DA_etsy

Ramya Balasubramaniam

3/5/2017

Loading the data and analyzing the structure

We load the data and create a small dataset on which we could work for calculating the statistical power

```
library(openxlsx)
library(Hmisc)
etsy_data <- read.xlsx("~/downloads/etsy_data/etsy_data_long.xlsx",
    1, detectDates = T)
Sales.weekdays <- weekdays(etsy_data$Sale.Date)
Sales.months <- months(etsy_data$Sale.Date)
etsy_data_wk_month <- cbind(Sales.weekdays, Sales.months, etsy_data)
etsy_DA <- etsy_data_wk_month[, c(1:3, 10, 20)]
str(etsy_DA)
```

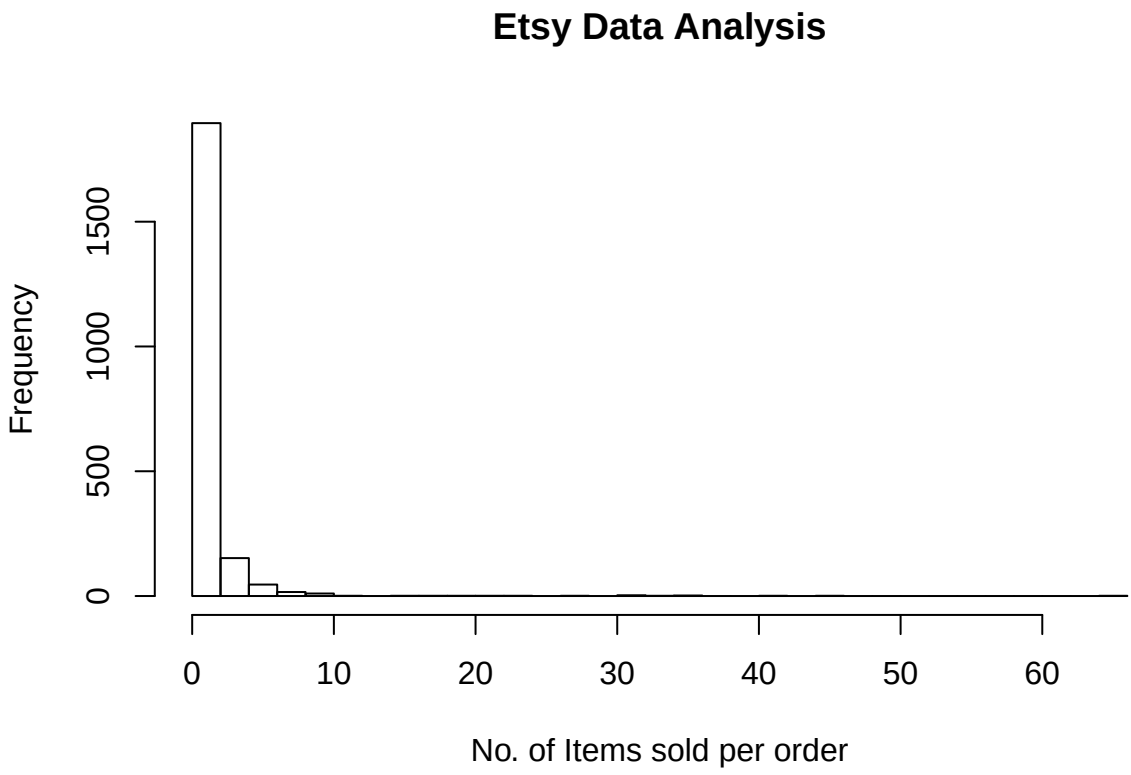
```
## 'data.frame':    2135 obs. of  5 variables:
##  $ Sales.weekdays : Factor w/ 7 levels "Friday","Monday",..: 1 1 5 7 6 2 2 2 2 2 ...
##  $ Sales.months   : Factor w/ 7 levels "April","December",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ Sale.Date      : Date, format: "2016-05-20" "2016-05-20" ...
##  $ Number.of.Items: num  2 1 1 6 1 1 1 1 2 1 ...
##  $ Order.Value    : num  14.15 8.75 16 33.9 5.75 ...
```

```
describe(etsy_DA)
```

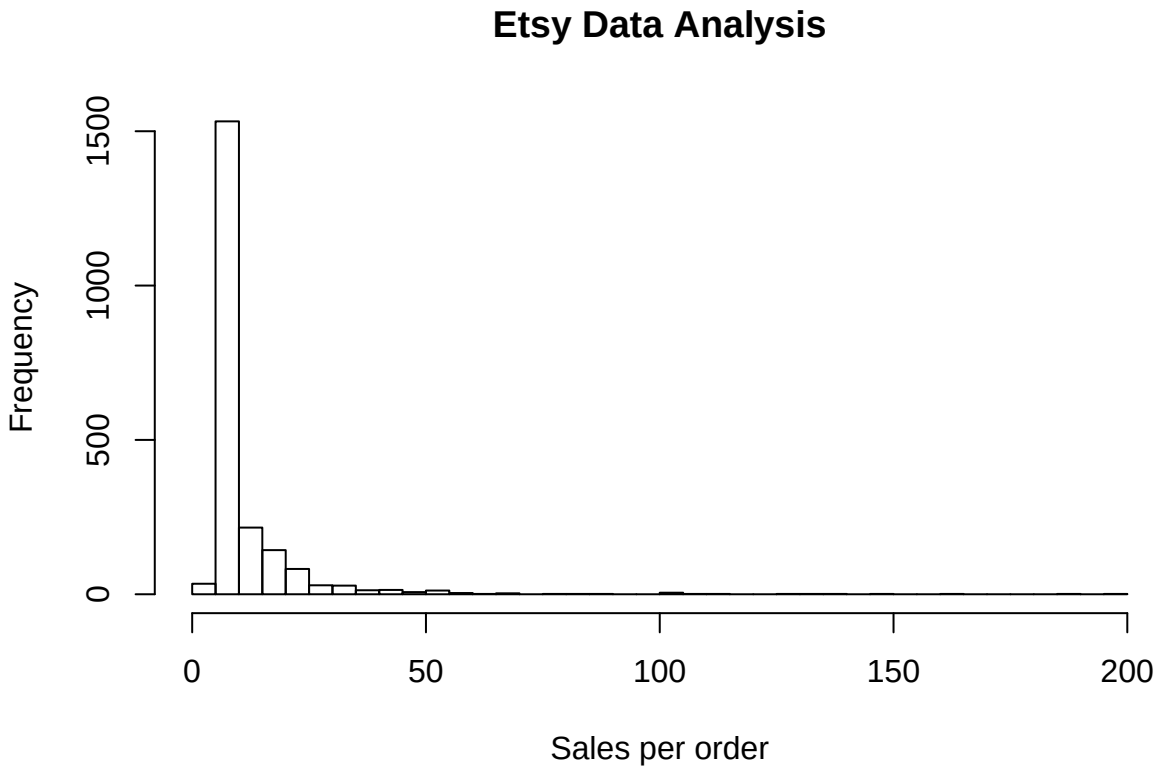
```
## etsy_DA
##
##  5  Variables      2135  Observations
## ---------------------------------------------------------------------------
## Sales.weekdays
##        n  missing distinct
##     2135        0        7
##
## Value         Friday    Monday  Saturday    Sunday  Thursday   Tuesday
## Frequency        295       369       281       299       322       316
## Proportion     0.138     0.173     0.132     0.140     0.151     0.148
##
## Value      Wednesday
## Frequency        253
## Proportion     0.119
## ---------------------------------------------------------------------------
## Sales.months
##        n  missing distinct
##     2135        0        7
##
## Value         April December February  January    March      May November
## Frequency       661      115      369      628      106       79      177
## Proportion    0.310    0.054    0.173    0.294    0.050    0.037    0.083
## ---------------------------------------------------------------------------
## Sale.Date
```

```
##        n  missing distinct
##     2135        0      191
##
## lowest : 2015-11-01 2015-11-02 2015-11-03 2015-11-04 2015-11-05
## highest: 2016-05-16 2016-05-17 2016-05-18 2016-05-19 2016-05-20
## ---------------------------------------------------------------------------
## Number.of.Items
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2135        0       24    0.573    1.706    1.247        1        1
##      .25      .50      .75      .90      .95
##        1        1        1        3        4
##
## lowest :  1  2  3  4  5, highest: 35 36 42 46 65
## ---------------------------------------------------------------------------
## Order.Value
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     2135        0      233    0.992    11.67    8.237     5.75     5.75
##      .25      .50      .75      .90      .95
##     6.50     8.20    11.50    19.71    28.95
##
## lowest :   0.20   3.00   4.75   4.95   5.25, highest: 137.80 147.60 164.00 185.45 198.65
## ---------------------------------------------------------------------------
```

```
hist(etsy_DA$Number.of.Items, xlab = "No. of Items sold per order",
    main = "Etsy Data Analysis", breaks = 30)
```




```
hist(etsy_DA$Order.Value, xlab = "Sales per order", main = "Etsy Data Analysis",
    breaks = 30)
```



Analysis of Data weekly basis

```
etsy_week_price <- tapply(etsy_DA$Order.Value, INDEX = etsy_DA$Sales.weekdays,
    FUN = mean)
sort(etsy_week_price)
```

```
##  Thursday    Monday    Friday    Sunday   Tuesday Wednesday  Saturday
##  10.67717  10.77425  11.14725  12.02753  12.03848  12.30202  13.17687
```

```
mean(etsy_week_price)
```

```
## [1] 11.7348
```

```
sd(etsy_week_price)
```

```
## [1] 0.9097285
```

```
etsy_week_item <- tapply(etsy_DA$Number.of.Items, INDEX = etsy_DA$Sales.weekdays,
    FUN = mean)
sort(etsy_week_item)
```

```
##  Thursday    Monday    Friday    Sunday Wednesday  Saturday   Tuesday
##  1.459627  1.566396  1.644068  1.722408  1.853755  1.868327  1.898734
```

```
mean(etsy_week_item)
```

```
## [1] 1.716188
```

```
sd(etsy_week_item)
```

```
## [1] 0.1677773
```

Analysis of Data monthly basis

```
etsy_mon_price <- tapply(etsy_DA$Order.Value, INDEX = etsy_DA$Sales.months,
    FUN = mean)
sort(etsy_mon_price)
```

```
## February    April      May  January    March December November
## 10.13924 11.04274 11.30506 11.57683 12.39717 12.45443 16.76469
```

```
mean(etsy_mon_price)
```

```
## [1] 12.24002
```

```
sd(etsy_mon_price)
```

```
## [1] 2.148799
```

```
etsy_mon_item <- tapply(etsy_DA$Number.of.Items, INDEX = etsy_DA$Sales.months,
    FUN = mean)
sort(etsy_mon_item)
```

```
##    April February      May    March  January December November
## 1.453858 1.457995 1.658228 1.707547 1.808917 1.878261 2.706215
```

```
mean(etsy_mon_item)
```

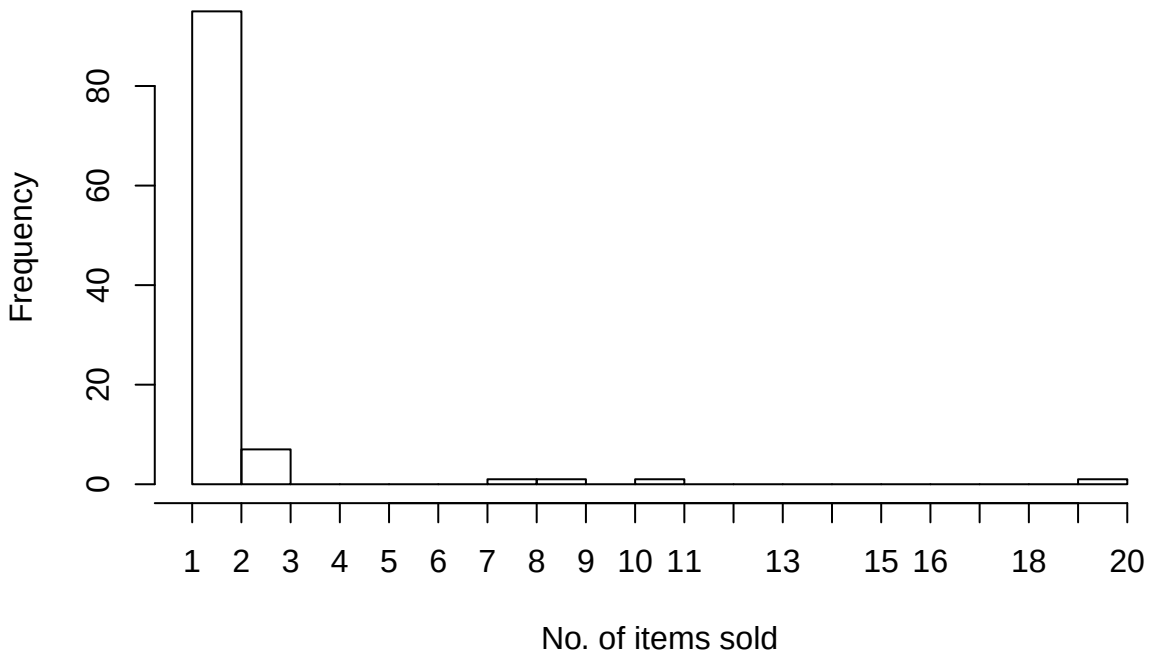
```
## [1] 1.810146
```

```
sd(etsy_mon_item)
```

```
## [1] 0.4266415
```

Analysis od Data in March

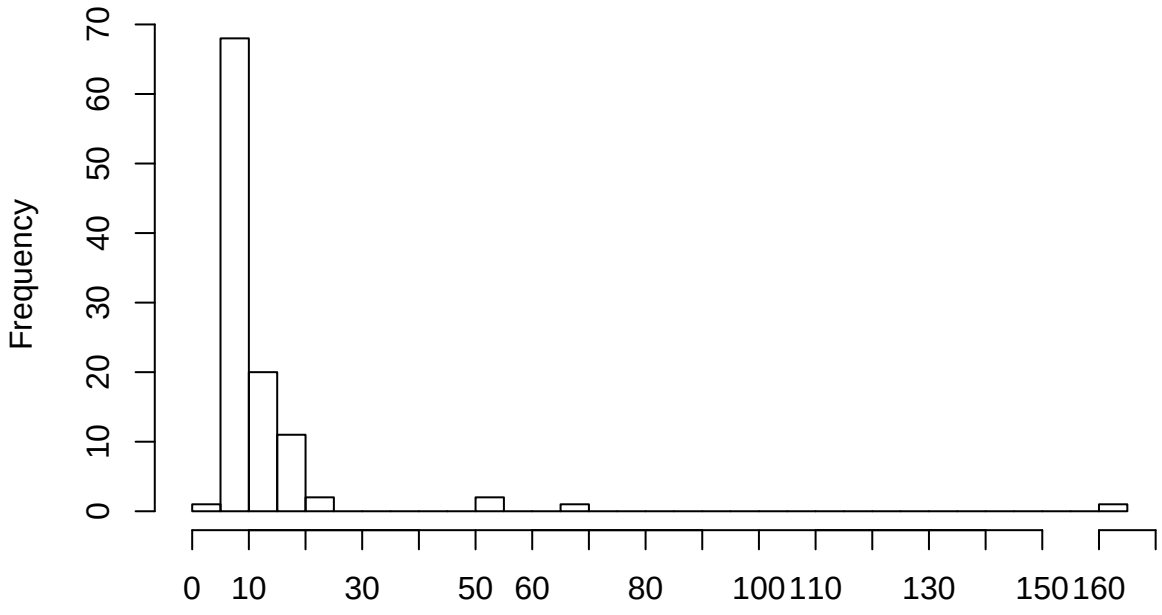
```
etsy_DA_march <- subset(etsy_DA, etsy_DA$Sales.months == "March")
hist(etsy_DA_march$Number.of.Items, xlab = "No. of items sold",
    breaks = 20, main = "Etsy data March")
axis(1, c(0:4, 6:9, 11:14, 16:19))
```




```
table(etsy_DA_march$Number.of.Items)
```

```
##
##  1  2  3  8  9 11 20
## 78 17  7  1  1  1  1
```

```
hist(etsy_DA_march$Order.Value, xlab = "Sales per order", breaks = 25,
    main = "Etsy data March")
axis(1, seq(10, 40, by = 10))
axis(1, seq(60, 90, by = 10))
axis(1, seq(110, 140, by = 10))
axis(1, seq(160, 170, by = 10))
```




```
etsy_march_price <- tapply(etsy_DA_march$Order.Value, INDEX = etsy_DA_march$Sales.weekdays,
    FUN = mean)
sort(etsy_march_price)
```

```
##    Friday  Saturday Wednesday    Monday   Tuesday    Sunday  Thursday
##  8.357692  9.334615  9.675000 10.763158 12.723333 13.703571 23.053571
```

```
mean(etsy_march_price)
```

```
## [1] 12.51585
```

```
sd(etsy_march_price)
```

```
## [1] 5.017473
```

```
etsy_march_item <- tapply(etsy_DA_march$Number.of.Items, INDEX = etsy_DA_march$Sales.weekdays,
    FUN = mean)
sort(etsy_march_item)
```

```
##    Friday Wednesday  Saturday    Monday    Sunday   Tuesday  Thursday
##  1.076923  1.333333  1.384615  1.526316  1.857143  1.866667  3.000000
```

```
mean(etsy_march_item)
```

```
## [1] 1.720714
```

```
sd(etsy_march_item)
```

```
## [1] 0.6313704
```

Comparison of weekly data for March and entire bimonthly data

```
sort(etsy_march_item)
```

```
##    Friday Wednesday  Saturday    Monday    Sunday   Tuesday  Thursday
##  1.076923  1.333333  1.384615  1.526316  1.857143  1.866667  3.000000
```

```
sort(etsy_week_item)
```

```
##  Thursday    Monday    Friday    Sunday Wednesday  Saturday   Tuesday
##  1.459627  1.566396  1.644068  1.722408  1.853755  1.868327  1.898734
```

```
sort(etsy_march_price)
```

```
##    Friday  Saturday Wednesday    Monday   Tuesday    Sunday  Thursday
##  8.357692  9.334615  9.675000 10.763158 12.723333 13.703571 23.053571
```

```
sort(etsy_week_price)
```

```
##  Thursday    Monday    Friday    Sunday   Tuesday Wednesday  Saturday
##  10.67717  10.77425  11.14725  12.02753  12.03848  12.30202  13.17687
```

```
etsy_march <- subset(etsy_DA, etsy_DA$Sales.months == "March")
dte = seq(as.Date("2016-03-01"), as.Date("2016-03-07"), by = 1)
etsy_march1 <- subset(etsy_march, etsy_march$Sale.Date %in% dte)
dte = seq(as.Date("2016-03-08"), as.Date("2016-03-14"), by = 1)
etsy_march2 <- subset(etsy_march, etsy_march$Sale.Date %in% dte)
dte = seq(as.Date("2016-03-15"), as.Date("2016-03-21"), by = 1)
etsy_march3 <- subset(etsy_march, etsy_march$Sale.Date %in% dte)
dte = seq(as.Date("2016-03-22"), as.Date("2016-03-28"), by = 1)
etsy_march4 <- subset(etsy_march, etsy_march$Sale.Date %in% dte)
tapply(etsy_march1$Order.Value, INDEX = etsy_march1$Sales.months,
    FUN = sum)
```

```
##    April December February  January    March      May November
##       NA       NA       NA       NA    165.8       NA       NA
```

```
tapply(etsy_march2$Order.Value, INDEX = etsy_march2$Sales.months,
    FUN = sum)
```

```
##    April December February  January    March      May November
##       NA       NA       NA       NA   483.25       NA       NA
```

```
tapply(etsy_march3$Order.Value, INDEX = etsy_march3$Sales.months,
    FUN = sum)
```

```
##    April December February  January    March      May November
##       NA       NA       NA       NA   145.45       NA       NA
```

```
tapply(etsy_march4$Order.Value, INDEX = etsy_march4$Sales.months,
    FUN = sum)
```

```
##    April December February  January    March      May November
##       NA       NA       NA       NA   218.25       NA       NA
```