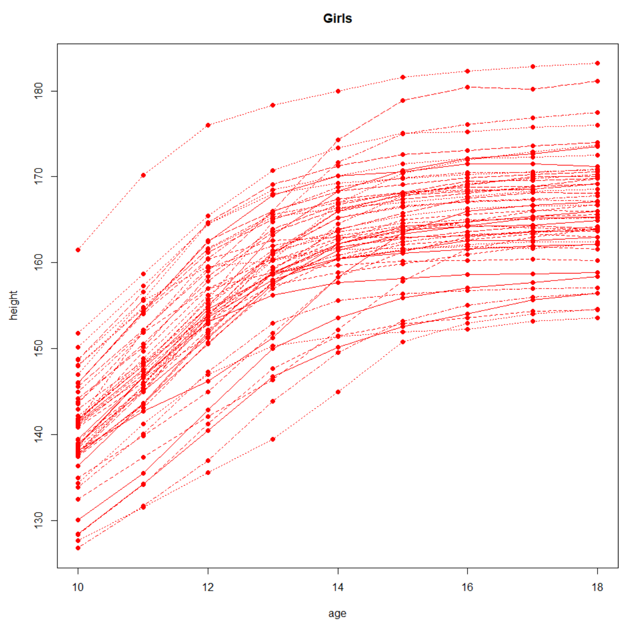
1. Consider the data set growthheight.txt related to growth profiles of sample of boys and girls:

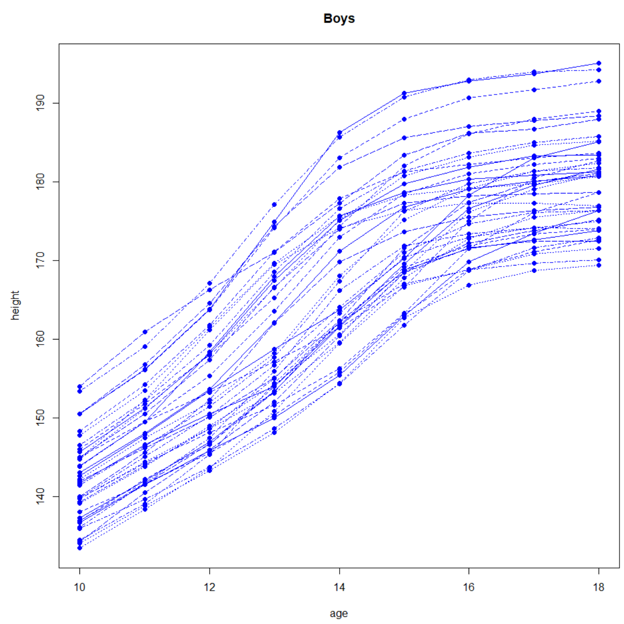
```
> head(data)
        gender   Y10   Y11   Y12   Y13   Y14   Y15   Y16   Y17   Y18
girl01    girl 138.6 146.8 153.1 156.2 157.7 158.2 158.6 158.7 158.9
girl02    girl 140.9 146.1 152.9 159.5 162.6 165.0 165.6 166.1 166.0
girl03    girl 148.8 155.6 159.6 160.3 161.6 161.7 161.9 161.7 162.2
girl04    girl 143.0 148.5 154.8 161.2 165.2 166.5 167.2 167.4 167.8
girl05    girl 141.0 147.0 153.0 161.0 166.0 168.0 169.0 170.0 170.0
> tail(data)
      gender   Y10   Y11   Y12   Y13   Y14   Y15   Y16   Y17   Y18
boy35    boy 154.0 160.9 166.3 171.0 177.3 183.4 186.2 186.7 188.0
boy36    boy 143.0 148.0 153.6 162.1 171.2 176.8 179.1 180.1 180.8
boy37    boy 148.3 154.2 161.8 171.1 177.9 181.3 182.2 183.1 183.7
boy38    boy 147.8 153.5 161.5 169.7 175.6 178.3 179.2 179.8 180.7
boy39    boy 139.8 145.1 150.2 156.7 164.1 171.0 174.7 176.1 176.4

A list containing the heights of 39 boys and 54 girls from
age 10 to 18 and the ages at which they were collected.

gender - values of boy and girl
Y10, Y11, ..., Y18 - the variables giving the heights in centimeters of
                     children at ages 10,11,12,...18.

Tuddenham, R. D., and Snyder, M. M. (1954)
"Physical growth of California boys and girls from birth to age 18",
University of California Publications in Child Development, 1, 183-364.
```





Denote variables as following $Y_1 = \text{Y10}, Y_2 = \text{Y11}, \ldots, Y_9 = \text{Y18}$ and $X = \texttt{gender}$ with index variable j associated to its values. Let us assume that the random vector (i.e, random vector for each row i)

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{i9} \end{pmatrix}$$

follows the normal distribution $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where the expected value vector $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{i9})'$ is modeled by the multivariate linear model

$$\mathcal{M}: \quad \boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{i9} \end{pmatrix} = \begin{pmatrix} \beta_{0_1} + \beta_{1_1} x_i \\ \beta_{0_2} + \beta_{1_2} x_i \\ \vdots \\ \beta_{0_9} + \beta_{1_9} x_i \end{pmatrix} = \begin{pmatrix} \beta_{0_1} & \beta_{1_1} \\ \beta_{0_2} & \beta_{1_2} \\ \vdots & \vdots \\ \beta_{0_9} & \beta_{1_9} \end{pmatrix} \begin{pmatrix} 1 \\ x_i \end{pmatrix} = \mathbf{B}'\mathbf{x}_i,$$

where x_i is a dummy variable getting value 1 when children is a girl and value 0 when a boy.

(a) Let us write the model for whole data as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E},$$

where, for each row of $\mathbf{E}$, it is assumed $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Calculate the maximum likelihood estimate $\widehat{\mathbf{B}}$ for the parameters $\mathbf{B}$.

(1 point)

(b) Find the unbiased estimate $\widehat{\boldsymbol{\Sigma}}$ for the covariance matrix $\boldsymbol{\Sigma}$.

(1 point)

(c) Predict the value of the new observation $\mathbf{y}_f$ in both cases of a child being girl (i.e., $x_f = 1$) and a boy (i.e., $x_f = 0$). Create also a graph that displays the predicted profiles for a girl and a boy in the same graph.

(1 point)

(d) Test at 5% significance level, is the explanatory variable $X = \texttt{gender}$ statistically significant variable in the model. Calculate the value of the test statistic.

(1 point)

(e) Consider the partitioned random vector

$$\mathbf{y}_* = \begin{pmatrix} \mathbf{y}_{*1} \\ \mathbf{y}_{*2} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1' \\ \mathbf{B}_2' \end{pmatrix} \mathbf{x}_* + \begin{pmatrix} \boldsymbol{\varepsilon}_{*1} \\ \boldsymbol{\varepsilon}_{*2} \end{pmatrix},$$

where $\mathbf{y}_{*1}$ contains the random variables $\mathbf{y}_{*1} = (y_{i_*1}, y_{i_*2}, y_{i_*3})'$. Predict the value of the random vector $\mathbf{y}_{*2} = (y_{i_*4}, y_{i_*5}, y_{i_*6}, y_{i_*7}, y_{i_*8}, y_{i_*9})'$ when $\mathbf{y}_{*1}$ and $\mathbf{x}_*$ have observed values

```
gender   Y10   Y11   Y12
  girl 148.7 156.6 164.7
```

Create also 80 % simultaneous (asymptotic) prediction intervals for elements of the random vector $\mathbf{y}_{*2}$.

(2 points)

2. Smith, Gnanadesikan, and Hughes (1962) provide data on characteristics of the urine of young men. The men are categorized into four groups based on their degree of obesity. The four variables given in consist of $X_1 = 10^3((\text{specific gravity}) - 1)$, $X_2 = \text{obesity}$, and three dependent variables $Y_1 = \text{pigment creatinine}$, $Y_2 = \text{chloride}$, and $Y_3 = \text{chlorine}$. The data can be found on the file excretory.txt.

Group I				Group II				Group III				Group IV			
x	y_1	y_2	y_3	x	y_1	y_2	y_3	x	y_1	y_2	y_3	x	y_1	y_2	y_3
24	17.6	5.15	7.5	31	18.1	9.00	14.5	18	17.0	4.55	1.9	32	12.5	2.90	22.5
32	13.4	5.75	7.1	23	19.7	5.30	12.5	10	12.5	2.65	0.7	25	8.7	3.00	19.5
17	20.3	4.35	2.3	32	16.9	9.85	8.0	33	21.5	6.50	8.3	28	9.4	3.40	1.3
30	22.3	7.55	4.0	20	23.7	3.60	4.9	25	22.2	4.85	9.3	27	15.0	5.40	20.0
30	20.5	8.50	2.0	18	19.2	4.05	0.2	35	13.0	8.75	13.0	23	12.9	4.45	1.0
27	18.5	10.25	2.0	23	18.0	4.40	3.6	33	13.0	5.20	18.3	25	12.1	4.30	5.0
25	12.1	5.95	16.8	31	14.8	7.15	12.0	31	10.9	4.75	10.5	26	13.2	5.00	3.0
30	12.0	6.30	14.5	28	15.6	7.25	5.2	34	12.0	5.85	14.5	34	11.5	3.40	5.1
28	10.1	5.45	0.9	21	16.2	5.30	10.2	16	22.8	2.85	3.3				
24	14.7	3.75	2.0	20	14.1	3.10	8.5	31	16.5	6.55	6.3				
26	14.8	5.10	0.4	15	17.5	2.40	9.6	28	18.4	6.60	4.9				
27	14.4	4.05	3.8	26	14.1	4.25	6.9								
				24	19.1	5.80	4.7								
				16	22.5	1.55	3.5								

Smith, H., Gnanadesikan, R., & Hughes, J. B. (1962). Multivariate analysis of variance (MANOVA). Biometrics, 18, 22?41.

Model the variables $Y_1 = \text{pigment creatinine}$, $Y_2 = \text{chloride}$, and $Y_3 = \text{chlorine}$ by the multivariate linear model moniulotteisella varianssimallilla

$$
\begin{aligned}
y_{i1} &= \beta_{0_1} + \beta_{1_1} x_{i1} + \alpha_{j_1} + \varepsilon_{i1} \\
y_{i2} &= \beta_{0_2} + \beta_{1_2} x_{i1} + \alpha_{j_2} + \varepsilon_{i2} \\
y_{i3} &= \beta_{0_3} + \beta_{1_3} x_{i1} + \alpha_{j_3} + \varepsilon_{i3},
\end{aligned}
$$

where parameters $\alpha_{j_1}, \alpha_{j_2}, \alpha_{j_3}$ are related to the categories of the variable $X_2 = \text{obesity}$. It is assumed that $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3})'$ follows th multivariate normal distribution $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_i = \mathbf{B}'\mathbf{x}_i$. The model for the data can be written also as

$$
\mathbf{Y} = \mathbf{XB} + \mathbf{E}.
$$

(a) Predict the value of the new observation $\mathbf{y}_f$, when

$$
x_{f1} = 35, \qquad x_{f2} = \text{“Group III”}.
$$

Create also 80 % simultaneous prediction intervals for elements of the random vector $\mathbf{y}_f$.

(2 points)

(b) Test the hypotheses

$$
\begin{aligned}
H_0 &: \beta_{1_1} = 0 \text{ and } \beta_{1_2} = 0, \\
H_1 &: \beta_{1_1} \neq 0 \text{ or } \beta_{1_2} \neq 0.
\end{aligned}
$$

(2 points)

(c) Test does the variable $Y_3 = \text{chlorine}$ contains any additional information about te parameters $\mathbf{B}$ beyond that is available in variables $Y_1 = \text{pigment creatinine}$, $Y_2 = \text{chloride}$.

(2 points)

3. (a) Let us go back the Question 1 and consider again the data set growthheight.txt. From the given graphs, it looks like that growth of girls stops around at the age of 16. After that, it seems that the average height level stays at the same also for the ages of 17 and 18. In considered model

$$\mathcal{M}: \quad \boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{i9} \end{pmatrix} = \begin{pmatrix} \beta_{0_1} + \beta_{1_1} x_i \\ \beta_{0_2} + \beta_{1_2} x_i \\ \vdots \\ \beta_{0_9} + \beta_{1_9} x_i \end{pmatrix} = \begin{pmatrix} \beta_{0_1} & \beta_{1_1} \\ \beta_{0_2} & \beta_{1_2} \\ \vdots & \vdots \\ \beta_{0_9} & \beta_{1_9} \end{pmatrix} \begin{pmatrix} 1 \\ x_i \end{pmatrix} = \mathbf{B}'\mathbf{x}_i,$$

the expected values $\mu_{i7}, \mu_{i8}, \mu_{i9}$ are related to the ages 16,17, and 18, respectively. Thus formally test the hypotheses

$$H_0 : \mu_{i_*7} = \mu_{i_*8} = \mu_{i_*9},$$
$$H_1 : \mu_{i_*7} \neq \mu_{i_*8} \neq \mu_{i_*9},$$

when the child i_* is girl. Report you R-code on performing the testing.

(3 points)

(b) Under the assumption of normality, the multivariate linear model $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$ can written in a form

$$\text{vec}(\mathbf{Y}) \sim N\left[(\mathbf{I} \otimes \mathbf{X})\,\text{vec}(\mathbf{B}), \boldsymbol{\Sigma} \otimes \mathbf{I}\right],$$

where $\text{Cov}(\text{vec}(\mathbf{Y})) = \text{Cov}(\text{vec}(\mathbf{E})) = \boldsymbol{\Sigma} \otimes \mathbf{I}$, and hence $\text{Cov}(\text{vec}(\mathbf{Y}')) = \text{Cov}(\text{vec}(\mathbf{E}')) = \mathbf{I} \otimes \boldsymbol{\Sigma}$. The fitted values of the model are $\mathbf{X}\widehat{\mathbf{B}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Show what form the covariance matrix of the vectorized fitted values $\text{vec}(\mathbf{X}\widehat{\mathbf{B}})$ has. That is, find a expression for the covariance matrix

$$\text{Cov}\left(\text{vec}(\mathbf{X}\widehat{\mathbf{B}})\right).$$

(3 points)