**Lecture 13: Tests of Hypotheses**

**[Ref #2]**

* Practical necessity
* Practical example of reasoning under uncertainty

Review of basic terminology

*Population* / *universe* assumed to be infinite

*Sampling*  see previous lecture

*Parameter*  pertains to the population, as we have seen  or 

*Statistic* obtained from a sample, as Xmean

*Estimation* of the value of a parameter from one or more samples

Test of a hypothesis

Essentially, we would like to verify whether a given hypothesis, say H, has an “acceptable probability” of being true.

We will verify this based on a *parameter* / *statistic*, say . In the examples we consider,  will be the population or sample mean [ or Xmean], but the standard deviation Xvar will play an important role.

Let θ0 represent a population parameter and let θ represent the corresponding sample statistic.

Often the hypothesis to be tested is of the general form: “There is no *significant* difference between 0 and ” ... that is, between  and Xmean.

Such a hypothesis is called the *null hypothesis*, H0. The logically complementary hypothesis is called the *alternative hypothesis*, H1.

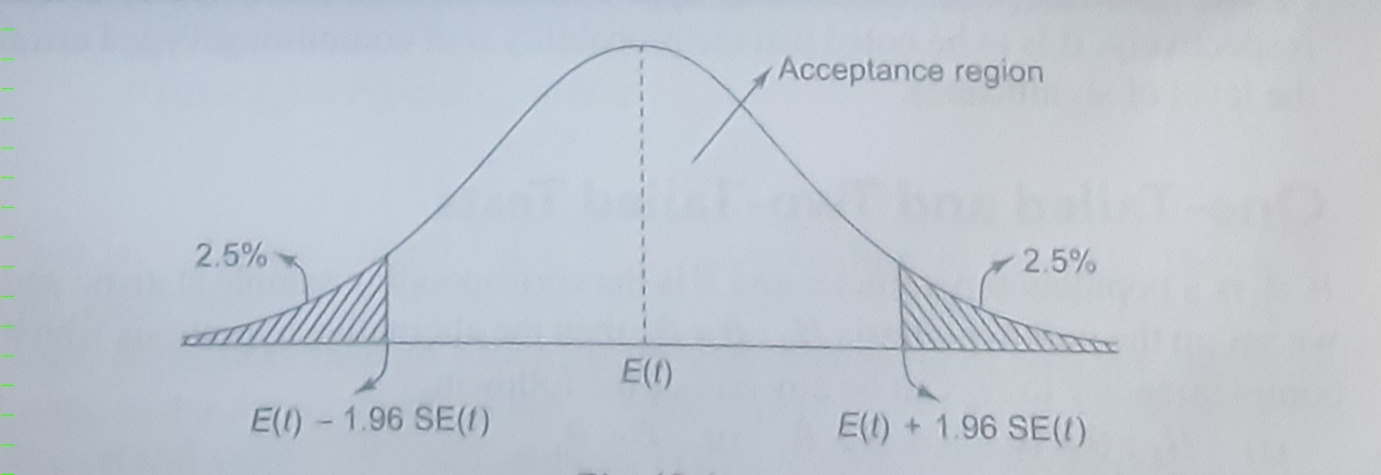
*Tests of hypotheses* involve decisions of whether to accept H0, or to reject H0 and accept H1.

*Tests of significance* involve decisions of whether the difference between 0 and  is significant.

Recall CLT: Exp[ Xmean ] = 

We need a test statistic which is, in the limit, N(0,1), standard normal.

Test statistic Z = [ Xmean –  ] / SQRT(Xvar)



In the diagram:

*t* stands for Xmean in our notation,

SE(*t*), *standard error* in *t*, stands for sample standard deviation = SQRT(Xvar)

The offsets 1.96\*SE(*t*) are chosen to yield area of *acceptance region* = 0.95.

Therefore area of the shaded *rejection* or *critical region* = 0.05.

We can easily verify that, if F(z) denotes the cdf of N(0,1), then:

F(1.96) – [1-F(1.96)] = 2\*F(1.96) – 1 = 0.95

***Level of significance*** (*LOS*)= area of rejection region, in percent. Thus, in the above diagram, the level of significance is 5%. The interval from E(*t*)-1.96\*SE(*t*) to E(*t*)+1.96\*SE(*t*) is here known as the ***95% confidence interval*** for E(*t*).

**Equivalently, LOS = maximum probability of false rejection.**

One of two types of error can occur while testing a hypothesis.

* **Type I error**: rejecting H0 when it is true; false rejection; “producer’s risk".
* **Type II error**: accepting H0 when it is false; false acceptance; “consumer’s risk".

H0 represents:  = 0. In theory, the alternative hypothesis H1 can be of one of THREE forms:

1. H1 :  <> 0
2. H1 :  > 0
3. H1 :  < 0

Why is this so?

Because *specifics* of the problem vary, depending on the application. The three types of tests are called, respectively, *two-tailed*, *right-tailed* and *left-tailed*.

Based on N(0,1), for various values of LOS, denoted by a, we have:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | LOS 1% | LOS 2% | LOS 5% | LOS 10% |
| Two-tailed test | **|*z*a| = 2.58** | **|*z*a| = 2.33** | **|*z*a| = 1.96** | **|*z*a| = 1.645** |
| Right-tailed test | ***z*a = 2.33** | ***z*a = 2.055** | ***z*a = 1.645** | ***z*a = 1.28** |
| Left-tailed test | ***z*a = -2.33** | ***z*a = -2.055** | ***z*a = -1.645** | ***z*a = -1.28** |

By the way, it should be noted that the standard normal distribution N(0,1) does not reach value ZERO except in the limit as *z* 🡪 + infinity.

However, at *z* = +6, the value of N(0,1) is ~ 6.076\*10-9.

**Bernoulli trials**

Let X represent the number of “successes" in *n* Bernoulli trials. By CLT, for large *n*, X follows the distribution N(*np*, sqrt(*npq*)), where *q* = 1-*p*. That is, for large *n*, binomial distribution may be approximated by normal distribution.

Therefore, the *proportion* of successes X/*n* follows N(*p*,sqrt(*pq*/*n*)).

**Example #1**

It is hypothesized that 20% of a manufactured product is of top quality. On one day, it is found that 50 units out of a production volume of 400 units are of top quality. With 5% level of significance, show that either the hypothesis is wrong, or the day’s production is not a representative sample.



[Meaning of ***sampling error***, ***systemic error***]

Here  represents the proportion of successes, with 0 = 0.2, and test statistic  = 0.125. Sample size *n* = 400.

We saw that, with *LOS* = 5%, for two-sided test, critical value *z*a = 1.96.

Standardized test statistic:

*z* = [ - 0]/sqrt(*pq*/*n*) = -0.075/ sqrt(*0.2x0.8*/400) = -3.75

Since *z* < -*z*a, we reject H0, or the assumption that the day's production was a representative sample; both cannot simultaneously be true.

OR we look for a systemic error!

Part 2

Re-do the previous part with n = 100, showing up 12 items of top quality. For given , 0, *p and q*, note that z is proportional to sqrt(*n*).

Part 3

Based on the above sample data, find the 95% confidence interval for the percentage P of top quality product. [Note: Now we must take P as being unknown.]

By definition, the required limits are defined by:

|p - P|/sqrt(pq/n) < 1.96

with p = 1/8, q = 7/8, n = 400.

Substituting these values and simplifying, we get:

0.093 < P < 0.157

Therefore, in percentage terms, the required 95% confidence interval is:

9.3% < P < 15.7%

**Example #2**

The average fatality rate for a specific category of hospitalized patients is known to be 17%. In a given hospital, out of 640 patients of this category, 63 died. At LOS = 1%, can this hospital be considered significantly better than average, for this category of patients?

Here P = 0.17, p = 63/640, n = 640.

Q = 1-P = 0.83.

H0 : p = P, the hospital is as good or bad as the average.

H1 : p < P, the hospital is significantly better than the average.

Left-tailed test must be used, with z = -2.33 (see table above).

Calculate z = [p - P]/sqrt(PQ/n) ..... = -4.96

Since |z| > 2.33, or z is to the left of z = -2.33, we reject H0 and accept H1. The hospital can justifiably claim to be better than average in this particular aspect.

**Example #3**

A salesman claims that *at most* 60% of shoppers at a store leave without making any purchase. A random sample 0f 50 shoppers shows that 35 of the shoppers left the store without making any purchase. At LOS = 0.05, are the sample results consistent with the salesman’s claim?

Here p = 0.7, P = 0.6, n = 50.

H0 : the sample result p is consistent with claim P.

H1 : the sample result is significantly worse, or p > P.

Right-tailed test is needed.

Calculate z = [p - P]/sqrt(PQ/n) ..... = 1.443

z = 1.645 (see table)

Since z < z, we conclude that there is no *significant difference* between the sample data and the salesman’s claim, at 5% LOS.

Part 2: Re-do previous part with n = 200, same p.