

ECN 3972

Lab 4

Import data “EDearn.xls” into R.

Make sure you call up the packages mosaic and psych:

```
library(mosaic)
```

```
library(psych)
```

We’re going to run several different linear models in R, adding independent variables each time.

Model 1:

$$earn_i = \beta_0 + \beta_1 Exp_i + \varepsilon_i$$

where:

$earn_i$ = earnings of person i

Exp_i = experience of person i

In R:

```
model1 = lm(Earn~Exp, data=EDearn)
```

```
summary(model1)
```

Now, we let’s add other independent variables. In this case, we will add a set of dummy variables that indicate the highest degree the person earned. We can choose from:

$LTHS_i = 1$ if person i ’s highest education is less than a high school degree; = 0 otherwise

$HS_i = 1$ if person i ’s highest education is a high school degree; = 0 otherwise

$BA_i = 1$ if person i ’s highest education is a bachelor’s degree; = 0 otherwise

$MA_i = 1$ if person i ’s highest education is a master’s degree; = 0 otherwise

$PHD_i = 1$ if person i ’s highest education is a doctorate degree; = 0 otherwise

We have to decide which of these to omit from our regression. **Remember that the omitted category is the category for which the other dummy variables are compared.**

Let’s omit $LTHS_i$ from the model:

Model 2:

$$earn_i = \beta_0 + \beta_1 Exp_i + \beta_2 HS_i + \beta_3 BA_i + \beta_4 MA_i + \beta_5 PHD_i + \varepsilon_i$$

where:

$earn_i$ = earnings of person i

Exp_i = experience of person i

HS_i = 1 if person i 's highest education is a high school degree; = 0 otherwise

BA_i = 1 if person i 's highest education is a bachelor's degree; = 0 otherwise

MA_i = 1 if person i 's highest education is a master's degree; = 0 otherwise

PHD_i = 1 if person i 's highest education is a doctorate degree; = 0 otherwise

In R:

```
model2 = lm(Earn~Exp + HS + BA + MA + PHD, data=EDearn)
```

```
summary(model2)
```

Let's try a different omitted category. Let's omit PHD_i . **Remember that this changes how we interpret our coefficients.**

Model 3:

$$earn_i = \beta_0 + \beta_1 Exp_i + \beta_2 LTHS_i + \beta_3 HS_i + \beta_4 BA_i + \beta_5 MA_i + \varepsilon_i$$

where:

$earn_i$ = earnings of person i

Exp_i = experience of person i

$LTHS_i$ = 1 if person i 's highest education is less than a high school degree; = 0 otherwise

HS_i = 1 if person i 's highest education is a high school degree; = 0 otherwise

BA_i = 1 if person i 's highest education is a bachelor's degree; = 0 otherwise

MA_i = 1 if person i 's highest education is a master's degree; = 0 otherwise

In R:

```
model3 = lm(Earn~Exp + LTHS + HS + BA + MA, data=EDearn)
```

```
summary(model3)
```