1. In a linear equation, what is the difference between a dependent variable and an independent variable?

The independent variable is the variable the experimenter manipulates or changes, and is assumed to have a direct effect on the **dependent** variable. The dependent variable is the variable being tested and measured in an experiment, and is 'dependent' on the independent variable.

2. What is the concept of simple linear regression? Give a specific example.

**Simple linear regression** is used to estimate the relationship between**two**[**quantitative variables**](https://www.scribbr.com/methodology/types-of-variables/#quantitative-vs-categorical). You can use simple linear regression when you want to know:

1. How strong the relationship is between two variables (e.g. the relationship between rainfall and soil erosion).
2. The value of the dependent variable at a certain value of the independent variable (e.g. the amount of soil erosion at a certain level of rainfall).

Example: You are a social researcher interested in the relationship between income and happiness. You survey 500 people whose incomes range from $15k to $75k and ask them to rank their happiness on a scale from 1 to 10.

Your independent variable (income) and dependent variable (happiness) are both quantitative, so you can do a regression analysis to see if there is a linear relationship between them.

3. In a linear regression, define the slope.

In a regression context, the slope is the heart and soul of the equation because **it tells you how much you can expect Y to change as X increases**. In general, the units for slope are the units of the Y variable per units of the X variable. It's a ratio of change in Y per change in X.

4. Determine the graph's slope, where the lower point on the line is represented as (3, 2) and the higher point is represented as (2, 2).

Graph slope is 2

5. In linear regression, what are the conditions for a positive slope?

A positive slope indicates that **the regression line rises as the y-axis values increase**, while a negative slope implies the line falls as y-axis values increase. Pick two points that fall on the regression line.

6. In linear regression, what are the conditions for a negative slope?

A negative slope means that **two variables are negatively related**; that is, when x increases, y decreases, and when x decreases, y increases. Graphically, a negative slope means that as the line on the line graph moves from left to right, the line falls.

7. What is multiple linear regression and how does it work?

Multiple linear regression refers to a **statistical technique that uses two or more independent variables to predict the outcome of a dependent variable**. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance.

Multiple regression is an extension of linear regression models **that allow predictions of systems with multiple independent variables**. It does this by simply adding more terms to the linear regression equation, with each term representing the impact of a different physical parameter.

8. In multiple linear regression, define the number of squares due to error.

9. In multiple linear regression, define the number of squares due to regression.

10. In a regression equation, what is multicollinearity?

Multicollinearity occurs **when two or more independent variables are highly correlated with one another in a regression model**. This means that an independent variable can be predicted from another independent variable in a regression model.

11. What is heteroskedasticity, and what does it mean?

Heteroskedasticity refers to **situations where the variance of the residuals is unequal over a range of measured values**. When running a regression analysis, heteroskedasticity results in an unequal scatter of the residuals (also known as the error term).

12. Describe the concept of ridge regression.

Ridge regression is **a model tuning method that is used to analyse any data that suffers from multicollinearity**. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values to be far away from the actual values.

13. Describe the concept of lasso regression.

Lasso regression is **a type of linear regression that uses shrinkage**. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). The acronym “LASSO” stands for Least Absolute Shrinkage and Selection Operator.

14. What is polynomial regression and how does it work?

Polynomial Regression is a form of Linear regression known as a special case of Multiple linear regression which **estimates the relationship as an nth degree polynomial**. Polynomial Regression is sensitive to outliers so the presence of one or two outliers can also badly affect the performance.

In statistics, polynomial regression is a **form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x**. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

15. Describe the basis function.

In mathematics, a basis function is **an element of a particular basis for a function space**. Every function in the function space can be represented as a linear combination of basis functions, just as every vector in a vector space can be represented as a linear combination of basis vectors.

16. Describe how logistic regression works.

Logistic regression is a statistical analysis **method used to predict a data value based on prior observations of a data set**. Based on historical data about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into a particular outcome category.

Logistic regression uses an **equation as the representation**, very much like linear regression. Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y).