**Employee Data Set**

**The following are the regression results:**

> model1 <- lm(Calories.Consumed ~ Weight.gained..grams.) # lm(Y ~ X)

>

> summary(model1)

Call:

lm(formula = Churn\_out\_rate ~ Salary\_hike)

Residuals:

Min 1Q Median 3Q Max

-3.804 -3.059 -1.819 2.430 8.072

Coefficients:

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

(Intercept) 244.36491 27.35194 8.934 1.96e-05 \*\*\*

Salary\_hike -0.10154 0.01618 -6.277 0.000239 \*\*\*

---

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

Residual standard error: 4.469 on 8 degrees of freedom

Multiple R-squared: 0.8312, Adjusted R-squared: 0.8101

F-statistic: 39.4 on 1 and 8 DF, p-value: 0.0002386

The R-Squared value without any transformation is 0.8312

**With Log Transformation:**

> model1 <- lm(Calories.Consumed ~ Weight.gained..grams.) # lm(Y ~ X)

>

> summary(model1)

Call:

lm(formula = Calories.Consumed ~ Weight.gained..grams.)

Residuals:

Min 1Q Median 3Q Max

-450.41 -115.03 -41.46 194.55 375.75

Coefficients:

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

(Intercept) 1577.201 100.541 15.69 2.33e-09 \*\*\*

Weight.gained..grams. 2.134 0.209 10.21 2.86e-07 \*\*\*

---

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

Residual standard error: 251.5 on 12 degrees of freedom

Multiple R-squared: 0.8968, Adjusted R-squared: 0.8882

F-statistic: 104.3 on 1 and 12 DF, p-value: 2.856e-07

> plot(log(Calories.Consumed),Weight.gained..grams.)

> cor(log(Calories.Consumed),Weight.gained..grams.)

[1] 0.8987253

>

> model1\_log <- lm(Calories.Consumed ~ log(Weight.gained..grams.)) # lm(Y ~ X)

>

> summary(model1\_log)

Call:

lm(formula = Calories.Consumed ~ log(Weight.gained..grams.))

Residuals:

Min 1Q Median 3Q Max

-445.19 -152.38 2.36 68.27 616.01

Coefficients:

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

(Intercept) -1911.12 464.19 -4.117 0.00143 \*\*

log(Weight.gained..grams.) 774.17 83.46 9.276 8.02e-07 \*\*\*

---

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

Residual standard error: 273.9 on 12 degrees of freedom

Multiple R-squared: 0.8776, Adjusted R-squared: 0.8674

F-statistic: 86.04 on 1 and 12 DF, p-value: 8.018e-07

The R-Squared value is 0.877. This value has reduced from 0.89

**Using Exponential Model:**

> model1\_exp <- lm(log(Calories.Consumed) ~ Weight.gained..grams.) #lm(log(Y) ~ X)

>

> summary(model1\_exp)

Call:

lm(formula = log(Calories.Consumed) ~ Weight.gained..grams.)

Residuals:

Min 1Q Median 3Q Max

-0.271626 -0.074071 0.009187 0.111771 0.196081

Coefficients:

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

(Intercept) 7.4068383 0.0577006 128.4 < 2e-16 \*\*\*

Weight.gained..grams. 0.0008517 0.0001200 7.1 1.25e-05 \*\*\*

---

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

Residual standard error: 0.1443 on 12 degrees of freedom

Multiple R-squared: 0.8077, Adjusted R-squared: 0.7917

F-statistic: 50.4 on 1 and 12 DF, p-value: 1.248e-05

The R-squared value has further reduced from 0.877 from the log transformation to exponential transformation providing the value 0.80

**Using Polynomial model with 2 degrees.**

model12degree <- lm(log(Weight.gained..grams.) ~ Calories.Consumed + I(Calories.Consumed\*Calories.Consumed))

>

> summary(model12degree)

Call:

lm(formula = log(Weight.gained..grams.) ~ Calories.Consumed +

I(Calories.Consumed \* Calories.Consumed))

Residuals:

Min 1Q Median 3Q Max

-0.86562 -0.10529 0.02403 0.13541 0.42759

Coefficients:

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

(Intercept) 2.829e+00 1.062e+00 2.664 0.022 \*

Calories.Consumed 1.142e-03 8.807e-04 1.297 0.221

I(Calories.Consumed \* Calories.Consumed) -1.675e-09 1.707e-07 -0.010 0.992

---

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

Residual standard error: 0.3461 on 11 degrees of freedom

Multiple R-squared: 0.8776, Adjusted R-squared: 0.8553

F-statistic: 39.44 on 2 and 11 DF, p-value: 9.611e-06

The R-squared value improved from 0.80 to 0.877

The regression model without any transformation is providing the best R-squared value in the case of Calories Consumed dataset where Weight gained is the dependent Variable “Y” and Calories Consumed is the independent variable “X”.