

STAT 636, Fall 2015 - Assignment 5
SOLUTIONS

1. For the (paired) effluent data in Table 6.1 from the textbook, let $\delta' = [\delta_1, \delta_2]$, where δ_1 is the mean difference between the commercial lab BOD measurements and the state lab BOD measurements, and δ_2 is the mean difference between the commercial lab SS measurements and the state lab SS measurements.

- (a) Test $H_0 : \delta = \mathbf{0}$ using the `HotellingsT2Test` function from the `DescTools` R package. Provide R code that matches the test statistic and p-value returned by `HotellingsT2Test`. What is your conclusion if testing at $\alpha = 0.01$?

SEE CODE BELOW. THE T^2 STATISTIC EQUALS 13.6393, CORRESPONDING TO A P-VALUE OF 0.0208. WE FAIL TO REJECT H_0 AT $\alpha = 0.01$.

- (b) Construct and plot a 99% confidence region for δ . Is $\delta = \mathbf{0}$ inside this region? Is this consistent with the results of part (a)?

SEE FIGURE 1. THE REGION DOES INCLUDE $\delta = \mathbf{0}$, AGREEING WITH THE T^2 TEST, AS EXPECTED.

- (c) Construct 99% simultaneous confidence intervals for δ_1 , δ_2 , and $\delta_1 - \delta_2$, using both the T^2 and Bonferroni methods. Compare the lengths of the T^2 intervals to those of the Bonferroni intervals.

THE INTERVALS ARE SHOWN IN THE TABLE BELOW. AS USUAL, THE BONFERRONI INTERVALS ARE NARROWER THAN THEIR T^2 COUNTERPARTS.

	T^2	BONFERRONI
δ_1	$[-27.33, 8.61]$	$[-25.65, 6.93]$
δ_2	$[-12.77, 39.32]$	$[-10.34, 36.88]$
$\delta_1 - \delta_2$	$[-49.38, 4.10]$	$[-46.88, 1.60]$

- (d) Sample 8 could arguably be considered an outlier. What is the p-value for testing $H_0 : \delta = \mathbf{0}$ when this sample is excluded?

EXCLUDING SAMPLE 8, THE T^2 P-VALUE IS 0.0375, AND OUR CONCLUSION REMAINS UNCHANGED.

2. For the data in Exercise 6.8 from the textbook ($p = 2$, $g = 3$, $n_1 = 5$, $n_2 = 3$, $n_3 = 4$):

- (a) Consider just treatments 2 and 3, and assume that the covariance matrix is the same for both treatments.

- i. Test $H_0 : \mu_2 - \mu_3 = \mathbf{0}$ with a two-sample T^2 statistic, at $\alpha = 0.01$. Do this using the `hotelling.test` function from the `Hotelling` R package, and provide R code that matches its output (statistic and p-value).

SEE CODE BELOW. THE P-VALUE IS 0.3177, SO WE FAIL TO REJECT H_0 .

- ii. Construct 99% Bonferroni simultaneous confidence intervals for the differences $\mu_{2i} - \mu_{3i}$, $i = 1, 2$.

THE CONFIDENCE INTERVALS ARE $[-5.61, 3.61]$ AND $[-3.16, 7.16]$ FOR TREATMENTS 2 AND 3, RESPECTIVELY.

(b) Now consider all three treatments.

i. Construct the one-way MANOVA table.

WE HAVE

$$\mathbf{B} = \begin{bmatrix} 36 & 48 \\ 48 & 84 \end{bmatrix}$$

WITH 2 DEGREES OF FREEDOM,

$$\mathbf{W} = \begin{bmatrix} 18 & -13 \\ -13 & 18 \end{bmatrix}$$

WITH 9 DEGREES OF FREEDOM, AND

$$\mathbf{B} + \mathbf{W} = \begin{bmatrix} 54 & 35 \\ 35 & 102 \end{bmatrix}$$

WITH 11 DEGREES OF FREEDOM.

ii. Use the Wilks' lambda statistic to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ at $\alpha = 0.01$. Do this using the `manova` function in R, and provide R code that matches its output (statistic and p-value). Note that to get the Wilks' statistic and p-value from `manova`, you can do the following (you will need to fill in the "..."):

```
summary(manova(...), test = 'Wilks')
```

THE VALUE OF WILK'S LAMBDA IS 0.0362, CORRESPONDING TO A P-VALUE OF 0.000013. WE REJECT H_0 AT $\alpha = 0.01$.

3. The use of the pooled sample covariance matrix in the two-sample T^2 test can be motivated by a likelihood argument. Give the likelihood function, $L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, for two independent samples of sizes n_1 and n_2 from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ populations, respectively. Show that this likelihood is maximized by the choices $\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{x}}_1$, $\hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{x}}_2$, and

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n_1 + n_2} [(n_1 - 1) \mathbf{S}_1 + (n_2 - 1) \mathbf{S}_2] = \left(\frac{n_1 + n_2 - 2}{n_1 + n_2} \right) \mathbf{S}_{\text{pooled}}$$

Hint: Refer to slides 34-40 in the Topic 4 notes and / or Section 4.3 of the textbook.

THE LIKELIHOOD IS

$$\begin{aligned} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= \left\{ \frac{1}{(2\pi)^{n_1 p/2} |\boldsymbol{\Sigma}|^{n_1/2}} e^{-\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{1j} - \boldsymbol{\mu}_1)/2} \right\} \times \\ &\quad \left\{ \frac{1}{\{2\pi\}^{n_2 p/2} |\boldsymbol{\Sigma}|^{n_2/2}} e^{-\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{2j} - \boldsymbol{\mu}_2)/2} \right\} \\ &\propto \left\{ \frac{1}{|\boldsymbol{\Sigma}|^{n_1/2}} e^{-\frac{1}{2} \text{TR}[\boldsymbol{\Sigma}^{-1} (\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)')] - n_1 (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1)} \right\} \times \\ &\quad \left\{ \frac{1}{|\boldsymbol{\Sigma}|^{n_2/2}} e^{-\frac{1}{2} \text{TR}[\boldsymbol{\Sigma}^{-1} (\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)')] - n_2 (\bar{\mathbf{x}}_2 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_2 - \boldsymbol{\mu}_2)} \right\} \end{aligned}$$

WHERE THE SECOND EQUALITY IS AN APPLICATION OF THE RESULTS ON SLIDE 37 OF THE TOPIC 4 SLIDES. NOW WE FOLLOW THE REASONING FROM SLIDE 39 OF THE TOPIC 4 SLIDES. TO MAXIMIZE THE LIKELIHOOD IN TERMS OF $\boldsymbol{\mu}_1$ AND $\boldsymbol{\mu}_2$, WE NEED TO MINIMIZE

$$n_1 (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1)$$

AND

$$n_2 (\bar{\mathbf{x}}_2 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_2 - \boldsymbol{\mu}_2)$$

SINCE $\boldsymbol{\Sigma}$ IS POSITIVE DEFINITE, $(\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_l - \boldsymbol{\mu}_l) > 0$ UNLESS $\boldsymbol{\mu}_l = \bar{\mathbf{x}}_l$, SO $\hat{\boldsymbol{\mu}}_l = \bar{\mathbf{x}}_l$, $l = 1, 2$. WE NOW NEED TO CHOOSE $\boldsymbol{\Sigma}$ TO MAXIMIZE

$$\begin{aligned} L(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \boldsymbol{\Sigma}) &= \frac{1}{|\boldsymbol{\Sigma}|^{(n_1+n_2)/2}} e^{-\frac{1}{2} \{ \text{TR}[\boldsymbol{\Sigma}^{-1} (\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)')] + \text{TR}[\boldsymbol{\Sigma}^{-1} (\sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)')] \}} \\ &= \frac{1}{|\boldsymbol{\Sigma}|^{(n_1+n_2)/2}} e^{-\frac{1}{2} \text{TR}[\boldsymbol{\Sigma}^{-1} (\sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)' + \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)')] } \end{aligned}$$

NOW WE APPLY THE FOLLOWING MAXIMIZATION RESULT: FOR A $p \times p$ POSITIVE DEFINITE MATRIX \mathbf{B} AND A SCALAR $b > 0$,

$$\frac{1}{|\boldsymbol{\Sigma}|^b} e^{-\text{TR}(\boldsymbol{\Sigma}^{-1} \mathbf{B})/2} \leq \frac{1}{|\mathbf{B}|^b} (2b)^{pb} e^{-pb}$$

FOR ALL $p \times p$ POSITIVE DEFINITE MATRICES $\boldsymbol{\Sigma}$, WITH EQUALITY HOLDING FOR $\boldsymbol{\Sigma} = (1/2b) \mathbf{B}$. TAKING $b = (n_1 + n_2)/2$ AND

$$\begin{aligned} \mathbf{B} &= \sum_{j=1}^{n_1} (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1) (\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)' + \sum_{j=1}^{n_2} (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2) (\mathbf{x}_{2j} - \bar{\mathbf{x}}_2)' \\ &= (n_1 - 1) \mathbf{S}_1 + (n_2 - 1) \mathbf{S}_2 \end{aligned}$$

WE HAVE

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n_1 + n_2} [(n_1 - 1) \mathbf{S}_1 + (n_2 - 1) \mathbf{S}_2]$$

AS CLAIMED.

4. Consider the data in Table 6.17 from the textbook. These represent a two-factor experiment on peanut crops. The two factors are (i) the geographical location of the crop (2 locations were considered) and (ii) the variety of peanut grown (three varieties were considered). We have 2 crops under each of the $2 \times 3 = 6$ factor combinations. For each crop, we have measurements of three weight variables: X_1 = total yield, X_2 = sound mature kernels, and X_3 = seed size. So, in terms of a two-way MANOVA model, $g = 2$, $b = 3$, and $n = 2$.

- (a) Construct the two-way MANOVA table.

WE HAVE

$$\text{SSP}_{\text{FAC 1}} = \begin{bmatrix} 0.70 & -10.66 & 7.13 \\ -10.66 & 162.07 & -108.41 \\ 7.13 & -108.41 & 72.52 \end{bmatrix}$$

WITH 1 DEGREE OF FREEDOM,

$$SSP_{\text{FAC } 2} = \begin{bmatrix} 196.11 & 365.18 & 42.63 \\ 365.18 & 1089.02 & 414.66 \\ 42.63 & 414.66 & 284.10 \end{bmatrix}$$

WITH 2 DEGREES OF FREEDOM,

$$SSP_{\text{INT}} = \begin{bmatrix} 205.10 & 363.67 & 107.79 \\ 363.67 & 780.69 & 254.22 \\ 107.79 & 254.22 & 85.95 \end{bmatrix}$$

WITH 2 DEGREES OF FREEDOM,

$$SSP_{\text{RES}} = \begin{bmatrix} 104.20 & 49.36 & 76.48 \\ 49.36 & 352.10 & 122.00 \\ 76.48 & 122.00 & 94.84 \end{bmatrix}$$

WITH 6 DEGREES OF FREEDOM, AND

$$SSP_{\text{COR}} = \begin{bmatrix} 506.12 & 767.56 & 234.02 \\ 767.56 & 2383.88 & 682.46 \\ 234.02 & 682.46 & 537.41 \end{bmatrix}$$

WITH 11 DEGREES OF FREEDOM.

- (b) Test for a location effect, a variety effect, and a location-variety interaction at $\alpha = 0.05$. Do this using the `manova` function in R, and provide R code that matches the Wilks' statistics it computes. Note that your p-values (computed according to the book / notes) will not match those of `manova`. The distributional results we have learned for two-way MANOVA are large-sample approximations. That said, how do your p-values compare to those of `manova`? Overall, what do you conclude about these data?

THE WILKS LAMBDA STATISTICS AND P-VALUES ARE:

	STATISTIC	P-VALUES	
		TEXTBOOK	R
LOCATION	0.1065	0.0205	0.0205
VARIETY	0.0124	0.0002	0.0019
INTERACTION	0.0743	0.0055	0.0508

OUR LARGE-SAMPLE P-VALUES ARE SIMILAR TO THOSE REPORTED BY R. THERE IS MODERATE EVIDENCE OF AN INTERACTION EFFECT AND RELATIVELY-STRONG EVIDENCE OF MAIN EFFECTS FOR BOTH LOCATION AND VARIETY. YOU COULD ALSO HAVE COMPUTED THE P-VALUES USING A LARGE-SAMPLE χ^2 APPROXIMATION; THAT IS FINE.

```

####
#### (1)
####

library(DescTools)
library(plotrix)

## Input data.
X <- as.matrix(read.delim("T6-1.DAT", header = FALSE, sep = ""))
n <- nrow(X)
p <- ncol(X) / 2
colnames(X) <- c("BOD_1", "SS_1", "BOD_2", "SS_2")

## Summary statistics.
d <- cbind(X[, 1] - X[, 3], X[, 2] - X[, 4])
d_bar <- colMeans(d)
S_d <- var(d)

##
## (a)
##

T2 <- n * t(d_bar - 0) %*% solve(S_d) %*% (d_bar - 0)
T2_scaled <- (n - p) / ((n - 1) * p) * T2
p_value <- 1 - pf(T2_scaled, p, n - p)

HotellingsT2Test(d, mu = rep(0, p))

##
## (b)
##

ee <- eigen(S_d)
lambda <- ee$values
ee <- ee$vectors
theta <- atan(ee[2, 1] / ee[1, 1]) * (360 / (2 * pi))
c2 <- ((n - 1) * p) / (n - p) * qf(0.99, p, n - p)

pdf("figures/1_ellipse.pdf")
plot(c(-40, 10), c(-20, 45), xlab = expression(delta[1]),
     ylab = expression(delta[2]), asp = 1, type = "n")
draw.ellipse(d_bar[1], d_bar[2], sqrt(c2 * lambda[1] / n), sqrt(c2 * lambda[2] / n),
             angle = theta, lwd = 2)
points(d_bar[1], d_bar[2], pch = 20, col = "blue")
points(0, 0, pch = 20, col = "red")

```

```

##
## (c)
##

## T^2 intervals.
T2_interval <- function(a) {
  t(a) %*% d_bar + c(-1, 1) * sqrt(c2) * sqrt(t(a) %*% S_d %*% a / n)
}

T2_1 <- T2_interval(c(1, 0))
T2_2 <- T2_interval(c(0, 1))
T2_diff <- T2_interval(c(1, -1))

## Bonferroni intervals.
Bonf_interval <- function(a) {
  t(a) %*% d_bar + c(-1, 1) * qt(1 - 0.01 / (2 * 3), n - 1) * sqrt(t(a) %*% S_d %*% a / n)
}

Bonf_1 <- Bonf_interval(c(1, 0))
Bonf_2 <- Bonf_interval(c(0, 1))
Bonf_diff <- Bonf_interval(c(1, -1))

## Add lines to ellipse plot.
lines(c(T2_1[1], T2_1[1]), c(-100, T2_2[2]), lty = 2, col = "red")
lines(c(T2_1[2], T2_1[2]), c(-100, T2_2[2]), lty = 2, col = "red")
lines(c(-100, T2_1[2]), c(T2_2[1], T2_2[1]), lty = 2, col = "red")
lines(c(-100, T2_1[2]), c(T2_2[2], T2_2[2]), lty = 2, col = "red")

lines(c(Bonf_1[1], Bonf_1[1]), c(-100, Bonf_2[2]), lty = 2, col = "blue")
lines(c(Bonf_1[2], Bonf_1[2]), c(-100, Bonf_2[2]), lty = 2, col = "blue")
lines(c(-100, Bonf_1[2]), c(Bonf_2[1], Bonf_2[1]), lty = 2, col = "blue")
lines(c(-100, Bonf_1[2]), c(Bonf_2[2], Bonf_2[2]), lty = 2, col = "blue")
dev.off()

##
## (d)
##

HotellingsT2Test(d[-8, ], mu = rep(0, p))

####
#### (2) Data from Exercise 6.8 in the textbook.
####

```

```

library(Hotelling)

p <- 2
g <- 3
n <- c(5, 3, 4)

X <- matrix(c(6, 5, 8, 4, 7, 3, 1, 2, 2, 5, 3, 2, 7, 9, 6, 9, 9, 3, 6, 3, 3, 1, 1, 3),
  ncol = p)
TX <- factor(rep(1:g, times = n))
X_df <- data.frame("TX" = TX, "X_1" = X[, 1], "X_2" = X[, 2])

##
## (a)
##

## Prep for only groups 2 and 3.
g_a <- 2
n_a <- n[-1]

X_a <- X[-(1:n[1]), ]
TX_a <- factor(rep(2:3, times = n_a))
X_df_a <- data.frame("TX" = TX_a, "X_1" = X_a[, 1], "X_2" = X_a[, 2])

## Summary statistics.
x_bar_1 <- by(X_a, TX_a, colMeans)
S_1 <- by(X_a, TX_a, var)
S_po <- ((n_a[1] - 1) * S_1[[1]] + (n_a[2] - 1) * S_1[[2]]) / (sum(n_a) - 2)

## Two-sample T^2 test.
T2 <- (1 / sum(1 / n_a)) * t(x_bar_1[[1]] - x_bar_1[[2]]) %*% solve(S_po) %*%
  (x_bar_1[[1]] - x_bar_1[[2]])
T2_scaled <- ((sum(n_a) - p - 1) / (p * (sum(n_a) - 2))) * T2
p_value <- 1 - pf(T2_scaled, p, sum(n_a) - p - 1)

print(hotelling.test(X_1 + X_2 ~ TX, data = X_df_a))

## 99% Bonferroni intervals on component differences.
x_bar_1[[1]][1] - x_bar_1[[2]][1] +
  c(-1, 1) * qt(1 - 0.01 / (2 * p), n_a[1] + n_a[2] - 2) *
  sqrt(sum(1 / n_a) * S_po[1, 1])
x_bar_1[[1]][2] - x_bar_1[[2]][2] +
  c(-1, 1) * qt(1 - 0.01 / (2 * p), n_a[1] + n_a[2] - 2) *
  sqrt(sum(1 / n_a) * S_po[2, 2])

##

```

```

## (b)
##

## Summary statistics.
x_bar <- colMeans(X)
x_bar_l <- by(X, TX, colMeans)
S_l <- by(X, TX, var)

## MANOVA calculations.
B <- n[1] * (x_bar_l[[1]] - x_bar) %*% t(x_bar_l[[1]] - x_bar) +
  n[2] * (x_bar_l[[2]] - x_bar) %*% t(x_bar_l[[2]] - x_bar) +
  n[3] * (x_bar_l[[3]] - x_bar) %*% t(x_bar_l[[3]] - x_bar)
W <- (n[1] - 1) * S_l[[1]] + (n[2] - 1) * S_l[[2]] + (n[3] - 1) * S_l[[3]]
d_f <- c(g - 1, sum(n) - g, sum(n) - 1)

Lambda <- det(W) / det(B + W)
Lambda_scaled <- ((sum(n) - p - 2) / p) * (1 - sqrt(Lambda)) / sqrt(Lambda)
p_value <- 1 - pf(Lambda_scaled, 2 * p, 2 * (sum(n) - p - 2))

fit <- manova(cbind(X_1, X_2) ~ TX, data = X_df)
summary(fit, test = "Wilks")

####
#### (4)
####

## Load data.
p <- 3
g <- 2
b <- 3
n <- 2

X_df <- read.table("T6-17.DAT", header = FALSE)
colnames(X_df) <- c("Location", "Variety", "X_1", "X_2", "X_3")
X_df$Location <- factor(X_df$Location)
X_df$Variety <- factor(X_df$Variety)

X <- as.matrix(X_df[, 3:5])

## Summary statistics.
x_bar <- colMeans(X)
x_bar_l <- by(X, X_df$Location, colMeans)
x_bar_k <- by(X, X_df$Variety, colMeans)

Loc_Var <- factor(rep(c("1_1", "2_1", "1_2", "2_2", "1_3", "2_3"), each = n))

```



```

x_bar_lk <- by(X, Loc_Var, colMeans)
S_lk <- by(X, Loc_Var, var)

## MANOVA calculations.
SSP_1 <- SSP_2 <- SSP_int <- matrix(0, nrow = p, ncol = p)

for(l in 1:g)
  SSP_1 <- SSP_1 + b * n * (x_bar_l[[l]] - x_bar) %*% t(x_bar_l[[l]] - x_bar)
for(k in 1:b)
  SSP_2 <- SSP_2 + g * n * (x_bar_k[[k]] - x_bar) %*% t(x_bar_k[[k]] - x_bar)
for(l in 1:g) {
  for(k in 1:b) {
    innards <- x_bar_lk[[k + (l - 1) * b]] - x_bar_l[[l]] - x_bar_k[[k]] + x_bar
    SSP_int <- SSP_int + n * innards %*% t(innards)
  }
}
SSP_res <- Reduce("+", S_lk)

## Testing interaction.
Lambda_int <- det(SSP_res) / det(SSP_int + SSP_res)
Lambda_scaled <- (((g * b * (n - 1) - p + 1) / 2) /
  ((abs((g - 1) * (b - 1) - p) + 1) / 2)) * (1 - Lambda_int) / Lambda_int
p_val_int <- 1 - pf(Lambda_scaled, abs((g - 1) * (b - 1) - p) + 1, g * b * (n - 1) - p + 1)

## Testing location main effect.
Lambda_1 <- det(SSP_res) / det(SSP_1 + SSP_res)
Lambda_scaled <- (((g * b * (n - 1) - p + 1) / 2) / ((abs((g - 1) - p) + 1) / 2)) *
  (1 - Lambda_1) / Lambda_1
p_val_1 <- 1 - pf(Lambda_scaled, abs((g - 1) - p) + 1, g * b * (n - 1) - p + 1)

## Testing variety main effect.
Lambda_2 <- det(SSP_res) / det(SSP_2 + SSP_res)
Lambda_scaled <- (((g * b * (n - 1) - p + 1) / 2) / ((abs((b - 1) - p) + 1) / 2)) *
  (1 - Lambda_2) / Lambda_2
p_val_2 <- 1 - pf(Lambda_scaled, abs((b - 1) - p) + 1, g * b * (n - 1) - p + 1)

summary(manova(cbind(X_1, X_2, X_3) ~ Location * Variety, data = X_df), test = "Wilks")

```

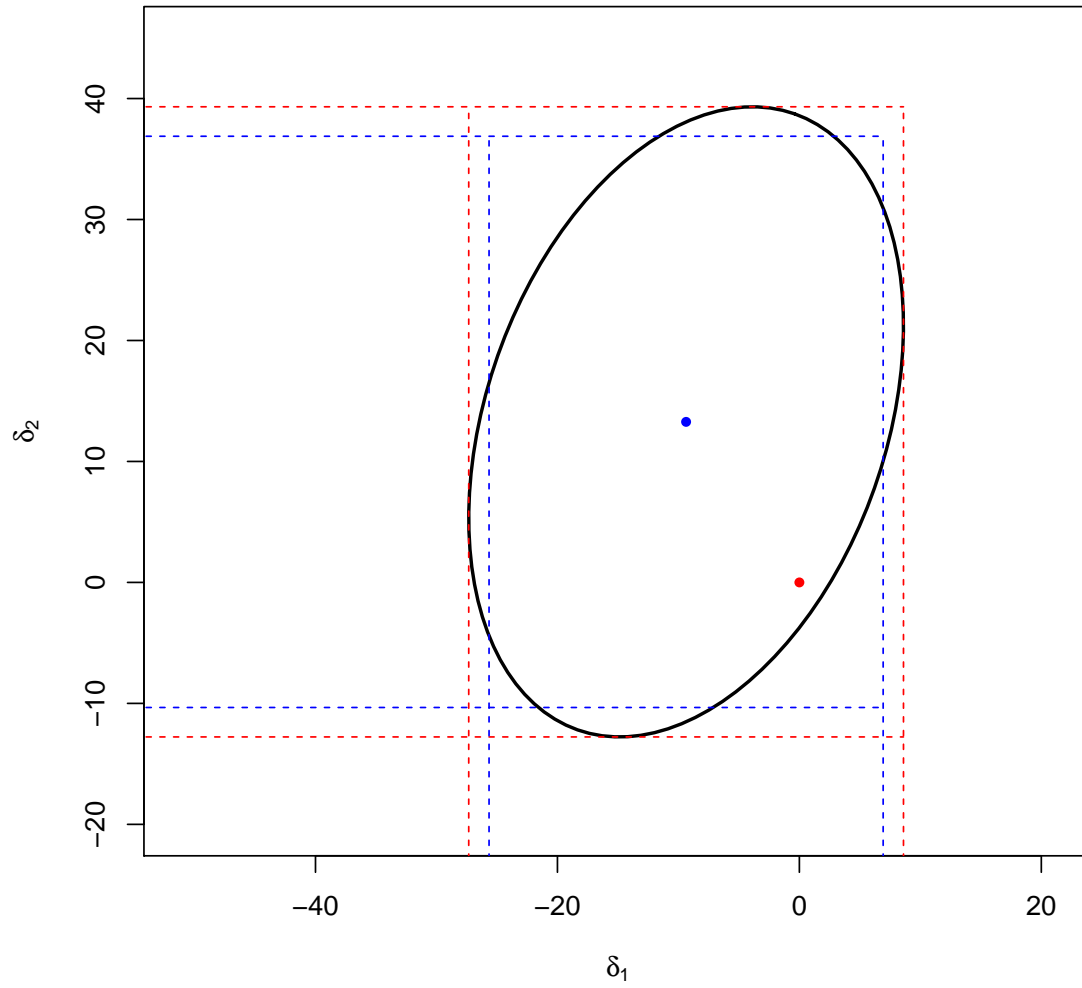


Figure 1: A 99% confidence region for $\boldsymbol{\delta}$ in number 1. The blue dot is at $\bar{\boldsymbol{\delta}}$, and the red dot is at $\boldsymbol{\mu}'_0 = [0, 0]$. The red dashed lines show 99% T^2 intervals, and the blue dashed lines show 99% Bonferroni intervals.