

## STAT-S 352 Problem Set 8

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

Tuesday, October 28, 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. In an October 2023 YouGov poll with 1000 respondents, 60% of those surveyed had a favorable opinion of Taylor Swift. (23% had an unfavorable opinion, while 17% didn't know.) For the purpose of this question, treat the 1000 respondents as a random sample of American adults, and assume that 600 of them said they had a favorable opinion of Taylor Swift.

Let  $p$  be the proportion of all American adults who have a favorable opinion of Taylor Swift.

- (a) Find the MLE for  $p$  and a 95% Wald confidence interval (CI) for  $p$ .
- (b) Find a 95% profile likelihood confidence interval (CI) for  $p$ , using `binom.profile`. Compare this interval with the Wald CI you got in part (a). Are these two intervals close? Explain the reason why they are close or not.
- (c) Plot the log-likelihood as a function of  $p$ , for values of  $p$  between 0.5 and 0.7. Draw one horizontal line at the maximum log-likelihood and another horizontal line at the log-likelihood at the lower and upper bound of the profile likelihood CI you found in part (b). Explain where/how the number 1.92 should be marked on this plot.
- (d) The logit transformation of  $p$  is

$$\text{logit}(p) = \log \frac{p}{1-p}$$

What is the MLE of  $\text{logit}(p)$ ?

- (e) Find a 95% profile likelihood confidence interval for  $\text{logit}(p)$ .

2. Does a horse's starting position affect the probability that it will win a race on a circular track? The following table lists the number of wins from each starting position (1 is closest to the inside rail) in 144 eight-horse races:

Starting position	1	2	3	4	5	6	7	8
Number of wins	29	19	18	25	17	10	15	11

Assuming that starting positions were randomly assigned, perform a  $G$ -test of the null hypothesis that a horse's starting position does not affect its chance of winning. Give a  $P$ -value and conclusion.

3. In a tomato plant breeding experiment by MacArthur (1931), the following proportions were expected:

- Tall cut-leaf: 9/16
- Tall potato-leaf: 3/16
- Dwarf cut-leaf: 3/16
- Dwarf potato-leaf: 1/16

The following counts of plants were observed:

- Tall cut-leaf: 926
- Tall potato-leaf: 288
- Dwarf cut-leaf: 293
- Dwarf potato-leaf: 104

Are the observations consistent with the expected proportions? Perform a  $G$ -test, giving a  $P$ -value and conclusion.

4. The file `laliga.txt` contains the number of goals scored in each of the 380 games of the 2021–22 Spanish La Liga soccer season (between 0 and 8.) We wish to model this data using a Poisson distribution with parameter  $\lambda$ .
- Let  $\hat{\lambda}$  be the MLE for  $\lambda$ . What is the numerical value of the MLE for  $\lambda$ ?
  - What is the variance of  $\hat{\lambda}$ ? (Write the mathematical notation/symbol.) What is the estimated variance of  $\hat{\lambda}$ ?
  - Find a 98% Wald confidence interval for  $\lambda$ . Find a 98% profile likelihood confidence interval for  $\lambda$ , using `plkhci`. Find a 98%  $t$  confidence interval for  $\lambda$ . Are the results similar?
  - Suppose we wish to test the null hypothesis that on average, the number of goals scored per game is 2.5. Perform a Wald test, a  $t$ -test and a likelihood ratio test of this hypothesis, and compare the results. For each test, give a test statistic, a  $P$ -value, and your conclusion.