MAT 2379, Introduction to Biostatistics

Chapter 11. Hypothesis Testing

Hypothesis testing is a statistical method that is used when one wants to gain support (or evidence) towards a desired statement, called *the research hypothesis* (denoted with H_1). The other hypothesis, which the researcher would like to reject is called the *null hypothesis* and is denoted with H_0. When you are using this method, formulate your statistical hypotheses in hopes of being able to reject H_0 and hence gain evidence towards H_1.

Hypotheses testing can result in two types of errors:

1. *Type I error* We have rejected hypothesis H_0 when H_0 was true.

2. *Type II error* We have failed to reject hypothesis H_0 when H_1 was true.

In class we will draw the table which contains all the 4 possible situations.

Statisticians working in the pharmaceutical industry encounter this all the time when they are trying to promote a new drug and to convince the population that this drug is efficient.

Example 1. (adapted from p. 190 of the the book "Biostatistics. How it works" by Selvin) The systolic blood pressure level in a certain hypertensive population is approximately equal to the value 130 mm Hg. A new drug is developed to reduce the systolic blood pressure levels in this population under the value 130. Set up the two hypotheses. Explain when the two errors occur.

Let μ denote the mean level of systolic blood pressure for the patients who were administered the drug. The null hypothesis (that we would like to reject) is that the drug didn't change anything:

$$H_0 : \mu = 130$$

The research hypothesis (that we would like to gain support for) is that μ is below 130:

$$H_1 : \mu < 130$$

A type I error occurs when we decide that the new drug has reduced the systolic blood pressure level in that population when, in fact, it has not.
A type II error occurs when we are unable to gain enough evidence that the new drug has reduced the systolic blood pressure level in that population when, in fact, it has.

To perform a test of hypotheses for μ, we need to consider 2 cases:

- the variance σ^2 is known (Section 11.1). We skip this case (it is never encountered in practice).

- the variance σ^2 is unknown (Section 11.2).

We also discuss hypothesis testing for the proportion p (Section 11.3).

11.2 Hypothesis Testing for the Mean: σ^2 unknown

In this section we will study the problem of testing a hypothesis on the mean μ of a population. There are three different cases. In all of these cases we will encounter a (known) numerical value μ_0 for the unknown mean, which is called *the null value*.

Case I $H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$

This is the case when we would like to gain evidence (or to convince) that the true mean μ of the population is larger than a numerical value μ_0. In order to do this, we calculate the difference $\bar{x} - \mu_0$ (where $\bar{x}$ is the mean of our observed sample) and we hope that it is a positive number, even a *large* positive number. If this is the case, then we reject H_0; otherwise, we say that we do not have enough evidence for rejecting H_0. This is called a *right-tailed test*.

How large the difference $\bar{x} - \mu_0$ should be? To answer this question, the difference $\bar{x} - \mu_0$ itself is not of big help. We have to calculate the ratio

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

and to remember that the random variable

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

has a T distribution with $n - 1$ degrees of freedom (by Theorem 9.4).

In this context, T_0 is called the *test statistic*, and t_0 is the observed value of the test statistic.

Using Table 17.4, we can calculate the probability that T takes a larger value (i.e it is more extreme) that the value t_0 that we have already observed. This probability is called the *p-value of the right-tailed test*:

$$p - \text{value} = P(T > t_0) = P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

The smaller the p-value, the less likely it becomes that H_0 is true.

For a given value α (called the significance level), we use the rule:

$$\text{if } p\text{-value} < \alpha, \quad \text{then we reject } H_0$$

Using this rule ensures that the probability of the Type I error is equal to α.

Example 2. (adapted from p.212 of "Biostatistics. How it works" by S. Selvin) The standard permissible exposure to benzene in the oil refining industry is 1 part per million (ppm). An average value larger that 1 ppm is felt to be dangerous. An industrial hygienist at a specific oil company measured the benzene exposure levels of 25 workers. He obtained a sample mean $\bar{x} = 1.03$ ppm with a sample standard deviation $s = 0.075$ ppm. Using this data, is there enough evidence that the level of exposure is larger than 1 ppm? Justify your answer based on the p-value.

Let μ be the average level of exposure. We are interested in testing

$$H_0 : \mu = 1 \qquad H_1 : \mu > 1$$

The test statistic is

$$T_0 = \frac{\bar{X} - 1}{S/5}$$

The observed value of the test statistic is

$$t_0 = \frac{\bar{x} - 1}{s/5} = \frac{1.03 - 1}{0.075/5} = 2.00$$

Hence, $p-\text{value} = P(T \geq 2.00)$ where T is a random variable with a T distribution with 24 degrees of freedom. To find this probability, we look in Table 17.4. We see that the value 2.00 falls between 1.711 and 2.064, with $P(T > 1.711) = 0.05$ and $P(T > 2.064) = 0.025$. We can conclude that

$$0.025 < p - \text{value} < 0.05.$$

Since p-value < 0.05, we reject H_0 (in favor of H_1). There is enough evidence that $\mu > 1$.

Case II $H_0 : \mu = \mu_0, \quad H_1 : \mu < \mu_0$

This is the case when we would like to gain evidence that the true mean μ of the population is smaller than a numerical value μ_0. In order to do this, we have to show that the difference $\bar{x} - \mu_0$ is a negative number, which is far from 0. If this is the case, then we reject H_0. This is called a *left-tailed test*.

How far this number should be from 0? To answer this question, we use the same test statistic as in Case I. We have to calculate the probability that a random variable T takes a value smaller (i.e more extreme) that the one that we already observed. This probability is called the *p-value of the left-tailed test*:

$$p - \text{value} = P(T < t_0) = P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} < \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

(Note that in this case, t_0 is negative.) The smaller the p-value, the less likely it is that H_0 is true.

Example 1. (continued) Suppose that the new drug has been administered to a random sample of 16 patients. The observed values for this sample yielded a sample mean $\bar{x} = 123.7$ and a sample standard deviation $s = 15.4$. Is there enough evidence that the drug is efficient in reducing the systolic blood pressure? Use the significance level $\alpha = 0.01$.

We want to test

$$H_0 : \mu = 130 \qquad H_1 : \mu < 130$$

The test statistic

$$T_0 = \frac{\bar{X} - 130}{S/\sqrt{16}}$$

has a T distribution with 15 degrees of freedom. The observed value of the test statistic is

$$t_0 = \frac{\bar{x} - 130}{s/\sqrt{16}} = \frac{123.7 - 130}{15.4/\sqrt{16}} = -1.636$$

Hence,

$$p - \text{value} = P(T < -1.636) = P(T > 1.636)$$

We now look in Table 17.4, on row 15. We see that the value 1.636 falls between 1.341 and 1.753, where

$$P(T > 1.341) = 0.10 \quad \text{and} \quad P(T > 1.753) = 0.05$$

Here T is a random variable with a T distribution with 15 degrees of freedom. We conclude that

$$p - \text{value is between } 0.05 \text{ and } 0.10$$

Since the p-value is greater than $\alpha = 0.01$, we fail to reject H_0. We conclude that there is not enough evidence that the new drug is efficient in reducing the systolic blood pressure.

Case III $H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$

This is the case when we would like to gain evidence that the true mean μ of the population "stays away" from a numerical value μ_0. In order to do this, we have to show that the difference $\bar{x} - \mu_0$ is either a large positive number or a small negative number. If this is the case, then we reject H_0. This is called a *two-tailed test*.

a) If $\bar{x} > \mu_0$, then the p-value of the two-tailed test is:

$$p - \text{value} = 2P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

b) If $\bar{x} < \mu_0$, then the p-value of the two-tailed test is:

$$p - \text{value} = 2P\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \leq \frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right)$$

Example 3. Measurements of blood viscosity were made on laboratory mice. A normal value should be close to 3.95. Researchers which are testing a new drug suspect that this could have modified their blood viscosity level. Levels which are either too small or too large are not acceptable. A sample of 9 mice yielded $\bar{x} = 4.25$ and $s = 0.6$. Is there enough evidence that the average level of viscosity is different than 3.95? Use the level $\alpha = 0.05$.

Let μ be the average viscosity level. We are interested in testing:

$$H_0 : \mu = 3.95 \quad H_1 : \mu \neq 3.95$$

The test statistic is

$$T_0 = \frac{\bar{X} - 3.95}{S/3}$$

The observed value of the test statistic is

$$t_0 = \frac{\bar{x} - 3.95}{s/3} = \frac{4.25 - 3.95}{0.6/3} = 1.5$$

Hence, we can conclude that

$$p\text{−value} = 2P(T \geq 1.5) \text{ is between } 0.10 \text{ and } 0.20.$$

where T has a T distribution with 8 degrees of freedom. (From Table 17.4 we see that $P(T \geq 1.397) = 0.10$ and $P(T \geq 1.860) = 0.05$.) Since p-value> 0.05, we cannot reject H_0. There is not enough evidence that μ is different that 3.95.

11.3 Hypothesis Testing for the Proportion

In this section we will test some hypotheses on the proportion p. The test statistic to be used in each of the three cases is:

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

which can be approximated by a standard normal random variable Z. The smaller the p-value, the less likely it becomes that H_0 is true. We reject H_0 (and we gain evidence for H_1), if the p-value is smaller than a level α.

$$\textbf{Case I}: H_0 : p = p_0, \quad H_1 : p > p_0, \quad p-\text{value} = P\left(Z \geq \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right)$$

$$\textbf{Case II}: H_0 : p = p_0, \quad H_1 : p < p_0, \quad p-\text{value} = P\left(Z \leq \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}\right)$$

$$\textbf{Case III}: H_0 : p = p_0, \quad H_1 : p \neq p_0, \quad p-\text{value} = \begin{cases} 2P\left(Z \geq \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right) & \text{if } \hat{p} \geq p_0 \\ 2P\left(Z \leq \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right) & \text{if } \hat{p} \leq p_0 \end{cases}$$

Example 4. A manufacturer claims that a new beetle trap attracts and kills more than 90% of the beetles that come close to it. We would like to test this claim. A sample of 900 beetles is available from which 825 were attracted to the trap. Is there enough evidence that supports the claim? (Use the significance level $\alpha = 0.05$.)

We want to test the following hypothesis:

$$H_0 : p = 0.9, \quad H_1 : p > 0.9$$

The point estimate of the percentage p is $\hat{p} = 825/900 = 0.917$. The value of the test statistic for this sample is

$$z_0 = \frac{\hat{p} - 0.9}{\sqrt{(0.9)(0.1)/900}} = \frac{0.917 - 0.9}{\sqrt{(0.9)(0.1)/900}} = 1.7$$

Using Table 17.3, we have:

$$p - \text{value} = P(Z \geq 1.7) = 0.0446$$

Since the p-value is smaller than 0.05, we reject H_0. There is evidence that the percentage p is higher than 0.9.