Tutorial 6

Hong Xiang Yue

05/04/2019

Part A

Part B

Hypothesis testing in a nutshell

- ▶ Given some data, a model and some assumptions, we want to make some sort of statement about what (probably) is happening in the real world
- ▶ Choose a null and alternative hypothesis and determine whether the data is more likely to come from the null or the alternative hypothesis

Intuitive examples

- ▶ Suppose you tossed a coin ten times in a row and got ten heads, you intuitively know that the coin probably isn't fair
- ▶ Ever played Monopoly and lost your temper? We all know that feeling when you keep landing on other people's properties but they never seem to land on yours. It just doesn't seem fair!!
- ▶ These are all examples of (adhoc) hypothesis testing

Figure 1: When the null hypothesis has been rejected

Formalising the examples

- ▶ For the coin toss example, we assume i.e.
  $H_0 : \Pr(heads) = Pr(tails) = 0.5$
- ▶ Under this assumption, the number of heads you get over ten tosses, X have a Binomial distribution
  $X \sim Binomial(n = 10, p = 0.5)$
- ▶ This means $\Pr(\text{ 10 heads or 10 tails }) = \frac{1}{2}^{10} \times 2 = 0.00195$
- ▶ Not impossible but a bit too unlikely
- ▶ More precisely if we set a threshold of what we consider unlikely, say 0.05, we can compare our probability, 0.00195 to it and make an objective conclusion
- ▶ Since $0.00195 < 0.05$ we conclude that a result of getting 10 head or tails in a row is too unlikely to come from a fair coin and therefore the coin is unfair

Hypothesis testing under the CLM

▶ E.1 and E.2 allow us to arrive at the OLS solution
▶ E.1-E.3 tells us that OLS is unbiased
▶ E.1-E.4 means that OLS is BLUE (Best Linear Unbiased Estimator)
▶ With the addition of E.5: $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2 I_n)$ i.e. the assumption of normality we also know that:

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$$

Hypothesis testing under the CLM

- If we want to test the hypothesis that $H_0 : \beta_j = b$
- We would calculate $\frac{\hat{\beta}_j - b}{se(\hat{\beta}_j)}$
- If it's unlikely to have come from t_{n-k-1} , then we would reject the null hypothesis in favour of the alternative
- When we test $H_0 : \beta_j = 0$, we're just interested in calculating the t-statistic $\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$

Question 1

Suppose we estimated an equation and got the results

$$\hat{y}_i = \underset{(3.1)}{5.4} + \underset{(1.5)}{3.2}\, x_i, \quad i = 1, 2, .., 22$$

$$R^2 = 0.26$$

1. Test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ at the 5% level
2. Construct a 95% confidence interval for β_1. How can we interpret this confidence interval?

https://flux.qa/LDPPHD

But what is a confidence interval?

A confidence interval is an estimated range of numbers in which we are reasonably sure the true parameter lies.

- ▶ This interval is random when we think of taking multiple samples and re-estimating the intervals. They will not be the same each time.
- ▶ Some may cover the true parameter, some may not.
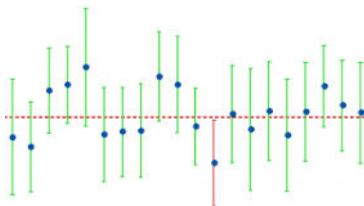- ▶ With a 95% confidence interval, it will cover the true parameter 95% of the time.

Figure 2:

3. Suppose you learn that x and y are independent. Would you be surprised? Why or why not?
4. Suppose that x and y are independent (i.e. $\beta_1 = 0$). If we took many samples of size $n = 22$ and re-estimated the equations and did the same hypothesis test as in part (a) for each sample, what proportion of tests would we expect to reject the true null? https://flux.qa/LDPPHD
5. What proportion of confidence intervals would we expect to contain $\beta_1 = 0$? https://flux.qa/LDPPHD

Question 2

To model the relationship between sleep, working hours, education and age we estimated this equation from a random sample of 706 adults

$$\hat{sleep} = 3638.25 - \underset{(0.017)}{0.148}\,totwrk - \underset{(5.88)}{11.13}educ + \underset{(1.45)}{2.2}\ age$$
$$\underset{(112.27)}{}$$

$$R^2 = 0.113, SSR = 123455057$$

Sleep and totwrk are measured in minutes and education and age are measured in years.

3. Test the hypothesis that adults do not trade-off sleep for work against the alternative that they do at the 1% level. What is the conclusion? https://flux.qa/LDPPHD

Question 2 continued

We also estimate this equation on the same sample:

$$\hat{sleep} = 3586.38 - 0.151\,totwrk$$
$$\quad\quad\quad (38.91) \quad\quad (0.017)$$

$$SSR = 124858119$$

4. Test the *joint* hypothesis that education and age have no effect on sleep time while holding work constant at the 5% level.
5. What is the R^2 of the second equation?

https://flux.qa/LDPPHD

6. Suppose you had the hypothesis that one year of education had the opposite effect of one year of age (all other things equal) i.e. $H_0 : \beta_2 = -\beta_3$. How would you test this hypothesis with an F-test?

7. If we wanted to keep the same null hypothesis but test the alternative hypothesis $H_1 : \beta_2 < -\beta_3$. Would we be able to test this using an F-test? How would we go about doing it?