Section 5.2: Sampling Distribution of Sample Mean

In Section 3.4 we looked at histograms of the sampling distribution of the sample average $\bar{x}$ and we saw that **averages are __________________ than individual observations** and **averages are __________________ than individual observations**.

We can represent $n$ observations as the random variables

$$x_1, x_2, ..., x_n$$

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + ... + x_n)$$

using the rules for combining independent random variables just as we did in Section 5.1, we find that for $\bar{x}$

$$\mu = \frac{1}{n} (\mu_1 + \mu_2 + ... + \mu_n)$$

and

$$\sigma^2 = \frac{1}{n^2} (\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$$

When the $X$'s represent a ___________ of size $n$ from a population, they all have the _____ distribution, so all the $\mu$'s are the same and all the $\sigma$'s are the same.

We can summarize the result of using the formulas above by:

<table>
<thead>
<tr>
<th>Mean and Standard Deviation of a Sample Mean, $\bar{x}$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suppose that $\bar{x}$ is the mean of a ___________ of size $n$ drawn from a large population that has mean $\mu$ and standard deviation $\sigma$. Then the mean of the sampling distribution of $\bar{x}$ is_________ and its standard deviation is ________________________</td>
</tr>
</tbody>
</table>

Example: ACT Scores

The scores on the ACT college entrance examination in a recent year followed a normal distribution with a mean $\mu = 18.6$ and a standard deviation of $\sigma = 5.9$.

Suppose we took a SRS of 50 students who took the test. What are the mean and standard deviation of the distribution of the sample mean score $\bar{x}$ of these 50 students
Sampling Distribution of a Sample Mean

- If the population we are sampling from is Normal(\(\mu, \sigma\)) then the sample mean \(\bar{x}\) based on \(n\) independent observations has a Normal(\(\mu\), \(\sigma\)) distribution.

Example: ACT Scores cont'd

(a) What is the probability that a single student, randomly chosen from those taking the ACT, scores 21 or higher?

(b) What is the probability that the mean score \(\bar{x}\) of a SRS of 50 students is 21 or higher? Will this probability be larger or smaller than the one calculated in (a)?

Note that any linear combination of independent normal random variables is __________________ _______ distributed.

Example:

\(X\) is the diameter of a cylinder with \(\mu_X = 5.526 cm, \sigma_X = 0.0004 cm\)

\(Y\) is the diameter of a piston with \(\mu_Y = 5.525 cm, \sigma_Y = 0.0003 cm\)

Let \(W = X - Y\) be the clearance between a randomly chosen cylinder and a randomly chosen piston.

\[\mu_W = \mu_X - \mu_Y, \quad \sigma_W = \sqrt{\sigma_X^2 - \sigma_Y^2}\]

Now assume that both \(X\) and \(Y\) have normal distributions. Then ___________________________.

Now let's find the probability of a positive clearance, i.e.

\[P(W \geq 0) = \]

However, if the population we sample from is not normal, then the distribution of \(\bar{x}\) is _________ ________________________

Fortunately, we have the Central Limit Theorem to the rescue!!

The **Central Limit Theorem (CLT)** states:

- Draw a _________ of size \(n\) from any population with mean \(\mu\) and a standard deviation \(\sigma\).

When \(n\) is _________, the sampling distribution of the sample mean \(\bar{x}\) is approximately normal:

\[\bar{x} \text{ is approximately Normal}(\mu, \sigma/\sqrt{n})\]
Q: How large is large enough?

A: It depends on the shape of the distribution we are sampling from. More observations are required if the shape of the population distribution is far from normal.

Rule of Thumb: CLT is usually applicable for \( n > \)______________.
The CLT for a continuous (exponential) distribution:
(a) n=1, (b) n=2, (c) n=10, (d) n=25

Example:
The number of accidents per week at a hazardous intersection varies with mean 2.2 and standard deviation 1.4. Since the number of accidents per week can only take on whole numbers (e.g. 0, 1, 2, 3 ...) it is certainly ____________________________.

(a) Let $\bar{X}$ be the mean number of accidents per week at the intersection during a year (52 weeks). What is the approximate distribution of $\bar{X}$ according to the CLT?

(b) What is the probability that $\bar{X}$ is less than 2?

Extensions of the Central Limit Theorem.
1. Sums or averages of ___________________________ are close to ___________ even if they do not have the same distribution and are not ___________________ (subject to some limitations). Thus, observed data for any variable that is a sum of many small influences will look ________________.
2. Combining the Central Limit Theorem with the linear combinations theorem. Suppose we have several ________________ that are each ____________________________ due to the CLT. Then, linear combinations of these ________________ will be ____________________________. 