

Stima per intervallo

- In ogni stima è insito un certo margine di errore che è opportuno misurare. Ciò può essere fatto tramite un *intervallo di confidenza*.
- Dato un campione (X_1, X_2, \dots, X_n) estratto da una popolazione di cui interessa stimare il parametro θ , (L_1, L_2) è un *intervallo di confidenza* per θ al $(1 - \alpha)100\%$ se

$$P_{\theta}(L_1 < \theta < L_2) = 1 - \alpha, \quad \forall \theta.$$

dove L_1 ed L_2 sono funzioni del campione e quindi sono quantità aleatorie cioè prima di estrarre il campione. Si richiede cioè che la probabilità che l'intervallo includa θ sia pari a $(1 - \alpha)$

- Il **coefficiente fiduciario** o **livello di confidenza** $1 - \alpha$ è solitamente pari al 90%, 95% o 99%.
- Ampiezza dell'intervallo: $A = L_2 - L_1$.

- **ESEMPIO:** In base ad un'indagine campionaria, l'intervallo di confidenza al 95% per il consumo settimanale medio μ di un certo prodotto risulta pari a $[150, 200]$.

Possiamo dire che il consumo settimanale medio assume un valore fra 150 e 200 con probabilità 0.95, cioè

$$P(150 < \mu < 200) = 0.95 ?$$

No!, la probabilità è riferita all'intervallo aleatorio

Infatti per costruire un intervallo di confidenza fissiamo una regola $[L_1(X_1, \dots, X_n), L_2(X_1, \dots, X_n)]$ in base a cui calcoleremo l'intervallo, in funzione del campione (X_1, \dots, X_n) .

Un volta estratto un campione (x_1, \dots, x_n) , calcolo gli estremi dell'intervallo come $[L_1(x_1, \dots, x_n), L_2(x_1, \dots, x_n)]$.

A priori posso dire che la probabilità di estrarre un campione tale da fornire un intervallo che includa μ è $(1 - \alpha)$.

A posteriori, confido di aver estratto uno dei campioni che forniscono un intervallo che include μ .

METODO della QUANTITA' PIVOTALE

Per costruire IC un metodo è quello della quantità pivotale o ausiliaria

Una QP è una quantità aleatoria

- funzione del parametro di interesse e del campione
- la cui distribuzione non dipende dal parametro
- e che sia invertibile rispetto al parametro

Intervalli di confidenza per la media per una popolazione normale con varianza nota

Se $X \sim N(\mu, \sigma^2)$, con σ^2 noto, una QP per IC per μ è

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

infatti abbiamo che:

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha,$$

dove $z_{\alpha/2}$ è il quantile di ordine $(1 - \alpha/2)$ della distribuzione normale standardizzata cioè

$P(Z < z_{\alpha/2}) = 1 - \alpha/2$ con $Z \sim N(0, 1)$ ed invertendo rispetto al parametro otteniamo

- intervallo di confidenza (*centrale*) per μ al $(1 - \alpha)100\%$

$$(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n}).$$

dove $z_{\alpha/2}$ è il quantile di ordine $(1 - \alpha/2)$ della distribuzione normale standardizzata. Questo quantile lo si trova sulle tavole.

Si dice *centrale* perchè questo intervallo è centrato intorno alla media campionaria

- Ampiezza dell'intervallo: $A = L_2 - L_1 = 2z_{\alpha/2}\sigma/\sqrt{n}$.

Notiamo che a parità di altre condizioni:

- l'ampiezza dell'intervallo **diminuisce** all'aumentare dell'ampiezza del campione (n)
- l'ampiezza dell'intervallo **aumenta** all'aumentare della varianza della popolazione (σ^2)
- l'ampiezza dell'intervallo **aumenta** all'aumentare del livello di confidenza $(1 - \alpha)$

- Intervallo *non centrale* per μ al $(1 - \alpha)100\%$

$$(\bar{X} - z_1\sigma/\sqrt{n}, \bar{X} + z_2\sigma/\sqrt{n})$$

dove $P(z_1 < Z < z_2) = 1 - \alpha$ con $Z \sim N(0, 1)$.

- L'ampiezza dell'intervallo non centrale $A = (z_2 - z_1)\sigma/\sqrt{n}$ è superiore a quella dell'intervallo centrale. Per questo motivo in pratica in generale si usano intervalli centrali

ESEMPIO: In uno studio sul grado di informatizzazione delle aziende di un certo settore, si decide di sottoporre un questionario ad un campione casuale di aziende, in cui si chiede fra l'altro di indicare la spesa per servizi informatici.

(a) Sulla base di studi precedenti, si ritiene che la varianza della spesa per servizi informatici nella popolazione in esame sia $\sigma^2 = 400$. Utilizzando tale informazione, e assumendo l'ipotesi di normalità, qual è l'ampiezza campionaria necessaria per ottenere un intervallo di confidenza di livello 0.95 per la spesa media, di lunghezza non superiore a 10?

(b) Ora immaginate di aver fissato $n = 64$ e di aver osservato $\sum_{i=1}^n x_i = 768$. Determinate un intervallo di confidenza al 95% per la spesa media in servizi informatici.

RISPOSTA

Ipotizziamo che la spesa per servizi informatici sia descritta da un numero aleatorio $X \sim N(\mu, \sigma^2 = 400)$.

Poichè si chiede un intervallo di livello 0.95, abbiamo che $1 - \alpha = 0.95$ quindi $\alpha = 0.05$ e di conseguenza $\alpha/2 = 0.025$ quindi il quantile che occorre è $z_{0.025} = 1.96$

(a) La lunghezza dell'intervallo è

$$A = 2 z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} .$$

Vogliamo determinare n in modo che $A \leq 10$, per cui risolviamo in n la disequaglianza

$$2 * 1.96 * \frac{20}{\sqrt{n}} < 10 ,$$

ottenendo $n \geq 62$.

(b) La spesa media campionaria è $\bar{x} = \frac{768}{64} = 12$, e l'intervallo di confidenza risulta

$$(\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$$

quindi

$$(12 - 1.96 * 20/\sqrt{64}, 12 + 1.96 * 20/\sqrt{64}).$$

ovvero (7.1, 16.9) questo intervallo ha ampiezza $A = 9.8$

- Se avessi scelto un livello di confidenza pari a 0.99 cosa cambia?

Cambia solo $z_{\alpha/2}$ che diventa $z_{0.005} = 2.57$ e quindi l'intervallo diventa: $(5.57, 18.42)$ con ampiezza $A = 12.85$ maggiore rispetto a prima

- Se la deviazione standard passa da 20 a 10 come cambia l'intervallo di confidenza?

L'intervallo diventa: $(8.79; 15.21)$ con ampiezza minore e pari a $A = 6.42$

Intervalli di confidenza per la media per una popolazione normale con varianza NON nota

Se $X \sim N(\mu, \sigma^2)$, con σ^2 non noto, si utilizza la quantità ausiliaria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim \mathcal{T}^{(n-1)},$$

dove $\mathcal{T}^{(n-1)}$ indica la distribuzione t-student con $(n-1)$ gradi di libertà .

- abbiamo quindi che

$$P\left(-t_{\alpha/2}^{(n-1)} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}^{(n-1)}\right) = 1 - \alpha,$$

dove $P\left(T < t_{\alpha/2}^{(n-1)}\right) = 1 - \alpha/2$ con $T \sim t(n-1)$.

- Intervallo di confidenza per μ al $(1 - \alpha)100\%$

$$\left(\bar{X} - t_{\alpha/2}^{(n-1)} S/\sqrt{n}, \bar{X} + t_{\alpha/2}^{(n-1)} S/\sqrt{n}\right).$$

- Ampiezza dell'intervallo: $A = 2t_{\alpha/2}^{(n-1)} S/\sqrt{n}$
- L'ampiezza dell'intervallo dipende ancora in modo direttamente proporzionale dal livello di confidenza

Notiamo che poichè le code della distribuzione Normale sono meno pesanti rispetto alle code della distribuzione T-Student, a parità di altre condizioni, l'ampiezza dell'intervallo di confidenza con varianza non nota è **maggiore** rispetto all'ampiezza dell'intervallo di confidenza con varianza nota.

Questo è intuitivo visto che, non avendo informazioni sulla varianza, ma dovendola stimare, mi aspetto di ottenere un intervallo meno preciso (e quindi con ampiezza più ampia).

Nota però che al crescere dell'ampiezza del campione aumentano i gradi di libertà della distribuzione T-Student usata per costruire l'intervallo. Ricorda inoltre che la distribuzione T-Student tende alla distribuzione Normale al crescere dei gradi di libertà quindi l'intervallo di confidenza basato sulla distribuzione T-Student tende a quello basato sulla distribuzione normale al crescere dell'ampiezza del campione

ESERCIZIO: Supponiamo che il fatturato delle aziende di un certo settore sia descritto da un numero aleatorio X con distribuzione normale $N(\mu, \sigma^2)$. Per un campione casuale di 8 imprese osserviamo i seguenti valori del fatturato: (40, 35, 80, 20, 60, 70, 75, 45)

(a) Stimare il fatturato medio e la varianza del fatturato nella popolazione, usando stimatori non distorti.

(b) Fornite un intervallo di confidenza al 95% per μ .

RISPOSTA

(a) Uno stimatore non distorto per il fatturato medio μ è la media campionaria

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n},$$

e la stima è $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 53.125$

Uno stimatore non distorto per la varianza del fatturato σ^2 è la varianza campionaria corretta

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

e la stima è pari a $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = 456.7$.

(b) Per costruire l'intervallo di confidenza, poichè la varianza della popolazione non è nota, usiamo la formula

$$(\bar{X} - t_{\alpha/2}^{(n-1)} S / \sqrt{n}, \bar{X} + t_{\alpha/2}^{(n-1)} S / \sqrt{n}).$$

dove la distribuzione T-Student ha $n - 1 = 7$ gradi di libertà quindi $t_{\alpha/2}^{(n-1)} = 2.36$ e otteniamo

$$(53.125 - 2.36 * \sqrt{456.7} / \sqrt{8}, 53.125 + 2.36 * \sqrt{456.7} / \sqrt{8}).$$

quindi l'intervallo di confidenza ha estremi: **(35.24, 71.00)**

Se invece il campione fosse stato solo di 5 unità , a parità di altre condizioni (e in particolare a parità di media e varianza campionaria), avrei

$$(53.125 - 2.78 * \sqrt{456.7} / \sqrt{5}, 53.125 + 2.78 * \sqrt{456.7} / \sqrt{5}).$$

quindi l'intervallo di confidenza ha estremi: **(26.6, 79.7)** e risulta più ampio del precedente

Intervalli di confidenza per la media per una popolazione qualsiasi nel caso di grandi campioni

- se l'ampiezza del campione è suffic. grande ($n \geq 30$) abbiamo che, anche se la popolazione da cui è estratto il campione non è normale:

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

In virtù del teorema del limite centrale.

- Quindi l'intervallo di confidenza asintotico per μ a livello $(1 - \alpha)100\%$ è

$$(\bar{X} - z_{\alpha/2}S/\sqrt{n}, \bar{X} + z_{\alpha/2}S/\sqrt{n})$$

NB: se la varianza della popolazione fosse nota al posto di S sostituisco σ

Intervalli di confidenza per la media per una

Bernoulliana per grandi campioni

- Se $X \sim \text{Bin}(1, p)$, nel caso di grandi campioni

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - p}{\sqrt{\bar{X}(1 - \bar{X})/n}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

- Intervallo di confidenza per p al $(1 - \alpha)100\%$

$$\left(\bar{X} - z_{\alpha/2}\sqrt{\bar{X}(1 - \bar{X})/n}, \bar{X} + z_{\alpha/2}\sqrt{\bar{X}(1 - \bar{X})/n}\right).$$

ESERCIZIO

Supponiamo che in un'intervista ad un campione casuale di 100 studenti delle università italiane venga chiesto un giudizio sulla recente riforma universitaria. 20 studenti rispondono "insoddisfatto", 60 rispondono "indifferente", 20 rispondono "soddisfatto".

(a) Fornite una stima puntuale della proporzione degli studenti delle università italiane che si ritengono soddisfatti della riforma

(b) Calcolate inoltre un intervallo di confidenza asintotico al 95% per tale proporzione

RISPOSTA

(a) Indichiamo con θ la proporzione da stimare. La stima è data dalla proporzione campionaria $\bar{x} = \frac{20}{100} = 0.2$.

(b) Essendo $n = 100$, possiamo fornire un intervallo di confidenza asintotico usando la formula:

$$\left(\bar{X} - z_{\alpha/2} \sqrt{\bar{X}(1 - \bar{X})/n}, \bar{X} + z_{\alpha/2} \sqrt{\bar{X}(1 - \bar{X})/n} \right).$$

$$\bar{x} = 0.2 \quad , \quad z_{\frac{\alpha}{2}} = 1.96 \quad , \quad \sqrt{\frac{\bar{x}(1 - \bar{x})}{100}} = 0.04 \quad ,$$

perciò l'intervallo di confidenza asintotico è **[0.1216, 0.2784]**

ESERCIZIO

Voglio stimare la proporzione p dei clienti di una assicurazione che sono interessati all'offerta di una nuova polizza. A questo scopo decido di sottoporre un questionario ad un campione casuale di clienti.

(a) Qual è il modello che adattereste per il fenomeno studiato?

(b) Cosa sapete dire sull'ampiezza campionaria necessaria affinché l'errore quadratico medio dello stimatore della proporzione di clienti interessati sia inferiore a 0.005?

(c) Ora immaginate che in un campione casuale di 100 clienti, la proporzione degli interessati alla nuova polizza sia risultata pari a 0.10. Stimare la proporzione p e lo scarto quadratico medio dello stimatore utilizzato.

(d) Determinate un intervallo di confidenza asintotico al 99% per p

RISPOSTA

(a) La popolazione è bernoulliana, $X \sim \text{bern}(p)$, e il parametro p rappresenta la proporzione dei clienti dell'assicurazione

interessati all'offerta della nuova polizza.

(b) Lo stimatore della proporzione dei clienti interessati è la proporzione campionaria $\hat{p} = \sum_{i=1}^n X_i/n$. Essendo non distorto, il suo errore quadratico medio coincide con la sua varianza, data da $\theta(1-\theta)/n$. Non conoscendo θ , ma essendo

$$\theta(1-\theta)/n \leq 1/(4n)$$

(disegnate il grafico della funzione $\theta(1-\theta)$ per convincervi di questo: si tratta di una parabola con concavità rivolta verso il basso e vertice in $\theta = 0.5$, il valore della funzione nel vertice è pari a $1/4$), poniamo

$$\frac{1}{4n} < 0.005$$

cioè $n > 50$.

(b) La stima della proporzione dei clienti interessati è $\hat{p} = 0.10$, e la stima dello scarto quadratico medio dello stimatore utilizzato è

$$\sqrt{\frac{0.1 \cdot 0.9}{100}} = 0.03$$

Da non confondere con l'errore quadratico medio (che in questo caso coincide con la varianza) stimato, che è pari a

0.009 ed effettivamente è minore di 0.005 come mi aspettavo
(avendo estratto un campione di ampiezza $100 > 50$)

(c) Dalla formula per gli intervalli di confidenza asintotici per popolazione Bernoulliana, in cui $\bar{x} = \hat{p} = 0.1$ e $z_{0.005} = 2.576$, si trova l'intervallo di confidenza asintotico **[0.0227 , 0.1773]**

Intervalli di confidenza per la differenza tra due medie

ESEMPIO

X è la VA che rappresenta il prezzo di vendita di un titolo.

Y è la VA che rappresenta il prezzo di acquisto di un titolo.

Sono interessato al guadagno medio cioè alla differenza fra il prezzo medio di acquisto e di vendita

- Se $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$, con σ_1^2 e σ_2^2 *note*,

$$P\left(-z_{\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < z_{\alpha/2}\right) = 1 - \alpha,$$

dove $P(Z < z_{\alpha/2}) = 1 - \alpha/2$ con $Z \sim N(0, 1)$.

- Intervallo di confidenza per $\mu_1 - \mu_2$ al $(1 - \alpha)100\%$

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{X} - \bar{Y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

- Se $X \sim N(\mu_1, \sigma^2)$ e $Y \sim N(\mu_2, \sigma^2)$, con σ^2 *non nota*,

$$P\left(-t_{\alpha/2}^{n_1+n_2-2} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_c^2(1/n_1 + 1/n_2)}} < t_{\alpha/2}^{n_1+n_2-2}\right) = 1-\alpha,$$

dove

$$S_c^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

e

$$S_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{n_1 - 1}$$

e

$$S_2^2 = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_2 - 1}$$

e

$$P\left(T < t_{\alpha/2}^{n_1+n_2-2}\right) = 1 - \alpha/2 \text{ con } T \sim t(n_1 + n_2 - 2).$$

- Intervallo di confidenza per $\mu_1 - \mu_2$ al $(1 - \alpha)100\%$

$$\left(\bar{X} - \bar{Y} - t_{\alpha/2}^{n_1+n_2-2} S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X} - \bar{Y} + t_{\alpha/2}^{n_1+n_2-2} S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right).$$

ESERCIZIO

Si è rilevato il prezzo (in euro) di un prodotto in un campione casuale di punti vendita, in due diverse città . Indichiamo con (x_1, \dots, x_{25}) e con (y_1, \dots, y_{22}) i dati campionari rilevati nelle due città e supponiamo che

$$\bar{x} = 3.5 \quad , \quad \sum_{i=1}^{25} (x_i - \bar{x})^2 = 2.7 \quad , \quad \bar{y} = 3.8 \quad , \quad \sum_{i=1}^{22} (y_i - \bar{y})^2 = 1.35$$

In quale città il prezzo presenta maggior variabilità ?

Devo stimare il coefficiente di variazione. In R i comandi per fare questo sono:

```
> 2.7/24
```

```
[1] 0.1125
```

```
> 1.35/21
```

```
[1] 0.06428571
```

```
> sqrt(0.1125)/3.5
```

```
[1] 0.09583148
```

```
> sqrt(0.064)/3.8
```

```
[1] 0.06657427
```

Abbiamo quindi che $CV_1 = 0.096$, questo è il coeff. di

variazione del prezzo stimato nella prima città mentre $CV_2 = 0.067$, questo è il coeff. di variazione del prezzo stimato nella seconda città quindi, sulla base del mio campione, risulta che il prezzo è più variabile nella prima città

Determinate un intervallo di confidenza a livello 0.95 per la differenza fra i prezzi medi di quel prodotto nelle due città, specificando le ipotesi che introducete.

RISPOSTA

Indichiamo con X e con Y i prezzi (aleatori) nelle due città e supponiamo che

$$X \sim N(\mu_X, \sigma_X^2) \quad , \quad Y \sim N(\mu_Y, \sigma_Y^2) \quad ,$$

e che X e Y siano indipendenti. Supponiamo anche che le due varianze σ_X^2 e σ_Y^2 siano uguali fra loro (ma non note)

Sotto queste ipotesi, per determinare l'intervallo di confidenza per $\mu_X - \mu_Y$, possiamo utilizzare la formula

$$\left(\bar{X} - \bar{Y} - t_{\alpha/2} S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X} - \bar{Y} + t_{\alpha/2} S_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right).$$

Abbiamo

$$n_1 = 25, n_2 = 22, n_1 + n_2 - 2 = 45$$

$$S_c^2 = (2.7 + 1.35)/(25 + 22 - 2) = 0.09 ,$$

$$S_c = 0.3 ,$$

$$t_{0.025}^{(45)} = 2.014 ,$$

perciò l'intervallo di confidenza per la differenza fra i prezzi medi risulta $[-0.477 ; -0.123]$ che è sempre negativo quindi il costo è in media maggiore nella seconda città e posso dire questo con un livello di confidenza pari al 95%.

- Per due *popolazioni qualsiasi*,

nel caso di *grandi campioni* (varianze non nec. uguali):

$$n_1 \rightarrow \infty$$

$$n_2 \rightarrow \infty$$

$$P\left(-z_{\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

- Intervallo di confidenza asintotico per $\mu_1 - \mu_2$

a livello $(1 - \alpha)100\%$

$$\left(\bar{X} - \bar{Y} - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, \bar{X} - \bar{Y} + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right).$$

- Se $X \sim \text{Bin}(1, p_1)$ e $Y \sim \text{Bin}(1, p_2)$, nel caso di *grandi campioni*,

$$P\left(-z_{\alpha/2} < \frac{(\bar{X} - \bar{Y}) - (p_1 - p_2)}{\hat{\sigma}_{\bar{X}-\bar{Y}}} < z_{\alpha/2}\right) \approx 1 - \alpha.$$

dove $\hat{\sigma}_{\bar{X}-\bar{Y}}^2 = \bar{X}(1 - \bar{X})/n_1 + \bar{Y}(1 - \bar{Y})/n_2$

- Intervallo di confidenza per $p_1 - p_2$ al $(1 - \alpha)100\%$

$$(\bar{X} - \bar{Y} - z_{\alpha/2}\hat{\sigma}_{\bar{X}-\bar{Y}}, \bar{X} - \bar{Y} + z_{\alpha/2}\hat{\sigma}_{\bar{X}-\bar{Y}}).$$

ESERCIZIO

Vogliamo studiare la proporzione di persone che votano per la Lega a Varese e a Roma. Estraggo un campione di 30 persone a Varese e 100 a Roma e chiedo se voteranno per la Lega alle prossime elezioni.

A Varese il 40% delle persone intervistate si dice favorevole alla Lega. A Roma solo il 5% delle persone intervistate si dice favorevole alla Lega.

Costruite un intervallo di confidenza per la differenza fra le due proporzioni ad un livello di confidenza pari al 95%.

Abbiamo che:

$$n_1 = 30, n_2 = 100;$$

$$\bar{X} = 0.4, \bar{Y} = 0.05;$$

$$z_{\alpha/2} = 1.96$$

In R i comandi per ottenere l'intervallo di confidenza sono:

```
> 0.35+1.96*sqrt(0.4*0.6/30+0.05*0.95/100)
```

```
[1] 0.5304371
```

```
> 0.35-1.96*sqrt(0.4*0.6/30+0.05*0.95/100)
```

```
[1] 0.1695629
```

Abbiamo quindi che l'intervallo di confidenza ha estremo inferiore pari a 17% ed estremo superiore pari a 53%. Essendo l'intervallo di confidenza tutto positivo concludo, con un livello di confidenza pari al 95%, che la proporzione di persone che votano per la Lega è maggiore a Varese rispetto a Roma.

Intervalli di confidenza per la varianza

- Se $X \sim N(\mu, \sigma^2)$, con μ e σ^2 non noti,

$$P\left(\chi_{1-\alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{\alpha/2}^2\right) = 1 - \alpha,$$

dove $\chi_{1-\alpha/2}^2$ e $\chi_{\alpha/2}^2$ sono i centili della distribuzione $\chi^2(n-$

1) tali che

$$P(\chi^2 \leq \chi_{1-\alpha/2}^2) = P(\chi^2 > \chi_{\alpha/2}^2) = \frac{\alpha}{2}.$$

- Intervallo di confidenza per σ^2 al $(1 - \alpha)100\%$

$$\left((n-1)S^2/\chi_{\alpha/2}^2, (n-1)S^2/\chi_{1-\alpha/2}^2\right).$$

Esercizio.

Estraggo un campione casuale di ampiezza 30 da una popolazione normale. Nel campione abbiamo che $\sum(X_i - \bar{X})^2 = 90$. Costruite un intervallo di confidenza a livello 95 % e uno a livello 99%.

A livello 95% abbiamo che i quantili della distribuzione χ^2 con 29 gradi di libertà sono:

$$\chi_{\alpha/2}^2 = 45.7, \quad \chi_{1-\alpha/2}^2 = 16$$

quindi

$$L_1 = 90/45.7 = 1.97$$

$$L_2 = 90/16 = 5.6$$

e quindi l'intervallo è [1.97; 5.6]

A livello 99% abbiamo che:

$$\chi_{\alpha/2}^2 = 52, \quad \chi_{1-\alpha/2}^2 = 13$$

$$L_1 = 90/52 = 1.7$$

$$L_2 = 90/13 = 6.9$$

quindi l'intervallo è [1.7; 6.9]

Come prevedibile il secondo intervallo di confidenza è più ampio del primo.