STAT/ME 424 (Fall 2022) Midterm Solutions

This exam consists of four (4) problems worth a total of 100 points. Answer as many parts as you can. Except for Problems 3 and 4, you must show work to get full credit.

1. A completely randomized experiment was conducted to compare seven treatments for their effectiveness in reducing scab disease in potatoes. The field plan and results are shown in Table 1. In each plot, the upper number (*in italics*) denotes the treatment, coded *1–7*, and the lower number gives an index of scabbiness of the potatoes in the plot: 100 potatoes were randomly sampled from the plot; for each potato 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.

Table 1: Field plan and results. In each plot, the upper number (in *italics*) is the treatment code and lower number is the yield (scabbiness index).

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

(a) Complete the following ANOVA table. *(15 points)*

Source	SS	Df	MS
Treatment			
Error	1122.87		
Total			

```
%                Df  Sum Sq Mean Sq F value  Pr(>F)
% Treatment      6  972.34  162.06  3.6081 0.01026 *
% Residuals     25 1122.87   44.91
```

(b) Is there any evidence that the differences in mean scabbiness are due to differences in the treatments? Justify your answer with an appropriate hypothesis test at the 0.05 level of significance. *(5 points)*

```
% Test is significant
```

(c) Estimate the standard deviation of each estimated treatment mean. *(5 points)*

```
% SE(Treatment 1) = sqrt(44.91/8) = 2.369335
% SE(Other treatments) = sqrt(44.91/4) = 3.350746
```

(d) Estimate the standard deviations of differences between pairs of estimated treatment means. *(5 points)*

```
% SE(Treatment 1 vs other) = sqrt(44.91*(1/8 + 1/4)) = 4.103809
% SE(Other treatment differences) = sqrt(44.91*(1/4 + 1/4)) = 4.738671
```

(e) The experimenter wants to know if the normal distribution assumption is satisfied for his ANOVA analysis. He shows you Figure 1, a normal probability plot of the data. What can you say to him? *(5 points)*

```
% The plot should be for residuals, not y data
```

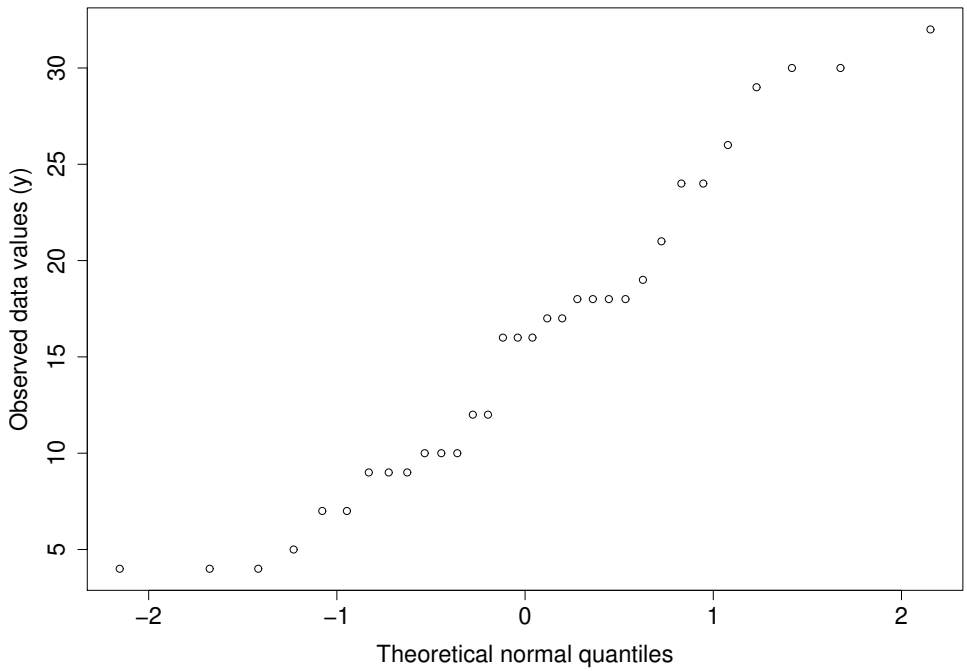

Figure 1: Normal probability plot of the data in Table 1

2. Answer each question briefly but clearly. *(5 points each)*

(a) State the two most important reasons for randomized assignment of treatments to subjects in a designed experiment.

```
% statistical basis for inference and removal of experimenter bias
```

(b) If an experiment is properly randomized, is it necessary to check the IIDN assumptions? Why or why not?

```
% yes if the F distribution is used
% no if randomization distribution is used
```

(c) At what stage in data analysis do we consider the use of transformations and what can they accomplish for us?

```
% If one or more of the model assumptions fail. Transformations
% enlarge the domain of application of linear methods.
```

(d) For what type of data is the family of power transformations y^λ inapplicable? Why?

```
% Negative valued-data
```

(e) What are the strengths and weaknesses of the Bonferroni method of simultaneous confidence intervals compared to the Scheffé method?

```
% Bonferroni is more general but conservative. Scheffe is
% restricted to contrasts.
```

(f) If the sample size of an experiment is doubled, how much do you expect the lengths of confidence intervals to shrink by?

```
% reciprocal of square root 2
```

(g) If the standard deviation σ in an experiment is halved (e.g., by using more precise equipment), how much do you expect the lengths of confidence intervals to shrink by?

```
% half
```

3. Which of the following statements (if stated by the manager in charge of a study) would lead you to choose a relatively large value of the significance level α in a hypothesis testing situation? No explanations needed. *(15 points)*

(a) "If the null hypothesis is true, we're in big trouble if we reach the wrong conclusion."

(b) "If the null hypothesis is false, we're in big trouble if we don't discover this fact."

(c) "A type II error is not particularly costly in this study."

(d) "A type I error is very costly in this study."

```
% (b)
```

4. Some researchers have conjectured that a stem-pitting disease in peach tree seedlings might be controlled with weed and soil treatment. An experiment was conducted to compare peach tree seedling growth with soil and weeds treated with one of two herbicides.

   In a field containing ten seedlings, five were randomly selected and assigned to receive herbicide A. The remainder received herbicide B. Soil and weeds for each seedling were treated with the appropriate herbicide and, at the end of the study period, the height (cm) was recorded for each seedling. The following results (sorted from largest to smallest) were obtained:

Herbicide A	87, 80, 80, 76, 73
Herbicide B	78, 77, 74, 68, 62

   A 90% confidence interval for the difference in the true mean heights for the two herbicides is found to be (0.3, 14.5). True or false (no explanations needed):

   (a) There is 0.9 probability that the difference in the true mean heights for the two herbicides is between 0.3 cm and 14.5 cm. *(5 points)*

   ```
   % False
   ```

   (b) The p-value for a test of the null hypothesis of equal means versus the alternative hypothesis of unequal means is greater than 0.1 because the interval does not contain 0. *(5 points)*

   ```
   % False
   ```

   (c) A 95% confidence interval would not include 0 either, because we would be even more confident of a difference between the two herbicides. *(5 points)*

   ```
   % False
   ```