More About Hypothesis Testing and Tests of Significance (Chapter 26.2-26.4)

1. Overview

Remember the basic idea about hypothesis testing: we make either make assumptions about the parameters or the parameters are known, and then we test to see if those parameters could have led to the sample outcome we observed. We then use a probability calculation to express the strength of our conclusions.

What we are doing in Chapters 26 is learning about a "method" for making decisions. This involves identifying the parameter and statistic, constructing a test, and then interpreting the results of the test to determine whether any differences between what was expected and the actual outcome could have happened by chance or is the difference meaningful?

Recall that we assume that the parameters are known, even if they are not. So for example, we might use a hypothetical box average or a hypothetical expected value. We might also use a hypothetical box standard deviation (or at least use the standard deviation from a sample and pretend that is a box standard devation).

2. An example

The Statistical Abstract of the United States reported that the net average earnings of all M.D.'s was $155,800 per year with a standard deviation of $23,400. Suppose the earnings are normally distributed.

A random sample of income tax returns of 9 M.D.'s practicing in rural communities showed that their net earnings to be:

93,700 110,500 173,600 123,300 136,800 142,700 129,900 153,400 140,200

Let's assume that the earnings of these M.D.'s follow a distribution that is approximately normal, use a 1% level of significance to test the claim that the mean earnings of all rural M.D.'s is less than the national average. If you know the data is normally distributed, you can use Z.

1) What is the population? What are the parameters?

All MDs in the US. Average income is $155,800 and the SD is $23,400

2) What is the sample? What are the statistics?

9 MDs. Sample Average= 133,789, Sample SD = $22,045

3) What is the null hypothesis? (See definitions, next section)

Rural average earnings are not different from US average earnings, that is $155,800

4) What is the alternative hypothesis? (See definitions, next section)

Rural average earnings are less than US average earnings. That is < \$155,800

5) What is the appropriate test statistic? (See definitions, next section)

Ask yourself...what do I know about the population?

I'll say, although the sample size is very small (9) but the population standard deviation is known (Chapter 26.6 deals with situations when population standard deviation is unknown – but that's beyond the scope of this course) and we know it's a normal population so a Z-test is appropriate here.

z = (observed value - expected value)/ appropriate measure of spread

(133789 - 155800) / SE of the average.

Why the SE of the average?

SE of the average = $\dfrac{\sqrt{9} * 23{,}400}{9}$ = 7,800 dollars .

So how far is 133789 from 155800? about a Z of -2.85.

6) What is the interpretation of the test statistic given the stated significance level of 1% (or p-value of 1%)?

First, what is being suggested is that the person asking the question will only consider p-values of 1% or less. In other words, returning to concepts in Chapters 17, 19, and 20 if there is less than a 1% chance of getting an average this low if we were expecting 155,800 I will reject the null hypothesis in favor of the alternative.

The p-value associated with a Z of -2.85 is about .22%. How did I get that? The probability of getting a Z between +2.85 and -2.85 is 99.56% (from table A 105). So the chance of being outside of that is (100 - 99.56) or .44%. Since the question stated the alternative as a one-tailed (or one-sided) we're only interested in the left side of the curve.

3. Issues Related To Hypothesis Testing

A. Parameters are usually unknown, therefore either hypothetical values are used for parameters OR values from samples might substitute for parameters. The preferred values in order of importance are (a) the value of the true parameter (b) a hypothetical value which is assumed to equal the true parameter value and (c) a value derived from the sample

B. If the sample size is small (less than 30 for most textbooks, less than 100 for this textbook) BUT the parameters are known AND the population is known to be normal AND the process could be repeated, a Z test can be used

C. If the sample size is small and you do not know whether the population is normal, a Z test should not be used even if you know the values of the parameters and even if you can repeat the process.

D. If the sample size is large and unbiased and if the process is repeatable, you can use the Z test even if you must make assumptions about the values of the parameters (typical).

E. Beware of situations when researchers have very large samples and declare everything to be “statistically significant”. If sample sizes become large enough, differences between “observed values” and “expected values” will test as “significant” even though they are very small.

F. Remember, that The SIGNIFICANCE LEVEL (or P-VALUE) 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.

4. Hypothesis Testing Summarized

A. Clearly identify the parameter and the outcome.

B. State the null hypothesis. This is what is being tested. A test of significance assesses the strength of evidence (outcomes) against the null hypothesis. Usually the null hypothesis is a statement of "no-effect" In the doctor example, we would say that the incomes of rural doctors are the same as urban ones.

C. State the alternative hypothesis. This is a counter to the null and a restatement of the null hypothesis. The counter comes from the new information that rural MDs may have lower incomes.

D. Calculate the test statistic. The significance test assesses the evidence by examining how far the test statistic falls from the proposed null. The test is Z when certain conditions are met (see Part 3, above).

E. The probability that is observed as a result of the Z test is called a P-VALUE. The smaller the p-value the stronger is the evidence against the null hypothesis.

F. On significance levels. Sometimes prior to calculating a score and finding it's p-value, we state in advance what we believe to be a decisive value of p. This is the significance level. 5% and 1% significance levels are most commonly used. If your p-value is as small or smaller than the significance level you have chosen then you would say that "the data is statistically significant at the ---- (e.g. 1% or 5% or some other) level."