1. Explain the linear regression algorithm in detail.

Ans:

linear regression is a machine learning algorithm that finds the best linear-fit relationship on any given data, between independent and dependent variables. It is mostly done by the Sum of Squared Residuals Method.

It follow simple straight line equation $y = mx + c$

2. What are the assumptions of linear regression regarding residuals?

Ans: Assumptions about the residuals:

1) Normality assumption: It is assumed that the error terms, ε (i), are normally distributed.
2) Zero mean assumption: It is assumed that the residuals have a mean value of zero, i.e., the error terms are normally distributed around zero.
3) Constant variance assumption: It is assumed that the residual terms have the same (but unknown) variance, $\sigma 2$. This assumption is also known as the assumption of homogeneity or homoscedasticity.
4) Independent error assumption: It is assumed that the residual terms are independent of each other, i.e. their pair-wise covariance is zero.

3. What is the coefficient of correlation and the coefficient of determination?

Ans: - Coefficient correlation:

The correlation coefficient is a statistical measure that calculates the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0. Generally in linear regression it is the slope which define the coefficient of correlation.

- Coefficient of determination:

The coefficient of determination can be thought of as a percent. It gives you an idea of how many data points fall within the results of the line formed by the regression equation. The higher the coefficient, the higher percentage of points the line passes through when the data points and line are plotted. If the coefficient is 0.80, then 80% of the points should fall within the regression line. Values of 1 or 0 would indicate the regression line represents all or none of the data, respectively. A higher coefficient is an indicator of a better goodness of fit for the observations.

The CoD can be negative, although this usually means that your model is a poor fit for your data. It can also become negative if you didn't set an intercept.

4. Explain the Anscombe's quartet in detail:

Ans: Anscombe's quartet comprises four data sets that have nearly identical simple descriptive statistics, yet have very different distributions and appear very different when graphed. Each dataset consists of eleven (x,y) points.

- he first scatter plot (top left) appears to be a simple linear relationship, corresponding to two variables correlated where y could be modelled as gaussian with mean linearly dependent on x.
- The second graph (top right) is not distributed normally; while a relationship between the two variables is obvious, it is not linear, and the Pearson correlation coefficient is not relevant. A more general regression and the corresponding coefficient of determination would be more appropriate.
- In the third graph (bottom left), the distribution is linear, but should have a different regression line (a robust regression would have been called for). The calculated regression is offset by the one outlier which exerts enough influence to lower the correlation coefficient from 1 to 0.816.
- Finally, the fourth graph (bottom right) shows an example when one high-leverage point is enough to produce a high correlation coefficient, even though the other data points do not indicate any relationship between the variables.

5. What is Pearson's R?

Ans: Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables; hence the modifier product-moment in the name.

6. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?

Ans:

1. Ease of interpretation.
2. Faster convergence for gradient descent method

Scaling only changes the coefficients and none of the other parameters like t-statistics, F-statistic, p-values and R-squared etc doesn't change at all.

Methods of scaling a variable:

1) Standardisation:

Standardisation basically bring all of the data into a std normal distribution with mean zero and sd as one(1).

$x = x - mean(x)/sd(x)$

2) MinMax scaling:

It brings all the data in the range of 0 and 1.

$x = x- min(x)/max(x) - min(x)$

7. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
   Ans: VIF as infinite indicates that there are multicollinearity in the dataset.

8. What is the Gauss-Markov theorem?
   Ans: The Gauss-Markov theorem states that if your linear regression model satisfies the first six classical assumptions, then ordinary least squares (OLS) regression produces unbiased estimates that have the smallest variance of all possible linear estimators.

9. Explain the gradient descent algorithm in detail:

Gradient descent algorithm is an iterative process that takes us to the minimum of a function(barring some caveats).
ref:
https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e

10. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.

    No Idea on this.