MSDS UNIT 3 HW 3

PART I

1. Back in Unit 1 we considered a study in which 4 different fertilizers were tested by researchers for their yield (in mm of growth) on a local grass: Red Fescue (Just for fun .. this is a real mountain grass! http://fescue.com/info/creepingred.html#.WIu6rbYrKu4). To conduct their study researchers had enough money to run three replicates of each fertilizer. They knew the red fescue was a mountain grass so they went out to the mountains and carefully fertilized plots of land as you see in the diagram below. The data is contained in the ***growth3*** data set.



Given the information you have available, run a simple ANOVA to test for the effects (if any) of the fertilizers. **From a previous study, we know that the yields from each fertilizer are normally distributed with equal variances. For now assume that independence is not a concern here. You may assume the assumptions are met for all questions in this homework.**

|  |
| --- |
| Deliverables: 1. ANOVA table. 2. Means Plot (Interaction Plot from SAS) 2. Conclusion for the ANOVA. 3. Confidence intervals with multiple comparison corrections for **SIGNIFANT** differences (between Fertilizers). 4. SAS Code: proc glm or Proc mixed code. Your answer should fit in the given box. The ANOVA table produced indicates the F value is 2.83 and the p value is 0.11. As a result, we would fail to reject the null hypothesis that there is difference in growth between the fertilizers that is not attributed to chance. Because there was only 1 class variable (fertilizer), the Type 1 and Type 3 output is identical. There is no interaction as a result of this model. A closer look into the confidence intervals also confirms all of the means fell within their confidence intervals and none of them were significant.    Fit Plot for Growth by Fertilizer |

1. Now assume that you get to thinking about it and realize that we may really be looking at three different environments here: Sunny, Wetlands and Mostly Shady. This data has been recorded in the ***growth4*** data set. Conduct a similar analysis as in 1 but now account for the environment variable (ENV).



Deliverables: 1. ANOVA table. 2. Means Plot (Interaction Plot from SAS) 2. Conclusion for the ANOVA (only step 6 is required.) 3. The researchers were specifically interested in (a) Inference on the main effect of the fertilizer. (b) If Fertilizer 4 performed significantly better in one environment than another. (c) Which fertilizer will perform best in the wetlands. Answer each QOI (question of interest) with a confidence interval and a 1 sentence conclusion. Your answer should fit in the space below (on this page.) 4. SAS Code: Proc glm or proc mixed statement. Answer should fit in the given box.

|  |
| --- |
| Based on the ANOVA statistical output, there is sufficient evidence to suggest that the fertilizer has a significant effect (p value < 0.0001) on the growth of the grass. Fertilizer 4 performed better than all other fertilizers, with its closest being fertilizer 2 (confidence interval between -4.91 and -3.76). Furthermore, it appears Fertilizer 4 also performed better in Wetlands than other environments, with the closest environment being SUNNY (confidence intervals between -8.25 and -7.25).  Interaction Plot for Growth by Fertilizer |

1. Consider the model you used in problem 2. Inspect the Means Plot (Interaction Plot). Does it look like there will be a significant interaction? Explain by interpreting what an interaction is and then comparing that to what you see in the plot.

|  |
| --- |
| An interaction is an instance where the effects of one variable has on the mean response also relies on the other. This may be classified as *Categorical\_Variable\_A x Categorical\_Variable\_B* in a multiple regression model. In a means plot, this would be visible by seeing the means cross or *interact* with each other. When we look at the Means Plot, we can see that the lines stack on each other but do not overlap. Because the lines never overlap, this would confirm there is no interaction between Environment and Fertilizer. |

1. Fit the full model that includes the interaction term? What do you notice? Why? Discuss.

|  |
| --- |
| The output doesn’t show an F-statistic or p value for any of the variables. This could be  attributed to the fact that the degrees of freedom is coming out at zero for the error, which  would also explain the r-squared of 1. In other words, this model is a perfect fit and there are  no errors. |

Extra credit: Repeat number 2. With a regression model instead of a 2 way ANOVA. Show that the results are the same.

Deliverables: 1. Parameter Estimate Table, 2. Written conclusion showing that the estimates from the regression analysis is equivalent to the ANOVA analysis. 3. SAS proc glm or proc mixed code.

PART II

Exercise 13.15 :

Blood–Brain Barrier. Analyze the effect of the design variables—sacriﬁce time and treatment—on the log of the ratio of brain count to liver count in the data set described in Section 11.1.2 (ﬁle: case1102). (a) Ignore the covariates and use an analysis of variance procedure to ﬁt the data. Fit a model that includes interaction terms; plot the residuals versus the ﬁtted values. (b) Test whether there is an interactive effect of treatment and sacriﬁce time. What are the F-statistic, the degrees of freedom, and the p-value? (c) If there are no interactive effects, test whether there are main effects of treatment and sacriﬁce time. (d) Complete the analysis by describing the effects of treatment and sacriﬁce time, either by estimating the appropriate contrasts or by using a regression procedure with indicator variables to model treatment (one indicator) and sacriﬁce time (three indicators).

Based on the results below, there is insufficient evidence to support an interaction between Time and Treatment in this instance. The interaction plot confirms there is no interaction and the p value from the ANOVA table (p value = 0.93) also indicates we would fail to reject the hypothesis the interaction effect is other than zero. Looking at the residual v. predicted plot, we can see evidence of what appears to be stratification in the data results between treatments. Using Treatment NS and Time 72 as the base, it appears there is insufficient evidence to support the null hypothesis, therefore would reject the null. The p values for all time and treatment parameters returned significant results, which would confirm the differences between the treatment and time (when compared to 72 hours and NS) are not zero.







