Homework 6
MTH 3220, Fall 2019
Due Tuesday, Oct. 22

Section in Book	**Problems**
10.2	Problem 1 (given below)
11.1	1*, 2**, 3***, 4***, 5, 6

* For **Problem 1** in **Section 11.1**, after calculating the test statistics F_A and F_B, you can obtain the p-values (F distribution upper tail areas) using the **pf()** function. For example, you should get $F_A = 7.16$, and so the p-value (from the $F(3, 6)$ distribution) is obtained by typing:

```
> pf(q = 7.16, df1 = 3, df2 = 6, lower.tail = FALSE)
```

** For **Problem 2** in **Section 11.1**, you're welcome to carry out the two-factor ANOVA using R. See the course website for the data set **ex_11_2.txt**. Then, after reading it into R using **read.table()**, type:

```
> my.anova <- aov(Corrosion ~ Coating + SoilType, data = my.data)
> summary(my.anova)
```

For **Part *b***, you can compute the grand mean $\bar{X}_{..}$ by typing:

```
> mean(my.data$Corrosion)
```

and you can compute the Factor A level means $\bar{X}_{1.}$, $\bar{X}_{2.}$, $\bar{X}_{3.}$, and $\bar{X}_{4.}$ by typing:

```
> aggregate(Corrosion ~ Coating, data = my.data, FUN = mean)
```

and the Factor B level means $\bar{X}_{.1}$, $\bar{X}_{.2}$, and $\bar{X}_{.3}$ by typing:

```
> aggregate(Corrosion ~ SoilType, data = my.data, FUN = mean)
```

*** For **Problems 3** and **4** in **Section 11.1**,, you may carry out the analysis using R. The data are in the files **ex_11_3.txt** and **ex_11_4.txt** on the course website. After reading the data into R using **read.table()** and carrying out the ANOVA using **aov()**, you can conduct ***Tukey's multiple comparison procedure*** by typing:

```
> TukeyHSD(my.anova)
```

Additional Problem (Next Page)

Problem 1 (**cont'd from HW 5**). Here are the data from the study to investigate the time (in minutes) taken to complete a task using three different word processing programs. The data are also in the file **ergonomics.txt**.

Group 1 Menu-Driven	**Group 2 Command-Driven**	**Group 3 Mixed**
12	14	10
15	11	8
12	12	10
10	11	7
11	13	9
13	14	8

Use **read.table()** (or **read.csv()**) to read to the data from **ergonomics.txt** into an R *data frame* named, say, **my.data**.

On **HW5**, you found statistically significant differences among the mean completion times for the three word processing programs via a one-factor ANOVA *F* test by typing:

```
> my.anova <- aov(Time ~ Group, data = my.data)
> summary(my.anova)
```

We want to check the assumptions required by the *F* test.

i. Write out the **one-factor ANOVA model** for the data in terms of the true means μ_1, μ_2, and μ_3 and the random error terms ε_{ij} (i.e. the **group means version** of the model). Be sure to state any **assumptions** associated with the ε_{ij}'s.

ii. State the **hypotheses** for a one-factor ANOVA *F* test in terms of the true means μ_1, μ_2, and μ_3.

iii. Now write out the **one-factor ANOVA model** in terms of the true grand mean μ, the treatment effects α_1, α_2, and α_3, and the random error terms ε_{ij} (i.e. the **treatment effects version** of the model), where

$$\mu = \frac{\mu_1+\mu_2+\mu_3}{3} \qquad \text{and} \qquad \alpha_i = \mu_i - \mu \quad \text{for } i = 1,2,3.$$

Be sure to state any **assumptions** associated with the ε_{ij}'s.

iv. State the **hypotheses** for a one-factor ANOVA *F* test in terms of the treatment effects α_1, α_2, and α_3.

v. Is the **normality assumption** required for the ANOVA *F* test met? Check by making a histogram and a normal probability plot of the residuals:

```
> hist(my.anova$residuals)
> qqnorm(my.anova$residuals)
> qqline(my.anova$residuals)
```

vi. Is the **constant standard deviation assumption** required for the ANOVA *F* test met? Check by plotting the residuals (*y*-axis) versus the fitted values (*x*-axis):

```
> plot(x = my.anova$fitted.values, y = my.anova$residuals)
> abline(h = 0)
```

vii. Show that the error sum of squares **SSE** (from the ANOVA table) is equal to the **sum** of **squared residuals**:

```
> sum(my.anova$residuals^2)
```

viii. Provide the numerical value of the **estimate**, $\sqrt{\text{MSE}}$, of the true (unknown) standard deviation $\boldsymbol{\sigma}$ of the N(0, σ) distribution of the random error term ε_{ij} in the ANOVA model.