

# Homework 5 Solutions

Note for there are multiple ways to complete this homework. The below represent the methods I used (today, as I write this), and alternative methods are certainly possible.

1. Load the *countyComplete* dataset from the *openintro* package.

- Use a `for()` loop to calculate the grand mean for the following variables: white, black, native, asian, pac\_isl, two\_plus\_races, hispanic, white\_not\_hispanic, hs\_grad, bachelors, housing\_units, home\_ownership, housing\_multi\_unit, median\_val\_owner\_occupied, households.
- Calculate the same values using an `apply` family loop.
- Conduct a test (i.e., not just visual comparison) to ensure the same means were produced by both methods.

```
library(openintro)
data(countyComplete)
d <- countyComplete

varLocs <- c(10:17, 21, 22, 25:29)

# for loop
means <- rep(NA, length(varLocs))
for(i in seq_along(means)) {
  means[i] <- mean(d[,varLocs[i]], na.rm = TRUE)
}

round(means, 2)
```

```
## [1]      82.89      8.93      2.03      1.17      0.15      1.98      8.28
## [8]      78.29     83.11     19.03    41904.15     73.26     12.33    132544.93
## [15]    36346.16
```

```
# apply family loop
means2 <- apply(d[,varLocs], 2, mean, na.rm = TRUE)
round(means2, 2)
```

```
##           white           black
##           82.89           8.93
##           native          asian
##           2.03           1.17
##           pac_isl       two_plus_races
##           0.15           1.98
##           hispanic     white_not_hispanic
##           8.28           78.29
##           hs_grad       bachelors
##           83.11          19.03
##           housing_units  home_ownership
##           41904.15       73.26
##           housing_multi_unit median_val_owner_occupied
##           12.33          132544.93
##           households
##           36346.16
```

```
table(means == means2)
```

```
##
## TRUE
## 15
```

2. Calculate the *difference* between the mean for each of the preceding variables and the mean for each state. Do so using only `for()` loops, and then again using only the `apply` family of loops.

```
## For loop method

stateMeanDiff <- matrix(#      NROWS                                NCOLS
                        rep(NA, length(unique(d$state)) * length(varLocs)),
                        ncol = length(varLocs)
                        )

s <- split(d[,varLocs], d$state)

for(i in seq_along(s)) {
  for(j in 1:ncol(s[[i]])) {
    stateMeanDiff[i, j] <- means[j] - mean(s[[i]][,j], na.rm = TRUE)
  }
}

# Make it prettier
dimnames(stateMeanDiff) <- list(names(s), names(s[[1]]))

# Show first six rows and columns
stateMeanDiff[1:6, 1:6]
```

```
##           white      black      native      asian      pac_isl
## Alabama    15.637508 -19.461987   1.3397469  0.5872629478  0.09289995
## Alaska     30.581821   7.737447 -30.6805824 -3.7267360228 -0.32376671
## Arizona    12.985269   7.103884 -11.7270192 -0.0002992412  0.01401107
## Arkansas     4.135936  -6.930783   1.3623141  0.5271782363  0.01764743
## California 12.811131   5.665033  -0.3305824 -5.4991498159 -0.13391997
## Colorado   -4.281919   7.401979   0.6356891  0.1101174255  0.06801107
##           two_plus_races
## Alabama         0.6272019
## Alaska        -5.5131996
## Arizona        -1.0835444
## Arkansas         0.3457889
## California     -2.5994065
## Colorado       -0.4469819
```

```
## apply family method
stateMeanDiff2 <- sapply(seq_along(varLocs), function(i) {
  means[i] - tapply(d[,varLocs[i]], d[,1], mean, na.rm = TRUE)
})

#Give column names
colnames(stateMeanDiff2) <- names(d[,varLocs])
stateMeanDiff2[1:6, 1:6]
```

```
##           white      black      native      asian      pac_isl
## Alabama    15.637508 -19.461987   1.3397469  0.5872629478  0.09289995
## Alaska     30.581821   7.737447 -30.6805824 -3.7267360228 -0.32376671
## Arizona     12.985269   7.103884 -11.7270192 -0.0002992412  0.01401107
## Arkansas     4.135936  -6.930783   1.3623141  0.5271782363  0.01764743
## California  12.811131   5.665033  -0.3305824 -5.4991498159 -0.13391997
## Colorado   -4.281919   7.401979   0.6356891  0.1101174255  0.06801107
##           two_plus_races
## Alabama         0.6272019
## Alaska        -5.5131996
## Arizona        -1.0835444
## Arkansas         0.3457889
## California     -2.5994065
## Colorado       -0.4469819
```

```
table(stateMeanDiff == stateMeanDiff2)
```

```
##
## TRUE
## 765
```

3. Load the *births* dataset. Use `by()` or `tapply()` to calculate the mean birth weight by the following: weeks, premature, sexBaby, smoke. Calculate the same means using `aggregate()`

- Use the output to report the following:
  - Mean birth weights across weeks for full term non-smokers.
  - Mean birth weights across weeks for premature babies from mothers who smoke.
  - Mean birth weights for weeks 38-40 for full term males across smokers
  - Mean birth weights across weeks for female babies of nonsmoker mothers, between premie and full term.

```
data(births)
d2 <- births

# by method

weightMeans <- by(d2$weight,
                  as.list(d2[,c(3, 4, 8, 9)]),
                  mean, na.rm = TRUE)
weightMeans[, "full term", , "nonsmoker"]
```

```
##           sexBaby
## weeks  female    male
## 26      NA      NA
## 28      NA      NA
## 29      NA      NA
## 32      NA      NA
## 33      NA      NA
## 34      NA      NA
## 35      NA      NA
## 36      NA      NA
## 37 7.090000 7.500000
## 38 7.174444 7.056667
```

```
##      39 7.272222 7.920000
##      40 7.723333 7.547333
##      41 7.752857 8.202000
##      42 6.500000 7.780000
##      43 7.810000      NA
##      44      NA 7.655000
```

```
weightMeans[ , "premie", "smoker"]
```

```
##      sexBaby
## weeks female  male
##      26      NA   NA
##      28      NA   NA
##      29      NA   NA
##      32  2.19 2.690
##      33      NA 3.220
##      34      NA   NA
##      35  5.50 4.500
##      36      NA 6.125
##      37      NA   NA
##      38      NA   NA
##      39      NA   NA
##      40      NA   NA
##      41      NA   NA
##      42      NA   NA
##      43      NA   NA
##      44      NA   NA
```

```
weightMeans[10:14, "full term", "male", ]
```

```
##      smoke
## weeks nonsmoker  smoker
##      38 7.056667 7.282500
##      39 7.920000 7.583333
##      40 7.547333 7.937500
##      41 8.202000 7.800000
##      42 7.780000 7.380000
```

```
weightMeans[ , , "female", "nonsmoker"]
```

```
##      premature
## weeks full term premie
##      26      NA   NA
##      28      NA 1.630
##      29      NA 2.630
##      32      NA 6.500
##      33      NA 5.690
##      34      NA   NA
##      35      NA   NA
##      36      NA 6.845
##      37 7.090000   NA
##      38 7.174444   NA
```

```
##    39  7.272222    NA
##    40  7.723333    NA
##    41  7.752857    NA
##    42  6.500000    NA
##    43  7.810000    NA
##    44      NA     NA
```

```
# aggregate method
```

```
weightMeans2 <- aggregate(weight ~ weeks + premature + sexBaby + smoke,
  data = d2, mean, na.rm = TRUE)
subset(weightMeans2, premature == "full term" & smoke == "nonsmoker")
```

```
##    weeks premature sexBaby    smoke    weight
## 1     37 full term  female nonsmoker 7.090000
## 2     38 full term  female nonsmoker 7.174444
## 3     39 full term  female nonsmoker 7.272222
## 4     40 full term  female nonsmoker 7.723333
## 5     41 full term  female nonsmoker 7.752857
## 6     42 full term  female nonsmoker 6.500000
## 7     43 full term  female nonsmoker 7.810000
## 13    37 full term   male nonsmoker 7.500000
## 14    38 full term   male nonsmoker 7.056667
## 15    39 full term   male nonsmoker 7.920000
## 16    40 full term   male nonsmoker 7.547333
## 17    41 full term   male nonsmoker 8.202000
## 18    42 full term   male nonsmoker 7.780000
## 19    44 full term   male nonsmoker 7.655000
```

```
subset(weightMeans2, premature == "premie" & smoke == "smoker")
```

```
##    weeks premature sexBaby    smoke    weight
## 31     32    premie  female smoker    2.190
## 32     35    premie  female smoker    5.500
## 40     32    premie   male smoker    2.690
## 41     33    premie   male smoker    3.220
## 42     35    premie   male smoker    4.500
## 43     36    premie   male smoker    6.125
```

```
subset(weightMeans2, weeks %in% 38:42 & premature == "full term" &
  sexBaby == "male")
```

```
##    weeks premature sexBaby    smoke    weight
## 14     38 full term   male nonsmoker 7.056667
## 15     39 full term   male nonsmoker 7.920000
## 16     40 full term   male nonsmoker 7.547333
## 17     41 full term   male nonsmoker 8.202000
## 18     42 full term   male nonsmoker 7.780000
## 34     38 full term   male    smoker 7.282500
## 35     39 full term   male    smoker 7.583333
## 36     40 full term   male    smoker 7.937500
## 37     41 full term   male    smoker 7.800000
## 38     42 full term   male    smoker 7.380000
```

```
subset(weightMeans2, sexBaby == "female" & smoke == "nonsmoker")
```

##	weeks	premature	sexBaby	smoke	weight
## 1	37	full term	female	nonsmoker	7.090000
## 2	38	full term	female	nonsmoker	7.174444
## 3	39	full term	female	nonsmoker	7.272222
## 4	40	full term	female	nonsmoker	7.723333
## 5	41	full term	female	nonsmoker	7.752857
## 6	42	full term	female	nonsmoker	6.500000
## 7	43	full term	female	nonsmoker	7.810000
## 8	28	premie	female	nonsmoker	1.630000
## 9	29	premie	female	nonsmoker	2.630000
## 10	32	premie	female	nonsmoker	6.500000
## 11	33	premie	female	nonsmoker	5.690000
## 12	36	premie	female	nonsmoker	6.845000