Stat501_Homework2

Kelby Kies

2/28/2021

Question 1:

Part A: Read in the file: you may use readxl or XLConnect or some other package. (I usually use readxl but you should use whatever package you find most comfortable to work with.) Store this dataset into a file called senators. [5 points]

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

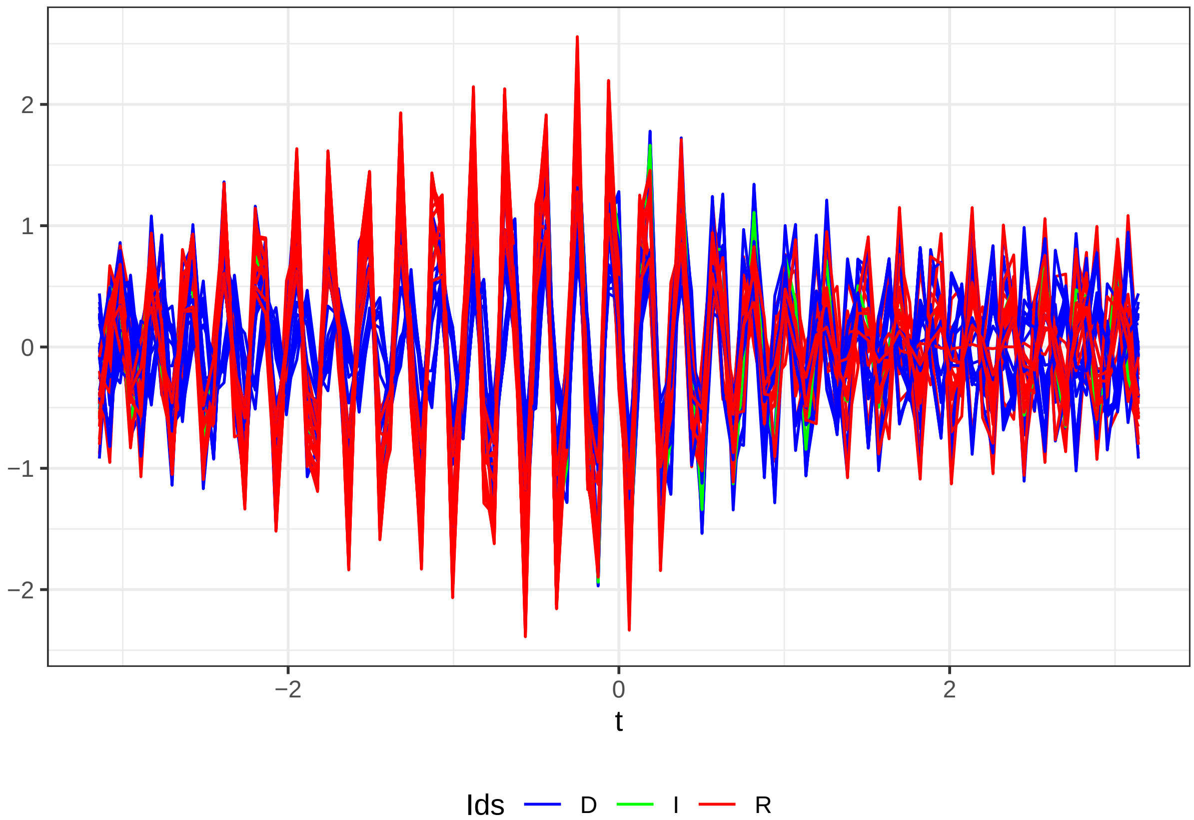
```
library(readxl)

senators <- read_excel("~/Desktop/stat_501/hw2/senate_voting_data.xls")[,3:102]
# Store in file
write.table(senators, '~/Desktop/senators_file.txt', append = FALSE, sep = " ", dec = ".",row.names = T
```

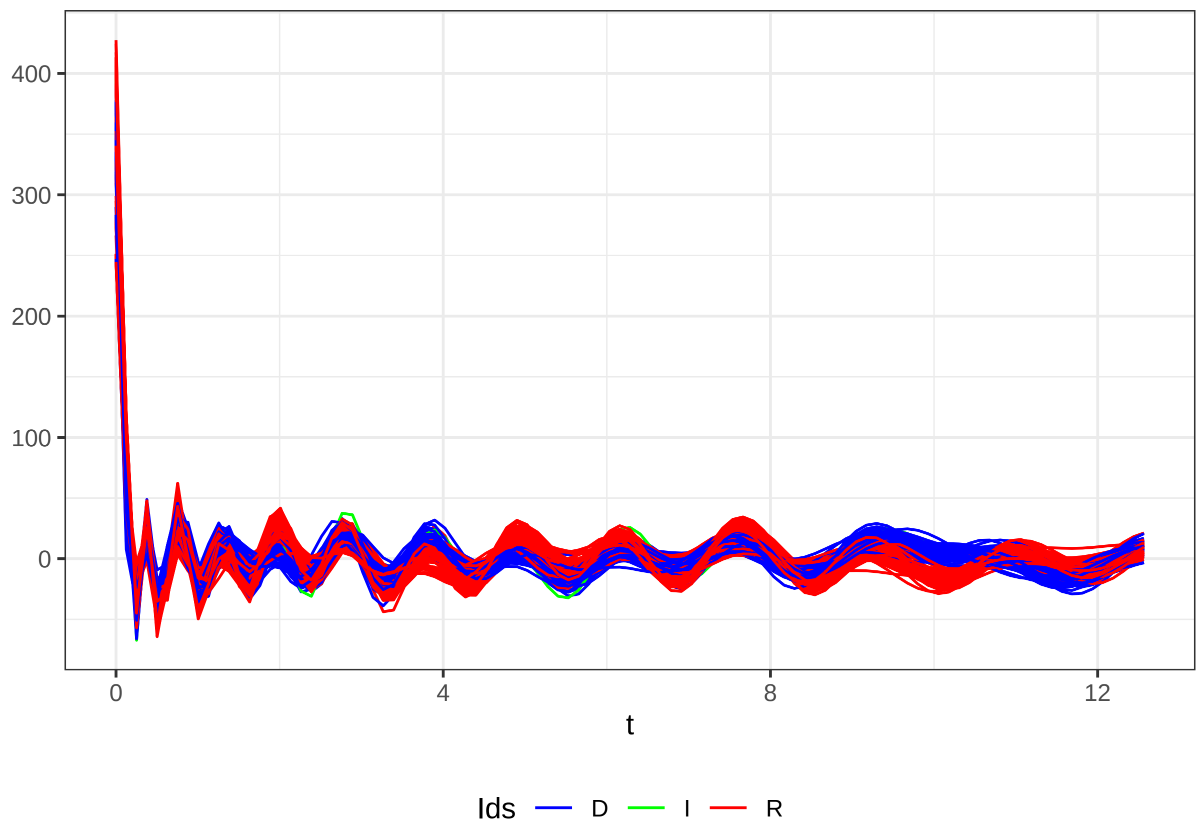
Part B: Using Andrews' curves, display the voting preferences of each senator. Use different colors for "Republican", "Democratic" and "Independent". Comment. [10 points]

Comment: Here I have created 2 Andrew's Curve plots that are colored based upon the political party of each senator. Each line represents a senator. The first plot is creating the plot using the type = 4 parameter and the second plot is using type = 3 parameter. The type parameters here is a different f(t).

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```





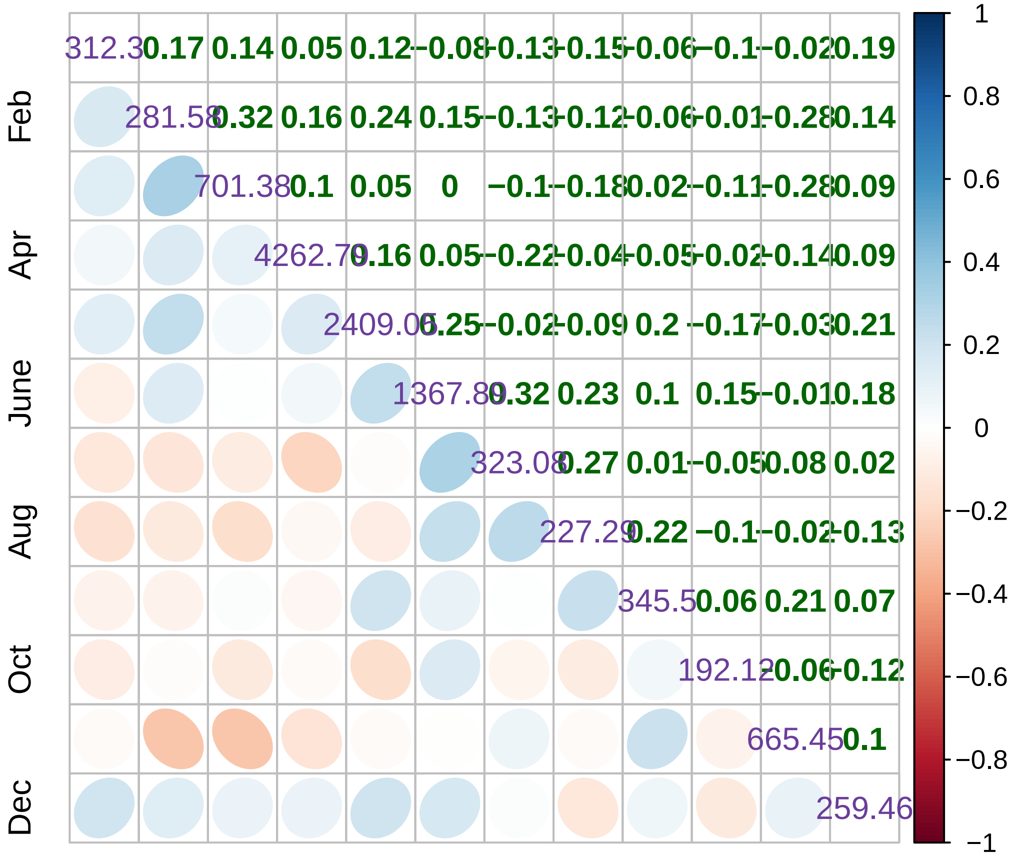
Comments on the first plot: This plot shows no big voting differences between the political parties. As time goes on it seems like the nice smooth trend becomes disrupted.

Question 2:

Part A:Formulate a correlation plot and discuss the general relationships,if any,between the month-to-month frequency of the number of tornadoes in the US. Are there any patterns? How about the variances? [10 points]

```
## Loading required package: corrplot
```

```
## corrplot 0.84 loaded
```

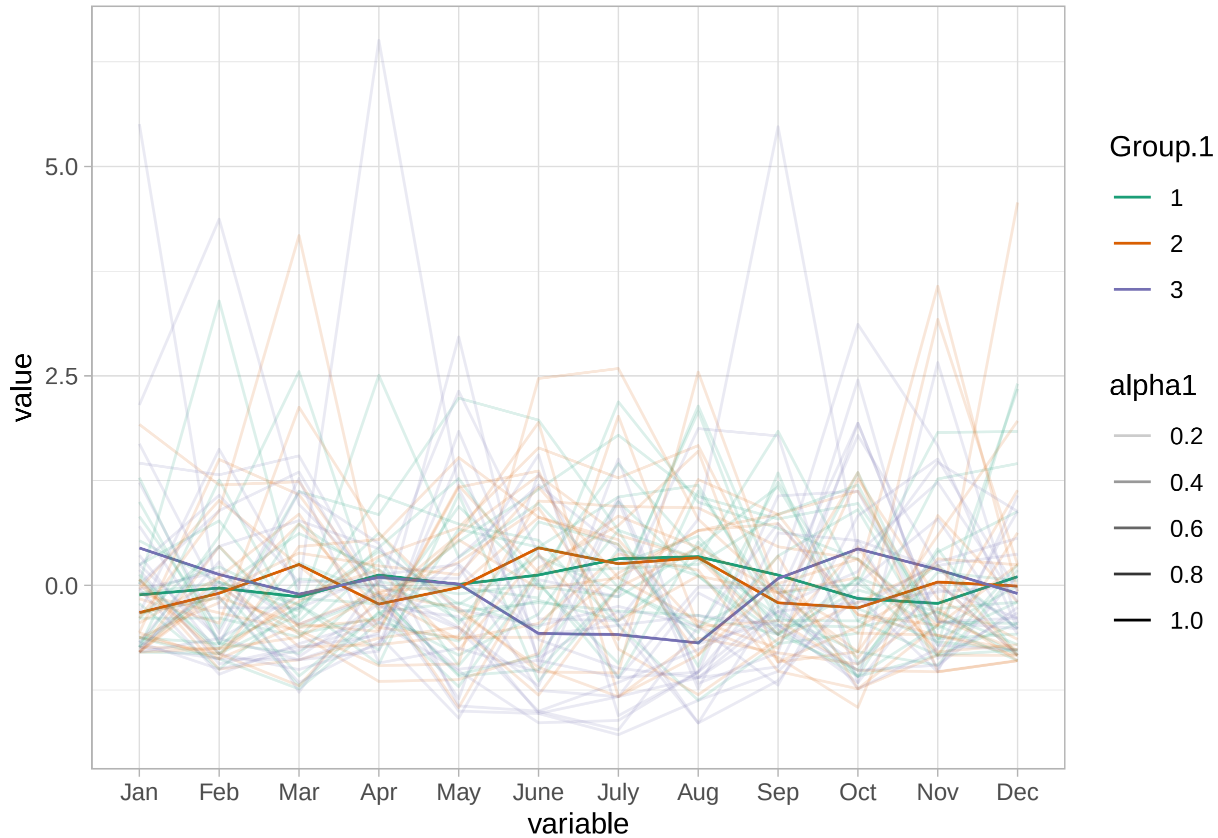


Are there any patterns? How about the variances? [10 points] To me it seems like the spring months (jan - april) have a negative correlation with the summer(june- sep) and even some fall months have a negative correlation. Similarly the variances increase during the summer months meaning that the amount of F1 tornados occuring during these months varies from year to year. The maximum variance happens in April with 4262.70 and the minimum happens in october with 192.1.

Part B: Divide the period into the three 20 year periods (1955-1974, 1975-1994, 1995-2014). Call these periods I, II and III respectively.

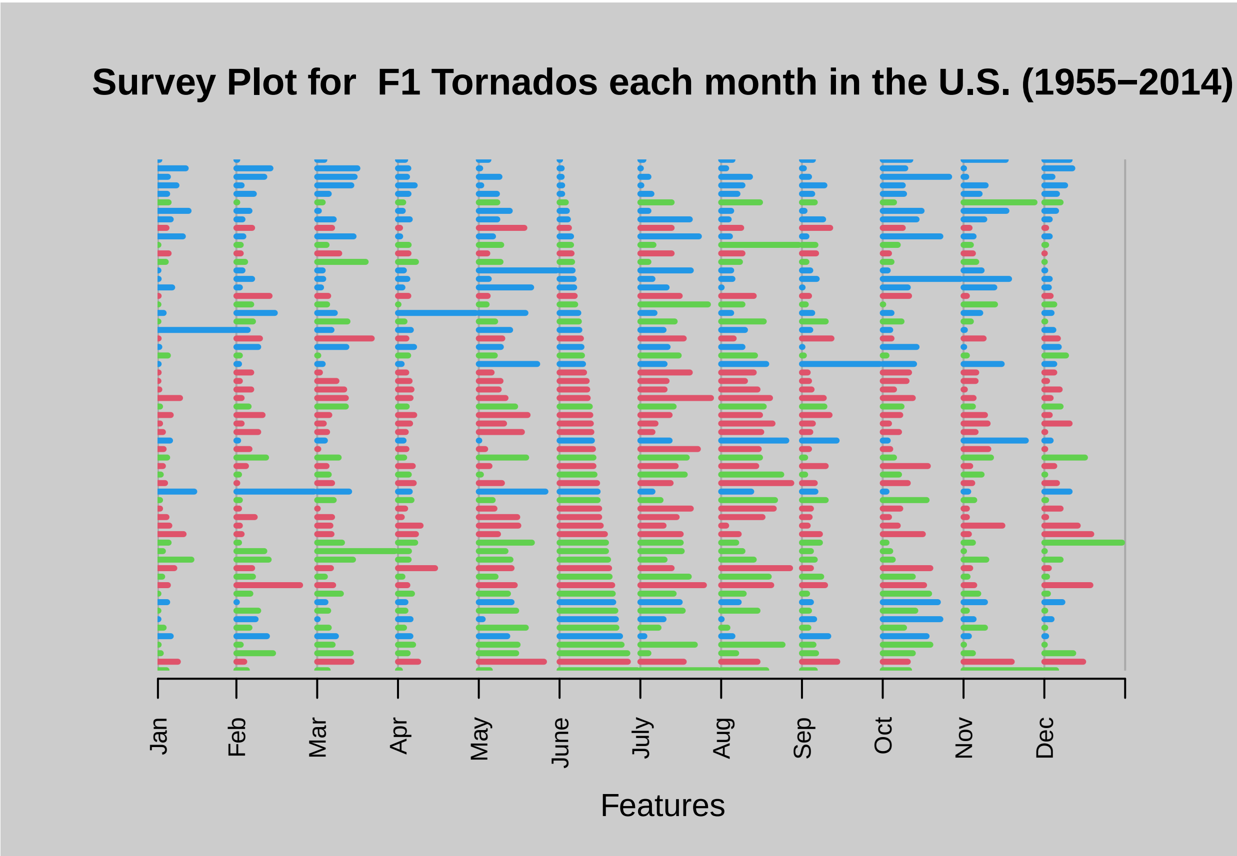
i.) For each period,construct a parallel-coordinates plot(use color for the different periods).Superimpose the mean for each group.

```
## Loading required package: GGally

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

## Loading required package: RColorBrewer
```



Are there any patterns that you can see? [10 points] There is not clear separation for most of the months of the year, however from May to Sep there is very clear separation between Period 1(1955-1974) & 2(1975-1994) and Period 3(1995-2014), which could suggest that the avg. number of F1 tornadoes happening from May to Sep has decreased over time.

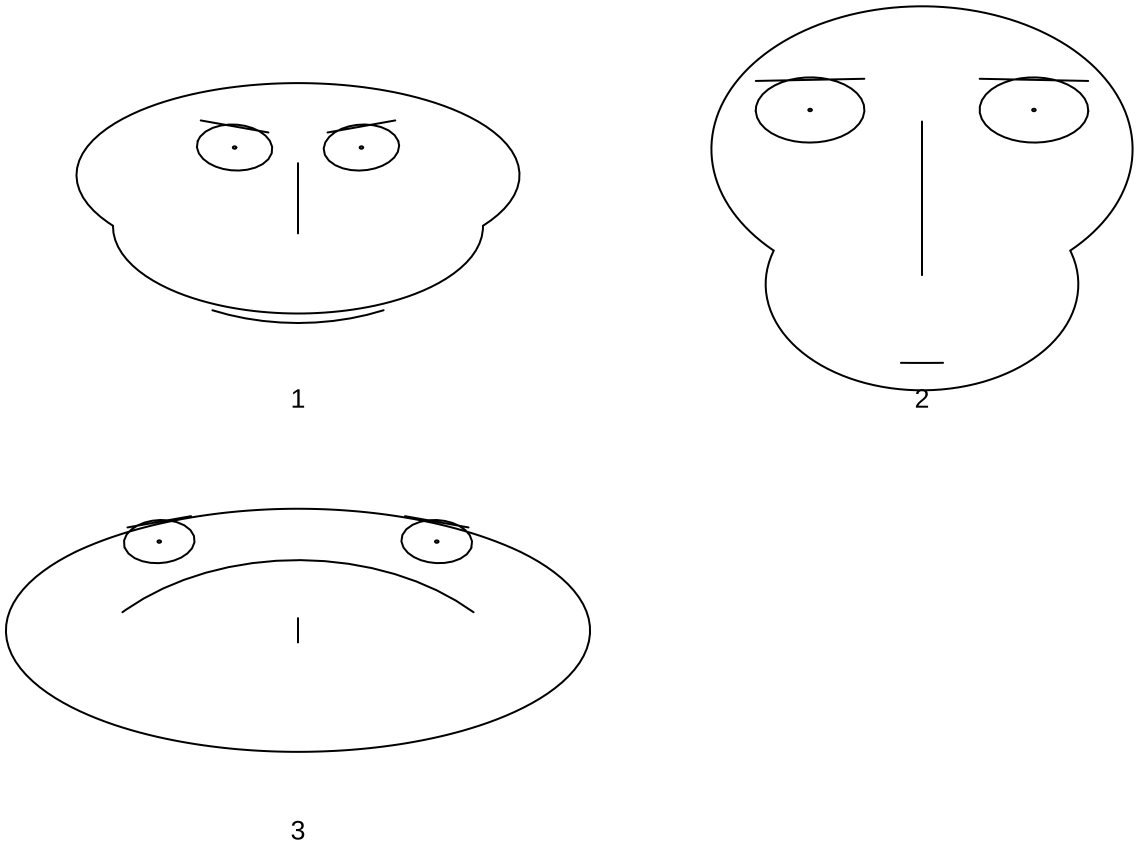
ii.) Display these observations using a survey plot.

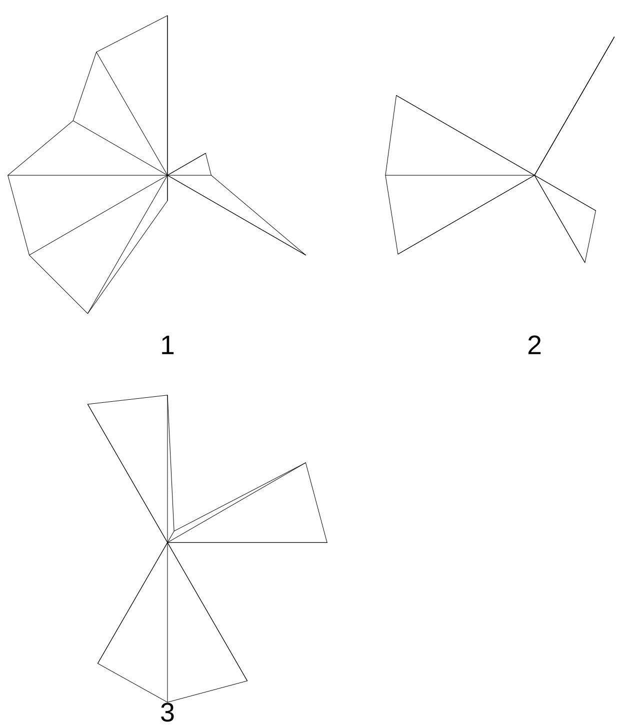


Are there any patterns that we can see? You may try out patterns with different months as the ordering variable and see which one provides the clearest separation. [10 points]

Ordered by the month of June! Generally as I looked at each month, the clearest separation was shown when I ordered by June. Here the blue period is mostly near the top of the graph, with red in the middle and finally green period.

iii.) Display the three group means using Chernoff faces and stars.





Are there any differences between these three means. [5 points] There are big differences between group 3 (1995-2014) and the other two groups based upon the cheroff faces. Based on the star plot, it looks like group 1 (1955-1974) has higher means for the most consecutive months out of all the groups. Group 2 (1975-1994) seems to have overall lower means.

Question 3A: Let $\underline{X} \sim N_p(\underline{\mu}, \underline{\Sigma})$

Show: $M_{\underline{X}}(\underline{t}) \equiv E[\exp(\underline{t}'\underline{X})] = \exp(\underline{t}'\underline{\mu} + \underline{t}'\underline{\Sigma}\,\underline{t}/2)$

Definitions/ Knowns:

- For a univariate $Z_i \sim N(0,1)$: $M_{z_i}(t_i) = E[e^{t_i z_i}]$ $\rightarrow E[x] = \int_a^b x f(x)$

$$= \int_{-\infty}^{\infty} e^{t_i z_i} \cdot \left(\frac{e^{(-z_i^2/2)}}{\sqrt{2\pi}} \right) dz_i$$

$$= e^{(t_i)^2/2} \cdot \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-(z_i - t_i)^2/2} dz_i}_{\text{evaluates to } 1}$$

$$= e^{(t_i)^2/2}$$

- When $\underline{Z}$ is a vector of random samples: $\underline{Z} = (z_1, \dots z_n)$
  $\underline{Z} \sim N(0,1)$

$$M_{\underline{z}}(t) = E[e^{\sum_{i=1}^n z_i t_i}] = E[\prod_{i=1}^n e^{z_i t_i}] = \prod_{i=1}^n E[e^{z_i t_i}] = \prod_{i=1}^n M_{z_i}(t_i) = \prod_{i=1}^n e^{t_i^2/2}$$

b/c z_i is a standardized normal we can assume ind.

$$= \exp\{\sum_{i=1}^n t_i^2/2\}$$
$$= \exp\{\underline{t}^2/2\} =$$
$$= \exp\{\underline{t}\,\underline{t}'/2\}$$

- $AA' = A'A = \Sigma^{-1}$

If we define $\underline{X} = A\underline{Z} + \underline{\mu}$ ← here $\underline{X}$ is a linear cobination of Z.

$M_{\underline{X}}(\underline{t}) = E[\exp\{\underline{X}\,\underline{t}'\}]$ ← sub for $\underline{X}$

$= E[\exp\{(A\underline{Z} + \underline{\mu})\,\underline{t}'\}] = E[\exp\{A\underline{Z}t' + \underline{\mu}t'\}]$ Nothing to do w/ Z so we can pull it out

$= \exp\{\underline{\mu} t'\} \cdot E[\exp\{\underline{A}\,\underline{Z} t'\}]$ ← $M_z(At')$

$= \exp\{\underline{\mu}\,\underline{t}'\} \cdot \exp\{(At)(At)'/2\}$

$= \exp\{\underline{\mu}\,\underline{t}'\} \cdot \exp\{(AA')(tt')/2\}$

$= \exp\{\underline{\mu}\,\underline{t}'\} \cdot \exp\{\Sigma^{-1} tt'/2\} = \exp\{\underline{\mu}'\underline{t}' + \frac{1}{2}\underline{t}\,\Sigma^{-1}\underline{t}'\}$ ✓

Question 3b: $\underline{X} \sim N_p(\mu, \underline{\Sigma})$

If we partition $\underline{X}$ ...

$$\underline{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \underline{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$\underline{X}_1 = (x_1 \dots x_q)$ $\qquad \underline{\mu}_1 = (\mu_1 \dots \mu_q)$

$\underline{X}_2 = (x_{q+1} \dots x_{p-q})$ $\qquad \mu_2 = (\mu_{q+1} \dots \mu_{p-q})$

Show that marginal distribution of $\underline{X}_1$ is $\underline{X}_1 \sim N_q(\mu_1, \Sigma_{11})$

If we can use similar logic to part A then we can prove this by writing $\underline{X}_1$ as a linear combination of $\underline{X}$ then we can show this true.

test: $\underline{x}_1 = B\underline{X}$

In the test, we will assume B is a vector that when multiplied by $\underline{X}$ will leave $\underline{X}_1$ remaining

ex: $\underline{X}_1 = \overset{B}{\begin{bmatrix} 1 & 0 \end{bmatrix}} \overset{X}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = 1 \cdot x_1 + 0 \cdot x_2 = x_1$

$1 \cdot 0$

$1 \times 2 \qquad 2 \times 2$

Σ'

Apply:

$\underline{X}_1 = B'\underline{X}$

$E[\underline{X}_1] = E[B'\underline{X}]$ → constant

$= B' E[\underline{X}]$ → this is defined above:

$= B' \underline{\mu}$ → B' is a constant that when multiplied w/ $\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ μ_1 remains.

$E[\underline{X}_1] = \mu_1$

Similarly for the $Var(\underline{X}_1)$

$Var(\underline{X}_1) = Var(B'\underline{X})$

$= B' Var(\underline{X})$ → Σ covariance matrix

$= B\underline{\Sigma}$

Question 3c: Show $x_1|x_2 \sim N_q(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

Show: $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}(x_2-\mu_2)$ $\quad var(x_1|x_2) = \Sigma_{11} - \Sigma_{12}\cdot\Sigma_{22}^{-1}\cdot\Sigma_{21}$

Knowns: $f(x_1|x_2) = \dfrac{f(x_1,x_2)}{f(x_2)}$

$$\Sigma^{-1} = \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}^{-1} = \begin{bmatrix}\Sigma^{11} & \Sigma^{12}\\ \Sigma^{21} & \Sigma^{22}\end{bmatrix}$$

$$\Sigma^{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\cdot\Sigma_{12}(\Sigma_{22} - A_{12}'\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12})^{-1}\Sigma_{12}'\,\Sigma_{11}^{-1}$$

$$\Sigma^{12} = (\Sigma^{21})' = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}$$

$$\Sigma^{22} = (\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1} =$$

$$= \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\cdot\Sigma_{12}'(\Sigma_{11}-\Sigma_{12}\cdot\Sigma_{22}^{-1}\cdot\Sigma_{12}')^{-1}\cdot\Sigma_{12}\cdot\Sigma_{22}^{-1}$$

$$f(x_1,x_2) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\cdot\exp\{-\tfrac{1}{2}\underbrace{(\underline{x}-\underline{\mu})'\cdot\Sigma^{-1}\cdot(\underline{x}-\underline{\mu})}\}$$

$$R(x_1,x_2) = (\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})$$

$$= ((x_1-\mu_1)', (x_2-\mu_2)')\begin{bmatrix}\Sigma^{11} & \Sigma^{12}\\ \Sigma^{21} & \Sigma^{22}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}$$

$$= (x_1-\mu_1)'\Sigma^{11}(x_1-\mu_1) + 2(x_1-\mu_1)'\Sigma^{12}(x_2-\mu_2) +$$
$$(x_2-\mu_2)'\Sigma^{22}(x_2-\mu_2)$$

Plug in for Σ^{-1}:

$$R(x_1,x_2) = (x_1-\mu_1)'[\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\cdot\Sigma_{12}(\Sigma_{22}-A_{12}'\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12})^{-1}\cdot\Sigma_{12}'\cdot\Sigma_{11}^{-1}](x_1-\mu_1)$$
$$-2(x_1-\mu_1)'[\Sigma_{11}^{-1}\cdot\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}]\cdot(x_2-\mu_2) +$$
$$+(x_2-\mu_2)'[(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{21})^{-1}]\cdot(x_2-\mu_2)$$

$R(x_1, x_2) = (x-\mu)^T \Sigma^{-1} (x-\mu)$

$A =$

$v = x_2 - \mu_2$

$R(x_1, x_2) =$

$$= \left((x_1-\mu_1)\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12} - (x_2-\mu_2)\right)^T (\Sigma_{22} \cdot A_{12}^T \Sigma_{11}^{-1}\Sigma_{12})^{-1} \cdot$$

$$\left((x_1-\mu_1)\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12} - (x_2-\mu_2)\right)$$

$(x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1)$

$$(x_1-\mu_1)\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12}^T(\Sigma_{22}\cdot A_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1} - (x_2-\mu_2)^T\cdot(\Sigma_{22}\cdot A_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}$$

$$\cdot(x_1-\mu_1)\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12} \qquad \cdot(x_2-\mu_2)$$

$$R(x_1,x_2) = \begin{cases} R_1(x_1,x_2) = (x_1-\mu_1)^T\Sigma_{11}^{-1}(x_1-\mu_1) + (x_1-\mu_1)\cdot\Sigma_{11}^{-1}\cdot\Sigma_{12}^T(\Sigma_{22}\cdot A_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\ \qquad (x_1-\mu_1)\Sigma_{11}^{-1}\Sigma_{12} \\ R_2(x_2) = -(x_2-\mu_2)^T(\Sigma_{22}-A_{12}^T\Sigma_{11}^{-1}\Sigma_{12})^{-1}\cdot x_2-\mu_2 \end{cases}$$

$$R(x_1, x_2) = R_1(x_1, x_2) + R_2(x_2)$$

$$f(x_1, x_2) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\{-\frac{1}{2}(R(x_1, x_2))$$

$$= \frac{1}{(2\pi)^{p/2} |\Sigma_{22}|^{1/2} \cdot |\Sigma_{11} - \Sigma_{12}^T \Sigma_{22}^{-1} \Sigma_{12}|^{1/2}} \exp\{-\frac{1}{2}(R(x_1, x_2))\}$$

$$= \frac{1}{(2\pi)^{p/2} \cdot |\Sigma_{11} - \Sigma_{12}^T \Sigma_{22}^{-1} \Sigma_{12}|^{1/2}} \cdot \exp\{-1/2\Big((x_1-\mu_1)^T \Sigma_{11}^{-1}(x_1-\mu_1) + (x_1-\mu_1)^T \cdot \Sigma_{11}^{-1} \Sigma_{12}^T (\Sigma_{22} - A_{12}^T \Sigma_{11}^{-1} \Sigma_{12}) \cdot (x_1-\mu_1) \Sigma^{-1} \Sigma_{12}$$

$$\cdot \frac{1}{(2\pi)^{q/2} \cdot |\Sigma_{22}|^{1/2}} \exp\{\frac{1}{2}\big(-(x_2-\mu_2)^T (\Sigma_{22} - A_{12}^T \Sigma_{11}^{-1} \Sigma_{12})^{-1} \cdot (x_2-\mu_2)\}$$

Define:

$$f_2(x_2) = \int f(x_1, x_2)\, dx_1 = \frac{1}{(2\pi)^{p/2} |\Sigma_{22}|^{1/2}} \exp\{-\frac{1}{2}(x_2-\mu_2) \cdot (\Sigma_{11} - A_{12}^T \Sigma_{22}^{-1} \Sigma_{12})^{-1} (x_2-\mu_2)\}$$

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{\exp\{-\frac{1}{2}((x_1-\mu_1)^T (\Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12}^T (\Sigma_{22} - A_{12}^T \Sigma_{11}^{-1} \Sigma_{12}) \cdot (x_1-\mu_1)}{(2\pi)^{p/2} \cdot |\Sigma_{11} - \Sigma_{12}^T \Sigma_{22}^{-1} \Sigma_{12}|^{1/2}}$$

I am stopping because I'm not entirely sure I have done the math right to this point. ~~Also, I a~~ So I don't want to derive $E[x_1|x_2]$ and $Var(x_1|x_2)$ if my $f(x_1|x_2)$ isn't correct.

I struggled with this problem. I referenced website (see below); but they derived for $x_2|x_1$. I tried to go back & figure it out, but I am not sure where I went wrong

Question 4. Let $\underline{Z} \sim N_p(0, I_p)$

$\underline{\Gamma}$: orthogonal matrix

$\underline{\Gamma}\underline{\Gamma}' = I_p$

Show that $\underline{W} = \underline{\Gamma}\underline{Z} \sim N_p(0, I_p)$:

Here $\underline{W} = \underline{\Gamma}\underline{Z}$ is a linear combination of Z_i

I attempted 2 solutions, but I'm not sure Solution #2 is valid.

Solution #1: Similar to 3A.

Before we had $X = AZ + \mu$ so

$$M_X(t) = E[\exp\{Xt'\}] = E[\exp\{AZ+\mu\}] = \exp\{\mu t' + \tfrac{1}{2} t \Sigma^{-1} t'\}$$

But in this case we have $\underline{W} = \Gamma Z$

$$M_W(t) = E[\exp\{Wt'\}] = E[\exp\{\underline{\Gamma}\underline{Z}t'\}] = M_Z(\Gamma t')$$

$$\cancel{E[\exp} = \exp\{(\underline{\Gamma}\underline{t})^2/2\} = \exp\{(\underline{\Gamma}t)(\underline{\Gamma}t)/2\}$$

$$= \exp\{(\underline{\Gamma}t)'(\underline{\Gamma}t)/2\}$$

$$= \exp\{(\underline{\Gamma}\underline{\Gamma}')(tt')/2\}$$

There is no μ term thus the mean is 0.

$$= \exp\{I_p(tt')/2\}$$

usually Σ^{-1}

$I_p^{-1} = I_p$

Solution #2:

$Z \sim N_p(0, I_p)$

$\underline{\Gamma}\underline{Z} \sim N_p(\Gamma \cdot 0, \Gamma I_p \Gamma')$ → similar to $AX \sim N(A\mu, A\Sigma A')$

$N_p(0, I_p \cdot I_p)$

$I_p \cdot I_p$ ex: $I_p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{pmatrix} 1\cdot 1 + 0\cdot 0 & 0\cdot 0 + 0\cdot 1 \\ 0\cdot 1 + 0\cdot 0 & 1\cdot 0 + 1\cdot 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_p$$

$\Gamma Z \sim N_p(0, I_p)$

↳ I'm not sure if I've done enough to prove this formally though.

Question 5:

Show that the identity matrix I_p is the one among all $p \times p$ dimensional positive definite matrices B that maximizes

$$f(B) = |nB|^{n/2} \exp\{-\frac{n}{2} \text{trace}(B)\}$$

How do we maximize a function? Find the MLE

Log likelihood:

$$\ln(f(B)) = \frac{n}{2} \ln(|nB|) + -\frac{n}{2} \text{trace}(B)$$

estimate $\hat{B}$

$$\frac{d}{d\Sigma} \ln|\Sigma| = \Sigma^{-1}$$

$$\frac{d}{d\Sigma} (\text{trace}(a\Sigma) \to a'$$

$a = I_p$ b/c $I_p \cdot B = B$

$$\frac{d}{dB}(\ln(f(B))) = \frac{n}{2} \cdot n \cdot (nB)^{-1} + -\frac{n}{2} \cdot I_p$$

Set equal to 0 & Solve for $\hat{B}$

$$0 = \frac{n^2}{2}(n\hat{B})^{-1} + \frac{-n}{2} \cdot I_p$$

$$0 = \frac{n}{2}(\cancel{n}(\cancel{n}\hat{B})^{-1} - I_p)$$

$$0 = n(n\hat{B})^{-1} - I_p$$

$$I_p = \cancel{\frac{n}{n}}\hat{B} \quad n \cdot n^{-1} = 1$$

Question 6: Let $\underline{X} \sim N_p(\underline{\mu}, \underline{\Sigma})$

Find the MLE of $\underline{\Sigma}$ & $\underline{\mu}$, given a sample $(\underline{X}_1 \dots \underline{X}_n) \sim N_p(\underline{\mu}, \underline{\Sigma})$

- use the method of derivatives of symmetric positive definite matrices & vectors

① Define likelihood:

$\theta = \{\underline{\mu}, \underline{\Sigma}\}$

$$L(\theta \mid \underline{X}_1, \dots \underline{X}_n) = f(\underline{X}_1 \dots \underline{X}_n \mid \theta) = \prod_{i=1}^{n} f(\underline{X}_i \mid \theta) = \prod_{i=1} \frac{1}{(2\pi)^{p/2} \cdot |\Sigma|^{1/2}} \exp\{-\tfrac{1}{2}(x_i-\mu)' \Sigma^{-1} (x_i-\mu)\}$$

$$= \frac{1}{2\pi^{(np/2)} \cdot |\Sigma|^{n/2}} \prod_{i=1}^{n} \exp\{-\tfrac{1}{2}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\}$$

$$= \frac{1}{2\pi^{(np/2)} \cdot |\Sigma|^{n/2}} \cdot \exp\{-\tfrac{1}{2}\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)'(x_i-\mu)\}$$

② Take the natural log ($\ln$):

$$\ln(L(\theta \mid \underline{X}_1, \dots \underline{X}_n)) = \ln(\ )$$

$$= -n/2(p \ln(2\pi) + \ln(|\Sigma|) + -\tfrac{1}{2} \cdot \Sigma^{-1} \cdot \sum_{i=1}^{n}(x_i-\mu)'(x-\mu)$$

③ Estimate Parameters:

$$\hat{\underline{\mu}} = \frac{d}{d\mu}\left[\ln(L(\theta \mid \underline{X}_1, \dots \underline{X}_n))\right] = \frac{d}{d\mu}\Big(\underbrace{-\tfrac{n}{2}(p \ln(2\pi) + \ln|\Sigma|)}_{\text{no } \mu = \text{constant}} + \frac{d}{d\mu}\left(-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)\Big)$$

$$= 0 + -\tfrac{1}{2} \cdot \Sigma^{-1} \sum_{i=1}^{n} \frac{d}{d\mu} \underbrace{(x_i-\mu) \cdot (x_i-\mu)'}$$

$$\frac{d}{d\mu}\begin{bmatrix} x_1^2-\mu \\ \vdots \\ x_n^2-\mu \end{bmatrix} \rightarrow \begin{bmatrix} 2x_1-\mu \\ \vdots \\ 2x_n-\mu \end{bmatrix}$$

$$= -\cancel{\tfrac{1}{2}} \cdot \Sigma^{-1} \cdot -\cancel{2} \sum_{i=1}^{n}(x_i-\mu)$$

$$= \Sigma^{-1} \cdot \sum_{i=1}^{n} x_i - \hat{\mu} = 0 \leftarrow \text{Solve for } \hat{\mu}$$

$$\sum_{i=1}^{n} x_i - n\hat{\mu} = 0$$

$$\sum_{i=1}^{n} x_i = n\hat{\mu}$$

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\mu}}$$

$$\hat{\Sigma} = \frac{d}{d\Sigma}\left[\ln(L(\theta \mid \bar{x}_1, \dots \bar{x}_n))\right] = \frac{d}{d\Sigma}\left(\frac{n}{2}(p\ln(2\pi) + \ln|\Sigma|\right) + \frac{d}{d\Sigma}\left(-\frac{1}{2}\cdot\Sigma^{-1}\cdot\sum_{i=1}^{n}(x_i-\mu)'(x_i-\mu)\right)$$

$$= \frac{d}{d\Sigma}\left(-\frac{n}{2}p\ln(2\pi)\right) - \frac{d}{d\Sigma}\left(+\frac{n}{2}\cdot\ln|\Sigma|\right) + -\frac{1}{2}\cdot\sum_{i=1}^{n}(x_i-\mu)'(x_i-\mu)\frac{d}{d\Sigma}(\Sigma^{-1})$$

$$= 0 + \frac{n}{2}\Sigma^{-1} + -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{d}{d\Sigma}((X_i-\mu)(X_i-\mu)'\cdot\Sigma^{-1})\right)$$

$$= -\frac{n}{2}\Sigma^{-1} + -\frac{1}{2}\cdot\sum_{i=1}^{n}\cdot -\Sigma^{-1}\cdot(X_i-\mu)\cdot(X_i-\mu)'\cdot\Sigma^{-1}$$

$$= -\frac{n}{2}\Sigma^{-1} + +\frac{1}{2}\cdot +\Sigma^{-1}\cdot\Sigma^{-1}\cdot\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'$$

$$= \frac{1}{2}\Sigma^{-1}\left(-n + \Sigma^{-1}\cdot\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'\right)$$

Set equal to 0 & Solve for $\hat{\Sigma}$:

$$0 = \frac{1}{2}\hat{\Sigma}^{-1}\left(-n + \hat{\Sigma}^{-1}\cdot\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'\right)$$

$$0 = -n + \hat{\Sigma}^{-1}\cdot\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'$$

$$\Sigma\cdot n = \hat{\Sigma}^{-1}\cdot\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'\cdot\Sigma^{-1}$$

$$\hat{\Sigma}n = \sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'$$

$$\boxed{\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)'} \rightarrow$$

Sample variance is usually defined as:

$$S_k^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ik} - \bar{X}_k)^2$$

but sometimes the denominator is an n instead of $n-1$ which can happen if the mean is known. In this case we are fixing the mean (it is known) in order to solve for Σ.