STATISTICAL METHODS FOR BUSINESS

UNIT 7: INFERENTIAL TOOLS. DISTRIBUTIONS ASSOCIATED WITH SAMPLING

7.1.- Distributions associated with the sampling process.

7.2.- Inferential processes and relevant distributions.

UNIT 7. GOALS

To describe the chi-square and Student's t distributions.
To calculate probabilities and percentiles (quantiles).
To apply the main pivotal statistics used in inferential processes on the mean, the proportion and the variance of a population.

UNIT 7: INFERENTIAL TOOLS. DISTRIBUTIONS ASSOCIATED WITH SAMPLING

7.1.- Distributions associated with the sampling process.

Probability models associated with sampling

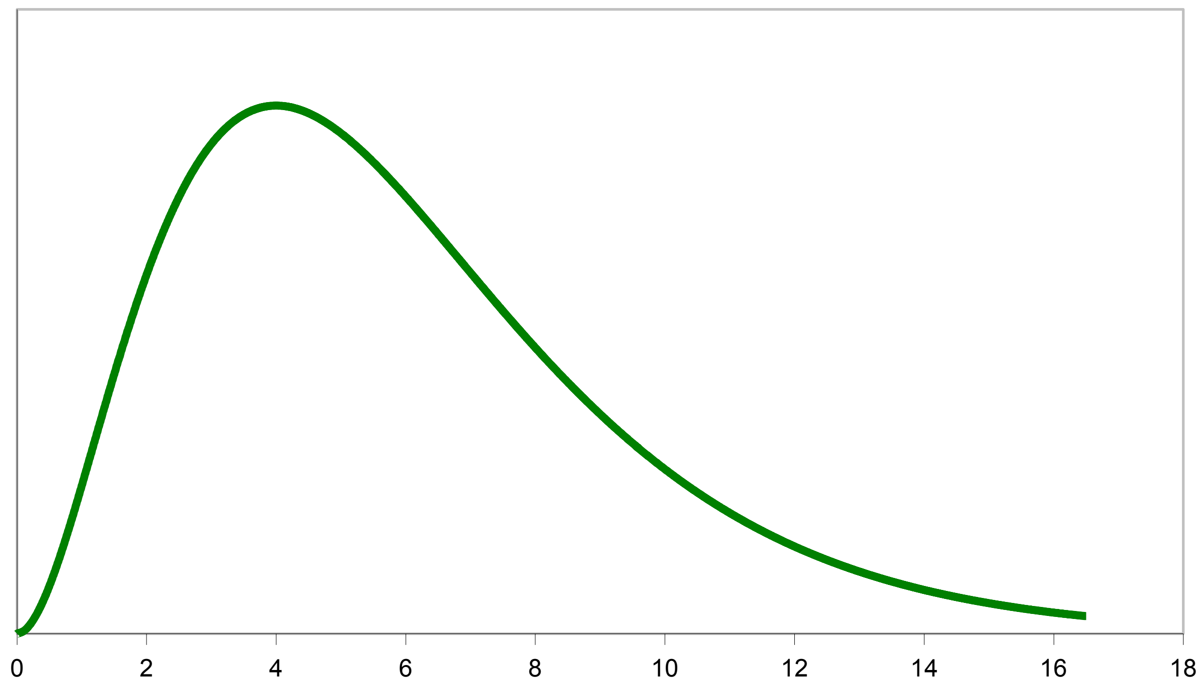
- **Normal distribución** $N(\mu, \sigma)$
- **Chi-squared distribution** χ^2_n
- **Student's *t* distribution** t_n

Chi-squared distribution

$X_1, \ldots, X_n$ INDEPENDIENT RVs with distribution N(0,1):

$$Z = \sum_{i=1}^{n} X_i^2 \approx \chi_n^2$$

Z has a chi-squared distribution with n degrees of freedom



Degrees of freedom

Expression	RANDOM VARIABLES	CONSTRAINTS	Degrees of freedom
$\sum_{i=1}^{n} X_i^2$	$X_1, X_2, \ldots, X_n$	None	**n**
$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$	$X_1, X_2, \ldots, X_n$ or $X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X}$	$\dfrac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$ $\Updownarrow$ one $\sum_{i=1}^{n} \left(X_i - \bar{X}\right) = 0$	**n-1**

The number of degrees of freedom in an expression may be interpreted as the number of values that may be arbitrarily chosen.

Chi-squared distribution

Reproductivity

Given two independent RVs $X \approx \chi^2_n$ $Y \approx \chi^2_m$

It holds:

$$\boxed{X + Y \approx \chi^2_{n+m}}$$

The chi-squared model is REPRODUCTIVE with respect to the number of degrees of freedom.

Fisher's Theorem

Given a simple random sample (X_1, ..., X_n) from a N(μ,σ) population, it holds:

- The sample mean $\bar{X}$ and the sample variance S^2 are independent random variables.

- The distribution of the random variable $\frac{(n-1)S^2}{\sigma^2}$ is chi-squared with n-1 degrees of freedom.

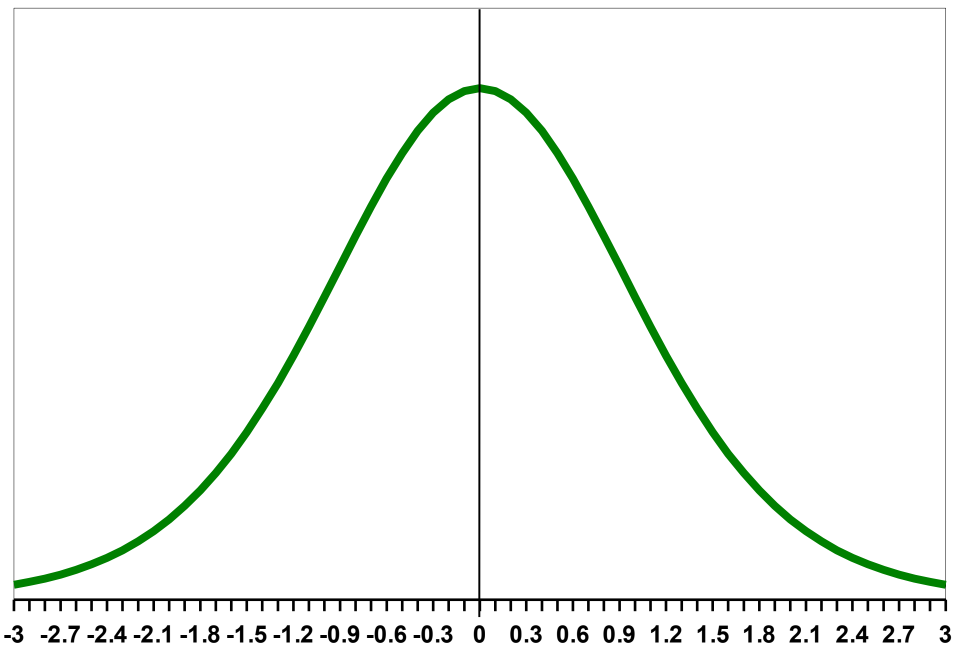
$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} \longleftarrow \textbf{SAMPLE VARIANCE}$$

Student's t distribution

X and Y INDEPENDIENT RVs $X \approx N(0,1)$ $Y \approx \chi^2_n$

Then $$t = \frac{X}{\sqrt{\frac{Y}{n}}} \approx t_n$$

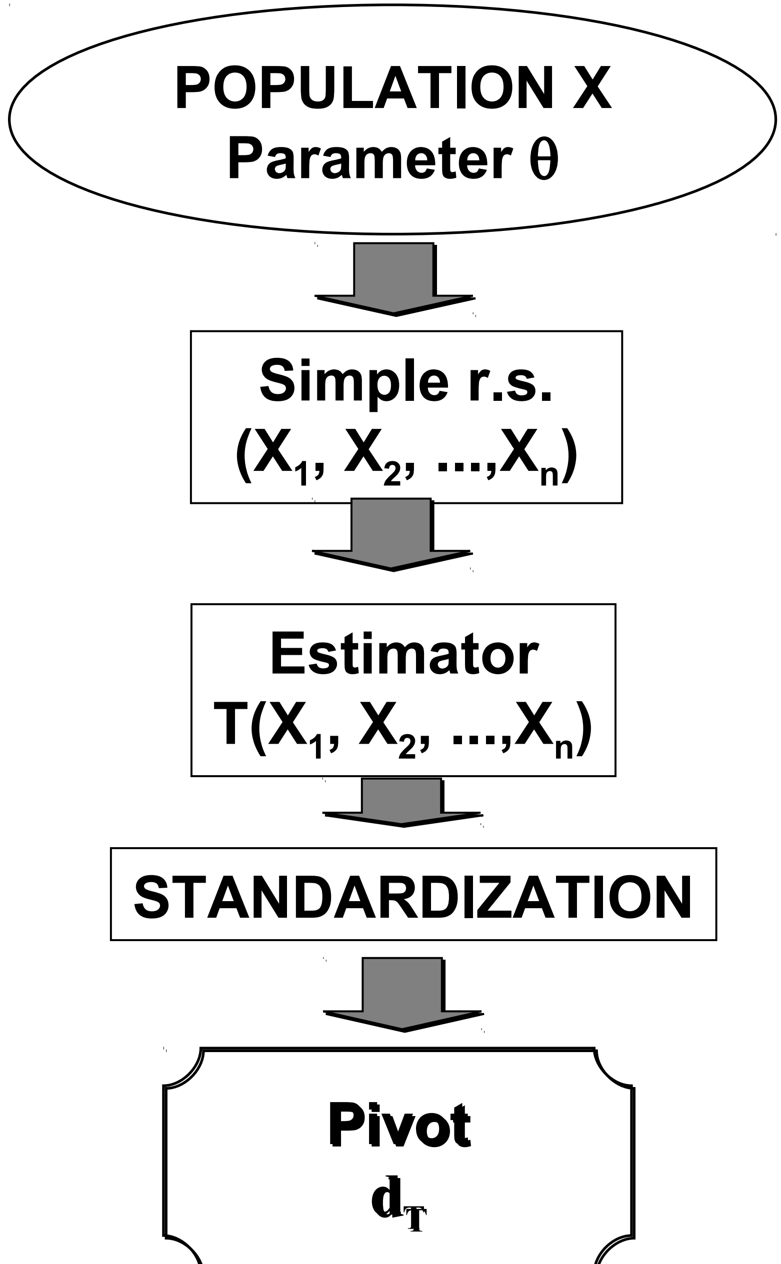
RV distributed as a Student's t with n freedom degrees.



UNIT 7: INFERENTIAL TOOLS. DISTRIBUTIONS ASSOCIATED WITH SAMPLING

7.2.- Inferential processes and relevant distributions.

Pivotal statistics associated with inferential processes



Pivotal statistics

- The estimator T summarizes the information that the sample contains on the unknown parameter θ.

- The pivot is obtained by standardizing the estimator. A pivot should have a known probability distribution, not depending on unknown parameters.

Estimator for the population mean

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

SAMPLE MEAN

$$E\left(\bar{X}\right) = \mu \qquad Var\left(\bar{X}\right) = \frac{\sigma^2}{n} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Unbiased estimator

Standard error

- If X≈N(μ,σ) then $\bar{X} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- If the distribution of X is unknown and n large $\bar{X} \to N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Inference on the population mean

ESTIMATOR $\longrightarrow$ $\bar{X}$

$$E\left(\bar{X}\right)=\mu \qquad Var\left(\bar{X}\right)=\frac{\sigma^2}{n}$$

PIVOT

$$\boxed{d_{\bar{X}}=\frac{\bar{X}-E\left(\bar{X}\right)}{\sqrt{Var\left(\bar{X}\right)}}=\frac{\bar{X}-\mu}{\sigma\sqrt{n}}}$$

Inference on the population mean

What is the probability distribution of the pivot?

NORMAL POPULATION
σ known

Reproductivity of normal model

NON-NORMAL POPULATION
σ known and n large

Central Limit Theorem

N(0,1)

$$d_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Unknown

NORMAL POPULATION
σ unknown

$$d_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

t_{n-1}

Estimator for the population proportion

$(X_1, \ldots, X_n)$ a s.r.s. with X_i being Bernoulli distributed:

$X_i=1$	If the characteristic under study is present.	$P(X_i=1)=p$
$X_i=0$	Otherwise.	$P(X_i=0)=1-p$

$$\hat{p}=\frac{X}{n}$$

SAMPLE PROPORTION

X: "Number of elements in the sample that have the characteristic of interest".

$$X \approx B(n,p)$$

$$E(\hat{p})=\frac{E(X)}{n}=\frac{np}{n}=p$$

Unbiased estimator

$$Var(\hat{p})=\frac{Var(X)}{n^2}=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}$$

Inference on the population proportion

PIVOT

$$d_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

n large

N(0,1)

Central Limit Theorem

$$d_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}}$$

p unknown

Estimator for the population variance

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

Plug-in estimator

$$E\left(S_n^2\right) = \frac{n-1}{n}\sigma^2$$

Biased estimator

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

SAMPLE VARIANCE

$$E(S^2) = E\left(\frac{n}{(n-1)} S_n^2\right) = \frac{n}{(n-1)} E(S_n^2) = \sigma^2$$

Unbiased estimator

Inference on the population variance

PIVOT

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

NORMAL POPULATION

Fisher Theorem

What parameters are we interested in?

Mean	Proportion	Variance
μ	p	σ^2

What estimators to use?

Sample mean	Sample proportion	Sample variance
$\bar{X}$	$\hat{p}$	S^2
Normal Student's t	Normal	Chi-squared

What probability distributions are relevant?