Normal Distribution, Known σ, Any Sample Size

- Under these assumptions, a $100(1-\alpha)\%$ CI of sample mean μ is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

General Distribution, Unknown σ, Large Sample

- Under these assumptions, an approximate $100(1-\alpha)\%$ CI of sample mean μ is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right)$$

- Notice a special case if the distribution is Bernoulli, we have a more accurate but very complicated formula.

Normal Distribution, Unknown σ, Any Sample Size

- Under these assumptions, a $100(1-\alpha)\%$ CI of sample mean μ is given by
$$\left(\bar{x} - t_{\alpha/2,n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right)$$
- Here we utilize the *t* distribution.

Hypothesis Testing

- A *statistical hypothesis*, or just *hypothesis*, is a claim or assertion either about the value of a single parameter (population characteristic or characteristic of a probability distribution), about the values of several parameters, or about the form of an entire probability distribution.

- A testing problem usually contains two hypotheses: the null hypothesis, denoted by H_0, is the claim that is initially assumed to be true (the "prior belief" claim). The alternative hypothesis, denoted by H_a, is the assertion that is contradictory to H_0.

- The null hypothesis will be rejected in favor of the alternative only if sample evidence suggests that H_0 is false. If the sample does not strongly contradict H_0, we will continue to believe in the truth of the null hypothesis. The two possible conclusions from a testing analysis are then *reject* H_0 or *fail to reject* H_0.

Examples

Ex. A factory claims that less than 10% of the components they produce are defective. A consumer group is skeptical of the claim and checks a random sample of 300 components and finds that 39 are defective. Is there evidence that more than 10% of all components made at the factory are defective?

$H_0: p \leq 0.10 \qquad H_a: p > 0.10$

Ex. We are interested in height of all Columbia students. In a sample of 12 students, the sample mean is 66.30 inches, and the sample s.d. is 4.35 inches. Should we reject the null hypothesis $H_0: \mu = 68$ vs $H_a: \mu \neq 68$?

Remarks

- In our treatment of hypothesis testing, H_0 will always be stated as an equality claim. If θ denotes the parameter of interest, the null hypothesis will have the form H_0: $\theta = \theta_0$.

- The alternative to the null hypothesis H_0: $\theta = \theta_0$ will usually look like one of the following three forms:

1. H_a: $\theta > \theta_0$ (in which case the implicit null hypothesis is $\theta \leq \theta_0$).
2. H_a: $\theta < \theta_0$ (in which case the implicit null hypothesis is $\theta \geq \theta_0$).
3. H_a: $\theta \neq \theta_0$.
4. H_a: $\theta = \theta_1 \neq \theta_0$ (simple alternative).

- The value θ_0 separates the alternative from the null and is called the null value. The null and alternative are not treated equivalently, once a statement is in the null hypothesis, we will not easily reject it unless we have enough evidence.

Motivating example

Ex. Suppose we have a biased coin, we believe that it has probability 95% of having a head in a flip. Alternatively, it could also have probability 5% of having a head. Can you design a simple test to see if the coin has probability 95% of having heads?

Simple alternative: $H_0 : p = 0.95$ $H_a : p = 0.05$

Test Procedures

A test procedure is specified by the following:

1. Find a test statistic, a function of the sample data on which the decision (reject H_0 or do not reject H_0).
2. Construct a rejection region, the set of all test statistic values for which H_0 will be rejected.

The null hypothesis will then be rejected if and only if the observed or computed test Statistic value falls in the rejection region.

Can you construct a test procedure for the previous example?

Example cont.

Ex. (Biased coin cont.) In order to test if $p = 0.95$ we decide to conduct one experiment. We are going to flip this biased coin once, if it comes out a head, we will accept the null hypothesis, if it comes out a tail, we will reject the null hypothesis.

Test statistic: X = outcome of the first flip (Bernoulli rv.)

Rejection region: $\{X: X = 0\}$

Any other test statistics?

What are the odds that we'll make a mistake in our decision?

Two types of errors

- Definition

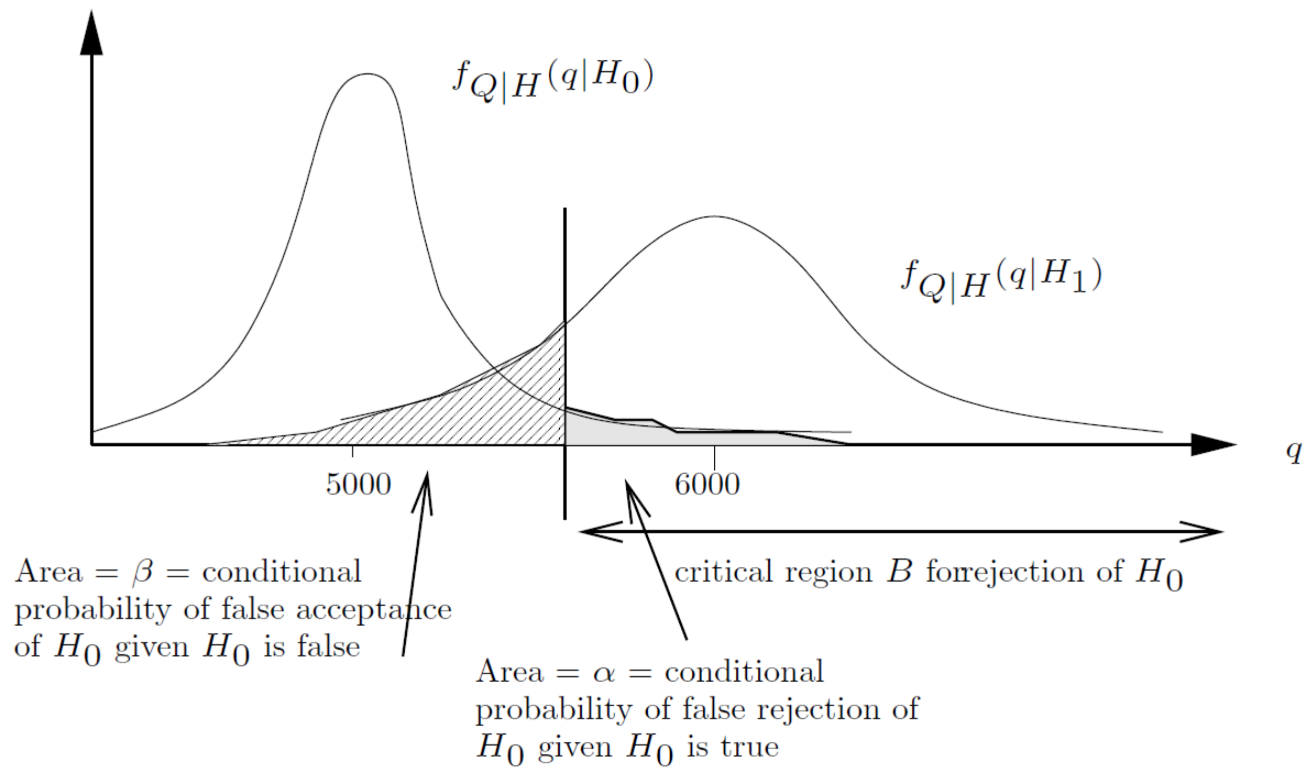
A type I error α consists of rejecting the null hypothesis H_0 when it is true.

A type II error β involves not rejecting H_0 when H_0 is false.

	Decide to accept	Decide to reject
Null is true	Right	Type I
Alternative is true	Type II	Right

Errors

- Choice of *α* is subjective. As move threshold to left, increase *α* and decrease *β*.



Example cont.

Ex. (Biased coin cont.) In order to test if $p = 0.95$ we decide to conduct one experiment. We are going to flip this biased coin once, if it comes out a head, we will accept the null hypothesis, if it comes out a tail, we will reject the null hypothesis. What are the two types of errors associated with this test procedure?

Criteria

- A good test will be aimed to make two types of errors, both α and β, as small as possible.

- Unfortunately, there is no rejection region that will simultaneously make both α and β small once the test statistic and sample size are fixed. Thus, a region must be chosen to effect a compromise between α and β.

- Because of the suggested guidelines for specifying and . A type I error is usually more serious than a type II error (we don't want to reject the null easily).

- In practice, people specify to the largest value that α can be tolerated and find a rejection region having that value of α. The resulting value of α is often referred to as the significance level of the test (0.1, 0.05, 0.01). The corresponding test procedure is called an α level test. The previous example was an exact 0.05-level test.