Exercise 8

0316213 Yu-Wen Pu

2018-05-13

```
knitr::opts_chunk$set(results = "hold")#, fig.retina = 2)
set.seed(1830)
```

19.11

```
x <- matrix(c(22, 187, 19, 74), byrow = TRUE, ncol = 2)
chisq.test(x, correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  x
## X-squared = 5.3804, df = 1, p-value = 0.02036
```

Yes. **ADD classification** and **remedial/nonremedial English enrollment** are not independent. ($\alpha = 0.05$)

19.12

a) See below.

```
(x <- matrix(c(22, 187, 2, 17, 1, 11, 3, 16,
               2, 9, 4, 7, 3, 8, 4, 6),
             byrow = TRUE, ncol = 2))
```

```
##      [,1] [,2]
## [1,]   22  187
## [2,]    2   17
## [3,]    1   11
## [4,]    3   16
## [5,]    2    9
## [6,]    4    7
## [7,]    3    8
## [8,]    4    6
```

```
(result <- chisq.test(x, correct = FALSE))
```

```
## Warning in chisq.test(x, correct = FALSE): Chi-squared approximation may be
## incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  x
## X-squared = 14.945, df = 7, p-value = 0.03671
```

b) **Exhibition of ADD-like behaviors** and **remedial/nonremedial English enrollment** are not independent. ($\alpha = 0.05$)

c) See below.

Expected frequencies:

```
result$expected
```

```
##           [,1]       [,2]
## [1,] 28.374172 180.625828
## [2,]  2.579470  16.420530
## [3,]  1.629139  10.370861
## [4,]  2.579470  16.420530
## [5,]  1.493377   9.506623
## [6,]  1.493377   9.506623
## [7,]  1.493377   9.506623
## [8,]  1.357616   8.642384
```

If the expected frequency is small, the observed frequency cannot be normally distributed around it. Thus the chi-square test may be invalid.

Solution: Use Fisher's Exact Test.

```
fisher.test(x)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  x
## p-value = 0.02791
## alternative hypothesis: two.sided
```

Got the same conclusion. ($\alpha = 0.05$)

19.13

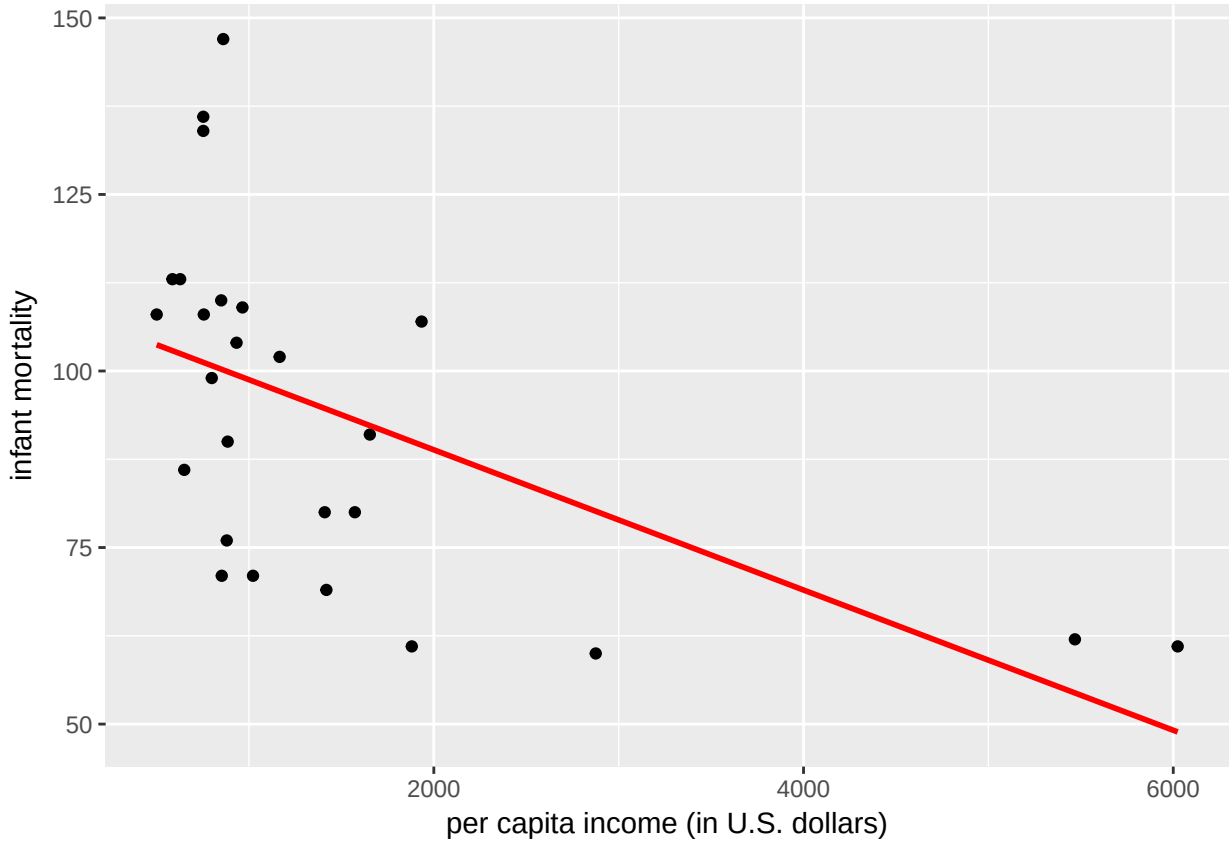
結果是隨機的，各類別平分 N (sample size, = 41)。**19.12** 的假設是 **Exhibition of ADD-like behaviors** 與 **remedial/nonremedial English enrollment** 彼此獨立。

9.1

```
d <- read.table("SubSaharanInfMort.dat", header = TRUE, sep = "",
                na.strings = "NA", fileEncoding = "ISO-8859-1")
head(d)
```

```
##               country infmort income youngmom oldmom toosoon contracept
## 1           Benin_Rep     104    933       16      5      17          3
## 2        Burkina_Faso     109    965       17      5      17          5
## 3            Cameroon      80   1573       21      4      25          7
## 4 Central_African_Rep     102   1166       22      5      26          3
## 5            Chad_Rep     110    850       21      3      24          1
## 6       Côte_d'Ivoire      91   1654       21      6      16          4
##   need
## 1   26
## 2   26
## 3   20
## 4   16
## 5   NA
## 6   28
```

```r
library(ggplot2)
ggplot(d, aes(x = income, y = infmort)) +
    geom_point() +
    geom_smooth(method = "lm", color = "red", se = FALSE) +
    labs(x = "per capita income (in U.S. dollars)", y = "infant mortality")
```



effect of outliers: 使負相關更明顯，r^2 更大。

9.2

```r
rownames(d) <- d$country
d$country <- NULL
d[is.na(d)] <- 0  # replace NA with 0
head(d)
```

```
##                     infmort income youngmom oldmom toosoon contracept need
## Benin_Rep               104    933       16      5      17          3   26
## Burkina_Faso            109    965       17      5      17          5   26
## Cameroon                 80   1573       21      4      25          7   20
## Central_African_Rep     102   1166       22      5      26          3   16
## Chad_Rep                110    850       21      3      24          1    0
## Côte_d'Ivoire            91   1654       21      6      16          4   28
```

```r
cor(d)
```

```
##                infmort      income    youngmom      oldmom     toosoon
## infmort     1.00000000 -0.55557166  0.22411749 -0.04923238  0.12327898
## income     -0.55557166  1.00000000  0.05712275  0.04150834 -0.14017945
```

```
## youngmom    0.22411749  0.05712275  1.00000000 -0.57804571  0.00764576
## oldmom     -0.04923238  0.04150834 -0.57804571  1.00000000 -0.19416761
## toosoon     0.12327898 -0.14017945  0.00764576 -0.19416761  1.00000000
## contracept -0.43990972  0.32949103  0.30185841 -0.14207945 -0.32022938
## need       -0.28939078  0.00712952 -0.41237107  0.38722342 -0.06756170
##             contracept        need
## infmort    -0.43990972 -0.28939078
## income      0.32949103  0.00712952
## youngmom    0.30185841 -0.41237107
## oldmom     -0.14207945  0.38722342
## toosoon    -0.32022938 -0.06756170
## contracept  1.00000000 -0.02441417
## need       -0.02441417  1.00000000
```

9.4

```
cor(d$infmort, d[, names(d) != "infmort"]) ** 2
```

```
##         income   youngmom      oldmom    toosoon contracept       need
## [1,] 0.3086599 0.05022865 0.002423828 0.01519771  0.1935206 0.08374702
```

Per capita income.

9.5

嬰兒死亡率與人均收入成負相關，人均收入愈高，嬰兒死亡率愈低；嬰兒死亡率也與避孕措施的使用率成負相關，避孕措施的使用率愈高，嬰兒死亡率愈低。

9.6

```
summary(d)
```

```
##     infmort           income          youngmom         oldmom
##  Min.   : 60.00   Min.   : 501.0   Min.   : 9.00   Min.   :3.000
##  1st Qu.: 72.25   1st Qu.: 766.8   1st Qu.:15.25   1st Qu.:4.000
##  Median : 95.00   Median : 909.0   Median :21.00   Median :5.000
##  Mean   : 94.15   Mean   :1464.6   Mean   :19.46   Mean   :5.115
##  3rd Qu.:108.75   3rd Qu.:1534.5   3rd Qu.:22.00   3rd Qu.:6.000
##  Max.   :147.00   Max.   :6024.0   Max.   :32.00   Max.   :7.000
##     toosoon        contracept         need
##  Min.   :12.00   Min.   : 1.00   Min.   : 0.00
##  1st Qu.:17.00   1st Qu.: 5.00   1st Qu.:20.50
##  Median :20.50   Median : 7.50   Median :25.00
##  Mean   :20.85   Mean   :11.42   Mean   :23.54
##  3rd Qu.:25.00   3rd Qu.:13.00   3rd Qu.:28.00
##  Max.   :31.00   Max.   :50.00   Max.   :36.00
```

在撒哈拉以南非洲（Sub-Saharan Africa），相較於年幼產婦的比例（中位數達 21%），高齡產婦的比例不高（中位數僅 5%），因此並非主要的風險來源；但在其他地區，相較於年幼產婦的比例，高齡產婦的比例偏高，因此會是一個風險因子。

9.7

如果能結合人均收入與避孕措施的使用率來預測，應該能預測得更準確，使負相關更明顯。

9.8

不能，相關性不表示因果關係！這是一個邏輯謬誤。