stat230-hw3-4

Yifan Zheng

3/2/2020

Lecture 7, Question 3: Examples of Simpson's Paradox

Example 1:

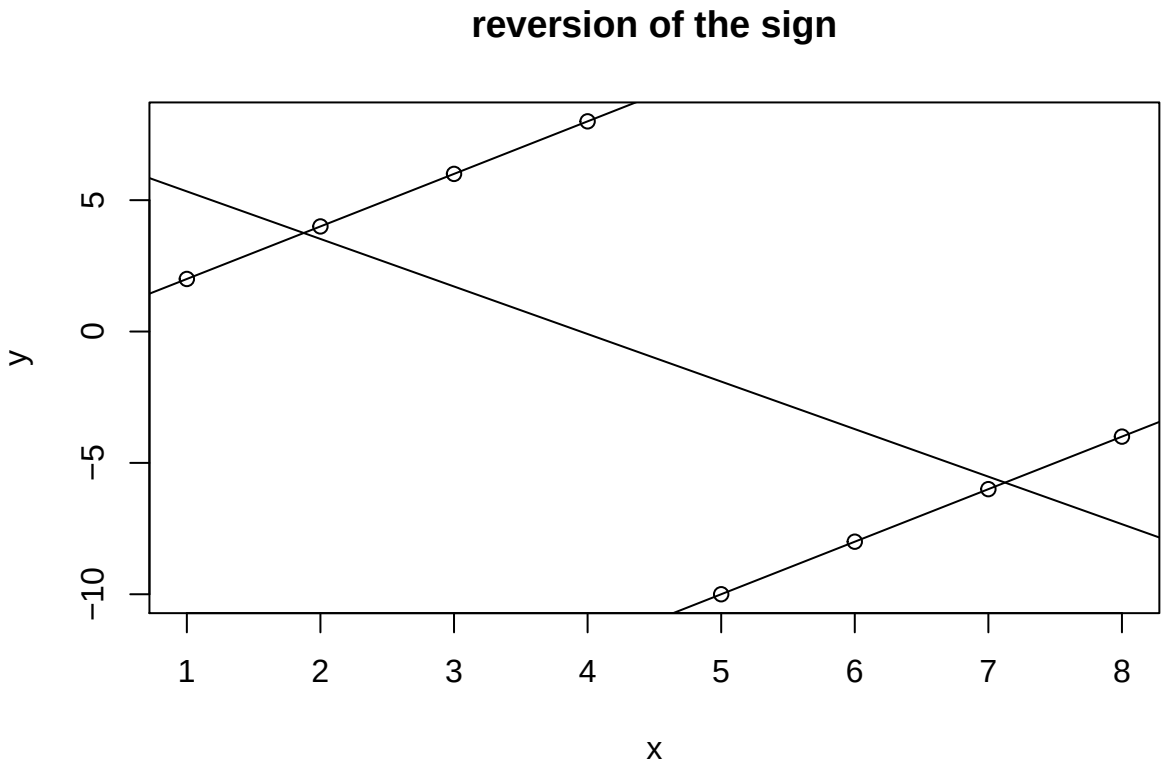
```
# a simple example
w = c(0,0,0,0,1,1,1,1)
x = 1:8
y = NULL
for (i in 1:8){
  if (w[i]==0){
    y[i] = 2*x[i]
  }else{
    y[i] = 2*x[i]-20
  }
}

fit.xw = lm(y~x+w)$coef
fit.xw
```

```
##   (Intercept)             x             w
##  1.256074e-15  2.000000e+00 -2.000000e+01
```

```
fit.x = lm(y~x)$coef
fit.x
```

```
## (Intercept)           x
##    7.142857   -1.809524
```

```
#plot
plot(x,y, main = "reversion of the sign")
abline(fit.x)
abline(fit.xw[1], fit.xw[2])
abline(fit.xw[1]+fit.xw[3], fit.xw[2])
```




```
M = cbind(y,x,w)
# mean matrix
mean_matrix = colMeans(M)
mean_matrix
```

```
##    y    x    w
## -1.0  4.5  0.5
```

```
# covariance matrix
cov_matrix = cov(M)
cov_matrix
```

```
##            y          x          w
## y  46.857143 -10.857143 -3.4285714
## x -10.857143   6.000000  1.1428571
## w  -3.428571   1.142857  0.2857143
```

Example 2:

```
# inspired by the example of uc berkeley gender bias

set.seed(123)

# recieve total 100 applications
```

```r
n=100

# w = 0 is department A, w = 1 is department B
w = rbinom(n,1,0.5)
x = NULL

# x = 0 is male and x = 1 is female
# simplify the different preference between male and female applicants
# male applicants are more interested in department B
# while female applicants are more interested in department A
for (i in 1:n){
  if (w[i] == 0){
    x[i] = rbinom(1,1,0.9)
  }else{
    x[i] = rbinom(1,1,0.1)
  }
}

# y is the probability of admitting an applicant
# department B has much higher admitted rate
y = (4*w+x)/5

fit.xw = lm(y~x+w)$coef
fit.xw
```

```
##  (Intercept)            x            w
## 5.551115e-16 2.000000e-01 8.000000e-01
```

```r
fit.x = lm(y~x)$coef
fit.x
```

```
## (Intercept)           x
##   0.7130435  -0.4241546
```

```r
M = cbind(y,x,w)
# mean matrix
mean_matrix = colMeans(M)
mean_matrix
```

```
##     y     x     w
## 0.484 0.540 0.470
```

```r
# covariance matrix
cov_matrix = cov(M)
cov_matrix
```

```
##            y          x          w
## y  0.1084283 -0.1064242  0.1621414
## x -0.1064242  0.2509091 -0.1957576
## w  0.1621414 -0.1957576  0.2516162
```

Example 3.

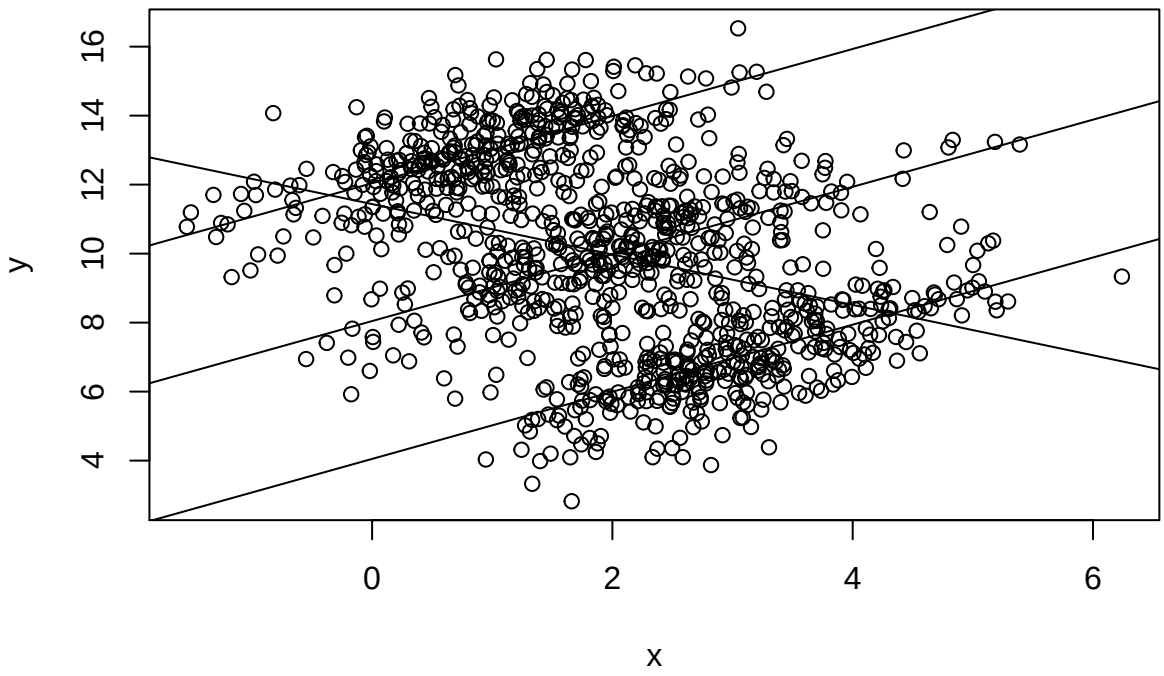
```
set.seed(123)
n=1000
# classify w into three types
w = c(rep(1,333),rep(2,333),rep(3,334))
x1 = rnorm(n,3,1)
x2 = rnorm(n,2,1)
x3 = rnorm(n,1,1)

x[1:333] = x1[1:333]
x[334:666] = x2[1:333]
x[667:1000] = x3[1:334]

y = 4*w+x+rnorm(n)

fit.xw = lm(y~x+w)$coef
fit.xw
```

```
## (Intercept)           x           w
##  0.05367225  0.97297802  3.99544753
```

```
fit.x = lm(y~x)$coef
fit.x
```

```
## (Intercept)           x
##  11.4353035  -0.7305048
```

```
#plot
plot(x,y, main = "reversion of the sign")
abline(fit.x)
#abline(fit.xw[1], fit.xw[2])
abline(fit.xw[1]+fit.xw[3], fit.xw[2])
abline(fit.xw[1]+2*fit.xw[3], fit.xw[2])
abline(fit.xw[1]+3*fit.xw[3], fit.xw[2])
```




```
M = cbind(y,x,w)
# mean matrix
mean_matrix = colMeans(M)
mean_matrix
```

```
##        y        x        w
## 9.982967 1.988127 2.001000
```

```
# covariance matrix
cov_matrix = cov(M)
cov_matrix
```

```
##           y          x          w
## y  7.696157 -1.2171645  1.9764306
## x -1.217165  1.6661964 -0.7103928
## w  1.976431 -0.7103928  0.6676667
```