1. Load the “menarche” dataset from the R MASS package using the following code. The dataset whether or not 3918 Polish girls of different ages have reached menarche (1/0 for Yes/No).

library("MASS")

menarche

# first, make a "long" version of the data

# - one row per person

menarchelong <- data.frame(age=NULL, menarche=NULL)

for(i in 1:25){

n <- menarche$Total[i]

r <- menarche$Menarche[i]

age <- menarche$Age[i]

menarchelong <- rbind(menarchelong, cbind(rep(age, n), rep(0:1, c(n-r,r))))

}

names(menarchelong) <- c("age","menarche")

1. What is the mean age? What percentage of girls in the dataset have reached menarche?
2. Plot the percentage of girls that reach menarche by age (use scatterplot). Hint: It is easier to do this using “menarche” instead of “menarchelong”.
3. Use glm to perform logistic regression of menarche on age. Repeat using the regress function in uwIntroStats. Are there any differences in the results?
4. Report the odds ratio and 95%CI for age.
5. What is the fitted *odds* and 95% confidence interval for 12 year old girls? What is the fitted *probability* for 12 year old girls?
6. Add the fitted probability line to the plot you made in #3. Hint: use the curve function with parameter add = TRUE.