MONT 107N – Understanding Randomness
Solutions for Midterm Examination – March 19, 2010

I. A gambler is playing roulette. On each spin of the wheel, he places a bet \$1 on *six numbers*. If one of his block of six numbers comes up, he will win \$5 on a \$1 bet. Otherwise, he loses \$1.

A) (10) Construct a box model for this bet and compute the average and SD of the box. (Recall: There are a total of 38 slots on the roulette wheel.)

Solution – The box should have 38 tickets (representing the possible results of the spin of the wheel), 6 of which have $+5$, and 32 of which have -1 values.

B) (10) If this bet is played 36 times, compute the EV and the SE for the total outcome (i.e. how much ahead or behind the gambler will be after the 36 plays).

Solution – We need the average and SD of the box first. The average of the box is $(6 \times 5 + 32 \times (-1))/38 = -2/38 = -1/19$. Using the "shortcut formula," the SD of the box is

$$(5 - (-1)) \times \sqrt{\frac{6}{38} \times \frac{32}{38}} \doteq 2.19$$

So the expected value is

$$\text{EV} = 36 \times \frac{-1}{19} = \frac{-36}{19} \doteq -1.89$$

and the standard error is

$$\text{SE} = \sqrt{36} \times 2.19 \doteq 13.13.$$

C) (5) What are the gamblers chances of ending up \$5 or more ahead after these 36 spins of the wheel?

Solution – In standard units $\frac{5-(-1.89)}{13.13} \doteq .525$. Using the entries for $z = .50$ and $z = .55$, we can estimate the chance as the area in the upper tail of the normal curve starting at $z = .525$. This is about

$$\frac{1}{2}\left(100\% - \frac{38.29\% + 41.77\%}{2}\right) \doteq 30\%$$

(It will also be acceptable to use either the value $A(.5) = 38.29\%$ or the value $A(.55) = 41.77\%$ here.)

II. Assume 100 draws with replacement are made from the following box.

1 2 3 3 7 8

A) (5) What box model would you use to model the number of 3s that are drawn?

Solution – We would use a 0/1 box with tickets marked $0, 0, 1, 1, 0, 0$.

B) (5) If 40 3's are drawn among the 100 numbers, what is the chance error?

Solution – The average of the box is $2/6 = 1/3$. So the EV for the number of 3's in 100 draws is $100 \times 1/3 \doteq 33.3$. The chance error is the difference $40 - 33.3 \doteq 6.7$.

C) (10) Would you consider the chance error from part B to be a large chance error, roughly an average chance error, or a small chance error? Explain your answer.

Solution – The SD of the 0/1 box is $\sqrt{1/3 \times 2/3} \doteq .47$. So the SE for 100 draws is $\sqrt{100} \times .47 \doteq 4.7$. So the chance error from part B is between 1SE and 2SE – bigger than usual, but not wildly improbable.

III. On November 6, 2007, the Gallup polling organization reported the results of a survey about the role of women in society in the three North African countries of Algeria, Morocco, and Tunisia. Among the results of the poll, 76% of Tunisian adults agreed that Women should be allowed to hold any job for which they are qualified outside the home. The Tunisian sample consisted of 912 adults.

A) (15) Based on this information, construct a 95% confidence interval for this estimate. (Assume for the purposes of this calculation that it was a simple random sample.)

Solution – The observed percent is the 76% given. Using the "bootstrap" we estimate the SE for the percent:

$$\text{SE} \doteq \sqrt{\frac{(.76)(.24)}{912}} \times 100\% \doteq 1.4\%$$

Hence the 95% confidence interval will be

$$76\% \pm 2 \times 1.4\% = 76\% \pm 2.8\%.$$

B) (5) Your answer in A should have used an estimate for the standard error for the precentage. What is the largest possible actual value for this standard error?
Solution – The largest the SE for the percent could be would occur if the actual population percent was 50%. This would give

$$\sqrt{\frac{(.5)(.5)}{912}} \times 100\% \doteq 1.7\%.$$

(Using this and being "conservative" would give a slightly wider confidence interval.)

IV. (15) An urn contains 1000 blue marbles and 500 red marbles. The process of drawing 100 marbles with replacement from the urn and recording the number of red marbles is repeated 10 times. The counts are

$$28, 29, 27, 31, 34, 29, 26, 32, 27, 30$$

Is there any reason to be suspicious about the results? Answer yes or no and back up your claim with calculations.

Solution – No, there is not any reason to be suspicious. Note that the EV for the number of reds in 100 draws from the urn would be $100 \times 1/3 \doteq 33.3$. So it is true that all but one of the actual numbers are smaller than that. *However*, the SE for the 100 draws is $\sqrt{100} \times \sqrt{2/3 \times 1/3} \doteq 4.7$. So 1 SE below the EV is about 28.6 and 2 SE's below the EV is about 23.9. So 5 of the 10 totals are within 1 SE of the EV, all of the others are within 2 SE's of the EV. Conclusion – this would be a somewhat unusual outcome, but nothing that should raise any undue suspicions. This is what the results of chance processes can look like!

V. Short answer.

A) (5) For a final project in a psychology class, two students want look at the connection between a person's long-term memory after hearing lists of words read aloud to them and the "emotional content" of those words. Student A says, "it is fine if we use the students in our class as the subjects for the study." Student B says, "it would be better to put an ad in the college newspaper to recruit the subjects." Are there possible problems with both of these methods? If not, which would be better? If so, can you suggest a clearly superior way of choosing the subjects?

Solution – Neither method for selecting the subjects is especially reliable because both are prone to *selection biases* of various kinds. The students in the class might have knowledge of the connection being studied. The students who respond to the ad might be special in some way that would affect the outcome of the test. A superior method would be to try to make a random sample of the students at the college, or even of all residents in the town (if possible).

B) (5) Why are we justified in using the normal curve to analyze the chances of sums of draws from a box even when the box itself does not have a normal distribution of values? Is there a limitation to keep in mind when doing this?

Solution – The justification comes from the Central Limit Theorem which says that the probability histogram for the sums will follow a shifted, scaled normal curve as long as the number of draws is large enough. This is the limitation – the normal approximation is only guaranteed to be good for large numbers of draws.

C) (5) True or False and explain: All estimated 95% confidence intervals for a population percentage computed from samples of the same size $n = 100$ from the same population will have the same width.

Solution – False. The estimated confidence intervals use the bootstrap method to estimate the SE. This will vary with the sample. So some intervals will be wider than others.

D) (5) True or False and explain: A simple random sample of size 300 is selected from the 27000 students at a large state university and a 95% confidence interval for the percentage is estimated as $74\% \pm 5\%$. These results show that it is impossible for 80% of the student population to be undergraduates.

Solution – False. Any percent in the confidence interval is a believable value for the percentage of undergraduates. But values outside the interval *are still possible.* (The idea is that that will not happen very often if we sample repeatedly and compute a confidence interval each time.)