

## Homework Assignment 1 (Due Wednesday, February 15, 2017)

The homework is due at 10.30am in the dropbox on the Course Plus page (you can find the dropbox under the 'Resources' tab in the upper right). For exercises involving R code, please knit a document from your R markdown (Rmd) file. Generate a single pdf file for your entire submission and give it a name that makes it identifiable (calling it `140.615.HW.Number.Lastname.Firstname` or similar).

1. The seed of Mendel's pea plants were either smooth or wrinkled, the result of a single gene with two alleles,  $A$  (smooth) and  $a$  (wrinkled), with  $A$  dominant to  $a$ , so that seed with genotype  $AA$  or  $Aa$  are smooth and seed with genotype  $aa$  are wrinkled.  $\rightarrow$  We cross two pure-breeding lines, one with smooth seed and genotype  $AA$  and the other with wrinkled seed and genotype  $aa$ , to obtain the  $F_1$  with genotype  $Aa$ .  $\rightarrow$  We self an  $F_1$  and pick a random  $F_2$  seed.  $\rightarrow$  We grow up the  $F_2$  and self it to obtain a random  $F_3$  seed. Calculate the following:
  - (a)  $\Pr(F_2 \text{ seed is smooth})$
  - (b)  $\Pr(F_2 \text{ seed has genotype } Aa)$
  - (c)  $\Pr(F_2 \text{ seed has genotype } Aa \mid \text{it is smooth})$
  - (d)  $\Pr(F_3 \text{ seed is smooth} \mid F_2 \text{ has genotype } AA)$
  - (e)  $\Pr(F_3 \text{ seed is smooth} \mid F_2 \text{ has genotype } Aa)$
  - (f)  $\Pr(F_3 \text{ seed is smooth})$
  - (g)  $\Pr(F_3 \text{ seed is smooth} \mid F_2 \text{ is smooth})$
  
2. Suppose that  $1/100,000$  bacterial cells contain a mutation providing resistance to substance A. Suppose I create a bunch of plates with 200,000 cells per plate. Let  $X$  be the number of cells on a plate that are resistant to substance A.
  - (a) What is the mean number of cells in a plate that are resistant to A?
  - (b) What is  $\text{SD}(X)$ ?
  - (c) Calculate  $\Pr(X = 0)$ .
  - (d) Calculate  $\Pr(X = 5)$ .
  - (e) Calculate  $\Pr(X > 2)$ .
  
3. Suppose  $X \sim \text{Normal}(\text{mean}=5, \text{SD}=3)$ . Calculate the following:
  - (a)  $\Pr(X < 6)$
  - (b)  $\Pr(X > 0)$
  - (c)  $\Pr(0 < X < 5)$
  - (d)  $\Pr(2 < X < 8)$
  - (e)  $\Pr(|X - 5| > 2)$

4. Suppose  $Y$  is a random variable with  $E(Y) = 30$  and  $SD(Y) = 5$ .
- (a) Let  $Z = (Y - 30)/5$ . Calculate  $E(Z)$  and  $SD(Z)$ .
  - (b) Let  $X = -Y$ . Calculate  $E(X)$  and  $SD(X)$ .
  - (c) Let  $R = 5 + Y/3$ . Calculate  $E(R)$  and  $SD(R)$ .
5. The Army Medical Corps is concerned about the intestinal disease  $X$ . From previous experience they know that soldiers suffering from the disease invariably harbor pathogenic organisms in their feces and that for all practical purposes every disease stool specimen contains these organisms. The organisms are never abundant, however, and thus only 20% of all slides prepared by the standard procedure contain some of them. (We assume that if an organism is present on a slide, it will be seen.) How many slides per stool specimen should the laboratory technicians prepare and examine to ensure that if a specimen is positive, it will be erroneously diagnosed negative in less than 1% of the cases (on the average)?
6. Problems 2.1-2.3. (p 9/10) from John Verzani's simpleR notes.
7. Problem 2.6. (p10) from John Verzani's simpleR notes. There is a typo in simpleR notes for this exercise. It should read: *Note, we use  $X_1$  to denote the first element of  $x$  (which is **1**) etc.*