Questions 7 - 11

Bookmark this page

Now, we are interested in estimating the following model:

$$\log(wage_i) = \beta_0 + \beta_1 black + \varepsilon_i$$

Question 7

1/1 point (graded)

Researcher A says that this model is not correctly specified. Researcher A suggests that the correct model should estimate the following equation (where $other\ race$ is a dummy variable equal to 1 when the person is not black):

$$log(wage_i) = \beta_0 + \beta_1 black + \beta_2 other\ race + \varepsilon_i$$

Researcher B claims that Researcher A is wrong, and that in this second model, it is not possible to separately identify $\beta_0,\ \beta_1,\ \text{and}\ \beta_2.$ Who do you agree with?

- a. Researcher A

b. Researcher B ✔

Explanation

The model proposed by Researcher A has the problem of multicollinearity. In particular we have that $other\ race + black = 1$ which is the vector we use to estimate the intercept β_0. Thus, Researcher B is right -- it is not possible to separately identify $\beta_0,\ \beta_1, \text{and}\ \beta_2$.

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Question 8

1.0/1.0 point (graded)

Assume that you don't have all the data. However, you know that the sample mean of the log wage for women who are not black is $\overline{y}_{other}$, and the sample mean of the log wage for black women is $\overline{y}_{black}$. What are the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ if we run this model using OLS?

a. We have that $\hat{\beta}_0 = \overline{y}_{black}$ and $\hat{\beta}_1 = \overline{y}_{other} - \overline{y}_{other}$

b. We have that $\hat{\beta}_0 = \overline{y}_{black}$ and $\hat{\beta}_1 = \overline{y}_{black} - \overline{y}_{other}$

[The Multivariate Linear Model](#)
due Nov 28, 2016 05:00 IST

Module 9: Homework
due Nov 21, 2016 05:00 IST

Exit Survey

- c. We have that $\hat{\beta}_0 = \overline{y}_{other}$ and $\hat{\beta}_1 = \overline{y}_{black} - \overline{y}_{other}$ ✔
- d. We have that $\hat{\beta}_0 = \overline{y}_{other}$ and $\hat{\beta}_1 = \overline{y}_{other} - \overline{y}_{other}$

Explanation

This was discussed in the lecture. In general, we have that since $\mathbb{E}\varepsilon_i = 0$

$$\mathbb{E}\left[lwage|black = 0\right] = \beta_0 + \beta_1 \times 0 + 0 = \beta_0$$

$$\mathbb{E}\left[lwage|black = 1\right] = \beta_0 + \beta_1 \times 1 + 0 = \beta_0 + \beta_1$$

Thus, the sample analogues must satisfy:

$$\overline{y}_{other} = \hat{\beta}_0$$

$$\overline{y}_{black} = \hat{\beta}_0 + \hat{\beta}_1 = \overline{y}_{other} + \hat{\beta}_1 \iff \hat{\beta}_1 = \overline{y}_{black} - \overline{y}_{other}$$

Submit

You have used 1 of 2 attempts

Question 9

2.0/2.0 points (graded)

Now, estimate this model by yourself using both the sample means approach and the regression approach with the command **lm**. (You should get the same results!)

For the following answers, please round to the third decimal place, i.e. if the solution is 0.23412, please round to 0.234, and if it is 0.23498, please round to 0.235.

What value did you find for $\hat{\beta}_0$?

1.912 ✔ **Answer:** 1.912

1.912

What value did you find for $\hat{\beta}_1$?

-0.166 ✔ **Answer:** -0.166

-0.166

Explanation

If we run the following code:

```
#dummy variables
meanother <- mean(nlsw88$lwage[nlsw88$black == 0])
meanblack <- mean(nlsw88$lwage[nlsw88$black == 1])
meanother
meanblack - meanother

dummymodel <- lm(lwage ~ black, data = nlsw88)
summary(dummymodel)
```

This is the output we get:

```
> #dummy variables
> meanother <- mean(nlsw88$lwage[nlsw88$black == 0])
> meanblack <- mean(nlsw88$lwage[nlsw88$black == 1])
> meanother
[1] 1.911614
> meanblack - meanother
[1] -0.1655357
>
> dummymodel <- lm(lwage ~ black, data = nlsw88)
> summary(dummymodel)

Call:
lm(formula = lwage ~ black, data = nlsw88)

Residuals:
     Min       1Q   Median       3Q      Max
-1.90667 -0.40290 -0.03418  0.37105  1.96129

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.91161    0.01398 136.739  < 2e-16 ***
black       -0.16554    0.02744  -6.033 1.88e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5701 on 2244 degrees of freedom
Multiple R-squared:  0.01596,	Adjusted R-squared:  0.01552
F-statistic: 36.39 on 1 and 2244 DF,  p-value: 1.88e-09
```

Submit You have used 1 of 2 attempts

Question 10

1.0/1.0 point (graded)

A critic is claiming that this doesn't prove that there are differences in the wage of black women and women of other races. You decide to conduct a test on the parameter β_1, where the null hypothesis is $\beta_1 = 0$. What is the value of the statistic of this test?

Please round to the third decimal place, i.e. if the solution is 0.23412, please round to 0.234, and if it is 0.23498, please round to 0.235.

-6.033 ✔ **Answer:** -6.033

-6.033

Explanation

As Sara discussed in lecture, we use a t-statistic to perform this test. The t-statistic is defined as:

$$\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{-0.1655357}{0.02744} = -6.033.$$

Submit You have used 1 of 2 attempts

Question 11

1.0/1.0 point (graded)

Would you reject this null hypothesis using a 99% level of confidence?

- a. Yes ✔
- b. No

Explanation

According to the R output, the p-value associated with this test is 1.88e-09. We can reject the null hypothesis at a 99% level of confidence.

Submit You have used 1 of 1 attempt