Sampling Distributions

Suppose we have some population distribution we are interested in understanding or using to make a decision. The truth (parameters) about the population is usually unknown. Usually the best we can do is to take a sample from the population and find the statistics for the sample. We can assume the population has a center, variability, and probabilities associated with the outcomes even though we do not know them and that the sample statistics estimate these values. The obvious question arises, “How are the sample statistics related to population parameters?” The main purpose of this activity is to answer this question.

In this activity we will explore sampling distributions. Specifically we will consider the sampling distribution of the sample mean given a sample of size n. Inferential statistics involves using sampling distributions to make decisions about the original population the sample is selected from.

Consider the data set $\{X: X \epsilon\ 1, 2, 3, 4, 5, 6\}$ (with all elements equally likely) as our population data set.

1. Can you think of a scenario in which you might get observations from 1 to 6 with all values being equally likely? This is our experiment. rolling a die with 6 sides

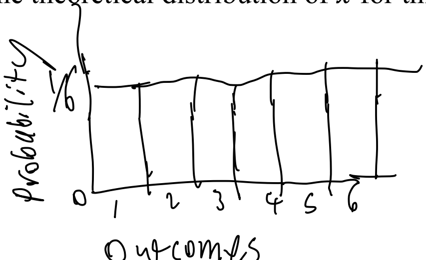
2. How many different outcomes are there for x?

   6

3. Find the probability that x will be 6, $P(x = 6)$. What is true about probability for each of the elements in the data set? $P(x=6) = 1/6$. The probability is the same for all outcomes

4. What type of distribution is this? Look at the distributions in note set 2 if you are having difficulty deciding this. Uniform probability

5. Sketch a graph of the theoretical distribution of x for this data set.



6. Since this is a population distribution of x we will be able to find μ, σ^2, and σ. You can find the formulas for these parameters in note set 2. Show the work necessary to find each.

$$\mu = \frac{(1+2+3+4+5+6)}{6} = \frac{21}{6} = 3.5$$

$$\sigma^2 = \frac{(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2}{6} = 0$$

$$\sigma^2 = 2.917 \qquad \sqrt{2.917} = 1.708 = \sigma$$

Instead of just a single sample, suppose we were to consider all the possible samples for a given sample of size n. The experiment now requires that we roll the die twice and find a **mean** instead of just rolling it once and making a single observation. Remember if we were to take a sample of size n, we can find the sample mean using the formula $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n}$. One sample of size $n = 2$ is rolling a 2 and then a 4. To find the mean we would sum 2 and 4 to get 6 and then divide by the sample size 2 to get $\bar{x} = \frac{6}{2}$ or $\bar{x} = 3$. Make sure you understand the denominator is 2 because the sample is size 2.

$1,1 = 1$

$1,2 = \frac{3}{2}$

$4,5 = \frac{9}{2} = 4.5$

$6,2 = \frac{8}{2} = 4$

1. Provide another possible example for the two values of x and find the sample mean. Clearly there are other possibilities for x and thus $\bar{x}$. Find a few more so that you begin getting ideas about the differences between x and $\bar{x}$. Once you have done this a few times answer the question, "In general is $x = \bar{x}$?". No

2. How many different combinations of x are there? In other words, how many samples of size 2 are there given the population distribution? From these samples you can calculate different sample means $\bar{x}$. How many different $\bar{x}'s$ are there? You may want to make a list of all the possibilities.

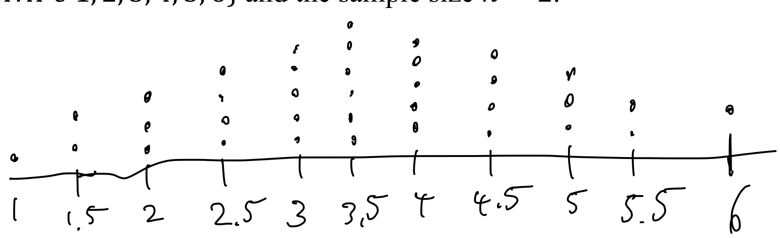
   $6^2 = 36$ different samples

   11 unique $\bar{x}$'s

3. Now let's think further about a distribution of possible sample means given a sample of size 2. Do you think the distribution of x and the distribution of $\bar{x}$ are going to be the same? What is the lowest possible $\bar{x}$? What is the highest possible $\bar{x}$? No, they are different.

   $\bar{X}_{min} = 1$ $\bar{X}_{max} = 6$

4. Using a dot plot, construct the theoretical sampling distribution of $\bar{x}$ given our population data set $\{X: X \epsilon\ 1, 2, 3, 4, 5, 6\}$ and the sample size $n = 2$.



5. What is the mean of this distribution above? Use the formula below to find it.

$$\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \cdots + \bar{x}_N}{N}$$

$\mu_{\bar{x}} = 3.5$

6. The theoretical distribution of $\bar{x}$ is also a probability distribution for $\bar{x}$ given a sample of size 2. Using the dot plot find the following and then compare these to $P(x)$, the original population probabilities.

$P(\bar{x} = 1) = 1/36 = .0278$	$P(x = 1) = 1/6 = .167$
$P(\bar{x} = 2.5) = 4/36 = .111$	$P(x = 2.5) = 0$
$P(\bar{x} = 3.5) = 6/36 = .167$	$P(x = 3.5) = 0$
$P(3 \le \bar{x} \le 4) = 16/36 = .444$	$P(3 \le x \le 4) = 2/6 = .333$

$\frac{7}{2}$ most likely $\bar{x}$

(min)	1,1 = $\frac{1}{2}$	2,1	3,1			
	1,2 = $\frac{3}{2}$	2,2	3,2			
	1,3 = $\frac{4}{2}$	2,3	3,3			
	1,4 = $\frac{5}{2}$	2,4	3,4			
	1,5 = $\frac{6}{2}$	2,5	3,5			
	1,6 = $\frac{7}{2}$	2,6 = $\frac{8}{2}$	3,6 = $\frac{9}{2}$	4,6 = $\frac{10}{2}$	5,6 = $\frac{11}{2}$	6,6 = $\frac{12}{2}$ (max)

$1 + 2(1.5) + 3(2) + 4(2.5) + 5(3) + 6(3.5) + 5(4) + 4(4.5) + 3(5) + 2(5.5) + 1(6)$

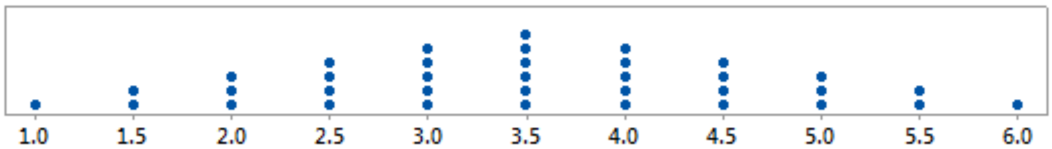
$1 + 3 + 6 + 10 + 15 + 21 + 20 + 18 + 15 + 11 + 6 = 126$

$$\frac{126}{36} = 3.5$$

Sampling Distribution Activity Part 2

Understanding the way Standard Deviation Measures Variation

Recall the theoretical sampling distribution of $\bar{x}$ with $n = 2$ below.



Also, recall the mean of this distribution $\left(\mu_{\bar{x}} = \frac{1+1.5+1.5+2+\cdots+5.5+5.5+6}{36}\right)$ was $\mu_{\bar{x}} = 3.5$. Further we inferred from this distribution that in general given some population $\mu_{\bar{x}} = \mu$.

1. This time we will be exploring the standard deviation of the sampling distribution ($\sigma_{\bar{x}}$). We will begin by finding the variance and the standard deviation of this distribution.

$$\sigma_{\bar{x}}^{\;2} = \sum_{i=1}^{N} \frac{(\bar{x}_i - \mu_{\bar{x}})^2}{N}$$

$$= \sum \frac{(1-3.5)^2+(1.5-3.5)^2+(1.5-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2}{36}$$

$$\sigma_{\bar{x}} = 3.5$$

2. We first need to get some ideas about the probabilities associated with specific means given a sample size 2. Using the sampling distribution dot plot above find and list $P(\bar{x} = k)$ for each sample mean.

$P(\bar{x} = 1) = \frac{1}{36} = .028$

$P(\bar{x} = 1.5) = \frac{2}{36}$

$P(\bar{x} = 2) = 3/36$

$P(\bar{x} = 2.5) = 4/36$

$P(\bar{x} = 3) = 5/36$

$P(\bar{x} = 3.5) = 6/36$

$P(\bar{x} = 4) = 5/36$

$P(\bar{x} = 4.5) = 4/36$

$P(\bar{x} = 5) = 3/36$

$P(\bar{x} = 5.5) = 2/36$

$P(\bar{x} = 6) = 1/36$

3. For the sampling distribution with $n = 2$, find $P(3 \le \bar{x} \le 4)$, $P(2.5 \le \bar{x} \le 4.5)$, and , $P(2 \le \bar{x} \le 5)$. We are going to compare these to the sampling distribution with $n = 3$.

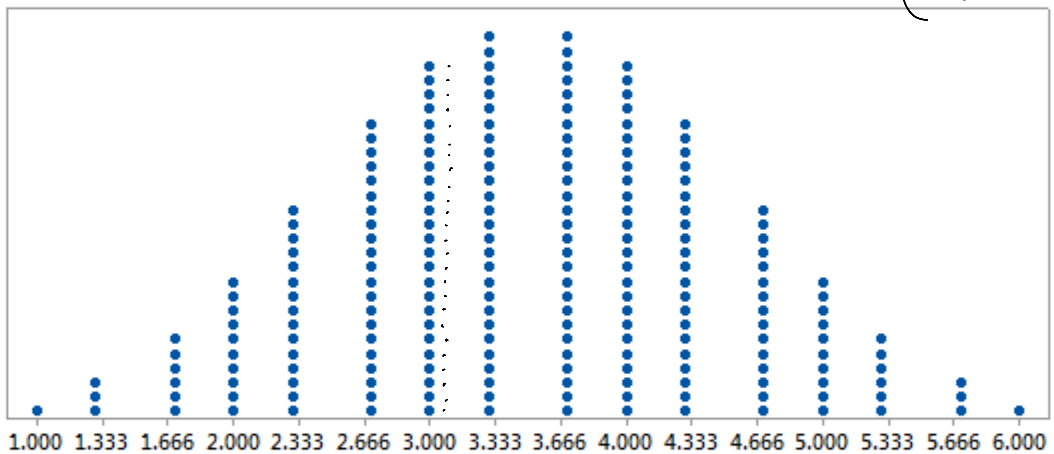
$P(3 \le \bar{x} \le 4) = \frac{16}{36} = 0.444$

$P(2.5 \le \bar{x} \le 4.5) = .667$

$P(2 \le \bar{x} \le 5) = .833$

The theoretical sampling distribution, given $\{X: X \epsilon\ 1, 2, 3, 4, 5, 6\}$, of $\bar{x}$ with $n = 3$ is below. Notice this sampling distribution has different possibilities for the sample means and quite a few more total sample means than the one with $n = 2$. How many sample means does this distribution have?

$(6)^3 = 216$



1. We will begin by finding the mean of this sampling distribution. Keep in mind that we do not need to actually perform the calculation because we know the population mean. Go ahead and fill in the formula below but do not attempt to perform the calculation.

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{N} \bar{X}_i}{N}$$

2. Next we will find the variance and standard deviation for this distribution. Again fill in the formula but do not attempt to perform the calculation.

$$\sigma_{\bar{x}}^{\ 2} = \sum_{i=1}^{N} \frac{(\bar{X}_i - \mu_{\bar{X}})^2}{N}$$

$$\sigma_{\bar{x}} = \sqrt{\sigma_{\bar{X}}^2}$$

3. Again we need to get some ideas about the probabilities associated with specific means given a sample size 3. Using the sampling distribution dot plot above find and list $P(\bar{x} = k)$ for each sample mean.

$P(\bar{X} = 1) = 1/216$ $P(\bar{X} = 2) = 10/216$ $P(\bar{x} = 3.666) = 27/216$

$P(\bar{X} = 1.333) = 3/216$ $P(\bar{X} = 2.333) = 15/216$ $P(\bar{x} = 4) = 25/216$

$P(\bar{x} = 1.666) = 5/216$ $P(\bar{x} = 2.666) = 21/216$ $P(\bar{x} = 4.333) = 21/216$

$P(\bar{x} = 3) = 25/216$ $P(\bar{x} = 4.666) = P(\bar{x} = 6 - 4.666)$ ... $P(\bar{x} = 6) = P(\bar{x} = 1)$

4. For the sampling distribution with $n = 3$, find $P(3 \leq \bar{x} \leq 4)$, $P(2.3 \leq \bar{x} \leq 4.6)$, and , $P(2 \leq \bar{x} \leq 5)$. Once you have done this begin comparing these values to the sampling distribution with $n = 2$.

$P(3 \leq \bar{x} \leq 4) = \frac{25 + 27 + 27 + 25}{216} = .481$ $P(2 \leq \bar{x} \leq 5) = .907$

$P(2.3 \leq \bar{x} \leq 4.6) = .815$

Notice as we increase the sample size the values for the mean closer to the middle of the distribution are becoming more likely while the values for the mean closer to the boundaries of the sampling distribution are becoming less likely.