STAT-S 352 Problem Set 5

Upload your answers through the Assignments tab on Canvas by 11:59 pm,

Monday, September 29, 2025.

Important Note: Answer all questions and include R code when necessary. In general in this course, give explanations and/or working for all answers unless otherwise stated. Show your work for full credit.

Reminder: As a student at IU, you are expected to uphold and maintain professional and academic honesty and integrity. Academic integrity violations include: cheating, fabrication, plagiarism, interference, violation of course rules, and facilitating academic dishonesty. When you submit an assignment with your name on it, you are signifying that the work contained therein is yours, unless otherwise cited or referenced. Any ideas or materials taken from another source must be fully acknowledged.

1. The data in the file `expo.txt` contains data from an Exponential distribution. I know this because I generated it. However, I've inconveniently forgotten what rate parameter λ I used.

   It may be useful to recall that an Exponential(λ) distribution has expected value $1/\lambda$ and variance $1/\lambda^2$.

   (a) By drawing graphs in R, verify that it's plausible that the data in `expo.txt` comes from an Exponential distribution.

   (b) In R, write the negative log-likelihood as a function of λ.

   (c) In R, plot the negative log-likelihood function for values of λ between 0.2 and 0.5. **Estimate** the MLE of λ from the plot. Write down your estimation and explanation.

   (d) **Find** the MLE of λ theoretically (by writing a notation and then computing it in R) and numerically in R (by using the optimization function you have learned from this class).

   (e) Use R to find the Method of Moments (MOM) estimate of λ.

   (f) In R, compute a 95% Wald confidence interval for λ. Show work.

   (g) Find a 95% profile likelihood confidence interval for λ in R. Is it close to the Wald confidence interval you found? Why or why not?

   Note: This question was modified from one of last year's midterm exam questions.

2. The file `mensweights.txt` contains the weights (in pounds) of a representative sample of 3654 American adult men. Your job is to model these weights.

   (a) In R, draw a histogram of the weights with bars of width 10 pounds (e.g with breaks 100, 110, 120, ..., 500.)

   (b) In R, fit a probability distribution to the data that captures the important features of the data (for example, the data is right-skewed, so your distribution should be right-skewed.) Which distribution? That's for you to decide. Estimate the parameters of your distribution using maximum likelihood estimation method if that's possible or the method of moments if MLE is not possible.

   (c) In R, simulate a sample of size 3654 from your distribution. Draw a histogram of your simulated data, again with bars of width 10. (You might want to fix your random seed so that you get the same results every time you run your code: e.g. put the line `set.seed(352)` just before you generate your random numbers.) Describe any similarities and differences you see between the distribution of the real data and the distribution of your simulated data.

   (d) In R, compute the sample mean of the real data and the sample mean of your simulated data. Are they close? (If not, you might need a better probability model.)

   (e) In R, compute the standard deviation (SD) of the real data and the standard deviation (SD) of your simulated data. Are they close? (Note they may not be super-close because the SDs can be noisy, but they shouldn't be completely different.)