Revision of Some Basic Mathematical Concepts

Summation Operator and Descriptive Statistics:

Suppose we have a sequence of n numbers with no apparent order, $x_1, x_2, ..., x_n$. Then

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$$

Property 1: For any constant,

$$\sum_{i=1}^{n} c = c + c + ... + c = nc$$

Property 2: For any constant,

$$\sum_{i=1}^{n} cx_i = cx_1 + cx_2 + ... + cx_n = c(x_1 + x_2 + ... + x_n) = c\sum_{i=1}^{n} x_i$$

Property 3: Let $\{(x_i, y_i): i = 1,2,...,n\}$ be a set of n pairs of numbers, and a and b be constants, then

$$\sum_{i=1}^{n}(ax_i + by_i) = \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} by_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$$

With these ideas, we can describe some commonly used descriptive statistics used to describe variables in our samples.

Mean or Average

$$\bar{x} = \frac{(x_1 + x_2 + ... + x_n)}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sum of Deviations from the Mean

There is an obvious property of the mean which might already see,

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i) - \sum_{i=1}^{n}\bar{x} = \sum_{i=1}^{n}(x_i) - n\bar{x} = \sum_{i=1}^{n}(x_i) - n\left[\frac{1}{n}\sum_{i=1}^{n}(x_i)\right] = 0$$

In words, this means that deviations from the mean always sum to 0.

Sum of Squared Deviations from the Mean

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}\left(x_i^{\ 2} - 2x_i\bar{x} + \bar{x}^2\right) = \sum_{i=1}^{n}\left(x_i^{\ 2}\right) - 2\bar{x}\sum_{i=1}^{n}(x_i) + n\bar{x}^2$$

$$= \sum_{i=1}^{n}\left(x_i^{\ 2}\right) - 2\bar{x}[n\bar{x}] + n\left(\bar{x}^2\right) = \sum_{i=1}^{n}\left(x_i^{\ 2}\right) - n\left(\bar{x}^2\right)$$

This can also be written as **Var(*x*)**, in words the variance of x.
This can be generalized to two variables, $\{(x_i, y_i): i = 1,2,...,n\}$.

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i y_i - x_i\bar{y} - y_i\bar{x} + \overline{xy}) = \sum_{i=1}^{n}(x_i y_i) - \bar{x}\sum_{i=1}^{n}(y_i) - \bar{y}\sum_{i=1}^{n}(x_i) + n\overline{xy}$$

$$= \sum_{i=1}^{n}(x_i y_i) - 2n\overline{xy} + n\overline{xy} = \sum_{i=1}^{n}(x_i y_i) - n(\overline{xy})$$

In this form, this can be expressed as **Cov(*x*,*y*)**, and is read as covariance between x and y.

Properties of Linear Functions

A linear function is of the form,

$$y = \beta_0 + \beta_1 x$$

Such a function is said to be linear because on a two dimensional diagram, it is depicted as a straight line, with β_0 the intersect, and β_1 the slope. In econometrics, β_1 also referred to as the marginal effect of x on y. Note that the **marginal effect** is constant. When there is more than one variable, such as in our funding example, we can write the linear function as follows,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

In this case, the coefficients/parameters the slope in the direction of its respective variables, and are referred to as the **partial effects** in this context.

Proportions and Percentages

When we perform empirical analysis, it is important to not only know the effects denominated in the measure of the variable since we want to know the relative impact, or we want to have a basis for comparison. After all, how great is an marginal effect of an additional year of education if it increases your income by $10,000. Consider this, what is the difference of such a marginal effect if you annual income if you terminated your education instead of taking an additional year is $1 million, as opposed to $50,000. If the latter, that additional year of education amounts to a proportionate change of 0.2 or 20%, while in terms of the former it implies a proportionate increase of 0.01 or 1% only!

To formalize the measure, consider for a variable, the initial quantum is x_0, while next level is x_1. Then the **proportionate change** is $\frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0}$, and in percentages (**percentage change**), it is $\frac{x_1 - x_0}{x_0} \times 100 = \frac{\Delta x}{x_0} \times 100 = \%\Delta x$.

There is another additional concept, percentage point change. We have denominated of variable x in terms of nominal currency. What if x were already in percentages? There is no change in how we think of percentage change as described above. However, our proportionate change now is altered a little. Consider our earlier example of funding and GPA. Suppose we find without the funding, the number of student with a 4.0 GPA was 5%, but after the funding experiment, we see that percentage increase to 8%. We say that funding increases the percentage of students who get a 4.0 GPA by 3 **percentage points**, or a 60% increase.

Quadratic Functions

What if we think the linear model is not a satisfactory model. An example of this is when we think of our wages over time. A priori, we know that the trajectory upon graduation from university, your income will maintain a steady increase. However, as we tend towards retirement, our income may begin to fall as a result of lower productivity. In which case, if we draw a diagram of the wage-time relationship, it will not follow a linear form. We can more accurately depict this relationship with a quadratic function such as the following,

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

where based on the above argument, β_1 (the slope) will be greater than 0 and β_2 (curvature) will be less than 0. **Why?** In general this equation says that x has a diminishing marginal effect on y. If however β_2 is greater than 0, x will have a increasing marginal effect on y. **Why?**

Natural Logarithmic Functions

Another way we can depict a diminishing marginal effect of x on y is through the use of the natural logarithmic function. That is

$$y = \beta_0 + \beta_1 \ln(x)$$

In this form, β_1 has a new interpretation, i.e. it is no longer the slope. Why? To see why,

$$\frac{\Delta y}{\Delta x} = \frac{\beta_1}{x} \Rightarrow \frac{\Delta y}{\Delta x / x} = \beta_1$$

So that we say that for every 1% change in x, leads to a β_1 change in y (measure in y's unit of measure). This is different from the linear or quadratic model, since there we say a unit (x's unit of measure) change in x, leads to a β_1 change in y (measure in y's unit of measure). Note that in this logarithmic form, it is easy to calculate the elasticity which is defined as $\frac{\Delta y / y}{\Delta x / x} = \frac{\Delta y}{\Delta x} \times \frac{x}{y}$.

You will often see the following log model,

$$\ln(y) = \beta_0 + \beta_1 \ln(x)$$

Here, β_1 is interpreted as elasticity, which is far more convenient than the former log model since there is no need for additional calculations. To see why,

$$\frac{\Delta y}{\Delta x}\frac{1}{y} = \frac{\beta_1}{x} \Rightarrow \frac{\Delta y / y}{\Delta x / x} = \beta_1$$

So this last model is often referred to as the **constant elasticity model**.

What if y were in logs, while x is not? How would you interpret β_1 then?

Differential Calculus
Differentiation:
The exposition on Differential Calculus here is necessarily sparse since this is not a course on calculus. See me if you want some additional references for self study.

1. **Constant Function Rule**: If the function is as follows, $y = f(x) = k$, where k is a constant in R^1. Then $\frac{dy}{dx} = \frac{df(x)}{dx} = \frac{dk}{dx} = f'(x) = f_x(x) = 0$. Where $f'(x)$ and $f_x(x)$ are just different way of describing the same process.
2. **Power Function Rule**: If $y = f(x) = kx^n$, where n is a real number in R^1, and k is as before. Then $\frac{dy}{dx} = \frac{df(x)}{dx} = \frac{dkx^n}{dx} = f'(x) = f_x(x) = knx^{n-1}$.
3. **Log-Function Rule**: Let $y = f(x) = \ln(x)$, then $\frac{dy}{dx} = \frac{df(x)}{dx} = \frac{d\ln(x)}{dx} = \frac{1}{x}$. And more generally let $y = \ln(g(x))$, then $\frac{dy}{dx} = \frac{d\ln(g(x))}{dx} = \frac{g_x(x)}{g(x)} = \frac{g'(x)}{g(x)}$.
4. **Exponential-Function Rule**: Let $y = f(x) = e^x$, then $\frac{dy}{dx} = \frac{df(x)}{dx} = \frac{de^x}{dx} = e^x$. And more generally let $y = e^{g(x)}$, then $\frac{dy}{dx} = \frac{de^{g(x)}}{dx} = g_x(x)e^{g(x)} = g'(x)e^{g(x)}$.
5. **Sum-Difference Rule**: Let there be two functions, $y = f(x)$ and $z = g(x)$, where f and g are different functions. Then
$$\frac{d[y \pm z]}{dx} = \frac{d[f(x) \pm g(x)]}{dx} = \frac{df(x)}{dx} \pm \frac{dg(x)}{dx} = f'(x) \pm g'(x) = f_x(x) \pm g_x(x)$$
6. **Product Rule**: For the same two functions, this rule says that
$$\begin{aligned} \frac{d[y \times z]}{dx} &= \frac{d[f(x) \times g(x)]}{dx} \\ &= f(x)\frac{dg(x)}{dx} + g(x)\frac{df(x)}{dx} \\ &= f(x)g'(x) + g(x)f'(x) \\ &= f(x)g_x(x) + g(x)f_x(x) \end{aligned}$$

7. **Quotient Rule**: For the same two functions, this rule says that

$$\frac{d\left[y/z\right]}{dx} = \frac{d\left[f(x)/g(x)\right]}{dx}$$
$$= \frac{g(x)\frac{df(x)}{dx} - f(x)\frac{dg(x)}{dx}}{\{g(x)\}^2}$$
$$= \frac{g(x)f_x(x) - f(x)g_x(x)}{\{g(x)\}^2}$$
$$= \frac{g(x)f'(x) - f(x)g'(x)}{\{g(x)\}^2}$$

8. **Chain Rule**: Let there be two functions such that $y = f(x)$ and $x = g(w)$. Then

$$\frac{dy}{dw} = \frac{dy}{dx}\frac{dx}{dw} = \frac{df(x)}{dx}\frac{dg(w)}{dw} = f_x(x)g_w(w) = f'(x)g'(w)$$

9. **Partial Differentiation**: So far we have worked with functions with one variable. Consider now instead now $y = f(x_1, x_2, ..., x_n)$. The differentiation operation remains the same, and is performed holding all other variables you are not differentiating with respect to constant. However, we write it a little differently.

$$\frac{\partial y}{\partial x_i} = \frac{\partial f(x_1, x_2, ..., x_i, ..., x_n)}{\partial x_i} = f_{x_i}(x_1, x_2, ..., x_i, ..., x_n)$$