RStudio Lab Week 9 – Regression revision

This lab will test your ability to apply some aspects of the regression techniques covered in this course. There is no pictorial for this week – this has been done on purpose so that you test to see that you can apply these techniques with less help. However, if you need to, you can refer back to the previous three weeks pictorials (attached to the quiz). Remember that the RStudio cheatsheet is always available.

There are three datasets in this lab: 'netball_goals.csv', 'accel.csv' and 'prob_purch.csv'.

The *netball_goals* dataset contains data on the shooting accuracy and number of goals scored by netball players.

The *acceleration* dataset contains data for a situation where we're trying to predict acceleration based on five predictor variables: velocity, mass, distance, force, and time. The 'acceleration' variable represents the outcome, and it's assumed to be related to the predictor variables.

The *purchase* dataset contains data for a scenario where we're trying to predict whether a customer will purchase a product based on their age and income.

Instructions:

1) Make sure that you have downloaded all three of the datasets.
2) Import the datasets into RStudio.
3) Fit a simple linear regression model to the *netball_goals* dataset to predict the number of goals scored from the shooting accuracy of players. Save your model as an object and print the summary of the model fit to the console. Predict the number of goals scored for a player with a shooting accuracy of 0.62.
4) Conduct a forwards model building procedure on the *accel* dataset. Fit the best model (based on AIC), save this model and print its summary to the console.
5) With the *prob_purch* dataset, convert the purchase variable to a factor. Then fit a logistic regression model using all the predictor variables. Predict the probability of a purchase for someone who is 30 years of age, and has an income of 33500.
6) Now answer the questions on Vula.