Section Q
Distribution of the Sample Mean and the Central Limit Theorem

Up to this point, the probabilities we have found have been based on individuals in a sample, but suppose we want to find probabilities based on the mean of a sample. In order for us to find these probabilities we need to know determine the sampling distribution of the sample mean. Knowing the sampling distribution of the sample mean will not only allow us to find probabilities, but it is the underlying concept that allows us to estimate the population mean and draw conclusions about the population mean which is what inferential statistics is all about.

Sampling Error : The error resulting from using a sample to estimate a population characteristic.

For a variable x and a given sample size n, the distribution of the variable $\overline{X}$ (all possible sample means of size n) is called the sampling distribution of the mean.

Note: The larger the sample size the smaller the sampling error tends to be in estimating a population mean, μ, by a sample mean $\overline{X}$.

Mean of $\overline{X}$: denoted $\mu_{\overline{X}}$

For samples of size n, the mean of the variable $\overline{X}$ equals the mean of the variable under consideration, i.e. $\mu_{\overline{X}} = \mu$, where $\mu_{\overline{X}}$ is the mean of variable $\overline{X}$ and μ is the population mean.

In other words, the mean of all possible sample means of size n equals the population mean.

Example:
The following data represent the ages of the winners (age, in years, at time of award given) of the Academy Award for Best Actress for the years 2012 – 2017.

2012: Meryl Streep	62
2013: Jennifer Lawrence	22
2014: Cate Blanchett	44
2015: Julianne Moore	54
2016: Brie Larson	26
2017: Emma Stone	28

a) Calculate the population mean, μ. $\boldsymbol{\mu = \frac{62+22+44+54+26+28}{6} = 39.333}$

b) The following table consisting of all possible samples with size n = 2 and calculate their corresponding means.

Sample	Mean	Sample	Mean	Sample	Mean	Sample	Mean	Sample	Mean	Sample	Mean
62, 62	**62**	22, 62	**42**	44, 62	**53**	54, 62	**58**	26, 62	**44**	28, 62	**45**
62, 22	**42**	22, 22	**22**	44, 22	**33**	54, 22	**38**	26, 22	**24**	28, 22	**25**
62, 44	**53**	22, 44	**33**	44, 44	**44**	54, 44	**49**	26, 44	**35**	28, 44	**36**
62, 54	**58**	22, 54	**38**	44, 54	**49**	54, 54	**54**	26, 54	**40**	28, 54	**41**
62, 26	**44**	22, 26	**24**	44, 26	**35**	54, 26	**40**	26, 26	**26**	28, 26	**27**
62, 28	**45**	22, 28	**25**	44, 28	**36**	54, 28	**41**	26, 28	**27**	28, 28	**28**

c) Calculate the mean of the sampling distribution of the mean, $\mu_{\bar{x}}$.

(i.e. calculate the mean of the sample means) **N = 36** $\mathbf{\sum x = 1416}$ $\mathbf{\mu_{\bar{x}} = \frac{1416}{36} = 39.333}$

d) What do you conclude about μ and $\mu_{\bar{x}}$? **They are equal.**

Note: The above example is exactly that an example, it is not a proof of $\mu_{\bar{x}} = \mu$.

<u>**Standard deviation of**</u> $\bar{X}$ denoted $\sigma_{\bar{x}}$

For samples of size n, the standard deviation of the variable $\bar{X}$ equals the standard deviation of the population under consideration divided by the square root of n.

So, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sigma_{\bar{x}}$ is smaller than σ

$\sigma_{\bar{x}}$ is sometimes called the **standard error** of the mean.

Example: Using the example above,

a) Calculate the population standard deviation, σ. N = 6 $\sum \bar{x} = 236$ $\sum \bar{x}^2 = 10640$

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \mu^2} = \sqrt{\frac{10640}{6} - 39.3333^2} = 15.0408$$

b) Calculate the standard deviation of the sampling distribution of the mean, $\sigma_{\bar{x}}$.
(i.e. calculate the standard deviation of the sample means) N = 36 $\sum \bar{x} = 1416$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum x^2}{N} - \mu_{\bar{x}}^2} = \sqrt{\frac{59768}{36} - 39.3333^2} = \mathbf{10.635}$$

c) Use the formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ to calculate $\sigma_{\bar{x}}$. $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15.0408}{\sqrt{2}} = \mathbf{10.635}$

d) What do you conclude from parts b and c? **The answers are equal, so** $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

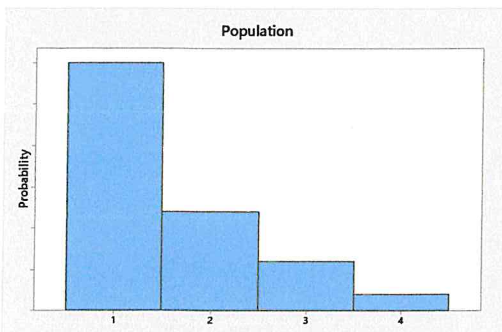
Again, the example is an example not a proof.

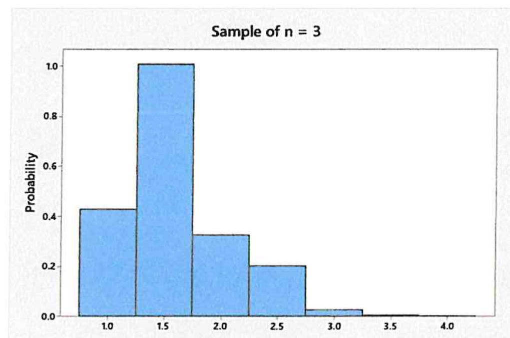
The Sampling Distribution of the Sample Mean for a Normally Distributed Variable

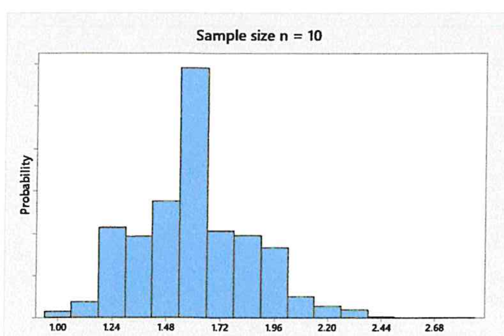
Suppose that a variable x of a population is normally distributed with a mean μ and a standard deviation σ. Then, for samples of size n, the variable $\overline{X}$ is also normally distributed and has mean μ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

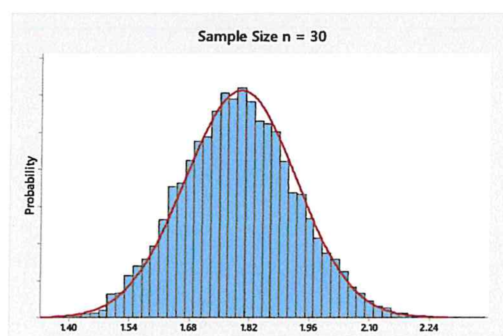
Of course, not all distributions are normal, but given certain conditions we can assume the variable $\overline{X}$ is approximately normally distributed regardless of the distribution of x. This leads to one of the most important theorems in statistics; the central limit theorem.

<u>Central Limit Theorem (CLT)</u> For a relatively large sample size (n > 30), the variable $\overline{X}$ is approximately normally distributed, regardless of the distribution of the variable under consideration. The approximation becomes better with increasing sample size.









The Sampling Distribution of the Sample Mean

Suppose that a variable x of a population has mean, μ and standard deviation, σ. Then, for samples of size n,

1) The mean of $\overline{X}$ equals the population mean, μ, in other words: $\mu_{\overline{X}} = \mu$

2) The standard deviation of $\overline{X}$ equals the population standard deviation divided by the square root of the sample size, in other words: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

3) If x is normally distributed, so is $\overline{X}$, regardless of sample size

4) If the sample size is large (n > 30), $\overline{X}$ is approximately normally distributed, regardless of the distribution of x.

Therefore, we can say, $\overline{X}$ is normally distributed with parameters $\mu_{\overline{X}}$ and $\sigma_{\overline{X}}$, where $\mu_{\overline{X}} = \mu$ and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$.

Note: Since the sampling distribution of the sample mean is normally under certain conditions you can use the normal approximation to find probabilities, therefore you need convert $\overline{X}$ to a z-score.

Converting $\overline{X}$ to a z-score: $Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$

Examples:

1) The times that college students spend studying per week have a distribution that is right skewed with a mean of 8.4 hours and a standard deviation of 2.7 hours. Suppose a random sample of 45 students is selected.

a) What is the sampling distribution of the mean number of hours these 45 students spend studying per week? (i.e. What is the sampling distribution of $\overline{X}$?)

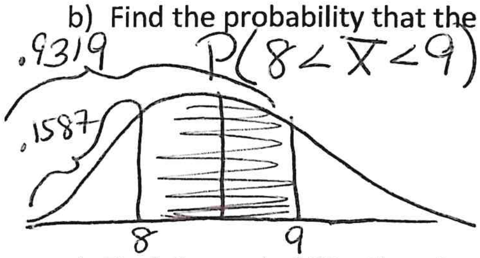
n= 45

μ = 8.4

σ = 2.7

$\overline{X}$ is normally distributed with $\mu_{\overline{X}} = 8.4$ and $\sigma_{\overline{X}} = \frac{2.7}{\sqrt{45}} = .402$

b) Find the probability that the mean time spent studying per week is between 8 and 9 hours.

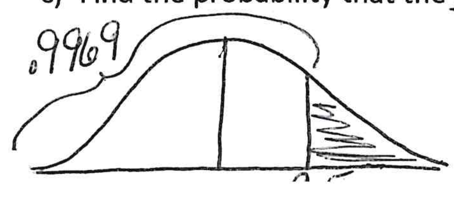


$P(8 < \overline{X} < 9) = .9319 - .1587 = .7732$

$Z = \frac{8 - 8.4}{.402} = -1.00 \Rightarrow .1587$

$Z = \frac{9 - 8.4}{.402} = 1.49 \Rightarrow .9319$

c) Find the probability that the mean time spent studying per week is greater than 9.5 hours.



$P(\overline{X} > 9.5) = 1 - .9969 = .0031$

$Z = \frac{9.5 - 8.4}{.402} = 2.74 \Rightarrow .9969$

2) At an urban hospital the weights of newborn babies are normally distributed, with a mean of 7.2 pounds and standard deviation of 1.2 pounds. Suppose a random sample of 30 is selected.

a) What is the sampling distribution of the mean weight of these newborn babies?

(i.e. What is the sampling distribution of $\overline{X}$?)

n= 30

μ = 7.2

σ = 1.2

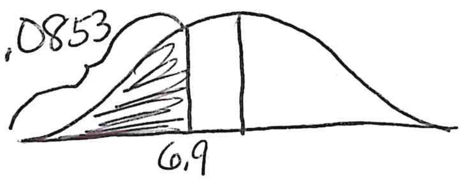
$\overline{X}$ is normally distributed

$\mu_{\overline{X}} = 7.2$

$\sigma_{\overline{X}} = \frac{1.2}{\sqrt{30}} = .219$

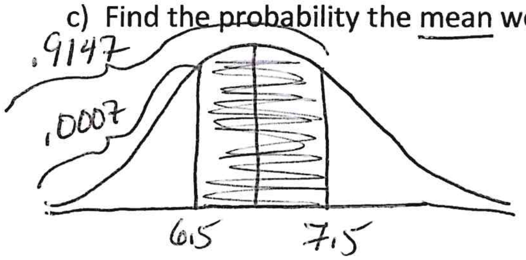
$\overline{X} \sim N(7.2, .219)$

b) Find the probability the mean weight is less than 6.9 pounds? $P(\overline{X}<6.9) = .0853$



$Z = \frac{6.9-7.2}{.219} = -1.37 \Rightarrow .0853$

c) Find the probability the mean weight is between 6.5 and 7.5 pounds? $P(6.5<\overline{X}<7.5) =$

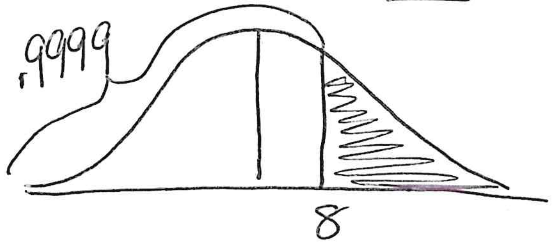


$Z = \frac{7.5-7.2}{.219} = 1.37 \Rightarrow .0007$

$Z = \frac{6.5-7.2}{.219} = -3.20 \Rightarrow .9147$

$.9147 - .0007 = .9140$

d) Find the probability the mean weight is greater than 8 pounds? $P(\overline{X}>8) = 1 - .9999 = .0001$



$Z = \frac{8-7.2}{.219} = 3.65 \Rightarrow .9999$

3) A battery manufacturer claims that the lifetime of a certain battery has a mean of 40 hours and a standard deviation of 5 hours. A simple random sample of 100 batteries is selected.

a) What is the sampling distribution of the mean life of the batteries?

(i.e. What is the sampling distribution of $\overline{X}$?)

n= 100

μ = 40

σ = 5

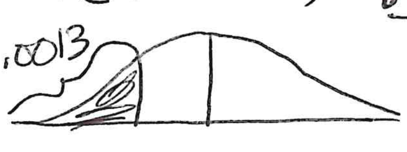
$\overline{X}$ is normally distributed

$\mu_{\overline{X}} = 40$ and $\sigma_{\overline{X}} = \frac{5}{\sqrt{100}} = .5$

$\overline{X} \sim N(40, .5)$

b) What is the probability the mean life is less than 38.5? Would this be unusual? yes, since $.0013 < .05$

$P(\overline{X}<38.5) = .0013$



$Z = \frac{38.5-40}{.5} = -3.00 \Rightarrow .0013$