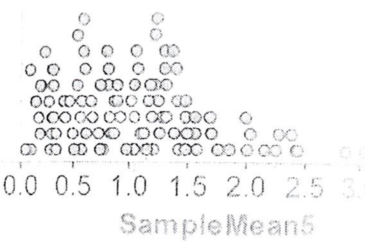
7.3 Lecture Notes & Examples

Section 7.3 - Sample Means (pp. 450-464)

Sample proportions arise most often when we are interested in *categorical* variables. But when we are looking at *quantitative* variables, we are interested in other statistics such as the median, mean or standard deviation. In this section we will study the *sampling distribution of the sample mean* $\bar{x}$.

Activity

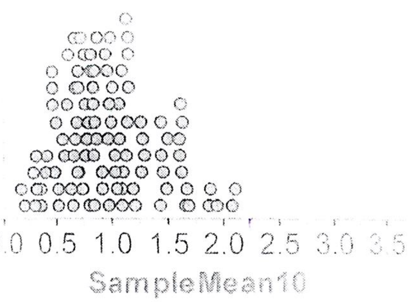
Let's go back to the hyena experiment. Suppose that we were now interested in determining the average age of the hyenas in the Croatan NF. A team first chose repeated samples of size 5 and determined the mean age of each sample. The distribution of sample means is shown at the right.



Describe the distribution:

Skewed right ≈ bimodal
Center ≈ 1.0
Range ≈ 0 to 3 (3)

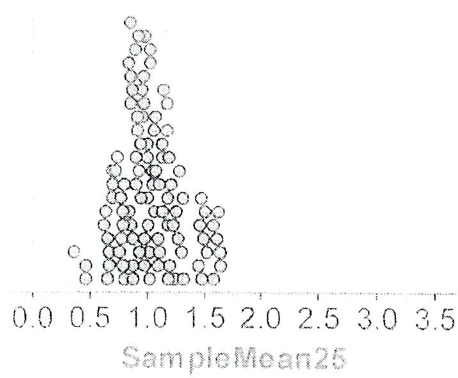
The team then took repeated samples of size 10. The distribution of sample means is shown at the right.



Describe the distribution:

more symmetric, unimodal
center ≈ 1.0
Range ≈ .25 to 2.25 (≈ 2)

Finally, the team took repeated samples of size 25. The distribution of sample means is shown at the right.



Describe the distribution:

More symmetric, unimodal
Center ≈ 1.0
Range ≈ .3 to 1.6 (≈1.3)

Summarize what happened to the center, shape, and spread as the sample size increased from 5 to 20.

As n increased, center stayed at 1, shape became more symmetric, spread decreased.

The Sampling Distribution of $\bar{x}$: Mean and Standard Deviation

Mean and Standard Deviation of the Sampling Distribution of $\bar{x}$

Suppose that $\bar{x}$ is the mean of an SRS of size n drawn from a large population with mean μ and standard deviation σ. Then

* On AP Formula Sheet *

- The **mean** of the sampling distribution of $\bar{x}$ is $\mu_{\bar{x}} = \mu$
- The **standard deviation** of the sampling distribution of $\bar{x}$ is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

as long as the *10% condition* is satisfied: $n \le (1/10)N$.

$10n \le N$

Population is at least 10 times sample size

The behavior of $\bar{x}$ in repeated samples is much like that of the sample proportion $\hat{p}$:

- The sample mean $\bar{x}$ is an unbiased estimator of the population with mean μ.
- Less variability in $\bar{x}$ with larger samples. σ decreases at a rate of $\sqrt{n}$
- * $\frac{\sigma}{\sqrt{n}}$ only when 10% condition is satisfied.

These facts about the mean and standard deviation of $\bar{x}$ are true *no matter what shape the population distribution has.*

Example - Suppose that the number of movies viewed in the last year by high school students has an average of 19.3 with a standard deviation of 15.8. Suppose we take an SRS of 100 high school students and calculate the mean number of movies viewed by the members of the sample.

(a) What is the mean of the sampling distribution of $\bar{x}$?

Since SRS $\mu_{\bar{x}} = \mu = 19.3$ movies

(b) What is the standard deviation of the sampling distribution of $\bar{x}$? Check that the 10 % condition is satisfied.

$10(n) \le$ Population

$10(100) \le$ Population of H.S. students

It is reasonable to assume there are more than 1000 H.S. students

$\therefore \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15.8}{\sqrt{100}} = \boxed{1.58 \text{ movies}}$

Sampling from a Normal Population

Sampling Distribution of a Sample Mean from a Normal Population

Suppose that a population is Normally distributed with mean μ and standard deviation σ. Then the sampling distribution of $\bar{x}$ has the Normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, provided the *10% condition* is met.

Example - At the P. Nutty Peanut Company, dry-roasted, shelled peanuts are placed in jars by a machine. The distribution of weights in the jars is approximately Normal with a mean of 16.1 ounces and a standard deviation of 0.15 ounces.

$\mu = 16.1$ oz $\quad \sigma = 0.15$ oz $\quad \sim N(16.1, .15)$

(a) Without doing any calculations, explain which is more likely: randomly selecting a single jar and finding that the contents weigh less than 16 ounces or randomly selecting 10 jars and finding that the average content weighs less than 16 ounces.

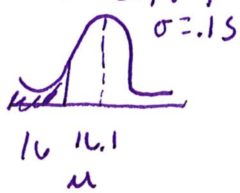
Averages are less variable than individual values, so you would expect the sample mean of 10 jars to be closer, on average to true mean of 16.1 oz. ∴ it is more likely that a single jar weighs less than 16 oz than for the average of 10 jars to be less than 16 oz.

(b) Find the probability of each event described above.

Let X = weight of randomly chosen jar $\sim N(16.1, 0.15)$

$P(X < 16) = \text{normcdf}(-\infty, 16, 16.1, 0.15)$

$\approx .2525$



✓10% for $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

10 (10) ≤ pop peanuts in jar

100 ≤ pop reasonable to assume

Let $\bar{x}$ = average weight of random sample of 10 jars. $\bar{x} \sim N\left(16.1, \frac{.15}{\sqrt{10}}\right)$

$\sigma = \frac{0.15}{\sqrt{10}} \approx .0474$ $\quad P(\bar{x} \le 16) = \text{normcdf}(-\infty, 16, 16.1, .0474)$

$\approx .0174$

Note: A common error on the AP Exam is that students often forget to divide by the square root of the sample size.

Check Your Understanding – Complete CYU on p. 456.

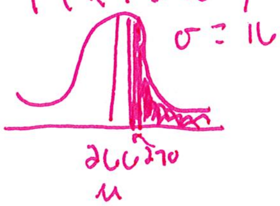
CHECK YOUR UNDERSTANDING

The length of human pregnancies from conception to birth varies according to a distribution that is approximately Normal with mean 266 days and standard deviation 16 days.

μ $\quad$ σ $\quad$ $\sim N(266, 16)$

1. Find the probability that a randomly chosen pregnant woman has a pregnancy that lasts for more than 270 days. Show your work. Suppose we choose an SRS of 6 pregnant women. Let X = the mean pregnancy length for the sample.

$P(X > 270) = \text{normcdf}(270, \infty, 266, 16) \approx .4013$



Probability that randomly chosen pregnant woman has pregnancy that lasts for more than 270 days ≈ 40.13 %

2. What is the mean of the sampling distribution of X? Explain

Since SRS, $\mu_{\bar{x}} = \mu = 266$ days

3. Compute the standard deviation of the sampling distribution of X. Check that the 10% condition is met.

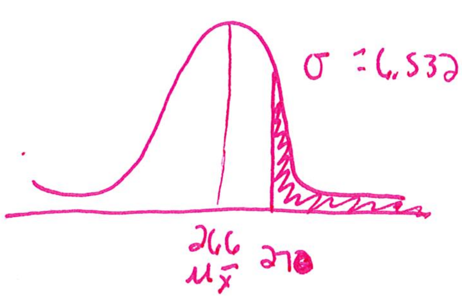
$n = 6$ $\quad$ $10(6) \leq$ population

Reasonable to assume there are more than 60 pregnant women in population ✓

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{16}{\sqrt{6}} \approx 6.532 \text{ days}$$

4. Find the probability that the mean pregnancy length for the women in the sample exceeds 270 days. Show your work

$P(\bar{x} > 270) \sim N(266, 6.532)$



$\text{Normcdf}(270, \infty, 266, 6.532) \approx .2701$

The probability that mean pregnancy lengths for the women exceed 270 days is ≈ 27.01 %.

The Central Limit Theorem

Many populations are not Normal. However, it turns out that if we take large enough samples from any population, the sampling distribution of the sample mean is Normally distributed. This result is known as the **Central Limit Theorem** and it will have a major impact on our ability to make inferences about populations.

> **The Central Limit Theorem (CLT)**
>
> Draw an SRS of size n from *any* population with mean μ and standard deviation σ. The **Central Limit Theorem (CLT)** says that when n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal.

The burning question that this theorem quickly solicits is how large does n have to be? Without getting into theory, the quick answer is n = 30. This leads us to now be able to state the *Normal condition* for sample means.

> **Normal Conditions for Sample Means**
>
> - If the population distribution is Normal, then so is the sampling distribution of $\bar{x}$. This is true no matter what the sample size n is.
> - If the population distribution is not Normal, the Central Limit Theorem tells us that the sampling distribution of $\bar{x}$ will be approximately Normal in most cases if $n \geq 30$.

Example - Suppose that the mean number of texts sent during a typical day ~~at~~ by a randomly selected high school student follows a right-skewed distribution with a mean of 15 and a standard deviation of 35. Assuming that the students at your school are typical texters, how likely is it that a random sample of 50 students will have sent more than a total of 1000 for a random sample of 50 high school students.

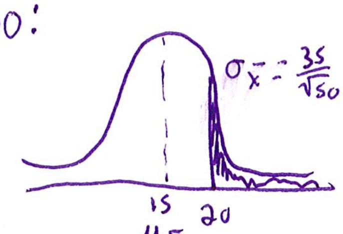
State: What is the probability that the total number of texts in the last 24 hours is greater than 1000 for a random sample of 50 high school students?

Plan: A total of 1000 texts among 50 students is the same as an average # of texts of 1000/50 = 20. We want to find $P(\bar{x} \geq 20)$ where $\bar{x}$ is the sample mean of texts. Since n is large ($50 \geq 30$), by the C.L.T. the distribution of $\bar{x}$ is ~ Normal. $N(15, \frac{35}{\sqrt{50}})$

$n \geq 30$

DO:

$P(\bar{x} \geq 20) \approx \text{normcdf}(20, \infty, 15, 4.9497) \approx .1562$



$\sigma_{\bar{x}} = \frac{35}{\sqrt{50}} \approx 4.9497$

Conclude: There is about a 16% chance that a random sample of 50 HS students will send more than 1000 texts in a day.

HW: Read pp. 450-461, do problems: p. 449 43-46; p. 461 49-63 odd, 65-68.