

Chapter 10

Analysis of Covariance

An analysis procedure for looking at group effects on a continuous outcome when some other continuous explanatory variable also has an effect on the outcome.

This chapter introduces several new important concepts including multiple regression, interaction, and use of indicator variables, then uses them to present a model appropriate for the setting of a quantitative outcome, and two explanatory variables, one categorical and one quantitative. Generally the main interest is in the effects of the categorical variable, and the quantitative explanatory variable is considered to be a “control” variable, such that power is improved if its value is controlled for. Using the principles explained here, it is relatively easy to extend the ideas to additional categorical and quantitative explanatory variables.

The term ANCOVA, analysis of covariance, is commonly used in this setting, although there is some variation in how the term is used. In some sense ANCOVA is a blending of ANOVA and regression.

10.1 Multiple regression

Before you can understand ANCOVA, you need to understand multiple regression. Multiple regression is a straightforward extension of simple regression from one to several quantitative explanatory variables (and also categorical variables as we will see in the section 10.4). For example, if we vary water, sunlight, and fertilizer to see their effects on plant growth, we have three quantitative explanatory variables.

In this case we write the structural model as

$$E(Y|x_1, x_2, x_3) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3.$$

Remember that $E(Y|x_1, x_2, x_3)$ is read as expected (i.e., average) value of Y (the outcome) given the values of the explanatory variables x_1 through x_3 . Here, x_1 is the amount of water, x_2 is the amount of sunlight, x_3 is the amount of fertilizer, β_0 is the intercept, and the other β s are all slopes. Of course we can have any number of explanatory variables as long as we have one β parameter corresponding to each explanatory variable.

Although the use of numeric subscripts for the different explanatory variables (x 's) and parameters (β 's) is quite common, I think that it is usually nicer to use meaningful mnemonic letters for the explanatory variables and corresponding text subscripts for the parameters to remove the necessity of remembering which number goes with which explanatory variable. Unless referring to variables in a completely generic way, I will avoid using numeric subscripts here (except for using β_0 to refer to the intercept). So the above structural equation is better written as

$$E(Y|W, S, F) = \beta_0 + \beta_W W + \beta_S S + \beta_F F.$$

In multiple regression, we still make the fixed- x assumption which indicates that each of the quantitative explanatory variables is measured with little or no imprecision. All of the error model assumptions also apply. These assumptions state that for all subjects that have the same levels of all explanatory variables the outcome is Normally distributed around the true mean (or that the errors are Normally distributed with mean zero), and that the variance, σ^2 , of the outcome around the true mean (or of the errors) is the same for every other set of values of the explanatory variables. And we assume that the errors are independent of each other.

Let's examine what the (no-interaction) multiple regression structural model is claiming, i.e., in what situations it might be plausible. By examining the equation for the multiple regression structural model you can see that the meaning of each slope coefficient is that it is the change in the mean outcome associated with (or caused by) a one-unit rise in the corresponding explanatory variable *when all of the other explanatory variables are held constant*.

We can see this by taking the approach of writing down the structural model equation then making it reflect specific cases. Here is how we find what happens to

the mean outcome when x_1 is fixed at, say 5, and x_2 at, say 10, and x_3 is allowed to vary.

$$\begin{aligned}E(Y|x_1, x_2, x_3) &= \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \\E(Y|x_1 = 5, x_2 = 10, x_3) &= \beta_0 + 5\beta_1 + 10\beta_2 + \beta_3x_3 \\E(Y|x_1 = 5, x_2 = 10, x_3) &= (\beta_0 + 5\beta_1 + 10\beta_2) + \beta_3x_3\end{aligned}$$

Because the β s are fixed (but unknown) constants, this equation tells us that when x_1 and x_2 are fixed at the specified values, the relationship between $E(Y)$ and x_3 can be represented on a plot with the outcome on the y-axis and x_3 on the x-axis as a straight line with slope β_3 and intercept equal to the number $\beta_0 + 5\beta_1 + 10\beta_2$. Similarly, we get the same slope with respect to x_3 for any combination of x_1 and x_2 , and this idea extends to changing any one explanatory variable when the others are held fixed.

From simplifying the structural model to specific cases we learn that the no-interaction multiple regression model claims that not only is there a linear relationship between $E(Y)$ and any x when the other x 's are held constant, it also implies that the effect of a given change in an x value does not depend on what the values of the other x variables are set to, as long as they are held constant. These relationships must be plausible in any given situation for the no-interaction multiple regression model to be considered. Some of these restrictions can be relaxed by including interactions (see below).

It is important to notice that the concept of changing the value of one explanatory variable while holding the others constant is meaningful in experiments, but generally not meaningful in observational studies. Therefore, interpretation of the slope coefficients in observational studies is fraught with difficulties and the potential for misrepresentation.

Multiple regression can occur in the experimental setting with two or more continuous explanatory variables, but it is perhaps more common to see one manipulated explanatory variable and one or more observed control variables. In that setting, inclusion of the control variables increases power, while the primary interpretation is focused on the experimental treatment variable. Control variables function in the same way as blocking variables (see 8.5) in that they affect the outcome but are not of primary interest, and for any specific value of the control variable, the variability in outcome associated with each value of the main experimental explanatory variable is reduced. Examples of control variables for many

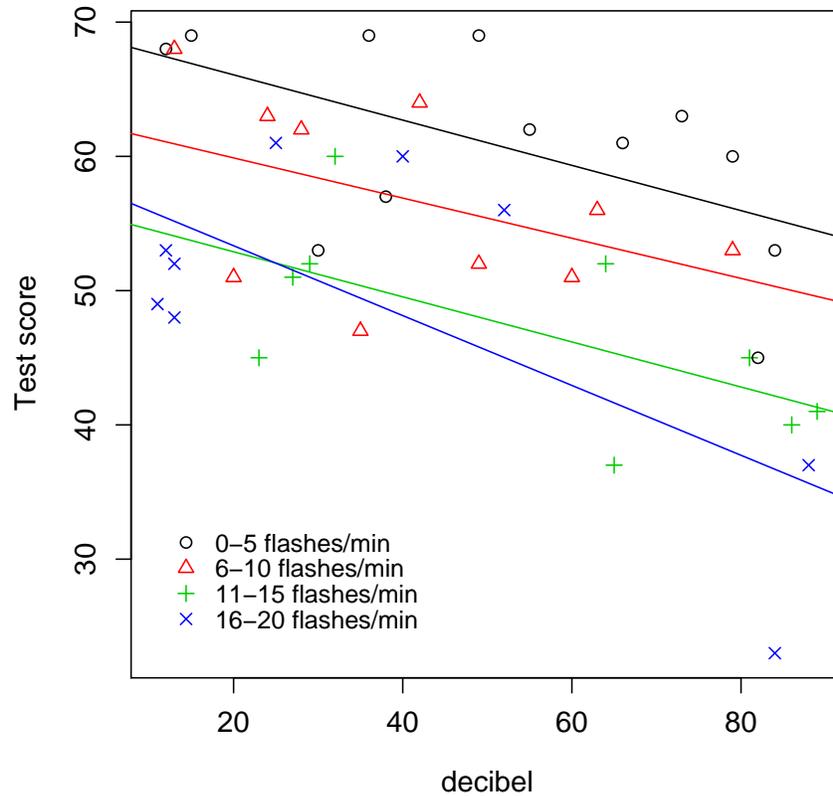


Figure 10.1: EDA for the distraction example.

psychological studies include things like ability (as determined by some auxiliary information) and age.

As an example of multiple regression with two manipulated quantitative variables, consider an analysis of the data of [MRdistract.dat](#) which is from a (fake) experiment testing the effects of both visual and auditory distractions on reading comprehension. The outcome is a reading comprehension test score administered after each subject reads an article in a room with various distractions. The test is scored from 0 to 100 with 100 being best. The subjects are exposed to auditory distractions that consist of recorded construction noise with the volume randomly set to vary between 10 and 90 decibels from subject to subject. The visual distraction is a flashing light at a fixed intensity but with frequency randomly set to between 1 and 20 times per minute.

	Unstandardized Coefficients		t	Sig.	95% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
(Constant)	74.688	3.260	22.910	<0.0005	68.083	81.294
db	-0.200	0.043	-4.695	<0.0005	-0.286	-0.114
freq	-1.118	0.208	-5.38	<0.0005	-1.539	-0.697

Table 10.1: Regression results for distraction experiment.

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.744	0.553	0.529	6.939

Table 10.2: Distraction experiment model summary.

Exploratory data analysis is difficult in the multiple regression setting because we need more than a two dimensional graph. For two explanatory variables and one outcome variable, programs like SPSS have a 3-dimensional plot (in SPSS try Graphs/ChartBuilder and choose the “Simple 3-D Scatter” template in the Scatter/Dot gallery; double click on the resulting plot and click the “Rotating 3-D Plot” toolbar button to make it “live” which allows you to rotate the plot so as to view it at different angles). For more than two explanatory variables, things get even more difficult. One approach that can help, but has some limitations, is to plot the outcome separately against each explanatory variable. For two explanatory variables, one variable can be temporarily demoted to categories (e.g., using the visual bander in SPSS), and then a plot like figure 10.1 is produced. Simple regression fit lines are added for each category. Here we can see that increasing the value of either explanatory variable tends to reduce the mean outcome. Although the fit lines are not parallel, with a little practice you will be able to see that given the uncertainty in setting their slopes from the data, they are actually consistent with parallel lines, which is an indication that no interaction is needed (see below for details).

The multiple regression results are shown in tables 10.1 10.2, and 10.3.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	22202.3	2	1101.1	22.9	<0.0005
Residual	1781.6	37	48.152		
Total	3983.9	39			

Table 10.3: Distraction experiment ANOVA.

Really important fact: There is an one-to-one relationship between the coefficients in the multiple regression output and the model equation for the mean of Y given the x's. There is exactly one term in the equation for each line in the coefficients table.

Here is an interpretation of the analysis of this experiment. (Computer reported numbers are rounded to a smaller, more reasonable number of decimal places – usually 3 significant figures.) A multiple regression analysis (additive model, i.e., with no interaction) was performed using sound distraction volume in decibels and visual distraction frequency in flashes per minute as explanatory variables, and test score as the outcome. Changes in both distraction types cause a statistically significant reduction in test scores. For each 10 db increase in noise level, the test score drops by 2.00 points ($p < 0.0005$, 95% CI=[1.14, 2.86]) at any fixed visual distraction level. For each per minute increase in the visual distraction blink rate, the test score drops by 1.12 points ($p < 0.0005$, 95% CI=[0.70, 1.54]) at any fixed auditory distraction value. About 53% of the variability in test scores is accounted for by taking the values of the two distractions into account. (This comes from adjusted R^2 .) The estimate of the standard deviation of test scores for any fixed combination of sound and light distraction is 6.9 points.

The validity of these conclusions is confirmed by the following assumption checks. The quantile-normal plot of the residuals confirms Normality of errors, the residual vs. fit plot confirms linearity and equal variance. (Subject 32 is a mild outlier with standardized residual of -2.3). The fixed-x assumption is met because the values of the distractions are precisely set by the experimenter. The independent errors assumption is met because separate subjects are used for each test, and the subjects were not allowed to collaborate.

It is also a good idea to further confirm linearity for each explanatory variable

with plots of each explanatory variable vs. the residuals. Those plots also look OK here.

One additional test should be performed before accepting the model and analysis discussed above for these data. We should test the “additivity” assumption which says that the effect (on the outcome) of a one-unit rise of one explanatory variable is the same at *every* fixed value of the other variable (and vice versa). The violation of this assumption usually takes the form of “interaction” which is the topic of the next section. The test needed is the p-value for the interaction term of a separate multiple regression model run with an interaction term.

One new interpretation is for the p-value of <0.0005 for the F statistic of 22.9 in the ANOVA table for the multiple regression. The p-value is for the null hypothesis that *all* of the slope parameters, but not the intercept parameter, are equal to zero. So for this experiment we reject $H_0 : \beta_V = \beta_A = 0$ (or better yet, $H_0 : \beta_{visual} = \beta_{auditory} = 0$)

Multiple regression is a direct extension of simple regression to multiple explanatory variables. Each new explanatory variable adds one term to the structural model.

10.2 Interaction

Interaction is a major concept in statistics that applies whenever there are two or more explanatory variables. Interaction is said to exist between two or more explanatory variables in their effect on an outcome. *Interaction is never between an explanatory variable and an outcome, or between levels of a single explanatory variable.* The term interaction applies to both quantitative and categorical explanatory variables. The definition of interaction is that the effect of a change in the level or value of one explanatory variable on the mean outcome *depends* on the level or value of another explanatory variable. Therefore interaction relates to the structural part of a statistical model.

In the absence of interaction, the effect *on the outcome* of any specific *change* in one explanatory variable, e.g., a one unit rise in a quantitative variable or a change from, e.g., level 3 to level 1 of a categorical variable, does not depend on

Setting	x_S	x_L	$E(Y)$	difference from baseline
1	2	4	$100-5(2)-3(4)=78$	
2	3	4	$100-5(3)-3(4)=73$	-5
3	2	6	$100-5(2)-3(6)=72$	-6
4	3	6	$100-5(3)-3(6)=67$	-11

Table 10.4: Demonstration of the additivity of $E(Y) = 100 - 5x_S - 3x_L$.

the level or value of the other explanatory variable(s), as long as they are held constant. This also tells us that, e.g., the effect on the outcome of changing from level 1 of explanatory variable 1 and level 3 of explanatory variable 2 to level 4 of explanatory variable 1 and level 2 of explanatory variable 2 is equal to the sum of the effects on the outcome of only changing variable 1 from level 1 to 4 plus the effect of only changing variable 2 from level 3 to 1. For this reason the lack of an interaction is called **additivity**. The distraction example of the previous section is an example of a multiple regression model for which additivity holds (and therefore there is no interaction of the two explanatory variables in their effects on the outcome).

A mathematic example may make this more clear. Consider a model with quantitative explanatory variables “decibels of distracting sound” and “frequency of light flashing”, represented by x_S and x_L respectively. Imagine that the parameters are actually known, so that we can use numbers instead of symbols for this example. The structural model demonstrated here is $E(Y) = 100 - 5x_S - 3x_L$. Sample calculations are shown in Table 10.4. Line 1 shows the arbitrary starting values $x_S = 2$, $x_L = 4$. The mean outcome is 78, which we can call the “baseline” for these calculations. If we leave the light level the same and change the sound to 3 (setting 2), the mean outcome drops by 5. If we return to $x_S = 2$, but change x_L to 6 (setting 3), then the mean outcome drops by 6. Because this is a non-interactive, i.e., additive, model we expect that the effect of simultaneously changing x_S from 2 to 3 and x_L from 4 to 6 will be a drop of $5+6=11$. As shown for setting 4, this is indeed so. This would not be true in a model with interaction.

Note that the component explanatory variables of an interaction and the lines containing these individual explanatory variables in the coefficient table of the multiple regression output, are referred to as **main effects**. In the presence of an interaction, when the signs of the coefficient estimates of the main effects are the

same, we use the term **synergy** if the interaction coefficient has the same sign. This indicates a “super-additive” effect, where the whole is more than the sum of the parts. If the interaction coefficient has opposite sign to the main effects, we use the term **antagonism** to indicate a “sub-additive” effects where simultaneous changes in both explanatory variables has less effect than the sum of the individual effects.

The key to understanding the concept of interaction, how to put it into a structural model, and how to interpret it, is to understand the construction of one or more new interaction variables from the existing explanatory variables. An interaction variable is created as the product of two (or more) explanatory variables. That is why some programs and textbooks use the notation “A*B” to refer to the interaction of explanatory variables A and B. Some other programs and textbooks use “A:B”. Some computer programs can automatically create interaction variables, and some require you to create them. (You can always create them yourself, even if the program has a mechanism for automatic creation.) Peculiarly, SPSS has the automatic mechanism for some types of analyses but not others.

The creation, use, and interpretation of interaction variables for two quantitative explanatory variables is discussed next. The extension to more than two variables is analogous but more complex. Interactions that include a categorical variable are discussed in the next section.

Consider an example of an experiment testing the effects of the dose of a drug (in mg) on the induction of lethargy in rats as measured by number of minutes that the rat spends resting or sleeping in a 4 hour period. Rats of different ages are used and age (in months) is used as a control variable. Data for this (fake) experiment are found in [lethargy.dat](#).

Figure 10.2 shows some EDA. Here the control variable, age, is again categorized, and regression fit lines are added to the plot for each level of the age categories. (Further analysis uses the complete, quantitative version of the age variable.) What you should see here is that the slope appears to change as the control variable changes. It looks like more drug causes more lethargy, and older rats are more lethargic at any dose. But what suggests interaction here is that the three fit lines are *not* parallel, so we get the (correct) impression that the effect of any dose increase on lethargy is *stronger* in old rats than in young rats.

In multiple regression with interaction we add the new (product) interaction variable(s) as additional explanatory variables. For the case with two explanatory

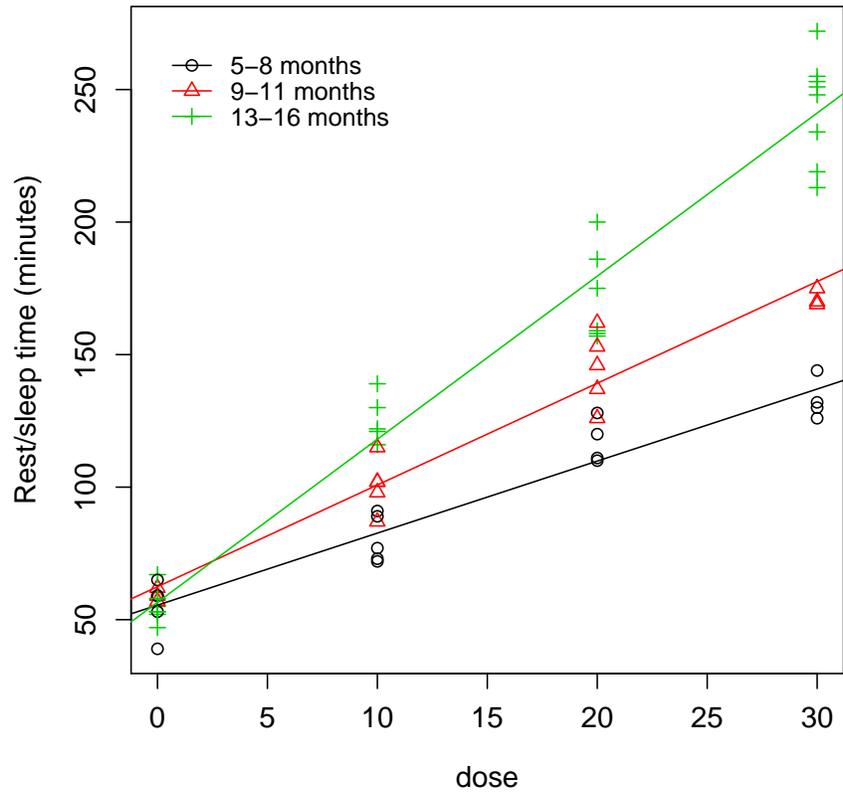


Figure 10.2: EDA for the lethargy example.

variable, this becomes

$$E(Y|x_1, x_2) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}(x_1 \cdot x_2)$$

where β_{12} is the single parameter that represents the interaction effect and $(x_1 \cdot x_2)$ can either be thought of as the single new interaction variable (data column) or as the product of the two individual explanatory variables.

Let's examine what the multiple regression with interaction model is claiming, i.e., in what situations it might be plausible. By examining the equation for the structural model you can see that the effect of a one unit change in either explanatory variable *depends* on the value of the other explanatory variable.

We can understand the details by taking the approach of writing down the model equation then making it reflect specific cases. Here, we use more meaningful variable names and parameter subscripts. Specifically, β_{d^*a} is the symbol for the single interaction parameter.

$$\begin{aligned} E(Y|\text{dose}, \text{age}) &= \beta_0 + \beta_{\text{dose}}\text{dose} + \beta_{\text{age}}\text{age} + \beta_{d^*a}\text{dose} \cdot \text{age} \\ E(Y|\text{dose}, \text{age} = a) &= \beta_0 + \beta_{\text{dose}}\text{dose} + a\beta_{\text{age}} + a\beta_{d^*a} \cdot \text{dose} \\ E(Y|\text{dose}, \text{age} = a) &= (\beta_0 + a\beta_{\text{age}}) + (\beta_{\text{dose}} + a\beta_{d^*a})\text{dose} \end{aligned}$$

Because the β s are fixed (unknown) constants, this equation tells us that when age is fixed at some particular number, a , the relationship between $E(Y)$ and dose is a straight line with intercept equal to the number $\beta_0 + a\beta_{\text{age}}$ and slope equal to the number $\beta_{\text{dose}} + a\beta_{d^*a}$. The key feature of the interaction is the fact that the slope with respect to dose *is different* for each value of a , i.e., for each age. A similar equation can be written for fixed dose and varying age. The conclusion is that the interaction model is one where the effects of any one-unit change in one explanatory variable while holding the other(s) constant is a change in the mean outcome, but the *size* (and maybe direction) of that change *depends* on the value(s) that the other explanatory variable(s) is/are set to.

Explaining the meaning of the interaction parameter in a multiple regression with continuous explanatory variables is difficult. Luckily, as we will see below, it is much easier in the simplest version of ANCOVA, where there is one categorical and one continuous explanatory variable.

The multiple regression results are shown in tables [10.5](#), [10.6](#), and [10.7](#).

	Unstandardized Coefficients		t	Sig.	95% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
(Constant)	48.995	5.493	8.919	<0.0005	37.991	59.999
Drug dose	0.398	0.282	1.410	0.164	-0.167	0.962
Rat age	0.759	0.500	1.517	0.135	-0.243	1.761
DoseAge IA	0.396	0.025	15.865	<0.0005	0.346	0.446

Table 10.5: Regression results for lethargy experiment.

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.992	0.985	0.984	7.883

Table 10.6: Lethargy experiment model summary.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	222249	3	1101.1	22.868	<0.0005
Residual	3480	56	48.152		
Total	225729	59			

Table 10.7: Lethargy experiment ANOVA.

Here is an interpretation of the analysis of this experiment written in language suitable for an exam answer. A multiple regression analysis including interaction was performed using drug dose in mg and rat age in months as explanatory variables, and minutes resting or sleeping during a 4 hour test period as the outcome. There is a significant interaction ($t=15.86$, $p<0.0005$) between dose and age in their effect on lethargy. (Therefore changes in either or both explanatory variables cause changes in the lethargy outcome.) Because the coefficient estimate for the interaction is of the same sign as the signs of the individual coefficients, it is easy to give a general idea about the effects of the explanatory variables on the outcome. Increases in both dose and age are associated with (cause, for dose) an increase in lethargy, and the effects are “super-additive” or “synergistic” in the sense that the effect of simultaneous fixed increases in both variables is more than the sum of the effects of the same increases made separately for each explanatory variable. We can also see that about 98% of the variability in resting/sleeping time is accounted for by taking the values of dose and age into account. The estimate of the standard deviation of resting/sleeping time for any fixed combination of dose and age is 7.9 minutes.

The validity of these conclusions is confirmed by the following assumption checks. The quantile-normal plot of the residuals confirms Normality of errors, the residual vs. fit plot confirms linearity and equal variance. The fixed-x assumption is met because the dose is precisely set by the experimenter and age is precisely observed. The independent errors assumption is met because separate subjects are used for each test, and the subjects were not allowed to collaborate. Linearity is further confirmed by plots of each explanatory variable vs. the residuals.

Note that the p-value for the interaction line of the regression results (coefficient) table tells us that the interaction is an important part of the model. Also note that the component explanatory variables of the interaction (main effects) are almost always included in a model if the interaction is included. In the presence of a significant interaction both explanatory variables must affect the outcome, so (except in certain special circumstances) *you should not interpret the p-values of the main effects if the interaction has a significant p-value*. On the other hand, if the interaction is not significant, generally the appropriate next step is to perform a new multiple regression analysis excluding the interaction term, i.e., run an additive model.

If we want to write prediction equations with numbers instead of symbols, we should use Y' or \hat{Y} on the left side, to indicate a “best estimate” rather than the

true but unknowable values represented by $E(Y)$ which depends on the β values. For this example, the prediction equation for resting/sleeping minutes for rats of age 12 months at any dose is

$$\hat{Y} = 49.0 + 0.398(\text{dose}) + 0.76(12) + 0.396(\text{dose} \cdot 12)$$

which is $\hat{Y} = 58.1 + 5.15(\text{dose})$.

Interaction between two explanatory variables is present when the effect of one on the outcome depends on the value of the other. Interaction is implemented in multiple regression by including a new explanatory variable that is the product of two existing explanatory variables. The model can be explained by writing equations for the relationship between one explanatory variable and the outcome for some fixed values of the other explanatory variable.

10.3 Categorical variables in multiple regression

To use a categorical variable with k levels in multiple regression we must re-code the data column as $k - 1$ new columns, each with only two different codes (most commonly we use 0 and 1). Variables that only take on the values 0 or 1 are called **indicator** or **dummy** variables. They should be considered as quantitative variables. and should be named to correspond to their “1” level.

An indicator variable is coded 0 for any case that does not match the variable name and 1 for any case that does match the variable name.

One level of the original categorical variable is designated the “baseline”. If there is a control or placebo, the baseline is usually set to that level. The baseline level does not have a corresponding variable in the new coding; instead subjects with that level of the categorical variable have 0’s in all of the new variables. Each new variable is coded to have a “1” for the level of the categorical variable that matches its name and a zero otherwise.

It is very important to realize that when new variables like these are constructed, they *replace* the original categorical variable when entering variables into a multiple regression analysis, so the original variables are no longer used at all. (The originals should not be erased, because they are useful for EDA, and because you want to be able to verify correct coding of the indicator variables.)

This scheme for constructing new variables insures appropriate multiple regression analysis of categorical explanatory variables. As mentioned above, sometimes you need to create these variables explicitly, and sometime a statistical program will create them for you, either explicitly or silently.

The choice of the baseline variable only affects the convenience of presentation of results and does not affect the interpretation of the model or the prediction of future values.

As an example consider a data set with a categorical variable for favorite condiment. The categories are ketchup, mustard, hot sauce, and other. If we arbitrarily choose ketchup as the baseline category we get a coding like this:

Level	Indicator Variable		
	mustard	hot sauce	other
ketchup	0	0	0
mustard	1	0	0
hot sauce	0	1	0
other	0	0	1

Note that this indicates, e.g., that every subject that likes mustard best has a 1 for their “mustard” variable, and zeros for their “hot sauce” and “other” variables.

As shown in the next section, this coding flexibly allows a model to have no restrictions on the relationships of population means when comparing levels of the categorical variable. It is important to understand that if we “accidentally” use a categorical variable, usually with values 1 through k , in a multiple regression, then we are inappropriately forcing the mean outcome to be ordered according to the levels of a nominal variable, *and* we are forcing these means to be equally spaced. Both of these problems are fixed by using indicator variable recoding.

To code the interaction between a categorical variable and a quantitative variable, we need to create another $k - 1$ new variables. These variables are the products of the $k - 1$ indicator variable(s) and the quantitative variable. Each of the resulting new data columns has zeros for all rows corresponding to all levels of the categorical variable except one (the one included in the name of the interaction

variable), and has the value of the quantitative variable for the rows corresponding to the named level.

Generally a model includes all or none of a set of indicator variables that correspond with a single categorical variable. The same goes for the $k - 1$ interaction variables corresponding to a given categorical variable and quantitative explanatory variable.

Categorical explanatory variables can be incorporated into multiple regression models by substituting $k - 1$ indicator variables for any k -level categorical variable. For an interaction between a categorical and a quantitative variable $k - 1$ product variables should be created.

10.4 ANCOVA

The term ANCOVA (analysis of covariance) is used somewhat differently by different analysts and computer programs, but the most common meaning, and the one we will use here, is for a multiple regression analysis in which there is at least one quantitative and one categorical explanatory variable. Usually the categorical variable is a treatment of primary interest, and the quantitative variable is a “control variable” of secondary interest, which is included to improve power (without sacrificing generalizability).

Consider a particular quantitative outcome and two or more treatments that we are comparing for their effects on the outcome. If we know one or more explanatory variables are suspected to both affect the outcome and to define groups of subjects that are more homogeneous in terms of their outcomes for any treatment, then we know that we can use the blocking principle to increase power. Ignoring the other explanatory variables and performing a simple ANOVA increases σ^2 and makes it harder to detect any real differences in treatment effects.

ANCOVA extends the idea of blocking to continuous explanatory variables, as long as a simple mathematical relationship (usually linear) holds between the control variable and the outcome.

10.4.1 ANCOVA with no interaction

An example will make this more concrete. The data in [mathtest.dat](#) come from a (fake) experiment testing the effects of two computer aided instruction (CAI) programs on performance on a math test. The programs are labeled A and B, where A is the control, older program, and B is suspected to be an improved version. We know that performance depends on general mathematical ability so the students math SAT is used as a control variable.

First let's look at t-test results, ignoring the SAT score. EDA shows a slightly higher mean math test score, but lower median for program B. A t-test shows no significant difference with $t=0.786$, $p=0.435$. It is worth noting that the CI for the mean difference between programs is $[-5.36, 12.30]$, so we are 95% confident that the effect of program B relative to the old program A is somewhere between lowering the mean score by 5 points and raising it by 12 points. The estimate of σ (square root of MSwithin from an ANOVA) is 17.1 test points.

EDA showing the relationship between math SAT (MSAT) and test score separately for each program is shown in figure [10.3](#). The steepness of the lines and the fact that the variation in y at any x is smaller than the overall variation in y for either program demonstrates the value of using MSAT as a control variable. The lines are roughly parallel, suggesting that an additive, no-interaction model is appropriate. The line for program B is higher than for program A, suggesting its superiority.

First it is a good idea to run an ANCOVA model with interaction to verify that the fit lines are parallel (the slopes are not statistically significantly different). This is done by running a multiple regression model that includes the explanatory variables ProgB, MSAT, and the interaction between them (i.e, the product variable). Note that we do not need to create a new set of indicator variables because there are only two levels of program, and the existing variable is already an indicator variable for program B. We do need to create the interaction variable in SPSS. The interaction p-value is 0.375 (not shown), so there is no evidence of a significant interaction (different slopes).

The results of the additive model (excluding the interaction) are shown in tables [10.8](#) [10.9](#), and [10.10](#).

Of primary interest is the estimate of the benefit of using program B over program A, which is 10 points ($t=2.40$, $p=0.020$) with a 95% confidence interval of 2 to 18 points. Somewhat surprisingly the estimate of σ , which now refers to

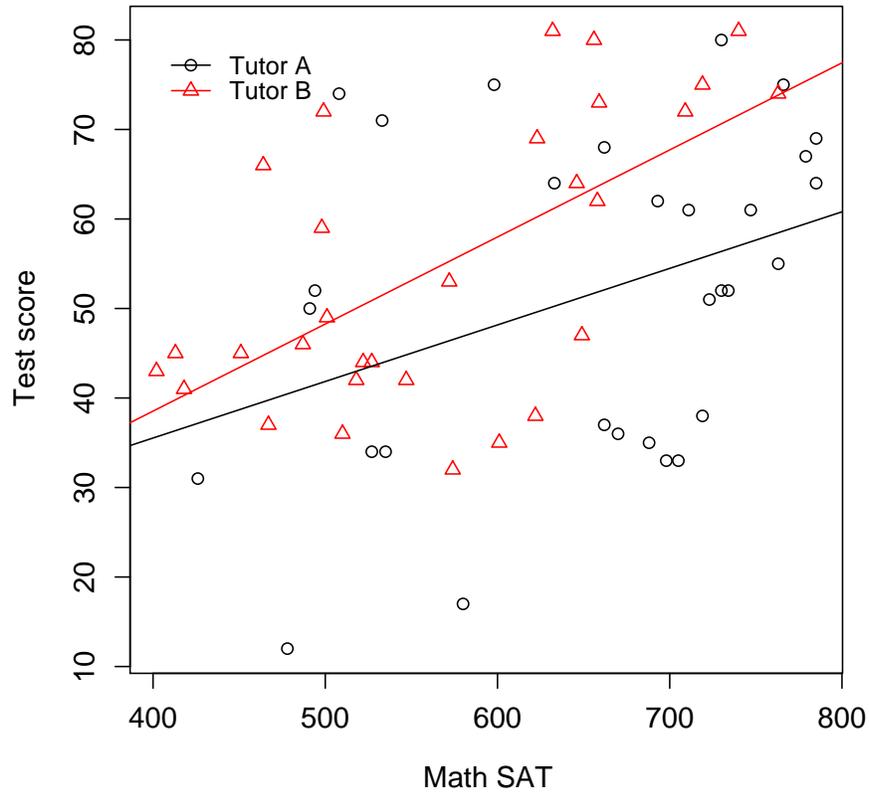


Figure 10.3: EDA for the math test / CAI example.

	Unstandardized Coefficients		t	Sig.	95% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
(Constant)	-0.270	12.698	-0.021	0.983	-25.696	25.157
ProgB	10.093	4.206	2.400	0.020	1.671	18.515
Math SAT	0.079	0.019	4.171	<0.0005	0.041	0.117

Table 10.8: Regression results for CAI experiment.

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.492	0.242	0.215	15.082

Table 10.9: CAI experiment model summary.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	4138	2	2069.0	0.095	<0.0005
Residual	12966	57	227.5		
Total	17104	59			

Table 10.10: CAI experiment ANOVA.

the standard deviation of test score for any combination of program *and* MSAT is only slightly reduced from 17.1 to 15.1 points. The ANCOVA model explains 22% of the variability in test scores (adjusted r-squared = 0.215), so there are probably some other important variables “out there” to be discovered.

Of minor interest is the fact that the “control” variable, math SAT score, is highly statistically significant ($t=4.17$, $p<0.0005$). Every 10 additional math SAT points is associated with a 0.4 to 1.2 point rise in test score.

In conclusion, program B improves test scores by a few points on average for students of all ability levels (as determined by MSAT scores).

This is a typical ANOVA story where the power to detect the effects of a treatment is improved by including one or more control and/or blocking variables, which are chosen by subject matter experts based on prior knowledge. In this case the effect of program B compared to control program A was detectable using MSAT in an ANCOVA, but not when ignoring it in the t-test.

The simplified model equations are shown here.

$$\begin{aligned}
 E(Y|\text{ProgB}, MSAT) &= \beta_0 + \beta_{\text{ProgB}}\text{ProgB} + \beta_{\text{MSAT}}MSAT \\
 \text{Program A: } E(Y|\text{ProgB} = 0, MSAT) &= \beta_0 + \beta_{\text{MSAT}}MSAT \\
 \text{Program B: } E(Y|\text{ProgB} = 1, MSAT) &= (\beta_0 + \beta_{\text{ProgB}}) + \beta_{\text{MSAT}}MSAT
 \end{aligned}$$

To be perfectly explicit, β_{MSAT} is the slope parameter for MSAT and β_{ProgB} is the parameter for the indicator variable ProgB. This parameter is technically a “slope”, but really determines a difference in intercept for program A vs. program B.

For the analysis of the data shown here, the predictions are:

$$\begin{aligned}\hat{Y}(\text{ProgB}, MSAT) &= -0.27 + 10.09\text{ProgB} + 0.08MSAT \\ \text{Program A: } \hat{Y}(\text{ProgB} = 0, MSAT) &= -0.27 + 0.08MSAT \\ \text{Program B: } \hat{Y}(\text{ProgB} = 1, MSAT) &= 9.82 + 0.08MSAT\end{aligned}$$

Note that although the intercept is a meaningless extrapolation to an impossible MSAT score of 0, we still need to use it in the prediction equation. Also note, that in this no-interaction model, the simplified equations for the different treatment levels have different intercepts, but the same slope.

ANCOVA with no interaction is used in the case of a quantitative outcome with both a categorical and a quantitative explanatory variable. The main use is for testing a treatment effect while using a quantitative control variable to gain power.

10.4.2 ANCOVA with interaction

It is also possible that a significant interaction between a control variable and treatment will occur, or that the quantitative explanatory variable is a variable of primary interest that interacts with the categorical explanatory variable. Often when we do an ANCOVA, we are “hoping” that there is no interaction because that indicates a more complicated reality, which is harder to explain. On the other hand sometimes a more complicated view of the world is just more interesting!

The multiple regression results shown in tables 10.11 and 10.12 refer to an experiment testing the effect of three different treatments (A, B and C) on a quantitative outcome, performance, which can range from 0 to 200 points, while controlling for skill variable S, which can range from 0 to 100 points. The data are available at [Performance.dat](#). EDA showing the relationship between skill and

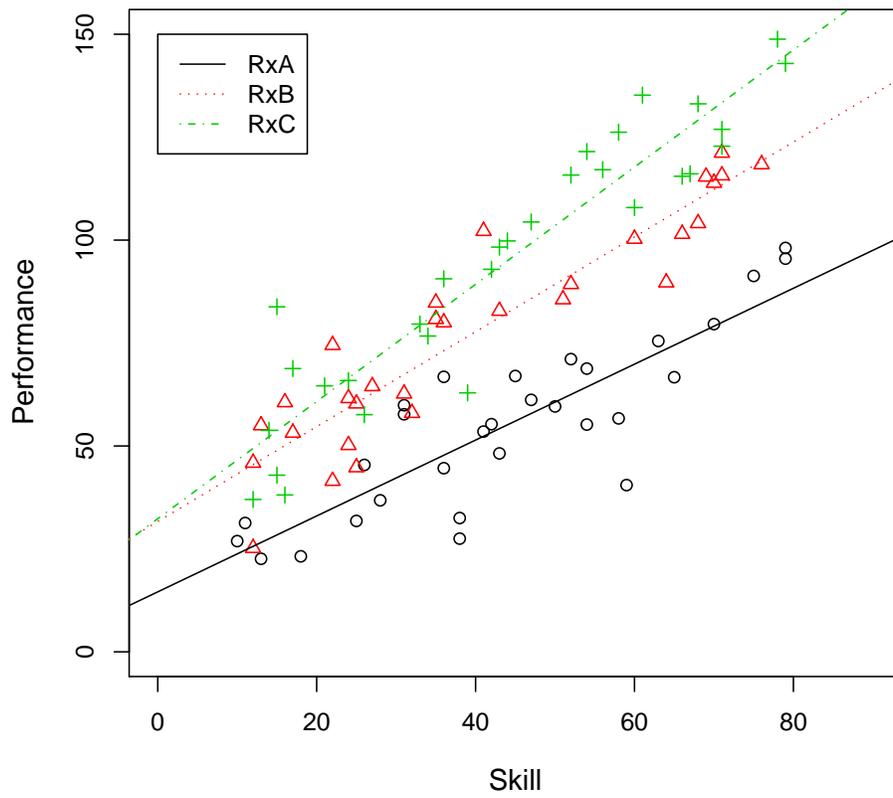


Figure 10.4: EDA for the performance ANCOVA example.

performance separately for each treatment is shown in figure 10.4. The treatment variable, called Rx, was recoded to $k - 1 = 2$ indicator variables, which we will call RxB and RxC, with level A as the baseline. Two interaction variables were created by multiplying S by RxB and S by RxC to create the single, two column interaction of Rx and S. Because it is logical and customary to consider the interaction between a continuous explanatory variable and a k level categorical explanatory variable, where $k > 2$, as a *single* interaction with $k - 1$ degrees of freedom and $k - 1$ lines in a coefficient table, we use a special procedure in SPSS (or other similar programs) to find a single p-value for the null hypothesis that model is additive vs. the alternative that there is an interaction. The SPSS procedure using the Linear Regression module is to use two “blocks” of independent variables, placing the main effects (here RxB, RxC, and Skill) into block 1, and the going to the “Next” block and placing the two interaction variables (here, RxB*S and RxC*S) into block 2. The optional statistic “R Squared Change” must also be selected.

The output that is labeled “Model Summary” (Table 10.11) and that is produced with the “R Squared Change” option is explained here. Lines are shown for two models. The first model is for the explanatory variables in block 1 only, i.e., the main effects, so it is for the additive ANCOVA model. The table shows that this model has an adjusted R^2 value of 0.863, and an estimate of 11.61 for the standard error of the estimate (σ). The second model adds the single 2 df interaction to produce the full interaction ANCOVA model with separate slopes for each treatment. The adjusted R^2 is larger suggesting that this is the better model. One good formal test of the necessity of using the more complex interaction model over just the additive model is the “F Change” test. Here the test has an F statistic of 6.36 with 2 and 84 df and a p-value of 0.003, so we reject the null hypothesis that the additive model is sufficient, and work only with the interaction model (model 2) for further interpretations. (The Model-1 “F Change test” is for the necessity of the additive model over an intercept-only model that predicts the intercept for all subjects.)

Using mnemonic labels for the parameters, the structural model that goes with this analysis (Model 2, with interaction) is

$$E(Y|Rx, S) = \beta_0 + \beta_{RxB}RxB + \beta_{RxC}RxC + \beta_S S + \beta_{RxB*S}RxB \cdot S + \beta_{RxC*S}RxC \cdot S$$

You should be able to construct this equation directly from the names of the explanatory variables in Table 10.12.

Using Table 10.12, the parameter estimates are $\beta_0 = 14.56$, $\beta_{RxB} = 17.10$, $\beta_{RxC} = 17.77$, $\beta_S = 0.92$, $\beta_{RxB*S} = 0.23$, and $\beta_{RxC*S} = 0.50$.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.931	0.867	0.863	11.61
2	0.941	0.885	0.878	10.95

Change Statistics					
Model	R Square Change	F Change	df1	df2	Sig. F Change
1	0.867	187.57	3	86	<0.0005
2	0.017	6.36	2	84	0.003

Table 10.11: Model summary results for generic experiment.

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	3.22	3.39	0.95	0.344
	RxB	27.30	3.01	9.08	<0.0005
	RxC	39.81	3.00	13.28	<0.0005
	S	1.18	0.06	19.60	<0.0005
2	(Constant)	14.56	5.00	2.91	0.005
	RxB	17.10	6.63	2.58	0.012
	RxC	17.77	6.83	2.60	0.011
	S	0.92	0.10	8.82	<0.0005
	RxB*S	0.23	0.14	1.16	0.108
	RxC*S	0.50	0.14	3.55	0.001

Table 10.12: Regression results for generic experiment.

To understand this complicated model, we need to write simplified equations:

$$\begin{aligned} \text{RxA: } E(Y|\text{Rx}=\text{A}, \text{S}) &= \beta_0 + \beta_{\text{S}}\text{S} \\ \text{RxB: } E(Y|\text{Rx}=\text{B}, \text{S}) &= (\beta_0 + \beta_{\text{RxB}}) + (\beta_{\text{S}} + \beta_{\text{RxB}*\text{S}})\text{S} \\ \text{RxC: } E(Y|\text{Rx}=\text{C}, \text{S}) &= (\beta_0 + \beta_{\text{RxC}}) + (\beta_{\text{S}} + \beta_{\text{RxC}*\text{S}})\text{S} \end{aligned}$$

Remember that these simplified equations are created by substituting in 0's and 1's for RxB and RxC (but not into parameter subscripts), and then fully simplifying the equations.

By examining these three equations we can fully understand the model. From the first equation we see that β_0 is the mean outcome for subjects given treatment A and who have $\text{S}=0$. (It is often worthwhile to “center” a variable like S by subtracting its mean from every value; then the intercept will refer to the mean of S, which is never an extrapolation.)

Again using the first equation we see that the interpretation of β_{S} is the slope of Y vs. S for subjects given treatment A.

From the second equation, the intercept for treatment B can be seen to be $(\beta_0 + \beta_{\text{RxB}})$, and this is the mean outcome when $\text{S}=0$ for subjects given treatment B. Therefore the interpretation of β_{RxB} is the *difference* in mean outcome when $\text{S}=0$ when comparing treatment B to treatment A (a positive parameter value would indicate a higher outcome for B than A, and a negative parameter value would indicate a lower outcome). Similarly, the interpretation of $\beta_{\text{RxB}*\text{S}}$ is the *change* in slope from treatment A to treatment B, where a positive $\beta_{\text{RxB}*\text{S}}$ means that the B slope is steeper than the A slope and a negative $\beta_{\text{RxB}*\text{S}}$ means that the B slope is less steep than the A slope.

The null hypotheses then have these specific meanings. $\beta_{\text{RxB}} = 0$ is a test of whether the intercepts differ for treatments A and B. $\beta_{\text{RxC}} = 0$ is a test of whether the intercepts differ for treatments A and C. $\beta_{\text{RxB}*\text{S}} = 0$ is a test of whether the slopes differ for treatments A and B. And $\beta_{\text{RxC}*\text{S}} = 0$ is a test of whether the slopes differ for treatments A and C.

Here is a full interpretation of the performance ANCOVA example. Notice that the interpretation can be thought of a description of the EDA plot which uses ANCOVA results to specify which observations one might make about the plot that are statistically verifiable.

Analysis of the data from the performance dataset shows that treatment and

skill interact in their effects on performance. Because skill levels of zero are a gross extrapolation, we should not interpret the intercepts.

If skill=0 were a meaningful, observed state, then we would say all of the things in this paragraph. The estimated mean performance for subjects with zero skill given treatment A is 14.6 points (a 95% CI would be more meaningful). If it were scientifically interesting, we could also say that this value of 14.6 is statistically different from zero ($t=2.91$, $df=84$, $p=0.005$). The intercepts for treatments B and C (mean performances when skill level is zero) are both statistically significantly different from the intercept for treatment A ($t=2.58, 2.60$, $df=84$, $p=0.012, 0.011$). The estimates are 17.1 and 17.8 points higher for B and C respectively compared to A (and again, CIs would be useful here).

We can also say that there is a statistically significant effect of skill on performance for subjects given treatment A ($t=8.82$, $p < 0.0005$). The best estimate is that the mean performance increases by 9.2 points for each 10 point increase in skill. The slope of performance vs. skill for treatment B is not statistically significantly different for that of treatment A ($t=1.15$, $p=0.108$). The slope of performance vs. skill for treatment C is statistically significantly different for that of treatment A ($t=3.55$, $p=0.001$). The best estimate is that the slope for subjects given treatment C is 0.50 higher than for treatment A (i.e., the mean change in performance for a 1 unit increase in skill is 0.50 points *more for treatment C than for treatment A*). We can also say that the best estimate for the slope of the effect of skill on performance for treatment C is $0.92+0.50=1.42$.

Additional testing, using methods we have not learned, can be performed to show that performance is better for treatments B and C than treatment A at all observed levels of skill.

In summary, increasing skill has a positive effect on performance for treatment A (of about 9 points per 10 point rise in skill level). Treatment B has a higher projected intercept than treatment A, and the effect of skill on subjects given treatment B is not statistically different from the effect on those given treatment A. Treatment C has a higher projected intercept than treatment A, and the effect of skill on subjects given treatment C is statistically different from the effect on those given treatment A (by about 5 additional points per 10 unit rise in skill).

If an ANCOVA has a significant interaction between the categorical and quantitative explanatory variables, then the slope of the equation relating the quantitative variable to the outcome differs for different levels of the categorical variable. The p-values for indicator variables test intercept differences from the baseline treatment, while the interaction p-values test slope differences from the baseline treatment.

10.5 Do it in SPSS

To create $k - 1$ indicator variables from a k -level categorical variable in SPSS, run Transform/RecodeIntoDifferentVariables, as shown in figure 5.16, $k - 1$ times. Each new variable name should match one of the non-baseline levels of the categorical variable. Each time you will set the old and new values (figure 5.17) to convert the named value to 1 and “all other values” to 0.

To create $k - 1$ interaction variables for the interaction between a k -level categorical variable and a quantitative variable, use Transform/Compute $k - 1$ times. Each new variable name should specify what two variables are being multiplied. A label with a “*”, “.” or the word “interaction” or abbreviation “I/A” along with the categorical level and quantitative name is a really good idea. The “Numeric Expression” (see figure 5.15) is just the product of the two variables, where “*” means multiply.

To perform multiple regression in any form, use the Analyze/Regression/Linear menu item (see figure 9.7), and put the outcome in the Dependent box. Then put all of the main effect explanatory variables in the Independent(s) box. Do **not** use the original categorical variable – use only the $k - 1$ corresponding indicator variables. If you want to model non-parallel lines, add the interaction variables as a second block of independent variables, and turn on the “R Square Change” option under “Statistics”. As in simple regression, add the option for CI’s for the estimates, and graphs of the normal probability plot and residual vs. fit plot. Generally, if the “F change test” for the interaction is greater than 0.05, use “Model 1”, the additive model, for interpretations. If it is ≤ 0.05 , use “Model 2”, the interaction model.