A. Chapter 26.1-26.4: Overview

The basic idea on chapter 26.1-26.4: we make assumptions about the value of the parameters, and then test to see if those assumptions could have led to the outcomes (statistics) we observed. We then use a probability calculation (a Z score and areas) to express the strength of our conclusions.

The basic question: Suppose you have a sample outcome (sample statistic) which is different from the expected value (population parameter) was the difference in the sample outcome we observed due to chance error or something else?

Let's walk through the example at the beginning of Chapter 26 together.

- A senator introduces a bill to simplify the tax code. His claim is the bill is revenue-neutral. Basically it won't change the amount of taxes the government collects, it just simplifies the law.
- A law can be evaluated. The IRS could SAMPLE from the POPULATION of all tax returns, figure out the effect the bill would have on these revenues, and then check to see if the bill is really revenue-neutral.
- In the example, the IRS samples 100 forms. The sample average comes out to -$219 which means that the government would have collected 219 fewer dollars from taxpayers. The sample standard deviation is $725.
- The senator's argument (issued through an aide) is that the SD is so large, $725, that an average of -219 is inconsequential.
- The IRS's argument is what you want to learn. To understand the -219 and the 725, you need to convert the sample SD to an SE for the sample average.
- Remember what the SE is, it is variation association with sample statistics. The SD is the variation in a given sample (e.g. the 725 here) or the variation in a population, or in a list (see Chapter 4).
- The IRS goes on to say, the senator may think/argue that the population parameter is $0, but the IRS thinks it's not zero and in fact it is also negative.

How do they figure that?

First, they calculate an SE for the sample average

$$\frac{\sqrt{100} * 725}{100} \approx \$72$$

Second, they set up a "test" and use the Z score as the "test statistic"

$$\frac{-219-0}{72} \approx -3.0 = Z$$

This test statistic Z is interpreted in this manner in chapter 26: that if the true parameter was zero dollars and the samples of size 100 have a variation (standard error) of $72 then the chance (probability) that you could have picked a sample of size 100 with a mean of -219 is about 0.1 of 1% which is the area to the left of -3 under the normal curve.

In Chapter 26, when using the Z score, you are always interested in the area in the "tail" which is as extreme or more extreme than your statistic. So in this case, we are looking at areas beyond –3 Z scores. In previous chapters, you learned to work with the formula for a Z score and the normal curve. In Chapter 26, it all comes together.

The basic question restated: one side thinks any observed difference between what you expect and what you get is REAL (so perhaps something is wrong with the expected value or with the sample process). The other side thinks the difference is just random chance error operating. If your observed value is too many STANDARD ERRORS away from the expected value, this is hard to explain by chance alone. For example, here, we are seeing a chance of .1% which is very small, this means you could have gotten this outcome randomly only once in 1000 samples. The number of standard errors away is called a Z score and the method you use to arrive at this score is a "test of significance"

B. Vocabulary

The **NULL HYPOTHESIS (26.2)** is that the observed results are purely due to chance alone. That is, any differences between the parameter (the expected value) and the observed (or actual) outcome are due to chance only. In this case, the null hypothesis is a statement about a parameter: the population average is 0 for the IRS example in 26.1.

The **ALTERNATIVE HYPOTHESIS (26.2)** is that the observed results are due to more than just chance. It implies that the NULL is not correct and any observed difference is real, not luck.

Usually, the **ALTERNATIVE** is what we're setting out to prove. The **NULL** is like a "straw man" that we wish to knock down.

The **TEST STATISTIC (26.3)** measures how different the observed results are from what we would expect to get if the null hypothesis were true. When using the normal curve, the test statistic is z,

$$Z = \frac{observed\_statistic - hypothesized\_value}{standard\_error}$$

All a Z does in Chapter 26 is tell you how many SEs away the observed statistic is from the expected (hypothesized) value when the expected (hypothesized) value is generated from the **NULL HYPOTHESIS**.

The **SIGNIFICANCE LEVEL (or P-VALUE) (26.3).** This is the chance of getting results as or more extreme than what we got, IF the null hypothesis were true. P-VALUE could also be called "probability value" and it is simply the area associated with the calculated Z.

p-values are always "if-then" statements:

"If the null hypothesis were true, then there would be a p% chance to get these kind of results."

So in our case: if the bill is truly revenue-neutral, there would be less than 0.1% chance to get a result of –219 from a sample of 100 returns.

If the p-value is less than 5%, we say the results are **STATISTICALLY SIGNIFICANT (26.4)**; if $p < 1\%$, the results are **HIGHLY STATISTICALLY SIGNIFICANT**. A "significant" result means that it would be unlikely to get such extreme observed values by chance alone.