MONT 105Q – Mathematical Journeys
Solutions for Practice Problems for Midterm Exam
March 17, 2016

Sample Mathematical Questions

Note: There are somewhat more questions and more parts to these questions than I could reasonably ask you to do in the actual exam (because of the limited time available). This should give you an idea of the range of possible topics and styles of questions I might include in the exam, though.

A. The following data set has $n = 9$: $23, 28, 40, 44, 47, 50, 51, 54, 55$

1) Find the "5-number" summary for this data set (min, 1st quartile, median, 3rd quartile, max).
Solution: The numbers are already in increasing order so we can see min = 23, the median is the "middle" element = 47, the max = 55 and then (according to the convention we used in class), $Q_1 = \frac{28+40}{2} = 34$ (the median of the 4 data items strictly below the overall median) and $Q_3 = \frac{51+54}{2} = 52.5$ (from the 4 data items strictly above the median).
2) Draw the corresponding "box and whisker" plot.
Solution: Should show a "box" extending from $Q_1 = 34$ to $Q_3 = 52.5$, with a vertical line at 47, plus "whiskers" extending from 23 to 34 and 52.5 to 55.
3) Construct a frequency histogram for the data using 4 equal bins on the range 20 to 60. Explain how you are treating data that falls at a bin boundary. (Any consistent method is OK.)
Solution: Say we decide to place data that falls at a bin boundary into the higher (right) bin each time. Then the four bins yield counts:

$$[20, 30) : 2$$
$$[30, 40) : 0$$
$$[40, 50) : 3$$
$$[50, 60] : 4$$

The histogram shows 4 boxes, the first of height 2, the second of height 0, the third of height 3, the last of height 4.
4) One measure of "skewness" of a data set is the quantity

$$B = \frac{Q_3 - 2 * \text{median} + Q_1}{Q_3 - Q_1}.$$

Compute that statistic for this data set. What does it seem to indicate? Does this seem reasonable from the box plot?
Solution: We get

$$B = \frac{52.5 - 2 * 47 + 34}{52.5 - 34} = -0.41$$

This indicates a slight skew to the left (i.e. toward lower values). That should seem reasonable since the median is farther from Q_1 than it is from Q_3, and the left "whisker" is also quite a bit longer.

5) Compute the SD of the data set. How many of the data values lie within two SD's of the mean?
Solution: To compute the SD we also need the mean:

$$\overline{x} = \frac{23+28+40+44+47+50+51+54+55}{9} = 43.\overline{5} \doteq 43.6$$

(rounding). (Note that this also confirms what we were saying in part 4 – the mean is smaller than the median which indicates a skew to the lower values.) Then from the basic formula

$$SD \doteq \sqrt{\frac{(23-43.6)^2+(28-43.6)^2+\cdots+(55-43.6)^2}{9}} \doteq 10.7$$

(rounded again). This could also be computed by the equivalent, more efficient, formula we mentioned in class:

$$SD = \sqrt{\frac{23^2+28^2+\cdots+55^2}{9} - (43.\overline{5})^2} \doteq 10.7$$

(Practical note: If you use the rounded value 43.6 here, you'll get something a bit smaller than the 10.7 – a rounding artifact.) The numbers within two standard deviations of the mean are (roughly) $43.6 - 2(10.7) \leq x \leq 43.6 + 2(10.7)$, or $22.2 \leq x \leq 65$. *All nine* of the numbers in the data set satisfy those inequalities.

B. The aces, kings, and queens from a standard deck of cards are removed and placed in a stack of twelve cards by themselves (there are three hearts, three spades, three diamonds and three clubs).

1) If a single card is selected at random from the stack, what is the probability that it is a heart?
Answer: $\frac{3}{12} = \frac{1}{4}$.
2) If a single card is selected at from from the stack, what is the conditional probability that the card is a queen, given that it is a spade?
Answer: $\frac{1}{3} = \frac{\frac{1}{12}}{\frac{3}{12}}$.
3) If a single card is selected from the stack, consider the two events: (1) the card is an ace, and (2) the card is a diamond. Are these independent?
Answer: Yes. $P(\text{ace}) = \frac{4}{12} = \frac{1}{3} = P(\text{ace}|\text{diamond})$.

3') Same question as 3) but on the stack of 11 cards obtained by removing the queen of diamonds.
Answer: No. $P(\text{ace}) = \frac{4}{11} \neq P(\text{ace}|\text{diamond}) = \frac{1}{2}$.

4) Three random draws *with replacement* are made from this stack of cards. What is the probability that all three cards are spades?

Answer: With replacement, there is a $\frac{3}{12}$ chance of getting a spade on each draw. So the probability of getting three spades is $\frac{3}{12} \cdot \frac{3}{12} \cdot \frac{3}{12} = \frac{1}{64}$.

5) Three random draws *without replacement* are made from the whole stack of 12 cards. What is the probability that all three are clubs?
   Answer: $\frac{3}{12} \cdot \frac{2}{11} \cdot \frac{1}{10}$.

C. (Short answer) Suppose that a researcher collects 80 individuals of the Atlantic surf clam. These clams can be found at levels down to about a meter in the sand, and larger clams tend to live at deeper levels. The researcher finds an average shell width 10.2 cm. Think of this as a sampling process.

1) What is the population? What is the sample?
   Answer: The population might be all clams of this genus (or possibly all the clams of this genus at this particular location, depending on what you might want to be studying). The sample is the 80 particular clams collected.
2) Is the 10.2 a value derived from particular data (a "statistic") or is it a property of the population as a whole (a "parameter")?
   Answer: It's a statistic since it's computed from the shell widths of those 80 individual clams.
3) Would the researcher *know* the population mean in this circumstance?
   Answer: Most probably not. The idea of doing the collection and measurements might be to estimate the population mean.
4) What additional information would the researcher need in order to understand the distribution of shell widths? Describe how that would be estimated and how that would be interpreted.
   Answer: In addition to the sample mean, the other big statistics to look at would be the sample SD, or other measures of the distribution such as the "five number summary" as in question A above.
5) What could we say about the distribution of sample means of samples of size 80 (even if the distribution of the individual shell widths was unknown)?
   Answer: With samples that large, the Central Limit Theorem would "kick in" and say that the sample means would have an approximately normal distribution with $\mu =$ population mean and SD $= \frac{\sigma}{\sqrt{80}}$, where $\sigma =$ population standard deviation.
6) If you knew that reasearcher was being lazy about digging and the clams he collected were all taken from sand levels no deeper than 10cm, would that be a *simple random sample*? What kind of bias might this process introduce?
   Answer: Not a simple random sample, because of a selection bias including fewer larger clams.

D. Suppose that a large data set of air temperature readings is normally distributed with $\mu = 18.6°$C and $\sigma = .2°$ C. The "z-score" of a temperature reading x is computed by

$$z = \frac{x - \mu}{\sigma}$$

1) What would be the z-score of a reading of $x = 17.9°$?

Answer:

$$z = \frac{17.9 - 18.6}{.2} = -3.5$$

(i.e. this x value is 3.5 standard deviations below the population mean. It would be very unusual to observe a value this small!)

2) What temperature reading would correspond to a z-score of 1.4?
Doing some easy algebra,

$$z = \frac{x - 18.6}{.2} = 1.4 \quad \text{when} \quad x = (1.4)(.2) + 18.6 = 18.88$$

3) Based on this information, if a temperature reading x is selected at random from the data set, what is the probability that $18.2° \leq x \leq 18.9°$?
Answer: This corresponds to the range of z-scores

$$\frac{18.2 - 18.6}{.2} \leq z \leq \frac{18.9 - 18.6}{.2},$$

or $-2 \leq z \leq 1.5$. From the standard normal table, the probability that z is in this range is

$$.5 - P(z > 2) + .5 - P(z > 1.5) = .5 - .0228 + .5 - .0668 = .9104$$

4) Based on this information, if a temperature reading x is selected at random from the data set, what is the probability that $x \geq 18.9°$?
Answer: .0668

5) Based on this information, if a temperature reading x is selected at random from the data set, what is the probability that $x \leq 18.2°$?
Answer: .0228

6) Based on this information, if $n = 10$ measurements are taken, what is the probability that exactly 5 of them are greater than 18.8?
Answer: The probability that any one $x > 18.8$ is

$$P(x > 18.8) = P\left(z > \frac{18.8 - 18.6}{.2}\right) = P(z > 1) = .1587$$

Out of the 10 measurements (assuming independence), probability that exactly 5 of them are > 18.8 is given by the binomial probability formula:

$$\binom{10}{5}(.1587)^5(.8413)^5$$

(It's OK to leave the answer in this form since it's clear where it comes from! If you really need a single number, this works out to about .01 (a 1 in 100 chance).