Proprietà della varianza

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{i=n}(x_i - \mu)^2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}(x_i - \bar{x})^2$$

$$\mathrm{Var}(x)$$

$$\mathrm{Var}(x+a) = \mathrm{Var}(x)$$

$$\mathrm{Var}(ax) = a^2\mathrm{Var}(x)$$

$$\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y) \iff x \text{ e } y \text{ sono indipendenti}$$

$$\mathrm{Var}(x-y) = ?$$

$$\mathrm{Var}(x \cdot y) = ?$$

$$\mathrm{Var}(\frac{x}{y}) = ?$$

$$\mathrm{Var}(f(x)) = ?$$

Proprietà della varianza

$$\mathrm{Var}(x+a) = \mathrm{Var}(x)$$

$$x = \{10, 12, 9, 11\} \quad \bar{x} = 10.5$$

$$s_x^2 = \frac{(10-10.5)^2 + (12-10.5)^2 + (9-10.5)^2 + (11-10.5)^2}{3} = \frac{5}{3}$$

$$y = x + 4 \to y = \{(10+4), (12+4), (9+4), (11+4)\} \quad \bar{y} = (10.5+4) = 14.5$$

$$s_y^2 = \frac{[(y_1+4) - (\bar{y}+4)]^2 + ...}{3} = \frac{5}{3}$$

Proprietà della varianza

$$\mathrm{Var}(ax) = a^2\mathrm{Var}(x)$$

$$x = \{10, 12, 9, 11\} \quad \bar{x} = 10.5$$

$$s_x^2 = \frac{(10-10.5)^2 + (12-10.5)^2 + (9-10.5)^2 + (11-10.5)^2}{3} = \frac{5}{3}$$

$$a = 4 \quad y = ax \to y = \{(a10), (a12), (a9), (a11)\} \quad \bar{y} = (a10.5) = 42$$

$$s_y^2 = \frac{(ay_1 - a\bar{y})^2 + ...}{3} = \frac{[a(y_1 - \bar{y})]^2 + ...}{3} = \frac{a^2 5}{3} = \frac{16 \cdot 5}{3} = \frac{80}{3}$$

Proprietà della varianza

Intermezzo: ma perché dovremmo darci la pena di studiare come calcolare la varianza nel caso di somme, differenze, prodotti e divisioni?

ecco due esempi molto comuni:

- misuriamo la risposta di un sistema biologico ad un trattamento. Eseguiamo diverse repliche dell'esperimento e calcoliamo la media e l'errore standard. Contemporaneamente misuriamo anche il background (risposta del sistema in assenza di trattamento) medio e la sua dispersione attorno al valore medio. Siamo interessati a sottrarre tale valore alla risposta del sistema al trattamento.

- vogliamo comparare tra loro diversi set di dati, ad esempio la risposta di diverse linee cellulari ad uno stesso trattamento, e vogliamo usare uno standard di riferimento come controllo. Il modo naturale di procedere è normalizzare i valori ottenuti con ciascuna linea per il valore di controllo.

Spesso capita di vedere (in tesi e lavori scientifici in genere) che in queste operazioni vengano dimenticati gli errori. Una misura senza l'errore associato NON SERVE A NULLA!

Proprietà della varianza

Partiamo dal seguente caso: una funzione trasforma il valor medio di una variabile nel valor medio di un'altra variabile.

$$\bar{x} \xrightarrow{\quad f \quad} \bar{y} = f(\bar{x})$$

Allora:

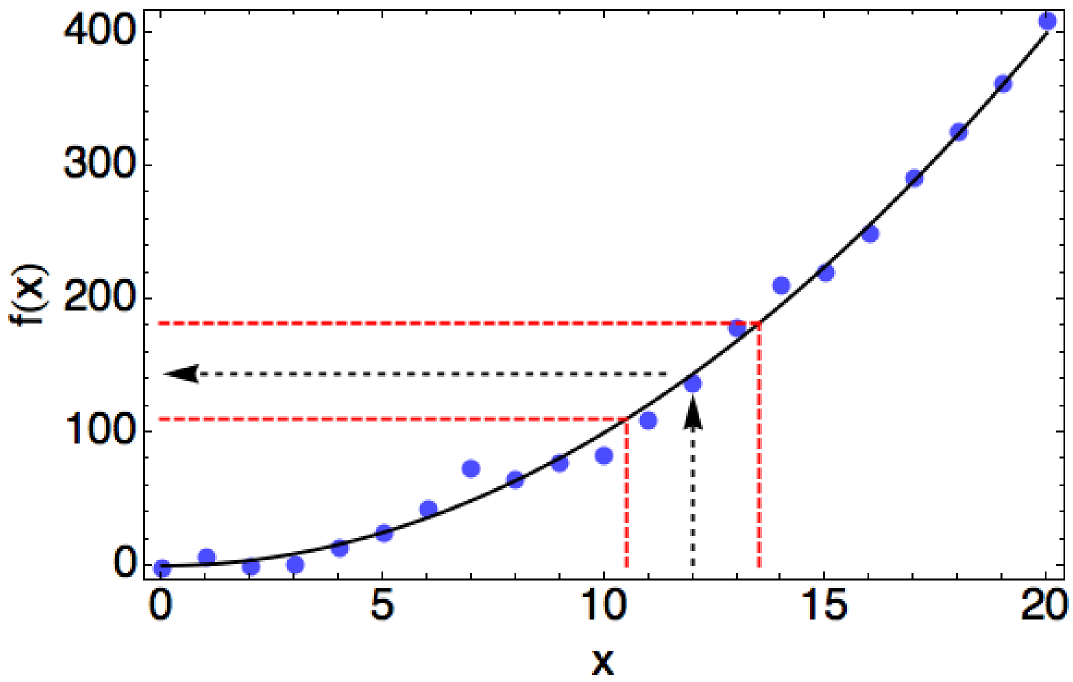
$$\sigma_{\bar{y}}^2 \simeq \left(\left. \frac{df(x)}{dx} \right|_{x=\bar{x}} \right)^2 \cdot \sigma_{\bar{x}}^2$$

derivata prima della funzione
calcolata nel valore medio di x

Proprietà della varianza

$$\bar{x} \pm \sigma_{\bar{x}} = 12 \pm 1.5$$

$$y = f(x) = x^2$$



$$\bar{y} = \bar{x}^2 = 12^2 = 144$$

$$\left.\frac{df(x)}{dx}\right|_{x=\bar{x}} = 2x|_{x=\bar{x}} = 2 \cdot 12 = 24$$

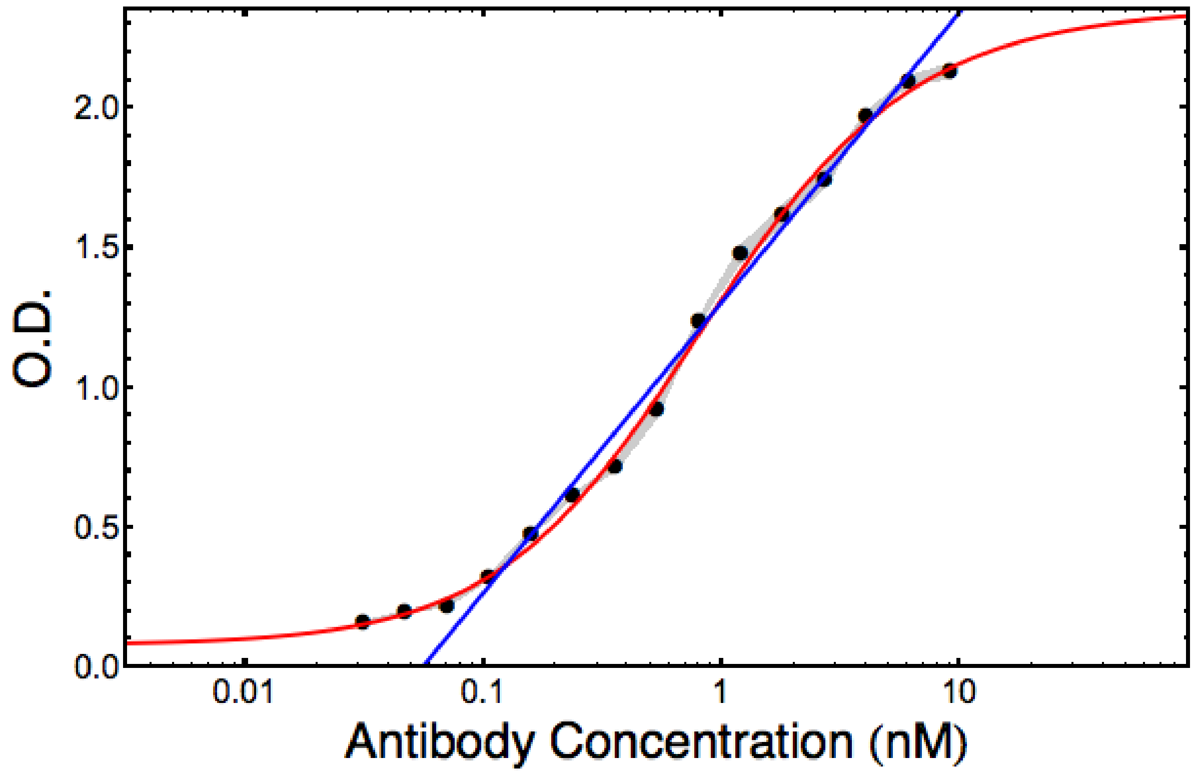
$$\sigma_{\bar{y}}^2 \simeq \left(\left.\frac{df(x)}{dx}\right|_{x=\bar{x}} \right)^2 \cdot \sigma_{\bar{x}}^2 = 24^2 \cdot 1.5^2 = 1296$$

$$\bar{y} \pm \sigma_{\bar{y}} = 144 \pm 36$$

$$\sigma_{\bar{y}} = \sqrt{\sigma_{\bar{y}}^2} = \sqrt{1296} = 36$$

Proprietà della varianza

In realtà nella pratica di laboratorio abbiamo spesso a che fare con il problema inverso:



es. reale di laboratorio

linea rossa: funzione logistica a 5 parametri

linea blu: approssimazione "lineare"

$$y = a_1 log(x) + a_2$$

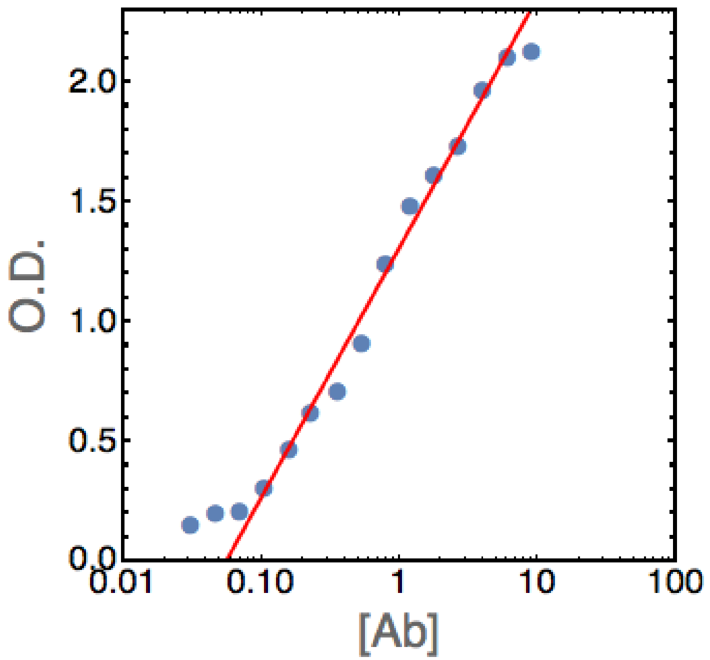
Proprietà della varianza

$$y = a_1 \log(x) + a_2 \quad a_1 = 0.453 \quad a_2 = 1.313$$

$$\bar{y} \pm \sigma_{\bar{y}} = 1.3 \pm 0.2 \ o.d.u.$$

$$\bar{x} = e^{\frac{y - a_2}{a_1}} = 0.97$$

$$\left.\frac{df(x)}{dx}\right|_{x=\bar{x}} = \left.\frac{a_1}{x}\right|_{x=\bar{x}} = \frac{0.453}{0.97} = 0.467$$



$$\sigma_{\bar{y}}^2 \simeq \left(\left.\frac{df(x)}{dx}\right|_{x=\bar{x}} \right)^2 \cdot \sigma_{\bar{x}}^2$$

$$0.2^2 \simeq 0.467^2 \cdot \sigma_{\bar{x}}^2 \rightarrow \sigma_{\bar{x}}^2 \simeq 0.184$$

$$\bar{x} \pm \sigma_{\bar{x}} = 0.97 \pm 0.43 \ nM$$

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} = \sqrt{0.184} = 0.43$$

Propagazione dell'errore: generale

Solo per variabili **indipendenti**

$$x_1, x_2 \ variabili \ casuali \ indipendenti$$

$$y = f(x_1, x_2)$$

$$\sigma_y^2 \simeq \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2$$

derivata parziale di f rispetto alla variabile x_1. Derivo la funzione considerando x_1 una variabile e x_2 costante

NB nelle prossime slide per semplicità $\simeq \rightarrow =$

Propagazione dell'errore: somma

x_1, x_2 *variabili casuali indipendenti*

$y = f(x_1, x_2) = a_1 x_1 + a_2 x_2 \quad a_1, a_2$ *costanti*

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2$$

$$\frac{\partial f}{\partial x_1} = a_1$$

$$\frac{\partial f}{\partial x_2} = a_2$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2 = a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2$$

$$se\ a_1^2 = a_2^2 = 1 \quad \sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 \quad e\ dunque\ \sigma_y = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}$$

Propagazione dell'errore: differenza

$$x_1, x_2 \; variabili \; casuali \; indipendenti$$

$$y = f(x_1, x_2) = a_1 x_1 - a_2 x_2 \quad a_1, a_2 \; costanti$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2$$

$$\frac{\partial f}{\partial x_1} = a_1$$

$$\frac{\partial f}{\partial x_2} = -a_2$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2 = a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2$$

$$se \; a_1^2 = a_2^2 = 1 \quad \sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 \quad e \; dunque \; \sigma_y = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2}$$

si noti l'implicazione!

Propagazione dell'errore: prodotto

$$x_1, x_2 \ variabili \ casuali \ indipendenti$$

$$y = f(x_1, x_2) = a_1 x_1 \cdot a_2 x_2 \quad a_1, a_2 \ costanti$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2$$

$$\frac{\partial f}{\partial x_1} = a_1 a_2 x_2$$

$$\frac{\partial f}{\partial x_2} = a_1 a_2 x_1$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2 = a_1^2 a_2^2 x_2^2 \sigma_{x_1}^2 + a_1^2 a_2^2 x_1^2 \sigma_{x_2}^2$$

$$se \ a_1^2 = a_2^2 = 1 \quad \sigma_y^2 = x_2^2 \sigma_{x_1}^2 + x_1^2 \sigma_{x_2}^2$$

Propagazione dell'errore: prodotto

$$se\ a_1^2 = a_2^2 = 1 \quad \sigma_y^2 = x_2^2\sigma_{x_1}^2 + x_1^2\sigma_{x_2}^2$$

trucco

$$\frac{\sigma_y^2}{x_1^2x_2^2} = \frac{x_2^2\sigma_{x_1}^2}{x_1^2x_2^2} + \frac{x_1^2\sigma_{x_2}^2}{x_1^2x_2^2} = \frac{\sigma_{x_1}^2}{x_1^2} + \frac{\sigma_{x_2}^2}{x_2^2} = \left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2$$

$$\sigma_y^2 = x_1^2x_2^2\left[\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2\right]$$

$$\sigma_y = x_1x_2\sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2}$$

Propagazione dell'errore: prodotto

$$y = x_1 x_2 \qquad \sigma_y = x_1 x_2 \sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2}$$

es.:

$$x_1 \pm \sigma_{x_1} = 10 \pm 2$$

$$x_2 \pm \sigma_{x_2} = 5 \pm 3$$

$$y = x_1 x_2 = 10 \cdot 5 = 50$$

$$\sigma_y = x_1 x_2 \sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2} = 50\sqrt{\left(\frac{2}{10}\right)^2 + \left(\frac{3}{5}\right)^2} = 10\sqrt{10} = 31.62$$

$$y \pm \sigma_y = 50 \pm 31.62$$

Propagazione dell'errore: divisione

x_1, x_2 *variabili casuali indipendenti*

$$y = f(x_1, x_2) = \frac{a_1 x_1}{a_2 x_2} \quad a_1, a_2 \; costanti$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2$$

$$\frac{\partial f}{\partial x_1} = \frac{a_1}{a_2 x_2} \qquad \frac{\partial f}{\partial x_2} = -\frac{a_1 x_1}{a_2 x_2^2}$$

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \cdot \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \cdot \sigma_{x_2}^2 = \left(\frac{a_1}{a_2 x_2}\right)^2 \sigma_{x_1}^2 + \left(-\frac{a_1 x_1}{a_2 x_2^2}\right)^2 \sigma_{x_2}^2$$

$$se \; a_1^2 = a_2^2 = 1 \quad \sigma_y^2 = \frac{1}{x_2^2}\sigma_{x_1}^2 + \frac{x_1^2}{x_2^4}\sigma_{x_2}^2$$

Propagazione dell'errore: divisione

$$se\ a_1^2 = a_2^2 = 1 \quad \sigma_y^2 = \frac{1}{x_2^2}\sigma_{x_1}^2 + \frac{x_1^2}{x_2^4}\sigma_{x_2}^2$$

trucco

$$\frac{x_2^2}{x_1^2}\sigma_y^2 = \frac{x_2^2}{x_1^2}\frac{1}{x_2^2}\sigma_{x_1}^2 + \frac{x_2^2}{x_1^2}\frac{x_1^2}{x_2^4}\sigma_{x_2}^2 = \frac{\sigma_{x_1}^2}{x_1^2} + \frac{\sigma_{x_2}^2}{x_2^2}$$

$$\sigma_y^2 = \frac{x_1^2}{x_2^2}\left[\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2\right]$$

$$\sigma_y = \frac{x_1}{x_2}\sqrt{\left[\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2\right]}$$

Propagazione dell'errore: divisione

$$\sigma_y = \frac{x_1}{x_2}\sqrt{\left[\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2\right]}$$

es.:

$$x_1 \pm \sigma_{x_1} = 10 \pm 2$$

$$x_2 \pm \sigma_{x_2} = 5 \pm 3$$

$$y = \frac{x_1}{x_2} = \frac{10}{5} = 2$$

$$\sigma_y = \frac{x_1}{x_2}\sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2} = 2\sqrt{\left(\frac{2}{10}\right)^2 + \left(\frac{3}{5}\right)^2} = 2\sqrt{\frac{2}{5}} = 1.26$$

$$y \pm \sigma_y = 2 \pm 1.26$$

Propagazione dell'errore

$$x_1 \pm \sigma_{x_1} = 10 \pm 2$$

$$x_2 \pm \sigma_{x_2} = 5 \pm 3$$

$$y = x_1 + x_2 \quad \sigma_y = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2} \rightarrow y \pm \sigma_y = 15 \pm 3.6$$

$$y = x_1 - x_2 \quad \sigma_y = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2} \rightarrow y \pm \sigma_y = 5 \pm 3.6$$

$$y = x_1 x_2 \qquad \sigma_y = x_1 x_2 \sqrt{\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2} \rightarrow y \pm \sigma_y = 50 \pm 31.62$$

$$y = \frac{x_1}{x_2} \quad \sigma_y = \frac{x_1}{x_2} \sqrt{\left[\left(\frac{\sigma_{x_1}}{x_1}\right)^2 + \left(\frac{\sigma_{x_2}}{x_2}\right)^2\right]} \rightarrow y \pm \sigma_y = 2 \pm 1.26$$

Propagazione dell'errore

$\frac{\sigma}{\bar{x}} \cdot 100 = CV$ coefficiente di variazione (percentuale)
o **errore relativo**

$x_1 \pm \sigma_{x_1} = 10 \pm 2 \quad \mathrm{CV} = 20\%$

$x_2 \pm \sigma_{x_2} = 5 \pm 3 \quad \mathrm{CV} = 60\%$

$$\text{somma} \quad y = x_1 + x_2 \rightarrow y \pm \sigma_y = 15 \pm 3.6 \quad \mathrm{CV} = 24\%$$

$$\text{differenza} \quad y = x_1 - x_2 \rightarrow y \pm \sigma_y = 5 \pm 3.6 \quad \mathrm{CV} = 72\%$$

$$\text{prodotto} \quad y = x_1 x_2 \rightarrow y \pm \sigma_y = 50 \pm 31.62 \quad \mathrm{CV} = 63.24\%$$

$$\text{divisione} \quad y = \frac{x_1}{x_2} \rightarrow y \pm \sigma_y = 2 \pm 1.26 \quad \mathrm{CV} = 63\%$$

Propagazione dell'errore

$$x_1, x_2, ..., x_i, ..., x_n \quad variabili\ aleatorie\ indipendenti$$

$$y = f(x_1, ..., x_n) = \frac{x_1 + x_2 + ... + x_n}{n}$$

$$se\ \sigma_1 = ... = \sigma_n = \sigma_x \quad allora \quad \sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_x^2 + ... + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_x^2$$

$$ma \quad \frac{\partial f}{\partial x_i} = \frac{1}{n} \quad e\ dunque \quad \sigma_y^2 = \underbrace{\frac{1}{n^2}\sigma_x^2 + ... + \frac{1}{n^2}\sigma_x^2}_{n\ volte}$$

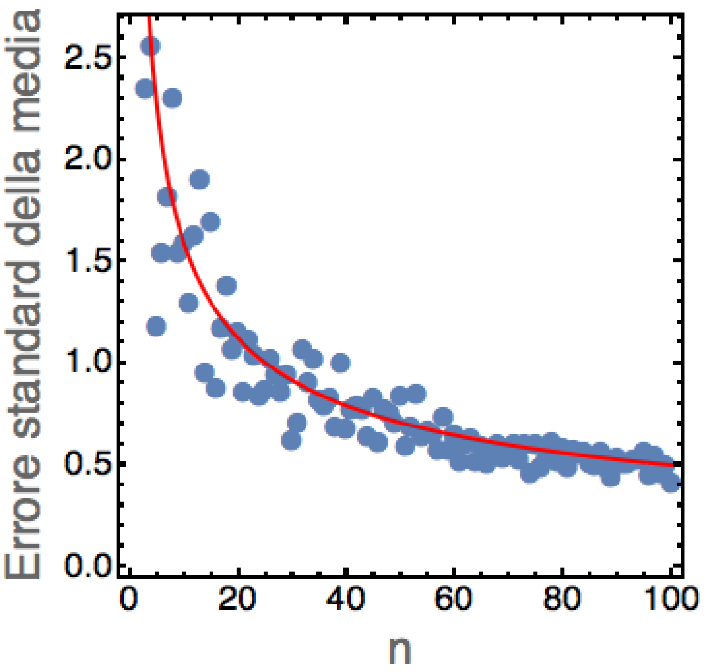
$$\sigma_y^2 = \frac{n}{n^2}\sigma_x^2 = \frac{\sigma_x^2}{n} \quad e \quad \sigma_y = \frac{\sigma_x}{\sqrt{n}}$$

implicazioni

Propagazione dell'errore

$$\sigma_y = \frac{\sigma_x}{\sqrt{n}}$$ implicazioni

prendo campioni casuali di n=3, 4,5,...,100 repliche da $\mathcal{N}(30, 5)$. Calcolo la media e l'errore standard della media:



Implicazioni: pianificazione

...ulteriori implicazioni.

Abbiamo visto che: $N(\mu, \sigma) \;\rightarrow\; x_1, \ldots, x_n \begin{cases} \bar{x} \\ \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} \end{cases}$

dunque, se $y = \bar{x}_1 - \bar{x}_2$

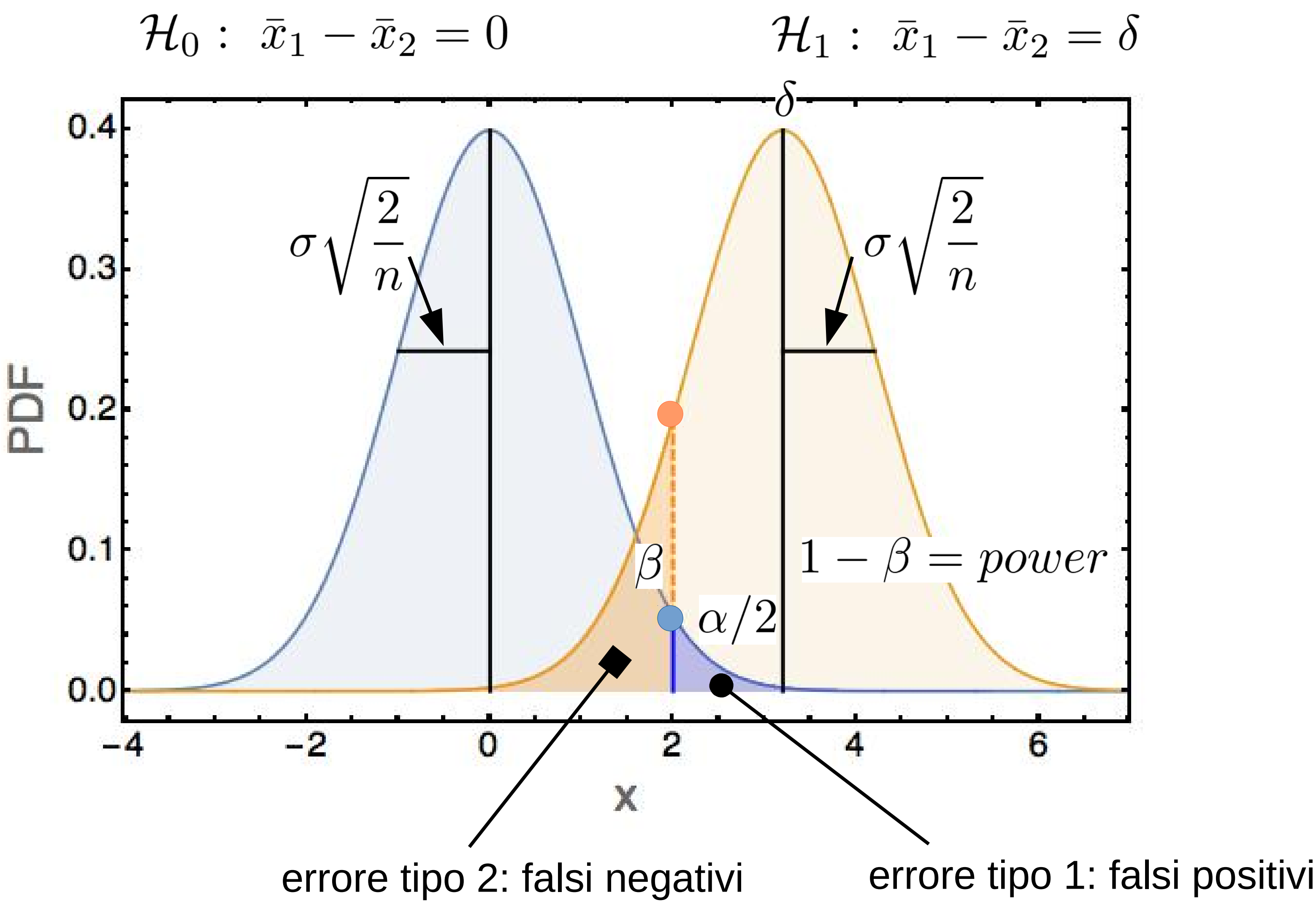
abbiamo che $\sigma_y = \sqrt{\dfrac{\sigma^2_{\bar{x}_1}}{n_1} + \dfrac{\sigma^2_{\bar{x}_2}}{n_2}}$ e per semplicità $\left.\begin{array}{c} \sigma^2_{\bar{x}_1} = \sigma^2_{\bar{x}_2} \\ n_1 = n_2 \end{array}\right\} \sigma_y = \sqrt{2\dfrac{\sigma^2}{n}}$

e ci aspettiamo che: $\dfrac{y}{\sigma_y} \;\rightarrow\; \mathcal{N}(0,1)$

$$\left.\begin{array}{l} \mathcal{H}_0 : \;\bar{x}_1 - \bar{x}_2 = 0 \\ \mathcal{H}_1 : \;\bar{x}_1 - \bar{x}_2 \neq 0 \end{array}\right\} \frac{y}{\sigma_y} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{2\frac{\sigma^2}{n}}}$$

dunque il test di significatività dipende da quanto distano tra loro le medie, da quanto è grande la dev. standard e **dalla dimensione dei campioni**

Implicazioni: pianificazione



Implicazioni: pianificazione

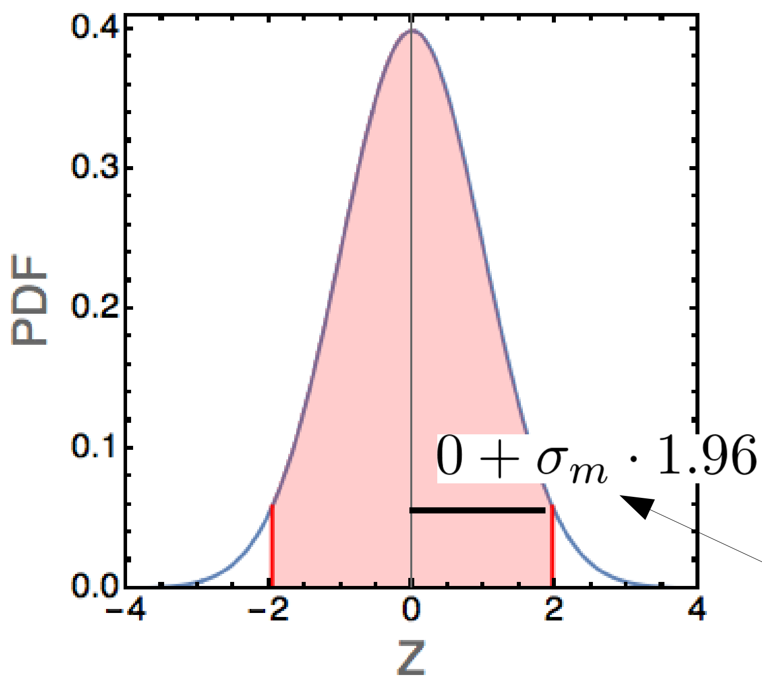
Ricordiamo che: $$Z = \frac{m-\mu}{\sigma_m}$$

$m - \mu =$ **scostamento della media del campione dalla media della popolazione**

dunque: $m - \mu = \sigma_m \cdot Z$

ma Z segue una distribuzione normale standard e quindi possiamo calcolare (o trovare nelle tabelle opportune) tutti i valori di probabilità che desideriamo per ogni valore di Z. Ad es. possiamo calcolare i valori critici di Z che definiscono l'intervallo all'interno del quale ricade il 95% dei valori di Z:

$$-1.96 \leq Z_{0.95} \leq 1.96$$



$$0 + \sigma_m \cdot 1.96$$

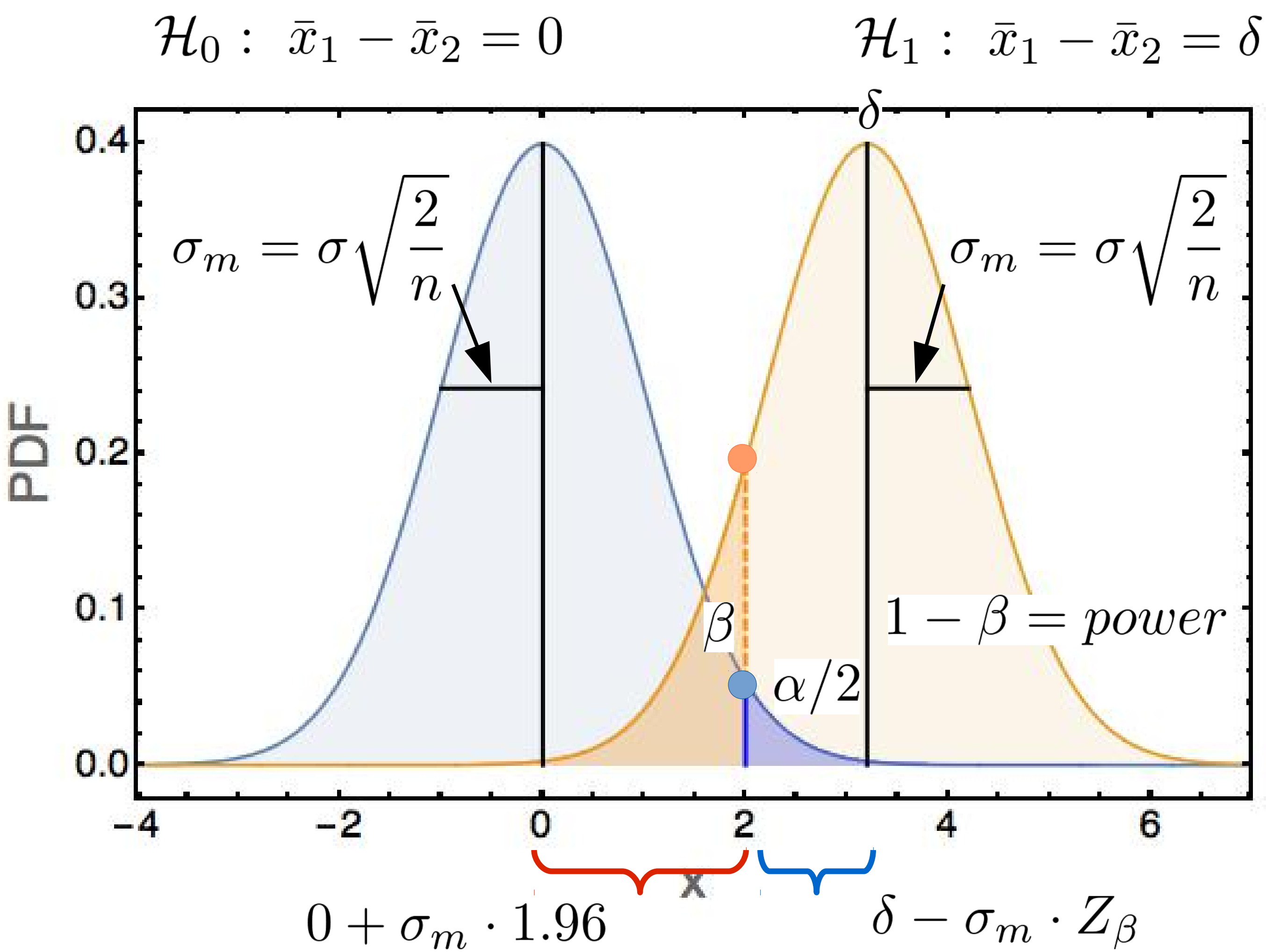
Dunque

$$-(\sigma_m \cdot 1.96) \leq m - \mu \leq \sigma_m \cdot 1.96$$

è l'intervallo che ci dice quanto la media campionaria si discosta al più del 95% dalla media vera:

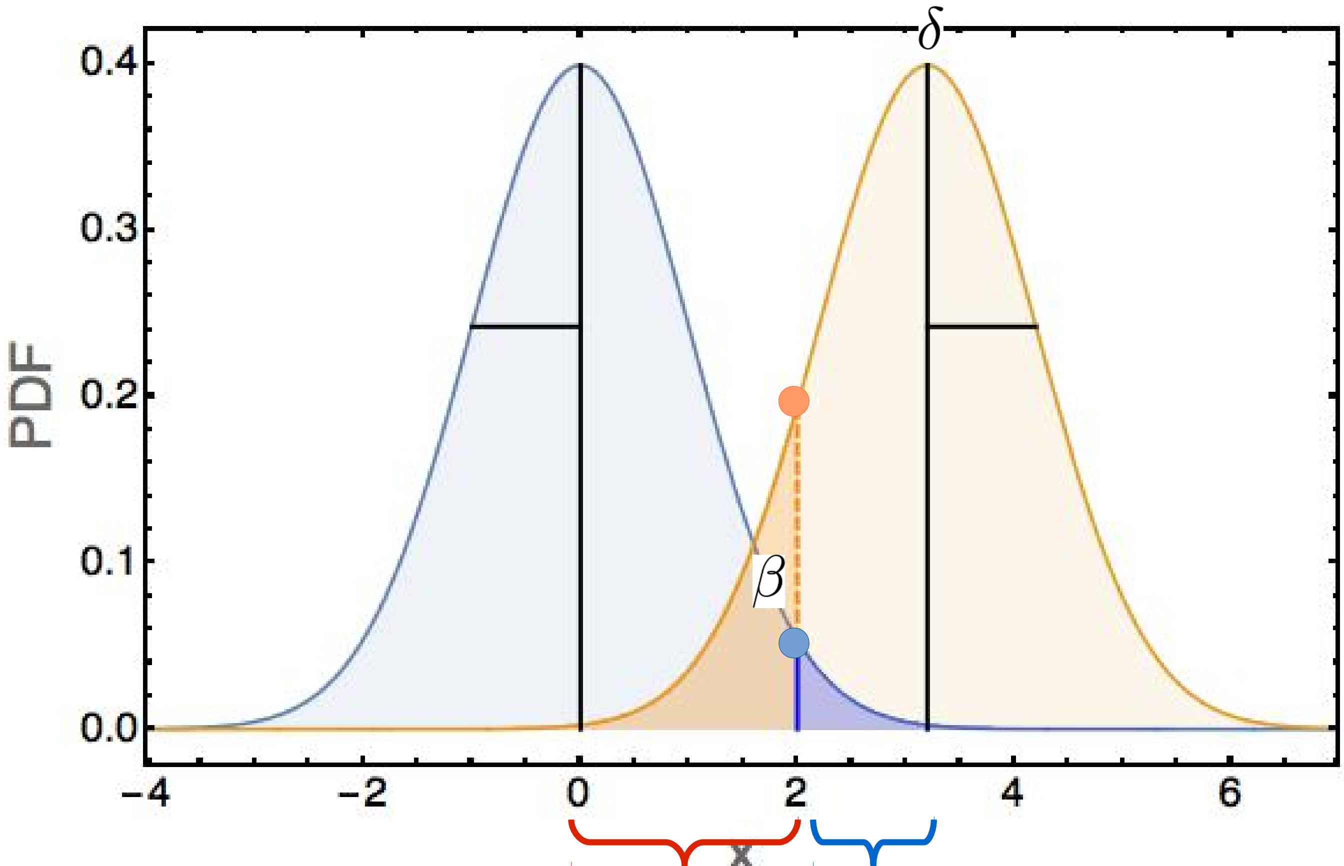
<u>**intervallo di confidenza (CI) al 95%**</u>

Implicazioni: pianificazione



dunque dobbiamo scegliere un valore critico di Z in modo che l'area β assuma un certo valore che noi vogliamo accettare. Che significa?

Implicazioni: pianificazione



$$0 + \sigma_m \cdot 1.96 \qquad \delta - \sigma_m \cdot Z_\beta$$

Che significa? Che accettiamo di rischiare di ottenere falsi negativi con una certa probabilità. Per convenzione si fissa tale valore al 20%, e il valore critico di Z è:

$$Z_{0.2} = 0.84 \quad P[Z \leq -0.84] = 0.2 \quad P[Z \geq 0.84] = 0.2$$

Implicazioni: pianificazione

$$0 + \sigma_m \cdot 1.96$$

Accettiamo di avere solo il 5% di falsi positivi.

$$\delta - \sigma_m \cdot 0.84$$

Accettiamo di avere il 20% di falsi negativi, cioè siamo in grado di discriminare i veri positivi nel 80%. Siamo in grado di rifiutare l'ipotesi nulla H_0 con una probabilità P=0.8: POTENZA

$$0 + \sigma_m \cdot 1.96 = \delta - \sigma_m \cdot 0.84$$

$$\sigma_m (1.96 + 0.84) = \delta$$

$$\sigma \sqrt{\frac{2(1.96 + 0.84)^2}{n}} = \delta$$

$$n = \frac{\sigma^2 2(1.96 + 0.84)^2}{\delta^2} \simeq \frac{16}{\left(\frac{\bar{x}_1 - \bar{x}_2}{\sigma}\right)^2}$$

Implicazioni: pianificazione

$$n = \frac{16}{\left(\frac{\bar{x}_1 - \bar{x}_2}{\sigma}\right)^2} = \frac{16}{\Delta^2}$$

Equazione di Leher

differenza standardizzata

es. il QI (scemenza) ha una distribuzione normale con media pari a 100 e deviazione standard pari a 20. Ci attendiamo che una popolazione sottoposta ad un certo trattamento si discosti dal valore medio di 10 punti, ad esempio un trattamento rincretinente ridurrebbe il QI medio da 100 a 90 punti. Quanti soggetti dobbiamo valutare accettando di avere il 5% di falsi positivi e una potenza del 80%, cioè di avere una probabilità del 80% di mettere in evidenza l'effetto (o di rifiutare l'ipotesi nulla nel 80% dei casi)?

$$n = \frac{16}{\left(\frac{\bar{x}_1 - \bar{x}_2}{\sigma}\right)^2} = \frac{16}{\left(\frac{100-90}{20}\right)^2} = 64 \; soggetti$$

Implicazioni: pianificazione

ovviamente posso fare i conti per avere una potenza del test più elevata, ma si noti che aumenta anche il numero di soggetti

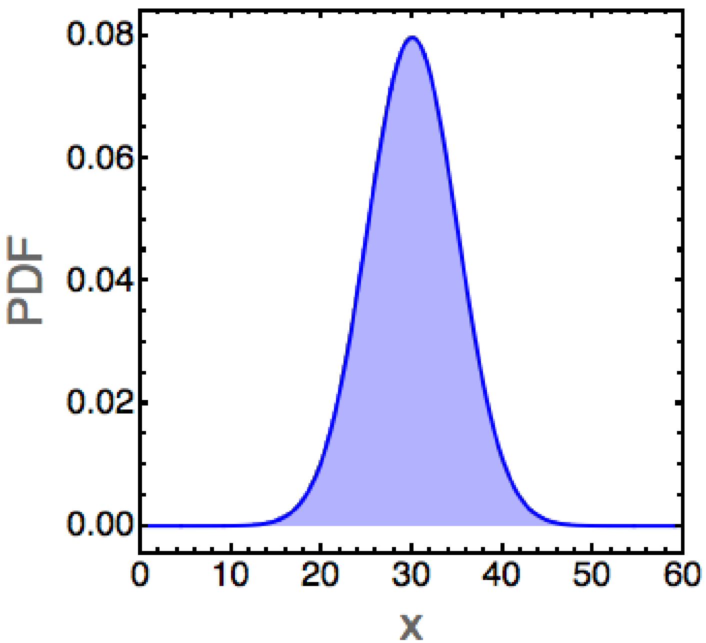
$$n = \frac{\sigma^2 \cdot 2 \cdot (1.96 + 0.84)^2}{\delta^2}$$

oppure rigirare l'equazione per ottenere una stima di ciascun parametro in funzione del valore ipotizzato per gli altri

$$n = \frac{2 \cdot (1.96 + 0.84)^2}{\left(\frac{\bar{x}_1 - \bar{x}_2}{\sigma}\right)^2}$$

Varianza: proprietà

Riprendiamo questo esperimento numerico:



$$\mu = 30 \quad a.u.$$
$$\sigma = 5 \quad a.u.$$

estrazione di campioni e
calcolo delle medie per
ogni campione

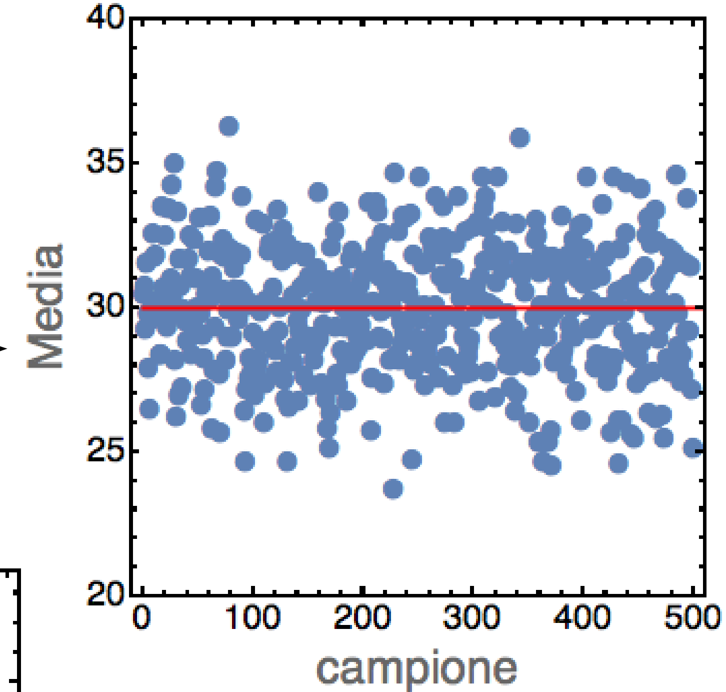
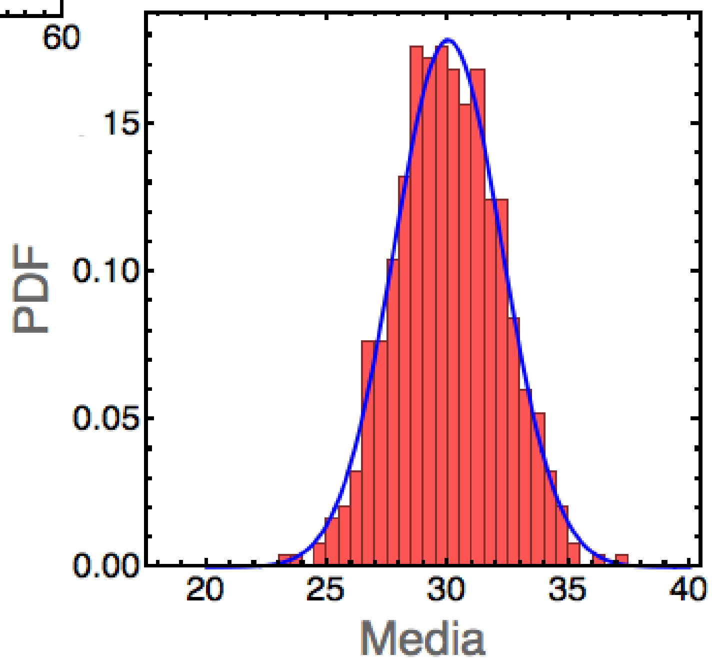
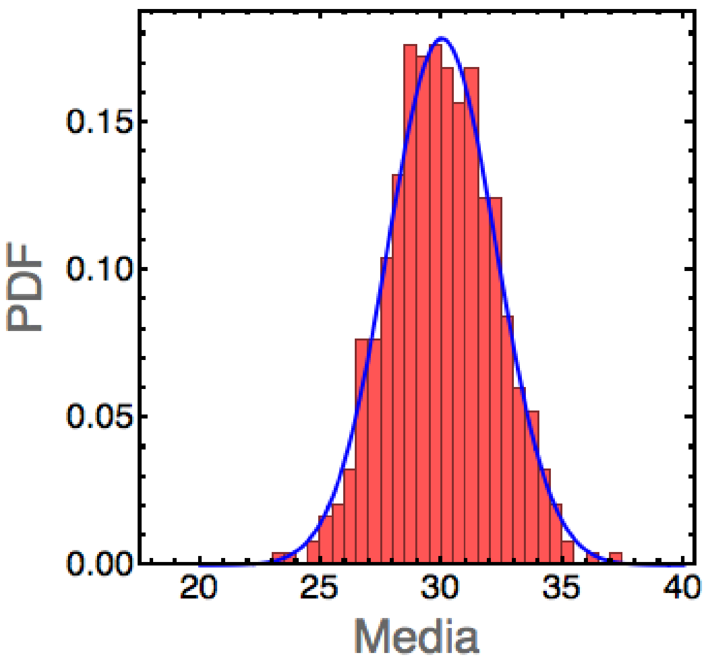


grafico della distribuzione
delle medie campionarie



$$\mu_m = 30 \quad a.u.$$
$$\sigma_m = \frac{\sigma}{\sqrt{n}} \quad a.u.$$

Varianza: proprietà



$$\mu_m = 30 \quad a.u.$$

$$\sigma_m = \frac{\sigma}{\sqrt{n}} \quad a.u.$$

Questo ci dice che le medie campionarie si comportano come una variabile aleatoria con distribuzione $\mathcal{N}(\mu_m, \sigma_m)$

E la varianza (calcolata per ogni campione) come si distribuisce?

$$\mathrm{Var(x)} = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2$$

varianza campionaria

Chi-quadro

In generale, se $x_1, x_2, x_3, ..., x_k$ sono variabili aleatorie **indipendenti** con distribuzione $\mathcal{N}(0,1)$ allora la variabile aleatoria

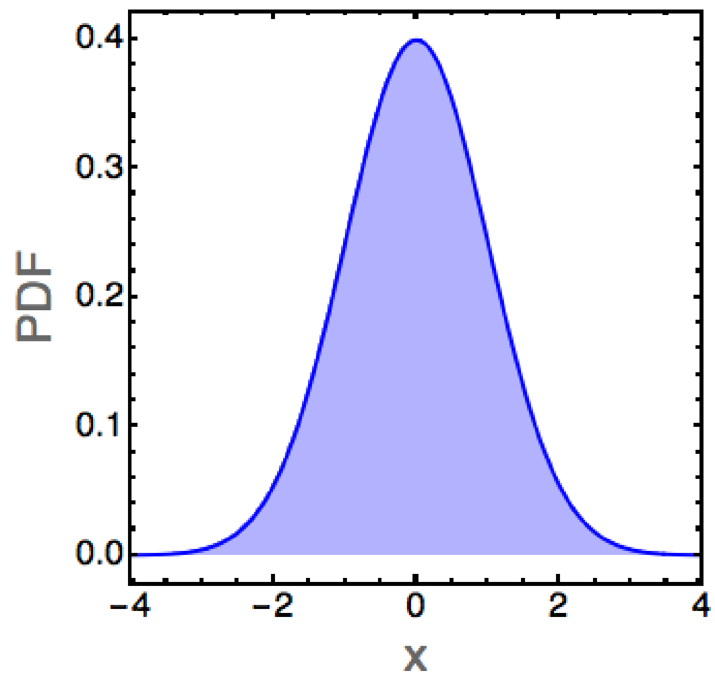
$$x^2 = x_1^2 + x_2^2 + ... + x_k^2 = \sum_{i=1}^{i=k} x_i^2$$

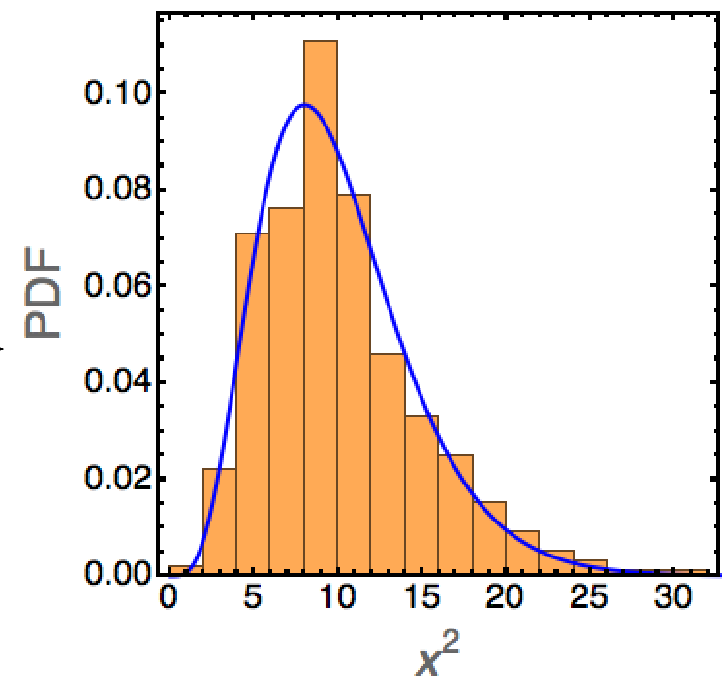
si distribuisce secondo una distribuzione Chi-quadro con k gradi di libertà

$$\chi^2(k)$$

Chi-quadro

es. numerico:





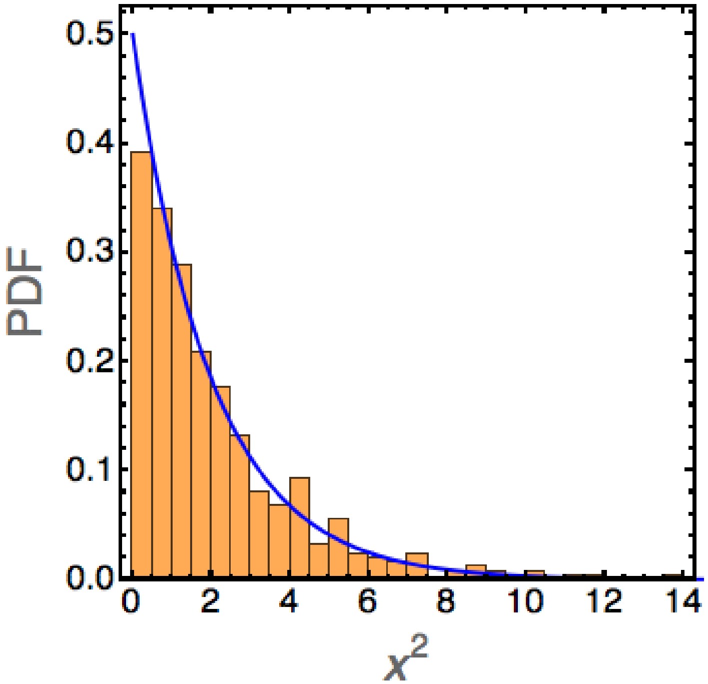
$$\mathcal{N}(0,1) \;\rightarrow\; \{x_1, x_2, ..., x_{10}\}$$

$$x^2 \;\rightarrow\; \chi^2(10)$$

nb. 500 campioni

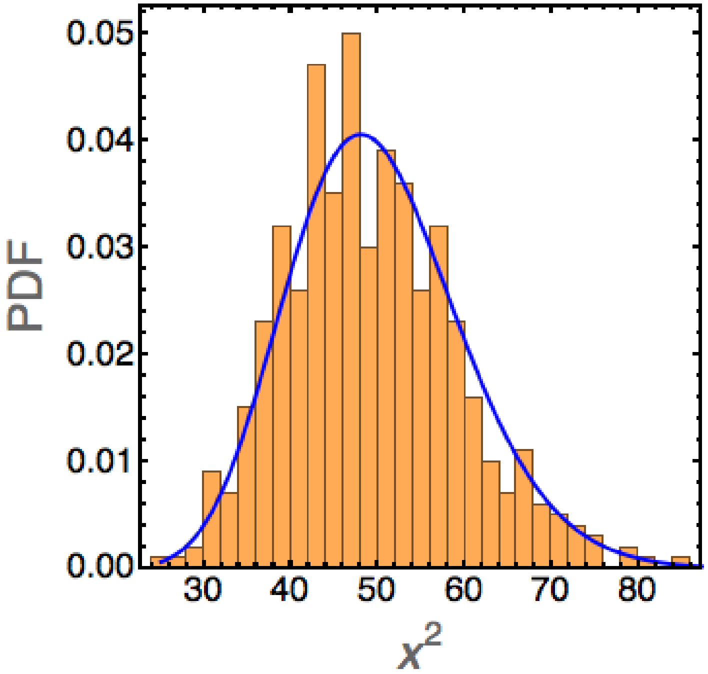
Chi-quadro

es. numerico:



$$\mathcal{N}(0,1) \;\rightarrow\; \{x_1, x_2\}$$

$$x^2 \;\rightarrow\; \chi^2(2)$$



$$\mathcal{N}(0,1) \;\rightarrow\; \{x_1, x_2 ..., x_{50}\}$$

$$x^2 \;\rightarrow\; \chi^2(50)$$

Varianza: proprietà

$$\mathrm{Var(x)} = \frac{1}{n-1}\sum_{i=1}^{i=n}(x_i - \bar{x})^2$$

varianza campionaria

ci attendiamo che $s^2 \rightarrow \propto \chi^2(k)$

e infatti:

$$\boxed{\frac{(n-1)s^2}{\sigma^2} \rightarrow \chi^2(n-1)}$$

questo risultato può essere usato per calcolare l'intervallo di confidenza della varianza esattamente come abbiamo fatto per la media utilizzando la distribuzione t di Student.

Attenzione: il metodo è molto sensibile all'ipotesi di normalità della variabile x di partenza

Varianza: proprietà

$$\frac{(n-1)s^2}{\sigma^2} \rightarrow \chi^2(n-1)$$

es. gli occhi di *Cyrtodiopsis dalmanni* si trovano in cima a peduncoli oculari, particolarmente sviluppati nei maschi a cui pare conferiscano un vantaggio riproduttivo. La distanza interoculare misurata a 9 maschi presi a caso dalla popolazione è:

$$c = \{8.69, 8.15, 9.25, 9.45, 8.96, 8.65, 8.43, 8.79, 8.63\}\ mm$$

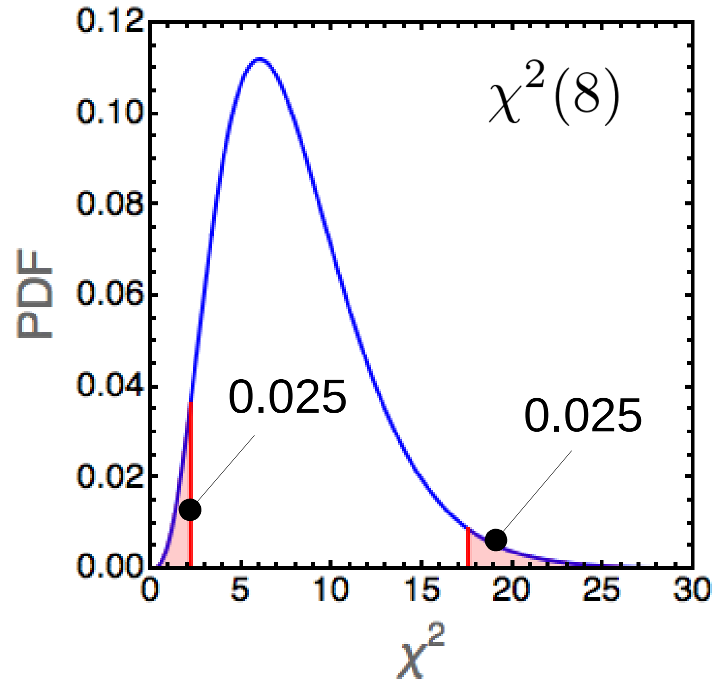
Calcolare l'intervallo di confidenza della varianza:

$$n = 9 \quad \bar{x} = 8.78\ mm \quad s = 0.40\ mm \quad s^2 = 0.16\ mm^2 \quad s_e = \frac{s}{\sqrt{n}} = \frac{0.40}{\sqrt{9}} = 0.134$$

$$\frac{(n-1)s^2}{\chi^2_{crit}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{crit}}$$

il problema è calcolare: perché?

Varianza: intervallo di confidenza



La distribuzione è asimmetrica e dunque i valori critici sono diversi:

$$\chi^2(8)_{0.025} = 17.535$$

$$\chi^2(8)_{1-0.025} = 2.18$$

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Varianza: intervallo di confidenza

$$n = 9 \quad \bar{x} = 8.78\ mm \quad s = 0.40\ mm \quad s^2 = 0.16\ mm^2 \quad s_e = \frac{s}{\sqrt{n}} = \frac{0.40}{\sqrt{9}} = 0.134$$

$$\frac{(n-1)s^2}{\chi^2_{crit}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{crit}}$$

$$\chi^2(8)_{0.025} = 17.535$$

$$\chi^2(8)_{1-0.025} = 2.18$$

$$\frac{8 \cdot 0.16}{17.535} \leq \sigma^2 \leq \frac{8 \cdot 0.16}{2.18}$$

$$0.073 \leq \sigma^2 \leq 0.59$$

$$CI_{s^2} = [0.073, 0.59] \qquad CI_s = [\sqrt{0.073}, \sqrt{0.59}] = [0.27, 0.77]$$

Roberto Chignola
Università di Verona
roberto.chignola@univr.it