

[illegible]

PRODUCER:

DR. MARIO

IMG CREDIT: ALEX RIEGERT-WATERS

Categorical Predictors

- Categorical Variable with **More Than Two** Options
- Suppose there are K Different Categories for X
- We Need $K-1$ Different Indicator Variables
- The Category Not Included is Called the **Reference Category**
- Intercept (Constant Term) Is the Mean of Reference Category

Categorical Predictors

- Suppose X has 4 Different Levels (A, B, C, D)
- We Create 3 Binary Indicator Variables

$$A = \begin{cases} 0 & \text{if not A} \\ 1 & \text{if A} \end{cases} \quad B = \begin{cases} 0 & \text{if not B} \\ 1 & \text{if B} \end{cases} \quad D = \begin{cases} 0 & \text{if not D} \\ 1 & \text{if D} \end{cases}$$

- Full Model Involves All Three Variables

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 D + \epsilon$$

- Individual t-Tests for A, B, D Are Testing Difference in Mean of Y Compared to the Reference Category C (When $A=B=D=0$)

Example: Predicting Birth Weight of Baby

- Data **Ncbirths** in **Stat2Data** Package
- Response Variable: *BirthWeightOz* = Birth Weight in Ounces
- Predictor Variables
 - *Weeks* = Completed Weeks of Gestation
 - *Plural* = {1 = Single, 2=Twins, 3=Triplets}
 - *MomRace* = {Black, Hispanic, Other, White}

Supplement for Lecture 18

- Fit Models Based Off Mom's Race
- Add in Weeks Variable to See Impact On Model
- Add in Plural Variable to See Impact on Model
- Stepwise Model Selection

Thank You

Make Reasonable Decisions

