Homework
Applied Logistic Regression

WEEK 4

Exercise 1:

The data in hyponatremia.dta derive from an epidemiological study of hyponatremia (a life-threatening condition) among runners of the 2002 Boston Marathon. Hyponatremia is defined as an electrolyte disturbance in which the serum sodium concentration is lower than normal (<135 mmol/l). The aim of the study was to determine whether a runner experienced hyponatremia and to identify the principal risk factors. Participants in the 2002 Boston Marathon completed a survey including demographic and anthropometric characteristics (Body Mass Index) one or two days before the race. After the race, runners provided a blood sample in order to measure their serum sodium concentration and completed a questionnaire detailing their urine output during the race. Prerace and postrace weights were also recorded.

a. Perform a logistic regression analysis with Stata using **nas135** as dependent variable and **female** as the only independent variable. Interpret the coefficients of the model.

b. Fit a model with **runtime** as the only independent variable. Interpret the coefficient for **runtime**.

c. Calculate the Odds Ratio for the variable **runtime** and interpret it.

d. Interpret the coefficient for the constant in the model with **runtime** as the only independent variable. Does it make sense? If not, what can you do to obtain a coefficient for the constant which is easily interpreted?

e. Calculate the Odds Ratio of hyponatremia of a runner who takes 2 hours more than another runner, and the corresponding 95% Confidence Interval

f. Fit a model with **female** and **runtime** as independent variables. Interpret both coefficients.

g. Compare the coefficients for **female** in the model with **female** as the only independent variable with that in the model that contains **female** and **runtime**. What is the percentage change in the coefficient of **female**?

h. Calculate the Odds Ratio of hyponatremia for a female compared to a male who completes the marathon in the same time.

i. What type of association do you expect between the variables **female** and **runtime**? Answer this question before looking at the data, only on the basis of

the observed change in the coefficient for **female** when **runtime** is entered into the model. Then make a box-plot of **runtime** by **female**.

j. Assess whether there is an interaction between **female** and **runtime**

k. Add to the model that contains **female** and **runtime** a dichotomous variable **wgain** which takes the value of 0 if **wtidff** ≤ 0, and the value of 1 if **wtidff** > 0. Test for interaction between **female** and **wgain**.

l. On the basis of the model with the interaction term, calculate the Odds Ratios of hyponatremia for males who gain weight as compared to those who don't. Repeat this exercise for a female. Interpret your findings.

m. Compare using the Likelihood Ratio test the model with **female** and **runtime** with a model with **female**, **runtime**, **wgain**, **urinat3p** and **bmi**. (Hint: the 2 models must be fitted on the same set of observations. Be aware of missing values in some of these variables). How many degrees of freedom does the test statistic have?