Homework1

Jon Campbell

```
library(tidyverse)
```

Part 1

Question 1

```
Y_obs <- c(8.62,1.48,8.93,9.57,2.65,7.3,.06,1.72,2.19,7.32,7.53,7.62)
Z <- c(rep(0,6), rep(1,6))
```

a)

```
Y_treat <- Y_obs[Z == 1]
Y_notreat <- Y_obs[Z == 0]
tstat_obs <- mean(Y_treat) - mean(Y_notreat)

ind_combos <- combn(12,6, simplify = FALSE)
tstats = c()
for (i in seq_along(ind_combos)) {
  Y_treat_mean <- mean(Y_obs[ind_combos[[i]]])
  Y_notreat_mean <- mean(Y_obs[-ind_combos[[i]]])
  tstats <- c(tstats, Y_treat_mean - Y_notreat_mean)
}

pval <- mean(abs(tstats) >= abs(tstat_obs))
pval
```

```
[1] 0.2705628
```

b)

```r
sampled_tstats <- sample(tstats, size = 1000, replace = TRUE)
pval_1000 <- mean(abs(sampled_tstats) >= abs(tstat_obs))
pval_1000
```

```
[1] 0.252
```

c)

```r
t.test(Y_treat, Y_notreat, var.equal = TRUE)$p.value
```

```
[1] 0.3367792
```

d)

Question 2

a)

```r
X <- c(1:6,1:6)
taus = c()
for (i in unique(X)) {
  tau = Y_obs[Z == 1 & X == i] - Y_obs[Z == 0 & X == i]
  taus = c(taus, tau)
}
tstat_obs = mean(taus)

tstats <- c()
for (i in seq_along(ind_combos)) {
  Z_perm = rep(0, 12)
  Z_perm[ind_combos[[i]]] = 1

  taus = c()
  for (j in unique(X)) {
    tau = Y_obs[Z_perm == 1 & X == j] - Y_obs[Z_perm == 0 & X == j]
    taus = c(taus, tau)
  }
  tstats = c(tstats,mean(taus))
}

pval <- mean(abs(tstats) >= abs(tstat_obs))
pval
```

```
[1] NA
```

b)

```
ind_combos[[70]]
```

```
[1]  1  2  3  7  9 11
```

c)

```
t.test(Y_treat, Y_notreat
       , var.equal = TRUE, paired = TRUE)
```

```

	Paired t-test

data:  Y_treat and Y_notreat
t = -0.99559, df = 5, p-value = 0.3652
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -7.229626  3.192959
sample estimates:
mean difference 
      -2.018333 
```

d)

Question 3

$$Y_i^{obs} = Z_iY_i(1)+(1-Z_i)Y_i(0)\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} Z_iY_i^{obs}-\frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)Y_i^{obs}\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} Z_iY_i(1)-\frac{1}{n_0}\sum_{i=1}^{n}(1-Z_i)Y_i(0)\text{Unde}$$

Question 4

Part 2

Question 1

```
pot_outcomes <- matrix(c(35, 40, 45 ,55, 55 ,55, 65, 70, 25, 30, 45, 55, 60, 65, 75, 80, 3
colnames(pot_outcomes) <- c("Y1_pot","Y0_pot")
sample_ind = combn(1:12, m =4, simplify = FALSE)
rand_assign_ind = combn(1:4, m =2, simplify = FALSE)
```

```
diffs <- matrix(NA, nrow = length(sample_ind), ncol = length(rand_assign_ind))

for (samp in seq_along(sample_ind)) {
  sample <- pot_outcomes[sample_ind[[samp]],]

  for (i in seq_along(rand_assign_ind)) {

    Y1_obs <- sample[rand_assign_ind[[i]],"Y1_pot"]
    Y0_obs <- sample[-rand_assign_ind[[i]],"Y0_pot"]

    difference <- mean(Y1_obs) - mean(Y0_obs)

    diffs[samp,i] <- difference
  }
}
var(as.vector(diffs))
```

```
[1] 228.9755
```

```
var1 = var(pot_outcomes[,"Y1_pot"])
var0 = var(pot_outcomes[,"Y0_pot"])

var01 = sum((pot_outcomes[,"Y1_pot"] -pot_outcomes[,"Y0_pot"] - (mean(pot_outcomes[,"Y1_po

var0/6+var1/6-var01/12
```

```
[1] 75.42614
```

Question 2

```r
diffs <- matrix(NA, nrow = length(sample_ind), ncol = length(rand_assign_ind))

for (samp in seq_along(sample_ind)) {
  for (i in seq_along(rand_assign_ind)) {

    sample <- pot_outcomes[sample_ind[[samp]],]
    Y1_obs <- sample[rand_assign_ind[[i]],"Y1_pot"]
    Y0_obs <- sample[-rand_assign_ind[[i]],"Y0_pot"]
    difference <- mean(Y1_obs) - mean(Y0_obs)

    diffs[samp,i] <- difference
  }
}
var(apply(diffs, MARGIN = 1,FUN = sum))
```

```
[1] 94.50911
```

Part 3

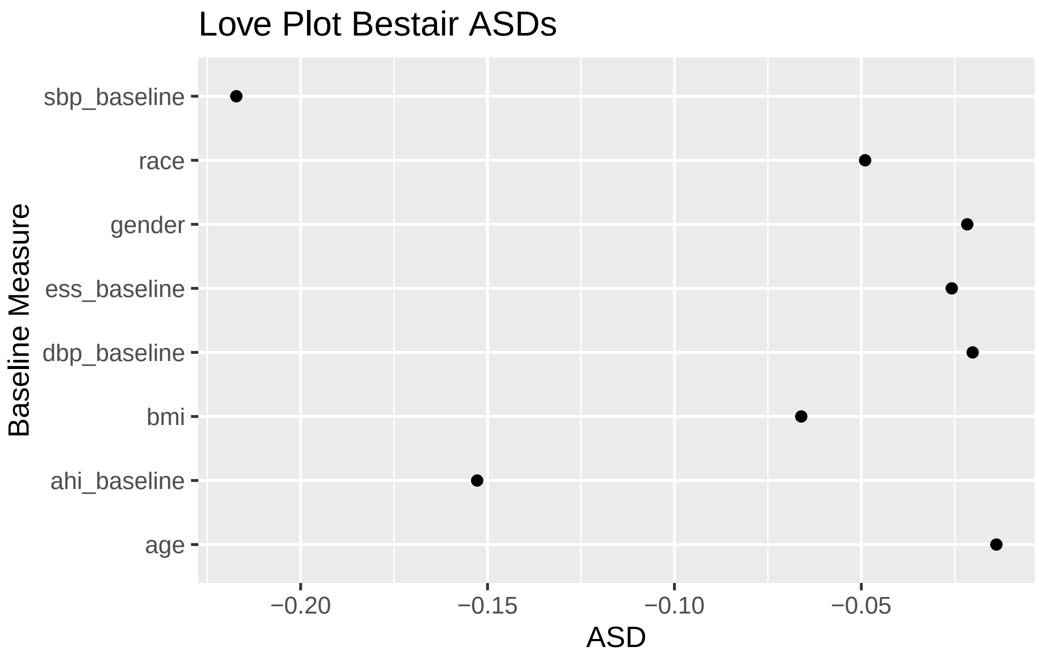
```r
bestair <- readxl::read_xlsx("bestair640-1.xlsx", sheet = "data")
for (i in seq_along(bestair)) {
  y = pull(bestair[,i])
  m = mean(y, na.rm = TRUE)
  y = ifelse(is.na(y), m, y)
  bestair[,i] = y
}
```

Question 1

```r
baselines <- c("gender","age","bmi",
 "race","sbp_baseline","dbp_baseline","ahi_baseline","ess_baseline")
ASDs = matrix(NA, nrow = 1, ncol = 8)
colnames(ASDs) <- baselines
Z <- bestair$treatment_arm
for (bl in baselines) {
  X <- pull(bestair[,bl])
  s1 <- var(X[Z==1])
  s0 <- var(X[Z==0])
  diff_sum <- sum(X*Z)/sum(Z)-sum(X*(1-Z))/sum(1-Z)
  asd <- diff_sum/sqrt(s1+s0)
  ASDs[,bl] <- asd

}
```

```
#love plot
asd_dat <- tibble(
  bls = baselines,
  asd = ASDs[1,]
)
ggplot(asd_dat,aes(x = asd, y = bls)) +
  geom_point() +
  labs(title = "Love Plot Bestair ASDs"
       ,y = "Baseline Measure"
       ,x = "ASD")
```



Question 2

```
Z <- bestair$treatment_arm
Y <- bestair$sbp_6mo
tau_unadj <- mean(Y[Z==1]) - mean(Y[Z==0])
tau_unadj
```

```
[1] -4.907177
```

```
bestair_centered <- bestair|>
  mutate(across(gender:ess_baseline, ~ .x-mean(.x)))

ancova1 <- lm(formula = sbp_6mo~.,data = bestair_centered)
tau_adj_anc1 <- ancova1$coefficients["treatment_arm"]

ancova2 <- lm(formula = sbp_6mo~.^2,data = bestair_centered)
tau_adj_anc2 <- ancova2$coefficients["treatment_arm"]
```

```
X_mat <- as.matrix(bestair_centered[,2:9])
as.vector(ancova1$residuals)
```

```
 [1]   0.72418463   5.87782418   2.02205186  -0.67970961   5.48543941
 [6]  11.12333493  22.03036046  10.40373557  16.00880076  -7.49903245
[11]  21.83336401   0.34417092  -2.77587678  -5.46146113   6.29798214
[16]   1.53501748  -1.69504331  -6.45803092 -15.68243254   8.21399912
[21] -15.30960343  -5.04025508  -7.17155008  -3.46508683  -5.34078380
[26]  -6.00645577  15.48489652  -1.73139629   4.81205921 -11.89994057
[31]  -1.35631817  -0.12539863 -10.92493221  -7.45379837   1.61531145
[36]   1.98766286 -11.96707802   4.32947706  -5.46926497   4.36435506
[41]   4.13981194 -13.31440187  -2.16750693   2.99973369  -1.26475631
[46]  -3.99403619 -16.52478518  -2.76970803   1.49838183   3.89688930
[51]   0.25809623   1.64161358   5.82966726  21.09775063  -9.38553208
[56]   9.00219896 -11.07159488  11.15178012   0.53924690 -13.11086054
[61]  11.47154172 -12.43857882   6.84561827  -5.95813641  12.80801506
[66] -13.32132353  -4.58483058 -21.71124152   4.51179466  15.55032001
[71]   6.38077123  -3.86445816  11.24286990   0.81960655  -2.86280479
[76]   2.69471525  22.96781764   5.77082371   0.11695790  -0.06314628
[81]   5.75139807  -5.22485105   1.18023905  -9.62798032   1.56231069
[86]   1.06411549   2.79881359   3.35448165   0.39576589  11.42904196
[91] -15.40328838   7.23520472   4.86064081   6.39067383  -6.70689714
```

```
 [96]  -9.12434444  11.34250922  -3.65727973   0.41435408  -9.10183514
[101] -12.54239142  -8.50610473  11.93262901   3.60166708 -13.79882297
[106]   9.35128949  -6.79121015  -0.92039034  -6.78934590   1.15694232
[111]   0.35542314   4.09807003   0.68423875   3.10956405   5.76107354
[116]  -0.34215833  12.44204229   1.56215718  -5.82690642  -5.81474801
[121] -16.27646759  -6.91003448   4.11894404  -8.39743233
```

```
hw_se_anc1 <-sqrt(car::hccm(ancova1
                             ,type = "hc2")["treatment_arm","treatment_arm"])
#sqrt(car::hccm(ancova2, type = "hc0")["treatment_arm","treatment_arm"])
```

Question 3

```
bestair_hyperten <- bestair_centered |>
  mutate(resist_hyperten = if_else(sbp_6mo>=130,1,0)) |>
  select(treatment_arm:ess_baseline,resist_hyperten)
```

a)

```
Z <- bestair$treatment_arm
Y <- bestair_hyperten$resist_hyperten
mean(Y[Z==1]) - mean(Y[Z==0])
```

```
[1] -0.2083442
```

```
bin_ols <- lm(resist_hyperten~., data = bestair_hyperten)
bin_ols$coefficients["treatment_arm"]
```

```
treatment_arm
   -0.1299632
```

```
bin_ols_inter <- lm(resist_hyperten~.^2, data = bestair_hyperten)
bin_ols_inter$coefficients["treatment_arm"]
```

```
treatment_arm
  -0.08518287
```

b)