PSTAT 126 HW5

Mujie Wang

2018/05/11

Problem 1

(a) b0 = 24697 is the expected salary in dollars for males. (Sex = 0, so the second term is -3340*0.) b1 = -3340 is the expected di erence in salaries between males and females, so females on average have a salary which is $3340 lower.

(b) This could occur if the salaries of the male professors are higher because they have been employed longer than the females professors. This model shows female salaries are actually $201 higher on average when controlling for the number of years employed. The males have been employed longer so their salaries appear arti cally larger if we do not account for years employed.

Problem 2

```r
library(faraway)
data(infmort)

mortality = infmort$mortality
income = infmort$income
region = infmort$region

y = log(mortality)
x1 = log(income)
x2 = region

#(a)
#The overall F-test is for the hypotheses H0 :b1 =b2 =b3 =0 vs H1:
#at least one of the three b's is not 0

mod.reduced = lm(y~1)
mod.full = lm(y~x1+x2+x2*x1)
anova(mod.reduced, mod.full)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ 1
## Model 2: y ~ x1 + x2 + x2 * x1
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    100 93.769
## 2     93 33.152  7    60.617 24.292 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# The p-value is very close to 0, so we reject the null hypothesis and
# conclude that at least one of the three terms is signi cant in the model.

#(b)
#Under this null hypothesis, both terms involving region are no longer in the
#model. In other words, it supposes that there is no regional e ect, and the
#effect of income on mortality is not a ected by region.
```

```r
#(c)
mod.reduced = lm(y~x1)
mode.full = lm(y~x1+x2+x1*x2)
anova(mod.reduced, mod.full)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x1
## Model 2: y ~ x1 + x2 + x2 * x1
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     99 46.685
## 2     93 33.152  6    13.533 6.3274 1.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#The p-value is very close to 0, so we reject the null hypothesis and
#conclude that there is indeed a regional e ect on mortality. The e ect of
#income (region) on mortality may be a ected by region (income).
```

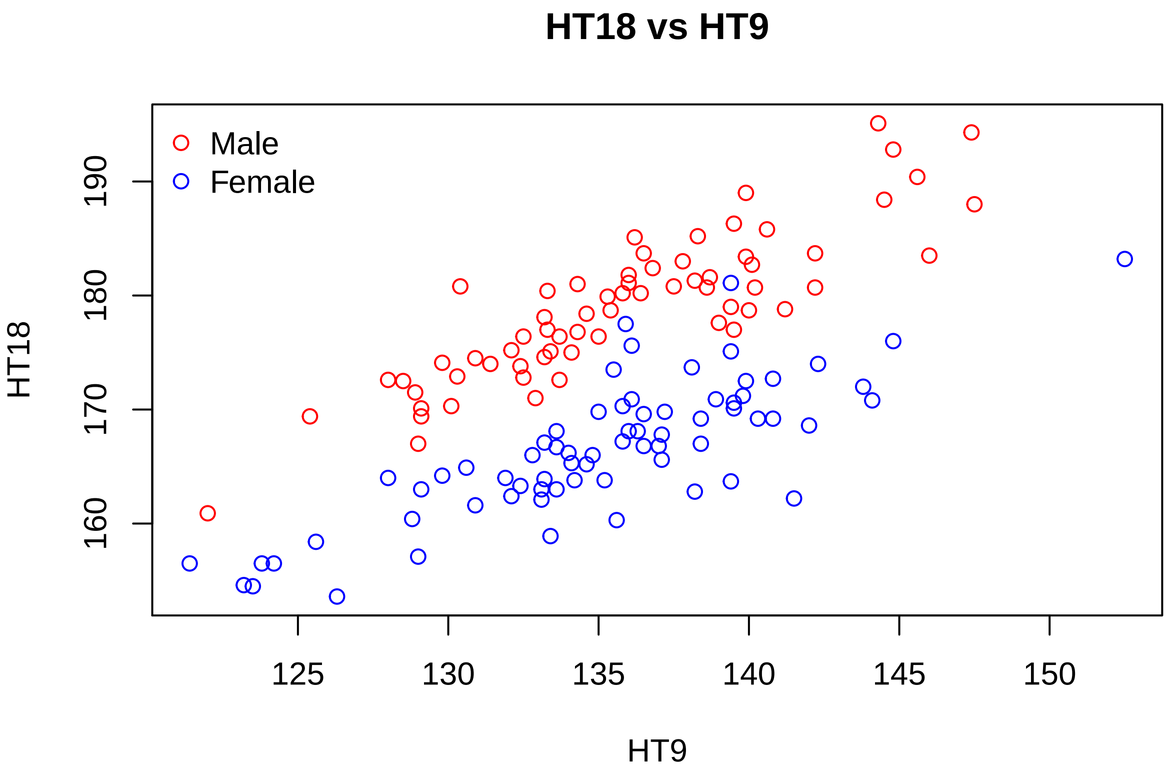
Problem 3

```r
library(alr4)
```

```
## Loading required package: car

## Warning: package 'car' was built under R version 3.4.4

## Loading required package: carData

## Warning: package 'carData' was built under R version 3.4.4

##
## Attaching package: 'car'

## The following objects are masked from 'package:faraway':
##
##     logit, vif

## Loading required package: effects

## Warning: package 'effects' was built under R version 3.4.4

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

##
## Attaching package: 'alr4'

## The following objects are masked from 'package:faraway':
##
##     cathedral, pipeline, twins
```

```r
data(BGSall)
HT18 = BGSall$HT18
HT9 = BGSall$HT9
Sex = BGSall$Sex

#(a)
plot(HT9, HT18, type = 'n', main = 'HT18 vs HT9')
points(HT9[Sex ==0], HT18[Sex == 0], col = 2)
```

```
points(HT9[Sex ==1], HT18[Sex == 1], col = 4)
legend('topleft', bty = 'n', col = c('red','blue'),c('Male', 'Female'), pch = 1)
```

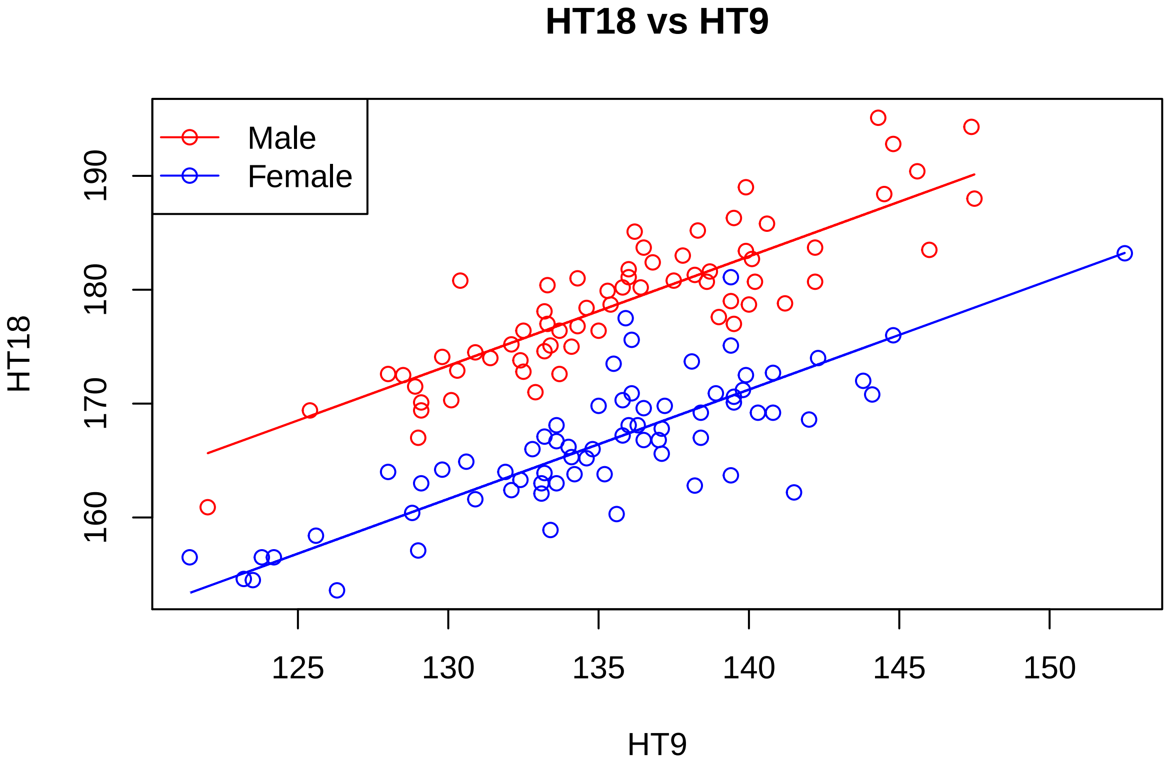



```
#As seen in the plot above, a linear relationship appears appropriate for
#both groups, and the slopes look similar between males and females.

#(b)
#The interval is computed below.
fit = lm(HT18 ~ HT9 + Sex)
confint(fit)
```

```
##                   2.5 %     97.5 %
## (Intercept)  34.0112360  63.023384
## HT9           0.8534845   1.066628
## Sex         -12.8635477 -10.528134
```

Hence, we are 95% con dent that, for any xed value of HT9, females are between 10.53 and 12.86 cm shorter than males.

```
#(c)
bs = coef(fit)
yhat1 = bs[1] + HT9[Sex == 0] * bs[2]
yhat2 = bs[1] + HT9[Sex == 1] * bs[2] + bs[3]

plot(HT9, HT18, type = 'n', main = 'HT18 vs HT9')
points(HT9[Sex == 0], HT18[Sex == 0], col = 2)
points(HT9[Sex == 1], HT18[Sex == 1], col = 4)
lines(HT9[Sex == 0], yhat1, col = 2)
lines(HT9[Sex == 1], yhat2, col = 4)
legend('topleft', col = c('red', 'blue'), c('Male', 'Female'),pch = 1, lty =c(1,1))
```


HT18 vs HT9
Male
Female
190
180
170
160
HT18
125
130
135
140
145
150
HT9