Chapter 15. Test of Significance: The Basics

The Reasoning of Tests of Significance

Note. Tests of significance are intimately related to confidence intervals. The "level of confidence" of Chapter 14 will be translated into a "level of significance" in this chapter. The basic idea of a significance test is: "*an outcome that would rarely happen if a claim were true is good evidence that the claim is not true.*" As a result, we will often make a claim (or "hypothesis"), take a sample, and find that the probability of the particular sample being drawn is small assuming the claim is true (this is a conditional probability). We then reject the claim. Therefore we will often gather data with an eye towards finding evidence *against* a claim.

Note. Throughout this chapter, we make the same assumptions of Chapter 14: we have a perfect simple random sample from a normally distributed population with standard deviation σ which we know.

Example. Exercise 15.1 (notice part (b) especially).

Stating Hypotheses

Definition. The statement being tested in a statistical test is called the **null hypothesis**. The test is designed to assess the strength of the evidence *against* the null hypothesis. Usually the hypothesis is a statement of "no effect" or "no difference." The claim about the population that we are trying to find evidence *for* is the **alternative hypothesis**. The alternative hypothesis is **one-sided** if it states that a parameter is *larger than* or *smaller than* the null hypothesis value. It is **two-sided** if it states the parameter is *different from* the null value.

Note. "We abbreviate the null hypothesis as H_0 and the alternative hypothesis as H_a. *Hypotheses always refer to a population, not to a particular outcome. Be sure to state* H_0 *and* H_a *in terms of population parameters. ... The hypotheses should express the hopes or suspicions we have before we see the data. It is cheating to first look at the data and then frame hypotheses to fit what the data show.*" (page 366)

Example S.15.1. Hypothesized Stooges.

A Stoogeologist thinks that in all of the 97 Three Stooges films with Curly as the third stooge, Moe eye poked Curly an average of 4 times per film. She plans on performing a statistical test to explore this idea. What is her null and alternative hypotheses?

Solution. The null hypothesis (always a hypothesis of equality) is $H_0 : \mu = 4$. The alternative hypothesis is $H_a : \mu \neq 4$. This is similar to Exercise 15.6 on page 367. We should comment that a statistical test never gives evidence for H_0 but can only give evidence against H_0. Therefore this test will not (statistically) validate the Stoogeologist's idea, but it is possible that the Stoogeologist's idea could be found to be statistically unlikely.

Test Statistics and P-Values

Note. A test-statistic is a measure of the distance of a parameter from its value as hypothesized by H_0 to its estimated value from a sample. The test-statistic is measured (in most cases) in units of sample standard deviations.

Definition. The probability, computed assuming that H_0 is true, that the test statistic would take a value as extreme or more extreme than that actually observed is called the **P-value** of the test. The smaller the P-value, the stronger the evidence against H_0 provided by the data.

Note. "*Failing to find evidence against H_0 means only that the data are consistent with H_0, not that we have clear evidence that H_0 is true.*" (page 370) The only outcomes of a statistical test are to either (1) reject H_0 (with a certain level of confidence as expressed by the P-value), or (2) fail to reject H_0 (at some desirable level of confidence). The logic of a statistical tests implies that "accept the null hypothesis" is not a possible conclusion.

Example S.15.2. P-Stooges.
Consider Example S.15.1 again. In this problem, we have

$$H_0 : \mu = 4 \text{ eye pokes/film and } H_a : \mu \neq 4 \text{ eye pokes/film.}$$

Suppose a sample of size $n = 15$ is taken and it is found that the sample mean is $\overline{x} = 3.2$ eye pokes/film. In order to find a P-value, we need to know the population standard deviation, so suppose $\sigma = 1.5$ (all of this data is hypothetical). Go to the *P-Value of a Test of Significance* applet provided by the textbook website. Input this data to find the P-value for this data. (HINT: The answer is $p = 0.0384$. This means that the null hypothesis is very improbable.)

Note. You can also use Minitab to find P-values. To perform the calculations mentioned in Example S.15.2, click on the `Stat` tab and on the pull down menu `Basic Statistics` and `1-Sample Z`. In the pop-up window, enter the data (`Sample size` [15], `Mean` [3.2], `Standard deviation` [1.5], and `Test mean` [4]). Minitab will output a P-value of 0.039.

Statistical Significance

Note. "We can compare the P-value with a fixed value that we regard as decisive. This amounts to announcing in advance how much evidence against H_0 we will insist on. The decisive value of P is called the **significance level**. We write it as α..." (page 371)

Definition. If the P-value is as small or smaller than α, we say that the data are **statistically significant at level** α.

Example S.15.3. Significant Stooges.
If we choose $\alpha = 0.05$ in Example S.15.2, then the data is "significant at level α" since $P = 0.0384 < 0.05 = \alpha$. This means that there is significant statistical evidence at the level $\alpha = 0.05$ that the null hypothesis is not true and that the mean number of eye pokes is not 4 eye pokes/film. (Notice that the data is not significant at a level of $\alpha = 0.01$.)

Tests for a Population Mean

Note. Again the text presents a 4-step process. This time it is for a test of significance:

State: What is the practical question that requires a statistical test?

Formulate: Identify the parameter and state null and alternative hypotheses.

Solve: Carry out the test in three phases:

(a) **Check the conditions** for the test you plan to use.

(b) Calculate the **test statistic**.

(c) Find the P**-value.**

Conclude: Return to the practical question to describe your results in this setting.

Note. Now for some real substance! This is the protocol for the z test for a population mean:
Draw a SRS of size n from a normal population that has unknown mean μ and known standard deviation σ. To **test the null hypothesis that μ has a specified value**,

$$H_0 : \mu = \mu_0$$

calculate the **one-sample z test statistic**

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}.$$

In terms of a variable Z having the standard normal distribution, the P-value for a test of H_0 against:

$$H_a : \mu > \mu_0 \text{ is } P(Z \geq z)$$
$$H_a : \mu < \mu_0 \text{ is } P(Z \leq z)$$
$$H_a : \mu \neq \mu_0 \text{ is } 2P(Z \geq |z|)$$

Note. We again mention the assumptions of the above test: (1) the *sample must be random*, (2) the *population is assumed to be normally distributed* (this may not be the case), and (3) the *standard deviation of the population is known* (this is really the most unrealistic part of the process).

Example S.15.4. Z Stooges.
Calculate the z test statistic for the data given in Example S.15.2. What is the probability interpretation of the P-value? How would you use the Standard Normal Probabilities (Table A) to estimate the P-value?

Solution. From Example S.15.2, we have $\overline{x} = 3.2$ eye pokes/film, $\mu_0 = 4$ eye pokes/film, $\sigma = 1.5$ eye pokes/film and $n = 15$. Therefore the z test statistic is

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.2 - 4}{1.5/\sqrt{15}} = -2.0656.$$

The P-value $2P(Z \geq |z|) = 2P(Z \geq 2.0656)$ is the probability that the null hypothesis is wrong (and the alternative hypothesis is correct). We can use Table A to find the P-value by looking up 2.07 in the table from which we find the entry 0.9808. Now this number represents $P(Z \leq 2.07)$, so $P(Z \geq 2.07) = 1 - 0.9808 = 0.0192$. So the P-value is $2P(Z \geq 2.07) = 2 \times 0.0192 = 0.0384$ (as mentioned in Example S.15.3—we are lucky that we did not suffer from any round-off error here!).

Tests from Confidence Intervals

Note. A level α two-sided significance tests rejects a hypothesis $H_0 : \mu = \mu_0$ exactly when the value μ_0 falls outside a level $1 - \alpha$ confidence interval for μ.

Examples. Exercises 15.40 and 15.48

rbg-3-13-2009