

R Notebook

This is an R Markdown (<http://rmarkdown.rstudio.com>) Notebook. When you execute code within the notebook, the results appear beneath the code.

```
dat=read.csv('Walmart_Data.csv')
str(dat)
```

```
## 'data.frame':    100 obs. of  6 variables:
## $ week      : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Sales     : int  586953 838022 861991 767198 777392 725924 701517 1027152 755625 445967
## ...
## $ Promotion: num  0.89 1.08 0.95 1.06 1.01 1.07 1.22 1.06 1.08 0.8 ...
## $ Feature  : num  0.87 0.84 1.12 0.95 1.06 1.09 1.03 1.08 0.99 0.88 ...
## $ Walmart  : Factor w/ 2 levels "No","Present": 1 1 1 1 1 1 1 1 1 1 ...
## $ Holiday  : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(dat)
```

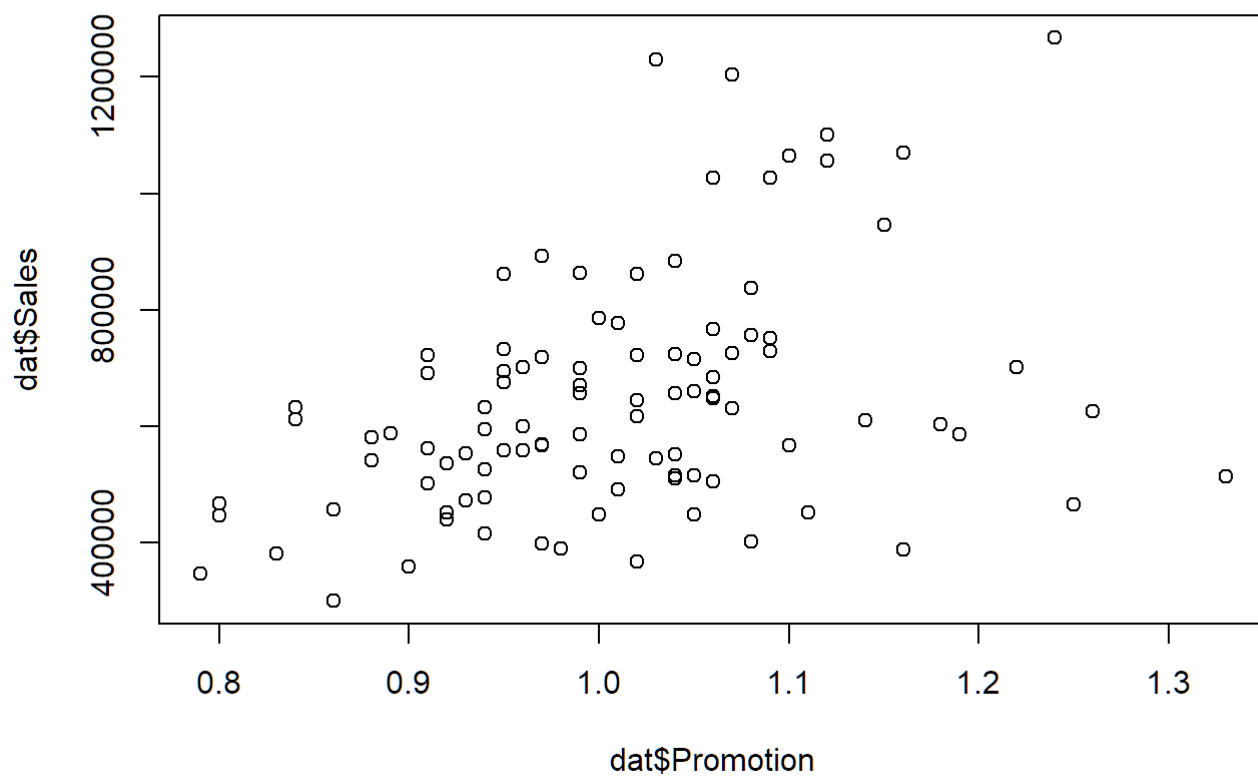
```
##      week      Sales      Promotion      Feature
## Min.   : 1.00   Min.   : 299359   Min.   :0.790   Min.   :0.780
## 1st Qu.: 25.75   1st Qu.: 512627   1st Qu.:0.940   1st Qu.:0.940
## Median : 50.50   Median : 610755   Median :1.010   Median :1.015
## Mean   : 50.50   Mean   : 644054   Mean   :1.011   Mean   :1.007
## 3rd Qu.: 75.25   3rd Qu.: 722809   3rd Qu.:1.062   3rd Qu.:1.080
## Max.   :100.00   Max.   :1267301   Max.   :1.330   Max.   :1.260
##      Walmart  Holiday
## No      :50   No :92
## Present:50   Yes: 8
##
##
##
##
```

```
dat$logSales=log(dat$Sales)
```

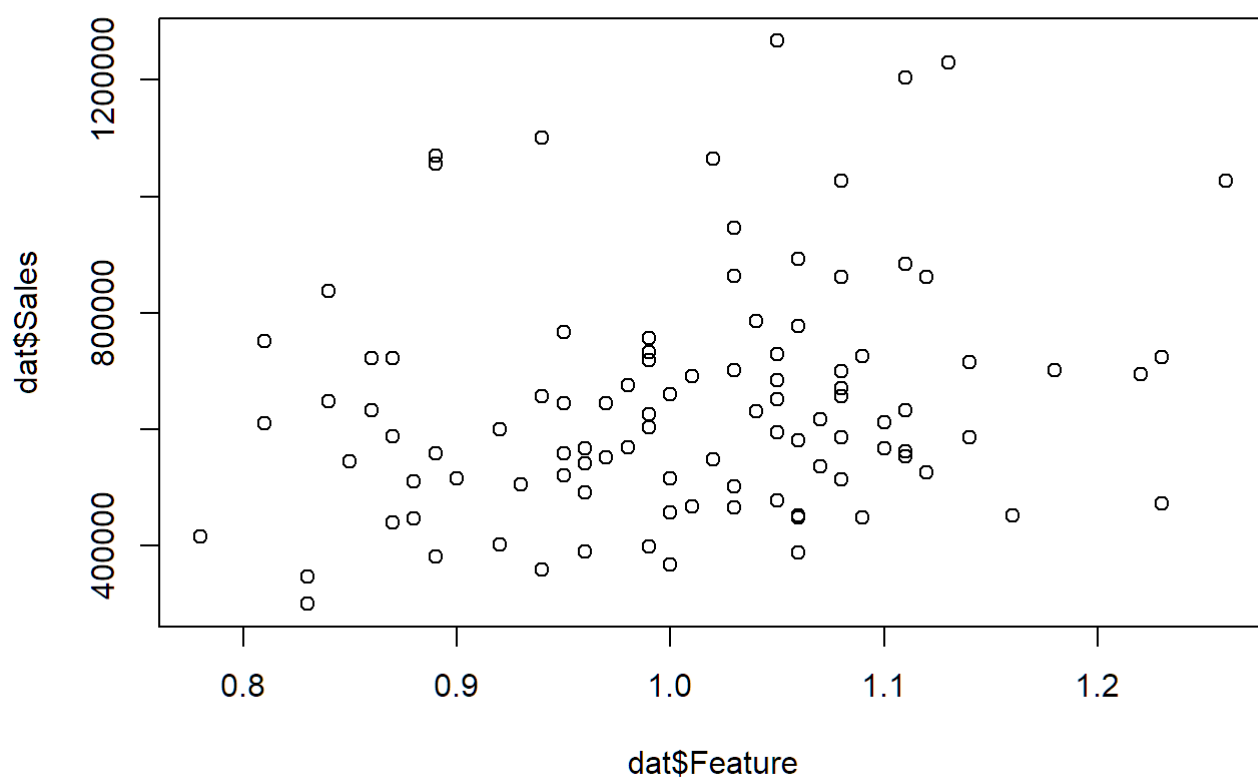
```
cor(dat[,2:4],use="complete.obs", method="pearson")
```

```
##           Sales  Promotion   Feature
## Sales      1.0000000 0.37739562 0.22438793
## Promotion  0.3773956 1.00000000 0.06513678
## Feature    0.2243879 0.06513678 1.00000000
```

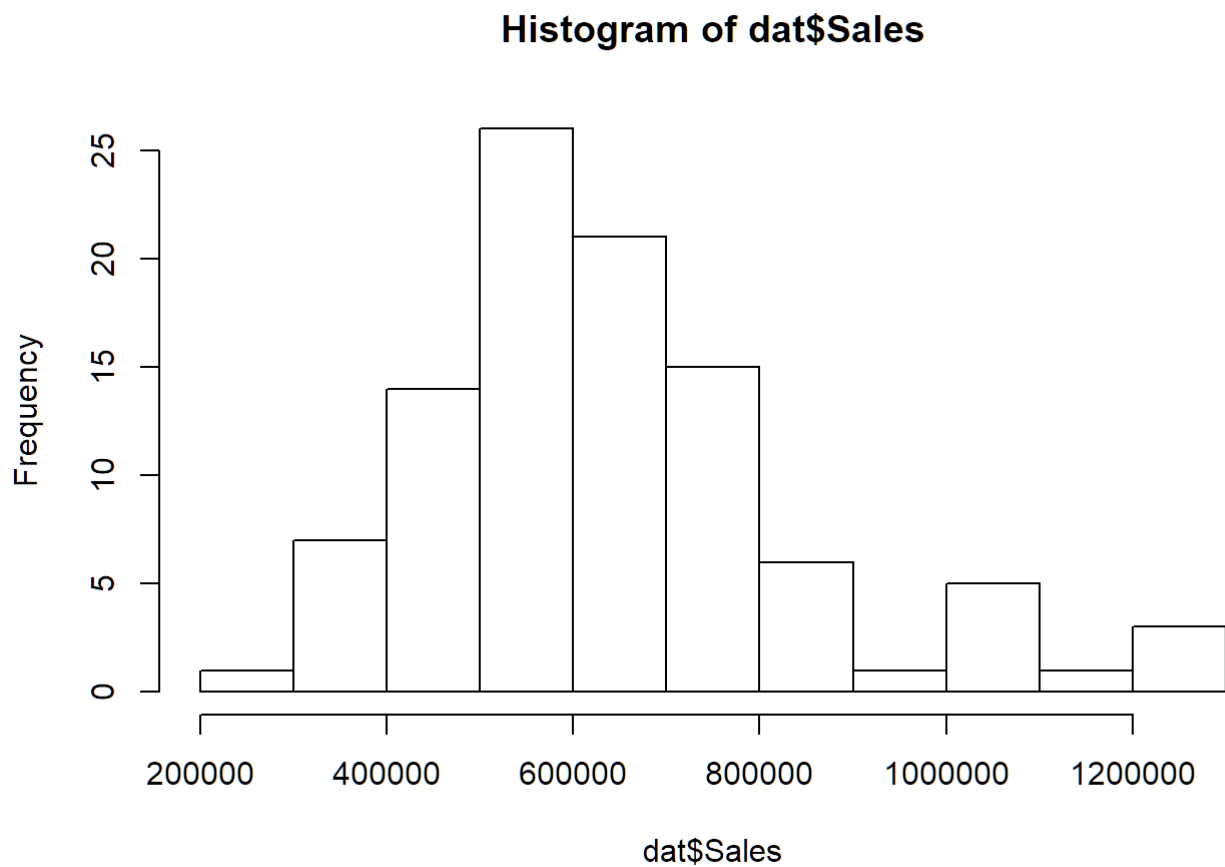
```
plot(dat$Promotion, dat$Sales)
```



```
plot(dat$Feature, dat$Sales)
```

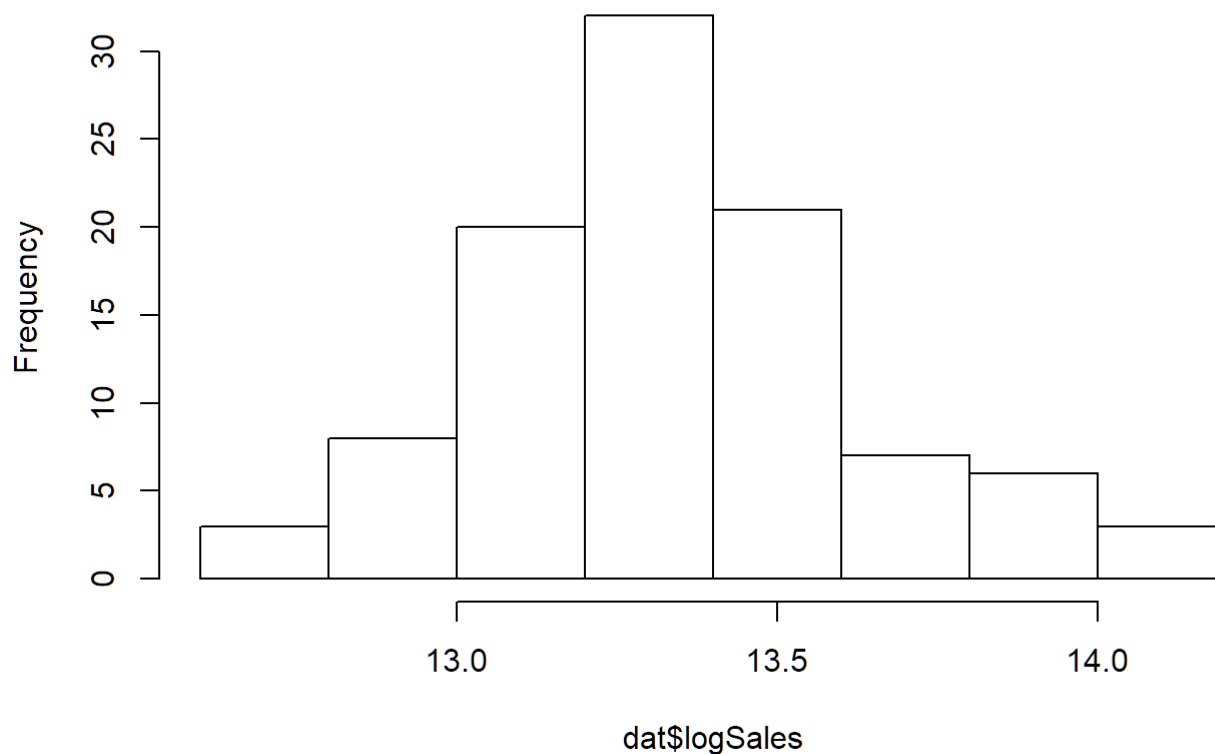


```
hist(dat$Sales)
```



```
hist(dat$logSales)
```

Histogram of dat\$logSales



```
a=lm(logSales~Promotion+Feature+Walmart+Holiday,data=dat)
summary(a)
```

```
##
## Call:
## lm(formula = logSales ~ Promotion + Feature + Walmart + Holiday,
##     data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45435 -0.15761 -0.00412  0.12948  0.46955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.85276    0.28826  41.119  < 2e-16 ***
## Promotion      0.84754    0.20635   4.107 8.48e-05 ***
## Feature        0.75076    0.20774   3.614 0.000485 ***
## WalmartPresent -0.31127    0.04233 -7.354 6.76e-11 ***
## HolidayYes     0.26004    0.07765   3.349 0.001164 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.21 on 95 degrees of freedom
## Multiple R-squared:  0.5206, Adjusted R-squared:  0.5004
## F-statistic: 25.79 on 4 and 95 DF, p-value: 1.76e-14
```

```
b=lm(logSales~Promotion+Feature+Walmart+Holiday+Holiday:Walmart+Holiday:Promotion,data=dat)
summary(b)
```

```
##
## Call:
## lm(formula = logSales ~ Promotion + Feature + Walmart + Holiday +
##      Holiday:Walmart + Holiday:Promotion, data = dat)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -0.44745 -0.14350  0.00013  0.11836  0.47639
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.9169     0.2994   39.806 < 2e-16 ***
## Promotion          0.7454     0.2236    3.333  0.00123 **
## Feature           0.7828     0.2099    3.729  0.00033 ***
## WalmartPresent   -0.2978     0.0439   -6.783 1.08e-09 ***
## HolidayYes       -0.1128     0.7428   -0.152  0.87961
## WalmartPresent:HolidayYes -0.1307     0.1887   -0.693  0.49034
## Promotion:HolidayYes  0.4330     0.6741    0.642  0.52219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2101 on 93 degrees of freedom
## Multiple R-squared:  0.5302, Adjusted R-squared:  0.4999
## F-statistic: 17.49 on 6 and 93 DF, p-value: 1.866e-13
```

AIC(a)

```
## [1] -21.45434
```

AIC(b)

```
## [1] -19.48443
```

BIC(a)

```
## [1] -5.823324
```

BIC(b)

```
## [1] 1.356929
```

```
sow.data=read.csv("CreditCard_SOW_Data.csv")
sow.data$ConsumerID=as.factor(sow.data$ConsumerID)
sow.data$logIncome=log(sow.data$Income)
sow.data$logSowRatio=log(sow.data$WalletShare/(1-sow.data$WalletShare))
str(sow.data)
```

```
## 'data.frame': 3600 obs. of 8 variables:
## $ ConsumerID : Factor w/ 300 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ History : int 55 55 55 55 55 55 55 55 55 55 ...
## $ Income : num 82000 82000 82000 82000 82000 82000 82000 82000 82000 82000 ...
## $ WalletShare: num 0.643 0.628 0.567 0.638 0.554 0.573 0.666 0.649 0.527 0.459 ...
## $ Promotion : num 0.5 0.2 1 0.8 0.7 1.1 0.9 0.6 0.1 0 ...
## $ Balance : int 836 467 1208 792 1215 1248 197 567 1190 1709 ...
## $ logIncome : num 11.3 11.3 11.3 11.3 11.3 ...
## $ logSowRatio: num 0.588 0.524 0.27 0.567 0.217 ...
```

```
lml=lm(logSowRatio~History+Balance+Promotion+History:Promotion+logIncome:Promotion, data=sow.data)
summary(lml)
```

```
##
## Call:
## lm(formula = logSowRatio ~ History + Balance + Promotion + History:Promotion +
##     logIncome:Promotion, data = sow.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59976 -0.14401  0.00153  0.13634  0.75883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.908e-02  1.603e-02   5.558 2.92e-08 ***
## History        1.039e-02  4.153e-04  25.027 < 2e-16 ***
## Balance       -4.959e-04  2.882e-06 -172.064 < 2e-16 ***
## Promotion      7.777e-01  1.888e-01   4.120 3.87e-05 ***
## History:Promotion -2.598e-03  5.722e-04  -4.541 5.79e-06 ***
## Promotion:logIncome -4.558e-02  1.651e-02  -2.760 0.00581 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2078 on 3594 degrees of freedom
## Multiple R-squared:  0.8984, Adjusted R-squared:  0.8982
## F-statistic: 6353 on 5 and 3594 DF, p-value: < 2.2e-16
```

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.0.0
```

```
## Loading required package: Matrix
```

```
rel=lmer(logSowRatio~History+Balance+Promotion+History:Promotion+logIncome:Promotion+(1+Promoti
on|ConsumerID), data=sow.data, REML=F, control=lmerControl(optimizer="Nelder_Mead"))
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
summary(rel)
```

```
## Linear mixed model fit by maximum likelihood ['lmerMod']
## Formula:
## logSowRatio ~ History + Balance + Promotion + History:Promotion +
##   logIncome:Promotion + (1 + Promotion | ConsumerID)
## Data: sow.data
## Control: lmerControl(optimizer = "Nelder_Mead")
##
##      AIC      BIC    logLik deviance df.resid
## -6532.1 -6470.2   3276.0  -6552.1     3590
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.1063 -0.6424  0.0049  0.6336  3.4532
##
## Random effects:
## Groups      Name      Variance Std.Dev. Corr
## ConsumerID (Intercept) 0.0359421 0.18958
##              Promotion  0.0005355 0.02314  0.06
## Residual              0.0066071 0.08128
## Number of obs: 3600, groups: ConsumerID, 300
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    9.595e-02  2.655e-02   3.613
## History        1.039e-02  7.135e-04  14.569
## Balance       -5.003e-04  1.799e-06 -278.110
## Promotion      6.129e-01  1.466e-01   4.181
## History:Promotion -2.571e-03  2.402e-04 -10.703
## Promotion:logIncome -3.110e-02  1.288e-02  -2.414
##
## Correlation of Fixed Effects:
##              (Intr) Histry Balanc Promtn Hstr:P
## History      -0.900
## Balance      -0.107 -0.001
## Promotion    -0.011  0.009  0.013
## Hstry:Prmtn  0.143 -0.159 -0.002 -0.153
## Prmtn:lgInc  0.001  0.000 -0.012 -0.998  0.099
## fit warnings:
## Some predictor variables are on very different scales: consider rescaling
```

```
AIC(lm1)
```

```
## [1] -1087.389
```

```
AIC(re1)
```

```
## [1] -6532.094
```

```
BIC(lm1)
```

```
## [1] -1044.069
```

```
BIC(re1)
```

```
## [1] -6470.207
```