# Analysis of MBA SALARIES

# NAME: Praveen Kumar

# EMAIL: praveenirodov@gmail.com

# COLLEGE : MNIT Jaipur

* I would be considering significant correlation when p-value is less than 0.05 .

setwd("C:/Users/IIM")

library("psych", lib.loc="~/R/win-library/3.4")

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

View(mba)

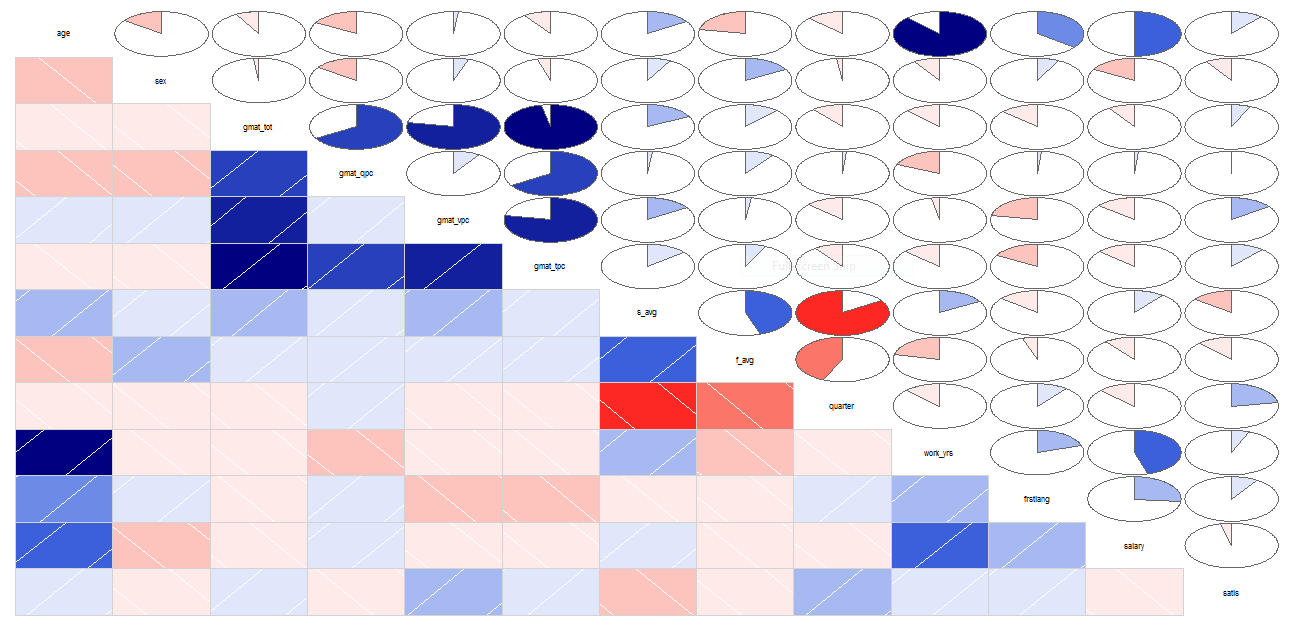
placed<-mba[which(mba$salary!=0 & mba$salary!=999 & mba$salary!=998),]

View(placed)

> library("corrgram", lib.loc="~/R/win-library/3.4")

> library("corrplot", lib.loc="~/R/win-library/3.4")

> corrgram(placed, order = TRUE, upper.panel = panel.pie)



> cor.test(placed$salary, placed$age)

Pearson's product-moment correlation

data: placed$salary and placed$age

t = 5.7968, df = 101, p-value = 7.748e-08

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.3388862 0.6320523

sample estimates:

cor

0.4996428

> mytable<-xtabs(~salary+sex, data=placed)

> mytable

sex

salary 1 2

64000 0 1

77000 1 0

78256 0 1

82000 0 1

85000 1 3

86000 0 2

88000 0 1

88500 1 0

90000 3 0

92000 2 1

93000 2 1

95000 4 3

96000 3 1

96500 1 0

97000 2 0

98000 6 4

99000 0 1

100000 4 5

100400 1 0

101000 0 2

101100 1 0

101600 1 0

102500 1 0

103000 1 0

104000 2 0

105000 11 0

106000 2 1

107000 1 0

107300 1 0

107500 1 0

108000 2 0

110000 0 1

112000 3 0

115000 5 0

118000 1 0

120000 3 1

126710 1 0

130000 1 0

145800 1 0

146000 1 0

162000 1 0

> chisq.test(mytable, simulate.p.value = TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: mytable

X-squared = 52.681, df = NA, p-value = 0.02999

> cor.test(placed$salary, placed$sex)

Pearson's product-moment correlation

data: placed$salary and placed$sex

t = -1.6948, df = 101, p-value = 0.0932

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.3485952 0.0281416

sample estimates:

cor

-0.1662887

> cor.test(placed$salary, placed$gmat\_tot)

Pearson's product-moment correlation

data: placed$salary and placed$gmat\_tot

t = -0.91501, df = 101, p-value = 0.3624

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.2792952 0.1046903

sample estimates:

cor

-0.09067141

> cor.test(placed$salary, placed$gmat\_qpc)

Pearson's product-moment correlation

data: placed$salary and placed$gmat\_qpc

t = 0.14213, df = 101, p-value = 0.8873

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.1798756 0.2070992

sample estimates:

cor

0.0141413

> cor.test(placed$salary, placed$gmat\_vpc)

Pearson's product-moment correlation

data: placed$salary and placed$gmat\_vpc

t = -1.3944, df = 101, p-value = 0.1663

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.32238270 0.05762499

sample estimates:

cor

-0.1374323

> cor.test(placed$salary, placed$gmat\_tpc)

Pearson's product-moment correlation

data: placed$salary and placed$gmat\_tpc

t = -1.3385, df = 101, p-value = 0.1837

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.31743251 0.06311945

sample estimates:

cor

-0.1320178

> cor.test(placed$salary, placed$s\_avg)

Pearson's product-moment correlation

data: placed$salary and placed$s\_avg

t = 1.0277, df = 101, p-value = 0.3065

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.09363639 0.28955576

sample estimates:

cor

0.1017317

> cor.test(placed$salary, placed$f\_avg)

Pearson's product-moment correlation

data: placed$salary and placed$f\_avg

t = -1.0717, df = 101, p-value = 0.2864

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.29353985 0.08931862

sample estimates:

cor

-0.106039

> mytable1<-xtabs(~salary+quarter, data=placed)

> chisq.test(mytable1,simulate.p.value = TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: mytable1

X-squared = 129.85, df = NA, p-value = 0.2624

> cor.test(placed$salary, placed$work\_yrs)

Pearson's product-moment correlation

data: placed$salary and placed$work\_yrs

t = 5.1303, df = 101, p-value = 1.403e-06

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.2863362 0.5957697

sample estimates:

cor

0.4546663

> mytable1<-xtabs(~salary+frstlang, data=placed)

> chisq.test(mytable1,simulate.p.value = TRUE)

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: mytable1

X-squared = 69.847, df = NA, p-value = 0.01549

> cor.test(placed$salary, placed$frstlang)

Pearson's product-moment correlation

data: placed$salary and placed$frstlang

t = 2.7846, df = 101, p-value = 0.0064

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.07749965 0.43791500

sample estimates:

cor

0.2670195

> cor.test(placed$salary,placed$satis)

Pearson's product-moment correlation

data: placed$salary and placed$satis

t = -0.40283, df = 101, p-value = 0.6879

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.2317788 0.1546729

sample estimates:

cor

-0.0400506

# NOW COMAPRING DIFFERENT REGRESSION MODELS :

> model<-lm(salary~.,data = placed)

> summary(model)

Call:

lm(formula = salary ~ ., data = placed)

Residuals:

Min 1Q Median 3Q Max

-26489 -7983 -373 5923 70602

Coefficients:

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

(Intercept) 78005.66 52981.93 1.472 0.1444

age 1750.65 1130.92 1.548 0.1251

sex -3584.07 3595.85 -0.997 0.3216

gmat\_tot 16.19 178.85 0.090 0.9281

gmat\_qpc 796.55 496.78 1.603 0.1123

gmat\_vpc 546.31 501.97 1.088 0.2794

gmat\_tpc -1457.09 714.94 -2.038 0.0445 \*

s\_avg -931.53 8240.31 -0.113 0.9102

f\_avg -2222.82 3894.57 -0.571 0.5696

quarter -2336.56 2721.89 -0.858 0.3929

work\_yrs 749.66 1135.90 0.660 0.5110

frstlang 7719.42 7373.27 1.047 0.2979

satis -1086.54 2157.76 -0.504 0.6158

---

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

Residual standard error: 15430 on 90 degrees of freedom

Multiple R-squared: 0.3422, Adjusted R-squared: 0.2545

F-statistic: 3.902 on 12 and 90 DF, p-value: 8.086e-05

> model1<-lm(salary~.-satis-quarter,data = placed) #As these failed the chi

square test

> summary(model1)

Call:

lm(formula = salary ~ . - satis - quarter, data = placed)

Residuals:

Min 1Q Median 3Q Max

-30627 -8168 -767 5445 70245

Coefficients:

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

(Intercept) 53005.0682 47814.1043 1.109 0.2705

age 1702.9961 1124.8254 1.514 0.1335

sex -3781.6672 3551.3887 -1.065 0.2897

gmat\_tot -0.2345 168.0014 -0.001 0.9989

gmat\_qpc 830.0941 488.3993 1.700 0.0926 .

gmat\_vpc 579.0159 488.6739 1.185 0.2391

gmat\_tpc -1465.3294 705.6349 -2.077 0.0406 \*

s\_avg 5117.8754 4987.7232 1.026 0.3075

f\_avg -1693.3454 3815.2794 -0.444 0.6582

work\_yrs 775.6177 1131.4164 0.686 0.4947

frstlang 7704.4426 7289.3746 1.057 0.2933

---

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

Residual standard error: 15370 on 92 degrees of freedom

Multiple R-squared: 0.3326, Adjusted R-squared: 0.26

F-statistic: 4.585 on 10 and 92 DF, p-value: 2.812e-05

> model2<-lm(salary~.-satis-quarter-sex,data = placed)

> summary(model2)

Call:

lm(formula = salary ~ . - satis - quarter - sex, data = placed)

Residuals:

Min 1Q Median 3Q Max

-32053 -8335 518 5365 67353

Coefficients:

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

(Intercept) 49095.92 47707.27 1.029 0.3061

age 1959.37 1099.55 1.782 0.0780 .

gmat\_tot -19.44 167.15 -0.116 0.9077

gmat\_qpc 892.41 485.23 1.839 0.0691 .

gmat\_vpc 601.11 488.58 1.230 0.2217

gmat\_tpc -1449.30 705.98 -2.053 0.0429 \*

s\_avg 4832.55 4984.10 0.970 0.3348

f\_avg -2020.71 3805.61 -0.531 0.5967

work\_yrs 625.70 1123.43 0.557 0.5789

frstlang 6153.44 7147.51 0.861 0.3915

---

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

Residual standard error: 15380 on 93 degrees of freedom

Multiple R-squared: 0.3244, Adjusted R-squared: 0.259

F-statistic: 4.961 on 9 and 93 DF, p-value: 1.909e-05

> model3<-lm(salary~.-satis-quarter-gmat\_tot,data = placed)

> summary(model3)

Call:

lm(formula = salary ~ . - satis - quarter - gmat\_tot, data = placed)

Residuals:

Min 1Q Median 3Q Max

-30627 -8166 -767 5446 70244

Coefficients:

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

(Intercept) 52953.1 29850.0 1.774 0.0793 .

age 1703.0 1118.8 1.522 0.1313

sex -3782.2 3511.8 -1.077 0.2843

gmat\_qpc 829.6 356.7 2.326 0.0222 \*

gmat\_vpc 578.6 357.9 1.616 0.1094

gmat\_tpc -1465.5 694.6 -2.110 0.0376 \*

s\_avg 5116.9 4908.0 1.043 0.2999

f\_avg -1694.1 3752.9 -0.451 0.6527

work\_yrs 775.6 1125.1 0.689 0.4923

frstlang 7703.4 7210.1 1.068 0.2881

---

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

Residual standard error: 15290 on 93 degrees of freedom

Multiple R-squared: 0.3326, Adjusted R-squared: 0.268

F-statistic: 5.149 on 9 and 93 DF, p-value: 1.172e-05

> model4<-lm(salary~.-satis-quarter-gmat\_tot-gmat\_qpc,data = placed)

> summary(model4)

Call:

lm(formula = salary ~ . - satis - quarter - gmat\_tot - gmat\_qpc,

data = placed)

Residuals:

Min 1Q Median 3Q Max

-30150 -8790 -2115 5222 79170

Coefficients:

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

(Intercept) 46510.92 30410.53 1.529 0.130

age 1816.28 1143.62 1.588 0.116

sex -4307.34 3585.84 -1.201 0.233

gmat\_vpc -164.97 164.83 -1.001 0.319

gmat\_tpc 57.86 236.82 0.244 0.808

s\_avg 4097.43 5001.77 0.819 0.415

f\_avg -862.16 3822.51 -0.226 0.822

work\_yrs 646.88 1149.83 0.563 0.575

frstlang 8589.77 7367.02 1.166 0.247

Residual standard error: 15640 on 94 degrees of freedom

Multiple R-squared: 0.2938, Adjusted R-squared: 0.2337

F-statistic: 4.887 on 8 and 94 DF, p-value: 4.609e-05

> model5<-lm(salary~.-satis-quarter-gmat\_tot-gmat\_vpc,data = placed)

> summary(model5)

Call:

lm(formula = salary ~ . - satis - quarter - gmat\_tot - gmat\_vpc,

data = placed)

Residuals:

Min 1Q Median 3Q Max

-28765 -9088 -1092 4900 75914

Coefficients:

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

(Intercept) 43718.0 29548.3 1.480 0.1423

age 1942.0 1118.4 1.736 0.0858 .

sex -3548.2 3538.8 -1.003 0.3186

gmat\_qpc 314.8 161.9 1.945 0.0548 .

gmat\_tpc -387.5 195.8 -1.980 0.0507 .

s\_avg 4531.7 4936.4 0.918 0.3610

f\_avg -1314.8 3777.6 -0.348 0.7286

work\_yrs 604.7 1129.7 0.535 0.5937

frstlang 6937.2 7256.0 0.956 0.3415

---

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

Residual standard error: 15420 on 94 degrees of freedom

Multiple R-squared: 0.3138, Adjusted R-squared: 0.2554

F-statistic: 5.374 on 8 and 94 DF, p-value: 1.428e-05

> model6<-lm(salary~.-satis-quarter-gmat\_tot-gmat\_tpc,data = placed)

> summary(model6)

Call:

lm(formula = salary ~ . - satis - quarter - gmat\_tot - gmat\_tpc,

data = placed)

Residuals:

Min 1Q Median 3Q Max

-28098 -8994 -1771 4642 79218

Coefficients:

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

(Intercept) 39256.4 29665.6 1.323 0.1889

age 1890.4 1135.5 1.665 0.0993 .

sex -3737.4 3575.7 -1.045 0.2986

gmat\_qpc 120.1 121.0 0.993 0.3234

gmat\_vpc -146.6 101.8 -1.439 0.1535

s\_avg 4006.6 4968.5 0.806 0.4220

f\_avg -1055.5 3808.8 -0.277 0.7823

work\_yrs 663.6 1144.3 0.580 0.5634

frstlang 7851.5 7340.9 1.070 0.2876

---

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

Residual standard error: 15570 on 94 degrees of freedom

Multiple R-squared: 0.3006, Adjusted R-squared: 0.2411

F-statistic: 5.051 on 8 and 94 DF, p-value: 3.102e-05

> model7<-lm(salary~.-satis-quarter-gmat\_tot-f\_avg,data = placed)

> summary(model7)

Call:

lm(formula = salary ~ . - satis - quarter - gmat\_tot - f\_avg,

data = placed)

Residuals:

Min 1Q Median 3Q Max

-30193 -8078 -1195 5213 69687

Coefficients:

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

(Intercept) 50129.3 29063.3 1.725 0.0878 .

age 1753.3 1108.5 1.582 0.1171

sex -3937.7 3480.1 -1.132 0.2607

gmat\_qpc 814.3 353.5 2.303 0.0235 \*

gmat\_vpc 568.4 355.7 1.598 0.1134

gmat\_tpc -1440.2 689.4 -2.089 0.0394 \*

s\_avg 3990.8 4208.9 0.948 0.3455

work\_yrs 811.0 1117.6 0.726 0.4698

frstlang 7398.5 7148.0 1.035 0.3033

---

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

Residual standard error: 15220 on 94 degrees of freedom

Multiple R-squared: 0.3311, Adjusted R-squared: 0.2742

F-statistic: 5.817 on 8 and 94 DF, p-value: 5.003e-06

> model8<-lm(salary~.-satis-quarter-gmat\_tot-f\_avg-s\_avg,data = placed)

> summary(model8)

Call:

lm(formula = salary ~ . - satis - quarter - gmat\_tot - f\_avg -

s\_avg, data = placed)

Residuals:

Min 1Q Median 3Q Max

-28358 -8231 -151 6073 67397

Coefficients:

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

(Intercept) 58275.1 27749.9 2.100 0.0384 \*

age 1875.9 1100.3 1.705 0.0915 .

sex -3474.3 3443.7 -1.009 0.3156

gmat\_qpc 798.3 352.9 2.262 0.0260 \*

gmat\_vpc 552.0 355.1 1.554 0.1234

gmat\_tpc -1389.8 687.0 -2.023 0.0459 \*

work\_yrs 812.6 1117.0 0.727 0.4687

frstlang 6095.3 7010.8 0.869 0.3868

---

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

Residual standard error: 15220 on 95 degrees of freedom

Multiple R-squared: 0.3247, ***Adjusted R-squared: 0.275***

F-statistic: 6.526 on 7 and 95 DF, p-value: 2.725e-06