Boston University Department of Mathematics and Statistics

MA 214

Applied Statistics

Lab Session 9: ANOVA

In this lab we will use the JMP software to explore ANOVA, the analysis of variance, and gain practice in first assessing the presence of an interaction effect in two-way ANOVA, and then in identifying the "best" treatments through the use of the Tukey multiple comparison procedure.

Preparation

This week we will be using a new data set called ratweight.jmp which is available on the Blackboard site under the Course Documents link. In this file, for a study conducted on different possible diets for rats, data are given for the diet type, diet amount, and weight gain (in grams) for 60 rats. SOURCE: Hoaglin, D., Mosteller, F., and Tukey, J. (1991). Fundamentals of Exploratory Analysis of Variance. Wiley, New York, page 100. http://www.itl.nist.gov/div898/wrkshops/stats_sci/ratfeed.dat

Variable Descriptions:

Columns	
x1	Diet amount (1 = high, 2 = low)
x2	Diet type (1 = beef, 2 = pork, 3 = cereal)
У	Rat weight gain (in grams)

Questions

- 1) Let us consider investigating the effects of the various treatments to the weight gain of the rats. Note that since we have two factors, "x1" and "x2", and that since "x1" has two levels and "x2" has three levels, we will have six total treatments. As a first step to determining whether there is a significant difference in the treatment means, let us fit a full factorial model for the data using JMP.
 - (a) Go to the "Analyze -> Fit Model" platform and place "y" in the "Y, Response" role, and then highlight both "x1" and "x2", and go to the "Macro" option and select "Full Factorial". You should now see both "x1",

"x2", and an interaction term "x1*x2" in the Construct Model Effects box. This time we will not select the "Minimal Report" option from the "Emphasis" command and instead keep the default "Effect Leverage" option. When you have finished setting up the model, click the "Run" button.

(b) Let's use the interaction plots as we have discussed in class to explore the possibility of an interaction between the diet type and the diet amount variables. To do this, go to the red triangle next to the "Response y" box, and select "Interaction Plots" from the "Factor Profiling" option. The plots will appear at the bottom left corner of the JMP output; note that you can drag the sides of the plot window to resize it. Based on what you see in the interaction plots and what we have discussed in class, do you believe that there could be a significant interaction effect between "x1" and "x2"? What else can you say about the effects of the two factors?

For the yx1 plot (weight gain vs diet amount), the lines start separately and then converge. This suggests that the effect of diet amount on weight gain may be different depending on the level of the other factor, but the differences become less pronounced as the levels of the other factor change.

For the yx2 plot (weight gain vs diet type), the lines cross each other. This is a clear indication of an interaction effect. The effect of diet type on weight gain depends on the level of the other factor (diet amount).

So it seems there could be a significant interaction effect between diet amount (x1) and diet type (x2). The effect of one factor on weight gain is not independent of the level of the other factor.

(c) Now scroll to the "Analysis of Variance" table and conduct a hypothesis test to determine whether or not there is a significant difference among the treatment means. Record the F-statistic for this test, its p-value, and your conclusion.

F test statistic: 4.3000

p-value: 0.0023

Conclusion: There is a significant difference among the treatment means. This means that at least one of the treatments (combinations of diet type and diet amount) has a significantly different effect on rat weight gain compared to the others.

(d) If you have concluded that there is a significant difference among the treatment means, then the natural follow-up question is: where are the significant differences? To answer this question, we must determine which effects, the two main effects of "x1" and "x2", or the interaction effect between them, are statistically significant. We will first focus on the interaction effect.

Scroll to the "Effect Tests" box and conduct a hypothesis test to determine whether there is a significant interaction effect between "x1" and "x2". Once again record the values of the F test statistic and p-value associated with this test, and use a significance level of 0.05 to make your conclusion. Does your conclusion here match your intuition from part (b)?

F test statistic: 2.7455

p-value: 0.0732

Conclusion: There is not a significant interaction effect between "x1" (diet amount) and "x2" (diet type) at the 0.05 significance level. This means that the effect of diet amount on rat weight gain does not depend on the type of diet, and vice versa, at least not to a degree that is statistically significant at this level.

This conclusion may not match the intuition from part (b) based on the interaction plots, which suggested a possible interaction effect. This discrepancy could be due to the fact that visual interpretations from plots are subjective and can sometimes be misleading, while the p-value provides a more objective measure of statistical significance. It's also possible that there is a true interaction effect, but it's not strong enough to be detected as statistically significant at the 0.05 level with this sample size.

2) Depending on your conclusion in part (d) above, it will be necessary to investigate either all of the treatments in the case where the interaction effect is significant, or only the treatments corresponding to the significant main effects in the case where the interaction effect is not significant.

(a) If you concluded that the interaction effect is significant, then go to the far-right part of the JMP output to the "x1*x2" Leverage Plot, and conduct the Tukey multiple comparison procedure by selecting "LSMeans Tukey HSD" from the red triangle. The results of the procedure will be given in a table labeled "LSMeans Differences Tukey HSD", and the significantly different treatment means will be highlighted in red. The results will be further summarized below the table in the usual way where the treatment means are ordered from greatest to least and different letters separate the groups which are significantly different.

Give a practical interpretation of the output of the Tukey multiple comparison procedure.

(b) On the other hand, if you concluded that the interaction effect is not significant, then return to the "Effect Tests" box on the left side of the JMP output and conduct hypothesis tests to determine which, if any, of the main effects are significant. For each of the significant main effects, repeat the steps outlined above and conduct the Tukey multiple comparison procedure on the treatments of each main effect by going to the appropriate Leverage Plot and selecting the "LSMeans Tukey HSD" option from the red triangle.

If you determine that a main effect is significant but you are unable to perform the Tukey multiple comparison procedure for it, then provide an explanation for why you are unable to do so. Summarize your conclusions in the space below and give a practical interpretation of the effects of the diet type and diet amount used in the study on the weight gain in rats.

Based on the results from part (d), the interaction effect between "x1" (diet amount) and "x2" (diet type) was not found to be statistically significant. Therefore, we should focus on the main effects of "x1" and "x2".

From the "Effect Tests" table, we can see the following:

For "x1" (diet amount), the F-statistic is 14.7666 and the p-value is 0.0003. Since the p-value is less than 0.05, the effect of diet amount on rat weight gain is statistically significant.

For "x2" (diet type), the F-statistic is 0.6211 and the p-value is 0.5411. Since the p-value is greater than 0.05, the effect of diet type on rat weight gain is not statistically significant. In this case, the factor "x1" (diet amount) has only two levels: high (1) and low (2). Since there are only two levels, a multiple comparison procedure like Tukey's is not necessary. The significant main effect of "x1" already tells us that there is a significant difference in rat weight gain between the high and low diet amounts.

In terms of practical interpretation, the significant main effect of "x1" (diet amount) suggests that the amount of diet given to the rats has a significant effect on their weight gain. Depending on the direction of the effect (which could be determined by comparing the means of weight gain for high and low diet amounts), it could mean that rats given a high amount of diet gain more weight (or less weight) compared to rats given a low amount of diet.

The non-significant main effect of "x2" (diet type) suggests that the type of diet (beef, pork, or cereal) does not have a significant effect on rat weight gain, at least not in this study with this sample size. This means that, regardless of the type of diet the rats were given, there was no significant difference in their weight gain.

The non-significant interaction effect between "x1" and "x2" suggests that the effect of diet amount on rat weight gain does not depend on the type of diet, and vice versa. This means that the impact of diet amount on weight gain is consistent across different types of diets, and the impact of diet type on weight gain is consistent across different amounts of diet.

In conclusion, based on the results of this study, diet amount appears to have a significant effect on rat weight gain, while diet type does not. Furthermore, the effects of diet amount and diet type on weight gain appear to be independent of each other.

3) a. Now let's go back to part 1(d), the test for interaction. At this time, carry out the test for interaction at 10% level of significance. What is your conclusion? What does this say about the level of interaction? Is it consistent with your observation of the interaction plots? If you wish, you can change the default alpha level from the Fit Model dialogue box by choosing Set Alpha =0.1 from the drop-down menu at the top right-hand corner.

In part 1(d), the p-value for the interaction effect between "x1" (diet amount) and "x2" (diet type) was 0.0732. This was not significant at the 0.05 level, leading to the conclusion that there was not a significant interaction effect at that level.

However, if we change the significance level to 0.10, the interpretation of the

p-value changes. A p-value of 0.0732 is less than 0.10, which means that the interaction effect is statistically significant at the 0.10 level.

Conclusion: There is a significant interaction effect between "x1" (diet amount) and "x2" (diet type) at the 0.10 significance level. This means that the effect of diet amount on rat weight gain does depend on the type of diet, and vice versa, to a degree that is statistically significant at this level.

This conclusion might be more consistent with the intuition from part (b) based on the interaction plots, which suggested a possible interaction effect. The discrepancy between this conclusion and the previous one at the 0.05 level highlights the importance of the chosen significance level in statistical hypothesis testing. By choosing a higher significance level (0.10 instead of 0.05), we are more willing to accept the risk of a Type I error, which is rejecting a true null hypothesis, to be more sensitive to detecting an effect if it exists.

b. Repeat part (2) based on the answer that you got in 3(a) and write your conclusion.

Given the conclusion from part 3(a) that the interaction effect between "x1" (diet amount) and "x2" (diet type) is significant at the 0.10 level, we would need to investigate all of the treatments, as the interaction effect is significant.

(a) However, after conducting the Tukey HSD test, no significant differences were found between the treatment groups. This suggests that while the interaction effect between diet amount and diet type is statistically significant, the differences between the specific combinations of diet amount and diet type are not large enough to be statistically significant when adjusted for multiple comparisons.

In practical terms, this means that while the effect of diet amount on rat weight gain does depend on the type of diet, and vice versa, the specific combinations of diet amount and diet type do not lead to significantly different weight gains in rats. This could be due to a variety of reasons, such as the inherent variability in rat weight gain, the specific levels of diet amount and diet type used in the study, or other factors not accounted for in the study.

(b) Since the interaction effect is significant, it's not necessary to test the main effects separately. The significant interaction effect indicates that the effect of diet amount on rat weight gain depends on the type of diet, and vice versa. Therefore, the effects of the diet amount and diet type cannot be considered independently of each other.

In terms of practical interpretation, the significant interaction effect suggests that the combination of diet amount and diet type has a significant effect on rat weight gain. For example, it could be that a high amount of a beef diet has a different effect on weight gain than a high amount of a pork or cereal diet. Similarly, the effect of a low amount of diet

could depend on whether the diet is beef, pork, or cereal.

In conclusion, the results of this study suggest that while the interaction between diet amount and diet type is statistically significant, the specific combinations of diet amount and diet type do not have significantly different effects on rat weight gain. This highlights the complexity of the effects of diet on weight gain and the importance of considering interaction effects in the analysis.