

Assignment 1

Zahlen Zbinden

add libraries

```
library(ICSNP)
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: ICS
```

```
library(rrcov)
```

```
## Loading required package: robustbase
```

```
## Scalable Robust Estimators with High Breakdown Point (version 1.7-4)
```

Read in the sweat data:

```
file_path <- "D:/RepoMan/osu/data/SweatData.csv"
```

```
sweat <- read.table(file = file_path, sep = ",", header = T)
```

1. Read in HRBPData.csv. save this as hrbp and display the first 6 rows

```
file_path <- "D:/RepoMan/osu/data/HRBPData.csv"
```

```
hrbp <- read.table(file = file_path, sep = ",", header = T)
```

```
head(hrbp)
```

```
##   HR DiaBP SysBP
## 1 82    88   137
## 2 72    79   127
## 3 74    87   146
## 4 71    84   126
## 5 60    81   134
## 6 87    81   132
```

2. Produce all pairwise scatter plots of the variables in the hrbp data, and include the sample mean as a red solid point and the null hypothesis $\mu_0 = [70, 80, 120]^t$ as solid green point.

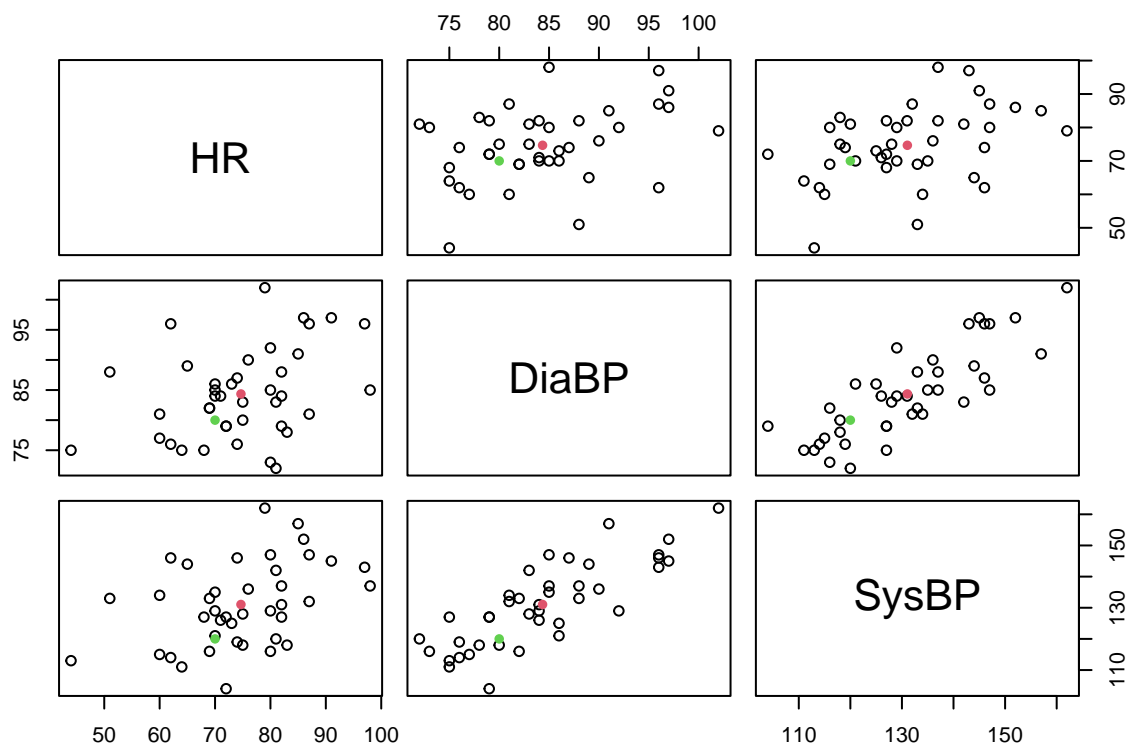
```
hrbp_mean <- colMeans(hrbp)
```

```
samp_mean <- matrix(c(70, 80, 120), nrow = 1)
```

```
colnames(samp_mean) <- c("HR", "DiaBP", "SysBP")
```

```
hrbp_aug <- rbind(hrbp, hrbp_mean, samp_mean)
```

```
pairs(
  hrbp_aug,
  col = rep(1:3, c(40, 1, 1)),
  pch = rep(c(1, 16), c(40, 3))
)
```



3. Using the hrbp data, calculate the univariate t-statistic to test that the mean vector is $\mu_0 = [70, 80, 120]^T$. What are the values of the resulting t-statistic?

First lets manually calculate all of the test statistics at the same time.

```
mu <- c(70, 80, 120)
hrbp_xbar <- apply(hrbp, 2, mean)
hrbp_var <- apply(hrbp, 2, var)
n <- nrow(hrbp)
p <- ncol(hrbp)
ts <- (hrbp_xbar - mu) / sqrt(hrbp_var / n)
ts
```

```
##      HR      DiaBP      SysBP
## 2.629580 3.700989 5.187670
```

To confirm our manual calculations lets compare the t stat for HR from the manual calcuations to the t stat from the function. If these align we will make the assumption that all of the manually calculated t stats are correct. As we can see they align, we got a t stat of 2.6296 from the manual calculations and a t stat of 2.6296 from the t.test function.

```
t.test(hrbp[, 1], mu = mu[1])
```

```
##
## One Sample t-test
##
## data:  hrbp[, 1]
## t = 2.6296, df = 39, p-value = 0.01217
## alternative hypothesis: true mean is not equal to 70
```

```
## 95 percent confidence interval:
## 71.07896 78.27104
## sample estimates:
## mean of x
## 74.675
```

4. Calculate the Bonferroni corrected p-values corresponding to the simultaneous univariate t-tests from the previous part.

```
alpha <- .05
alpha_star <- .05 / p
crit_t <- qt(alpha_star / 2, df = n - 1, lower.tail = F)
crit_t
```

```
## [1] 2.501658
```

```
any(abs(ts) > crit_t)
```

```
## [1] TRUE
```

5. Based on the simultaneous univariate t-tests, would you reject the hypothesis that the population mean vector is $\mu_0 = [70, 80, 120]^T$? Yes I would reject the null hypothesis as at least one of the test statistics was larger than the critical value used with the Bonferroni method.
6. Using the hrbp data, perform Hotelling's T^2 test by hand, with $H_0 : \mu_0 = [70, 80, 120]^T$.
 - a. What is the value of the scaled test statistic?
 - i. 8.89
 - b. What is the resulting p-value?
 - i. .000144
 - c. What is the decision: would you reject or fail to reject the null hypothesis?
 - i. Since the p-value is less than $\alpha = .05$ we reject the null hypothesis

```
n <- nrow(hrbp)
p <- ncol(hrbp)
hrbp_means <- colMeans(hrbp)
mu <- c(70, 80, 120)
s_cov <- cov(hrbp)

T <- n * t((hrbp_means - mu)) %*% solve(s_cov) %*% (hrbp_means - mu)
T_scale <- ((n - p) / ((n - 1) * p)) * T
T
```

```
## [1]
```

```
## [1,] 28.11156
```

```
T_scale
```

```
## [1]
```

```
## [1,] 8.889982
```

```
p <- 1 - pf(T_scale, df1 = p, df2 = n - p)
```

```
p
```

```
## [1]
```

```
## [1,] 0.0001443427
```

7. Using the hrbp data, perform Hotelling's T^2 test using the `T2.test()` function to test the null hypothesis $\mu_0 = [70, 80, 120]^T$. Do the results match what you got by hand?

- a. The results from this function mirror what I got by hand, both conclude that we reject the null hypothesis in favor of the alternative.

```
T2.test(hrbp, mu = mu)
```

```
##
## One-sample Hotelling test
##
## data:  hrbp
## T2 = 28.112, F = 8.890, df1 = 3, df2 = 37, p-value = 0.0001443
## alternative hypothesis: true mean vector is not equal to (70, 80, 120)'
##
## sample estimates:
##              HR DiaBP  SysBP
## mean x-vector 74.675 84.35 131.05
```

8. Read in HRBPPairedData.csv Save this data as hrbpPaired and display the first six rows.

```
hrbpPaired <- read.csv("D:/RepoMan/osu/data/SweatPairedData.csv")
head(hrbpPaired)
```

```
## Rate.Before Sodium.Before Potassium.Before Rate.After Sodium.After
## 1      3.7      48.5      9.3      2.9      37.5
## 2      5.7      65.1      8.0      7.3      64.6
## 3      3.8      47.2     10.9      5.2      51.7
## 4      3.2      53.2     12.0      4.6      52.1
## 5      3.1      55.5      9.7      5.2      61.6
## 6      4.6      36.1      7.9      3.8      32.1
## Potassium.After
## 1      8.1
## 2      9.1
## 3     14.5
## 4     13.4
## 5      9.9
## 6      7.1
```

9. Compute the difference vectors, and use these differences to perform simultaneous univariate tests of the null hypothesis that mean heart rate, mean diastolic blood pressure, and mean systolic blood pressure are the same before and after walking a mile.
- What are the values of the univariate t-statistic?
 - (1.291, -1.216, 1.081)
 - What are the values of the Bonferroni corrected univariate p-values?
 - (.6366, .7164, .8799)
 - Based on these results, what would your decision be at level .05?
 - Based on the Bonferroni corrected p-values we would fail to reject the null hypothesis that heart rates are the same before and after walking a mile

```
hrbp_difs <- hrbpPaired[, 4:6] - hrbpPaired[, 1:3]
n <- nrow(hrbp_difs)
p <- ncol(hrbp_difs)
mu <- rep(0, 3)
hrbp_difs_mean <- apply(hrbp_difs, 2, mean)
hrbp_difs_var <- apply(hrbp_difs, 2, var)

tstats <- sqrt(n) * (hrbp_difs_mean - mu) / sqrt(hrbp_difs_var)
pvals <- 2 * (1 - pt(abs(tstats), n - 1))
pvals_bonf <- pvals * p
```

```
tstats

##      Rate.After      Sodium.After Potassium.After
##      1.290994      -1.216218      1.080809
```

```
pvals

##      Rate.After      Sodium.After Potassium.After
##      0.2121933      0.2387961      0.2933104
```

```
pvals_bonf

##      Rate.After      Sodium.After Potassium.After
##      0.6365798      0.7163882      0.8799312
```

10. Perform Hotelling's T^2 test on the paired data.
 - a. What is the value of the scaled test statistic?
 - i. 5.2754
 - b. What is the resulting p-value?
 - i. .009353
 - c. What is the decision: would you reject or fail to reject the null hypothesis?
 - i. Based on the p-value which is less than .05 we reject the null hypothesis in favor of the alternative.
 - d. How does this compare to the result you got from the simultaneous univariate tests?
 - i. In the simultaneous univariate tests we failed to reject the null hypothesis, and in the multivariate analysis we rejected the null hypothesis.

```
T2.test(hrbp_difs, mu = mu)
```

```
##
## One-sample Hotelling test
##
## data:  hrbp_difs
## T2 = 17.6881, F = 5.2754, df1 = 3, df2 = 17, p-value = 0.009353
## alternative hypothesis: true mean vector is not equal to (0, 0, 0)'
##
## sample estimates:
##              Rate.After Sodium.After Potassium.After
## mean x-vector      0.35      -1.655      0.425
```

11. Read in HRRepMeasData.csv save this data as hrRepMeas and display the first six rows.

```
hrRepMeas <- read.csv("D:/RepoMan/osu/data/HRRepMeasData.csv")
head(hrRepMeas)
```

```
##   GBBS Friends SNL News WorldSeries
## 1   89      78 66   74           86
## 2   75      67 69   74           86
## 3   61      65 69   69           77
## 4   74      74 62   66           82
## 5   59      71 68   66           84
## 6   67      65 78   74           78
```

12. Construct a contrast matrix and compute the contrast observations corresponding to that contrast matrix. Display the first six rows of your contrast observations

```
contrast <- rbind(
  c(-1, 1, 0, 0, 0),
  c(-1, 0, 1, 0, 0),
```

```

      c(-1, 0, 0, 1, 0),
      c(-1, 0, 0, 0, 1)
    )
hrRepMeas_means <- apply(hrRepMeas, 2, mean)

contrast_obj <- as.matrix(hrRepMeas) %*% t(contrast)
head(contrast_obj)

```

```

##      [,1] [,2] [,3] [,4]
## [1,] -11 -23 -15 -3
## [2,] -8  -6  -1  11
## [3,]  4   8   8  16
## [4,]  0 -12  -8   8
## [5,] 12   9   7  25
## [6,] -2  11   7  11

```

13. Perform Hotelling's T^2 test on these contrast observations to test that the average heart rate is the same as people watch each of these five TV shows.
- What is the value of the scaled test statistic?
 - 1.4395
 - What is the resulting p-value?
 - .241
 - What is the decision: would you reject or fail to reject the hypothesis?
 - As the p-value is greater than .05 i would fail to reject the null hypothesis.

```

p <- ncol(contrast_obj)
T2.test(contrast_obj, mu = rep(0, p))

```

```

##
## One-sample Hotelling test
##
## data: contrast_obj
## T2 = 6.2380, F = 1.4395, df1 = 4, df2 = 36, p-value = 0.241
## alternative hypothesis: true mean vector is not equal to (0, 0, 0, 0)'
##
## sample estimates:
##      [,1] [,2] [,3] [,4]
## mean x-vector -0.525 -0.975  2.2  1.475

```

14. Construct a different contrast matrix, and again compute the contrast observations and again perform Hotelling's T^2 test on these contrast observations to test that the average heart rate is the same as people watch each of these five tv shows. Do you get the same results as in the previous part?
- The same outcome is achieved with the different contrast matrix, only the mean x-vector has changed, and it is just multiplied by -1.

```

contrast <- rbind(
  c(1, -1, 0, 0, 0),
  c(1, 0, -1, 0, 0),
  c(1, 0, 0, -1, 0),
  c(1, 0, 0, 0, -1)
)
hrRepMeas_means <- apply(hrRepMeas, 2, mean)

contrast_obj <- as.matrix(hrRepMeas) %*% t(contrast)
head(contrast_obj)

```

```
##      [,1] [,2] [,3] [,4]
## [1,]   11   23   15    3
## [2,]    8    6    1  -11
## [3,]   -4   -8   -8  -16
## [4,]    0   12    8   -8
## [5,]  -12   -9   -7  -25
## [6,]    2  -11   -7  -11
```

```
p <- ncol(contrast_obj)
T2.test(contrast_obj, mu = rep(0, p))
```

```
##
## One-sample Hotelling test
##
## data: contrast_obj
## T2 = 6.2380, F = 1.4395, df1 = 4, df2 = 36, p-value = 0.241
## alternative hypothesis: true mean vector is not equal to (0, 0, 0, 0)'
##
## sample estimates:
##      [,1] [,2] [,3] [,4]
## mean x-vector 0.525 0.975 -2.2 -1.475
```