STAT 477/577 - Homework Assignment 7
Due Thursday, March 29, 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 of March 29. 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 31-st). No submissions will be accepted past this date.

Homework problems.

1. In lecture, we discussed results from one of the surveys in the baby pictures study that looked at the effect of Dr. Nettleton's son wearing a hat in the picture. In the problem below, we will analyze data from another survey in this study. In this survey, judges (students) were asked to identify which baby was Dr. Froelich's twice, once when she was pictured as an adult and and then when she was pictured as a baby. Similar to the survey we studied in class, judges were not allowed to revisit their response to the previous question. The pictures used in the study are located in the file **babypictures.pdf**. The babies selected by the 123 judges on the two questions is given in the table below. You will recall that Baby C is the correct baby in the analysis.

Baby Selected With Adult Parent Picture	Baby Selected With Baby Parent Picture				Total
	A	B	C	D	
A	7	1	4	0	12
B	18	12	8	5	43
C	17	7	22	8	54
D	3	2	2	7	14
Total	45	22	36	20	123

   (a) Using McNemar's test, determine whether there is a difference in the proportion of correct responses between the two questions. Do the judges perform differently based on which picture is given? *Hint: Collapse the table above to a 2 x 2 table in order to answer the question.*

   (b) Obtain a 95% confidence interval for the difference in the proportion of correct responses between the two questions. Interpret this confidence interval in the context of the problem.

   (c) Using the extension of McNemar's test, conduct a test of marginal homogeneity for the baby selected. Clearly write down the null and alternative hypotheses, the value of the test statistic and the p-value (which you should obtain using R), and a conclusion within the context of the problem.

   (d) Describe the differences you find in the marginal distributions between the baby selected when the parent was pictured as an adult versus when the parent was pictured as a baby.

2. In educational assessment, open-ended questions (called free-response questions) tend to provide more information about learning than multiple choice questions. However, these open-ended questions are more difficult and time consuming to grade than multiple choice questions. In an on-going project, researchers at a large university are studying the accuracy of computer-scored open-ended responses by comparing them to person-scored responses. In this example, the computer and a person scored the same 1,011 student responses to the same open-ended question. Each question was scored as either a 1 = minimal understanding of concept, 2 = moderate understanding of concept, or 3 = full understanding of concept. The table below gives the cross-classification table of scores and the full data set with all 1,011 rows can be found in the file **Scores.csv**.

Person-scored	Computer-scored		
	1	2	3
1	439	26	20
2	48	113	33
3	38	29	265

(a) Calculate Cohen's kappa for these data.

(b) Calculate weighted Cohen's kappa using the squared weight function for these data.

(c) Based on your answers to part (a) and (b), does there seem to be agreement between the computer scored and person scored responses? Explain your answer.

3. During the 1970's, 80's and 90's, Gene Siskel and Roger Ebert worked as film critics for the Chicago Tribune and Chicago Sun Times, respectively. In their syndicated TV show, *At the Movies*, they presented their reviews of recently released movies using a Thumbs Up/Thumbs Down system. This system also allowed them to give a mixed review. Their reviews for 160 movies from April 1995 through September 1996 are given in the data file **movies.csv**.

(a) Use R to calculate the contingency table for Siskel and Ebert's reviews.

(b) On what proportion of these movies did Siskel and Ebert agree on their review?

(c) Calculate Cohen's kappa for these data.

(d) Calculate weighted Cohen's kappa for these data using the squared weight function.

(e) Based on your answers to parts (c) and (d), do Siskel and Ebert seem to agree on their ratings for these movies? Explain your answer.