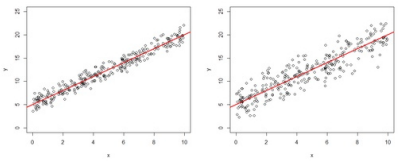
Previous Item

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 $\varepsilon_i, i = 1, \ldots, n$ are constant.
- [ ] (iv) The error terms $\varepsilon_i, i = 1, \ldots, n$ are independent.
- [ ] (v) The variance of the error term ε is the same for all values of the explanatory variables.
- [ ] (vi) The error terms $\varepsilon_i, i = 1, \ldots, n$ 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
```

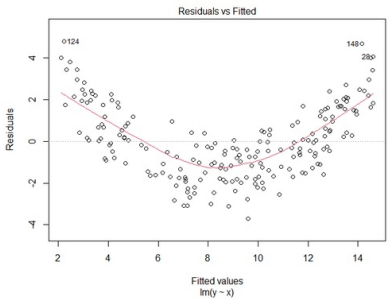
output for **(d)**

(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!