**Salary Hike Data Set**

**The following are the regression results:**

Summary Statistics

> summary(salary\_data)

YearsExperience Salary

Min. : 1.100 Min. : 37731

1st Qu.: 3.200 1st Qu.: 56721

Median : 4.700 Median : 65237

Mean : 5.313 Mean : 76003

3rd Qu.: 7.700 3rd Qu.:100545

Max. :10.500 Max. :122391

> reg <- lm(Salary ~ YearsExperience) # lm(Y ~ X)

>

> summary(reg)

Call:

lm(formula = Salary ~ YearsExperience)

Residuals:

Min 1Q Median 3Q Max

-7958.0 -4088.5 -459.9 3372.6 11448.0

Coefficients:

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

(Intercept) 25792.2 2273.1 11.35 5.51e-12 \*\*\*

YearsExperience 9450.0 378.8 24.95 < 2e-16 \*\*\*

---

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

Residual standard error: 5788 on 28 degrees of freedom

Multiple R-squared: 0.957, Adjusted R-squared: 0.9554

F-statistic: 622.5 on 1 and 28 DF, p-value: < 2.2e-16

>

The R-Squared value without any transformation is 0.957

**With Log Transformation:**

> reg\_log <- lm(Salary ~ log(YearsExperience)) # lm(Y ~ X)

>

> summary(reg\_log)

Call:

lm(formula = Salary ~ log(YearsExperience))

Residuals:

Min 1Q Median 3Q Max

-15392.6 -7523.0 559.7 6336.1 20629.8

Coefficients:

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

(Intercept) 14928 5156 2.895 0.00727 \*\*

log(YearsExperience) 40582 3172 12.792 3.25e-13 \*\*\*

---

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

Residual standard error: 10660 on 28 degrees of freedom

Multiple R-squared: 0.8539, Adjusted R-squared: 0.8487

F-statistic: 163.6 on 1 and 28 DF, p-value: 3.25e-13

> The R-Squared value from log transformation is 0.8539

**Using Exponential Model:**

> reg\_exp <- lm(log(Salary) ~ YearsExperience) #lm(log(Salary) ~ YearsExperience)

>

> summary(reg\_exp)

Call:

lm(formula = log(Salary) ~ YearsExperience)

Residuals:

Min 1Q Median 3Q Max

-0.18949 -0.06946 -0.01068 0.06932 0.19029

Coefficients:

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

(Intercept) 10.507402 0.038443 273.33 <2e-16 \*\*\*

YearsExperience 0.125453 0.006406 19.59 <2e-16 \*\*\*

---

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

Residual standard error: 0.09789 on 28 degrees of freedom

Multiple R-squared: 0.932, Adjusted R-squared: 0.9295

F-statistic: 383.6 on 1 and 28 DF, p-value: < 2.2e-16

> The R-Squared value from log transformation is 0.932

**Using Polynomial model with 2 degrees.**

> reg2degree <- lm(log(Salary) ~ YearsExperience+ I(YearsExperience\*YearsExperience))

>

> summary(reg2degree)

Call:

lm(formula = log(Salary) ~ YearsExperience + I(YearsExperience \*

YearsExperience))

Residuals:

Min 1Q Median 3Q Max

-0.156176 -0.052355 -0.000915 0.048548 0.156817

Coefficients:

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

(Intercept) 10.336852 0.066962 154.368 < 2e-16 \*\*\*

YearsExperience 0.202382 0.026625 7.601 3.55e-08 \*\*\*

I(YearsExperience \* YearsExperience) -0.006614 0.002236 -2.957 0.00638 \*\*

---

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

Residual standard error: 0.08664 on 27 degrees of freedom

Multiple R-squared: 0.9486, Adjusted R-squared: 0.9448

F-statistic: 249.2 on 2 and 27 DF, p-value: < 2.2e-16

>

The R-Squared value from exponential transformation is 0.9486

**Polynomial model with 3 degree**

> reg3degree<-lm(log(Salary)~YearsExperience+ I(YearsExperience\*YearsExperience) + I(YearsExperience\*YearsExperience\*YearsExperience))

>

> summary(reg3degree)

Call:

lm(formula = log(Salary) ~ YearsExperience + I(YearsExperience \*

YearsExperience) + I(YearsExperience \* YearsExperience \*

YearsExperience))

Residuals:

Min 1Q Median 3Q Max

-0.151492 -0.064233 -0.005077 0.041586 0.174288

Coefficients:

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

(Intercept) 10.448472 0.112072 93.230 <2e-16 \*\*\*

YearsExperience 0.114325 0.075993 1.504 0.145

I(YearsExperience \* YearsExperience) 0.011769 0.015042 0.782 0.441

I(YearsExperience \* YearsExperience \* YearsExperience) -0.001085 0.000878 -1.236 0.228

---

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

Residual standard error: 0.08581 on 26 degrees of freedom

Multiple R-squared: 0.9515, Adjusted R-squared: 0.9459

F-statistic: 169.9 on 3 and 26 DF, p-value: < 2.2e-16

>

The R-squared value increased from 0.9486 to 0.9515

The best model is obtained with the R-squared value of 0.957 without any transformations.

The polynomial model with 3 degrees provides a R-squared value of 0.9515 but he p value of the independent variable i.e YearsExperience is beyond 0.05. This means that the variable is not significate enough to predict salary hikes in this model with 3 degrees transformation.