DYNEXITE 01:14:18

Next Item

Item 1 **10 points**

Consider a linear model $\boldsymbol{Y} = B\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ according to Definition I.4.2 with design matrix

$$B = \begin{pmatrix} \frac{1}{3} & 1 \\ \frac{2}{3} & b_{22} \\ \frac{2}{3} & b_{32} \end{pmatrix} \in \mathbb{R}^{3\times 2}$$

parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)' \in \mathbb{R}^2$ and error term $\boldsymbol{\varepsilon}$ with $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 I_3$, where $\sigma > 0$ is unknown.

For the following tasks, give your answers by filling the results into the blanks. If necessary, numerical values have to be rounded to three decimal places.

(a) Check if there exists $c \in \mathbb{R}$ with $b_{22} = c$ and $b_{32} = -c$, such that the corresponding least-squares estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is uniquely determined. **2 points**

2 points

If there exists a unique $c \in \mathbb{R}$, then give its value in the space below. If such an $c \in \mathbb{R}$ does not exist, then type "NA" (without quotation marks). If $\widehat{\boldsymbol{\beta}}$ is uniquely determined for all $c \in \mathbb{R}$, then type "R" (without quotation marks).

2

(b) Check if there exists $c \in \mathbb{R}$ with $b_{22} = b_{32} = c$, such that $\boldsymbol{\beta}$ is **not** identifiable. **2 points**

2 points

If there exists a unique $c \in \mathbb{R}$, then give its value in the space below. If such an $c \in \mathbb{R}$ does not exist, then type "NA" (without quotation marks). If $\boldsymbol{\beta}$ is identifiable for all $c \in \mathbb{R}$, then type "R" (without quotation marks).

2

For the following two parts (c) and (d), assume $b_{22} = 0$ and $b_{32} = 1$. Futher assume that the design matrix B possesses the following QR-decomposition:

$$B = \begin{pmatrix} \frac{1}{3} & 1 \\ \frac{2}{3} & 0 \\ \frac{2}{3} & 1 \end{pmatrix} = QR \quad \text{with} \quad Q = \frac{1}{3}\begin{pmatrix} 1 & 2 \\ 2 & -2 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

(c) Calculate the corresponding least-squares estimate $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \widehat{\beta}_2)'$ of $\boldsymbol{\beta}$ for the observation $\boldsymbol{y} = (1,1,0)'$ of $\boldsymbol{Y}$. **3 points**

1.5 points

Give the value of $\widehat{\beta}_1$.

2

1.5 points

Give the value of $\widehat{\beta}_2$.

2

(d) For $\sigma = 1$, determine $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$. **3 points**

1 point

Give the value of c_{11}.

2

1 point

Give the value of c_{12}.

2

1 point

Give the value of c_{22}.

2

Next Item

All answers have been saved!

OVERVIEW 1 2 3 4 SUBMISSION

DYNEXITE 01:14:30

Previous Item | Next Item

Item 2 **10 points**

For given measurements

$$x_1 = 1\,,\; x_2 = 0\,,\; x_3 = -1\,,\; x_4 = 2\,,$$

consider the following polynomial regression model according to I.5.8:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^3 + \varepsilon_i\,,\; i \in \{1,2,3,4\}\,,$$

with (unknown) parameters $\beta_0, \beta_1, \beta_2 \in \mathbb{R}$ and stochastically independent, indentically distributed error terms $\varepsilon_i \sim N(0,\sigma^2)\,,\, i \in \{1,2,3,4\}$, where $\sigma > 0$ is also unknown. Then, it holds:

$$(B'B)^{-1} \approx \begin{pmatrix} 0.33333333 & 0.05555556 & -0.05555556 \\ 0.05555556 & 0.92592593 & -0.25925926 \\ -0.05555556 & -0.25925926 & 0.09259259 \end{pmatrix}$$

with B denoting the corresponding design matrix of the given polynomial regression model.

Further, consider the observation $\boldsymbol{y} = (0,3,0,-3)'$ of $\boldsymbol{Y} = (Y_1,Y_2,Y_3,Y_4)'$ with resulting least-squares estimate

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 1 \\ \frac{2}{3} \\ -\frac{2}{3} \end{pmatrix}$$

of the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$.

For the following tasks, give your answers by filling the results into the blanks. If necessary, numerical values have to be rounded to three decimal places.

(a) For the given polynomial regression model, consider the following testing problem

$$H_0 : (\beta_1,\beta_2)' = (0,0)' \quad \longleftrightarrow \quad H_1 : (\beta_1,\beta_2)' \neq (0,0)'\,.$$

For this testing problem, the corresponding decision rule based on the F-distribution can be formulated as follows:

$$\text{Rejection of } H_0 : (\beta_1,\beta_2)' = (0,0)' \text{ at significance level } \alpha \in (0,1)\,,\text{ if } \frac{\mathrm{SSR}/\mathrm{df1}}{\mathrm{SSE}/\mathrm{df2}} > F_{1-\alpha}(\mathrm{df1},\mathrm{df2})$$

with the quantities SSR, SSE and the degrees of freedom $\mathrm{df1}$ and $\mathrm{df2}$ being appropriately defined and with $F_{1-\alpha}(\mathrm{df1},\mathrm{df2})$ denoting the $(1-\alpha)-$quantile of the $F(\mathrm{df1},\mathrm{df2})$-distribution.

(i) Determine the degrees of freedom $\mathrm{df1}$ and $\mathrm{df2}$. 1 point

Give the value of $\mathrm{df1}$. 0.5 points

2

Give the value of $\mathrm{df2}$. 0.5 points

2

(ii) Calculate the values of SSR and SSE for the given observation $\boldsymbol{y}$. **Hint: In order to derive these values, it is *not* neccessary to calculate any orthogonal projector.** 5 points

Give the value of SSR. 2.5 points

2

Give the value of SSE. 2.5 points

2

(b) Now, for the given polynomial regression model, consider the following testing problem

$$H_0 : \beta_1 = 0 \quad \longleftrightarrow \quad H_1 : \beta_1 \neq 0\,.$$

(i) For the reduced regression model associated with H_0, consider the corresponding design matrix

$$B_0 = \begin{pmatrix} 1 & b_{12} \\ 1 & b_{22} \\ 1 & b_{32} \\ 1 & b_{42} \end{pmatrix} \in \mathbb{R}^{4\times 2}\,.$$

Determine the last entry of the second column of B_0. 1 point

Give the value of b_{42}. 1 point

2

(ii) For the testing problem considered in part (b), the corresponding decision rule based on the t-distribution can be formulated as follows:

$$\text{Rejection of } H_0 : \beta_1 = 0 \text{ at significance level } \alpha \in (0,1)\,,\text{ if } \frac{|\widehat{\beta}_1|}{\sqrt{6\,c}} > q(\alpha)\,.$$

Here, $\widehat{\beta}_1$ denotes the least-squares estimate of the parameter β_1, $c > 0$ is some appropriate quantity and $q(\alpha)$ denotes the quantile of the corresponding t-distribution.

Determine the quantity c and the quantile $q(\alpha)$ for $\alpha = 0{,}1$. 3 points

1 point

Give the value of $q(\alpha)$, **rounded to threee decimal places**.

2

2 points

Give the value of c, **rounded to threee decimal places**.

2

Previous Item

Next Item

All answers have been saved!

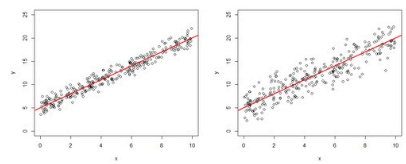
← Previous Item

Item 4 **10 points**

In the following task, the corresponding points are given if **all** answers are chosen correctly. **Multiple choices are possible.**

(a) Observe the two scatterplots (see below) of randomly generated data, where on the x-axis the explanatory variable is displayed and on the y-axis the response variable. We call the regression line of the left figure model 1 and the other, on the right figure, model 2. Choose the **right** answer/answers. 2 points

- [x] (i) Model 1 has higher sum of residuals than model 2.
- [ ] (ii) Model 1 has lower sum of residuals than model 2.
- [ ] (iii) Both models have the same sum of residuals.
- [ ] (iv) The sum of residuals is always zero.
- [ ] (v) None of the statements (i)-(iv) is correct.



scatterplot for **(a)**

(b) Which of the following assumptions is **not** an assumption for a normal linear model with fixed effects? 1 point

- [x] (i) Uncorrelated explanatory variables
- [ ] (ii) Linear relationship between the response variable and the explanatory variables
- [ ] (iii) Constant variance of the errors
- [ ] (iv) Normally distributed explanatory variables
- [ ] (v) The response and the explanatory variables are jointly multivariate normally distributed

(c) Which of the following assumptions is **not** an assumption about the error term ε in a normal linear model ? 1 point

- [x] (i) The error term ε follows a normal distribution.
- [ ] (ii) The expected value of the error term ε is equal to one.
- [ ] (iii) The error terms $\varepsilon_i, i = 1, \dots, n$ are constant.
- [ ] (iv) The error terms $\varepsilon_i, i = 1, \dots, n$ are independent.
- [ ] (v) The variance of the error term ε is the same for all values of the explanatory variables.
- [ ] (vi) The error terms ε_i, $i = 1, \dots, n$ are uncorrelated.

(d) Consider the following two linear models. First, a linear model with response variable Y and two explanatory variables X_1, X_2. Second, a linear model with response variable Y and explanatory variable X_1. Both models are fitted on the same dataset, providing the `R` output that can be found below. Answer the following questions. 2 points

(i) What is the marginal effect of X_1 on the response variable? 0.5 points

2

(ii) Fixing X_2, what is the conditional effect of X_1 on the response variable? 0.5 points

2

(iii) Consider the model $Y = a + bX_1 + cX_2$. Test $H_0 : c = 0$ versus $H_1 : c \neq 0$. On a significance level of $\alpha = 0.05$, would you reject the null hypothesis? 0.5 points reject H_0

0.5 points

Give the respective p-value of (iii)

2

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.7573 -1.6734 -0.1721  1.5349  7.0208 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.8106     0.8252   3.406  0.00096 ***
x1            2.0803     0.1927  10.793  < 2e-16 ***
x2           -0.4112     0.3894  -1.056  0.29366    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.214 on 97 degrees of freedom
Multiple R-squared:  0.5479,	Adjusted R-squared:  0.5386 
F-statistic: 58.77 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = y ~ x1)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5701 -1.5483 -0.1202  1.4137  7.0602 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.6088     0.8032   3.248  0.00159 ** 
x1            2.0797     0.1928  10.784  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.215 on 98 degrees of freedom
Multiple R-squared:  0.5427,	Adjusted R-squared:  0.538 
F-statistic: 116.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

output for **(d)**

(e) Suppose that you have one explanatory variable X and a response variable Y where you fit a linear model based on n observations. Assume that the rank of the design matrix is 2. Which of the following statements are **wrong**?

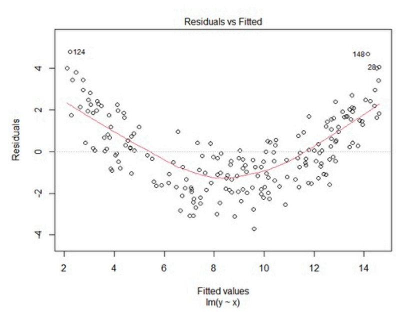
2 points

- [x] (i) The SSE is always larger than SST
- [ ] (ii) If we have $y_i = c \in \mathbb{R}$ for all $i = 1, \ldots n$, then SSE$= 0$
- [ ] (iii) The SSR can never be equal to one
- [ ] (iv) The SSR can never be equal to zero

(f) Suppose that you fit a simple linear model with explanatory variable X and response variable Y. Now, one of the diagnostic plot looks as follows (see below). Choose the **right** answer/answers.

2 points

- [x] (i) The plot indicates that the data contains observations with high leverage.
- [ ] (ii) The plot indicates that we have non-linearity in the data.
- [ ] (iii) The plot indicates that the error terms have a non-zero expected value.
- [ ] (iv) The plot indicates that the data contains outliers.
- [ ] (v) The plot indicates that we have a non-constant variance of the error terms.



diagnostic plot of **(f)**

Previous
Item

All answers have been saved!

DYNEXITE 01:14:47

Previous Item | Next Item

Item 3 **10 points**

For values of $x \in \mathbb{R}$ suppose the linear model $(Y|X=x) \sim \mathcal{N}(\mu(x), \sigma^2)$ holds with

$$E(Y|X=x) = \mu(x) = 1 + 2x - x^2$$

and $\sigma = 2$. Generate $n = 50$ observations of $X \sim \mathcal{N}(0,1)$ and calculate the values $\mu(x_i) = 1 + 2x_i - x_i^2, i = 1, \ldots, n$, and, having that, sample the resulting values $\boldsymbol{y} = (y_1, \ldots, y_n)$ for the response Y by using the following `R` code (it is important to use the specific seed):

```
set.seed(2022)
n=50
x=rnorm(n)
mu=1+2*x-x*x
y=mu+rnorm(n,sd=2)
```

(a) What is the sample standard deviation of the sample $\boldsymbol{x} = (x_1, \ldots, x_n)$? Further, give the sample mean values (i) for $\mu(\boldsymbol{x}) = (\mu(x_1), \ldots, \mu(x_n))$ and (ii) for $\boldsymbol{y} = (y_1, \ldots, y_n)$. **(requested precision: 2 digits)** 1.5 points

sample standard deviation of $\boldsymbol{x}$ 0.5 points

2

sample mean value (i) 0.5 points

2

sample mean value (ii) 0.5 points

2

(b) Fit the model $\mu^{(0)}(x) = \beta_1 x$. What is the resulting estimate of the regression coefficient? **(requested precision: 4 digits)** 1 point

$\hat{\beta}_1$ 1 point

2

(c) Fit the model $\mu^{(1)}(x) = \beta_0 + \beta_1 x + \beta_2 x^2$. What are the resulting estimates of the regression coefficients of x and x^2? **(requested precision: 4 digits)** 1 point

$\hat{\beta}_1$ 0.5 points

2

$\hat{\beta}_2$ 0.5 points

2

(d) For both models in (b) and (c), calculate $\delta_j := \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}^{(j)}(x_i) - \mu(x_i))^2$ for $j \in \{0,1\}$, with $j=0$ and $j=1$ corresponding to the models in (b) and (c), respectively. **(requested precision: 4 digits)** 1 point

δ_0 0.5 points

2

δ_1 0.5 points

2

(e) Based on the values of $\delta_j, j \in \{0,1\}$, calculated in (d), which model do you prefer? 0.5 points

Deselect

- Model in (c)
- Model in (b)

(f) For both models in (b) and (c), use the Shapiro-Wilk test to test on level $\alpha = 0.05$, if there is evidence against the assumption of normally distributed residuals. Give the resulting p-values and your decision whether there is evidence or not. **(requested precision: 4 digits)** 4 points

p-value for model in (b) 1 point

2

1 point

decision for model in (b)

reject H_0

1 point

p-value for model in (c)

2

1 point

decision for model in (c)

reject H_0

(g) In the setting of (c), test for the parameter β_2 the hypotheses
$H_0 : \beta_2 = 0 \quad \textbf{versus} \quad H_1 : \beta_2 \neq 0$ on the significance level $\alpha = 0.05$. Do you reject H_0? Give the corresponding p-value. **(requested precision: 4 digits)**

1 point

0.5 points

p-value

2

0.5 points

decision

reject H_0

Previous
Item

Next
Item

All answers have been saved!

OVERVIEW

1 2 3 4

SUBMISSION