Lab 10: Bootstrap Bonus Lab

Frank Woodling

May 12, 2016

1.

```
GPA <- read.delim("C:/Users/Frank/Desktop/STAT 3480/Lab 10/GPA.txt")


attach(GPA)

summary(lm(CollGPA~SAT))
```

```
##
## Call:
## lm(formula = CollGPA ~ SAT)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46421 -0.47381 -0.00147  0.29138  1.69570
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.1518923  0.3079925   0.493    0.623
## SAT         0.0018020  0.0002968   6.071 2.42e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.642 on 98 degrees of freedom
## Multiple R-squared:  0.2733, Adjusted R-squared:  0.2659
## F-statistic: 36.85 on 1 and 98 DF,  p-value: 2.417e-08
```

The equattion is Y = 0.0018*SAT + 0.1520. For each 0.00018020 increase in the SAT score their GPA will increase by 1.

2.

```
### create our data and calculate thetahat, the slope of the regression line
oursample = GPA
thetahat = lm(CollGPA ~ SAT, data=oursample)$coeff[2]
thetahat
```

```
##        SAT
## 0.00180201
```

```r
thetahat.b = rep(NA,1000)
for (i in 1:1000) {
### draw the bootstrap sample and calculate thetahat.b
index = 1:100
bootindex = sample(index, 100, replace=T)
bootsample = oursample[bootindex,]
thetahat.b[i] = lm(CollGPA ~ SAT, data=bootsample)$coeff[2]
}


### draw the bootstrap sample
index = 1:100
bootindex = sample(index, 100, replace=T)
bootsample = oursample[bootindex,]
bootsample
```

```
##       CollGPA  SAT
## 37       2.46 1090
## 3        3.75 1466
## 31       3.50 1034
## 48       2.24 1158
## 27       1.61  644
## 32       3.18 1202
## 37.1     2.46 1090
## 100      0.89  864
## 36       1.57 1038
## 49       0.45  676
## 31.1     3.50 1034
## 100.1    0.89  864
## 46       2.03  886
## 57       1.80  814
## 97       2.64 1304
## 36.1     1.57 1038
## 26       1.77  744
## 73       1.87  954
## 31.2     3.50 1034
## 46.1     2.03  886
## 7        1.38 1058
## 26.1     1.77  744
## 86       1.99 1182
## 98       2.08 1212
## 35       1.54  952
## 60       3.44 1424
## 68       0.97  776
## 8        1.50 1008
## 92       2.15  400
## 33       2.39 1018
## 58       1.29  778
## 74       2.00 1000
## 44       2.10 1222
## 1        2.04 1070
## 45       1.40 1120
## 39       2.11 1096
```

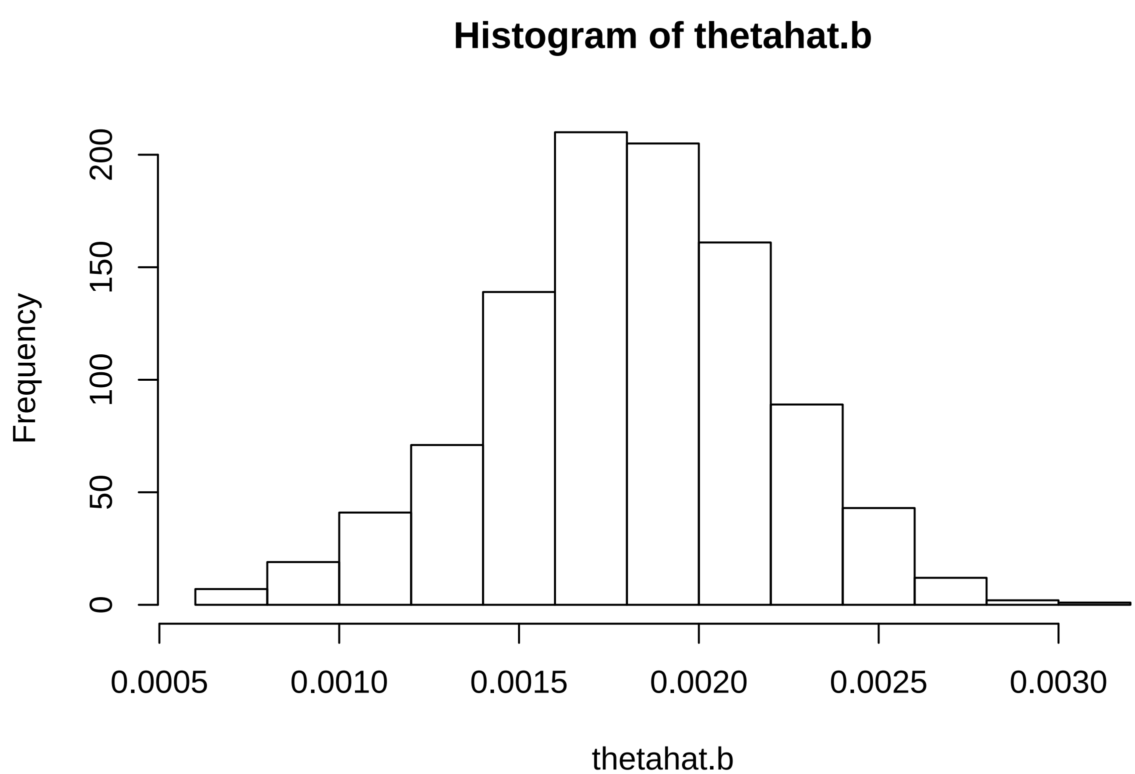
```
## 10       4.01 1200
## 4        1.10  706
## 100.2    0.89  864
## 80       1.88  856
## 62       2.06 1056
## 100.3    0.89  864
## 20       2.05 1054
## 4.1      1.10  706
## 37.2     2.46 1090
## 94       2.29  776
## 9        1.38 1104
## 39.1     2.11 1096
## 16       0.81  790
## 32.1     3.18 1202
## 68.1     0.97  776
## 96       1.80  772
## 62.1     2.06 1056
## 65       2.00  852
## 40       2.04 1114
## 72       3.09 1084
## 79       2.01 1000
## 28       0.99  842
## 64       1.80 1352
## 1.1      2.04 1070
## 89       3.02 1374
## 73.1     1.87  954
## 65.1     2.00  852
## 81       1.64  798
## 94.1     2.29  776
## 45.1     1.40 1120
## 32.2     3.18 1202
## 3.1      3.75 1466
## 52       2.56 1264
## 20.1     2.05 1054
## 23       0.38  456
## 60.1     3.44 1424
## 88       1.79  910
## 80.1     1.88  856
## 65.2     2.00  852
## 20.2     2.05 1054
## 95       2.39 1134
## 39.2     2.11 1096
## 38       2.42  694
## 93       1.46  998
## 19       2.00 1046
## 12       1.29  848
## 28.1     0.99  842
## 3.2      3.75 1466
## 14       3.11 1246
## 32.3     3.18 1202
## 97.1     2.64 1304
## 54       2.92 1292
## 47       1.99 1126
## 98.1     2.08 1212
```

```
## 1.2       2.04 1070
## 22        2.55  940
## 70        1.31 1232
## 9.1       1.38 1104
## 40.1      2.04 1114
## 48.1      2.24 1158
## 59        1.68  800
## 70.1      1.31 1232
## 6         0.05  756
## 33.1      2.39 1018
```

We can look at the bootindex and seee if the same index is repeated twice. In my case I see both 1 and 6 repeated.

3.

```r
hist(thetahat.b)
```




```r
quantile(thetahat.b, .025); quantile(thetahat.b, .975)
```

```
##         2.5%
## 0.0009990322
```

```
##       97.5%
## 0.002554662
```

Out of an infinite amount of confidence intervals 95% of bootstrap intervals will fall between 0.001126614 and 0.002549624.

4.

It would be equal to .05, so 95% of intervals would bracket the true mean.

5.

```r
quantile(thetahat.b, .005); quantile(thetahat.b, .995)
```

```
##         0.5%
## 0.0007898025
```

```
##       99.5%
## 0.002716075
```

The interval is between 0.0009297782 and 0.002757923 so the p-value would have to be less than 0.01. We can say than of an infinite amount of bootstraps 99% are between that interval.

6.

It would have to contain 99% if all sample's means to be a 99% confidence interval.

7.

```r
detach(GPA)
GPAfull <- read.csv("C:/Users/Frank/Desktop/STAT 3480/Lab 10/GPA_full.txt", sep="")
attach(GPAfull)

summary(lm(CollGPA ~ SAT + HSGPA))
```

```
##
## Call:
## lm(formula = CollGPA ~ SAT + HSGPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12153 -0.44120  0.00954  0.38198  1.80356
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0881312  0.2866638  -0.307 0.759169
## SAT          0.0012167  0.0003011   4.041 0.000107 ***
## HSGPA        0.4071133  0.0905946   4.494 1.94e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.587 on 97 degrees of freedom
## Multiple R-squared:  0.3985, Adjusted R-squared:  0.3861
## F-statistic: 32.13 on 2 and 97 DF,  p-value: 1.963e-11
```

I am getting some strange regression results. It lists 30 different coefficients for SAT. No matter how I change it there are a ton of different SAT coefficients.

8.

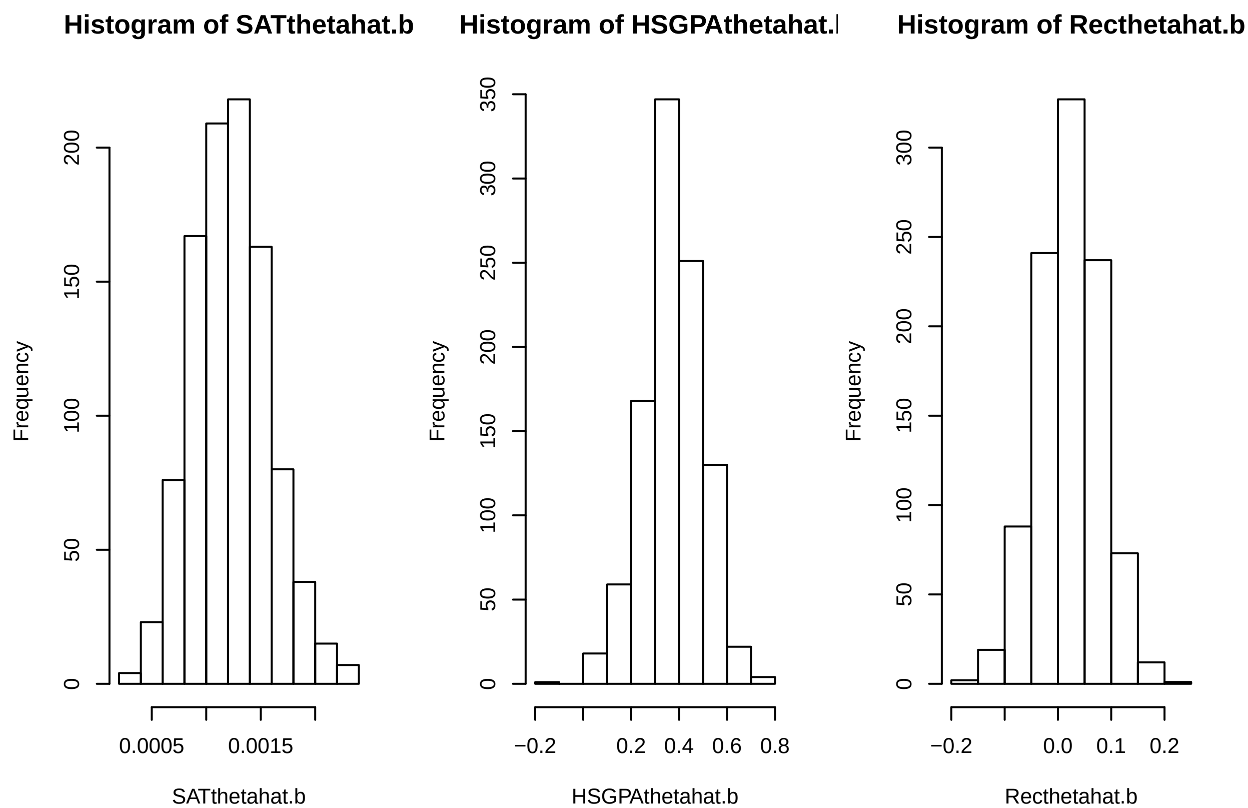
```
### create our data
oursample = GPAfull
SATthetahat = lm(CollGPA ~ SAT + HSGPA + Rec, data=oursample)$coeff[2]
HSGPAthetahat = lm(CollGPA ~ SAT + HSGPA + Rec, data=oursample)$coeff[3]
Recthetahat = lm(CollGPA ~ SAT + HSGPA + Rec, data=oursample)$coeff[4]
SATthetahat; HSGPAthetahat; Recthetahat
```

```
##        SAT
## 0.00122693
```

```
##     HSGPA
## 0.3763511
```

```
##        Rec
## 0.02268425
```

```
SATthetahat.b = rep(NA,1000); HSGPAthetahat.b = rep(NA,1000); Recthetahat.b = rep(NA,1000)
for (i in 1:1000) {
### draw the bootstrap sample and calculate thetahat.b
index = 1:100
bootindex = sample(index, 100, replace=T)
bootsample = oursample[bootindex,]
SATthetahat.b[i] = lm(CollGPA ~ SAT + HSGPA + Rec, data=bootsample)$coeff[2]
HSGPAthetahat.b[i] = lm(CollGPA ~ SAT + HSGPA + Rec, data=bootsample)$coeff[3]
Recthetahat.b[i] = lm(CollGPA ~ SAT + HSGPA + Rec, data=bootsample)$coeff[4]
}
par(mfrow=c(1,3))
hist(SATthetahat.b); hist(HSGPAthetahat.b); hist(Recthetahat.b)
```




```
quantile(SATthetahat.b, .025); quantile(SATthetahat.b, .975)
```

```
##         2.5%
## 0.0005926126
```

```
##       97.5%
## 0.001965745
```

```
quantile(HSGPAthetahat.b, .025); quantile(HSGPAthetahat.b, .975)
```

```
##      2.5%
## 0.1152664
```

```
##     97.5%
## 0.6020128
```

```
quantile(Recthetahat.b, .025); quantile(Recthetahat.b, .975)
```

```
##        2.5%
## -0.09657798
```

```
##     97.5%
## 0.1327022
```

9.

In only one interval does the confidence interval bracket 0.

10.

```
quantile(SATthetahat.b, .025);
```

```
##          2.5%
## 0.0005926126
```

```
quantile(SATthetahat.b, .975)
```

```
##        97.5%
## 0.001965745
```

```
quantile(HSGPAthetahat.b, .025); quantile(HSGPAthetahat.b, .975)
```

```
##      2.5%
## 0.1152664
```

```
##     97.5%
## 0.6020128
```

```
quantile(Recthetahat.b, .025);
```

```
##         2.5%
## -0.09657798
```

```
quantile(Recthetahat.b, .975)
```

```
##     97.5%
## 0.1327022
```

This lab does not give enough information to answer this problem. As we reduce the confidence level our confidence interval decreases.

Summary

```
fit = lm(CollGPA ~ SAT + HSGPA + Rec)
summary(fit)
```

```
##
## Call:
## lm(formula = CollGPA ~ SAT + HSGPA + Rec)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0979 -0.4407 -0.0094  0.3859  1.7606
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.1532639  0.3229381  -0.475 0.636156
## SAT          0.0012269  0.0003032   4.046 0.000105 ***
## HSGPA        0.3763511  0.1142615   3.294 0.001385 **
## Rec          0.0226843  0.0509817   0.445 0.657358
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5895 on 96 degrees of freedom
## Multiple R-squared:  0.3997, Adjusted R-squared:  0.381
## F-statistic: 21.31 on 3 and 96 DF,  p-value: 1.16e-10
```

```
predict(fit, GPAfull, interval="confidence")
```

```
##          fit       lwr      upr
## 1   2.029438 1.9062765 2.152600
## 2   2.801006 2.5324864 3.069525
## 3   3.166493 2.8208280 3.512159
## 4   1.383267 1.1689481 1.597585
## 5   2.632882 2.3301070 2.935657
## 6   0.966544 0.6625307 1.270557
## 7   1.325683 0.9762142 1.675152
## 8   1.983683 1.7570605 2.210306
## 9   2.055997 1.8837369 2.228257
## 10  2.249362 1.9989950 2.499729
## 11  1.906483 1.6882295 2.124737
## 12  1.459536 1.2440628 1.675010
## 13  1.703847 1.5484201 1.859273
## 14  2.685812 2.4619446 2.909680
## 15  2.099849 1.9188338 2.280865
## 16  1.907945 1.6540019 2.161888
## 17  1.823032 1.6634523 1.982611
## 18  2.974871 2.6798643 3.269879
## 19  1.845688 1.6913035 2.000073
## 20  1.968306 1.8065544 2.130057
## 21  2.190352 1.9952264 2.385478
## 22  1.655211 1.4430171 1.867405
## 23  1.144484 0.7936183 1.495349
## 24  2.139090 1.8835455 2.394634
## 25  1.926094 1.5468307 2.305358
## 26  1.166547 0.8825383 1.450556
## 27  1.548165 1.2874613 1.808868
## 28  1.644114 1.3922640 1.895963
## 29  1.675407 1.5284741 1.822340
## 30  1.737939 1.4611458 2.014732
```

```
## 31  2.636872 2.2337953 3.039949
## 32  2.345697 2.1780895 2.513305
## 33  1.864023 1.7319784 1.996068
## 34  2.119239 1.9552107 2.283267
## 35  1.621009 1.4125927 1.829424
## 36  1.828243 1.6725970 1.983888
## 37  2.332580 2.1756511 2.489509
## 38  1.485315 1.2728235 1.697807
## 39  2.351232 2.1909372 2.511527
## 40  2.136113 2.0040219 2.268203
## 41  2.359365 2.1681513 2.550580
## 42  2.025633 1.8073414 2.243925
## 43  1.869586 1.6784414 2.060731
## 44  2.535727 2.2355182 2.835936
## 45  2.124657 1.9915407 2.257772
## 46  1.641749 1.4832926 1.800206
## 47  2.452123 2.2611431 2.643103
## 48  2.020636 1.8384621 2.202811
## 49  1.493442 1.2502544 1.736630
## 50  2.503640 2.3027586 2.704522
## 51  2.336330 2.1885686 2.484091
## 52  2.549829 2.3635673 2.736092
## 53  1.908786 1.6790486 2.138523
## 54  2.361930 2.1267293 2.597131
## 55  1.457689 1.1854867 1.729891
## 56  1.497315 1.2393072 1.755324
## 57  1.715448 1.5307397 1.900155
## 58  1.520428 1.2640245 1.776832
## 59  1.280108 0.9950050 1.565211
## 60  3.054849 2.7635165 3.346182
## 61  2.281242 2.0489036 2.513581
## 62  1.413657 1.1062307 1.721083
## 63  2.405479 2.1229894 2.687969
## 64  2.183494 1.8886571 2.478330
## 65  1.649062 1.5000834 1.798041
## 66  2.142151 1.9900970 2.294204
## 67  2.211693 2.0267201 2.396666
## 68  1.476679 1.2890497 1.664308
## 69  1.700545 1.5168936 1.884196
## 70  1.968519 1.7019298 2.235108
## 71  1.708968 1.3338622 2.084073
## 72  1.926080 1.7645807 2.087579
## 73  1.593251 1.3081044 1.878398
## 74  2.206896 1.9237292 2.490063
## 75  1.937370 1.7761955 2.098545
## 76  1.739876 1.5553217 1.924430
## 77  1.533386 1.3626996 1.704073
## 78  2.529351 2.3289141 2.729787
## 79  1.868386 1.6917100 2.045063
## 80  1.604941 1.4398331 1.770049
## 81  1.593995 1.4020629 1.785928
## 82  2.875601 2.6035453 3.147656
## 83  1.279404 0.9876508 1.571158
## 84  2.106790 1.9059995 2.307581
```

```
## 85  2.000289 1.8566365 2.143940
## 86  2.551042 2.2955405 2.806544
## 87  2.041508 1.9060800 2.176936
## 88  1.829469 1.6699714 1.988967
## 89  3.336189 2.9437301 3.728647
## 90  1.915671 1.7490471 2.082296
## 91  2.782732 2.5118956 3.053569
## 92  1.117174 0.7375910 1.496757
## 93  2.057975 1.8557375 2.260212
## 94  1.736567 1.5369556 1.936179
## 95  2.292580 2.0946400 2.490521
## 96  1.501879 1.3101607 1.693597
## 97  2.286535 2.0334694 2.539601
## 98  2.376681 2.1361680 2.617194
## 99  1.656376 1.4637833 1.848968
## 100 1.403794 1.1229962 1.684591
```

SAT scores, high school GPA, and number of positive recommendation letters are all positive indicators of college GPA.

CI(lm(CollGPA ~ SAT + HSGPA + Rec))