## *1. Statistics review:*

Check for understanding—basic statistics (first try for yourself, then I'll put you in pairs)

| Symbol | Name | Formula |
|---|---|---|
| $\bar{x}$ | | |
| $var(x)\ or\ \sigma_x^2$ | | |
| $cov(x, y)$ | | |

Here is a population consisting of 4 people, 2 with x=1 and 2 with x=2:

| x (education) | u (spunk) |
|---|---|
| 1 | 8 |
| 1 | -10 |
| 2 | -4 |
| 2 | 6 |

The expectation of $u$, which we call $E(u)$, is: ____________. (Calculate the number)

The CONDITIONAL expectation of $u$, given $x = 1$, which we call $E(u|x = 1)$, is: ____________.

The CONDITIONAL expectation of $u$, given $x = 2$, which we call $E(u|x = 2)$, is: ____________.

**The point:** When we're asked for the expectation of $u$, we're being asked our best guess of what $u$ will be if we pick a person at random from the entire population. When we're asked for the CONDITIONAL expectation of $u$ given $x = 1$, we're being asked for our best guess of what $u$ will be if we pick a person at random only from the population with $x = 1$. That's all a conditional expectation is.

## *2. Mechanics of regression:*

Remember that OLS fits a line through a set of points by minimizing the sum of squared vertical distances from the regression line to the data points ("errors" or "residuals"). Let's write out the formulae for the OLS parameters one more time:

| | |
|---|---|
| $\hat{\beta}_1 =$ | $\hat{\beta}_0 =$ |

**The point:** These are just formulae derived mechanically from our criterion of minimizing squared errors. We can always do this with a set of data. What we want to know is when it makes sense to do this. That's where we have to start making assumptions, like in the previous lecture.

### *3. Interpreting $\beta_1$ (slope):*

To get more comfortable with logarithms in regression, let's do an example:
We want to see how food consumption (y) measured in $/year is related to household income (x) measured in $/year. How would we interpret each of the following regressions?

| Name | Functional Form | Interpretation in Words |
|---|---|---|
| linear ("constant returns") | $y = \beta_0 + \beta_1 x$ | Ceteris paribus, when income increases by _____________, food consumption increases by ________________. |
| log ("decreasing returns") | $y = \beta_0 + \beta_1 logx$ | Ceteris paribus, when income increases by _____________, food consumption increases by ________________. |
| log-linear ("increasing returns") | $logy = \beta_0 + \beta_1 x$ | Ceteris paribus, when income increases by _____________, food consumption increases by ________________. |
| log-log ("constant elasticity") | $logy = \beta_0 + \beta_1 logx$ | Ceteris paribus, when income increases by _____________, food consumption increases by ________________. |

If you don't like logarithms, that is okay. Just remember how to interpret regressions when logs are present.

### *4. When is ceteris really paribus?:*

We know that we can mechanically draw the OLS line just by using our formulae above, no assumptions required. But that doesn't allow us to use our results to make *ceteris paribus* statements like "all else equal, when income increases by $1/year, food consumption increases by $0.70/year." Now we'll see the minimum assumptions we must make in order for these statements to be valid.

**Assumption 1 (hidden in plain sight):**

The relationship between the dependent and independent variables is actually linear:

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

where *y* or *x* could be the logarithm of something.

**Assumption 2 (very weak, not a big deal):**

The expected value of the unobservable variable *u* is zero:

$$E(u) = 0$$

Like we heard in lecture, we can just re-center *u* in our minds to make it mean-zero, and that will do the trick.

**Assumption 3 (big deal):**

No matter what *x* is, we expect *u* to be zero:

$$E(u|x) = 0$$

This goes back to the discussion of conditional expectations. We're saying that the observable variable (*x*) and the unobservable variables (*u*) really have nothing to do with each other. Can you think of any situation where that assumption is likely to be true?

***Without these assumptions holding, making "ceteris paribus" statements involving $\beta_1$ is not justified!***