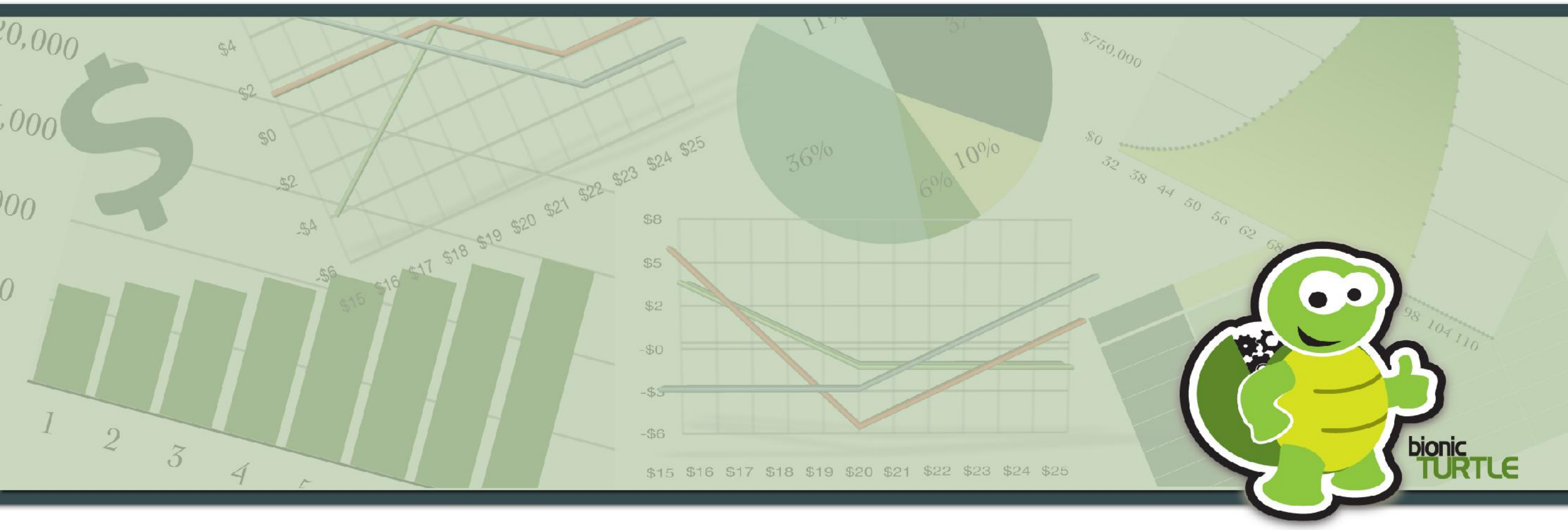


P1.T2. Quantitative Analysis

Chapter 6: Hypothesis Testing

Bionic Turtle FRM Study Notes

Chapter 6: Hypothesis Testing

Chapter 6: Hypothesis Testing

- **Construct an appropriate null hypothesis and alternative hypothesis and distinguish between the two.**
- **Differentiate between a one-sided and a two-sided test and identify when to use each test.**
- **Explain the difference between Type I and Type II errors and how these relate to the size and power of a test.**
- **Understand how a hypothesis test and a confidence interval are related.**
- **Explain what the p-value of a hypothesis test measures.**
- **Interpret the results of hypothesis tests with a specific level of confidence.**
- **Identify the steps to test a hypothesis about the difference between two population means.**
- **Explain the problem of multiple testing and how it can bias results.**

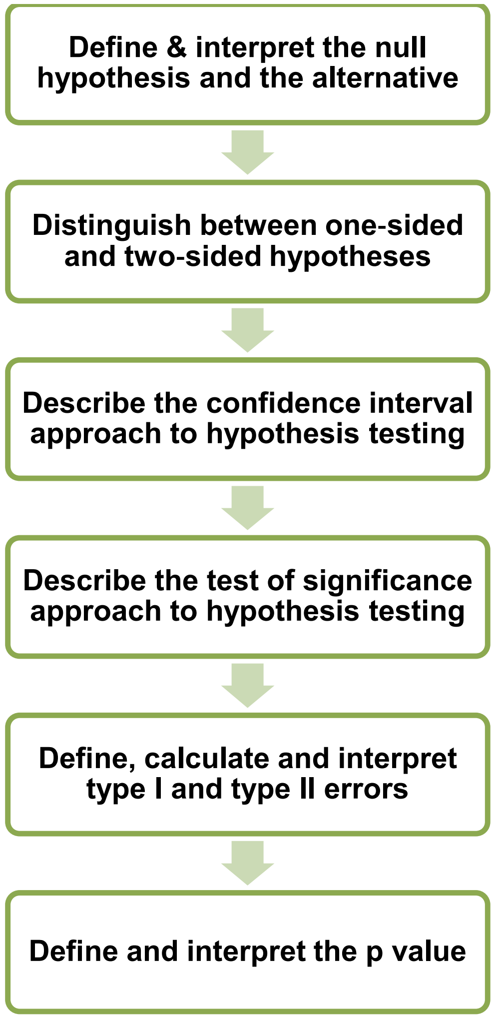
Selected key concepts:

- The null hypothesis must contain the equals sign, correct nulls are: $H(0): \mu = c$, $H(0) \leq c$, or $H(0) \geq c$.
- While the majority of hypothesis test are two-tailed (and this can be your default assumption) risk-type questions often seek a one-tailed test: pay careful attention to the wording of the question. A good question must make it clear which is sought.
- The sample mean test statistic is $(X - \mu)/SE$, where standard error (SE) is given by $\sigma/\text{sqrt}(n)$. If we do not know the population variance, the student's t (t test) is appropriate.
- When the sample size is large ($n > 30$), the normal Z distribution (Z test statistic) can be used to approximate the student's t. At 95.0% confidence, the large sample critical value is about 2.0. This is the large sample "neighborhood" of 2.0 to 3.0: for example, you do not need a lookup table to reject a (sample mean) test statistic of 7.3 or 9.5, prima facie these values are too large (too far away from the null) and we can safely reject the null.
- The **p-value is the exact significance level**: it is the lowest significance level at which the null hypothesis can be rejected. For example, a p-value of 3.5% implies rejection at 95.0% confidence (ie, 5.0% > 3.5%) but "acceptance" (failure to reject) at 99.0% confidence (ie, 1.0% < 3.5%).

Please Note: This Study Note begins immediately with several *fully illustrated* examples. There are three ways to test a hypothesis: significance test, confidence interval and/or p value. We think it's easier to examine examples than to debug abstract terms and formulas, even if the concepts are initially unfamiliar. Don't worry, it will all be explained!

The general framework (prior to first learning objective)

This chapter concerns the details of hypothesis testing, and its general framework (the basic steps) are shown here:



In mathematical terms, there is only a *single* mechanic involved here. But when it comes to an interpretation of the observed statistic (and of the *difference* between the observation and a hypothesized null value) there are *three different approaches* to the hypothesis test. Because they are merely different approaches to the same underlying mechanics, they **must always lead to consistent (aka, similar) conclusions**.

- Construct a confidence interval
- Conduct a significance test
- Compute a p-value

Test of a sample mean: Normal Z versus Student's t (prior to first learning objective)

As a practical matter, many of our hypothesis tests are tests of a sample mean. In testing a sample mean (regardless of which of the three approaches we use: confidence interval, significance test, and/or p-value), we will need to choose between the normal and student's t distribution. In regard to this choice, there are three essential questions:

- **Are we sampling from a normal distribution?** Realistically, we probably cannot make this assumption, or we just do not know. However, this is a common academic or exam-type assumption that is often made for the sake of convenience.
- **Do we know the population variance?** In most realistic cases, we do not know the population variance. Instead, *we use the sample variance* as an estimate of the population variance.
- **Is the sample small (< 30) or large (≥ 30)?**

If we are sampling from a normal distribution, and if we do know the population variance, then regardless of the sample size, we are justified in using the normal distribution and our test statistic is given by:

$$\text{normal Z} = \frac{\overline{X} - \mu_0}{\frac{\sigma_X}{\sqrt{n}}}$$

However, in most cases, we do not know the population variance such that our appropriate distribution is the student's t distribution (with degrees of freedom equal to sample size minus one; df = n - 1) and our test statistic is given by:

$$\text{student's t} = \frac{\overline{X} - \mu_0}{\frac{S_X}{\sqrt{n}}}$$

The magical central limit theorem (CLT) ensures this student's t is appropriate for large samples even when we do not know whether the underlying distribution is normal. You read correctly: we can assume the **student's t for a test of the sample mean for any large sample** and we require no distributional assumption. Because the student's t converges toward the normal as sample size increases ($n \rightarrow \infty$), the normal Z is often used to approximate the student's t. That's why you will see the normal Z deviates used for sample mean tests of large (read: realistic) datasets (But because the population's variance is unknown, this use case is actually using the normal to approximate the student's t.).

Okay, so when are we stuck? You can probably deduce that we are only stuck when both the sample size is small (because the CLT isn't getting enough help, so to speak) and when we do not know the distribution is normal.