ALGES
ADVANCED LABORATORY FOR
GEOSTATISTICAL SUPERCOMPUTING

Lab 4: Comparative statistics

Statistical and Geostatistical Data Analysis.

Assistant Professor: Fabián Soto F.
Chair Professor: Xavier Emery

Ingeniería de Minas
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Comparative Test

Student, Fisher y ANOVA

ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Comparative test:

- Student test: Compares two means
- Fisher test: Compares two variances
- ANOVA: Compares multiple means

Student test:

- Data distribution may not be exactly Gaussian but the tests are ROBUST.

There are 4 types:

- Compares the mean against a specific value
- Compares 2 means with equal variance
- Compares 2 means with different variances
- Compares 2 means for paired data

Student test

Compares the mean against a specific value

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

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

Degrees of freedom: $n - 1$

Student test

Compares means of populations
with the same variance

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$\boldsymbol{H_0 : \mu = \mu_1 = \mu_2}$$
$$\boldsymbol{H_1 : \mu_1 \neq \mu_2}$$

$$T = \frac{\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2}{\boldsymbol{S}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Degrees of freedom $n_1 + n_2 - 2$

Student test

Compares means of populations with different variances

$$\boldsymbol{H_0: \mu = \mu_1 = \mu_2}$$
$$\boldsymbol{H_1: \mu_1 \neq \mu_2}$$

$$S = \sqrt{\frac{(n_1 \; - 1)S_1^2 + (n_2 \; - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$T = \frac{\bar{\boldsymbol{X}}_{\mathbf{1}} - \bar{\boldsymbol{X}}_{\mathbf{2}}}{\boldsymbol{S}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Degrees of freedom:

$$\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^{\mathbf{2}}}{\left(\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}\right)}$$

Student test

Compares means for paired data

Se define $\boldsymbol{D = X_1 - X_2}$

$$\boldsymbol{H_0: \mu = \mu_0}$$
$$\boldsymbol{H_1: \mu \neq \mu_0}$$

$$\boldsymbol{T = \frac{\bar{D}}{S/\sqrt{n}}}$$

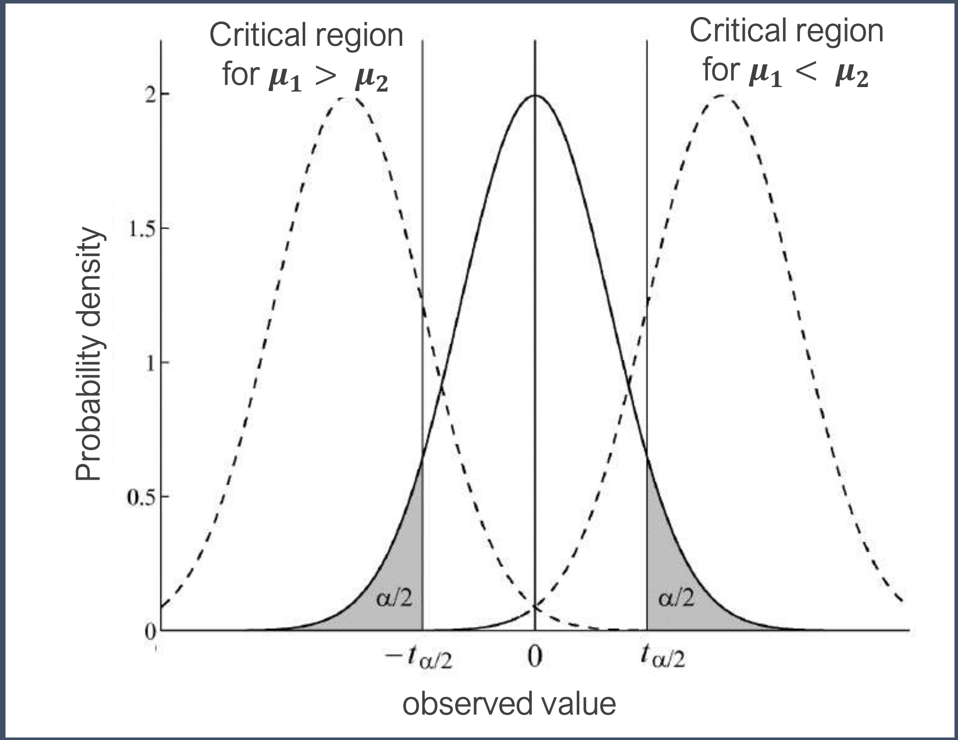
Degrees of freedom: $\boldsymbol{n - 1}$

Student test

Depending on the alternative hypothesis chosen, a one-way test or two-way test will be used.

Alternative hypothesis : $\boldsymbol{H_1}: \boldsymbol{\mu_1} > \boldsymbol{\mu_2}$
(One-way test)

Alternative hypothesis : $\boldsymbol{H_1}: \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$
(Two-way test)



Fisher test

Compares two variances

Consider independent Gaussian variables with $S_1^2 \geq S_2^2$ variances.

$$\boldsymbol{H_0}: \sigma_1^2 = \sigma_2^2 = \sigma^2$$
$$\boldsymbol{H_1}: \sigma_1^2 \neq \sigma_2^2$$

$$F = \frac{S_1^2}{S_2^2}$$

Degrees of freedom $n_1 - 1\ y\ n_2 - 1$

Selecting the sample size

$$n \geq 2\left(\frac{S z_{\alpha}}{\delta}\right)^{2}$$

With

- z_{α} : Gaussian threshold (one-way or two-way, depending on the alternative hypothesis) for a significance level $1-\alpha$.
- S: expected standard deviation.
- δ : minimum difference that should be detected by the test.

Once the value of n has been calculated, one can replace z_{α} by the Student threshold with $2n-2$ degrees of freedom for the same significance level $1-\alpha$, then recalculate the value of n.

Problems

ALGES
ADVANCED LABORATORY FOR
GEOSTATISTICAL SUPERCOMPUTING

Problem 1:

- As a chief metallurgist, you receive the proposal of a seller who claims that sulfide minerals recovery in the flotation process could increase from 89% to 95%, thanks to a new reagent. The proposal seems interesting enough to make an evaluation. To this end, the plant metallurgist performs an experiment with the new reagent for 6 consecutive days and compares the results with those of the 20 days previous to the experiment. (File *"Reactivo planta de flotación.xls"*).

Average recovery:

- It is asked to test:

1. Whether or not the average recovery with the new reagent is equal to 95%.
2. Whether or not the recovery variances are the same with the two reagents.
3. Whether or not the average recoveries are the same with the two reagents.

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING

Problem 2:

- Using one hundred duplicated data, sent to two different laboratories, it is of interest to know whether or not there is a significant difference between the laboratories. (File*"Duplicados.xls"*).

Problem 3 (proposed):

- One has a set of diamond drill hole and blast hole data, with measurements of the total copper grade (in %). (File *"Leyes de cobre.xls"*)
- It is of interest to design a test in order to decide whether or not the two types of measurements have the same quality.