Math 146 6.4 - The Central Limit Theorem

THE CENTRAL LIMIT THEOREM

I. For all samples of size $n > 30$, drawn from *any* population with a mean μ and a standard deviation σ, the sampling distribution of $\bar{x}$ (the sample means) can be approximated by a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$. (The larger the sample size, the better the approximation since the standard deviation is getting smaller.)

II. If the population itself is normally distributed, the sampling distribution of sample means is normally distributed for any sample size .

If the *population* has mean μ and standard deviation σ, **and** SRS's of the same size n are selected from the population, then :

Case 1: If the population distribution **is** normally distributed, then the distribution of $\bar{x}$ for **any** sample size n is normally distributed and:

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{and} \qquad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

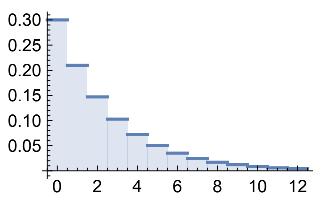
Case 2: If the population is **not** normally distributed, then for $n > 30$, the distribution of $\bar{x}$ is normally distributed with:

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{and} \qquad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

Case 3: If $n \le 30$ and the population is **not** normal, the distribution of $\bar{x}$ cannot be approximated by a normal distribution, and other methods of analysis must be used. :(

COMPUTER DEMO

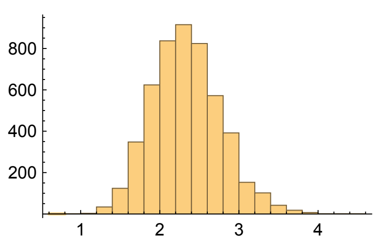
The following is a geometric probability distribution with parameter $\lambda = 0.3$ (it's a discrete distribution, and the probability of events greater than 12 are nearly 0). You can see that the distribution is **far** from normal.



Suppose we take a random sample of size $n = 40$ from this distribution. We would expect a lot more smaller numbers than larger numbers. Here is a random sample from this distribution and the mean of the sample.

7	4	1	1	0	2	0	5	2	3	1	2	7	7	0	3	1	1	0	2
0	1	0	7	2	7	8	2	1	2	0	1	1	3	1	1	4	0	0	0

and the mean if the sample is $\bar{x} = 2.25$. Now, suppose we do this 5000 times and obtain 5000 different sample means, and then make a histogram for the distribution of the sample means. This is what we get:



Notice how the distribution of the sample means looks fairly normal, nothing like the original distribution. If we increase our sample size n, the distribution will even look more normal! For the 5000 sample means above I calculated $\mu_{\bar{x}} = 2.323$, and $\sigma_{\bar{x}} = 0.434$. The actual population mean and standard deviation of the geometric distribution is $\mu = 2.333$ and $\sigma = 2.789$. The mean of the sampling distribution, $\mu_{\bar{x}}$, and the population mean, μ, are nearly identical. AND, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.789}{\sqrt{40}} \approx 0.441$ which again is nearly identical to our calculated $\sigma_{\bar{x}}$.

That is the Central Limit Theorem!!

EXAMPLE 1 A.C. Nielsen reported that children between the ages of 2 and 5 watch an average of 25.2 hours of television per week. Assume the distribution is normally distributed and the standard deviation is 3 hours.

a) Find the probability that if a single child is randomly selected, the mean number of hours they watch television is greater than 27 hours. (Draw a normal curve.)

b) If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the sample mean of all 20 children is greater than 27 hours of television watched.

EXAMPLE 2 The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months.

a) Assuming the ages are normally distributed, find the probability that a randomly selected vehicle will have an age between 90 and 100 months.

b) If a random sample of 40 vehicles is selected, find the probability that the mean of their age is between 90 and 100 months.

c) Find the probability that a sample of 10 cars has a mean age of more than 9 years.

EXAMPLE 3 If, under a certain assumption, the probability of an observed event is extremely small, say $p < 0.05$, we conclude that the assumption is probably not correct. A machine fills 12-oz water bottles in a bottling plant, and has a standard deviation of 0.2 oz. If the machine is determined to be over-filling bottles, the machine line needs to be shut down and recalibrated. Let's assume the machine is filling bottles with a mean volume of 12 oz.

a) Assuming a normal distribution, a single bottle of water is tested and found to have a mean of 12.3 oz. Should the bottling line be shut down and recalibrated?

b) A sample of 20 bottles is found to have a mean of 11.5 oz. Is recalibration necessary?

c) A sample of 50 bottles is found to have a mean of 12.06 oz. Is recalibration necessary.
