Dataset and variables

Dataset with 880 observations on Vitamin D and Osteoporosis levels from NHANES in the cycles 2007-2008 and 2009-2010 (dataset: VitaD_Osteo.xlsx).

The dataset consists of the following 6 variables:

`age`: age (years)

`sex`: sex (2=women, 1=men)

`vitD_group`: groups of serum levels of vitamin D [0=Deficiency (<30mg/ml), 1= Inadequacy (30-50 mg/ml), 2= Sufficiency (>50)]

`Calcium`: serum levels of calcium (mg/ml)

`Osteop`: osteoporosis (1=yes, 0=no).

Questions

a) Run a simple logistic regression model to examine whether age is a risk factor for osteoporosis. Interpret the results. Express the result for every 5 years increase of age.
b) Run a simple logistic regression model to examine whether sex is a risk factor for osteoporosis. Interpret the results.
c) Run a multiple Logistic regression model against all explanatory variables with $p<0.2$ in the univariable analysis. Interpret the results.
d) Run a multiple Logistic regression model including the non-significant variables too.
e) Compare models generated by b) and c) interpret the results. Which is the best fitting model?
f) Check the diagnostics of the best fitting model.
g) Present the results in a table with all the necessary information.