ØAMET4100 · Spring 2019
Lecture Note 1B

Instructor: Fenella Carpena

January 10, 2019

This lecture note provides a review of Statistics (Stock & Watson, Chapter 3). This lecture note is not intended to be a comprehensive review of lecture or the textbook, since there is a lot more material than we have time to cover. However, I have tried to focus on the concepts which I believe are necessary to be successful in our class.

1   Estimation of the Population Mean

- As mentioned previously, a key theme of this course is that we want to say something about the population using a *sample* of that population. Now, we will introduce statistical methods that we can apply using the sample to reach conclusions about population. These statistical methods are: (1) estimation, (2) hypothesis testing, and (3) confidence intervals. We will use these three approaches throughout the course.

- Suppose you want to know the population mean of Y (i.e., the value μ_Y). How can we use the sample to estimate this population parameter? One way is to use the sample mean $\overline{Y}$. Another way is to just use the first observation Y_1. Still another way is to consider the weighted average in which the observations are alternately weighted by 1/2 and 3/2, namely, $\tilde{Y} = \frac{1}{n}(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \ldots + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n)$. All of these (i.e., $\overline{Y}$, Y_1, and $\tilde{Y}$) are called **estimators**.

- An **estimator** is a procedure for using sample data to compute an educated guess of the value of a population parameter (such as the population mean). Note that estimators are random variables. Can you explain why?

- A "good" estimator of the population mean should have the following three properties:

  - **Unbiasedness**. Let $\hat{\mu}_Y$ denote some estimator of μ_Y. Mathematically, unbiasedness is stated as $E(\hat{\mu}_Y) = \mu_Y$. In words, unbiasedness implies that the mean of the estimator must equal the true population parameter. For example, if $\tilde{Y}$ is an unbiased estimator for the population mean, then $E(\tilde{Y}) = \mu_Y$.

  - **Consistency**. When the sample size is large, the uncertainty about the value of μ_Y arising from random variations in the sample is very small. Stated more precisely, consistency means that the probability that $\hat{\mu}_Y$ is within a small interval of the true value of μ_Y approaches one as the sample size increases. Mathematically, $\hat{\mu}_Y \xrightarrow{p} \mu_Y$.

  - **Efficiency.** Efficiency entails a comparison of estimators. Suppose you had two candidate estimators, both of which are unbiased. If the first estimator has a smaller variance than the second estimator, then the first is said to be more efficient.

- Consider again the sample mean $\overline{Y}$. Let's see if this estimator satisfies the above three properties.

  - It is unbiased because $E(\overline{Y}) = \mu_Y$. Can you show this on your own?

  - It is consistent by the law of large numbers.

  - What about efficiency? Note that efficiency is based on comparing the variance across estimators. You can show that $var(\overline{Y})$ is smaller than the $var(Y_1)$ and $var(\tilde{Y})$. Hence, $\overline{Y}$ is more efficient than the other two estimators.

- In fact, $\overline{Y}$ is the **Best Linear Unbiased Estimator (BLUE)** for the population mean. BLUE means that out of all the estimators that are unbiased and are linear functions of $Y_1, \ldots, Y_n$, $\overline{Y}$ has the smallest variance. In addition, $\overline{Y}$ is the least squares estimator of μ_Y. We will return to these points later on in the course.

- The table below summarizes the various unknown parameters related to $\overline{Y}$ that we will encounter later (see Section 3 of these notes) and the estimators that we use for them. Note that as usual we have $Y_1, \ldots, Y_n$ i.i.d. with mean μ_Y and variance σ_Y^2.

Notation	What it represents	Estimator
μ_Y	$E(Y_i)$	$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ (sample average)
σ_Y^2	$var(Y_i)$	s_Y^2 (sample variance) Note: $s_Y = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$
$\frac{\sigma_Y}{\sqrt{n}}$	$sd(\overline{Y})$	$\frac{s_Y}{\sqrt{n}}$ (standard error of $\overline{Y}$)

Note that μ_Y and σ_Y^2 are characteristics of the **population distribution** (i.e., the distribution from which we drew the sample $Y_1, \ldots, Y_n$), while $\frac{\sigma_Y}{\sqrt{n}}$ is a characteristic of the **sampling distribution** (i.e., the distribution of the sample statistic $\overline{Y}$). The standard error of $\overline{Y}$ is used as estimator for the standard deviation of $\overline{Y}$ if σ_Y is unknown.

2   Basic Concepts of Hypothesis Testing

- **Hypothesis testing** refers to a set of procedures used to weigh the "strength" of statistical evidence. In particular, we use these tests to determine whether the evidence from our sample is "sufficiently strong" to support one hypothesis over another.

  ○ For example, suppose you were interested in determining whether a coin is fair or biased. If you toss the coin 100 times and find 51 heads, you might conclude that the coin is probably fair. If you find 90 heads, that would provide much more evidence that the coin is probably biased. But if you find 58 heads, what would you conclude? What is the threshold that we need before we can conclude that the coin is probably biased? Hypothesis testing helps us to figure that out.

- The approach requires a **null hypothesis** H_0 and an **alternative hypothesis** H_1.

  ○ The **null hypothesis** H_0 is the starting point of statistical hypothesis testing.

  ○ The **alternative hypothesis** H_1 is the hypothesis that contradicts the assertion of the null hypothesis.

  ○ An important feature to note about hypothesis testing is that H_0 and H_1 are not treated in the same way: H_0 is given the benefit of the doubt, and H_1 has the burden of proof. This framework is analogous to the presumption of innocence in the law: the defendant is presumed innocent until proven guilty beyond a reasonable doubt. Similarly, H_0 is presumed to be true unless the data strongly suggest otherwise.

- A hypothesis test can be either a **one-sided** or a **two-sided** test.

  ○ A **one-sided** test involves $>$ and $<$ in the alternative hypothesis, e.g., $H_1 : \mu > \mu_0$ or $H_1 : \mu < \mu_0$

○ A **two-sided** test involves $\neq$ in the alternative hypothesis, e.g., $H_1 : \mu \neq \mu_0$

- The conclusion of a hypothesis test is either: (1) we **reject** H_0, or (2) we **fail to reject** H_0. This wording is important for the following reasons.

  ○ When we conduct a hypothesis test, we are **NOT** testing whether H_0 or H_1 is true; rather, we are asking whether the available data is significant evidence to cast doubt on H_0 in favor of H_1.

  ○ Rejecting H_0 means that the data lend support to H_1. It does **NOT** mean that we have proven H_1 is true; rather, it means that the data from our sample are inconsistent with H_0, so the evidence in our data favor H_1.

  ○ Failing to reject H_0 means that the data lend support to H_0. It does does **NOT** mean that we have proven H_0 is true; rather it means that the data are consistent with H_0, so they lend support to H_0. Lending support is not the same as proving H_0; this is why we say "fail to reject H_0" instead of "accept H_0."

- Because every hypothesis test concludes by either rejecting or failing to reject the null, there are 2 ways of drawing incorrect conclusions.

  ○ **Type I error**: rejecting H_0 when H_0 is true (false positive).

  ○ **Type II error**: sticking with H_0 when H_1 is true (false negative).

  ○ Ideally, we would like the probability of both types of errors to be small. However, in practice, we cannot simultaneously reduce the chances of both Type I and Type II errors. For example, if we make it harder to reject H_0, then we are less likely to commit Type I error, but more likely to commit a Type II error.

- How do we conclude whether we reject or fail to reject the null hypothesis? We can use 3 methods: (1) **t-statistic**, (2) **p-value**, and (3) **confidence intervals**. All methods will arrive at the same conclusion. The next section discusses the procedures in more detail.

- **Why do we even care about hypothesis tests?** Let's tie this question in with what we've learned so far in this class.

  ○ We want to answer questions about the **population.**

  ○ But collecting data from the entire population is infeasible or impractical, we take a simple random sample (which as you recall, results in i.i.d observations).

  ○ Using data from the sample, we can calculate a **sample statistic**, e.g., the sample average $\overline{Y}$.

  ○ But because **there is only one population but many possible samples,** the value of the sample statistic we obtain would differ from sample to sample. This variability is called **sampling variation.**

  ○ Sampling variation also implies that even if we observed a sample average of, say, $\overline{Y} = 145$ in our sample, this does not mean that the population average height μ is equal to 145 as well.

  ○ Hence, we need to use statistical methods, in this case hypothesis testing, so that we can reach a conclusion about the population average μ.

3 Two-Sided Hypothesis Tests

In this class, we will be using hypothesis tests for means and differences of means. For the remainder of these notes, I will focus on **two-sided** hypothesis tests, and please refer to the textbook for one-sided hypothesis tests. I will also assume that we have a large sample size, so that by the Central Limit Theorem, we can use the normal distribution.

The following table summarizes the notation that we use and the different characteristics we need to know to be able to carry out a hypothesis test.

	Mean	**Differences of Means**
Population parameter	μ	μ_1, μ_2 Note: The subscript 1 and 2 refer to groups, e.g., 1 could be men, and 2 could be women.
Sample statistic	$\overline{Y}$	$\overline{Y}_1 - \overline{Y}_2$ Note: The order of subtraction follows the same order of the parameters in the null hypothesis.
Standard Error (SE) of sample statistic	$s_Y / \sqrt{n}$ Note: $s_Y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ Note: This assumes that sample group 1 and sample group 2 are from independent populations.
t-statistic	$t = \frac{\overline{Y} - \mu_{Y,0}}{SE(\overline{Y})}$ Note: $\mu_{Y,0}$ is the hypothesized value in the null hypothesis, e.g., if $H_0 : \mu = 4$, then $\mu_{Y,0} = 4$	$t = \frac{\overline{X}_1 - \overline{X}_2 - d_0}{SE(\overline{X}_1 - \overline{X}_2)}$ Note: d_0 is the hypothesized value in the null hypothesis

Note that the population parameter is usually what we want to say something about but it is unknown. The sample statistic is what we observe in the data.

3.1 Two-Sided Hypothesis Test Using the t-statistic

The **t-statistic** is the number of standard errors that separate the sample statistic (e.g., $\overline{Y}$, or $\overline{Y}_1 - \overline{Y}_2$) from the hypothesized value in the null hypothesis (e.g., $\mu_{Y,0}$ or d_0).

Here is a general "recipe" for conducting a hypothesis test using the t-statistic. You do not have to use this recipe, but I provide it here in case it is helpful.

- **Step 1:** Choose a level of significance α. This is our tolerance for a Type I error. Note: This α might also just be given to you.

- **Step 2:** Formulate the null and alternative hypothesis.

  - Important: The null/alternative hypothesis involve the population parameter μ, **NOT** the sample statistic $\overline{Y}$.

- **Step 3:** Calculate the t-statistic (see table above for the formulas).

  - Note that the t-stat has the following format:

$$\frac{sample\ statistic - value\ from\ hypothesis}{se(sample\ statistic)}$$

- **Step 4:** Find the "two-sided critical value" in the sampling distribution that corresponds to α. Assuming that our sample size is large enough, we can use the normal table to find this critical value.

○ The two-sided critical value is the number $z_{\alpha/2}$ from the normal table such that $P(|Z| > z_{\alpha/2}) = \alpha$.

○ In words, $z_{\alpha/2}$ is basically a positive number $z > 0$ such that the area to the right of z is $\alpha/2$, and the area to the left if $-z$ is $\alpha/2$ (so the total sum of the two areas is α).

- **Step 5:** Compare the *t-statistic* (from Step 3) with the *two-sided critical value* (from Step 4), and apply the following rejection rule.

  ○ **Reject** H_0 if $|t\text{-}stat| >$ *two-sided critical value*. Otherwise, we **fail to reject** H_0**.**

  ○ If we reject H_0, for example with $\alpha = 0.05$, then we say that our result is **statistically significant** at the 5% level.

  ○ Note: The rejection rule for the two-sided test takes the absolute value of the t-stat. This is because for a two-sided test, we are looking for evidence against the null in both the left and right tails, i.e., either *t-stat* $< -1 *$ *two-sided critical value* or *t-stat* $>$ *two-sided critical value*. Taking the absolute value covers both cases.

Tip: I would recommend memorizing the two-sided critical values from the normal table for α levels 1%, 5% 10%, since these are the most common values of α when carrying out a hypothesis test.

- $\alpha = 0.10$: $z_{0.10/2} = 1.64$

- $\alpha = 0.05$: $z_{0.05/2} = 1.96$

- $\alpha = 0.01$: $z_{0.01/2} = 2.58$

3.2 Two-Sided Hypothesis Test Using the p-value

The **p-value** is the probability of drawing a value of the sample statistic that is at least as extreme as what we observe in our sample, if the null hypothesis is true. It is also the smallest α level at which H_0 can be rejected. We can think of the p-value as the weight of evidence against H_0, that is, small p-values mean that H_0 is not plausible.

Here is a general "recipe" for conducting a hypothesis test using the p-value. You do not have to use this recipe, but I provide it here in case it is helpful.

- **Step 1 to Step 3:** Follow the same steps as in the previous section. Make a note of the t-statistic that you calculated in Step 3, since you will be needing this in the next step.

- **Step 4:** Find the following probabilities using the normal table, which give us the p-value.

  ○ *p-value* $= P(|Z| > |t\text{-}stat|) = 2 \cdot P(Z > |t\text{-}stat|) = 2 \cdot P(Z < -|t\text{-}stat|)$. In words, this means that we need to take the absolute value of the t-stat, find the area to the right of the absolute value of the t-stat, or equivalently the area to the left of $-|t\text{-}stat|$, then multiply that area by 2 to find the *p*-value.

- **Step 5:** Compare the p-value obtained from Step 4 with the level of α from Step 1. We **reject** H_0 if p-value $< \alpha$. Otherwise, we **fail to reject** H_0.

  ○ A potentially useful phrase to help you remember the rejection rule: "If the *p*-value is low, the null has to go."

  ○ If we reject H_0, for example with $\alpha = 0.05$, then we say that our result is **statistically significant** at the 5%.

3.3 Two-Sided Hypothesis Test Using the Confidence Interval

A **confidence interval** (CI) is a range of plausible values of a parameter (e.g., μ) based on a sample from the population (e.g., $\overline{Y}$).

Here is a general "recipe" for conducting a hypothesis test using the confidence interval. You do not have to use this recipe, but I provide it here in case it is helpful.

- **Step 1 to Step 3.** As before, choose significance level α, formulate the null and alternative hypotheses, and find the two-sided critical value. The important ingredient you will need here is the two-sided critical value.

- **Step 4.** Construct the $(100 - \alpha)\%$ confidence interval using the formula:

$$\underbrace{sample\ statistic}_{\text{e.g., } \overline{Y} \text{ or } \overline{Y}_1 - \overline{Y}_2} \pm \underbrace{critical\ value}_{z_{\alpha/2}} * \underbrace{SE(sample\ statistic)}_{\text{see formula in table above}}$$

- **Step 5.** Check if the hypothesized value (e.g. $\mu_{Y,0}$, d_0) is contained in the CI you constructed in Step 4, then apply the following rejection rule:

  ○ If the hypothesized value is outside the CI, we **reject** H_0.

  ○ Otherwise, we **fail to reject** H_0.

3.4 Equivalent Inferences for t-statistic, p-value, and Confidence Interval

- As mentioned above, all three methods will reach the same conclusion.

- The difference between the three lies in the "types" of information they contain.

  ○ The hypothesis test using t-stat tells us whether we reject or fail to reject only *one* particular hypothesized value.

  ○ The p-value gives us the smallest significance level α for which we can reject the null.

  ○ The CI gives a list of all hypothesized values we would reject (i.e., those outside the CI) or fail to reject (i.e., those within the CI).