Learning Goals:

- Determine the mean and standard deviation of a distribution of sample means.
- Describe conditions necessary for use of the normal curve to model a distribution of sample means.
- Estimate probabilities using a normal model of the sampling distribution.

Introduction:

We bet you can guess what comes next … we just finished running simulations to observe how the sample means from random samples behave, so now we will create a mathematical model to summarize what we observed.

When can a normal curve be used to model the distribution of sample means?

It really depends on the population's distribution. For small samples, the distribution of sample means will have a similar shape to the population's distribution. Therefore, if a normal curve models the population's distribution well, a normal curve can also be used to model the distribution of sample means, regardless of sample size. But if the population's distribution is skewed, a normal curve is a good fit for the distribution of sample means only for large samples.

The general guideline is that samples of size 30 will have a fairly normal distribution regardless of the shape of the population's distribution.

What is the mean and standard deviation of the distribution of sample means?

If the population has a mean of μ and a standard deviation of σ, then the theoretical distribution of sample means has a mean of μ and a smaller standard deviation given by $\frac{\sigma}{\sqrt{n}}$, where n is the sample size. In theory, this is always true (regardless of the shape of the sampling distribution).

How do we calculate z-scores?

Recall that the z-score is $\frac{sample\ statistic - population\ parameter}{standard\ error}$.

The z-score for an individual measurement is $z = \frac{x-\mu}{\sigma}$

The z-score for a sample mean is $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

1) **Example:** The World Health Organization (WHO) monitors many variables to assess a population's overall health. One of these variables is birth weight. A low birth weight is defined as 2500 grams or less.

   Suppose that babies in a town had a mean birth weight of 3,500 grams with a standard deviation of 500 grams in 2005. This year, a random sample of 25 babies has a mean weight of 3,400 grams. Obviously, this sample weighs less on average than the population of babies in the town in 2005. A decrease in the town's mean birth weight could indicate a decline in overall health of the town.

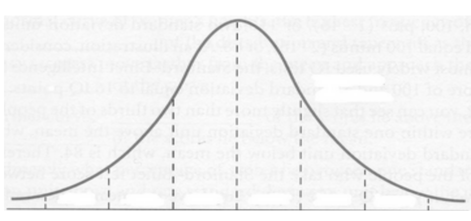
   Are differences this large expected in random sampling from a population with a mean birth weight of 3,500 grams? What is the probability that a random sample of 25 babies will have a mean birth weight of 3,400 grams or less?

   We assume that the variability in individual birth weights is the same this year as it was in 2005. In general, body measurements in a large population can be modeled by a normal curve.

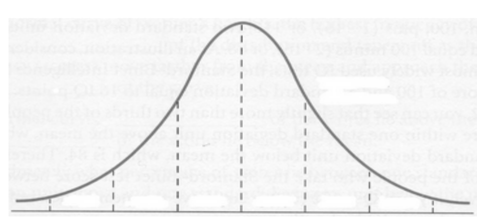
   a) Identify the following: μ , σ , $\bar{x}$, and n from the information given above.

   b) Verify that a normal curve can be used to model the distribution of sample means for this situation.

   c) Label the mean and mean±SD in the normal model for the population and for the sampling distribution.

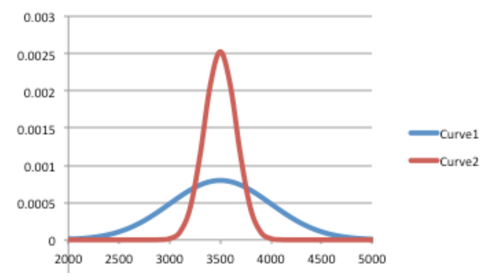


Birth Weights Individual Babies (grams)



Sample mean birth weight
Random samples of 25 babies

Here are the two normal models drawn on the same scale. Which is the population and which is the sampling distribution? How do you know?



d) What is the z-score for a baby that weighs 3400 grams? What is the z-score for a sample of babies with a mean birth weight of 3400 grams? Why do your answers make sense when you look at the normal curves in (c)?

e) What is the probability that a random sample of 25 babies weighs 3,400 grams or less? (Shade the area representing the probability in the appropriate normal curve in (c) and give your estimate.)

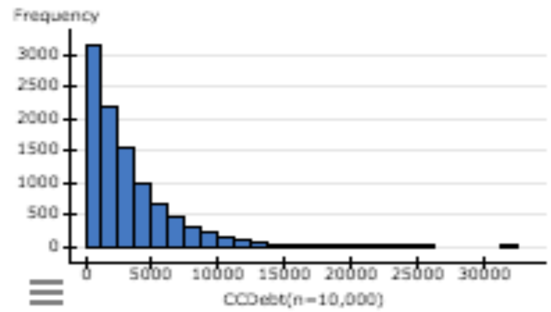
f) Is the difference between 3,400g and 3,500g statistically significant? Or is this difference what we expect to see in random sampling when the population has a mean of 3,500g? How do you know?

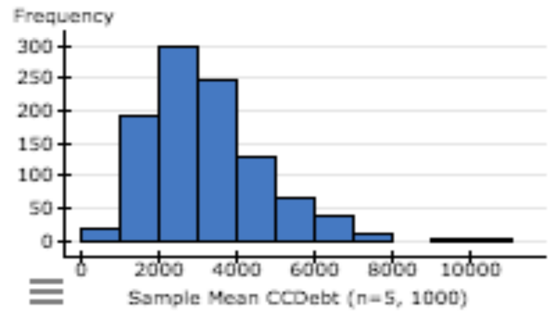
Group work

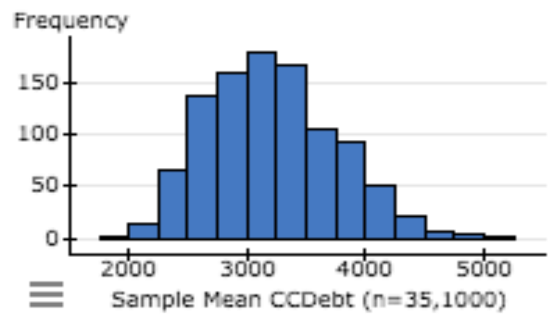
2) According to Debt.org, college students (undergraduates) have an average of $3,200 in credit card debt. https://www.debt.org/students/debt/

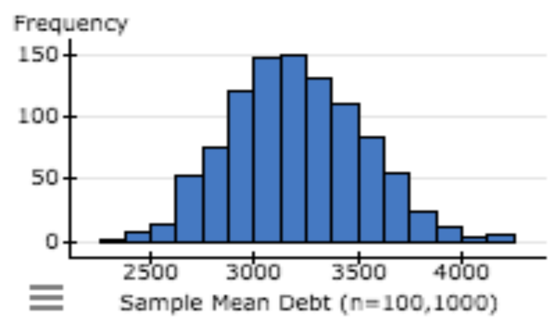
   The top histogram is a hypothetical distribution of credit card debt for 10,000 college students with μ = $3200 and $\sigma = \$3000$. The distribution is strongly skewed to the right with a few students who have large amounts of credit card debt.

   The other three histograms are sampling distributions. Each sampling distribution contains the sample means from 1,000 samples. One histogram is composed of samples of 5 college students. The samples sizes are 35 and 100 for the other two histograms.









   a) Locate the bin containing an individual student who has $4766 in credit card debt.

   b) Five college students are randomly selected. Their credit card debts are $1321, $5951, $782, $3365, $1504. Locate the bin in the appropriate histogram containing the mean debt for these 5 students.

   c) How do these four histograms illustrate the conditions for the use of a normal curve to model a distribution of sample means?

d) We want to determine the probability that a random sample of 20 college students from this population has an average credit card debit of $3500 or more. Which can we use to find the probability? Circle all that will work.

- Empirical Rule
- StatCrunch Normal Calculator
- OLI Z-score calculator
- Simulation
- none of these

e) Is it unusual for a random sample of 100 college students to have an average credit card debit of $3500 or more? What is the probability? (Show your work.)

3) Which of the following questions can be answered based on a normal model? Explain. (You do not need to answer the questions, just determine if a normal model is appropriate.)

a) According to the growth charts produced by the World Health Organization, one-month-old girls have a mean head circumference of 36.55cm and a standard deviation of 1.17cm. In general, body measurements in a large population can be modeled by a normal curve.

In a study of health conditions in a county with a high poverty rate, researchers find that a random sample of 25 one-month-old girls have a mean head circumference of 36cm. Does this sample provide strong evidence that the mean head circumference for the population of one-month-old girls in this county is unusually small?

b) According to the US Census Bureau 2014 Annual Social and Economic Supplement, the mean household income in the United States was $72,641. Previous studies suggest that the standard deviation is about $35,000. Income data is skewed strongly to the right.

We are interested in determining whether the mean household income is higher in our county. We randomly sample 25 households and determine that the mean income is $65,000. Does this sample provide strong evidence that the mean income is lower in our county?

4) At Starbucks the advertised nutrition facts say that a Tall Chai Latte with nonfat milk has 32 grams of sugar, which is equivalent to 8 sugar cubes. Of course, there will be some variability in sugar content. Suppose that the standard deviation is about 0.1 grams. Engineers monitoring production processes assume that measurements are normally distributed.

   A quality control engineer randomly samples 10 Tall Chai Lattes and finds a mean sugar content of 32.07 grams. Is the difference between 32.07 and 32 statistically significant for a sample of 10?

   a) Explain why we can use a normal curve to model the distribution of sample means despite the fact that the samples only contain 10 lattes.

   b) What is the z-score for the engineer's sample? What does the z-score tell us?

   c) Is the sample statistically significant? Support your answer using probability.