Kaleb Burris
STAT F300 Statistics, Dr. Short
Worksheet 2, due Monday January 23, 4:15pm

---

1. The difference between countably infinite and uncountable infinite sets. For each of the following sets, state whether it is countably infinite or uncountably infinite. (All are infinite.)

(a) $[0,1] \cup [3,5]$: $\boxed{\text{Uncountably infinite}}$

(b) $\ldots,-3,-2,-1,0,1,2,3,\cdots$: $\boxed{\text{Countably infinite}}$

(c) $0, 0.1, 0.2, 0.3, 0.4, \cdots$: $\boxed{\text{Countably infinite}}$

(d) $[0,1] \cap [0.5, 1.5]$: $\boxed{\text{Uncountably infinite}}$

(e) $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \cdots$: $\boxed{\text{Countably infinite}}$

(f) The rational numbers, i.e. all numbers that can be written as $\frac{n}{m}$, where n and m are integers and $m \neq 0$: $\boxed{\text{Countably infinite}}$

2. Without doing any computations, which of the following data sets has the smallest standard deviation (sd), and which has the largest sd? Explain; your answer should include terms such as "center" and "spread", and how far are many or most observations from the mean of the data. Note that each data set has 10 observations. It might help to draw a dot plot for each and use it to help explain how you know your answer is correct (without ever finding the value of the sd's)

A: {2, 2, 2, 2, 2, 20, 20, 20, 20, 20}

B: {2, 11, 11, 11, 11, 11, 11, 11, 11, 20}

C: {2, 4, 6, 8, 10, 12, 14, 16, 18, 20}

Standard deviation is the root of the sum of the square (those cancel out) of each value minus the "average value". Practically, it's how far the data set is spread out from a central mean, the larger the difference between each value in the data set, the larger the standard deviation. Because of this, data set B would have the smallest standard deviation as it's made up of almost all 11s and 2 relatively in-between outliers, so the sum part of sd would be relatively small. Data set A would have the largest standard of deviation as it's made up of two halves of numbers that are relatively far apart. The range of numbers used for the data sets is [2,20], and A is made up entirely half-and-half of the extreme values that would accumulate a sum larger than data set A or C.

3. Give an example of a concrete population. Give an example of a probability question and an inferential statistics question for this population.

A concrete population: The number of chew toys my dog as destroyed.

Probability: If I buy my dog a bulk box of 120 dog toys at the beginning of the year with a probability that my dog destroys a toy at 2.3% daily, how many toys is my dog destroying a month? Will I have any toys left at the end of the year?

Inferential: If I observe that the 10th toy is destroyed after 40 days, what is the chance a toy is destroyed daily? How many would I then expect to be destroyed in total by the end of the year?

4. Give an example of a hypothetical population. Give an example of a probability question and an inferential statistics question for this population.

Hypothetical Population: All toys my dog will destroy in her lifetime.

Probability Question: What is the daily chance any toy exposed to my dog will be destroyed over her lifetime if I bought her 3000 toys and only 40 remain after 11 years?

Inferential Question: If the average weight of a toy is 2.5 lbs, and she commits actions resulting in a standard deviation of 1.2 lbs, what can be said about the average weight of each toy after she's done with it?