Determinazione di n per l'intervallo di confidenza di una proporzione.

Sia *P* la proporzione di elementi della popolazione che hanno un determinato attributo e *(1-P)* la proporzione di quelli che non l'hanno. Possiamo determinare l'intervallo di confidenza per la proporzione sfruttando l'approssimazione alla normale:

$$p - z_{\alpha/2}\sqrt{\frac{(1-n/N)\,p(1-p)}{n-1}} \le P \le p + z_{\alpha/2}\sqrt{\frac{(1-n/N)\,p(1-p)}{n-1}}$$

dove n è la numerosità campionaria e N la numerosità della popolazione.

Il valore di n dipende dal livello di precisione che richiediamo, in genere determinato dall'ampiezza massima dell'intervallo di confidenza.

Sia 2D l'ampiezza massima ammissibile per l'intervallo di confidenza; quanto deve valere n affinché nel 95% dei casi il vero valore del parametro sia compreso in un intervallo di ampiezza pari a 2D?

Per avere la soluzione basta porre la differenza fra i due estremi dell'intervallo pari a 2D:

per variabile dicotomica non ho bisogno di conoscere la stima della varianza, se on so nulla di p posso considereare la max varibilita

$$2D = 2z_{\alpha/2}\sqrt{\frac{(1-n/N)\,p(1-p)}{n-1}}$$, da cui

$$n = \frac{N}{\frac{N \cdot D^2}{z^2_{\alpha/2} S^2}+1} = \frac{z^2_{\alpha/2} S^2}{D^2+\frac{z^2_{\alpha/2} S^2}{N}} = \frac{n_0}{1+\frac{n_0}{N}},$$

dove $S^2 = \frac{np(1-p)}{n-1}$ e $n_0 = \frac{z^2_{\alpha/2} S^2}{D^2}$

p è la stima di P, ma in questo caso non è nota, poiché l'indagine deve ancora essere effettuata, per cui è necessario fare delle ipotesi sul possibile valore di P.

Se n << N, possiamo semplificare l'espressione come:

$$n \cong n_0 = \frac{z^2_{\alpha/2}\, p(1-p)}{D^2}$$

ulterioiri ipotesi rispetto a quello gia visto

Nel caso di assenza totale di ipotesi, si può porre P=0.5, la situazione di massima variabilità (e quindi di massima numerosità). In tal caso:

ipotizzo di non conoscere nulla cioe max varianza = 0.25 xke 0.5

$$n = \frac{1{,}96^2 \cdot 0{,}25}{D^2} \approx \frac{1}{D^2}$$

Il valore di n dipende prevalentemente dalla precisione richiesta: essendo il livello di precisione un numero minore di 1 posto al denominatore (e elevato al quadrato), fa aumentare rapidamente n:

di solito D^2 = 3%

=10% ipotizzo un errore del 30% avro un int che va dal 20 al 40%
infatti 2*D, infatti le n sono piccole xke ce tanto errore

numerosità ottimale al variare di p

p	D	n
0.1	0.1	35
0.2	0.1	61
0.3	0.1	81
0.4	0.1	92
0.5	0.1	96
0.6	0.1	92
0.7	0.1	81
0.8	0.1	61
0.9	0.1	35

p	D	n
0.1	0.05	138
0.2	0.05	246
0.3	0.05	323
0.4	0.05	369
0.5	0.05	384
0.6	0.05	369
0.7	0.05	323
0.8	0.05	246
0.9	0.05	138

p	D	n
0.1	0.025	553
0.2	0.025	983
0.3	0.025	1291
0.4	0.025	1475
0.5	0.025	1537
0.6	0.025	1475
0.7	0.025	1291
0.8	0.025	983
0.9	0.025	553

In mancanza di informazioni sul possibile valore di *P* nella popolazione, si può assumere *P=1/2*, ottenendo così il valore massimo di *n*

effetto di p n ha il max in 0.5=p
infatti per una v dicotomica ipotizzo la max variabilita
ma si deve cercare di utilizzare altri metodi come dati precedenti o aiuti da esperti ad es nella medicina