Arun Sangar

Hyunwook Park

CPSC 375

Homework 1

1.
- a.
  - i. 5
  - ii. ncol(esoph)
- b.
  - i. 88
  - ii. nrow(esoph)
- c.
  - i. 3 categorical: agegp/alcgp/tobgp
    2 numeric: ncases/ncontrols
  - ii. str(esoph)
- d.
  - i. Levels: 25-34 < 35-44 < 45-54 < 55-64 < 65-74 < 75+
  - ii. esoph$agegp
- e.
  - i. 0
  - ii. esoph[10,4]
- f.
  - i.   agegp   alcgp   tobgp ncases ncontrols
      25-34 0-39g/day 0-9g/day   0   40
  - ii. esoph[1,]
- g.
  - i.   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
  - ii. esoph[1:5,]
- h.
  - i.   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
```
   ii.   esoph[esoph[,1] == '45-54',]

i.

   i.   [1] 46 18 10 4 38 21 15 7 16 14 5 4 4 4 3 4

   ii.   esoph[esoph[,1] == '45-54',5]

2.

   a.

      i.   2.272727

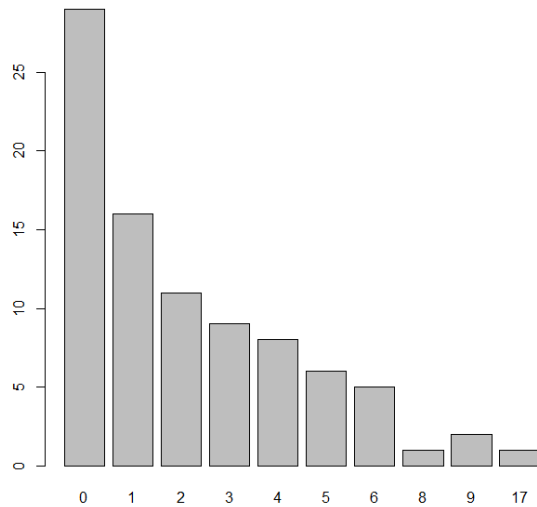   ii.   mean(esoph$ncases)

   b.

      i.   [1] 2.875

     ii.   mean(esoph[esoph[,1] == '45-54',4])

   c.

      i.
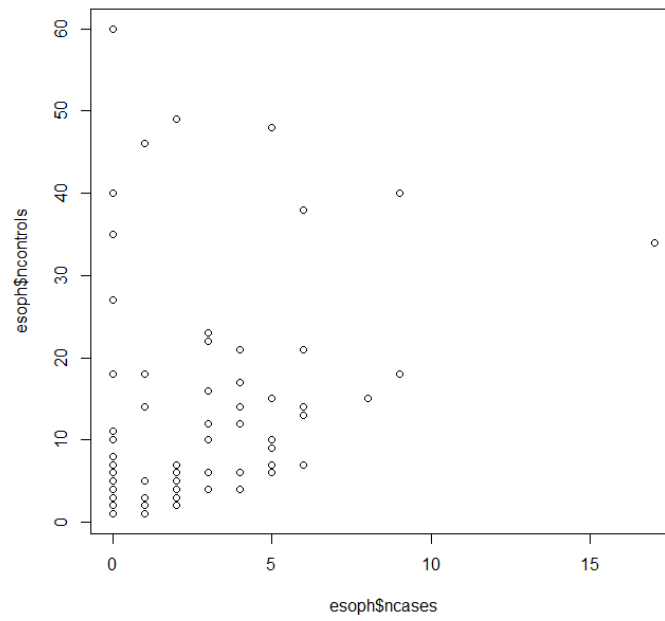
     ii.   barplot(table(esoph[,4]))

   d.

      i.   No, because the data set has many 0 values thus the values of ncases are not symmetrically distributed around the mean.
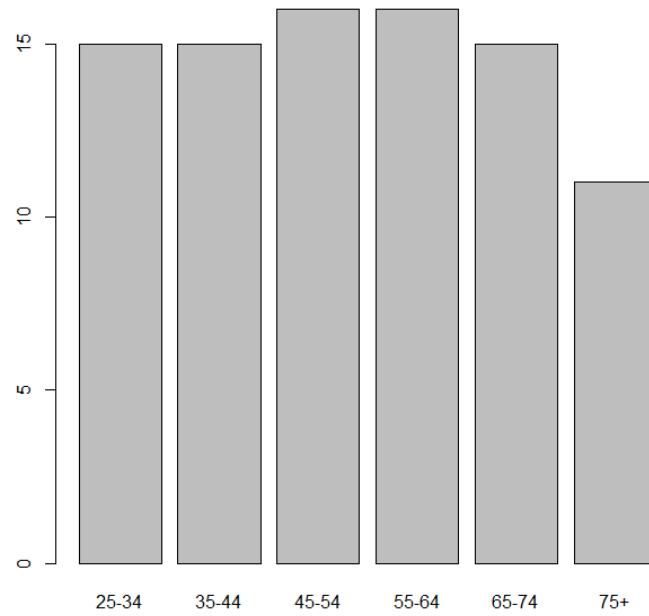
e.

   i.

  ii.   plot(esoph$ncases,esoph$ncontrols)

f.

   i.

  ii.   barplot(table(esoph[,1]))

g.

    i.
  ii.   plot(esoph$agegp,esoph$ncases)