**Console Output with plots**

**BY**

**Anjali Jha**

**IIT Kharagpur**

**Email: anjalijha167@gmail.com**

**Contact: 9672076045**

> mba.df<-read.csv(paste("MBA Starting Salaries Data.csv",sep = ""))

> View(mba.df)

> library(psych)

> summary(mba.df)

age sex gmat\_tot gmat\_qpc gmat\_vpc gmat\_tpc

Min. :22 Min. :1.00 Min. :450 Min. :28 Min. :16 Min. : 0

1st Qu.:25 1st Qu.:1.00 1st Qu.:580 1st Qu.:72 1st Qu.:71 1st Qu.:78

Median :27 Median :1.00 Median :620 Median :83 Median :81 Median :87

Mean :27 Mean :1.25 Mean :619 Mean :81 Mean :78 Mean :84

3rd Qu.:29 3rd Qu.:1.00 3rd Qu.:660 3rd Qu.:93 3rd Qu.:91 3rd Qu.:94

Max. :48 Max. :2.00 Max. :790 Max. :99 Max. :99 Max. :99

s\_avg f\_avg quarter work\_yrs frstlang

Min. :2.0 Min. :0.0 Min. :1.0 Min. : 0.0 Min. :1.00

1st Qu.:2.7 1st Qu.:2.8 1st Qu.:1.2 1st Qu.: 2.0 1st Qu.:1.00

Median :3.0 Median :3.0 Median :2.0 Median : 3.0 Median :1.00

Mean :3.0 Mean :3.1 Mean :2.5 Mean : 3.9 Mean :1.12

3rd Qu.:3.3 3rd Qu.:3.2 3rd Qu.:3.0 3rd Qu.: 4.0 3rd Qu.:1.00

Max. :4.0 Max. :4.0 Max. :4.0 Max. :22.0 Max. :2.00

salary satis

Min. : 0 Min. : 1

1st Qu.: 0 1st Qu.: 5

Median : 999 Median : 6

Mean : 39026 Mean :172

3rd Qu.: 97000 3rd Qu.: 7

Max. :220000 Max. :998

> describe(mba.df$age)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 27 3.7 27 27 3 22 48 26 2.2 6.4 0.22

> describe(mba.df$sex)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 1.2 0.43 1 1.2 0 1 2 1 1.2 -0.66 0.03

> describe(mba.df$gmat\_tot)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 619 58 620 619 59 450 790 340 -0.01 0.06 3.5

> describe(mba.df$gmat\_qpc)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 81 15 83 82 15 28 99 71 -0.92 0.3 0.9

> describe(mba.df$gmat\_vpc)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 78 17 81 80 15 16 99 83 -1 0.74 1

> describe(mba.df$gmat\_tpc)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 84 14 87 86 12 0 99 99 -2.3 9 0.85

> describe(mba.df$s\_avg)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 3 0.38 3 3 0.44 2 4 2 -0.06 -0.38 0.02

> describe(mba.df$f\_avg)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 3.1 0.53 3 3.1 0.37 0 4 4 -2.1 11 0.03

> describe(mba.df$quarter)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 2.5 1.1 2 2.5 1.5 1 4 3 0.02 -1.4 0.07

> describe(mba.df$work\_yrs)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 3.9 3.2 3 3.3 1.5 0 22 22 2.8 9.8 0.2

> describe(mba.df$salary)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 39026 50952 999 33608 1481 0 220000 220000 0.7 -1.0 3078

> describe(mba.df$satis)

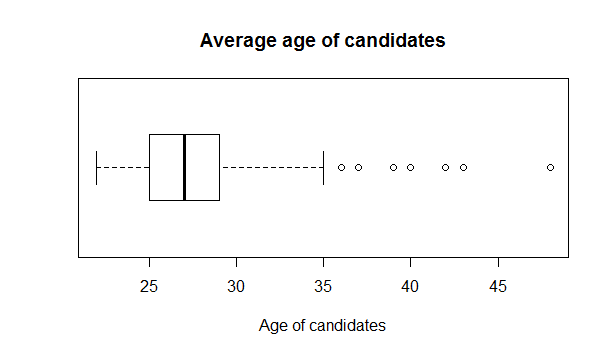
vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 274 172 372 6 92 1.5 1 998 997 1.8 1.1 22

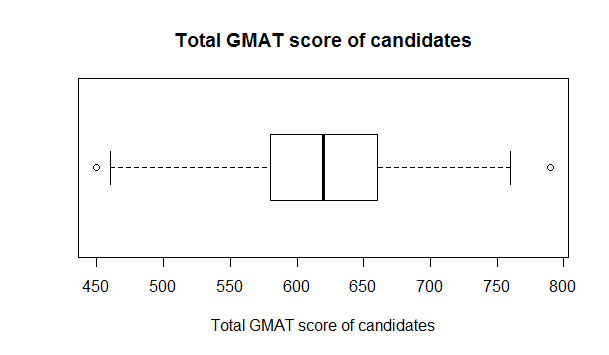
mbap.df<-mba.df[which(mba.df$salary>999),]

> mbanp.df<-mba.df[which(mba.df$salary==0),]

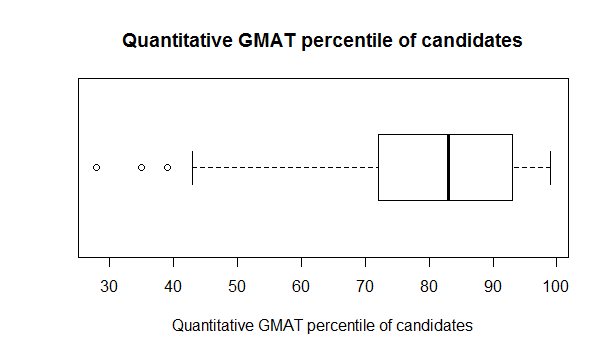
> boxplot(mba.df$age,horizontal=TRUE,xlab="Age of candidates",main ="Average age of candidates")

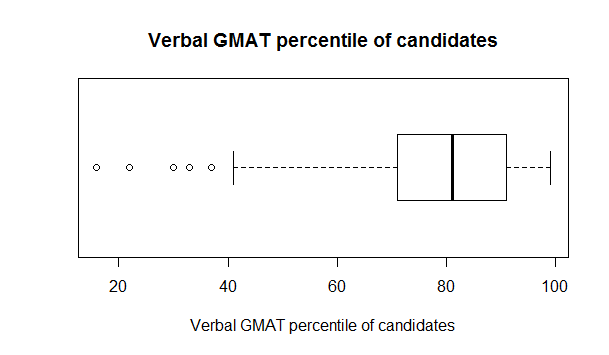


|  |
| --- |
| > boxplot(mba.df$gmat\_tot,horizontal=TRUE,xlab="Total GMAT score of candidates",main ="Total GMAT score of candidates") |
|  |
| |  | | --- | |  | |

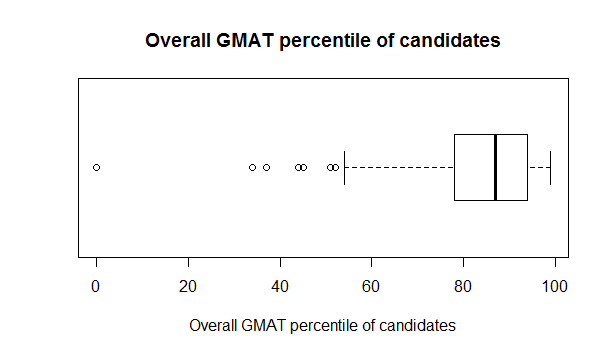


boxplot(mba.df$gmat\_qpc,horizontal=TRUE,xlab="Quantitative GMAT percentile of candidates",main ="Quantitative GMAT percentile of candidates")

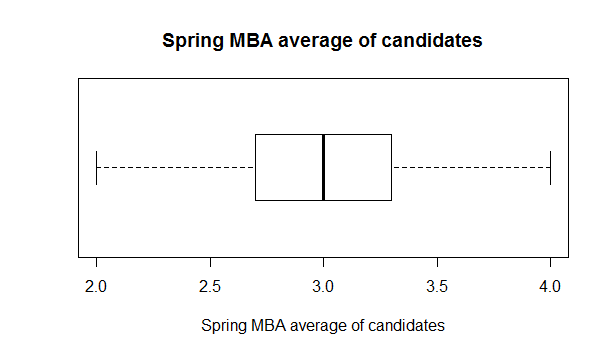
boxplot(mba.df$gmat\_vpc,horizontal=TRUE,xlab="Verbal GMAT percentile of candidates",main ="Verbal GMAT percentile of candidates")



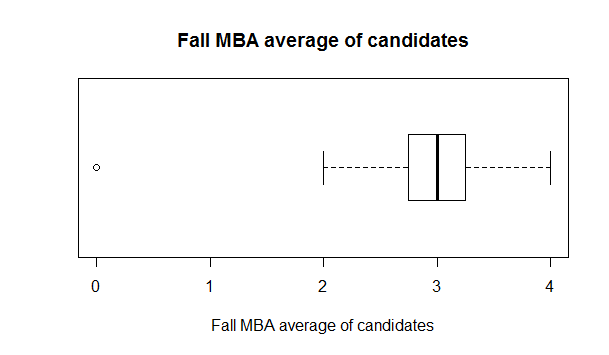
|  |
| --- |
| > boxplot(mba.df$gmat\_tpc,horizontal=TRUE,xlab="Overall GMAT percentile of candidates",main ="Overall GMAT percentile of candidates") |
|  |
| |  | | --- | |  | |



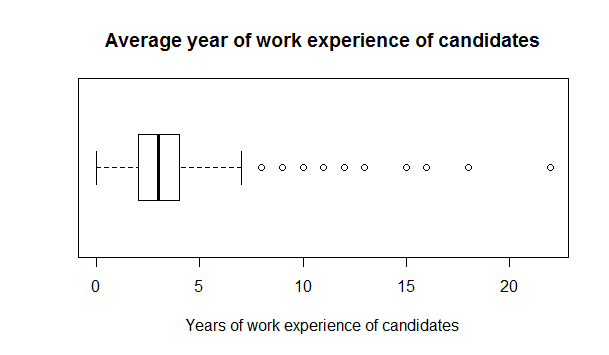
|  |
| --- |
| > boxplot(mba.df$s\_avg,horizontal=TRUE,xlab="Spring MBA average of candidates",main ="Spring MBA average of candidates") |
|  |
| |  | | --- | |  | |



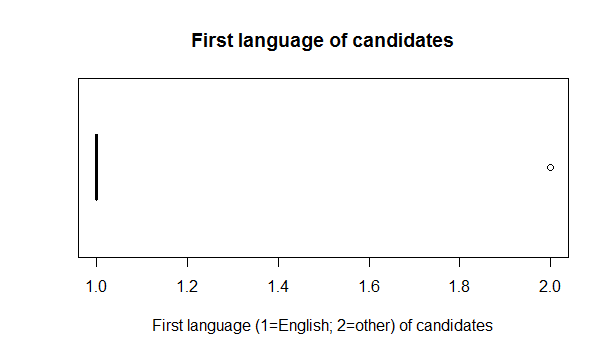
|  |
| --- |
| > boxplot(mba.df$f\_avg,horizontal=TRUE,xlab="Fall MBA average of candidates",main ="Fall MBA average of candidates") |
|  |
| |  | | --- | |  | |



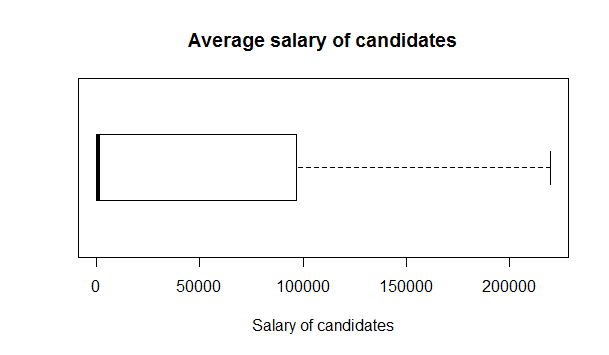
|  |
| --- |
| > boxplot(mba.df$work\_yrs,horizontal=TRUE,xlab="Years of work experience of candidates",main ="Average year of work experience of candidates") |
|  |
| |  | | --- | |  | |



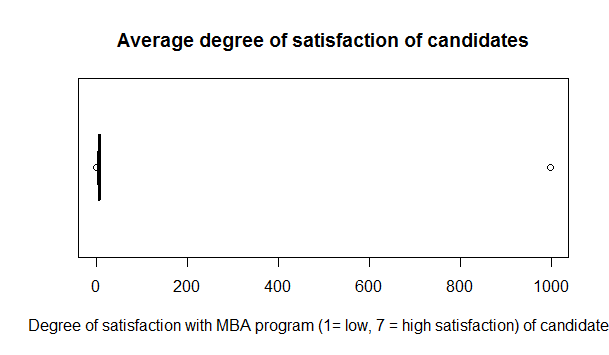
|  |
| --- |
| > boxplot(mba.df$frstlang,horizontal=TRUE,xlab="First language (1=English; 2=other) of candidates",main ="First language of candidates") |
|  |
| |  | | --- | |  | |



|  |
| --- |
| > boxplot(mba.df$salary,horizontal=TRUE,xlab="Salary of candidates",main ="Average salary of candidates") |
|  |
| |  | | --- | |  | |

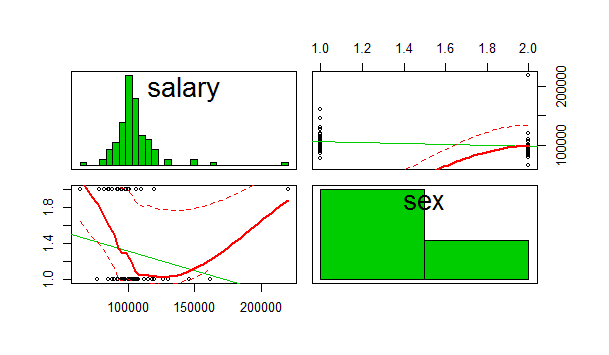


|  |
| --- |
| > boxplot(mba.df$satis,horizontal=TRUE,xlab="Degree of satisfaction with MBA program (1= low, 7 = high satisfaction) of candidates",main ="Average degree of satisfaction of candidates") |
|  |
| |  | | --- | |  | |

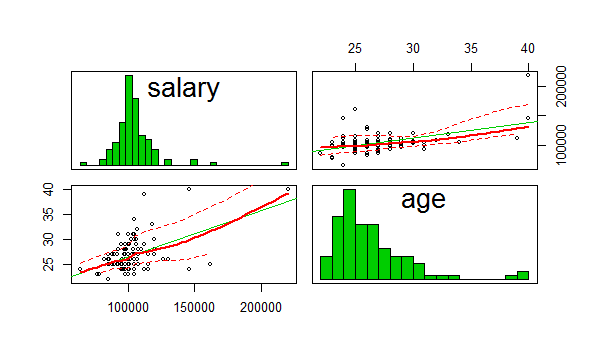
library(car)

> attach(mbap.df)

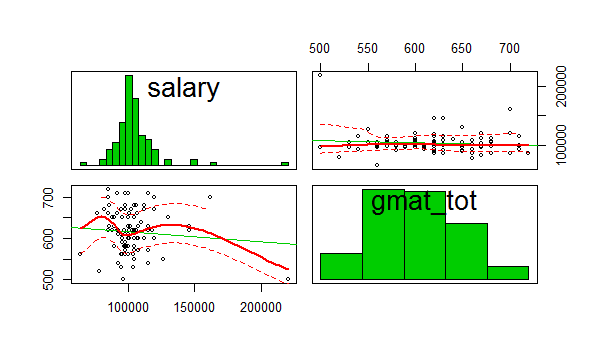
|  |
| --- |
| > scatterplotMatrix(formula=~salary+sex,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



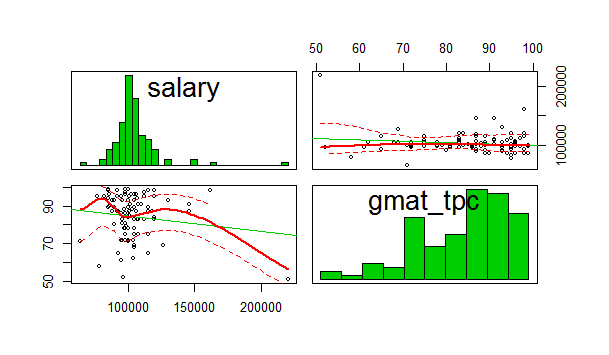
|  |
| --- |
| > scatterplotMatrix(formula=~salary+age,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



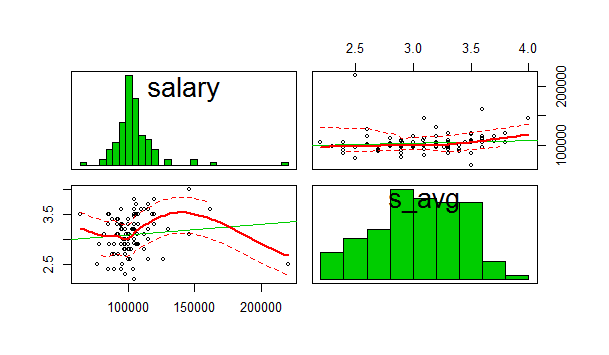
|  |
| --- |
| > scatterplotMatrix(formula=~salary+gmat\_tot,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



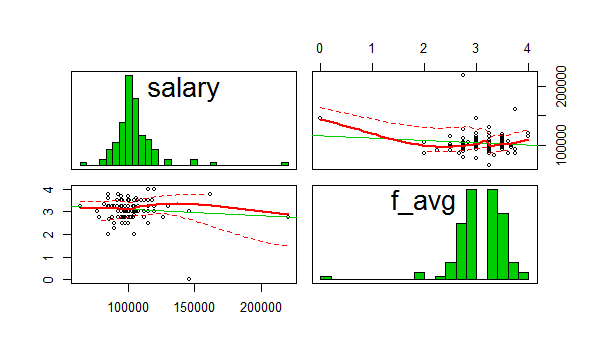
|  |
| --- |
| > scatterplotMatrix(formula=~salary+gmat\_tpc,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



|  |
| --- |
| > scatterplotMatrix(formula=~salary+s\_avg,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



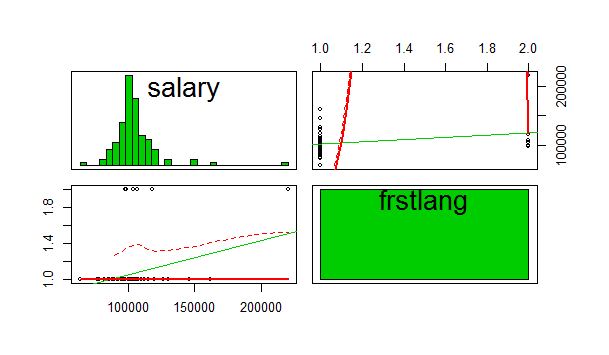
|  |
| --- |
| > scatterplotMatrix(formula=~salary+f\_avg,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



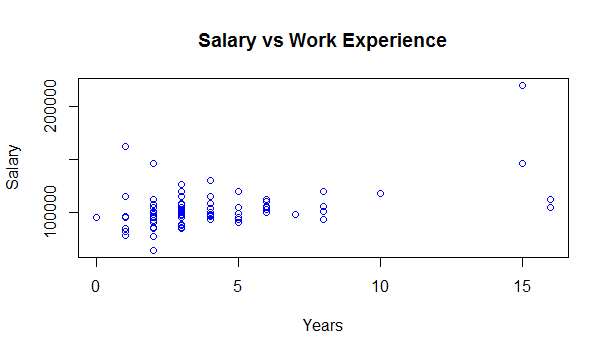
|  |
| --- |
| > scatterplotMatrix(formula=~salary+work\_yrs,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



|  |
| --- |
| > scatterplotMatrix(formula=~salary+frstlang,cex=0.6,diagonal="histogram") |
|  |
| |  | | --- | |  | |



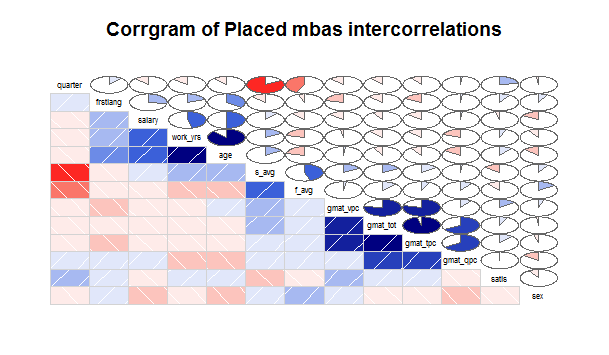
|  |
| --- |
| > plot(work\_yrs,salary, col="blue", main="Salary vs Work Experience", xlab="Years", ylab="Salary") |
|  |
| |  | | --- | |  | |

library(corrgram)

> attach(mbap.df)

corrgram(mbap.df,order=TRUE,lower.panel = panel.shade,upper.panel = panel.pie,

+ text.panel = panel.txt,main="Corrgram of Placed mbas intercorrelations")

cov(mbap.df)

age sex gmat\_tot gmat\_qpc gmat\_vpc gmat\_tpc s\_avg f\_avg

age 10.70 -2.2e-01 -1.3e+01 -7.228 9.5e-01 -3.5e+00 0.194 -3.5e-01

sex -0.22 2.1e-01 -4.6e-01 -0.908 4.0e-01 -2.4e-01 0.014 3.7e-02

gmat\_tot -13.05 -4.6e-01 2.6e+03 452.143 6.4e+02 5.4e+02 3.300 3.0e+00

gmat\_qpc -7.23 -9.1e-01 4.5e+02 179.180 2.0e+01 9.7e+01 0.078 6.4e-01

gmat\_vpc 0.95 4.0e-01 6.4e+02 20.458 2.6e+02 1.4e+02 0.969 1.8e-01

gmat\_tpc -3.46 -2.4e-01 5.4e+02 97.036 1.4e+02 1.2e+02 0.581 3.8e-01

s\_avg 0.19 1.4e-02 3.3e+00 0.078 9.7e-01 5.8e-01 0.143 8.2e-02

f\_avg -0.35 3.7e-02 3.0e+00 0.643 1.8e-01 3.8e-01 0.082 2.4e-01

quarter -0.46 -1.1e-02 -6.0e+00 0.190 -2.3e+00 -1.2e+00 -0.356 -2.4e-01

work\_yrs 8.67 -1.3e-01 -1.9e+01 -7.362 -1.4e+00 -4.4e+00 0.186 -3.2e-01

frstlang 0.29 8.8e-03 -1.7e+00 0.048 -8.9e-01 -4.6e-01 -0.013 -6.2e-03

salary 29210.52 -1.4e+03 -8.2e+04 3382.438 -4.0e+04 -2.6e+04 688.020 -9.2e+02

satis 0.28 -3.3e-02 2.6e+00 -0.042 1.9e+00 1.0e+00 -0.043 -4.5e-02

quarter work\_yrs frstlang salary satis

age -4.6e-01 8.67 2.9e-01 2.9e+04 0.278

sex -1.1e-02 -0.13 8.8e-03 -1.4e+03 -0.033

gmat\_tot -6.0e+00 -18.74 -1.7e+00 -8.2e+04 2.571

gmat\_qpc 1.9e-01 -7.36 4.8e-02 3.4e+03 -0.042

gmat\_vpc -2.3e+00 -1.37 -8.9e-01 -4.0e+04 1.880

gmat\_tpc -1.2e+00 -4.39 -4.6e-01 -2.6e+04 1.003

s\_avg -3.6e-01 0.19 -1.3e-02 6.9e+02 -0.043

f\_avg -2.4e-01 -0.32 -6.2e-03 -9.2e+02 -0.045

quarter 1.3e+00 -0.43 3.1e-02 -2.6e+03 0.198

work\_yrs -4.3e-01 9.06 1.5e-01 2.4e+04 0.149

frstlang 3.1e-02 0.15 6.4e-02 1.2e+03 0.018

salary -2.6e+03 24458.20 1.2e+03 3.2e+08 -560.658

satis 2.0e-01 0.15 1.8e-02 -5.6e+02 0.614

cor(mbap.df)

age sex gmat\_tot gmat\_qpc gmat\_vpc gmat\_tpc s\_avg f\_avg quarter

age 1.000 -0.144 -0.079 -0.165 0.018 -0.096 0.157 -0.217 -0.126

sex -0.144 1.000 -0.020 -0.147 0.053 -0.047 0.081 0.166 -0.021

gmat\_tot -0.079 -0.020 1.000 0.666 0.780 0.967 0.172 0.122 -0.106

gmat\_qpc -0.165 -0.147 0.666 1.000 0.095 0.659 0.015 0.098 0.013

gmat\_vpc 0.018 0.053 0.780 0.095 1.000 0.784 0.159 0.023 -0.129

gmat\_tpc -0.096 -0.047 0.967 0.659 0.784 1.000 0.139 0.071 -0.100

s\_avg 0.157 0.081 0.172 0.015 0.159 0.139 1.000 0.446 -0.840

f\_avg -0.217 0.166 0.122 0.098 0.023 0.071 0.446 1.000 -0.431

quarter -0.126 -0.021 -0.106 0.013 -0.129 -0.100 -0.840 -0.431 1.000

work\_yrs 0.881 -0.092 -0.123 -0.183 -0.028 -0.132 0.163 -0.216 -0.129

frstlang 0.350 0.075 -0.132 0.014 -0.218 -0.164 -0.138 -0.051 0.110

salary 0.500 -0.166 -0.091 0.014 -0.137 -0.132 0.102 -0.106 -0.128

satis 0.108 -0.092 0.065 -0.004 0.149 0.116 -0.144 -0.118 0.225

work\_yrs frstlang salary satis

age 0.881 0.350 0.500 0.108

sex -0.092 0.075 -0.166 -0.092

gmat\_tot -0.123 -0.132 -0.091 0.065

gmat\_qpc -0.183 0.014 0.014 -0.004

gmat\_vpc -0.028 -0.218 -0.137 0.149

gmat\_tpc -0.132 -0.164 -0.132 0.116

s\_avg 0.163 -0.138 0.102 -0.144

f\_avg -0.216 -0.051 -0.106 -0.118

quarter -0.129 0.110 -0.128 0.225

work\_yrs 1.000 0.196 0.455 0.063

frstlang 0.196 1.000 0.267 0.090

salary 0.455 0.267 1.000 -0.040

satis 0.063 0.090 -0.040 1.000

> cor.test(salary,work\_yrs)

Pearson's product-moment correlation

data: salary and work\_yrs

t = 5, df = 100, p-value = 1e-06

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.29 0.60

sample estimates:

cor

0.45

cor.test(salary, s\_avg)

Pearson's product-moment correlation

data: salary and s\_avg

t = 1, df = 100, p-value = 0.3

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.094 0.290

sample estimates:

cor

0.1

> cor.test(salary, frstlang)

Pearson's product-moment correlation

data: salary and frstlang

t = 3, df = 100, p-value = 0.006

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.077 0.438

sample estimates:

cor

0.27

> cor.test(salary,sex)

Pearson's product-moment correlation

data: salary and sex

t = -2, df = 100, p-value = 0.09

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.349 0.028

sample estimates:

cor

-0.17

> cor.test(salary,satis)

Pearson's product-moment correlation

data: salary and satis

t = -0.4, df = 100, p-value = 0.7

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.23 0.15

sample estimates:

cor

-0.04

job <- subset(mba.df, salary>1000, select = sex:salary)

> unemployed.sub <- subset(mba.df, salary<1, select = sex:salary)

> reg<-lm(salary~sex+gmat\_tot+gmat\_qpc+gmat\_tpc+s\_avg+f\_avg+work\_yrs+frstlang+quarter-1,data=job)

> summary(reg)

Call:

lm(formula = salary ~ sex + gmat\_tot + gmat\_qpc + gmat\_tpc +

s\_avg + f\_avg + work\_yrs + frstlang + quarter - 1, data = job)

Residuals:

Min 1Q Median 3Q Max

-30956 -7975 -464 6697 72127

Coefficients:

Estimate Std. Error t value Pr(>|t|)

sex -5095 3561 -1.43 0.15584

gmat\_tot 316 106 2.96 0.00386 \*\*

gmat\_qpc 218 164 1.32 0.18889

gmat\_tpc -1610 521 -3.09 0.00265 \*\*

s\_avg 6204 6873 0.90 0.36902

f\_avg -2605 3913 -0.67 0.50722

work\_yrs 2261 586 3.86 0.00021 \*\*\*

frstlang 14074 6521 2.16 0.03346 \*

quarter -269 2210 -0.12 0.90349

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15700 on 94 degrees of freedom

Multiple R-squared: 0.979, Adjusted R-squared: 0.977

F-statistic: 495 on 9 and 94 DF, p-value: <2e-16

> reg1<-lm(salary~gmat\_tot+gmat\_tpc+work\_yrs-1,data = job)

> summary(reg1)

Call:

lm(formula = salary ~ gmat\_tot + gmat\_tpc + work\_yrs - 1, data = job)

Residuals:

Min 1Q Median 3Q Max

-36183 -7320 -418 6366 72801

Coefficients:

Estimate Std. Error t value Pr(>|t|)

gmat\_tot 393.0 49.3 7.97 2.6e-12 \*\*\*

gmat\_tpc -1764.9 352.2 -5.01 2.3e-06 \*\*\*

work\_yrs 2715.1 532.1 5.10 1.6e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16100 on 100 degrees of freedom

Multiple R-squared: 0.977, Adjusted R-squared: 0.976

F-statistic: 1.41e+03 on 3 and 100 DF, p-value: <2e-16

> sex#Model 2 gives better r square so considering it

[1] 2 2 2 2 1 2 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1

[41] 1 1 1 1 2 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 2 2 1 1 1 1 2 2 1 1 1

[81] 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 2

> library(UsingR)

Loading required package: HistData

Loading required package: Hmisc

Loading required package: lattice

Loading required package: survival

Loading required package: Formula

Loading required package: ggplot2

Attaching package: ‘ggplot2’

The following object is masked from ‘mtcars’:

mpg

The following objects are masked from ‘package:psych’:

%+%, alpha

Attaching package: ‘Hmisc’

The following object is masked from ‘package:psych’:

describe

The following objects are masked from ‘package:base’:

format.pval, round.POSIXt, trunc.POSIXt, units

Attaching package: ‘UsingR’

The following object is masked from ‘package:survival’:

cancer

The following objects are masked from ‘package:psych’:

galton, headtail

> t.test(mba.df$salary~mba.df$sex)

Welch Two Sample t-test

data: mba.df$salary by mba.df$sex

t = -1, df = 100, p-value = 0.3

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-22412 6197

sample estimates:

mean in group 1 mean in group 2

37014 45121

> t.test(mba.df$salary~mba.df$frstlang)

Welch Two Sample t-test

data: mba.df$salary by mba.df$frstlang

t = 1, df = 40, p-value = 0.2

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-6698 34123

sample estimates:

mean in group 1 mean in group 2

40627 26915

> mytable <- xtabs(~salary, data = job)

> chisq.test(mytable)

Chi-squared test for given probabilities

data: mytable

X-squared = 100, df = 40, p-value = 1e-07

Warning message:

In chisq.test(mytable) : Chi-squared approximation may be incorrect