CSSS 510: Lab 1

Logistics & R Refresher

2017-9-29

Logistics

- **Lab Sessions**: Fri, 3:30-5:20pm in Smith 105
  - Emphasis on application of material from lecture using examples; clarification and extention of lecture material; Q & A for homeworks and lectures
  - Materials will be available on the **course website** and my **Github** on Wednesday evening

- **Office Hours**: Tues and Thurs, 3:30-4:20pm in Smith 220
  - Available for trouble shooting and questions about homework and lecture materials

- **Homeworks**: 5-6 homework assignments due every 2 weeks or so
  - Should be done using R or R Studio with write up in LaTeX
  - Using R Studio with R Markdown is the simplest way to do this (*Please* do not handwrite your homeworks or do them in MS Word)
  - We will use two of Chris's packages extensively: simcf and tile

- When this course is over, you should be able to do the following (and more) using R:
  - Fit a logistic regression model using both glm and "by hand" using optim, extract parameters of interest, and interpreted these as probabilities
  - Compute predicted probabilities for counterfactuals values of $\boldsymbol{x}$ and use simulation to find the expected values and confidence intervals of $\hat{\pi}$ across these counterfactuals
  - Use cross-validation to assess the predictive accuracy of several models and also compare these models across a variety of in-sample goodness of fit tests
  - Fit a variety of bounded and unbounded count models that address for overdispersion
  - Use one of several algorithms to impute missing data

- The course moves fast: you should at least be comfortable doing the following for the homework assignments and project
  - data wrangling (tidying and transforming data)
  - importing and exporting data sets
  - generating plots of your data and results
  - writing basic functions and loops for repeated procedures

- Fortunately, for those of you new to R, there are many resources to get you up to speed
  - Zuur et al. (2009), Chapter 1-5
  - Wickham and Groleman (2017)

R Refresher

Vectors

Create the following vectors

1. vector.1 : 1,2,3,4,5,6,6,6,6,6
2. vector.2: 10 randomly drawn numbers from a normal distribution with a mean 10 and a s.d. of 1
3. vector.3: Results of 10 single binomial trials with a probability of 0.4
4. vector.4: For 100 binomial observations with 5 trials for each observation with a probability of 0.4

```
#Clear memory
rm(list=ls())

vector.1 <- c(seq(1,5,1), rep(6,5))

vector.2 <- rnorm(10, 10, 1)

#help?
?rnorm

vector.3 <- rbinom(10, 1, 0.4)

vector.4 <- rbinom(100, 5, 0.4)
```

5. Check what type of data vector.2 is
6. Round up vector.2 to two decimal place

```
is.character(vector.2)
```

```
## [1] FALSE
```

```
mode(vector.2)
```

```
## [1] "numeric"
```

```
round(vector.2, 2)
```

```
##  [1]  9.19 11.36  9.41  9.42 11.11  9.68  8.09  9.17 12.73 12.04
```

Matrices

7. matrix.1: Create 5 by 5 matrix containing all NAs
8. Assign matrix.1 the row names (a,b,c,d,e) and the column names (1,2,3,4,5)
9. Replace the NAs in the first columne of matrix.1 with Inf

```
matrix.1<-matrix(NA, nrow=5, ncol=5)

rownames(matrix.1)<-c("a","b","c","d","e")
colnames(matrix.1)<-c(1,2,3,4,5)
```

```
matrix.1[,1]<-Inf
```

Lists

10. Create a list that contains vector.1, vector.2, and matrix.1
11. Locate vector.2 from the list

```
list.1<-list(vector.1, vector.2, vector.3, matrix.1)
names(list.1)<-c("vector.1", "vector.2", "vector.3", "matrix.1")

list.1[[2]]
```

```
##  [1]  9.187539 11.360581  9.408068  9.421426 11.110139  9.684095  8.090118
##  [8]  9.174738 12.734239 12.040247
```

```
list.1$vector.2
```

```
##  [1]  9.187539 11.360581  9.408068  9.421426 11.110139  9.684095  8.090118
##  [8]  9.174738 12.734239 12.040247
```

Data frames

Data frames are a special type of list in which each row has same length. It is also a matrix like object, yet its elements - unlike elements in a matrix - doesn't have to be of same type. Most of the data we use are in data frames.

12. Open Lab1data.csv in R
13. Is it a data frame? Is it a matrix?
14. Check the names and summary statistics of the data
15. Remove observations with missing values
16. Plot GDP per capita (on the x-axis) and polity2 (on the y-axis)
17. Create a new variable called "democracy". Assign 0 to countries with negative value or zero polity2 score, and assign 1 to countries with positive score.
18. Use a loop to do the same recoding

```
library(foreign)

setwd("/Users/danielyoo/CSSS-POLS-510-MLE/Lab1Notes")

data<-read.csv("Lab1data.csv", header=T)


is.data.frame(data) #Yes!
```

```
## [1] TRUE
```

```
is.matrix(data) #No
```

```
## [1] FALSE
```

```r
is.character(data$Year)
```

```
## [1] FALSE
```

```r
data$Year<-as.character(data$Year)

names(data)
```

```
## [1] "country"
## [2] "Year"
## [3] "GDP.per.capita.PPP.current.international"
## [4] "polity2"
```

```r
summary(data)
```

```
##                 country         Year
##  Afghanistan        :  11   Length:1914
##  Albania            :  11   Class :character
##  Algeria            :  11   Mode  :character
##  Andorra            :  11
##  Angola             :  11
##  Antigua and Barbuda:  11
##  (Other)            :1848
##  GDP.per.capita.PPP.current.international    polity2
##  Min.   :  219.2                          Min.   :-10.000
##  1st Qu.: 1625.0                          1st Qu.: -4.000
##  Median : 4299.2                          Median :  5.000
##  Mean   : 7874.9                          Mean   :  2.431
##  3rd Qu.: 9818.6                          3rd Qu.:  8.000
##  Max.   :91712.3                          Max.   : 10.000
##  NA's   :373                              NA's   :542
```

```r
unique(data$country) # observations on 174 countries
```

```
##   [1] Antigua and Barbuda           Afghanistan
##   [3] Albania                       Algeria
##   [5] Andorra                       Angola
##   [7] Argentina                     Armenia
##   [9] Aruba                         Azerbaijan
##  [11] Bahrain                       Barbados
##  [13] Benin                         Burkina Faso
##  [15] Bahamas, The                  Bhutan
##  [17] Belarus                       Belize
##  [19] Bangladesh                    Bolivia
##  [21] Bosnia and Herzegovina        Botswana
##  [23] Brazil                        Brunei Darussalam
##  [25] Burundi                       Bulgaria
##  [27] Cambodia                      Cameroon
##  [29] Cape Verde                    Cote d'Ivoire
##  [31] Central African Republic      Chad
##  [33] Chile                         China
##  [35] Colombia                      Comoros
##  [37] Congo, Rep.                   Costa Rica
##  [39] Croatia                       Cuba
##  [41] Cyprus                        Czech Republic
##  [43] Djibouti                      Dominica
```

```
##  [45] Dominican Republic            Congo, Dem. Rep.
##  [47] Vietnam                       Ecuador
##  [49] Egypt, Arab Rep.              Equatorial Guinea
##  [51] Eritrea                       Estonia
##  [53] Ethiopia                      Timor-Leste
##  [55] Fiji                          Micronesia, Fed. Sts.
##  [57] Gabon                         Gambia, The
##  [59] Ghana                         Guinea-Bissau
##  [61] Georgia                       Grenada
##  [63] Guatemala                     Guinea
##  [65] Guyana                        Haiti
##  [67] Hongkong                      Honduras
##  [69] Hungary                       India
##  [71] Indonesia                     Iran, Islamic Rep.
##  [73] Iraq                          Israel
##  [75] Jamaica                       Jordan
##  [77] Kenya                         Kiribati
##  [79] Kosovo                        Kuwait
##  [81] Kyrgyz Republic               Kazakhstan
##  [83] Lao PDR                       Latvia
##  [85] Liberia                       Lebanon
##  [87] Lesotho                       Libya
##  [89] Liechtenstein                 Lithuania
##  [91] Mauritania                    Macedonia, FYR
##  [93] Maldives                      Madagascar
##  [95] Malaysia                      Mauritius
##  [97] Malawi                        Mayotte
##  [99] Mexico                        Moldova
## [101] Mali                          Malta
## [103] Monaco                        Montenegro
## [105] Mongolia                      Morocco
## [107] Marshall Islands              Myanmar
## [109] Mozambique                    Namibia
## [111] Nepal                         Nicaragua
## [113] Nigeria                       Niger
## [115] Netherlands Antilles          Oman
## [117] Pakistan                      Palau
## [119] Panama                        Paraguay
## [121] Peru                          Philippines
## [123] Palestinian Adm. Areas        Papua New Guinea
## [125] Poland                        Korea, Dem. Rep.
## [127] Qatar                         Korea, Rep.
## [129] Romania                       Russian Federation
## [131] Rwanda                        South Africa
## [133] El Salvador                   Saudi Arabia
## [135] Senegal                       Seychelles
## [137] Sierra Leone                  Singapore
## [139] St. Kitts and Nevis           Slovak Republic
## [141] St. Lucia                     Slovenia
## [143] San Marino                    Solomon Islands
## [145] Somalia                       Sri Lanka
## [147] Sao Tome and Principe         Sudan
## [149] Suriname                      St. Vincent and the Grenadines
## [151] Swaziland                     Syrian Arab Republic
```

```
## [153] Tajikistan                    Tanzania
## [155] Thailand                      Turkmenistan
## [157] Togo                          Tonga
## [159] Trinidad and Tobago           Tunisia
## [161] Turkey                        Tuvalu
## [163] United Arab Emirates          Uganda
## [165] Ukraine                       Uruguay
## [167] Uzbekistan                    Vanuatu
## [169] Venezuela, RB                 Samoa
## [171] Yemen, Rep.                   Serbia
## [173] Zambia                        Zimbabwe
## 174 Levels: Afghanistan Albania Algeria Andorra ... Zimbabwe
```

```
tapply(data$country, data$Year, length)
```

```
## 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  174  174  174  174  174  174  174  174  174  174  174
```

```
tapply(data$Year, data$country, length)
```

```
##                    Afghanistan                        Albania
##                             11                             11
##                        Algeria                        Andorra
##                             11                             11
##                         Angola             Antigua and Barbuda
##                             11                             11
##                      Argentina                        Armenia
##                             11                             11
##                          Aruba                     Azerbaijan
##                             11                             11
##                  Bahamas, The                        Bahrain
##                             11                             11
##                     Bangladesh                       Barbados
##                             11                             11
##                        Belarus                         Belize
##                             11                             11
##                          Benin                         Bhutan
##                             11                             11
##                        Bolivia          Bosnia and Herzegovina
##                             11                             11
##                       Botswana                         Brazil
##                             11                             11
##               Brunei Darussalam                       Bulgaria
##                             11                             11
##                   Burkina Faso                        Burundi
##                             11                             11
##                       Cambodia                       Cameroon
##                             11                             11
##                     Cape Verde       Central African Republic
##                             11                             11
##                           Chad                          Chile
##                             11                             11
##                          China                       Colombia
##                             11                             11
##                        Comoros               Congo, Dem. Rep.
```

```
##                        11                        11
##              Congo, Rep.                Costa Rica
##                        11                        11
##             Cote d'Ivoire                   Croatia
##                        11                        11
##                      Cuba                    Cyprus
##                        11                        11
##            Czech Republic                  Djibouti
##                        11                        11
##                  Dominica        Dominican Republic
##                        11                        11
##                   Ecuador          Egypt, Arab Rep.
##                        11                        11
##               El Salvador         Equatorial Guinea
##                        11                        11
##                   Eritrea                   Estonia
##                        11                        11
##                  Ethiopia                      Fiji
##                        11                        11
##                     Gabon               Gambia, The
##                        11                        11
##                   Georgia                     Ghana
##                        11                        11
##                   Grenada                 Guatemala
##                        11                        11
##                    Guinea             Guinea-Bissau
##                        11                        11
##                    Guyana                     Haiti
##                        11                        11
##                  Honduras                  Hongkong
##                        11                        11
##                   Hungary                     India
##                        11                        11
##                 Indonesia          Iran, Islamic Rep.
##                        11                        11
##                      Iraq                    Israel
##                        11                        11
##                   Jamaica                    Jordan
##                        11                        11
##                Kazakhstan                     Kenya
##                        11                        11
##                  Kiribati          Korea, Dem. Rep.
##                        11                        11
##               Korea, Rep.                    Kosovo
##                        11                        11
##                    Kuwait            Kyrgyz Republic
##                        11                        11
##                   Lao PDR                    Latvia
##                        11                        11
##                   Lebanon                   Lesotho
##                        11                        11
##                   Liberia                     Libya
##                        11                        11
##             Liechtenstein                 Lithuania
```

```
##                          11                          11
##               Macedonia, FYR                  Madagascar
##                          11                          11
##                      Malawi                    Malaysia
##                          11                          11
##                    Maldives                        Mali
##                          11                          11
##                       Malta             Marshall Islands
##                          11                          11
##                  Mauritania                   Mauritius
##                          11                          11
##                     Mayotte                      Mexico
##                          11                          11
##         Micronesia, Fed. Sts.                     Moldova
##                          11                          11
##                      Monaco                    Mongolia
##                          11                          11
##                  Montenegro                     Morocco
##                          11                          11
##                  Mozambique                     Myanmar
##                          11                          11
##                     Namibia                       Nepal
##                          11                          11
##          Netherlands Antilles                   Nicaragua
##                          11                          11
##                       Niger                     Nigeria
##                          11                          11
##                        Oman                    Pakistan
##                          11                          11
##                       Palau        Palestinian Adm. Areas
##                          11                          11
##                      Panama              Papua New Guinea
##                          11                          11
##                    Paraguay                        Peru
##                          11                          11
##                 Philippines                      Poland
##                          11                          11
##                       Qatar                     Romania
##                          11                          11
##            Russian Federation                      Rwanda
##                          11                          11
##                       Samoa                   San Marino
##                          11                          11
##         Sao Tome and Principe                Saudi Arabia
##                          11                          11
##                     Senegal                      Serbia
##                          11                          11
##                  Seychelles                Sierra Leone
##                          11                          11
##                   Singapore               Slovak Republic
##                          11                          11
##                    Slovenia               Solomon Islands
##                          11                          11
##                     Somalia                South Africa
```

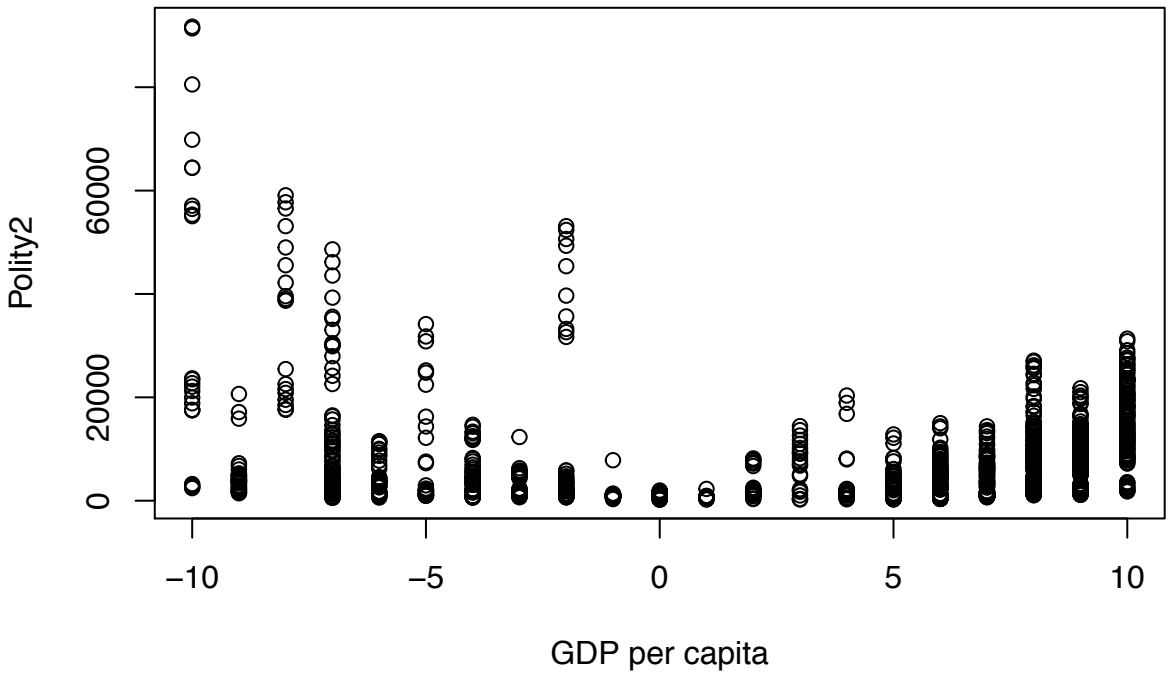
```
##                           11                                11
##                    Sri Lanka             St. Kitts and Nevis
##                           11                                11
##                    St. Lucia St. Vincent and the Grenadines
##                           11                                11
##                        Sudan                          Suriname
##                           11                                11
##                    Swaziland            Syrian Arab Republic
##                           11                                11
##                   Tajikistan                         Tanzania
##                           11                                11
##                     Thailand                      Timor-Leste
##                           11                                11
##                         Togo                            Tonga
##                           11                                11
##           Trinidad and Tobago                         Tunisia
##                           11                                11
##                       Turkey                      Turkmenistan
##                           11                                11
##                       Tuvalu                           Uganda
##                           11                                11
##                      Ukraine            United Arab Emirates
##                           11                                11
##                      Uruguay                       Uzbekistan
##                           11                                11
##                      Vanuatu                     Venezuela, RB
##                           11                                11
##                      Vietnam                      Yemen, Rep.
##                           11                                11
##                       Zambia                         Zimbabwe
##                           11                                11
```

```
data<-na.omit(data) # listwise deletion!!

dim(data)
```

```
## [1] 1305    4
```

```
attach(data)
plot(polity2, GDP.per.capita.PPP.current.international, ylab="Polity2", xlab="GDP per capita")
```




```
data$democracy[data$polity2>0]<-1
data$democracy[data$polity2<0|data$polity2==0]<-0
summary(data$democracy)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.6322  1.0000  1.0000
```

```
data$democracy.2<-rep(NA, length(data$polity2)) # 1305

for (i in 1:length(data$polity2)) {
    if (data$polity2[i]>0) data$democracy.2[i]<-1
     else data$democracy.2[i]<-0
     }

head(cbind(data$democracy, data$democracy.2))
```

```
##      [,1] [,2]
## [1,]    1    1
## [2,]    1    1
## [3,]    1    1
## [4,]    1    1
## [5,]    1    1
## [6,]    1    1
```

```
#rbind
```

Data frames

19. Subset the data frame to show only the country name and GDP per capita
20. Rearrange the columns of the data frame ascending by polity score
21. Show only values of GDP per capita for South Africa from 2002 to 2008

22. Create a new variable that takes the first letter of the country and attaches it to the year of observation
23. Find the mean of GDP per capita for each year of observation

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
head(select(data, country, GDP.per.capita.PPP.current.international))
```

```
##    country GDP.per.capita.PPP.current.international
## 23 Albania                                 4259.308
## 24 Albania                                 4658.009
## 25 Albania                                 4860.035
## 26 Albania                                 5230.007
## 27 Albania                                 5673.623
## 28 Albania                                 6161.608
```

```
head(data[, c(1,3)])
```

```
##    country GDP.per.capita.PPP.current.international
## 23 Albania                                 4259.308
## 24 Albania                                 4658.009
## 25 Albania                                 4860.035
## 26 Albania                                 5230.007
## 27 Albania                                 5673.623
## 28 Albania                                 6161.608
```

```
head(data.frame(data$country, data$GDP.per.capita.PPP.current.international))
```

```
##   data.country data.GDP.per.capita.PPP.current.international
## 1      Albania                                       4259.308
## 2      Albania                                       4658.009
## 3      Albania                                       4860.035
## 4      Albania                                       5230.007
## 5      Albania                                       5673.623
## 6      Albania                                       6161.608
```

```
head(arrange(data, polity2))
```

```
##   country Year GDP.per.capita.PPP.current.international polity2 democracy
## 1  Bhutan 2000                                2436.943     -10         0
## 2  Bhutan 2001                                2587.442     -10         0
## 3  Bhutan 2002                                2775.398     -10         0
## 4  Bhutan 2003                                2984.397     -10         0
## 5  Bhutan 2004                                3219.421     -10         0
## 6   Qatar 2000                               55053.515     -10         0
##   democracy.2
## 1           0
```

```
## 2           0
## 3           0
## 4           0
## 5           0
## 6           0
```

```
head(data[order(data$polity2),])
```

```
##      country Year GDP.per.capita.PPP.current.international polity2
## 166   Bhutan 2000                                 2436.943     -10
## 167   Bhutan 2001                                 2587.442     -10
## 168   Bhutan 2002                                 2775.398     -10
## 169   Bhutan 2003                                 2984.397     -10
## 170   Bhutan 2004                                 3219.421     -10
## 1387   Qatar 2000                                55053.515     -10
##      democracy democracy.2
## 166          0           0
## 167          0           0
## 168          0           0
## 169          0           0
## 170          0           0
## 1387         0           0
```

```
head(filter(data, country==c("South Africa"), Year>=2002 & Year<=2008))
```

```
##        country Year GDP.per.capita.PPP.current.international polity2
## 1 South Africa 2002                                 7244.218       9
## 2 South Africa 2003                                 7522.254       9
## 3 South Africa 2004                                 7992.767       9
## 4 South Africa 2005                                 8596.831       9
## 5 South Africa 2006                                 9269.283       9
## 6 South Africa 2007                                10002.543       9
##   democracy democracy.2
## 1         1           1
## 2         1           1
## 3         1           1
## 4         1           1
## 5         1           1
## 6         1           1
```

```
head(subset(data, data$country==c("South Africa") & data$Year>=2002 & Year<=2008))
```

```
##           country Year GDP.per.capita.PPP.current.international polity2
## 1444 South Africa 2002                                 7244.218       9
## 1445 South Africa 2003                                 7522.254       9
## 1446 South Africa 2004                                 7992.767       9
## 1447 South Africa 2005                                 8596.831       9
## 1448 South Africa 2006                                 9269.283       9
## 1449 South Africa 2007                                10002.543       9
##      democracy democracy.2
## 1444         1           1
## 1445         1           1
## 1446         1           1
## 1447         1           1
## 1448         1           1
## 1449         1           1
```

```
head(mutate(data, paste(substring(data$country, 1, 1), data$Year, sep="")))
```

```
##   country Year GDP.per.capita.PPP.current.international polity2 democracy
## 1 Albania 2000                                4259.308       5         1
## 2 Albania 2001                                4658.009       5         1
## 3 Albania 2002                                4860.035       7         1
## 4 Albania 2003                                5230.007       7         1
## 5 Albania 2004                                5673.623       7         1
## 6 Albania 2005                                6161.608       9         1
##   democracy.2 paste(substring(data$country, 1, 1), ...
## 1           1                                    A2000
## 2           1                                    A2001
## 3           1                                    A2002
## 4           1                                    A2003
## 5           1                                    A2004
## 6           1                                    A2005
```

```
data%>%
  group_by(Year)%>%
  summarize(mean(GDP.per.capita.PPP.current.international, na.rm=T)
            )
```

```
## # A tibble: 10 × 2
##    Year `mean(GDP.per.capita.PPP.current.inter...`
##    <chr>                                     <dbl>
## 1   2000                                  5757.223
## 2   2001                                  5976.854
## 3   2002                                  6167.580
## 4   2003                                  6597.168
## 5   2004                                  7157.506
## 6   2005                                  7712.546
## 7   2006                                  8416.708
## 8   2007                                  9218.926
## 9   2008                                  9566.308
## 10  2009                                  9113.082
```