William Kuhns

5/10/19

1. A. wage in thousands = .015 + .00028 \* YrsEdu

B. wage = 15 + .28/52 \* WeeksofEdu

C. wage in hundreds = .15 + .0028/180 \* DaysofEdu

2) A. B\_0 hat means that a person has a 50% chance to pass the class if they never show up. B\_1 hat means for each time a person shows up there chance goes up by 1%.

B. Pass113 = .5 + .01(10) = .6 or 60 % chance of passing

C. Let’s say student A goes to 5 classes and student B goes to 10. Pass113 for A is .55 and Pass113 for B it is .65. The difference is .1 or 10%.

D. pass113 = .5 + .03 \* Attendance in weeks, this is assume there are 3 classes a week

pass113 = .5 + .07 \* Attendance in weeks, this is assume there are 7 classes a week

3) A. B\_0 hat means that a person with no experience should expect 3.31 percent change in wage. B\_1 hat means for every year of experience an additional .09 percent change in wage is added.

B. ln(wage) = 3.31 + 0.0075 \* experienceInMonths, Now for every additional month of experience an additional .0075 percent change in wage is added.

C. B\_0 hat means a person with no change in experience should expect a 10.2 wage. B\_1 hat means that a percent change experience will increase the wage by 2.

D. B\_0 hat means that wage will increase by 2.17 given no change in percent experience. B\_1 hat shows that a percent change in experience gives way to .5 increase in percent wage. Assuming this plus sign is a typo.

4) A. Spurious correlation is when an unknown variable makes it look like two variables are interdependent.

B. “A spurious correlation is a relationship between two variables that appear to have interdependence or association with each other but actually do not. Spurious correlation is often caused by a third factor that is not apparent at the time of examination.” -Investopedia

C. When Y and X are swapped. X becomes the dependent or effect. Y becomes the independent or cause.

D. “Reverse causality means that X and Y are associated, but not in the way you would expect. Instead of X causing a change in Y, it is really the other way around: **Y is causing changes in X**. In epidemiology, it’s when the exposure-disease process is reversed; In other words, the exposure causes the risk factor. “ -dataScienceCentral

E. R^2 is the improvement that linear regression has at guessing.

F. “measures how much better the regression line predicts the true values of y than just the mean, i.e. what proportion of the variation in Y can be explained by X, between 0 and 1” -from slides

G. R^2 is the improvement that linear regression has at guessing. Compared to the mean.

“The coefficients estimate the trends while R-squared represents the scatter around the regression line.

The interpretations of the significant variables are the same for both high and low R-squared models.

Low R-squared values are problematic when you need precise predictions.” -minitab

5) A. The linear regression proceeding through zero, zero needs to make logical sense given the context.

B. Blood alcohol level on car accidents, both are reasonable to be zero. Books read on education level, both are reasonable to be zero, or least very low.

C. Hypothesis test the linear regression over the origin over the regular linear regression.

6) A. “Regress wages [south] “-stata

95% conf. Intervaul = -1.436198 to -.1189207

B. t\_value=-1.40 95% conf. Intervaul = -1.043632 to .1756805, the T value does not fall within the CI. The data suggest that education is lower relative to the other groups sampled.

. regress wage south educ

Source | SS df MS Number of obs = 519

-------------+---------------------------------- F(2, 516) = 50.82

Model | 1155.84794 2 577.923971 Prob > F =0.0000

Residual | 5867.76595 516 11.3716394 R-squared =0.1646

-------------+---------------------------------- Adj R-squared = 0.1613

Total | 7023.61389 518 13.5591002 Root MSE = 3.3722

------------------------------------------------------------------------------

wage | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

south | -.4339759 .3103253 -1.40 0.163 -1.043632 .1756805

educ | .524723 .0537554 9.76 0.000 .4191166 .6303295

\_cons | -.5431464 .7116492 -0.76 0.446 -1.941232 .8549397

------------------------------------------------------------------------------

C.

. regress educ exper

Source | SS df MS Number of obs = 519

-------------+---------------------------------- F(1, 517) = 52.13

Model | 365.132886 1 365.132886 Prob > F =0.0000

Residual | 3621.46056 517 7.00475931 R-squared =0.0916

-------------+---------------------------------- Adj R-squared = 0.0898

Total | 3986.59345 518 7.69612635 Root MSE =2.6467

------------------------------------------------------------------------------

educ | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

exper | -.0619827 .008585 -7.22 0.000 -.0788485 -.0451168

\_cons | 13.60569 .1870568 72.74 0.000 13.2382 13.97317

------------------------------------------------------------------------------

. reg3 educ exper

Three-stage least-squares regression

--------------------------------------------------------------------------

Equation Obs Parms RMSE "R-sq" chi2 P

--------------------------------------------------------------------------

educ 519 1 2.641546 0.0916 52.33 0.0000

--------------------------------------------------------------------------

------------------------------------------------------------------------------

educ | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

educ |

exper | -.0619827 .0085685 -7.23 0.000 -.0787765 -.0451888

\_cons | 13.60569 .1866961 72.88 0.000 13.23977 13.97161

------------------------------------------------------------------------------

Endogenous variables: educ

Exogenous variables: exper

------------------------------------------------------------------------------

D. Binary variable of rural or city.