Notes on regression analysis

1. Basics in regression analysis – key concepts (actual implementation is more complicated)
   A. Collect data
   B. Plot data on graph, draw a line through the middle of the scatter of points
   C. Determine the intercept and slope(s) of the line (called regression coefficient(s) – or parameter(s))
   D. Use statistical theory and data to determine "margin of error," "standard error," "t-statistic" for each regression coefficient

2. Example: class survey – results are displayed as an equation or in a table
   A. as an equation:
      \[
      \text{earnings per hour} = -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 5.28 \text{ GPA}
      \]
      \[
      \text{standard error:} \quad (22.94) \quad (1.65) \quad (0.88) \quad (2.25)
      \]
      \[
      \text{t-statistic:} \quad [0.58] \quad [2.12] \quad [0.69] \quad [2.34]
      \]
   B. in a table:
      \[
      \begin{array}{lll}
      \text{dependent variable} & \text{is} & \text{earnings per hour} \\
      \text{independent variable} & \text{regression coefficient} & \text{(t-statistic in parentheses)} \\
      \text{female} & -3.51 & (2.12) \\
      \text{age} & 0.61 & (0.69) \\
      \text{GPA} & 5.28 & (2.34)
      \end{array}
      \]
      … where Female = 0 for males, = 1 for females
      Age is measured in years (22, 23, …)
      GPA is measured in units on a four-point scale (0, 1.22, 3.45, etc.)

3. Interpreting regression coefficients: The regression coefficient on a particular independent variable represents the effect on the dependent variable of an increase of one unit in that independent variable, with all other independent variables remaining unchanged.

First consider the coefficient for sex, which is -3.51. This means that changing the sex variable from 0 (male) to 1 (female), with age and GPA remaining unchanged, would reduce earnings by 3.51 (and this effect is statistically significant, as we will see later):

   for a male, Female = 0, so, for a male,
   \[
   \text{earnings per hour} = -13.23 - 3.51 \times 0 + 0.61 \text{ Age} + 5.28 \text{ GPA} \\
   = -13.23 + 0.61 \text{ Age} + 5.28 \text{ GPA}
   \]

   for a female, Female = 1, so, for a female,
   \[
   \text{earnings per hour} = -13.23 - 3.51 \times 1 + 0.61 \text{ Age} + 5.28 \text{ GPA} \\
   = -13.23 - 3.51 + 0.61 \text{ Age} + 5.28 \text{ GPA}
   \]

Thus, when sex changes from 0 (male) to 1 (female), earnings per hour will change by -3.51, "other things being equal" (i.e., provided both Age and GPA remain the same).
3. **Interpreting regression coefficients** (continued):

Now consider the coefficient for Age, which is +0.61. This means that changing the age variable by one year, with sex and GPA remaining unchanged, would raise earnings by 0.61 (but this effect is not statistically significant, as we will see later):

\[
\text{earnings per hour} = -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 5.28 \text{ GPA}
\]

Adding one year to Age, with Female and GPA remaining the same, will mean that…

\[
\begin{align*}
\text{earnings per hour} &= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 5.28 \text{ GPA} \\
&= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 0.61 + 5.28 \text{ GPA} \\
&= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 0.61 + 5.28 \text{ GPA} \\
&= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 0.61 + 5.28 \text{ GPA} \\
&= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 0.61 + 5.28 \text{ GPA} \\
&= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 0.61 + 5.28 \text{ GPA} \\
\end{align*}
\]

(i.e., 0.61 larger than before)

Thus, when Age changes by 1 unit, earnings per hour will change by +0.61, "other things being equal" (i.e., with Female and GPA remaining the same).

Finally, consider the coefficient for GPA, which is +5.28. This means that changing the GPA variable by one unit, with sex and age remaining unchanged, would raise earnings by 5.28 (and this effect is statistically significant, as we will see later):

\[
\text{earnings per hour} = -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 5.28 \text{ GPA}
\]

Adding one year to GPA, with Female and Age remaining the same, will mean that…

\[
\begin{align*}
\text{earnings per hour} &= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 5.28 \text{ GPA} \\
&= -13.23 - 3.51 \text{ Female} + 0.61 \text{ Age} + 5.28 \text{ GPA}
\end{align*}
\]

(i.e., 5.28 larger than before)

Thus, when GPA changes by 1 unit, earnings per hour will change by 5.28, "other things being equal" (i.e., with Female and Age remaining the same).

4. **Statistical significance**: margin of error, standard error, t-statistic

Regression coefficients (or, equivalently, regression coefficients) are just estimates. Like all estimates, they are subject to variation and are subject to a "margin of error." This "margin of error" is computed based on statistical theory and the data.

standard error = a measure of the variability in estimates of a regression coefficient

"margin of error" = ± {2 × standard error} or, more precisely, ± {1.96 × standard error}

t-statistic = | regression coefficient | ÷ standard error

if t-statistic ≥ 2 (more precisely, 1.96), then coefficient is "statistically significant"

(it's unlikely we'd get an estimate of this coefficient at least this large, if the true value of the coefficient were actually zero)

if t-statistic < 2 (or, more precisely, 1.96), then coefficient is "not statistically significant"

(it's not unlikely we'd get an estimate of this coefficient at least this large, if the true value of the coefficient were actually zero)
More on statistical significance calculations

To understand the logic of statistical significance calculations, consider a simple example involving opinion polls. We conduct a poll of 1,000 people and ask how many of them favor Candidate A. We record the answers, and find that 52.1 percent of the people in this poll favor candidate A and 47.9 percent favor candidate B. (Thus, according to this poll, candidate A leads by 52.1 – 47.9 = 4.2 percentage points.) We conduct another poll of 1,000 people and find that in this second poll, 54.3 percent of the people favor candidate A and 45.7 percent favor candidate B (so that, in this second poll, A's lead over B is 8.6 percentage points.) We conduct another poll, and another, and so on, until we have the results of thousands of polls, each of which has surveyed 1,000 people at exactly the same moment.

Statistical theory tells us the following: (1) the distribution of the results of all of these polls will be centered on the percentage of the people in the population as a whole who favor A; and (2) the distribution of the results will look like a bell-shaped ("normal") curve.

One of the characteristics of the normal curve is that 2.5% of the observations are located in the left "tail" of the distribution (i.e., well below the mean or "average" value); and another 2.5% are located in the right tail (i.e., well above the mean). More precisely, if we measure the results not in percentage terms but in units of "standard errors," precisely 2.5% of the observations will be 1.96 standard errors below the mean; and another 2.5% will be 1.96 standard errors above the mean. (See below. It's conventional to use the Greek letter σ, which is the lower-case "sigma," as the symbol for the standard error. The value of the standard error is calculated from the data, using formulas determined by rules of statistical inference.) In other words, it is rare for observations to be 1.96 (or, roughly speaking, 2.0) or more standard errors away from the mean – this occurs only about 5% of the time. Similarly, most of the time (95% of the time, to be precise), the poll will produce a result that is less than 1.96 standard errors away from the mean.

![Figure 2.12](image)

**Figure 2.12.** Approximate probability or area in the tail portions in the normal curve. The shaded area represents the probability of observing a value 1.96 standard deviations or more from the mean and has area .05. The right half of the figure presents the probabilities of observing a value exceeding the mean by .67, 1.65, 1.96, 2.58 and 3 standard deviations. The symmetry of the normal curve implies that the probability of exceeding the mean by x or more standard deviations equals the probability of being below the mean by x or more standard deviations.

Now suppose we want to use the results of a single opinion poll to determine how the population feels about candidate A. (Thus, we are no longer imagining that we can take thousands upon thousands of such polls.) Suppose that, in this single poll, 57.1 percent of the people support candidate A. We know that, simply as the result of chance, there is some possibility that we could get a poll result that is at least this much different from 50.0 percent even if, in the population as a whole, people are divided 50-50 over the two candidates. The logic of statistical significance then proceeds as follows:

We start by assuming that the population is evenly divided between the two candidates. Equivalently, we assume that there is no difference between the degree of support for the two candidates in the population. (This assumption of "no difference" is usually called the "null hypothesis," and it is the starting point for all significance testing.) Next, we ask, "How likely is it that, if the population were evenly divided between the two candidates, we could nevertheless get a poll result like ours – one that is 7.1 percentage points away from 50.0 percent?" Next, we use the data to calculate the standard error of the population mean, and we measure the difference between our poll's result and the assumed 50-50 split using units of standard errors. For example, if one standard error is 3.0 percentage points, then the 7.1 percentage-point difference in this poll (57.1 – 50.0) is equivalent to 7.1 ÷ 3.0 = 2.33 standard errors.

If this result is equal or greater than 1.96 (or approximately 2.0), then this difference is called "statistically significant." In other words, in this case, if the population were truly evenly divided between the two candidates (so that each candidate enjoyed the support of 50.0 percent of the population), it is very unlikely (there is less than a 5% probability) that, simply as a result of sampling variation or "chance" a poll would show that one candidate's support differs from 50.0 percent by as much as this candidate's support does. (In other words, if the population were truly evenly divided, it is very unlikely that, simply because of sampling variation, an opinion poll would show that one candidate is ahead by at least 7.1 percentage points – which is what we actually see here.)

On the other hand, suppose that the difference between the poll's result and the assumed 50-50 split is equivalent to something less than 1.96 standard errors. For example, suppose again that 57.1 percent of the people favor candidate A, but that one standard error is equal to 5.0 percentage points. Then, expressed in terms of standard error units, 7.1 percentage points (= the extent to which our poll's result departs from a 50-50 split) is 7.1 ÷ 5.0 = 1.42, which is less than 1.96. In this case, the difference between our poll result and an assumed 50-50 split is called "not statistically significant." In other words, in this case, if the population were truly evenly divided between the two candidates, it is not unlikely (There is better than a 5% probability) that, simply as a result of sampling variation ("chance"), we would get a poll result showing a lead for one candidate that is as different from 50% as this one.

In the context of regression analysis, the statistical significance of a regression coefficient is usually assessed using the t-statistic. This is simply the ratio of the regression coefficient to its standard error:

\[
t = \frac{\text{regression coefficient}}{\text{standard error}}
\]

The regression coefficient is usually written as \( \beta \); the standard error of the coefficient is usually written as \( \sigma \), so we can also write

\[
t = \frac{\beta}{\sigma}
\]

(The standard error is calculated using the data and formulas derived using statistical theory. Intuitively, if there is a lot of variation in the data and/or if the number of observations in the analysis is low, then \( \sigma \) will tend to be large.) So, for example, if a regression coefficient estimate is 5.00 and its standard error is 2.50, then the t-statistic for that regression estimate is \( t = 5.00 / 2.50 = 2.00 \). Finally, as above, if the t-statistic is equal to or greater than 1.96, then the regression coefficient is said to be "statistically significant."
In intuitive terms, it may be helpful to think of significance testing as a kind of courtroom exercise. We begin by adopting a "null hypothesis," i.e., that there is no difference between two things. (In the example here, we assumed that there is no difference in support for the candidates.) In courtroom terms, this would be like an assumption of "innocent until proven guilty." We then consider the statistical evidence. If the evidence seems very far away from the null hypothesis (in terms of "standard deviation" units), we reject the null hypothesis on the grounds that the evidence isn't really consistent with the assumption of no difference. Likewise, if the courtroom evidence seems very far away from the assumption of innocence, we reject the assumption of innocence.

A few caveats and qualifications are appropriate here.

First, statistical significance is not the same as "significance" in the ordinary, everyday sense. In everyday speech, "significant" simply means "big," "sizeable," etc. In contrast, even a difference that is small can nevertheless be "significant" in a statistical sense. Likewise, a difference that is "large" in the ordinary language sense will not necessarily be significant in a statistical sense. For example, we might sample ten men and four women and find that the sex difference in pay in this sample is $5,000 per year. This is certainly significant in the everyday sense ($5,000 is a lot of money!), but it might not be statistically significant, because it's possible that mere sampling variation alone could produce a difference of at least this size.

Second, it is not correct to say that a statistically significant difference is one for which there is less than a 5% probability that it was caused by chance; and it is completely incorrect to say that a statistically significant difference is one for which there is more than a 95% probability that it occurred due to something other than chance. Neither of these statements correctly characterizes what a statistical significance test does.

Statistical significance testing proceeds as follows. We begin by adopting a "null hypothesis" of a zero difference in the underlying population. (For example, if we are testing for a difference between the average pay of men and women, we assume no sex difference in pay for the underlying population; if we are testing for a difference in support for two candidates, we assume a 50-50 split in the underlying population; and so on.) Next, we ask: if there were a zero difference in the underlying population, how likely is it that, solely as a result of chance, we would get a difference that is at least as far away from zero as is the difference in our actual sample?

If the difference obtained in our sample is "far away" from zero (i.e., is at least 1.96 standard errors away from zero), then the difference is "statistically significant" – i.e., chance alone would produce such a result only very infrequently, if the null hypothesis is correct. In other words, less than 5% of the time, chance would produce a result at least as far away from zero as our result is, if the null hypothesis of zero difference were correct; and more than 95% of the time, chance alone would produce a result that is closer to zero than ours is – again, if the null hypothesis were correct.

**Notes on indicator/dichotomous/binary/dummy variables**

Dummy variables are equal to 1 if the observation is in a specific category, and equal to 0 otherwise.

- **sex:**
  - F = 1 if person is female, = 0 otherwise (i.e., if male)
  - M = 1 if person is male, = 0 otherwise (i.e., if female)

The category equaling zero – the "otherwise" category – is often called the "reference category." (E.g., for the dummy variable called F, above, *male* is the reference category.)

Dummy variables act like a "switch," changing a characteristic into something else. Dummy variables can also provide a convenient way to combine several different equations into a single equation.
Case 1: Two equations differing only in terms of the intercept (have the same slope)

<table>
<thead>
<tr>
<th>two equations:</th>
<th>written as a single equation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>men:</td>
<td>S = 0.15 + 0.20 X</td>
</tr>
<tr>
<td>women:</td>
<td>S = 0.05 + 0.20 X</td>
</tr>
<tr>
<td>using M dummy:</td>
<td>S = 0.05 + 0.20 X + 0.10 M</td>
</tr>
<tr>
<td>using F dummy:</td>
<td>S = 0.15 + 0.20 X – 0.10 F</td>
</tr>
</tbody>
</table>

For example, using an M dummy, consider the single equation above and to the right:

- For women, M = 0, so equation is \( S = 0.05 + 0.20 X + 0.10 \times 0 = 0.05 + 0.20 X \)
- For men, M = 1, so equation is \( S = 0.15 + 0.20 X + 0.10 \times 1 = 0.15 + 0.20 X \)

Thus, we have written both equations as a single equation using the M dummy – the one equation with the M dummy is exactly the same as the two separate equations shown above and to the left.

This is equally true of the single equation using the F dummy (can you verify this?). Thus, it doesn't matter whether we use an M dummy (making women the reference category) or an F dummy (making men the reference category). Using the M dummy tells us that the male intercept is 0.10 higher than the female intercept; using the F dummy tells us that the female intercept is 0.10 lower than the male intercept. Thus, either approach yields exactly the same information.

Case 2: Two equations with different intercepts and also different slopes

<table>
<thead>
<tr>
<th>two equations:</th>
<th>written as a single equation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>men:</td>
<td>S = 0.15 + 0.40 X</td>
</tr>
<tr>
<td>women:</td>
<td>S = 0.05 + 0.21 X</td>
</tr>
<tr>
<td>using M dummy:</td>
<td>S = 0.05 + 0.21 X + 0.10 M × M</td>
</tr>
<tr>
<td>using F dummy:</td>
<td>S = 0.10 + 0.40 X – 0.10 F × F</td>
</tr>
</tbody>
</table>

For example, using an M dummy:

- For women, M = 0, so equation is \( S = 0.05 + 0.21 X + 0.10 \times 0 + 0.19 X \times 0 = 0.05 + 0.21 X \)
- For men, M = 1, so equation is \( S = 0.05 + 0.21 X + 0.10 \times 1 + 0.19 X \times 1 = 0.15 + 0.40 X \)

Note that, in the single equations up and to the right, the term \( X \times M \) and the term \( X \times F \) involve the product of two variables (either \( X \) and \( M \), or \( X \) and \( F \)). These are "interaction terms," and we'll say more about them later on.

Dummy variables involving multiple categories

The above examples refer to dummy variables representing exactly two categories (e.g., male/female, before/after, pass/fail, under 21/21 or older, etc.). Note that we need only one dummy variable to represent the state of being, or not being, in either of these two categories. A "one" refers to an observation that is in one category (e.g., male), and a "zero" for the same variable refers to an observation that is in the only other ("reference") category (e.g., female).

More generally, a variable that has \( K \) mutually-exclusive and exhaustive categories can be represented using \( K-1 \) dummy variables; and the state of being in the "reference" category is indicated by a situation in which all of the \( K-1 \) variables equal zero. ("Mutually-exclusive" means that each person, or object, can be in one and only one of the categories; "exhaustive" means that the categories cover all of the possible outcomes for that variable.)

For example, suppose there are three kinds of college majors: math/science, social science, and humanities. (Note that these categories must be mutually-exclusive, so that there are no double majors; and must also be exhaustive, so that all possible majors fit into one of these three categories.) Since there are three categories (\( K = 3 \)), we need \( K - 1 = 2 \) dummy variables to represent everyone's major:
\[ D_1 = 1 \text{ if majoring in math/science} \]
\[ D_1 = 0 \text{ otherwise (i.e., if majoring in social science or humanities)} \]

\[ D_2 = 1 \text{ if majoring in social science} \]
\[ D_2 = 0 \text{ otherwise (i.e., if majoring in math/science or humanities)} \]

How we represent the fact of majoring in humanities, the third category? If \( D_1 = 0 \) and \( D_2 = 0 \), then we know this is not a major in math/science, and we also know this is not a major in social science – so it must be a major in humanities. (Just the same logic applies to a single dummy variable: if \( M = 0 \), we know this is not a male, so it must be a female. Thus, with two sex categories, \( K = 2 \), and so we only need one dummy variable \((K - 1 = 1)\) to represent the fact of being in either of the two categories.)

Thus, for example, a regression equation for salary \((S)\) that took account of age \((A)\), sex \((M)\), and college major (using the dummy variables \( D_1 \) and \( D_2 \)) might look like this:

\[ S = 20000 + 500 A + 450 M + 1200 D_1 - 400 D_2 \]

Note that this setup assumes that there may be different intercepts (but not different slopes) for the salary lines for the different majors:

for math/science, \( D_1 = 1 \) and \( D_2 = 0 \), so \[ S = 20000 + 500 A + 450 M + 1200 \]
\[ = 21200 + 500 A + 450 M \]

for social science, \( D_1 = 0 \) and \( D_2 = 1 \), so \[ S = 20000 + 500 A + 450 M - 400 \]
\[ = 19600 + 500 A + 450 M \]

for humanities, \( D_1 = 0 \) and \( D_2 = 0 \), so \[ S = 20000 + 500 A + 450 M \]

Note that interpreting the coefficients for these dummy variables \((1200 \text{ for } D_1, -400 \text{ for } D_2)\) can be a little tricky. To begin with, remember that the regression coefficient for a particular independent variable always tells us the effect on the dependent variable of an increase of one unit in that independent variable, with all other independent variables remaining unchanged. Thus, for example, the coefficient of 1200 on the dummy variable for \( D_1 \) tells us the effect on salary \((S)\) of increasing \( D_1 \) by one unit, with all other variables unchanged.

Now, how is it possible to increase \( D \) by one unit while leaving all of the other independent variables unchanged? We do this only by increasing \( D_1 \) from zero to 1, while leaving \( A, M \) and \( D_2 \) unchanged. Note, however, that if we increase \( D_1 \) from zero to 1 but do not change \( D_2 \), we must be changing major from humanities \((\text{with } D_1 = 0 \text{ and } D_2 = 0)\) to math/science \((\text{with } D_1 = 1 \text{ and } D_2 = 0)\). So the coefficient on \( D_1 \), 1200, represents the effect on salary of switching major from humanities to math/science, other things \((A \text{ and } M)\) being equal.

Note that we would get the same result if we assume identical values for \( A \) and \( M \) and then compare average salaries for math/science and humanities majors: from the above,

math/science equation is \[ S = 21200 + 500 A + 450 M \]

humanities equation is \[ S = 20000 + 500 A + 450 M \]

If we assume that \( A \) and \( M \) are the same in both equations, and then subtract the two equations, there will get \( 21200 - 20000 = 1200 \) as the difference in pay between the two majors, other things \((A \text{ and } M)\) being equal. (Of course, this is the same result we just got by looking directly at the coefficient for \( D_1 \).)

Likewise, the effect on pay (other things being equal) of changing major from humanities to social science is -400. Note that this is the effect of switching \( D_2 \) from 0 to 1, while leaving \( D_1 \) unchanged (and equal to zero).
Finally, how about other kinds of "major" effects – for example, what is the effect of changing major from social science to math/science, other things being equal? Note that this is a little more complicated, because in this case we are turning off the "switch" for being a social science major and turning on the "switch" for being a math/science major. More precisely, we have to change $D_2$ from 1 to zero (because we are no longer a social science major) and also change $D_1$ from 0 to 1 (because we are turning into a math/science major). Switching $D_2$ from 1 to 0 reduces $D_2$ by one unit, and therefore changes pay by $\times -1 = +400$. Switching $D_1$ from 0 to 1 raises $D_1$ by one unit, and therefore changes pay by a further $+1200 \times +1 = +1200$. So the effect of changing major from social science to math/science (other things being equal) is $400 + 1200 = +1600$.

Allowing regression slopes (as well as intercepts) to depend on multi-category dummy variables

We previously saw that it is straightforward to allow regression slopes (as well as intercepts) to depend on a two-category dummy variable (we considered the case in which a slope in a regression depends on a variable for sex, $M$). The extension to multi-category dummy variables is straightforward. For example, suppose that salary grows with age ($A$) and depends on sex ($M$) at different rates for persons with different majors. To allow for this possibility, you can run a regression that includes so-called "interaction terms" which represent the multiplication of the dummy variables with the other variables, age ($A$) and sex ($M$). Such a regression might look like this:

$$S = 18900 + 500 A - 100 A \times D_1 + 50 A \times D_2$$
$$+ 450 M + 50 M \times D_1 + 150 M \times D_2$$
$$+ 800 D_1 + 300 D_2$$

For example, for a math/science major, $D_1 = 1$ and $D_2 = 0$, so the equation above can be written more simply as

$$S = 19700 + 400 A + 500 M$$

whereas for a social science major, $D_1 = 0$ and $D_2 = 1$, so the equation above can be written more simply as

$$S = 19200 + 550 A + 600 M$$

Finally, for a humanities major, $D_1 = 0$ and $D_2 = 0$, so the equation above can be written more simply as

$$S = 18900 + 500 A + 450 M$$

**Interaction terms**

In some of the regressions above, there has been a "variable" that is actually the product of two variables. (For example, in the equation for salary, $S$, at the top of this page, there are several such variables: $A \times D_1$, $A \times D_2$, $M \times D_1$, and $150 M \times D_2$. Variables that are themselves the product of two (or more) variables are usually called interaction terms. Including such interaction terms in a regression is a way of allowing for the possibility that the variables included within the interaction term "interact," either magnifying or reducing their joint effect.

The label on prescription medication often contains a warning against consuming certain things (including other medicines) while taking the medication, because these other things may interact adversely with the prescription medication itself, causing an outcome that is dramatically worse than the outcome when either the medication or the other thing is consumed separately, without the other.

Likewise, for example, field of study and mathematical aptitude may interact in terms of their effect on salary: high mathematical aptitude may produce only a small effect on pay if one majors in the humanities, but it may produce a much larger (positive) effect on pay if one majors in the sciences.
Likewise, time spent studying today may interact in a highly positive way with time spent studying last week.

To make the notion of interactions more precise, consider the following regression for test score, \( T \), taking account of time spent studying in week 1 (\( S_1 \)) and time spent studying in week 2 (\( S_2 \)):

\[
T = 45.02 + 3.43 S_1 + 4.17 S_2 + 1.21 (S_1 \times S_2 )
\]

In this regression, the interaction term is, of course, \((S_1 \times S_2 )\). The coefficient on the interaction term is positive (and equal to 1.21), so, intuitively, it seems clear that this is telling us that studying more in either week is somehow "magnified" or increased if we have studied more in the other week. For example, consider the effect of a one-unit increase in just \( S_2 \), time spent studying in week 2, with no change in \( S_1 \). (This is the "other things being equal" effect of an increase in \( S_2 \).) The regression tells us that the new value of \( T \), test score, will be

\[
\text{new } T = 45.02 + 3.43 S_1 + 4.17 (S_2 + 1) + 1.21 (S_1 \times (S_2 + 1))
\]

\[
= 45.02 + 3.43 S_1 + 4.17 S_2 + 1.21 (S_1 \times S_2 ) + 4.17 + 1.21 S_1
\]

(a)

The original value of \( T \) is simply

\[
\text{original } T = 45.02 + 3.43 S_1 + 4.17 S_2 + 1.21 (S_1 \times S_2 ) \quad (b)
\]

Of course, the change in the value of \( T \) is the difference between these two expressions, (a) and (b), and it is already clear that this change is equal to the two terms in boldface type in (a) above, i.e.,

\[
\text{new } T – \text{original } T = 4.17 + 1.21 S_1 \quad (c)
\]

which is the effect of an increase in \( S_2 \), with \( S_1 \) constant.

Now note from (c) that this effect of \( S_2 \) depends on the size of \( S_1 \), the other "study time" variable! In other words, \( S_1 \) and \( S_2 \) interact, and in a good way: the effect of extra \( S_2 \) will be bigger if \( S_1 \) is large rather than small. For example, from (c),

if \( S_1 = 0 \), effect of one more unit of \( S_2 \) on \( T \) is \( 4.17 + 1.21 S_1 = 4.17 + 1.21 \times 0 = 4.17 \)

if \( S_1 = 1 \), effect of one more unit of \( S_2 \) on \( T \) is \( 4.17 + 1.21 S_1 = 4.17 + 1.21 \times 1 = 4.17 \)

if \( S_1 = 2 \), effect of one more unit of \( S_2 \) on \( T \) is \( 4.17 + 1.21 S_1 = 4.17 + 1.21 \times 2 = 4.17 \)

…and so on.

As you can see, in these results, the effect of additional \( S_2 \) will be larger if \( S_1 \) is large – there is a positive interaction between the two variables. Can you calculate the effect of an extra unit of \( S_1 \) on \( T \), and show that it is also positive, and depends on \( S_2 \)? Can you tell a story about two variables that might have a negative interaction in terms of their effect on test score?