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

II appello di Statistica per Ingegneria Fisica
20 Luglio 2018

Cognome, Nome e Numero di matricola:

Problema 1. Da quando l'inchiesta-scoop del giornalista televisivo Kent Brockman ha svelato che 1 hg di Krusty Burger (l'hamburger più venduto in tutta Springfield) contiene mediamente ben 300 mg di colesterolo, la catena di fast-food di Krusty il Clown ha subito un vertiginoso crollo delle vendite. Nel tentativo di arginare la fuga dei clienti, Krusty ha ideato il nuovo 'Mother Nature Burger', un hamburger completamente vegetariano che, secondo la campagna pubblicitaria di lancio, dovrebbe contenere una quantità media di colesterolo nettamente inferiore al Krusty Burger.
Tuttavia, conoscendo la non proprio brillante reputazione di Krusty, prima di credere che il nuovo hamburger sia effettivamente migliore di quello vecchio, la piccola Lisa Simpson fa analizzare dal Prof. Frink un campione di 7 Mother Nature Burger diversi, trovando queste concentrazioni di colesterolo (in mg/hg):

$$241.8 \qquad 337.5 \qquad 258.2 \qquad 279.3 \qquad 290.9 \qquad 321.6 \qquad 297.8\,.$$

Si può assumere che le 7 misure precedenti siano le realizzazioni di un campione aleatorio gaussiano.
Inoltre, Lisa studia approfonditamente l'inchiesta di Brockman, scoprendovi che la concentrazione di colesterolo nel vecchio Krusty Burger, oltre ad avere una media di 300 mg/hg, aveva anche una varianza pari a $1600\,(\text{mg/hg})^2$.

(a) Stabilite con un opportuno test di significatività del 5% se la varianza della concentrazione di colesterolo nel Mother Nature Burger si può assumere uguale a quella nota nel vecchio Krusty Burger.

(b) Scrivete le ipotesi statistiche H_0 e H_1 di un test volto a dimostrare che la concentrazione media di colesterolo nel Mother Nature Burger è inferiore a quella nota nel vecchio Krusty Burger. In tale test, l'errore più grave deve essere considerare migliore il nuovo hamburger quando in realtà non lo è.

Per fare un test per le ipotesi H_0 e H_1 del punto (b), Lisa decide di utilizzare la regola:

$$\text{considero il nuovo hamburger migliore di quello vecchio se } \bar{X} < 295\,\text{mg/hg}\,,$$

dove $\bar{X}$ è la media campionaria delle 7 misure di Lisa.

(c) Qual è il livello di significatività del test così impostato da Lisa?

(d) Cosa conclude Lisa in base al suo test e ai suoi dati? Voi condividete la conclusione di Lisa? Perché?

(e) Scrivete voi la regola di un test di livello α arbitrario per decidere tra le ipotesi H_0 e H_1 del punto (b), e calcolatene il p-value coi dati di Lisa. Secondo voi, c'è evidenza che il nuovo hamburger sia effettivamente migliore di quello vecchio?

Risultati.
(a) Bisogna fare un test per le ipotesi

$$H_0 : \sigma^2 = 1600 =: \sigma_0^2 \qquad \text{vs.} \qquad H_1 : \sigma^2 \neq \sigma_0^2\,,$$

dove $1600\,(\text{mg/hg})^2$ è la varianza nota della concentrazione di colesterolo nel vecchio Krusty Burger. Poiché le 7 misure $X_1, \dots, X_7$ costituiscono un campione gaussiano, come regola per un test di livello α si può usare

$$\text{“rifiuto } H_0 \text{ se} \quad \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{\frac{\alpha}{2}}(n-1) \quad \text{oppure} \quad \frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{1-\frac{\alpha}{2}}(n-1)\,\text{”}\,,$$

dove S^2 è la varianza campionaria delle 7 misure. Coi dati assegnati, otteniamo le realizzazioni

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{241.8 + 337.5 + \ldots + 297.8}{7} = 289.5857$$
$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}\right) = \frac{1}{7-1}\left(241.8^2 + 337.5^2 + \ldots + 297.8^2 - 7 \cdot 289.5857\right) = 1\,127.371\,.$$

Di conseguenza, la realizzazione della statistica test è

$$x_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(7-1)\cdot 1\,127.371}{40^2} = 4.2276\,,$$

che al livello $\alpha = 5\%$ va confrontata coi quantili

$$\chi^2_{\frac{\alpha}{2}}(n-1) = \chi^2_{0.025}(6) = 1.2373 \qquad \chi^2_{1-\frac{\alpha}{2}}(n-1) = \chi^2_{0.975}(6) = 14.4494\,.$$

Poiché $1.2373 < 4.2276 < 14.4494$, dobbiamo accettare H_0 e concludere che al 5% di significatività non c'è evidenza che $\sigma^2 \neq 1600\,(\text{mg/hg})^2$.

(b) Indichiamo con μ la concentrazione media di colesterolo dei Mother Nature Burger. Vogliamo dimostrare che $\mu < 300\,\text{mg/hg} \Rightarrow$ mettiamo questa affermazione come ipotesi alternativa:

$$H_0 : \mu = 300 =: \mu_0 \qquad \text{vs.} \qquad H_1 : \mu < \mu_0\,. \tag{$*$}$$

L'errore più grave in un test è sempre accettare H_1 quando in realtà è vera H_0 (= errore di prima specie). Con le ipotesi precedenti, tale errore è equivalente a considerare migliori i nuovi hamburger (= accettare $H_1 : \mu < \mu_0$) quando in realtà non lo sono (= è vera $H_0 : \mu = \mu_0$). Ciò è proprio quanto richiesto.

(c) Con le ipotesi statistiche $(*)$, abbiamo l'uguaglianza di eventi

"considero il nuovo hamburger migliore di quello vecchio" = "accetto H_1" = "rifiuto H_0"

Di conseguenza, la regola del test di Lisa è

"rifiuto H_0 se $\bar{X} < 295$"\,.

Con tale regola, il livello di significatività è la probabilità di errore di prima specie

$$\begin{aligned}
\text{significatività} &= \mathbb{P}_{H_0 \text{ vera}}(\text{"rifiuto } H_0\text{"}) = \mathbb{P}_{\mu=300}\left(\bar{X} < 295\right) \\
&= \mathbb{P}_{\mu=300}\left(\underbrace{\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}}_{\sim N(0,1)} < \frac{295-300}{\sqrt{1600/7}}\right) \qquad \text{poiché, per un campione normale, } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \\
&= \Phi\left(\frac{295-300}{\sqrt{1600/7}}\right) = \Phi(-0.3307) = 1 - \Phi(0.3307) = 1 - 0.62930 \\
&= 37.070\%\,,
\end{aligned}$$

dove abbiamo usato il fatto che, per quanto visto al punto (a), possiamo assumere $\sigma^2 = 1600$.

(d) Abbiamo già calcolato $\bar{x} = 289.5857$. Dal momento che $289.5857 < 295$, Lisa conclude il suo test rifiutando H_0, cioè accettando l'ipotesi alternativa che i nuovi hamburger siano migliori di quelli vecchi. Tuttavia, abbiamo visto al punto precedente che la probabilità di errore di prima specie (= il livello di significatività) del test di Lisa è insolitamente alta (37.070% è molto maggiore dell'usuale 5%). La conclusione di Lisa pertanto non è condivisibile.
Una giustificazione alternativa è che col suo test Lisa ha trovato che il p-value dei dati è minore del 37.070%. La disuguaglianza p-value $< 37.070\%$ non garantisce tuttavia che il p-value sia abbastanza piccolo da poter rifiutare H_0 agli usuali livelli di significatività.

(e) La regola di un test con livello di significatività α per le ipotesi $(*)$ è

$$\text{"rifiuto } H_0 \text{ se } Z_0 := \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n} < -z_{1-\alpha}\text{"} \tag{$**$}$$

Con $\mu_0 = 300$, $\sigma = \sqrt{1600}$, $n = 7$ e $\bar{x} = 289.5857$, abbiamo

$$z_0 = \frac{289.5857 - 300}{\sqrt{1600}}\sqrt{7} = -0.689\,.$$

Il p-value è il valore di α che fa ottenere l'uguaglianza nella regola (**), cioè

$$z_0 \equiv -z_{1-\alpha} \quad \Leftrightarrow \quad \Phi(z_0) \equiv \Phi(-z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha}) = 1 - (1-\alpha) = \alpha\,.$$

Pertanto,

$$p\text{-value} = \Phi(z_0) = \Phi(-0.689) = 1 - \Phi(0.689) = 1 - 0.75490 = 0.24510 = 24.510\%\,.$$

Con un p-value così alto (ben maggiore dell'usale livello di significatività del 5%) non c'è nessuna evidenza a favore di H_1, e dunque non possiamo rigettare l'ipotesi nulla che il nuovo hamburger abbia una concentrazione media di colesterolo uguale a quello vecchio.

Problema 2. Il ritardo (in minuti) dell'autobus che prendo ogni mattina per raggiungere il Politecnico può essere modellizzato con una variabile aleatoria assolutamente continua avente densità uniforme sull'intervallo $[0, \theta]$. Non conosco però quale sia il massimo ritardo possibile θ. Per darne una stima, ho dunque deciso di registrare i ritardi X_1, X_2, X_3, X_4, X_5 che si verificano in 5 mattine consecutive. Naturalmente, le variabili aleatorie $X_1, \dots, X_5$ si possono considerare tutte indipendenti tra loro.
Ora però sono indeciso su quale scegliere fra i tre seguenti *stimatori del parametro* θ:

$$Z = \frac{1}{5}\sum_{i=1}^{5} X_i \qquad T = \frac{2}{5}\sum_{i=1}^{5} X_i \qquad U = 2X_1 + X_2 + X_3 - (X_4 + X_5)\,.$$

(a) Determinate le distorsioni di Z, di T e di U come stimatori di θ.

(b) Determinate gli errori quadratici medi di Z, di T e di U come stimatori di θ.

(c) Quale dei tre stimatori è preferibile per stimare θ? Perché?

(d) Chiamiamo $\hat{\Theta}$ lo stimatore scelto nel punto precedente. Ricordandovi l'espressione dell'errore quadratico medio $\mathrm{mse}(\hat{\Theta}; \theta)$ trovata al punto (b), proponete uno stimatore *esattamente* non distorto per il parametro $\mathrm{mse}(\hat{\Theta}; \theta)$. Se non ci riuscite, proponetene uno che sia almeno *approssimativamente* non distorto. In ogni caso, giustificate opportunamente la vostra risposta.

(e) Ho registrato i ritardi nelle 5 mattine da lunedì a venerdì scorsi, ottenendo questi valori (in minuti):

17.5 30.1 9.9 43.2 21.8

Coi dati precedenti, fornite una stima di θ e dell'errore quadratico medio dello stimatore utilizzato.

Risultati.

(a) Osserviamo che

$$Z = \bar{X} \qquad T = 2Z\,.$$

Usando il formulario, abbiamo

$$\mathbb{E}[X_i] = \frac{0+\theta}{2} = \frac{\theta}{2} \qquad \forall i\,,$$

e quindi

$$\begin{aligned}
&\mathbb{E}[Z] = \mathbb{E}\left[\bar{X}\right] = \mathbb{E}[X_i] = \frac{\theta}{2}\\
&\mathbb{E}[T] = \mathbb{E}[2Z] = 2\,\mathbb{E}[Z] = 2\,\frac{\theta}{2} = \theta\\
&\mathbb{E}[U] = \mathbb{E}[2X_1 + X_2 + X_3 - (X_4 + X_5)] = 2\,\mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] - (\mathbb{E}[X_4] + \mathbb{E}[X_5])\\
&\qquad = 2\,\frac{\theta}{2} + \frac{\theta}{2} + \frac{\theta}{2} - \left(\frac{\theta}{2} + \frac{\theta}{2}\right) = \theta\,.
\end{aligned}$$

Le distorsioni dunque sono

$$b(Z;\theta) = \mathbb{E}[Z] - \theta = -\frac{\theta}{2} \qquad b(T;\theta) = \mathbb{E}[T] - \theta = 0 = b(U;\theta)$$

(b) Usiamo la formula

$$\mathrm{mse}(\hat{\Theta};\theta) = \mathrm{var}(\hat{\Theta}) + b(\hat{\Theta};\theta)^2$$

Quindi, dobbiamo prima calcolare le varianze degli stimatori:

$$\begin{aligned}
&\mathrm{var}(Z) = \mathrm{var}(\bar{X}) = \frac{\mathrm{var}(X_i)}{5} = \frac{\theta^2/12}{5} = \frac{\theta^2}{60}\\
&\mathrm{var}(T) = \mathrm{var}(2Z) = 2^2\mathrm{var}(Z) = 4\,\frac{\theta^2}{60} = \frac{\theta^2}{15}\\
&\mathrm{var}(U) = \mathrm{var}(2X_1 + X_2 + X_3 - (X_4 + X_5))\\
&\qquad = 2^2\mathrm{var}(X_1) + \mathrm{var}(X_2) + \mathrm{var}(X_3) + (-1)^2\mathrm{var}(X_4) + (-1)^2\mathrm{var}(X_5) \qquad \text{indipendenza delle } X_i \text{ e quadraticità di var}\\
&\qquad = 8\,\mathrm{var}(X_i) = \frac{2\theta^2}{3}\,,
\end{aligned}$$

dove abbiamo ricavato dal formulario

$$\mathrm{var}(X_i) = \frac{(\theta - 0)^2}{12} = \frac{\theta^2}{12}\,.$$

Di conseguenza,

$$\begin{aligned}
\mathrm{mse}(Z;\theta) &= \frac{\theta^2}{60} + \left(-\frac{\theta}{2}\right)^2 = \frac{4}{15}\,\theta^2 \\
\mathrm{mse}(T;\theta) &= \frac{\theta^2}{15} + 0^2 = \frac{1}{15}\,\theta^2 \\
\mathrm{mse}(U;\theta) &= \frac{2\theta^2}{3} + 0^2 = \frac{2}{3}\,\theta^2\,.
\end{aligned}$$

(c) Per ogni θ, per quanto visto al punto (b) abbiamo

$$\mathrm{mse}(T;\theta) < \mathrm{mse}(Z;\theta) < \mathrm{mse}(U;\theta)\,,$$

e quindi T è lo stimatore col minor errore quadratico medio. Dal momento che T è anche non distorto, è senza dubbio lo stimatore migliore dei tre.

(d) Posto $\hat\Theta = T = 2\bar X$, abbiamo già visto che

$$\mathrm{mse}(\hat\Theta;\theta) = \frac{1}{15}\,\theta^2 = \frac{12}{15}\cdot\frac{\theta^2}{12} = \frac{12}{15}\mathrm{var}(X_i) = \frac{4}{5}\,\mathrm{var}(X_i)\,.$$

Ora, uno stimatore esattamente non distorto del parametro $\mathrm{var}(X_i)$ è la varianza campionaria

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar X)^2 = \frac{1}{4}\sum_{i=1}^{5}(X_i - \bar X)^2$$

Per la linearità della media, uno stimatore *esattamente* non distorto di $\mathrm{mse}(\hat\Theta;\theta) = \frac{4}{5}\,\mathrm{var}(X_i)$ è quindi

$$\widehat{\mathrm{MSE}} := \frac{4}{5}\,S^2 = \frac{1}{5}\sum_{i=1}^{5}(X_i - \bar X)^2\,.$$

Altrimenti, dal momento che $\hat\Theta$ è uno stimatore esattamente non distorto del parametro θ e $\mathrm{mse}(\hat\Theta;\theta) = \frac{1}{15}\,\theta^2$, dal metodo delta segue che la variabile aleatoria

$$\widehat{\mathrm{MSE}}' := \frac{1}{15}\,\hat\Theta^2 = \frac{4}{15}\,\bar X^2$$

è uno stimatore *approssimativamente* non distorto di $\mathrm{mse}(\hat\Theta;\theta)$.

(e) Coi dati assegnati,

$$\bar x = \frac{17.5 + \ldots + 21.8}{5} = 24.5 \qquad\qquad s^2 = \frac{1}{5-1}\left(17.5^2 + \ldots + 21.8^2 - 5\cdot 24.5\right) = 162.625\,.$$

Dunque, usando $\hat\Theta$ e $\widehat{\mathrm{MSE}}$ per stimare rispettivamente θ e $\mathrm{mse}(\hat\Theta;\theta)$, ricaviamo le stime

$$\hat\theta = 2\bar x = 2\cdot 24.5 = 49\,\mathrm{min} \qquad\qquad \widehat{\mathrm{mse}} = \frac{4}{5}\,s^2 = \frac{4}{5}\,162.625 = 130.1\,\mathrm{min}^2\,.$$

Se invece usiamo lo stimatore approssimativamente non distorto $\widehat{\mathrm{MSE}}'$, otteniamo la seguente stima di $\mathrm{mse}(\hat\Theta;\theta)$:

$$\widehat{\mathrm{mse}}' = \frac{4}{15}\cdot 24.5^2 = 160.067\,\mathrm{min}^2\,.$$

Problema 3. In uno studio clinico si vuole indagare quali fattori influenzino il diametro della carotide in una coorte di pazienti affetti da scompenso cardiaco. Vengono quindi raccolti i dati relativi a 156 persone e le variabili misurate sono: il `diametro` della carotide dell'individuo in mm (misurato sempre nella stessa posizione); la `pressione` media dell'individuo; il `peso` dell'individuo.
Si vuole studiare il diametro della carotide in relazione alle altre variabili.
Per tutti i modelli considerati si ritengano soddisfatte le ipotesi sui residui.

(a) Si scriva il modello di regressione lineare gaussiano utilizzato per descrivere il `diametro` tramite `pressione` e `peso`.

(b) Il modello è globalmente significativo? Perché?

(c) Sulla base dell'output in Figura 1, è consigliabile ridurre il modello? Argomentare la risposta impostando un opportuno test.

In Figura 2 sono riportati gli output di due possibili modelli ridotti, ottenuti a partire dal modello completo impostato al punto (a).

(d) Si scrivano le equazioni dei modelli ridotti, e si scelga, tra i tre modelli proposti quello che si ritiene il migliore, tenendo conto di quanto osservato al punto (c) e della variabilità spiegata.

(e) Si costruisca un intervallo di previsione al 95% per il diametro di un nuovo individuo che abbia `pressione` $= 110$ e `peso` $= 55$, sapendo che $\overline{\texttt{pressione}} = 96.24$ e $\overline{\texttt{peso}} = 78.51$. *(Suggerimento: se non si è in grado di ricavare s_{xx} si usi $s_{xx} = 100$)*

Risultati.

(a) Modello 1: $\texttt{diametro}_i = \beta_0 + \beta_1 \texttt{pressione}_i + \beta_2 \texttt{peso} + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$

(b) Il modello è globalmente significativo. Infatti è possibile impostare un F-test:

$$H_0 : \beta_1 = \beta_2 = 0 \qquad \text{vs.} \qquad H_1 : \exists i = 1, 2 \text{ t.c. } \beta_i \neq 0$$

e dall'output di R vediamo che il p-value del test è praticamente zero.

(c) Impostiamo un test su β_2: $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$. Dall'output abbiamo che il p-value del test è pari a 0.917, per cui non è possibile rifiutare l'ipotesi nulla. Quindi è consigliabile ridurre il modello togliendo il regressore peso.

(d)

$$\text{Modello 2: } \texttt{diametro}_i = \beta_0 + \beta_1 \texttt{pressione}_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$
$$\text{Modello 3: } \texttt{diametro}_i = \beta_0 + \beta_1 \texttt{peso} + \varepsilon_i$$

Per quanto riguarda la variabilità spiegata il modello 2 è sicuramente migliore del modello 3, infatti ha un R^2 di 0.9125 contro lo 0.1317 del modello 3. L'R^2 del modello completo e del modello 2 sono uguali, ed entrambi i modelli sono ottimi in termini di variabilità spiegata. Per quanto detto al punto (b) il modello migliore risulta quindi il modello 2, dato che il modello 1 ha un regressore non significativo.

(e) Abbiamo: $s_{xx} = \frac{\hat{\sigma}^2}{\text{se}(\beta_1)^2} = \frac{0.1056^2}{0.0007618^2} = 19\,215.24$. Usando la formula dell'intervallo di previsione bilatero per il modello 2 si ottiene

$$\begin{aligned} &IP_{Y(\texttt{pressione}=110)}(95\%) \\ &= \left(\hat{\beta}_0 + \hat{\beta}_1 \cdot \texttt{pressione}^* \pm t_{\frac{1+\gamma}{2}}(n-2)\, \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(\texttt{pressione}^* - \overline{\texttt{pressione}})^2}{s_{xx}}} \right) \\ &= \left(6.8264346 + 0.0305282 \cdot 110 \pm 1.96 \cdot 0.1056 \cdot \sqrt{1 + \frac{1}{156} + \frac{(110 - 96.24)^2}{19\,215.24}} \right) = (9.9759\,,\, 10.3932)\,. \end{aligned}$$

dove abbiamo usato

$$t_{\frac{1+\gamma}{2}}(n-2) = t_{\frac{1+0.95}{2}}(156-2) \simeq z_{\frac{1+0.95}{2}} = 1.96\,.$$

```
Call:
lm(formula = diametro ~ pressione + peso)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36602 -0.06359  0.00047  0.07169  0.25521

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.827e+00  7.414e-02  92.077   <2e-16 ***
pressione    3.056e-02  8.271e-04  36.948   <2e-16 ***
peso        -4.526e-05  4.359e-04  -0.104    0.917
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.106 on 153 degrees of freedom
Multiple R-squared:  0.9125,    Adjusted R-squared:  0.9113
F-statistic: 797.7 on 2 and 153 DF,  p-value: < 2.2e-16
```

Figura 1: Summary modello completo

```
Call:
lm(formula = diametro ~ pressione)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36689 -0.06421  0.00016  0.07227  0.25624

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.8264346  0.0738078   92.49   <2e-16 ***
pressione   0.0305282  0.0007618   40.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1056 on 154 degrees of freedom
Multiple R-squared:  0.9125,    Adjusted R-squared:  0.9119
F-statistic:  1606 on 1 and 154 DF,  p-value: < 2.2e-16
Call:
lm(formula = diametro ~ peso)

Residuals:
     Min       1Q   Median       3Q      Max
-0.93713 -0.23427  0.00326  0.23303  0.90299

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.284715   0.102788  90.329  < 2e-16 ***
peso        0.006112   0.001265   4.834 3.22e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3327 on 154 degrees of freedom
Multiple R-squared:  0.1317,    Adjusted R-squared:  0.1261
F-statistic: 23.36 on 1 and 154 DF,  p-value: 3.217e-06
```

Figura 2: Summary modelli ridotti