Exercise Set 5
Binary Regression and Non-linear Optimisation with R

Murshid Saqlain (msq@du.se)
Deadline March 17th, 2019.

1. We will use the data Mroz in the package car for this exercise. It includes data about labor-force participation among married women in the US.

Variable	*Description*	*Remarks*
lfp	wife's labor-force participation	factor: no, yes
k5	number of children ages 5 and younger	0-3, few 3's
k618	number of children ages 6 to 18	0-8, few > 5
age	wife's age in years	30-60, single years
wc	wife's college attendance	factor: no, yes
hc	husband's college attendance	factor: no, yes
lwg	log of wife's estimated wage rate	see text, pg 235
inc	family income excluding wife's income	$1,000s

(a) Use a logit regression to predict the probability to work given age, number of children 5 years old or younger, number of children 6 to 18 years old and a factor indicating if the woman has college attendance. Thus, run the binary regression to obtain estimates of the coefficients on these variables. Comment on how these variables affect the probability that a married woman works.

(b) Sue is 30 years old, has college attendance and one child, Tom, he is 5 years old. What is the predicted probability that Sue works given the logit regression?

(c) If Sue had another child who is a 15 year old girl, what would be the new predicted probability? How does this compare to the answer in (b)? Comment on whether you expected this change and why.

(d) Betty is 25 years old, without college attendance but with a 3 year old girl. What is the predicted probability to work for Betty?

(e) If Betty had college attendance instead, what is her new predicted probability to work? How does this compare to the answer in (d)? Comment on whether you expected this change and why.

(f) Now add the variables *hc*, *inc* and *lwg* to the logit regression. What does a higher *inc* imply for a married woman's probability to work according to the regression? Explain in words why this makes sense (or if you don't think it makes sense: explain).

2. We will compare different models and do hypothesis testing in this exercise using the data Mroz again.

(a) Test the following hypothesis:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

That is, the coefficients of all the variables are 0.

(b) Test the hypothesis that labor-force participation does not depend on the number of young children versus the alternative that it depends jointly on the number of young children and the number of older children. State the change in deviance, degrees of freedom, and the corresponding p-value. Comment on your result.

(c) Test the hypothesis that labor-force participation does not depend on the wife's college attendance versus the alternative that it does depend on it. State the change in deviance, degrees of freedom, and the corresponding p-value. Comment on your result.