

# Hypothesis Tests

## Reading

- Section 4.3

## Practice Problems

**4.6.3 (Page 209)** 4.17, 4.18, 4.19, 4.20, 4.21, 4.22, 4.24, 4.25, 4.26, 4.27, 4.28, 4.31, 4.32

## Notes

### Hypothesis Tests in general

Hypothesis Testing has a ubiquitous presence in applied statistics and related disciplines. Almost every study out there is accompanied by appropriate hypothesis tests. We start by describing some standard terminology associated with hypothesis tests.

### Null and Alternative Hypotheses

In every hypothesis test there are two statements, the null hypothesis and the alternative hypothesis.

- The **null hypothesis** is an assertion we are making that we are trying to disprove. So it is a statement about the population, that is claimed to be true. We are trying to determine if there is sufficient evidence against it. It is denoted by  $H_0$ .
- The **alternative hypothesis** is the contradictory assertion. It is what we will accept as truth if we reject the null hypothesis. It is denoted by  $H_a$ .

It is vital to keep in mind, that the conclusion/decision from the test must be one of the following:

- You have enough evidence to reject  $H_0$  in favor of  $H_a$ , or
- You do not have enough evidence to reject  $H_0$ . This does not mean that it is true, only that we don't have enough evidence to conclude that it is not.

You can *never* prove the null hypothesis. You can only reject it in favor of the alternative.

For example, say we want to test a new diet pill, to see if it is effective.

- We could set up an study where we have people follow the recommended diet for say a month, and then compare the difference in their weight.
- We could then say that the null hypothesis  $H_0$  is that on average people don't lose weight, so the population mean of the differences between their new weight to their previous weight is  $\mu \geq 0$ .
- The alternative hypothesis would then be that  $\mu < 0$ , i.e. that on average people lose weight.

## Types of Errors

In every hypothesis test there are two kinds of errors that we might make.

- **Type I errors** occur when we reject the null hypothesis when it actually was true.
- **Type II errors** occur when we end up not rejecting the null hypothesis, even though it actually was false.

The probabilities of making those errors are related to how we set up our test and what parameters and bounds we use. We will see some of this in a bit.

We have seen these types of errors before, when we were looking at medical tests.

As a general principle, we want to keep both of those errors to a minimum. Some times however, one of the two types is more critical and we might be willing to allow for a higher chance of error in the other type in order to make the chance of error in the first type smaller.

## Hypothesis Test for population mean

We will now focus specifically on tests for the population mean  $\mu$ . We assume that we are in the IID setting, and that  $n$  is sufficiently large for us to use the Central Limit Theorem. Therefore we are working under the assumption that:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

There are three types of hypotheses tests we may encounter:

- The null hypothesis is that the population mean  $\mu$  equals a certain value  $\mu_0$  (some specific number). The alternative hypothesis would then simply say that  $\mu \neq \mu_0$ . We usually summarize this as follows:

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_a : & \mu \neq \mu_0 \end{cases}$$

- The null hypothesis is that the population mean  $\mu$  is at least a certain value  $\mu_0$  (some specific number). The alternative hypothesis would then simply say that  $\mu < \mu_0$ . So:

$$\begin{cases} H_0 : & \mu \geq \mu_0 \\ H_a : & \mu < \mu_0 \end{cases}$$

- The null hypothesis is that the population mean  $\mu$  is at most a certain value  $\mu_0$  (some specific number). The alternative hypothesis would then simply say that  $\mu > \mu_0$ . We usually summarize this as follows:

$$\begin{cases} H_0 : & \mu \leq \mu_0 \\ H_a : & \mu > \mu_0 \end{cases}$$

Depending on the wording of the question, you have to choose the appropriate test. The first one, with the so-called “two-sided alternative”, is the default choice.

The logic behind the hypothesis tests goes like this:

- Suppose the null hypothesis were true. In particular,  $\mu = \mu_0$  is a pivotal value for it.
- With that assumed value for  $\mu$ , our sample mean  $\bar{x}$  comes from a fully known distribution  $N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$ .
- We now ask ourselves: How *extreme* is the value  $\bar{x}$  we obtained from our sample, compared to what that distribution suggests? in other words, *what are the chances that  $\bar{x}$  would be as far away from  $\mu_0$  as it appears to be from our sample?*
- This translates to wanting to compute an area under the curve in a normal distribution, and we usually convert the  $\bar{x}$  into a  $z$  first, via:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In this context this value  $z$  is called the **test statistic**. It has a known distribution under the assumption that the null hypothesis is true.

- The area we compute is interpreted as *the chances, if the null hypothesis were correct, that we would see the kind of extreme value we saw*. We call this the **P-value**.

- A small  $P$ -value means that if the null hypothesis were correct, then the kind of sample we just witnessed would have been very unlikely.
- But since we did just witness that sample, we have to either accept that we just witnessed something rare or to accept that the null hypothesis is incorrect, and to reject it in favor of the alternative.
- This is the essence of the hypothesis test: We end up rejecting the null hypothesis on the basis that the kind of data we saw would have been very rare if the null hypothesis were true.

How we compute the  $P$ -value depends on which of the three types of hypothesis tests we are considering. We will phrase our answers in terms of  $z$ .

- In the alternative:

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_a : & \mu \neq \mu_0 \end{cases}$$

We compute the area that is above  $|z|$  or below  $-|z|$ .

- In the alternative:

$$\begin{cases} H_0 : & \mu \geq \mu_0 \\ H_a : & \mu < \mu_0 \end{cases}$$

We compute the area that is below  $z$ . If  $z > 0$  this would be bad (large), and that is a good thing as a  $\bar{x}$  that is greater than  $\mu_0$  should not help us at all in proving the alternative  $\mu < \mu_0$ .

- In the alternative:

$$\begin{cases} H_0 : & \mu \leq \mu_0 \\ H_a : & \mu > \mu_0 \end{cases}$$

We compute the area that is above  $z$ . If  $z < 0$  this would be bad (large), and that is a good thing as a  $\bar{x}$  that is less than  $\mu_0$  should not help us at all in proving the alternative  $\mu > \mu_0$ .

## An Example

The college claims that the average GPA of our students is at least 3.1. Suppose we know that the standard deviation of all Hanover GPAs is 0.55. We will use a sample of 50 students. We find the mean in our sample to be only 2.9. We would like to see if this is enough evidence to reject the college's claim. We therefore set a hypothesis test, one-sided since the college claimed "at least":

$$\begin{cases} H_0 : \mu \geq 3.1 \\ H_a : \mu < 3.1 \end{cases}$$

To carry out the test, we would compute

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.9 - 3.1}{0.5/\sqrt{50}} = -2.83$$

We then compute the area below that value and we find it to be 0.0023. Therefore there is a less than 0.25% chance that we would see such a low mean if the college's claim was correct. As this is extremely unlikely, we can consider it very strong evidence against the college's claim (null hypothesis). We would therefore reject the null hypothesis, and instead suggest that the true mean is smaller. In fact we may provide a confidence interval.

## Significance Level

When setting up a hypothesis test, it is important to agree beforehand how small a percentage we would consider strong enough evidence. This is called the **significance level** and is denoted by  $\alpha$ . For example we might decide that we want  $\alpha = 0.05$ , meaning that we will reject the null hypothesis as long as the  $P$ -value we compute is less than that.

A significance level of 5% is fairly typical.