Please submit an Rmd file and the knitted pdf file showing all your work for each of the following problems. Use a significance level of 0.05 unless stated otherwise.

Plasma ferritin concentration study

In these tasks, you will assess the effect of a collection of explanatory variables on the plasma ferritin concentration (`Ferr`) in 202 Australian athletes. The file "Sports Data CW 2021.csv" contains the data on the plasma ferritin concentration as well as a selection of demographic variables of 202 male and female athletes. In particular, the data set comprises observations on the following eleven variables:

Variable	Description
Sport	Types of sport
Sex	Male or female
LBM	Lean body mass
RCC	Red cell count
WCC	White cell count
Hc (%)	Hematocrit (Hc) is the volume percentage (vol%) of red blood cells in blood. It is normally 47% ± 5% for men and 42% ± 5% for women.
Hg (g/dl)	Hemoglobin (Hg) is the protein contained in red blood cells that is responsible for delivery of oxygen to the tissues. The normal Hg level for males is 14 to 18 g/dl; that for females is 12 to 16 g/dl.
BMI	Body mass index = weight/height^2
SSF (mm)	Sum of skin folds
% Bfat	% body fat
Ferr (μmol/L)	Plasma ferritin concentration

Task 1

Test if plasma ferritin concentration differs between male and female athletes. Make sure that the assumptions of the test are satisfied. State null and alternative hypotheses and your conclusion.

Task 2

Randomly divide the dataset into two sets, training ($n1 = 141$) and testing ($n2 = 61$). Use the training dataset to:

> (a) Write down the equation of a regression model with `Ferr` as the response and other variables as predictors except for the `Sport` variable.

(b) Fit the model in (a) and remove insignificant predictors until all the variables in the model are statistically significant. Is a full model better than a smaller model? Use an appropriate test or score to support your argument.

(c) Check the linear regression assumptions for the model fitted in part (b). Do the assumptions hold for your model?