College of the Holy Cross, Spring Semester 2019
MONT 105N – Analyzing Environmental Data
Solutions for Midterm Exam, March 25

I. Both parts of this question deal with the data set:

$$6 \quad 5 \quad 11 \quad 7 \quad 5 \quad 10 \quad 16 \quad 10 \quad 9$$

A) (15) Find the 5-number summary, and draw the "box and whisker plot." (Use the inclusive method for the quartiles.)

Solution In order, the numbers are

$$5 \le 5 < 6 < 7 < 9 < 10 \le 10 < 11 < 16$$

We have min $= 5$, max $= 16$, and $Q_1 = 6$, median $= 9$, $Q_3 = 10$. The box extends from 6 to 10 with a vertical bar at 9. The whiskers extend to 5 on the left and 16 on the right

B) (15) Given that the SD of this data set is $S \doteq 3.52$, how many data values are within two SD's of the mean?

Solution: We need to compute

$$\overline{x} = \frac{5+5+6+7+9+10+10+11+16}{9} = \frac{79}{9} \doteq 8.78$$

We have $\overline{x} - 2S = 1.72$ and $\overline{x} + 2S \doteq 15.84$. The interval between these two endpoints contains *eight out of nine*, or about 88.9% of the data values. This is consistent with Chebyshev's Theorem, which says *at least* $\left(1 - \frac{1}{2^2}\right) \times 100\% = 75\%$ of the data values must lie within 2 SD's of the mean.

II. A weighted (biased) coin is tossed repeatedly. On any one toss, the probability of a head is $p = .44$.

A) (10) What is the probability of getting 2 heads and 8 tails in $n = 10$ tosses?

Solution: Using the binomial p.m.f., if X is the number out of the 10 tosses on which a head comes up,

$$P(X = 2) = \binom{10}{2}(.44)^2(.56)^8 \doteq .084$$

or about an 8.4% chance.

B) (10) Is your answer in A the same as the probability of drawing 2 red and 8 black balls on 10 draws from an urn containing a total of 44 red and 56 black balls, if the draws are done *without replacement*? Explain.

Solution: No, it is not the same. If the balls are drawn without replacement, the draws are not independent and the binomial p.m.f. formula does not apply. (We need the hypergeometric formula then:

$$P(X = 2) = \frac{\binom{44}{2} \cdot \binom{56}{8}}{\binom{100}{10}} \doteq .0776$$

C) (5) *Given that the first 3 tosses come up heads*, what is the probability that the first tails comes up on the 7th toss? (Note: This is a conditional probability.)

Solution We want $\frac{(.44)^3(.56)^3(.44)}{(.44)^3} = (.56)^3(.44) \doteq .0773$

III. Males over 20 years old in the U.S. at present have an average height of 178 cm (about 5' 10"), with $SD = 10$ cm. Their average weight is 76 kg and with $SD = 12$ kg.

A) (10) If a male over 20 years old is selected at random, what is the probability that he has height between 178 and 190 cm (between about 5' 10" and about 6' 3")?

Solution: We standardize using the mean and SD for the height and use the normal table:

$$P(178 < X < 190) = P\left(0 < Z < \frac{190 - 178}{10}\right) = P(0 < Z < 1.2) = .3849$$

(about 38.5% chance).

B) (10) What is the probability that a male over 20 years old, selected at random, has a weight between 70 and 90 kg (between about 154.5 lb and about 198 lb)?

Solution: Now we standardize using the mean and SD for the weight:

$$P(70 < X < 90) = P\left(70 - 7612 < Z < \frac{90 - 76}{12}\right) \doteq P(-.5 < Z < .86)$$

By the symmetry of the normal curve, this equals

$$P(0 < Z < .5) + P(0 < Z < .86) = .1915 + .3051 = .4966$$

C) (5) (True/False and Explain): It would be reasonable to compute the probability that a male 20 years old, selected at random has a height between 178 and 190 cm *and* a weight between 70 and 90 kg by *multiplying your answers* from A and B.

Solution: False: It would not be reasonable because doing that is assuming that the height and the weight are *independent*, whereas this is clearly not true. Larger heights are *correlated* with larger weights in the sense that taller bodies also tend to have higher weights.

Essay (20)

One very common everyday situation where we are given information phrased in probabilistic terms is when weather forecasts for a given geographical area include an estimated *probability of precipitation* for some time period in the future. But what does one of those probabilities mean, really? What is the exact formula used to calculate a probability of precipitation? Explain by indicating exactly what it means, for instance, if a weather forecast says there is an 80% chance of rain in Worcester tomorrow. Can it mean more than one thing? What does the National Oceanographic and Atmospheric Administration (NOAA) say about this and what is the justification they give for doing it this way? Do you think just the number gives an adequate indication to an average person? Would it help to educate people better about how these probabilities are actually computed? Is there a better way to indicate how likely certain weather conditions might be?

Model Answer: The probability of precipitation, PoP, is defined as a product:

$$\mathrm{PoP} = C \times A,$$

where C = the confidence or estimated probability that precipitation will occur *somewhere* in the forecast area, and where A = the percent of the area that will receive measureable precipitation, if it occurs at all. For instance if the forecast said there was an 80% chance of rain in Worcester tomorrow, that could mean:

- The forecaster estimated $C = .8$ or was 80% certain that rain would occur, and if it does, the rain will happen over $A = 1$ or 100% of the city, or
- The forecaster estimated $C = 1$ or was 100% certain that rain would occur, but if it does, it would only rain over $A = .8\%$ or 80% of the city, or
- (Among infinitely many other possibilities), the forecaster estimated $C = .9$ or was 90% certain rain would occur and if it does, it would cover about $A \doteq .889$ or about 89% of the area.

The justification offered for doing it this way is that forecasts are usually made for rather large geographical areas and weather can vary a lot from point to point within these areas. The forecaster is expressing a combination of degree of confidence in the assertion that the given precipitation event will happen, and an estimate of the area it will cover if it does. (Personal opinion): Unless they are specifically told how this works, though, I really doubt that most people understand this, and many ("non-math-people") might not understand it even if they are told(!) It might be better (i.e. more informative and clearer) to report the C and A values separately rather than multiplying them!