**HOMEWORK 9**

*“After the ANOVA”, proportions, and odds*

Reading: The “After ANOVA” material is in *STAT2*, Sections 5.6, 5.7, 8.2, and 8.3. The material on proportions and odds is in OpenIntroStats, section 6.1. The Chi-square test material is in OpenIntroStats, section 6.4 and the 2nd half of *STAT2*, section 11,4

* Round all numbers to 3 decimal places unless otherwise specified.

Part 1

Fantasy baseball – problem 5.66 with new questions. Read the study description given in problem 5.66. The data are in FantasyBaseball.jmp. Our analysis will focus on **Time**, the time required for a player to make a choice. Ignore the round variable in the data set.

The first few questions concern the need for and choice of a transformation. Start by calculating each player’s mean and standard deviation of Time. This should create a new JMP dataset with 8 rows of data.

1. Use the ratio of standard deviations criterion to decide whether or not there is a concern about unequal variances. *104.166/25.830 > 2. False*

2. log (natural log, as always) transform each player’s mean and standard deviation, then regress Y = log sd on X = log mean.

Report the estimated regression slope: *0.669*

3. This estimated slope is most consistent with which of these commonly used transformations?

No transformation; Slope = 0

Square-root transformation; Slope = 0.5

Log transformation; Slope = 1

Reciprocal transformation; Slope = 2

Note: make sure you can explain your choice.

**Use the log(Y) transformation for subsequent questions, no matter how you answered question 3.**

The rest of the questions concern “after the ANOVA” evaluations of differences between players. The data were collected because the person organizing the game suspected that some players were markedly slower or faster than others. Fit the one-way ANOVA using log(Time) as the response variable. The p-value for the F test is really small, < 0.0001; you don’t need to report anything about that ANOVA. We focus on “after the ANOVA”.

The organizer had two specific a-priori questions about players and groups of players.

The organizer suspected that DR was significantly slower than MF. Estimate the difference in log time between DR and MF, as DR – MF.

4. Report that estimated difference: *.544*

5. And report the appropriate p-value for the test of no difference between DR and MF:

(Remember, this is an a-priori comparison) *.0442*

The organizer knows that DR, DJ, and RL are the three newest players of Fantasy Baseball and AR and BK are the two most experienced players. That suggests comparing the average of DR, DJ, and RL to the average of AR and BK. Estimate that difference, as newest players – most experienced.

6. Report that estimated difference: *-.352*

7. And report the appropriate p-value for the test of no difference between the two groups:

(Remember, this is an a-priori comparison) *.0439*

We now repeat the analyses ignoring any a-priori questions. The organizer simply wants to know “which players differed from which others?”. Use the appropriate analysis to evaluate all pairwise differences, when there is no a-priori question. If you ask “which players differed from which others?”:

8. is DR significantly different from MF? The p-value for that comparison is < 0.05 (T/F)

*P > .05, F*

9. is DR significantly different from AR? The p-value for that comparison is < 0.05 (T/F)

*P > .05, F*

Part 2

Pines – Problem 11.22 with new questions. The data are in pines600.jmp. The text of problem 11.22 provides a description of the data set. We will use **deer97** as the response variable. This has the value of 1 when the tree was deer browsed by 1997 and 0 if not. The pine trees were planted at either a 10 foot spacing (**spacing**  = 10) or 15 foot spacing (**spacing**  = 15).

10. Calculate and report the overall probability that a tree is deer browsed: *.09*

11. Removed from this week’s HW

12. Calculate and report the probability that a tree planted at 10 foot spacing is deer browsed: *.0633*

13. and the probability that a tree planted at 15 foot spacing is deer browsed: *.117*

14. Removed from this week’s HW

Calculate and report the standard errors of the probabilities that a tree is deer browsed for the 10 foot and the 15 foot spacings.

15. 10 foot spacing: *.0141*

16. 15 foot spacing: *.0185*

17. You should have found that the standard error was smaller for the 10 foot spacing. Check all correct reasons for why it is smaller.

10 foot is shorter than 15 foot

The sample size is larger for the 10 foot spacing

The probability of deer browse is smaller for the 10 foot spacing

No reason; it just happened

18. Calculate the odds that a tree planted at 10 foot spacing is deer browsed: *14.798*

19. Report the odds ratio calculated by JMP Fit Y by X / Odds ratio: *1.953*

20.The odds ratio reported by JMP is based on the odds of deer browse, but JMP Fit Y by X / Odds ratio doesn’t tell you which is the numerator and which is the denominator. Is the odds ratio that JMP reports calculated as :

the odds in the 15 foot spacing divided by the odds in the 10 foot spacing

the odds in the 10 foot spacing divided by the odds in the 15 foot spacing

21. What would be the value of the odds ratio if both spacings had the same probability of deer browse? *1*

Have JMP calculate the 95% confidence interval for the odds ratio in question 19. Report:

22. the lower CI bound: 1.090

23. the upper CI bound: 3.500

24. Use the 95% confidence interval for the odds ratio to decide on the approximate p-value for the test of the null hypothesis that the two spacings have the same probability of deer browse. Is that p-value:

less than 5% (i.e. there is evidence of a difference in probabilities)

equal to 5%

more than 5% (i.e. there is no evidence of a difference in probabilities)