ALGES
ADVANCED LABORATORY FOR
GEOSTATISTICAL SUPERCOMPUTING

Lab 2: Inference

Statistical and Geostatistical Data Analysis.

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

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

Statistical inference

Summary

Concept summary

- Statistical Inference
- It has the objective of knowing the properties of a population through information observed on a sample level.

- Properties of an estimator
  - Optimality
  - Accuracy (unbiasedness)
  - Sufficiency
  - Consistency

Confidence Interval

- Let $\boldsymbol{X}$ be a Gaussian variable with unknown expectation $\boldsymbol{\mu}$ and known variance $\boldsymbol{\sigma^2}$.

$$Prob\left(-1.96 < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < 1.96\right) = 0.95$$

Probability interval

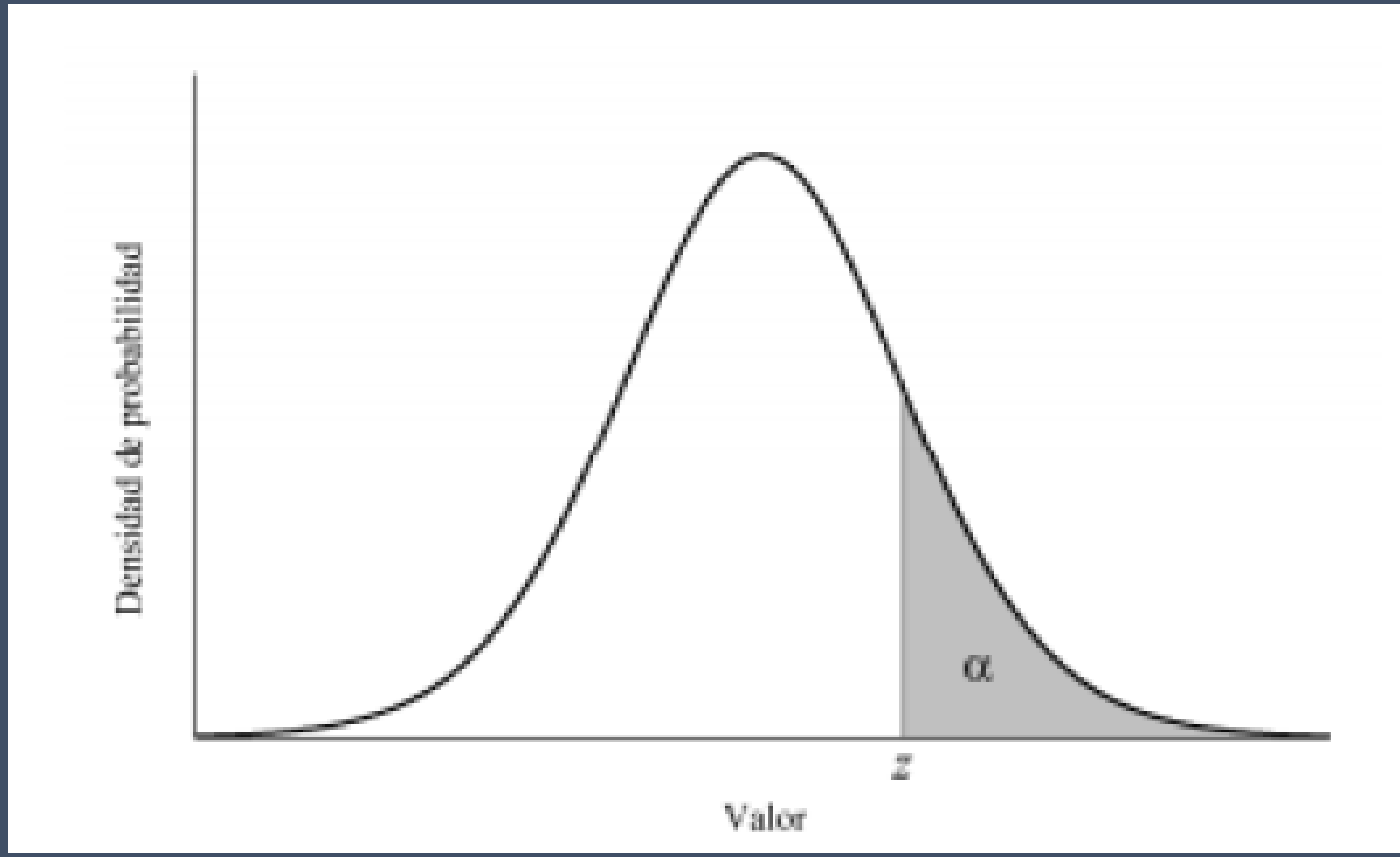
$$Prob\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Confidence interval

$$Prob\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Gaussian Distribution

- Standard Normal Distribution N(0,1)



- Values with probability α of being exceeded.

α	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0005
z	0.675	0.84	1.03	1.28	1.64	1.96	2.32	2.57	3.27

Confidence Interval

- Let $\boldsymbol{X}$ be a Gaussian variable with unknown expectation $\boldsymbol{\mu}$ and unknown variance $\boldsymbol{\sigma^2}$.

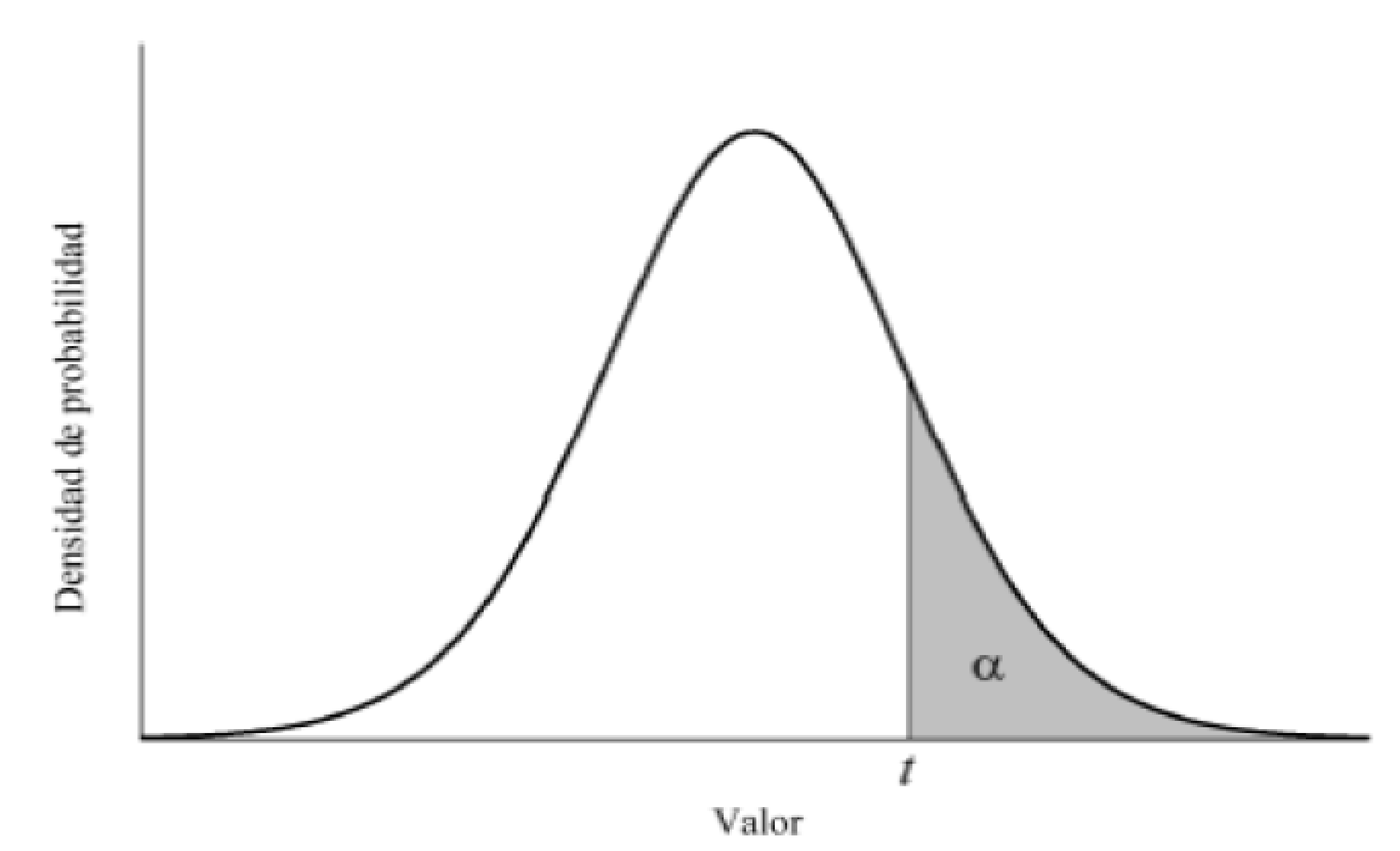
$$Prob\left(\bar{X} - t_{n-1,\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

- With $\boldsymbol{t_{n-1,\alpha/2}}$ value of the Student distribution with n – 1 degrees of freedom and probability $\boldsymbol{\alpha}$ of being exceeded.

Student Distribution

- Student distribution T_n with n degrees of freedom

- Values with probability α of being exceeded



α / n	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437

Estimation of a proportion

- $\mathbf{n}$ is the size of the sample
- If $\mathbf{n} < \mathbf{50}$

$$Prob\left(f - 1.96\sqrt{\frac{f(1-f)}{n}} < p < f + 1.96\sqrt{\frac{f(1-f)}{n}}\right) = 0.95$$

Estimation of a proportion

- If $\mathbf{n \geq 50}$

$$Prob(p_{min} < p < p_{max}) = 1 - \alpha$$

$$\left\{\begin{array}{l} p_{min} = \dfrac{n}{n + U_\alpha^2}\left(f + \dfrac{U_\alpha^2}{2n} - U_\alpha\sqrt{\dfrac{f(1+f)}{n} + \dfrac{U_\alpha^2}{4n^2}}\right) \\ p_{max} = \dfrac{n}{n + U_\alpha^2}\left(f + \dfrac{U_\alpha^2}{2n} + U_\alpha\sqrt{\dfrac{f(1+f)}{n} + \dfrac{U_\alpha^2}{4n^2}}\right) \end{array}\right. = 0.95$$

- Approximation when $\boldsymbol{n} \rightarrow \infty$

$$\left\{\begin{array}{l} p_{min} = f - U_\alpha\sqrt{\dfrac{f(1+f)}{n}} \\ p_{max} = f + U_\alpha\sqrt{\dfrac{f(1+f)}{n}} \end{array}\right.$$

Estimation of a proportion

- If $\mathbf{n \geq 50}$

$$Prob(p_{min} < p < p_{max}) = 1 - \alpha$$

$$\begin{cases} p_{min} = \dfrac{n}{n + U_\alpha^2}\left(f + \dfrac{U_\alpha^2}{2n} - U_\alpha\sqrt{\dfrac{f(1+f)}{n} + \dfrac{U_\alpha^2}{4n^2}}\right) \\ p_{max} = \dfrac{n}{n + U_\alpha^2}\left(f + \dfrac{U_\alpha^2}{2n} + U_\alpha\sqrt{\dfrac{f(1+f)}{n} + \dfrac{U_\alpha^2}{4n^2}}\right) \end{cases} = 0.95$$

1. Factor 1
2. Factor 2
3. Factor 3
4. Factor 4

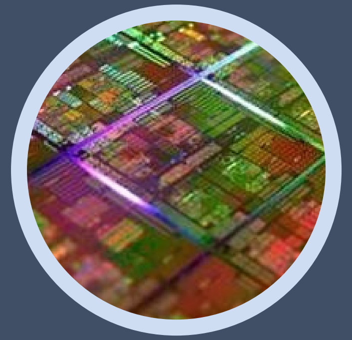
Problems

Problem 1:

- One has 9 core samples with total copper grade assays (in %):
  0.52 0.63 0.70 0.47 0.39 0.12 0.21 0.55 1.38
- Determine a confidence interval for the average copper grade of the population from which the samples were taken. Assume that the grade variance of the population is equal to 0.15.

Problem 1:

- One has 9 core samples with total copper grade assays (in %):
  0.52 0.63 0.70 0.47 0.39 0.12 0.21 0.55 1.38
- Determine a confidence interval for the average copper grade of the population from which the samples were taken. Assume that the grade variance of the population is equal to 0.15.



Problem 2:

- The same previous exercise, but in this case assuming that the grade variance is unknown.

Problem 1:

- One has 9 core samples with total copper grade assays (in %):
  0.52 0.63 0.70 0.47 0.39 0.12 0.21 0.55 1.38
- Determine a confidence interval for the average copper grade of the population from which the samples were taken. Assume that the grade variance of the population is equal to 0.15.

Problem 2:

- The same previous exercise, but in this case assuming that the grade variance is unknown.

Problem 3:

- One has 1000 draw points in a block cave mine. From 50 observed draw points, 13 had some failure.
- ¿How many draw points (in total) in the entire mine could be failing?

ALGES
ADVANCED LABORATORY FOR GEOSTATISTICAL SUPERCOMPUTING