Politecnico di Milano - Scuola di Ingegneria Industriale e dell'Informazione

IV appello di Statistica per Ingegneria Fisica
23 Gennaio 2019

Cognome, Nome e Numero di matricola:

Problema 1. In condizioni d'utilizzo standard, l'autonomia di uno smartphone di marca ACME (cioè il tempo che esso impiega a esaurire una ricarica al 100% della batteria) è una variabile aleatoria X con media 7.50 ore e varianza 4.00 ore^2, entrambe note e dichiarate dal produttore. Molti acquirenti si sono però lamentati perché secondo loro 7.50 ore è un'autonomia media troppo breve. Di conseguenza, sperando di risolvere il problema, la ACME ha sviluppato e rilasciato un aggiornamento software che – a suo dire – aumenta sensibilmente il valor medio di X, mantenendone allo stesso tempo inalterata la varianza.

Una rivista di elettronica ha messo alla prova l'efficacia dell'aggiornamento, cronometrando le ore di autonomia di 40 smartphone ACME su cui era stato installato il nuovo software. I risultati $X_1, \ldots, X_{40}$ sono stati raccolti nel vettore `data` e poi elaborati con R, ottenendo l'output seguente:

```
> data
 [1]  6.85  6.54  8.11 10.03 10.98  8.69  6.62  8.27  6.60  8.37
[11]  9.82 10.34  8.52  7.12  6.82  6.76  7.51  7.28  6.40  6.69
[21] 12.99  6.78  6.91 11.02  7.78  8.21  8.53  7.31  7.79  7.04
[31]  8.94  8.09  6.79  9.17  5.91  7.02  9.40 13.59  7.34  8.90

> mean(data)
[1] 8.19575

> shapiro.test(data)

        Shapiro-Wilk normality test

data:  data
W = 0.87109, p-value = 0.0003054

> qqnorm(data); qqline(data)
```

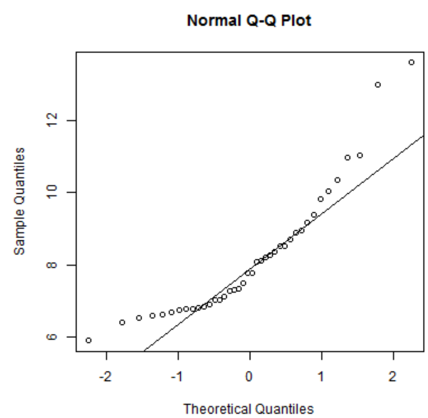


Figura 1: Console di R coi comandi utilizzati e il relativo output

Nel seguito, si assuma che X abbia media μ incognita e varianza $\sigma^2 = 4.00$ ore^2 nota. Si assuma inoltre che $X_1, \ldots, X_{40}$ sia un campione aleatorio estratto dalla stessa densità di X.

(a) È ragionevole supporre che l'autonomia di uno smartphone ACME abbia densità gaussiana? Perché?

(b) Impostate un test opportuno per decidere se i dati dimostrino un aumento evidente dell'autonomia media negli smartphone col software aggiornato. Cosa concludete al livello di significatività del 5%?

(c) Calcolate il p-value del test precedente.

Si supponga ora che $Y_1, \ldots, Y_{40}$ sia un altro campione aleatorio, indipendente dal precedente, ma *estratto di nuovo dalla stessa densità di* X. Si indichi con $\bar{Y}$ la media campionaria di questo nuovo campione.

(d) A vostra scelta, calcolate *solo una* delle seguenti probabilità (eventualmente in modo approssimato):

$$\mathbb{P}\left(|Y_{34} - \mu| > 3.00 \text{ ore}\right) \qquad \text{oppure} \qquad \mathbb{P}\left(\left|\bar{Y} - \mu\right| > 0.30 \text{ ore}\right)$$

Spiegate anche brevemente per quale motivo non avete preferito calcolare l'altra probabilità.

(e) *Facoltativo:* La probabilità che *non* avete calcolato al punto precedente, potrebbe essere maggiore del 60%? Giustificate adeguatamente la risposta.

Risultati.

(a) No. Infatti, il normal Q-Q plot dei dati non risulta abbastanza allineato lungo la retta della qqline, e questo si riflette nel p-value molto basso del test di Shapiro-Wilks (p-value $= 0.0003054 = 0.3054\%$).

(b) Vogliamo stabilire se dai dati c'è evidenza che $\mu > 7.50$ ore $\Rightarrow$ mettiamo questa affermazione nell'ipotesi nulla. Bisogna dunque fare un test per le ipotesi

$$H_0 : \mu = 7.50\,\text{ore} =: \mu_0 \quad (\text{oppure } H_0 : \mu \le \mu_0) \qquad \text{contro} \qquad H_1 : \mu > \mu_0\,.$$

Trattandosi di un campione numeroso ($n > 30$), possiamo applicare il TLC e concludere che la media campionaria $\bar{X}$ ha approssimativamente densità gaussiana. Di conseguenza, possiamo fare uno Z-test per un campione non gaussiano ma numeroso. La regola al livello di significatività α è

$$\text{“ rifiuto } H_0 \text{ se trovo } Z_0 := \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} > z_{1-\alpha}\text{”}\,.$$

La realizzazione della media campionaria coi dati cronometrati dalla rivista è $\bar{x} = 8.19575$, come si può leggere dall'output di R (vedi comando `mean`). La corrispondente realizzazione della statistica test è

$$z_0 := \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{8.19575 - 7.50}{\sqrt{4.00}}\sqrt{40} = 2.200\,.$$

Confrontando questo valore col quantile al livello $\alpha = 5\%$

$$z_{1-\alpha} = z_{1-0.05} = z_{0.95} = 1.645\,,$$

troviamo che $z_0 = 2.200 > 1.645 = z_{1-\alpha}$, e dunque dobbiamo rifiutare H_0. Ne concludiamo che al 5% di significatività c'è evidenza che l'aggiornamento software effettivamente funzioni.

(c) Ricordando l'espressione della regione di rifiuto trovata al punto precedente, il p-value si ricava imponendo l'uguaglianza

$$\begin{aligned} z_0 \equiv z_{1-p\text{-value}} \quad &\Leftrightarrow \quad \Phi(z_0) \equiv \Phi(z_{1-p\text{-value}}) = 1 - p\text{-value} \\ &\Leftrightarrow \quad p\text{-value} = 1 - \Phi(z_0) = 1 - \Phi(2.200) = 1 - 0.98610 = 1.390\%\,. \end{aligned}$$

Tale valore è basso, e dimostra che c'è forte evidenza a favore di H_1.

(d) Non conosciamo la densità delle Y_i, e non possiamo assumere che essa sia gaussiana per quanto visto al punto (a). Dunque non possiamo calcolare $\mathbb{P}(|Y_{34} - \mu| > 3.00)$. Invece, per il TLC abbiamo

$$\bar{Y} \approx N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(\mu, \frac{4.00}{40}\right) = N(\mu, 0.10)\,,$$

e dunque possiamo (approssimativamente) calcolare

$$\begin{aligned} \mathbb{P}\left(\left|\bar{Y} - \mu\right| > 0.30\right) &= 1 - \mathbb{P}\left(\left|\bar{Y} - \mathbb{E}\left[\bar{Y}\right]\right| \le 0.30\right) = 1 - \mathbb{P}\left(\left|\frac{\bar{Y} - \mathbb{E}\left[\bar{Y}\right]}{\sqrt{\mathrm{Var}(\bar{Y})}}\right| \le \frac{0.30}{\sqrt{0.10}}\right) \\ &= 1 - \mathbb{P}\left(-0.9487 \le \frac{\bar{Y} - \mathbb{E}\left[\bar{Y}\right]}{\sqrt{\mathrm{Var}(\bar{Y})}} \le 0.9487\right) \simeq 1 - [\Phi(0.9487) - \Phi(-0.9487)] \\ &= 1 - [2\,\Phi(0.9487) - 1] = 2\,(1 - 0.82894) = 34.212\%\,. \end{aligned}$$

(e) Possiamo trovare un limite superiore per la probabilità $\mathbb{P}(|Y_{34} - \mu| > 3.00)$ utilizzando la disuguaglianza di Chebyshev:

$$\mathbb{P}(|Y_{34} - \mathbb{E}[Y_{34}]| > \varepsilon) \le \frac{\mathrm{Var}(Y_{34})}{\varepsilon^2} \quad \Rightarrow \quad \mathbb{P}(|Y_{34} - \mu| > 3.00) \le \frac{\sigma^2}{3.00^2} = \frac{4.00}{3.00^2} = 0.4444 = 44.44\%\,.$$

Dunque, la probabilità $\mathbb{P}(|Y_{34} - \mu| > 3.00)$ *non può* essere maggiore del 60%.

Problema 2. Quasi tutte le ricerche sull'effetto serra e sulle sue conseguenze per il riscaldamento globale indicano che negli ultimi 300 anni la concentrazione media di CO_2 nell'atmosfera terrestre è aumentata di più di 100 parti per milione (ppm). Alcuni studiosi negazionisti mettono in dubbio però i dati ufficiali, e sostengono che l'aumento di CO_2 sia stato sensibilmente minore di quanto comunemente accettato.

Per dimostrare la loro tesi, i negazionisti hanno recentemente finanziato una spedizione al Polo Sud con lo scopo di effettuare alcuni carotaggi del ghiaccio antartico. Confrontando la concentrazione di CO_2 nelle bollicine d'aria intrappolate vicino alla superficie con quella nelle bollicine situate più in profondità, essi sperano infatti di trovare prove evidenti a loro favore.

Sono stati analizzati 10 diversi campioni d'aria ottenuti in questo modo: i primi 6 sono recenti, e gli altri 4 risalgono invece a circa 300 anni fa. Si può supporre che i risultati delle corrispondenti misure costituiscano due campioni aleatori *gaussiani* e *indipendenti*, ciascuno con densità centrata sulla concentrazione media di CO_2 della rispettiva epoca. Ecco i valori trovati (in ppm):

concentrazioni attuali	342	355	378	398	384	343
concentrazioni di 300 anni fa	333	298	251	264		

(a) Con un opportuno test al livello di significatività del 5%, verificate se le varianze dei due campioni si possono considerare uguali.

(b) Quali sono le condizioni di applicabilità del test del punto (a)? Esse sono soddisfatte oppure no?

(c) Impostate un opportuno test volto a dimostrare che *vi è evidenza a favore della tesi dei negazionisti*, e cioè che l'aumento medio della concentrazione di CO_2 negli ultimi 300 anni sia stato *sensibilmente inferiore ai* 100 ppm sostenuti dal resto della comunità scientifica. Per farlo, scrivete correttamente le ipotesi nulla e alternativa, la statistica test e la regola di rifiuto.

(d) Quali sono le condizioni di applicabilità del test del punto (c)? Esse sono soddisfatte oppure no?

(e) Al livello di significatività del 5%, il test del punto (c) dà ragione alla tesi dei negazionisti? La sua conclusione è debole o forte?

(f) Ricavate un intervallo in cui cade il p-value del test del punto (c).

Risultati.

(a) Dobbiamo fare un F-test per il rapporto delle varianze. Esso è applicabile se le due serie di misure $X_1, \ldots, X_6$ e $Y_1, \ldots, Y_4$ costituiscono due campioni aleatori gaussiani e indipendenti, condizione che è soddisfatta per quanto detto nel testo dell'esercizio. Le ipotesi statistiche che vogliamo mettere a confronto sono

$$H_0 : \sigma_X^2 = \sigma_Y^2 \qquad \text{contro} \qquad H_1 : \sigma_X^2 \neq \sigma_Y^2 \,,$$

dove $\sigma_X^2 = \mathrm{Var}(X_i)$ e $\sigma_Y^2 = \mathrm{Var}(Y_j)$ sono le varianze incognite dei due campioni. La regola del test è

$$\text{“ rifiuto } H_0 \text{ se trovo } F_0 > f_{1-\frac{\alpha}{2}}(n_X - 1, n_Y - 1) \text{ oppure } F_0 < f_{\frac{\alpha}{2}}(n_X - 1, n_Y - 1)\text{ ”},$$

dove F_0 è la statistica test

$$F_0 := \frac{S_X^2}{S_Y^2} \,,$$

n_X, n_Y sono le numerosità dei due campioni, e S_X^2, S_Y^2 sono le rispettive varianze campionarie. Dai dati ricaviamo le realizzazioni delle medie e varianze campionarie

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{342 + \ldots + 343}{6} = 366.667 \\
s_X^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right) = \frac{1}{6-1}\left(342^2 + \ldots + 343^2 - 6 \cdot 366.667^2\right) \\
&= 543.067 \\
\bar{y} &= \frac{333 + \ldots + 264}{4} = 286.5 \\
s_X^2 &= \frac{1}{4-1}\left(333^2 + \ldots + 264^2 - 4 \cdot 286.5^2\right) = 1353.667\,.
\end{aligned}$$

Di conseguenza,

$$f_0 = \frac{543.067}{1353.667} = 0.4012\,.$$

Al livello di significatività $\alpha = 5\%$, questo valore va confrontato coi quantili

$$f_{1-\frac{\alpha}{2}}(n_X - 1, n_Y - 1) = f_{1-\frac{0.05}{2}}(6-1, 4-1) = f_{0.975}(5,3) = 14.885$$
$$f_{\frac{\alpha}{2}}(n_X - 1, n_Y - 1) = f_{\frac{0.05}{2}}(6-1, 4-1) = f_{0.025}(5,3) = \frac{1}{f_{0.975}(3,5)} = \frac{1}{7.764} = 0.1288\,.$$

Dal momento che $0.1288 < f_0 = 0.4012 < 14.885$, *non* possiamo rifiutare H_0. Otteniamo quindi la conclusione *debole* che le varianze dei due campioni si possono considerare uguali.

(b) Abbiamo già anticipato nel punto precedente che il test è applicabile se le due serie di misure costituiscono due campioni aleatori gaussiani e indipendenti, e che questa condizione è soddisfatta in virtù di quanto affermato nel testo del problema.

(c) La tesi dei negazionisti è che $\mu_X - \mu_Y < 100$ ppm. Se vogliamo dimostrarla, cioè se cerchiamo evidenza forte in suo favore, dobbiamo metterla nell'ipotesi alternativa del test. Pertanto, le ipotesi statistiche corrette sono

$$H_0 : \mu_X - \mu_Y \geq 100 \text{ ppm} =: \delta_0 \quad (\text{oppure } H_0 : \mu_X - \mu_Y = \delta_0) \qquad \text{contro} \qquad H_1 : \mu_X - \mu_Y < \delta_0\,.$$

Poiché i due campioni sono entrambi gaussiani e indipendenti, e abbiamo verificato al punto (a) che le loro varianze possono considerarsi uguali, valgono le condizioni per fare un T-test. La regola è

$$\text{“ rifiuto } H_0 \text{ se trovo } T_0 := \frac{\bar{X} - \bar{Y} - \delta_0}{S_p\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} < -t_{1-\alpha}(n_X + n_Y - 2)\text{”}\,,$$

dove abbiamo usato la varianza pooled

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}\,.$$

(d) Per quanto già detto nel punto precedente, le condizioni di applicabilità del test sono le stesse dell'F-test del punto (a), con in più la richiesta che le varianze incognite dei due campioni siano uguali. In virtù del test del punto (a), anche quest'ultima condizione è soddisfatta.

(e) Coi dati trovati,

$$s_p^2 = \frac{(6-1)\cdot 543.067 + (4-1)\cdot 1353.667}{6+4-2} = 847.042 \quad \Rightarrow \quad t_0 = \frac{366.667 - 286.5 - 100}{\sqrt{847.042}\cdot\sqrt{\frac{1}{6}+\frac{1}{4}}} = -1.0557\,.$$

Al livello $\alpha = 5\%$, dobbiamo confrontare t_0 con

$$-t_{1-\alpha}(n_X + n_Y - 2) = -t_{1-0.05}(6+4-2) = -t_{0.95}(8) = -1.8595\,.$$

Dal momento che $t_0 = -1.0557 \not< -t_{1-\alpha}(n_X + n_Y - 2) = -1.8595$, *non* possiamo rifiutare H_0. Perciò, otteniamo la conclusione *debole* che non c'è nessuna evidenza a favore della tesi dei negazionisti.

(f) Il p-value dei dati è il valore di α per cui nella regola di rifiuto vale l'uguaglianza

$$t_0 \equiv -t_{1-\alpha}(n_X + n_Y - 2) = -t_{1-\alpha}(8) \quad \Leftrightarrow \quad t_{1-\alpha}(8) = -t_0 = -(-1.0557) = 1.0557\,.$$

Dalle tavole della t di Student ricaviamo che

$$t_{0.80}(8) = 0.8889 < 1.0557 < t_{0.85}(8) = 1.1081 \quad \Rightarrow \quad 0.80 < 1-\alpha < 0.85 \quad \Leftrightarrow \quad 0.15 < \alpha < 0.20\,.$$

Perciò, $0.15 < p\text{-value} < 0.20$.

Problema 3. La ACME è interessata a studiare quanto si usura la batteria dei suoi cellulari dopo un anno di utilizzo. Per questo motivo, vengono esaminati 42 dispositivi, tutti muniti della stessa batteria. Vengono misurate la percentuale di usura della batteria (Y), il tempo di utilizzo medio giornaliero (TU) e la dimensione in pollici dello schermo (DS). Viene impostato un modello empirico lineare gaussiano con responso la percentuale di usura Y e predittori TU e DS senza interazione tra i regressori. Vengono impostati successivamente anche due modelli lineari semplici, uno utilizzando come unico predittore TU, e l'altro utilizzando come unico predittore DS.

(a) Scrivere la relazione tra le variabili ipotizzata dai tre modelli.

(b) L'ipotesi di omoschedasticità è soddisfatta per i tre modelli?

(c) I tre modelli sono globalmente significativi?

(d) I singoli regressori nei tre modelli sono tutti significativi?

Sulla base delle risposte precedenti, scegliere il modello migliore fra i tre proposti.

(e) Stimare puntualmente l'aumento di usura medio della batteria, quando il tempo di utilizzo aumenta di un'ora.

(f) Con un opportuno test, verificare se c'è evidenza che, quando il tempo di utilizzo aumenta di un'ora, l'aumento di usura medio della batteria è maggiore di 1.

Risultati.

(a)
- modello completo: $Y = \beta_0 + \beta_1\, TU + \beta_2\, DS + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2),$
- modello ridotto 1: $Y = \beta_0 + \beta_1\, TU + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2),$
- modello ridotto 2: $Y = \beta_0 + \beta_1\, DS + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2).$

(b) L'ipotesi di omoschedasticità è soddisfatta per tutti i modelli, come si può vedere dagli scatterplot dei residui che non presentano particolari pattern.

(c) Per rispondere alla domanda occorre fare un F-test sui coefficienti dei regressori. Per quanto riguarda il modello completo, il test ha un p-value minore di $2.2 \cdot 10^{-16}$. Di conseguenza, tale modello è significativo. Per quanto riguarda i modelli ridotti, il test equivale al test sul singolo regressore di ciascun modello. In particolare, per il modello con solo il regressore 'tempo di utilizzo', si ha un p-value minore di $2.2 \cdot 10^{-16}$, mentre per quello con solo la 'dimensione dello schermo' si ha 0.1873. Il modello con solo la dimensione dello schermo risulta pertanto non significativo.

(d) Per quanto riguarda i modelli di regressione semplice, abbiamo già visto nel punto precedente che il p-value dell'F-test e del T-test sull'unico regressore coincidono. Quindi, nel modello ridotto 1 il regressore è significativo, mentre nel modello ridotto 2 il regressore non è significativo. Per verificare la significatività dei singoli regressori nel modello completo bisogna guardare i p-value dei due T-test. Dal summary, vediamo che il p-value del T-test su β_1 è minore di $2.2 \cdot 10^{-16}$ e di conseguenza il tempo di utilizzo è un regressore significativo, mentre quello su β_2 è 0.388, e quindi la dimensione dello schermo non è un regressore significativo.

(e) Dato che nel modello completo la dimensione dello schermo non è significativa, è opportuno ridurre il modello togliendo tale regressore. In questo modo otteniamo il modello ridotto 1, in cui l'unico regressore è significativo. Quindi il modello migliore fra i tre considerati è il modello ridotto 1. La stima per l'aumento medio di usura quando il tempo ti utilizzo aumenta di un'ora è data da $\hat{\beta}_1 = 3.1093$.

(f) Dobbiamo fare un test su β_1 per il modello ridotto 1:

$$H_0 : \beta_1 \leq 1 =: \beta_{1,0} \qquad \text{contro} \qquad H_1 : \beta_1 > \beta_{1,0}\,.$$

La regola di rifiuto al livello α è

$$\text{“ rifiuto } H_0 \text{ se trovo } T_0 := \frac{\hat{\beta}_1 - \beta_{1,0}}{\mathrm{se}(\hat{\beta}_1)} > t_{1-\alpha}(n-k-1)\text{ ”}\,,$$

dove $n = 42$ è il numero di dati e $k = 1$ quello dei predittori. Calcolo il p-value:

$$t_0 = \frac{3.1093 - 1}{0.1819} = 11.59593 \equiv t_{1-p\text{-value}}(40)\,.$$

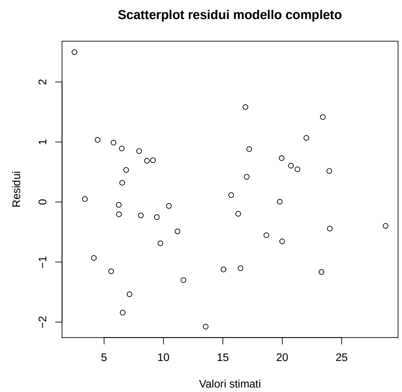
Sulle tavole vediamo che $11.59593 \gg 3.5510 = t_{0.9995}(40)$. Di conseguenza, $1 - p$-value $\gg 0.9995$, da cui p-value $\ll 0.0005$. Siamo pertanto indotti a rifiutare H_0 a tutti i livelli di significatività sensati. In altre parole, c'è una grandissima evidenza che β_1 sia maggiore di 1 (conclusione forte).

```
Call:
lm(formula = Usura ~ TempoUtilizzo + DimensioneSchermo)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4898 -1.6657 -0.0548  1.8365  6.6026

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.0760     5.1308   0.405    0.688
TempoUtilizzo       3.0817     0.1852  16.643   <2e-16 ***
DimensioneSchermo  -0.8145     0.9324  -0.874    0.388
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.646 on 39 degrees of freedom
Multiple R-squared:  0.8819,    Adjusted R-squared:  0.8758
F-statistic: 145.6 on 2 and 39 DF,  p-value: < 2.2e-16
```

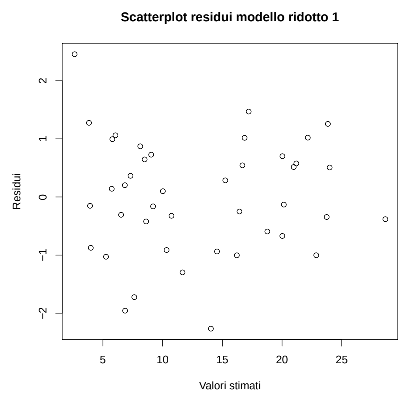



```
Call:
lm(formula = Usura ~ TempoUtilizzo)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9791 -1.7186 -0.0419  1.8152  6.4819

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.3212     0.9923  -2.339   0.0244 *
TempoUtilizzo   3.1093     0.1819  17.093   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.638 on 40 degrees of freedom
Multiple R-squared:  0.8796,    Adjusted R-squared:  0.8766
F-statistic: 292.2 on 1 and 40 DF,  p-value: < 2.2e-16
```




```
Call:
lm(formula = Usura ~ DimensioneSchermo)

Residuals:
    Min      1Q  Median      3Q     Max
-11.796  -5.007  -2.670   4.957  15.414

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         31.262     13.552   2.307   0.0263 *
DimensioneSchermo   -3.464      2.582  -1.341   0.1873
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.436 on 40 degrees of freedom
Multiple R-squared:  0.04305,   Adjusted R-squared:  0.01913
F-statistic:   1.8 on 1 and 40 DF,  p-value: 0.1873
```

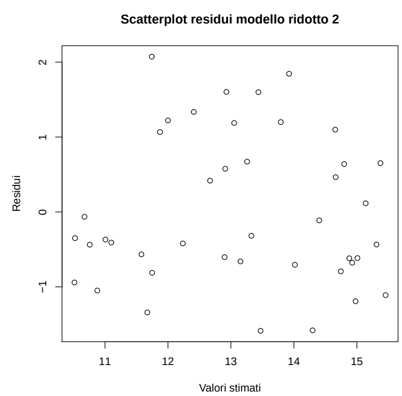


Figura 2: Summary e scatterplot dei residui per i tre modelli