

MATH 185 – Homework 1
Due Wednesday, 04/06/2016, by 11:59 PM

Send your code to math185ucsd@gmail.com. Follow the following format exactly. For Homework 1, in subject line write “MATH 185 (HW 1)” and nothing else in the body. There should only be one file attached, named `hw1-lastname-firstname.R`. Make sure your code is clean, commented and running. Keep your code simple, using packages only if really necessary. If your code does not run, include an explanation of what is going on.

Problem 1. (Earthquakes) Consider the `earthquakes-2014` dataset (on the cloud). This is taken from a catalog data from the Southern California Earthquake Data Center (SCEDC) holdings for the year 2014. The data include local, regional, and quarry-blast events with epicenters between latitudes 32 degrees S and 37 degrees N and longitudes between -122 degrees W and -114 degrees E. See <http://www.data.scec.org> for more information. To read the data, use `load` and the data frame is called `dat`.

- A. Focus on events of magnitude at least 2. Use the function `subset` to extract the corresponding observations. When this is done, produce a table of counts for the number of events in each month, and then draw a relevant plot.
- B. Were earthquakes more prevalent some months out of the year 2014? Formalize this question into a test of hypothesis (what is the null?) and perform a test. Name the test and specify how you choose to calibrate it. After you perform the test, offer some brief comments.

Problem 2. (UC Berkeley admissions) In 1973, the Graduate Division at UC Berkeley received a number of applications. Ignoring incomplete applications, there were 8442 male applicants of whom 3738 were admitted, compared with 4321 female applicants of whom 1494 were admitted.

- A. Based on these data, can we test whether the applicant’s gender influenced the admission decision?
- B. Perform a test of association. Write down the null hypothesis you are testing and conclude after performing the test.

Problem 3. (UC Berkeley admissions (detailed)) A subset of these data is available in R as `UCBAdmissions` (already loaded). These are the admission decisions for the six largest departments. Notice that we now have not 1 but 6 contingency tables.

- A. Produce a (2 by 2) contingency table grouping all these departments and rerun the analysis of Problem 2 on that.
- B. Find a way to plot the `UCBAdmissions` dataset in an informative way. Is this graphical investigation congruent with your previous findings?
- C. (Optional) Suggest a test that is better able to utilize these more detailed data (than a test of association based on the pooled data). Make sure you properly introduce the test statistic, explaining why it is relevant, and say how you calibrate it.