AKAIKE'S INFORMATION CRITERION (AIC)

- AIC compares the quality of set of statistical model to each other.
- Suppose we have 10 variables, we will create different different models. One model will contain all 10 variables, one model will contain 9 variable, another 8, different combination different model.
- AIC will take each model and rank them from best to worst and AIC will choose the best model from the dataset. If all of the models are poor, it will choose the best of bad bunch.
- AIC score rewards the model that acheive a high goodness-of-fit score and penalize them if they become overcomplex.

  Model 1 - AIC < Model 2 - AIC i.e, model 1 is better than model 2.
  (less than)
- Model which have the lowest AIC value should be choose.

EXAMPLE →

In a study of how hours spent studying and test format (multiple choice vs written answers) affect test score, you create two models:

Model 1. Final test score in response to hours spent studying.

Model 2. Final test score in response to hous spent studying + test format.

Model 1 → $r^2 = 0.45$ with p value less than 0.5.

Model 2 → $r^2 = 0.46$ with p value less than 0.5

Model 2 is slightly better than Model 1 but is it worth to add another parameter to get this small differen? Check AIC value.

When we ran AIC test to find out, which shows that model 1 AIC is lesser than model 2 AIC because model require less information to predict with almost the exact same level of precision.

BAYESIAN INFORMATION CRITERION (BIC) / SCHWARZ CRITERION

→ When fitting the model, it is possible to increase the likelihood by adding the parameter, but doing so may result in overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than AIC.

→ In BIC score also, we choose the model which have lowest BIC score.

→ Compare to AIC, AIC stahstic penalize complex model less maning that it may put more emphasis on model performance on training dataset and in turn select more complex models.

→ A doonside of BIC is that for simpler model, less representative training datasch, it is more likely to choose models that are too simple.