Henry Lankin
Gui Larangeira
January 17, 2017
STAT 6305
Dr. Zhou

HW 1: 14.1, 14.5, 14.7, 14.16

14.1

EXERCISES
Engin.

14.1 Researchers ran a quality control study to evaluate the quality of plastic irrigation pipes. The study design involved a total of 24 pipes, with 12 pipes randomly selected from each of two manufacturing plants. The compressive strength was recorded at five locations on each of the pipes. The pipes were manufactured under one of two water temperatures and one of three types of hardeners. The experimental conditions are as follows:

Pipe No.	Plant	Temperature (°F)	Hardener	Pipe No.	Plant	Temperature (°F)	Hardener
1	1	200	H_1	13	1	200	H_3
2	1	175	H_2	14	1	175	H_3
3	2	200	H_1	15	2	200	H_3
4	2	175	H_2	16	2	175	H_3
5	1	200	H_1	17	1	200	H_2
6	1	175	H_2	18	1	175	H_1
7	2	200	H_1	19	2	200	H_2
8	2	175	H_2	20	2	175	H_1
9	1	200	H_3	21	1	200	H_2
10	1	175	H_3	22	1	175	H_1
11	2	200	H_3	23	2	200	H_2
12	2	175	H_3	24	2	175	H_1

Identify each of the following components of the experimental design.

a. factors
b. factor levels
c. blocks
d. experimental units
e. measurement units
f. replications
g. covariates
h. treatments

a. 3 factors: Plant, Temperature (°F), Hardener
b. Plant has 2 levels: Plant 1, Plant 2
   Temperature has 2 factor levels: $175, 200$
   Hardener has 3 factor levels: H_1, H_2, H_3
c. Blocks: none
d. Experimental units: 24 pipes
e. Measurement units: 5 locations measured on each of the 24 pipes totaling 120 measurement units
f. 2 replications: 2 for each combination of plant, temperature, and hardener
g. Covariates: could be thickness, weight, and length of each pipe

h. Treatments: 12 treatments listed in the table below

Treatment	Plant	Temperature	Hardener	Replications
1	1	175	H_1	pipes 18, 22
2	1	200	H_1	pipes 1, 5
3	2	175	H_1	pipes 20, 24
4	2	200	H_1	pipes 3, 7
5	1	175	H_2	pipes 2, 6
6	1	200	H_2	pipes 17, 21
7	2	175	H_2	pipes 4, 8
8	2	200	H_2	pipes 19, 23
9	1	175	H_3	pipes 10, 14
10	1	200	H_3	pipes 9, 13
11	2	175	H_3	pipes 12, 16
12	2	200	H_3	pipes 11, 15

14.5

Eng. **14.5** The production manager of a large investment casting firm is studying different methods to increase productivity in the workforce of the company. The process engineer and personnel in the human resource department develop three new incentive plans (B, C, D) for which they will design a study to compare the incentive plans with the current plan (plan A). Twenty workers are randomly assigned to each of the four plans. The response variable is the total number of units produced by each worker during one month on the incentive plans. The data are given here along with the output from Minitab.

	Incentive Plan			
Rep	**A**	**B**	**C**	**D**
1	422	521	437	582
2	431	545	422	639
3	784	600	473	735
4	711	406	478	800
5	641	563	397	853
6	709	361	944	748
7	344	387	394	622
8	599	700	890	514
9	511	348	488	714
10	381	944	521	627
11	349	545	387	548
12	387	337	633	644
13	394	427	627	736
14	621	771	444	528
15	328	752	1,467	595
16	636	810	828	572
17	388	406	644	627
18	901	537	1,154	546
19	394	816	430	701
20	350	369	508	664
Mean	514.1	557.2	628.3	649.8
St Dev	171.8	184.4	290.2	93.1

```
One-way ANOVA: C5 versus C6

Source    DF        SS     MS     F      P
Plan       3    236758  78919  2.02  0.118
Error     76   2972125  39107
Total     79   3208883

S = 197.8   R-Sq = 7.38%   R-Sq(adj) = 3.72%
```

a. State the null and alternative hypotheses being tested by the *F*-statistic in the AOV table.
b. Is there significant evidence ($\alpha = .05$) that the mean output associated with the four incentive plans is different?
c. Use Tukey's HSD procedure to identify the pairs of incentive plans which have different mean output.

a. Hypotheses being tested:

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_a: At\ least\ one\ of\ the\ population\ mean\ units\ produced$$
$$under\ a\ plan\ differs\ from\ the\ others$$

b. No, there is not significant evidence that at least one is different from the others since the p-value $= 0.118 > 0.05$.

c. Use Fisher's LSD procedure to identify the pairs of incentive plans which have different mean output.

```
# given information about the experiment
n <- 20
alpha <- 0.05
# point estimates of the means
incentive.means <- c(514.1, 557.2, 628.3, 649.8)
# from ANOVA output
MSE <- 39107
df.errors <- 76
# Fisher's LSD procedure
sd.errors <- sqrt(MSE)
t.value <- qt(alpha/2, df.errors, lower.tail = FALSE)
LSD <- t.value*sd.errors*sqrt(2/n)
        [1] "LSD = 124.550421"
diffCombn.matrix <- combn(incentive.means, 2)
means.diff <- diffCombn.matrix[2,] - diffCombn.matrix[1,]
[1] "Pairs to subtract listed for each possible combination by column:"
             [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
        [1,] 514.1 514.1 514.1 557.2 557.2 628.3
        [2,] 557.2 628.3 649.8 628.3 649.8 649.8
[1] "Differences for each pair in relative order to the matrix above:"
        [1]  43.1 114.2 135.7  71.1  92.6  21.5
```

We see that only the difference between plan d and plan a, $\bar{x}_D - \bar{x}_A = 135.7$, is larger than the LSD, $LSD = 124.6$. But, since H_0 was not rejected in part b, the LSD procedure would not be performed. If we ignored the result of the hypothesis test, the LSD procedure would imply that there is a significant difference between plans A and D.

14.7

Bio. **14.7** A research specialist for a large seafood company plans to investigate bacterial growth on oysters and mussels subjected to three different storage temperatures. Nine cold storage units are available. She plans to use three storage units for each of the three temperatures. One package of oysters and one package of mussels will be stored in each of the storage units for 2 weeks. At the end of the storage period, the packages will be removed and the bacterial count made for two samples from each package. The "treatment" factors of interest are temperature (levels: 0, 5, 10°C) and seafood (levels: oysters, mussels). She will also record the bacterial count for each package prior to placing seafood in the cooler. Identify each of the following components of the experimental design.

a. factors
b. factor levels
c. blocks
d. experimental units
e. measurement units
f. replications
g. covariates
h. treatments

a. 2 factors: temperature, seafood
b. Temperature has 3 factor levels: $0, 5, 10°\mathrm{C}$
   Seafood has 2 factor levels: oysters, mussels
c. Blocks: none
d. Experimental units: the package of oysters and package of mussels in each of the 9 storage containers
e. Measurement units: the 2 samples taken from each package, totaling 36 measurement units
f. 3 Replications: Each treatment will be applied in three different storage containers
g. Covariates: the bacterial count in each package prior to performing the experiment, the weight of each package of seafood, number of oysters and mussels in each package
h. 6 treatments: (oyster, 0), (oyster, 5), (oyster 10), (mussel, 0), (mussel, 5), (mussel, 10)

14.16

Bus. **14.16** A computer magazine wants to rate four software programs used to prepare annual federal income tax forms based on the amount of time needed to complete the form. The study will select individuals who have incomes less than $100,000 and who itemize their deductions. Determine how many individuals would be needed for each software program to declare a difference in the average completion times at the $\alpha = .05$ level of significance with a power of .90 if the difference between any pair of means is greater than 30 minutes. From previous studies using similar software, the standard deviation in completion time is thought to be about 12.25 minutes.

Given information:

$$\alpha = 0.05, \beta = 0.10, t = 4$$
$$D = 30, \hat{\sigma} = 12.25$$

The desired power is $Power = 1 - \beta$. We first obtain the parameters to use in Table 14 from the textbook,

$$\phi = \sqrt{\frac{rD^2}{2t\hat{\sigma}^2}} = \sqrt{\frac{r(30)^2}{2(4)(12.25)^2}}$$
$$\phi = 0.866\sqrt{r}$$

and,

$$v_1 = t - 1 = 4 - 1 = 3$$
$$v_2 = t(r-1) = 4(r-1)$$

By trial and error, we get the following table,

r	$v_2 = 4(r-1)$	$\phi = 0.866\sqrt{r}$	$Power = 1 - \beta$
3	8	1.50	0.48
7	24	2.29	0.97
5	16	1.94	0.84
6	20	2.12	0.90

```
## Alternative method: calculating the power table values using R

# values of replications to test
r <-c(3, 7, 5, 6)

# degrees of freedom for F-dist
v.1 <- t-1
v.2 <- t*(r-1)
```

```
df.1 <- v.1
df.2 <- v.2

# We use the non-centrality parameter of the F-distribution instead of the psi parameter of the
book
ncp <- r*D^2/(2*sd^2)
F.alpha <- qf(1-alpha, df.1, df.2)

# The power is P(F > F(alpha, t−1,t(r−1),ncp))
power <- 1- pf(F.alpha, df.1, df.2, ncp = ncp)

# Form the table as a data.frame

phi.value <- sqrt(ncp)
table.power <- data.frame(r, v.2, phi.value, power)
```

r	v.2	ncp	power
3	8	8.996252	0.4869220
7	24	20.991254	0.9567155
5	16	14.993753	0.8301800
6	20	17.992503	0.9120138

From the chart above, we see that with a minimum of 6 replications the desired power of the test, $power = 0.90$, will be most closely attained.