Econometric\_regression

name

2023-12-10

#Loading the data

library(readxl)  
df<- read\_excel("2013lifeexpectancy.xls")  
View(df)

# Simple linear Model

library(knitr)  
# Fit the simple linear regression model  
model\_simple <- lm(Life\_Expectancy ~ Alcohol, data = df)  
  
# Print the summary of the model  
summary(model\_simple)

##   
## Call:  
## lm(formula = Life\_Expectancy ~ Alcohol, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.278 -3.917 1.060 5.467 13.242   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 67.7305 0.7564 89.55 < 2e-16 \*\*\*  
## Alcohol 0.9441 0.1320 7.15 2.5e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.131 on 169 degrees of freedom  
## Multiple R-squared: 0.2322, Adjusted R-squared: 0.2277   
## F-statistic: 51.12 on 1 and 169 DF, p-value: 2.498e-11

library(modelsummary)  
modelsummary(model\_simple, output = "markdown")

|  | (1) |
| --- | --- |
| (Intercept) | 67.731 |
|  | (0.756) |
| Alcohol | 0.944 |
|  | (0.132) |
| Num.Obs. | 171 |
| R2 | 0.232 |
| R2 Adj. | 0.228 |
| AIC | 1161.1 |
| BIC | 1170.5 |
| Log.Lik. | -577.543 |
| F | 51.123 |
| RMSE | 7.09 |

# Multilinear regression model I

# Fit the multiple linear regression model  
model\_multiple <- lm(Life\_Expectancy ~ Alcohol + Schooling + BMI, data = df)  
  
# Print the summary of the model  
summary(model\_multiple)

##   
## Call:  
## lm(formula = Life\_Expectancy ~ Alcohol + Schooling + BMI, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.3144 -2.8698 0.3882 3.0997 14.7307   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42.24690 1.94126 21.763 < 2e-16 \*\*\*  
## Alcohol 0.04870 0.11092 0.439 0.66116   
## Schooling 2.05413 0.17730 11.586 < 2e-16 \*\*\*  
## BMI 0.06135 0.02083 2.946 0.00368 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.873 on 167 degrees of freedom  
## Multiple R-squared: 0.6457, Adjusted R-squared: 0.6393   
## F-statistic: 101.5 on 3 and 167 DF, p-value: < 2.2e-16

modelsummary(model\_multiple, output = "markdown")

|  | (1) |
| --- | --- |
| (Intercept) | 42.247 |
|  | (1.941) |
| Alcohol | 0.049 |
|  | (0.111) |
| Schooling | 2.054 |
|  | (0.177) |
| BMI | 0.061 |
|  | (0.021) |
| Num.Obs. | 171 |
| R2 | 0.646 |
| R2 Adj. | 0.639 |
| AIC | 1032.8 |
| BIC | 1048.5 |
| Log.Lik. | -511.421 |
| F | 101.456 |
| RMSE | 4.82 |

# F-test

# Fit the simple model without new variables  
model\_simple <- lm(Life\_Expectancy ~ Alcohol, data = df)  
  
# Perform ANOVA test to compare models  
anova\_result <- anova(model\_simple, model\_multiple)  
print(anova\_result)

## Analysis of Variance Table  
##   
## Model 1: Life\_Expectancy ~ Alcohol  
## Model 2: Life\_Expectancy ~ Alcohol + Schooling + BMI  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 169 8592.7   
## 2 167 3965.2 2 4627.5 97.447 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Print the ANOVA results using kable  
kable(anova\_result, format = "markdown", caption = "ANOVA Test Results for Model Comparison")

ANOVA Test Results for Model Comparison

| Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
| --- | --- | --- | --- | --- | --- |
| 169 | 8592.719 | NA | NA | NA | NA |
| 167 | 3965.213 | 2 | 4627.506 | 97.44667 | 0 |

# Multilinear regression model II

# Fit the multiple linear regression model with a quadratic term for Schooling  
model\_quadratic <- lm(Life\_Expectancy ~ Alcohol + Schooling + I((Schooling)^2) + BMI, data = df)  
  
# Print the summary of the model  
summary(model\_quadratic)

##   
## Call:  
## lm(formula = Life\_Expectancy ~ Alcohol + Schooling + I((Schooling)^2) +   
## BMI, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.0643 -2.8445 0.4559 3.0709 13.6225   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 45.89101 5.49711 8.348 2.59e-14 \*\*\*  
## Alcohol 0.03353 0.11313 0.296 0.7673   
## Schooling 1.44135 0.88267 1.633 0.1044   
## I((Schooling)^2) 0.02452 0.03460 0.709 0.4795   
## BMI 0.06261 0.02093 2.991 0.0032 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.88 on 166 degrees of freedom  
## Multiple R-squared: 0.6468, Adjusted R-squared: 0.6383   
## F-statistic: 75.99 on 4 and 166 DF, p-value: < 2.2e-16

modelsummary(model\_quadratic, output = "markdown")

|  | (1) |
| --- | --- |
| (Intercept) | 45.891 |
|  | (5.497) |
| Alcohol | 0.034 |
|  | (0.113) |
| Schooling | 1.441 |
|  | (0.883) |
| I((Schooling)^2) | 0.025 |
|  | (0.035) |
| BMI | 0.063 |
|  | (0.021) |
| Num.Obs. | 171 |
| R2 | 0.647 |
| R2 Adj. | 0.638 |
| AIC | 1034.3 |
| BIC | 1053.2 |
| Log.Lik. | -511.162 |
| F | 75.990 |
| RMSE | 4.81 |

# Logarithmic model

# Fit the logarithmic model  
log\_model <- lm(log(Life\_Expectancy) ~ Alcohol + log(Schooling) + BMI, data = df)  
  
# Print the summary of the logarithmic model  
summary(log\_model)

##   
## Call:  
## lm(formula = log(Life\_Expectancy) ~ Alcohol + log(Schooling) +   
## BMI, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.284775 -0.037745 0.005375 0.046391 0.278812   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.4008213 0.0730552 46.551 < 2e-16 \*\*\*  
## Alcohol 0.0016833 0.0016777 1.003 0.31715   
## log(Schooling) 0.3211605 0.0315668 10.174 < 2e-16 \*\*\*  
## BMI 0.0010192 0.0003244 3.142 0.00199 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07588 on 167 degrees of freedom  
## Multiple R-squared: 0.5967, Adjusted R-squared: 0.5895   
## F-statistic: 82.37 on 3 and 167 DF, p-value: < 2.2e-16