|  |
| --- |
| > library(readr)  > cal\_consumed <- read.csv(file.choose())  > View(cal\_consumed)  > cal\_consumed  Weight.gained..grams. Calories.Consumed  1 108 1500  2 200 2300  3 900 3400  4 200 2200  5 300 2500  6 110 1600  7 128 1400  8 62 1900  9 600 2800  10 1100 3900  11 100 1670  12 150 1900  13 350 2700  14 700 3000  > # Exploratory data analysis #  > summary(cal\_consumed)  Weight.gained..grams. Calories.Consumed  Min. : 62.0 Min. :1400  1st Qu.: 114.5 1st Qu.:1728  Median : 200.0 Median :2250  Mean : 357.7 Mean :2341  3rd Qu.: 537.5 3rd Qu.:2775  Max. :1100.0 Max. :3900  > # Scatter plot #  > plot(cal\_consumed$Weight.gained..grams., cal\_consumed$Calories.Consumed) # plot(X,Y)  > attach(cal\_consumed)  > # Correlation Coefficient (r) #  > cor(Waight.gained..grams., Calories.Consumed)  > # Correlation Coefficient (r) #  > cor(Weight.gained..grams., Calories.Consumed)  [1] 0.946991  > # Simple Linear Regression Model #  > reg <- lm(Calories.Consumed ~ Weight.gained..grams.) #lm(Y ~ X)  > summary(reg)  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  > pred <- predict(reg)  > pred  1 2 3 4 5 6 7 8 9 10  1807.718 2004.085 3498.181 2004.085 2217.528 1811.987 1850.407 1709.535 2857.854 3925.066  11 12 13 14  1790.643 1897.364 2324.249 3071.297  > reg$residuals  1 2 3 4 5 6 7  -307.718381 295.914706 -98.181364 195.914706 282.472411 -211.987227 -450.406841  8 9 10 11 12 13 14  190.465075 -57.854477 -25.065955 -120.642998 2.635854 375.751263 -71.296772  > sum(reg$residuals)  [1] -1.847411e-13  > mean(reg$residuals)  [1] -1.320075e-14  > sqrt(sum(reg$residuals^2)/nrow(cal\_consumed)) #RMSE  [1] 232.8335  > sqrt(mean(reg$residuals^2))  [1] 232.8335  > confint(reg,level = 0.95)  2.5 % 97.5 %  (Intercept) 1358.141455 1796.259949  Weight.gained..grams. 1.678994 2.589852  > predict(reg,interval="predict")  fit lwr upr  1 1807.718 1229.249 2386.187  2 2004.085 1432.376 2575.795  3 3498.181 2879.564 4116.799  4 2004.085 1432.376 2575.795  5 2217.528 1649.740 2785.316  6 1811.987 1233.697 2390.278  7 1850.407 1273.660 2427.154  8 1709.535 1126.585 2292.485  9 2857.854 2280.041 3435.668  10 3925.066 3264.781 4585.351  11 1790.643 1211.447 2369.839  12 1897.364 1322.350 2472.378  13 2324.249 1757.059 2891.439  14 3071.297 2483.085 3659.509  Warning message:  In predict.lm(reg, interval = "predict") :  predictions on current data refer to \_future\_ responses  > predict  standardGeneric for "predict" defined from package "stats"  function (object, ...)  standardGeneric("predict")  <environment: 0x000001923d025518>  Methods may be defined for arguments: object  Use showMethods("predict") for currently available ones.  > ## ggplot for adding regresion line for data ##  > library(ggplot2)  > ggplot(data = cal\_consumed, aes(x = Weight.gained..grams., y = Calories.Consumed)) +  + geom\_point(color='blue') +  + geom\_line(color='red',data = cal\_consumed, aes(x=Weight.gained..grams., y=pred))  > ## Logrithamic Model/ Transformation ##  > # x = log(Weight.gained..grams.); y = Calories.Consumed  > plot(log(Weight.gained..grams.), Calories.Consumed)  > cor(log(Weight.gained..grams.), Calories.Consumed)  [1] 0.9368037  > reg\_log <- lm(Calories.Consumed ~ log(Weight.gained..grams.)) ## lm(Y ~ X)  > summary(reg\_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  > predict(reg\_log)  1 2 3 4 5 6 7 8 9 10  1713.658 2190.693 3355.110 2190.693 2504.593 1727.863 1845.189 1283.994 3041.209 3510.464  11 12 13 14  1654.077 1967.977 2623.933 3160.549  > reg\_log$residuals  1 2 3 4 5 6 7  -213.657759 109.307231 44.890293 9.307231 -4.593132 -127.863176 -445.189099  8 9 10 11 12 13 14  616.006117 -241.209345 389.536348 15.923444 -67.976918 76.067489 -160.548724  > sqrt(sum(reg\_log$residuals^2)/nrow(cal\_consumed)) ###RMSE  [1] 253.558  > confint(reg\_log, level = 0.95)  2.5 % 97.5 %  (Intercept) -2922.5030 -899.7458  log(Weight.gained..grams.) 592.3259 956.0212  > predict(reg\_log,interval = "confidence")  fit lwr upr  1 1713.658 1496.567 1930.749  2 2190.693 2027.366 2354.020  3 3355.110 3068.390 3641.830  4 2190.693 2027.366 2354.020  5 2504.593 2340.533 2668.653  6 1727.863 1513.022 1942.704  7 1845.189 1647.751 2042.627  8 1283.994 988.960 1579.028  9 3041.209 2812.064 3270.355  10 3510.464 3192.769 3828.158  11 1654.077 1427.257 1880.896  12 1967.977 1786.044 2149.910  13 2623.933 2451.133 2796.732  14 3160.549 2910.512 3410.585  > ## Exponential Transformation ##  > ## x =Weight.gained..grams. and y = log(Calories\_Consumed)  > plot(Weight.gained..grams., log(Calories.Consumed))  > cor(Weight.gained..grams., log(Calories.Consumed))  [1] 0.8987253  > reg\_exp <- lm(log(Calories.Consumed) ~ Weight.gained..grams.) ## lm(log(Y) ~ X)  > summary(reg\_exp)  Call:  lm(formula = log(Calories.Consumed) ~ Weight.gained..grams.)  Residuals:  Min 1Q Median 3Q Max  -0.271626 -0.074071 0.009187 0.111771 0.196081  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 7.4068383 0.0577006 128.4 < 2e-16 \*\*\*  Weight.gained..grams. 0.0008517 0.0001200 7.1 1.25e-05 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 0.1443 on 12 degrees of freedom  **Multiple R-squared: 0.8077, Adjusted R-squared: 0.7917**  F-statistic: 50.4 on 1 and 12 DF, p-value: 1.248e-05  > reg\_exp$residuals  1 2 3 4 5 6 7  -0.185599143 0.163490516 -0.041817626 0.119038753 0.161704346 -0.122763977 -0.271625570  8 9 10 11 12 13 14  0.089966814 0.019529696 -0.074952062 -0.071427202 0.015019168 0.196081498 0.003354788  > sqrt(mean(reg\_exp$residuals^2))  [1] 0.1336239  > logat <- predict(reg\_exp)  > logat  1 2 3 4 5 6 7 8 9 10  7.498820 7.577174 8.173348 7.577174 7.662342 7.500523 7.515853 7.459642 7.917845 8.343684  11 12 13 14  7.492006 7.534590 7.704926 8.003013  > at <- exp(logat)  > error = cal\_consumed$Calories.Consumed -Calories.Consumed  > error  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0  > sqrt(sum(error^2)/nrow(cal\_consumed))  [1] 0  > confint(reg\_exp,level = 0.95)  2.5 % 97.5 %  (Intercept) 7.2811195444 7.53255711  Weight.gained..grams. 0.0005903057 0.00111305  > predict(reg\_exp,interval = "confidence")  fit lwr upr  1 7.498820 7.392407 7.605232  2 7.577174 7.483564 7.670784  3 8.173348 8.008566 8.338131  4 7.577174 7.483564 7.670784  5 7.662342 7.576953 7.747730  6 7.500523 7.394430 7.606615  7 7.515853 7.412565 7.619142  8 7.459642 7.345460 7.573825  9 7.917845 7.812612 8.023078  10 8.343684 8.132249 8.555118  11 7.492006 7.384299 7.599714  12 7.534590 7.434535 7.634645  13 7.704926 7.620856 7.788995  14 8.003013 7.880263 8.125762 |
|  |
| |  | | --- | | > | |