Case Study 3

WS 2020/2021

Deadline: November 29nd, 2020@ 23:55

1 Data Description

The dataset `WD1` contains information on monthly wages (`wage`) of $N = 663$ US workers from 1980, reported in USD. The following variables are available as explanatory variables:

- `hours`: average weekly hours worked
- `educ`: years of education
- `exper`: years of work experience
- `tenure`: years with current employer
- `age`: age in years
- `IQ`: IQ score
- `sibs`: number of siblings
- `brthord`: birth order (`brthord` is one for a first-born child, two for a second-born child, and so on)
- `meduc`: years of education of mother
- `feduc`: years of education of father

2 Descriptive statistics

Answer the following questions by inspecting the data through histograms and/or summary statistics:

2.1

What is the average and median wage? What does the distribution of wages look like?

2.2

What is the proportion of workers working more than 40 hours a week?

2.3

What is the number of years of education that is most common among the workers?

2.4

Is *single child* the most frequent sibling pattern?

3 Data modelling

Analyze the effect of the working conditions as well as the demographic variables on `wage` using a multiple regression model.

3.1

Now, regress the variables on the log(*wage*) (`lwage`) and report the output for the fitted model. How large is the coefficient of determination and how can it be interpreted qualitatively?

3.2

Interpret the effect of years of education.

3.3

Interpret the effect of `brthord`. Be particularly careful about the sign.

3.4

Who achieves the largest average effect on `wage` with three additional years of education, ceteris paribus: the workers' mothers, their fathers or the workers themselves?

3.5

Test the following hypotheses (report the hypotheses, significance level, test statistic, p-value and interpretation)

3.5.1

An additional year of education has no influence on wages, ceteris paribus.

3.5.2

The birth order has no influence on wages, ceteris paribus.

3.5.3

The variables which have no significant influence on wages individually also do not influence wages jointly. That is, they can be jointly excluded from the model, ceteris paribus.

3.5.4

The birth order, number of siblings and mother's education years can be jointly excluded from the model, ceteris paribus.

3.5.5

The effect of mother's and father's years of education on wages is the same, ceteris paribus.

3.5.6

Do the results motivate you to continue your studies at WU? Why or why not?

4 Simulation Study

The purpose of the following simulation study is to illustrate the asymptotic behaviour of a linear regression model.

4.1

Simulate data $(y_i, x_i)_{i=1,\dots,N}$ with $N_1 = 10, N_2 = 100, N_3 = 1000$ from the simple linear regression model

$$Y = \beta_0 + \beta_1 X + u$$

where the error term u is independent of X and u has a normal distribution with mean 0 and variance $\sigma^2 = 4$. Moreover, X has a uniform distribution on $[-3, 3]$ and the true parameters are $\beta_0 = -1$, $\beta_2 = 0.2$.

Provide your code and the model output of the fitted regression models.

Remark: If you work in R, please set your seed to 1, using the command `set.seed(1)`.

4.2

Report 95%-confidence intervals for β_0 and β_1 for the three sample sizes. What effect can you observe? Make a link to the consistency result of the OLS-estimator.

Case Study 3

WS 2020/2021

Deadline: November 29nd, 2020@ 23:55

4.3

Perform similar experiments to observe the effects of the following on confidence intervals: (a) the variance of the error terms σ^2 varies, (b) the level of confidence, $1 - \alpha$, varies.