8.1: Distribution of the Sample Mean

Suppose we want to estimate some characteristic of a population; also suppose that it is unrealistic to survey or measure the entire population. Therefore, we select a sample from that population, and we survey or measure the characteristic of interest. If we record the values of the characteristic for the sample, we can compute a sample statistic: the mean, standard deviation, proportion, etc.

Recall: A *parameter* is a numerical summary of a population; a *statistic* is a numerical summary of a sample. (For example, the population mean and population standard deviation are parameters; the sample mean and sample standard deviation are statistics.)

The *sampling distribution of a statistic* is the probability distribution of all possible values for that statistic computed from all possible samples of fixed size *n*.

The *sampling distribution of the sample mean* is the probability distribution of all possible values for that the sample mean $\bar{x}$ computed from all possible samples of fixed size *n*.

Example 1: Consider a population composed of all possible rolls of a single six-sided die. Calculate the means for all possible samples of two dice rolls. Use a dotplot or histogram to illustrate the shape of the sampling distribution of the sample means.

Simulations:

Rice Virtual Lab in Statistics: http://onlinestatbook.com/rvls/index.html
(A public domain resource created by Dr. David Lane of Rice University)
http://onlinestatbook.com/stat_sim/sampling_dist/index.html

http://www.rossmanchance.com/
(Created by Dr. Allen Rossman and Dr. Beth Chance of California Polytechnic State University – San Luis Obispo; used with their permission.)
http://www.rossmanchance.com/applets/OneSample.html

These examples illustrate the Central Limit Theorem, an extremely important result. Much of the usefulness of statistics depends on this theorem.

The Central Limit Theorem:

The Central Limit Theorem

For a large population and sufficiently large sample sizes, the sampling distribution of the Sample mean of a random variable will be approximately normal.

This is true regardless of the shape of the underlying distribution of the random variable (i.e., regardless of whether the variable is normally distributed in the population from which the samples are drawn).

What is a "sufficiently large sample size" for the Central Limit Theorem to be true?

For a variable that is normally distributed in the underlying population: The Central Limit Theorem will be true for any sample size.

For a variable that is NOT normally distributed in the underlying population: As the sample size increases, the sampling distribution of the sample mean will become closer and closer to a normal distribution (i.e., it will become more bell-shaped). The less closely the underlying distribution (in the population) resembles a normal distribution, the larger the sample size requirement.

Rule of Thumb: According to a frequently-used (and conservative) rule of thumb, a sample size of at least 30 is sufficient to assume that the Central Limit Theorem is true, if the variable is not normally distributed in the population.

Mean and standard deviation of the sampling distribution of the sample means:

Mean and Standard Error of the Sampling Distribution:

Suppose a random variable X has population mean μ and population standard deviation σ.

Then, for samples of size n, the sampling distribution of the sample means will have:

Mean: $\mu_{\bar{x}} = \mu$

Standard deviation: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

The standard deviation of the sample mean $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ is called the *standard error of the mean.*

Note: The variance in the population will be n times as large as the variance of the sampling distribution of the sample means (i.e., $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$ is equivalent to $n\sigma_{\bar{x}}^2 = \sigma^2$).

Calculating the probability of a sample's results:

Example 2: Suppose a coffee maker dispenses servings of coffee that are normally distributed with a mean size of 10 ounces and a standard deviation of 0.3 ounces. Find the probability that:

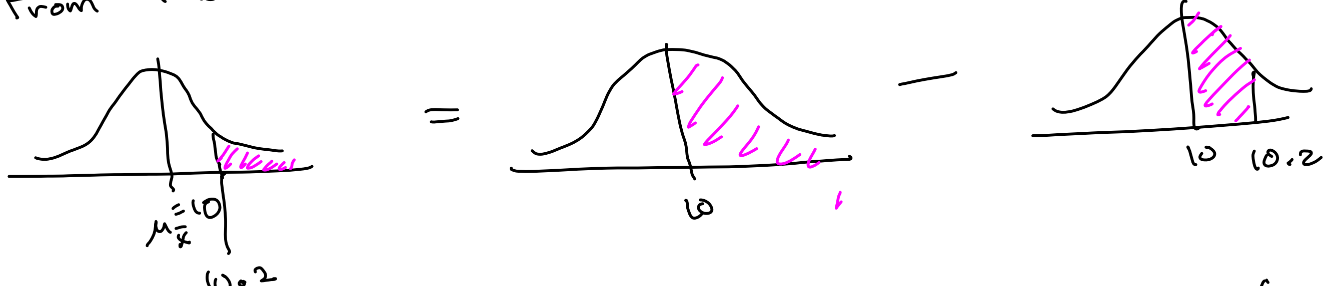
a) A sample of 8 cups of coffee will result in a mean greater than 10.2 ounces.
b) A sample of 8 cups of coffee will result in a mean between 9.8 and 10.2 ounces.
c) A sample of 8 cups of coffee will result in a mean below 9.8 ounces.
d) An individual serving of coffee is less than 9.8 ounces.
e) A sample of 12 cups of coffee will result in a mean below 9.8 ounces.

a) $\mu_{\bar{x}} = 10$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{8}} = 0.106$$

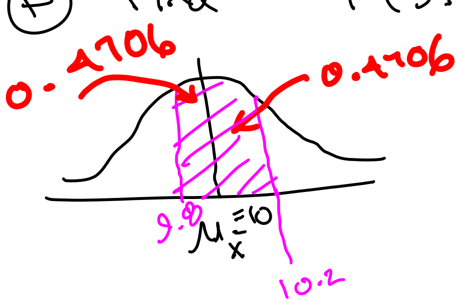
$$\text{Z-score} = z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{10.2 - 10}{0.106} = 1.89$$

From Table Handout, Area $= P(0 < z < 1.89) = 0.4706$



$$P(\bar{x} > 10.2) = 0.5 - 0.4706 = \boxed{0.0294}$$

b) Find $P(9.8 < \bar{x} < 10.2)$.



Note: 9.8 is 0.2 units below $\mu_{\bar{x}} = 10$
10.2 is 0.2 units above $\mu_{\bar{x}} = 10$.

So, from symmetry,

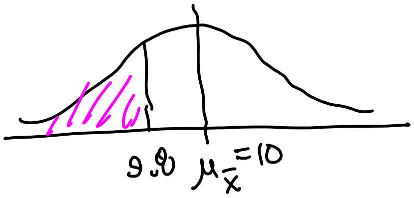
$$P(9.8 < \bar{x} < 10.2) = 2(0.4706) = \boxed{0.9412}$$

next page

Ex 1: (c) Find $P(\bar{x} < 9.8)$

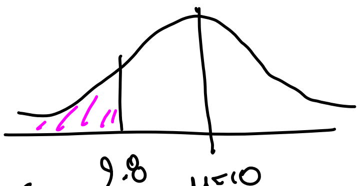
From symmetry, this will be the same area as (a).

So, $P(\bar{x} < 9.8) = \boxed{0.0294}$

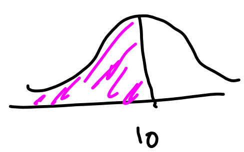


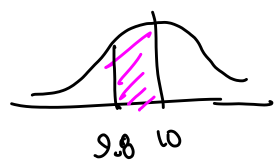
(d) Find prob. that an individual cup of coffee is below 9.8 oz.

$$z_{9.8} = \frac{x-\mu}{\sigma} = \frac{9.8-10}{0.3} = -0.67$$



From table, $P(0\ z < 0.67) = 0.2486$

= 

− 

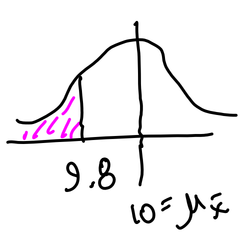
$P(x<9.8) = 0.5 - 0.2486 = \boxed{0.2514}$

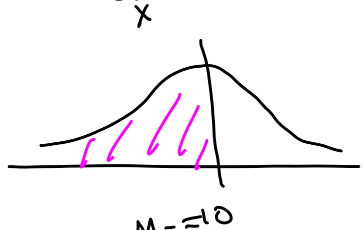
(e) For a sample of 12 cups, what is the prob. that the mean will be below 9.8?

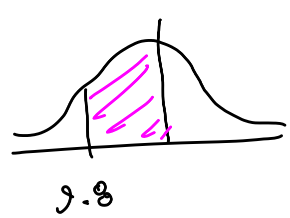
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{12}} = \frac{0.3}{\sqrt{12}} = 0.0866$$

From table, Area = 0.4896

Find Zscore: $z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{9.8-10}{0.0866} = 2.31$



= 

− 

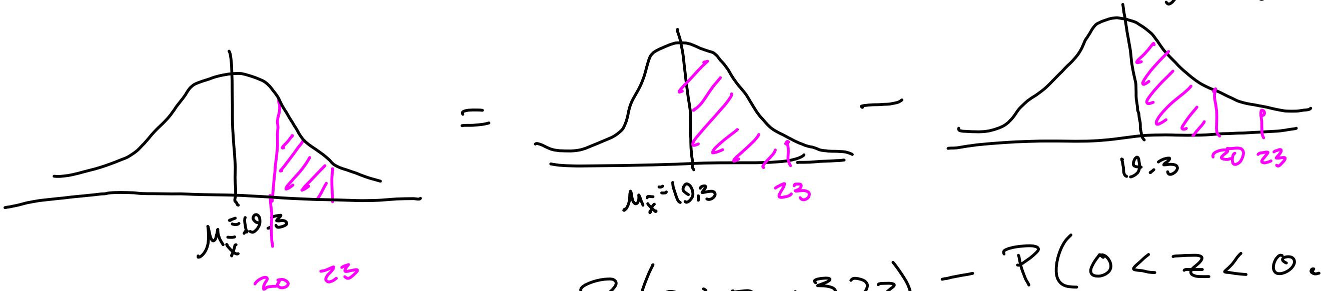
$P(\bar{x}<9.8) = 0.5 - 0.4896 = \boxed{0.0104}$

Example 3: Suppose that a pizza company's records showed that the mean delivery time is 19.3 minutes and that the standard deviation of the delivery times is 6.8 minutes.

a) What sample size is required for us to be able to calculate probabilities using the normal model?
b) What is the probability that a sample of 35 orders has a mean delivery time between 20 and 23 minutes?
c) What is the probability that a sample of 35 orders has a mean delivery time over 26 minutes?
d) Calculate the value for which there is a 0.05 probability that the mean delivery time is below that value (for a sample of 35).

(a) Rule of thumb is that a sample of at least 30 is needed.

(b) standard error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 1.1494$. Mean of the sampling distribution is $\mu_{\bar{x}} = \mu = 19.3$



$$P(20 < \bar{x} < 23) = P(0 < z < 3.22) - P(0 < z < 0.61)$$

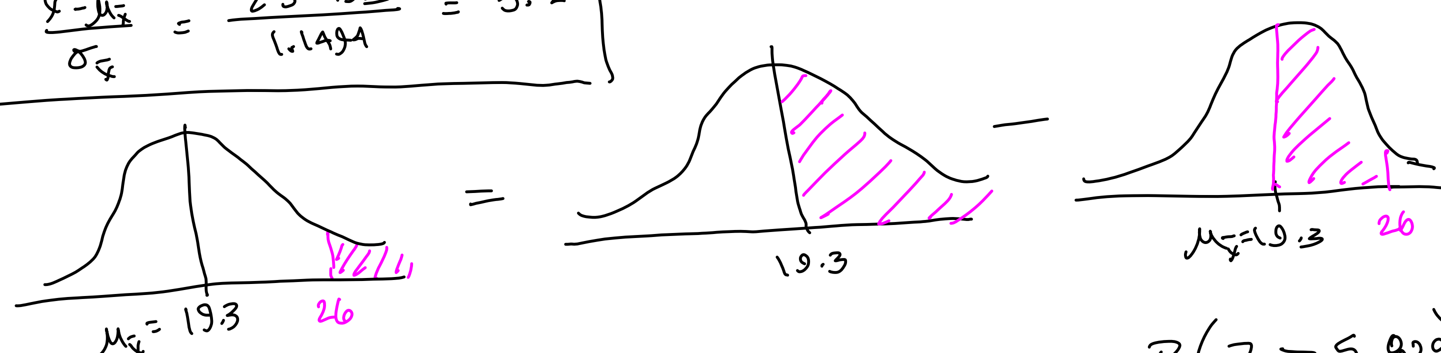
$$= 0.4994 - 0.2291$$

$$= \boxed{0.2703}$$

$$z_{20} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{20 - 19.3}{1.1494} = 0.61$$

$$z_{23} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{23 - 19.3}{1.1494} = 3.22$$

(c)



$$P(\bar{x} > 26) = 0.5 - P(z > 5.829)$$

$$z_{26} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{26 - 19.3}{1.494} = 5.829$$

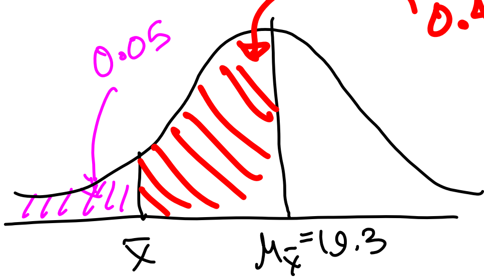
See next page

(c) The z-value 5.829 is literally "off the chart"! This means that the area between 0 and $z = 5.829$ (or between 19.3 and 26) is so close to 0.5 that it would round to 0.5000 if we round it to 4 decimal places.

So, the area to the right of 26 is very tiny, and must be beyond the 4th decimal place

Conclusion: $\boxed{P(\bar{x} > 26) < 0.0001}$

(d)



(because the left half must be 0.50) $0.05 + 0.45 = 0.50$

we want to find the z-value for which 0.05 of the area is to the left. then we'll work backwards to get the $\bar{x}$.

What value in the z-table is closest to 0.45?

Looks like we're halfway between 0.4495 and 0.4505, corresponding to $z = 1.64$ and $z = 1.65$.

So, let's use $z = 1.645$. But we want $z = -1.645$, since we're below the mean.

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$-1.645 = \frac{\bar{x} - 19.3}{1.1494}$$

$$-1.645(1.1494) = \bar{x} - 19.3 \Rightarrow -1.645(1.1494) + 19.3 = \bar{x}$$

$$\Rightarrow \boxed{\bar{x} = 16.84}$$