

Exercise for more practice

Type 2 Diabetes in women as a function of Age, Number of pregnancies, Plasma glucose concentration, Diastolic blood pressure, and Body mass index.

Dataset: women-diabetes.xlsx (found in the e-learning page for Logistic Regression)

Variables included in the dataset:

Diabetes: 0=No, 1=Yes

Age: Age of the individual in years

times_pregnant: Number of times pregnant

PGC-2h: Plasma glucose concentration at 2 hours in an oral glucose tolerance test

Diastolic_Blood_Pressure: Diastolic blood pressure (mm Hg)

BMI_categories: 1=Normal, 2=Overweight, 3=Obese

- a) Run a simple logistic regression model to examine whether age is a risk factor for diabetes (model 1). Interpret the results. Find the regression equation for this model.
- b) Express the result for every 10 years increase of age.
- c) Run a multiple logistic regression model to examine whether age, plasma glucose concentration at 2 hours in an oral glucose tolerance test, number of times pregnant and BMI are associated with diabetes (model 2). Interpret the results.
- d) Run a multiple logistic regression model to examine whether age, plasma glucose concentration at 2 hours in an oral glucose tolerance test, number of times pregnant, diastolic blood pressure and BMI are associated with diabetes (model 3). Interpret the results.
- e) Run model 3 again excluding any non-significant variables (model 4).
- f) Compare models 1 and 2. Interpret the results.
- g) Compare models 2 and 4. Interpret the results.