Final project 2019

The attached `hurrican356.csv` included the track data of 356 hurricanes in the North Atlantic area since 1989. For all the storms, their location (longitude & latitude) and maximum wind speed were recorded every 6 hours. The data includes the following variables

1. **ID**: ID of the hurricans
2. **Season**: In which **year** the hurricane occurred
3. **Month**: In which **month** the hurricane occurred
4. **Nature**: Nature of the hurricane

- ET: Extra Tropical
- DS: Disturbance
- NR: Not Rated
- SS: Sub Tropical
- TS: Tropical Storm

5. **time**: dates and time of the record

6. **Latitude** and **Longitude**: The location of a hurricane check point
7. **Wind.kt** Maximum wind speed (in Knot) at each check point

Researchers want to develop a model to predict the hurricane trajectory.

Load and plot the hurrican data

```
library(ggplot2)
dt= read.csv("/Users/yw2148/Dropbox/Teaching/Teaching-computing/My Teaching Files/big data computing/1_
ggplot(data=dt, aes(x = Longitude, y = Latitude)) +
  stat_summary_2d(data = dt, aes(x = Longitude, y = Latitude, z = dt$Wind.kt), fun = median, binwidth =
library(data.table)
dt <- as.data.table(dt)
summary(dt)
```

Overlay the hurrican data in the world map

```
library(maps)
map <- ggplot(data = dt, aes(x = Longitude, y = Latitude)) +
  geom_polygon(data = map_data(map = 'world'), aes(x = long, y = lat, group = group))
map +
  stat_summary_2d(data = dt, aes(x = Longitude, y = Latitude, z = dt$Wind.kt), fun = median, binwidth =
  ggtitle(paste0("Atlantic Windstorm mean knot"))
```

Additional Plots

```
map <- ggplot(dt, aes(x = Longitude, y = Latitude, group = ID)) +
  geom_polygon(data = map_data("world"),
               aes(x = long, y = lat, group = group),
               fill = "gray25", colour = "gray10", size = 0.2) +
  geom_path(data = dt, aes(group = ID, colour = Wind.kt), size = 0.5) +
  xlim(-138, -20) + ylim(3, 55) +
  labs(x = "", y = "", colour = "Wind \n(knots)") +
  theme(panel.background = element_rect(fill = "gray10", colour = "gray30"),
        axis.text.x = element_blank(), axis.text.y = element_blank(),
```

```
        axis.ticks = element_blank(), panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())

seasonrange <- paste(range(dt[, Season]), collapse=" - ")

map + ggtitle(paste("Atlantic named Windstorm Trajectories (",
                     seasonrange, ")\n"))
```

Show hurricance tracks by month

```
mapMonth <- map + facet_wrap(~ Month) +
  ggtitle(paste("Atlantic named Windstorm Trajectories by Month (",
                 seasonrange, ")\n"))
mapMonth
```

Problem 1 (60 points)

Let t be time (in hours) since a hurricane began, and For each hurrican i, we denote $\{Y_{i,1}(t), Y_{i,2}(t), Y_{i,3}(t)\}, j = 1,2,3$ be the latitude, longitude, and wind speed at time t. We consider the following model

$$Y_{i,j}(t+6) = \mu_{i,j}(t) + \rho_j Y_{i,j}(t) + \epsilon_{i,j}(t)$$

where $\mu_{i,j}(t)$ is the funtional mean, and the errors $(\epsilon_{i,1}(t), \epsilon_{i,2}(t), \epsilon_{i,3}(t))$ follows a multivariate normal distributions with mean zero and covariance matrix Σ, independent across t. We further assume that the mean functions $\mu_{i,j}(t)$ can be written as

$$\mu_{i,j}(t) = \beta_{0,j} + x_{i,1}(t)\beta_{1,j} + +x_{i,2}\beta_{2,j} + x_{i,3}\beta_{3,j} + \sum_{k=1}^{3}\beta_{3+k,j}\Delta_{i,k}(t-6)$$

where $x_{i,1}(t)$, ranging from 0 to 365, is the day of year at time t, $x_{i,2}$ is the calenda year of the hurrican, and $x_{i,3}$ is the type of hurrican, and

$$\Delta_{i,k}(t-6) = Y_{i,k}(t) - Y_{i,k}(t-6), k = 1,2,3$$

are the change of latitude, longitude, and wind speed between $t-6$ and t.

Prior distribution

We assume the prior distribution of $\boldsymbol{\beta} = (\beta_{k,j})_{k=0,\ldots,6,j=1,2,3}$, $\pi(\boldsymbol{\beta})$ is jointly normal with mean 0 and variance $diag(1,p)$.

$\pi(\rho_j)$ follows a trucated normal $N_{[0,1]}(0.5, 1/5)$

$\pi(\Sigma^{-1})$ follows a $Wishart(3, diag(0.1, 3))$

Your to-do-list:

1. Randomly select 80% hurricanes and develop an MCMC algorithm to estiamte the posterior mean of the model parameters.
2. Apply your model to track the remaining 20% hurricans, and evaluate how well your model could predict and track these hurricanes.
3. Write a summary to report your findings.

Problem 2 (40 points)

The attached `hurricanoutcome2.csv` recorded the damages and death caused by 46 hurricanes in the U.S. The variables include

1. **ID**: ID of the hurricans
2. **Season**: In which **year** the hurricane occurred
3. **Month**: In which **month** the hurricane occurred
4. **Nature**: Nature of the hurricane

- ET: Extra Tropical
- DS: Disturbance
- NR: Not Rated
- SS: Sub Tropical
- TS: Tropical Storm

5. **Damage**: Financial loss (in Billion U.S. dollars) caused by hurricanes
6. **Deaths**: Number of death caused by hurricanes
7. **Maxspeed**: Maximum recorded wind speed of the hurricane
8. **Meanspeed**: average wind speed of the hurricane
9. **Maxpressure**: Maximum recorded central pressure of the hurricane
10. **Meanpressure**: average central pressure of the hurricane
11. **Hours**: Duration of the hurricane in hours
12. **Total.Pop**: Total affected population

13. **Percent.Poor**: % affected population that reside in poor countres (i.e. GDP per Capita <= 10,000)

14. **Percent.USA**: % affected population that reside in the United States

Please propose a model to investigate which characteristics of the hurricanes are associated with damage and deaths, and how well they could predict the hurricane induced damage and deaths. Propose estimation method/algorithm, interpret your estimated models, and evaluate the prediction power of your estimated models.