Tutorial Problems

Question 1

In a 3×2 factorial experiment on the control of *Agropyron repens* (quack-grass), Zick (1956) used 3 rates of spray (0, 4 and 8 pounds per acre) of maleic hydrazide in combination with 2 periods of delay (3 and 10 days) in cultivation after spraying. Fifty two days after spraying, the number of quack-grass shoots per square foot was recorded for each plot. It was believed reasonable to set up a standard linear model for the square roots of these numbers. The square roots are as follows: Periods of delay Rates of Spray

Period of delay (days)	Rate of Spray (lbs/acre)	Square roots of quack-grass shoots (/square ft)			
	0	15.7	14.6	16.5	14.7
3	4	9.8	14.6	11.9	12.4
	8	7.9	10.3	9.7	9.6
	0	18.0	17.4	15.1	14.4
10	4	13.6	10.6	11.8	13.3
	8	8.8	8.2	11.3	11.2

Consider the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

$i = 1, 2, 3, j = 1, 2, k = 1, 2, 3, 4$ where

- $\epsilon_{ijk} \sim NID(0, \sigma^2)$,
- α_i corresponds to the effect of Rate of spray i,
- β_j corresponds to the effect of Period j, and
- the $(\alpha\beta)_{ij}$ denote the Rate $\times$ Period interaction terms

(a) Organise the data given in the above table in a tidy form.

(b) Confirm that $y_{1\bullet\bullet} = 126.4, y_{2\bullet\bullet} = 98, y_{3\bullet\bullet} = 77, y_{\bullet 1\bullet} = 147.7, y_{\bullet 2\bullet} = 153.7$ and $y_{\bullet\bullet\bullet} = 301.4$.

(c) Given that the treatment sum of squares is given as 155.6533 and the sample variance of the response, s_y^2 is 8.505145, construct the analysis of variance table corresponding to the above model. Determine if the above model can be simplified. (Lecture 18-19)

(d) Now consider the one-way ANOVA model with factor `Rate` as

$$Y_{ijk} = \mu + \alpha_i + \epsilon_{ijk}, \quad \text{where } \epsilon_{ijk} \sim NID(0, \sigma^2),$$

to answer this question. Since the factor Rate is a numerical factor with 3 levels, Lecture 17 says that the TSS of this one-way ANOVA model is the same as the TSS of order (3-1) polynomial regression. Use a contrast with coefficient $c_i = n_i(x_i - \bar{x})$ to calculate the linear regression component of the Rate of spray sum of squares in the polynomial regression. Is a linear regression model appropriate in this case? (Lecture 17)

(e) Given that $S_{xy} = -197.6$ and $S_{xx} = 256$, estimate the parameters in

$$Y_{ijk} = \beta_0 + \beta_1 x_i + \epsilon_{ijk}, where \epsilon_{ijk} \sim NID(0, \sigma^2),$$

where x_i is the rates of spray. (Lecture 3 and assumed knowledge)

```
c(qf(0.05, 1, 18), qf(0.05, 2, 18), qf(0.05, 1, 20), qf(0.05, 2, 20))
```

```
[1] 0.004043292 0.051439739 0.004032045 0.051425070
```

```
c(qf(0.95, 1, 18), qf(0.95, 2, 18), qf(0.95, 1, 20), qf(0.95, 2, 20))
```

```
[1] 4.413873 3.554557 4.351244 3.492828
```

Question 2

Wilkinson (1954) reports the results of a study on the influence of time of bleeding and diethylstilbestrol (estrogen compound) on plasma phospholipid in lambs. Five lambs were assigned at random to each of the four treatment groups.

Totals of 5 values in each treatment group.

	Time (Factor A)		
Factor B	am	pm	Totals
No estrogen	66.39	182.67	249.06
Estrogen	96.80	139.06	235.86
Totals	163.19	321.73	484.92

The total sum of squares for the data set of 20 values is 1919.33.

(a) Construct the analysis of variance table for this factorial design, analyse the data and comment. (Lecture 18-19)

(b) Using the two-way ANOVA model, Construct a 95% confidence interval for the difference in mean plasma phospholipid levels for lambs receiving diethylstilbestrol and those not.

```
qf(c(0.05,0.95),1,16)
```

```
[1] 0.004057389 4.493998478
```

```
qt(0.975,16)
```

```
[1] 2.119905
```


Question 3

A completely randomized experiment was conducted to compare seven treatments for their effectiveness in reducing scab disease in potatoes. The field plan is shown below. The upper figure in each plot denotes the treatment, coded 1 to 7. The lower figure denotes an index of scabbiness of potatoes in that plot: 100 potatoes were randomly sampled from the plot, for each one the percentage of the surface area infected with scabs was assessed by eye and recorded, and the average of these 100 percentages was calculated to give the scabbiness index.

2 9	1 12	6 18	4 10	6 24	7 17	5 30	3 16
1 10	5 7	4 4	3 10	5 21	1 24	1 29	6 12
2 9	7 7	3 18	1 30	3 18	7 16	2 16	4 4
5 9	1 18	7 17	6 19	1 32	4 5	1 26	2 4

(a) What is the experimental unit, observational unit and treatment of the above experiment? How many experimental units, observational units and treatments are there? (Lecture 23)

(b) Suppose that y is a vector of response and trt is a factor that denotes the treatment. Given the below information only, complete the analysis of variance table below by filling out the ??s. (Lecture 14-15, 24)

```r
tapply(y, trt, sum)
```

```
## 1 2 3 4 5 6 7
## 181 38 62 23 67 73 57
```

```r
table(trt)
```

```
## trt
## 1 2 3 4 5 6 7
## 8 4 4 4 4 4 4
```

```r
sum(y)
```

```
## [1] 501
```

```r
length(y)
```

```
## [1] 32
```

```r
var(y)
```

```
## [1] 67.5877
##
## Response: y
##         Df   Sum Sq   Mean Sq   F value   Pr(>F)
## trt     ??     ??       ??        ??       ###
```

(c) Is there any evidence that the mean scabbiness is different according to different treatments? Justify your answer. (Lecture 14-15, 24)

```
qf(0.95, 6, 25)
```

```
[1] 2.49041
```

```
qf(0.95, 1, 30)
```

```
[1] 4.170877
```

(d) Estimate the mean scabbiness produced by each treatment. (Lecture 14-15, 24)

(e) What is the standard error of the differences between treatment 1 mean and the average of treatment 2 and 3 means? (Lecture 14-15, 24)

Computer Problems

Question 1

An anesthesiologist made a comparative study of the effects of acupuncture and codeine on postoperative dental pain in male subjects. The four treatments (`Tmt`) were

- A : placebo capsule (sugar) & placebo acupuncture (inactive points)
- B : codeine capsule & placebo acupuncture
- C : placebo capsule & acupuncture
- D : codeine capsule & acupuncture

Thirty-two subjects were grouped into eight blocks (`Tol`) of four according to an initial evaluation of their level of pain tolerance. The subjects in each block were then randomly assigned to the four treatments. Pain relief scores (`Pain`) were obtained for all subjects two hours after dental treatment. Data were collected on a double blind basis. The data is given in the file `pain.txt`.

(a) Assume you have to plan this experiment yourself. Generate an allocation list for a complete randomized block design with 8 blocks and four treatments. This part does not use the data but only its dimensions (sample size: 32; variables: block and treatment).

(b) Assume that after the experiment was run by the anesthesiologist, with a design that was similarly generated as what you just produced in (a), you are given the data in pain.txt. Try to understand the structure of the data by using `str`, `summary` or `skim`, i.e. think about if you have to change the variable type of any of the data columns?

(c) Obtain the residuals for the randomized block model and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings? Comment on any violation of underlying assumptions, otherwise confirm that assumptions seem reasonable.

(d) Produce an interaction plot. Comment on how parallel the plots are, also taking into account the variability of each calculated "treatment mean" (in the sense that a treatment mean is the average, $\bar{Y}_{ij\bullet}$, of all observations that received level i of factor A and level j of factor B.

(e) Obtain the analysis of variance table (without interactions).

(f) Calculate a 95% confidence interval for $\alpha_B???\alpha_C$ and use this to test if this pairwise contrast is zero.