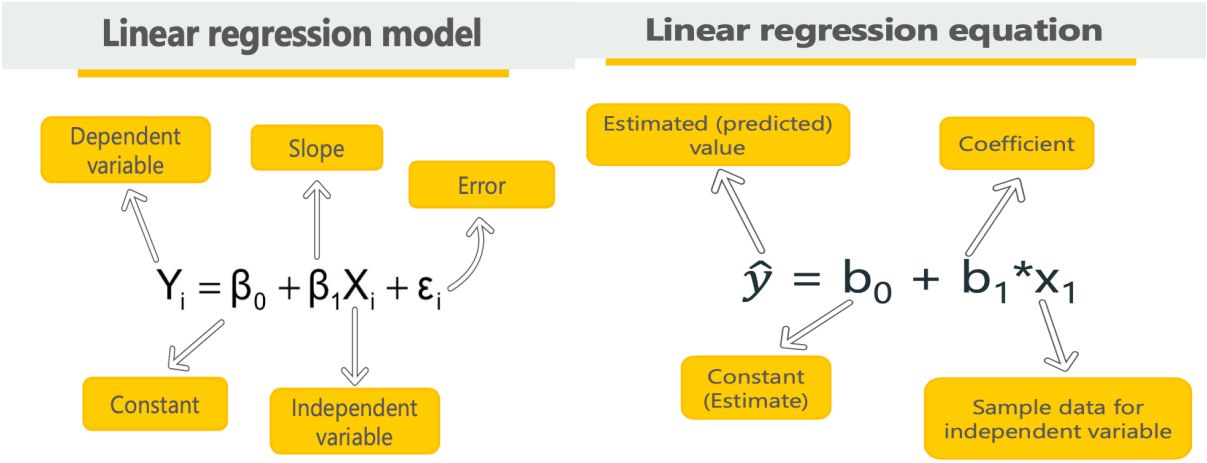
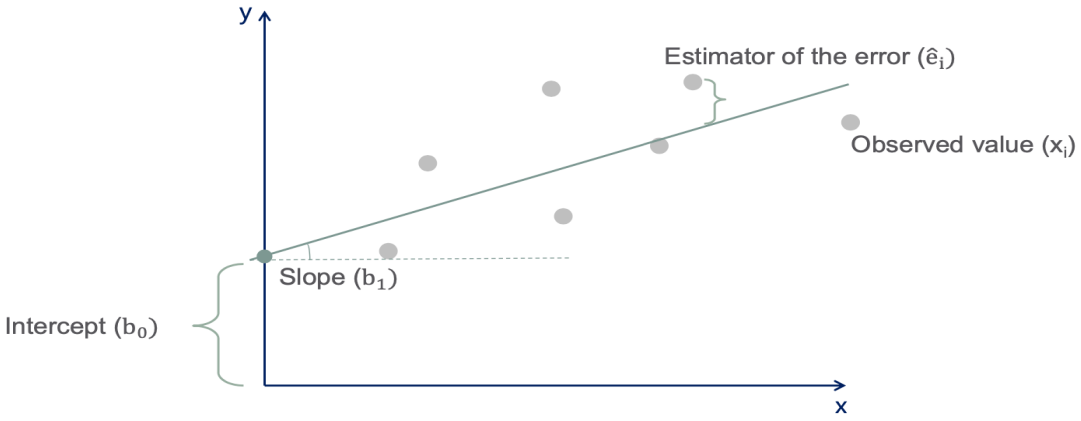
Linear regression and Summary of Error Metrics

Linear regression

A linear regression is a linear approximation of a causal relationship between two or more variables.
Regression models are highly valuable, as they are one of the most common ways to make inferences and predictions.
In simple terms, Linear regression attempts to model the relationship between two variables by fitting a linear equation (= a straight line) to the observed data. One variable is considered to be an explanatory variable or independent variable (e.g. your income), and the other is considered to be a dependent variable (e.g. your expenses).

As many other statistical techniques, regression models help us make predictions about the population based on sample data.





Ordinary least squares
OLS (ordinary least squares) is one of the most common methods for estimating the linear regression equation. However, its simplicity implies that it cannot be always used. Therefore, all OLS regression assumptions should be met before we can rely on this method of estimation.

OLS regression assumptions

1) **Linearity**
   The specified model must represent a linear relationship.
   The relationship between independent variable and dependent variable should be Linear
2) **No multicollinearity**
   No predictor variable should be perfectly (or almost perfectly) explained by the other predictors.
3) **homoscedasticity**
   The variance of the residuals(errors) should be consistent across observations.
4) **Normality**
   Residual should be normally distributed.

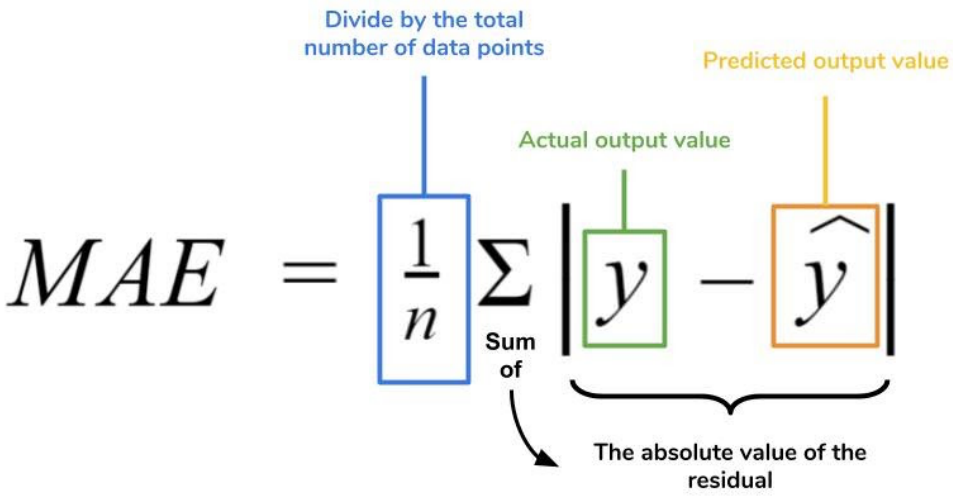
Error metrics
The model will produce an output given any input or set of inputs, we can then check these estimated outputs(y^) against the actual values(y) that we tried to predict. We call the difference between the actual value and the model's estimate a residual. (ε)

We can calculate the residual for every point in our data set, and each of these residuals will be of use in assessment. If our collection of residuals are small, it implies that the model that produced them does a good job at predicting our output of interest. Conversely, if these residuals are generally large, it implies that the model is a poor estimator.

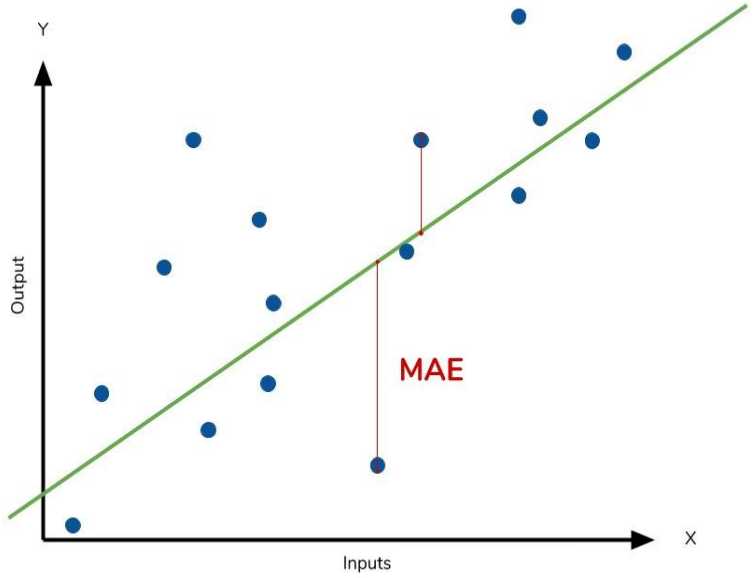
The various error metrics used to evaluate the results of the prediction are :

- Mean Absolute Error (MAE)
- Mean Square Error (MSE)
- Root mean square error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Mean Percentage Error
- R^2 or Coefficient of Determination.

Mean absolute error
The mean absolute error (MAE) is calculated residual for every data point, taking only the absolute value of each so that negative and positive residuals do not cancel out. We then take the average of all these residuals. Effectively, MAE describes the *typical* magnitude of the residuals.



The picture below is a graphical description of the MAE. The green line represents our model's predictions, and the blue points represent our data.



Because we use the absolute value of the residual, the MAE does not indicate underperformance or overperformance of the model (whether or not the model under or overshoots actual data).
small MAE suggests the model is great at prediction, while a large MAE suggests that your model may have trouble in certain areas. A MAE of 0 means that your model is a perfect predictor of the outputs (but this will almost never happen).
While the MAE is easily interpretable, using the absolute value of the residual often is not as desirable as squaring this difference.

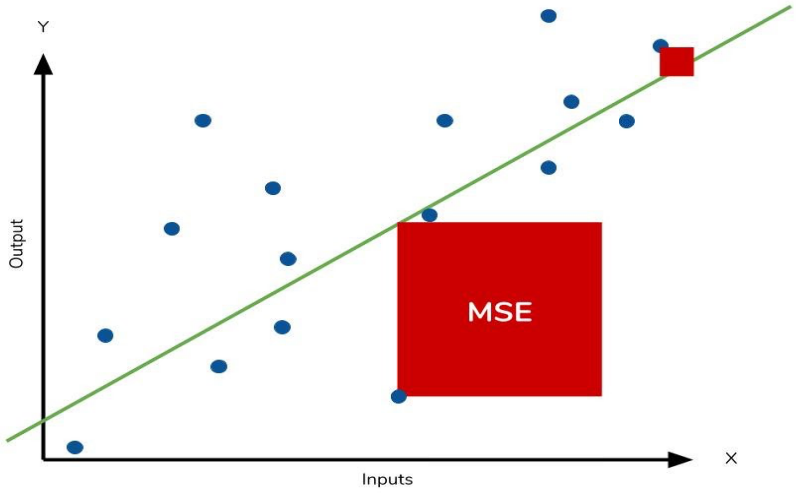
Mean square error

The mean square error (MSE) is just like the MAE, but squares the difference before summing them all instead of using the absolute value.

$$MSE = \frac{1}{n}\Sigma\underbrace{\left(y - \hat{y}\right)^2}_{\text{The square of the difference between actual and predicted}}$$

Because we are squaring the difference, the MSE will almost always be bigger than the MAE.

The outliers in our data will contribute to much higher total error in the MSE than they would the MAE. The following picture graphically demonstrates what an individual residual in the MSE might look like.



Outliers will produce these exponentially larger differences, and it is our job to judge how we should approach them.

Root mean squared error (RMSE)

The root mean squared error (RMSE) - as the name suggests, it is the square root of the MSE. Because the MSE is squared, its units do not match that of the original output. Researchers will often use RMSE to convert the error metric back into similar units, making interpretation easier. Since the MSE and RMSE both square the residual, they are similarly affected by outliers. The RMSE is analogous to the standard deviation (MSE to variance) and is a measure of how large your residuals are spread out. Both MAE and MSE can range from 0 to positive infinity, so as both of these measures get higher, it becomes harder to interpret how well your model is performing.

Mean absolute percentage error

Another way we can summarize the collection of residuals is by using percentages so that each prediction is scaled against the value it's supposed to estimate.

The mean absolute percentage error (MAPE) is the percentage equivalent of MAE. The equation looks just like that of MAE, but with adjustments to convert everything into percentages.

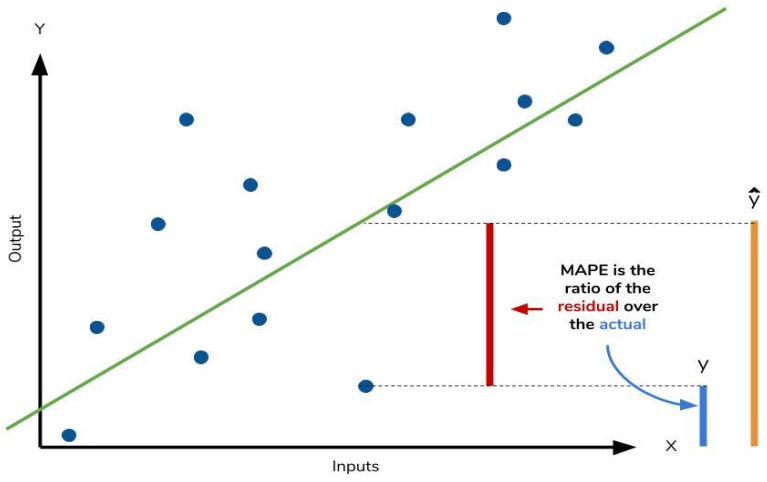
Multiplying by 100% converts to percentage

The residual

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

Each residual is scaled against the actual value

Just as MAE is the average magnitude of error produced by your model, the MAPE is how far the model's predictions are off from their corresponding outputs on average. Like MAE, MAPE also has a clear interpretation since percentages are easier for people to conceptualize. Both MAPE and MAE are robust to the effects of outliers thanks to the use of absolute value.



Finally, the MAPE is biased towards predictions that are systematically less than the actual values themselves. That is to say, MAPE will be lower when the prediction is lower than the

actual compared to a prediction that is higher by the same amount.

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

$\hat{y}$ is smaller than the actual value	$\hat{y}$ is greater than the actual value
n = 1 $\hat{y}$ = 10 y = 20	n = 1 $\hat{y}$ = 20 y = 10
MAPE = 50%	MAPE = 100%

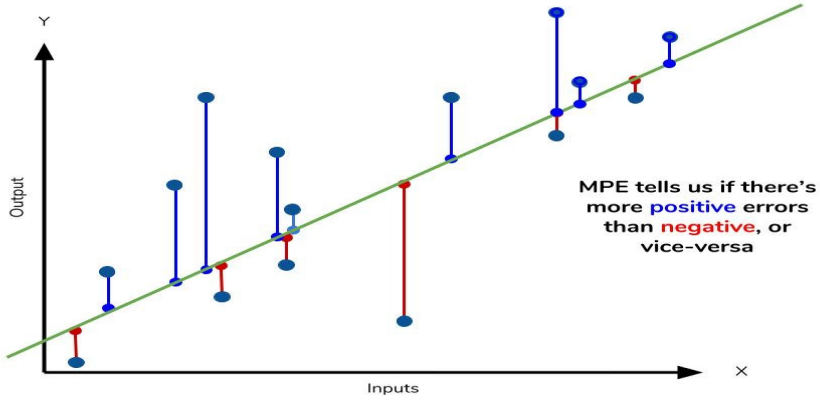
We have a measure similar to MAPE in the form of the mean percentage error. While the absolute value in MAPE eliminates any negative values, the mean percentage error incorporates both positive and negative errors into its calculation.

Mean percentage error

The mean percentage error (MPE) equation is exactly like that of MAPE. The only difference is that it lacks the absolute value operation.

$$MPE = \frac{100\%}{n} \sum \left(\frac{y - \hat{y}}{y} \right)$$

Even though the MPE lacks the absolute value operation, it is actually its absence that makes MPE useful. Since positive and negative errors will cancel out, we cannot make any statements about how well the model predictions perform overall. However, if there are more negative or positive errors, this bias will show up in the MPE. Unlike MAE and MAPE, MPE is useful to us because it allows us to see if our model systematically underestimates (more negative error) or overestimates (positive error).



R² (Coefficient of Determination)

The coefficient of determination, denoted R2 and pronounced “R squared”, is the proportion of the variance in the dependent variable (the response) that is explained (i.e., predictable) from the independent variables (the predictors).

It is an “absolute” index of *goodness-of-fit*, ranging from 0 to 1, and can be used for model performance assessment or models comparison.

The metric helps us to compare our current model with a constant baseline and tells us how much our model is better. The constant baseline is chosen by taking the mean of the data and drawing a line at the mean. R^2 is a scale-free score that implies it doesn't matter whether the values are too large or too small, the R^2 will always be less than or equal to 1.

$$R^2 = 1 - \frac{\text{MSE(model)}}{\text{MSE(baseline)}}$$

Adjusted R^2: Adjusted R^2 depicts the same meaning as R^2 but is an improvement of it. R^2 suffers from the problem that the scores improve on increasing terms even though the model is not improving which may misguide the researcher. Adjusted R^2 is always lower than R^2 as it adjusts for the increasing predictors and only shows improvement if there is a real improvement.

$$R_a^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) \times (1 - R^2)\right]$$

where:

n = number of observations

k = number of independent variables

R_a^2 = adjusted R^2