ba1

Rithin Gujja

22/10/2019

```
## 1a ##
pnorm(700,mean=494,sd=100,lower.tail = FALSE)
```

```
## [1] 0.01969927
```

```
## 1b ##
pnorm(450,mean=494,sd=100)-pnorm(350,mean=494,sd=100)
```

```
## [1] 0.2550349
```

2

```
z<-qnorm(0.8665)
mean<-449-(z*36)
mean
```

```
## [1] 409.0401
```

3

```
kent<-c(59,68,78,60)
la<-c(90,82,78,75)
kentm<-mean(kent)
lam<-mean(la)
yy1<-sum((kent-kentm)*(la-lam))
yy2<-sqrt(sum((kent-kentm)*(kent-kentm)))*sqrt(sum((la-lam)*(la-lam)))
yy1/yy2
```

```
## [1] -0.3566049
```

4

```
getwd()
```

```
## [1] "C:/Users/RithinG/Desktop/fall 19/BA"
```

```
ba1<-read.csv("Online_Retail.csv")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

f1<-summarise(group_by(ba1,Country),Count=n())
f2<-as.data.frame(f1)
f3<-select(f2,Count)
percent<-(f3/sum(f3))*100
f4<-cbind(f2,percent)
names(f4)[3]<-"percentage"
filter(f4,percent>=1)

##          Country  Count percentage
## 1           EIRE   8196   1.512431
## 2         France   8557   1.579047
## 3        Germany   9495   1.752139
## 4 United Kingdom 495478  91.431956
```

5

```
Transactionvalue<-ba1$Quantity*ba1$UnitPrice
ba2<-cbind(ba1,Transactionvalue)
head(ba2)

##   InvoiceNo StockCode                          Description Quantity
## 1    536365    85123A  WHITE HANGING HEART T-LIGHT HOLDER        6
## 2    536365     71053                 WHITE METAL LANTERN        6
## 3    536365    84406B      CREAM CUPID HEARTS COAT HANGER        8
## 4    536365    84029G KNITTED UNION FLAG HOT WATER BOTTLE        6
## 5    536365    84029E      RED WOOLLY HOTTIE WHITE HEART.        6
## 6    536365     22752        SET 7 BABUSHKA NESTING BOXES        2
##        InvoiceDate UnitPrice CustomerID        Country Transactionvalue
## 1 12-01-2010 08:26      2.55      17850 United Kingdom            15.30
## 2 12-01-2010 08:26      3.39      17850 United Kingdom            20.34
## 3 12-01-2010 08:26      2.75      17850 United Kingdom            22.00
## 4 12-01-2010 08:26      3.39      17850 United Kingdom            20.34
## 5 12-01-2010 08:26      3.39      17850 United Kingdom            20.34
## 6 12-01-2010 08:26      7.65      17850 United Kingdom            15.30
```

6

```
f5<-
(summarise(group_by(ba2,Country),totalsum=sum(Transactionvalue))%>%filter(tot
alsum>130000))
f6<-as.data.frame(f5)
f6$Country

## [1] Australia      EIRE           France         Germany
## [5] Netherlands    United Kingdom
## 38 Levels: Australia Austria Bahrain Belgium Brazil ... USA
```

7

```
Temp=strptime(ba1$InvoiceDate,format='%m/%d/%Y %H:%M',tz='GMT')
ba1$New_Invoice_Date <- as.Date(Temp)
ba1$New_Invoice_Date[20000]- ba1$New_Invoice_Date[10]

## Time difference of NA days

ba1$Invoice_Day_Week= weekdays(ba1$New_Invoice_Date)
ba1$New_Invoice_Hour = as.numeric(format(Temp, "%H"))
ba1$New_Invoice_Month = as.numeric(format(Temp, "%m"))
```

7a

```
c7<-summarise(group_by(ba1,ba1$Invoice_Day_Week),count=n())
c8<-as.data.frame(c7)
c8$count/sum(c8$count)*100

## [1]  7.612164 10.594768  7.041773 10.756788 11.096697  9.909228 42.988583

               ## or ##
 tapply(ba1$Quantity,ba1$Invoice_Day_Week,NROW) / NROW(ba1$Quantity) * 100

##    Friday    Monday    Sunday  Thursday   Tuesday Wednesday
##  7.612164 10.594768  7.041773 10.756788 11.096697  9.909228
```

7b

```
d7<-summarise(group_by(ba1,ba1$Invoice_Day_Week),sum=sum(Quantity))
d8<-as.data.frame(d7)
d8$sum/sum(d8$sum)*100

## [1]  8.374446  9.614852  5.430382 12.148268 11.296255 10.789479 42.346318

              ## or ##
 tapply(ba1$Quantity,ba1$Invoice_Day_Week,sum) / sum(ba1$Quantity) * 100

##    Friday    Monday    Sunday  Thursday   Tuesday Wednesday
##  8.374446  9.614852  5.430382 12.148268 11.296255 10.789479
```

7c

```
e7<-summarise(group_by(ba1,ba1$New_Invoice_Month),sum=sum(Quantity))
e8<-as.data.frame(e7)
e8$sum/sum(e8$sum)*100

##  [1]  3.617054  3.393020  4.698181  3.399009  4.216133  3.855422  5.030455
##  [8]  4.145814  7.233065  6.619498  8.650330  2.795700 42.346318

               ## or ##
 tapply(ba1$Quantity,ba1$New_Invoice_Month,sum) / sum(ba1$Quantity) * 100

##        1        2        3        4        5        6        7        8
## 3.617054 3.393020 4.698181 3.399009 4.216133 3.855422 5.030455 4.145814
```

```
##        9       10       11       12
## 7.233065 6.619498 8.650330 2.795700
```

7d

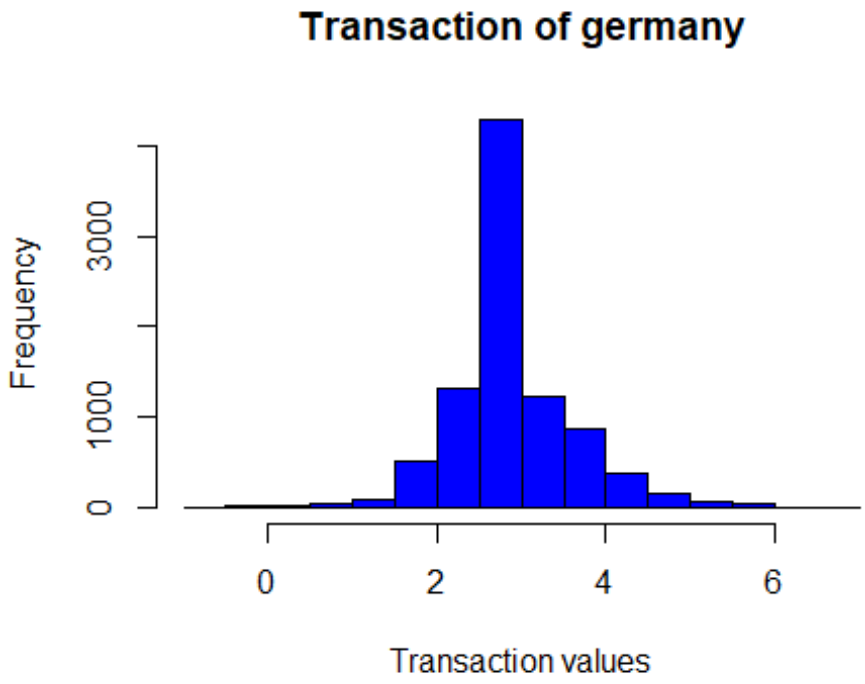
```
f7<-ba1%>%filter(Country=="Australia")
f8<-summarise(group_by(f7,Country),high=max(Quantity))
f9<-as.data.frame(f8)
f10<-filter(f7,Quantity==1152)
select(f10,Invoice_Day_Week)

##   Invoice_Day_Week
## 1             <NA>
```

7e

```
## install.packages("zoo") ##
library(zoo)

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

r <- table(ba1$New_Invoice_Hour)
r

##
##     6     7     8     9    10    11    12    13    14    15    16    17
##    41   203  4990 18454 27330 32882 43788 44127 39320 48619 26484 15730
##    18    19    20
##  4779  1761   442

rollapply(r, 2, sum)

##  [1]   244  5193 23444 45784 60212 76670 87915 83447 87939 75103 42214
## [12] 20509  6540  2203
```

8

```
library(ISLR)
f7<-select(ba2,8,9)%>%filter(Country=="Germany")
hist(log(f7$Transactionvalue),main = "Transaction of germany",
xlab="Transaction values",col = "blue")

## Warning in log(f7$Transactionvalue): NaNs produced
```



9

```
c9 <- tapply(ba2$Transactionvalue, ba2$CustomerID, length)
c9[which.max(c9)]

## 17841
##  7983

d9 <- tapply(ba2$Transactionvalue, ba2$CustomerID, sum)
d9[which.max(d9)]

##  14646
## 279489
```

10

```
colMeans(is.na(ba1))*100

##         InvoiceNo         StockCode       Description          Quantity
##           0.00000           0.00000           0.00000           0.00000
##       InvoiceDate         UnitPrice        CustomerID           Country
##           0.00000           0.00000          24.92669           0.00000
##  New_Invoice_Date  Invoice_Day_Week  New_Invoice_Hour New_Invoice_Month
##          42.98858          42.98858          42.98858          42.98858
```

11

```
fun1<-function(x){
  z<-sum(is.na(x))
```

```
  return(z)}
tapply(ba2$CustomerID,ba2$Country,fun1)
```

```
##            Australia              Austria              Bahrain
##                    0                    0                    2
##              Belgium               Brazil               Canada
##                    0                    0                    0
##      Channel Islands               Cyprus       Czech Republic
##                    0                    0                    0
##              Denmark                 EIRE   European Community
##                    0                  711                    0
##              Finland               France              Germany
##                    0                   66                    0
##               Greece            Hong Kong              Iceland
##                    0                  288                    0
##               Israel                Italy                Japan
##                   47                    0                    0
##              Lebanon            Lithuania                Malta
##                    0                    0                    0
##          Netherlands               Norway               Poland
##                    0                    0                    0
##             Portugal                  RSA         Saudi Arabia
##                   39                    0                    0
##            Singapore                Spain               Sweden
##                    0                    0                    0
##          Switzerland United Arab Emirates       United Kingdom
##                  125                    0               133600
##          Unspecified                  USA
##                  202                    0
```

```
            ## or ##
b11<-summarise(group_by(ba2,Country),Navalues=fun1(CustomerID))
as.data.frame(b11)
```

```
##                 Country Navalues
## 1             Australia        0
## 2               Austria        0
## 3               Bahrain        2
## 4               Belgium        0
## 5                Brazil        0
## 6                Canada        0
## 7       Channel Islands        0
## 8                Cyprus        0
## 9        Czech Republic        0
## 10              Denmark        0
## 11                 EIRE      711
## 12   European Community        0
## 13              Finland        0
## 14               France       66
## 15              Germany        0
```

```
## 16               Greece        0
## 17            Hong Kong      288
## 18              Iceland        0
## 19               Israel       47
## 20                Italy        0
## 21                Japan        0
## 22              Lebanon        0
## 23            Lithuania        0
## 24                Malta        0
## 25          Netherlands        0
## 26               Norway        0
## 27               Poland        0
## 28             Portugal       39
## 29                  RSA        0
## 30         Saudi Arabia        0
## 31            Singapore        0
## 32                Spain        0
## 33               Sweden        0
## 34          Switzerland      125
## 35 United Arab Emirates        0
## 36       United Kingdom   133600
## 37          Unspecified      202
## 38                  USA        0
```

12

13

```
b13<-ba2%>%filter(Country=="France")
b14<-filter(b13,Quantity<0)
nrow(b14)/nrow(b13)*100 ## with respect to french customers ##

## [1] 1.741264

nrow(b14)/nrow(ba2)*100 ## with respect to total customers ##

## [1] 0.02749539
```

14

```
b14<-summarise(group_by(ba2,Description),highvalue=sum(Transactionvalue))
c14<-as.data.frame(b14)
c14[which.max(c14$highvalue),]

##         Description highvalue
## 1140 DOTCOM POSTAGE  206245.5
```

15

```
length(unique(ba2$CustomerID))

## [1] 4373
```