Iniziato	mercoledì, 5 luglio 2023, 16:56
Stato	Completato
Terminato	martedì, 11 luglio 2023, 13:48
Tempo impiegato	5 giorni 20 ore
Valutazione	Non ancora valutato

undefinedG

Domanda 1

Risposta errata

Punteggio ottenuto -0,20 su 1,00

A Gaussian linear model has been fitted to investigate the effect of the categorical regressor `x.cat` (with 3 categories: A, B, C and D) on the dependent variable `y`, using a sample of n units. The output obtained using the `lm` fuction is reported in the following:

```
> summary(M1)

Call:
lm(formula = y ~ x.cat - 1)

Residuals:
   Min     1Q Median     3Q    Max 
-8.588 -2.523 -0.170  2.638 10.338 

Coefficients:
       Estimate Std. Error t value Pr(>|t|)    
x.catA   4.1297     0.9628   4.289 5.98e-05 ***
x.catB   1.4439     0.7458   1.936  0.05713 .  
x.catC   2.8431     0.8338   3.410  0.00111 ** 
x.catD   2.9955     1.1792   2.540  0.01344 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.729 on 66 degrees of freedom
Multiple R-squared:  0.3787,    Adjusted R-squared:  0.341 
F-statistic: 10.06 on 4 and 66 DF,  p-value: 2.038e-06
```

Suppose that one is interested in testing whether `x.cat` has an overall significant effect on `y`. Which of the following matrix $\mathbf{K}$ can be used (jointly with a vector $\mathbf{t}$ composed of 0 with an appropriate number of rows) to express the corresponding null hypothesis H_0 as a system of linear constraints on the parameters of the fitted model?

Scegli un'alternativa:

a.
$$\mathbf{K} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

b.
$$\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

c.
$$\mathbf{K} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Risposta errata.

La risposta corretta è: $\mathbf{K} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

Domanda 2

Risposta errata

Punteggio ottenuto -0,20 su 1,00

Consider the Gaussian linear regression model

$$Y|\mathbf{X} \sim MVN_n\left(\mathbf{X}\beta, \sigma^2\mathbf{I}_n\right).$$

and suppose that one is interested in the system of linear hypotheses

$$H_0 : \mathbf{K}\beta = \mathbf{t}.$$

Which of the following expressions can be used to compute $\hat{\mathbf{b}}_{H_0}$, the maximum likelihood estimate for β in the constrained parametric subspace $\{\mathbf{b} : \mathbf{K}\mathbf{b} = \mathbf{t}\} \in \mathrm{R}^{p+1}$?

Scegli un'alternativa:

a. $\left(\mathbf{K}\hat{\mathbf{b}} - \mathbf{t}\right)^\top \left[\mathbf{K}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\right]^{-1}\left(\mathbf{K}\hat{\mathbf{b}} - \mathbf{t}\right)$

b. $\hat{\mathbf{b}} - \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\left[\mathbf{K}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\right]^{-1}\left(\mathbf{K}\hat{\mathbf{b}} - \mathbf{t}\right)$

c. $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\left[\mathbf{K}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\right]^{-1}\left(\mathbf{K}\hat{\mathbf{b}} - \mathbf{t}\right)$

Risposta errata.

La risposta corretta è: $\hat{\mathbf{b}} - \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\left[\mathbf{K}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{K}^\top\right]^{-1}\left(\mathbf{K}\hat{\mathbf{b}} - \mathbf{t}\right)$

Domanda 3

Parzialmente corretta

Punteggio ottenuto 0,50 su 1,00

Suppose that one is considering a set of candidate parametric statistical models for the same random sample $\boldsymbol{Y}$. In order to select the optimal one, which of the following quantities should be minimized?

Note that more than one option can be selected

Scegli una o più alternative:

L a. $-\ln L(\hat{\theta}) + \frac{k}{2}\ln n$

L b. $2\ln L(\hat{\theta}) - \ln n$

L c. $2\ln L(\hat{\theta}) + k\ln n$

K d. $-2\ln L(\hat{\theta}) + k\ln n$

Risposta parzialmente esatta.

Hai selezionato correttamente 1.
Le risposte corrette sono:
$-2\ln L(\hat{\theta}) + k\ln n$,

$-\ln L(\hat{\theta}) + \frac{k}{2}\ln n$

Domanda 4

Risposta errata

Punteggio ottenuto 0,00 su 1,00

Suppose that the `R` dataframe `dataset` contains information on a dependent variable `y` and a set of candidate regressor `x1`, `x2`, `x3`, `x4`, `x5`, `x6` and `x7`, where `x1`, `x2` and `x3` are numerical regressors, `x4` is a categorical regressor with 2 categories, `x5` and `x6` are categorical regressors with 3 categories and `x7` is a categorical regressor with 4 categories.

Consider the model `M1`, fitted using the following `R` instruction:

```
M1<-lm(y~x1+x2+x6, data=dataset)
```

Which of the following models can be compared with `M1` using R^2?

Note that more than one option can be selected

Scegli una o più alternative:

K a. `lm(y~x1+x2+x3, data=dataset)`

L b. `lm(y~x3+x7, data=dataset)`

L c. `lm(y~x5+x6, data=dataset)`

L d. `lm(y~x1+x2+x3+x4, data=dataset)`

L e. `lm(y~x4+x5+x7, data=dataset)`

Risposta errata.

Le risposte corrette sono:

`lm(y~x3+x7, data=dataset)`,

`lm(y~x1+x2+x3+x4, data=dataset)`,

`lm(y~x5+x6, data=dataset)`

Domanda 5

Risposta errata

Punteggio ottenuto -0,25 su 1,00

Consider a Guassin linear regression model

$Y \mid \mathbf{X} \sim MVN_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n).$

Which of the following expressions can be used to compute the expected Fisher information matrix for the parameter vector $\boldsymbol{\beta}$?

Scegli un'alternativa:

a. $-\frac{\mathbf{X}^\top \mathbf{X}}{\sigma^2}$

b. $\frac{\mathbf{X}^\top \mathbf{X}}{\sigma^2}$

c. $\sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$

d. $\frac{\mathbf{X}^\top [\mathbf{y} - \mathbf{X}\boldsymbol{\beta}]}{\sigma^2}$

Risposta errata.

La risposta corretta è:
$\frac{\mathbf{X}^\top \mathbf{X}}{\sigma^2}$

Domanda 6

Completo

Punteggio max.: 2,00

Suppose that the following Gaussian linear model

$Y_i | \mathbf{x}_i \sim N(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}, \sigma^2)$ independent $i = 1\ldots, n$

has been fitted in `R`. This is a part of the resulting output:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.08039    0.06461  63.152   <2e-16
x1           0.05506    0.08077   0.682   0.4987
x2          -0.16193    0.08101  -1.999   0.0514

Residual standard error: 0.4472 on 47 degrees of freedom
Multiple R-squared:  0.083,   Adjusted R-squared:  0.044
F-statistic: 2.134 on 2 and 47 DF,  p-value: 0.1297
```

Is it correct to state that, considering a significance level $\alpha = 0.05$, at least one of the two regressors has a significant effect on the dependent variable? Motivate your answer.

No, because the F-statistic does not show a significant value

No, it is not correct. Since the p-value associated with the F test statistic reported in the last line of the output is larger that 0.05, the linear independence hypothesis should not be rejected at a significance level 0.05. This implies that none of the regressors has significant effect on the dependent variable.

Precedente

Vai a...

Successivo