Code	**LOGISTIC_REGRESSION.PY**
Author	Nathaniel Heatwole, PhD (heatwolen@gmail.com) (GitHub) (LinkedIn)
Summary	Uses logistic regression to predict probability of passing an exam using number of hours studied (using both sklearn and statsmodels).
Methods/ process	Logistic regression: - Extends the concepts of linear regression to model a binary quantity (0/1). - Functions as a binary classifier, and outputs the probability that Y = 1. - Basis is the logistic function, which takes the form of an s-curve, with dual asymptotes at zero and one. - Argument of the logistic function (X) is a linear equation containing one or multiple variables (continuous or categorical). - Accordingly, the logistic function maps an unbounded quantity (X) onto the bounded probability space (0-1). - Coefficients for the linear component (X) are often estimated using maximum likelihood.[1] - Model fit typically assessed using prevalence of true/false positives. - Extensions of logistic regression have been developed for non-binary outcome variables (Y), or those that can take on more than two possible levels.[2]
Training data	Exam data – synthetic data on whether 20 students passed an exam and number of hours studied (from Wikipedia).
Output	Plot: - Predicted probability (sklearn, statsmodels) Summary: - Regression coefficients and performance (sklearn, statsmodels)
Result	The predictions from the two logistic regression modules (sklearn and statsmodels) are similar, although the statsmodels fit is slightly better, as assessed using log-likelihood (maximum value sought).

[1] Informally, this means choosing the model parameters so as to maximize the likelihood that the particular input data would be observed. More rigorously, in this context, it means selecting the s-curve parameters (location, steepness) so as to minimize the aggregate distance between the s-curve and the true/actual outcome (0 or 1), considering all points in the training data.

[2] If there is a "natural" order (monotonicity) to the groups (e.g., small/medium/large), ordinal logistic regression is useful. Otherwise (e.g., favorite color), multinomial logistic regression relaxes the monotonicity assumption, and allows the regression to select the group order.