

Problem Set 8

(due: Monday week 10, 11:00am)

Submission Instructions:

Online submission per Canvas file upload only. Only pdf-files accepted. Your solution must be in your own handwriting. You can write on paper and use a suitable app on your phone to scan and upload a pdf-file, or you can use a stylus-type pen on your tablet-type device. List everyone you worked with.

Exercises

Provide transparent derivations. Justify steps that are not obvious. Use self sufficient proofs. Make reasonable assumptions where necessary.

Consider the *normal regression model*

$$Y_i = \beta X_i + e_i,$$

where X_i and Y_i are scalars and $e_i|X_i \sim N(0, \sigma_e^2)$. This model makes the strong assumption that the conditional error distribution is exactly normal.

- (a) Derive the conditional distribution of the OLS estimator for β , that is come up with a statement like

$$\hat{\beta}^{\text{OLS}}|(X_1, \dots, X_N) \sim \dots,$$

where you determine the “...” part.

Notice that you are deriving the *exact* distribution here, not the approximate one. You do not need to use asymptotic theory because of the normality assumption imposed on the error distribution.

Read the primer on hypothesis testing on the next page to answer the following exercises.

- (b) Propose a sensible test of $H_0 : \beta = 0$ with an exact size of 5%. (Please assume that you know the value of σ_e^2 .)

Aside:

It is easy to propose a non-sensible test: draw a number randomly from $\{1, 2, 3, \dots, 20\}$ and reject the null if that number equals 8. If the null is true you will reject it with a probability of $1/20 = 5\%$. Unfortunately, the power of that test is also 5%, which is sad.

- (c) Derive the power function $p(\beta)$ for the normal regression model. Let $H_0 : \beta = 0$. Sketch $p(\beta)$ either by hand or computer. This does not need to be very precise.

Primer on Hypothesis Testing

Let me illustrate everything with the simple univariate model $Y_i \sim N(\beta, 1)$.

You do not know β , the value of the coefficient that generates Y_i . But you have available a random sample $Y_i, i = 1, \dots, N$ and you would like to know if $\beta \stackrel{?}{=} 0$.

You formulate that question by way of hypotheses:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

You conduct a statistical test:

Definition. Let $T := g(Y_1, \dots, Y_N)$ be some statistic. A **statistical test** is a decision rule about a null hypothesis that compares T to some critical value c and decides to

reject H_0 if $|T| > c$

not reject H_0 if $|T| \leq c$

A simple and silly example would be a decision rule that uses $T = \bar{Y}$ and decides $\beta \neq 0$ (reject H_0) whenever $|T| > 42$.

A more conventional example is the decision rule that uses $T = \bar{Y}/\text{se}(\bar{Y})$ and decides $\beta \neq 0$ (reject H_0) whenever $|T| > 1.96$. This is the simple t -test.

There are two types of errors that you can make with a statistical test:

- type I error: to reject H_0 when it is true
- type II error: not to reject H_0 when it is false

Ideally we want to minimize the probability of these two errors. Usually, however, the two goals of minimizing these errors stand in contradiction to each other, and the conventional approach is to fix the type I error at some low level (it is seen as the worse of the two errors) and then cross your fingers (and close your eyes) for the type II error.

The power function is a convenient way to graph the type II error as a function of the true value of β . Because the Y_i are determined by the value of β , your statistic T also is determined by the value of β . Denote the distribution of T under the particular value of β by $\mathbb{P}_\beta(T)$. If you know $\mathbb{P}_\beta(T)$ then you can ask yourself questions such as: *what is the probability of rejecting H_0 for particular values of β ?* You would answer this question by deriving the distribution $\mathbb{P}_\beta(|T| > c)$. This is the idea behind the *power function*:

$$p(\beta) := \mathbb{P}_\beta(|T| > c).$$

The power function answers this question: For a given value of β , what is the probability that I reject my null hypothesis? When $H_0 : \beta = 0$ then $p(0)$ will give you the probability of a type I error (size), in all other cases it will give you 1 minus the probability of a type II error.