

STAT 477/577 - Homework Assignment 5

Due Tuesday, March 8, by 11:59 PM

General homework guidelines: All homework assignments should be submitted using Canvas. Please submit your answers separately, for each of the problems, as set-up in the Canvas submission portal. You are allowed to either type in your answers, as well as submit graphs directly, within each submission item. You are also allowed to submit a scanned copy of your answer, as long as the answers are submitted separately for each question, as instructed. You can either scan your answers (ask if you don't have access to a scanner or do not know how to use your phone to do so), or just submit a picture of your answer. Please note that if we can't read your answer, we won't be able to award any (partial) credit.

You have one attempt to submit your answers. If technical issues appear and your submission portal has closed for some reason, please email Prof. Caragea explaining the situation and requesting permission to resubmit. Please note that such requests must be made before the deadline.

For full credit, please make sure you submit your HW by the deadline. A late submission is possible, with a 20% penalty, as long as it is turned in before the end of the day on Thursday following the due day (for this HW, it is March 10-th). No submissions will be accepted past this date.

Homework problems.

1. [8pts.] On the night of April 14, 1912, the luxury liner *RMS Titanic* hit an iceberg and sank in the North Atlantic Ocean. In the popular movie about this disaster, first class passengers (i.e. those with the most expensive tickets) appeared to be able to get to the life boats, while third class passengers (i.e. those with the less expensive tickets) were kept away. Is there truth to this appearance? Was the proportion of passengers rescued different for each class of ticket? The data are located in the file **Titanic.csv**.
 - (a) Use R to create a contingency table and mosaic plot for the two variables and write a description of their relationship.
 - (b) Use R to conduct a test of hypothesis to determine if the proportion of passengers rescued was the same across all groups. Make sure to include the null and alternative hypotheses, test statistic, p -value, and conclusion.
 - (c) Conduct all the pairwise hypothesis tests for the proportion of passengers rescued for the four groups. Which groups appear to have a significantly different proportion rescued? Please include a connected letters report with your answer (as we did in the notes).
 - (d) Was the perception from the movie correct: the proportion of passengers rescued related to class of ticket? Explain your answer.
2. [8pts.] In 1996, in the General Social Survey of 1,895 adults in the United States conducted by the National Opinion Research Center, respondents were asked about their attitudes towards premarital sex. The question asked was **When is premarital sex wrong?** and the possible answers were **Always Wrong, Almost Always Wrong, Sometimes Wrong, Not Wrong at All**. People's attitudes about social behaviors tend to be related to other more general background variables about the individual. Among other questions, respondents were asked about one such variable, their religious affiliation. Possible answers were **Catholic, Protestant, Jewish, Other, None**. The data are located in the file **GSS.csv**.
 - (a) Use R to create a contingency table and mosaic plot for the two variables and write a description of their relationship.
 - (b) Use R to calculate the conditional distribution of attitude towards premarital sex given religious affiliation is Catholic.
 - (c) Use R to calculate the conditional distribution of attitude towards premarital sex given religious affiliation is Protestant.
 - (d) Use R to conduct a test of hypothesis to determine if attitude towards premarital sex is the same for all five groups.

- (e) In conducting the hypothesis test, at least one of the cells has an expected value less than 5. Explain how you would collapse categories in order to make the expected values for all cells greater than or equal to 5.
3. [9pts.] Aspirin used to be thought of only as a pill to help with pain relief. However, in recent years, researchers have been looking at the effectiveness of daily aspirin therapy in preventing heart disease, Alzheimer's disease and cancer. The Physicians Health Study Research Group at Harvard Medical School conducted a five-year randomized experiment about whether regular aspirin intake reduces deaths from heart disease. Study participants were randomly assigned to one of two groups: taking an aspirin tablet every other day or taking a placebo tablet every other day. After five years, the number of heart attacks in each group were determined. The data are located in the file **aspirin.csv**.
- (a) Use R to create a contingency table and mosaic plot for the two variables and write a description of their relationship.
 - (b) Use the output from part (a) to calculate the sample proportion of heart attacks for the two groups.
 - (c) Use R to conduct a test of hypothesis to determine if the proportion of physicians having a heart attack is lower in the aspirin group than in the placebo group.
 - (d) Calculate a 95% confidence interval for the difference in the proportion of physicians having heart attacks between the aspirin group and the placebo group. Interpret this confidence interval.
 - (e) Estimate the relative risk of having a heart attack when taking aspirin therapy versus a placebo. Find a 95% confidence interval for the relative risk and interpret this confidence interval.
 - (f) Estimate the odds ratio of having a heart attack when taking aspirin therapy versus a placebo. Find a 95% confidence interval for the odds ratio and interpret this confidence interval.