Esame di Probabilità e Statistica [3231]

Esame di Calcolo delle Probabilità e Statistica [2959]

Corso di Studi di Ingegneria Gestionale (D.M.270/04) (L)

Dipartimento di Meccanica, Matematica e Management
Politecnico di Bari

Cognome: ____________________ Docente: Gianluca Orlando
Nome: ____________________ Appello: settembre 2023 - I
Matricola: ____________________ Data: 05/09/2023

Tempo massimo: 2 ore.

Esercizio 1. (6 punti) Ci si aspetta che la resa percentuale di una reazione chimica sia correlata con la temperatura a cui avviene la reazione. Nella seguente tabella sono rappresentati i dati di un campione:

temperatura ($°C$)	45	55	60	70	85	90
resa (%)	43	48	51	57	66	68

1. Rappresentare i dati in uno scatterplot.
2. Determinare (derivando le formule) la retta di regressione lineare e rappresentarla.
3. Determinare il coefficiente di correlazione lineare.

Soluzione. 1. Segue lo scatterplot (con la retta di regressione lineare):

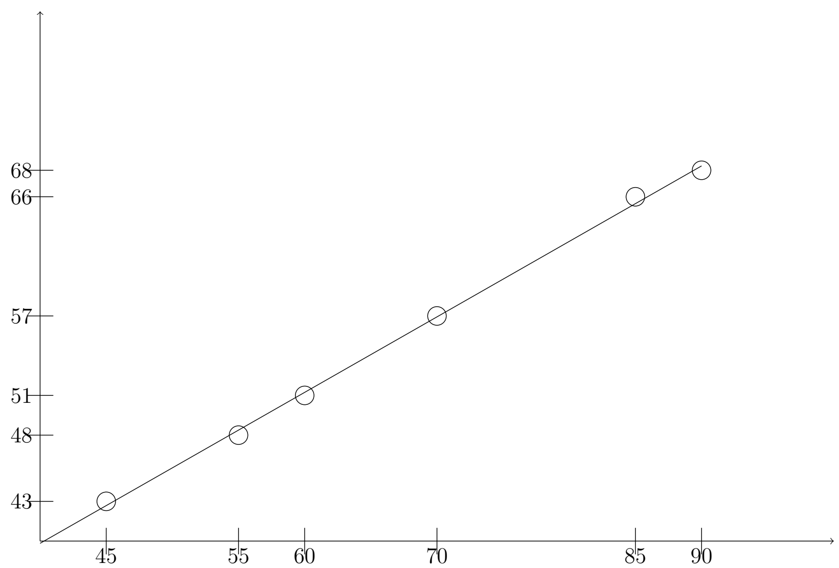

Figura 1: Scatterplot e retta di regressione lineare.

2. Denotiamo con $(x_1, y_1), \dots, (x_n, y_n)$, $n = 6$, i dati del campione. Cerchiamo la retta di equazione

$$y = ax + b$$

che meglio approssima i dati, utilizzando il metodo dei minimi quadrati. Vogliamo minimizzare l'errore

$$r(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \,.$$

Imponiamo che il gradiente rispetto ad (a, b) sia nullo, ovvero,

$$\begin{aligned} 0 = \partial_a r(a,b) &= -2 \sum_{i=1}^{n} (y_i - ax_i - b) x_i = -2 \sum_{i=1}^{n} (x_i y_i - a x_i^2 - b x_i) \,, \\ 0 = \partial_b r(a,b) &= -2 \sum_{i=1}^{n} (y_i - ax_i - b) \end{aligned}$$

Dalla seconda equazione segue che

$$nb = \sum_{i=1}^{n} (y_i - ax_i) \implies b = \overline{y} - a\overline{x} \,.$$

Sostituendo nella prima,

$$\begin{aligned} \sum_{i=1}^{n} (x_i y_i - a x_i^2 - b x_i) = 0 &\implies \sum_{i=1}^{n} (x_i y_i - a x_i^2 - x_i \overline{y} + a \overline{x} x_i) = 0 \\ &\implies a\Big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \Big) = \sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y} \\ &\implies a = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \,. \end{aligned}$$

Completiamo la tabella con i valori necessari a calcolare a e b:

							somma
x_i	45	55	60	70	85	90	405
y_i	43	48	51	57	66	68	333
x_i^2	2025	3025	3600	4900	7225	8100	28875
y_i^2	1849	2304	2601	3249	4356	4624	18983
$x_i y_i$	1935	2640	3060	3990	5610	6120	23355

Pertanto $\overline{x} = 405/6 = 67.5$ e $\overline{y} = 55.5$. Segue che

$$\begin{aligned} a &= \frac{23355 - 6 \cdot 67.5 \cdot 55.5}{28875 - 6 \cdot 67.5^2} = \frac{877.5}{1537.5} \simeq 0.57 \,, \\ b &= 55.5 - 0.57 \cdot 67.5 \simeq 17.03 \,, \end{aligned}$$

ovvero, la retta di regressione lineare ha equazione

$$y = 0.57x + 17.03 \,.$$

3. Per calcolare il coefficiente di correlazione lineare usiamo la formula

$$\begin{aligned} \rho_{x,y} &= \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}\sqrt{\sum_{i=1}^{n} y_i^2 - n\overline{y}^2}} = \frac{23355 - 6 \cdot 67.5 \cdot 55.5}{\sqrt{28875 - 6 \cdot 67.5^2}\sqrt{18983 - 6 \cdot 55.5^2}} \\ &= \frac{877.5}{\sqrt{1537.5}\sqrt{501.5}} \simeq 0.9993 \,. \end{aligned}$$

Esercizio 2. (7 punti) Un'azienda produce *cookies* con gocce di cioccolato. Si assuma che il numero di gocce di cioccolato in un *cookie* sia distribuito con legge di Poisson e che, in media, un *cookie* abbia 10 gocce di cioccolato al suo interno. Si assuma che i numeri di gocce di cioccolato in *cookies* diversi siano indipendenti.

1. Qual è la probabilità che il numero di gocce di cioccolato in un *cookie* sia compreso tra 8 e 12 (estremi inclusi)?

2. Qual è la probabilità che in due *cookies* ci siano in tutto esattamente 30 gocce di cioccolato?

Una confezione di *cookies* contiene 6 *cookies*.

3. Compriamo una confezione. Qual è la probabilità che almeno 3 *cookies* nella confezione contengano ciascuno un numero gocce di cioccolato compreso tra 8 e 12?

4. In media, in una confezione quanti *cookies* contengono un numero gocce di cioccolato compreso tra 8 e 12?

Soluzione. Consideriamo la variabile aleatoria

$$X = \text{“numero di gocce di cioccolato in un } \textit{cookie}\text{''} \sim \mathrm{P}(\lambda)\,.$$

Ricordiamo che

$$\mathbb{E}(X) = \lambda$$

per una variabile distribuita con legge di Poisson, quindi $\lambda = 10$ dalle ipotesi del problema. Ricordiamo che

$$\mathbb{P}(\{X = k\}) = e^{-\lambda}\frac{\lambda^k}{k!}\,.$$

1. Ci viene chiesto di calcolare

$$\begin{aligned}
&\mathbb{P}(\{8 \le X \le 12\}) \\
&= \mathbb{P}(\{X = 8\}) + \mathbb{P}(\{X = 9\}) + \mathbb{P}(\{X = 10\}) + \mathbb{P}(\{X = 11\}) + \mathbb{P}(\{X = 12\}) \\
&= e^{-10}\Big(\frac{10^8}{8!} + \frac{10^9}{9!} + \frac{10^{10}}{10!} + \frac{10^{11}}{11!} + \frac{10^{12}}{12!}\Big) \simeq 57.13\%
\end{aligned}$$

2. Consideriamo la variabile aleatoria

$$Y = \text{“numero di gocce di cioccolato nel secondo } \textit{cookie}\text{''} \sim \mathrm{P}(\lambda)\,.$$

La traccia dice che X e Y sono indipendenti. Quindi, per una proprietà delle variabili aleatorie distribuite con legge di Poisson, abbiamo che la somma è ancora distribuita con legge di Poisson con parametro dato dalla somma dei parametri, ovvero $X + Y \sim \mathrm{P}(2\lambda) = \mathrm{P}(20)$. Possiamo allora calcolare

$$\mathbb{P}(\{X + Y = 30\}) = e^{-20}\frac{20^{30}}{30!} = 0.8\%\,.$$

3. Consideriamo come successo l'evento “il numero di gocce di cioccolato nel *cookie* è compreso tra 8 e 12”. Questo successo ha probabilità

$$p = \mathbb{P}(\{8 \le X \le 12\}) = 57.13\%\,.$$

Consideriamo la variabile aleatoria

$$Z = \text{“numero di successi in 6 tentativi”} \sim \mathrm{B}(6, p)\,.$$

Questa è distribuita come una legge binomiale perché i tentativi sono indipendenti.

Ci viene chiesto di calcolare

$$\begin{aligned}\mathbb{P}(\{Z \geq 3\}) &= 1 - \mathbb{P}(\{Z \leq 2\}) = 1 - \mathbb{P}(\{Z = 0\}) - \mathbb{P}(\{Z = 1\}) - \mathbb{P}(\{Z = 2\}) \\ &= 1 - \binom{6}{0} p^0 (1-p)^6 - \binom{6}{1} p^1 (1-p)^5 - \binom{6}{2} p^2 (1-p)^4 = 77.88\%\,.\end{aligned}$$

4. Il valore atteso della binomiale è $\mathbb{E}(Z) = np = 6 \cdot 57.13\% \simeq 3.43$.

Esercizio 3. (7 punti) Ad un ristorante il tempo di attesa al tavolo dal momento dell'ordine all'arrivo del piatto dipende dal piatto ordinato.

- Se si ordina la pizza, il tempo di attesa è distribuito con legge esponenziale con media 15 minuti.
- Se si ordinano gli spaghetti all'assassina, il tempo di attesa è distribuito con legge esponenziale con una deviazione standard di 30 minuti.
- Il 60% delle persone ordina la pizza, mentre il 40% delle persone ordina gli spaghetti all'assassina.

Al tuo tavolo siete due persone. Si assumano i tempi di preparazione dei vari piatti indipendenti.

1. Tu ordini una pizza. Qual è la probabilità che la pizza sia pronta in un tempo compreso tra i 10 e i 20 minuti?
2. Il/la tuo/a partner ordina spaghetti all'assassina. Qual è la probabilità che gli spaghetti siano pronti dopo 40 minuti?
3. Il servizio porterà al tavolo la pizza e gli spaghetti insieme (quindi il tempo di attesa totale dipenderà da quale preparazione impiega più tempo ad essere pronta). Qual è la probabilità di dover aspettare più di 40 minuti perché arrivino entrambi i piatti?
4. Una persona sola al tavolo a fianco si sta lamentando del servizio perché aspetta da 40 minuti e ancora non è arrivato il suo piatto. Qual è la probabilità che abbia ordinato una pizza?

Soluzione. Consideriamo le seguenti variabili aleatorie:

$$X_{\text{pizza}} = \text{“tempo di attesa se si ordina la pizza”} \sim \mathrm{Exp}(\lambda)$$

$$X_{\text{spagh}} = \text{“tempo di attesa se si ordinano gli spaghetti all'assassina”} \sim \mathrm{Exp}(\mu)$$

$$X = \text{“tempo di attesa (senza conoscere l'ordine)”}$$

$$Y \sim \mathrm{Be}(60\%)\,,\ \text{dove } \{Y = 1\} \text{ se viene ordinata pizza}$$

1. Ricordiamo che per la legge esponenziale si ha $\mathbb{E}(X_{\text{pizza}}) = \frac{1}{\lambda} = 15$ da cui segue che $\lambda = \frac{1}{15}$. Viene chiesto di calcolare

$$\begin{aligned}\mathbb{P}(\{10 \leq X_{\text{pizza}} \leq 20\}) &= \int_{10}^{20} \lambda e^{-\lambda x}\, \mathrm{d}x = \Big[-e^{-\lambda x} \Big]_{x=10}^{x=20} = e^{-10\lambda} - e^{-20\lambda} \\ &= e^{-\frac{10}{15}} - e^{-\frac{20}{15}} \simeq 24.98\%\,.\end{aligned}$$

2. Ricordiamo che per la legge esponenziale si ha $\mathrm{Var}(X_{\text{spagh}}) = \frac{1}{\mu^2}$. La traccia fornisce il valore della deviazione standard, quindi $\frac{1}{\mu^2} = 30^2$, da cui $\mu = \frac{1}{30}$. Possiamo allora calcolare

$$\mathbb{P}(\{X_{\text{spagh}} \geq 40\}) = e^{-40\mu} = e^{-\frac{40}{30}} \simeq 26.36\% \,.$$

3. Il tempo di attesa sarà dato dal massimo tra X_{pizza} e X_{spagh}, poiché bisogna aspettare che entrambi i piatti siano pronti. Ci viene chiesto di calcolare

$$\mathbb{P}(\{\max\{X_{\text{pizza}}, X_{\text{spagh}}\} \geq 40\}) \,.$$

Conviene calcolare la probabilità dell'evento complementare. Infatti $\max\{X_{\text{pizza}}, X_{\text{spagh}}\} < 40$ quando entrambi $X_{\text{pizza}} < 40$ e $X_{\text{spagh}} < 40$. Quindi, usando l'indipendenza

$$\begin{aligned}
&\mathbb{P}(\{\max\{X_{\text{pizza}}, X_{\text{spagh}}\} \geq 40\}) = 1 - \mathbb{P}(\{\max\{X_{\text{pizza}}, X_{\text{spagh}}\} < 40\}) \\
&= 1 - \mathbb{P}(\{X_{\text{pizza}} < 40\} \cap \{X_{\text{spagh}} < 40\}) = 1 - \mathbb{P}(\{X_{\text{pizza}} < 40\})\mathbb{P}(\{X_{\text{spagh}} < 40\}) \\
&= 1 - (1 - e^{-\frac{40}{15}})(1 - e^{-\frac{40}{30}}) \simeq 31.48\% \,.
\end{aligned}$$

4. Sappiamo che si è verificato l'evento $X \geq 40$. Calcoliamo

$$\begin{aligned}
&\mathbb{P}(\{Y = 1\}|\{X \geq 40\}) = \frac{\mathbb{P}(\{Y = 1\} \cap \{X \geq 40\})}{\mathbb{P}(\{X \geq 40\})} = \frac{\mathbb{P}(\{X \geq 40\}|\{Y = 1\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{X \geq 40\})} \\
&= \frac{\mathbb{P}(\{X_{\text{pizza}} \geq 40\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{X \geq 40\}|\{Y = 1\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{X \geq 40\}|\{Y = 0\})\mathbb{P}(\{Y = 0\})} \\
&= \frac{\mathbb{P}(\{X_{\text{pizza}} \geq 40\})\mathbb{P}(\{Y = 1\})}{\mathbb{P}(\{X_{\text{pizza}} \geq 40\})\mathbb{P}(\{Y = 1\}) + \mathbb{P}(\{X_{\text{spagh}} \geq 40\})\mathbb{P}(\{Y = 0\})} \\
&= \frac{e^{-\frac{40}{15}} 60\%}{e^{-\frac{40}{15}} 60\% + e^{-\frac{40}{30}} 40\%} \simeq 28.34\% \,.
\end{aligned}$$

Esercizio 4. (8 punti) Un responsabile di controllo di qualità misura la distanza tra le piastre degli scambiatori di calore prodotti da un'azienda. Le misurazioni su un campione di 20 scambiatori di calore forniscono una distanza media calcolata sul campione di 4.0 mm e una deviazione standard calcolata sul campione di 0.1 mm. Si assuma che la distribuzione della distanza tra le piastre sia normale.

1. Calcolare sui dati un intervallo di confidenza al 95% bilaterale per la distanza media.

2. Calcolare sui dati un intervallo di confidenza al 90% bilaterale per la varianza della distanza.

Derivare il calcolo degli intervalli di confidenza in entrambi i punti, motivando i passaggi.

Soluzione. Abbiamo a che fare con una popolazione $X \sim \mathcal{N}(\mu, \sigma^2)$ e un campione casuale $X_1, \ldots, X_n$ estratto dalla popolazione con $n = 20$. La media μ e la varianza σ^2 non sono note.
1. Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \leq \mu \leq V_n\}) \,.$$

dove $\beta = 95\%$. Per stimare μ sfrutteremo lo stimatore media campionaria $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché

$X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $T_{n-1} = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$ ha distribuzione t-Student con $n-1 = 19$ gradi di libertà. Allora

$$\begin{aligned}\beta = \mathbb{P}(\{U_n \le \mu \le V_n\}) &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\ &= \mathbb{P}\Big(\Big\{\frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} \le T_{n-1} \le \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) \\ &= 1 - \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) - \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big).\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) + \mathbb{P}\Big(\Big\{T_{n-1} > \frac{\overline{X}_n - U_n}{S_n/\sqrt{n}}\Big\}\Big) = 1 - \beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \mathbb{P}\Big(\Big\{T_{n-1} < \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}}\Big\}\Big) = \frac{\alpha}{2}\,.$$

Definiamo $t_{n-1,\alpha/2}$ come il punto tale che

$$\mathbb{P}(\{T_{n-1} \ge t_{n-1,\alpha/2}) = \frac{\alpha}{2}\,.$$

Scegliendo

$$\begin{aligned}\frac{\overline{X}_n - U_n}{S_n/\sqrt{n}} &= t_{n-1,\alpha/2} \implies U_n = \overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\,, \\ \frac{\overline{X}_n - V_n}{S_n/\sqrt{n}} &= -t_{n-1,\alpha/2} \implies V_n = \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\,,\end{aligned}$$

si ottiene la condizione che definisce l'IC. In conclusione

$$\Big[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}, \overline{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1,\alpha/2}\Big]$$

è un intervallo di confidenza bilaterale per μ con livello di confidenza $\beta = 1 - \alpha$.

Calcoliamo la realizzazione dell'IC sui dati. La traccia fornisce la media calcolata sui dati $\overline{x}_n = 4.0$ e la deviazione standard calcolata sui dati $s_n = 0.1$. Calcoliamo, utilizzando la tabella,

$$t_{n-1,\alpha/2} = t_{19,0.025} \simeq 2.093\,.$$

Segue che l'IC calcolato sui dati è

$$\Big[4.0 - \frac{0.1}{\sqrt{20}} 2.093, 4.0 + \frac{0.1}{\sqrt{20}} 2.093\Big] \simeq [3.95, 4.05]\,.$$

2. Dalla definizione di IC si ha che

$$\beta = \mathbb{P}(\{U_n \le \sigma^2 \le V_n\})\,,$$

con $\beta = 90\%$. Per stimare σ^2 sfrutteremo lo stimatore varianza campionaria $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Poiché $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ sono indipendenti, $Q_{n-1} = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ ha distribuzione chi-quadro con $n-1 = 19$ gradi di libertà. Allora

$$\begin{aligned}\beta = \mathbb{P}(\{U_n \le \sigma^2 \le V_n\}) &= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \le \frac{(n-1)S_n^2}{\sigma^2} \le \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\ &= \mathbb{P}\Big(\Big\{\frac{(n-1)S_n^2}{V_n} \le Q_{n-1} \le \frac{(n-1)S_n^2}{U_n}\Big\}\Big) \\ &= 1 - \mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) - \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big).\end{aligned}$$

Segue che

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) + \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = 1-\beta = \alpha\,.$$

Decidiamo di equipartire α:

$$\mathbb{P}\Big(\Big\{Q_{n-1} < \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = \mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,,$$

da cui

$$\mathbb{P}\Big(\Big\{Q_{n-1} > \frac{(n-1)S_n^2}{U_n}\Big\}\Big) = \frac{\alpha}{2}\,, \quad \mathbb{P}\Big(\Big\{Q_{n-1} \geq \frac{(n-1)S_n^2}{V_n}\Big\}\Big) = 1-\frac{\alpha}{2}\,.$$

Definiamo $\chi^2_{n-1,\alpha/2}$ e $\chi^2_{n-1,1-\alpha/2}$ come i punti tali che

$$\mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,\alpha/2}\}) = \frac{\alpha}{2}\,, \quad \mathbb{P}(\{Q_{n-1} \geq \chi^2_{n-1,1-\alpha/2}\}) = 1-\frac{\alpha}{2}\,.$$

Scegliendo

$$\frac{(n-1)S_n^2}{U_n} = \chi^2_{n-1,\alpha/2} \implies U_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}\,,$$

$$\frac{(n-1)S_n^2}{V_n} = \chi^2_{n-1,1-\alpha/2} \implies V_n = \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\,,$$

si ottiene la condizione che definisce l'IC. In conclusione

$$\Big[\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\Big]$$

è un intervallo di confidenza bilaterale per σ^2 con livello di confidenza $\beta = 1-\alpha$.

Per calcolare la realizzazione dell'IC sui dati, usiamo la varianza campionaria $s_n^2 = 0.1^2 = 0.01$ e i valori delle tavole

$$\chi^2_{19,0.05} = 30.144\,, \quad \chi^2_{19,0.095} = 10.117\,.$$

Segue che

$$\Big[\frac{19\cdot 0.01}{30.144}, \frac{19\cdot 0.01}{10.117}\Big] \simeq [0.0063, 0.0188]\,.$$