# KIN 610 - Lab 3

**Questions**

1. What is the primary purpose of conducting a regression analysis?
2. In simple regression analysis, what is the role of the independent variable and the dependent variable?
3. How can you determine the strength and direction of the relationship between the independent and dependent variables in a simple regression analysis?
4. What is the difference between simple and multiple regression analysis?
5. How do you interpret the R-squared value in a regression analysis?
6. In the simple regression analysis predicting VO2max based on Age, what is the direction of the relationship between Age and VO2max?
7. In the multiple regression analysis predicting VO2max, which variable(s) had a significant effect on VO2max?
8. In the multiple regression analysis predicting VO2max, which variable(s) had the strongest effect on VO2max?
9. How can multicollinearity affect the results of a multiple regression analysis?

**Answers**

1. Regression analysis is conducted to explain the relationship between variables. It can also help to predict outcomes and changes in a variable based on its association with other variables, estimate a variable's value using the values of other known variables, and manage the impact of variables while exploring their relationship.
2. In the context of regression analysis, the "independent variable" is the predictor variable, and its value is not influenced by other variables in your study. The "dependent variable" is the outcome variable (variable being predicted) and its value changes in response to changes in the independent variable. To express this relationship, we use the following equation: $Y = a + bX + \epsilon$, where Y represents the dependent variable and X represents the independent/explanatory variable.
3. In a simple regression analysis, the correlation coefficient determines the strength and direction of the relationship between the independent and dependent variables. This coefficient is between -1 and 1, indicating the similarity of measurements across a dataset for multiple variables. It reflects the relationship's strength and direction between the variables.
4. Simple regression establishes the relationship between two variables, whereas multiple regression estimates the relationship between one dependent variable and two or more independent variables.
5. The R-squared value also called the coefficient of determination, is a statistical measure that shows how much of the variance in a dependent variable is explained by one or more independent variables in a regression model. R-squared is commonly used to assess how well the regression model explains the observed data. For instance, if the R-squared is 60%, the regression model accounts for 60% of the variability observed in the target variable. A higher R-squared value indicates that the model explains more variability. Do you have any other questions?
6. The directions of the relationship is **negative** - reefer to the scatterplot and the r value in the Matrix table.
7. The only variable that had a significant effect on the outcome variable is AGE - refer to the Model Coefficients table.
8. The variable that had the strongest correlation with the outcome variable was AGE - refer to the Correlation Matrix table.
9. Multicollinearity happens when two or more independent variables in a multiple regression model are strongly correlated. This can create issues when fitting and interpreting the regression model because it becomes challenging for the model to accurately estimate the relationship between each independent variable and the dependent variable. Furthermore, multicollinearity can cause some of the significant variables being studied to become statistically insignificant.

# Simple Linear Regression

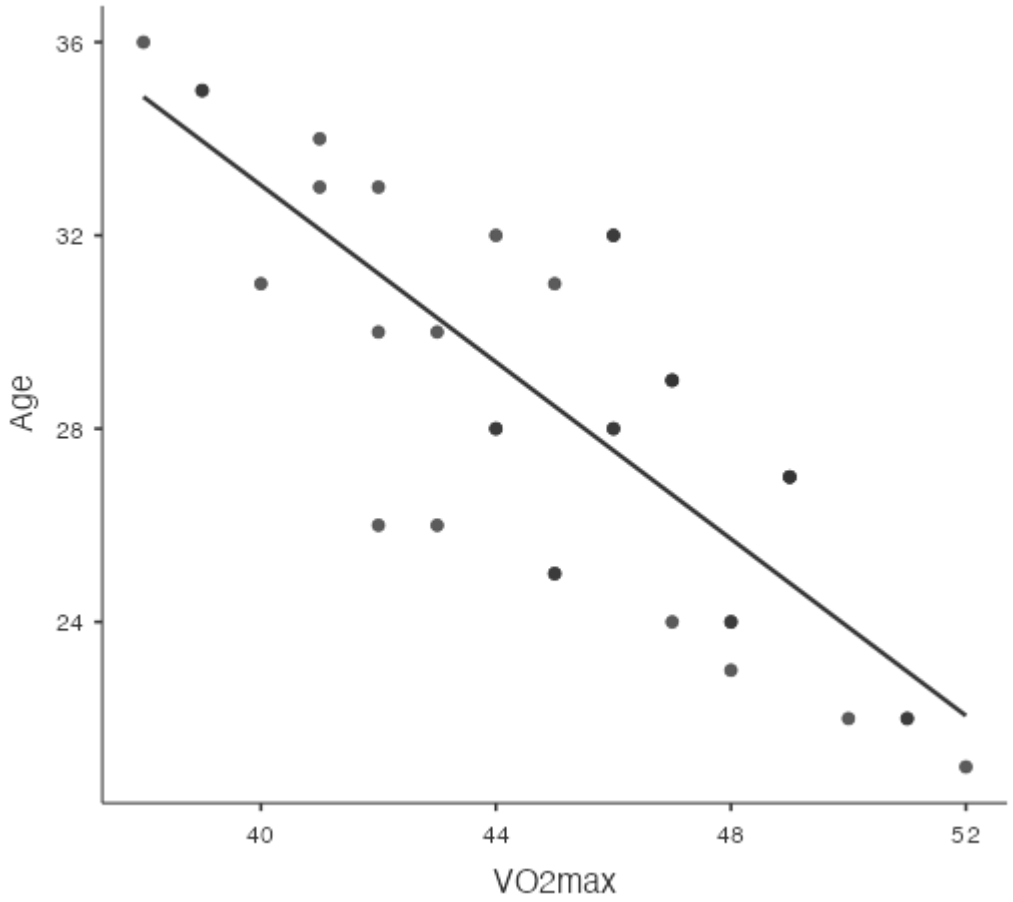
Model Fit Measures

| Model | R | $R^2$ |
|---|---|---|
| 1 | 0.809 | 0.654 |

Model Coefficients - VO2max

| Predictor | Estimate | SE | t | p |
|---|---|---|---|---|
| Intercept | 65.431 | 2.5812 | 25.35 | <.001 |
| Age | -0.715 | 0.0905 | -7.90 | <.001 |

# Scatterplot




## Multiple Linear Regression

Model Fit Measures

| Model | R | $R^2$ |
|---|---|---|
| 1 | 0.859 | 0.738 |

Model Coefficients - VO2max

| Predictor | Estimate | SE | t | p |
|---|---|---|---|---|
| Intercept | 274.858 | 123.693 | 2.22 | 0.034 |
| Age | -0.798 | 0.139 | -5.74 | <.001 |
| Height | -1.161 | 0.710 | -1.63 | 0.113 |
| Weight | 1.295 | 0.842 | 1.54 | 0.134 |
| BMI | -4.156 | 2.621 | -1.59 | 0.123 |

## Strenght of correlations

Correlation Matrix

| | VO2max | Age | Height | Weight | BMI |
|---|---|---|---|---|---|
| VO2max | — | | | | |
| Age | -0.809 | — | | | |
| Height | -0.229 | -0.024 | — | | |
| Weight | -0.357 | 0.141 | 0.978 | — | |
| BMI | -0.574 | 0.454 | 0.752 | 0.870 | — |

# References


**[1]** The jamovi project (2022). *jamovi*. (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org.

**[2]** R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).