MAT 2377 (Spring 2011)
Hypothesis Testing - Sections 9-1 to 9-3 and Section 9-5

Hypothesis testing is a procedure that leads us to decide if experimental data supports a hypothesis concerning population(s) parameter(s). We will consider hypotheses concerning a population mean μ or a population proportion p.

Stating the Hypotheses: Often the researcher would to verify a change in the unknown parameter under new experimental conditions. For example, a manufacturer of a new fiberglass tire claims that the mean life of the new tires are greater than the mean life of tires using the old manufacturing process. The previous mean life was $65,000$ km.

Let μ denote the mean life of the new tires. The no change hypothesis (that we will call the **null hypothesis**) is $H_0 : \mu = 65,000$ and the claim or research hypothesis (that we will call the **alternative hypothesis**) is $H_1 : \mu > 65,000$.

We want to test

$$H_0 : \ \mu = 65,000 \quad \text{against} \quad H_1 : \mu > 65,000.$$

Now we consider an example involving a proportion. Suppose that we would like to test the hypothesis that the proportion of defective items produced at a particular plant is $p = 2\%$. Then, we would test

$$H_0 : \ p = 0.02 \quad \text{against} \quad H_1 : p \neq 0.02.$$

Null Hypothesis: The null hypothesis will always be a simple statement concerning the unknown parameter θ. That is, it is a statement of the form $\theta = \theta_0$, where θ_0 is some real number. For example, $H_0 : \mu = 65,000$ or $H_0 : p = 0.02$. The value of the parameter in the null hypothesis will be the boundary value of the parameter from the alternative hypothesis.

Alternative Hypothesis: The alternative hypothesis will be a composite statement concerning θ. It is often the research hypothesis, i.e. the hypothesis that we would like to support with the data. We will consider three types of alternatives: (θ is the unknown parameter and θ_0 is some real number)

$$H_1 : \theta < \theta_0 \quad \text{is a left-sided alternative;}$$

$$H_1 : \theta > \theta_0 \quad \text{is a right-sided alternative;}$$

$$H_1 : \theta \neq \theta_0 \quad \text{is a two-sided alternative.}$$

Collecting Evidence: We select a random sample of n observations and compute a point estimate for the unknown parameter θ.

Example 1 : [Tire Example] We collect a random sample of $n = 45$ of the new fiberglass tires and observe an lifetime of $65,158.7$ km. This is a point estimate for the true mean lifetime of such tires. Note that this evidence is in favour of the alternative hypothesis that $\mu > 65,000$. However we should not yet state that the data supports H_1.

Suppose that we decide to say that the data support $H_1 : \mu > 65,000$ if $\overline{x} > 65,000$. Now suppose that $H_0 : \mu = 65,000$ is true, what are our chances that we will say that the data support H_1. Well

$$P(\overline{X} > 65,000) \approx 1 - \Phi(0) = .5$$

So there is a 50% chance that we say that the data support H_1 when in fact H_0 is true. We need to come up with a way to properly analyze the evidence.

Definitions:

- A **test statistic** is a statistic that is used to test hypotheses.
- The **critical region** of the test statistic is a set of possible values of the test statistic such that if the observed of the test statistic falls in the critical region we will reject H_0 and accept H_1.
- If we reject H_0 when H_0 is true, we say that we have committed an error of **type I** and

$$\alpha = P(\text{type I error}) = P(\text{ reject } H_0 \text{ when } H_0 \text{ is true})$$

- If the observed value of the test statistic does not fall in the critical region, then we fail to reject H_0. If we fail to reject H_0 when H_0 is false, then we say that we have committed an error of **type II** and

$$\begin{aligned} \beta(\theta_1) &= P(\text{type II error}) \\ &= P(\text{ fail to reject } H_0 \text{ when } \theta = \theta_1 \in H_1) \end{aligned}$$

Example 2: [**Tire Example Continued**] Suppose that the population standard deviation is $\sigma = 1000$ km. We will use $\overline{X}$ as a test statistic. Recall that we want to test

$$H_0 : \mu > 65,000 \quad \text{and} \quad H_1 : \mu > 65,000.$$

The sample size is $n = 45$.

For the following critical region: $\overline{x} > 65,400$.
(i) Compute the probability of committing an error of type I.
(ii) If the true mean life is $\mu = 66,000$, then compute the probability of committing an error of type II.

Remarks:

- When working with a critical region, often in practice, we fix the probability of making an error of type I, i.e. α. We then call, α the **level of significance** of the test. Furthermore, α should be small, but not too small, some common values are $\alpha = 10\%$, $\alpha = 5\%$ or $\alpha = 1\%$.
- We control the chances of making an error of type II with the sample size. We will discuss this later.

Test Statistic for a hypothesis concerning μ: It will be easier to construct a critical region when using a standardized test statistic. Depending on the experimental conditions we will use one of three test statistics.

(i) **Conditions:**

- the population is normal or $n \geq 30$;
- σ is known.

We will use the following test statistic:

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

which follows a $N(0,1)$ distribution if $\mu = \mu_0$.

(ii) **Conditions:**

- $n \geq 40$;
- σ is unknown.

We will use the following statistic

$$Z_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

which follows a $N(0,1)$ distribution if $\mu = \mu_0$.

(ii) **Conditions:**

- the population is normal;
- σ is unknown.

We will use the following statistic

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

which follows a t distribution with $\nu = n - 1$ degrees of freedom, if $\mu = \mu_0$.

Critical Region: Suppose that the test statistic is Z_0 which follows a standard normal distribution under H_0. Let z_0 be the observed value of Z_0.

(i) [**Right-Sided Alternative**] $H_1 : \mu > \mu_0$**.** The critical region is

$$z_0 > z_\alpha.$$

(ii) [**Left-Sided Alternative**] $H_1 : \mu < \mu_0$**.** The critical region is

$$z_0 < -z_\alpha.$$

(iii) [**Two-Sided Alternative**] $H_1 : \mu \neq \mu_0$**.** The critical region is

$$z_0 < -z_{\alpha/2} \quad \text{or} \quad z_0 > z_{\alpha/2}.$$

Critical Region: Suppose that the test statistic is T_0 which follows a t distribution with $\nu = n - 1$ degrees of freedom under H_0. Let t_0 be the observed value of T_0.

(i) [**Right-Sided Alternative**] $H_1 : \mu > \mu_0$**.** The critical region is

$$t_0 > t_{\alpha,n-1}.$$

(ii) [**Left-Sided Alternative**] $H_1 : \mu < \mu_0$**.** The critical region is

$$t_0 < -t_{\alpha,n-1}.$$

(iii) [**Two-Sided Alternative**] $H_1 : \mu \neq \mu_0$**.** The critical region is

$$t_0 < -t_{\alpha/2,n-1} \quad \text{or} \quad t_0 > t_{\alpha/2,n-1}.$$

Example 3: We would like to test the claim that the mean life of the new fiberglass tires is greater than $65,000$ km at a level of significance of 5%. A sample of $n = 50$ tires (measured in 1000 km) yielded a

$$\sum_{i=1}^{50} x_i = 2,859.3 \quad \text{and} \quad \sum_{i=1}^{50} x_i^2 = 163,910.1$$

What are the conclusions of the test?

Example 4: A company manufactures 6-meter tubes. We randomly selected $n = 10$ tubes and computed $\overline{x} = 5.7$m and $s = 0.2$m. Can we conclude that the mean population length is not 6m at a level of significance of 5%? Assume that the population is normally distributed.

P-value Method: The modern approach to hypothesis testing is to use a p-value instead of a critical region. A p-value is the probability of observing a value of the test statistic as or more extreme as the observed value in favour of H_1 assuming that H_0 is true.

We compute the p-value as follows.

If the test statistic is Z_0 which follows a $N(0,1)$ when H_0 is true. Then the p-value of the test is

$$P = \begin{cases} 2\,[1-\Phi(|z_0|)], & \text{for a two-sided alternative} \\ 1-\Phi(z_0), & \text{for a right-sided alternative} \\ \Phi(z_0), & \text{for a left-sided alternative} \end{cases}$$

If the test statistic is T_0 which follows a t distribution with $\nu = n-1$ degrees of freedom, when H_0 is true. Let T be a t random variable with $n-1$ degrees of freedom, then the p-value of the test is

$$P = \begin{cases} 2\,P(T > |t_0|), & \text{for a two-sided alternative} \\ P(T > t_0), & \text{for a right-sided alternative} \\ P(T < t_0), & \text{for a left-sided alternative} \end{cases}$$

Decision Rule: $P < \alpha$ is equivalent to the observed value falling in the critical region. Hence, we will reject H_0, if $P < \alpha$.

Example 4: Answer Question 3 using the p-value method.

Example 5:
(a) Answer Question 4 using the p-value method.
(b) Refer to Question 4, but suppose that we would like to test

$$H_0: \ \mu = 6 \quad \text{against} \quad H_1: \ \mu < 6.$$

Does the data support the alternative hypothesis at $\alpha = 5\%$?

Type II error and Choice of Sample Size: We will only discuss type II error and the choice of sample size for a normal population with σ known.

Suppose that we would like to find the probability of making a type II error when we assume that the true mean is μ_1. Define

$$\delta = \mu_1 - \mu_0,$$

when testing the null hypothesis $H_0: \ \mu = \mu_0$. The test statistic

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} + \frac{\delta}{\sigma/\sqrt{n}}.$$

If the population mean is $\mu = \mu_1$, then $\overline{X}$ follows a $N(\mu_1, \sigma^2/n)$ distribution and thus

$$Z_0 \quad \text{follows a } N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right) \quad \text{distribution.}$$

Example 6: Refer to Example 4. Suppose that we are testing $H_0 : \mu = 6$ against $H_1 : \mu \neq 6$ at $\alpha = 5\%$. The population is normal. We will suppose that we know the population standard deviation $\sigma = 0.2$. If the sample size is $n = 10$, what is the probability of committing an error of type II when the true mean is 6.5.

Sample Size: To control the probability of error of type II, that is β, at $\mu_1 = \mu_0 + \delta$, we can choose an appropriate sample size.

For a two-sided alternative: We require the following sample size

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \, \sigma^2}{\delta^2}.$$

For a one-sided alternative: We require the following sample size

$$n = \frac{(z_\alpha + z_\beta)^2 \, \sigma^2}{\delta^2}.$$

Example 7: Suppose that we are testing $H_0 : \mu = 6$ against $H_1 : \mu \neq 6$ at $\alpha = 5\%$. The population is normal with $\sigma = 0.2$. If we want to design the experiment such that if $\mu = 6.1$, then the probability of failing to reject H_0 is 5%. Determine the required sample size.

Hypothesis testing concerning a proportion p

Suppose that we want to test the null hypothesis $H_0 : p = p_0$, where p is an unknown population proportion. We will use a test statistic based upon the sample proportion $\widehat{P}$. The test statistic is

$$Z_0 = \frac{\widehat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}}.$$

It follows approximately a $N(0, 1)$ distribution when

$$n\, p_0 \geq 5 \quad \text{and} \quad n\,(1 - p_0) \geq 5.$$

Critical Region: Let z_0 be the observed value of Z_0.

(i) [**Right-Sided Alternative**] $H_1 : p > p_0$**.** The critical region is

$$z_0 > z_\alpha.$$

(ii) [**Left-Sided Alternative**] $H_1 : p < p_0$**.** The critical region is

$$z_0 < -z_\alpha.$$

(iii) [**Two-Sided Alternative**] $H_1 : p \neq p_0$**.** The critical region is

$$z_0 < -z_{\alpha/2} \quad \text{or} \quad z_0 > z_{\alpha/2}.$$

Example 8: Suppose that we would like to test the hypothesis that the proportion of defective items produced at a particular plant is $p = 2\%$. From $n = 500$ random selected items there are 8 which are defective. Do the data suggest that $p \neq .02$ at $\alpha = 5\%$?