

Logistic Regression

Mini-Assignment - MTH 361 A/B - Spring 2023

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

I. Introduction to Logistic Regression

Materials

The exercises below are derived from the textbook [OpenIntro Statistics \(4th edition\)](#) by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr.

Exercises

1. **Challenger disaster, Part I.** On January 28, 1986, a routine launch was anticipated for the Challenger space shuttle. Seventy-three seconds into the flight, disaster happened: the shuttle broke apart, killing all seven crew members on board. An investigation into the cause of the disaster focused on a critical seal called an O-ring, and it is believed that damage to these O-rings during a shuttle launch may be related to the ambient temperature during the launch. The table below summarizes observational data on O-rings for 23 shuttle missions, where the mission order is based on the temperature at the time of the launch. *Temp* gives the temperature in Fahrenheit, *Damaged* represents the number of damaged O-rings, and *Undamaged* represents the number of O-rings that were not damaged.

Shuttle Mission	1	2	3	4	5	6	7	8	9	10	11	12
Temperature	53	57	58	63	66	67	67	67	68	69	70	70
Damaged	5	1	1	1	0	0	0	0	0	0	1	0
Undamaged	1	5	5	5	6	6	6	6	6	6	5	6

Shuttle Mission	13	14	15	16	17	18	19	20	21	22	23
Temperature	70	70	72	73	75	75	76	76	78	79	81
Damaged	1	0	0	0	0	1	0	0	0	0	0
Undamaged	5	6	6	6	6	5	6	6	6	6	6

- a. Each column of the table above represents a different shuttle mission. Examine these data and describe what you observe with respect to the relationship between temperatures and damaged O-rings.
- b. Failures have been coded as 1 for a damaged O-ring and 0 for an undamaged O-ring, and a logistic regression model was fit to these data. A summary of this model is given below. Describe the key components of this summary table in words.

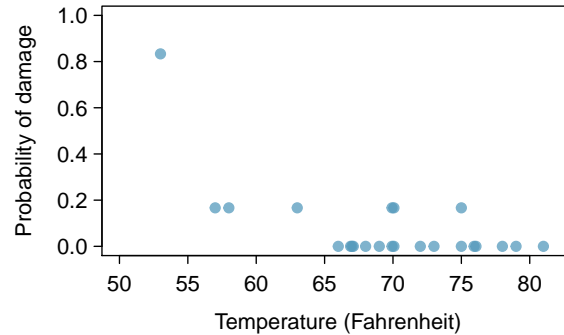
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	11.6630	3.2963	3.54	0.0004
Temperature	-0.2162	0.0532	-4.07	0.0000

- c. Write out the logistic model using the point estimates of the model parameters.
- d. Based on the model, do you think concerns regarding O-rings are justified? Explain.

2. **Logistic regression fact checking.** Determine which of the following statements are true and false. For each statement that is false, explain why it is false.

- a. Suppose we consider the first two observations based on a logistic regression model, where the first variable in observation~1 takes a value of $x_1 = 6$ and observation~2 has $x_1 = 4$. Suppose we realized we made an error for these two observations, and the first observation was actually $x_1 = 7$ (instead of~6) and the second observation actually had $x_1 = 5$ (instead of~4). Then the predicted probability from the logistic regression model would increase the same amount for each observation after we correct these variables.
- b. When using a logistic regression model, it is impossible for the model to predict a probability that is negative or a probability that is greater than 1.
- c. Because logistic regression predicts probabilities of outcomes, observations used to build a logistic regression model need not be independent.
- d. When fitting logistic regression, we typically complete model selection using adjusted R^2 .

3. (Outstanding Question) **Challenger disaster, Part II.** Exercise (1) introduced us to O-rings that were identified as a plausible explanation for the breakup of the Challenger space shuttle 73 seconds into takeoff in 1986. The investigation found that the ambient temperature at the time of the shuttle launch was closely related to the damage of O-rings, which are a critical component of the shuttle. See this earlier exercise if you would like to browse the original data.



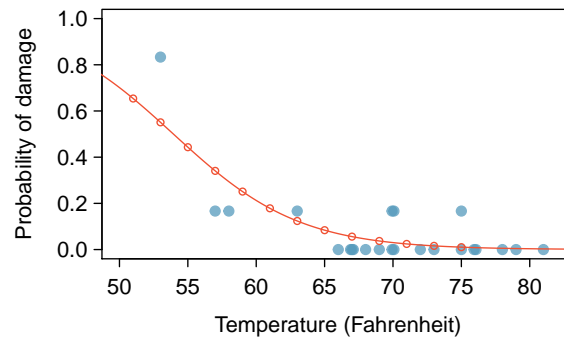
- a. The data provided in the previous exercise are shown in the plot. The logistic model fit to these data may be written as

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = 11.6630 - 0.2162 \times \text{Temperature}$$

where \hat{p} is the model-estimated probability that an O-ring will become damaged. Use the model to calculate the probability that an O-ring will become damaged at each of the following ambient temperatures: 51, 53, and 55 degrees Fahrenheit. The model-estimated probabilities for several additional ambient temperatures are provided below, where subscripts indicate the temperature:

$$\begin{array}{llll} \hat{p}_{57} = 0.341 & \hat{p}_{59} = 0.251 & \hat{p}_{61} = 0.179 & \hat{p}_{63} = 0.124 \\ \hat{p}_{65} = 0.084 & \hat{p}_{67} = 0.056 & \hat{p}_{69} = 0.037 & \hat{p}_{71} = 0.024 \end{array}$$

- b. Describe the model-estimated probabilities from part (a) on the plot shown below.



- c. Describe any concerns you may have regarding applying logistic regression in this application, and note any assumptions that are required to accept the model's validity.