

## Applied Statistics HW 15

1. We are about to conduct a survey on campus to determine if there is a equal percentage of males and females. Let us suppose for the moment that there is, so that 50% of the students are females, and that our campus has 1200 students altogether. In order to get our data, we select 60 students at random, and compute how many of them are females. Denote by  $X$  the number of females we end up with, and with  $\hat{p}$  the percent of females we end up with, in our group of 60.
  - a. Can we assume that  $X$  and  $\hat{p}$  fit into the binomial setting? Explain all the steps you had to follow and the assumptions you had to make. Can we further assume they are approximately normally distributed?
  - b. What are the mean and standard deviation of  $X$  and  $\hat{p}$ ?

- c. What are the chances, that we would see at least 70% females in our survey?
- d. What are the chances, that we would see a female percentage between 58% and 62%?
- e. Provide a range that would include the middle 90% of the female percentages we are likely to see.

2. Recently there was a class action law suit against Walmart. Walmart was accused on a gender bias on their hiring practices when it came to store managers. The basic math to support that was the following: two-thirds of all Walmart employees are women. However, only 14% of the 8000 Walmart store managers are female. We will explore the mathematics behind this discrepancy. To start with, let us make the assumption that the store managers are selected among all Walmart employees, and that every Walmart employee has an equal chance of being selected for store manager. Then the number  $X$  of store managers that are female will follow a binomial distribution, with  $n = 8000$  and  $p = 2/3$ . Correspondingly, the sample proportion  $\hat{p}$  of store managers that are female is related to that binomial.

a. What do you think of the independence assumption needed for the binomial distribution in this case? Don't forget to also check the sizes of the sample and the population.

b. What are the mean and standard deviation of  $\hat{p}$ ? Is  $n$  large enough for us to assume that  $\hat{p}$  follows an approximately normal distribution?

- c. What are the chances, that  $\hat{p}$  will be at least 65%? How improbable does that 14% seem?
- d. Provide a plausible range for the possible values of  $\hat{p}$  (you will need to specify what “plausible” means for you, i.e. how likely the provided range is to be correct).

- e. Let us suppose that there were only 50 stores. How likely is it, that the percent of female store managers is at most 30%? How does having only 50 stores alter the assumptions we made earlier and the numbers we computed earlier?
- f. Continuing with the assumption that there were only 50 stores, we would like to find a value, such that there is only a 5% chance for the sample proportion to be below this value. What would that value be?