Varianza y Prueba de Hipotesis

Varianza: La varianza se define como la dispersión de los datos alrededor de la media:

$$Var(X) = E(X - EX)^2$$

$x_1, x_2, \ldots, x_n$

$$E[x] = \frac{\sum_{i=1}^{m} x_i}{m} \qquad Var(x) = \frac{\sum (x_i - \mu)^2}{m} \; ; \quad \mu = \frac{\sum x_i}{m}$$

Covarianza: Sean X y Y variables aleatorias. La covarianza entre X e Y se define como:

$$\text{Cov}(X, Y) = E((X - EX)(Y - EY))$$

$$Cov(x, x) = E((x - EX)(x - Ex)) = E((X - EX)^2)$$

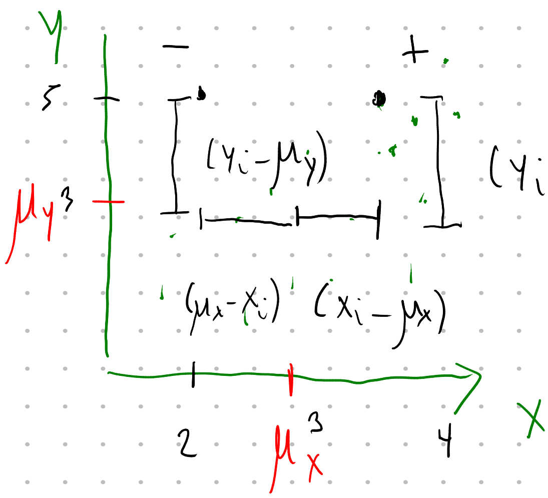
$$= Var(x)$$

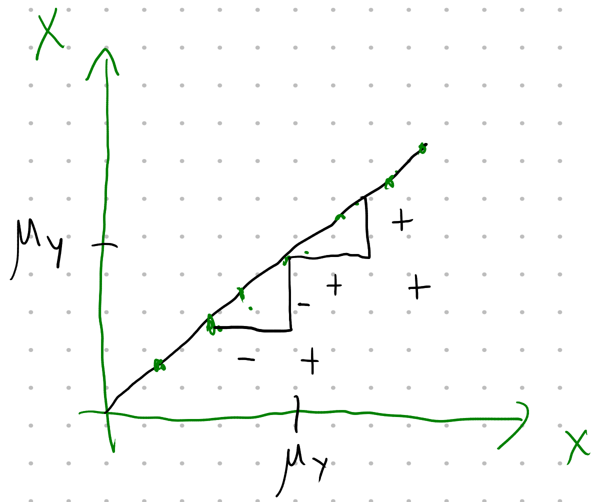
$$= E((X_i - \mu_x)(Y_i - \mu_y))$$

$$Cov(x, y) = \frac{\sum_{i=1}^{m} (X_i - \mu_x)(Y_i - \mu_y)}{m} \; ; \quad \mu_x = \frac{\sum x_i}{m}$$

$$\mu_y = \frac{\sum Y_i}{m}$$

$$Cov(X,Y) = \frac{\sum_{i=1}^{m}(X_i-\mu_x)(Y_i-\mu_y)}{m}$$

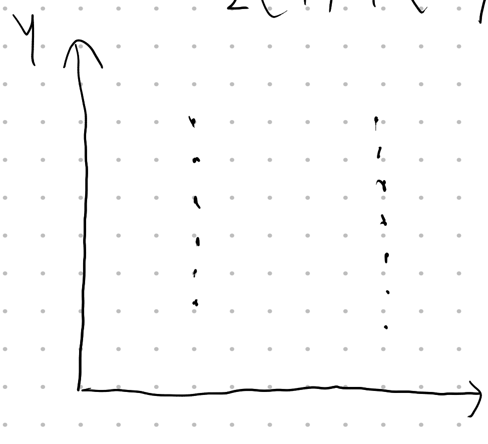




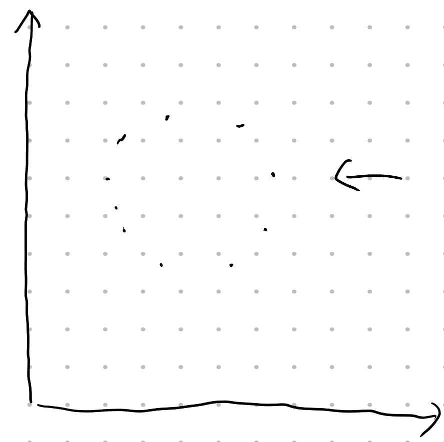
$$(Y_1-\mu_y)(X_1-\mu_x) + (X_2-\mu_x)(Y_2-\mu_y)$$

$$(5-3)(3-2) + (5-3)(4-3)$$

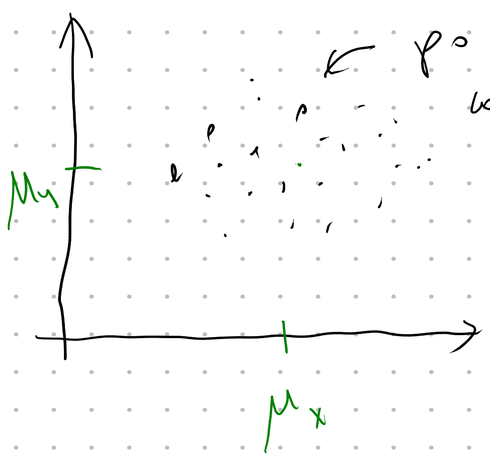
$$2(1)+(2)(-1)=0$$



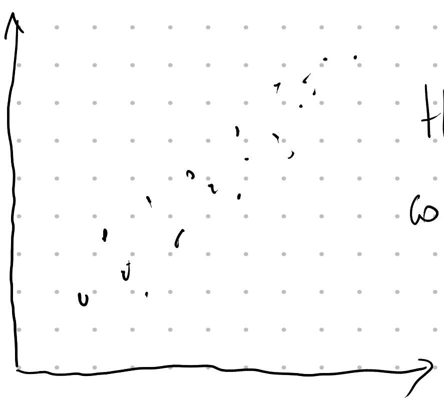
← 0 Cov



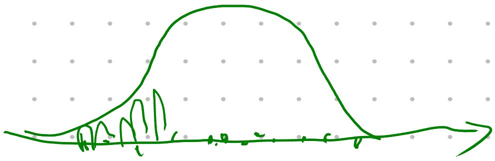
← 0 Cov

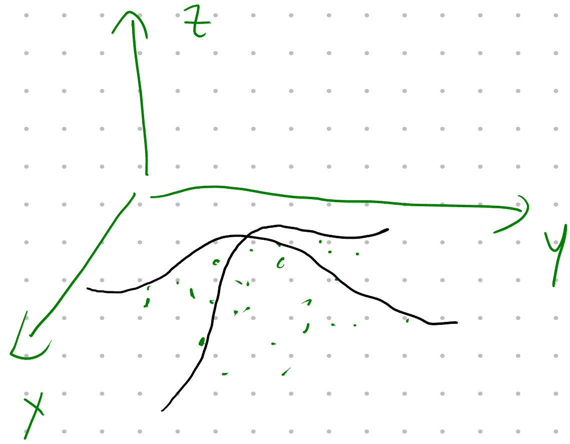


← poca covarianza



Harta covarianza





$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

No hay correlación $cov(x,y)=0 \rightarrow f(x,y)=f(x)f(y)$

$$f(x,y)=\frac{1}{\sqrt{2\pi\sigma_x^2}}\,\frac{1}{\sqrt{2\pi\sigma_y^2}}\, e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}\, e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}$$

$$e^{-\frac{1}{2\sigma_x^2}(x-\mu_x)^2-\frac{1}{2\sigma_x^2}(y-\mu_y)^2}$$

Si $cov(x,y)\neq 0 \qquad f(x,y)\neq f(x)f(y)$

Matriz de covarianza

$$\Sigma=\begin{bmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y)\end{bmatrix} \qquad \vec{\mu}=\begin{bmatrix}\mu_x \\ \mu_y\end{bmatrix}$$

The probability density function for `multivariate_normal` is

$$f(x)=\frac{1}{\sqrt{(2\pi)^k\det\Sigma}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right),$$

$$\vec{x}_i=\begin{bmatrix}x_i \\ y_i\end{bmatrix} \qquad (\vec{x}-\vec{\mu})_i=\begin{bmatrix}x_i-\mu_x \\ x_i-\mu_y\end{bmatrix} \qquad (x-y)^T_i=\begin{bmatrix}x_i-\mu_x & y_i-\mu_y\end{bmatrix}$$

$$\langle\psi|H|\psi\rangle$$

vector fila conjugado

vector columna

$$(x-\mu)^T\Sigma^{-1}(\vec{x}-\vec{\mu})=\sum\begin{bmatrix}x_i-\mu_x & y_i-\mu_y\end{bmatrix}\begin{bmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y)\end{bmatrix}\begin{bmatrix}x_i-\mu_x \\ y_i-\mu_y\end{bmatrix}$$

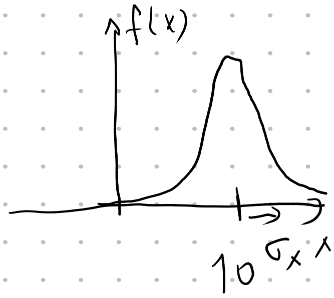
$$(1,2)(2,2)(2,1)=(1,1)$$

Teoria del error

Medición.

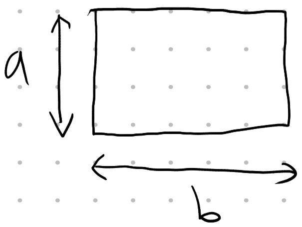
$X_1, X_2, \dots, X_n$ $\qquad$ σ_x: Desviación Est

Reporto los datos $\qquad$ σ_x^2: Varianza



$\bar{X} \pm \sigma_x$

$\bar{X} \pm \Delta X$ $\qquad$ $\sigma_x \approx \Delta X$



$\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}$

$\mu_a \pm \sigma_a \;,\; \mu_b \pm \sigma_b$

$\bar{a} \pm \Delta a \;,\; \bar{b} \pm \Delta b$

$c = a + b = \frac{P}{2}$ $\qquad$ $\mu_c = \frac{\sum a_i + b_i}{m} = \frac{\sum a_i}{m} + \frac{\sum b_i}{m}$

$\mu_c = \mu_a + \mu_b$

$\bar{c} = \bar{a} + \bar{b}$ $\qquad$ $\Delta c \approx \Delta a + \Delta b$

$\Delta c \approx \sqrt{(\Delta a)^2 + (\Delta b)^2}$

$$\sigma_c^2 = E[(C - E[X])^2] = \frac{1}{m}\sum_{i=1}^{m}(c_i - \mu_c)^2 \qquad \mu_c = \frac{\sum c_i}{m}$$

$c_i = a_i + b_i$ $\qquad$ $\mu_c = \mu_a + \mu_b$

$$\sigma_c^2 = \frac{1}{m}\sum(a_i + b_i - \mu_a - \mu_b)^2 = \frac{1}{m}\sum\left((a_i - \mu_a) + (b_i - \mu_b)\right)^2$$

$$= \frac{1}{m}\sum\Big[\underbrace{(a_i - \mu_a)^2}_{(1)} + \underbrace{(b_i - \mu_b)^2}_{(2)} + \underbrace{2(a_i - \mu_a)(b_i - \mu_b)}_{(3)}\Big]$$

$\sigma_c^2 = \sigma_a^2 + \sigma_b^2 + 2\,Cov(a,b) \leftrightarrow Var(A,B) = Var(A) + Var(B) + 2\,Cov(A,B)$

$\sigma_c^2 \approx \sigma_a^2 + \sigma_b^2 \leftarrow$ Para la teoria del error.

$$\boxed{\Delta c^2 \approx (\Delta a)^2 + (\Delta b)^2}$$

$c = a - b \Rightarrow \quad \sigma_c^2 = \sigma_a^2 + \sigma_b^2$

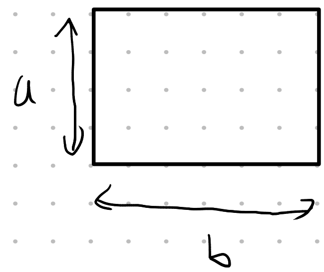
para la suma y resta $\{ \; (\Delta c)^2 \approx (\Delta a)^2 + (\Delta b)^2$

Regla del producto y del cociente.

$$\frac{(\Delta c)^2}{c^2} = \frac{(\Delta a)^2}{a^2} + \frac{(\Delta b)^2}{b^2}$$

$$\Delta c = c\sqrt{\left(\frac{\Delta a}{a}\right)^2 + \left(\frac{\Delta b}{b}\right)^2}$$

$\{a_1, \dots, a_n\}, \{b_1, \dots, b_n\}$



Area: $c = ab$

$$\mu_c = \frac{1}{m}\sum (x_i y_i)$$

$$\mu_a \mu_b = \left(\frac{1}{m}\sum x_i\right)\left(\frac{1}{m}\sum y_i\right)$$

$$= \frac{1}{m^2}\left(\sum x_i\right)\left(\sum y_i\right) = \frac{1}{m^2}(x_1 + \dots + x_n)(y_1 + \dots + y_n)$$

$$= \frac{1}{m^2}\sum x_{ij} = \frac{1}{m^2}\left[\sum_{i=1}^{m} x_i y_i + \sum_{i \neq j} x_i y_j\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} c_i$$

$$\mu_c = \mu_a \mu_b$$

$$Var(c) = Var(a, b) = Var(a) Var(b) + var(a) \mu_b^2 + var(b) \mu_a^2$$

$$\sigma_c^2 = \sigma_a^2 \sigma_b^2 + \sigma_a^2 \mu_b^2 + \sigma_b^2 \mu_a^2$$

$$(\Delta c)^2 = (\Delta a)^2 (\Delta b)^2 + (\Delta a)^2 \bar{b}^2 + (\Delta b)^2 \bar{a}^2$$

$$(\Delta a)^2 (\Delta b)^2 << (\Delta a)^2 \bar{b} + (\Delta b)^2 \bar{a}^2$$

$$\left[\sigma_c^2 \simeq \sigma_a^2 \mu_b^2 + \sigma_b^2 \mu_a^2 \right] \frac{1}{\mu_a^2 \mu_b^2}$$

$$\frac{\sigma_c^2}{\mu_a^2 \mu_b^2} \simeq \frac{\sigma_a^2}{\mu_a^2} + \frac{\sigma_b^2}{\mu_b^2}$$

$$\frac{\sigma_c^2}{\bar{c}^2} = \frac{\sigma_a^2}{\bar{a}^2} + \frac{\sigma_b^2}{\bar{b}^2}$$

Formula para el producto y el cociente.

$$x = x_0 + v_0 t$$

Datos: $\bar{x}_0, \sigma_0, \bar{v}, \sigma_v, \bar{t}, \sigma_t$

$$\left[\frac{\delta(v_0 t)}{(v_0 t)}\right]^2 = \frac{\sigma_v^2}{\bar{v}^2} + \frac{\sigma_t^2}{\bar{t}^2}$$

$$\delta(v_0 t) = v_0 t \sqrt{\left(\frac{\sigma_v}{\bar{v}}\right)^2 + \left(\frac{\sigma_t}{\bar{t}}\right)^2}$$

$$(\delta x)^2 = (\delta x_0)^2 + \delta(v_0 t)^2$$

$$\sigma_x = \sqrt{(\sigma_{x_0})^2 + (v_0 t)^2\left[\left(\frac{\sigma_v}{\bar{v}}\right)^2 + \left(\frac{\sigma_t}{\bar{t}}\right)^2\right]}$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X,Y)$$

$$X_1, X_2, X_3, \ldots, X_n$$

$$\mathrm{Var}\left(\sum X_i\right) = \sum_{i=1}^{m} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Var}(X+Y+Z) = \mathrm{Var}(X) + \mathrm{Var}(Y+Z) + \mathrm{Cov}(X, Y+Z)$$

$$\sigma_{\bar{X}}^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \cdots + \sigma_{x_n}^2$$

$$f(x) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right),$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} & & \\ & \Sigma^{-1} & \\ & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\Sigma = U \Lambda U^{+}$$

$$= \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix}$$

$$\Sigma^{-1} = U \Lambda^{-1} U^{T} = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} 1/\lambda_1 & \\ & 1/\lambda_2 \end{bmatrix} \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix}$$

$$\Sigma^{-1}\Sigma = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} 1/\lambda_1 & \\ & 1/\lambda_2 \end{bmatrix} \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix} \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix}$$

$$= \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \begin{bmatrix} 1/\lambda_1 & \\ & 1/\lambda_2 \end{bmatrix} I \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix}$$

$$\begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} I \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix} = I \qquad UU^{T} = U^{T}U = I$$

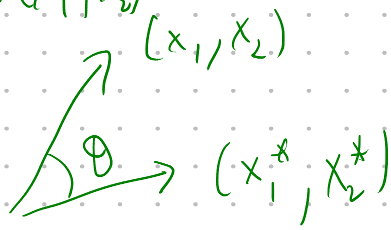
$$\Sigma^{-1} = U \Lambda U^{T}$$

$$U^{T} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$$

$$\Sigma^{-1} = [x_1 \;\; x_2]\, U \Lambda U^{T} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\rightarrow \operatorname{cov}(x_1^*, x_2^*)$$

$$[x_1^* \;\; x_2^*]\, \Lambda \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{bmatrix}$$



- - - - - - - - - - - - - - - - - -

$$\operatorname{Var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y)$$

$$\operatorname{var}(X+Y) = \operatorname{Var}(X) + \operatorname{var}(Y)$$ Si X, Y no están correlacionadas

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y$$

$$\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{var}(X_1) + \cdots + \operatorname{Var}(X_n)$$

$$\sigma^2_{\Sigma x} = \sum_{i=1}^{m} \sigma^2_{x_i}$$

$$Var\left(\frac{X_1+\cdots+X_n}{N}\right)=\frac{1}{N^2}Var(X_1+\cdots+X_n)$$

$$Var(cX)=\frac{\sum(cx_i-c\mu)^2}{N}=c^2\frac{\sum(x_i-\mu)^2}{N}$$

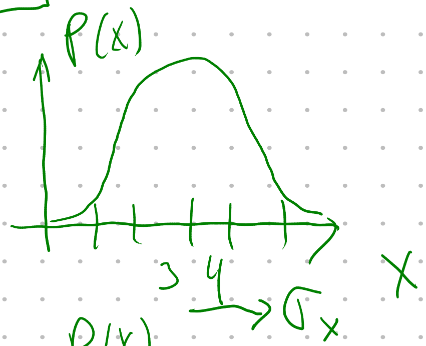
$$Var(cX)=c^2\,Var(X)$$

$$Var\left(\frac{\sum X_i}{N}\right)=\frac{1}{N^2}Var\left(\sum X_i\right)=\frac{1}{N^2}\sum Var(X_i)$$

$$\underbrace{X, X, X, \dots, X}_{N},\ \underbrace{\sigma_x, \sigma_x, \dots, \sigma_x}_{N}$$

$$Var\left(\frac{\sum X_i}{N}\right)=\frac{1}{N^2}N\,Var(X)=\frac{1}{N}Var(X)$$

$$\sigma_\mu^2=\frac{\sigma_x^2}{N} \Rightarrow \boxed{\sigma_\mu=\frac{\sigma_x}{\sqrt{N}}}$$

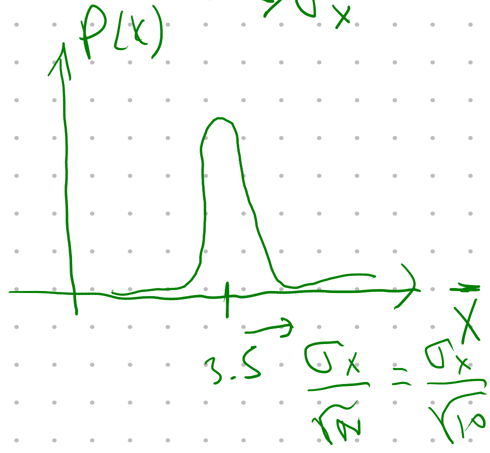


1, 2, 4, 2, 6,

Lanzo el dado 10 veces

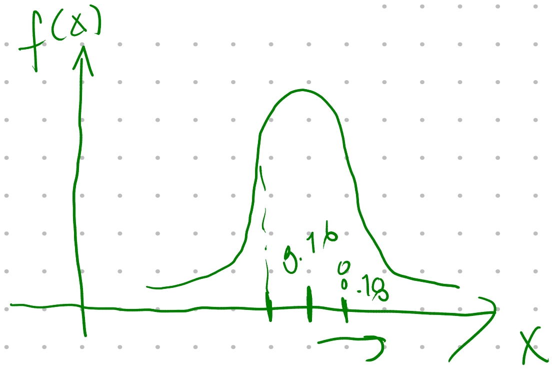
Hago diez veces 10 lanzamientos

$\bar{X}_1, \bar{X}_2, \dots, \bar{X}_{10}$



p-value: Considere una $\bar{X}$ el promedio de una muestra de n datos que viene de una distribución de μ desconocida y σ conocida. Se quiere probar la hipotesis de que $\bar{X}$ tiene viene de una distribución con promedio μ_0. Sea $\Phi(x)$ la función acumulativa de la función normal $\mathcal{N}(0,1)$. El p-value se define como,

$$2 * \Phi\left(\left|\frac{\bar{X}-\mu_o}{\sigma/\sqrt{n}}\right|\right) \tag{2}$$



$\sigma = 0.2$

1 Muestra: 8.18

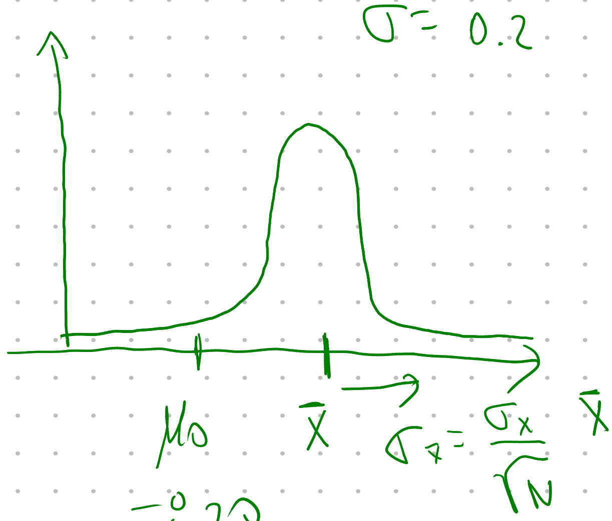
2 Muestra: 8.16

⋮

10 Muestras:

$\mu = \mu_0$

$\mu_0 - \bar{X} >> 0$ No son comparables

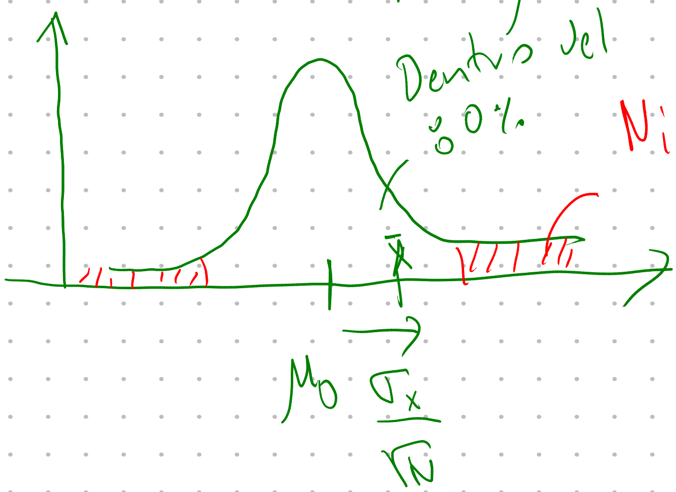


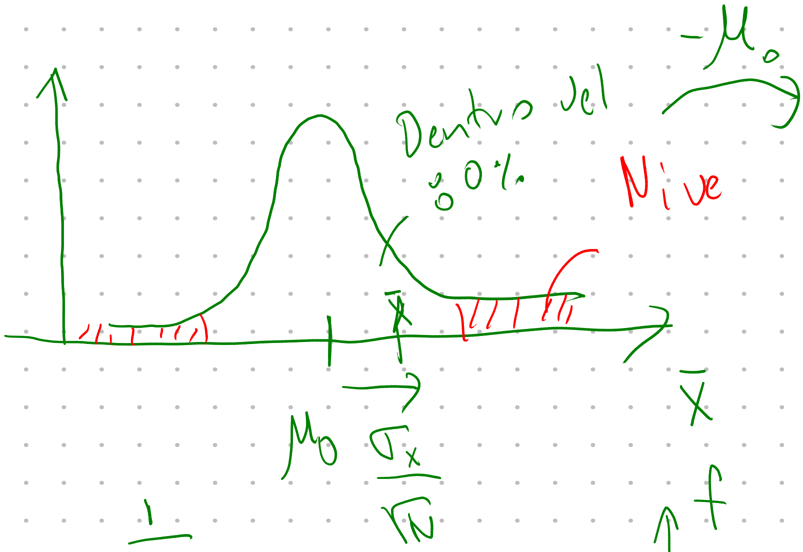
$0 < \bar{X} - \mu < 1$ Se puede decir que $\mu = \bar{X}$

$\mu = \bar{X}$

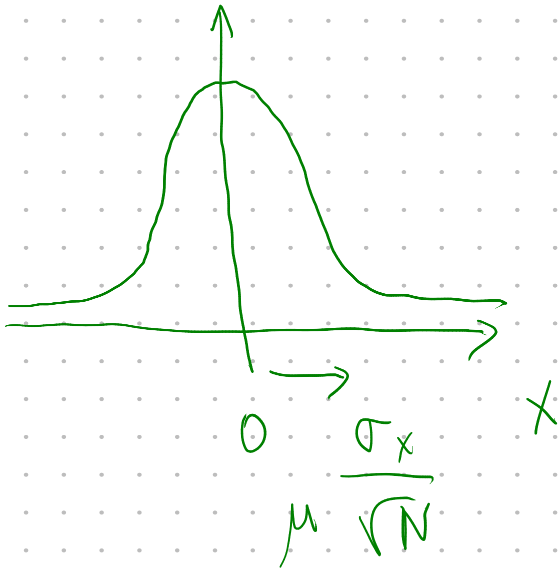
Dentro del 90%

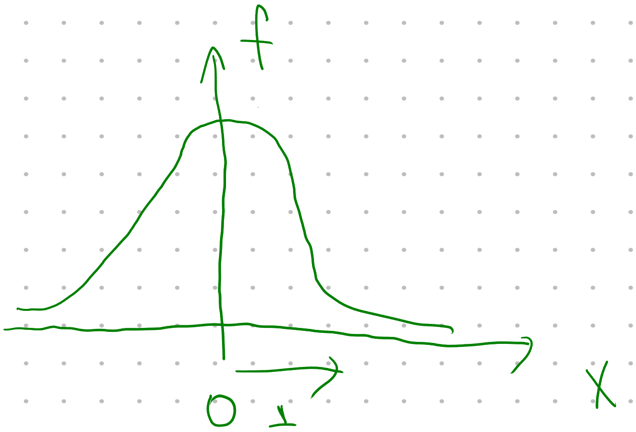
Nivel de significancia 10%.





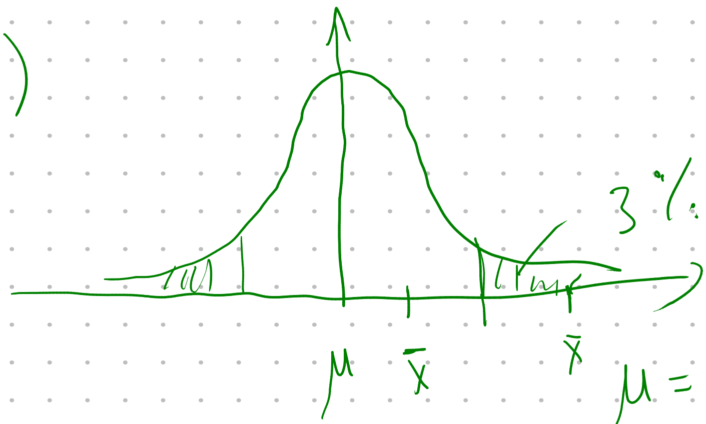
$\rightarrow \frac{1}{\frac{\sigma_x}{\sqrt{N}}}$





$$\bar{X} \rightarrow \bar{X} - \mu_0 \rightarrow \frac{\bar{X} - \mu_0}{\sigma_x / \sqrt{N}} = p\text{-value}$$

$N(0, 1)$



$\mu = \bar{x}$ $\mu_0 \neq \bar{x}$

Example

In a certain chemical process, it is very important that a particular solution that is to be used as a reactant have a pH of exactly 8.20. μ A method for determining pH that is available for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and with a standard deviation of .02. σ Suppose 10 independent measurements yielded the following pH values:

8.18	8.17
8.16	8.15
8.17	8.21
8.22	8.16
8.19	8.18

$$\rightarrow \frac{\sum x_i}{N}$$

(a) What conclusion can be drawn at the $\alpha = .10$ level of significance?
(b) What about at the $\alpha = .05$ level of significance?