

Problem Set 1: Starting off slow

Introduction

To get you familiar with the format of this course, we will start slow with an easy assignment. You can complete the statistical analyses in this assignment with any tool you'd like—there are a variety of appropriate online statistical test calculators available that would be appropriate for this data.

Data description

This contingency table is from a real dataset on diabetes in Pima Indian women.

Diabetic pedigree function	Not diabetic	Diabetic
Under 0.3	60	17
Over 0.3	72	51

The question

Are people in this study more likely to have diabetes if their diabetic pedigree function number is above 0.3 (indicating a greater prevalence of diabetes in their family history)? What statistical test did you use? What is the resulting p-value?

The report

Develop a report (I recommend a Word (or other text editor) document) for your problem set that includes 1) a description of the statistic test used, 2) where you found your calculator, 3) your 'Results' from your statistical test, including descriptions of answers to the questions above. Save your report as a pdf file and submit your report through the course 2GW site.

Due date

Day 7 Week 1

Data Source

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C. and Johannes, R. S. (1988) Using the ADAP learning algorithm to forecast the onset of diabetes

mellitus. In Proceedings of the Symposium on Computer Applications in Medical Care (Washington, 1988), ed. R. A. Greenes, pp. 261–265. Los Alamitos, CA: IEEE Computer Society Press.