

## SAMPLING DISTRIBUTIONS

The sampling distribution of the mean:

**CONSIDER THE FOLLOWING VERY, VERY SIMPLISTIC EXAMPLE:**

Suppose that in a population consisting of 5 elements and one wishes to take a random sample of 2. There are 10 possible samples which might be selected.

Population

(N=5)

- 1
- 2
- 3
- 4
- 5

Since we know the entire population, we can compute the population parameters:

$$\mu = 3.0 \quad \sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}} = \sqrt{2} = 1.41$$

[Note: N, not n-1. This is the formula for computing the *population* standard deviation,  $\sigma$ .]

All possible samples of size  $n=2$ :

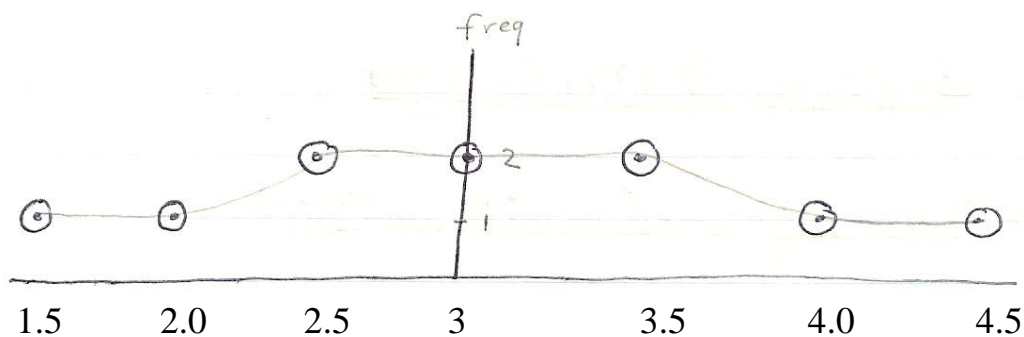
Sample	Mean ( $\bar{X}$ )
(1,2)	1.5
(1,3)	2.0
(1,4)	2.5
(1,5)	3.0
(2,3)	2.5
(2,4)	3.0
(2,5)	3.5
(3,4)	3.5
(3,5)	4.0
(4,5)	4.5
	30.0

The average of all the possible sample means  $E(\bar{X}) = 30.0 / 10 = 3.0$ .

$E(\bar{X}) = \mu$ . This property is called unbiasedness.

Also notice that the variability in the sample means is smaller than the variability in the original population. For this simple example it is best to use the (population) range as a measure of dispersion:  $5 - 1 = 4$  vs.  $4.5 - 1.5 = 3$ . If the sample size were larger, the range of the sample means would be even smaller.

Notice that if we graph the distribution of all possible sample means, it looks almost normal – symmetric, mean=median=mode, etc.



## Central Limit Theorem

If a population distribution is nonnormal, the sampling distribution of  $\bar{X}$  may be considered to be approximately normal for large samples.

In our example, above,  $n$  was only 2 and we saw the beginnings of a normal sampling distribution. Usually we require  $n \geq 30$  to use the central limit theorem; some statisticians require samples of about 50.

This “hypothetical” sampling distribution of the mean, as  $n$  gets large, has the following properties:

- (1)  $E(\bar{X}) = \mu$ .  
It has a mean equal to the population mean.
- (2) It has a standard deviation (called the standard error of the mean,  $\sigma_{\bar{x}}$ , equal to the population standard deviation divided by  $\sqrt{n}$ .
- (3) It is normally distributed.

This means that, for large samples, the sampling distribution of the mean ( $\bar{x}$ ) can be approximated by a normal distribution with

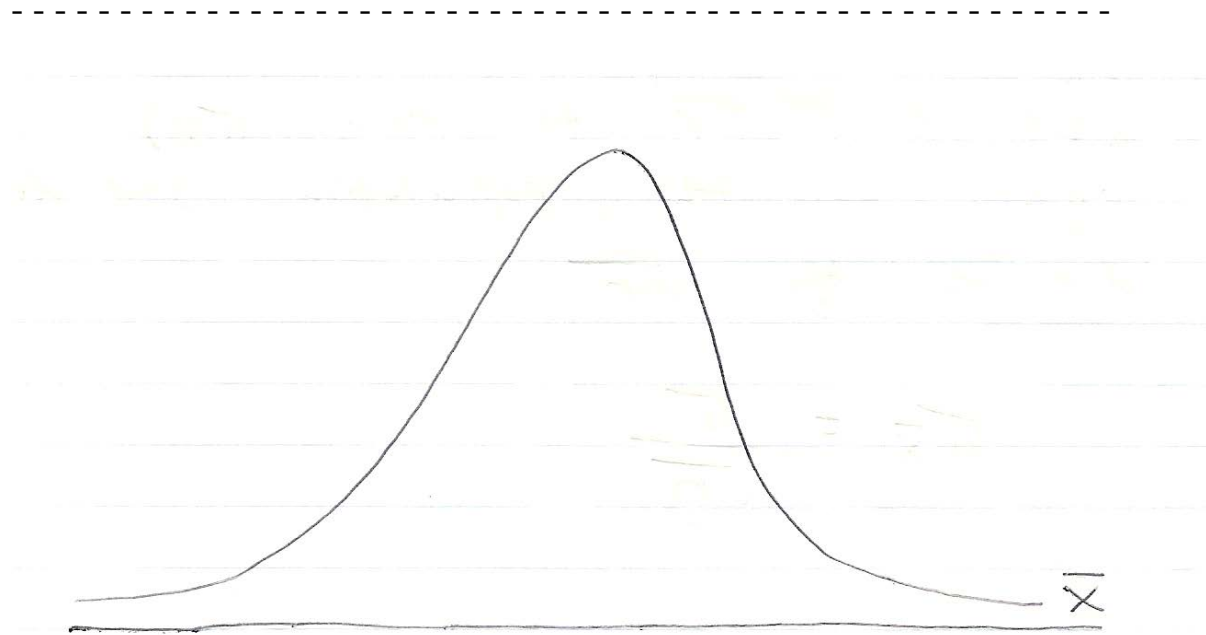
$$E(\bar{x}) = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

-----  
NOTE: When sampling from a finite population without replacement:

$$\left[ \begin{array}{l} \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{n-1}} \\ \text{where } \sqrt{\frac{N-n}{n-1}} \text{ is the finite population correction factor.} \end{array} \right]$$

However, if  $n$  is a very small fraction of  $N$ , this term is about 1.



$\mu$   
The Sampling Distribution of the Mean

Since this is a normal distribution, we can standardize it (transform to  $Z$ ) just like any other normal distribution.

$$Z = \frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} = Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

If  $n$  is large, say 30 or more (some prefer using 50 as a cutoff), use  $s$  as an unbiased estimate of  $\sigma$ .

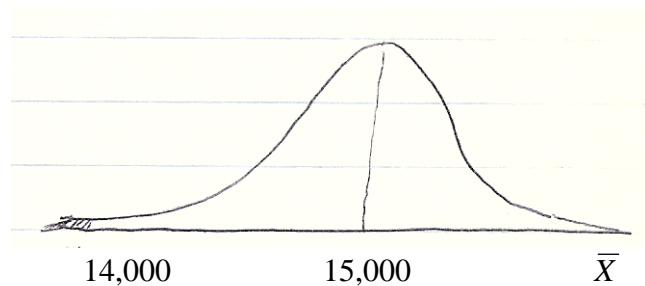
$$Z = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Example:

A manufacturer claims that its bulbs have a mean life of 15,000 hours and a standard deviation of 1,000 hours. A quality control expert takes a random sample of 100 bulbs and finds a mean of  $\bar{X} = 14,000$  hours. Should the manufacturer revise the claim?

We don't know the distribution of the lifetimes of the bulbs. But we DO know the distribution of the mean life of a (large enough) sample of bulbs.

DRAW A PICTURE!



$$Z = \frac{14,000 - 15,000}{\frac{1,000}{\sqrt{100}}} = \frac{-1,000}{100} = -10$$

That's **10** standard deviations away from the mean! Yes, the claim should definitely be revised.

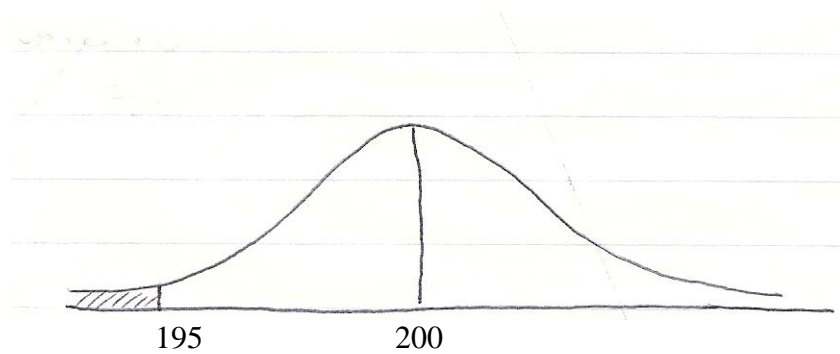
prob < .000000

Example:

Suppose you have steel chains with an average breaking strength of  $\mu=200$  lbs with a  $\sigma=10$  lbs, and you take a sample of  $n=100$  chains.

What is the probability that the sample mean breaking strength will be:

- (a) 195 lbs or less
- (b) 201 lbs or more



$$(a) \quad Z = \frac{195 - 200}{\frac{10}{\sqrt{100}}} = -5 \quad \text{prob} < .0000$$

$$(b) \quad Z = \frac{201 - 200}{\frac{10}{\sqrt{100}}} = 1 \quad \text{prob} = .1587$$

This is an **important example** that illustrates the difference between two situations: (1) Situation 1: We are examining individual observations when the population mean and population standard deviation are known. (2) Situation 2: We are examining a sample mean (or means) when the population mean and population standard deviation are known. Do not confuse the two. What you have to know is that the standard deviation of the sample mean is not the same as the standard deviation of an individual observation from the population. Sample means, as explained above, have a smaller standard deviation. In fact, as the sample size gets larger, the standard deviation of the sample mean (known as the standard error of the mean) gets smaller and smaller. With a very large sample, the distribution of the sample means has a narrow range around the population mean.

Example: In a large automobile manufacturing company, the mean life of hybrid motors is normally distributed with a mean of 100,000 miles and a standard deviation of 10,000 miles.

- (a) What is the probability that a randomly selected hybrid motor has a life between 90,000 miles and 110,000 miles per year?  
(b) If a random sample of 100 motors is selected, what is the probability that the sample mean will be below 98,000 miles per year?

Note that to solve the answer to (a), one simply converts to Z by using the following formula:

$$Z = (X_i - \mu) / \sigma$$

$$Z = \frac{90,000 - 100,000}{10,000} = -1$$

$$Z = \frac{110,000 - 100,000}{10,000} = +1$$

Thus, we have to find how much area lies between  $-1$  and  $+1$  of the Z-distribution (answer = .6826)

With (b), we are looking at the sampling distribution of the mean. Sample means follow a different distribution and to convert to Z, we use the following formula.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

$$Z = \frac{98,000 - 100,000}{10,000 / \sqrt{100}} = -2$$

Thus, the answer is  $.5 - .4772 = .0228$ .

Example:

Farm-raised herring have an average weight ( $\mu$ ) of 50 pounds and a population standard deviation of 4 pounds. (a) What proportion of herring from this fishery will weigh less than 44 pounds? (b) A random sample of 25 herring is taken at this fishery? What is the probability that the sample mean will be below 48.60 pounds?

Hint: In (b) we are working with the sampling distribution of the mean. The standard error of the mean is not the same as the population standard deviation.

Solutions, herring problem:

(a)  $Z = -1.50$  probability =  $.50 - .4332 = .0668$

(b)  $Z = -1.75$  probability =  $.50 - .4599 = .0401$



Example:

A candy manufacturer produces bags of jelly beans. The weight of a bag of jelly beans is normally distributed with a mean of 12 ounces and a standard deviation of 0.4 ounces.

(a) What is the probability that a randomly selected bag weighs between 11.62 and 12.3 ounces?

(b) 96% of the bags of jelly beans weight more than \_\_\_ ounces ?

If a random sample of 16 bags of jelly beans is selected ...

(c) .... what is the probability that the sample mean will be between 11.62 and 12.3 ounces?

(d) .... what is the probability that the sample mean will be above 12.2 ounces?