10 November

Tuesday, November 10, 2020 11:28 AM

Regression w/ dummy variables → categorical variables
→ binary variables

$wage = \beta_0 + \beta_1(female) + \beta_2(educ) + \varepsilon$

female = 1

if male, $wage_{male} = \beta_0 + \beta_2(educ) + \varepsilon_m$

if female, $wage_{fem} = \beta_0 + \beta_1 + \beta_2(educ) + \varepsilon_f$

for given error term,

$wage_F - wage_m = \beta_1$

β_1 = difference in wage for F and M given equal education

What about in log form?

log(price) = −1.35 + .168(lotsize)
+ .707(log(sqFt))
+ .027(bedrooms)
+ .054(colonial) → dummy for style → 1 if colonial

for given lot size, sq ft, and bedrooms, colonial sells for 5.4% higher

Dummies w/ multiple categories

↳ more than 2 categories

$wage = \beta_0 + \beta_1(educ) + \beta_2(exp) + \beta_3(hotness)$

↳ homely, plain, average, good, stricking

aggregate the categories
below average
above average
average → reference category

Men → $\log(wage) = \beta_0 - .164(below) + .16(above) + \beta_3(educ) + \varepsilon$

below average men earn 16.4% less than average if all else is equal

F-test

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_k = 0$

None of the explanatory variables have an effect on y

H_a: at least on β is different from zero

overall significance of model