NOTE METODOLOGICHE

1. Calcolo dell'intervallo di confidenza

Nell'esempio postato da Gaspy, abbiamo a che fare con un C. porphyropus di cui vengono prese in considerazione le dimensioni (lunghezza e larghezza) di N= 25 spore.

I parametri campionari rilevati sono i seguenti:

$$\overline{X} = (media)8{,}33$$
$$S_X = (s.q.m.)0{,}438$$
$$\overline{Y} = (media)5{,}32$$
$$S_y = (s.q.m.)0{,}348$$

Da qui si fa discendere l'intervallo di confidenza : 7,9 - 8,8 x 5-5,7 μm.

Questi risultati si sono ottenuti da :

$$\overline{X} \pm S_X ; \overline{Y} \pm S_y$$

cioè :

8,33-0,438= 7,892 (arr. **7,9**) ; 8,33+0,438 = 8,768 (arr. **8,8**)

Stesso criterio per la larghezza.

Questo, in presunzione del fatto che la media più una volta lo scarto quadratico medio contengano il 68% dei valori (per essere precisi il 68,27%). In realtà il 68% indica la densità di probabilità di una ***distribuzione normale standardizzata***

con media $\mu = 0$e varianza σ^2 =1

L'intervallo di confidenza è legato alla *formulazione di rischio* α, *dove* α*assume valori* = *0.001; 0,05; 0,1.*

Può assumerne anche altri, a piacere; però le codifiche internazionali e le relative tavole di contingenza, prevedono questi tre livelli di significatività.

La formulazione di rischio serve a determinare la significatività dei dati; ed afferisce all'errore di primo tipo

Poi c'è la formulazione di rischio β(potenza del test) che ci fornisce informazioni sull'errore di secondo tipo.

Senza rendere eccessivamente complicata la spiegazione, l'intervallo di confidenza, relativo ai dati dell'esempio, si calcola nel modo seguente: si eguaglia la media a zero (stiamo parlando di curva normale standardizzata); da cui la formula:

$$\overline{X} = \mu \pm Z_\alpha \frac{\sigma}{\sqrt{n}}$$

dove:

$\overline{X}$ è la media del campione

μè la media vera della popolazione

Z_αè il valore tabulato della curva normale standardizzata alla rispettiva significatività α

σ è lo scarto quadratico medio della popolazione

n è il numero delle unità statistiche osservate

Con $\alpha = 0,05$(conf. 95%) Z=1,96 (valore desunto dalle tavole statistiche).

Dunque si avrà:

$$\overline{X} = 0 \pm 1,96\frac{0,438}{\sqrt{25}} = 0 \pm 1,96 \cdot 0,0876 = 0,172$$

Dunque l'intervallo di confidenza sarà :

8,33 $\pm 0,172$= 8,18 - 8,51μm

In realtà questo risultato è ancora inesatto; perchè abbiamo preso in considerazione l'ipotetico scarto quadratico medio della popolazione che, a ben guardare è incognito; mentre è noto lo s.q.m. del campione.

Quando la varianza è incognita, l'intervallo di confidenza si ottiene col t di Student.

Salto la dimostrazione e inserisco solo il risultato:

8,15- 8,52 μm

Utilizzando il medesimo criterio di calcolo, l'intervallo di confidenza della laghezza sara:

5,18- 5,47 μm