市场数据分析 Fall 2018

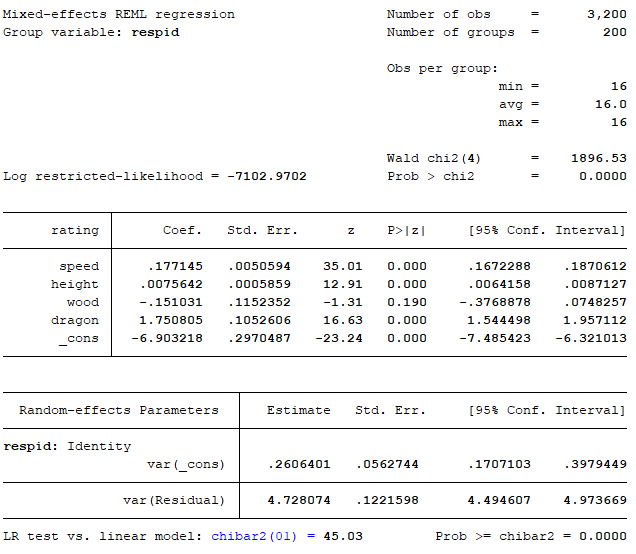
Ying Lei

Problem Set 2

Printed version due in 1 week (in class on 10/29)

(100 points)

1. [15 pts] Suppose that satisfy the least squares assumptions. A random sample of size is drawn and yields
   1. [10pts] Test vs. at the 5% significance level. Explain clearly how you reach your test conclusion.
   2. [5pts] Construct a 95% confidence interval for .
2. [25 pts] A park designer did a conjoint analysis survey with 200 respondents who each rate the same set of 16 roller coaster profiles. Each respondent is asked to rate from a scale of 1-10 for 16 coasters with various possible levels of maximum speed (40, 50, 60 or 70mph), height (200, 300, or 400ft), construction type (wood or steel), and theme (dragon or eagle). For example: On a 10-point scale, where 10 is the best and 1 is the worst, how would you rate a roller coaster that is made of wood, is 400ft high, has a maximum speed of 50mph, with a dragon theme? Regressing consumers’ ratings on coaster features using random effects multilevel linear regression model, we get the contribution by each feature to consumer satisfaction of roller coasters.



* 1. What conclusion can you draw about the coefficient of *speed* in the regression?
  2. Construct a 95% confidence interval for the effect of roller coaster speed.
  3. Is height a significant feature for roller coasters?
  4. Which construction type would pick if you are the park designer? Why?
  5. Which theme would you pick if you are the park designer? Why?

**Empirical Exercises**

**(You are required to print out your log file and submit with your homework. 10 Points off if log file is missing.)**

1. [30 pts] The data file **CPS12** contains data for full-time, full-year workers, age 25–34, with a high school diploma or B.A./B.S. as their highest degree. A detailed description is given in **CPS12\_Description**. In this exercise, you will investigate the relationship between a worker’s age and earnings for different sub-groups of people.
   1. Construct a two-way graph with two fitted lines of *AHE* on *Age*, one for the group of female workers and the other for male workers. (Note: Put two fitted lines in the same graph, give clear legends and axis titles.) What conclusion can you get from the graph?
   2. Run a regression of average hourly earnings (*AHE*) on gender (*Female*). What’s the coefficient of Female? How to interpret it?
   3. Get the mean of average hourly earnings for female workers, and the mean of average hourly earnings for male workers. What’s the difference between this two numbers? Compare this difference with your answer in (b).
   4. First run a regression of average hourly earnings (*AHE*) on age (*Age*) and gender (*Female*). Then run a regression of average hourly earnings (*AHE*) on age (*Age*), gender (*Female*) and education (*Bachelor*). Check the coefficients of *Age* and *Female*. Which one has obviously changed after including the education variable (*Bachelor*)? What might be the reason of this change? (Hint: Check the correlation between *Age* and *Bachelor* and the correlation between *Female* and *Bachelor*.)
   5. Interpret the coefficients of *Age*, *Female* and *Bachelor* in the regression of *AHE* on these three variables.
2. [30 pts] Keep using the data file **CPS12**. In this exercise, you will investigate the relationship between a worker’s age and earnings. (Generally, older workers have more job experience, leading to higher productivity and higher earnings.)
   1. Run a regression of average hourly earnings (*AHE*) on age (*Age*), gender (*Female*), and education (*Bachelor*). If *Age* increases from 25 to 26, how are earnings expected to change? If *Age* increases from 33 to 34, how are earnings expected to change?
   2. Run a regression of the logarithm of average hourly earnings, ln(*AHE*), on *Age*, *Female*, and *Bachelor*. If *Age* increases from 25 to 26, how are earnings expected to change? If *Age* increases from 33 to 34, how are earnings expected to change?

Ln(AHE) = a Age + b Female + c Bachelor

系数a 的意义：年龄变动1个单位，Ahe变动（1/a）%。

* 1. Run a regression of the logarithm of average hourly earnings, ln(*AHE*), on ln(*Age*), *Female*, and *Bachelor*. If *Age* increases from 25 to 26, how are earnings expected to change? If *Age* increases from 33 to 34, how are earnings expected to change?

Ln(AHE) = a ln(Age) + b Female + c Bachelor

系数a 的意义：年龄变化a%，收入增长1%

* 1. Run a regression of the logarithm of average hourly earnings, ln(*AHE*), on *Age*, *Age*2, *Female*, and *Bachelor*. If *Age* increases from 25 to 26, how are earnings expected to change? If *Age* increases from 33 to 34, how are earnings expected to change?

Ln(AHE) = a1 Age + a2Age^2 + b Female + c Bachelor

* 1. Plot the regression relation between *Age* and ln(*AHE*) from (b), (c), and (d) for males with a high school diploma. Describe the similarities and differences between the estimated regression functions. Which regression (b, c, d) do you prefer?