Arun Sangar

Hyunwook Park

CPSC 375

Homework 1

* + 1. 5
    2. ncol(esoph)
    3. 88
    4. nrow(esoph)
    5. 3 categorical: agegp/alcgp/tobgp

2 numeric: ncases/ncontrols

* + 1. str(esoph)
    2. Levels: 25-34 < 35-44 < 45-54 < 55-64 < 65-74 < 75+
    3. esoph$agegp
    4. 0
    5. esoph[10,4]
    6. agegp alcgp tobgp ncases ncontrols

25-34 0-39g/day 0-9g/day 0 40

* + 1. esoph[1,]
    2. agegp alcgp tobgp ncases ncontrols

1 25-34 0-39g/day 0-9g/day 0 40

2 25-34 0-39g/day 10-19 0 10

3 25-34 0-39g/day 20-29 0 6

4 25-34 0-39g/day 30+ 0 5

5 25-34 40-79 0-9g/day 0 27

* + 1. esoph[1:5,]
    2. agegp alcgp tobgp ncases ncontrols

31 45-54 0-39g/day 0-9g/day 1 46

32 45-54 0-39g/day 10-19 0 18

33 45-54 0-39g/day 20-29 0 10

34 45-54 0-39g/day 30+ 0 4

35 45-54 40-79 0-9g/day 6 38

36 45-54 40-79 10-19 4 21

37 45-54 40-79 20-29 5 15

38 45-54 40-79 30+ 5 7

39 45-54 80-119 0-9g/day 3 16

40 45-54 80-119 10-19 6 14

41 45-54 80-119 20-29 1 5

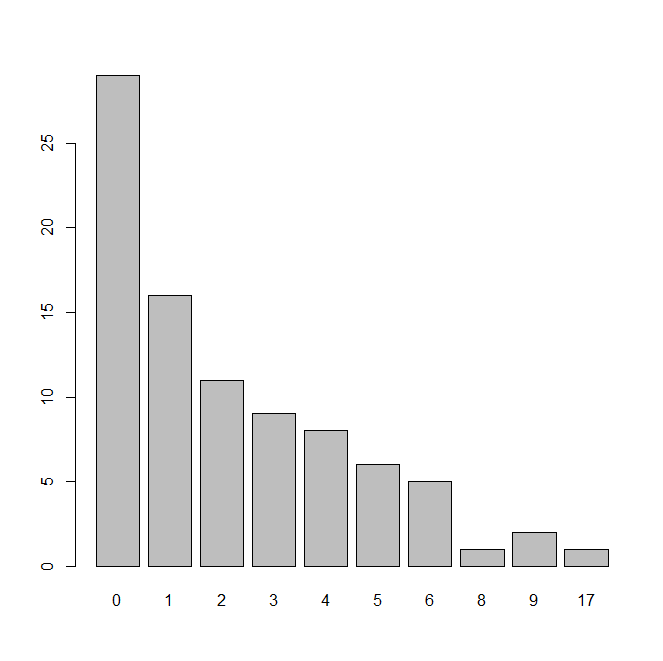
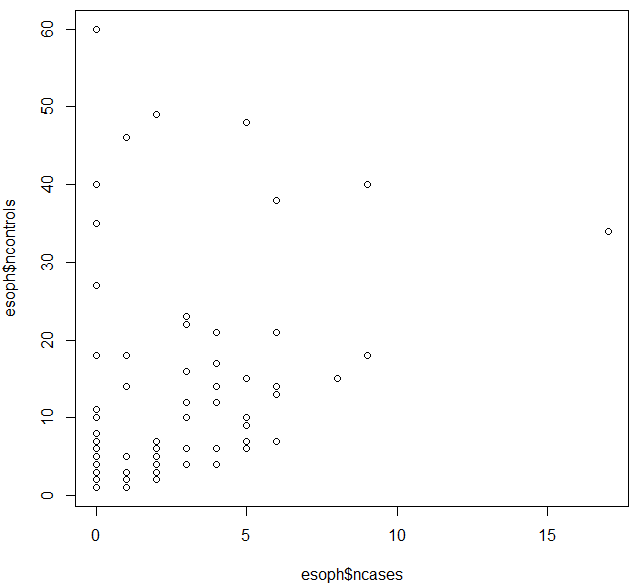
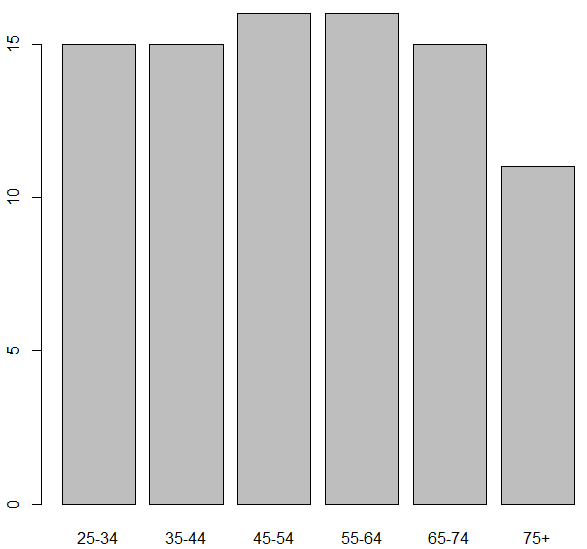
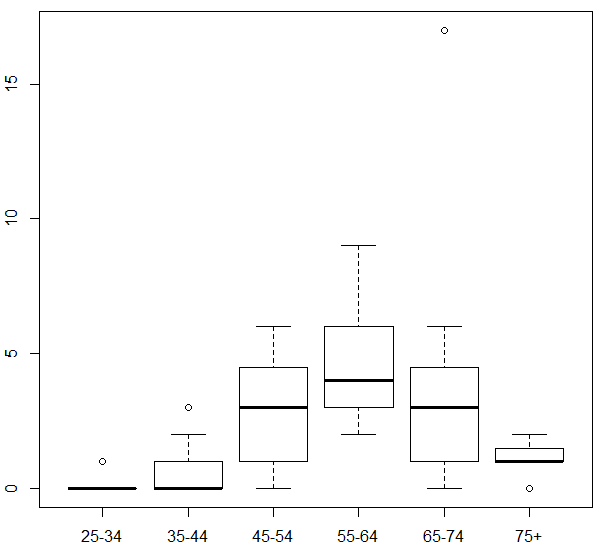
42 45-54 80-119 30+ 2 4

43 45-54 120+ 0-9g/day 4 4

44 45-54 120+ 10-19 3 4

45 45-54 120+ 20-29 2 3

46 45-54 120+ 30+ 4 4

* + 1. esoph[esoph[,1] == '45-54',]
    2. [1] 46 18 10 4 38 21 15 7 16 14 5 4 4 4 3 4
    3. esoph[esoph[,1] == '45-54',5]
    4. 2.272727
    5. mean(esoph$ncases)
    6. [1] 2.875
    7. mean(esoph[esoph[,1] == '45-54',4])
    8. 
    9. > barplot(table(esoph[,4]))
    10. No, because the data set has many 0 values thus the values of ncases are not symmetrically distributed around the mean.
    11. 
    12. Plot(esoph$ncases,esoph$ncontrols)
    13. 
    14. barplot(table(esoph[,1]))
    15. 
    16. plot(esoph$agegp,esoph$ncases)