A Note on Testing Hypotheses

Econ 620

Spring, 1999

1. Null hypothesis vs. alternative hypothesis

Suppose that we have data $y = (y_1, \cdots, y_n)$ and the data is generated by the true probability distribution P_{μ_0}, from a family of probability distribution P_μ indexed by $\mu \in \Theta$. We can partition the parameter space Θ into two subsets, Θ_0 and Θ_A. Now, consider the following two hypothesis;
$$H_0; \mu_0 \in \Theta_0 \qquad H_A; \mu_0 \in \Theta_A$$
H_0 is called the **null hypothesis** and H_A is called the **alternative hypothesis**. The union of null and alternative hypothesis defines a hypothesis $H \in \Theta = \Theta_0 \cup \Theta_A$ called the **maintained hypothesis**.

² A hypothesis is called simple if it completely specify the probability distribution and otherwise composite.

Example 1 *Suppose that we observe data* $y = (y_1, \cdots, y_n)$: *If we are willing to assume that the data set is a random sample from* $N(\mu, 10)$ *where* $\mu \in \{1, 2\}$. *We want to check which value of* μ *is consistent with data. Then, we can formulate the hypotheses;*
$$H_0; \mu = 1 \qquad H_A; \mu = 2$$
Here, both null and alternative hypotheses are simple since mean and variance are sufficient to completely specify a normal distribution. The maintained hypothesis in this case is that $H; \mu \in \{1, 2\}$: *If we assume that the data set is a random sample from* $N(\mu, 10)$ *where* $\mu = R$: *We can formulate the following hypotheses;*
$$H_0; \mu = 1 \qquad H_A; \mu \neq 1$$
The null hypothesis is simple but the alternative hypothesis is composite since the alternative hypothesis includes infinite numbers of normal distributions. The maintained hypothesis in this case is that $H; \mu = R$: *Alternatively, we can formulate different hypotheses such as*
$$H_0; \mu \geq 1 \qquad H_A; \mu < 1$$
In this example, both null and alternative hypotheses are composite and again the maintained hypothesis is that $H; \mu = R$:

2. The null hypothesis

Null hypotheses can arise for consideration in a number of different ways, the main ones

being as follows;

- H_0 may corresponds to the prediction of some scientific(economic) theory or to some model of the system thought quite likely to be true or nearly so.
- H_0 may represent some simple set of circumstances which, in the absence of evidence to the contrary, we wish to assume holds. For example, the null hypothesis might assert the ineffectiveness of newly-developed medicine for AIDS. We want to play safe by assuming ineffectiveness unless we can find a significant evidence against our presumption.
- H_0 may assert complete absence of structure in some sense. So long as the data are consistent with the null hypothesis it can not be justified to claim that the data provide clear evidence in favor of some particular kind of structure. Testing joint significance of slope coefficients in linear regression model is an example.

3. Type I and type II errors

There are four possible cases once we take an action in testing hypotheses - correctly accept the null, correctly reject the null, wrongly accept the null and wrongly reject the null. We don't have any concern with the first two cases. A good test should avoid or minimize the possibilities of the last two cases.

- Type I error is the probability that we reject the null hypothesis when it is true;

$$\alpha = P\left[\text{reject } H_0 \mid H_0 \text{ is true}\right]$$

- Type II error is the probability that we do not reject the null hypothesis when it is not true;

$$\beta = P\left[\text{do not reject } H_0 \mid H_A \text{ is true}\right]$$

- We call α, the probability that we reject the null hypothesis when it is true, **size** of the test
- We call $(1 - \beta)$, the probability that we reject the null hypothesis when the alternative hypothesis is true, the **power** of the test.

The ultimate goal in designing a test statistic is to minimize the size and maximize the power as much as possible. Unfortunately, it is quite easy to prove that we can not design a test which has **both** the minimum size **and** the maximum power. Here is an intuitive example why it is the case. Consider minimum size first. What is the test which has the minimum size? It is a test with which we always accept the null hypothesis no matter what we observe from the data. Since we always accept the null hypothesis, the size of the test is 0 which is the minimum possible size ; $P\left[\text{reject } H_0 \mid H_0 \text{ is true}\right] = 0$, note that we never reject. However, what is the power of this test? It is a pathetic test as far as power is concerned - power of the test is 0; $1 - P\left[\text{do not reject } H_0 \mid H_A \text{ is true}\right] = 1 - 1 = 0$. Now consider the opposite case - the test with which we always reject the null hypothesis. Power is great - it is 1, maximum possible power. But, the size of the test also is 1-again pathetic.

We have to compromise between the two conflicting goals. Convention in testing procedure is to fix the size at a arbitrary prespecified level and search for a test which maximizes the power - even this is impossible in most cases.

² **Critical region** of a test is the area where we commit the type I error.

4. Test procedure and test statistic

The next question naturally arising is that how we can actually test the hypotheses. An obvious answer is that the test, whatever it is, should be based on the observed data. The data set itself is too lousy to determine the plausibility of the null hypothesis. We need a kind of summary measure of the data, which should be handy but retain relevant information on the true data generating process.

² Let $t = t(y)$ be a function of the observations and let $T = t(Y)$ be the corresponding random variable. We call T a test statistic for the testing of H_0 if

(i) the distribution of T when H_0 is true is knwon, at least approximately(asymptotically).

(ii) the larger the value of t the stronger the evidence of departure from H_0 of the type
it is required to test

After getting the distribution of test statistic under the null hypothesis, we now need a decision rule to determine whether the null hypothesis is consistent with the observed data. For given observations y we can calculate $t = t_{obs} = t(y)$; say, and identify a **critical region** for a given size of the test. If the value of the test statistic falls into the critical region, we reject the null hypothesis.

To sum up the test procedure;

1. Set up the null and alternative hypotheses
2. Design a test statistic which has some good properties
3. Find the distribution of test statistic under the null hypothesis - exact or asymptotic distribution
4. Identify the critical region for the null distribution with a given size of the test
5. Calculate the test statistic using the observed data
6. Check whether or not the value of the test statistic falls into the critical region.