Homework 7

Animesh Sengupta

10/25/2022

```r
setwd("/Users/animeshsengupta/Work Directory/DACSS/STAT625/Homeworks")
library(alr4)  # loads the installed package into the workspace so you can use it
```

```
## Loading required package: car

## Loading required package: carData

## Loading required package: effects

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

```r
library(summarytools)
library(ggplot2)
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

```r
library(splines)
library(boot)
```

```
##
## Attaching package: 'boot'

## The following object is masked from 'package:car':
##
##     logit
```

```
library(sandwich)
```

Answer 7.3

7.3.1

I believe that is not true , we need to provide lower weights to observations from oversampled subpopulation. The reason for assigning lower weights is because oversampled population have a higher sampling probability due to bootstrapping. Sampling probability and raw weights have an inverse relationship , hence higher sampling probability , lower the sampling weights.

7.3.2

Irrespective of wether the observations are from the responders or nonresponders, when combining them they will eventually increase the size of observation because theyre treated the same. The weight will increase since adding more observations will only increase the number of observations.

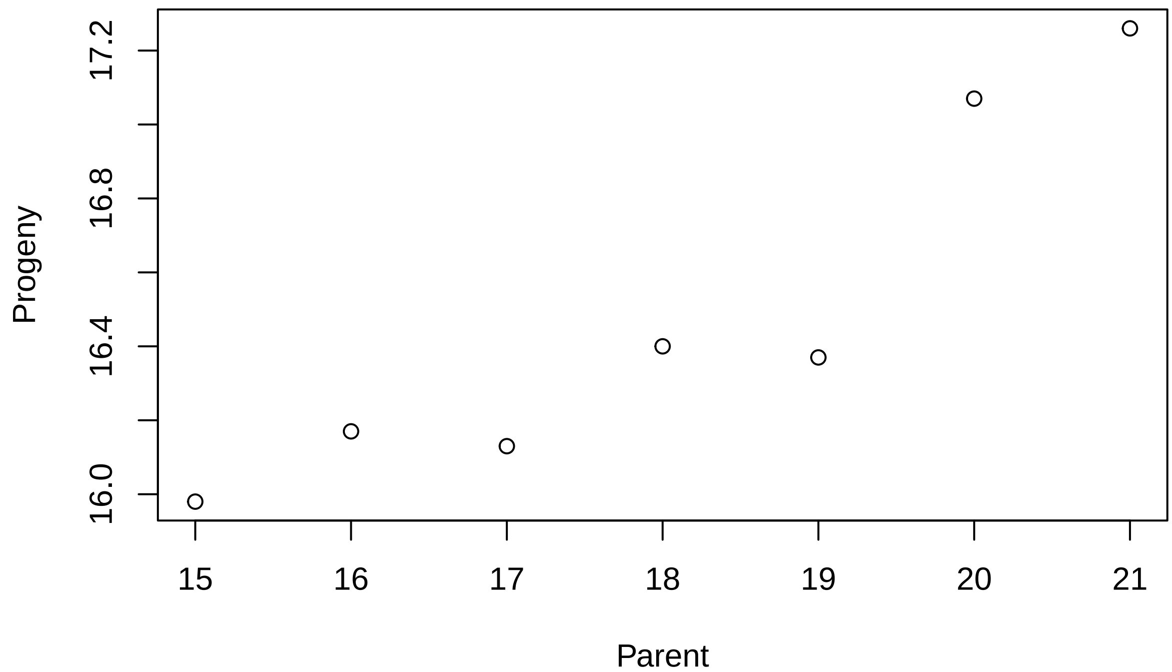
Answer 7.7

7.7.1

```
colnames(galtonpeas)
```

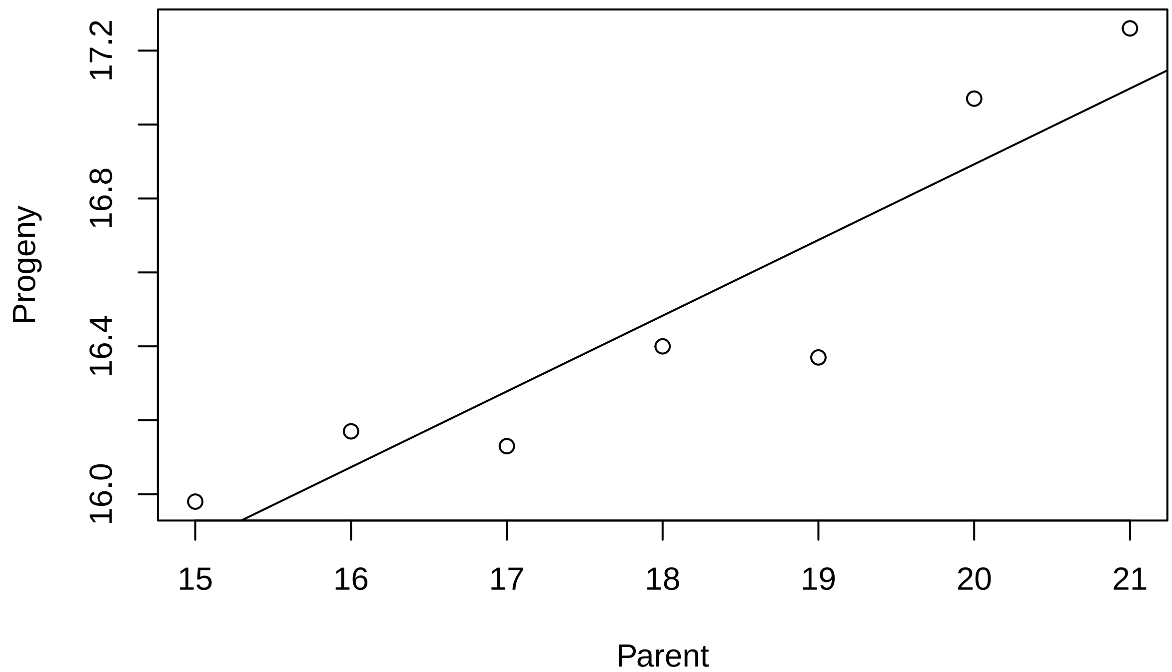
```
## [1] "Parent"  "Progeny" "SD"
```

```
plot(Progeny~Parent,galtonpeas)
```



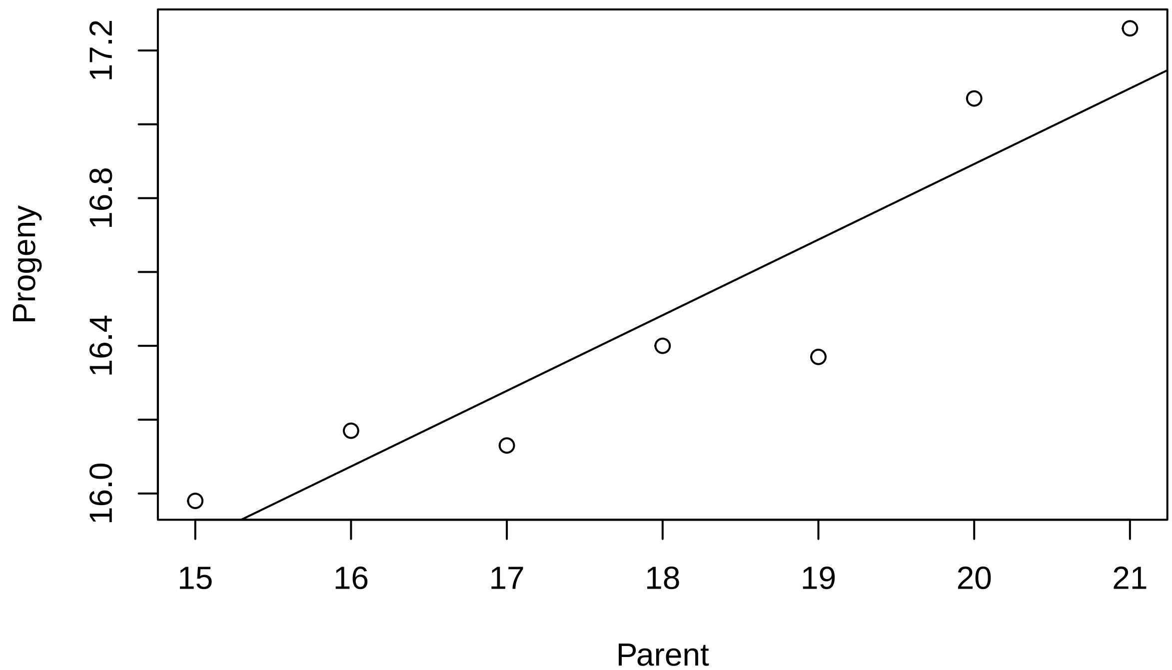
7.7.2

```
gplm=lm(Progeny~Parent,data=galtonpeas,weights=1/SD^2)
plot(Progeny~Parent,galtonpeas)
abline(gplm)
```



7.7.3

```
gplm=lm(Progeny~Parent,data=galtonpeas,weights=1/SD^2)
plot(Progeny~Parent,galtonpeas)
abline(gplm)
```



7.7.3

The experimental biases should decrease the slope and it would increase the estimates of error since it may increase the variance.

answer 7.9

7.9.1

```
cint<-t.test(log(UN11$fertility))$conf.int
cint
```

```
## [1] 0.8500350 0.9744333
## attr(,"conf.level")
## [1] 0.95
```

```
cint2<-rbind(cint,exp(cint))
cint2
```

```
##          [,1]      [,2]
## cint 0.850035 0.9744333
##      2.339729 2.6496653
```

###7.9.2

```
set.seed(12345)
median_data<-function(d,i){
  median(d[i])
}
b<-boot(UN11$fertility,median_data,R=1000)
boot.ci(b)
```

```
## Warning in boot.ci(b): bootstrap variances needed for studentized intervals
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = b)
##
## Intervals :
## Level      Normal              Basic
## 95%   ( 2.092,  2.419 )   ( 2.094,  2.389 )
##
## Level     Percentile            BCa
## 95%   ( 2.135,  2.430 )   ( 2.098,  2.422 )
## Calculations and Intervals on Original Scale
```

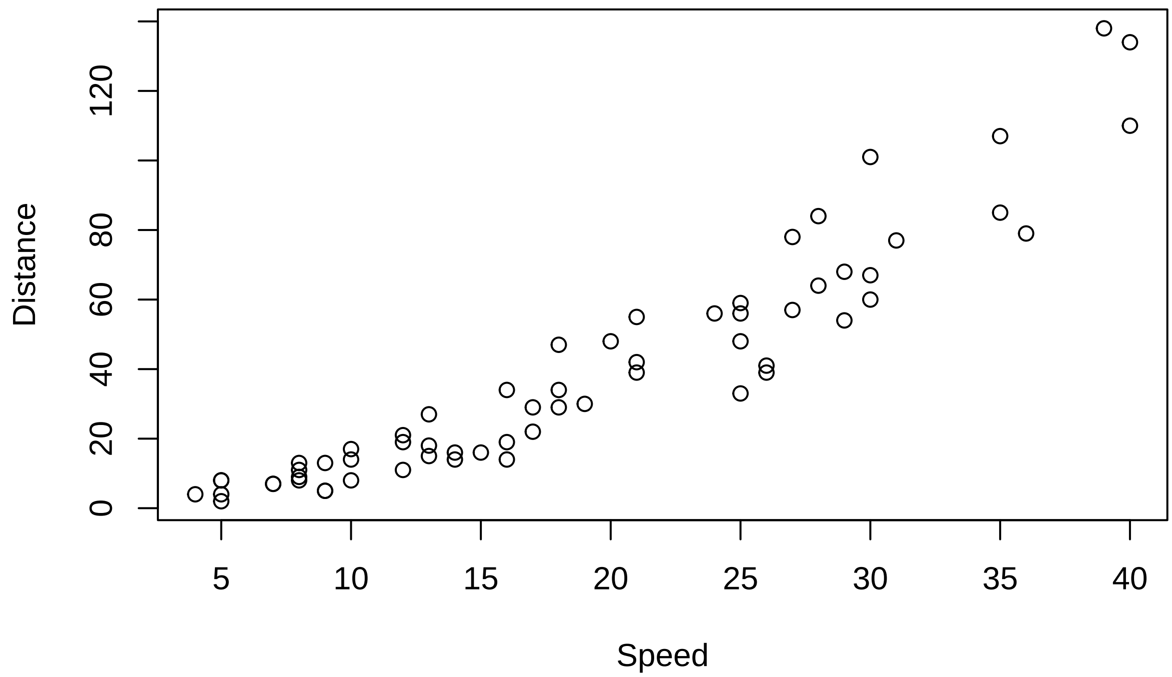
7.6

###7.6.1

```
colnames(stopping)
```

```
## [1] "Speed"    "Distance"
```

```
plot(Distance ~ Speed,data=stopping)
```



As seen on the graph, there is no linear relationship pattern between the two. The scatterploty tends to curve upwards hence the only conclusion that can be made, is that the regression fit will be polynomial in quadratic.

7.6.2

```
dlm<-lm(Distance~I(Speed^2),data=stopping)
summary(dlm)
```

```
##
## Call:
## lm(formula = Distance ~ I(Speed^2), data = stopping)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.3816  -4.9930  -0.3455   4.2631  28.3327
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.134768   1.848709   2.777   0.0073 **
## I(Speed^2)  0.075036   0.002979  25.192   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.89 on 60 degrees of freedom
```

```
## Multiple R-squared:  0.9136, Adjusted R-squared:  0.9122
## F-statistic: 634.6 on 1 and 60 DF,  p-value: < 2.2e-16
```

```
l1<-leveneTest(stopping$Distance,stopping$Speed)
```

```
## Warning in leveneTest.default(stopping$Distance, stopping$Speed): stopping$Speed
## coerced to factor.
```

```
l1
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group 27  1.0159 0.4773
##       34
```

7.6.3

```
dlm2<-lm(Distance~I(Speed^2),data=stopping,weights = 1/Speed)
summary(dlm2)
```

```
##
## Call:
## lm(formula = Distance ~ I(Speed^2), data = stopping, weights = 1/Speed)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0513 -1.2538 -0.1001  1.1759  5.0976
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.378403   1.180031    3.71 0.000455 ***
## I(Speed^2)  0.076334   0.002875   26.55  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.013 on 60 degrees of freedom
## Multiple R-squared:  0.9216, Adjusted R-squared:  0.9203
## F-statistic: 704.9 on 1 and 60 DF,  p-value: < 2.2e-16
```

7.6.4

```
sandwich(dlm)
```

```
##              (Intercept)    I(Speed^2)
## (Intercept)  1.722187130 -3.454315e-03
## I(Speed^2)  -0.003454315  1.422851e-05
```

7.11

$d(E(Y/X)/dX_1 = \beta_1 + 2\beta_3 X_1 + \beta_5 X_2$ $d(E(Y/X)/dX_2 = \beta_2 + 2\beta_4 X_2 + \beta_5 X_1$

solving for zero gives us in the above two equations gives us: $X1 = (\beta_2\beta_5 - 2\beta_1\beta_4)/(4\beta_3\beta_4 - \beta_5^2)$ $X2 = (\beta_1\beta_5 - 2\beta_2\beta_3)/(4\beta_3\beta_4 - \beta_5^2)$

```
clm=lm(Y~X1+X2+I(X2^2)+I(X1^2)+X1:X2,data=cakes)
clm$coefficients
```

```
##  (Intercept)           X1           X2      I(X2^2)      I(X1^2)        X1:X2
## -2204.484987    25.917558     9.918267    -0.011950    -0.156875    -0.041625
```

```
x1<-(clm$coefficients[2]*clm$coefficients[5]-2*clm$coefficients[1]*clm$coefficients[4])/(4*clm$coeffici
x2<-(clm$coefficients[1]*clm$coefficients[5]-2*clm$coefficients[2]*clm$coefficients[3])/(4*clm$coeffici
```

7.13

```
colnames(Transact)
```

```
## [1] "t1"   "t2"   "time"
```

```
tlm<-lm(time~t1+t2,Transact)
deltaMethod(tlm,"t1/t2")
```

```
##       Estimate      SE   2.5 % 97.5 %
## t1/t2  2.68465 0.31899 2.05945 3.3099
```