MONT 105N – Analyzing Environmental Data
Answers for Midterm Exam Review Questions
March 13, 2020

1) A random sample of 1000 college first-year students in the US is selected and their ages are determined.
   a) What would you expect was the sample mean age, to the nearest whole number?
   Solution: The sample mean is probably somewhere between 17 and 18, and I would suspect it's closer to 18. Note: This takes into account "nontraditional students" too because they form a relatively small fraction of the total college first-year population.
   b) Give the formula that would be used to compute the sample SD, calling the ages $x_1, \ldots, x_{1000}$.
   Solution: Letting the sample mean be $\overline{x}$, the formula is

$$S = \sqrt{\frac{1}{999}\sum_{i=1}^{1000}(x_i - \overline{x})^2}$$
$$= \sqrt{\frac{1}{999}\left((x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_{1000} - \overline{x})^2\right)}$$

   c) Would you expect the sample SD to be about 0.1 year, 1 year, or 10 years? Explain.
   Solution: It makes no sense for the SD to be as big as 10 years since that would leave the possibililty of a significant number of students being 8 years old or less. I would accept either 0.1 year or 1 year as a reasonable estimate. Almost all first year students would be 17, 18, or 19 years old (at least at liberal arts colleges). But I would expect the SD is closer to 0.1, since 18 would be the most common age.

2) Find the 5-number summary, and draw the "box and whisker plot" for the data set

$$5 \quad 3 \quad 7 \quad 8 \quad 8 \quad 10 \quad 1 \quad 2 \quad 4$$

   (Use the inclusive method for the quartiles.)

Solution: Rearranging the numbers to non-decreasing order we have

$$1 < 2 < 3 < 4 < 5 < 7 < 8 < 8 < 10$$

The min is 1, the max is 10, the median is 5. The first quartile is the median of the first half of the list (including the median):

$$1 < 2 < 3 < 4 < 5$$

so $Q_1 = 3$. Similarly, $Q_3 = 8$. The box extends from 3 to 8 with a vertical bar at 5. "The whiskers" extend to 1 on the left and to 10 on the right.

3) Eight birds – two robins, two woodpeckers, two thrushes, and two blackbirds are placed in a covered cage. Three random draws of a single bird are made from the cage *with replacement*.
   a) Are the three draws *independent* in this case? Why or why not?
      Solution: Yes, since the draws are done with replacement. The probabilities for each of the birds on each of the draws are the same, and knowing how the last draw turned out does not change those probabilities.
   b) What are the chances that all three birds are robins?
      Solution: Since $2/8 = 1/4$ of the birds are robins and the draws are independent, this is $(1/4)(1/4)(1/4) = 1/64$.
   c) What are the chances that none of the birds is a robin?
      Solution: We have a $6/8 = 3/4$ chance of picking a robin each time so the probability is $(3/4)(3/4)(3/4) = (27/64)$.
   d) What are the chances that there is a something other than a robin among the three birds?
      Solution: This is the complementary probability to the one computed in part (a): It equals $1 - (1/64) = 63/64$.
   e) (More challenging) What would change in your answers to a,b,c,d if the draws were done *without replacement*?
      Solution: (a) No, the probabilities change from one draw to the next depending on the outcomes of the previous draws. So without replacement the draws are *not independent*. (b) Since there are only 2 robins to start, there is no way to pick 3 robins if the draws are done without replacement. The probability is zero. (c) This is $(6/8)(5/7)(4/6) = 5/14$. Explanation: You have a $(6/8)$ chance of picking a non-robin the first time. After that, there are 5 non-robins and 7 birds left, so on the second draw, you have a $5/7$ chance of drawing a non-robin. Then after those two draws there are 4 non-robins and 6 birds left, so the chance of picking a non-robin on the last draw is $5/14$. Note that the factors after the first one here are technically *conditional probabilities*. For instance the $5/7 = P(non - robin\ on\ second \mid non - robin\ on\ first)$ and the $4/6$ is $P(non - robin\ on\ third \mid non - robins\ on\ first\ and\ second)$

4) A coin is tossed repeatedly.
   a) What is the chance of getting 7 heads and 3 tails in ten tosses if the heads and tails are equally likely?
      Solution: We should assume the tosses are independent, so this is a binomial situation. The probability is

$$\binom{10}{7}(1/2)^7(1/2)^3 = \frac{120}{1024} = \frac{15}{128} \doteq .117$$

      (that is, about a 12% chance).

b) *Given that the first 3 tosses come up heads*, what is the chance that the first tails comes up on the 7th toss if heads and tails are equally likely? (Note: This is a conditional probability.)
Solution: The chance that the first three tosses come up heads are

$$(1/2)(1/2)(1/2) = (1/2)^3 = 1/8$$

The chance that the first three tosses come up heads *and* then three more tosses come up heads before the first tails are $(1/2)^6 \cdot (1/2) = 1/128$. Hence the conditional probability we are asking for is

$$\frac{(1/2)^7}{(1/2)^3} = (1/2)^4 = 1/16.$$

c) How do your answers to (a) and (b) change if you know that on each toss, the probability of a head is .6 and the probability of a tail is .4 (that is the coin is weighted)?
Solution: In (a), the probability changes to

$$\binom{10}{7}(.6)^7(.4)^3 \doteq .215$$

In (b) the probability changes to

$$\frac{(.6)^6 \cdot (.4)}{(.6)^3} = (.6)^3 \cdot (.4) = .0864$$

In both (b) and here, note that the form is $\frac{q^6 p}{q^3} = q^3 p$, where p is the probability of a tail and q is the probability of a head. In other words, this is the same as ignoring the first three tosses, starting from the fourth, and asking what is the probability that the first tails comes on the fourth toss after the first three.

5) Assume that everyone starts school at age 5 and stays in school at least until age 16. True or False and explain:

a) The median years in school would be greater than 11.
Solution: True: the number of years in school would be at least 11 for everyone, so the median years in school would have to be larger. Recall the median is the number such that half the people have that number of years in school, and half the people have that number of years in school or more.

b) The average years in school would be greater than the median.
Solution: Almost certainly true. The median is probably somewhere around 13 or 14 (that is, finished high school and some college). But there are some people who finish college and get one or more graduate degrees, ending up with $12 + 4 + 5 = 21$ years in school or more. The presence of those people in the average would tend to pull the average up larger than the median. (This is positively skewed distribution.)

6) One ticket is drawn at random from each of the two boxes: (i) 1 2 3 4 5 and (ii) 1 2 3 4 5 6 . The tickets in the first box are blue and those in the second box are green, so you can tell which box the ticket came from.

a) What is the probability that one of the numbers is 2 and the other is 5?
Solution: There are $5 \cdot 6 = 30$ pairs (m, n) representing the draws from the two boxes, with $1 \leq m \leq 5$ and $1 \leq n \leq 6$. "At random" should mean that all of those are equally likely. There are two pairs as stated: $(2, 5)$ and $(5, 2)$ so the probability is $2/30 = 1/15 \doteq .067$.

b) What is the probability that the sum of the numbers is 7?
Solution: That can happen in these 5 ways: $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2)$. So the probability is $5/30 = 1/6 \doteq .167$.

c) What is the probability that one numbers is (strictly) bigger than twice the other? (Hint: List all the cases where that is true and count!)
Solution: This happens for (and only for)

$$(1, 3), (1, 4), (1, 5), (1, 6), (2, 5), (2, 6), (3, 1), (4, 1), (5, 1), (5, 2), (6, 1), (6, 2)$$

(I listed these by thinking of all possible values of the first entry of the pair–the draw from the first box, then seeing which draws from the second box would satisfy the condition that one number was strictly more than twice the other.) The probability is $12/30 = 2/5 = .4$.

7) Body lengths of a population of immature lobsters are *uniformly distributed* on the interval 15 to 22 cm.

a) If a single lobster is selected at random, what is the probability that its body length is between 19 and 20 cm?
Solution: The corresponding uniform pdf has value 1/7 for lengths between 15 and 22 inclusive, but zero for all other lengths. The probability is

$$\frac{20 - 19}{22 - 15} = \frac{1}{7}$$

(This is the area under the graph of the pdf for $19 \leq x \leq 20$.

b) If 6 lobsters are selected with replacement from the population, what is the probability that 4 of them have body lengths between 19 and 20?
Solution: Binomial:

$$\binom{6}{4} (1/7)^4 (6/7)^2 \doteq .0046$$

(about .5%, so very unlikely).

8) In a large class, the average score on a test was 75 out of 100 and the SD was 10. The scores followed a normal distribution.

a) A single test paper is selected randomly out of the pile. What is the probability that the score was between 75 and 85?

b) Using the standard normal table and standardizing as we discussed in class,

$$P(75 < X < 85) = P\left(\frac{75-75}{10} < Z < \frac{85-75}{10}\right) = P(0 < Z < 1) = .3413.$$

b) Same question as a) but the score is between 60 and 80?
Solution: Same method:

$$P(60 < X < 80) = P\left(\frac{60-75}{10} < Z < \frac{80-75}{10}\right)$$
$$= P(-1.5 < Z < .5)$$

$\& = .4332 + .1915 = .6247$

c) What was the 80th percentile score (the score such that 80% had scores less than or equal to that, and only 20% had scores greater than or equal to that)? Round to the nearest whole number.
Solution: We are looking for the score c such that

$$P(X \leq c) = P\left(Z \leq \frac{c-75}{10}\right) = .80$$

From the standard normal table the closest z to having $P(Z < c) = .8$ comes from $c = .84$ (note we are including the .5 for negative values of Z too, so we are looking for an area equal to $.8 - .5 = .3$ and $z = .84$ gives an area .2995 which is close enough to .3 for our purposes.) Then $\frac{c-75}{10} = .84$ implies $c = 75 + 8.4 = 83.4$, or 83 after rounding to the nearest whole number.

9) George is playing roulette (where, in the US version, there are 18 red numbers, 18 greens, and 2 additional slots, for a total of 38 possible ways for the ball to land). Assume the roulette wheel is not biased or "loaded" so that each of the 38 outcomes is equally likely. George starts with \$100. For his first 50 spins of the wheel, he bets \$1 on 17, his favorite number (he can win if either 17 red or 17 green comes up). Unfortunately, he loses each time, losing \$50 in the process. Ever the optimist, George decides his luck must change on the next spin of the wheel because 17 is "due" (it hasn't come up 50 spins in a row, after all). So he bets his remaining \$50 on 17 one more time. Donald's significant other Laura, who always brings George back to earth, says he is a fool and that he's likely to lose again. Is George right, is Laura right, or are neither of them right? Explain your answer.
Solution: George has fallen prey to the "gambler's fallacy." Even though 50 consecutive non-17's have come up, the probability of a 17 on the next roll is still the same as it is on every roll: $2/38 = 1/19$. There is no such thing as "being due for a 17" in this setting because the rolls are independent.