

SI 544 Stats Final, Fall 2006

Lada Adamic
Dec 12, 2006

This is your final exam. Turn in all R code, label your plots, and put in a sentence or two highlighting the results of each step of your analysis. Mention where appropriate things like p-values, what your null-hypothesis is, and whether you reject it or accept it. For each statistical method you choose, briefly justify your selection and comment on whether the assumptions needed in order to be able to use the method are met. This exam is open book/lecture notes/solution sets/web. You are to work on this without the assistance of others. Good luck!

1 Tennis balls

You are a tennis instructor. You mail-ordered a box of tennis balls. There seems to have been an error, because the box you received claims to contain “magic” tennis balls, proven to improve your accuracy. These “magic” tennis balls still claim to meet the standard requirements, including weight.

A. (15 pts) The box says that the balls should weigh 57.5 grams with a standard deviation of 0.5 grams. You have a high precision digital scale handy and have a few minutes to kill before your next class, so you decide to weigh 10 balls to see if they meet specifications. These weights are saved in the file `tennisballweights.txt`. What is your null hypothesis? Can you reject it?

B. (10 pts) If the balls are actually 0.5 grams heavier on average, how many balls would you need to weigh in order to detect the difference at the $\alpha = 0.05$ level with power $\beta = 0.80$. Comment on the result with respect to your analysis in 1A, assuming that the sample in actuality was drawn from a box of balls that were 0.5 grams heavier on average.

C. (30 pts) You have three classes to teach that day: beginner, intermediate, and master. You decide to test how well your students do with the “magic” tennis balls, so you design an experiment. In each class of 24, you randomly assign 12 people to practice their serves on one court with the magic tennis balls and the other 12 practice on another court with regular tennis balls. You record two things about each student’s serves: how fast they are on average, and how often the student succeeded in hitting a cone placed on the court. The file `tennisdata.txt` has the data, with 4 columns, each row corresponding to one student: **classLevel**, **typeOfBall**, **avServeSpeed**, **numConesHit**.

Do an analysis of variance to figure out how the average serve speed depends on the class level and the type of ball. Do the same for the accuracy of the serve (the number of times the cone was hit). Draw the corresponding interaction plots. Are the magic tennis balls really magic?

D. (20 pts) You'd like to quantify the relationship between the average speed of a student's serve and their accuracy in hitting a cone. Using only the scores of those students who were using the regular balls, fit a linear regression. Plot the data as well as your regression line. Is there a relationship between the two variables?

Extra credit(4pts) Calculate a prediction interval, such that if a student hits the cone 20 out of 40 times, you can predict with 95% confidence that their average serve speed will be within this interval.

2 Money can't buy you happiness, or can it?

(25 pts) The Pew Spyware survey asked the respondents about their income, reported in the column **INC** of the Spyware.sav data file (remember, it is in SPSS format). You would like to know if people who are better off financially are happier, or at least happier with the state of the country. You remember that the very first question, **Q1** was: "Overall, are you satisfied or dissatisfied with the way things are going in this country today?". Use both a plot and a statistical test to discuss the relationship between the two variables.