In-class Worksheet 12

1. Suppose a population has a distribution with $\mu = 72$ and $\sigma = 8$.

   (a) If we know nothing about the original population distribution and random samples of size $n = 16$ are selected, why can't we say anything about the $\bar{x}$ distribution of sample means?

   The sample size is too small to be certain what the resulting $\bar{x}$ distribution will look like. In order to apply the Central Limit Theorem, which asserts that the $\bar{x}$ distribution is approximately normal, the sample size needs to be $n \geq 30$.

   (b) If we know that the original population distribution is normal, then what can we say about the $\bar{x}$ distribution of random samples of $n = 16$. In this case, find $P(68 \leq \bar{x} \leq 73)$.

   When the original population distribution is normal, then so will the $\bar{x}$ distribution, regardless of the size of the samples.

   The $\bar{x}$ distribution still has a mean $\mu_{\bar{x}} = \mu = 72$, but the standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{16}} = 2$.

   $P(68 \leq \bar{x} \leq 73) = \texttt{normalcdf}(68, 73, 72, 2) \approx 0.6687124058$.

2. The heights of 18-year-old men are approximately normally distributed, with a mean of 68 inches and a standard deviation of 3 inches.

   (a) What is the probability that an 18-year-old man selected at random is between 67 and 69 inches tall?

   $P(67 < x < 69) = \texttt{normalcdf}(67, 69, 68, 3) \approx 0.2611171926$

   (b) If a random sample of nine 18-year-old men are selected at random, what is the probability that the mean height of the sample $\bar{x}$ is between 67 and 69 inches?

   In this case, it is not the value of the random variable x that we are considering, but rather the value of the mean $\bar{x}$ of samples of size $n = 9$.

   Because the distribution of x is normal, the distribution of $\bar{x}$ will be too. Further, the mean of the $\bar{x}$ distribution is still 68, but the standard deviation of that distribution is $\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{9}} = 1$.

   $P(67 < \bar{x} < 69) = \texttt{normalcdf}(67, 69, 68, 1) \approx 0.6826894809$

   (c) Explain why the probability in part (b) is so much higher than the probability in part (a) even though they both refer to the same interval of heights (67 to 69 inches).

   The standard deviation of the sample mean distribution is smaller than that of the original distribution, so there is less variance in the values. That is, the mean height of a randomly selected sample of men is more likely to be close to the mean height of all men than a single randomly selected man's height is.

3. Let x be a random variable that represents white blood cell count per cubic milliliter of whole blood. Assume that x has a distribution that is approximately normal with a mean of $\mu = 7500$ and estimated standard deviation of $\sigma = 1750$. A test result of $x < 3500$ is an indication of leukopenia. This indicates bone marrow repression that may be the result of a viral infection.

   (a) What is the probability that, on a single test, x is less than 3500?

   $P(x < 3500) = \texttt{normalcdf}(-1E99, 3500, 7500, 1750) \approx 0.01113549 \approx 1.11\%$

   (b) Suppose a doctor uses the average $\bar{x}$ for two tests taken about a week apart. What can we say about the probability distribution of $\bar{x}$? What is the probability that $\bar{x} < 3500$?

   Because x has a distribution that is approximately normal, the distribution of the averages of the samples (even with size as small as $n = 2$) is also normal. The mean of the distribution of averages is still 7500, however, the standard deviation is reduced to $\frac{\sigma}{\sqrt{n}} = \frac{1750}{\sqrt{2}}$.

   $P(\bar{x} < 3500) = \texttt{normalcdf}(-1E99, 3500, 7500, 1750/\sqrt{(2)}) = 0.0006135861 \approx 0.061\%$

   (c) Repeat part (b) but with $n = 3$ tests.

   Again, the distribution of the averages of three test results is normal, with mean 7500 and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{1750}{\sqrt{3}}$.

   $P(\bar{x} < 3500) = \texttt{normalcdf}(-1E99, 3500, 7500, 1750\sqrt{(3)}) = 0.00003763633 \approx 0.0038\%$

   (d) How did the probabilities change as n increased? What do such results imply about a patient that has $\bar{x} < 3500$ for 3 those tests?

   Although it is possible for an isolated random test to indicate a low count when absolutely nothing is wrong, the greater n is, it becomes less and less likely that the average of the tests will be less than 3500 when absolutely nothing is wrong.

   According to part (c), the likelihood that the average of 3 tests is less than 3500 just by random chance is only 0.00003763633. So, for a patient with such results, the probability is extremely high that they do, in fact, have leukopenia.

4. Assume that IQ scores are normally distributed, with a standard deviation of 15 points and a mean of 100 points. If 10 people are chosen at random, what is the probability that the sample mean of their IQ scores will not differ from the population mean by more than 2 points?

   In this case, we are looking for the probability that the sample mean $\bar{x}$ falls between 98 and 102 points. Because the distribution of IQ scores is normal, the Central Limit Theorem can be applied with $\mu_{\bar{x}} = 100$ and $\sigma_{\bar{x}} = \frac{15}{\sqrt{10}}$.

   So, $P(98 \le \bar{x} \le 102) = \texttt{normalcdf}(98, 102, 100, \frac{15}{\sqrt{10}}) \approx 0.3267099507 \approx 32.67\%$.

5. A large tank of fish from a hatchery is being delivered to a lake. The hatchery claims that the mean length of fish in the tank is 15 inches, and the standard deviation is 2 inches. A random sample of 36 fish is taken from the tank. Let $\bar{x}$ be the mean length of the sample. What is the probability that $\bar{x}$ is within 0.5 inches of the claimed population mean?

   Although there is no indication that the distribution of lengths of the fish is normal, the sample size is greater than 30, so the Central Limit Theorem applies. So, we can apply the `normalcdf` function with $\mu_{\bar{x}} = 15$ and $\sigma_{\bar{x}} = 2/\sqrt{36} = 1/3$.

   `normalcdf(14.5, 15.5, 15, 1/3)` $\approx 0.8663855426 \approx 86.64\%$