Homework 7
MTH 3220, Fall 2019
Due Tuesday, Nov. 5

Section in Book	**Problems**
11.2	16 (skip part *e*), 17*, 18 (skip part *b*)**, 20***
11.3	Problem 1 (given below), 27 (skip part *d*), 29, 31****

* For **Problem 17**, do **Part *a* by hand** and **show your work**. You can get the p-value using the **pf()** function in R. **Part *b*** may be done using the **aov()** function in R. The data are on the course website in the file **ex_11_17.txt**.

```
> my.anova <- aov(CastingHardness ~ Sand + CarbonFiber
                        + Sand:CarbonFiber, data = my.data)
> summary(my.anova)
```

Part *c* is asking for an interaction plot, which you can make in R by typing something like this:

```
> interaction.plot(x.factor = my.data$Sand,
                   trace.factor = my.data$CarbonFiber,
                   response = my.data$CastingHardness)
```

** **Problem 18** may be done using the **aov()** function in R, as above. Check the course website for the data (**ex_11_18.txt**). For **Part *c***, the factor level means can be obtained via something like:

```
>  aggregate(Yield ~ Formulation, data = my.data, FUN = mean)
>  aggregate(Yield ~ Speed, data = my.data, FUN = mean)
```

For **Part *d***, the fitted values and residuals are obtained via **my.anova$fitted.values** and **my.anova$residuals**, and for **Part *e***, the plot can be made using:

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

*** For **Problem 20**, check the course website for the data (**ex_11_20.txt**). For **Part *a***, the boxplots can be made by passing a *formula* to **boxplot()**, for example:

```
> boxplot(MPa ~ Adhesive + Condition + Adhesive:Condition, data =
my.data)
```

For **Part *b***, you may use the **aov()** function in R, as above. For **Part *c***, you can add a column named **Treatment** indicating the four treatments to your data frame:

```
> my.data$Treatment <- c("SBPDry", "SBPDry", … "OBPMoist")
```

and then carry out the **one-factor ANOVA**, followed by **Tukey's procedure**, using:

```
> my.anova <- aov(MPa ~ Treatment, data = my.data)
> TukeyHSD(my.anova)
```

Note that the result is the same as if you had performed Tukey's procedure on the **group means** in a **two-factor ANOVA** by typing:

```
> my.anova <- aov(MPa ~ Adhesive + Condition + Adhesive:Condition,
                  data = my.data)
> TukeyHSD(my.anova)
```

**** **Problem 31** may be done in R using **aov()**. The data are on the course website (**ex_11_31.txt**). Because there's only one observation per group, there aren't enough data to fit the full model *and* estimate σ, so you should **leave the three-factor interaction out** of the model:

```
> my.anova <- aov(NickelWt ~ Power + Speed + PasteThickness
                  + Power:Speed + Power:PasteThickness
                  + Speed:PasteThickness, data = my.data)
```

Note that the SSE and MSE for this model are the same as the SSAB and MSAB that you'd get if you fit the full model (try it!).

For **Part *d***, you can type (using **my.anova** that *doesn't* include the three-factor interaction):

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

then refer to the **$PasteThickness** part of the output.

Additional Problem

1. A three factor experiment is conducted to investigate the yields of three different varieties of rice grown under controlled greenhouse conditions. All three **varieties** (Factor B) were grown under two levels of **fertilizer** (Factor C) and two levels of **sunshine** (Factor A).

   The response variable is the rice yield, and there were four *replicates* of the experiment (i.e. four observations per cell). The data are shown below and are also in the file **rice.txt**.

Rice Variety 1 (j = 1)

	Low Fertilizer (k=1)	High Fertilizer (k=2)
Low Sunshine (i=1)	86	109
	115	87
	83	91
	70	97
High Sunshine (i=2)	78	99
	63	108
	91	116
	80	93

Rice Variety 2 (j=2)

	Low Fertilizer (k=1)	High Fertilizer (k=2)
Low Sunshine (i=1)	102	104
	86	114
	68	99
	85	113
High Sunshine (i=2)	92	85
	78	117
	83	99
	87	99

Rice Variety 3 (j = 3)

	Low Fertilizer (k=1)	High Fertilizer (k=2)
Low Sunshine (i=1)	75	99
	9	70
	68	80
	95	109
High Sunshine (i=2)	103	116
	115	125
	95	114
	118	111

Read the data into an R *data frame* named, say, **my.data** using **read.table()**.

i. Write out the **full** three-factor **ANOVA model** for the data. Be sure to **state any assumptions** associated with the random error term in the model.

ii. State the three sets of **hypotheses** for **main effects** that are tested by the three-factor ANOVA F tests in terms of the model parameters α_i, β_j, and δ_k.

iii. State the three sets of **hypotheses** for **two-factor interaction effects** that are tested by the three-factor ANOVA F tests in terms of the model parameters γ_{ij}^{AB}, γ_{ik}^{AC}, and γ_{jk}^{BC}.

iv. State the null and alternative **hypotheses** for **three-factor interaction effect** that are tested by the three-factor ANOVA *F* tests in terms of the model parameters γ_{ijk}^{ABC}.

v. Carry out the three-factor ANOVA and report the **ANOVA table**, using either:

```
> my.anova <- aov(Yield  ~ Sunshine + Variety + Fertilizer
                  + Sunshine:Variety + Sunshine:Fertilizer
                  + Variety:Fertilizer
                  + Sunshine:Variety:Fertilizer,
                   data = my.data)
```

or:

```
> my.anova <- aov(Yield  ~ Sunshine*Variety*Fertilizer,
                 data = my.data)
```

followed by:

```
> summary(my.anova)
```

vi. Based on the ANOVA *F* tests, answer the following questions:

a) Is the three-factor interaction effect statistically significant? Give an **F statistic** value and **p-value** to support your answer.

b) Recall that we only proceed with the test for a lower-order effect if it isn't involved in a significant higher-order interaction. Based on the result of the F test for the three-factor interaction, **is it reasonable to proceed** with the tests for two-factor interactions? Explain.

c) Is fertilizer more effective with some varieties of rice than others? Give an **F statistic** value and **p-value** to support your answer.

d) Does the effect of fertilizer differ depending on how much sunlight there is? Give an **F statistic** value and **p-value** to support your answer.

e) Does the amount of sunlight affect the yield of some rice varieties more than others? Give an **F statistic** value and **p-value** to support your answer.

f) We only proceed with the test for a lower-order effect if it isn't involved in a significant higher-order interaction. Based on the results of the F tests for the two-factor and three-factor interactions, **is it reasonable to proceed** with the tests for all three main effects? If not, which ones should we proceed with? Explain.

g) Does fertilizer have an effect on the rice yield? Give an **F statistic** value and **p-value** to support your answer.

h) Does variety of rice have ***any*** effect on the yield? Explain.

i) Does sunshine have ***any*** effect on the yield? Explain.

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

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

viii. Check the **constant standard deviation assumption** by plotting the residuals (**my.anova$residuals**) on the y axis versus the fitted (or predicted) values (**my.anova$fitted.values**) on the x axis. After making the plot, add a horizontal line to it at $y = 0$ by typing:

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