

obbligatorio - n. iscrizione sulla lista se non ve lo ricordate
siete fritti; o no?

il presente elaborato si compone di 5 (cinque) pagine

Cognome _____ Nome _____ matr.n. _____

Avvertimento: nello svolgimento degli esercizi se una quantità è codificata dal simbolo x continuare a chiamarla x , e se si chiama W non chiamarla X , e se si chiama t_i non chiamarla x_i , e se si chiama \bar{T}_n non indicarla con \bar{X}_n . Se volete scrivere dipendenti non scrivete indipendenti, se dovete sottrarre non sommate. Si fanno troppi errori di distrazione. Mi raccomando: concentrazione.

Si consiglia di lavorare con 3 decimali, arrotondando opportunamente.

Problema 1. Maya lavora per il teatro Orion, dove sono stati già venduti tutti i 500 biglietti per la prima de La Dea Scarlatta. Maya tuttavia sa che, su 1000 posti venduti, tipicamente ne rimane 1 vuoto.

Sono ammessi sia calcoli esatti, sia calcoli semplificati basati su opportune approssimazioni, sia mere limitazioni basate sulla disuguaglianza di Chebyshev.

2 1.1 Quale distribuzione ha X_k , la variabile che indica se il k -esimo posto in teatro rimane vuoto?
 $X_k \sim B(p)$, dove $p = 1/1000 = 0.001$.

2 1.2 Quale relazione ha Y , il numero totale di posti vuoti, con le variabili X_k ? Quale distribuzione ha Y ?

$$Y = \sum_{k=1}^{500} X_k \sim B(500, 0.001).$$

2 1.3 Qual è il numero atteso di posti vuoti?

$$\mathbb{E}[Y] = 500 \cdot 0.001 = 0.5.$$

2 1.4 Qual è la probabilità che Maya possa far sedere in teatro la sua amica Ayumi, rimasta senza biglietto?

$$\mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(Y = 0) = 1 - 0.999^{500} = 0.3936.$$

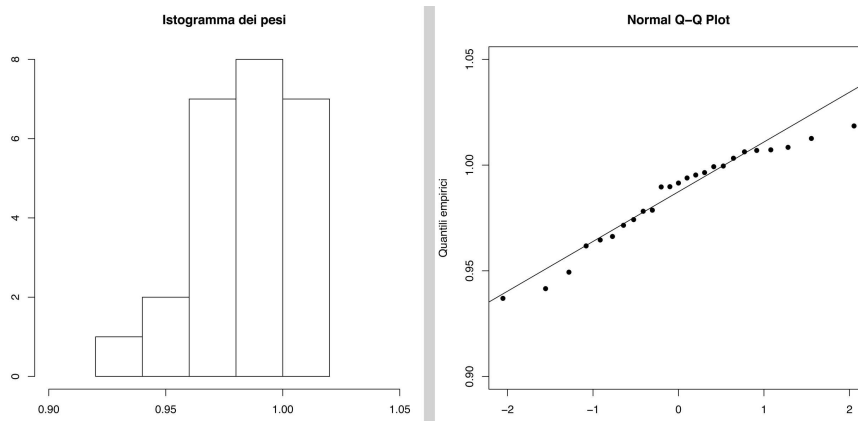
1 1.5 Qual è il numero più probabile di posti vuoti?

Dato $\mathbb{P}(Y = 0) = 0.6064 > 0.5$, per forza di cose il valore più probabile è $Y = 0$.

1 1.6 Qual è la probabilità che rimanga vuoto almeno mezzo teatro?

$$\mathbb{P}(Y \geq 250) = \mathbb{P}(Y - 0.5 \geq 249.5) = \mathbb{P}[|Y - 0.5| \geq 249.5] \leq \frac{\text{var}(Y)}{(249.5)^2} = 8 \cdot 10^{-6}.$$

Problema 2. Il proprietario di una gelateria vuole essere sicuro che i suoi dipendenti riempiano le vaschette “grandi” con una quantità di gelato mediamente inferiore ad 1 kg. Ha quindi registrato i pesi x_k delle ultime 25 vaschette “grandi” vendute, trovando dati così distribuiti



Istogramma (sinistra) e NPP (destra) dei dati

con una media campionaria ed una deviazione standard campionaria pari a

$$\bar{x}_{25} = 0.9857 \text{ kg}, \quad s_{25} = 0.0225 \text{ kg}.$$

1 **2.1** Fornire una stima puntuale della media e della deviazione standard del peso delle vaschette “grandi” vendute dalla gelateria.

Siano μ e σ la media e la deviazione standard del peso delle vaschette “grandi” vendute dalla gelateria.

$$\hat{\mu} = \bar{x}_{25} = 0.9857 \text{ kg}, \quad \hat{\sigma} = s_{25} = 0.0225 \text{ kg}.$$

3 **2.2** Fornire, se possibile, una stima intervallare al 99% della media e della deviazione standard del peso delle vaschette “grandi” vendute dalla gelateria.

Anche se il campione non è particolarmente numeroso ($n = 25 < 30$), il QQplot conferma la normalità della popolazione dei pesi delle vaschette “grandi” vendute dalla gelateria. Quindi

IC al 99% per μ :

$$\bar{x}_{25} \pm \frac{s_{25}}{\sqrt{25}} t_{24}(0.995) = 0.9857 \pm 0.0126 = (0.9731; 0.9983)$$

IC al 99% per σ^2 :

$$\left(\frac{24}{\chi_{24}^2(0,995)} s_{25}^2, \frac{24}{\chi_{24}^2(0,995)} s_{25}^2 \right) = (0.000267; 0.00123)$$

IC al 99% per σ :

$$(0.0163; 0.0351)$$

Non rassicurato da mere stime, il proprietario vi chiede di “quantificare” con un p-value l’evidenza statistica con cui le 25 vaschette controllate possono provare che il peso medio di tutte le vaschette “grandi” vendute sia inferiore ad 1 kg.

2 **2.3** Impostare un opportuno test statistico per rispondere al quesito del proprietario, specificando: ipotesi nulla, ipotesi alternativa, regione critica di livello α , condizioni di applicabilità del test.

$$H_0 : \mu \geq 1 \quad \text{vs} \quad H_1 : \mu < 1$$

$$R_\alpha : \bar{x}_{25} < 1 - \frac{s_{25}}{\sqrt{25}} t_{24}(\gamma)$$

che si può utilizzare (con livello sotto controllo) se il campione proviene da una popolazione è normale (il nostro caso) o se è numeroso.

2 **2.4** Calcolare il p-value dei dati raccolti.
Con le tavole

$$t_{24}(0.995) = 2.797 < \frac{1 - \bar{x}_{25}}{s_{25}} \sqrt{25} = 3.18$$

p-value < 0.05

2 mentre con conto esatto: p-value = 0.002.

2.5 Trarre le debite conclusioni. Sono forti o deboli?

Il p-value è inferiore agli usuali livelli di significatività per cui si può concludere che si rifiuta H_0 cioè si accetta $\mu < 1$ (conclusione forte).

Problema 3.

La Electron S.p.A. è un'azienda che produce componenti elettroniche, che sta testando un nuovo tipo di processori. La temperatura dei processori va tenuta monitorata, in modo da non comprometterne il funzionamento. In particolare, la ditta è interessata alla relazione tra la temperatura dei processori Y , la temperatura esterna T e il numero di processi che sta gestendo il processore N . Vengono raccolti i dati relativi a 20 processori, ottenendo i seguenti dati di sintesi (media e varianza campionaria di ogni variabile):

$$\bar{y} = 57.1628, s_y^2 = 64.0205, \quad \bar{t} = 29.1386, s_t^2 = 20.0575, \quad \bar{n} = 6.3, s_n^2 = 12.6421$$

Vengono proposti gli output di R del modello di regressione, il grafico dei residui e i p-value del test di Shapiro-Wilk per i residui per ciascuno dei seguenti modelli:

Modello 1

$$Y = \beta_0 + \beta_1 T + \beta_2 N + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

```
call:
lm(formula = Y ~ T + N)

Residuals:
    Min       1Q   Median       3Q      Max
-2.68410 -1.06505 -0.01521  0.98577  2.15461

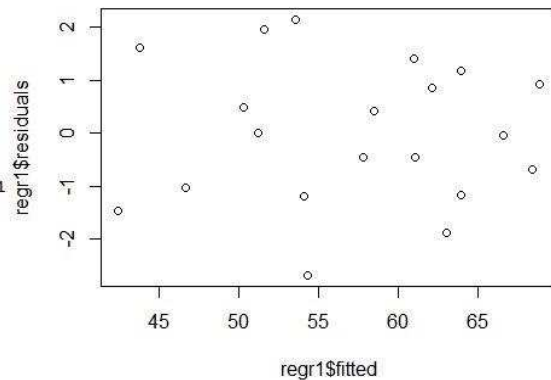
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.57371    2.36586   1.933   0.070 .
T            1.77796    0.07404  24.012 1.48e-14 ***
N            0.12409    0.09326   1.331   0.201
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.417 on 17 degrees of freedom
Multiple R-squared:  0.9719, Adjusted R-squared:  0.9686
F-statistic: 294.2 on 2 and 17 DF,  p-value: 6.476e-14

> shapiro.test(regr1$residuals)

Shapiro-wilk normality test

data:  regr1$residuals
W = 0.9777, p-value = 0.9017
```



Output dell'analisi per il modello 1

Modello 2

$$Y = \beta_0 + \beta_1 T + \epsilon \quad \epsilon \sim N(0, \sigma^2)$$

```
call:
lm(formula = Y ~ T)

Residuals:
    Min       1Q   Median       3Q      Max
-3.3658 -0.6532  0.2742  0.9621  1.8334

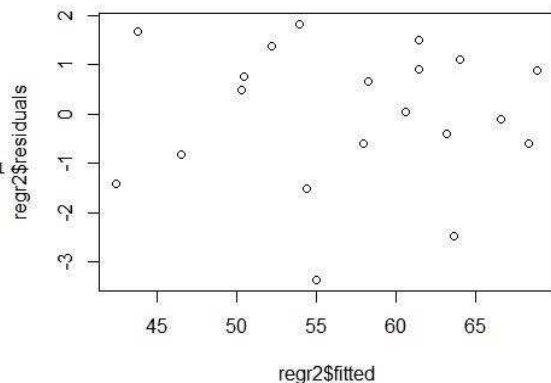
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.91778    2.18463   2.709  0.0144 *
T            1.75866    0.07415  23.719 4.98e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 18 degrees of freedom
Multiple R-squared:  0.969, Adjusted R-squared:  0.9673
F-statistic: 562.6 on 1 and 18 DF,  p-value: 4.979e-15

> shapiro.test(regr2$residuals)

Shapiro-wilk normality test

data:  regr2$residuals
W = 0.9375, p-value = 0.2148
```



Output dell'analisi per il modello 2

4 **3.1** *Si commenti la bontà dei modelli proposti e si scelga quello che meglio descrive il problema in esame.*
 Il Modello 1 ha un R^2 adjusted molto elevato, i residui non presentano particolari trend. Considerando il p-value del test di Shapiro-Wilks, l'ipotesi della normalità dei residui non è rifiutata a tutti i livelli usuali. Pertanto è possibile considerare i test di significatività proposti nell'output di R. Il modello è globalmente significativo (p-value: $6.476 \cdot 10^{-14}$), ma il coefficiente β_2 risulta non significativamente diverso da 0 (p-value: 0.201), e anche l'intercetta del modello potrebbe non essere significativamente diversa da zero (p-value: 0.070). Sarebbe opportuno eliminare il predittore N dal modello. Il Modello 2, ottenuto proprio eliminando N , presenta le stesse buone caratteristiche del Modello 1, ma in questo caso tutti i predittori risultano significativi. È un modello più semplice del Modello 1, avendo un regressore in meno, ma ha le stesse performances. Per questi motivi è opportuno scegliere il Modello 2.

2 **3.2** *Si scriva l'equazione di regressione stimata per il modello prescelto.*
 L'equazione stimata per il modello 2 è la seguente:

$$\hat{Y} = 5.91778 + 1.75866 \times T$$

2 **3.3** *Un processore dello stesso tipo di quelli testati viene montato su una videocamera da installare all'esterno. Sapendo che oggi la temperatura esterna è di $33^\circ C$ e che il processore dovrà gestire 10 processi, calcolare la stima puntuale della temperatura del processore, utilizzando il modello scelto.*
 La previsione puntuale per la temperatura del processore è

$$\hat{Y}_0 = \hat{Y}_{|T=33} = 5.91778 + 1.75866 \times 33 = 63.95356^\circ C.$$

2 **3.4** *Calcolare un intervallo di previsione al 95% per la temperatura del processore alle condizioni di **3.3**.*
 L'intervallo di previsione cercato è:

$$\hat{Y}_0 \pm t_{n-2}(0.975)\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(T_0 - \bar{t})^2}{S_{tt}}},$$

con $S_{tt} = s_t^2(n - 1)$, da cui l'intervallo cercato: [60.78094; 67.12618].