PURDUE UNIVERSITY

IE 33000: Probability and Statistics in Engineering II (Fall 2022)
School of Industrial Engineering, Purdue University

Homework 6

Instruction: There are 4 problems in total.

1) Problems 1-4 – 25 points

Due November 14, 2022 (11:59 pm)

For all problems, provide both hand-written solutions and R codes (for 30 points bonus), wherever applicable.

Problem 1

Consider the following computer output for an experiment. The factor was tested over four levels.

Source	DF	SS	MS	F	P-value
Factor	? $\alpha-1$	? SS_T	330.4716 $\frac{SS_T}{\alpha-1}$	4.42 $\frac{MS_T}{MS_E}$	?
Error	? $\alpha(n-1)$	? SS_E	?		
Total	31 $\alpha n-1$	? SS_T+SS_E			

(a) How many replicates did the experimenter use?

(b) Fill in the missing information in the ANOVA table. Use bounds for the P-value if you use the tables, not a calculator nor R.

(c) What conclusions can you draw about differences in the factor-level means?

a) treatments = 4

$n = df + 1$
$= 31 + 1$
$= 32$

$32/4 = 8$

b) From F-table: p-val = 0.0115

$4.42 = \frac{330.4716}{MS_E}$

$MS_E = 74.767$

$DF_F = 4-1$
$= 3$

$DF_E = 31 - 3$
$= 28$

$3 = \alpha - 1$
$\alpha = 4$

$\frac{SST}{3} = 330.4716$

$SST = 991.41$

$SSE = MS_E \cdot DF_E$
$= 74.767 \cdot 28$
$= 2093.48$

$SS_{Total} = SS_T + SS_E$
$= 3084.89$

c) Using H_0: no difference in means
H_1: difference in means

We say Fstat $= 4.42 \Rightarrow$ p-val $= 0.0115$ which is $< \alpha = 0.05$

So we reject H_0 and say there are differences in the factor level means.

Problem 2

A research study described an experiment to determine the effect of C_2F_6 flow rate on the uniformity of the etch on a silicon wafer used in integrated circuit manufacturing. Three flow rates are used in the experiment, and the resulting uniformity (in percent) for six replicates follows.

C_2F_6 Flow (SCCM)	Observations						
	1	2	3	4	5	6	AVG ($\bar{Y}$)
125	2.7	4.6	2.6	3.0	3.2	3.8	3.32
160	4.9	4.6	5.0	4.2	3.6	4.2	4.42
200	4.6	3.4	2.9	3.5	4.1	5.1	3.93

(a) Does C_2F_6 flow rate affect etch uniformity? Perform the analysis of variance using $\alpha = 0.05$.

(b) Apply Fisher's LSD method with $\alpha = 0.01$ and determine which levels of the factor differ.

$\bar{Y} = \dfrac{\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3}{3} = 3.89$

$SS = 6 \cdot \left[(\bar{y}_1 - \bar{y})^2 + (\bar{y}_2 - \bar{y})^2 + (\bar{y}_3 - \bar{y})^2 \right]$

$= 6 \cdot [0.327 + 0.279 + 0.002]$

$SS = 3.65$

$SS_T = \sum_j^3 \sum_i^6 (Y_{ij} - \bar{Y})^2$

$= (2.7 - 3.89)^2 + (4.9 - 3.89)^2 + (4.6 - 3.89)^2 \ldots + (3.8 - 3.89)^2 + (4.2 - 3.89)^2 + (5.1 - 3.89)^2$

$SS_T = 11.28$

$SS_E = SS_T - SS$
$= 11.28 - 3.65$
$= 7.63$

$DF_T = 18 - 1$
$= 17$

$DF_F = 3 - 1$
$= 2$

$DF_E = 18 - 3$
$= 15$

$MS_F = \dfrac{SS}{DF_F} = 1.83$

$MS_E = \dfrac{SS_E}{DF_E} = 0.509$

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_1 \neq \mu_2 \neq \mu_3$ (or)

$F_{stat} = \dfrac{MS_F}{MS_E} = \dfrac{1.83}{0.509}$

$F_{stat} = 3.60$

$F_{(2,15)(0.05)} = 3.68$

Since F_{stat} (3.6) is less than F_{crit} (3.68) we fail to reject and say flow rates don't affect uniformity.

B) $LSD = t_{\frac{0.01}{2}, (18-1)} \cdot \sqrt{\dfrac{2 \cdot MS_E}{n}}$

$2.898 \cdot 0.238$

$= 0.690$

SCCM 125 and 160

Problem 3

An article investigated four different methods of preparing the superconducting compound PbMo6S8. The authors contend that the presence of oxygen during the preparation process affects the material's superconducting transition temperature T_c. Preparation methods 1 and 2 use techniques that are designed to eliminate the presence of oxygen, and methods 3 and 4 allow oxygen to be present. Five observations on T_c (in °K) were made for each method, and the results are as follows:

Preparation Method	Transition Temperature T_c (°K)					Avg ($\bar{Y}$)
1	14.8	14.8	14.7	14.8	14.9	14.8
2	14.6	15.0	14.9	14.8	14.7	14.8
3	12.7	11.6	12.4	12.7	12.1	12.3
4	14.2	14.4	14.4	12.2	11.7	13.4

(a) Is there evidence to support the claim that the presence of oxygen during preparation affects the mean transition temperature? Use $\alpha = 0.05$.
(b) What is the P-value for the F-test in part (a)?
(c) Find a 95% confidence interval on mean Tc when method 1 is used to prepare the material.
(d) Apply Fisher's LSD method with $\alpha = 0.05$ and determine which levels of the factor differ.

a) $\bar{Y} = 13.83$ $\quad SS = 5 \cdot \left[(\bar{y}_1 - \bar{y})^2 + (\bar{y}_2 - \bar{y})^2 + (\bar{y}_3 - \bar{y})^2 + (\bar{y}_4 - \bar{y})^2 \right]$

$= 22.04$

$SS_T = \sum_j^4 \sum_i^5 (Y_{ij} - \bar{Y})^2$

$= 30.07$

$SS_E = SS_T - SS$
$= 8.03$

$DF_T = 20 - 1$
$= 19$

$DF_E = 20 - 4$
$= 16$

$DF_F = 4 - 1$
$= 3$

$MS_F = \frac{SS}{DF_F}$
$= 7.35$

$MS_E = \frac{SS_E}{DF_E}$
$= 0.502$

$F_{stat} = \frac{MS_F}{MS_E}$
$= \frac{7.35}{0.502}$

$F_{stat} = 14.85$

$F_{crit} = F_{(3,16)(0.05)}$
$= 3.24$

Since Fstat is greater than F_{crit} we say there is evidence that mean is equal across all 4 prep methods

b) From R pf(14.85, 3,16)

$p = 6.94e^{-5}$

c) $\bar{y} \pm t_{\frac{\alpha}{2},(n-1)a} \cdot \sqrt{\frac{MS_E}{n}}$

CI: $14.8 \pm 2.12 \cdot \sqrt{\frac{0.502}{4}}$

CI: $13.95 \le \mu_1 \le 15.65$

d) $LSD = t_{\frac{0.01}{2},(19)} \cdot \sqrt{\frac{2 \cdot MS_E}{n}}$

$= 2.861 \cdot \sqrt{\frac{2 \cdot 0.502}{20}}$

$= 0.641$

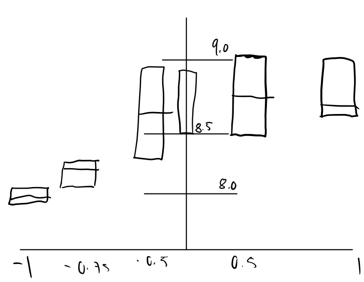
Problem 4

An article reported a study on the effects of additives on final polymer properties. In this case, polyurethane additives were referred to as cross-linkers. The average domain spacing was the measurement of the polymer property. The data are as follows:

Cross-Linker Level	Domain Spacing (nm)					
−1	8.2	8	8.2	7.9	8.1	8
−0.75	8.3	8.4	8.3	8.2	8.3	8.1
−0.5	8.9	8.7	8.9	8.4	8.3	8.5
0	8.5	8.7	8.7	8.7	8.8	8.8
0.5	8.8	9.1	9.0	8.7	8.9	8.5
1	8.6	8.5	8.6	8.7	8.8	8.8

(a) Is there a difference in the cross-linker level? Draw comparative box plots and perform an analysis of variance. Use $\alpha = 0.05$.
(b) Find the P-value of the test. Estimate the variability due to random error.
(c) Plot average domain spacing against cross-linker level and interpret the results.
(d) Apply Fisher's LSD method with $\alpha = 0.05$ and determine which levels of the factor differ.

a)



yes there a difference
a gradual increase then
slight decrease

c)

$a = 6$
$n = 36$

$\bar{y}_1 = 8.13$ $\bar{y}_4 = 8.7$
$\bar{y}_2 = 8.26$ $\bar{y}_5 = 8.83$
$\bar{y}_3 = 8.62$ $\bar{y}_6 = 8.6$
$\bar{Y} = 8.53$

$SS = 6 \cdot \left[(\bar{y}_1 - \bar{y})^2 + (\bar{y}_2 - \bar{y})^2 + (\bar{y}_3 - \bar{y})^2 + (\bar{y}_4 - \bar{y})^2 + (\bar{y}_5 - \bar{y})^2 \right]$
$= 2.189$

$SS_T = \sum_j^6 \sum_i^6 (Y_{ij} - \bar{Y})^2$
$= 3.408$

$SS_E = SS_T - SS$
$= 1.22$

$MS_F = \frac{2.189}{5}$
$= 0.438$

$MS_E = \frac{1.22}{30}$
$= 0.041$

$F_{stat} = \frac{0.438}{0.041}$
$= 10.683$

$F_{crit} = F_{(5, 30, 0.05)}$
$= 2.53$

Since $F_{stat} > F_{crit}$ we reject
null hyp. and say the

p value $= 5.79 \cdot 10^{-6}$

$\bar{y}_1 = 8.13$ $\bar{y}_4 = 8.7$
$\bar{y}_2 = 8.26$ $\bar{y}_5 = 8.83$
$\bar{y}_3 = 8.62$ $\bar{y}_6 = 8.6$

d) $LSD = t_{\frac{0.05}{2}, (35)} \cdot \sqrt{\frac{2 \cdot MS_E}{n}}$
$= 2.03 \cdot 0.047$
$= 0.09688$

```
###Problem 1###
MSf = 330.4716
Fstat = 4.45
DFt = 31
DFf = 4-1
DFe = 31-3
pval = pf(Fstat,DFf,DFe, lower.tail = FALSE)
MSe = MSf / Fstat
SSf = MSf * DFf
SSe = MSe * DFe
SSt = SSf + SSe

###Problem 2###
vals <-c(2.6,2.7,3,3.2,3.8,4.6,3.6,4.2,4.2,4.6,4.9,5,2.9,3.4,3.5,4.1,4.6,5.1)
groups <- c(rep("125", 6), rep("160", 6), rep("200", 6))
df<-data.frame(vals, groups)
result_anova <- aov(vals ~ groups, data = df)
summary(result_anova)
library(agricolae)
fisher <- LSD.test(result_anova,"groups"))

###Problem 3###
vals_3 <-c(14.8, 14.8, 14.7, 14.8, 14.9, 14.6, 15.0, 14.9, 14.8, 14.7, 12.7,
11.6,
         12.4, 12.7, 12.1, 14.2, 14.4, 14.4, 12.2, 11.7)
groups_3 <- c(rep("1", 5), rep("2", 5), rep("3", 5), rep("4",5))
df_3<-data.frame(vals_3, groups_3)
result_anova_3 <- aov(vals_3 ~ groups_3, data = df_3)
summary(result_anova_3)
library(agricolae)
fisher_3 <- LSD.test(result_anova_3,"groups_3")

###Problem 4###
vals_4 <- c(8.2, 8, 8.2, 7.9, 8.1, 8, 8.3, 8.4, 8.3, 8.2, 8.3, 8.1, 8.9, 8.7,
8.9,
          8.4, 8.3, 8.5, 8.5, 8.7, 8.7, 8.7, 8.8, 8.8, 8.8, 9.1, 9.0, 8.7,
8.9,
          8.5, 8.6, 8.5, 8.6, 8.7, 8.8, 8.8)
groups_4 <- c(rep("-1", 6), rep("-0.75", 6), rep("-0.5", 6), rep("0",6),
rep("0.5",6), rep("1",6))
df_4<-data.frame(vals_4, groups_4)
dim(vals_4) <- c(6,6)
colnames(vals_4)<- c("-1", "-0.75", "-0.5", "0", "0.5", "1")
boxplot(vals_4)
val_avg_4 <- colMeans(vals_4)
plot(val_avg_4)
result_anova_4 <- aov(vals_4 ~ groups_4, data = df_4)
summary(result_anova_4)
library(agricolae)
fisher_4 <- LSD.test(result_anova_4,"groups_4"))
```