Homework 2

Due on Gradescope 9/13/2016 at 4:00PM (Before Lecture)

1. Theory Meets Data, Section 2.3, Problem 1: Consider a list of numbers $x = \{x_1, x_2, \cdots, x_n\}$
   (a) If all the entries of x are the same, then what is the variance of this list?
   (b) Suppose some proportion p of the numbers in the list are 1, and the remaining $1-p$ proportion of the numbers are 0. For instance, if the list had 10 numbers and $p = 0.4$, then 4 of the numbers would be 1 and the remaining 6 would be 0. Show that the standard deviation s_x of the list is $\sqrt{p(1-p)}$.

2. Theory Meets Data, Section 2.3, Problem 2: Suppose we have a list $x = \{x_1, x_2, \cdots, x_n\}$ and constants a and b. Let $\bar{x}$ be the mean of the list, and s_x the standard deviation. In what follows, we will be creating new lists y by using x, a, and b. The notation $y = f(x)$ means that $y_i = f(x_i)$ for each i such that $1 \leq i \leq n$.
   (a) What is the standard deviation of $y = ax$, in terms of a, s_x, and $\bar{x}$?
   (b) What is the standard deviation of $y = x + b$, in terms of b, s_x, and $\bar{x}$?
   (c) What is the standard deviation of $y = ax + b$, in terms of a, b, s_x, and $\bar{x}$?

3. Freebie Problem, See Theory Meets Data Section 2.2: Suppose we have a list $x = \{x_1, x_2, \cdots, x_n\}$. Beginning from the definition of the variance of the list given in class,

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

   show that the variance can be written equivalently as

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2.$$

4. Theory Meets Data, Section 2.3, Problem 3: Suppose we have a class consisting of n students. This class has two sections, A and B. Section A has m students, and section B has $n - m$ students. In the two parts below, you will find the form of the variance derived in the last problem to be quite useful (called the "computational form" in the book).
   (a) Let $n = 100$ and suppose Section A has 70 students. Section A's students have an average score of 60 with a standard deviation of 10. Section B's students have an average score of 89 with a standard deviation of 6. Find the mean and standard deviation of student scores across the entire class.
   (b) Suppose section A has n students and B has $n - m$ students. Write the mean and standard deviation of section A as $\bar{A}$ and s_A, respectively, and the mean and standard deviation of section B as $\bar{B}$ and s_B, respectively. Find the mean and standard deviation of student scores across the entire class in terms of n, m, $\bar{A}$, $\bar{B}$, s_A, and s_B.

5. Theory Meets Data, Section 3.3, Problem 1 Suppose a list of numbers $x = \{x_1, \cdots, x_n\}$ has mean $\bar{x}$ and standard deviation s_x. We say that a number y is within z standard deviations of the mean if $\bar{x} - zs_x < y < \bar{x} + zs_x$.

   (a) Let c be the smallest number of standard deviations away from $\bar{x}$ we must go to ensure that the range $(\bar{x} - cs_x, \bar{x} + cs_x)$ contains at least 50% of the entries in x. What is c?

   (b) Suppose that we constructed a list of all of the BART travel times from Berkeley to San Francisco in the last year. The mean of this list is 38 minutes, and the standard deviation of the list is 4 minutes. If you want to make the claim "At least 90% of BART rides from Berkeley to San Francisco in the last year took between L and U minutes", what numbers could you substitute for L and U?