

Homework 05
Joseph Blubaugh
jblubau1@tamu.edu
STAT 636-720

1)

a)

```
library(DescTools)
library(Hotelling)

effluent = read.table("T6-1.DAT", quote="\\"", comment.char="")
colnames(effluent) = c("Comm_BOD", "Comm_SS", "State_BOD", "State_SS")

## Grp 1 -> Commercial
## Grp 2 -> State

effluent2 = data.frame(
  BOD = c(effluent$Comm_BOD, effluent$State_BOD),
  SS = c(effluent$Comm_SS, effluent$State_SS),
  GRP = c(rep(1, 11), rep(2, 11))
)

X1_bar = colMeans(effluent2[1:11, 1:2])
X2_bar = colMeans(effluent2[12:22, 1:2])
Var1 = var(effluent2[1:11, 1:2])
Var2 = var(effluent2[12:22, 1:2])
n1 = n2 = 11

## Manually calculating values
S.pooled = ((n1 - 1) * Var1 + (n2 - 1) * Var2) / (n1 + n2 - 2)
T2 = (X1_bar - X2_bar) %*% solve(S.pooled * (1 / n1 + 1 / n2)) %*% (X1_bar - X2_bar)
P.value = 1 - pf((n1 + n2 - 2 - 1) / ((n1 + n2 - 2) * 2) * T2, 2, n1 + n2 - 2 - 1)
Critical.Value = qf(0.95, 2, n1 + n2 - 2 - 1)

## Manually Calculated Values
data.frame(T2, P.value, Critical.Value)

      T2      P.value Critical.Value
1 12.6648 0.009462901      3.521893

## Matching the T2 Statistic from the Tests
T2 * (n1 + n2 - 2 - 1) / ((n1 + n2 - 2) * 2)

      [,1]
[1,] 6.015782

(mdl = hotelling.test(BOD + SS ~ GRP, data = effluent2, pair = c(1,2)))
```

```
Test stat: 6.0158
Numerator df: 2
Denominator df: 19
P-value: 0.009463
```

Based on the Hotellings T Test we can conclude that the difference in means are statistically different so we would reject the null hypothesis.

b)

```
library(plotrix)

## Samples
n = n1 = n2 = 11

## Means
(mu.Comm = colMeans(effluent)[1:2]); (mu.State = colMeans(effluent)[3:4])

Comm_BOD  Comm_SS
25.27273  46.45455

State_BOD  State_SS
34.63636   33.18182

## Covariances
(S.Comm = cov(effluent[1:2])); (S.State = cov(effluent[3:4]))

      Comm_BOD  Comm_SS
Comm_BOD 387.4182  489.3636
Comm_SS  489.3636 1014.0727

      State_BOD State_SS
State_BOD 109.2545 120.3727
State_SS  120.3727 363.7636

## Pooled Variance
S.po = ((n1 - 1) * S.Comm + (n2 - 1) * S.State) / (n1 + n2 - 2)

## Eigen Values
lambda = eigen(S.po)$values; ee = eigen(S.po)$vectors

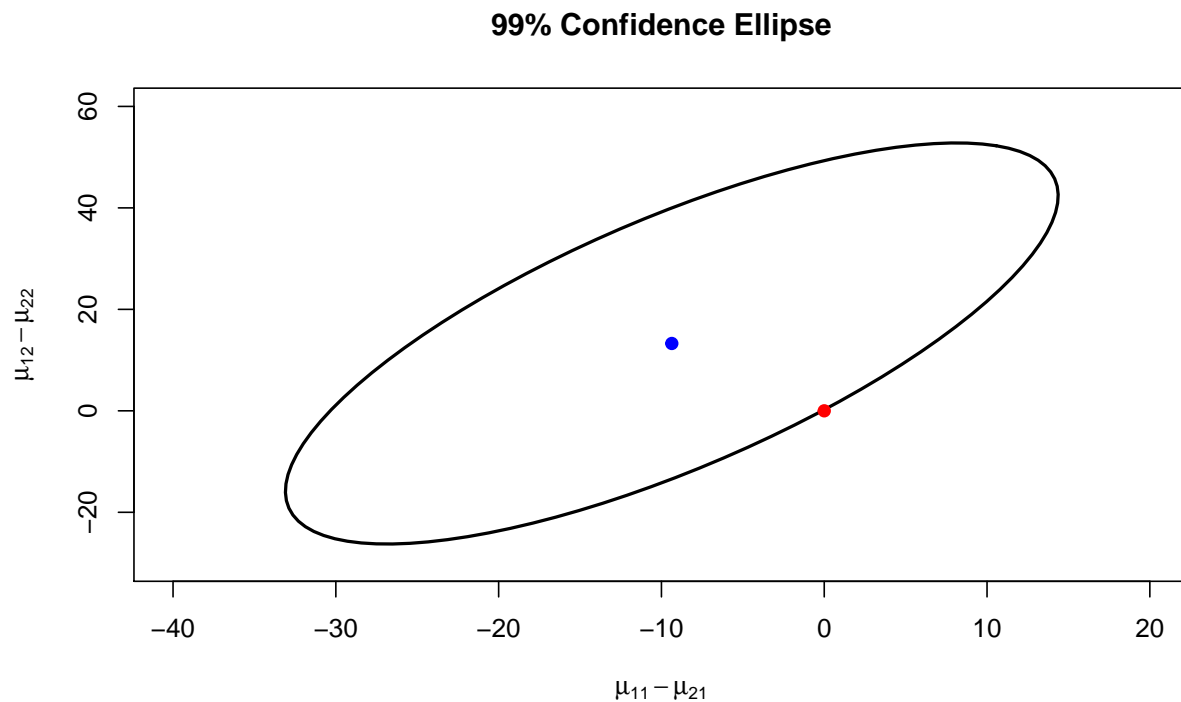
## 95% confidence ellipse for mu.Comm - mu.State
theta = atan(ee[2, 1] / ee[1, 1]) * 57.2957795
```

```
## c2
c2 = (n1 + n2 - 2) * 2 / (n1 + n2 - 2 - 1) * qf(.99, 2, n1 + n2 - 2 - 1)

## Plot
plot(c(-40, 20), c(-30, 60), type = "n", main = "99% Confidence Ellipse",
      xlab = expression(mu[11] - mu[21]),
      ylab = expression(mu[12] - mu[22]))

draw.ellipse(mu.Comm[1] - mu.State[1], mu.Comm[2] - mu.State[2],
             sqrt(c2 * lambda[1] * (1/n1 + 1/n2)),
             sqrt(c2 * lambda[2] * (1/n1 + 1/n2)),
             angle = theta, lwd = 2)

points(mu.Comm[1] - mu.State[1], mu.Comm[2] - mu.State[2], pch = 19, col = "blue")
points(0, 0, col = "red", pch = 19)
```



Our Critical value of 12.4755348 is less than the statistical distance to the origin 12.6648. This would indicate that (0,0) sits just outside the 99% confidence perimeter which supports the results from the previous test that the difference in the means are not equal to 0

c) The Bonferroni simultaneous intervals are shorter than the lengths of the T2 simultaneous intervals

```
diff = data.frame(d1 = effluent$Comm_BOD - effluent$State_BOD,
                  d2 = effluent$Comm_SS - effluent$State_SS)

## d1 99% T2 confidence interval
colMeans(diff)[1] + c(-1, 1) * sqrt((n - 1)*2/(n - 2) * qf(1 - .01, 2, n-2)) *
  sqrt(cov(diff)[1,1] / n)

[1] -27.332879  8.605607

## d2 99% T2 confidence interval
colMeans(diff)[2] + c(-1, 1) * sqrt((n - 1)*2/(n - 2) * qf(1 - .01, 2, n-2)) *
  sqrt(cov(diff)[2,2] / n)

[1] -12.77289  39.31835

## d1 99% Bonferroni confidence interval
colMeans(diff)[1] + c(-1, 1) * qt(1 - .01/(2*2), df = n-1) * sqrt(cov(diff)[1,1]/n)

[1] -24.606317  5.879045

## d2 99% Bonferroni confidence interval
colMeans(diff)[2] + c(-1, 1) * qt(1 - .01/(2*2), df = n-1) * sqrt(cov(diff)[2,2]/n)

[1] -8.820862 35.366317
```

d) With the outlier remove our pvalue is greater than our critical value, we fail to reject that $d = 0$.

```
effluent3 = effluent2[-c(8, 19), ]

X1_bar = colMeans(effluent3[1:10, 1:2])
X2_bar = colMeans(effluent3[11:20, 1:2])
Var1 = var(effluent3[1:10, 1:2])
Var2 = var(effluent3[11:20, 1:2])
n1 = n2 = 10

S.pooled = ((n1 - 1) * Var1 + (n2 - 1) * Var2) / (n1 + n2 - 2)
T2 = (X1_bar - X2_bar) %*% solve(S.pooled * (1 / n1 + 1 / n2)) %*% (X1_bar - X2_bar)
(P.value = 1 - pf((n1 + n2 - 2 - 1) / ((n1 + n2 - 2) * 2) * T2, 2, n1 + n2 - 2 - 1))

[,1]
[1,] 0.01860841
```

2)

a)

```
## i)
trt = data.frame(
  Treatment = c(1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3),
  X.1 = c(6, 5, 8, 4, 7, 3, 1, 2, 2, 5, 3, 2),
  X.2 = c(7, 9, 6, 9, 9, 3, 6, 3, 3, 1, 1, 3)
)

n2 = 3; n3 = 4

## Sample Means
TRT2_bar = colMeans(trt[6:8, 2:3]); TRT3_bar = colMeans(trt[9:12, 2:3])

## Sample Variance
Var2 = var(trt[6:8, 2:3]); Var3 = var(trt[9:12, 2:3])

## Pooled Variance
S.pooled = ((n2 - 1) * Var2 + (n3 - 1) * Var3) / (n2 + n3 - 2)

## Test Statistic
T2 = t(TRT2_bar - TRT3_bar) %>% solve((1/n2 + 1/n3) * S.pooled) %>%
  (TRT2_bar - TRT3_bar)
(Tst = (T2 * (n2 + n3 - 2 - 1)/((n2 + n3 - 2)*2)))

      [,1]
[1,] 1.548387

## PValue
1 - pf(Tst, 2, n2 + n3 - 2 - 1)

      [,1]
[1,] 0.317686

(md1 = hotelling.test(X.1 + X.2 ~ Treatment, data = trt[6:12, ]))

Test stat: 1.5484
Numerator df: 2
Denominator df: 4
P-value: 0.3177
```

```

## ii)
## Bonferroini Interval
## Difference in X.1
(TRT2_bar[1] - TRT3_bar[1]) + c(-1, 1) * qt(1 - .01/(2*2), df = n2 + n3 - 1) *
  sqrt(((1/n2) + 1/n3) * S.pooled[1,1])

[1] -5.170451  3.170451

## Difference in X.2
(TRT2_bar[2] - TRT3_bar[2]) + c(-1, 1) * qt(1 - .01/(2*2), df = n2 + n3 - 1) *
  sqrt(((1/n2) + 1/n3) * S.pooled[2,2])

[1] -2.662706  6.662706

b)

n1 = 5; n2 = 3; n3 = 4

## Sum of Squares Response 1
trt1.X1 = sum(n1 * (mean(trt[1:5, 2]) - mean(trt[, 2]))^2)
trt2.X1 = sum(n2 * (mean(trt[6:8, 2]) - mean(trt[, 2]))^2)
trt3.X1 = sum(n3 * (mean(trt[9:12, 2]) - mean(trt[, 2]))^2)
SS.R1 = sum(trt1.X1, trt2.X1, trt3.X1)

## Sum of Squares Response 2
trt1.X2 = sum(n1 * (mean(trt[1:5, 3]) - mean(trt[, 3]))^2)
trt2.X2 = sum(n2 * (mean(trt[6:8, 3]) - mean(trt[, 3]))^2)
trt3.X2 = sum(n3 * (mean(trt[9:12, 3]) - mean(trt[, 3]))^2)
SS.R2 = sum(trt1.X2, trt2.X2, trt3.X2)

## Residual SS Response 1
trt1.X1.RS = sum((trt[1:5, 2] - mean(trt[1:5, 2]))^2)
trt2.X1.RS = sum((trt[6:8, 2] - mean(trt[6:8, 2]))^2)
trt3.X1.RS = sum((trt[9:12, 2] - mean(trt[9:12, 2]))^2)
RS.R1 = sum(trt1.X1.RS, trt2.X1.RS, trt3.X1.RS)

## Residual SS Response 2
trt1.X2.RS = sum((trt[1:5, 3] - mean(trt[1:5, 3]))^2)
trt2.X2.RS = sum((trt[6:8, 3] - mean(trt[6:8, 3]))^2)
trt3.X2.RS = sum((trt[9:12, 3] - mean(trt[9:12, 3]))^2)
RS.R2 = sum(trt1.X2.RS, trt2.X2.RS, trt3.X2.RS)

## Compile one way MANOVA
MANOVA = matrix(c(SS.R1, SS.R2, 3-1, RS.R1, RS.R2, 4 + 2 + 3), nrow = 3)

```

```

row.names(MANOVA) = c("Response.1", "Response.2", "Deg of F")
colnames(MANOVA) = c("Treatment", "Residual")

MANOVA; (mdl = manova(cbind(X.1, X.2) ~ factor(Treatment), data = trt))

      Treatment Residual
Response.1      36      18
Response.2      84      18
Deg of F         2       9

Call:
manova(cbind(X.1, X.2) ~ factor(Treatment), data = trt)

Terms:
      factor(Treatment) Residuals
resp 1                36      18
resp 2                84      18
Deg. of Freedom         2       9

Residual standard errors: 1.414214 1.414214
Estimated effects may be unbalanced

## Wilks Lambda
B = matrix(c(SS.R1, 48, 48, SS.R2), nrow = 2)
W = matrix(c(RS.R1, -13, -13, RS.R2), nrow = 2)
(WL = det(W) / det(B + W))

[1] 0.03618959

## F Statistic and PValue
## p = 2, g = 3
(12 - 3 - 1)/(3 - 1) * ( (1 - sqrt(WL))/sqrt(WL)); 1 - pf(17.02644, 2, 8)

[1] 17.02656

[1] 0.001309716

summary(mdl, test = "Wilks")

      Df    Wilks approx F num Df den Df    Pr(>F)
factor(Treatment)  2 0.03619   17.027     4    16 1.283e-05 ***
Residuals          9

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```


3)

$$\begin{aligned}
L(\mu_1, \mu_2, \Sigma) &= \prod_{i=1}^n \left[\frac{1}{(2\pi)^p/2|\Sigma|^{1/2}} e^{-(x_i - \mu)' \Sigma^{-1} (x_i - \mu)/2} \right] \\
&= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-tr[\Sigma^{-1} (\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' + \frac{1}{n_1 + n_2} (\bar{x} - \mu)(\bar{x} - \mu)')] } \\
&= \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{tr \left[\Sigma^{-1} \left(\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' \right) \right] + \frac{1}{n_1 + n_2} (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu)} \\
&= \frac{1}{(2\pi)^n} e^{-n} \frac{1}{|\hat{\Sigma}_{pooled}|^{.5}}
\end{aligned}$$

4)

```

peanut = read.table("T6-17.DAT", quote="\\"", comment.char="")
colnames(peanut) = c("Location", "Variety", "Yield", "SMK", "Size")
attach(peanut)

peanut$Location = as.factor(peanut$Location);
peanut$Variety = as.factor(peanut$Variety)

n = 2; p = 3; g = 2; b = 3

## Summary statistics.
x_bar = colMeans(peanut[, 3:5])
x_bar_lk = rbind(
  colMeans(peanut[Location == 1 & Variety == 5, 3:5]),
  colMeans(peanut[Location == 2 & Variety == 5, 3:5]),
  colMeans(peanut[Location == 1 & Variety == 6, 3:5]),
  colMeans(peanut[Location == 2 & Variety == 6, 3:5]),
  colMeans(peanut[Location == 1 & Variety == 8, 3:5]),
  colMeans(peanut[Location == 2 & Variety == 8, 3:5])
)

x_bar_l_dot = rbind(
  colMeans(peanut[Location == 1, 3:5]),
  colMeans(peanut[Location == 2, 3:5])
)

x_bar_dot_k = rbind(
  colMeans(peanut[Variety == 5, 3:5]),

```

```

colMeans(peanut[Variety == 6, 3:5]),
colMeans(peanut[Variety == 8, 3:5])
)

## Components for MANOVA.
SSP_cor = SSP_Location = SSP_Variety = SSP_int = SSP_res =
  matrix(0, nrow = p, ncol = p)

for(l in 1:g) {
  SSP_Location = SSP_Location + b * n *
    t(x_bar_l_dot[l, , drop = FALSE] - x_bar) %*%
    (x_bar_l_dot[l, , drop = FALSE] - x_bar)
  SSP_Variety = SSP_Variety + g * n *
    t(x_bar_dot_k[l, , drop = FALSE] - x_bar) %*%
    (x_bar_dot_k[l, , drop = FALSE] - x_bar)

  for(k in 1:b) {
    SSP_int = SSP_int + n *
      t(x_bar_lk[(l - 1) * 2 + k, , drop = FALSE] -
        x_bar_l_dot[l, , drop = FALSE] -
        x_bar_dot_k[k, , drop = FALSE] + x_bar) %*%
      (x_bar_lk[(l - 1) * 2 + k, , drop = FALSE] -
        x_bar_l_dot[l, , drop = FALSE] -
        x_bar_dot_k[k, , drop = FALSE] + x_bar)

    for(r in 1:n) {
      SSP_res = SSP_res +
        t(as.matrix(peanut[(l - 1) * 2 * n + (k - 1) * n + r, 3:5]) -
          x_bar_lk[(l - 1) * 2 + k, , drop = FALSE]) %*%
        (as.matrix(peanut[(l - 1) * 2 * n + (k - 1) * n + r, 3:5]) -
          x_bar_lk[(l - 1) * 2 + k, , drop = FALSE])

      SSP_cor = SSP_cor +
        t(as.matrix(peanut[(l - 1) * 2 * n + (k - 1) * n + r, 3:5]) - x_bar) %*%
        (as.matrix(peanut[(l - 1) * 2 * n + (k - 1) * n + r, 3:5]) - x_bar)
    }
  }
}

SS = cbind(diag(SSP_Location), diag(SSP_Variety), diag(SSP_int), diag(SSP_res))
SS = rbind(SS, matrix(c(1, 1, 1, 8), nrow = 1))
row.names(SS)[4] = 'DF'
colnames(SS) = c("Location", "Variety", "Interaction", "Residuals")

```

```
## Location Effect
Lambda = det(SSP_res) / det(SSP_Location + SSP_res)
1 - pf((((g * b * (n - 1) - p + 1) / 2) / ((abs((g - 1) - p) + 1) / 2)) *
  (1 - Lambda) / Lambda, abs((g - 1) - p) + 1, g * b * (n - 1) - p + 1)

[1] 0.02514749

## Variety Effect
Lambda = det(SSP_res) / det(SSP_Variety + SSP_res)
1 - pf((((g * b * (n - 1) - p + 1) / 2) / ((abs((b - 1) - p) + 1) / 2)) *
  (1 - Lambda) / Lambda, abs((b - 1) - p) + 1, g * b * (n - 1) - p + 1)

[1] 0.000683825

## Interaction Effect
Lambda = det(SSP_res) / det(SSP_int + SSP_res)
1 - pf(
  (((g * b * (n - 1) - p + 1) / 2) / ((abs((g - 1) * (b - 1) - p) + 1) / 2)) *
  (1 - Lambda) / Lambda, abs((g - 1) * (b - 1) - p) + 1, g * b * (n - 1) - p + 1)

[1] 8.054312e-07

SS

      Location  Variety Interaction Residuals
Yield  0.7008333 195.39250    741.1850    128.285
SMK    162.0675000 763.21250   3879.9150    369.925
Size   72.5208333  96.86806    473.3267    100.415
DF      1.0000000   1.00000     1.0000     8.000

summary(manova(cbind(Yield, SMK, Size) ~ Location*Variety), test = "Wilks")

      Df    Wilks approx F num Df den Df  Pr(>F)
Location    1 0.41236    2.8502     3     6 0.12727
Variety      1 0.33915    3.8971     3     6 0.07362 .
Location:Variety 1 0.54380    1.6778     3     6 0.26968
Residuals      8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Start by looking at the interaction between Location and Variety. Based on an alpha of .05 I would conclude that there is not a significant effect between Variety and Location. Location is also no significant as a main effect, however Variety is a significant main effect.