Homework 2 Exercise 2

Chutian Chen cc4515; Congcheng Yan cy2550; Mingrui Liu ml4404

2020/3/5

Question 1

```r
cov1 #Complete case analysis
```

```
##        x1     x2     x3     x4     x5
## x1 216.30  -7.50  45.05  77.65  94.50
## x2  -7.50 221.50 117.50  77.00 226.75
## x3  45.05 117.50 157.30  85.90 242.00
## x4  77.65  77.00  85.90  75.20 132.25
## x5  94.50 226.75 242.00 132.25 422.00
```

```r
cov2 #Available case analysis
```

```
##            x1         x2        x3        x4       x5
## x1 121.363636   4.563636  35.79091  42.12727  94.5000
## x2   4.563636 179.134199 112.26840 114.60173 172.5000
## x3  35.790909 112.268398 151.48918 125.96537 182.3727
## x4  42.127273 114.601732 125.96537 153.56061 142.8636
## x5  94.500000 172.500000 182.37273 142.86364 294.5636
```

```r
cov3 #Mean imputation
```

```
##          x1        x2        x3        x4        x5
## x1 57.79221   2.17316  17.04329  20.06061  21.50138
## x2  2.17316 179.13420 112.26840 114.60173  82.14286
## x3 17.04329 112.26840 151.48918 125.96537  86.84416
## x4 20.06061 114.60173 125.96537 153.56061  68.03030
## x5 21.50138  82.14286  86.84416  68.03030 140.26840
```

```r
cov4 #Mean imputation with the bootstrap
```

```
##            x1          x2        x3        x4        x5
## x1 53.1586057  -0.1568703  14.22982  17.16402  17.23394
## x2 -0.1568703 162.7418831  98.32494 101.40327  70.06172
## x3 14.2298190  98.3249351 135.81292 112.41628  77.80261
## x4 17.1640157 101.4032684 112.41628 139.39108  62.57353
## x5 17.2339410  70.0617213  77.80261  62.57353 124.33218
```

```r
cov_em #The EM-algorithm
```

```
##          x1        x2        x3        x4        x5
## x1 81.02002  17.23897  58.73639  58.82235  87.66657
## x2 17.23897 179.13420 112.26840 114.60173 154.20420
## x3 58.73639 112.26840 151.48918 125.96537 197.58721
## x4 58.82235 114.60173 125.96537 153.56061 155.13043
## x5 87.66657 154.20420 197.58721 155.13043 286.57374
```

The covariance matrix of Mean imputation and Mean imputation with the bootstrap are quite similar. The covariance matrix of Complete case analysis is the largest one.

Question 2

Since $\sqrt{n}\left(\hat{\lambda}_1 - \lambda_1\right) \to N\left(0, 2\lambda^2\right)$, we have $\hat{\lambda}_1 \sim N\left(\lambda_1, \frac{2\lambda^2}{n}\right)$. So,

$$P\left[-Z_{1-\frac{\alpha}{2}} \le \frac{\hat{\lambda}_1 - \lambda_1}{\sqrt{\frac{2\hat{\lambda}_1^2}{n}}} \le Z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$-\frac{Z_{1-\frac{\alpha}{2}}\sqrt{2}\hat{\lambda}_1}{\sqrt{n}} \le \hat{\lambda}_1 - \lambda_1 \le \frac{Z_{1-\frac{\alpha}{2}}\sqrt{2}\hat{\lambda}_1}{\sqrt{n}}$$

$$\hat{\lambda}_1 - \frac{Z_{1-\frac{\alpha}{2}}\sqrt{2}\hat{\lambda}_1}{\sqrt{n}} \le \lambda_1 \le \hat{\lambda}_1 + \frac{Z_{1-\frac{\alpha}{2}}\sqrt{2}\hat{\lambda}_1}{\sqrt{n}}$$

```r
get_interval(max(eigen(cov1)$value))#Complete case analysis
```

```
## [1] "Left: 313.861953022843 Right: 1220.7736242501"
```

```r
get_interval(max(eigen(cov2)$value))#Available case analysis
```

```
## [1] "Left: 268.200413918438 Right: 1043.17196834864"
```

```r
get_interval(max(eigen(cov3)$value))#Mean imputation
```

```
## [1] "Left: 187.434302867348 Right: 729.030234523234"
```

```r
get_interval(max(eigen(cov4)$value))#Mean imputation with the bootstrap
```

```
## [1] "Left: 166.894421217937 Right: 649.139870236248"
```

```r
get_interval(max(eigen(cov_em)$value))#The EM-algorithm
```

```
## [1] "Left: 271.304422988596 Right: 1055.24508637326"
```

The interval of Mean imputation and Mean imputation with the bootstrap are similar. Available case analysis and The EM-algorithm are also close. We need to see the full data to check their performance.

Question 3

```r
cov6 #full data
```

```
##            mechanics   vectors   algebra  analysis statistics
## mechanics   305.7680 127.22257 101.57941 106.27273  117.40491
## vectors     127.2226 172.84222  85.15726  94.67294   99.01202
## algebra     101.5794  85.15726 112.88597 112.11338  121.87056
## analysis    106.2727  94.67294 112.11338 220.38036  155.53553
## statistics  117.4049  99.01202 121.87056 155.53553  297.75536
```

```r
get_interval(max(eigen(cov6)$value))
```

```
## [1] "Left: 281.004776384484 Right: 1092.97484449639"
```

Compared each method with the full data, the EM-algorithm is the best fit for the data. Since the size of the sample is small (only 22), it would make more sense if we test on a larger sample.

Question 4

$$
\begin{aligned}
\ell_i\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{x}_i\right) &= -\frac{p}{2}\log 2\pi - \frac{1}{2}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right) \\
&= -\frac{p}{2}\log 2\pi - \frac{1}{2}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\operatorname{tr}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)\left(\mathbf{x}_i-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}
\end{aligned}
$$

We have the derivative partial

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\mu}}\ell_i\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{x}_i\right) &= -\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right) \\
\frac{\partial}{\partial\boldsymbol{\Sigma}}\ell_i\left(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{x}_i\right) &= -\frac{1}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_i-\boldsymbol{\mu}\right)\left(\mathbf{x}_i-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1} \\
&= \frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\left(\mathbf{x}_i-\boldsymbol{\mu}\right)\left(\mathbf{x}_i-\boldsymbol{\mu}\right)^T - \boldsymbol{\Sigma}\right)\boldsymbol{\Sigma}^{-1}
\end{aligned}
$$

We can easily derive the equation

$$\mu^{(k+1)}:\sum_{i=1}^{n}\left(\hat{X}_i-\mu\right)=0$$

For the second equation,with

$$E\left(X_{im}|X_{io}\right)=\mu_{im}^{(k)}+\Sigma_{imo}^{(k)}\left(\Sigma_{ioo}^{(k)}\right)^{-1}\left(X_{io}-\mu_{io}^{(k)}\right)$$

$$\mathbf{C}_{imm}^{(k)}=\Sigma_{imm}^{(k)}-\Sigma_{imo}^{(k)}\left(\Sigma_{ioo}^{(k)}\right)^{-1}\Sigma_{iom}^{(k)}=E\left(X_{im}X_{im}'|X_{io}\right)$$

We can get

$$\sum_{i=1}^{n}\Sigma^{-1}\left(\Sigma-\left(\boldsymbol{\mu}-\hat{\mathbf{x}}_i\right)\left(\boldsymbol{\mu}-\hat{\mathbf{x}}_i\right)^T-\mathbf{C}_i\right)\boldsymbol{\Sigma}^{-1}=0$$

$$\Sigma^{(k+1)}:\sum_{i=1}^{n}\left(\Sigma-\left(\hat{X}_i-\mu\right)\left(\hat{X}_i-\mu\right)^T-\mathbf{C}_i^{(k)}\right)=0$$