STAT-S 352 Problem Set 7

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

Tuesday, October 21, 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 file `womensheights.txt` contains the heights (in cm) of what is supposed to be a representative sample of 3766 American adult women from the 2009 NHANES survey done by the CDC. The Normal distribution is usually a good model for heights; we'll parameterize this model using the mean μ and variance σ^2. Review the MLE Normality Theorem, and the Negative binomial examples (Chapter 3 slides, pp. 32-39) and Week 8 R files. Answer the following questions with the <u>computational approach</u>, i.e., using `optim()` for numerical optimization.

    (a) Find $\hat{\mu}$ and $\hat{\sigma}^2$, the MLEs for μ and σ^2, from the output of `optim()`.

    (b) Compute the estimated variance matrix for the MLEs from the output of `optim()`.

    (c) Compute the estimated standard error ($\hat{se}$) of the MLEs.

    (d) Compute a 99% Wald confidence interval for μ, the average height of American adult women, using the MLE and $\hat{se}$ of the MLE from (a) and (c).

    (e) Compute a 99% Wald confidence interval for σ^2, the variance of American adult women's weights, using the MLE and $\hat{se}$ of the MLE from (a) and (c).

    (f) How would you find a 99% confidence interval for σ, the standard deviation of heights of American adult women?

    (g) According to Wikipedia, the average height of American adult women is 161.3 cm (2015-2018). Perform a Wald test using the MLE and $\hat{se}$ of the MLE from (a) and (c), and give a P-value.

2. The data set `faithful` included in R contains the variable `eruption`, which gives eruption times (in minutes) of the Old Faithful geyser in Yellowstone National Park.

   (a) Using R, model the eruption times as a mixture of two Normal distributions. Give the MLEs for the five required parameters (probability of being in the first Normal; mean and SD of the first Normal; mean and SD of the second Normal), as well as standard errors and 95% confidence intervals for all five parameters.

   (Note: Study Chapter 3 slides (pp. 49-51) and textbook page 46 or Millar 3.3.4.1. The code should be very similar. One tricky part might be choosing good initial values for `optim()`; it may help to draw a density plot of the data to help you guess these.)

   (b) Using a Wald test, test the hypothesis that the two Normal distributions are equally likely; i.e. that $p = 0.5$.

3. (Chapter 2 Model Fitting Review.) The file `hurricanes.csv` contains data on Atlantic hurricanes that hit the continental U.S. between 1950 and 2012. We are interested in the variable `alldeaths`, which gives total deaths in the continental U.S. due to each hurricane. Fit the following models using maximum likelihood, and give the parameter estimates.

   (a) Poisson

   (b) Negative binomial

   (c) Zero-inflated Poisson

   (d) Zero-inflated negative binomial