

# Problem Set 7

## Hypothesis Testing Foundations

1. **Identify the parameter** For each of the following situations, state whether the parameter of interest is a mean or a proportion. It may be helpful to examine whether individual responses are numerical or categorical.
  - a. In a survey, one hundred college students are asked how many hours per week they spend on the Internet.
  - b. In a survey, one hundred college students are asked: “What percentage of the time you spend on the Internet is part of your course work?”
  - c. In a survey, one hundred college students are asked whether or not they cited information from Wikipedia in their papers.
  - d. In a survey, one hundred college students are asked what percentage of their total weekly spending is on alcoholic beverages.
2. **Hypotheses.** For each of the research statements below, note whether it represents a null hypothesis claim or an alternative hypothesis claim.
  - a. The number of hours that grade-school children spend doing homework predicts their future success on standardized tests.
  - b. King cheetahs on average run the same speed as standard spotted cheetahs.
  - c. For a particular student, the probability of correctly answering a 5-option multiple choice test is larger than 0.2 (i.e., better than guessing).
  - d. The mean length of African elephant tusks has changed over the last 100 years.
3. **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.<sup>1</sup> [Graham:2010]

treatment	No	Yes	Total
Pioglitazone	154,592	5,386	159,978
Rosiglitazone	65,000	2,593	67,593
Total	219,592	7,979	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.

---

<sup>1</sup>The **avandia** data used in this exercise can be found in the **openintro** R package.

- 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 100 times and each time record the difference between the proportions of cards that say “Yes” in the Rosiglitazone and Pioglitazone groups. Use the histogram of these differences in proportions to answer the following questions.
    - i. What are the claims being tested?
    - ii. Compared to the number calculated in part (b), which would provide more support for the alternative hypothesis, *higher* or *lower* proportion of 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?

