S520 Final Exam: Takehome

Upload your typed answers as a PDF file or Word document through the Assignments tab on Canvas by 11:59pm, Tuesday, April 30th, 2019

<u>Instructions and warnings:</u>

- **You may consult books and use computers, but may not discuss the problems with anyone.**

- Write explanations for all your answers. Answers alone will not get credit. For questions where you use R, you must give R code, but the code alone is not a sufficient explanation. You can use `lm` directly if you want, just make sure you understand what the outputs tell you.

- Conclusions should not just be reject or do not reject; they should be substantive conclusions about the issues under study.

<u>Example:</u>

- "We reject the null hypothesis": **NOT MANY POINTS**

- "The P-value is 0.005. This means the data gives strong evidence that three-toed sloths have more toes than two-toed sloths.: **LOTS OF POINTS**

1. (6 points) The number of misprints on 200 randomly selected pages from a newspaper within a year were recorded:

Number of misprints per page	0	1	2	3	4	5	6	7	8	9	10
Frequency	9	23	40	30	31	26	19	10	5	4	3

Does the number of misprints per page follow a Poisson distribution? Show your steps to perform an appropriate test of the null hypothesis that the data follow a Poisson distribution. Report the test statistic, P-value and state your conclusion.

2. The file `examanxiety.txt` on Canvas contains information on a number of variables measured on a sample of 103 students taking a math exam:

   - `Code`: a label for the individual in the sample (not scientically interesting.)
   - `Revise`: hours spent revising for the math exam.
   - `Exam`: score on a math exam on a scale from 0 to 100.
   - `Anxiety`: "math anxiety" on a scale from 0 to 100 (100 is most anxious.)
   - `Gender`: female or male.

   Say your data `examanxiety.txt` is saved in D drive, then you can read and define the data in R as follows:

   `exam = read.table("D:/examanxiety.txt", header=TRUE)`

   Assume the data is a random sample from a larger population of students.

   (a) (3 points) Let anxiety be your x-variable and exam score be your y-variable. Assume a simple linear regression is valid. Find the regression line to predict exam score from anxiety.

   (b) (3 points) Carefully explain what your regression line means (interpretation for the intercept, slope, and the $\hat{y}$), do not just paste R output.