

Statistics Homework, Part II

Please write up your answers (neatly!), and turn in next Wednesday at the beginning of class.

1. Suppose we have n real numbers:

$$\{x_1, x_2, \dots, x_n\}$$

Define a function $F(k)$:

$$F(k) = (x_1 - k)^2 + (x_2 - k)^2 + \dots + (x_n - k)^2$$

You might notice that this is indeed a function of the one variable k , since x_1, \dots, x_n are fixed real numbers.

Show that the k that minimizes F is the sample mean.

2. A Data Investigation¹

The University of California at Berkeley was charged with having discriminated against women in their graduate admissions process for the Fall quarter of 1973. The following table summarizes the data:

	Men	Women
Accepted	533	113
Denied	665	336
Total	665	336

- (a) Calculate the proportion of Men applicants who were accepted, and the proportion of Women who were were accepted. Is there evidence that men were accepted at a much higher rate than women?
- (b) Does it look like Berkeley was guilty of gender discrimination? Explain.

¹Rossmann and Chance, "Investigating Statistical Concepts, Applications and Methods", 2006

- (c) The following table breaks down the data in more detail- There were two graduate programs which we label A and F :

	Men Accepted	Men Denied	Women Accepted	Women Denied
Program A	511	314	89	19
Program F	22	351	24	317
Total	533	665	113	336

Within each program, calculate the proportion of men who were accepted and the proportion of women who were accepted. Did men have the higher rate of acceptance in both programs? Does this seem consistent with your previous answer?

- (d) There should appear to be a paradox here- This is called *Simpson's Paradox*. Explain why it was not reasonable to average the acceptance rates for the two programs to find the overall acceptance rate for women.
3. Construct your own Simpson's Paradox: Show that it is possible for one softball player (make up a name starting with A), to have a higher proportion of hits than another (make up a name starting with B) in July and August, and yet A can have a lower proportion of hits for the two months combined.

Hints: Give each player the same number of at-bats, say 200 for the two months combined. Try to make the differences in the players' proportion of hits as large as you can (don't worry about using realistic numbers).