MONT 107N – Understanding Randomness
Selected Solutions for Problem Set 3
March 14, 2010

Chapter 20 Review Problems

3) a) There should be $50,000$ tickets in the box – one for each of the forms in the whole "population."

b) Since we are counting/classifying, this will be a 0/1 box; tickets with 1's will represent forms with gross income over \$50,000, and tickets with 0's will represent forms with gross incomes under that level.

c) The SD of the box is (using the "shortcut formula")

$$\text{SD of box} = \sqrt{(.2) \times (.8)} = \sqrt{.16} = .4,$$

so this is False.

d) The 900 draws from the box will represent the sample, so this is True.

e) The number of forms in the sample with gross income over \$50,000 is the sum of the tickets drawn from the box, so EV $= 900 \times .2 = 180$, and the SE for the sum is $\sqrt{900} \times .4 = 12$. The number of forms in the sample with gross incomes over \$50,000 will be about 180, plus or minus 12 or so.

Now $12/900 \times 100\% \doteq 1.33\%$ is the SE for the percent. The percent of forms in the sample with gross incomes over \$50,000 will be about 20% give or take 1.33% or so.

Finally, we want to estimate the chance that a particular sample will have between 19% and 21% of forms with gross income over \$50,000. Note that $1\% = .75 \times 1.33$. So 1% corresponds to .75 in standard units (multiples of the SE for the percent). So this chance will be estimated by the area under the normal curve from $-.75$ to $.75$. This is about 55%.

f) This cannot be done with the given information. There is no information about the percentage of forms with gross incomes over \$75,000.

5) The total weight for a random group of 50 people is like the sum of a random draw of 50 tickets from a box representing the whole population of people. The ticket for each person has his or her weight in pounds. The given information shows that the average of the box is 150, and the SD of the box is 35. Hence the EV for the sum of the draws is $50 \times 150 = 7500$ pounds and the SE for the sum is $\sqrt{50} \times 35 \doteq 247.5$ pounds. The chance that a fully-loaded elevator will be carrying more than 4 tons $= 8000$ pounds is estimated as the area under the normal curve with

$$z > \frac{8000 - 7500}{247.5} \doteq 2.02,$$

which is a bit over 2%. (Since overloading is not all that unlikely, maybe a sturdier elevator is indicated. You don't want the elevator to fail roughly every 50 times it used with a full load of people!!)

6) Only option (ii) is correct here. It is important to realize that the size of the sample relative to the size of the population *is essentially irrelevant* for questions like this(!) The *absolute size of the sample* is what matters because of the formula for the SE for the percentage:

$$\sqrt{\frac{pq}{n}} \times 100\%,$$

where $p =$ observed proportion, $q = 1 - p$, and n is the size of the sample. The larger the sample is, the smaller the SE is (no matter what the size of the population the sample is drawn from is).

Chapter 21 Review Exercises

5) a) Think of a (big!) 0/1 box with one 1 ticket for each teenager who knows that Chaucer wrote the *Canterbury Tales*, and a 0 ticket for each teenager who does not know that fact. Then the data is like the result of drawing a random sample of size 6000 from the box, and we observe a percentage of 36.1% of 1s in the sample. Using the bootstrap, the SE is estimated by

$$\sqrt{\frac{(.361)(.639)}{6000}} \times 100\% = .6\%$$

(that's .6 of 1%). Hence the 95% is estimated as

$$36.1\% \pm 2 \times (.6\%) = 36.1\% \pm 1.2\%.$$

b) This part is similar. (There is no reason why one cannot repeat the calculations for the percentage of the population of teenagers who know who Thomas Edison was.) The answer is $95.2\% \pm 0.6\%$.