**Data Analytics based on Salary Dataset**

**Summary data about Salary data set**

> summary(Salaries)

X rank discipline yrs.since.phd yrs.service

Min. : 1 AssocProf: 64 A:181 Min. : 1.00 Min. : 0.00

1st Qu.:100 AsstProf : 67 B:216 1st Qu.:12.00 1st Qu.: 7.00

Median :199 Prof :266 Median :21.00 Median :16.00

Mean :199 Mean :22.31 Mean :17.61

3rd Qu.:298 3rd Qu.:32.00 3rd Qu.:27.00

Max. :397 Max. :56.00 Max. :60.00

sex salary

Female: 39 Min. : 57800

Male :358 1st Qu.: 91000

Median :107300

Mean :113706

3rd Qu.:134185

Max. :231545

As seen above lecturers’ ***years since PhD*** minimum values and maximum values found as 1 and 56 respectively. The first and third quartile values of the series can be identified as **12** and **32** respectively. The median and mean values of the series identifiable as **21** and **22.31** respectively. According to the analysis of the data it is confirmed that there are higher numbers of well experienced lecturers available for the university who are mastered and pioneered in specific and specialized subject domains. This is a good indicator of quality of academia of the university.

According to the above statistics ***years of service*** of the lecturers, minimum values and maximum values found as **0** and **60** respectively. The first and third quartile values of the series can be identified as **7** and **27** respectively. The median and mean values of the series identifiable as **16** and **17.61** respectively. The results indicates the university provide opportunities for young academia to involve with the faculty while giving them the confidence of working with the university for a long time period. The results further reveals the university always attempt to keep highly experienced faculty having more than a decade of academic exposure with them to deliver best and quality education for students.

By considering the above results, the minimum and maximum annual ***salaries*** of the lecturers found as $ **57800** and $ **231545** respectively. The first and third quartile values of the series can be identified as $ **91000** and $ **134185** respectively. The median and mean values of the series identifiable as $ **107300** and $ **113706** respectively. The statistics reveals that the university offer better remuneration packages for their well experienced academic panel since better the academia served the better the results university get in terms of learning satisfaction of the students, academic quality, academic excellence in projects ,number of excellent researches and research publications done annually, reputation among other universities locally and globally etc. .

1. **Statistical Hypothetical Testing**: **Normality tests for Samples : Salary Dataset : Professors’ Rank**

* Significant level (α) = **0.05**
* Confidence level =**95%**

(\*) Whether or not professor **ranks** are normally distributed?

H­0: Professor **rank** values are normally distributed

H­1: Professor **rank** values are not normally distributed

> ad.test(rank)

Anderson-Darling normality test

data: Salaries\_new$rank

A = 66.705, p-value < 2.2e-16

**Decision: p-value=2.2e-16< α=0.05 => Reject H0 at 5% significant level**

> lillie.test(rank)

Lilliefors (Kolmogorov-Smirnov) normality test

data: Salaries\_new$rank

D = 0.41213, p-value < 2.2e-16

**Decision: p-value<2.2e-16< α=0.05 => Reject H0 at 5% significant level**

|  |
| --- |
| > shapiro.test(rank)  Shapiro-Wilk normality test  data: Salaries\_new$rank  W = 0.64149, p-value < 2.2e-16 |
|  |
| |  | | --- | | **Decision: p-value=2.2e-16< α=0.05 => Reject H0 at 5% significant level** | |

**Normality Test Conclusion**: According to the **Anderson darling test**, **Lillie test** & **Shapiro test** there are **no enough** evidence to claim that professors’ **Rank** values are normally distributed at a 5% significance level.

Graphical Analysis of Salary Data

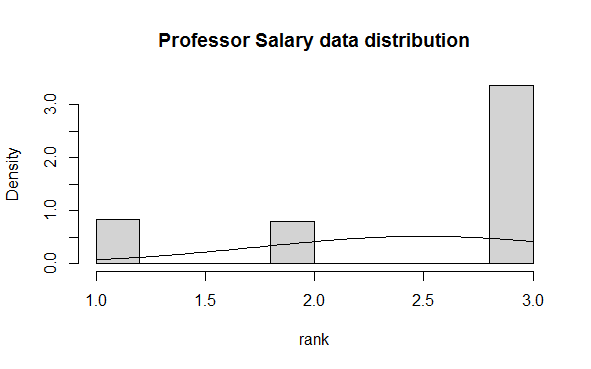


Figure 2: Monte Carlo graphical simulations -Rank

Interpretation of the Monte Carlo graph

As seen on (Figure 2) there is **no** **clear symmetrically distributed graph (bell curve)** is identifiable. Therefore the rank values can be considered as **not normally** distributed at 5% significance level.

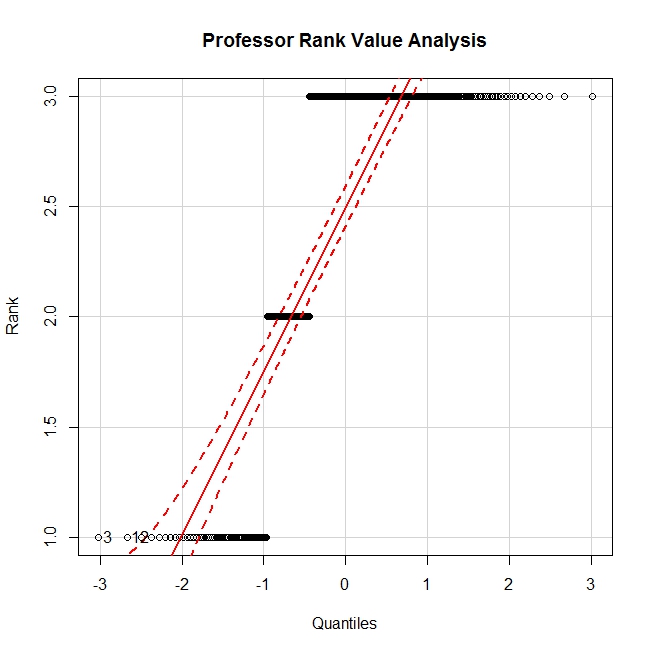


Figure 3: Quantile Comparison -Rank

Interpretation of the Quantile Comparison Graph

As seen on (Figure 3) the majority of rank values are **not within the envelope** of the quantile comparison graph. Therefore it can be concluded that professors’ rank values are not normally distributed at a 5% significance level.

1. **Statistical Hypothetical Testing**: **Normality tests for Samples ; Dataset : Salaries**

If **α – value** > **P-Value** 🡪 Reject H0

If **α – value** <  **P-Value** 🡪 Accept H0

* Significant level (α) = **0.05 (5%)**
* Confidence level =**0.95(95%)**
* Variable: salary

(\*) **Whether or not professor *salaries* are normally distributed?**

H­0: Professors’ **salaries** are normally distributed

H­1: Professors’ **salaries** are not normally distributed

> ad.test(salary)

Anderson-Darling normality test

data: salary

A = 4.1472, p-value = 2.508e-10

**Decision: p-value=** 2.508e-10**< α=0.05 => Reject H0 at 5% significant level**

> lillie.test(salary)

Lilliefors (Kolmogorov-Smirnov) normality

test

data: salary

D = 0.09086, p-value = 2.354e-08

**Decision: p-value=**2.354e-08**< α=0.05 => Reject H0 at 5% significant level**

> shapiro.test(salary)

Shapiro-Wilk normality test

data: salary

W = 0.95988, p-value = 6.076e-09

**Decision: p-value=**6.076e-09**< α=0.05 => Reject H0 at 5% significant level**

**Normality Test Conclusion**: According to the **Anderson darling test(p=**2.508e-10**)**, **Lillie test(p=**2.354e-08**)** and **Shapiro test(p=**6.076e-09**)** there are **no enough** evidences to claim that professors’ salaries are normally distributed **at a 5%** **significance level.**

**Hint:** Summary stat

Salary

Min. : 57800

1st Qu.: 91000

Median :107300

Mean :113706

3rd Qu.:134185

Max. :231545

Graphical Analysis of Salary Data

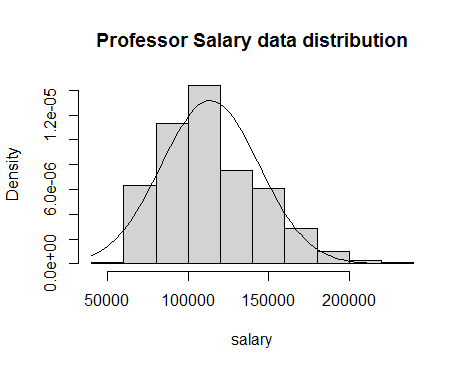


Figure 1: Monte Carlo graphical simulations of professors’ Salary

Interpretation of the Monte Carlo graph

As seen on (Figure 1) there is a slight **right skewedness** of data values identifiable. Therefore the salary data can be considered as **not normally** distributed at 5% significance level.

Graphical Analysis of Salary Data

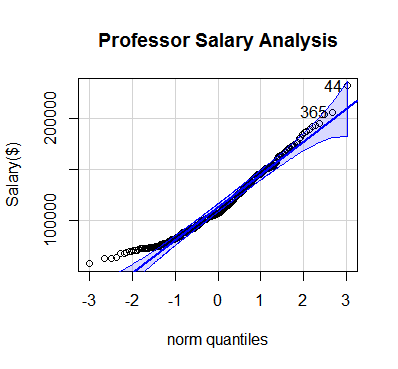


Figure 2: Quintiles Comparison Test

As seen on (Figure 2.0) the quintile comparison test indicate that considerable numbers of salary values are not within the envelope of the quantile comparison graph as well as not closer by located or overlapped on the absolute line. Therefore it can be concluded that professors’ salary data are not normally distributed at a 5% significance level.

1. **Statistical Hypothetical Testing**: **Normality tests for Samples : Salary Dataset : Professors’ years since PhD**

If α – value > **P-Value** 🡪 Reject H0

If α – value <  **P-Value** 🡪 Accept H0

* Significant level (α) = **0.05**
* Confidence level =**95%**

(\*) Whether or not professor years since PhD values are normally distributed?

H­0: Professors’ *Years since PhD* values are normally distributed

H­1: Professors’ *Years since PhD* values are not normally distributed

> ad.test(yrs.since.phd)

Anderson-Darling normality test

data: yrs.since.phd

A = 2.995, p-value = 1.565e-07

**Decision: p-value=**1.565e-07**< α=0.05 => Reject H0 at 5% significant level**

> lillie.test(yrs.since.phd)

Lilliefors (Kolmogorov-Smirnov) normality test

data: yrs.since.phd

D = 0.069888, p-value = 8.011e-05

**Decision: p-value=**8.011e-05**< α=0.05 => Reject H0 at 5% significant level**

> shapiro.test(yrs.since.phd)

Shapiro-Wilk normality test

data: yrs.since.phd

W = 0.96957, p-value = 2.328e-07

**Decision: p-value=**2.328e-07**< α=0.05 => Reject H0 at 5% significant level**

**Normality Test Conclusion**: According to the **Anderson darling test(p=**1.565e-07), **Lillie test(p-**8.011e-05)and **Shapiro test(p-**2.328e-07) there are **no enough** evidences to claim those *professors’ years* *since PhD values* are normally distributed **at a 5%** **significance level.**

**Hint:** Summary stat

yrs.since.phd

Min. : 1.00

1st Qu.: 12.00

Median : 21.00

Mean : 22.31

3rd Qu.: 32.00

Max. : 56.00

Graphical Analysis

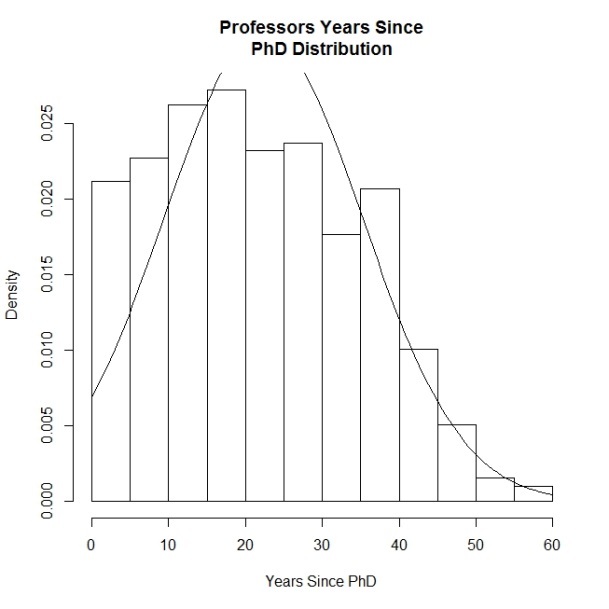


Figure 2: Monte Carlo graphical simulations of professors’ years since PhD

As seen on (Figure 2) there is a slight **right skewedness** of the bell curve identifiable. Therefore, the professors’ years since PhD values can be considered as **not normally** distributed at 5% significance level.

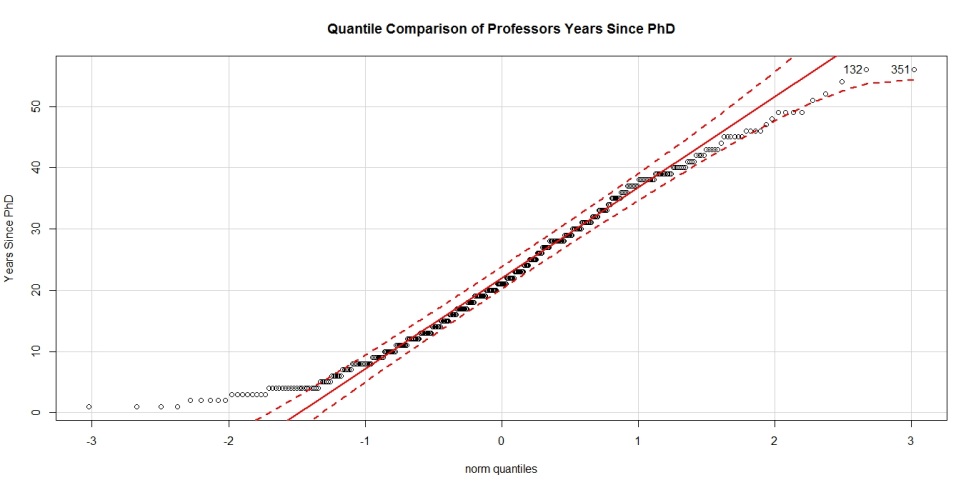


Figure 3: Quantile comparison graph of professors’ years since PhD

As seen on (Figure 3) the quantile comparison test indicate that considerable number of years since PhD values are not within the envelope of the graph. Therefore it can be concluded that professors’ years since PhD values are not normal at a 5% significance level.

1. **Statistical Hypothetical Testing**: **Normality tests for Samples : Salary Dataset : Professors’ years of service**

* Significant level (α) = **0.05**
* Confidence level =**95%**

(\*) Whether or not professors’ **years of service** values are normally distributed?

H­0: Professors’ **years of service** values are normally distributed

H­1: Professors’ **years of service** values are not normally distributed

> ad.test(yrs.service)

Anderson-Darling normality test

data: yrs.service

A = 6.3468, p-value = 1.325e-15

**Decision: p-value=**1.325e-15**< α=0.05 => Reject H0 at 5% significant level**

> lillie.test(yrs.service)

Lilliefors (Kolmogorov-Smirnov) normality test

data: yrs.service

D = 0.12017, p-value = 4.513e-15

**Decision: p-value=**4.513e-15**< α=0.05 => Reject H0 at 5% significant level**

> shapiro.test(yrs.service)

Shapiro-Wilk normality test

data: yrs.service

W = 0.94183, p-value = 2.337e-11

**Decision: p-value=**2.337e-11**< α=0.05 => Reject H0 at 5% significant level**

**Normality Test Conclusion**: According to the **Anderson darling test(p=**1.325e-15**)**, **Lillie test(**p=4.513e-15**)** and **Shapiro test(p=**2.337e-11**)** there are **no enough** evidences to claim those professors’ years of service values are normally distributed **at a 5%** **significance level.**

Graphical Analysis

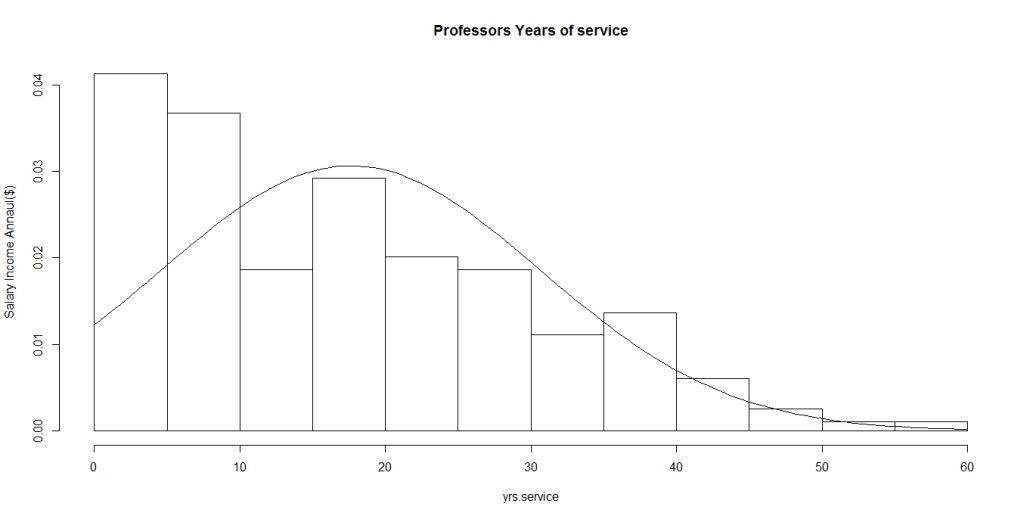


Figure 3: Monte Carlo graphical simulations of professors’ years of service

As seen on (Figure 3) there is a clear **right skewedness** of the bell curve identifiable. Therefore, the professors’ years of experience values can be considered as **not normally** distributed at 5% significance.

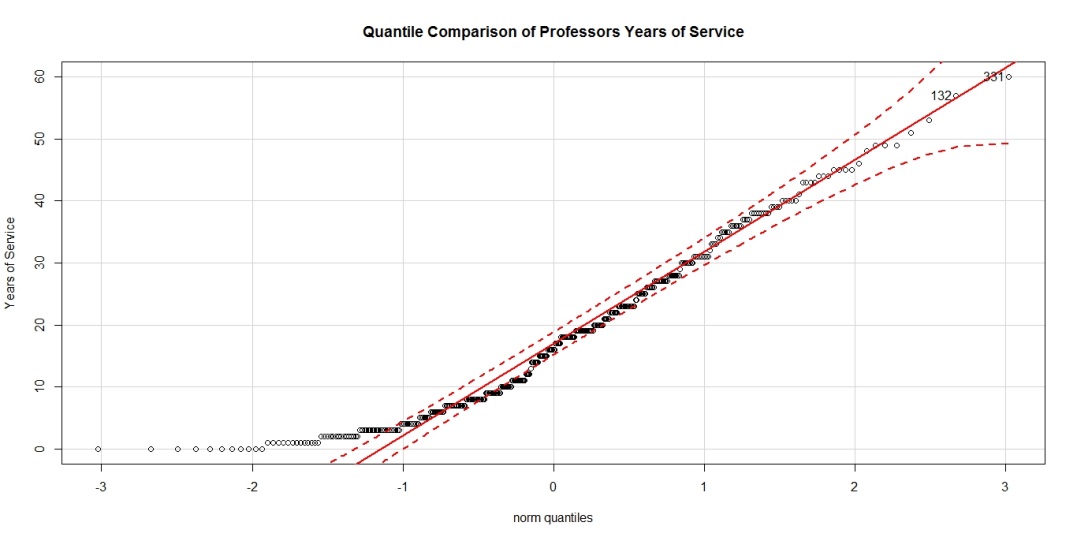


Figure 4: Quantile comparison graph of professors’ years since PhD

As seen on (Figure 3) the quantile comparison test indicate that considerable number of years of experience values are not within the envelope of the graph. Therefore, it can be concluded that professors’ years of experience values are not normally distributed at a 5% significance level.

1. **Statistical Hypothetical Testing**: **Bartlett’s Variance Test For** Samples**: Salary Dataset**

Sample fields**: *Rank*** and ***Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

(\*)Professor **salary** variancesare similar or not basedon **rank?**

H0: Ϭ2 professor salary = Ϭ2Associateprofessor salary= Ϭ2Assistantprofessor salary

H1: Ϭ2 professor salary ≠ Ϭ2Associateprofessor salary≠ Ϭ2Assistantprofessor salary

Rcmdr> with(Salaries, tapply(salary, rank, var, na.rm=TRUE))

AssocProf AsstProf Prof

191315921 66816117 768324944

Rcmdr> bartlett.test(salary ~ rank, data=Salaries)

Bartlett test of homogeneity of variances

data: salary by rank

Bartlett's K-squared = 122.49, df = 2, p-value < 2.2e-16

**Decision: p-value=2.2e-16< α=0.05 => Reject H0 at 5% significant level**

**Variance Test Conclusion**: According to the **Bartlett’s variance test** there is **no enough** evidence to claim that **variances of salaries** on ***professor’s ranks*** are equal at a 5% significance level. Therefore, based on the results following modified H1 can be accepted at a 5% significance level.

H1: Ϭ2 professor salary > Ϭ2Associateprofessor salary> Ϭ2Assistantprofessor salary

……………………………………………………………………………………………………………………………………………………………..

Alternative:

**Statistical Hypothetical Testing**: **One-Way ANOVA Test For** Samples**: Salary Dataset**

Sample fields**: *Rank and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*)**Professor ***salary*** variances are similar or not based on ***rank*?**

H0: Ϭ2 professor salary = Ϭ2Associateprofessor salary= Ϭ2Assistantprofessor salary

H1: Ϭ2 professor salary ≠ Ϭ2Associateprofessor salary≠ Ϭ2Assistantprofessor salary

Rcmdr> summary(AnovaModel.1)

Df Sum Sq Mean Sq F value Pr(>F)

rank 2 1.432e+11 7.162e+10 128.2 <2e-16 \*\*\*

Residuals 394 2.201e+11 5.586e+08

---

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

Rcmdr> with(Salaries, numSummary(salary, groups=rank, statistics=c("mean",

Assoc.

Asst.

Rcmdr+ "sd")))

mean sd data:n

Prof.

AssocProf 93876.44 13831.700 64

AsstProf 80775.99 8174.113 67

Prof 126772.11 27718.675 266

**Decision: p-value=2.2e-16< α=0.05 => Reject H0 at 5% significant level**

**One-Way ANOVA Test Conclusion**: According to the **ANOVA test (**p=2e-16)there is **no enough** evidence to claim that **variances of salaries** on ***professor’s ranks*** are equal at a 5% significance level.

Therefore, based on the results the following modified H1 can be accepted at a 5% significance level.

H1: Ϭ2 professor salary > Ϭ2Associateprofessor salary> Ϭ2Assistantprofessor salary

Graphical Analysis of Salary Data based on variance values

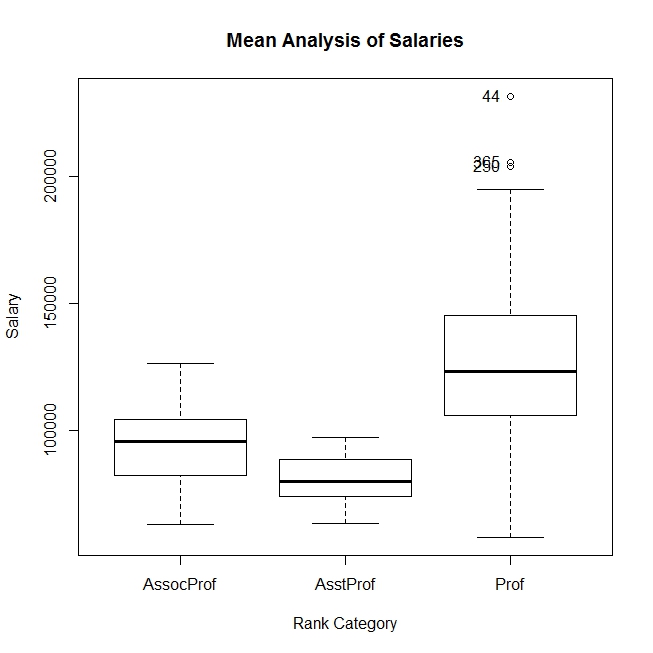


Figure 6: Boxplot Analysis-Professor Salary

Graph interpretation

According to the boxplot graph as seen on (Figure 6) it is clear that professors’ variances of salaries are different based on their respective ranks at a 5% significance level. The highest salary variance found for professors whereas least salary variance found for Assistant professors and associate professors found under 2nd highest salary variance category at a 5% significance level.

**Statistical Hypothetical Testing**: **Bartlett’s Variance Test For** Samples**: Salary Dataset**

Sample fields**: *Rank and Salary***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**Professor salary variances are similar or not based on rank?**

H0: Ϭ2 professor salary = Ϭ2Associateprofessor salary= Ϭ2Assistantprofessor salary

H1: Ϭ2 professor salary ≠ Ϭ2Associateprofessor salary≠ Ϭ2Assistantprofessor salary

Rcmdr> bartlett.test(salary ~ rank, data=Salaries)

Bartlett test of homogeneity of variances

data: salary by rank

Bartlett's K-squared = 122.49, df = 2, p-value < 2.2e-16

Rcmdr> with(Salaries, tapply(salary, rank, var, na.rm=TRUE))

AssocProf AsstProf Prof

191315921 66816117 768324944

**Decision: p-value=2.2e-16< α=0.05 => Reject H0 at 5% significant level**

**Bartlett’s Variance Test Conclusion**: According to the **Bartlett’s Variance test** (p=2.2e-16)there is **no enough** evidence to claim that **variances of salaries** on ***professors’ ranks*** are equal at a 5% significance level.

Therefore based on the results following modified H1 can be accepted at a 5% significance level.

H1: Ϭ2 professor salary > Ϭ2Associateprofessor salary> Ϭ2Assistantprofessor salary

**Statistical Hypothetical Testing**: **Levene’s Variance Test For** Samples**: Salary Dataset**

Sample fields**: *Rank*** *and* ***Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

(\*)Professor **salary** variances are similar or not based on **rank?**

H0: Ϭ2 professor salary = Ϭ2Associateprofessor salary= Ϭ2Assistantprofessor salary

H1: Ϭ2 professor salary ≠ Ϭ2Associateprofessor salary≠ Ϭ2Assistantprofessor salary

Rcmdr> leveneTest(salary ~ rank, data=Salaries, center="median")

Levene's Test for Homogeneity of Variance (center = "median")

Df F value Pr(>F)

group 2 38.711 4.477e-16 \*\*\*

394

---

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

Rcmdr> with(Salaries, tapply(salary, rank, var, na.rm=TRUE))

AssocProf AsstProf Prof

191315921 66816117 768324944

**Decision: p-value=2.2e-16< α=0.05 => Reject H0 at 5% significant level**

**Levene’s Variance Test Conclusion**: According to the **Levene’s Variance test** there is **no enough** evidence to claim that **variances of salaries** on ***professors’ ranks*** are equal at a 5% significance level.

Therefore, based on the results the following modified H1 can be accepted at a 5% significance level.

H1: Ϭ2 professor salary > Ϭ2Associateprofessor salary> Ϭ2Assistantprofessor salary

1. **Statistical Hypothetical MEAN** Testing: ***One-Way ANOVA test*** for Samples**: Salary dataset**

Sample field**: *Professors’ rank and Professors’ salary***

Assoc.

Asst.

* Significant level (α) = **0.05**
* Confidence level =**95%**

Prof.

(\*) **Whether or not professors’ salary mean values are equal based on rank**?

H0: µ professor salary = µ Associateprofessor Salary= µ Assistantprofessor Salary

H1: µ professor salary ≠ µ Associateprofessor Salary≠ µ Assistantprofessor Salary

Asst.

Rcmdr> AnovaSalRankModel.1 <- aov(salary ~ rank, data=Salaries)

Rcmdr> summary(AnovaSalRankModel.1)

Assoc.

Df Sum Sq Mean Sq F value Pr(>F)

rank 2 1.432e+11 7.162e+10 128.2 <2e-16 \*\*\*

Prof.

Residuals 394 2.201e+11 5.586e+08

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

Rcmdr> with(Salaries, numSummary(salary, groups=rank, statistics=c("mean", "sd")))

mean sd data:n

AssocProf 93876.44 13831.700 64

AsstProf 80775.99 8174.113 67

Prof 126772.11 27718.675 266

**Decision: p-value=**2e-16**< α=0.05 => Reject H0 at 5% significant level**

**Mean Analysis Conclusion**:

According to the **one-Way ANOVA test (p**<2e-16) there is **no** **statistical** evidence to claim that **Mean values of Salaries** on ***Professors’ ranks*** are same at a 5% significance level.

Therefore H0: µ professor salary = µ Associateprofessor Salary= µ Assistantprofessor Salary is **not accepted** at a significant level of 5% .After analyzing aforementioned statistical results it can be confirmed that Mean value of ***Professors’ salaries*** *on* ***Professors’ ranks*** are not equal.

Therefore, below modified H1 is accepted.

H1: µ professor Salary > µ Associateprofessor Salary > µ Assistantprofessor Salary

Graphical Analysis of Salary based on Ranks

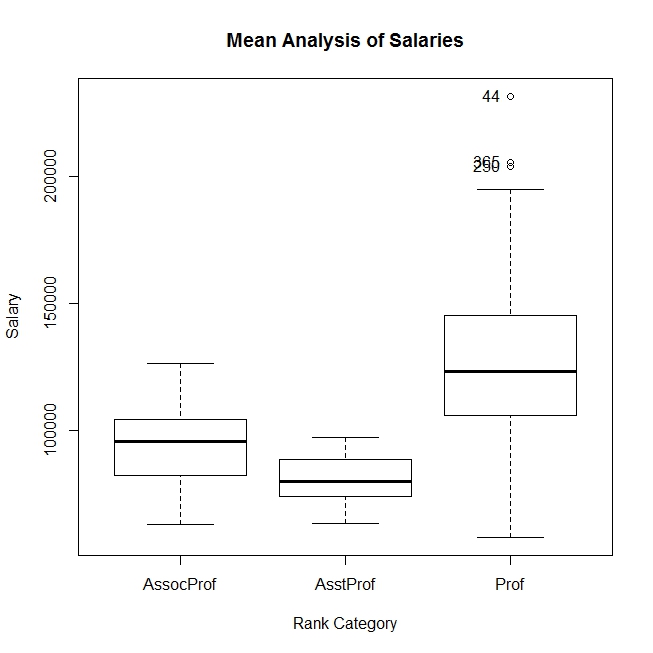


Figure 6: Boxplot Analysis-Professor Salary

Graph interpretation

According to the boxplot graph as seen on (Figure 6) it is clear that professors’ mean salaries are different based on their respective ranks at a 5% significance level. The highest salary mean found for professors’ whereas least salary means found for assistant professors’ and associate professors’ found as 2nd highest salary mean at a 5% significance level. More professors exists in the upper mean range of the graph where more associate professors available in the lower range of the average salary.

As selection of professors for the universities based their research and development involvement, experience, contribution and involvement with the industry, financial benefits are changed. As an example in USA most of the university professors need to bring new projects to the university in collaboration with industry leading companies (Nawarathne, 2018).

1. **Statistical Hypothetical Variance Testing**: **One way ANOVA - Test For** Samples**: Salary Dataset**

Sample fields**: *Discipline and Salary***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors salary variances are similar or not based on discipline?**

H0: Ϭ2 A salary = Ϭ2B salary

H1: Ϭ2 A salary ≠ Ϭ2B salary

Rcmdr> summary(AnovaModel.5)

Df Sum Sq Mean Sq F value Pr(>F)

discipline 1 8.851e+09 8.851e+09 9.863 0.00181 \*\*

Residuals 395 3.544e+11 8.973e+08

---

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

Rcmdr> with(Salaries, numSummary(salary, groups=discipline, statistics=c("mean", "sd")))

mean sd data:n

A 108548.4 30538.15 181

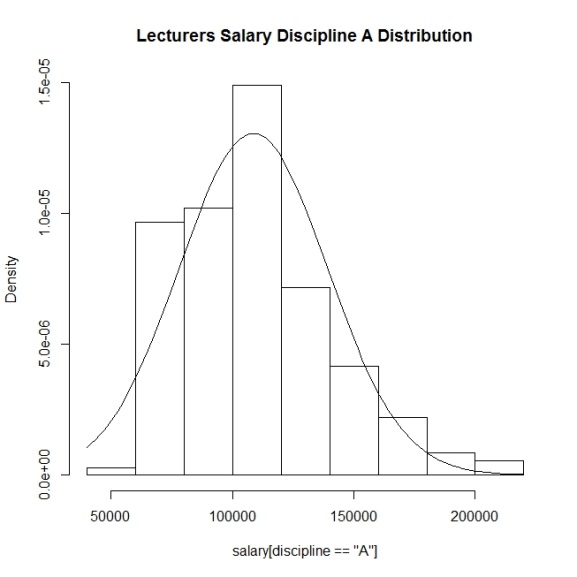
B 118028.7 29459.14 216

**Decision: p-value=**0.00181 **< α=0.05 => Reject H0 at 5% significant level**

**One-Way ANOVA Test Conclusion**: According to the **One-Way ANOVA test (p=**0.00181)there is **no enough** evidence to claim that **variances of salaries** on ***discipline*** are equal at a 5% significance level.

Therefore based on the results following modified H1 can be accepted at a 5% significance level.

H1: Ϭ2 A salary > Ϭ2B salary

Graphical Analysis

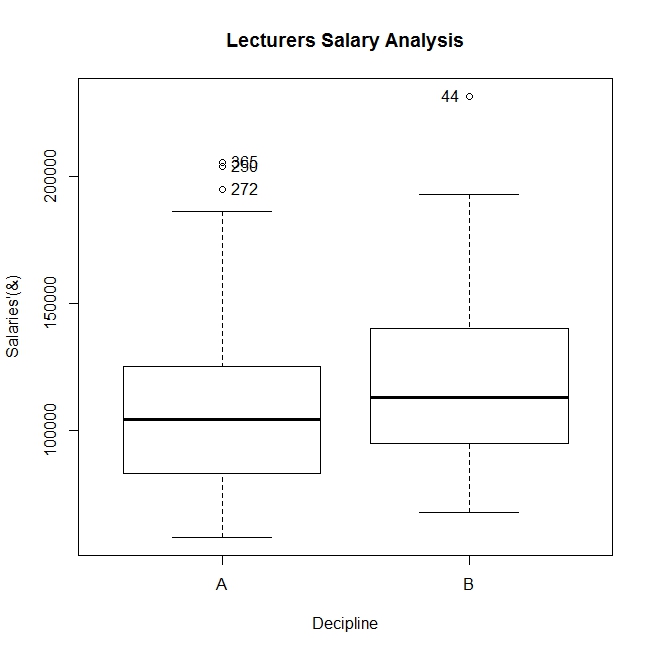
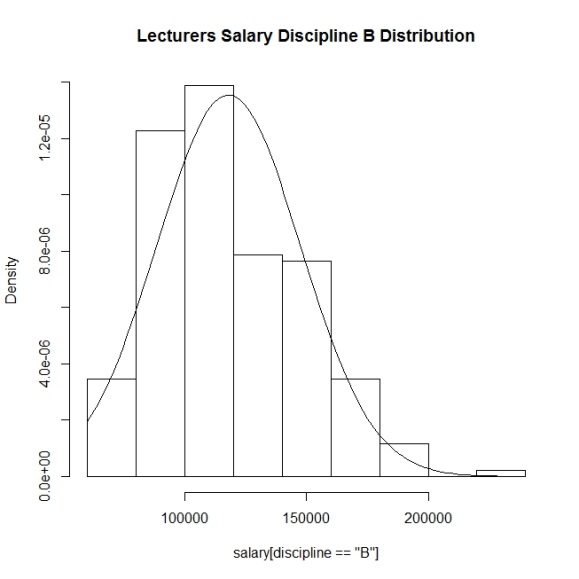


Figure 3: Salary Analysis on Discipline

Graph interpretation:

According to the boxplot graph as seen on (Figure 4) it is clear that professors’ variances of salaries are different based on discipline at a 5% significance level. The highest salary variance found for professors ‘A’ discipline whereas least salary variance found for ‘B’ discipline at a 5% significance level.

**Analysis of variance of Professors’ salary** based on discipline using **two variance F-test**

Sample fields**: *Discipline and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salary variances are similar or not based on discipline?**

H0: Ϭ2 A salary ≥ Ϭ2B salary

H1: Ϭ2 A salary < Ϭ2B salary

Rcmdr> var.test(salary ~ discipline, alternative='two.sided', conf.level=.95, Rcmdr+ data=Salaries)

F test to compare two variances

data: salary by discipline

F = 1.0746, num df = 180, denom df = 215, p-value = 0.6117

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.8129081 1.4255631

sample estimates:

ratio of variances

1.074596

**Decision: p-value=**0.6117**> α=0.05 => Accept H0 at 5% significant level**

**Two variance F- Test Conclusion**: According to the **Two variance F- Test (p=**0.6117)there is **enough** evidence to claim that **variances of salaries** of professors in discipline A ***are*** greater than professors in discipline B at a 5% significance level.

**Analysis of variance of Professors’ salary** based on discipline using **Bartlett’s Test**

Sample fields**: *Discipline and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salary variances are similar or not based on discipline?**

H0: Ϭ2 A salary ≥ Ϭ2B salary

H1: Ϭ2 A salary < Ϭ2B salary

Rcmdr> bartlett.test(salary ~ discipline, data=Salaries)

Bartlett test of homogeneity of variances

data: salary by discipline

Bartlett's K-squared = 0.2534, df = 1, p-value = 0.6147

Rcmdr> with(Salaries, tapply(salary, discipline, var, na.rm=TRUE))

A B

932578340 867840647

**Decision: p-value=**0.6147**> α=0.05 => Accept H0 at 5% significant level**

**Bartletts- Test Conclusion**: According to the **Bartlett’s Test (**= 0.6147**)** there is **no enough** evidence to claim that **variances of salaries** on ***discipline*** are equal at a 5% significance level.

**Analysis of variance of Professors’ salary** based on discipline using **Levene’s Test**

Sample fields**: *Discipline and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salary variances are similar or not based on discipline?**

H0: Ϭ2 A salary ≥ Ϭ2B salary

H1: Ϭ2 A salary < Ϭ2B salary

Rcmdr> leveneTest(salary ~ discipline, data=Salaries, center="median")

Levene's Test for Homogeneity of Variance (center = "median")

Df F value Pr(>F)

group 1 0.0458 0.8306

395

Rcmdr> with(Salaries, tapply(salary, discipline, var, na.rm=TRUE))

A B

932578340 867840647

**Decision: p-value=**0.8306**> α=0.05 => Accept H0 at 5% significant level**

**Levene’s- Test Conclusion**: According to the **Levene’s Test (p=**0.8306**)** there is **no enough** evidence to claim that **variances of salaries** on ***discipline*** are equal at a 5% significance level.

**Analysis of variance of Professors’ salary** based on gender using **Levene’s Test**

Sample fields**: *sex and Salary***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salary variances are similar or not based on sex?**

H0: Ϭ2 Male salary ≥ Ϭ2Femal salary

H1: Ϭ2 Male salary < Ϭ2 Female salary

Rcmdr> with(Salaries, tapply(salary, sex, var, na.rm=TRUE))

Female Male

673512912 926406546

Rcmdr> leveneTest(salary ~ sex, data=Salaries, center="median")

Levene's Test for Homogeneity of Variance (center = "median")

Df F value Pr(>F)

group 1 0.8401 0.3599

395

**Decision: p-value=**0.3599**> α=0.05 => Accept H0 at 5% significant level**

**Levene’s- Test Conclusion**: According to the **Levene’s Test (p=**0.3599**)** there is **no enough** evidence to claim that **variances of salaries** on ***sex*** are equal at a 5% significance level.

**Analysis of variance of Professors’ salary** based on gender using **Bartlett Test**

Sample fields**: *sex and Salary***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salary variances are similar or not based on sex?**

H0: Ϭ2 Male salary ≥ Ϭ2Femal salary

H1: Ϭ2 Male salary < Ϭ2 Female salary

> my\_modelBar\_sex<-bartlett.test(salary~sex,data=Salaries)

> my\_modelBar\_sex%>%tidy()

# A tibble: 1 x 4

statistic p.value parameter method

<dbl> <dbl> <dbl> <chr>

1 1.59 0.208 1 Bartlett test of homogeneity of variances

> with(Salaries, tapply(salary, sex, var, na.rm=TRUE))

Female Male

673512912 926406546

**Decision: p-value=**0.208 **> α=0.05 => Accept H0 at 5% significant level**

**Bartlett’s- Test Conclusion**: According to the **Bartlett’s Test (p=**0.208**)** there is **no enough** evidence to claim that **variances of salaries** on ***sex*** are equal at a 5% significance level.

1. **Statistical Hypothetical Mean Testing**: **One way ANOVA - Test For** Samples**: Salary Dataset**

Sample fields**: *Discipline and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salaries are similar or not based on discipline?**

H0: µ A salary = µ B salary

H1: µ A salary ≠ µ B salary

Rcmdr> summary(AnovaModel.2)

Df Sum Sq Mean Sq F value Pr(>F)

discipline 1 8.851e+09 8.851e+09 9.863 0.00181 \*\*

Residuals 395 3.544e+11 8.973e+08

---

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

Rcmdr> with(Salaries, numSummary(salary, groups=discipline, statistics=c("mean",

Rcmdr+ "sd")))

mean sd data:n

A 108548.4 30538.15 181

B 118028.7 29459.14 216

**Decision: p-value=**0.00181 **< α=0.05 => Reject H0 at 5% significant level**

**One-Way ANOVA Test Conclusion**: According to the **One-Way ANOVA test (*p=****0.00181*)there is **no enough** evidence to claim that **professors’ salaries mean values** on ***discipline*** are equal at a 5% significance level.

Therefore based on the results following modified H1 can be accepted at a 5% significance level.

H1: µ A salary < µ B salary

Graphical Analysis

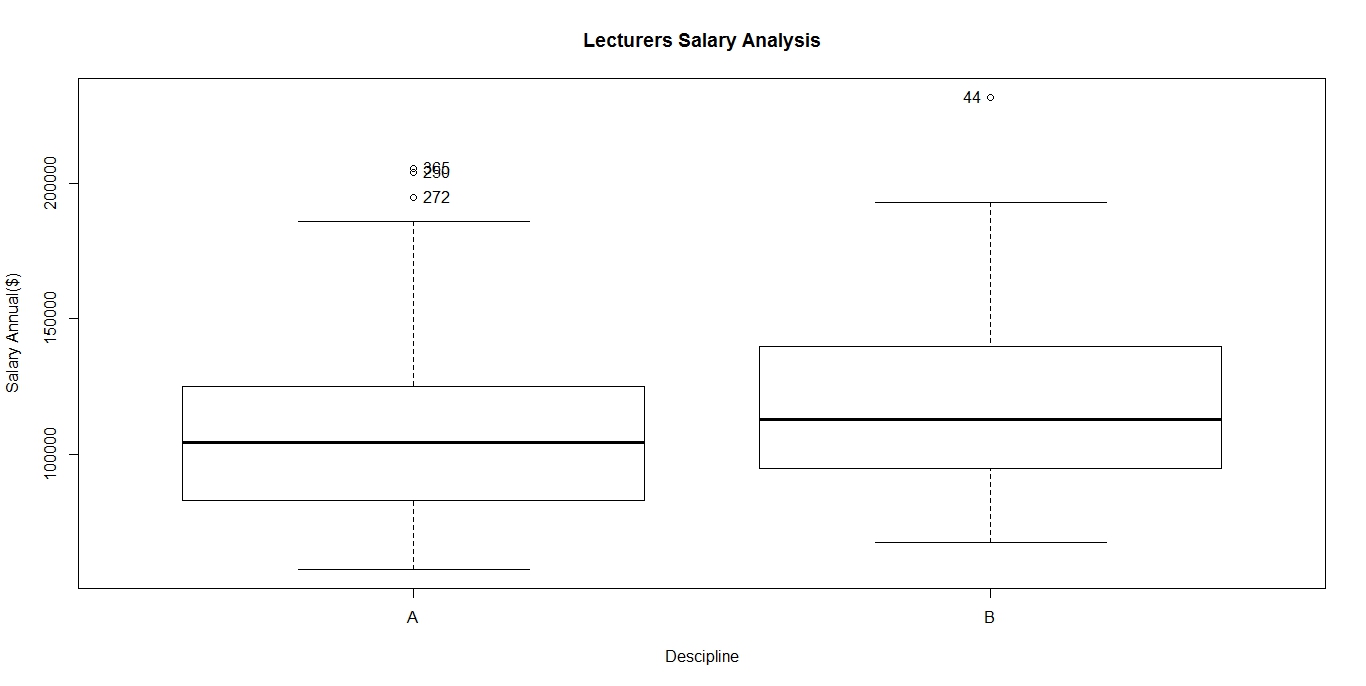


Figure 7: Salary Analysis on Discipline

Graph interpretation

According to the boxplot graph as seen on (Figure 4) it is clear that professors’ mean values of salaries are different based on discipline at a 5% significance level. The highest salary mean value found for B discipline professors whereas least salary mean value found for A discipline at a 5% significance level.

1. Statistical Hypothetical Testing: **Independent Sample t - Test** ForSamples**: Salary Dataset**

Sample fields**: *Discipline and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salaries are similar or not based on discipline?**

H0: µ A salary = µ B salary

H1: µ A salary ≠ µ B salary

Rcmdr> t.test(salary~discipline, alternative='two.sided', conf.level=.95,

Rcmdr+ var.equal=FALSE, data=Salaries)

Welch Two Sample t-test

data: salary by discipline

t = -3.1306, df = 377.83, p-value = 0.00188

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

95 percent confidence interval:

-15434.549 -3525.978

sample estimates:

mean in group A mean in group B

108548.4 118028.7

**Decision: p-value=**0.00188**< α=0.05 => Reject H0 at 5% significant level**

**Independent sample t- Test Conclusion**: According to the Independent sample t- **test (p=**0.00188)there is **no enough** evidence to claim that **salaries of professors’** on ***discipline*** are equal at a 5% significance level.

Therefore, based on the results following modified H1 can be accepted at a 5% significance level.

H1: µ A salary < µ B salary

Graphical Analysis

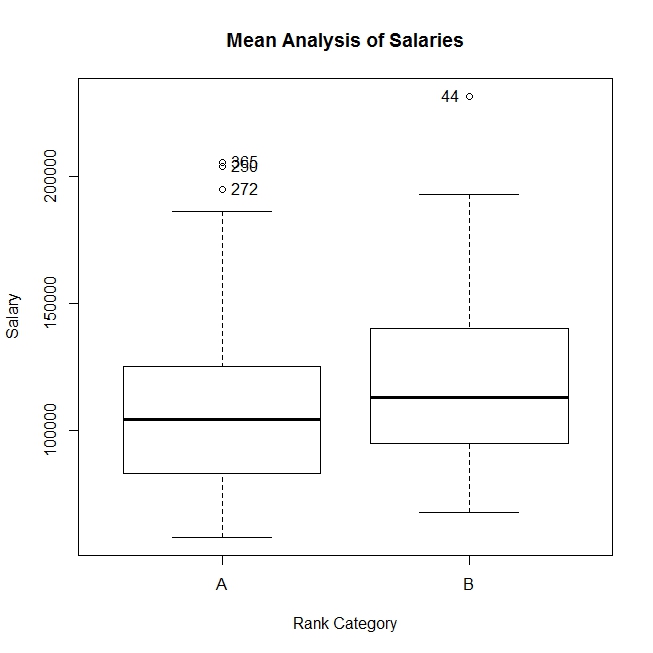


Figure 7: Salary Analysis on Discipline

Graph interpretation

According to the boxplot graph as seen on (Figure 7) it is clear that professors’ means of salaries are different based on discipline at a 5% significance level. The highest salary mean found for professors ‘*B’ discipline* whereas *least salary mean* found for ‘*A’ discipline* at a 5% significance level. ‘B’ discipline higher numbers of professionals are available in the upper range of the mean.

As per the results it is evident that more demand is available for Discipline B. Therefore, those professors earn better remuneration packages.

1. **Statistical Hypothetical Testing**: **Independent sample T - test For** Samples**: Salary Dataset**

Sample fields**: *Discipline and Salaries***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professor salary Means are similar or not based on discipline?**

H0: µ A salary = µ B salary

H1: µ A salary ≠ µ B salary

Rcmdr> t.test(salary~discipline, alternative='two.sided', conf.level=.95,

Rcmdr+ var.equal=FALSE, data=Salaries)

Welch Two Sample t-test

data: salary by discipline

t = -3.1306, df = 377.83, p-value = 0.00188

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

95 percent confidence interval:

-15434.549 -3525.978

sample estimates:

mean in group A mean in group B

108548.4 118028.7

**Decision: p-value=**0.00181 **< α=0.05 => Reject H0 at 5% significant level**

**Independent sample T- test Conclusion**: According to the **T- test (p=**0.00188s)there is **no enough** evidence to claim that **means of salaries of professors’** on ***discipline*** are equal at a 5% significance level.

Therefore based on the results following modified H1 can be accepted at a 5% significance level.

H1: µ A salary < µ B salary

1. Statistical Hypothetical Testing: **Independent Sample t - Test** ForSamples**: Salary Dataset**

Sample fields**: *Sex and Salary***

* Significant level (α) = **0.05**
* Confidence level =**95%**

**(\*) Professors’ salaries are similar or not based on sex?**

H0: µ Male salary = µ Female salary

H1: µ Male salary ≠ µ Female salary

Rcmdr> t.test(salary~sex, alternative='two.sided', conf.level=.95, var.equal=FALSE,

Rcmdr+ data=Salaries)

Welch Two Sample t-test

data: salary by sex

t = -3.1615, df = 50.122, p-value = 0.002664

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

95 percent confidence interval:

-23037.916 -5138.102

sample estimates:

mean in group Female mean in group Male

101002.4 115090.4

**Decision: p-value=**0.002664**< α=0.05 => Reject H0 at 5% significant level**

**Independent sample T- test Conclusion**: According to the **T- test (p=**0.002664)there is **no enough** evidence to claim that mean **salaries of professors’** on ***sex*** are equal at a 5% significance level.

Therefore, based on the results following modified H1 can be accepted at a 5% significance level.

H1: µ Female salary < µ Male salary

Graphical Analysis

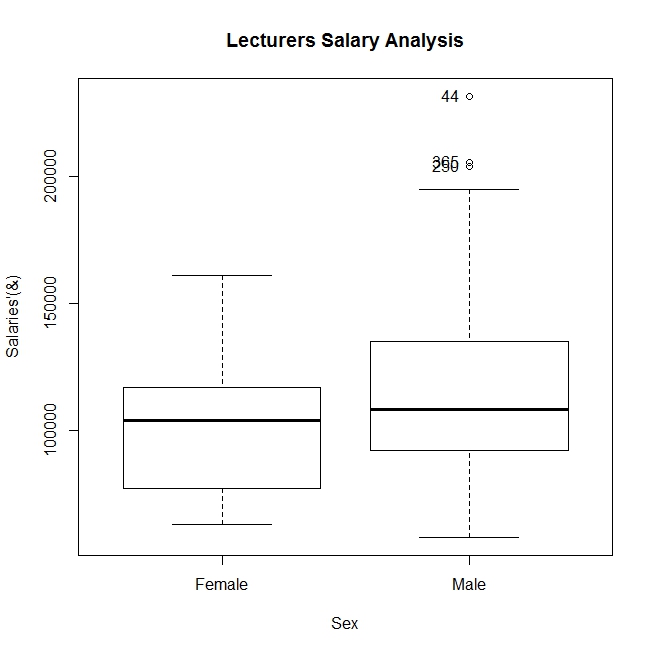


Figure 7: Salary Analysis on Sex

**Graph interpretation:**

According to the boxplot graph as seen on (Figure 7) it is clear that professors’ salaries’ mean values are different based on sex at a 5% significance level. The *highest* *salary mean* found for professors in *male’s category* whereas *least salary means* found for *Female’s category* at a 5% significance level. Further, it is noticeable that female lecturers majority earns less than the average salary whereas male lecturers majority earns more than the average salaries at a 5% significant level.

As Female individuals having more responsibilities apart from what is given by the profession itself, there are incapability associate with full time contribution unlike Male candidates. Some projects require physical strength apart from intellectual contribution for what Male individuals are much suited. Therefore, overall there is a possibility of having performance fluctuations between Male Professors and Female Professors. As a result Male Professors benefits possible to increase than Female ones.

1. Statistical Hypothetical Testing: **Single Sample t – Test/One sample t-test** ForSamples**: Salary Dataset**

Sample fields**: *Years Since PhD (from the given SAMPLE data), Years Since PhD (from assumed POPULATION data), POPLUATION=> All USA UNIVERSITIES***

* Significant level (α) = **0.05**
* Confidence level =**95%**

\*) Whether or not Professors Years Since PhD values from POPULATION data equals to 30?

H0: Professors Years Since PhD values from POPULATION data = 30

H1: Professors Years Since PhD values from POPULATION data ≠30

Rcmdr> with(Salaries, (t.test(yrs.since.phd, alternative='two.sided', mu=30, conf.level=.95)))

One Sample t-test

data: yrs.since.phd

t = -11.882, df = 396, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 30

95 percent confidence interval:

21.04331 23.58641

sample estimates:

mean of x

22.31486

**Decision: p-value=**2.2e-16**< α=0.05 => Reject H0 at 5% significant level**

**Conclusion of Single Sample T-test :** As per the results of the single Sample t-Test(P< 2.2e-16)it is clear that Population parameter mean value of years since PhD not equals to 30 at a 5% significance level.

As per the results the mean value should be available in between 21.04331 and 23.58641. The final results indicate the years since PhD estimated mean value would be 22.31486 at a 5% significance level.

\*) Whether or not Professors Years Since PhD values from POPULATION data equals to 22?

H0: Professors Years Since PhD values from POPULATION data = 22

H1: Professors Years Since PhD values from POPULATION data ≠22

Rcmdr> with(Salaries, (t.test(yrs.since.phd, alternative='two.sided', mu=22, conf.level=.95)))

One Sample t-test

data: yrs.since.phd

t = 0.48681, df = 396, p-value = 0.6267

alternative hypothesis: true mean is not equal to 22

95 percent confidence interval:

21.04331 23.58641

sample estimates:

mean of x

22.31486

**Decision: p-value=**0.6267**> α=0.05 => Accept H0 at 5% significant level**

**Conclusion of Single Sample T-test :** As per the results of the single Sample t-Test(P=0.6267)it is clear that Population parameter mean value of years since PhD equals to 22 at a 5% significance level.

1. Statistical Hypothetical Testing: **Multi-Way ANOVA(MANOVA)** ForSamples**: Salary Dataset**

Sample fields**: *salary, rank, discipline*** *and* ***sex***

* Significant level (α) = **0.05**
* Confidence level =**95%**

\*) Whether or not Professors salaries mean are equal based on rank, discipline and sex?

H0: Professors salaries mean are equal based on rank, discipline and sex

H1: Professors salaries mean are not equal based on rank, discipline and sex

Or

\*) Whether or not Professors salaries mean are equal based on rank, discipline and sex?

H0: µProfessor Salary (discipline, sex) = µAssociate Professor Salary (discipline, sex) = µAssistance Professor Salary(discipline, sex)

H1: µProfessor Salary (discipline, sex) ≠ µAssociate Professor Salary (discipline, sex) ≠ µAssistance Professor Salary(discipline, sex)

Rcmdr> AnovaModel.3 <- lm(salary ~ discipline\*rank\*sex, data=Salaries, contrasts=list(discipline

Rcmdr+ ="contr.Sum", rank ="contr.Sum", sex ="contr.Sum"))

Rcmdr> Anova(AnovaModel.3)

Anova Table (Type II tests)

Response: salary

Sum Sq Df F value Pr(>F)

discipline 18474779335 1 35.6269 0.000000005428 \*\*\*

rank 145243807629 2 140.0446 < 2.2e-16 \*\*\*

sex 758756669 1 1.4632 0.2272

discipline:rank 474830765 2 0.4578 0.6330

discipline:sex 461974122 1 0.8909 0.3458

rank:sex 218493774 2 0.2107 0.8101

discipline:rank:sex 132392998 2 0.1277 0.8802

Residuals 199646647445 385

---

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

Rcmdr> Tapply(salary ~ discipline + rank + sex, mean, na.action=na.omit, data=Salaries) # means

, , sex = Female

rank

discipline AssocProf AsstProf Prof

A 72128.50 72933.33 109631.9

B 99435.67 84189.80 131836.2

, , sex = Male

rank

discipline AssocProf AsstProf Prof

A 85048.86 74269.61 120619.3

B 101621.53 84647.08 133518.4

Rcmdr> Tapply(salary ~ discipline + rank + sex, sd, na.action=na.omit, data=Salaries) # std. deviations

, , sex = Female

rank

discipline AssocProf AsstProf Prof

A 6402.716 5463.210 15094.59

B 14086.476 9792.119 17504.27

, , sex = Male

rank

discipline AssocProf AsstProf Prof

A 10611.885 4580.125 28504.88

B 9607.891 6900.293 26514.29

Rcmdr> xtabs(~ discipline + rank + sex, data=Salaries) # counts

, , sex = Female

rank

discipline AssocProf AsstProf Prof

A 4 6 8

B 6 5 10

, , sex = Male

rank

discipline AssocProf AsstProf Prof

A 22 18 123

B 32 38 125

**Decision: based on the above results out of three P-values 2 p-values indicate results < α=0.05. Therefore Reject H0 at 5% significant level**

**Mean Analysis Conclusion**:

According to the **Multi-Way ANOVA (MANOVA) test** there is **no** **statistical** evidence to claim that **Mean values of Salaries** on ***Professors’ ranks, discipline and Sex*** are same at a 5% significance level. As per the multi-dimensional results reveals that professors earn highest salaries and associate professor earn second highest salaries and assistant professors earn third highest salaries. There is only one case found that Assistant Professors salaries are higher than Associate professors where Gender is Female and Discipline is A at a 5% significance level. Overall the analysis reveals that female professor categories earn significant lesser salaries that of male professor at a 5 % significance level.

**Statistical Hypothetical** Testing: ***Spearman test -***for correlation analysis of Samples**: Salary dataset**

Sample field**:** *Professors’ rank, professors’ salary*

* Significant level (α) = **0.05**
* Confidence level =**95%**
* Let **ῤ** =the true population correlation coefficient between *Professors’ rank* and *professor’s salary*

(\*) Whether or not there **is a correlation** foundbetween **professors’ rank** and **salary?**

**H0:** There is **no correlation (ῤ =0)** between *Professors’ rank* and *professor’s salary*

**H1:** There is a **correlation (ῤ ≠0)** between *Professors’ rank* and *professor’s salary*

> cor.test(salary,rank,method='spearman',alternative = "two.sided")

Warning in cor.test.default(salary, rank, method = "spearman", alternative = "two.sided") :

Cannot compute exact p-value with ties

Spearman's rank correlation rho

data: salary and rank

S = 3133686, p-value < 2.2e-16

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.6995045

**Decision: p-value=**2.2e-16**< α=0.05 => Reject H0 at 5% significant level**

Correlation Test Conclusion

According to the ***spearman rank correlation-tests*** (< 2.2e-16) for correlation analysis of Samples there is **a statistical** **evidence** to claim that **a moderate direct correlation** (ῤ =0.70) **exists** between ***professors’ rank*** *and* ***professors’* salary** at a 5% significance level. According to the finding it is required to do a regression analysis to find the model of the relationship.

The professors’ ***rank*** is very much influential factor to determine fringe benefits offered by the university to relevant professors. The professor rank increments happen based on qualifications such as no of doctorates, post doctorates or any equivalent, no of research publications done, national and international conferences attended, Research and development projects associated and how much contribution rendered to the University etc.

> summary(sal\_model\_1)

Call:

lm(formula = salary ~ rank, data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-68055 -17655 -1141 12916 105690

Coefficients:

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

(Intercept) 52777 4087 12.91 <2e-16 \*\*\*

rank 24359 1562 15.59 <2e-16 \*\*\*

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

Residual standard error: 23860 on 395 degrees of freedom

Multiple R-squared: 0.381, Adjusted R-squared: 0.3794

F-statistic: 243.1 on 1 and 395 DF, p-value: < 2.2e-16

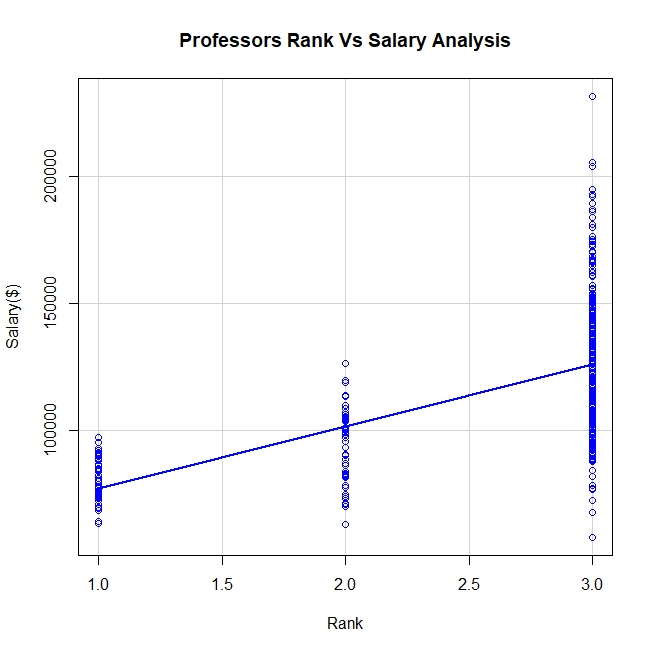
Graphical Analysis

Figure 4: Professor Salary Analysis Scatterplot

As seen on (Figure 5) it is clearly identifiable that scatter plot values position according to a positive trend. Therefore, it is confirmed that there is a **positive (direct) moderate** relationship (ῤ =0.70) available between professors’ salary and professor rank at a 5% significance level.

**Statistical Regression** Testing: Simple ***Linear Regression-tests*** for modeling of Samples**: Salaries ;rank** and **salary**

* + Sample dataset: Salaries
  + Significant level (α): 0.05 (5%)
  + Confidence level: 95%

Liner general model

> lm(salary~rank)

Call:

lm(formula = salary ~ rank)

Coefficients:

(Intercept) rank

52777 24359

Final Linear model

Graphical Representation – Professors’ Salary Data and Rank

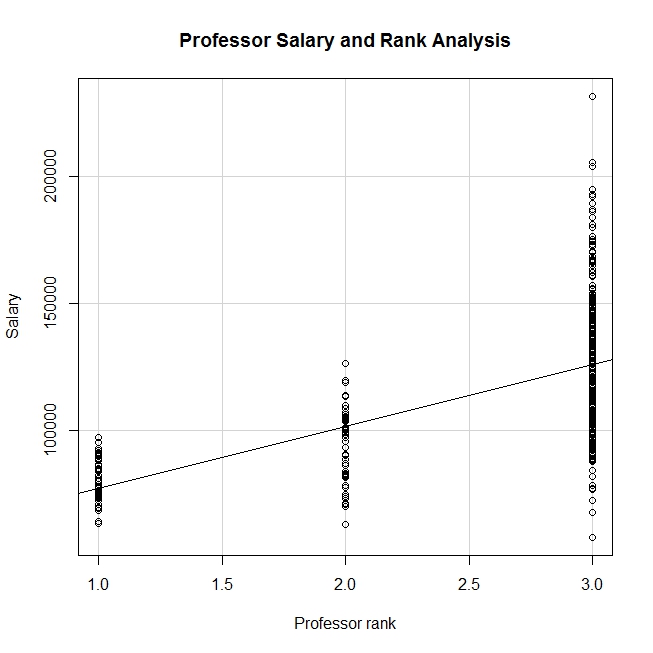


Figure 6: Professor Salary Analysis Scatterplot with abline

Interpretation of the model

According to the linear model generated by regression analysis as seen on(Figure 6), it is clear to identify values are arranged along with the absolute line creating a linear positive trend. When **professor’s rank** increase by **24359** **units**, in proportion to that professor’s **salary** increase by **one unit**. Here professor’s salary is further determined by the constant value **52777**. Therefore, the model clearly indicates that professors’ salary is determined by professors’ rank at a 5% significance level. The model accuracy becomes higher as R2= 0.381 and adjusted R2 = 0.3794 but the model residuals indicate still high values. When we consider about this model can be confirmed as a better model as accuracy of the model available somewhere around 38%. Therefore, much better salary predications can be identified with this model if only rank factor is considered.

**Statistical Hypothetical** Testing: ***Pearson product moment-tests*** for correlation analysis of Samples**: Salary dataset**

Sample field**:** *Professors’ years since PhD, professors’ salary*

* Significant level (α) = **0.05**
* Confidence level =**95%**
* Let **ῤ** =the true population correlation coefficient(*degree*) between ***years since PhD*** and *professors’ salary*

**(\*)** Whether or not **there is a correlation** found **between** professors’ ***years since PhD*** and **salary?**

**H0:** There is no correlation (**ῤ =0)** between *Professors’* ***years since PhD*** and *professors’* ***salary***

**H1:** There is a correlation (**ῤ ≠0)** between *Professors’* ***years since PhD*** and *professors’* ***salary***

Rcmdr> with(Salaries\_new, cor.test(salary, yrs.since.phd, alternative="two.sided",

Rcmdr+ method="pearson"))

Pearson's product-moment correlation

data: salary and yrs.since.phd

t = 9.1775, df = 395, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.3346160 0.4971402

sample estimates:

cor

0.4192311

> summary(sal\_model\_2)

Call:

lm(formula = salary ~ yrs.since.phd, data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-84171 -19432 -2858 16086 102383

Coefficients:

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

(Intercept) 91718.7 2765.8 33.162 <2e-16 \*\*\*

yrs.since.phd 985.3 107.4 9.177 <2e-16 \*\*\*

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

Residual standard error: 27530 on 395 degrees of freedom

Multiple R-squared: 0.1758, Adjusted R-squared: 0.1737

F-statistic: 84.23 on 1 and 395 DF, p-value: < 2.2e-16

**Decision: p-value=**2.2e-16**< α=0.05 => Reject H0 at 5% significant level**

Correlation Test Conclusion

According to the ***Pearson product moment-tests*** for correlation analysis of Samples there is **a statistical** **evidence** to claim that **a moderate direct correlation (ῤ** =0.42) **exists** between ***professors’ years since PhD*** *and* ***professors’* salary** at a 5% significance level. According to the finding it is required to do a *regression analysis* to find the model of the relationship.

Graphical Analysis

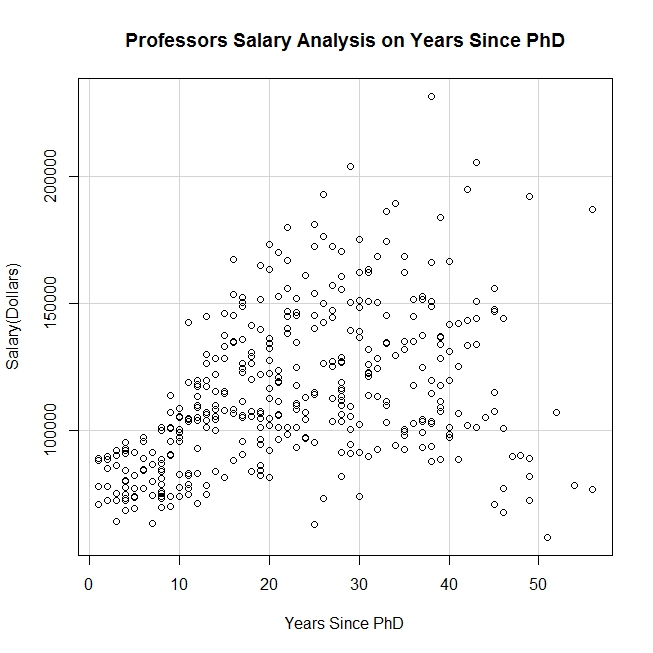
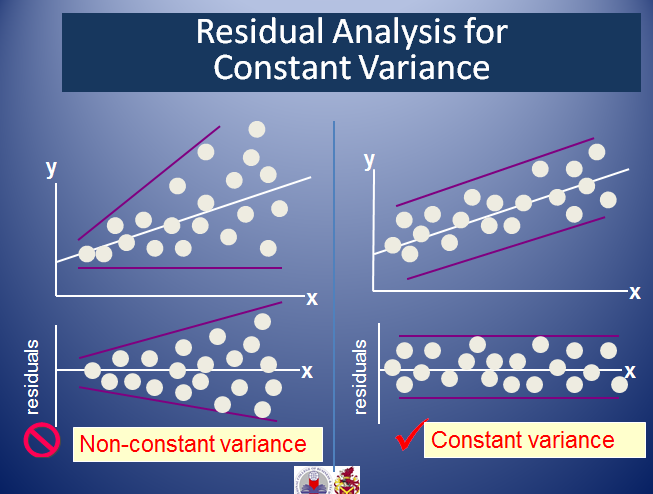


Figure 7: Professor Salary Analysis on Years since PhD Scatterplot

As seen on (Figure 7) it is clearly identifiable that scatter plot values position according to a *positive* Non constant variance trend with *moderate level* dispersion. Therefore, it is confirmed that there is a **positive (direct) moderate** relationship ( ῤ =0.42)available between professors’ salary and professors’ years since PhD at a 5% significance level.

**Statistical Regression** Testing: Simple ***Linear Regression-tests*** for modeling of Samples**: Salaries ;Years Since PhD** and **salary**

* + Sample dataset: Salaries
  + Significant level (α): 0.05 (5%)
  + Confidence level: 95%

Liner general model

> lm(salary~yrs.since.phd)

Call:

lm(formula = salary ~ yrs.since.phd)

Coefficients:

(Intercept) yrs.since.phd

91718.7 985.3

Final liner model

**Graphical Representation**

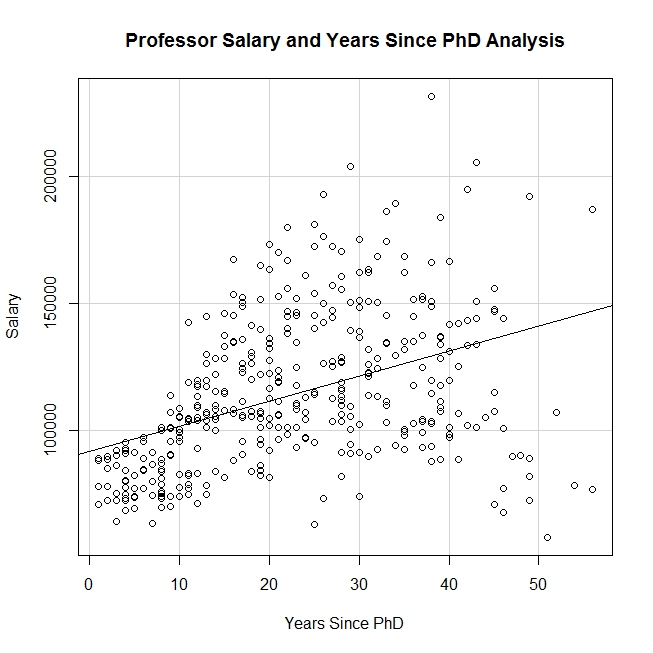


Figure 8: Professor Salary Analysis on Years since PhD Scatterplot

Interpretation of the model

According to the linear model generated by regression analysis as seen on(Figure 8), it is clear to identify that values are arranged along with the absolute line creating a linear positive moderate trend. When **professors’ years since PhD values** increase by 985 **units**, in proportion to that professors’ **salary** increase by **one unit**. Here professors’ salary is further determined by the constant value **91719**. Therefore, the model clearly indicates that professors’ salary is determined by professors’ years since PhD at a 5% significance level. The model accuracy becomes low as R2=0.1758 and adjusted R2 = 0.1737. The model residuals also indicate high values and that further confirms model accuracy is low. Therefore, possible errors in salary predications can be identified with this model.

**Statistical Hypothetical** Testing: ***Pearson product moment-tests*** for correlation analysis of Samples**: Salary dataset**

Sample field**:** *Professors’ years’ service, professor’s salary*

* Significant level (α) = **0.05**
* Confidence level =**95%**
* Let **ῤ** =the true population correlation coefficient between *Professors’ rank* and *professors’ salary*

\*Whether or not there is **a correlation exists** betweenprofessors’ **salary** and **years of service?**

**H0:** There is *no correlation (****ῤ =0)*** between *Professors’* ***years’ service*** and *professor’s* ***salary***

**H1:** There is *a correlation (****ῤ ≠0)***between *Professors’* ***years’ service*** and *professor’s* ***salary***

Let **ῤ** =the true population correlation coefficient between *Professors’* ***years’ service*** and *professor’s* ***salary***

Rcmdr> with(Salaries\_2, cor.test(salary, yrs.service, alternative="two.sided",

Rcmdr+ method="pearson"))

Pearson's product-moment correlation

data: salary and yrs.service

t = 7.0602, df = 395, p-value = 7.529e-12

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.2443740 0.4193506

sample estimates:

cor

0.3347447

**Decision: p-value=**7.529e-12**< α=0.05 => Reject H0 at 5% significant level**

Correlation Test Conclusion

According to the ***Pearson product moment-test (p=***7.529e-12)for correlation analysis of Samples there is **a statistical** **evidence** to claim that **a moderate direct correlation (ῤ** =0.335) **exists** between ***professors’ years’ service*** *and* ***professors’* salary** at a 5% significance level. According to the finding it is required to do a regression analysis to find the model of the relationship.

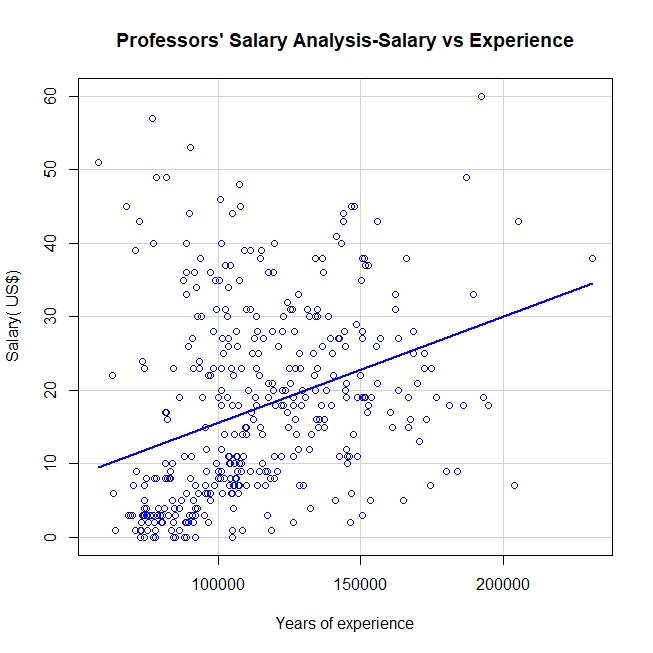
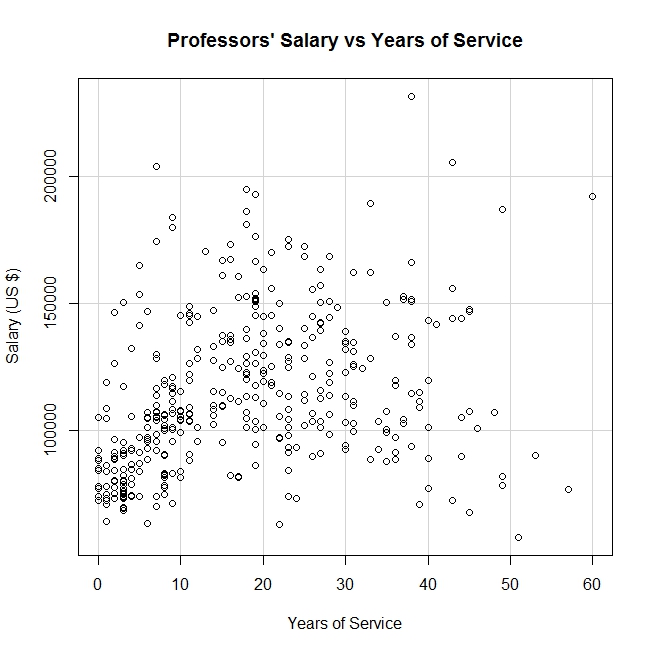
**Graphical Representation**

Figure 8 : Professors’ Salary vs Years of Service Scatterplot Graph

Figure : Professor Salary Vs Years of Experience Analysis

As seen on (Figure 8, Figure 5) it is clearly identifiable that scatter plot values position according to a positive non constant variance trend with moderate level data dispersion. Therefore, it is confirmed that there is a **positive (direct) moderate** relationship (ῤ =0.335) available between professors’ salary and professors’ years of experience at a 5% significance level.

**Statistical Hypothetical** Testing: Simple ***Linear Regression-tests*** for modeling of Samples**: Salaries ; years’ service** and **salary**

* + Sample dataset: Salaries
  + Significant level (α): 0.05 (5%)
  + Confidence level: 95%

Liner general model

Rcmdr> summary(RegModel.2)

Call:

lm(formula = salary ~ yrs.service, data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-81933 -20511 -3776 16417 101947

Coefficients:

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

(Intercept) 99974.7 2416.6 41.37 < 2e-16 \*\*\*

yrs.service 779.6 110.4 7.06 7.53e-12 \*\*\*

---

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

Residual standard error: 28580 on 395 degrees of freedom

Multiple R-squared: 0.1121, Adjusted R-squared: 0.1098

F-statistic: 49.85 on 1 and 395 DF, p-value: 7.529e-12

Final Liner model

Graphical Representation – Professors’ Salary Data and Years’ service

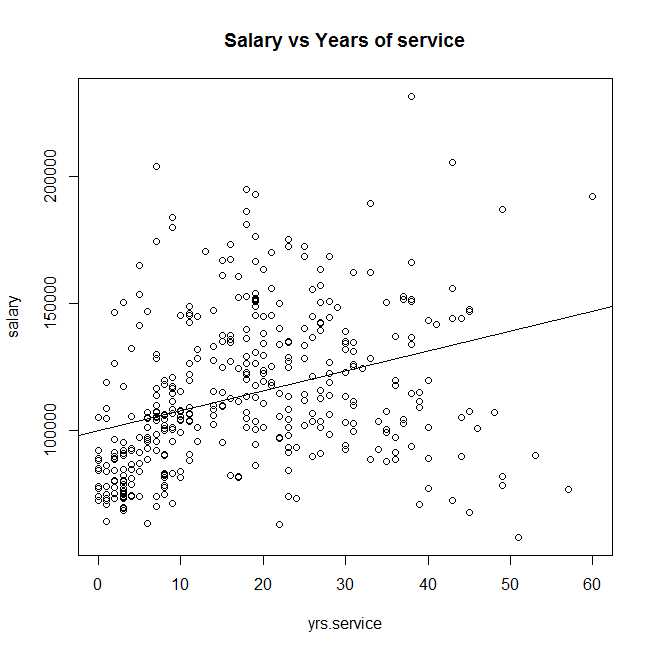


Figure Salary vs Years of Service

Interpretation of the model

According to the linear model generated by regression analysis, it is clear to identify that when **professor’s years of service** increases by **780** **units**, in proportion to that professor’s **salary** increases by **one unit**. Here the professor’s salary calculation is further determined by the constant value **99975** Therefore, the model clearly indicates that professor’s salary is determined by professor’s years of service at a 5% significance level. The model accuracy becomes low as R2= 0.1121 and adjusted R2 = 0.1098. The model residuals also indicate high values and that further confirms model accuracy is low. Therefore, possible errors in salary predications can be identified with this model.

**Statistical Regression** Testing: ***Multiple*** ***Linear Regression-tests*** for modeling of Samples**: Salaries ; salary ,Rank, years since PHD** and **years’ service**

* + Sample dataset: Salaries
  + Significant level (α): 0.05 (5%)
  + Confidence level: 95%

Multiple Liner general models

> lm(salary~rank+yrs.since.phd+yrs.service)

Call:

lm(formula = salary ~ rank + yrs.since.phd + yrs.service)

Coefficients:

(Intercept) rank

150690.3 -24716.6

yrs.since.phd yrs.service

304.0 -381.7

---------🡪

Rcmdr> summary(LinearModel.2)

Call:

lm(formula = salary ~ rank + yrs.since.phd + yrs.service, data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-64210 -17594 -1071 13635 108525

Coefficients:

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

(Intercept) 51823.9 4196.1 12.351 <2e-16 \*\*\*

rank 24716.6 2198.1 11.244 <2e-16 \*\*\*

yrs.since.phd 304.0 250.1 1.215 0.2249

yrs.service -381.7 222.7 -1.714 0.0873 .

---

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

Multiple R-squared: 0.3859, Adjusted R-squared: 0.3812

F-statistic: 82.32 on 3 and 393 DF, p-value: < 2.2e-16

Multiple Liner general model

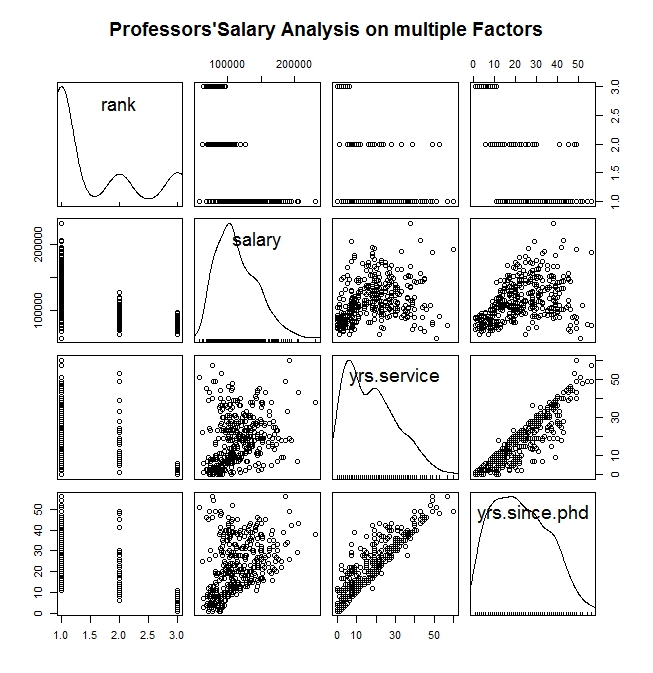
Graphical Representation – Professors’ Salary, Rank, years since PHD and years of service

Figure 7: Professor Salary analysis scatterplot matrix

Interpretation of the model

As seen on (Figure 7) the multiple linear general model developed by multiple regression analysis, it is clear to identify that when **professors’ rank proportionately** increase by **units** , **professors’ years since PhD** increases **by**  **units** and **professors’ years of service inversely** increases **by**  **units** in proportion to that professor’s **salary** increases **by one unit.** Here the professors’ salary calculation is further influenced by the constant value. Therefore, the model clearly indicates that professors’ salaries are determined by *professors’ rank*, *professors’ years since PhD* and *professors’ years of service* at a 5% significance level.

According to the standard error values and t-values of independent variables, it is confirmed that years of service (Std.Error:222.7) is the most crucial factor to be considered in Salary determination of a professor. The years since PhD (Std.Error:250.1) can be considered as the second most important factor in salary calculation. Professors’ rank(Std.Error:2198.1) is the least important factor to consider since assigning numbers for rank is stochastic in nature from context to context and university to university.

The model accuracy becomes higher as R2= 0.3859 and adjusted R2 = 0.3812 but the model residuals indicate high values. When we consider about the other models developed, this model can be confirmed as the best model as accuracy is higher than others. Therefore, much better salary predications can be identified with this model.

**Extra Knowledge Section**

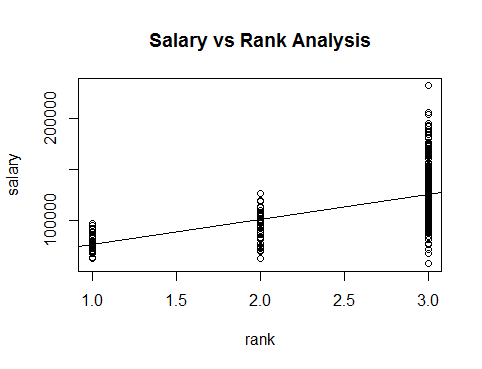
**Statistical Regression** Testing: Model Comparison based on ***Linear Regression vs Non Linear regression (Linear Transformation)*** for modeling of Samples**: Salaries ; salary ,Rank, years since PHD** and **years’ service**

Simple Linear Regression

**Statistical Regression** Testing: Simple ***Linear Regression-tests*** for modeling of Samples**: Salaries ;rank** and **salary**

* + Sample dataset: Salaries
  + Significant level (α): 0.05 (5%)
  + Confidence level: 95%

Liner general model

> summary(model\_sal\_rank)

Call:

lm(formula = salary ~ rank, data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-68055 -17655 -1141 12916 105690

Coefficients:

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

(Intercept) 52777 4087 12.91 <2e-16 \*\*\*

rank 24359 1562 15.59 <2e-16 \*\*\*

---

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

Residual standard error: 23860 on 395 degrees of freedom

Multiple R-squared: 0.381, Adjusted R-squared: 0.3794

F-statistic: 243.1 on 1 and 395 DF, p-value: < 2.2e-16

Final Linear model

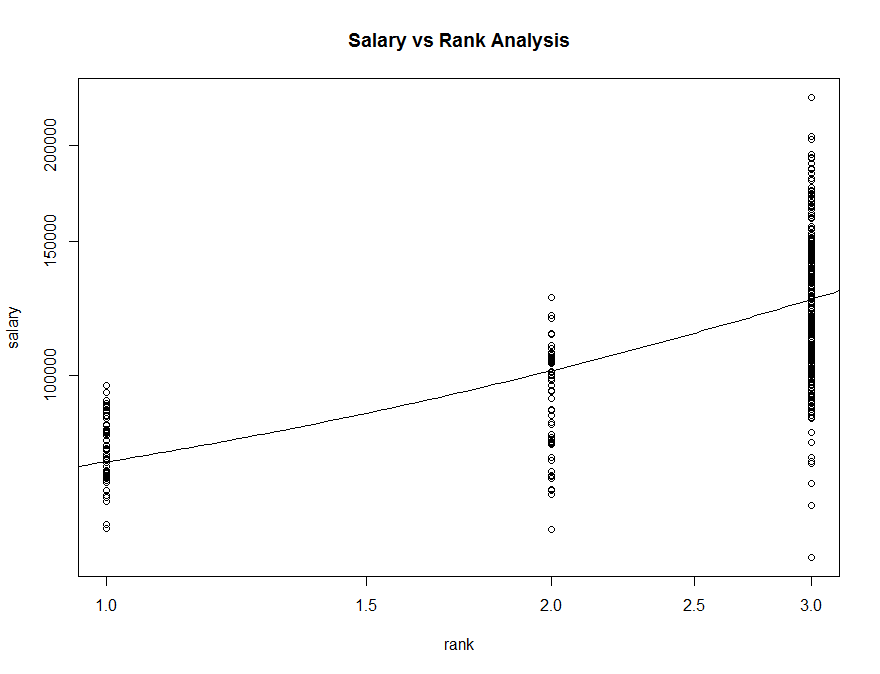
**Statistical Regression** Testing: Simple Non-***Linear Regression-tests*** for modeling of Samples**: Salaries ;rank** and **salary**

* + Sample dataset: Salaries
  + Significant level (α): 0.05 (5%)
  + Confidence level: 95%

Non-Liner general model-log(y)

model\_sal\_rank\_logy<-lm(log(salary)~rank,data=Salaries\_2)

> summary(model\_sal\_rank\_logy)

Call:

lm(formula = log(salary) ~ rank, data = Salaries\_2)

> summary(model\_sal\_rank\_logy)

Call:

lm(formula = log(salary) ~ rank, data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-0.75531 -0.13481 0.00391 0.13423 0.63247

Coefficients:

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

(Intercept) 11.04188 0.03325 332.05 <2e-16 \*\*\*

rank 0.22606 0.01271 17.78 <2e-16 \*\*\*

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

Residual standard error: 0.1941 on 395 degrees of freedom

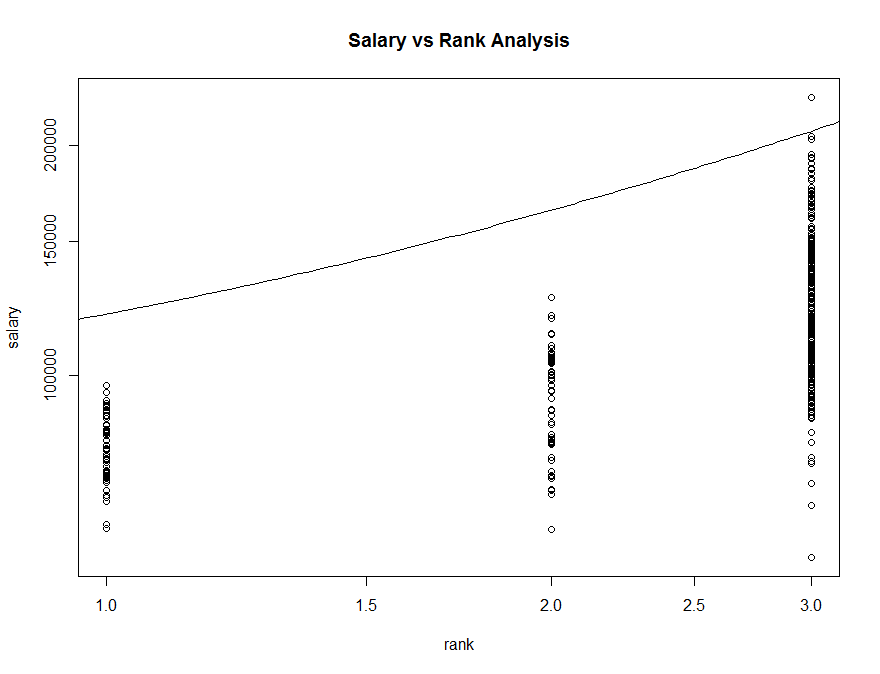
Multiple R-squared: 0.4447, Adjusted R-squared: 0.4432

F-statistic: 316.3 on 1 and 395 DF, p-value: < 2.2e-16

Final Non-Liner general model

Non-Liner general model-log(x)

> summary(model\_sal\_rank\_logx)

Call:

lm(formula = salary ~ log(rank), data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-67001 -17501 -2063 13744 106744

Coefficients:

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

(Intercept) 76198 2807 27.14 <2e-16 \*\*\*

log(rank) 44240 2982 14.84 <2e-16 \*\*\*

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

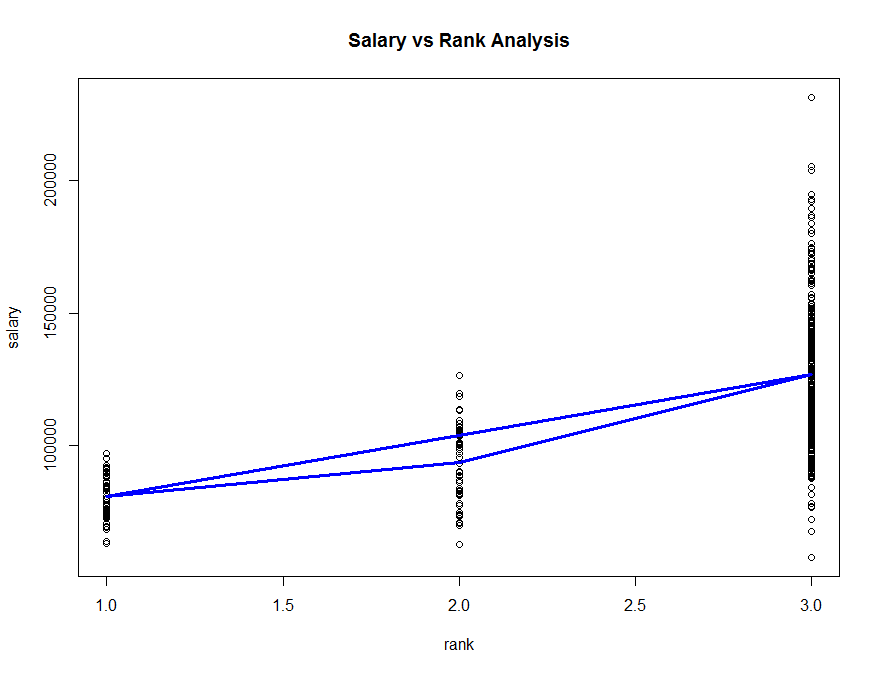
Residual standard error: 24300 on 395 degrees of freedom

Multiple R-squared: 0.3578, Adjusted R-squared: 0.3562

F-statistic: 220.1 on 1 and 395 DF, p-value: < 2.2e-16

Non-Liner general model-Polynomial-Second order

> summary(model\_sal\_rank\_poly2)

Call:

lm(formula = salary ~ poly(rank, 2), data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-68972 -16376 -1580 11755 104773

Coefficients:

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

(Intercept) 113707 1186 95.86 < 2e-16 \*\*\*

poly(rank, 2)1 372028 23634 15.74 < 2e-16 \*\*\*

poly(rank, 2)2 69474 23634 2.94 0.00348 \*\*

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

Residual standard error: 23630 on 394 degrees of freedom

Multiple R-squared: 0.3943, Adjusted R-squared: 0.3912

F-statistic: 128.2 on 2 and 394 DF, p-value: < 2.2e-16

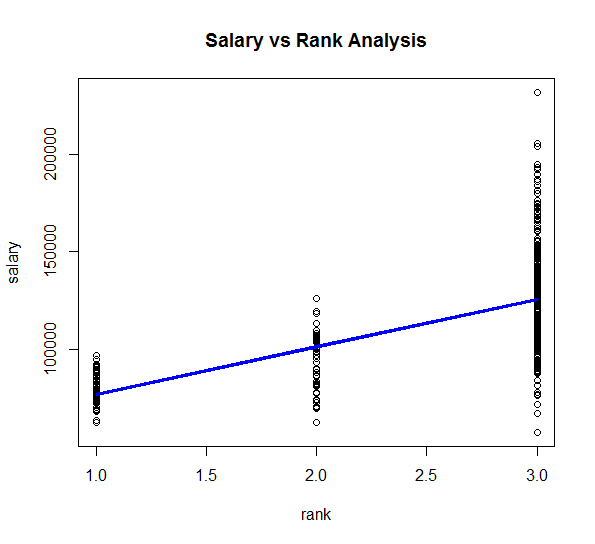
Non-Liner general model-Polynomial-Second order

Non-Liner general model-Polynomial-Third order

This model does not work

Non-Liner general model-Polynomial-one order ( This equals to Simple Linear Model)

> summary(model\_sal\_rank\_poly1)

Call:

lm(formula = salary ~ poly(rank, 1), data = Salaries\_2)

Residuals:

Min 1Q Median 3Q Max

-68055 -17655 -1141 12916 105690

Coefficients:

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

(Intercept) 113707 1198 94.95 <2e-16 \*\*\*

poly(rank, 1) 372028 23861 15.59 <2e-16 \*\*\*

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

Residual standard error: 23860 on 395 degrees of freedom

Multiple R-squared: 0.381, Adjusted R-squared: 0.3794

F-statistic: 243.1 on 1 and 395 DF, p-value: < 2.2e-16

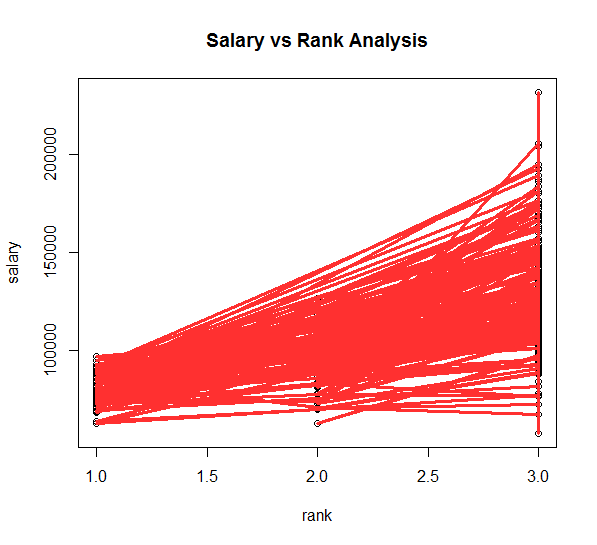
Final Non-Liner general model-Polynomial-one order ( This equals to Simple Linear Model)

Generalized Linear Model-Poisson(sqrt)

> summary(model\_sal\_rank\_gm\_po)

Call:

glm(formula = salary ~ rank, family = poisson(sqrt), data = Salaries\_2)

Deviance Residuals:

Min 1Q Median 3Q Max

-214.90 -53.00 -3.29 39.90 266.27

Coefficients:

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

(Intercept) 241.49361 0.08565 2820 <2e-16 \*\*\*

rank 37.76712 0.03274 1154 <2e-16 \*\*\*

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 3091659 on 396 degrees of freedom

Residual deviance: 1796716 on 395 degrees of freedom

AIC: 1802058

Number of Fisher Scoring iterations: 4

Conclusion of whole analysis

According to the data analysis done with the support of salaries dataset , the normality test rest results indicated none of variables salary ,rank, years since PhD and years of service are normally distributed at a 5% significance level. According to the Anderson darling test(p=2.508e-10),lillifore test(p= 2.354e-08) and Shapiro wilk test(p=6.076e-09) of salary indicated values are not normally distributed at 5% significance level. According to the Anderson darling test(p=1.565e-07),lillifore test(p= 8.011e-05) and Shapiro wilk test(p=2.328e-07) of **years since PhD** indicated values are not normally distributed at 5% significance level. In compliance with According to the Anderson darling test(p=1.325e-15),lillifore test(p= 4.513e-15) and Shapiro wilk test(p=2.337e-11) of **years of service** indicated values are not normally distributed at 5% significance level. In compliance with According to the Anderson darling test(p=< 2.2e-16),lillifore test(p= 2.2e-16) and Shapiro wilk test(p=2.2e-16) of **rank** indicated values are not normally distributed at 5% significance level.

According to the One Way - **ANOVA test (**p=2e-16)for variance analysis of salaries there is **no enough** evidence to claim that **variances of salaries** on ***professors’ ranks*** are equal at a 5% significance level. The results further revealed that the highest salary variance found for professors whereas least salary variance found for Assistant professors and associate professors found under 2nd highest salary variance category at a 5% significance level.

The One **Way -ANOVA** test**(p**<2e-16) for the mean analysis of salaries revealed that there is **no enough** evidence to claim that **mean of salaries** on ***professors’ ranks*** are equal at a 5% significance level. The results further revealed that the highest salary mean found for professors’ whereas least salary means found for assistant professors’ and associate professors’ found as 2nd highest salary mean at a 5% significance level.

According to the Pearson product moment test (*p-value < 2.2e-16*) for correlation analysis of Salary and Rank, it was clear to find out there is a **moderate direct correlation** (ῤ =0.62) between the two variables. As per the Spearman Correlation analysis of the two variables it was possible to recognize a strong direct correlation (ῤ =0.70) at a 5% significant level. When Salary and Years since PhD Pearson product moment test considered, it was clear to find out a **moderate direct correlation** (**ῤ** =0.42) between the variables. Further, according to the according to the Pearson product moment test value of salary and years of experience, it was clear to find out there is a **moderate direct correlation** (**ῤ** =0.335) between the two variables. Therefore, the full spectrum of correlation analysis revealed that there is clear association exist between professors’ salary and rank, years since PhD and years of experience at a 5 % significance level.

As per the results of simple linear regression analysis based on salary response variable and the other explanatory variables such as rank, Years since PhD and Years of service, it was able to find highest model fitness value (R2= 0.381) and (Adjusted R2 = 0.3794) for

model. As per the multiple linear regression analysis based on the mentioned variables it was clear to find out a model with a fitness value(R-squared: 0.3859) and (Adjusted R-squared: 0.3812) for

model showing clear evidences to pick multiple linear regression model as the best model out of the two models for decision making purpose of professors’ compensation management at a 5% significant level.

By considering the overall findings of the research it is very important consider professors rank, their years since PhD experience and their overall academic experience when deciding suitable remuneration packages and fringe benefits. Further, it very important to motivate themselves to further associate with academic excellence oriented service to give best out of best for the university students. It is very important to get rid of all type of biasness such as gender, discipline in taking decisions on compensation management in higher education sector.