Statistical Inference

4

Outline

1. Introduction
2. Confidence intervals
   - Sample size
3. Hypothesis testing
   - Generalities
   - Critical region and *p*-value
   - Confidence intervals and hypotesis tests
   - Particular tests

Introduction

Statistical Inference is a prediction process that leads us to conclusions about a population from information extracted from a sample.

We have studied a number of distribution models depending on one or several parameters, now we will learn how can we estimate such parameters.

Introduction

A statistical hypothesis is a conjecture about a population distribution. This hypothesis is to be tested using the information from a sample.

Interval-valued estimators of parameters (Confidence Intervals) are also considered here.

Basic concepts

- A random sample is composed by n independent random variables with the same distribution $X_1, X_2, \ldots, X_n$

- A statistic is any transformation (function) of the observations of the random sample. Consequently, it is a random variable $f(X_1, X_2, \ldots, X_n)$.

Basic concepts

- An estimator of a parameter θ is any function from the sample

$$\hat{\theta} = f(X_1, X_2, \ldots, X_n)$$

that assumes values close to the one of θ.

It is a statistic, and it is used to estimate θ.

It is said to be unbiased if

$$\mathrm{E}[\hat{\theta}] = \theta$$

Distribution of the sample mean

- The sample mean

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is a natural estimator of the population mean μ.

It is unbiased with variance σ^2/n, where σ is the standard deviation of X.

Distribution of the sample mean

In accordance with the CLT, for any distribution of *X*, whenever the sample size *n* is sufficiently large

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx \mathrm{N}(0,1)$$

Distribution of the sample mean

- Distribution of the sample proportion.

We denote by p the population proportion of individuals with a certain characteristic. The r.v. X that assumes value 1 on individuals with the characteristic and 0 on the remaining individuals follows a B(1,p) distribution

$$\hat{p} = \frac{\text{number of individuals from the sample with the characteristic}}{\text{sample size}} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

$$\mathrm{E}[\hat{p}] = p \ ; \ \mathrm{Var}[\hat{p}] = \frac{p(1-p)}{n}$$

If $n>30$ and $np(1-p)>5$, we can apply the CLT.

Distribution of the sample variance

- The sample variance is an unbiased estimator of the population variance (σ^2)

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2}{n-1}$$

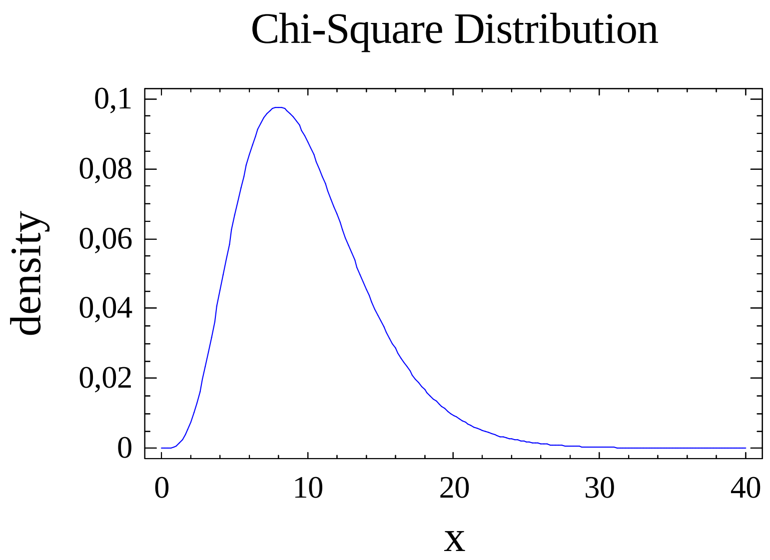
Addendum: Chi-square distribution

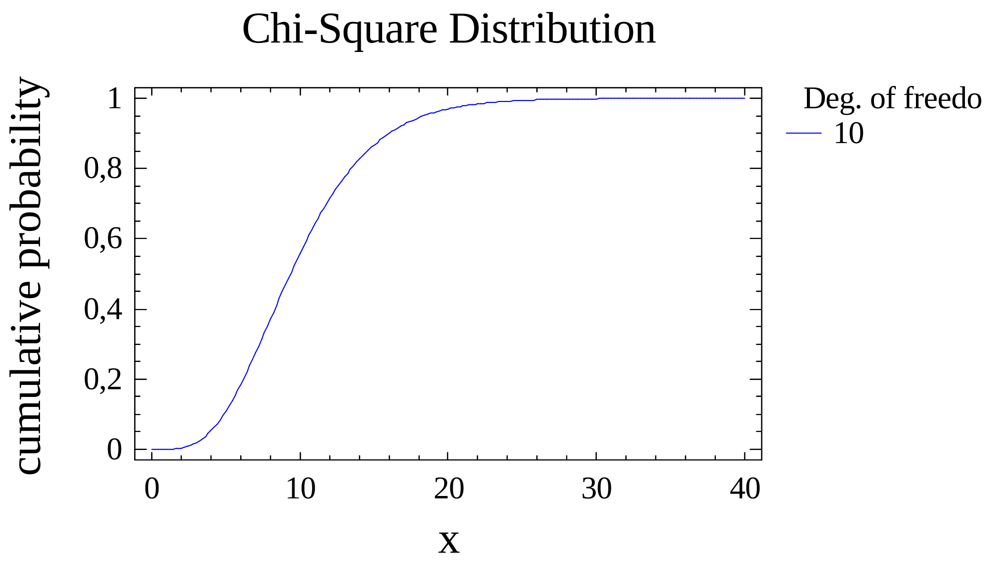
Given n independent standard Normal random variables $X_1,X_2,\ldots,X_n$, the random variable $Y=X_1^2+X_2^2+\ldots+X_n^2$ follows a Chi-square distribution with n degrees of freedom,

$$Y \sim \chi_n^2$$

$$\mathrm{E}[Y] = n \;\; ; \;\; \mathrm{Var}[Y] = 2n$$

Addendum: Chi-square distribution





Sampling distributions for normal populations

- Distribution of the sample variance. When the sample is taken from a normal population,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

We have then

$$\mathrm{Var}[S^2] = 2\sigma^4/(n-1)^2$$

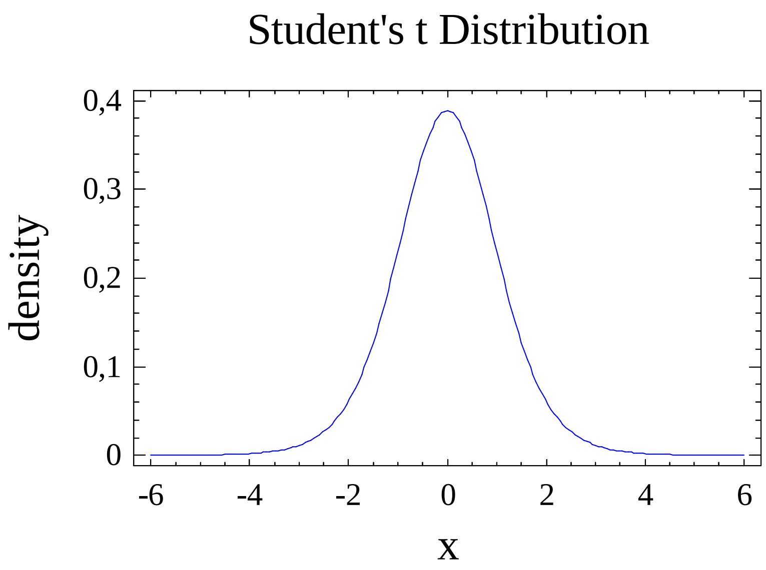
Addendum: Student's *t* distribution

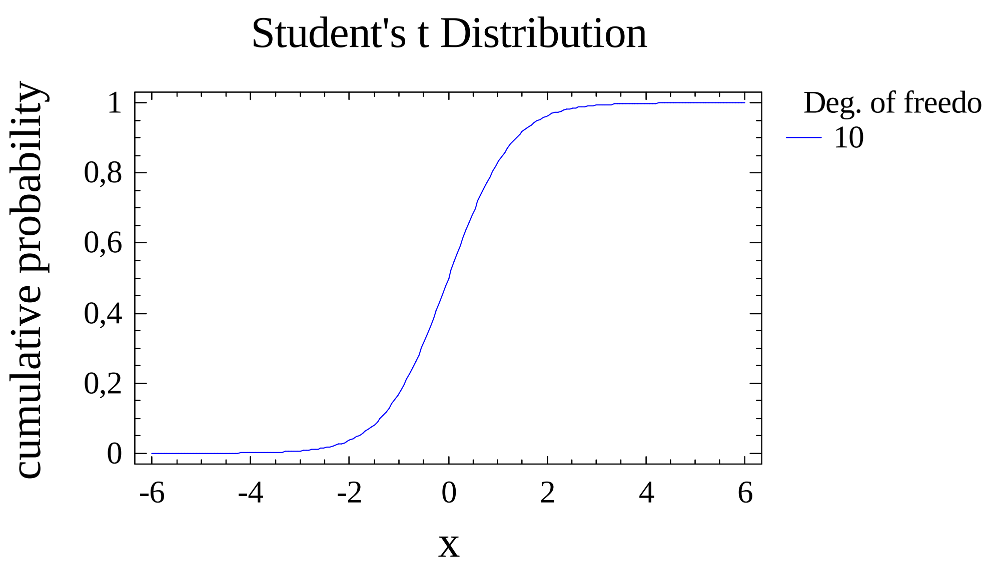
Given a standard normal random variable X and Y indep. of X following a chi-square distribution with n degrees of freedom, the random variable $X/(Y/n)^{1/2}$ follows a distribution t with n degrees of freedom

$$Z = \frac{X}{\sqrt{Y/n}} \sim t_n$$

$$E[Z] = 0 \text{ if } n \geq 2 \;;\; Var[Z] = n/(n-2) \text{ if } n \geq 3$$

Addendum: Student's *t* distribution





Sampling distributions for normal populations

- Distribution of the sampe mean with unknown variance. When the sample is taken from a normal population with unknown variance, we replace it by the sample variance to obtain

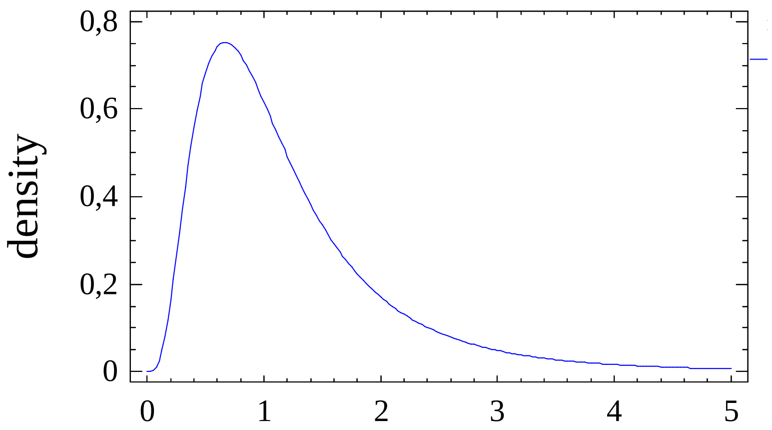
$$\frac{\overline{X}-\mu}{\sqrt{S^2/n}} \sim t_{n-1}$$

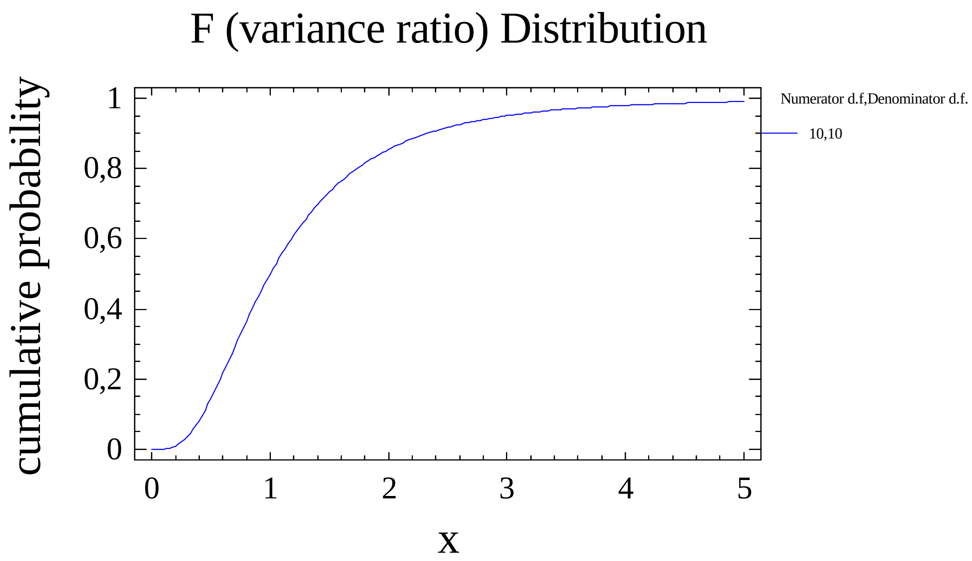
Addendum: Fisher's F distribution

Given two independent chi-square random variables X and Y such that X has n_1 degrees of freedom and Y has n_2 degrees of freedom, the random variable $(X/n_1)/(Y/n_2)$ follows a F distribution with n_1 and n_2 degrees of freedom

$$Z = \frac{X / n_1}{Y / n_2} \sim \mathrm{F}_{n_1, n_2}$$

Addendum: Fisher's F distribution





Sampling distributions for normal populations

- Variance ratio. When two independent samples (sample from X and sample from Y) of respective sizes n and m are taken from two normal populations, the ratio of their sample variances follows an F distribution with $n–1$ and $m–1$ degrees of freedom

$$\frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} \sim F_{n-1,m-1}$$

Outline

1. Introduction
2. Confidence intervals
   - Sample size
3. Hypothesis testing
   - Generalities
   - Critical region and *p*-value
   - Confidence intervals and hypotesis tests
   - Particular tests

Confidence Intervals

- For any fixed probability $1-\alpha$ and a random sample $X_1, X_2, \ldots, X_n$ we build a (random) interval containing a parameter θ with probability $1-\alpha$. The extreme values of the interval will be based on estimators of θ.

A Confidence Interval with confidence coefficient (or level) $1-\alpha$ is obtained sustituting the estimators of θ by its estimation.

Confidence Intervals

- Building a Confidence Interval.

Our aim is to find two statistics such that

$$P\left(\hat{\theta}_1(X_1, X_2, \ldots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \ldots, X_n)\right) = 1 - \alpha$$

Condier any estimator of θ with a known distribution (e.g. normal), then we have

$$P\left(\hat{\theta} - \sigma_{\hat{\theta}} z_{\alpha/2} < \theta < \hat{\theta} + \sigma_{\hat{\theta}} z_{\alpha/2}\right) = 1 - \alpha \quad \text{if} \quad \hat{\theta} \sim N(\theta, \sigma_{\hat{\theta}})$$

Confindence Intervals

- A Confindence Interval for θ with confidence level $1-\alpha$ built from a statistic with a normal distribution will be of the following type:

$$\theta \in \left[\hat{\theta} - \sigma_{\hat{\theta}} z_{\alpha/2}, \hat{\theta} + \sigma_{\hat{\theta}} z_{\alpha/2}\right]$$

where $P(Z > z_{\alpha/2}) = \alpha/2$ for $Z \sim N(0,1)$.

Confidence Intervals

- Choice of sample size. A confidence interval for θ is of the form

$$\left[\hat{\theta} - a, \hat{\theta} + b\right]$$

where a and b depend on

1. The confidence level ;
2. The variance of the estimator of θ;
3. The sample size .

The variance of the estimator depends on the sample size.

Confidence Intervals

- Choice of sample size.

For any fixed estimation error (length of the confidence interval), it is possible to determine the sample size that is needed by solving the equation

$$a+b = A \ ,$$

where L is the desired length for the CI.

Confidence Intervals (Examples)

1. The lifetime of an electronic device of a certain kind is normally distributed with $\sigma=1$ year. The sample mean of the lifetime of $n=10$ devices was 3.5 years. Build a 90% CI on the mean lifetime of the electronic devices of the former kind.
2. Out of a sample of $n=100$ University students, 86 of them owned a laptop. Build a 95% CI on the proportion of students owning a laptop.

Confidence Intervals (Examples)

3. The compressive strength of concrete is normally distributed. In a sample of size n=20, we have obtained a sample mean of 2450psi and a sample variance of 1600psi^2. Build a 99% CI on the mean compressive strength of concrete.
4. The percentage of titanium in an alloy used in aerospace castings is known to be normally distributed. It has been measured in n=51 randomly selected parts and the sample standard deviation obtained was s=0.37. Build a 95% CI on σ^2.

Outline

1. Introduction
2. Confidence intervals
   - Sample size
3. Hypothesis testing
   - Generalities
   - Critical region and *p*-value
   - Confidence intervals and hypotesis tests
   - Particular tests

Testing statistical hypothesis

Hypothesis test are used to check the truth of a statement about a parameter using information from a sample.

The null hypothesis (H_0) is the hypothesis (statement) that we wish to test

- 'the mean compressive strength of concrete is μ_0', H_0: $\mu = \mu_0$

The null hypothesis is only rejected when there is strong evidence against it (the sample is strongly inconsistent with it).

Testing statistical hypothesis

- The alternative hypothesis (H_1) is the statement to be accepted when rejecting H_0
  - $H_1: \mu \neq \mu_0$
  - $H_1: \mu < \mu_0$

We will only reject the null hypothesis when the sample is strongly inconsistent with it. Equivalently, when the sample information shows a strong evidence favourable to H_1.

Testing statistical hypothesis

- Types of tests
  - Two-sided: when the estimate of the parameter is very different from the objective value (larger or smaller), we will reject the null hypothesis

$$H_0: \mu = \mu_0 \; ; \; H_1: \mu \neq \mu_0$$

  - One-sided: we will reject the null hypothesis only when the estimate is much smaller (or much larger) than the objective value

$$H_0: \mu = \mu_0 \; ; \; H_1: \mu < \mu_0$$

$$\text{Equiv. to} \quad H_0: \mu \geq \mu_0 \; ; \; H_1: \mu < \mu_0$$

Testing statistical hypothesis

- When we wonder wheter we can firmly state something (guarantee it), we must write in the alternative hypothesis.
  - Q: ‘Can we state that the mean compressive strength of concrete is less than μ_0?’

$$H_0: \mu = \mu_0 \; ; \; H_1: \mu < \mu_0$$

- When we wonder whether we should reject something or not, we must write it in the null hypothesis.
  - Q: ‘The supplier states that the mean compressive strength of concrete is μ_0. Can we reject his statement?’

$$H_0: \mu = \mu_0 \; ; \; H_1: \mu \neq \mu$$

  - Q: ‘The supplier suspects that the mean compressive strength of concrete is greater than μ_0. Is there enough sample evidence to support his suspicion?’

$$H_0: \mu = \mu_0 \; ; \; H_1: \mu > \mu_0$$

Testing statistical hypothesis

- Error types
  - Type I error: Rejecting the null hipothesis (H_0) when it is true,

    $$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 | H_0)$$

    this error is crucial, α (significance level) is fixed.
  - Type II error: Failing to reject the null hypothesis (H_0) when it is false,

    $$\beta = P(\text{Type II error}) = P(\text{fail to reject } H_0 | H_1)$$

    this second error is not that important.

Test Statistic

- A test statistic is based on an estimator of our parameter of interest θ and, under the assumption that the null hypothesis is true, its distribution is known.

if $\quad H_0 : \theta = \theta_0 \quad$ we are interested in the distance $d(\hat{\theta}, \theta_0)$

- The values of the test statistic that lead us to rejection of the null hypothesis depend on the alternative hypothesis.

Test Statistic

- In the test

$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

Reject H_0 for large and small values of the statistic

- In the test

$$H_0: \theta \leq \theta_0 \qquad H_0: \theta = \theta_0$$
$$H_1: \theta > \theta_0 \qquad H_1: \theta > \theta_0$$

Reject H_0 for large values of the statistic

$\hat{\theta} - \theta_0$ sign does matter.

Rejection region

- The set of values of the test statistic that are strongly inconsistent with the null hypothesis and lead us to its rejection depend on the significance level α.
- The significance level α is the probability of type I error that we agree to assume.

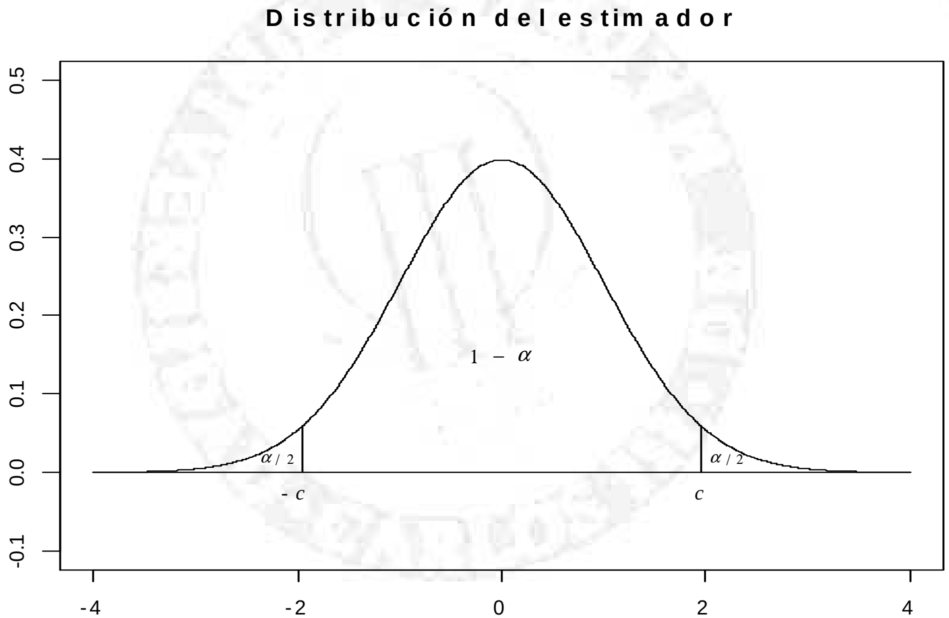
Commonly $\alpha = 0.01$ or 0.05

Rejection region

- For a given α, we have P(reject $H_0|H_0$) = α

Distribution of the estimator of θ known (under H_0)

In the test H_0: $\theta = \theta_0$; H_1: $\theta \neq \theta_0$



Reject H_0 if

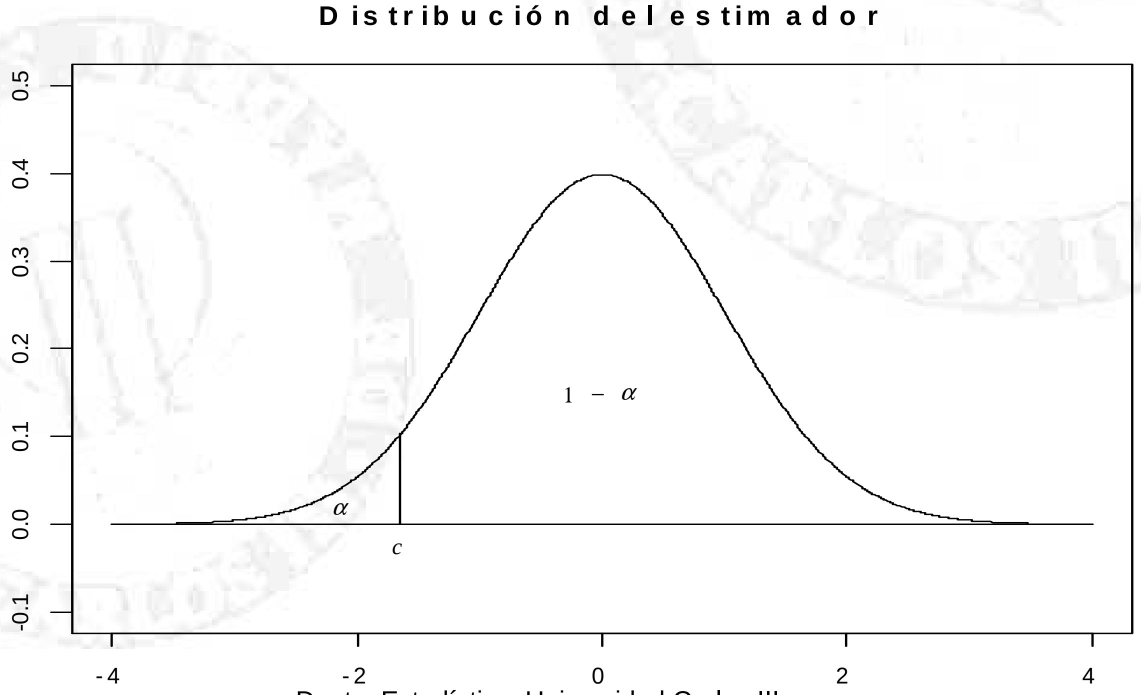
$$d(\hat{\theta},\theta_0) = \left|\hat{\theta} - \theta_0\right| > c$$

The value c determines the rejection region.

Rejection region

In the test H_0: $\theta = \theta_0$; H_1: $\theta < \theta_0$

Reject H_0 if $d(\hat{\theta}, \theta_0) = \hat{\theta} - \theta_0 < c$



Procedure for Hypothesis Tests

- From the problem context, identify the parameter of interest
- State the null hypothesis H_0
- Specify an appropriate alternative hypothesis H_1
- Choose a significance level α
- Determine an appropriate test statistic
- State the rejection region for the statistic
- Compute any necessary sample quantities, substitute these into the equation for the test statistic
- Decide whether or not H_0 should be rejected and report that in the problem context

p-value

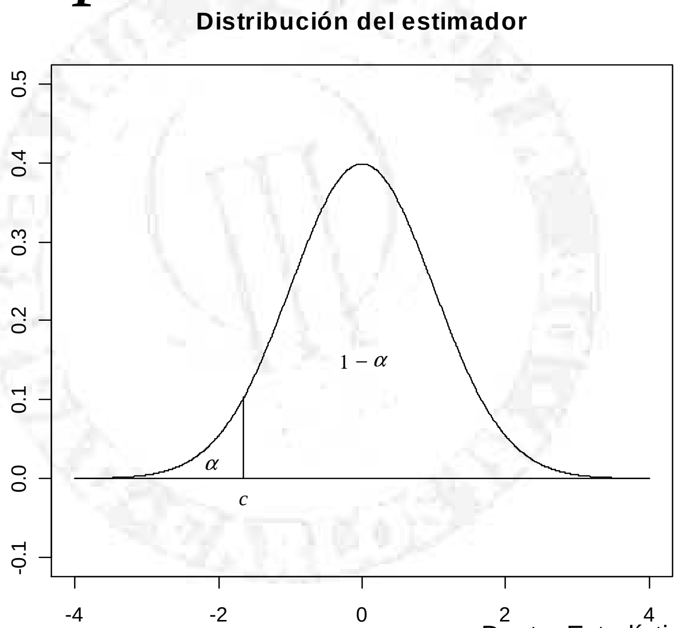
The *p*-value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

It is the probability of a random sample that is more inconsistent with H_0 than ours (assuming the truth of H_0).

p-value

The smaller the *p*-value is, the more confident we are that the null hypothesis is wrong.

For a *p*-value smaller than 0.05, H_0 is rejected.



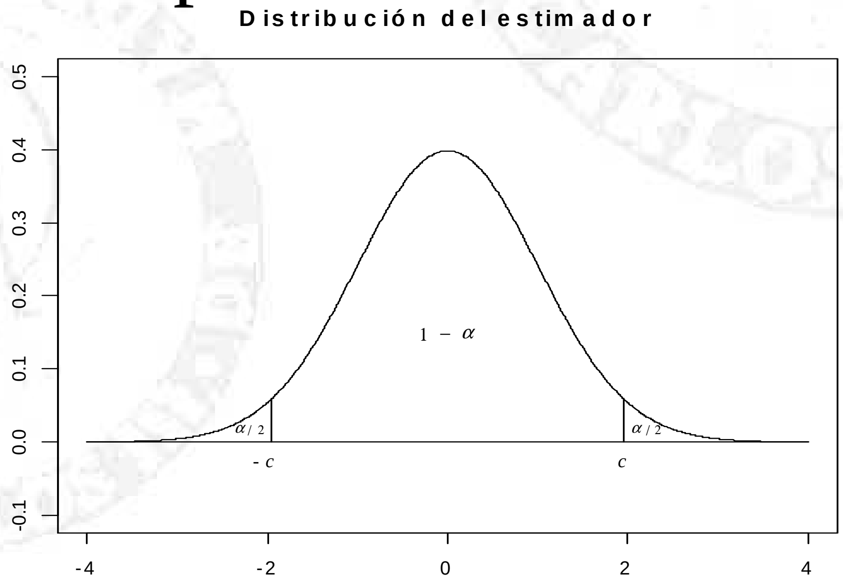
In the test

H_0: $\theta = \theta_0$; H_1: $\theta < \theta_0$

the *p*-value is α such that the statistic c assumes the same value as in our sample.

p-value

In the test H_0: $\theta = \theta_0$; H_1: $\theta \neq \theta_0$ the p-value is α such that the statistic c (or $-c$) assumes the same value as in our sample.



CIs versus Hypothesis tests

- Given a two-sided test

$$H_0\colon \theta = \theta_0 \;;\quad H_1\colon \theta \neq \theta_0$$

with significance level α, the null hypothesis is rejected if θ_0 does not belong to the Confidence Interval for θ with confidence level $1-\alpha$.

Particular tests

- Test on the mean of a normal population (or large sample) with known variance
  - Null hypothesis. H_0: $\mu = \mu_0$
  - Alternative hypothesis. H_1: $\mu \neq \mu_0$
    - Reject H_0 if $\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}$
  - Alternative hypothesis. H_1: $\mu < \mu_0$
    - Reject H_0 if $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_{\alpha}$
  - Alternative. H_1: $\mu > \mu_0$
    - Reject H_0 if $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > z_{\alpha}$

Particular Confindence Intervals

- Confidence Interval on the mean of a normal population (or large sample) with known variance

$$\mu \in \left[\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n}\,,\bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n}\right]$$

confindence level $1-\alpha$,

where $P(Z > z_{\alpha}) = \alpha$ if $Z \sim N(0,1)$

Particular tests

- Test on a proportion
  - Null hypothesis. H_0: $p = p_0$
  - Alternative hypothesis. H_1: $p \neq p_0$
    - Reject H_0 if $\left|\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}\right| > z_{\alpha/2}$
  - Alternative hypothesis. H_1: $p < p_0$
    - Reject H_0 if $\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} < -z_{\alpha}$
  - Alternative hypothesis. H_1: $p > p_0$
    - Reject H_0 if $\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} > z_{\alpha}$

Particular Confidence Intervals

- Confidence Interval on a proportion

$$p \in \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

confidence level $1-\alpha$,

where $P(Z > z_\alpha) = \alpha$ if Z~N(0,1)

Particular tests

- Test on the mean of a normal population with unknown variance
  - Null hypothesis. H_0: $\mu = \mu_0$
  - Alternative hypothesis. H_1: $\mu \neq \mu_0$
    - Reject H_0 if $\left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| > t_{n-1,\alpha/2}$
  - Alternative hypothesis. H_1: $\mu < \mu_0$
    - Reject H_0 if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$
  - Alternative hypothesis. H_1: $\mu > \mu_0$
    - Reject H_0 if $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$

Particular Confidence Intervals

- Confindence Interval on the mean of a normal population with unknown variance.

$$\mu \in \left[\bar{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \right]$$

confidence level $1-\alpha$,

where $P(X > t_{n,\alpha}) = \alpha$ if $X \sim t_n$

Particular tests

- Test on the variance of a normal population

  - Null hypothesis. H_0: $\sigma^2 = \sigma_0^2$
  - Alternative hypothesis. H_1: $\sigma^2 \neq \sigma_0^2$
    - Reject H_0 if $\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha/2}$ or $\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha/2}$
  - Alternative hypothesis. H_1: $\sigma^2 < \sigma_0^2$
    - Reject H_0 if $\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha}$
  - Alternative hypothesis. H_1: $\sigma^2 > \sigma_0^2$
    - Reject H_0 if $\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha}$

Particular Confidence Intervals

- Confidence Interval on the variance of a normal population

$$\sigma^2 \in \left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \right]$$

confidence level $1-\alpha$,

where $P(X > \chi^2_{n,\alpha}) = \alpha$ if $X \sim \chi^2_n$

Particular tests

- Test on the difference in means of two normal populations (large samples) with known variances
  - Null hypothesis. H_0: $\mu_1 = \mu_2$
  - Alternative hypothesis. H_1: $\mu_1 \neq \mu_2$
    - Reject H_0 if $$\left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \right| > z_{\alpha/2}$$
  - Alternative hypothesis. H_1: $\mu_1 < \mu_2$
    - Reject H_0 if $$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} < -z_{\alpha}$$
  - Alternative hypothesis. H_1: $\mu_1 > \mu_2$
    - Reject H_0 if $$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} > z_{\alpha}$$

Particular Confidence Intervals

- Confidence Interval on the difference in means of two normal populations (or large samples) with known variances

$$\mu_1 - \mu_2 \in \left[\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

conficence level $1-\alpha$,

where $P(Z > z_\alpha) = \alpha$ if $Z \sim N(0,1)$

Particular tests

$$\hat{p}_0 = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

- Test on the difference in proportions of two populations (independent samples)
  - Null hypothesis. H_0: $p_1 = p_2$
  - Alternative hypothesis. H_1: $p_1 \neq p_2$
    - Reject H_0 if $$\left| \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1-\hat{p}_0)(1/n_1 + 1/n_2)}} \right| > z_{\alpha/2}$$
  - Alternative hypothesis. H_1: $p_1 < p_2$
    - Reject H_0 if $$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1-\hat{p}_0)(1/n_1 + 1/n_2)}} < -z_{\alpha}$$
  - Alternative hypothesis. H_1: $p_1 > p_2$
    - Reject H_0 if $$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1-\hat{p}_0)(1/n_1 + 1/n_2)}} > z_{\alpha}$$

Particular Confidence Intervals

- Confidence Interval on the difference in proportions for two populations (independent samples).

$$p_1 - p_2 \in \left[\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right]$$

confidence level $1-\alpha$,

where $P(Z > z_\alpha) = \alpha$ if Z~N(0,1)

Particular tests

- Approximate test on the difference in means of two normal populations with unknown unequal variances (indep. samples)
  - Null hypothesis. H_0: $\mu_1 = \mu_2$
  - Alternative hypothesis. H_1: $\mu_1 \neq \mu_2$
    - Reject H_0 if $$\left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}\right| > z_{\alpha/2}$$
  - Alternative hypothesis. H_1: $\mu_1 < \mu_2$
    - Reject H_0 if $$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} < -z_{\alpha}$$
  - Alternative hypothesis. H_1: $\mu_1 > \mu_2$
    - Reject H_0 if $$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} > z_{\alpha}$$

Particular Confidence Intervals

- Approximate Confidence Interval on the difference in means of two normal populations, unknown unequal variances (indep. samples)

$$\mu_1 - \mu_2 \in \left[\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

confidence level $1-\alpha$,

where $P(Z > z_\alpha) = \alpha$ if Z~N(0,1)

Particular tests

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Test on the difference in means of two normal populations with unknown equal variances
  - Null hypothesis. H_0: $\mu_1 = \mu_2$
  - Alternative hypothesis. H_1: $\mu_1 \neq \mu_2$
    - Reject H_0 if $$\left| \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{1/n_1 + 1/n_2}} \right| > t_{n_1+n_2-2,\alpha/2}$$
  - Alternative hypothesis. H_1: $\mu_1 < \mu_2$
    - Reject H_0 if $$\frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{1/n_1 + 1/n_2}} < -t_{n_1+n_2-2,\alpha}$$
  - Alternative hypothesis. H_1: $\mu_1 > \mu_2$
    - Reject H_0 if $$\frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{1/n_1 + 1/n_2}} > t_{n_1+n_2-2,\alpha}$$

Particular Confidence Intervals

- Confidence Interval on the difference in means of two normal populations with unknown equal variances (indep. samples)

$$\mu_1 - \mu_2 \in \left[\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2,\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2,\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

confidence level $1-\alpha$,

where $P(X > t_{n,\alpha}) = \alpha$ if $X \sim t_n$

Particular tests

$$s_d^2 = \frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}$$

- Test on the difference in means of two normal populations (paired data), $d = x_1 - x_2$
  - Null hypothesis. H_0: $\mu_d = 0$
  - Alternative hypothesis. H_1: $\mu_d \neq 0$
    - Reject H_0 if $$\left| \frac{\bar{d}}{s_d / \sqrt{n}} \right| > t_{n-1,\alpha/2}$$
  - Alternative hypothesis. H_1: $\mu_d < 0$
    - Reject H_0 if $$\frac{\bar{d}}{s_d / \sqrt{n}} < -t_{n-1,\alpha}$$
  - Alternative hypothesis. H_1: $\mu_d > 0$
    - Reject H_0 if $$\frac{\bar{d}}{s_d / \sqrt{n}} > t_{n-1,\alpha}$$

Particular Confidence Intervals

- Confidence Interval on the difference in means of two normal populations with unknown variances (paired data)

$$\mu_1 - \mu_2 \in \left[\bar{d} - t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}} \right]$$

confidence level $1-\alpha$,

where $P(X > t_{n,\alpha}) = \alpha$ if $X \sim t_n$

Particular tests

- Test on the ratio of the variances of two normal populations
  - Null hypothesis. H_0: $\sigma_1^2 = \sigma_2^2$
  - Alternative hypothesis. H_1: $\sigma_1^2 \neq \sigma_2^2$
    - Reject H_0 if $\frac{S_1^2}{S_2^2} < F_{n_1-1,n_2-1,1-\alpha/2}$ or $\frac{S_1^2}{S_2^2} > F_{n_1-1,n_2-1,\alpha/2}$
  - Alternative hypothesis. H_1: $\sigma_1^2 < \sigma_2^2$
    - Reject H_0 if $\frac{S_1^2}{S_2^2} < F_{n_1-1,n_2-1,1-\alpha}$
  - Alternative hypothesis. H_1: $\sigma_1^2 > \sigma_2^2$
    - Reject H_0 if $\frac{S_1^2}{S_2^2} > F_{n_1-1,n_2-1,\alpha}$

Particular Confidence Intervals

- Confidence Interval on the variance ratio of two normal populations.

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left[\frac{s_1^2}{s_2^2} F_{n_2-1,n_1-1,1-\alpha/2}, \frac{s_1^2}{s_2^2} F_{n_2-1,n_1-1,\alpha/2} \right]$$

confidence level $1-\alpha$,

where $P(X > F_{n_1-1,n_2-1,\alpha}) = \alpha$ if $X \sim F_{n_1-1,n_2-1}$

$F_{n_2-1,n_1-1,1-\alpha} = 1/F_{n_1-1,n_2-1,\alpha}$

Particular tests

$$\hat{\lambda} = \bar{x} \quad ; \quad \hat{\sigma}^2_{\hat{\lambda}} = \bar{x}/n$$

- Test on the λ in a Poisson model
  - Null hypothesis. H_0: $\lambda = \lambda_0$
  - Alternative hypothesis. H_1: $\lambda \neq \lambda_0$
    - Reject H_0 if $\left| \dfrac{\bar{x} - \lambda_0}{\sqrt{\bar{x}/n}} \right| > z_{\alpha/2}$
  - Alternative hypothesis. H_1: $\lambda < \lambda_0$
    - Reject H_0 if $\dfrac{\bar{x} - \lambda_0}{\sqrt{\bar{x}/n}} < -z_{\alpha}$
  - Alternative hypothesis. H_1: $\lambda > \lambda_0$
    - Reject H_0 if $\dfrac{\bar{x} - \lambda_0}{\sqrt{\bar{x}/n}} > z_{\alpha}$

Particular Confidence Intervals

- Approximate Confidence Interval on the λ in a Poisson population

$$\lambda \in \left[\bar{x} - z_{\alpha/2}\sqrt{\bar{x}/n}, \bar{x} + z_{\alpha/2}\sqrt{\bar{x}/n}\right]$$

confidence level $1-\alpha$,

where $P(Z > z_{\alpha}) = \alpha$ if $Z \sim N(0,1)$