

Exercise 5

Riccardo De Bin and Vindi Jurinovic

November 18, 2015

1. Turn the variable `smokstat` into a binary variable (put the non-smokers and the people who almost never smoked into one group).
2. Do you expect smokers to have a higher cancer prevalence than non-smokers? Test if there is a significant difference in cancer prevalence between smokers and non-smokers (using the variable from 1.). If so, which group has a higher prevalence? Did you get the result you expected?
3. Do the same in the subgroup of people who are between 20 and 49 years old. What do you see now? How can you explain the results?
4. Turn the variable `diab_lft` into a binary variable (put diabetes and pre-diabetes together). Test if there is a significant difference in diabetes prevalence between overweight and normal-weighted people.
5. Test if overweight people have a higher prevalence of heart diseases. Does your finding confirm the hypothesis that overweight is a risk factor for heart disease?