Frequentist Inference, pt. II

ECON 3640–001

Marcio Santetti
Spring 2022

Motivation

Housekeeping

Notes based on `Keller (2009)`:

- Chapter **11**, sections `11.1` and `11.2`.

Motivation

Last time, we were introduced to the concept of **confidence intervals**.

Given how fragile a **point** estimator is, producing inferences about a population parameter through an **interval** allows for a more flexible interpretation of a statistic of interest.

Now, we move on to a **second** inferential approach:

- **Hypothesis testing**

This procedure serves for determining whether there is *enough statistical evidence* to confirm a belief/hypothesis about a parameter of interest.

A *nonstatistical* application of Hypothesis Testing

A *nonstatistical* application of Hypothesis Testing

When a person is accused of a crime, they face a **trial**.

The prosecution presents the case, and a jury must make a decision, **based on the evidence** presented.

In fact, what the jury conducts is a test of **different hypotheses**:

- *Prior hypothesis*: the defendant is **not guilty**.
- *Alternative hypothesis*: the defendant is **guilty**.

A *nonstatistical* application of Hypothesis Testing

The jury **does not** know which hypothesis is correct.

Their base will be the **evidence** presented by both prosecution and defense.

In the end, there are only two possible decisions:

- **Convict** or
- **Acquit**.

Back to Statistics

Statistical hypothesis testing

The *same reasoning* follows for Statistics:

- The **prior hypothesis** is called the **null hypothesis** (H_0);
- The **alternative hypothesis** is called the research or **alternative hypothesis** (H_1 or H_a).

Putting the *trial* example in statistical notation:

- H_0: the defendant is **not guilty**.
- H_a: the defendant is **guilty**.

The hypothesis of the defendant being guilty (H_a) is what we are **actually** testing, since any defendant enters the trial as *innocent*, until proven otherwise.

- That is why this is our **alternative** hypothesis!

Statistical hypothesis testing

If the jury decides to **convict**, they are **rejecting** the *null* hypothesis in favor or the *alternative*.

In other words, there was **enough evidence** to conclude that the defendant was *guilty*.

If the jury **acquits**, they are **not rejecting** the *null* hypothesis.

That is, there was **not enough evidence** to conclude that the defendant was *guilty*.

- **Note**: in Statistics, we do not say that we **accept** the null hypothesis. This would mean that the defendant is *innocent*, but we are only able to say that they are not guilty.

Type I and Type II errors

Type I and Type II errors

There are **two** possible errors when working with Hypothesis Testing:

- *Type I error*: when we **reject a true** null hypothesis (H_0);
- *Type II error*: when we **do not reject a false** alternative hypothesis (H_a).

How do these errors apply in the *trial* context?

Type I and Type II errors

The **probability of a Type I error** is denoted by α, the significance level of a statistical test.

The **probability of a Type II error** is denoted by β.

Any attempt of reducing one probability will increase the odds of the other type of error.

Decision	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct decision
Do not reject H_0	Correct decision	Type II error

A quick summary

To summarize, a couple of remarks:

- The testing procedure begins with the assumption that the null hypothesis (H_0) is **true**;
- The goal is to determine whether there is enough evidence to infer that the alternative hypothesis (H_a) is true.

Stating hypotheses

Stating hypotheses

The **first step** when doing hypothesis testing is to **state** the *null* and *alternative* hypotheses, H_0 and H_a, respectively.

Let us exercise that by practicing with an **example**.

Recall the inventory example from the last lecture.

Now, suppose the manager does not want to estimate the exact (or closest) mean inventory level (μ), but rather test whether this value is **different from** 350 computers.

- Is there enough evidence to conclude that μ is **not equal to 350 computers**?

As an important **first note**, Hypothesis Testing always tests values for **population parameters**.

Then, the next step is to know **what** population parameter the problem at hand is referring to.

Stating hypotheses

Now, consider the following **change** in the research question for this example:

- Is there enough evidence to conclude that μ is greater than 350?

The z test

The z test

After the hypotheses are properly stated, what do we do?

As a simplfying assumption, we will continue to assume that the population standard deviation (σ) is known, while μ is not.

- We will **relax** this hypothesis soon.

Another example:

A manager is considering establishing a new billing system for customers. After some analyses, they determined that the new system will be cost-effective only if the mean monthly account is more than US$ 170.00. A random sample of 400 monthly accounts is drawn, for which the sample mean is US$ 178.00. The manager assumes that these accounts are normally distributed, with a standard deviation of US$ 65.00. Can the manager conclude from this that the new system will be cost-effective? Also, they assume a confidence level of 95%.

The z test

After stating the null and alternative hypotheses, we need to calculate a **test statistic**.

Recall the **standardization** method for a sample statistic:

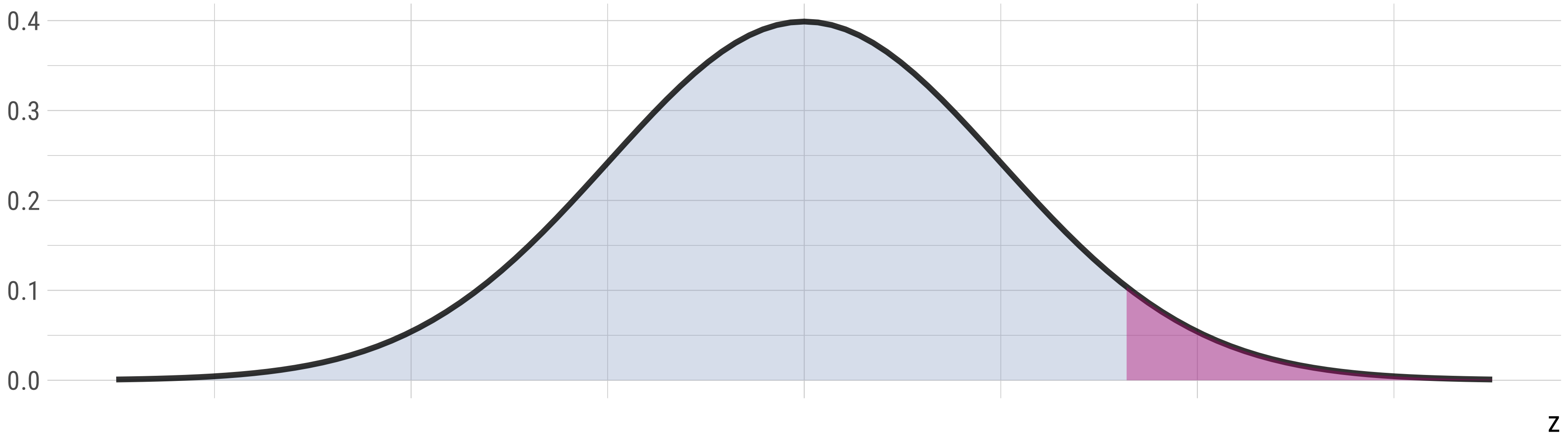
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

For hypothesis testing purposes, the above is also known as a ***z* test**.

The z test

After obtaining the z value, let us now make use of the confidence level $(1 - \alpha)$ of 95% assumed by the manager.

This value will be of use to establish a **threshold (critical) value** in a Standard Normal curve:



The z test

The **shaded area** is called the **rejection region**.

If a z statistic falls **within** the rejection region, our inference is to **reject the null hypothesis**.

In case the z value falls **outside** this region, then we **do not reject** the null hypothesis.

- So what is our decision from the example?

The p-value method

The p-value method

We may also produce inferences using **p-values** instead of critical values.

> The *p-value* of a statistical test is the probability of observing a test statistic *at least as extreme* as the one which has been computed, given that H_0 is true.

- What is the p-value in our example?

```r
1 - pnorm(q = 2.46, mean = 0, sd = 1)
```

```
#> [1] 0.006946851
```

The p-value method

A p-value of .0069 implies that there is a .69% probability of observing a sample mean at least as large as US$ 178 when the population mean is US$ 170.

In other words, this value says that we have a pretty good sample statistic for our Hypothesis Testing interests.

However, such interpretation is **almost never used** in practice when considering p-values.

Instead, it is *more convenient* to compare p-values with the test's significance level (α):

- If the p-value is **less** than the significance value, we **reject the null hypothesis**;
- If the p-value is **greater** than the significance value, we **do not reject the null hypothesis**.

The p-value method

Consider, for example, a p-value of **.001**.

This number says that we will only start not to reject the null when the significance level is **lower** than .001.

- Therefore, this means that it would be really **unlikely** not to reject the null hypothesis in such situation.

We can consider the following **ranges** and **significance** features for p-values:

- $p < .01$: **highly significant** test, *overwhelming* evidence to infer that H_a is true;
- $.01 \leq p \leq .05$: **significant** test, *strong* evidence to infer that H_a is true;
- $.05 < p \leq .10$: **weakly significant** test;
- $p > .10$: **little or no** evidence that H_a is true.

One- and two-tailed tests

One- and two-tailed tests

As you may have already noticed, the typical Normal distribution density curve has two "tails."

Depending on the sign present in the alternative hypothesis (H_a), our test may have **one or two** rejection regions.

Whenever the sign in the alternative hypothesis is either "<" or ">," we have a **one-tailed test**.

- For the former case, the rejection region lies on the *left* tail of the bell curve, whereas for the latter, the rejection region is located on the *right* tail.

One- and two-tailed tests

A two-tailed test will take place whenever the *not equal to* sign ($\neq$) is present in the alternative hypothesis.

- This happens because, assuming this sign, the value of our parameter may lie either on the *right* or on the *left* tail.

For two-tailed test, we simply divide the significance level (α) by 2.

- Just as with **confidence intervals**!

Next time: Inference when σ is unknown