

# Homework 1 (BA)

*Nitish Dabas (nd1292), Harsh Yadav (hy1217), Kush Shah (ks4437)*

*September 21, 2017*

## 1. Load the given data “contribution.csv”

```
setwd("/Users/nitishdabas/Desktop/hw1")
df <- read.csv("contribution.csv", header=TRUE, stringsAsFactors=FALSE)
```

## 2. How many observations and variables are there in this dataset?

```
dim(df)

## [1] 1230  11
```

Ans. 1230 Observations of 11 Variables.

## 3. Are there any missing values and anomaly values in each of the variable in the dataset?

```
colSums(is.na(df))
```

##	Gender	Class.Year	Marital.Status	Major
##	0	0	0	0
##	Next.Degree	FY04Giving	FY03Giving	FY02Giving
##	0	0	0	0
##	FY01Giving	FY00Giving	AttendanceEvent	
##	0	0	0	

Ans. No

## 4. What year(s) does the dataset cover?

```
unique(df$Class.Year)

## [1] 1957 1967 1977 1987 1997
```

Ans. 1957 1967 1977 1987 1997

## 5. Which year has the most alumni making contribution?

```
df.y0 <- subset(df, FY00Giving >0)
nrow(df.y0)
```

```
## [1] 539
```

```
df.y1 <- subset(df, FY01Giving >0)
nrow(df.y1)
```

```
## [1] 600
```

```
df.y2 <- subset(df, FY02Giving >0)
nrow(df.y2)
```

```
## [1] 548
```

```
df.y3 <- subset(df, FY03Giving >0)
nrow(df.y3)
```

```
## [1] 531
```

```
df.y4 <- subset(df, FY04Giving >0)
nrow(df.y4)
```

```
## [1] 507
```

Ans. Year FY01Giving (Calculated the number of almunis contributing each year by excluding '0' contributions and then selecting the maximum number of alumnis)

## 6. Which year has the largest average contribution?

```
df.m_y0 <- mean(df$FY00Giving)
df.m_y0      #Calculated the mean contribution of all years to select the largest value
```

```
## [1] 169.1818
```

```
df.m_y1 <- mean(df$FY01Giving)
df.m_y1
```

```
## [1] 276.5289
```

```
df.m_y2 <- mean(df$FY02Giving)
df.m_y2
```

```
## [1] 133.4584
```

```
df.m_y3 <- mean(df$FY03Giving)
df.m_y3
```

```
## [1] 241.4746
```

```
df.m_y4 <- mean(df$FY04Giving)
df.m_y4
```

```
## [1] 159.3999
```

Ans. FY01Givings

7. Which year has the largest 90th percentile contribution?

```
df.p_y0 <- quantile(df$FY00Giving, probs=0.9)
df.p_y0                                     #Calculated 90th percentile of each year to select the largest value

## 90%
## 200

df.p_y1 <- quantile(df$FY01Giving, probs=0.9)
df.p_y1

## 90%
## 210.5

df.p_y2 <- quantile(df$FY02Giving, probs=0.9)
df.p_y2

## 90%
## 200

df.p_y3 <- quantile(df$FY03Giving, probs=0.9)
df.p_y3

## 90%
## 200

df.p_y4 <- quantile(df$FY04Giving, probs=0.9)
df.p_y4

## 90%
## 207
```

Ans. FY01Giving

8. What are the standard deviation of the contribution amount in each of this years?

```
df.sd_y0 <- sd(df$FY00Giving)
df.sd_y0                                     #Calculated and printed the std of each other

## [1] 1170.638

df.sd_y1 <- sd(df$FY01Giving)
df.sd_y1

## [1] 4663.706

df.sd_y2 <- sd(df$FY02Giving)
df.sd_y2

## [1] 638.1267

df.sd_y3 <- sd(df$FY03Giving)
df.sd_y3
```

```
## [1] 2062.392
df.sd_y4 <- sd(df$FY04Giving)
df.sd_y4
```

```
## [1] 792.0376
```

**Ans. FY00Giving - 1170.638**

```
FY01Giving - 4663.706
FY02Giving - 638.1267
FY03Giving - 2062.392
FY04Giving - 792.0376
```

**9. Which degree contributed the most in the latest year in the data?**

```
degreelist<-c(unique(df$Next.Degree)) #Creates a list of all Degrees
j=0; #Initialize a variable to 0
sum_total <- c() # Initialize an empty list
for (degreename in degreelist) #Loop creates a list of total contribution of all degrees
{
  j=j+1;
  sum_total[j] <- (sum(subset(df$FY04Giving, df$Next.Degree==degreename)))
}
degreedata <- data.frame(degreelist, sum_total) #Created a new data frame
print(subset(degreedata$degreelist , sum_total==max(degreedata$sum_total)))

## [1] JD
## 50 Levels: AA BA BAE BD BFA BN BS BSE2 BSN DC DDS DMD DO DO2 DP JD ... UNKD
#Prints the degree with most contribution
```

**Ans. “JD”** (This result can also be deduced by printing out the total contribution of each degree along with its name and then manually searching for most contribution)

10. Which major contributed the most in the latest year in the data?

```
majorlist<-c(unique(df$Major)) #Creates a list of all Majors
i=0; #Initialize a variable to 0
sum_total2 <- c() # Initialize an empty list
for (majorname in majorlist) #Loop creates a list of total contribution of all majors
{
  i=i+1;
  sum_total2[i] <- (sum(subset(df$FY04Giving, df$Major==majorname)))
}
majordata <- data.frame(majorlist, sum_total2) #Created a new data frames
print (subset(majordata$majorlist , sum_total2==max(majordata$sum_total2)))

## [1] History
## 46 Levels: American Studies Anthropology Art Biology ... Zoology
#Prints the major with most contribution
```

Ans. “History” (This result can also be deduced by printing out the total contribution of each major along with its name and then manually searching for highest contribution)