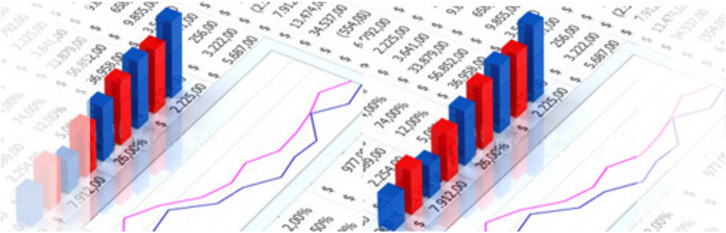
MTMS.01.099 Mathematical Statistics

Lecture 12. Hypothesis testing. Power function. Approximation of Normal distribution and application to Binomial distribution

Tõnu Kollo



Fall 2016

Hypothesis Testing

- A statistical hypothesis, or simply a **hypothesis**, is an assumption about a population parameter or population distribution.
- **Hypothesis testing** is the procedure whereby we decide to "reject" or "fail to reject" a hypothesis.
- **Null hypothesis** H_0: This is the hypothesis (assumption) under investigation or the statement being tested. The null hypothesis is a statement that "there is no effect", "there is no difference", or "there is no change". The possible outcomes in testing a null hypothesis are "reject" or "fail to reject".
- **Alternate hypothesis** H_1: This is a statement you will adopt if there is strong evidence (sample data) against the null hypothesis. A statistical test is designed to assess the strength of the evidence (data) against the null hypothesis.

Hypothesis Testing (2)

- **Fail to Reject** H_0: We never say we "accept or prove H_0" - we can only say we "fail to reject" it or "stay with H_0". Failing to reject H_0 means there is NOT enough evidence in the data and in the test to justify rejecting H_0. So, we retain the H_0 knowing we have not proven it true beyond all doubt.
- **Rejecting** H_0: This means there IS significant evidence in the data and in the test to justify rejecting H_0. When H_0 is rejected we adopt H_1. When we adopt H_1 we know that we can be occasionally wrong, the same time we say that H_1 is proved.

Hypothesis Testing (3)

- **Type I error** occurs when we reject a true null hypothesis H_0. We denote by α the probability of making Type I error.
- **Type II error** occurs when we "fail to reject" a false null hypothesis H_0. We denote by β the probability of making Type II error.
  For a given sample size reducing the probability of a type I error increases the probability of a type II error, and vice versa.
- **The level of significance** α is the probability we are willing to risk rejecting H_0 when it is true. Typically $\alpha = 0.05$ is used.

Types of error

The possible outcomes are:

	H_0 is true	H_1 is true
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Test statistic, Critical region, Test of significance

Definition

A **test statistic** is a numerical function of the data whose value determines the result of the test. The function itself is generally denoted $T = T(\boldsymbol{X})$, where $\boldsymbol{X}$ is theoretical sample, i.e. represents the data.

After being evaluated for the sample data $\boldsymbol{x}$, the result is called an observed test statistic and is written in lowercase, $t = T(\boldsymbol{x})$.

Definition

The set of all test statistic values for which H_0 will be rejected, is said to be **critical region**, denoted by C.

Definition

Test of significance is the control rule, which leads to rejection of H_0 if the test statistic value falls in the critical region C:

Example 1. The drying time of a paint under specified test conditions is known to be rv $X \sim N(75, 9^2)$. Chemists have proposed a new additive designed to decrease average drying time. It is believed that drying times with this additive will remain normally distributed with standard deviation 9. Let μ denote the true average drying time when the additive is used. The appropriate hypotheses are :

$$H_0 : \mu \geq 75,$$
$$H_1 : \mu < 75.$$

Experimental data is to consist of drying times from $n = 25$ test specimens. A reasonable rejection region has the form $\overline{x} < c$, where the cut-off value c is suitably chosen. Consider the choice $c = 70.8$, so that the test procedure consists of test statistic $\overline{X}$ and rejection region $\overline{x} < 70.8$. Based on data $n = 25$ the following decision is made: (test 1) if $\overline{x} < 70.8$, then reject H_0. Find both Type I and II error of this kind of test.

Example 1

Example 1 continues. Calculation of α and β now involves a routine standardization of $\overline{X}$ followed by reference to the standard normal probabilities:

$$\begin{aligned} \alpha &= P(\text{Type I error}) = P(H_0 \text{ is rejected when it is true}) \\ &= P(\overline{X} < 70.8 \text{ when } \overline{X} \sim N(75, 1.8^2)) \\ &= \Phi\left(\frac{70.8 - 75}{1.8}\right) = \Phi(-2.33) = 0.01 \end{aligned}$$

Example 1

Example 1 continues. Calculation of α and β now involves a routine standardization of $\overline{X}$ followed by reference to the standard normal probabilities:

$$\begin{aligned}
\alpha &= P(\text{Type I error}) = P(H_0 \text{ is rejected when it is true}) \\
&= P(\overline{X} < 70.8 \text{ when } \overline{X} \sim N(75, 1.8^2)) \\
&= \Phi\left(\frac{70.8 - 75}{1.8}\right) = \Phi(-2.33) = 0.01
\end{aligned}$$

$$\begin{aligned}
\beta(72) &= P(\text{Type II error when } \mu = 72) \\
&= P(H_0 \text{ is not rejected when it is false because } \mu = 72) \\
&= P(\overline{X} \geq 70.8 \text{ when } \overline{X} \sim N(72, 1.8^2)) \\
&= 1 - \Phi\left(\frac{70.8 - 72}{1.8}\right) = 1 - \Phi(-0.67) = 1 - 0.2514 = 0.74
\end{aligned}$$

Example 1

Example 1 continues. Calculation of α and β now involves a routine standardization of $\overline{X}$ followed by reference to the standard normal probabilities:

$$\begin{aligned}
\alpha &= P(\text{Type I error}) = P(H_0 \text{ is rejected when it is true}) \\
&= P(\overline{X} < 70.8 \text{ when } \overline{X} \sim N(75, 1.8^2)) \\
&= \Phi\left(\frac{70.8 - 75}{1.8}\right) = \Phi(-2.33) = 0.01
\end{aligned}$$

$$\begin{aligned}
\beta(72) &= P(\text{Type II error when } \mu = 72) \\
&= P(H_0 \text{ is not rejected when it is false because } \mu = 72) \\
&= P(\overline{X} \geq 70.8 \text{ when } \overline{X} \sim N(72, 1.8^2)) \\
&= 1 - \Phi\left(\frac{70.8 - 72}{1.8}\right) = 1 - \Phi(-0.67) = 1 - 0.2514 = 0.74
\end{aligned}$$

$$\beta(70) = 1 - \Phi\left(\frac{70.8 - 70}{1.8}\right) = 0.33, \ \beta(67) = 0.0174$$

Example 1

Example 1 continues. The use of cut-off value $c \leq 70.8$ resulted in a very small value of α (0.01) but rather large β's. Consider the same experiment and test statistic $\overline{X}$ with the new rejection region $\overline{x} < 72$; (test 2) if $\overline{x} < 72$ then reject H_0. What are the Type I and II error probabilities with this kind of test? Which test to prefer?

- You can construct many tests for testing some particular hypotheses.
- A test is good, if its Type I error is small and if H_0 is rejected with large probability in case H_1 is true.
- It is also necessary to control the Type II error.
- For comparing tests the **power function** is used.

The Power function

For the moment we suppose that the null hypothesis consists of one value, which we call θ_0.

Definition

The **power function** $h(\theta)$ of a test is the probability of rejecting H_0 when the value θ is the correct value of the parameter in the parameter space:

$$h(\theta) = P(H_0 \text{ is rejected if } \theta \text{ is the correct value of the parameter})$$

or just

$$h(\theta) = P(T(\boldsymbol{X}) \in C|\theta).$$

So, if $\beta(\theta) = P(\text{fail to reject } H_0|\theta)$, then

$$h(\theta) = 1 - \beta(\theta).$$

A test is good if: $h(\theta)$ is large for all $\theta \in H_1$ and $h(\theta)$ is small for $\theta \in H_0$.

The Power function

The properties of the power function:

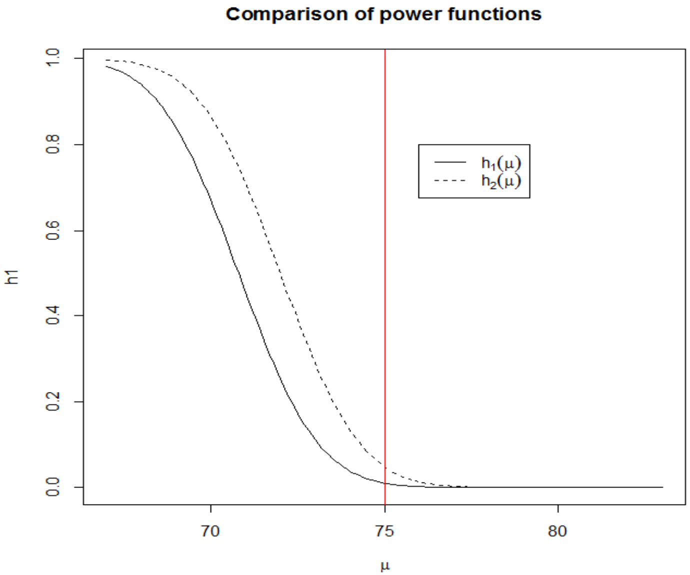
- $0 \leq h(\theta) \leq 1$,
- The level of significance, e.g. the probability of making a mistake when rejecting H_0 is expressed from the power function:

  $$sup_{\theta \in H_0} h(\theta) = \alpha,$$

  if H_0 is a simple hypothesis, then

  $$h(\theta_0) = \alpha.$$

Example 1. Comparison of Power functions



Relationship of α and β

Proposition

Suppose an experiment and a sample size are fixed and a test statistic is chosen. Then decreasing the size of the rejection region to obtain a smaller value of α results in a larger value of β for any particular parameter value consistent with H_1.

This proposition says that **once the test statistic and *n* are fixed, there is no rejection region that will simultaneously make both α and all β's small**. A region must be chosen to effect a compromise between α and β.

The Power function

If you still feel confused, check this online course page, where you can find some examples of power function (with some nice illustrations ☺)

Find the page here: Click on me! ☺

Hypotheses about the mean of Normal distribution

In Example 1 we stated the hypotheses about the mean of Normal distribution. The General Algorithm for testing hypotheses about **mean**:

- We have a random sample $x_1, \ldots, x_n$.
- Consider point estimate $\hat{\mu} = \overline{x}$.
- Corresponding test statistic is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\ X_i \sim N(\mu, \sigma^2),\ independent$$

- In case H_0 is true,

$$X_i \sim N(\mu_0, \sigma^2),\ \overline{X} \sim N(\mu_0, \frac{\sigma^2}{n}).$$

Hypotheses about the mean of Normal distribution

- Form test statistics U and V, which are in case H_0 is true distributed as follows:

$$U = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1), \text{if } \sigma \text{ is known.}$$

$$V = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \sim t(f), f = n-1, \text{if } \sigma \text{ is unknown.}$$

- Control rule (test) depends on: hypotheses being tested and available information about σ.

1) σ known	2) σ unknown
$H_1 : \mu \neq \mu_0$ test: $\lvert u\rvert > \lambda_{\frac{\alpha}{2}} \xrightarrow{+}$ reject H_0	$H_1 : \mu \neq \mu_0$ test: $\lvert v\rvert > t_{\frac{\alpha}{2}}(f) \xrightarrow{+}$ reject H_0
$H_1 : \mu < \mu_0$ test: $u < -\lambda_\alpha \xrightarrow{+}$ reject H_0	$H_1 : \mu < \mu_0$ test: $v < -t_\alpha(f) \xrightarrow{+}$ reject H_0
$H_1 : \mu > \mu_0$ test: $u > \lambda_\alpha \xrightarrow{+}$ reject H_0	$H_1 : \mu > \mu_0$ test: $v > t_\alpha(f) \xrightarrow{+}$ reject H_0

Hypotheses about the difference of two Population mean (independent samples)

Let us have two independent samples $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$, with corresponding distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and the following hypotheses:

$$H_0 : \mu_1 = \mu_2 \Leftrightarrow \mu_1 - \mu_2 = 0,$$

$$H_1 : \mu_1 \neq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \neq 0.$$

Algorithm for testing hypotheses:

- Form an estimate for the difference of parameters, $\hat{\mu}_1 - \hat{\mu}_2 = \overline{x} - \overline{y}$.
- Corresponding test statistic

$$T(\boldsymbol{X} - \boldsymbol{Y}) = \frac{\overline{X} - \overline{Y} - E(\overline{X} - \overline{Y})}{\sqrt{Var(\overline{X} - \overline{Y})}} \sim F,$$

where $E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$.

Hypotheses about the difference of two Population means (independent samples)

- Distribution F depends on available information about variances σ_1^2 and σ_s^2.
  - If variances are **known**, then $Var(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ and

$$T(\boldsymbol{X} - \boldsymbol{Y}) = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1),$$

  - If variances are **unknown**, but **equal**, then $Var(\overline{X} - \overline{Y}) = s^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$ and

$$T(\boldsymbol{X} - \boldsymbol{Y}) = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

    where

$$s = \sqrt{\frac{\sum_{i=1}^{n_1}(x_i - \overline{x})^2 + \sum_{i=1}^{n_2}(y_i - \overline{y})^2}{n_1 + n_2 - 2}}.$$

Hypotheses about the difference of two Population means (independent samples)

- If there is **no information** about σ_1 and σ_2, then $\widehat{Var}(\overline{X} - \overline{Y}) = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$ and

$$T(\boldsymbol{X} - \boldsymbol{Y}) = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(f) \approx N(0, 1),$$

where

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_i - \overline{x})^2 \text{ and } s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (y_i - \overline{y})^2,$$

$$f = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right\rfloor$$

Hypotheses about the difference of two Population means (independent samples)

And final step of the Algorithm:

- Test: if $|t(x, y)| > q_{\alpha/2}$ then reject H_0. Here $q_{\alpha/2}$ is the complement quantile of $N(0, 1)$, $t(n_1 + n_2 - 2)$ or $t(f)$ distribution.

Using the Normal Approximation

We have showed beforehand how the normal approximation can be used for interval estimation. Hypothesis testing can be performed in a similar way.

- Let us have a random sample $x_1, \ldots, x_n$ from $F(\theta)$, where θ is unknown. F is not a Normal distribution.
- Consider hypotheses

$$H_0 : \theta = \theta_0,$$

$$H_1 : \theta \neq \theta_0.$$

- Consider a consistent point estimate of θ, $\hat{\theta}$.
- If H_0 is true, then in case $n \to \infty$,

$$U = \frac{\hat{\boldsymbol{\theta}} - \theta_0}{\sqrt{Var(\hat{\boldsymbol{\theta}})}} \longrightarrow N(0,1), \quad V = \frac{\hat{\boldsymbol{\theta}} - \theta_0}{\sqrt{\widehat{Var}(\hat{\boldsymbol{\theta}})}} \longrightarrow N(0,1).$$

Using the Normal Approximation

- In case of **large sample *n* we can still use the Normal distribution** quantiles for testing. For given hypotheses, we test if
$$|u| > \lambda_{\alpha/2} \text{ or } |v| > \lambda_{\alpha/2}.$$
- If the inequality above holds, then reject H_0.
- Remember, the level of significance of the test is now **approximately** α (because test statistic is only approximately normally distributed).
- In practice, **usually *t*-quantile is used for testing**, because
  - $t_\alpha(n-1) > \lambda_\alpha$ and in case $n \longrightarrow \infty$, $t_\alpha(n-1) \longrightarrow \lambda_\alpha$
  - Using *t*-quantile decreases the probability of Type I error, because it is more difficult to reject H_0. But in case of failure of rejecting H_0, the probability of making a mistake is smaller when using quantiles of Normal distribution.

Application to the Binomial Distributions

Let p denote the proportion of individuals or objects in a population who possess a specified property (e.g. cars with manual transmissions or smokers who smoke a filter cigarette).
Let us have a random sample $x_1, \dots, x_n$. We are interested in testing the following hypotheses:

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

Consider the point estimate $\widehat{p}$ of parameter p.
The distribution of corresponding estimator $\hat{\boldsymbol{p}}$ is approximately Normal when n increases $((np_0 \geq 10), n(1-p_0) \geq 10)$. So in case of large sample

$$U = \frac{\hat{\boldsymbol{p}} - E\hat{\boldsymbol{p}}}{\sqrt{Var\hat{\boldsymbol{p}}}} \stackrel{H_0 \text{true}}{=} \frac{\hat{\boldsymbol{p}} - p_0}{\sqrt{p_0(1-p_0)/n}} \dot{\sim} N(0,1)$$

Test: $|u| > \lambda_{\alpha/2}$. Type I error is approximately α.

Application to the Binomial Distributions. Example

Example 1. A food industry plans to replace the standard production method by a new one. It is then important to know if the flavour of the product is changed. A so-called triangle test is performed. Each of 500 persons tastes, in random order, three packets of the product, two of which are manufactured according to the standard method and one according to the new method. Each person is asked to select the packet which tastes differently from the two others.
200 persons select the packet with new production. If there is no difference between the tastes, the probability that the new method is selected is 1/3; if there is a difference, we will have $p > 1/3$. Hence our hypotheses are

$$H_0 : p = 1/3$$
$$H_1 : p > 1/3.$$

The firm is satisfied with $\alpha = 0.05$, $\lambda_{0.05} = 1.64$. Compute

$$u = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{\frac{200}{500} - \frac{1}{3}}{\frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{500}} = 3.2$$

Since $3.2 > 1.64$, we can reject H_0 and say that there is difference between the tastes