

## Lab 7: Logistic regression

### Question 1:

The dataset **Heart study.xlsx** contains information from a case-control study with 98 patients who've been diagnosed with CHD (cases) and 100 patients who have not been diagnosed with CHD (controls).

The dataset contains the following variables:

Variable	Description
RANDID	Random ID
GENDER	Gender (1=male, 2= female)
TOTCHOL	Total serum cholesterol (mg/dl) at first examination
AGE	Age (years) at first examination
SYSBP	Systolic bloodpressure
CURSMOKE	Current smoke at start of study (1= yes, 0 = no)
CHD	Diagnosis of CHD in first 10 years (1= yes, 0 = no)

- (a) Use crosstab (two-way frequency table) and Chi-square test to examine the relationship between CHD and GENDER. Compute the odds ratio that compares the odds of Males developing CHD compared to Females using the Crosstab.
- (b) Use logistic regression to compute the odds ratio and 95% confidence interval of CHD for smokers compared to non-smokers.
- (c) Use logistic the odds ratio and 95% confidence interval of CHD for males compared to females
- (d) Fit a logistic regression model to CHD with the independent variables GENDER, TOTCHOL, AGE, SYSBP and CURSMOKE. What are your conclusions from this analysis?

**Question 2:**

A sample of 1594 car buyers is given in the Excel data file **Badcar.xlsx**. The objective of the study is to identify factors associated with a bad car buy (i.e a car that has faults). The following variables are contained in the dataset.

Variable Name	Description
RefId	ID
IsBadBuy	Whether or not the car was a bad buy (1 = yes, 0 = no)
VehicleAge	Age of the car
Dealer	Type of dealer (seller) (0 = Independent, 1 = Franchised)
VehOdo	Number of Kilometers (in 10,000s) on the odometer
VehBCost	Cost of the car when new (in 1000s)

- (i) Fit a logistic regression model to the data with IsBadBuy as the dependent variables and VehicleAge, Dealer, VehOdo and VehBCost as independent variables.
- (ii) State and interpret the adjusted odds ratios with associated 95% confidence intervals for each of the significant variables in the model. What conclusions can you draw from this model?