

# STA 3180 Statistical Modelling: Regression

## Regression Lecture Notes

### Introduction:

Regression is a statistical technique used to analyze the relationship between one or more independent variables (also known as predictors) and a dependent variable (also known as the response variable). Regression models are used to explain, predict, and forecast a wide variety of phenomena.

### Key Concepts:

1. **Linear Regression:** Linear regression is a type of regression analysis that models the relationship between a dependent variable (response variable) and one or more independent variables (predictor variables). It assumes that the relationship between the two variables is linear, meaning that the change in the dependent variable is directly proportional to the change in the independent variable.
2. **Multiple Linear Regression:** Multiple linear regression is a type of regression analysis that models the relationship between a dependent variable and two or more independent variables. It is similar to linear regression, except that it takes into account the effects of multiple independent variables on the dependent variable.
3. **Logistic Regression:** Logistic regression is a type of regression analysis that models the relationship between a dependent variable and one or more independent variables. Unlike linear regression, which assumes a linear relationship between the dependent and independent variables, logistic regression assumes a logistic (s-shaped) relationship between the dependent and independent variables.
4. **Polynomial Regression:** Polynomial regression is a type of regression analysis that models the relationship between a dependent variable and one or more independent variables using a polynomial equation. It is similar to linear regression, except that it takes into account the effects of higher-order terms (such as squares and cubes) of the independent variables on the dependent variable.
5. **Stepwise Regression:** Stepwise regression is a type of regression analysis that models the relationship between a dependent variable and one or more independent variables by adding or removing variables from the model in a stepwise fashion. It is used to identify the most important independent variables that have the greatest effect on the dependent variable.

### Definitions:

1. **Dependent Variable (Response Variable):** The dependent variable (also known as the response variable) is the variable that is being predicted or explained by the regression model.
2. **Independent Variable (Predictor Variable):** The independent variable (also known as the predictor variable) is the variable that is used to predict or explain the dependent variable.
3. **Linear Relationship:** A linear relationship is a type of relationship between two variables in which the change in the dependent variable is directly proportional to the change in the independent variable.

4. Logistic Relationship: A logistic relationship is a type of relationship between two variables in which the change in the dependent variable is not directly proportional to the change in the independent variable.

5. Polynomial Equation: A polynomial equation is an equation that contains one or more terms that are raised to a power greater than one.

Rules:

1. Linear regression models assume that the relationship between the dependent and independent variables is linear.

2. Multiple linear regression models assume that the relationship between the dependent and independent variables is linear.

3. Logistic regression models assume that the relationship between the dependent and independent variables is logistic (s-shaped).

4. Polynomial regression models assume that the relationship between the dependent and independent variables is non-linear.

5. Stepwise regression models use a stepwise approach to identify the most important independent variables that have the greatest effect on the dependent variable.

Examples:

1. Linear Regression: A researcher wants to understand the relationship between the amount of time a student spends studying for an exam and their final grade. The researcher could use linear regression to analyze the data and determine if there is a linear relationship between the two variables.

2. Multiple Linear Regression: A researcher wants to understand the relationship between a student's GPA, SAT score, and college major on their chances of getting accepted into a prestigious university. The researcher could use multiple linear regression to analyze the data and determine if there is a linear relationship between the three variables.

3. Logistic Regression: A researcher wants to understand the relationship between a person's age, gender, and income on their likelihood of voting in an election. The researcher could use logistic regression to analyze the data and determine if there is a logistic (s-shaped) relationship between the three variables.

4. Polynomial Regression: A researcher wants to understand the relationship between a person's height and weight. The researcher could use polynomial regression to analyze the data and determine if there is a non-linear relationship between the two variables.

5. Stepwise Regression: A researcher wants to understand the relationship between a person's age, gender, income, and education level on their likelihood of getting a job. The researcher could use stepwise regression to analyze the data and determine which of the four variables has the greatest effect on the dependent variable.