

## Notes on Hypothesis Testing, Type I and Type II Errors

### Part 1. Hypothesis Testing

Suppose that a medical firm develops a new medicine that it claims will lead to a higher mean cure rate. Suppose the old cure rate was  $\mu_0$ . The firm claims that the new mean rate is  $\mu_1 > \mu_0$ . How can the firm verify their claim? By using Hypothesis Testing.

**Null Hypothesis:** The position that must get the benefit of doubt. This is usually the conventional wisdom. In this case,  
 $H_0: \mu \leq \mu_0$

**Alternative Hypothesis:** The claim we seek to prove, often called the Research Hypothesis.  
 $H_a: \mu > \mu_0$

**Aim of the Researcher:** Find evidence in favor of  $H_a$ . So he will reject  $H_0$  if the calculated sample mean is high relative to  $\mu_0$ .

**Important Step:** Assume:  $\mu = \mu_0$ . This is done to give the maximum chance for  $H_0$  to be true if we get a high sample mean. Remember that the question of rejecting the null only arises if the sample mean is high.

**Next:** Assuming  $\mu = \mu_0$ , calculate the probability of observing a sample mean of  $\bar{x}$  or higher. This is because even after giving the null so much benefit of doubt, if  $\Pr(\bar{X}_n \geq \bar{x})$  is low, chances are that the null hypothesis doesn't hold.

Essentially, we are asking what is the probability that we observe the sample mean that we actually see, conditional on the null being true. If this probability is very low, then that is evidence that the null hypothesis does not hold.

Basically, the probability we want to compute is:

$$\Pr[\bar{X}_n \geq \bar{x} | H_0 \text{ true}] = \Pr[\bar{X}_n \geq \bar{x} | \mu = \mu_0] = \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right] = p$$

This is the **p-value**. If  $p < \alpha$ , then reject  $H_0$ .

So what is  $\alpha$ ?  $\alpha$  is a predetermined probability level that specifies the proportion of times we are willing to reject  $H_0$  when  $H_0$  is true. If we reject  $H_0$  when  $H_0$  is true, we are making an error, so

$\alpha$  = proportion of times we are willing to make this error

How high  $\alpha$  is depends on how costly (in terms of reputation, monetary cost, etc.) this error is likely to be. If the cost is very high, we should set  $\alpha$  very low. Typically  $\alpha = 0.01, 0.025, 0.05$  or  $0.1$ .

Suppose  $p > \alpha$ . Should we reject  $H_0$ ? NO!! Suppose  $p = 0.06$  and  $\alpha = 0.05$ . Then  $p = 0.06$  implies that  $\bar{X}_n \geq \bar{x}$  6% of the time that  $H_0$  is true. So if we reject  $H_0$  because  $\bar{X}_n \geq \bar{x}$ , we will be committing an error 6% of the time. But  $\alpha = 0.05$  implies that we are only willing to make these mistakes 5% of the time, so we should not reject  $H_0$ .

### Alternative Method to Conduct Hypothesis Tests

We can also use critical values to conduct hypothesis tests. Here, we calculate a critical value,  $c$ , such that

$$\Pr[\bar{X}_n \geq c \mid \mu = \mu_0] = \alpha \Leftrightarrow \Pr\left[\frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right] = \alpha \Leftrightarrow \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} = z_\alpha$$

$$\Leftrightarrow c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Then, if  $\bar{X}_n \geq c$ , we can reject  $H_0$ . The interpretation of  $\bar{X}_n \geq c$  is that even though  $H_0$  is true, by rejecting  $H_0$  we are committing an error less than  $\alpha$  proportion of times. This is within tolerable levels.

**EXAMPLE 1.** A store wants to install a new billing system that will be cost effective only if your monthly account exceeds \$170. In a sample of 400 accounts, the sample mean monthly account was \$178. Suppose  $\sigma = \$65$  and  $\alpha = 0.05$ . Should the new system be installed?

$$H_0: \mu \leq 170$$

$$H_a: \mu > 170$$

First, work out the p-value:

$$\begin{aligned} \Pr[\bar{X}_n > 178 \mid H_0 \text{ true}] &= \Pr[\bar{X}_n > 178 \mid \mu = 170] = \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{178 - 170}{\frac{65}{\sqrt{400}}}\right] \\ &= \Pr[Z > 2.46] = 0.0069 < 0.05 \end{aligned}$$

Since the p-value is lower than  $\alpha$ , we can reject the null hypothesis and install the new system.

The interpretation is as follows. Given  $\mu \leq 170$ , the probability of observing a sample mean of 178 is 0.69%. So it is reasonable to conclude that  $\mu > 170$ . This conclusion may be wrong, but it will only be wrong less than 1% of the time. Hence there is little risk in rejecting  $H_0$ .

Alternatively, we could use the critical value approach to compute the rejection region. So we need to find  $c$  such that:

$$\Pr[\bar{X}_n \geq c \mid \mu = 170] = \alpha = 0.05 \Leftrightarrow \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c - 170}{\frac{65}{\sqrt{400}}}\right] = 0.05$$

$$\Leftrightarrow \frac{c - 170}{\frac{65}{\sqrt{400}}} = z_{0.05} = 1.645$$

$$\Leftrightarrow c = 170 + 1.645 \frac{65}{\sqrt{400}} = 175.35$$

Then, since  $\bar{X}_n = 178 > 175.35$ , we can reject  $H_0$ .

What we just did was a **Right-Tailed Test** ( $H_0: \mu \leq \mu_0$ ). A **Left-Tailed Test** is very similar. Then,

$$H_0: \mu \geq \mu_0$$

$$H_a: \mu < \mu_0$$

And so we reject  $H_0$  if the sample mean is too LOW.

The construction of a p-value is very similar to before:

$$\Pr[\bar{X}_n < \bar{x} | H_0 \text{ true}] = \Pr[\bar{X}_n < \bar{x} | \mu = \mu_0] = \Pr\left[\frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right] = p$$

And we reject  $H_0$  if  $p < \alpha$ . The only difference in a Left-Tailed Test is that we want  $\Pr[\bar{X}_n < \bar{x} | H_0 \text{ true}]$  instead of  $\Pr[\bar{X}_n \geq \bar{x} | H_0 \text{ true}]$  in the Right-Tailed Test case (because low values of the sample mean provide evidence against the null).

We can also use the critical value approach, where the critical value is  $c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$  (instead of

$c = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$  in the Right-Tailed Test).

**EXAMPLE 2.** Say that  $H_0: \mu \geq 22$  and  $H_a: \mu < 22$ . Suppose that  $\sigma = 6$ ,  $\alpha = 0.1$ ,  $n = 220$  and  $\bar{X}_n = 21.63$ . Then the p-value is:

$$\Pr[\bar{X}_n < 21.63 | \mu = 22] = \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{21.63 - 22}{\frac{6}{\sqrt{220}}}\right] = \Pr[Z < -0.92] = 0.1788$$

Since  $p > \alpha$ , we cannot reject the null hypothesis that  $\mu \geq 22$ .

We can also use the critical value approach to construct the rejection region.

$$c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = 22 - 1.645 \frac{6}{\sqrt{220}} = 21.3346$$

Then, since  $\bar{X}_n = 21.63 > 21.3346$ , we cannot reject the null.

These were both One-Tailed Tests. Finally, we consider a **Two-Tailed Test**.

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

In the case of a Two-Tailed Test, we reject the null hypothesis if the sample mean turns out to be too big OR too small.

The calculation of the p-value is slightly different to a One-Tailed Test. Now, we want to calculate

$$\Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < -\left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)\right] + \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right] = 2 \Pr\left[Z > \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}\right] = p$$

Then, reject  $H_0$  if  $p < \alpha$ .

We can also use critical values to derive the rejection region, as before. Now however, there must be two critical values:

$$c_1 = \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$c_2 = \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

So we reject the null if:

$$\bar{X}_n \notin \left[ \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] \Leftrightarrow \left| \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \geq z_{\frac{\alpha}{2}}$$

**EXAMPLE 3.**  $H_0: \mu = 17.09, H_a: \mu \neq 17.09$ . Suppose that  $\sigma = 3.87, \alpha = 0.05, n = 100$  and  $\bar{X}_n = 17.55$ .

So let's first compute the p-value:

$$\text{So, } p = 2 \Pr \left[ Z > \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \right] = 2 \Pr \left[ Z > \frac{17.55 - 17.09}{\frac{3.87}{\sqrt{10}}} \right] = 2 \Pr[Z > 1.19] = 2(0.117) = 0.234$$

Since  $p = 0.234 > 0.05 = \alpha$ , we cannot reject the null.

Alternatively, we can find the rejection region using critical values:

$$c_1 = \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 17.09 + 1.96 \frac{3.87}{10} = 17.09 + 0.75852 = 17.84852$$

$$c_2 = \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 17.09 - 1.96 \frac{3.87}{10} = 17.09 - 0.75852 = 16.33148$$

Then since  $\bar{X}_n = 17.55 \in [16.33148, 17.84852]$ , we cannot reject the null.

### Part 2. Type I and Type II Errors

	<b>H<sub>0</sub> True</b>	<b>H<sub>0</sub> False</b>
<b>Accept H<sub>0</sub></b>	OK	Type II Error
<b>Reject H<sub>0</sub></b>	Type I Error	OK

Type I errors are very serious – you do not want to make wild claims and be proven wrong later.  $\alpha$  in our previous discussion is the probability of making a Type I error. Hypothesis tests are designed to make  $\alpha$  low, since we choose a low value for  $\alpha$ ! Type II errors occur then the null hypothesis is false but we fail to reject it. These errors are not as serious, but we would like to avoid them.

There is a trade-off between Type I and Type II errors. If we do not change the sample size, the only way to reduce Type II errors is by increasing the probability of making Type I errors. However, both can be reduced by increasing  $n$ .

Type II errors are always computed for a given  $\alpha$ . Think of the probability of NOT making a Type II error. To compute this, we need to specify a value for  $\mu, \mu_a$ , which we (the researcher) believe is true instead of  $\mu_0$ . Then the probability of not committing a Type II error is:

$$\begin{aligned} p(\mu_a) &= \Pr[\text{Reject } H_0 \text{ at } \alpha \text{ significance} | H_0 \text{ False}] \\ &= \Pr[\bar{X}_n \text{ falls in the rejection region} | \mu = \mu_a] \end{aligned}$$

So we have that:

$$\begin{aligned} \Pr[\text{Type II Error}] &= 1 - p(\mu_a) \\ p(\mu_a) &= \text{Power of the test given } \mu = \mu_a. \end{aligned}$$

EXAMPLE 4. Suppose the old manufacturing process produces 80 units per hour. We want to evaluate the claim that a new manufacturing process produces 85 units per hour. Let  $\alpha = 0.05$ ,  $\sigma = 17.19$  and  $n = 50$ . Hence,

$$H_0: \mu \leq 80$$

$$H_a: \mu > 80$$

So first, we compute the rejection region.

$$\Pr[\bar{X}_n \geq c \mid \mu = 80] = \alpha = 0.05 \Leftrightarrow \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c - 80}{\frac{17.19}{\sqrt{50}}}\right] = 0.05$$

$$\Leftrightarrow \frac{c - 80}{\frac{17.19}{\sqrt{50}}} = z_{0.05} = 1.645$$

$$\Leftrightarrow c = 80 + 1.645 \frac{17.19}{\sqrt{50}} \approx 84$$

So we reject the null if  $\bar{X}_n \geq 84$ .

Next, find the probability of a Type II error given  $\mu_a = 85$ .

$p(85)$  = Probability not committing a Type II error

$$= \Pr[\bar{X}_{50} \geq 84 \mid \mu = 85] = \Pr\left[\frac{\bar{X}_{50} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{84 - 85}{\frac{17.19}{\sqrt{50}}}\right] = 0.6595$$

So the power of this test is 0.6595, so the probability of a Type II error is  $1 - 0.6595 = 0.3405$ .

Hence, even if the claim is correct, the statistical test will fail to show it 34% of the time. We can reduce the probability of a Type II error by increasing  $n$ . Now, suppose  $n = 100$ .

The rejection region is:

$$\Pr[\bar{X}_n \geq c \mid \mu = 80] = \alpha = 0.05 \Leftrightarrow \Pr\left[\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c - 80}{\frac{17.19}{\sqrt{100}}}\right] = 0.05$$

$$\Leftrightarrow \frac{c - 80}{\frac{17.19}{\sqrt{100}}} = z_{0.05} = 1.645 \Leftrightarrow c = 80 + 1.645 \frac{17.19}{\sqrt{100}} \approx 82.83$$

$$\text{So the power is: } p(85) = \Pr[\bar{X}_{100} \geq 82.83 \mid \mu = 85] = \Pr\left[\frac{\bar{X}_{100} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{82.83 - 85}{\frac{17.19}{\sqrt{100}}}\right] = 0.8962$$

So the probability of a Type II error is  $1 - p(85) = 1 - 0.8962 = 0.1038$ . Hence, doubling the sample size reduced the probability of a Type II error from 34% to 10%.