STAT-S631 – Exam 1 – Take-home

Due October 13, 2017 at 5pm

Intructions:

- Please start your solutions with the following statement: "On my honor, I have not had any form of communication about this exam with any other individual (including other students, teaching assistants, instructors, etc.)" and sign by hand (if work was done partially or completely by hand and scanned) or typing your name (if work was done all typed).
- Use the file `takehome1.txt`. These data contains the following variables:
  - `population`: Total population in thousands (e.g. 3200 is 3.2 million).
  - `nonwhite`: Percent nonwhite population.
  - `density`: Population per square mile.
  - `crime`: Crime rate per 100,000.
- While in order to get full credit you need the appropriate output/results, you need to pay special attention to explanations/interpretations, when required. Correct output/results with no interpretation would receive zero points, and incomplete or incorrect explanations/interpretations will also be penalized.
- The exam does not have a restriction of space, but you are strongly encouraged to have succinct answers. Lengthy answers are neither required nor desired (and they are prone to contain incorrect statements that would be penalized).
- For each question, attach all the syntax used and any relevant output.
- Submit your solutions to Canvas as a single PDF file no later than 5pm on Friday, October 13th. Do not wait until the very last minute since late submissions will not be graded.

Questions

1. (2 point) What is the probability that a randomly selected city has a crime rate higher than 3200 (per 100000 people)? Is the crime rate normally distributed? Explain.
2. (6 points) Simple linear regression: Regress `crime` on `population`. Use `log` transformation on your regressor(s) if you deem appropriate.
   a. (3 points) Describe visually the relationship between your response and your regressor(s). Do you find any clear violations to the linear model assumptions? Explain.
   b. (3 points) Interpret the estimate $\hat{\beta}_1$, explain the underlying hypothesis test assumed in the `R` output for β_1, and interpret the corresponding output results.
3. (8 points) Multiple regression: Regress `crime` on all three predictors. Use `log` transformation on your regressor(s) if you deem appropriate.
   b. (3 points) Interpret the estimate $\hat{\beta}_1$, explain the underlying hypothesis test assumed in the `R` output for β_1, and interpret the corresponding output results.
   c. (3 points) Compare the effect that `population` has on `crime` between the models in part 2 and part 3 and explain why this effect changes.
   d. (2 points) Compare R^2 from parts 2 and 3. What can you conclude?
4. (4 points) In your opinion, and only based on this data set and the theory covered in class so far, what would be the most adequate linear model used to explain changes in crime. Using this model, answer the following questions.
   a. (2 points) Why is this your chosen model? Explain.
   b. (1 points) Obtain and interpret a 98% confidence interval for β_1.
   c. (1 points) If I want to be 99% confident, what could be the crime rate for a city with a population of 1.15 million people?