DaigleInClassLabWk11D3.R

2011home

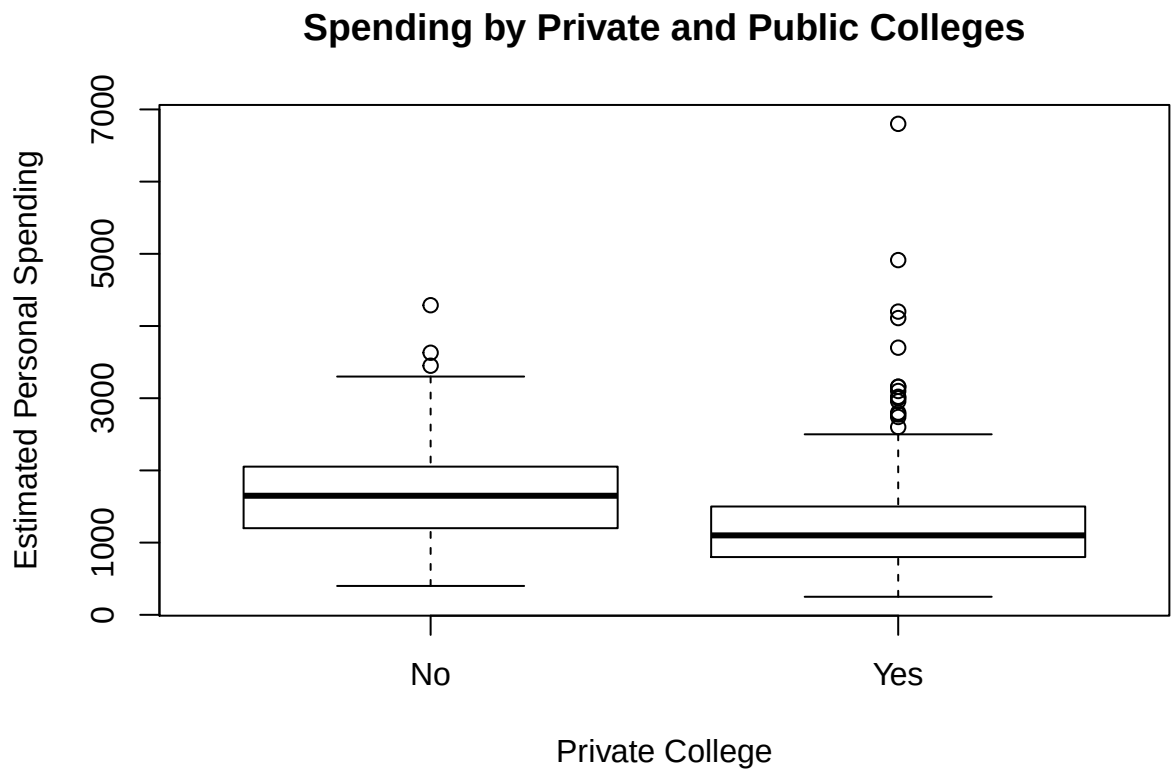
Sun Apr 8 21:59:53 2018

```
# Chris Daigle
# Week11 Day 3 - 6 April

# college <- read.csv("/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/
college <-
  read.csv(
    "/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/DataSets
  )

# College data: Demographic characteristics, tuition, and more of USA colleges
# Private: Public/Private indicator
# Apps: Number of applications recieved
# Enroll: Number of new students enrolled
# Top10perc: New students from top 10% of high school class
# Top25perc: New students from top 25% of high school class
# F.Undergrad: Number of ful-time undergraduates
# P.Undergrad: Number of part-time undergraduates
# Outstate: Out-of-state tuition
# Room.Board: Room and board costs
# Books: Estimated book costs
# Personal: Estimated personal spending
# PhD: Percent of faculty with Ph.D.'s
# Terminal: Percent of faculty with terminal degree
# S.F.Ratio: Student/faculty ratio
# perc.alumni: Percent of alumni who donate
# Expend: Instructional expenditure per student
# Grad.Rate: Graduation rate

# 1. Compare the distributions of "personal" between private school and public school.
# For this, you can first draw box plots and do 2 sample t test.
head(college)
```

```
##                             X Private Apps Accept Enroll Top10perc
## 1 Abilene Christian University     Yes 1660   1232    721        23
## 2           Adelphi University     Yes 2186   1924    512        16
## 3               Adrian College     Yes 1428   1097    336        22
## 4          Agnes Scott College     Yes  417    349    137        60
## 5    Alaska Pacific University     Yes  193    146     55        16
## 6            Albertson College     Yes  587    479    158        38
##   Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD
## 1        52        2885         537     7440       3300   450     2200  70
## 2        29        2683        1227    12280       6450   750     1500  29
## 3        50        1036          99    11250       3750   400     1165  53
## 4        89         510          63    12960       5450   450      875  92
## 5        44         249         869     7560       4120   800     1500  76
## 6        62         678          41    13500       3335   500      675  67
##   Terminal S.F.Ratio perc.alumni Expend Grad.Rate
```

```
## 1       78      18.1          12   7041        60
## 2       30      12.2          16  10527        56
## 3       66      12.9          30   8735        54
## 4       97       7.7          37  19016        59
## 5       72      11.9           2  10922        15
## 6       73       9.4          11   9727        55
```

```
boxplot(
  college$Personal ~ college$Private,
  xlab = "Private College",
  ylab = "Estimated Personal Spending",
  main = 'Spending by Private and Public Colleges'
)
```




```
t.test(college$Personal[college$Private == 'No'], college$Personal[college$Private == "Yes"], var.equal
```

```
##
##  Two Sample t-test
##
## data:  college$Personal[college$Private == "No"] and college$Personal[college$Private == "Yes"]
## t = 8.8991, df = 775, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  360.5093 564.5715
## sample estimates:
## mean of x mean of y
##  1676.981  1214.441
```

```
# 2. Divide the colleges into three groups based on Top10perc. Make the group sizes to be
#the same with each other. (I interpret this as "evenly split the Top10perc group into
# thirds and have those be subgroups")
```

```r
head(college)
```

```
##                              X Private Apps Accept Enroll Top10perc
## 1 Abilene Christian University     Yes 1660   1232    721        23
## 2           Adelphi University     Yes 2186   1924    512        16
## 3               Adrian College     Yes 1428   1097    336        22
## 4          Agnes Scott College     Yes  417    349    137        60
## 5    Alaska Pacific University     Yes  193    146     55        16
## 6            Albertson College     Yes  587    479    158        38
##   Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books Personal PhD
## 1        52        2885         537     7440       3300   450     2200  70
## 2        29        2683        1227    12280       6450   750     1500  29
## 3        50        1036          99    11250       3750   400     1165  53
## 4        89         510          63    12960       5450   450      875  92
## 5        44         249         869     7560       4120   800     1500  76
## 6        62         678          41    13500       3335   500      675  67
##   Terminal S.F.Ratio perc.alumni Expend Grad.Rate
## 1       78      18.1          12   7041        60
## 2       30      12.2          16  10527        56
## 3       66      12.9          30   8735        54
## 4       97       7.7          37  19016        59
## 5       72      11.9           2  10922        15
## 6       73       9.4          11   9727        55
```

```r
# private <- college[college$Private == 'Yes',]
# public <- college[college$Private == 'No',]
#
# private1 <- subset(private, Personal <= quantile(private$Personal, probs = seq(0,1, 1/3), na.rm = TRU
# private2 <- subset(private, Personal > quantile(private$Personal, probs =
# seq(0,1, 1/3), na.rm = TRUE)[2] & Personal <= quantile(private$Personal, probs
# = seq(0,1, 1/3), na.rm = TRUE)[3])
# private3 <- subset(private, Personal > quantile(private$Personal, probs = seq(0,1, 1/3), na.rm = TRUE
#
# quantile(public$Personal, probs = seq(0, 1, 1/3), na.rm = TRUE)
# public1 <- subset(public, Personal <= 1368)
# public2 <- subset(public, Personal > 1368 & Personal <=1951)
# public3 <- subset(public, Personal > 1951)
#
# college1 <- subset(college, Personal <= quantile(college$Personal, probs =
# seq(0,1, 1/3), na.rm = TRUE)[2]) college2 <- subset(college, Personal >
# quantile(college$Personal, probs = seq(0,1, 1/3), na.rm = TRUE)[2] & Personal
# <= quantile(college$Personal, probs = seq(0,1, 1/3), na.rm = TRUE)[3])
# college3 <- subset(college, Personal > quantile(college$Personal, probs =
# seq(0,1, 1/3), na.rm = TRUE)[3])

# I realize now this was not the desired result, here is the answer:

college <- college[order(college$Top10perc), ]
head(college)
```

```
##                                X Private Apps Accept Enroll Top10perc
## 101    Center for Creative Studies     Yes  601    396    203         1
## 198  Fayetteville State University      No 1455   1064    452         1
## 411      North Adams State College      No 1563   1005    240         1
```

```
## 390                 Morris College     Yes  882    730    330         2
## 713      Virginia State University      No 2996   2440    704         2
## 113 Christopher Newport University      No  883    766    428         3
##     Top25perc F.Undergrad P.Undergrad Outstate Room.Board Books Personal
## 101        20         525         323    11230       6643  2340      620
## 198        16        2632         617     6806       2550   350      766
## 411        19        1380         136     5542       4330   500     1000
## 390        13         926          12     4515       2550   850     2100
## 713        30        3006         338     5587       4845   500      600
## 113        37        2910        1749     7860       4750   525     1889
##     PhD Terminal S.F.Ratio perc.alumni Expend Grad.Rate
## 101   8       58       6.8           4  13025        47
## 198  75       75      15.1          10   6972        24
## 411  65       71      14.2          17   6562        57
## 390  53       60      18.6          34   6990        60
## 713  61       63      16.0          11   5733        31
## 113  80       82      21.2          16   4639        48
```

```
college1 <- college[1:259,]
college2 <- college[260:518,]
college3 <- college[519:777,]

# Compare the mean of "Grad.Rate" among these three groups. If you conclude there is any difference, id

mean1 <- summary(college1$Grad.Rate)[4]
mean1
```

```
##     Mean
## 56.64865
```

```
mean2 <- summary(college2$Grad.Rate)[4]
mean2
```

```
##     Mean
## 64.74903
```

```
mean3 <- summary(college3$Grad.Rate)[4]
mean3
```

```
##     Mean
## 74.99228
```

```
# Put simply, each of these are unequal. No t-test is needed.
```