Andrew Gard - equitable.equations@gmail.com

The Central Limit Theorem

The Central Limit Theorem describes the sampling distribution of the sample mean, $\overline{x}$. Make sure you understand sampling distributions before proceeding!

The Central Limit Theorem describes the sampling distribution of the sample mean, $\overline{x}$. Make sure you understand sampling distributions before proceeding!

The Central Limit Theorem. Suppose simple random samples of size n are drawn from a population with mean μ and standard deviation σ. If n is large (usually 30 is sufficient), then the sampling distribution of the sample mean is approximately normal.

The Central Limit Theorem describes the sampling distribution of the sample mean, $\overline{x}$. Make sure you understand sampling distributions before proceeding!

The Central Limit Theorem. Suppose simple random samples of size n are drawn from a population with mean μ and standard deviation σ. If n is large (usually 30 is sufficient), then the sampling distribution of the sample mean is approximately normal. If the population distribution is normal, then the sampling distribution of $\overline{x}$ is exactly normal, regardless of n.

The Central Limit Theorem describes the sampling distribution of the sample mean, $\overline{x}$. Make sure you understand sampling distributions before proceeding!

The Central Limit Theorem. Suppose simple random samples of size n are drawn from a population with mean μ and standard deviation σ. If n is large (usually 30 is sufficient), then the sampling distribution of the sample mean is approximately normal. If the population distribution is normal, then the sampling distribution of $\overline{x}$ is exactly normal, regardless of n.

Moreover, it is always the case that

$$\mu_{\overline{x}} = \mu$$

The Central Limit Theorem describes the sampling distribution of the sample mean, $\overline{x}$. Make sure you understand sampling distributions before proceeding!

The Central Limit Theorem. Suppose simple random samples of size n are drawn from a population with mean μ and standard deviation σ. If n is large (usually 30 is sufficient), then the sampling distribution of the sample mean is approximately normal. If the population distribution is normal, then the sampling distribution of $\overline{x}$ is exactly normal, regardless of n.

Moreover, it is always the case that

$$\begin{aligned} \mu_{\overline{x}} &= \mu \\ \sigma_{\overline{x}} &= \sigma/\sqrt{n} \end{aligned}$$

The Central Limit Theorem describes the sampling distribution of the sample mean, $\overline{x}$. Make sure you understand sampling distributions before proceeding!

The Central Limit Theorem. Suppose simple random samples of size n are drawn from a population with mean μ and standard deviation σ. If n is large (usually 30 is sufficient), then the sampling distribution of the sample mean is approximately normal. If the population distribution is normal, then the sampling distribution of $\overline{x}$ is exactly normal, regardless of n.

Moreover, it is always the case that

$$\begin{aligned} \mu_{\overline{x}} &= \mu \\ \sigma_{\overline{x}} &= \sigma/\sqrt{n} \end{aligned}$$

So the Central Limit Theorem says that the distribution of $\overline{x}$ is always approximately $N\left(\mu, \sigma^2/n\right)$.

Example 1. The length of calls to a tech help line have mean $\mu = 2$ minutes and standard deviation $\sigma = 3$ minutes. Find the probability that 40 randomly-selected calls have mean length less than 2.5 minutes.

Example 1. The length of calls to a tech help line have mean $\mu = 2$ minutes and standard deviation $\sigma = 3$ minutes. Find the probability that 40 randomly-selected calls have mean length less than 2.5 minutes.

The 40 calls represent a random sample from the population of all calls to the center. By the Central Limit Theorem, the sample mean $\overline{x}$ is approximately normal with mean 2 and standard deviation $3/\sqrt{40}$.

Example 1. The length of calls to a tech help line have mean $\mu = 2$ minutes and standard deviation $\sigma = 3$ minutes. Find the probability that 40 randomly-selected calls have mean length less than 2.5 minutes.

The 40 calls represent a random sample from the population of all calls to the center. By the Central Limit Theorem, the sample mean $\overline{x}$ is approximately normal with mean 2 and standard deviation $3/\sqrt{40}$. We need to find $P(\overline{x} < 2.5)$ in this distribution.

Example 1. The length of calls to a tech help line have mean $\mu = 2$ minutes and standard deviation $\sigma = 3$ minutes. Find the probability that 40 randomly-selected calls have mean length less than 2.5 minutes.

The 40 calls represent a random sample from the population of all calls to the center. By the Central Limit Theorem, the sample mean $\overline{x}$ is approximately normal with mean 2 and standard deviation $3/\sqrt{40}$. We need to find $P(\overline{x} < 2.5)$ in this distribution.

$$P(\overline{x} < 2.5) \text{ in } N(2, 9/40) \quad = \quad P\left(z < \frac{2.5 - 2}{3/\sqrt{40}}\right) \text{ in } N(0,1)$$

Example 1. The length of calls to a tech help line have mean $\mu = 2$ minutes and standard deviation $\sigma = 3$ minutes. Find the probability that 40 randomly-selected calls have mean length less than 2.5 minutes.

The 40 calls represent a random sample from the population of all calls to the center. By the Central Limit Theorem, the sample mean $\overline{x}$ is approximately normal with mean 2 and standard deviation $3/\sqrt{40}$. We need to find $P(\overline{x} < 2.5)$ in this distribution.

$$
\begin{aligned}
P(\overline{x} < 2.5) \text{ in } N(2, 9/40) &= P\left(z < \frac{2.5 - 2}{3/\sqrt{40}}\right) \text{ in } N(0,1) \\
&= P(z < 1.05)
\end{aligned}
$$

Example 1. The length of calls to a tech help line have mean $\mu = 2$ minutes and standard deviation $\sigma = 3$ minutes. Find the probability that 40 randomly-selected calls have mean length less than 2.5 minutes.

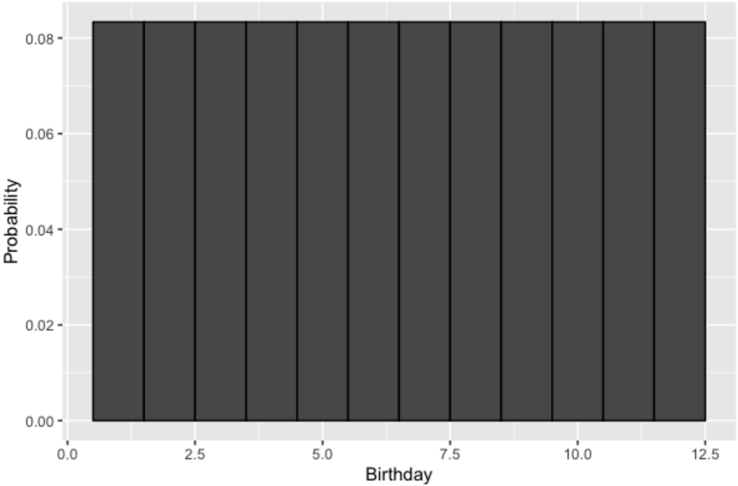
The 40 calls represent a random sample from the population of all calls to the center. By the Central Limit Theorem, the sample mean $\overline{x}$ is approximately normal with mean 2 and standard deviation $3/\sqrt{40}$. We need to find $P(\overline{x} < 2.5)$ in this distribution.

$$\begin{aligned}
P(\overline{x} < 2.5) \text{ in } N(2, 9/40) &= P\left(z < \frac{2.5 - 2}{3/\sqrt{40}}\right) \text{ in } N(0,1) \\
&= P(z < 1.05) \\
&= 85.3\%
\end{aligned}$$

Normal probabilities like this should be calculated using technology. In *R*, for instance, the commands pnorm(1.05) and pnorm(2.5, 2, 3/sqrt(40)) both do the job.

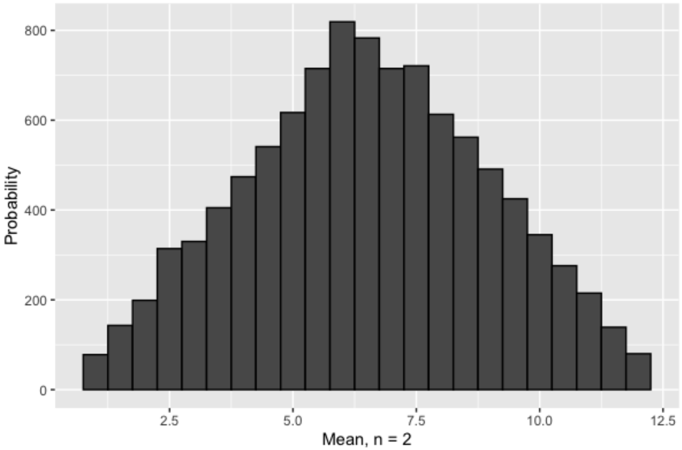
Example 2. A random number generator produces random integers from 1-12 with equal probability.

Example 2. A random number generator produces random integers from 1-12 with equal probability. The probability histogram for the results is shown below.



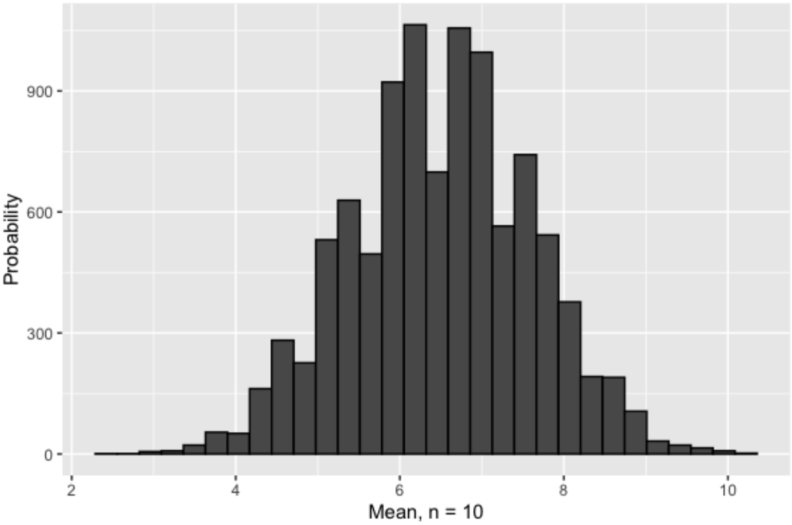
Next, we use the random number generator twice and take the mean, $\overline{x}$. The result is again a number between 1 and 12.

Next, we use the random number generator twice and take the mean, $\bar{x}$. The result is again a number between 1 and 12. If we do this, say, 1000 times and plot the results, we get a histogram like this:

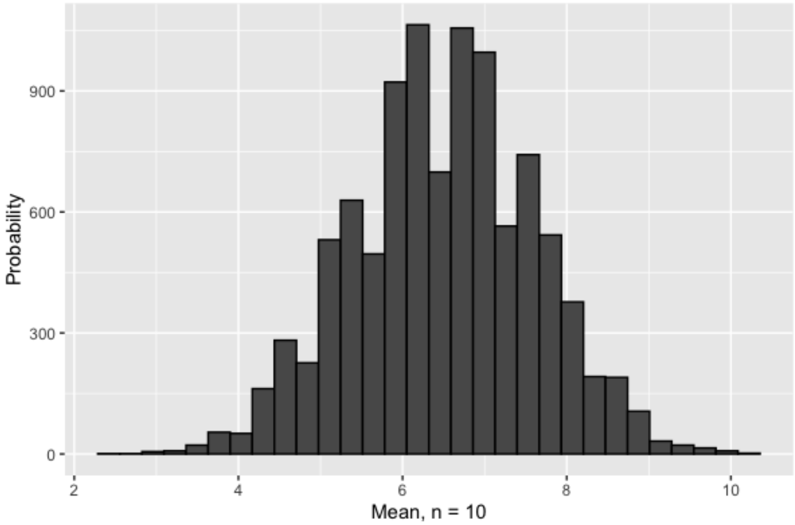


Results near 6.5 are more likely, while 1's and 12's are rare.

Now take 1000 samples of size $n = 10$.

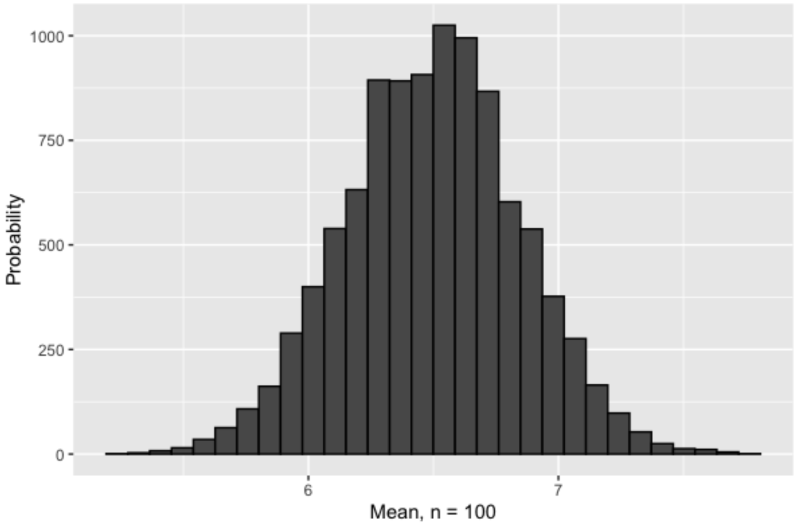


Now take 1000 samples of size $n = 10$.



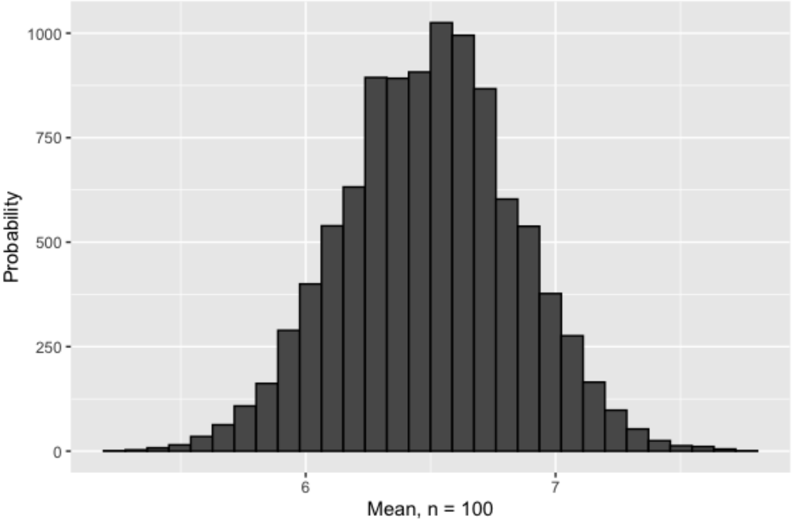
The histogram is beginning to have a vaguely bell shape. Note also that nearly all of the results lie between 4 and 9.

Using 1000 samples of size 100:



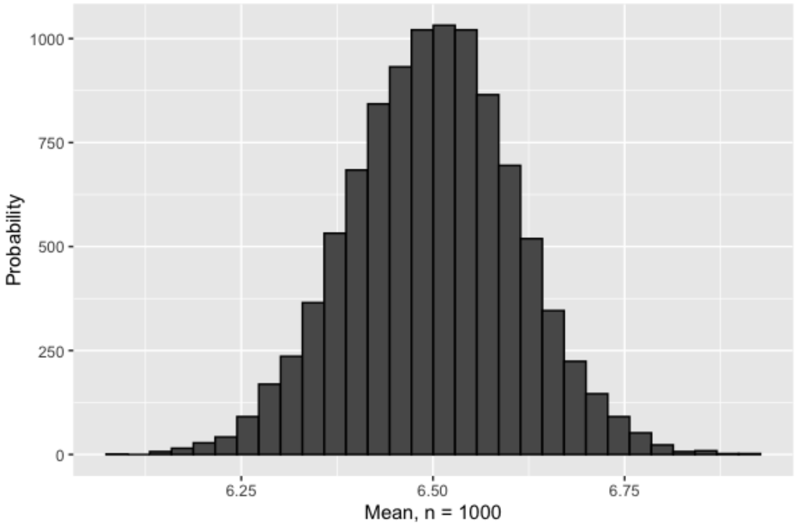
The bell shape is clear.

Using 1000 samples of size 100:



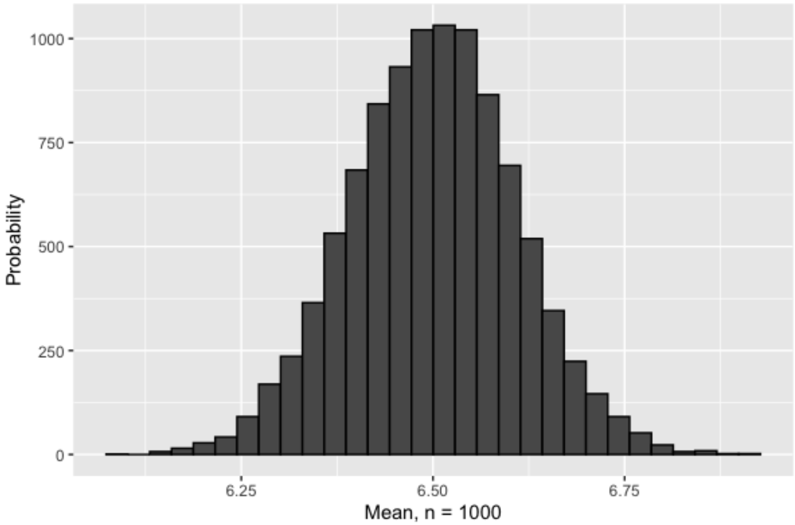
The bell shape is clear. Now, most of the data lies between 6 and 7.

Finally, let's use 1000 samples of size 1000:



A normal curve fits these results exceptionally well.

Finally, let's use 1000 samples of size 1000:



A normal curve fits these results exceptionally well. Most of the results lie between 6.25 and 6.75

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

For any n, the mean and standard deviation of $\overline{x}$ are μ and $\sigma/\sqrt{n}$, respectively.

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

For any n, the mean and standard deviation of $\overline{x}$ are μ and $\sigma/\sqrt{n}$, respectively.

$$n = 2: \qquad \mu_{\overline{x}} = 6.5, \qquad \sigma_{\overline{x}} \approx 2.44$$

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

For any n, the mean and standard deviation of $\overline{x}$ are μ and $\sigma/\sqrt{n}$, respectively.

$$\begin{aligned} n = 2: &\qquad \mu_{\overline{x}} = 6.5, \qquad \sigma_{\overline{x}} \approx 2.44 \\ n = 10: &\qquad \mu_{\overline{x}} = 6.5, \qquad \sigma_{\overline{x}} \approx 1.09 \end{aligned}$$

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

For any n, the mean and standard deviation of $\overline{x}$ are μ and $\sigma/\sqrt{n}$, respectively.

$$\begin{aligned} n = 2: &\qquad \mu_{\overline{x}} = 6.5, \qquad \sigma_{\overline{x}} \approx 2.44 \\ n = 10: &\qquad \mu_{\overline{x}} = 6.5, \qquad \sigma_{\overline{x}} \approx 1.09 \\ n = 100: &\qquad \mu_{\overline{x}} = 6.5, \qquad \sigma_{\overline{x}} \approx 0.35 \end{aligned}$$

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

For any n, the mean and standard deviation of $\overline{x}$ are μ and $\sigma/\sqrt{n}$, respectively.

$$\begin{aligned} n = 2: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 2.44 \\ n = 10: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 1.09 \\ n = 100: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 0.35 \\ n = 1000: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 0.11 \end{aligned}$$

Let's finish this example by actually computing the mean and standard deviation of each of these sampling distributions. If we randomly select a single number between 1 and 12, each value has a 1/12 chance of being picked. The mean and standard deviation of this probability distribution are

$$\mu = 6.50 \quad \text{and} \quad \sigma = 3.45$$

For any n, the mean and standard deviation of $\overline{x}$ are μ and $\sigma/\sqrt{n}$, respectively.

$$\begin{aligned}
n = 2: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 2.44 \\
n = 10: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 1.09 \\
n = 100: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 0.35 \\
n = 1000: &\qquad \mu_{\overline{x}} = 6.5, &\qquad \sigma_{\overline{x}} \approx 0.11
\end{aligned}$$

Roughly speaking, as n increases, the sample mean $\overline{x}$ becomes a more reliable estimator of μ.

Example 3. A distilled-water dispenser vends 'gallons' of water according to a normal distribution with mean 1.03 gallons and standard deviation .02 gallons. What is the probability that a single 'gallon' dispensed is actually under 1 gallon? What is the probability that 10 dispensed 'gallons' average less than 1 gallon each?

Example 3. A distilled-water dispenser vends 'gallons' of water according to a normal distribution with mean 1.03 gallons and standard deviation .02 gallons. What is the probability that a single 'gallon' dispensed is actually under 1 gallon? What is the probability that 10 dispensed 'gallons' average less than 1 gallon each?

For an individual 'gallon' x,

$$P(x < 1.00) \text{ in } N\left(1.03, .02^2\right) \quad = \quad P\left(z < \frac{1 - 1.03}{.02}\right) \text{ in } N\left(0, 1\right)$$

Example 3. A distilled-water dispenser vends 'gallons' of water according to a normal distribution with mean 1.03 gallons and standard deviation .02 gallons. What is the probability that a single 'gallon' dispensed is actually under 1 gallon? What is the probability that 10 dispensed 'gallons' average less than 1 gallon each?

For an individual 'gallon' x,

$$\begin{aligned} P(x < 1.00) \text{ in } N\left(1.03, .02^2\right) &= P\left(z < \frac{1 - 1.03}{.02}\right) \text{ in } N(0,1) \\ &= P(z < -1.50) \end{aligned}$$

Example 3. A distilled-water dispenser vends 'gallons' of water according to a normal distribution with mean 1.03 gallons and standard deviation .02 gallons. What is the probability that a single 'gallon' dispensed is actually under 1 gallon? What is the probability that 10 dispensed 'gallons' average less than 1 gallon each?

For an individual 'gallon' x,

$$\begin{aligned} P(x < 1.00) \text{ in } N\left(1.03, .02^2\right) &= P\left(z < \frac{1 - 1.03}{.02}\right) \text{ in } N(0,1) \\ &= P(z < -1.50) \\ &= 6.68\% \end{aligned}$$

On the other hand, when $n = 10$, the sample mean has the distribution $N(1.03, .02^2/10)$, by the Central Limit Theorem.

On the other hand, when $n = 10$, the sample mean has the distribution $N(1.03, .02^2/10)$, by the Central Limit Theorem. Since the original distribution is normal, this distribution is exact.

On the other hand, when $n = 10$, the sample mean has the distribution $N(1.03, .02^2/10)$, by the Central Limit Theorem. Since the original distribution is normal, this distribution is exact.

$$P(\overline{x} < 1.00) \text{ in } N\left(1.03, .02^2/10\right) \quad = \quad P\left(z < \frac{1 - 1.03}{.02/\sqrt{10}}\right) \text{ in } N\left(0, 1\right)$$

On the other hand, when $n = 10$, the sample mean has the distribution $N(1.03, .02^2/10)$, by the Central Limit Theorem. Since the original distribution is normal, this distribution is exact.

$$
\begin{aligned}
P(\overline{x} < 1.00) \text{ in } N\left(1.03, .02^2/10\right) &= P\left(z < \frac{1 - 1.03}{.02/\sqrt{10}}\right) \text{ in } N\left(0, 1\right) \\
&= P(z < -4.74)
\end{aligned}
$$

On the other hand, when $n = 10$, the sample mean has the distribution $N(1.03, .02^2/10)$, by the Central Limit Theorem. Since the original distribution is normal, this distribution is exact.

$$
\begin{aligned}
P(\overline{x} < 1.00) \text{ in } N\left(1.03, .02^2/10\right) &= P\left(z < \frac{1 - 1.03}{.02/\sqrt{10}}\right) \text{ in } N\left(0, 1\right) \\
&= P(z < -4.74) \\
&= .0001\%
\end{aligned}
$$

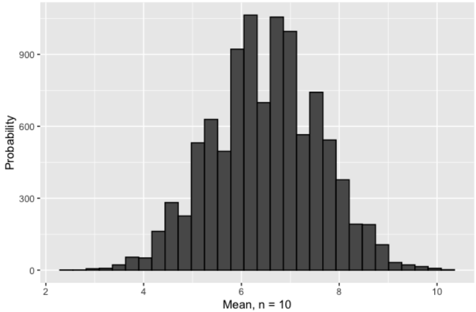
We conclude that while it's somewhat unlikely (6.68% chance) that a single 'gallon' will be underfilled, it would be extremely unusual if the mean of 10 'gallons' were less than 1.

The Central Limit Theorem says that the sampling distribution of $\overline{x}$ is approximately normal for large n, and suggests that $n = 30$ counts as large.

The Central Limit Theorem says that the sampling distribution of $\overline{x}$ is approximately normal for large n, and suggests that $n = 30$ counts as large. In fact, $n = 30$ is usually very conservative. When the distribution being sampled from isn't highly skewed, a much smaller n is acceptable.

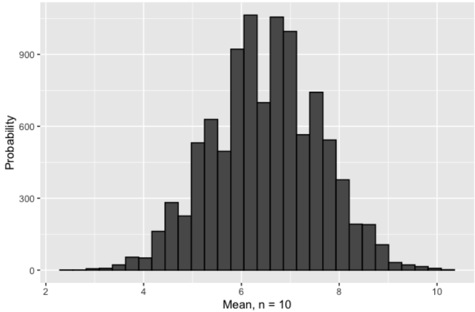
The Central Limit Theorem says that the sampling distribution of $\overline{x}$ is approximately normal for large n, and suggests that $n = 30$ counts as large. In fact, $n = 30$ is usually very conservative. When the distribution being sampled from isn't highly skewed, a much smaller n is acceptable.

In example 2, for instance, where we selected numbers between 1 and 12 at random, the normal shape was clear even for $n = 10$.



The Central Limit Theorem says that the sampling distribution of $\overline{x}$ is approximately normal for large n, and suggests that $n = 30$ counts as large. In fact, $n = 30$ is usually very conservative. When the distribution being sampled from isn't highly skewed, a much smaller n is acceptable.

In example 2, for instance, where we selected numbers between 1 and 12 at random, the normal shape was clear even for $n = 10$.



The more skewed the distribution, the larger the sample size needed.