MLR

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_k x_k + \varepsilon$$

β_0 = intercept (value of y when $x = 0$)

$\beta_1 \dots \beta_k$ - slope parameters for $x_1 \dots x_k$.

β_1 - effect of x_1 on y when all other factors are held constant.

(OR) change in y for unit change in x

Assumptions

1. $y \mid x_1, x_2 \ldots x_k$ is normally distributed

2. Observations are independent

3. $E(\varepsilon \mid x_1, x_2 \ldots x_k) = 0$

4. No correlation b/w ε and x's

5. No linear relation b/w independent variables.

6. Homoscedasticity (Constant Variance)

$$Var(\varepsilon \mid x_1 \ldots x_k) = \sigma^2$$

Goodness of fit of MLR

$$R^2 = \frac{SSR}{SST}$$

total variation explained by the model

$0 \le R^2 \le 1$ (OR) $0\% \le R^2 \le 100\%$

Higher the R^2 the better the fit

No relation (linear) between X's

$$trips = \beta_0 + \beta_1(\#\text{adults}) + \beta_2(\#\text{ of kids}) + \beta_3(\text{hhld size}) + \beta_4(\text{educ level}) + \varepsilon$$

$$\boxed{\text{hhld size} = \#\text{adults} + \#\text{ of kids}}$$

$$\text{Consumption} = \beta_0 + \beta_1(\text{income}) + \beta_2(\text{income})^2 + \varepsilon$$

$$\frac{\Delta \text{cons}}{\Delta \text{inc}} = \beta_1 + 2\beta_2(\text{income})$$

intercept

slopes

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k$$

$$H_0 : \beta_1 = 0$$

$$\text{wage} = \beta_0 + \beta_1(\text{educ})$$

	coefficient	Std error	t-stat	p-value
Constant	10			
educ	0.50			0.001

If $p < 0.05$ reject H_0

Regression with Dummy Variables

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

female $= 1$ if person is female

$= 0$ if person is male

If person is male

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

If person is female

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

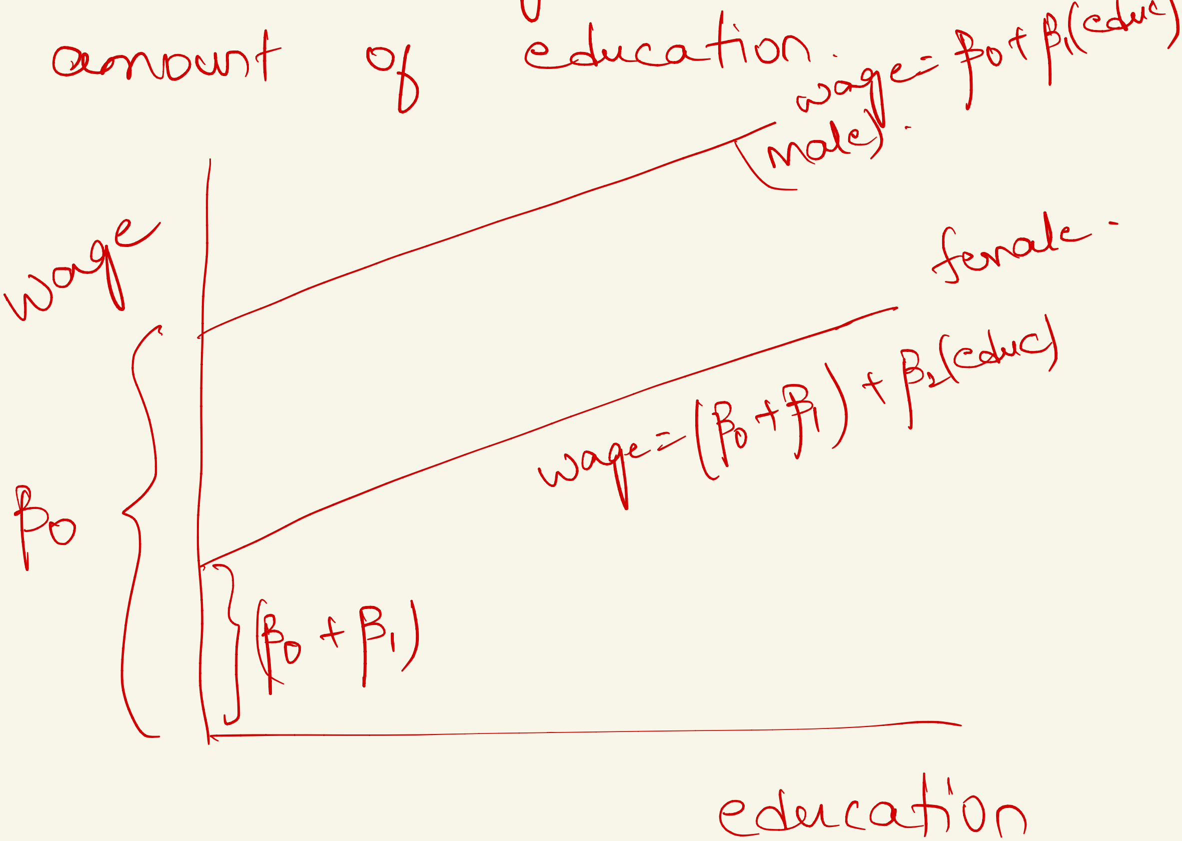
For given education and error term

$$wage_{female} - wage_{male}$$

$$= \cancel{\beta_0} + \beta_1 + \cancel{\beta_2(educ)} + \cancel{\varepsilon_f} - \left[\cancel{\beta_0} + \cancel{\beta_2(educ)} + \cancel{\varepsilon_m}\right]$$

$$= \beta_1$$

β_1 represents the difference in hourly wage between females and males given the same amount of education.



$$\log(\text{Price}) = -1.35 + 0.168(\text{lot size}) + 0.707(\log(\text{sqfootage})) + 0.027(\#\text{ of bedrooms}) + 0.054(\text{colonial})$$

Colonial – dummy variable for style of the house.

= 1 if house style is colonial

= 0 if " " " not colonial

For given lot size, square footage and number of bedrooms a colonial house sells for 5.4% higher price.

Regression with Multiple Categories.

Categorical Variables with more than 2 categories.

$$\boxed{wage = \beta_0 + \beta_1(educ) + \beta_2(experience) + \beta_3(\text{physical attractiveness})}$$

Physical attractiveness

below average
- homely
- Plain

average

above average
- good looking
- strikingly beautiful (or) handsome

You cannot include all the 3 categories. One should be left out.

below average, average, above average

average - <u>reference</u> category

Men

$$\log(wage) = \beta_0 - 0.164(\text{below avg}) + 0.016(\text{above avg}) + \beta_3(educ) + \varepsilon$$

Men with below average looks earn 16.4% less than those with average looks when everything else is the same.

Men with above average looks earn 1.6% more than those with average looks when everything else is the same.

$$\log(wage) = \beta_0 + \beta_1(ethnicity)$$

ethnicity = white

Asian

Hispanic

African American.

$$\log(wage) = \beta_0 + \beta_1(white) + \beta_2(Hispanic) + \beta_3(educ) + \ldots$$

0,1,0

$$\log(wage) = \beta_0 + \beta_1(Asian) + \beta_3(educ) + \ldots$$

$$\begin{array}{c}\text{Wage} \\ 50 \\ 100 \\ 120\end{array} \quad \begin{array}{ccc}\text{Asian} & \text{White} & \text{Hispanic} \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1\end{array} \quad \begin{array}{c} \\ = 1 \\ 1 \\ \\ \end{array}$$

β_1

F-test

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

None of the explanatory variables has an effect on y.

H_A : Atleast one of the β's is different from zero.

Overall significance of regression model.

If we reject H_0 : evidence that any of the independent variables has an effect on y

If we fail to reject H_0 : No evidence that independent variables affect y