LECTURE 8: HYPOTHESIS TESTING

2nd of December 2015

Hypothesis testing

2ND OF DECEMBER 2015 1 / 23

Consider the following hypothetical situation:

From previous experience we know that the birth weights of babies in England are Normally distributed with a mean of 3000g and a standard deviation of 500g.

We think that maybe babies in Australia have a mean birth weight greater than 3000g and we would like to test this hypothesis.

NULL AND ALTERNATIVE HYPOTHESIS

The main hypothesis that we are most interested in is the **research hypothesis**, denoted H_1 , that the mean birth weight of Australian babies is greater than 3000g.

The other hypothesis is the **null hypothesis**, denoted H_0 , that the mean birth weight is equall to 3000g.

In summary:

 $H_0: \mu = 3000g$ $H_1: \mu > 3000g$

The research hypothesis is often called the alternative hypothesis.

NULL AND ALTERNATIVE HYPOTHESIS

We start with the research hypothesis and "set up" the null hypothesis to be directly counter to what we hope to show.

We then try to show that, in light of our collected data, the null hypothesis is false.

We do this by calculating **the probability of the data if the null hypothesis is true**.

- If this probability is very small it suggests that the null hypothesis is false.
- If this probability is large it suggests that there is not enough evidence to reject the null hypothesis.

Once we have set up our null and alternative hypothesis we can collect a sample of data.

Imagine that we take a sample of 44 babies from Australia, measure their birth weights and we observe that the sample mean of these 44 weights is $\bar{X}=3275.955g.$

We now want to calculate the probability of obtaining a sample with mean as large as 3275.955 by chance under the assumption of the null hypothesis H_0 .

We know from previous lectures that,

If $X_1, X_2, \ldots X_n$ are n independent and identically distributed random variables from a $N(\mu, \sigma^2)$ then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Thus, under the assumption of the null hypothesis the sample mean of 44 values from a $N(3000,500^2)$ is

$$\bar{X} \sim N\left(3000, \frac{500^2}{44}\right) = N\left(3000, 5681.818\right) \ .$$

Now we can calculate the probability of obtaining a sample with a mean as large as 3275.955 using standardization.

 $\overline{X} \sim N(3000, 5681.818)$

 $Z \sim N(0, 1)$



$$P\left(\bar{X} > 3275.955\right) = P\left(\frac{\bar{X} - 3000}{75.378} > \frac{3275.955 - 3000}{75.378}\right)$$
$$= P(Z > 3.66) = 0.00015$$

This probability is called the **p-value** of the test.

In this case the p-value is very low: the probability of the data is very low if we assume the null hypothesis is true.

But how low does this probability has to be before we can conclude that the null hypothesis is false?

Convention: choose a **level of significance** before the experiment that dictates how low the p-value should be before we reject the null hypothesis.

It is common to choose a significance level of 0.01 or 0.05.

We conclude that there is significant evidence against the null hypothesis if the p-value is less than or equal to 0.01 (or 0.05).

In our baby weight example, the p-value is 0.00015 which is lower than $0.01. \label{eq:constraint}$

In this case, we conclude that

"there is significant evidence against the null hypothesis at the 0.01 level."

Another way of saying this is that

"we reject the null hypothesis at the 0.01 level."

If the p-value for the test were much larger, say $0.23,\,{\rm then}$ we could conclude that

"the evidence against the null hypothesis is not significant at the 0.01 level."

Another way of saying this is that

"we cannot reject the null hypothesis at the 0.01 level."

Test statistics and critical regions

We can also compare the test statistic under the null hypothesis

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = 3.66 \; .$$

to a critical region of values defined such that

- if the test statistic lies in this region then we will reject H_0 ,
- if the test statistic lies outside this region then we will not reject H_0 .



OVERVIEW OF HYPOTHESIS TESTING

Version1:

- Begin with a research (alternative) hypothesis and decide upon a level of significance for the test.
- Set up the null hypothesis.
- Ollect a sample of data.
- O Calculate a test statistic from the sample of data.
- Ompare the test statistic to its sampling distribution under the null hypothesis by calculating the p-value.
- Reject the null hypothesis if the p-value is less than the level of significance. Otherwise, retain the null hypothesis.

OVERVIEW OF HYPOTHESIS TESTING

Version2:

- Begin with a research (alternative) hypothesis and decide upon a level of significance for the test.
- Set up the null hypothesis.
- Collect a sample of data.
- Calculate a test statistic from the sample of data.
- Ompare the test statistic to its sampling distribution under the null hypothesis by calculating the critical region for the test.
- Reject the null hypothesis if the test statistic lies in the critical region. Otherwise, retain the null hypothesis.

ONE SAMPLE TEST FOR A PROPORTION

Suppose that a university claims to admit equal number of state and public school students.

We have a research hypothesis that the university tends to admit more public school students so we interview 500 first year students and discover that 267 came from public school.

We want to test our hypothesis at the 0.05 level.

We first write down our null and alternative hypotheses regarding the population proportion p of public school students:

 $H_0: p = 0.5$ $H_1: p > 0.5$

Using our sample of data we obtain an estimate of p as $\hat{p} = \frac{267}{500} = 0.534$.

We know that, among n = 500 interviewed students,

number of students from public school $\ \sim B(n,p) \approx N\left(np,np(1-p)
ight)$.

Here we are interested at the proportion (and not the number) of students that came from public school. Approximatively

proportion of students from public school $\sim N\left(p, \frac{p(1-p)}{n}\right)$.

In this situation, the test statistic used is

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

where p is the proportion dictated by the null hypothesis H_0 and n is the size of our sample.

In our example, the value of the test statistic is

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.534 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{500}}} = 1.52$$



The test statistic does not lie in the critical region so we conclude that the evidence against the null hypothesis **is not** significant at the 0.05 level.

Similarly, we can compute the probability of obtaining an estimated proportion \hat{p} as large as 0.534 under the hypothesis H_0 :

p-value =
$$P(\hat{p} > 0.534) \approx P(Z > 1.52) = 0.064$$
 where $Z \sim N(0, 1)$

which is larger than the critical level 0.05.

In the first example, we wanted to test the research hypothesis that mean birth weight of Australian babies was higher than 3000g. We had

 $\begin{array}{l} {\sf H}_0: \ \mu = {\sf 3000g} \\ {\sf H}_1: \ \mu > {\sf 3000g} \end{array}$

Instead, our research hypothesis could be that the mean birth weight of Australian babies is different from 300g. In this case we have

 $H_0: \mu = 3000g$ $H_1: \mu \neq 3000g$

This is an example of a **two-sided test** as opposed to the previous examples which were **one-sided tests**.

As before we would calculate our test statistic as $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = 3.66$.

The p-value is different, though.

We are not looking at the probability that Z is only higher than 3.66 (in the positive direction), but that Z is at least this further away from 3.66 in either direction:

$$\begin{split} P(|Z| > 3.66) &= P(Z > 3.66) + P(Z < -3.66) \\ &= 2P(Z > 3.66) \; . \end{split}$$

In general:

the two-sided p-value is always twice as big as the one-sided p-value.

Hypothesis testing and confidence interval

In the previous lecture, we have constructed confidence intervals for the mean μ of a normal distribution $N(\mu,\sigma^2)$ based on n observations as

$$\left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}}\right)$$

where z is such that

$$P\left(\bar{X} - z\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z\frac{\sigma}{\sqrt{n}}\right) = \alpha$$

for a confidence level α .

You may notice that we do a lot of the same things to carry out a statistical test than we do to compute a confidence interval:

- we compute a standard error
- and look up a value on a standard normal table.

For instance, in the baby weight example we might have expressed our uncertainty about the average Australian birthweight in the form of a confidence interval. The calculation would have been just slightly different:

• we start with the mean observed birthweight in the Australian sample:

$$\bar{X} = 3275.955g$$

• the standard error is $\sigma/\sqrt{44}$, where σ is the (unknown) SD of Australian birth weights. Since we don't know σ , we substitute the SD of the sample, which is 528g. So we use

$$SE = \frac{528}{\sqrt{44}} = 80g.$$

Then a 95% confidence interval for the mean birthweight of Australian babies is

 $(3275.955 - 1.96 \times 80, 3275.955 + 1.96 \times 80) = (3120, 3432)g$

This is consistent with the observation that the Australian birth weights is significantly greater than the the UK average 3000g, at the 0.05 significance level.

SUMMARY

- Framework of hypothesis testing
- Two ways to operate: computing a p-value or through a confidence region
- Examples for normal distribution or proportions
- One-sided and two sided tests
- Link between hypothesis testing and confidence interval

Reminder: the lecture notes contain more details and more examples; they are available on my website.

Happy Break

Hypothesis testing

2ND OF DECEMBER 2015 23 / 23