

# HW 5

Due: Thursday, October 24th by 11:59 pm

## Instructions

- Create your own .Rmd file called last\_first\_hw5.Rmd, replacing “last” and “first” with your first and last name.
  - **In the setup chunk, make sure to globally set echo = TRUE, message = FALSE, warning = FALSE.**
  - Type up your solutions to the following questions. Do NOT include the questions in your document, only your solutions.
  - Upload a pdf/doc AND Rmd.
- 1) The survey dataset from the MASS package contains survey data from a group of students from the University of Adelaide, including variables such as gender, smoking status, and exercise habits.

To access the dataset you will need to install the MASS package.

```
Unset
install.packages("MASS")
library(MASS)

data("survey")
```

- a) Create a new variable in the dataset called no\_smoke that is a binary factor variable to indicate whether or not a student is a regular smoker or not.
- b) Using the dataset, conduct a five-step hypothesis test with a p-value to determine if there is a difference in the proportion of non-smokers for females compared to males. Show your work.
- 2) The next dataset contains auction data from Ebay for the game Mario Kart for the Nintendo Wii. This data was collected in early October 2009 and is available in the openintro package. Using a confidence interval (instead of a p-value), determine if there is a significant difference in the price of new versus used Mario Kart games for Nintendo Wii's.

```
Unset
library(openintro)
data("markiokart")
```

- a) Write the null and alternative hypotheses in symbols AND in words.
- b) Check both of the necessary conditions to perform the hypothesis test.

- c) You will see in part (b) that the normality condition is not satisfied. Investigate the outliers. Do you think we should remove them from the dataset? Explain why and comment on the implications of removing data in this scenario.
  - d) Regardless of your decision in part (c), remove the outliers. Compute a confidence interval.
  - e) Make a decision based on your confidence interval and write a conclusion in the context of the problem.
- 3) In a clinical study, researchers are comparing the effectiveness of two vaccines in preventing a viral infection. In a trial of 1,200 participants, 200 out of 600 who received Vaccine A were infected, while 150 out of 600 who received Vaccine B were infected.
  - a) Perform a 5-step hypothesis test for the difference in proportions to determine if there is a significant difference in infection rates between the two vaccines at the **1% significance level**.
  - b) Discuss statistical significance versus practice importance in this example based on your results in part (a).
  - c) Imagine the trial only included 48 participants in each group. If the proportion of patients who were infected in each vaccine group were the same, does the result of the hypothesis test change? Prove your result. (You do not need to perform a full hypothesis test)
  - d) Evaluate the impact of small differences in sample proportions when large sample sizes are involved versus small sample sizes. Use your above answers to justify your conclusion.
- 4) Was there anything you found difficult with this homework? What topics (if any) do you feel you still need more work on?
- 5) Give yourself a rating for this assignment using the EMRN rubric.

E - Excellent; M - Meeting expectations; R - Revision needed; N - Not accessible