Trattazione statistica dei dati

Criterio di Chauvenet

Criterio oggettivo per scartare misure sospette in una serie di N dati misurati $x_1, x_2, \dots, x_N$

$$10.1,\ 10.0,\ 10.2,\ 10.3,\ 10.2,\ 10.1,\ 10.0,\ 10.2,\ 10.1,\ 10.2,\ 10.1,\ 10.3,\ 11.0,\ 11.5,\ 11.6$$

Primo passo (se possibile): ripetere le misure per vedere se le "anomalie" si ripresentano

Secondo passo (se il primo non è possibile o non ha dato risultati):

a) si determinano media e deviazione standard → $\boxed{\overline{x} \simeq 10.4\ mm\ , \quad \sigma_x \simeq 0.5\ mm}$

b) si determina la deviazione dalla media della misura più sospetta → $\boxed{d = 1.2\ mm \text{ ovvero } 2.4\ \sigma_x}$

c) si determina la probabilità di ottenere misure che differiscano dal valore "vero" almeno di questa quantità →

$$\begin{aligned} P(\text{al di fuori di } 2.4\ \sigma_x) &= 1 - P(\text{entro } 2.4\ \sigma_x) = \\ &= 1 - 0.984 = 0.016 \simeq \tfrac{1}{63} \end{aligned}$$

d) si applica il Criterio di Chauvenet che dice:
"Se il numero atteso di misure con deviazione rispetto al valore medio maggiore o uguale a quello della misura sospetta è minore di 0.5, la misura sospetta può essere scartata"

e) si scarta la misura sospetta e si riparte dal punto a) → NOOOO!

Il Criterio di Chauvenet può essere applicato 1 sola volta e la sua applicazione deve portare allo scarto di un numero ridotto di misure!

Trattazione statistica dei dati

Propagazione degli errori nel caso di incertezze accidentali e indipendenti tra loro

a) $z = x + A$ con A costante nota

probabilità di ottenere un certo valore x_i

$$P_x(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x_i - X)^2}{2\sigma_x^2}} dx$$

probabilità di avere il corrispondente valore $z_i = x_i + A$ di z

$$P_z(z_i) = P_x(z_i - A) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{[(z_i - A) - X]^2}{2\sigma_x^2}} dx = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{[z_i - (X + A)]^2}{2\sigma_x^2}} dz$$

valore vero $Z = X + A$

larghezza σ_x

b) $z = B\,x$ con B costante nota

probabilità di ottenere un certo valore $z_i = B\,x_i$

$$P_z(z_i) = P_x(z_i/B) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{[(z_i/B) - X]^2}{2\sigma_x^2}} dx = \frac{1}{\sqrt{2\pi}\,B\,\sigma_x} e^{-\frac{[z_i - (BX)]^2}{2\,B^2\,\sigma_x^2}} dz$$

valore vero $Z = BX$

larghezza $B\sigma_x$

Trattazione statistica dei dati

Propagazione degli errori nel caso di incertezze accidentali e indipendenti tra loro

c) $z = x + y$

Per le due grandezze misurate direttamente avremo

$$P_x(x_i) = \frac{1}{\sqrt{2\pi}\ \sigma_x}\ e^{-\frac{(x_i - X)^2}{2\sigma_x^2}}\, dx \quad \text{e} \quad P_y(y_i) = \frac{1}{\sqrt{2\pi}\ \sigma_y}\ e^{-\frac{(y_i - Y)^2}{2\sigma_y^2}}\, dy$$

probabilità di ottenere un certo valore $z_i = x_i + y_i$

$$P_z(z_i) = P_x(x_i)\ P_y(y_i) = \frac{1}{2\pi\ \sigma_x\ \sigma_y}\ e^{-\frac{1}{2}\left[\frac{(x_i - X)^2}{\sigma_x^2} + \frac{(y_i - Y)^2}{\sigma_y^2}\right]}\, dx\ dy$$

ma

$$\frac{(x_i - X)^2}{\sigma_x^2} + \frac{(y_i - Y)^2}{\sigma_y^2} = \frac{(x_i - X + y_i - Y)^2}{\sigma_x^2 + \sigma_y^2} + \frac{\left[\sigma_y^2\ (x_i - X)\ -\ \sigma_x^2\ (y_i - Y)\right]^2}{\sigma_x^2\ \sigma_y^2\ (\sigma_x^2 + \sigma_y^2)} = \frac{(x_i - X + y_i - Y)^2}{\sigma_x^2 + \sigma_y^2} + \alpha^2$$

da cui segue

$$P_z(z_i) = \frac{1}{2\pi\ \sigma_x\ \sigma_y}\ e^{-\frac{1}{2}\left[\frac{(x_i - X + y_i - Y)^2}{\sigma_x^2 + \sigma_y^2}\right]}\ e^{-\frac{\alpha^2}{2}}\, dx\ dy = \frac{1}{2\pi\ \sigma_x\ \sigma_y}\ e^{-\frac{1}{2}\left[\frac{(z_i - X - Y)^2}{\sigma_x^2 + \sigma_y^2}\right]}\ e^{-\frac{\alpha^2}{2}}\, dx\ dy$$

Sommando su tutte le possibili coppie x_i, y_i la cui somma vale z_i

$$P_z(z_i) \propto \frac{1}{\sqrt{2\pi}\ \sqrt{\sigma_x^2 + \sigma_y^2}}\ e^{-\frac{1}{2}\left[\frac{(z_i - X - Y)^2}{\sigma_x{}^2 + \sigma_y{}^2}\right]}$$

valore vero $X + Y$

larghezza $\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$.

Trattazione statistica dei dati

Propagazione degli errori nel caso di incertezze accidentali e indipendenti tra loro

d) caso generale $z = z(x, y)$

le incertezze relative sulle grandezze misurate direttamente $\sigma_x/\overline{x}$ e $\sigma_y/\overline{y}$ siano molto minori di 1

$$z_i = z(x_i, y_i) = z(X,Y) + \left(\frac{\partial z}{\partial x}\right)_{X,Y} (x_i - X) + \left(\frac{\partial z}{\partial y}\right)_{X,Y} (y_i - Y)$$

costante A

B costante

B costante

variabile distribuita normalmente intorno al valore vero 0 con parametro di larghezza σ_x σ_y

Combinando i tre termini

i valori z_i sono distribuiti normalmente intorno al valore vero $z(X,Y)$

parametro di larghezza

$$\sigma_z = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2_{X,Y} \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2_{X,Y} \sigma_y^2}$$

Trattazione statistica dei dati

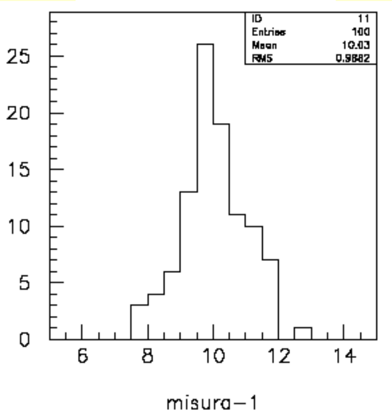
Deviazione standard della media

100 misure della grandezza x	X_{m1}, σ_{x1}	
altre 100 misure → totale (t) 200 misure	X_{m2}, σ_{x2}	$X_{tm2} \approx X_{m1}$, $\sigma_{tx2} \approx \sigma_{x1}$
. . .	. . .	. . .
altre 100 misure → totale (t) 5000 misure	X_{m50}, σ_{x50}	$X_{tm50} \approx X_{m1}$, $\sigma_{tx50} \approx \sigma_{x1}$

???

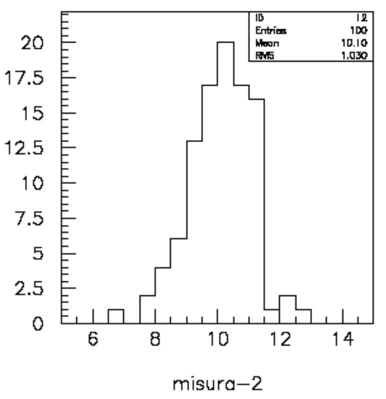
Misura della grandezza x (cm):

100 misure



x_m = 10.03 cm
σ_x = 0.97 cm

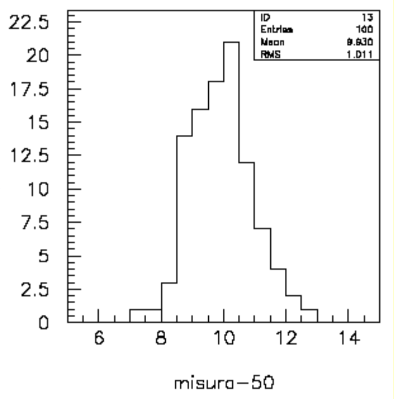
100 misure



x_m = 10.10 cm
σ_x = 1.03 cm

→ 3 .... 49 →

100 misure

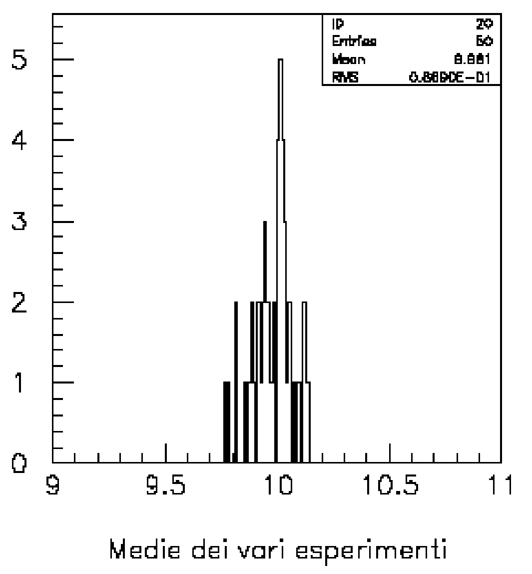


x_m = 9.93 cm
σ_x = 1.01 cm

Trattazione statistica dei dati

Deviazione standard della media

Distribuzione dei valori medi



x_m = 9.98 cm

σ_{x_m} = 0.09 cm

≈ 1/10 cm

la media è distribuita normalmente intorno al valore vero X con

deviazione standard della media

$$\sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}}$$

Incertezza con la quale il valor medio misurato stima statisticamente il valore vero

risultato della misura $\overline{x} \pm \sigma_{\overline{x}}$

Ma attentione!!!
(errori sistematici)

Trattazione statistica dei dati

Livello di confidenza

confrontare tra loro due misure di una grandezza z

$$z_1 \pm \sigma_1 \qquad\qquad z_2 \pm \sigma_2$$

La deviazione τ tra le due misure è anch'essa distribuita normalmente

$$\tau_M = | z_1 - z_2 | \quad \text{e} \quad \sigma_\tau = \sqrt{\left(\frac{\partial\tau}{\partial z_1}\right)^2 \sigma_1^2 + \left(\frac{\partial\tau}{\partial z_2}\right)^2 \sigma_2^2} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

In generale τ_M sarà diverso da zero → valore corrispondente all'accordo perfetto

$$t = | \tau_M / \sigma_\tau | \qquad P(\text{al di fuori di } t\sigma_\tau) = 1 - P(\text{entro } t\sigma_\tau) = 1 - \int_{-t\sigma_\tau}^{+t\sigma_\tau} p(\tau)d\tau$$

dove $p(\tau)$ è la densità di probabilità gaussiana associata alla deviazione τ.

$P(\text{al di fuori di } t\sigma_\tau)$ grande → due misure sono → consistenti tra loro

$P(\text{al di fuori di } t\sigma_\tau)$ piccola → due misure sono → inconsistenti tra loro

Il limite di accettabilità detto livello di confidenza convenzionalmente posto al 5%.

Trattazione statistica dei dati

Media pesata

$$x_1 \pm \sigma_1 \qquad x_2 \pm \sigma_2 \qquad x_3 \pm \sigma_3$$

tutte le misure siano consistenti tra loro

$$P_x(x_1) \propto \frac{1}{\sigma_1}\, e^{-\frac{(x_1-X)^2}{2\sigma_1^2}} \qquad P_x(x_2) \propto \frac{1}{\sigma_2}\, e^{-\frac{(x_2-X)^2}{2\sigma_2^2}} \qquad P_x(x_3) \propto \frac{1}{\sigma_3}\, e^{-\frac{(x_3-X)^2}{2\sigma_3^2}}$$

probabilità di ottenere la terna di misure

$$P_x(x_1,\ x_2,\ x_3) = P_x(x_1)\ P_x(x_2)\ P_x(x_3) \propto \frac{1}{\sigma_1\ \sigma_2\ \sigma_3}\, e^{-\frac{\chi^2}{2}}$$

con

$$\chi^2 = \left(\frac{x_1 - X}{\sigma_1}\right)^2 + \left(\frac{x_2 - X}{\sigma_2}\right)^2 + \left(\frac{x_3 - X}{\sigma_3}\right)^2$$

principio di massima verosimiglianza

$$\frac{\partial \chi^2}{\partial X} = 0 \qquad \Rightarrow \qquad \boxed{X_M = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} + \frac{x_3}{\sigma_3^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}}}$$

Media pesata
con pesi →

$$w_1 = \frac{1}{\sigma_1^2} \qquad w_2 = \frac{1}{\sigma_2^2} \qquad w_3 = \frac{1}{\sigma_3^2}$$

$$\boxed{\sigma_{X_M} = \sqrt{\sum_i \left(\frac{\partial X_M}{\partial x_i}\right)^2 \sigma_i^2} = \frac{1}{\sqrt{\sum_i \frac{1}{\sigma_i^2}}} = \frac{1}{\sqrt{\sum_i w_i}}}$$

Trattazione statistica dei dati

Covarianza nella propagazione degli errori

$z = z(x, y)$ → N misure $(x_1,\ y_1,\ x_2,\ y_2,\ \ldots,\ x_N,\ y_N)$ → $\overline{x},\ \overline{y}$ e $\sigma_x,\ \sigma_y$

In corrispondenza delle N coppie $x_i,\ y_i$ potremo determinare N valori di z_i

incertezze relative nelle misure dirette siano $\ll 1$

$$z_i = z(x_i, y_i) = z(\overline{x},\ \overline{y}) + \left(\frac{\partial z}{\partial x}\right)_{\overline{x},\overline{y}} (x_i - \overline{x}) + \left(\frac{\partial z}{\partial y}\right)_{\overline{x},\overline{y}} (y_i - \overline{y})$$

$$\overline{z} = \frac{\sum_{i=1}^{N} z_i}{N} = \frac{1}{N}\sum_{i=1}^{N}\left[z(\overline{x},\overline{y}) + \left(\frac{\partial z}{\partial x}\right)_{\overline{x},\overline{y}} (x_i - \overline{x}) + \left(\frac{\partial z}{\partial y}\right)_{\overline{x},\overline{y}} (y_i - \overline{y})\right] = z(\overline{x},\overline{y})$$

$$\sigma_z^2 = \frac{1}{N-1}\sum_{i=1}^{N}(z_i - \overline{z})^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left[\left(\frac{\partial z}{\partial x}\right)_{\overline{x},\overline{y}} (x_i - \overline{x}) + \left(\frac{\partial z}{\partial y}\right)_{\overline{x},\overline{y}} (y_i - \overline{y})\right]^2 =$$
$$= \boxed{\left(\frac{\partial z}{\partial x}\right)^2_{\overline{x},\overline{y}} \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2_{\overline{x},\overline{y}} \sigma_y^2 + 2\left(\frac{\partial z}{\partial x}\right)_{\overline{x},\overline{y}}\left(\frac{\partial z}{\partial y}\right)_{\overline{x},\overline{y}} \sigma_{xy}}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})$$ → covarianza

Trattazione statistica dei dati

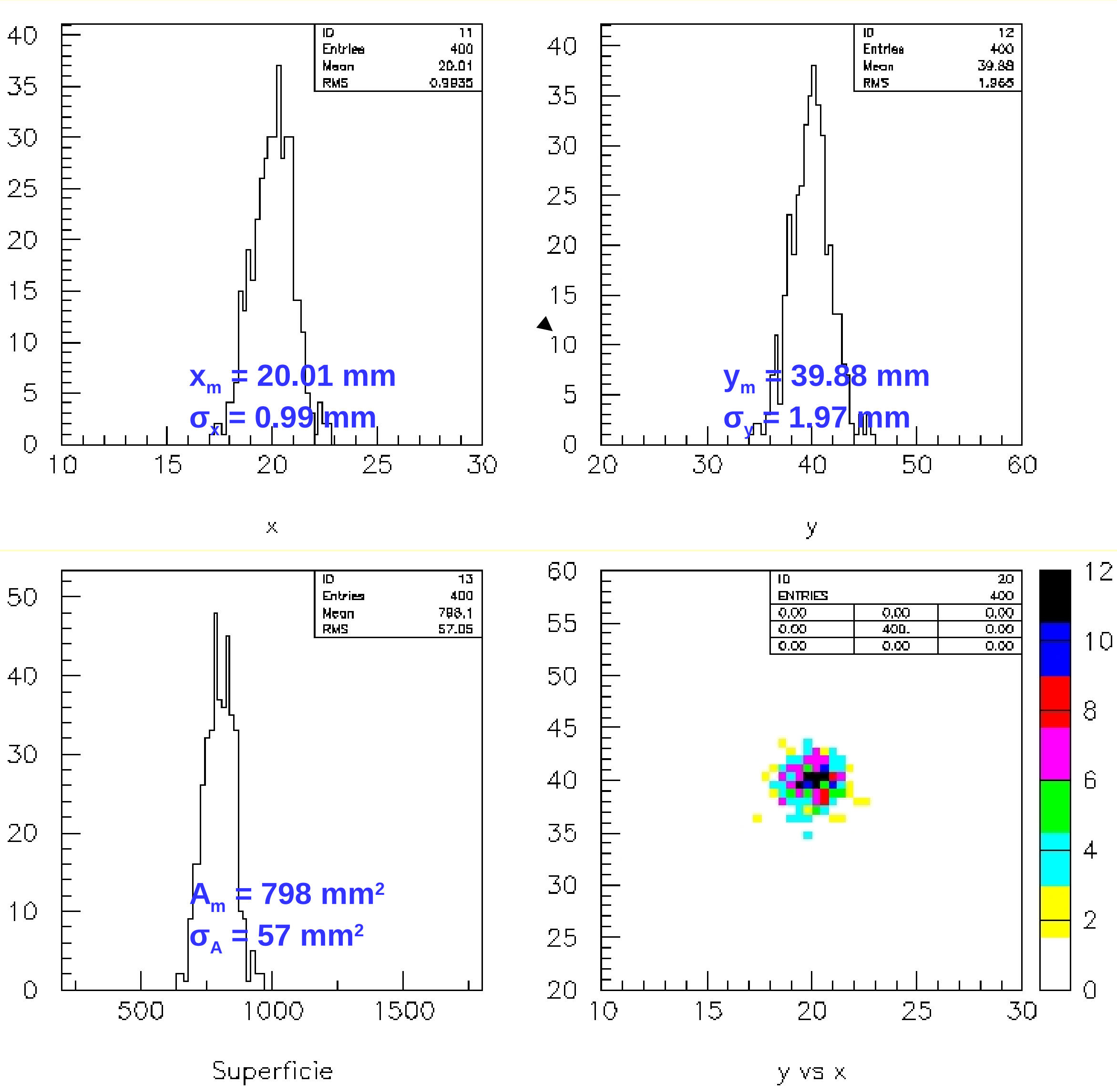
Covarianza nella propagazione degli errori

Misura della superficie di un rettangolo di lati x, y → A = x*y

x e y sono misurate direttamente e sono indipendenti tra loro

$$\sigma_z = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2_{X,Y} \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2_{X,Y} \sigma_y^2}$$

$20.01^2 * 1.97^2 + 39.88^2 * 0.99^2 = 57^2$



Trattazione statistica dei dati

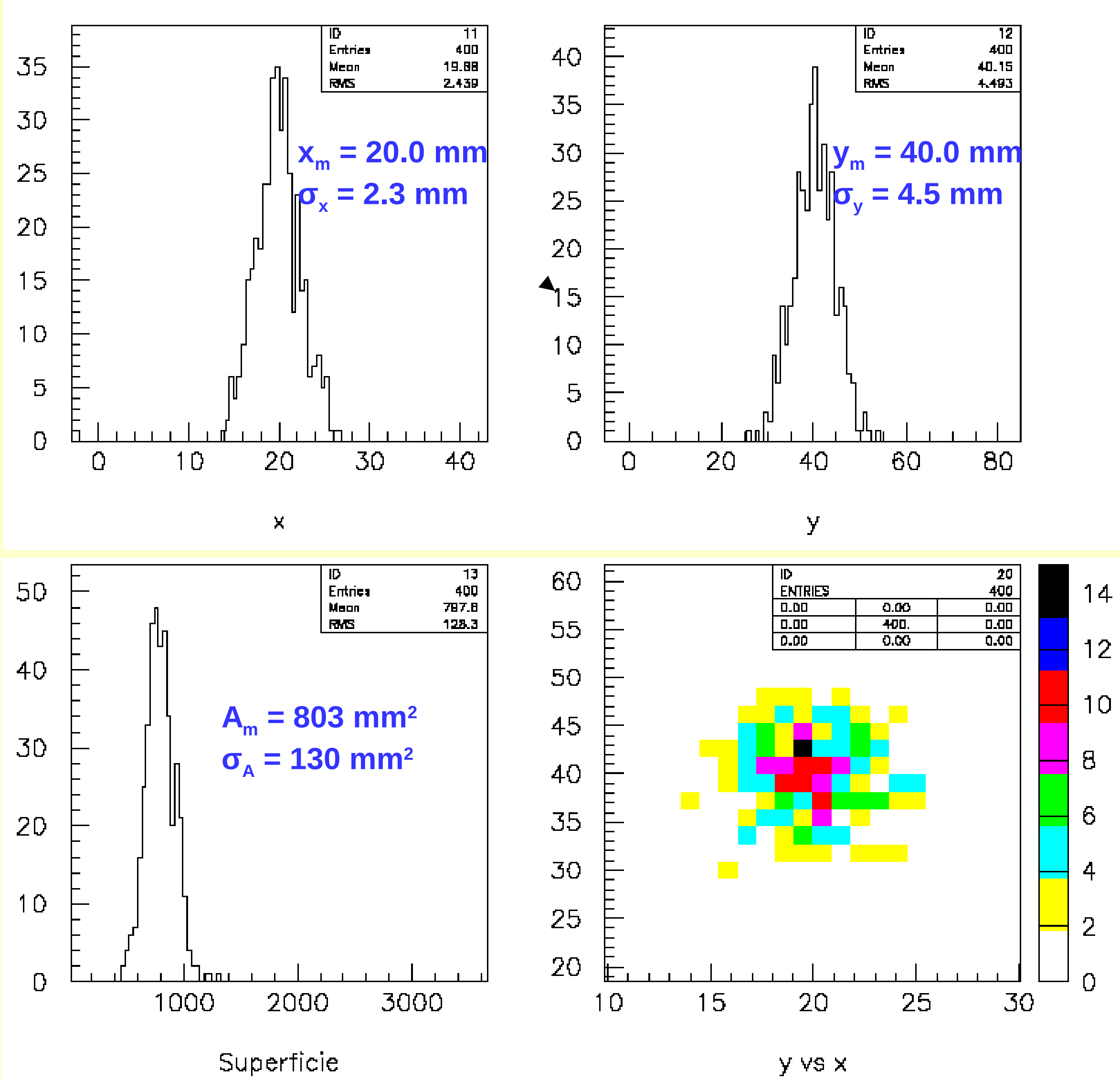
Covarianza nella propagazione degli errori

Durante le misure si ha una variazione gaussiana di temperatura che produce un'elongazione relativa di ± 10% (σ_T) su x e y

Tale variazione aumenta σ_x e σ_y ; se ciò influenzasse le misure di x e y in maniera indipendente avremmo:

$$\sigma_z = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2_{X,Y} \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2_{X,Y} \sigma_y^2}$$

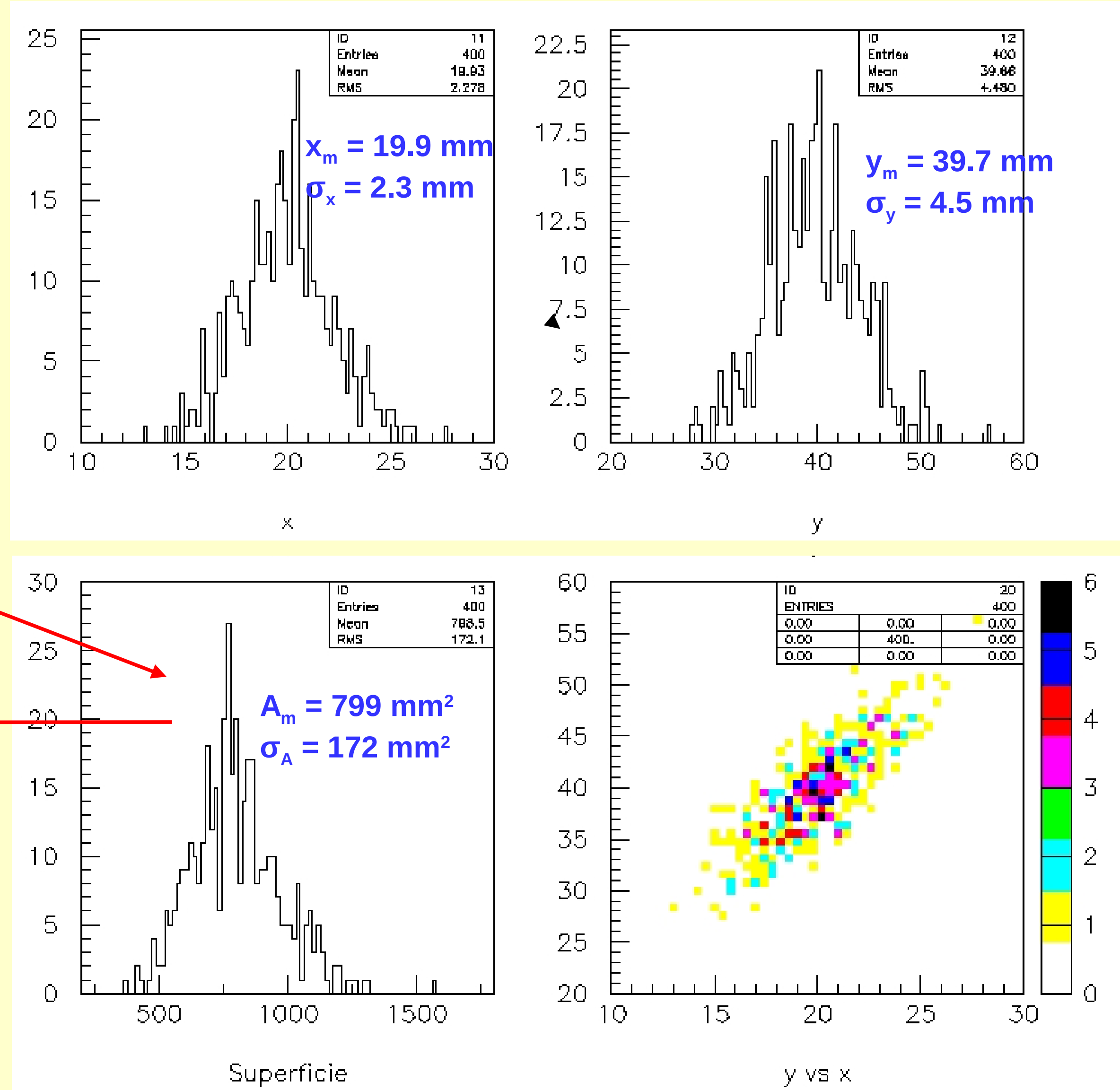
$$\sigma_A^2 = 20^2 * 4.5^2 + 40^2 * 2.3^2 = 129^2$$



Trattazione statistica dei dati

Covarianza nella propagazione degli errori

Più realisticamente la variazione gaussiana di temperatura influenzerà in maniera correlata le due lunghezze x e y rendendo le fluttuazioni sulle loro misure non indipendenti. Infatti



$$\boxed{\sigma_A^2 = 172^2 > 129^2}$$

Ma 39.7*2.3+19.9*4.5 ≈181!!!

$$\sigma_z \leq \left|\frac{\partial z}{\partial x}\right|_{\bar{x},\bar{y}} \sigma_x + \left|\frac{\partial z}{\partial y}\right|_{\bar{x},\bar{y}} \sigma_y$$