

# Independence Assumption

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. Case Studies: Side Effects of Avandia and Heart Transplants

### Materials

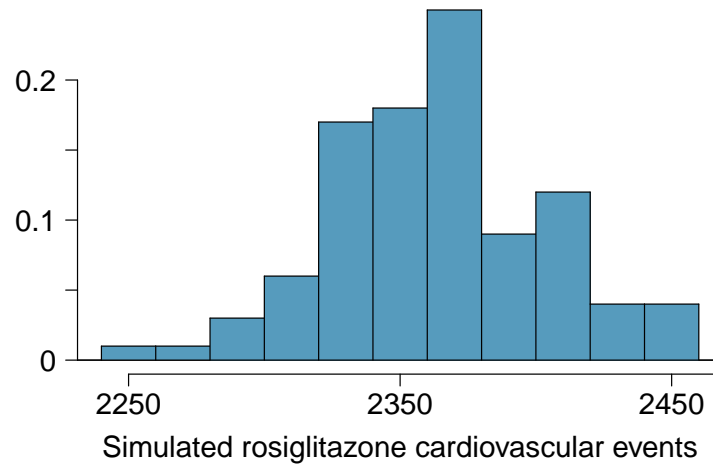
The exercises below are derived from the textbook [OpenIntro Statistics \(4th edition\)](#) by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr.

### Exercises

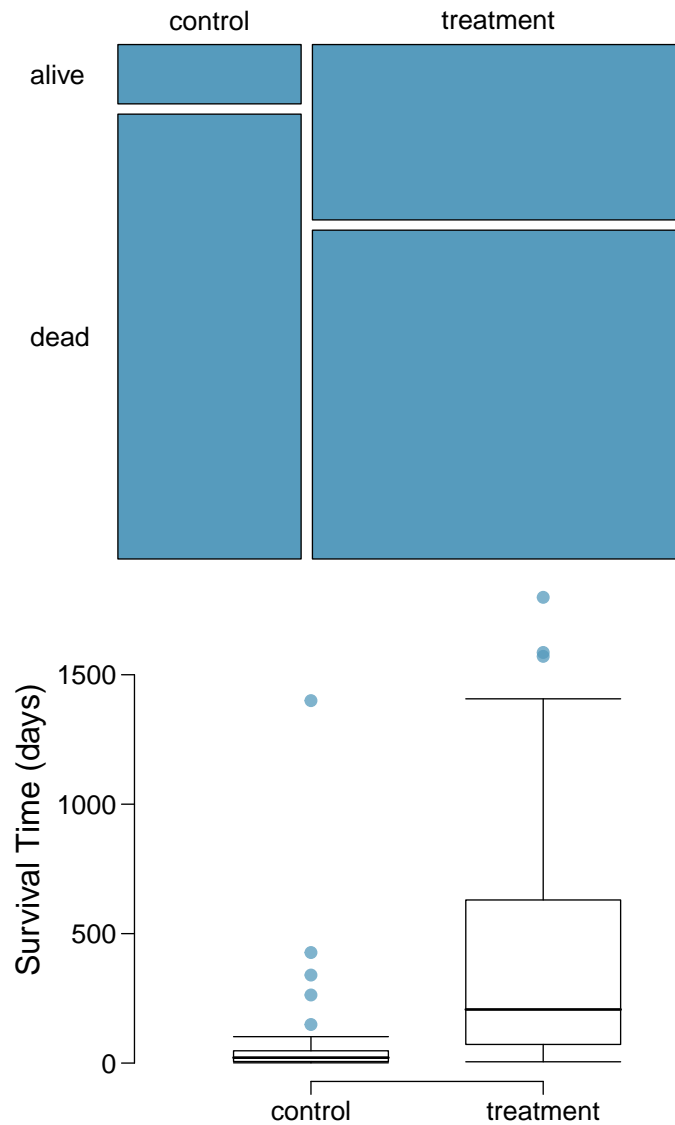
1. **Side effects of Avandia.** Rosiglitazone is the active ingredient in the controversial type-2 diabetes medicine Avandia and has been linked to an increased risk of serious cardiovascular problems such as stroke, heart failure, and death. A common alternative treatment is pioglitazone, the active ingredient in a diabetes medicine called Actos. In a nationwide retrospective observational study of 227,571 Medicare beneficiaries aged 65 years or older, it was found that 2,593 of the 67,593 patients using rosiglitazone and 5,386 of the 159,978 using pioglitazone had serious cardiovascular problems. These data are summarized in the contingency table below. ([Graham et al., 2010](#))

	<i>Cardiovascular problems</i>		Total
	Yes	No	
<i>Treatment</i>			
Rosiglitazone	2,593	65,000	67,593
Pioglitazone	5,386	154,592	159,978
Total	7,979	219,592	227,571

- a. Determine if each of the following statements is true or false. If false, explain why. Be careful, the reasoning may be wrong even if the statement's conclusion is correct. In such cases, the statement should be considered false.
  - i. Since more patients on pioglitazone had cardiovascular problems (5,386 vs. 2,593), we can conclude that the rate of cardiovascular problems for those on a pioglitazone treatment is higher.
  - ii. The data suggest that diabetic patients who are taking rosiglitazone are more likely to have cardiovascular problems since the rate of incidence was  $(2,593 / 67,593 = 0.038)$  3.8% for patients on this treatment, while it was only  $(5,386 / 159,978 = 0.034)$  3.4% for patients on pioglitazone.
  - iii. The fact that the rate of incidence is higher for the rosiglitazone group proves that rosiglitazone causes serious cardiovascular problems.
  - iv. Based on the information provided so far, we cannot tell if the difference between the rates of incidences is due to a relationship between the two variables or due to chance.
- b. What proportion of all patients had cardiovascular problems?
- c. If the type of treatment and having cardiovascular problems were independent, about how many patients in the rosiglitazone group would we expect to have had cardiovascular problems?
- d. We can investigate the relationship between outcome and treatment in this study using a randomization technique. While in reality we would carry out the simulations required for randomization using statistical software, suppose we actually simulate using index cards. In order to simulate from the independence model, which states that the outcomes were independent of the treatment, we write whether or not each patient had a cardiovascular problem on cards, shuffled all the cards together, then deal them into two groups of size 67,593 and 159,978. We repeat this simulation 1,000 times and each time record the number of people in the rosiglitazone group who had cardiovascular problems. Use the relative frequency histogram of these counts to answer (i)-(iii).
  - i. What are the claims being tested?
  - ii. Compared to the number calculated in part c, which would provide more support for the alternative hypothesis, *more* or *fewer* patients with cardiovascular problems in the rosiglitazone group?
  - iii. What do the simulation results suggest about the relationship between taking rosiglitazone and having cardiovascular problems in diabetic patients?



2. (Outstanding Question) **Heart transplants** The Stanford University Heart Transplant Study was conducted to determine whether an experimental heart transplant program increased lifespan. Each patient entering the program was designated an official heart transplant candidate, meaning that he was gravely ill and would most likely benefit from a new heart. Some patients got a transplant and some did not. The variable **transplant** indicates which group the patients were in; patients in the treatment group got a transplant and those in the control group did not. Of the 34 patients in the control group, 30 died. Of the 69 people in the treatment group, 45 died. Another variable called **survived** was used to indicate whether or not the patient was alive at the end of the study. (Turnbull et al., 1974)

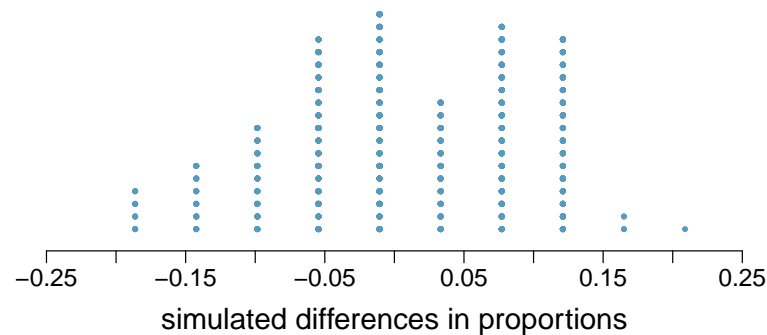


- Based on the mosaic plot, is survival independent of whether or not the patient got a transplant? Explain your reasoning.
- What do the box plots below suggest about the efficacy (effectiveness) of the heart transplant treatment?
- What proportion of patients in the treatment group and what proportion of patients in the control group died?
- One approach for investigating whether or not the treatment is effective is to use a randomization technique.

- i. What are the claims being tested?
- ii. The paragraph below describes the set up for such approach, if we were to do it without using statistical software. Fill in the blanks with a number or phrase, whichever is appropriate.

We write *alive* on [BLANK] cards representing patients who were alive at the end of the study, and *dead* on [BLANK] cards representing patients who were not. Then, we shuffle these cards and split them into two groups: one group of size [BLANK] representing treatment, and another group of size [BLANK] representing control. We calculate the difference between the proportion of *dead* cards in the treatment and control groups (treatment - control) and record this value. We repeat this 100 times to build a distribution centered at [BLANK]. Lastly, we calculate the fraction of simulations where the simulated differences in proportions are [BLANK]. If this fraction is low, we conclude that it is unlikely to have observed such an outcome by chance and that the null hypothesis should be rejected in favor of the alternative.

- iii. What do the simulation results shown below suggest about the effectiveness of the transplant program?



**References**

- Graham, D. J., Ouellet-Hellstrom, R., MaCurdy, T. E., Ali, F., Sholley, C., Worrall, C., & Kelman, J. A. (2010). Risk of acute myocardial infarction, stroke, heart failure, and death in elderly medicare patients treated with rosiglitazone or pioglitazone. *JAMA*, 304(4), 411.
- Turnbull, B., Brown, B., & Hu, M. (1974). *Journal of the American Statistical Association*, 69, 74–80.