

Name:

SID:

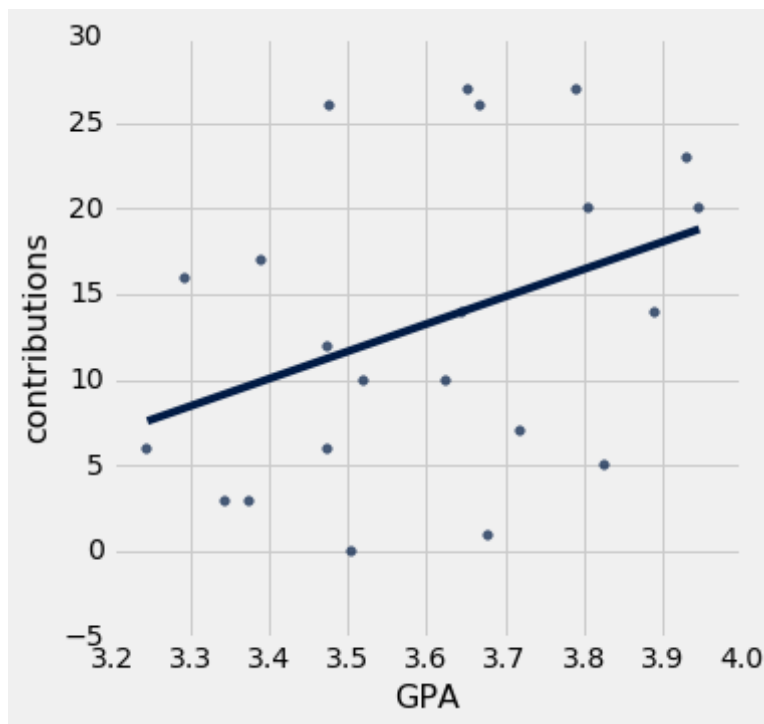
In the StudentLife Study at Dartmouth College, “passive and automatic sensing data from the phones of a class of 48 Dartmouth students over a 10 week term [were collected] to assess their mental health (e.g., depression, loneliness, stress), academic performance (grades across all their classes, term GPA and cumulative GPA) and behavioral trends (e.g., how stress, sleep, visits to the gym, etc. change in response to college workload – i.e., assignments, midterms, finals – as the term progresses).”

Part of that study included Piazza usage, which was recorded for 30 of the 48 students in the class.

A table called `students` has one row for each of these 30 students and the following columns:

- `GPA`: grade point average of the student
- `contributions`: The number of contributions, including posts and follow-up discussions
- `answers`: The number of answers posted

Analysis 1. A scatter plot of students with GPA above 3.1 versus their contributions appears below.



Analysis 2. Among the 30 students, there are 14 who answered at least one question; 16 did not.

Note: If you're interested, the `StudentLife.ipynb` file included with lab12 contains the data.

Problem 1 Estimation

Assuming these 30 were selected uniformly at random from all 48 students in the class, how would you compute a 95% confidence interval for the standard deviation of the GPA for all students in the class?

Problem 2 Regression

Among these 30 students, how would you determine whether their GPA and their number of Piazza contributions are linearly related?

Problem 3 Regression Inference

Assuming these 30 were selected uniformly at random from all 48 students in the class, how would you determine whether or not their GPA and their number of Piazza contributions are positively linearly related for all students in the class?

Problem 4 Comparison

You're interested in comparing the GPAs for two groups: those who did and did not answer at least one question on Piazza. Which of the following hypotheses could you test using the `students` table and how?

- a) The difference between the average GPAs of the two groups among these 30 students is at least 0.2.
- b) The difference between the average GPAs of the two groups among these 30 students is due to chance.
- c) Answering questions on Piazza increases the GPA of students in this class.

Problem 5 Randomness

Rival researchers claims that the data have been falsified! They believe that in fact every student gave at least one answer on Piazza, but the researchers flipped a fair coin for each student and only recorded his/her number of answers if it came up tails. For heads, they just wrote 0 answers.

How would you determine whether the evidence in the `students` table is consistent with this controversial claim?

- a) What hypothesis would you evaluate?
- b) What test statistic would you use?
- c) How would you compute the probability distribution of the test statistic under the hypothesis?
- d) What observation would you compare to this probability distribution?

Problem 6 Median (Optional)

If you don't know how the 30 students were chosen from the 48 in the class, could you compute an interval that certainly contained the median GPA of the class as a whole? If so, how?