Intervallo di Probabilità e di Confidenza

Intervallo di probabilità

Grazie al teorema del limite centrale, se la v.c. X ha $E(x)=\mu$ e $Var=\sigma^2$ si ha (per $n\rightarrow\infty$) che:

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\Downarrow$$

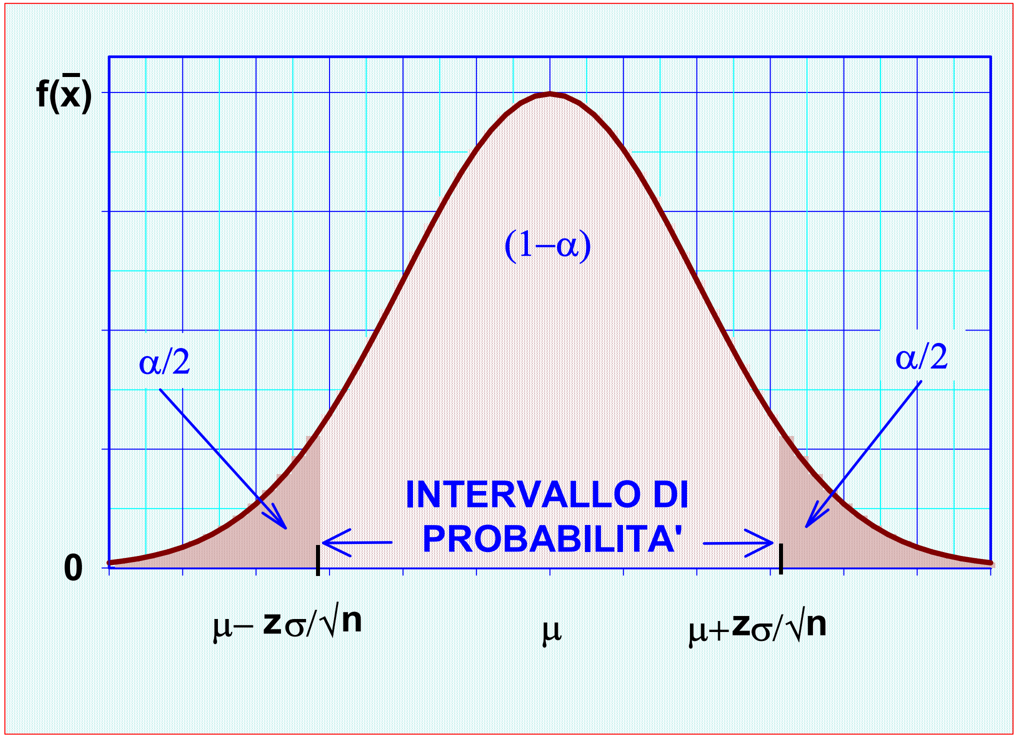
Possiamo calcolare la probabilità che una data media campionaria ha di appartenere ad un certo intorno della **media vera μ** della variabile casuale x

Intervallo di probabilità

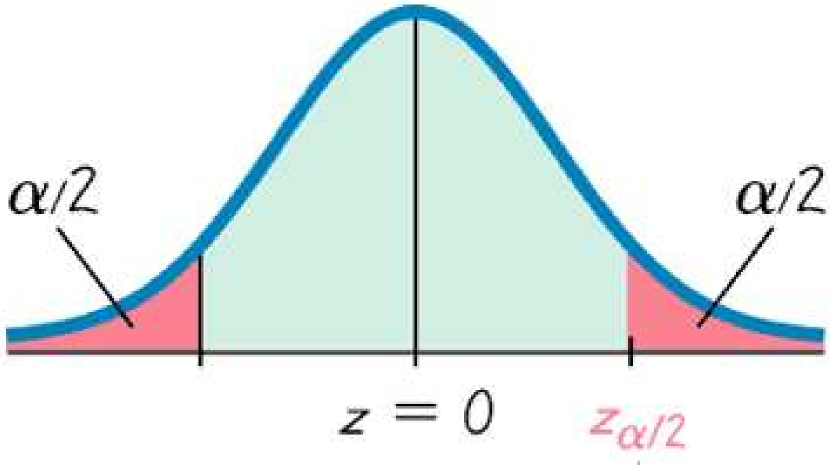
Consideriamo l'intervallo

$\mu \pm z_{\alpha/2}\, \sigma/\sqrt{n}$ simmetrico

attorno alla media vera μ



Nota



$$z_{\alpha/2}: P(Z > z_{\frac{\alpha}{2}}) = \alpha/2$$

Intervallo di probabilità

In base alla distribuzione campionaria di $\bar{x}$ la probabilità di ottenere una media campionaria appartenente a tale intervallo è $(1-\alpha)$, ovvero

$$p\ \{\mu - z_{\alpha/2}\sigma/\sqrt{n} < \bar{x} < \mu + z_{\alpha/2}\sigma/\sqrt{n}\ \} = 1 - \alpha$$

e tale intervallo è detto

INTERVALLO DI PROBABILITÀ PER LA MEDIA CAMPIONARIA

- Intervallo di ampiezza nota$=2\Delta=2z_{\alpha/2}\sigma/\sqrt{n}$
- Centrato sull'ignoto parametro μ

Intervallo di confidenza

Con semplici passaggi algebrici sulle due equazioni, possiamo esprimere l'intervallo in termini del valore osservato $\bar{x}$:

$$p \{ \bar{x} - z_{\alpha/2}\sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \}=1 - \alpha$$

- Intervallo di ampiezza nota=$2\Delta=2z_{\alpha/2}\sigma/\sqrt{n}$
- Centrato sulla media campionaria nota $\bar{x}$

Tale intervallo è detto

INTERVALLO DI CONFIDENZA PER IL PARAMETRO μ

Intervallo di confidenza

Poiché dipende dalla variabile $\bar{\mathbf{X}}$:

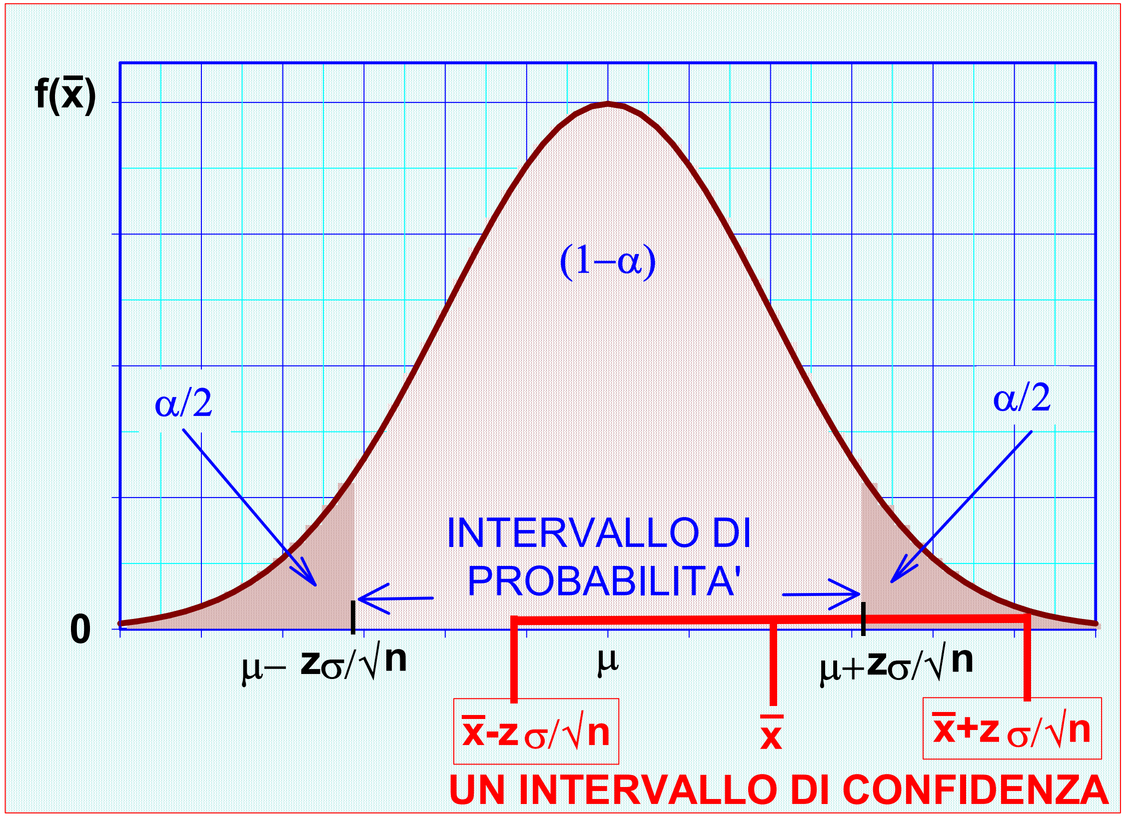
- Varia casualmente attorno al parametro μ
- Ha probabilità pari a (1- α) di includere il parametro μ della variabile **X**

Si noti che:

È **impossibile risalire** da una stima campionaria **al vero valore del parametro μ** di un universo, ma …

… è possibile determinare in base a tale stima un **intervallo** che abbia una **prefissata probabilità** (1- α) **di includere il parametro μ**

Intervallo di probabilità e di confidenza



Intervallo di confidenza

L'**ampiezza** ($2\Delta=2z_{\alpha/2}\sigma/\sqrt{n}$) dell'IC :

esprime l'**indeterminazione** con cui è noto il valore di μ.

L'ampiezza **non dipende** dal valore di μ.

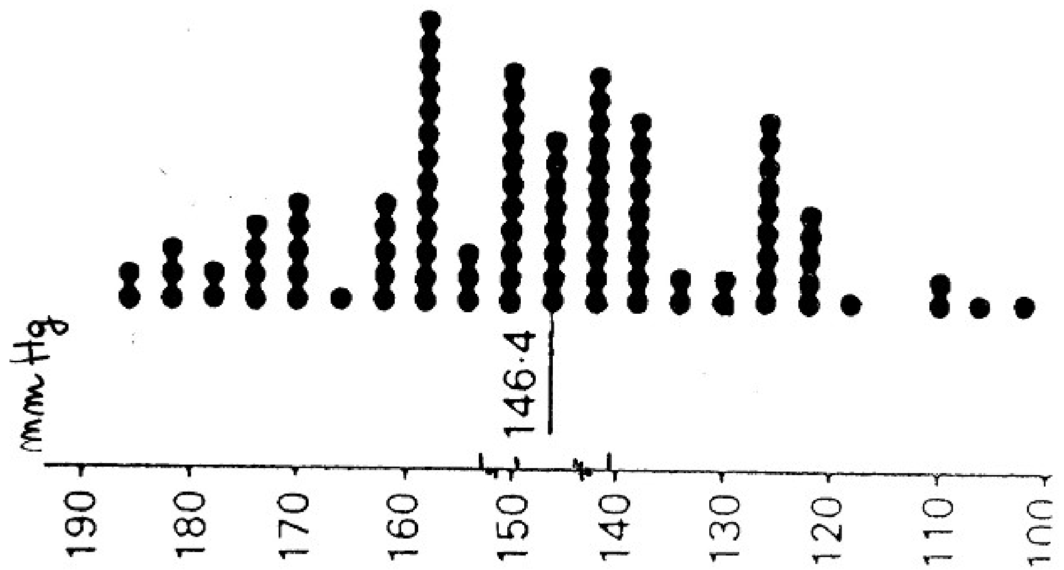
A parità del livello di confidenza $1-\alpha$:

- maggiore è l'ampiezza dell'IC, minore è la precisione della stima campionaria di μ,
- l'ampiezza dipende dall'**errore standard** ($\sigma/\sqrt{n}$), che a sua volta dipende dalla **dimensione** (n) del **campione** e quindi
- diminuisce al crescere di n, ovvero, al crescere di n la precisione aumenta

All'aumentare del grado di confidenza (es 99% invece di 95%) l'ampiezza dell'IC aumenta e la precisione diminuisce

Esercizio 1

I valori di pressione sistolica misurata su 100 individui diabetici di età compresa fra i 40-49 anni sono riportati nel grafico che segue.



La media campionaria è risultata pari a 146.4 mmHg.

Quesiti

1. Nell'ipotesi in cui la ds (deviazione standard) dell'intera popolazione sia nota e pari a 22 mmHg, si determini l'Intervallo di Confidenza (IC) al 95% della media campionaria.

2. Determinare l'IC associato a un grado di confidenza del 99%. Confrontare i risultati ottenuti nel punto precedente.

Risposte

1. $$\bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} = 146.4 \pm 1.96 \frac{22}{\sqrt{100}}$$

$$IC_{95\%} = (142.09; 150.71)$$

2. $$\bar{x} \pm z_{0.005} \frac{\sigma}{\sqrt{n}} = 146.4 \pm 2.58 \frac{22}{\sqrt{100}}$$

$$IC_{99\%} = (140.73; 152.07)$$

Il problema della stima

Effetto osservato

↓

Variabilità casuale

↓

Stima d'intervallo

E' la migliore informazione che abbiamo sull'effetto vero

A causa di questa l'effetto osservato non è uguale a quello vero

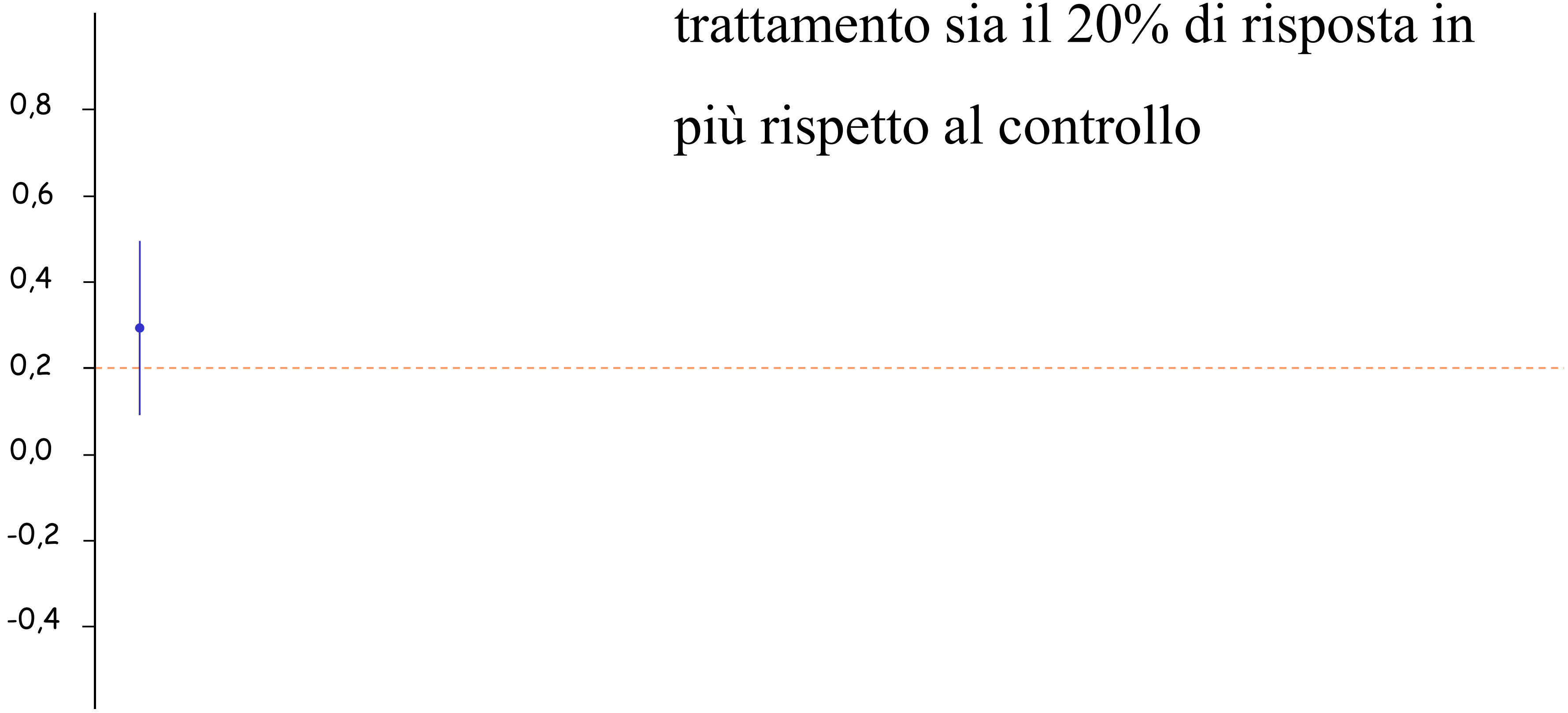
Posso però ricavare un intervallo di valori plausibili per l'effetto vero

Gli intervalli di confidenza

- Indicano un intervallo di valori all'interno del quale il ricercatore conclude, con una specificata probabilità, che sia compreso l'effetto vero del trattamento
- Forniscono informazioni sull'entità dell'effetto e sulla precisione della stima con un livello di 'confidenza' prestabilito (tipicamente 95%)

Gli intervalli di confidenza

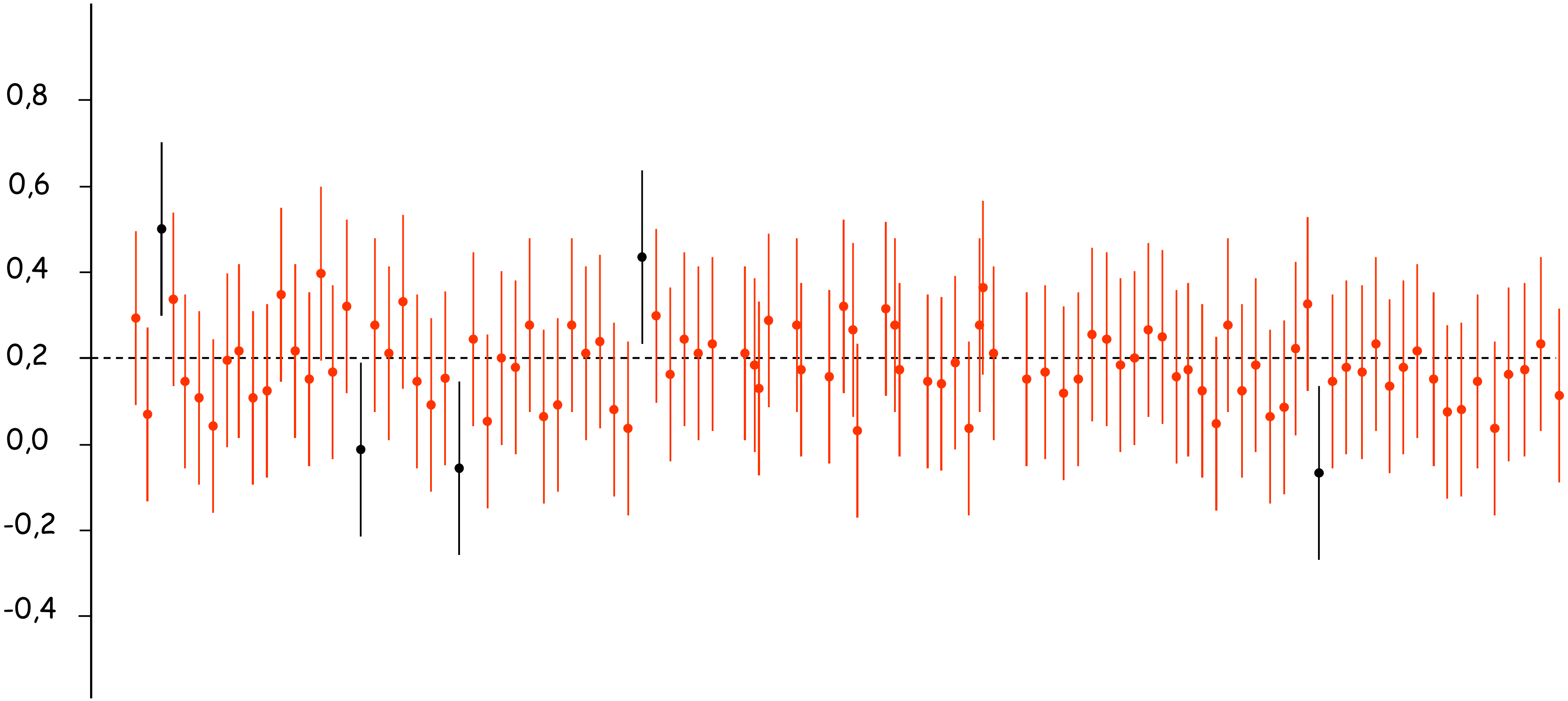
Ipotizziamo che l'effetto vero di un trattamento sia il 20% di risposta in più rispetto al controllo



Gli intervalli di confidenza

Gli intervalli di confidenza

Gli intervalli di confidenza



Nel 95% circa dei campioni possibili l'intervallo di confidenza comprende l'effetto 'vero' (δ = 0,2)

Gli intervalli di confidenza

- Maggiore è l'ampiezza dell' I.C. minore è la precisione della stima

- L'ampiezza dell' I.C., e quindi la precisione della stima, varia con la numerosità dello studio e il grado di confidenza desiderato

    - All'aumentare della numerosità l'ampiezza dell' I.C. diminuisce e la precisione aumenta

    - All'aumentare del grado di confidenza (es. 99% invece di 95%) l'ampiezza dell' I.C. aumenta e la precisione diminuisce

Dimensione del campione

Data σ

è possibile calcolare la

dimensione del campione

necessaria per ottenere

un intervallo di confidenza **(1- α)** e ampiezza **2Δ**

$\Delta = z_{\alpha/2}\sigma/\sqrt{n}$ $\Rightarrow$ $n = (z_{\alpha/2}\sigma/\Delta)^2$

Esempio

- Si vuole stimare il valore medio μ dell'uricemia in una popolazione maschile: è noto che in tale popolazione la dispersione dell'uricemia è $\sigma = 1.1$ mg/dl. Si richiede che la confidenza sia del 95% e che l'indeterminazione Δ non ecceda 0.35 mg/dl.

1) Calcolare la numerosità campionaria necessaria per soddisfare le condizioni richieste

$$n=(z_{\alpha/2}\sigma/\Delta)^2$$

$$n = (1.96*1.1/0.35)^2 = 37.9 \approx 40 \text{ soggetti}$$

Servono almeno 40 soggetti per ottenere una stima con la precisione prescelta

2) Si supponga di aver estratto un campione casuale di 40 soggetti dalla popolazione e di aver determinato il valore di uricemia di ognuno dei 40 soggetti. La media di tali valori risulta $\bar{x}$ =5.55 mg/dl).

Si ricavi l'intervallo di confidenza della media al 95%

$$I.C._{95\%} = \bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

$$I.C._{95\%} = 5.55 \pm 1.96 \times \frac{1.1}{\sqrt{40}} = 5.55 \pm 0.34 = (5.21, 5.89)$$

3) Sullo stesso campione, con $\bar{x} = 5.55$ mg/dl, di ricavi l'IC al 90%

$$I.C._{90\%} = \bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

$$I.C._{90\%} = 5.55 \pm 1.64 \times \frac{1.1}{\sqrt{40}} = 5.55 \pm 0.29 = (5.26, 5.84)$$

L'intervallo di confidenza ha minor ampiezza ma siamo meno certi di prima che questo intervallo "catturi" il valor vero

Conclusioni

1. Possiamo dire, con una buona confidenza (del 95%) che **l'ignoto parametro μ è compreso tra 5.21 e 5.89 mg/dl**

2. Si noti che: μ (il valor vero) è compreso oppure no nell'intervallo ….. non ci sono altre possibilità!

   Ciò che noi diciamo è che **siamo abbastanza certi** che **questo intervallo "catturi"** il valor vero

3. **L'affermazione che l'intervallo contiene il valor vero potrebbe anche essere falsa,** ma con probabilità del 5%, quindi piuttosto bassa