Estadística II
Chapter 2. Basic ideas in hypothesis testing

Academic year 2010/11

Chapter 2. Basic ideas in hypothesis testing

Contents

- Definition of a test of hypothesis.
- The null and alternative hypotheses.
- Type I and Type II errors.
- Power of a test.
- Definition of a p-value.
- Test of hypothesis: procedure.

Chapter 2. Basic ideas in hypothesis testing

Learning objectives

- Familiarity with the basic ideas in hypothesis testing.
- Ability to formulate the null and alternative hypotheses of a test.
- Ability to interpret the results of a test of hypothesis using the p-value.

Recommended reading

- Meyer, P. “Probabilidad y aplicaciones estadísticas” (1992)
  - Chapter 15
- Newbold, P. “Estadística para los negocios y la economía” (1997)
  - Chapter 9
- Peña, “Regresión y análisis de experimentos” (2005)
  - Chapter 10

Test of hypothesis

In Statistics, a test of hypothesis is a procedure that allows us to determine if a given hypothesis (claim, statement) about a population, is true or false, based on the sample information.

Test of hypothesis

Examples of hypotheses

- It is thought that in the upcoming local elections in Getefe, two seats will be taken by the political party called 'Vientos de Pueblo'. Parties participating in the elections want to check this claim for the possible pre-election deals.

- A company receives a large shipment of items. The shipment is accepted if no more that 5% of the items are nonconforming. How to decide whether to accept or to reject the shipment without examining all the items?

- A researcher wants to know if a new tax reform will be equally received by men and women. How can he check this?

Test of hypothesis

Examples of hypothesis tests (cont.):

All of these examples have something in common:

a) A hypothesis about the population is formed.

b) But, the conclusions about the validity of the hypothesis will be based on the sample.

Null and alternative hypotheses

Null hypothesis H_0 is the hypothesis we wish to test. The subscript 'null' means that H_0, represents the claim that we maintain, unless the data indicates otherwise, therefore it is 'neutral', it reflects the 'current state of affair'. We either 'reject' the null or 'fail to reject'. A test of hypothesis CANNOT prove the null hypothesis.

Null and alternative hypotheses

The null hypothesis is tested against another claim about the population, called the alternative hypothesis, denoted by H_1. The alternative hypothesis can be given explicitly or not. In the latter case, we typically take it as the opposite statement to H_0.

Null and alternative hypotheses

Example We have two coins, one is fair (that is, the probability of head/tails is $1/2$) and one is loaded (with probability of head, p, of $3/4$). We randomly choose one of the two coins for playing, but before we begin our game we are allowed to toss the selected coin twice. The objective is to test the hypothesis (H_0) that $p = 1/2$.

Null and alternative hypotheses

(**a**) Write the sample space:

$$\mathcal{X} = \{(H,H),(H,T),(T,H),(T,T)\}.$$

(**b**) Which of the above outcomes would provide an evidence against the null hypothesis?

$$R = \{(H,H)\}.$$

(**c**) We decide to reject H_0 if we observe R. What kind of errors can we be making here? Find their corresponding probabilities.

Error	Description	Probability
Type I	Reject H_0 if H_0 is true	$(1/2)^2$
Type II	Fail to reject H_0 if H_0 is false	$1-(3/4)^2$

Null and alternative hypotheses

Definition

A **hypothesis test** *is any partition of the sample space, $\mathcal{X}$, in two disjoint regions: critical region (also called rejection region), R, and the acceptance region, $R^c = \mathcal{X} \setminus R$.*

Null and alternative hypotheses

(**d**) Based on (**b**), propose a test of hypothesis to test $H_0: \; p = 1/2$.

$$\mathcal{X} = R \cup R^c = \{(H, H)\} \cup \{(H, T), (T, H), (T, T)\}.$$

(**e**) Propose another test.

$$\mathcal{X} = R_* \cup R_*^c = \emptyset \cup \{(H, H), (H, T), (T, H), (T, T)\}.$$

(**f**) Calculate the probabilities of Type I and Type II errors for the test in (**e**).

Error	Description	Probability
Type I	Reject H_0 if H_0 is true	0
Type II	Fail to reject H_0 if H_0 is false	1

Null and alternative hypotheses

Types of tests

- Simple hypothesis: $H_0 : \theta = \theta_0$.
- Composite hypothesis: $H_0 : \theta \in \Theta_0$ and Θ_0 contains more than one element.
  - One-sided hypothesis:
    - $$H_0 : \theta \leq \theta_0 \quad vs \quad H_A : \theta > \theta_0 \quad .$$
    - $$H_0 : \theta \geq \theta_0 \quad vs \quad H_A : \theta < \theta_0 \quad .$$
  - Two-sided hypothesis:
    - $$H_0 : \theta = \theta_0 \quad vs \quad H_A : \theta \neq \theta_0 \quad .$$
    - $$H_0 : \theta \in [\theta_1, \theta_2] \quad vs \quad H_A : \theta \notin [\theta_1, \theta_2] \quad .$$

Null and alternative hypotheses

Types of tests

Example

- Simple null vs simple alternative: $H_0 : p = 1/2 \quad vs \quad H_A : p = 3/4$.
- One-sided test: $\begin{cases} H_0 : p \leq 1/2 \quad vs \quad H_A : p > 1/2 \\ H_0 : p \geq 1/2 \quad vs \quad H_A : p < 1/2 \end{cases}$.
- Two-sided test: $H_0 : p = 1/2 \quad vs \quad H_A : p \neq 1/2$.

Null and alternative hypotheses

Example

- It is thought that in the upcoming local elections in Getefe, two seats will be taken by the political party called 'Vientos de Pueblo'. Parties participating in the elections want to check this claim for the possible pre-election deals. (NE = Number of Elected from 'Vientos de Pueblo'.)
  $\boxed{H_0 : NE = 2 \quad vs \quad NE \neq 2}$.

- A company receives a large shipment of items. The shipment is accepted if no more that 5% of the items are nonconforming. How to decide whether to accept of to reject the shipment without examining all the items?
  $\boxed{H_0 : p \leq 0.05 \quad vs \quad H_A : p > 0.05}$.

- A researcher wants to know if a new tax reform will be equally received by men and women. How can he check this?
  $\boxed{H_0 : p_M = p_W \quad vs \quad H_A : p_M \neq p_W}$.

The procedure - Type I error

- In the context of a problem, fix a level for the probability of committing Type I error, called **significance level of the test**, α.
- Exclude all the tests whose critical region, R, does not satisfy the condition:

$$\Pr\{R|H_0\} \leq \alpha$$

- Among the remaining tests, choose the one that **minimizes** the probability of Type II error.

Neyman–Pearson

The procedure - power of the test

If the null and alternative can be expressed in terms of a parameter $\theta \in \Theta$, that is, if $H_0 : \theta \in \Theta_0$ and $H_A : \theta \in \Theta_A$, then the **power function of the test** with the critical region R, is the probability of rejecting H_0 if the value of the parameter is θ:

$$\beta(\theta) = \Pr\{R|\theta\}.$$

Remark 1: A hypothesis test has level α if $\beta(\theta) \leq \alpha$ for $\theta \in \Theta_0$.
Remark 2: When $\theta \in \Theta_A$, the probability of Type II error is: $1 - \beta(\theta)$.

Example

(**i**) Let's assume that instead of the two coins we have infinitely many coins, each with the probability of head, p. Calculate the power of the test with the critical region $R = \{(H, H)\}$.

The procedure

In a parametric problem where the hypotheses can be expressed in terms of parameters, $H_0 : \theta \in \Theta_0$ and $H_A : \theta \in \Theta_A$:

1. In the context of the problem fix the **significance level of a test**, α.
2. Exclude all tests with the critical region, R, that does not satisfy the condition:
$$\Pr\{R|\theta\} \leq \alpha,$$
for all $\theta \in \Theta_0$.
3. Among the remaining ones, choose one that **maximizes** the power of the test for $\theta \in \Theta_A$.

Probability of Type II error $= 1 - \beta(\theta)$

The procedure

1. In the hypothesis test the null hypothesis is treated as (*neutral*).
2. We can fix the probability of Type I error as small as we wish, BUT this will increase the probability of Type II error.
3. A test of hypothesis can reject the null hypothesis.
4. A test of hypothesis CANNOT prove the null hypothesis.
5. The acceptance of the null hypothesis should be interpreted as the lack of evidence in the sample data to disprove it.
6. On the contrary, if we reject the null we are pretty sure ($\Pr(R|H_0) \leq \alpha$) that H_0 is false and thus we automatically accept the alternative.

p-value (observed significance level)

Definition

The **p-value** *of the test is the smallest significance level which would lead to the rejection of the null hypothesis.*

- The p-value is the probability (calculated assuming that H_0 is true) of observing a test statistic at least as extreme as the one we actually observed.
- The p-value depends on the sample $(x_1, x_2, \ldots, x_n)$.
- Can be regarded as the amount of evidence in favour of the null hypothesis coming from the sample data:
  - If the p-value is **smaller** than the significance level of the test α, then the support for H_0 is scarce and the null is rejected.
  - If the p-value is **bigger (or equal)** than the significance level of the test α, then the support for H_0 is sufficient and the null is not rejected.

p-value (observed significance level)

Calculating the p-value

- One-sided test: $H_0 : \theta \leq \theta_0 \quad vs \quad H_A : \theta > \theta_0$
  - Observed test statistic: $T(x_1, x_2, \ldots, x_n) = t$.
  - Rejection region: $R = \{T(x_1, x_2, \ldots, x_n) > k\}$.
  - p-value $= \Pr(T \geq t \,|\theta = \theta_0)$.
- One-sided test: $H_0 : \theta \geq \theta_0 \quad vs \quad H_A : \theta < \theta_0$
  - Observed test statistic: $T(x_1, x_2, \ldots, x_n) = t$.
  - Rejection region: $R = \{T(x_1, x_2, \ldots, x_n) < k\}$.
  - p-value $= \Pr(T \leq t \,|\theta = \theta_0)$.

p-value

Calculating the p-value

- Two-sided test: $H_0 : \theta = \theta_0 \quad vs \quad H_A : \theta \neq \theta_0$
  - Observed test statistic: $T(x_1, x_2, \ldots, x_n) = t$.
  - Rejection region:
    $R = \{T(x_1, x_2, \ldots, x_n) < k_1\} \cup \{T(x_1, x_2, \ldots, x_n) > k_2\}$.
  - p-value $= \min\big(2\,\Pr(T \leq t\,|\theta = \theta_0); 2\,\Pr(T \geq t\,|\theta = \theta_0)\big)$.

p-value

Example

Let X represent 'investment returns (from a given sector of economy) after a strong increase of euro against dollar'. It is thought that the mean of this variable is 15. An economist thinks that the mean return has changed, thus he takes a sample of 9 investments (of the sector) for which the sample mean return is 15.308 and sample variance (quasi-variance) is 0.193.

a) Carry out a test of hypothesis at 5% level to test economist's claim. Write down all the assumptions needed.

b) Based on part **a)**, justify if the 95% confidence interval for the population mean (centered at $\overline{x}$) would contain 15.

c) Bound the p-value. If we performed the test at 10% level, would we reject or fail to reject the null? Justify briefly.

p-value

a) Assumptions: $X_1, ..., X_n$ are iid (or SRS = simple random sample) from $\mathcal{N}(\mu, \sigma)$.

$$H_0 : \mu = 15 \qquad vs \qquad H_1 : \mu \neq 15$$

Reject H_0 at 5% level if $\left|\frac{\overline{x}-\mu_0}{\frac{s_x}{\sqrt{n}}}\right| > t_{n-1;\frac{\alpha}{2}}$

We have $\left|\frac{15.308-15}{\frac{0.44}{\sqrt{9}}}\right| = 2.1 \; > \; t_{8;0.025} = 2.306.$ $\Rightarrow$ **???**

b) Because of duality between the confidence intervals and hypothesis testing, 15 (value hypothesized in H_0) would belong to the 95% confidence interval for μ.

c) p-value $= 2P\,(t_8 > 2.1)$.

p-value $\in (0.05, 0.1)$, hence at $\alpha = 0.1$ level we would reject H_0.

Hypothesis testing on a normal population

Test for μ:

$H_0 : \mu = \mu_0$ (σ conocida); $R = \left\{ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$

$H_0 : \mu = \mu_0$ (σ desconocida); $R = \left\{ |\bar{x} - \mu_0| > t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \leq \mu_0$ (σ conocida); $R = \left\{ \bar{x} - \mu_0 > z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\}$

$H_0 : \mu \leq \mu_0$ (σ desconocida); $R = \left\{ \bar{x} - \mu_0 > t_{n-1;\alpha} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \geq \mu_0$ (σ conocida); $R = \left\{ \bar{x} - \mu_0 < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$

$H_0 : \mu \geq \mu_0$ (σ desconocida); $R = \left\{ \bar{x} - \mu_0 < t_{n-1;1-\alpha} \frac{s}{\sqrt{n}} \right\}$

Hypothesis testing on a normal population

Test for σ:

$H_0 : \sigma = \sigma_0; \qquad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 \notin \left[\chi^2_{n-1;1-\alpha/2} \; , \; \chi^2_{n-1;\alpha/2} \right] \right\}$

$H_0 : \sigma \leq \sigma_0; \qquad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 > \chi^2_{n-1;\alpha} \right\}$

$H_0 : \sigma \geq \sigma_0; \qquad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 < \chi^2_{n-1;1-\alpha} \right\}$

Test for proportions

$H_0 : p = p_0; \qquad R = \left\{ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

$H_0 : p \leq p_0; \qquad R = \left\{ \bar{x} - p_0 > z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

$H_0 : p \geq p_0; \qquad R = \left\{ \bar{x} - p_0 < z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$