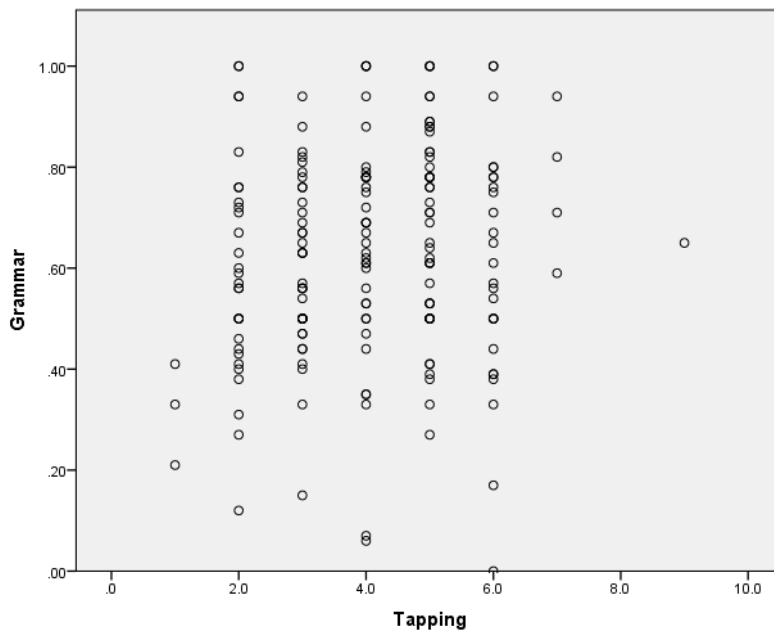


Lecture notes on Multiple Linear Regression

J. Ganger 2019 / SRCD

Section 1: Simple Linear Regression: One independent variable (X) and one dependent variable (Y)

The goal of linear regression is to specify the linear relationship between two variables, X and Y. Let's think about this visually with the scatter plot below, which plots two variables from a language study. It shows GRAMMAR (a measure of passive voice comprehension) against TAPPING (imitating sequences on a xylophone). We might want to know if a spatial memory measure (tapping) is predicts grammar score. Is there a linear relationship between these two variables? Try drawing it.

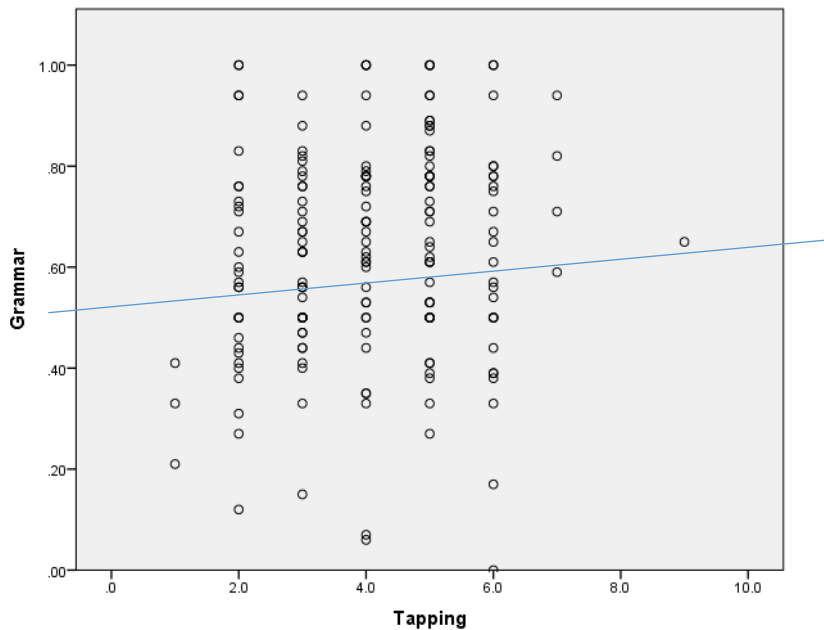


Y = Grammar (percent correct passive voice comprehension); X = Tapping (imitating sequences on small xylophone)

On the plot below is a line based loosely on the line specified a little later in this handout by SPSS (I drew by hand). We can use the formula below to describe the line. B_0 is the y-intercept (the value of y when $x = 0$). B_1 is the slope of the line.

$$Y = B_0 + B_1X_1 + e^1$$

The goal is to find values of B_0 (the y-intercept) and B_1 (the slope of the line) that fit the actual points as closely as possible. This is known as “Regressing Y on X.”



What does “as closely as possible” mean? There are different methods, but the most common is Ordinary Least Squares, which minimizes the sum of the squares of the differences between the *predicted* values and the *observed* values. The predicted values are the values of Y along the line for any given value of X. The observed values are the actual values of Y at that same X.

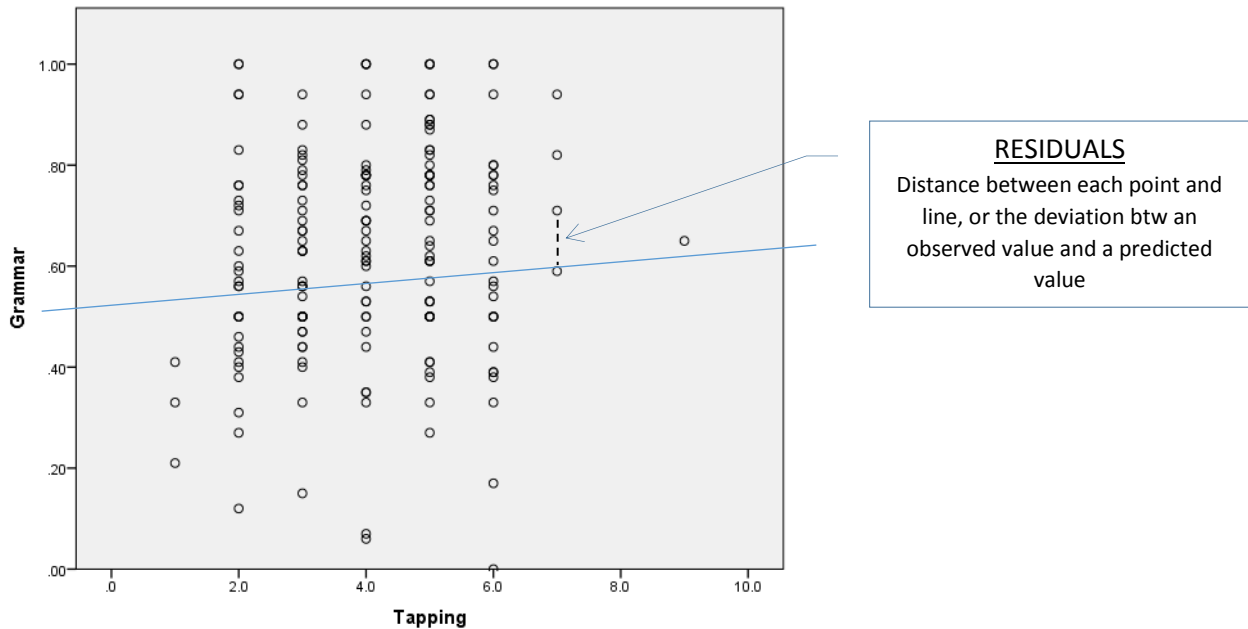
The regression algorithm finds the values of B_0 and B_1 that yield the smallest sum of those squared deviations.

Some Terminology

Regressor: another term for an independent variable in regression.

Regressing y on x: Using x to predict values of Y. In other words, x is the independent variable and Y the dependent variable.

¹ e is error. I will leave it out from now on but you should insert it mentally in every linear equation.



The residuals are the deviations between the observed values and the predicted values, or $(Y_i - \hat{Y}_i)$ s. The bigger they are, the worse a job we did of fitting the line, but even the best fitting line may still leave large residuals if the Xs do not predict Y very well. Those residuals tell us how well the Xs predicted Y, overall.

Key Statistics

B_1 is the slope of the line. How do we interpret it? If $B_1 = .021$...

For one unit increase in X, there is an expected increase of .021 in Y. In terms of our variables, that means that for an increase of one point in TAPPING score, there is an expected increase of .021 in the child's GRAMMAR score.

Beta (β) is the standardized version of B. If $\beta = .146$...

For one standard deviation increase in X (TAPPING score), there is an expected increase of .146 standard deviations increase in Y (GRAMMAR).

R²

R^2 is an estimate of the amount variance in Y that the Xs have accounted for, or the opposite of how large the residuals--the $(Y_i - \hat{Y}_i)$ s--are. There are several ways to define / estimate it.

R^2 is the correlation between Ys and \hat{Y} s. The higher R^2 is, the smaller the residuals, or the closer the fit of the line to the actual data points.

R^2 is also the proportion of the SS of regression over the total SS:

$$\frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

R^2 is also the sum of the products of the correlations (between each IV and the DV) and the beta coefficients of each IV, or:

$$\sum r_i \beta_i$$

Finally, *Adjusted $R^2 = 1 - (1 - R^2) \left(\frac{N-1}{N-k-1} \right)$* where k = number of regressors

t

Each B can also be tested for significance. No matter what the actual value of B, we can still ask whether it makes a statistically significant contribution. That is, we test the null hypothesis that each independent variable is unrelated to the dependent variable and we do this by asking whether the size of its B is significantly different from zero. Another way of saying this is, did we do a better job of fitting the line with that regressor (X) than we did without it? If the t is significant, then the regressor (X) contributes significantly to the fit and we can reject the null hypothesis that it is unrelated.

Section 2: Multiple Linear Regression with Two or More Independent Variables

We can extend this process to any number of Xs. We will use two Xs as an example:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

This time, we need to fit all the β s at once². The kicker is that each one takes the others into account.

We will explore this by running each variable separately in a single-regressor equation like we did in the first section, then running a regression with both to see how the coefficients change. We will compare the following three models:

Model 1: $Y = \beta_0 + \beta_1 X_1$

Model 2: $Y = \beta_0 + \beta_2 X_2$

Model 3: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

(where Y = GRAMMAR; X_1 = TAPPING; X_2 = VERBALMEM)

² $(X'X)B = X'Y \rightarrow B = (X'X)^{-1} X'Y$

	β_1 (TAPPING)	β_2 (VERBALMEM)	R ²
Model 1	.146 ($t = 1.98, p = .049$)		.021
Model 2		.365*** ($t = 5.25, p < .001$)	.133
Model 3	.036 (ns)	.354*** ($t = 4.824, p < .001$)	.135

Why did they change?

Explanation 1: Holding other variables constant.

One way to think about multiple regression (i.e., having more than one independent variable in the regression) is that each β_j tells you the change in Y for a unit change in X_j *while holding the other regressors constant*. For Model 3 in our example, that means for a fixed value of TAPPING (say, TAPPING score = 4.1, the mean), a change of one standard deviation in VERBALMEM is expected to result in .354 standard deviations change GRAMMAR score. In other words, *taking out the effect* of TAPPING, .354 is the (independent) contribution of VERBAL MEMORY.

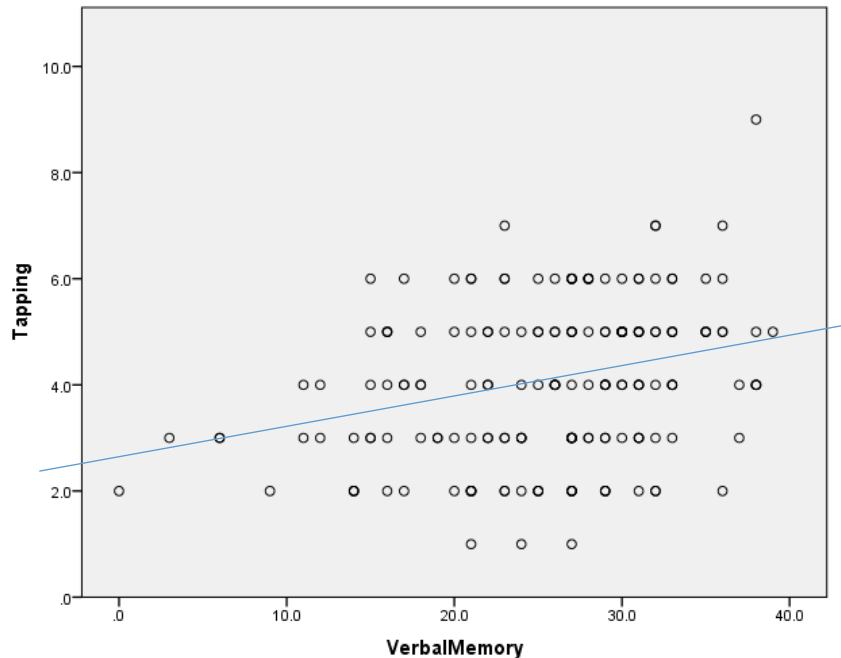
We can make the corresponding statement about the other variable, X_1 , in our example (TAPPING): In Model 3, if VERBAL MEMORY is held constant, then one standard deviation change in TAPPING predicts .036 units change in GRAMMAR (not significant).

How can it be that an independent variable that has a significant first order correlation and a significant β on its own ($\beta_1 = .146$ in Model 1) is now reduced to the point where it does not make a significant contribution? This is in fact the WHOLE POINT of multiple regression. We want to know if TAPPING made a significant contribution to GRAMMAR independent of other factors, such as VERBAL MEMORY. We would conclude at this point that it does not.

Explanation 2: Independent variance

Another way to think about multiple regression is that each independent variable X_j is first regressed on all the others and only the residuals are used in the regression with Y . Personally, I find this description more intuitive.

In our Model 3 above, this would mean that we first regress X_1 on X_2 :



Imagine the actual residuals: $(Y_i - \hat{Y}_i)$'s. (They are $-.44010, .37069, .18147, .18147, -1.00775, -.13390, .86610, -.19697, .80303, -1.32312, -.32312, -1.32312, -1.32312, etc.$) This is the variance left in X_1 (TAPPING) that cannot be explained by X_2 (VERBAL MEMORY). It is this potentially unique variance left in TAPPING that could explain GRAMMAR independently of VERBAL MEMORY.

To see how much of GRAMMAR these residuals can predict, we could now do a regression with GRAMMAR as the DV, or Y , and RESIDS(TAPPING) as the independent variable, that is:

$$\text{GRAMMAR} = B_0 + B_1(\text{RESIDS(TAPPING regressed on VERBAL MEMORY)})$$

And we would obtain the same results (for Tapping) as in Model 3 above.

In regressions with more independent variables, you could do this one variable at a time (regressed on all the others) to derive all the B s.