Q3 – AIC

$$AIC = 2p - 2ln(\hat{L})$$

a. Einstein referred to the scientific term – generalization. Generalization describes the model's ability to learn and adapt properly instead of just memorizing. We wish to have a simple model as possible, yet not to simple so we do not lose information.

b. P is the number of estimated parameters. Usually, the performance of the model rises with number of parameters, however its complexity rises as well. A high value of p will most likely increase the odds of overfitting (low bias – high variance). In general, we want to maximize the log likelihood function to get the best maximum likelihood estimation (MLE). For a given number of parameters p, AIC will be lower for higher (desirable) values of $\ln(\hat{L})$.

c. The two options are overfitting and underfitting. Too many parameters will probably bring us to an overfitting situation, whereas a too simple model will cause underfitting.

d. AIC is a measurement to how good is the balance of a certain model. Small values indicate on better balance than high ones. If we want to choose a model to apply on a dataset, it is possible to calculate AIC for all the options and pick the one with the lowest score. It is rather intuitive, since increasing the parameters' number (2p) will improve the model but will make it more complex as well. In addition, high values of $2\ln(\hat{L})$ are desirable (according to section b) and will lower the AIC. To conclude – the lower the better.