Formative Assessment #1

Topics:

- ANOVA vs regression
- Predicted slopes (the three-step process)

Q1: Are there any topics that we've covered so far (including general linear models) that you'd like to review more in class?

- Link functions
- Centering predictors

3 Parts of Generalized Linear Models

1. Non-Normal Conditional Distribution of y_i ⟸ **2. Link Function** = **3. Linear Predictor of Fixed (and Random) Effects**

2. **Link Function = $g(\cdot)$**: How the conditional mean to be predicted is transformed so that the model directly predicts an **unbounded** outcome

Agresti (2015): "…the *link function,* connects the random component with the linear predictor." (p.3)

Random component = conditional distribution of the outcome

Linear predictor: the model

$$\eta_i = \sum_{j=1}^{p} \beta_j x_{ij}, \quad i = 1, \ldots, n. \qquad g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij}, \quad i = 1, \ldots, n.$$

$$\eta_i = g(\mu_i)$$

Intercept ="You are Here" Sign of Data

With **original** years of education:

x_i =**educ,** y_i = **income**

$$y_i = \beta_0 + \beta_1(x_i) + e_i$$

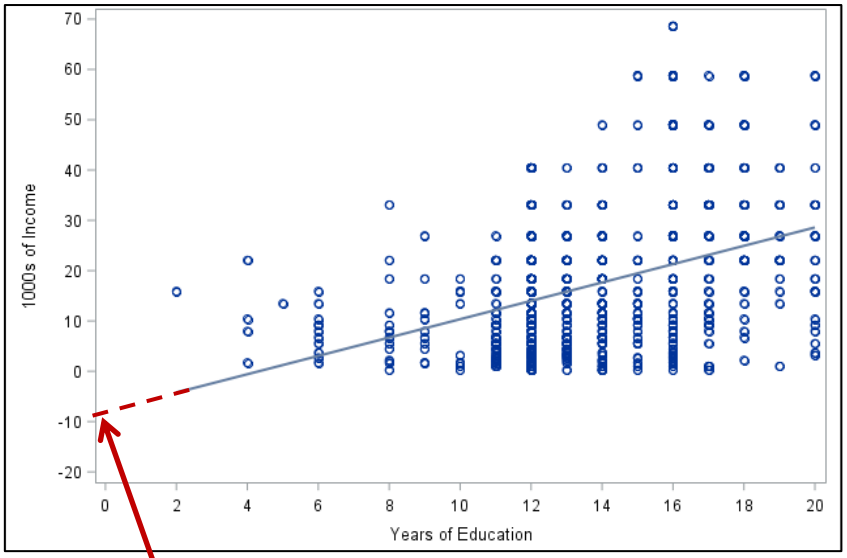
$$y_i = -7.89 + 1.82(x_i) + e_i$$

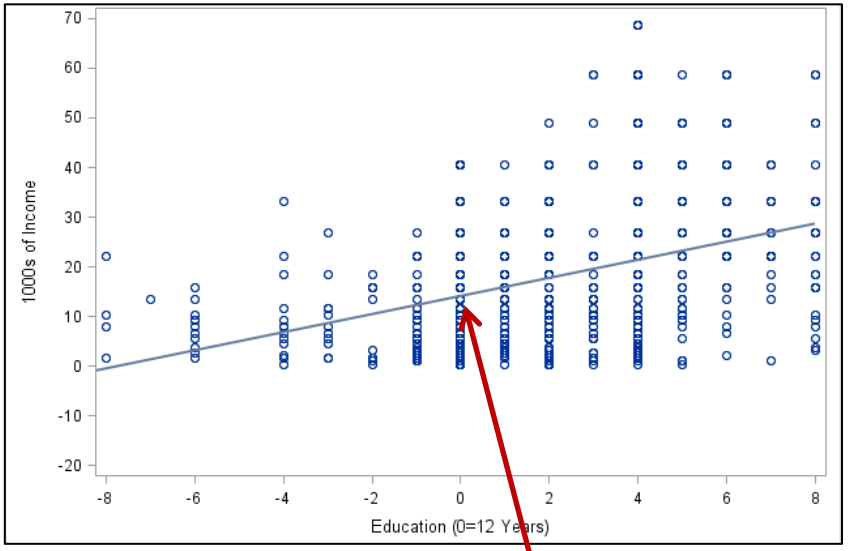
With **centered education** ($c = 12$):

x_i =**educ**-12, y_i = **income**

$$y_i = \beta_0 + \beta_1(x_i) + e_i$$

$$y_i = 14.00 + 1.82(x_i) + e_i$$





Intercept β_0

There is no *wrong* way to center, only *weird*. **Center so x_i=0 is meaningful.**

Intercept β_0

Q2: "Regression is for prediction, whereas ANOVA is for testing mean differences"

Flavors of General Linear Models

- Unlike any other family of statistical models, **the same General Linear Model is called different names** (often taught in different classes) based on **what kind of predictor variables** are included:
  - One quantitative predictor? "Simple (linear) regression"
  - 2+ quantitative predictors? "Multiple (linear) regression"
  - One categorical predictor with two groups? "Independent *t*-test"
  - One categorical predictor with 3+ groups? "One-way ANOVA"
  - 2+ categorical predictors (with interactions between them)? "Two-way (or more-way) ANOVA"
  - 2+ categorical predictors (with interactions between them) and 1+ quantitative predictors (without interactions with the categorical predictors)? "Two (or more)-way ANCOVA"
  - Whatever combination is necessary? "Multiple regression"
- These distinctions only serve to confuse people an what is **just one model**, the General Linear Model

Q2: "Regression is for prediction, whereas ANOVA is for testing mean differences"

- ANOVA is a special case of regression.
- The choice depends on the specific research questions and the type of data [the predictors] we have.

Q2: "Regression is for prediction, whereas ANOVA is for testing mean differences"

- ANOVA is a special case of regression.
- The choice depends on the specific research questions and the type of data [the predictors] we have.
- …you can get the same information from a regression or an ANOVA, so it doesn't matter which you run.
- …both [are] used for prediction, but with different kinds of predictor variables.

Q2: Write the equation for the predicted G slope for someone with an F=2

$$y_i = \beta_0 + \beta_1(F_i) + \beta_2(G_i) + \beta_3(H_i) + \beta_4(I_i) + \beta_5(F_i)(G_i) + \beta_6(F_i)(H_i) + e_i$$

G slope = beta2(G) + beta5(F)(G)

[beta2 + beta5(2)](G)

Q3: Write the equation for the predicted F slope for someone with an G=3, H=4, and I=5

$$y_i = \beta_0 + \beta_1(F_i) + \beta_2(G_i) + \beta_3(H_i) + \beta_4(I_i) + \beta_5(F_i)(G_i) + \beta_6(F_i)(H_i) + e_i$$

F slope = beta1(F) + beta5(F)(G) + beta6(F)(H)

[beta1 + beta5 (3) + beta6 (4)](F)

Q4: Write the equation for the predicted I slope for someone with an F=2, G=3, and H=4

$$y_i = \beta_0 + \beta_1(F_i) + \beta_2(G_i) + \beta_3(H_i) + \beta_4(I_i) + \beta_5(F_i)(G_i) + \beta_6(F_i)(H_i) + e_i$$

I slope = beta4(1)

Formative Assessment #1

Topics:

- ANOVA vs regression
- Predicted slopes (the three-step process)