Assignment sheet 10

November 28, 2018

Assignment 1 (Extreme cases $p \in \{0, n\}$)**.** Consider a linear model $y = X\beta + \varepsilon$ where $X_{n\times p}$ is full column-rank and $\varepsilon \sim N_n(0, \sigma^2 I)$. Assume $n = p$. What are the estimators for β et σ^2, the errors and the fitted values ? Comment on it. Do the same for the cases $p = 0$ and $p = 1$.

Assignment 2 (Diagnostic)**.** a) Figure 1 represents the the standardized residuals plotted against the fitted values for 4 different set of x_i's. For every case, discuss the fit of the model and explain briefly how you could fix mis-fits, if any are present.

b) Figure 2 represents 4 gaussian Q-Q plots. In every case, the covariates do not come from a Gaussian distribution. Actually, they are generated from distributions with

  i) tails heavier than gaussian ;
  ii) tails lighter than gaussian ;
  iii) positive *skewness* ;
  iv) negative skewness.

Match each case i)–iv) with a Q-Q plot from Figure 2 and comment.

Assignment 3 (confidence and prediction intervals)**.** The following table gives the estimators, the standard errors and the correlations for the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ fitted to $n = 13$ cement values from the example given in the course.

```
             Estimate   SE          Correlations of Estimates
(Intercept)  48.19      3.913           (Intercept)    x1     x2
         x1   1.70      0.205       x1      -0.736
         x2   0.66      0.044       x2      -0.416   -0.203
         x3   0.25      0.185       x3      -0.828    0.822  -0.089
```

a) Explain how R calculates the standard errors and the correlations that appear in the above table.

b) What is the prediction, under this model, of y when $x_1 = x_2 = x_3 = 1$? By how much it will change if instead $x_1 = 5$? And if $x_1 = x_2 = 5$?

c) Using only the above information and the following quantiles $t_9(0.975) = 2.262$, $t_9(0.95) = 1.833$ of the student distribution with 9 degrees of freedom, calculate under this model the confidence intervals for β_0, β_1, β_2 et β_3 at significance level $\alpha = 0.95$. Calculate a 0.9 confidence interval for $\beta_2 - \beta_3$.

Assignment 4. We fit the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$ to the cement dataset from the course ($n = 13$). R gives the following table :

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 48.19363    3.91330  12.315 6.17e-07 ***
x1           1.69589    0.20458   8.290 1.66e-05 ***
x2           0.65691    0.04423  14.851 1.23e-07 ***
x3           0.25002    0.18471   1.354    0.209
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

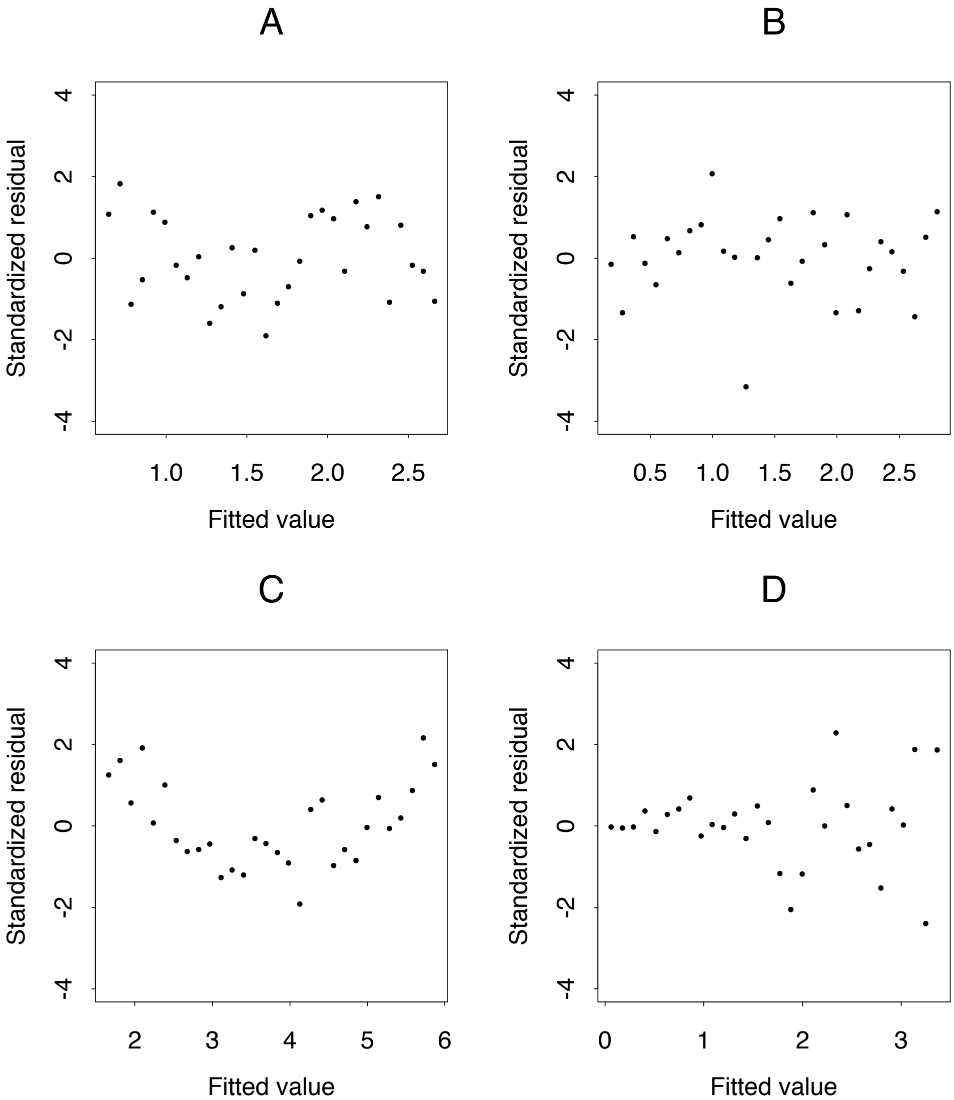

Figure 1 – Standardised residuals vs fitted values, Gaussian models

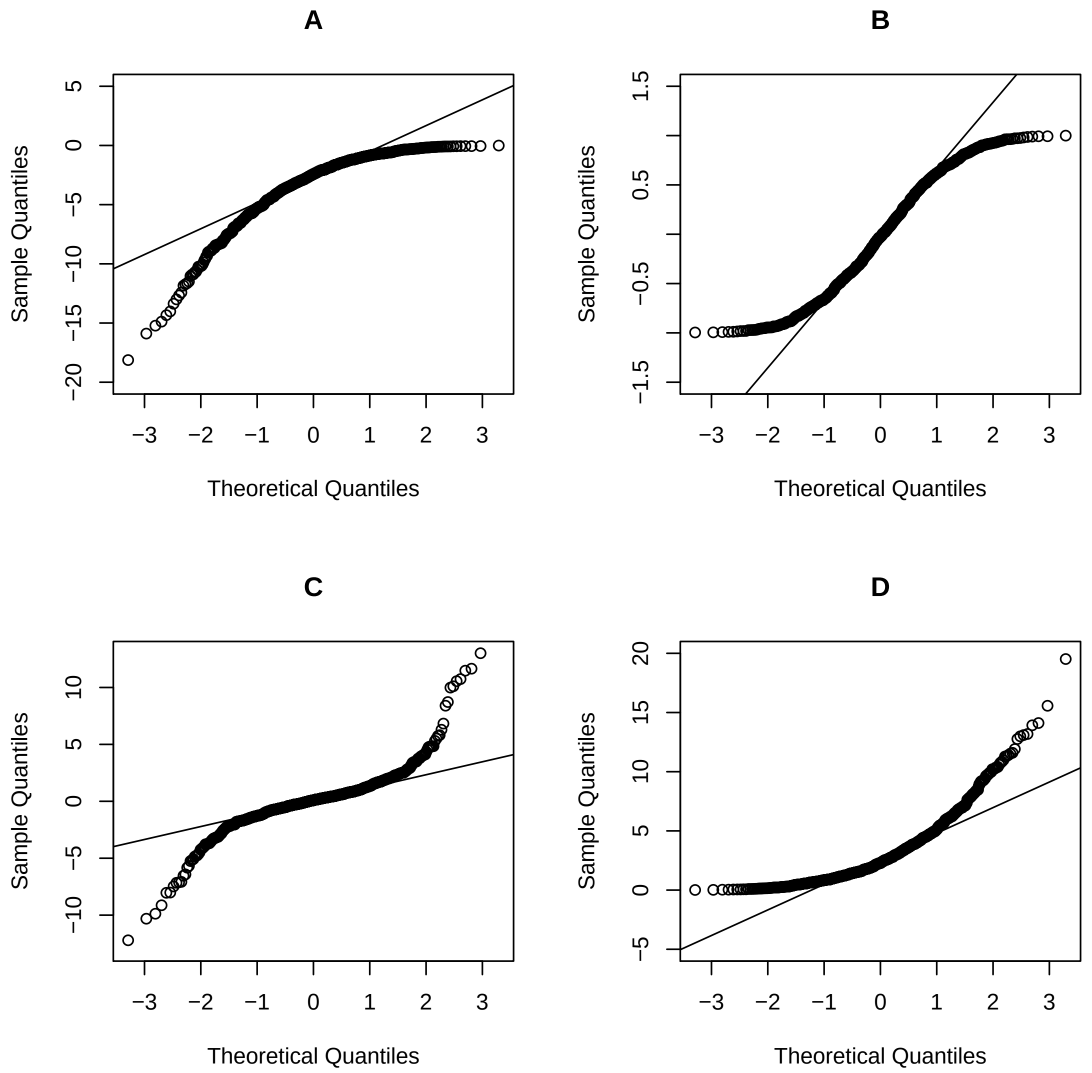

Figure 2 – Normal Q-Q plots for non Normal covariates.

a) Explain in detail how R calculates the values in the columns "`t value`" and "`Pr(>|t|)`". What do these values mean ? Comment the observed numbers in the table.

b) Knowing that $\widehat{\operatorname{corr}}(\hat{\beta}_2, \hat{\beta}_3) = -0.08911$, what is the p-value for the null hypothesis $\beta_2 - \beta_3 = 0$? For a 0.05 test, can we reject the null hypothesis ?

Assignment 5 (best design)**.** Consider the linear model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, \ldots, n,$$

where $\beta_0, \beta_1 \in \mathbb{R}$, $\mathbb{E}[\epsilon] = 0$ and $\operatorname{Var} \epsilon = \sigma^2 I_n$ (and $n \geq 2$). This is called *simple linear regrssion.*

(i). Write down the design matrix for this model and give a necessary and sufficient condition for it to be of full rank.

(ii). Find the covariance matrix of the least squares estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)^T$.

(iii). Suppose that you can choose all the $x_1, \ldots, x_n$ as you wish, but constrained to be in $[-1, 1]$. How would you choose them in order to minimise the variance of $\widehat{\beta}_1$?