AP Statistics
Unit 05 – Distributions
Day 04 Notes – CLT

Name Key
Period____________

What happens when the population distribution is not Normal? How can we meet the Normal/Large condition to perform Normal model calculations?

Central Limit Theorem (CLT): Draw an SRS of size n from any population of size N with mean μ and finite standard deviation σ. When n is sufficiently large, the sampling distribution of $\bar{x}$ is approximately Normal.

- The sampling distribution of ANY mean becomes more nearly Normal as the sample size grows

ASSUMPTIONS AND CONDITIONS

1. 10% Condition: When a sample is drawn without replacement, $n \leq \frac{1}{10}N$.
2. Normal/Large Condition: Regardless of the shape of the population distribution, the sampling distribution of $\bar{x}$ will be approximately Normal if $n \geq 30$.

NOTATIONS:

n= sample size vs. N = population size

μ = mean of the population

σ = standard deviation of the population

$\bar{x}$= mean of the sample

$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ = standard deviation of the sample

$\sigma_{\bar{x}}$

SAMPLING DISTRIBUTION MODEL FOR A MEAN (CLT)

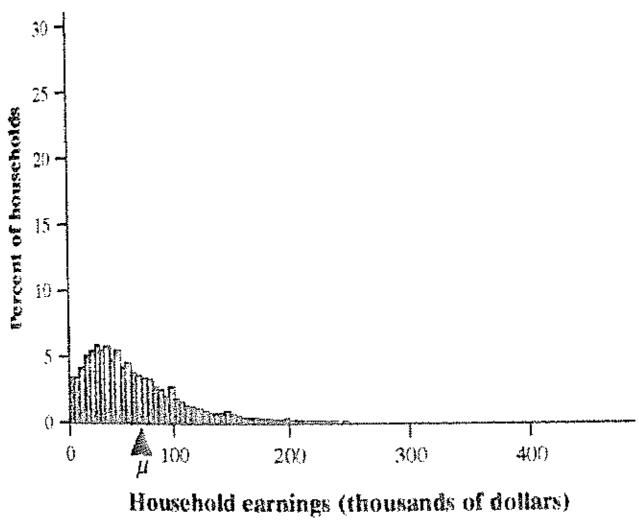
- If the population distribution is Normal, then so is the sampling distribution of $\bar{x}$. This is true no matter what the sample size n is.
- If the population distribution is **NOT** Normal, the central limit theorem tells us that the sampling distribution of $\bar{x}$ will be approximately Normal in most cases if $n \geq 30$.
- The larger the sample used, the more closely the sampling distribution is to a Normal distribution.

Example: Making Money

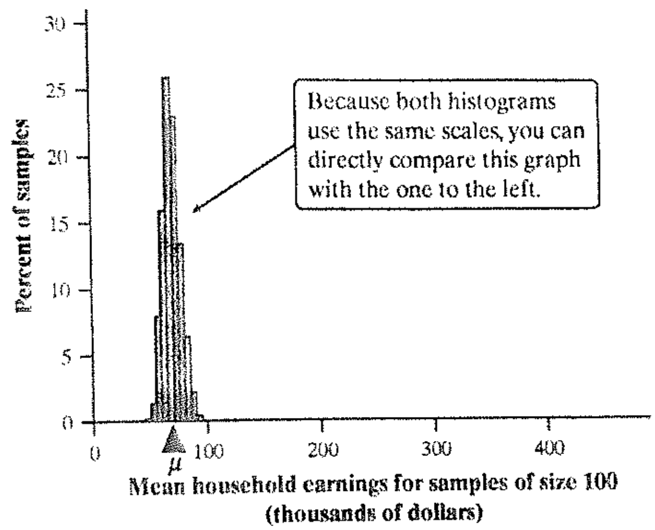
Figure (a) below is a histogram of earnings of a population of 61,742 households that had earned income greater than zero in a recent year.

As we expect, the distribution of earned incomes is strongly skewed to the right and very spread out. The right tail of the distribution is even longer than the histogram shows because there are too few high incomes for their bars to be visible on this scale. We cut off the earnings scale at $400,000 to save space). The mean earnings for these 61,742 households was μ = $69,750.

Take an SRS of 100 households. The mean earnings in this sample is $\bar{x}$ = $66,807. That's less than the mean of the population. Take another SRS of size 100. The means for this sample is $\bar{x}$ = $70,820. That's higher than the mean of the population. What would happen if we did this many times? Figure (b) below is a histogram of the mean earnings for 500 samples, each of size n = 100. The scales in both figures are the same, for easy comparison. Although the distribution of individual earnings is skewed and very spread out, the distribution of sample means is roughly symmetric and much less spread out. Both distributions are centered at $69,750.



(a)



(b)

EXAMPLE 1:

A college physical education department asked a random sample of 200 female students to self-report their heights and weights, but the percentage of students with body mass indexes over 25 seems suspiciously low. One possible explanation may be that the respondents "shaded" their weights down a bit. The CDC reports that the mean weight of 18-year-old women is 143.74 lbs, with a standard deviation of 51.54 lbs, but these 200 randomly selected women reported a mean weight of only 140 lb.

Based on the CLT, does the mean weight in this sample seem exceptionally low, or might this just be a random sample-to-sample variation?

STATE: We want to find the probability that the mean weight of an SRS of 200 women is 140 lbs or less.

$P(\bar{x} \leq 140) = ?$ let X = the weight of a randomly selected woman (female college student)

PLAN:

1. 10% Condition: $n = 200$ $200 \cdot 10 = 2000 <$ all female college students ✓

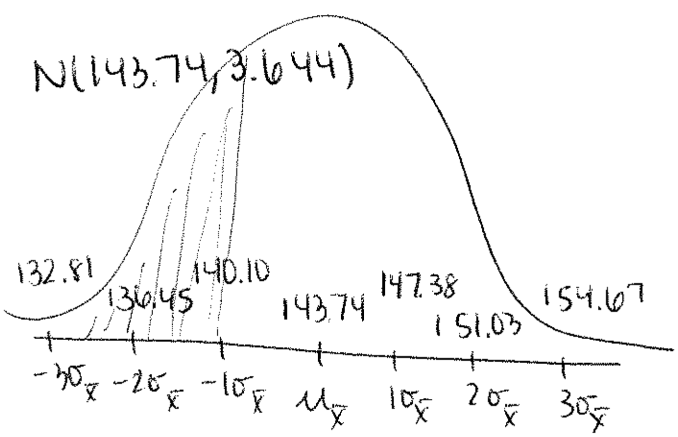
so $\sigma_{\bar{x}} = \frac{51.54}{\sqrt{200}} = 3.644$ lbs

2. Normal/Large Condition:

we don't know what the population distribution looks like, but $n = 200 \geq 30$ ✓ so the sampling distribution of $\bar{x}$ is approximately Normal.

DO:

$P(\bar{x} \leq 140) = P(z \leq -1.026) = 0.1524$



$$z = \frac{140 - 143.74}{3.644} = -1.026$$

CONCLUDE: There is a 15.24% chance that the mean weight of 200 randomly selected women is 140 lbs or less. This is a reasonable probability (> 5%) so it might just be random sample-to-sample variation.

EXAMPLE 2:

The Centers of Disease Control and Prevention reports that the mean weight of adult men in the US is 190 lbs with a standard deviation of 59 lbs, and is approximately Normal. An elevator in our building has a weight limit of 10 persons or 2500 lb. What's the probability that if 10 random men get on the elevator, they will overload its weight limit?

STATE: We want to find the probability that if 10 random men get on the elevator, they will overload its weight limit.

$P(10\bar{x} > 2500) = ?$ or $P(\bar{x} > 250) = ?$ let X = the weight of a randomly selected adult US man on the elevator

PLAN:

1. 10% Condition: $n = 10$ $\quad 10 \cdot 10 = 100 <$ all adult US men who ride elevators ✓

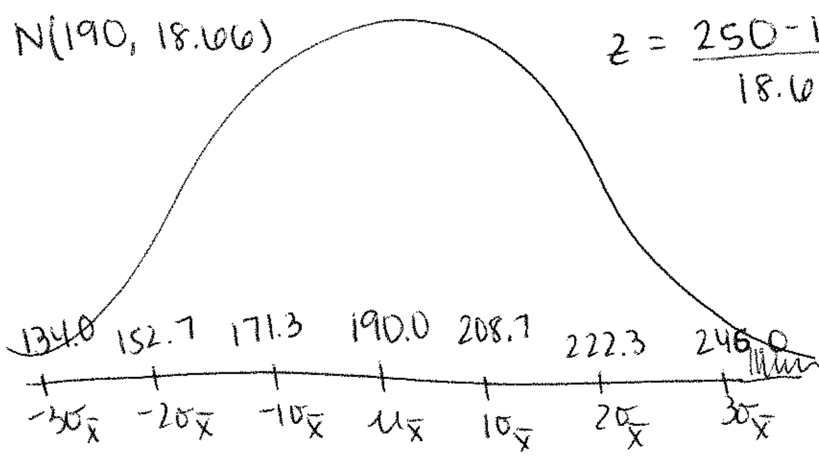
   so $\sigma_{\bar{x}} = 59/\sqrt{10} = 18.66$ lbs

2. Normal/Large Condition: the population distribution is approximately Normal ✓ so the sampling distribution of $\bar{x}$ is approx. Normal

DO:

$P(\bar{x} > 250) = P(z > 3.216) = 0.0006503$

$z = \frac{250 - 190}{18.66} = 3.216$



CONCLUDE: There is a 0.06503% chance that if 10 random men get on the elevator in our building, they will overload the weight limit.

EXAMPLE 3:

The population distribution of battery lifetimes has mean of 17 hours and a standard ~~deviation of 0.8 hours.~~ 30 batteries were randomly selected and had a mean life of 16.7. If the process is working properly, what's the probability of getting a random sample of 30 batteries in which the sample mean lifetime is 16.7 hours or less?

STATE: We want to find the probability that a random sample of thirty batteries will have a mean lifetime of 16.7 hour or less. $P(\bar{x} \leq 16.7) = ?$ let x = the lifetime of a randomly selected battery

PLAN: 10% condition: $n = 30$ $30 \cdot 10 = 300 <$ all batteries ✓

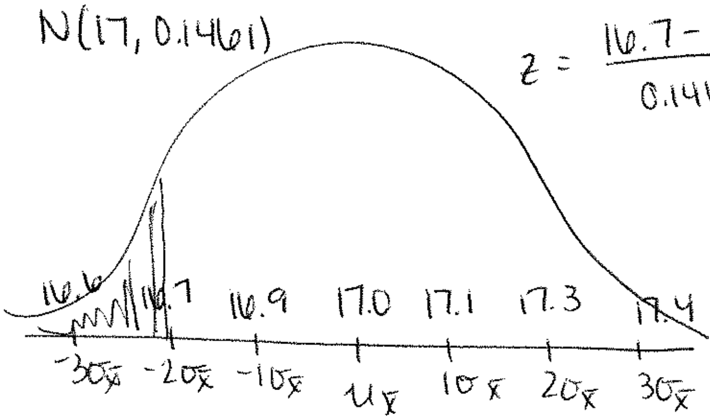
so $\sigma_{\bar{x}} = 0.8/\sqrt{30} = 0.1461$ hours

Normal/Large: $n = 30 \geq 30$ ✓ so the sampling distribution of $\bar{x}$ is approximately Normal

DO: $P(\bar{x} \leq 16.7) = P(z \leq -2.054) = 0.01999$

N(17, 0.1461)

$$z = \frac{16.7 - 17.0}{0.1461} = -2.054$$



CONCLUDE: There is a 1.999% chance that a random sample of 30 batteries has a mean lifetime of 16.7 hours or less.

EXAMPLE 4:

Your company has a contract to perform maintenance on thousands of air-conditioning units in a large city. Based on service records from the past year, the time (in hours) that a tech requires to complete the work has a mean of 1 hour and a standard deviation of 1 hour. In the coming week, your company will service an SRS of 70 air-conditioning units in the city. You plan to budget an average of 1.1 hours per unit for a technician to complete the work. Will this be enough? What is the probability that the average maintenance time for 70 units exceeds 1.1 hours?

STATE: We want to find the probability that the mean maintenance for 70 AC units exceeds 1.1 hours.

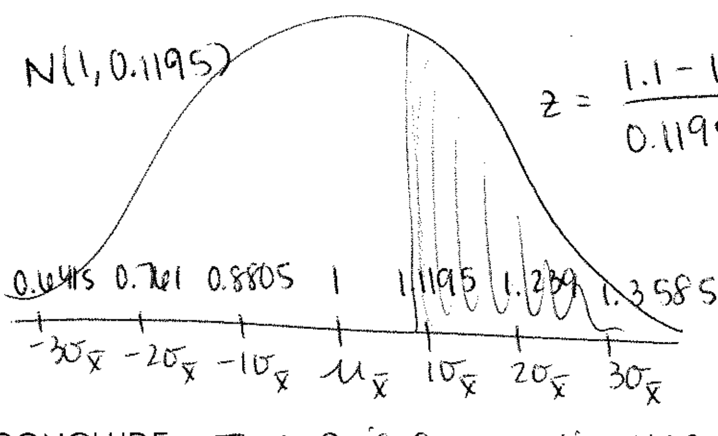
$P(\bar{x} > 1.1) = ?$ let X = the time it takes to complete maintenance on a randomly selected AC unit.

PLAN: 10% condition: $n = 70$ $70 \cdot 10 = 700 <$ all AC units in a large city ✓

so $\sigma_{\bar{x}} = 1/\sqrt{70} = 0.1195$ hours

Normal/Large: $n = 70 \geq 30$ ✓ so the sampling distribution of $\bar{x}$ is approximately Normal.

DO: $P(\bar{x} > 1.1) = P(z > 0.837) = 20.14\%$



$$z = \frac{1.1 - 1}{0.1195} = 0.837$$

CONCLUDE: There is a 20.14% chance that the mean maintenance time for 70 randomly selected AC units exceeds 1.1 hours.