Time for a quick game…
The Stats Are Right!
Rules:

- Room gets divided down the middle
- Class views a question at a time on the slide.
- The side of the room that answers first gets a point.
- No student can answer a question for their 'side of the room' more than once. It must be a different person each time.
- 8 questions total: all to do with the Circulatory Arrest peri-operative study
- 1ish minute per question
- Good luck!

The Stats are Right!

- **Question 1: True of False**

- In order to evaluate effect modification of birthweight by diagnosis in modeling the outcome PDI, you need to include the variable birthweight*diagnosis, but you don't need to include the variables birthweight or diagnosis.

Answer: False

The Stats are Right!

- **Question 2: True of False**

- Cook's Distance incorporates mainly 'leverage' in its calculation of overall influence.

Answer: False

The Stats are Right!

- **Question 3: True of False**

- In order to determine the potential presence of alternatives to linearity in the association between circulatory arrest (in minutes) and PDI score, we could include a polynomial version of circulatory arrest (in minutes) in our model that already includes circulatory arrest (minutes), and if it's significant we could then infer that there is evidence of nonlinearity in the relationship with PDI.

Answer: True

The Stats are Right!

- **Question 4: Short Answer**

- What's another way to evaluate the question of linearity of an independent variable (predictor) in your model that is a different method than the one employed in the previous question? (ie don't use any model coefficients)

Answer: Several include Scatterplot, Lowess, residual analysis—there are many

The Stats are Right!

- **Question 5: True or False**

- We would assess whether the variable 'diagnosis' confounds the relationship between circulatory arrest (in minutes) and PDI score by including the following variable in our model: diagnosis*circulatory arrest (minutes), then checking for a >10% change in the coefficient of circulatory arrest (in minutes).

Answer: False—the variable as defined above is an interaction term designed to assess potential effect modification, NOT confounding. Check course notes on how we check for potential confounding (ie Most importantly, is there clinical reason to suspect diagnosis to be a confounder? Also, compare coefficient for CA minutes both with and without diagnosis in the model…is change >10% rule of thumb?).

The Stats are Right!

- **Question 6: Choose one**

- Let's say we 'binned' birthweight into 4 categories. If we are interested in determining whether or not there is a birthweight trend with outcome PDI, would we want to treat our new categorical birthweight variable as nominal or ordinal?

- Bonus point: List a downside to binning!

Answer: ordinal since we are wanting to assess trend

Bonus Answer: Loss of power and information, may not define 'bins' in meaningful enough way inadvertently, etc

The Stats are Right!

- **Question 7: True or False**

- We will use ANOVA to consider an association between PDI and diagnosis.

- Bonus point: What does ANOVA stand for?

Answer: True--diagnosis is a categorical variable and PDI is considered continuous, so as long as our normality assumptions are met (required by ANOVA as well as linear regression) for such variables, and n is large enough, ANOVA is appropriate.

Bonus Answer: Analysis of Variance

The Stats are Right!

- **Question 8: Shortish Answer**

- Why would we choose a linear model instead of ANOVA in evaluating a possible association between diagnosis and PDI score? (please list 2 reasons)

Answer: (1) First, regression is the statistical model that you use to predict a continuous outcome on the basis of one or more continuous or categorical predictor variables. In contrast, ANOVA is the statistical model that you use to predict a continuous outcome on the basis of one or more categorical predictor variables. Our diagnosis variable is categorical, so ANOVA is appropriate when we simply want to compare treatment groups, but it just depends on what questions you are asking. The main reason we would choose regression here is if we want to include other potential *continuos* covariates in the model with diagnosis, or include numerous additional covariates in general—we can look then at confounding and effect modification and simply build a more extensive (and potentially meaningful) model; (2) Also to assess potential trend relationship and be able to use coefficient estimates from the regression in order to draw inferences about that relationship (ie for one level increase in diagnosis, PDI score changes by how much, etc). ANOVA is more often used for group comparisons, and not so much assessing trend in relationship. (3) Additional reasons include that with linear regression we then can do more diagnostics on residuals, and employ the regression framework and machinery.

The Stats are Right!

- **Thanks for playing!**
- **Now back to flexible modeling --**