Ginoza, Jin, Lovejoy

1. Introduction: briefly give meaning to your project – what is the background and why is this project important?

We are taking data from the Equity in Athletics Survey, sample year 2017-2018, from the U.S. Department of Education Office of Postsecondary Education and analyzing salaries of head coaches from 4-year colleges and institutions.

Our analysis will focus on differences between salaries of head coaches at public vs. private institutions; salaries of male head coaches vs. female head coaches; salaries of head coaches from various division classifications; and ….. *[NOTE: I’m just listing out things we’ve talked about, not necessarily keeping them all]*

2. Hypotheses: clearly state the hypotheses being tested.

a.

b.

c. : There is not an interaction between Classification and Gender.

: There is an interaction between Classification and Gender.

d. (where 1, 2, …, 18 represent the 18 different division classifications (e.g. NCAA Division I, NCAA Division II, etc.)

e. Post-hoc: Pairwise comparisons

for some

f. Examine the relation between head coach salary (y, in thousands of dollars) and the number of athletes. Test the significance of the linear regression

g.

3. Methods: state and explain the methodology used to test the hypotheses.

a. Test for , F-test

b. Wilcoxon Rank Sum Test for , independent samples, and/or

*Note: Results from a) show that there is significant evidence to suggest a difference in variances, so we proceed with the non-parametric test here.*

c. Two-way ANOVA with classification and gender as predictors for head coach salary to check for an interaction.

d. A BFL will be run to check the constant variance assumption. If the normality assumptions are violated, a Kruskal-Wallis will be run between the classifications.

e. If main effects were detected, either a Tukey’s W pairwise comparison or the Kruskal-Wallis post-hoc procedure will be used to determine which classifications are different from each other.

e. t-test for the coefficient

f. Lack of fit analysis

4. Results: describe the data as well as analysis results (order of results should correspond to hypotheses listed.

a. There is sufficient evidence to suggest a difference between the variances of salaries from head coaches of public institutions vs. private institutions.

b.

c. Do the residuals need to be normal even if there are interactions? Does Kruskal-Wallis care about interactions? I think we didn’t learn a nonparametric alternative for a two-way ANOVA.

d.

e.

5. Conclusion: restate the results in terms of the larger picture; state limitations and opportunities for future research.

General question:

The data is given as: school, classification, sector, number of male participants, number of female participants, average head coach salary males, average head coach salary females. Should we separate this out into two separate cases per school, one row for males data and one row for females data and add a variable named gender?

And in that case, do we always need to include or adjust for gender, eg. In the public vs. private?

Separately, the head coach data is reported as an average. We do have the number of head coaches, so we could presumably weight the data appropriately, but for the sake of this project, is it okay to use just the average?