

# 4630 Assignment 2

Ravish Kamath: 213893664

06 December, 2022

## Question 1

Posten (1962) performed an experiment with 2 factors: velocity (2 levels:  $V_1$  and  $V_2$ ) and lubricants (three types:  $L_1$ ,  $L_2$ , and  $L_3$ ). The ultimate torque  $x_1$  and the ultimate strain  $x_2$  of homogeneous pieces of bar steel are measured at each treatment combinations. The data are given below:

##	Lubricant	Velocity	torque	strain
## 1	L1	V1	7.80	90.4
## 2	L1	V1	7.10	88.9
## 3	L1	V1	7.89	85.9
## 4	L1	V1	7.82	88.8
## 5	L1	V2	7.12	85.1
## 6	L1	V2	7.06	89.0
## 7	L1	V2	7.45	75.9
## 8	L1	V2	7.45	77.9
## 9	L2	V1	9.00	82.5
## 10	L2	V1	8.43	92.4
## 11	L2	V1	7.65	82.4
## 12	L2	V1	7.70	87.4
## 13	L2	V2	8.19	66.0
## 14	L2	V2	8.25	74.5
## 15	L2	V2	7.45	83.1
## 16	L2	V2	7.45	86.4
## 17	L3	V1	7.60	94.1
## 18	L3	V1	7.00	86.6
## 19	L3	V1	7.82	85.9
## 20	L3	V1	7.80	88.8
## 21	L3	V2	7.06	81.2
## 22	L3	V2	7.04	79.9
## 23	L3	V2	7.52	86.4
## 24	L3	V2	7.70	76.4

- State clearly the model in terms of the overall mean, main effects and interaction effects. This should include all the necessary assumptions and constraints such that we can answer the rest of the questions.
- Obtain all the necessary sum of squares.
- Is there any evidence that treatment effect exists?
- Regardless of the answer in part (c), test for interaction effect and then test for main effect.

## Solution

### Part A

$$\begin{aligned} E(X_{lkr}) &= \mu + \tau_l + \beta_k + \gamma_{lk} + \epsilon_{lkr} \\ &= \mu + \text{Lubricant}_l + \text{Velocity}_k + (\text{Lubricant} \times \text{Velocity})_{lk} + \epsilon_{lkr} \end{aligned}$$

where:

$$l = L_1, L_2, L_3$$

$$k = V_1, V_2$$

$$r = 1, \dots, 4$$

$\mu$  represents the overall mean of the model.

$\text{Lubricant}_l$  represents the  $L^{th}$  fixed lubricant effect of factor 1

$\text{Velocity}_k$  represents the  $k^{th}$  fixed velocity effect of factor 2

$(\text{Lubricant} \times \text{Velocity})_{lk}$  represents the interaction effect between lubricant and velocity at the  $lk^{th}$  level.

Assumptions:  $\epsilon_{lkr} \stackrel{iid}{\sim} N_2(0, \Sigma)$  and  $\sum_{l=L_1}^{L_3} = \sum_{k=V_1}^{V_2} \beta_k = \sum_{l=L_1}^{L_3} \gamma_{lk} = \sum_{k=V_1}^{V_2} \gamma_{lk} = 0$ .

### Part B

Running it through SAS, we get:

$$SSE = \begin{bmatrix} 3.1532 & -14.9535 \\ -14.9535 & 535.9725 \end{bmatrix} \quad SSLubricant = \begin{bmatrix} 1.6927 & -9.6989 \\ -9.6989 & 56.3233 \end{bmatrix} \quad SSVelocity = \begin{bmatrix} 0.6240 & 14.8834 \\ 14.8834 & 354.9704 \end{bmatrix}$$

$$\begin{aligned} SSInteraction &= \begin{bmatrix} 0.0290 & -0.102 \\ -0.102 & 4.7233 \end{bmatrix} & SSTreatment &= SSLubricant + SSVelocity + SSInteraction \\ & & &= \begin{bmatrix} 2.3457 & 5.0825 \\ 5.0825 & 416.0170 \end{bmatrix} \end{aligned}$$

$$SSTotal = \begin{bmatrix} 5.4989 & -9.8710 \\ -9.8710 & 951.9895 \end{bmatrix}$$

### Part C

Let  $a = 3, b = 2, p = 2, n = 4$ .

```
SSE = matrix(c(3.1532, -14.9535, -14.9535, 535.9725), nrow = 2, ncol = 2, byrow = T)
SSLub = matrix(c(1.6927, -9.6989, -9.6989, 56.3233), nrow = 2, ncol = 2, byrow = T)
SSVel = matrix(c(0.6240, 14.8834, 14.8834, 354.9704), nrow = 2, ncol = 2, byrow = T)
SSInt = matrix(c(0.0290, -0.102, -0.102, 4.7233), nrow = 2, ncol = 2, byrow = T)
SStr = SSLub + SSVel + SSInt
wilkslam = det(SSE)/det(SSE + SStr)
wilkslam
```

```
## [1] 0.2854371
```

```
a = 3
b = 2
p = 2
n = 4
#chi-squared observed
chi_obs = -(a*b*(n-1) - ((p+1) - (a*b - 1))/2)*log(wilkslam)
chi_obs
```

```
## [1] 23.82094
```

```
#P-Value
```

```
pchisq(chi_obs, df = 10, lower.tail = F)
```

```
## [1] 0.00809013
```

Since it is a really small p-value, implying that based off the data, **there is evidence** that treatment effect exists.

## Part D

MANOVA Tests for the Hypothesis of No Overall Lub Effect H = Type III SSCP Matrix for Lub E = Error SSCP Matrix  S=2 M=-0.5 N=7.5		
Statistic	Value	P-Value
Wilks' Lambda	0.64815752	0.1082
Pillai's Trace	0.35240258	0.1188
Hotelling-Lawley Trace	0.54197071	0.0988
Roy's Greatest Root	0.54037156	0.0790

Figure 1: Lubrication Effect

Based off SAS, testing for existence of lubricant effect gives us a **0.1082 p-value**. This is a large p-value hence this implies that lubrication effect is **not significant**.

MANOVA Tests for the Hypothesis of No Overall Vel Effect H = Type III SSCP Matrix for Vel E = Error SSCP Matrix  S=1 M=0 N=7.5		
Statistic	Value	P-Value
Wilks' Lambda	0.43574673	0.0009
Pillai's Trace	0.56425327	0.0009
Hotelling-Lawley Trace	1.29491109	0.0009
Roy's Greatest Root	1.29491109	0.0009

Figure 2: Velocity Effect

Based of SAS, testing for existence of velocity effect gives **0.0009 p-value**, which is quite small. This implies that velocity effect **is significant**.

MANOVA Tests for the Hypothesis of No Overall Lub*Vel Effect H = Type III SSCP Matrix for Lub*Vel E = Error SSCP Matrix  S=2 M=-0.5 N=7.5		
Statistic	Value	P-Value
Wilks' Lambda	0.98157464	0.9881
Pillai's Trace	0.01851016	0.9910
Hotelling-Lawley Trace	0.01868483	0.9882
Roy's Greatest Root	0.01028339	0.9950

Figure 3: Interaction Effect

Based of SAS, testing for existence of interaction effect gives **0.9881 p-value**. This is a large value which means that interaction effect is **not significant**.

## Question 2

Using the data set given in Question 1 and ignoring the velocity factor.

- Is there any evidence that lubricant effect exists?
- Obtain the 95% confidence ellipsoid for the mean difference between the  $L_1$  and  $L_3$ .
- Is there evidence of heterogeneity in variance?

## Solution

### Part A

```
n1_indices = which(df$Lubricant == 'L1')
n1 = dim(df[n1_indices,])[1]
n2_indices = which(df$Lubricant == 'L2')
n2 = dim(df[n2_indices,])[1]
n3_indices = which(df$Lubricant == 'L3')
n3 = dim(df[n3_indices,])[1]
grp_obs = c(n1, n2, n3)
n = sum(grp_obs)
g = nlevels(factor(df$Lubricant))
p = 2

#Finding the means for each group and overall mean
L1mean = as.vector(colMeans(df[n1_indices[1]:tail(n1_indices, n = 1), 3:4]))
L2mean = as.vector(colMeans(df[n2_indices[1]:tail(n2_indices, n = 1), 3:4]))
L3mean = as.vector(colMeans(df[n3_indices[1]:tail(n3_indices, n = 1), 3:4]))
lub_group_mean = cbind(L1mean, L2mean, L3mean)
mean = (n1*L1mean + n2*L2mean + n3*L3mean)/(n1 + n2 + n3)

#Finding the sample covariance for each lubricant
X1 = as.matrix(df[n1_indices[1]:tail(n1_indices, n = 1), 3:4])
X2 = as.matrix(df[n2_indices[1]:tail(n2_indices, n = 1), 3:4])
X3 = as.matrix(df[n3_indices[1]:tail(n3_indices, n = 1), 3:4])
S1 = 1/(n1 - 1)*(t(X1)%*%X1 - n1*colMeans(X1)%*%t(colMeans(X1)))
S2 = 1/(n2 - 1)*(t(X2)%*%X2 - n2*colMeans(X2)%*%t(colMeans(X2)))
S3 = 1/(n3 - 1)*(t(X3)%*%X3 - n3*colMeans(X3)%*%t(colMeans(X3)))
S = list(S1, S2, S3)

#Calculate the Within Matrix
W = matrix(0, p, p)
Wnew = matrix(0, p, p)
for (i in 1:g){
  Wnew = as.matrix(lapply(S[i], '*', (grp_obs[i] - 1)))
  Wnew = matrix(unlist(Wnew), ncol = 2, byrow = T)
  W = W + Wnew
}
print(W)

##           [,1]      [,2]
## [1,]  3.806237 -0.172125
## [2,] -0.172125 895.666250
```

```
#Calculating the Between Matrix
B = matrix(0,p,p)
Bnew = matrix(0,p,p)
for (i in 1:g){
  Bnew = grp_obs[i]*(lub_group_mean[,i] - mean)%*%t(lub_group_mean[,i] - mean)
  B = B + Bnew
}
print(B)
```

```
##           [,1]      [,2]
## [1,]  1.692658 -9.698917
## [2,] -9.698917 56.323333
```

```
#Wilks Lambda and finding p-value
wilkslam = det(W)/det(W + B)
wilkslam
```

```
## [1] 0.6635755
```

```
fobs = (n - g - 1)/(g - 1)*((1 - sqrt(wilkslam))/sqrt(wilkslam))
fobs
```

```
## [1] 2.275942
```

```
pvalue = pf(fobs, df1 = 2*(g-1),df2 = 2*(n - g - 1), lower.tail = F)
pvalue
```

```
## [1] 0.07793588
```

P-value is **0.07**. If we have our  $\alpha = 0.05$ , then there is **no evidence** to reject  $H_0$ . Hence this implies that based on the data, the lubricant effect **does not exist**.

## Part B

```
#two sample difference
tau1_hat = L1mean - mean
tau2_hat = L2mean - mean
tau3_hat = L3mean - mean
tau_vec = matrix(data = c(tau1_hat, tau2_hat, tau3_hat), nco = 3, byrow = F)
A = c(1, 0, -1)
tau_hat = tau_vec%*%A
tau_hat
```

```
##           [,1]
## [1,] 0.01875
## [2,] 0.32500
```

```
#Finding the pooled variance for L1 and L3
Spooled_L1andL3 = ((n1 - 1)*S1+(n3 - 1)*S3)/(n1 + n3 - 2 )
```

```
#Variance for tau_hat
var_tau_hat = (1/n1 + 1/n3)*Spooled_L1andL3
var_tau_hat
```

```
##           torque      strain
## torque 0.03048281 0.0810067
## strain 0.08100670 7.5951339
```

```

#95% Ellipsoid
plot(tau_hat[1], tau_hat[2], type="p", xlim=c(-2, 2),
     ylim=c(-2, 2), xlab="L1mean", ylab="L3mean",
     main = '95% C.I. for mean difference between L1 and L3')

tau1 = matrix(seq(-2, 2, 0.05), ncol=1, byrow=T)
ntau1 = nrow(tau1)
tau3 = matrix(seq(-2, 2, 0.05), ncol=1, byrow=T)
ntau3 = nrow(tau3)

for (i in 1:ntau1) {
  for (j in 1:ntau3) {
    tau = matrix(c(tau1[i, 1], tau3[j, 1]), ncol=1, byrow=T)
    Tsq_obs = t((tau_hat - tau))%*%solve(var_tau_hat)%*(tau_hat-tau)
    Fcomp = c( (n1 + n3 - 2) - p + 1)/((n1 + n3 - 2)*p) * Tsq_obs
    Fcrit = qf(0.05, p, n1 + n3 - p - 1)
    if (Fcomp < Fcrit) points(tau1[i, 1], tau3[j, 1], pch="*")
  }
}
points(tau_hat[1], tau_hat[2], col='red')

```

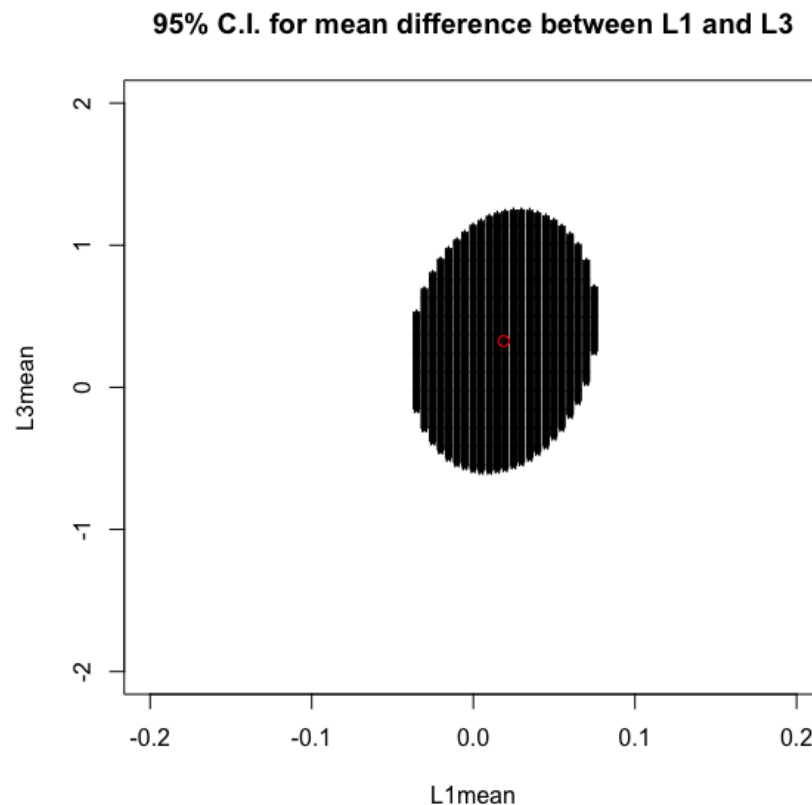


Figure 4: 95 percent Ellipsoid



## Part C

```
#Taking the determinants of the sample co variances
```

```
detS1 = det(S1)
```

```
detS2 = det(S2)
```

```
detS3 = det(S3)
```

```
detS = c(detS1, detS2, detS3)
```

```
#Finding the Spooled and its determinant
```

```
Spooled = W/(n1 + n2 + n3 - g)
```

```
detSpooled = det(Spooled)
```

```
detSpooled
```

```
## [1] 7.73036
```

```
#Doing the Box Test for equality of co variances
```

```
grp_obs = c(n1, n2 ,n3)
```

```
lambda = rep(1,1)
```

```
lambda_new = rep(0,1)
```

```
for (i in 1:g){
```

```
  lambda_new = (detS[i]/detSpooled)^((grp_obs[i]-1)/2)
```

```
  lambda = lambda*lambda_new
```

```
}
```

```
print(lambda)
```

```
## [1] 0.1140832
```

```
M = -2*log(lambda)
```

```
M
```

```
## [1] 4.341655
```

```
u = (sum(1/(grp_obs - 1)) - 1/(sum(grp_obs - 1)))*((2*p^2+3*p-1)/(6*(p+1)*(g-1)))
```

```
u
```

```
## [1] 0.1375661
```

```
C = (1 - u)*M
```

```
pchisq(C, df = ((1+p)*p*(g-1))/2, lower.tail = F)
```

```
## [1] 0.7112208
```

Since the p-value is large, then this implies that based off the data, there is **no evidence** of heterogeneity in the variance.

### Question 3

To compare two types of coating for resistance to corrosion, 15 pieces of pipe were coated with each type of coating. Two pipes, one with each type of coating, were buried together and left for the same length of time at 14 loactions. Corrosion for the coating was measured by two variables:

$x_1$  = maximum depth of pit in thousandths of an inch

$x_2$  = number of pits

The data are:

##	Coating	x1	x2
## 1	1	73	31
## 2	1	43	19
## 3	1	47	22
## 4	1	53	26
## 5	1	58	36
## 6	1	47	30
## 7	1	52	29
## 8	1	38	36
## 9	1	61	34
## 10	1	56	33
## 11	1	56	19
## 12	1	34	19
## 13	1	55	26
## 14	1	65	15
## 15	1	75	18
## 16	2	51	35
## 17	2	41	14
## 18	2	43	19
## 19	2	41	29
## 20	2	47	34
## 21	2	32	26
## 22	2	24	19
## 23	2	43	37
## 24	2	53	24
## 25	2	52	27
## 26	2	57	14
## 27	2	44	19
## 28	2	57	30
## 29	2	40	7
## 30	2	68	13

Do the two coatings differ significantly in their effect on corrosion? Clearly state the needed assumptions for your analysis.

### Solution

#### Assumptions:

- (1)  $X_{l,1}, \dots, X_{l,15}$  is a random sample size from a population with  $\mu_l$ , where  $l = Coating_1, Coating_2$ . The random samples from different populations and independent.
- (2) All population have a common variance matrix  $\Sigma$ .
- (3) Each population is multivariate normal.

Furthermore, let  $H_0 : \mu_1 = \mu_2$  and  $H_a : \mu_1 \neq \mu_2$

```

n1_indices = which(df$Coating == 1)
n1 = dim(df[n1_indices,])[1]
n2_indices = which(df$Coating == 2)
n2 = dim(df[n2_indices,])[1]
grp_obs = c(n1, n2)
n = grp_obs[1]
p = 2
g = nlevels(factor(df$Coating))

#Difference of Mean
C1 = df[n1_indices, 2:3]
C2 = df[n2_indices, 2:3]
d = C1 - C2
d_bar = colMeans(d)
d_bar

##          x1          x2
## 8.000000 3.066667

# sample Variance covariance matrix
S = cov(d)
S

##          x1          x2
## x1 121.57143 17.07143
## x2 17.07143 21.78095

HotellingT = n*t(d_bar)%*%solve(S)%*%d_bar
HotellingT

##          [,1]
## [1,] 10.8189

F_obs = (n-p)/((n-1)*p)*HotellingT
F_obs

##          [,1]
## [1,] 5.02306

pvalue = pf(F_obs, df1 = p, df2 = 15 - p, lower.tail = F)
pvalue

##          [,1]
## [1,] 0.02419613

```

Since p-value is small, we can reject  $H_0$ , which implies that based off the data, the two coating **do differ significantly** in their effect on corrosion.

### Question 4

Let  $x_1, \dots, x_n$  be a sample from  $N_2(\mu, \Sigma)$  where  $\Sigma$  is a diagonal matrix with  $\sigma_1^2$  and  $\sigma_2^2$  be the diagonal entries. Derive the likelihood ratio statistics for testing  $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ .

### Solution