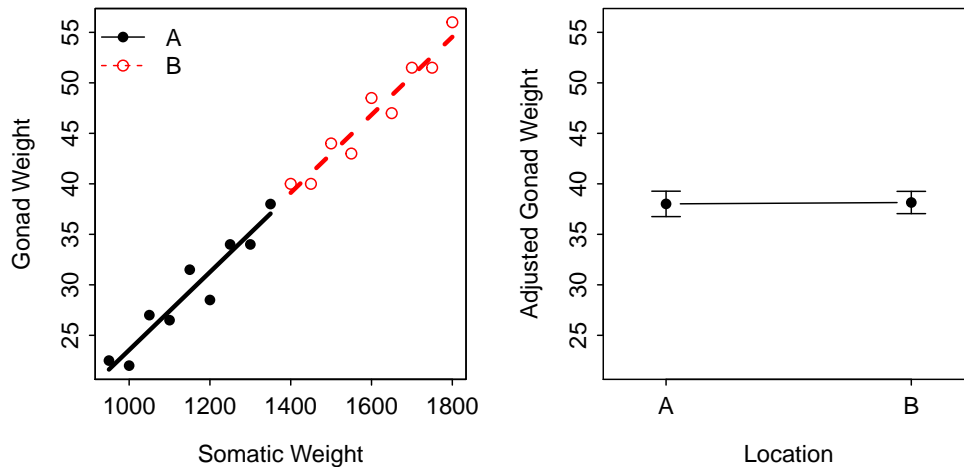

Figure 5.16. Fitted line plot for the regression of gonad weight on somatic weight by location (Left) and means plot for the adjusted mean gonad gain (at the mean somatic weight) by location (Right).

way ANOVA analysis, the variability that can be explained by the covariate was included in both the "explained" variability of the *loc* variable and the "unexplained" residual variability. The portion included in the *loc* variable was very large making that variable appear very significant. However, when the covariate was included in the ANCOVA analysis it was revealed that the SS due to the *loc* variable was negligible (=0.02; Table 5.10). Thus, nearly all of the SS attributable to *loc* in the one-way ANOVA results was really attributable to the covariate of somatic weight. In other words, the use of the ANCOVA in this example helped properly attribute the cause of explained variability.

Table 5.10. One-way ANCOVA results for mean gonad gain adjusted for somatic weight by location.

```
          Df  Sum Sq Mean Sq  F value    Pr(>F)
wt.somat   1 1824.71 1824.71 722.2189  4.21e-14
loc        1    0.02    0.02   0.0083    0.9284
Residuals 15   37.90    2.53
```

> ◇ **An ANCOVA may remove the effect of a covariate on a response variable that is confounding the effect of the factor on the response variable. In these instances a significant difference in unadjusted treatment means may disappear when analyzing adjusted treatment means.**

This example also illustrates how the inclusion of a "strong" covariate substantially reduces the residual variance. This is illustrated by the substantial reduction in $MS_{Residual}$ from Table 5.9 to Table 5.10 and the marked decrease in the width of confidence intervals from Figure 5.15-Left to Figure 5.16-Right.

In the fisheries literature it is very common to construct a contrived response variable called the gonadosomatic index (GSI). The GSI is the ratio of gonad weight to somatic weight and is an attempt to remove the effect of somatic weight on gonad weight by putting the gonad weight on a "per somatic weight" basis. A one-way ANOVA of the GSI variable, however, indicated that there was still a significant difference in the mean GSI between the two locations (p=0.0002). However, a scatterplot of the GSI versus somatic weight shows that there is still a strong relationship between GSI and somatic weight (Figure 5.17). In other words, once again, the effect of the covariate, somatic weight, on the response variable, gonad weight, has not been effectively factored out or removed.
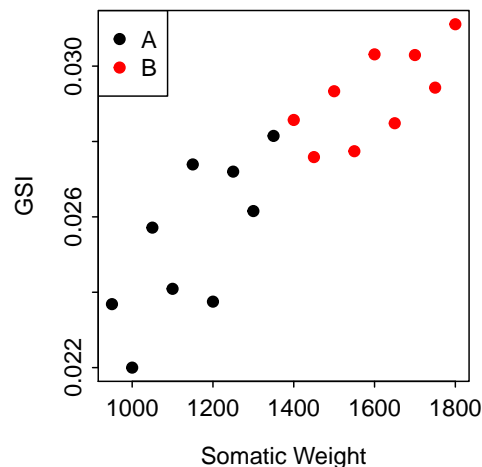

Figure 5.17. Scatterplot of gonadosomatic index (GSI) versus somatic weight by location.

Advantages of ANCOVA

The three major advantages of ANCOVA were illustrated in the two hypothetical examples above. First, ANCOVA, in contrast to ratios of the response to the covariate, effectively removes the relationship between the response and the covariate. Second, the removal of this relationship, if a relationship exists, will reduce the residual error of the model. In other words, more of the variability inherent in the response variable is

explained. This increases power and the likelihood of detecting a difference among treatments if a treatment really exists. Third, the removal of the relationship between the covariate and the response allows the researcher to compare groups on an "equal footing." In other words, the differences between the groups can be compared at a constant value of the covariate.

A main thing to remember about ANCOVA is that it is a very powerful method that is really only a subset of the models that have formed the bulk of this module. An ANCOVA refers to the situation of describing the difference in intercepts when the slopes are equal. The model fitting that formed the bulk of this module is not restricted to the case of equal slopes.