Sampling Distributions and the Central Limit Theorem (CLT)

Sahir Bhatnagar and James Hanley

EPIB 607
Department of Epidemiology, Biostatistics, and Occupational Health
McGill University

sahir.bhatnagar@mcgill.ca
https://sahirbhatnagar.com/EPIB607/

September 20, 2018

Parameters, Samples, and Statistics

Parameters, Samples, and Statistics

- **Paramter**: An unknown numerical constant pertaining to a population/universe, or in a statistical model.
  - μ: population mean π: population proportion

Parameters, Samples, and Statistics

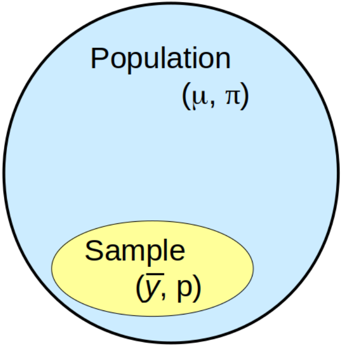
- **Paramter**: An unknown numerical constant pertaining to a population/universe, or in a statistical model.
  - μ: population mean π: population proportion
- **Statistic**: A numerical quantity calculated from a sample. The empirical counterpart of the parameter, used to *estimate* it.

Parameters, Samples, and Statistics

- **Paramter**: An unknown numerical constant pertaining to a population/universe, or in a statistical model.
  - μ: population mean $\qquad$ π: population proportion
- **Statistic**: A numerical quantity calculated from a sample. The empirical counterpart of the parameter, used to *estimate* it.
  - $\bar{y}$: sample mean $\qquad$ p: sample proportion

Parameters, Samples, and Statistics

- **Paramter**: An unknown numerical constant pertaining to a population/universe, or in a statistical model.
  - μ: population mean $\quad\quad\quad\quad$ π: population proportion
- **Statistic**: A numerical quantity calculated from a sample. The empirical counterpart of the parameter, used to *estimate* it.
  - $\bar{y}$: sample mean $\quad\quad\quad\quad$ p: sample proportion



Examples

Proportions:

- Proportion of Earth's surface covered by water
- Proportion who saw a medical doctor last year
- Proportion of Québécois who don't have a family doctor

Examples

Proportions:

- Proportion of Earth's surface covered by water
- Proportion who saw a medical doctor last year
- Proportion of Québécois who don't have a family doctor

Means:

- Mean depth in n randomly selected ocean locations
- Mean household size in n randomly selected households.
- Median number of persons under-5 in a sample of n households

Samples must be random

- The validity of inference will depend on the way that the sample was collected. If a sample was collected badly, no amount of statistical sophistication can rescue the study.

Samples must be random

- The validity of inference will depend on the way that the sample was collected. If a sample was collected badly, no amount of statistical sophistication can rescue the study.

- Samples should be **random**. That is, there should be no systematic set of characteristics that is related to the scientific question of interest that causes some people to be more likely to be sampled than others. The simplest type of randomization selects members from the population with equal probability (a uniform distribution).

Samples must be random

- The validity of inference will depend on the way that the sample was collected. If a sample was collected badly, no amount of statistical sophistication can rescue the study.

- Samples should be **random**. That is, there should be no systematic set of characteristics that is related to the scientific question of interest that causes some people to be more likely to be sampled than others. The simplest type of randomization selects members from the population with equal probability (a uniform distribution).

- When conducting a study, it is always better to seek statistical advice sooner rather than later. Get a statistician involved at the *planning* stage of the study... by the analysis stage, it may be too late!

Samples must be random - No cheating!

Do not cheat by

Samples must be random - No cheating!

Do not cheat by

- Taking 5 people from the *same* household to estimate
  - proportion of Québécois who don't have a family doctor
  - who saw a medical doctor last year
  - average rent

Samples must be random - No cheating!

Do not cheat by

- Taking 5 people from the *same* household to estimate
  - proportion of Québécois who don't have a family doctor
  - who saw a medical doctor last year
  - average rent

- Sampling the depth of the ocean *only around Montreal* to estimate
  - proportion of Earth's surface covered by water

Collecting data takes effort

In general

- The larger the sample $\rightarrow$ the more accurate the estimate (if sampling is done correctly)

Collecting data takes effort

In general

- The larger the sample $\rightarrow$ the more accurate the estimate (if sampling is done correctly)

CAVEAT

- Collecting more data takes effort and money!
- We will also soon discover the curse of the $\sqrt{n}$

Collecting data takes effort

In general

- The larger the sample $\rightarrow$ the more accurate the estimate (if sampling is done correctly)

Collecting data takes effort

In general

- The larger the sample $\rightarrow$ the more accurate the estimate (if sampling is done correctly)

CAVEAT

- Collecting more data takes effort and money!
- We will also soon discover the curse of the $\sqrt{n}$

Sampling Distributions

Sampling Distributions

- Given a sample of n observations from a population, we will be calculating estimates of the population mean, proportion, standard deviation, and various other population characteristics (parameters)

Sampling Distributions

- Given a sample of n observations from a population, we will be calculating estimates of the population mean, proportion, standard deviation, and various other population characteristics (parameters)

- Prior to obtaining data, there is uncertainty as to which of all possible samples will occur

Sampling Distributions

- Given a sample of n observations from a population, we will be calculating estimates of the population mean, proportion, standard deviation, and various other population characteristics (parameters)

- Prior to obtaining data, there is uncertainty as to which of all possible samples will occur

- Because of this, estimates such as $\bar{y}$ (the sample mean) will vary from one sample to another

Sampling Distributions

- The behavior of such estimates in many samples of equal size is described by what are called **sampling distributions**

Sampling Distributions

- The behavior of such estimates in many samples of equal size is described by what are called **sampling distributions**
- B&M definition: The sampling distribution of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population.

Why are sampling distributions important?

- They tell us how far from the target (true value of the parameter) our statistical *shot* at it (i.e. the statistic calculated form a sample) is likely to be, or, to have been.

Why are sampling distributions important?

- They tell us how far from the target (true value of the parameter) our statistical *shot* at it (i.e. the statistic calculated form a sample) is likely to be, or, to have been.

- Thus, they are used in confidence intervals for parameters. Specific sampling distributions (based on a null value for the parameter) are also used in statistical tests of hypotheses.

Exercise 1: How Deep is the Ocean?

- We will get a sense of what a sampling distribution is in Exercise 1

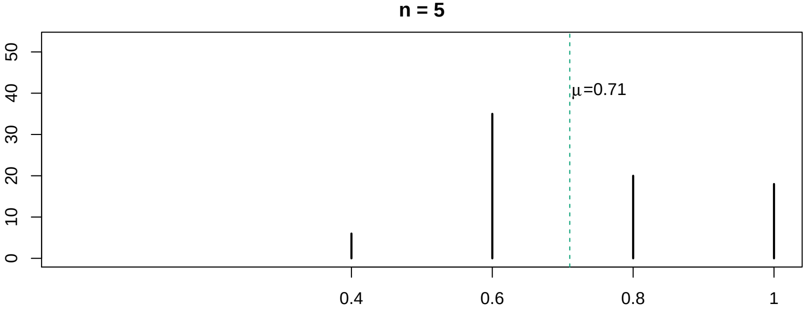
Exercise 1: How Deep is the Ocean?

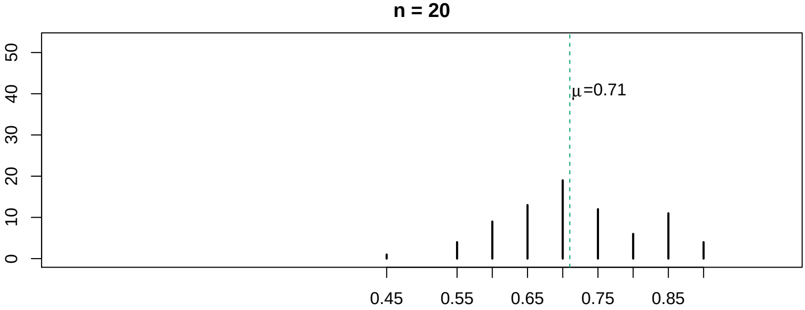
- We will get a sense of what a sampling distribution is in Exercise 1

- CAVEAT: This is a luxury using a toy example. In actual studies, we only get one shot!

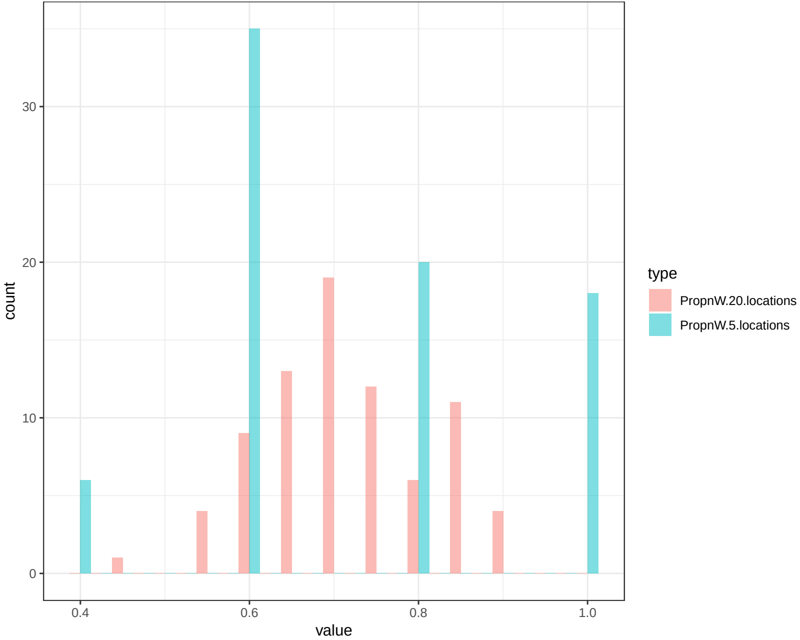
Exercise 1 Results

Sampling distribution: proportion covered by water

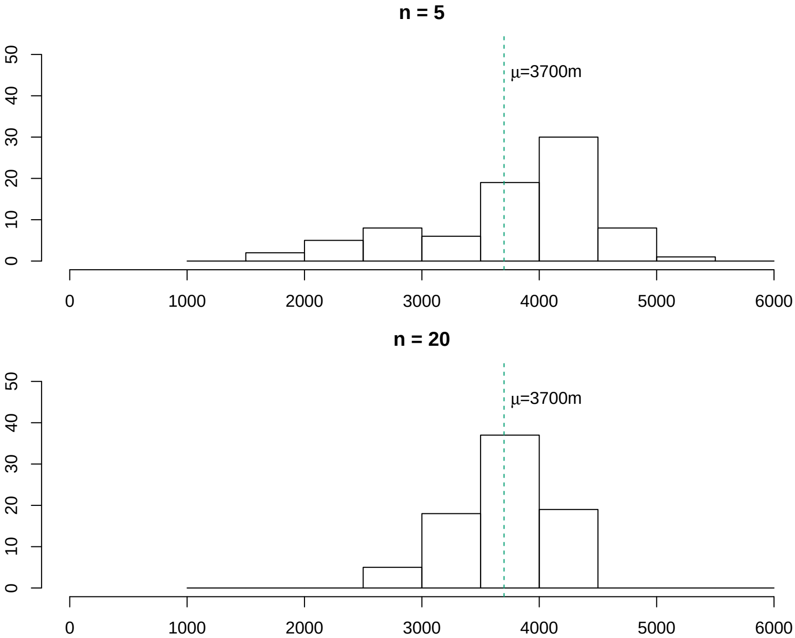




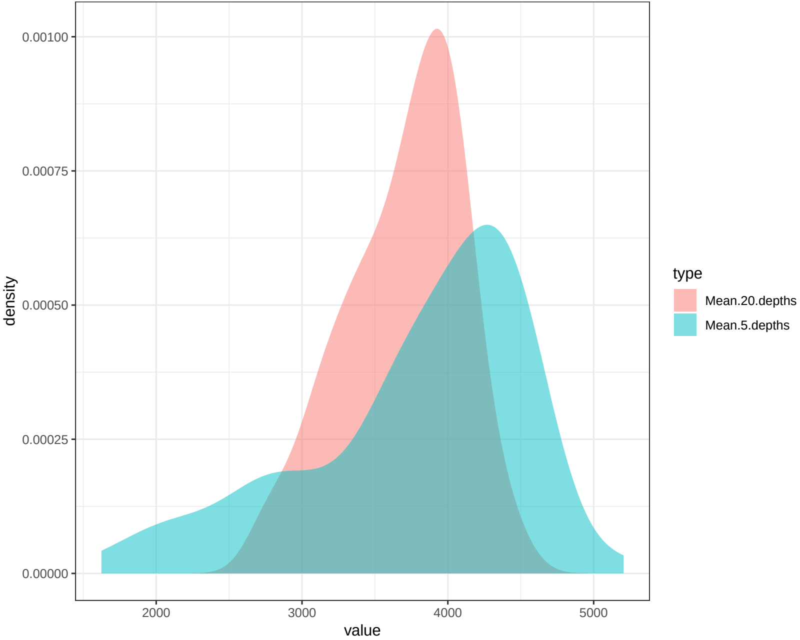
Sampling distribution: proportion covered by water



Sampling distribution: mean depth of the ocean



Sampling distribution: mean depth of the ocean



Key takeaways and next steps

1. We’ve been exclusively talking about point estimates

Key takeaways and next steps

1. We've been exclusively talking about point estimates
2. How confident are we about these point estimates?

Key takeaways and next steps

1. We've been exclusively talking about point estimates
2. How confident are we about these point estimates?
3. Thought experiment: Estimate the average temperature in Montreal in August over the past 100 years.

Key takeaways and next steps

1. We've been exclusively talking about point estimates
2. How confident are we about these point estimates?
3. Thought experiment: Estimate the average temperature in Montreal in August over the past 100 years.
4. We're going into stat territory now.

Normal Curves and Calculations

The Normal (Gaussian) distribution

What is it?

- A distribution that describes continuous (numerical) data
- Can also be used to approximate discrete data distributions
- Range is (technically) infinite, though the probability of seeing very large or very small values is extremely tiny
- Fully described by only two parameters, the mean and variance (μ and σ^2)
- NOTE: Baldi & Moore (and **R**) use the short-hand: $X \sim \mathcal{N}(\mu, \sigma)$, denoting the normal distribution as a function of the mean and *standard deviation*. This is not standard; many texts instead write $X \sim \mathcal{N}(\mu, \sigma^2)$. Be careful of this!

The Normal (Gaussian) distribution

Carl Gauss was a German mathematician who developed a number of important advances in statistics such as the method of least squares.

The Normal distribution

Where do Normal data come from?

- Natural processes
  - Blood pressure
  - Height
  - Weight

The Normal distribution

Where do Normal data come from?

- Natural processes
  - Blood pressure
  - Height
  - Weight

- "Man-made" (or derived)
  - Binomial (proportion) and Poisson (count) data are approximately Normal under certain conditions
  - Sums and means of random variables (Central Limit Theorem)
  - Data can sometimes be made to look Normal via transformations (squares, logs, etc)

The Normal distribution

For Normal data, we can use ~~the Gaussian tables~~ R to answer the questions:

- What is the probability that a single observation X is
  - greater than X^*?
  - less than X^*?
  - between X_L^* and X_U^*?
- That is, we can find out information about the percent distribution of X as a function of thresholds X^*, or X_L^* and X_U^*.

The Normal distribution

For Normal data, we can use ~~the Gaussian tables~~ R to answer the questions:

- What is the probability that a single observation X is
  - greater than X^*?
  - less than X^*?
  - between X_L^* and X_U^*?
- That is, we can find out information about the percent distribution of X as a function of thresholds X^*, or X_L^* and X_U^*.
- We can also use ~~the Normal tables~~ R to find out information about thresholds X^* that will contain particular percentages of the data. I.e., we can find what threshold values will
  - Exclude the lower ω^*% of a population
  - Exclude the upper ω^*% of a population
  - Contain the middle ω^*% of a population

The Normal distribution

We can use ~~the Gaussian tables~~ R to answer these questions **no matter what the values of** μ and σ^2.

That is, the % of the Normal distribution falling between $X_L^* = \mu - m_1\sigma$ and $X_U^* = \mu + m_2\sigma$ where m_1, m_2 are any multiples **remains the same** for any μ and σ.

How so??

The Normal distribution

We can use ~~the Gaussian tables~~ R to answer these questions **no matter what the values of** μ and σ^2.

That is, the % of the Normal distribution falling between $X_L^* = \mu - m_1\sigma$ and $X_U^* = \mu + m_2\sigma$ where m_1, m_2 are any multiples **remains the same** for any μ and σ.

How so??

Because we can **standardize** any $X \sim \mathcal{N}(\mu, \sigma)$ to find $Z \sim \mathcal{N}(0, 1)$

The Normal distribution

An illustration using IQ scores, which we presume have a $\mathcal{N}(100, 13)$ distribution of scores.

Q1: What percentage of scores are **above** 130?
Two steps:

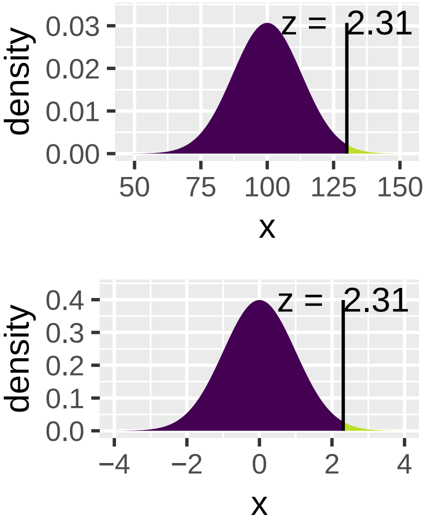
1. Change of location from $\mu_X = 100$ to $\mu_Z = 0$
2. Change of scale from $\sigma_X = 13$ to $\sigma_Z = 1$

Together, this gives us

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{130 - 100}{13} = 2.31$$

The Normal distribution

The position of X=130 in a $\mathcal{N}(100, 13)$ distribution is the same as the place of $Z = 2.31$ on the $\mathcal{N}(0, 1)$, which we call the **standardized** Normal distribution (or Z-distribution).



The Normal distribution

How are the values in the Normal tables found?

Normal density:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2}$$

Probabilities found by integration (area under the Normal curve):

$$P(a \leq x \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} dx$$

The Normal distribution

(The percent above $X = 130$) = (% above $Z = 2.31$) =1.04%

How do we know this? We look at the lower tail probability of 2.31 [i.e., the % below 2.31], and then subtract it from 1:

1. $P(X < 130) = P(Z < 2.31) = 0.9896$
2. $P(X > 130) = 1 - P(X < 130) = 0.0104$

So 130 is the 98.96th percentile of a $\mathcal{N}$(100,13) distribution.

Reminder about percentiles and quantiles

- Quantile
  - Any set of data, arranged in ascending or descending order, can be divided into various parts, also known as partitions or subsets, regulated by quantiles.
  - Quantile is a generic term for those values that divide the set into partitions of size n, so that each part represents $1/n$ of the set.
  - Quantiles are not the partition itself. They are the numbers that define the partition.
  - You can think of them as a sort of numeric boundary.

Reminder about percentiles and quantiles

- **Quantile**
  - Any set of data, arranged in ascending or descending order, can be divided into various parts, also known as partitions or subsets, regulated by quantiles.
  - Quantile is a generic term for those values that divide the set into partitions of size n, so that each part represents $1/n$ of the set.
  - Quantiles are not the partition itself. They are the numbers that define the partition.
  - You can think of them as a sort of numeric boundary.

- **Percentile**
  - Percentiles are quite similar to quantiles: they split your set, but only into two partitions.
  - For a generic kth percentile, the lower partition contains k% of the data, and the upper partition contains the rest of the data, which amounts to 100 - k %, because the total amount of data is 100%.
  - Of course k can be any number between 0 and 100.

More about percentiles and quantiles

- In class, we will find ourselves asking for the quantiles of a distribution.
- Percentiles go from 0 to 100
- Quantiles go from any number to any number
- Percentiles are examples of quantiles and you might find some people use them interchangeably (though this may not always be correct since quantiles can take on any value, positive or negative).
- **In particular**, R uses the term quantiles.
- In the previous example, we saw that $P(Z < 2.31) = 0.9896$. In R, 2.31 is called the quantile .

The Normal distribution

(The percent above $X = 130$) = (% above $Z = 2.31$) =1.04%

But wait!! The standard Normal is symmetric about 0, so we can do this another way... The % **above** 2.31 is equal to the % **below** -2.31:

$$P(X > 130) = P(Z > 2.31)$$

The Normal distribution

(The percent above $X = 130$) = (% above $Z = 2.31$) =1.04%

But wait!! The standard Normal is symmetric about 0, so we can do this another way... The % **above** 2.31 is equal to the % **below** -2.31:

$$P(X > 130) = P(Z > 2.31)$$
$$\Rightarrow P(Z > 2.31) = P(Z < -2.31)$$

The Normal distribution

(The percent above $X = 130$) = (% above $Z = 2.31$) =1.04%

But wait!! The standard Normal is symmetric about 0, so we can do this another way... The % **above** 2.31 is equal to the % **below** -2.31:

$$
\begin{aligned}
&P(X > 130) = P(Z > 2.31) \\
&\qquad \Rightarrow P(Z > 2.31) = P(Z < -2.31) \\
&\qquad \Rightarrow P(X > 130) = P(Z < -2.31) = 0.0104
\end{aligned}
$$

So 130 is the 98.96^{th} percentile of a $\mathcal{N}(130, 13)$ distribution. What is the 1.04^{th} percentile?

The Normal distribution

(The percent above $X = 130$) = (% above $Z = 2.31$) =1.04%

But wait!! The standard Normal is symmetric about 0, so we can do this another way... The % **above** 2.31 is equal to the % **below** -2.31:

$$P(X > 130) = P(Z > 2.31)$$
$$\Rightarrow P(Z > 2.31) = P(Z < -2.31)$$
$$\Rightarrow P(X > 130) = P(Z < -2.31) = 0.0104$$

So 130 is the 98.96^{th} percentile of a $\mathcal{N}(130, 13)$ distribution. What is the 1.04^{th} percentile?

Transform from $Z = -2.31$ back to X:

$$X = \sigma Z + \mu = 13(-2.31) + 100 = 69.97.$$

For probabilities we use *pnorm*

```r
stats::pnorm(q = 130, mean = 100, sd = 13)
```

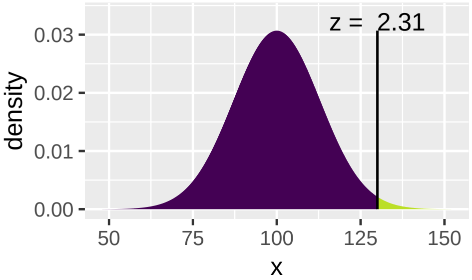
```
## [1] 0.9894919
```

For probabilities we use *pnorm*

```
stats::pnorm(q = 130, mean = 100, sd = 13)
```

```
## [1] 0.9894919
```

```
mosaic::xpnorm(q = 130, mean = 100, sd = 13)
```

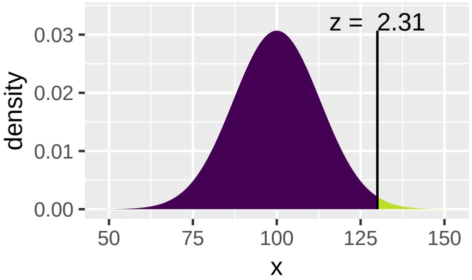



```
## [1] 0.9894919
```

For probabilities we use *pnorm*

```
stats::pnorm(q = 130, mean = 100, sd = 13)

## [1] 0.9894919
```

```
mosaic::xpnorm(q = 130, mean = 100, sd = 13)
```




```
## [1] 0.9894919
```

- `pnorm` returns the integral from $-\infty$ to q for a $\mathcal{N}(\mu, \sigma)$
- `pnorm` goes from *quantiles* (think *Z* scores) to probabilities

For quantiles we use *qnorm*

```
stats::qnorm(p = 0.0104, mean = 100, sd = 13)
```

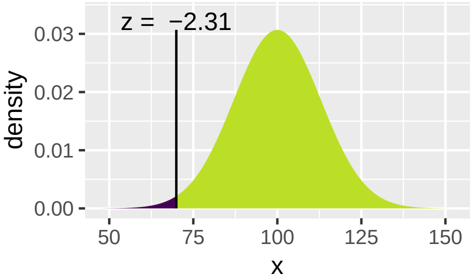
```
## [1] 69.94926
```

For quantiles we use *qnorm*

```r
stats::qnorm(p = 0.0104, mean = 100, sd = 13)
```

```
## [1] 69.94926
```

```r
mosaic::xqnorm(p = 0.0104, mean = 100, sd = 13)
```



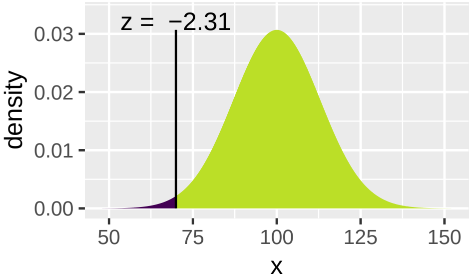

```
## [1] 69.94926
```

For quantiles we use *qnorm*

```
stats::qnorm(p = 0.0104, mean = 100, sd = 13)
```

```
## [1] 69.94926
```

```
mosaic::xqnorm(p = 0.0104, mean = 100, sd = 13)
```




```
## [1] 69.94926
```

- `qnorm` answers the question: What is the Z-score of the *p*th percentile of the normal distribution?
- `qnorm` goes from *probabilities* to quantiles

The Normal distribution

Q2: What is the probability of seeing an IQ score **as extreme as** (think highly unusual) 130?

1. Again, we find that $X = 130$ is the same percentile of the IQ Normal distribution as $Z = 2.31$ is of the standard Normal.

The Normal distribution

Q2: What is the probability of seeing an IQ score **as extreme as** (think highly unusual) 130?

1. Again, we find that $X = 130$ is the same percentile of the IQ Normal distribution as $Z = 2.31$ is of the standard Normal.
2. To see what scores are as extreme, we want to know the probability that $Z >$2.31 or that $Z <$-2.31.

The Normal distribution

Q2: What is the probability of seeing an IQ score **as extreme as** (think highly unusual) 130?

1. Again, we find that $X = 130$ is the same percentile of the IQ Normal distribution as $Z = 2.31$ is of the standard Normal.
2. To see what scores are as extreme, we want to know the probability that $Z >$2.31 or that $Z <$-2.31.
3. As we saw previously, $P(Z > 2.31) = P(Z < -2.31) = 0.0104$, so the probability of seeing an IQ as extreme or more so than 130 is $2 \times 0.0104 = 0.0208$.

Finding tail probabilities

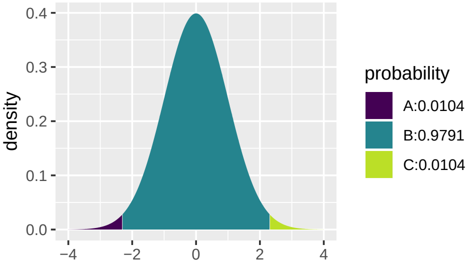
```
# lower.tail = TRUE is the default
stats::pnorm(q = -2.31, mean = 0, sd = 1, lower.tail = TRUE) +
stats::pnorm(q = 2.31, mean = 0, sd = 1, lower.tail = FALSE)

## [1] 0.02088815
```

Finding tail probabilities

```
# lower.tail = TRUE is the default
stats::pnorm(q = -2.31, mean = 0, sd = 1, lower.tail = TRUE) +
stats::pnorm(q = 2.31, mean = 0, sd = 1, lower.tail = FALSE)
```

```
## [1] 0.02088815
```

```
mosaic::xpnorm(q = c(-2.31,2.31), mean = 0, sd = 1)
```




```
## [1] 0.01044408 0.98955592
```

The Normal distribution

Q3: What is the 75^{th} percentile of the IQ scores distribution?
We now have to reverse the sequence of steps:

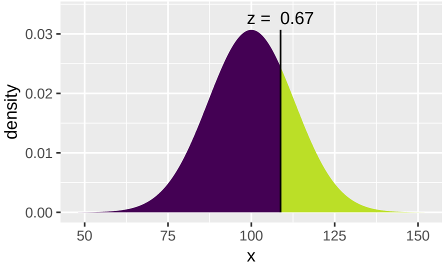
- Ask yourself: What Z value corresponds to a probability of 0.75? Should you use `pnorm` or `qnorm`?

The Normal distribution

Q3: What is the 75^{th} percentile of the IQ scores distribution?
We now have to reverse the sequence of steps:

- Ask yourself: What Z value corresponds to a probability of 0.75? Should you use `pnorm` or `qnorm`?

```
mosaic::xqnorm(p = 0.75, mean = 100, sd = 13)
```




```
## [1] 108.7684
```

This gives us that 75% of the IQ scores fall below 108.8.

Empirical Rule or 68-95-99.7% Rule

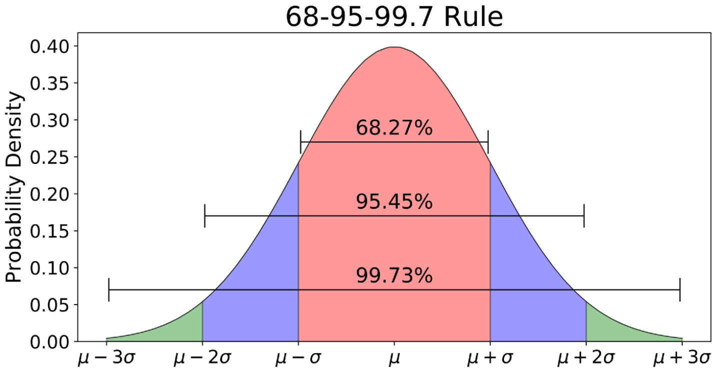
In any normal distribution with mean μ and standard deviation σ^2:

- Approximately 68% of the data fall within one standard deviation of the mean.
- Approximately 95% of the data fall within two standard deviations of the mean.
- Approximately 99.7% of the data fall within three standard deviations of the mean.

Demo of Empirical Rule

```r
pacman::p_load(mosaic)
pacman::p_load(manipulate)

mNorm <- function(mean = 0, sd = 1) {
  lo <- mean - 5 * sd
  hi <- mean + 5 * sd
  manipulate(
        xpnorm(c(A,B), mean, sd, verbose = FALSE, invisible = TRUE),
        A = slider(lo, hi, initial = mean - sd),
        B = slider(lo, hi, initial = mean + sd)
 )
}
mNorm(mean = 0, sd = 1)
```

Empirical Rule or 68-95-99.7% Rule



Properties of Normal random variables

Special properties of the Normal distribution:

- In X is a Normal random variables, then so is $a + bX$.

Properties of Normal random variables

Special properties of the Normal distribution:

- In X is a Normal random variables, then so is $a + bX$.
- If X and Y are two Normal random variables, then $X + Y$ is a Normal random variable. What is the mean and variance of this new random variable?

Properties of Normal random variables

Special properties of the Normal distribution:

- In X is a Normal random variables, then so is $a + bX$.
- If X and Y are two Normal random variables, then $X + Y$ is a Normal random variable. What is the mean and variance of this new random variable?
- If X and Y are two Normal random variables and $\rho_{XY} = 0$, then X and Y are independent.

Properties of Normal random variables

Example: Let $Y_1, ..., Y_n \sim \mathcal{N}(\mu, \sigma)$, and let each Y_i be independent of the others.

Then $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ has what distribution?

Properties of Normal random variables

Example: Let $Y_1, ..., Y_n \sim \mathcal{N}(\mu, \sigma)$, and let each Y_i be independent of the others.

Then $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ has what distribution?

- The sum of Normal random variables is Normal, so $\overline{Y}$ is a Normal random variable.

Properties of Normal random variables

Example: Let $Y_1, ..., Y_n \sim \mathcal{N}(\mu, \sigma)$, and let each Y_i be independent of the others.

Then $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ has what distribution?

- The sum of Normal random variables is Normal, so $\overline{Y}$ is a Normal random variable.
- $E(\overline{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$
- $Var(\overline{Y}) = Var(\frac{1}{n}\sum_{i=1}^{n} Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \sigma^2/n.$
- Standard Error of $\overline{Y}$ = $\sqrt{Var(\overline{Y})} = \sigma/\sqrt{n}$

Central Limit Theorem

Properties of the sample mean: The Central Limit Theorem (CLT)

The sampling distribution of $\overline{X}$ is Normal if X is Normal. What probability distribution does the sample mean follow if X is not Normal?

Properties of the sample mean: The Central Limit Theorem (CLT)

The sampling distribution of $\overline{X}$ is Normal if X is Normal. What probability distribution does the sample mean follow if X is not Normal?

As sample size increases, the distribution of $\overline{X}$ becomes closer to a Normal distribution, no matter what the distribution of sampled variable X!

(This is true as long as the distribution has a finite variance.)

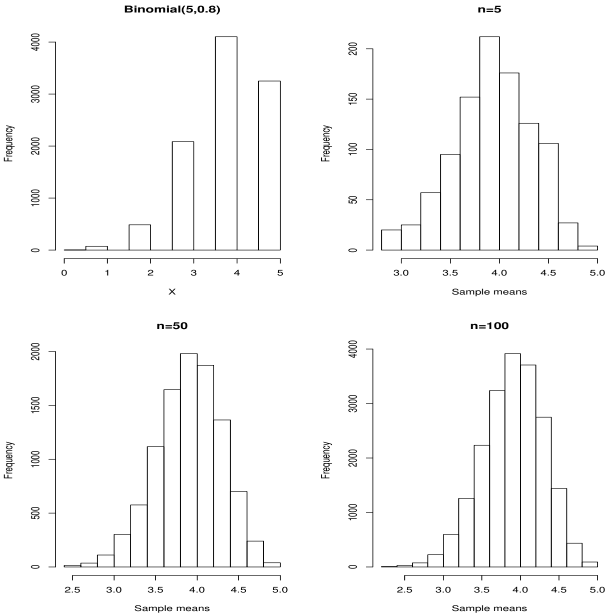
The Central Limit Theorem (CLT)

The Central Limit Theorem (CLT) tells us that, no matter the distribution of X, if σ^2 is finite,

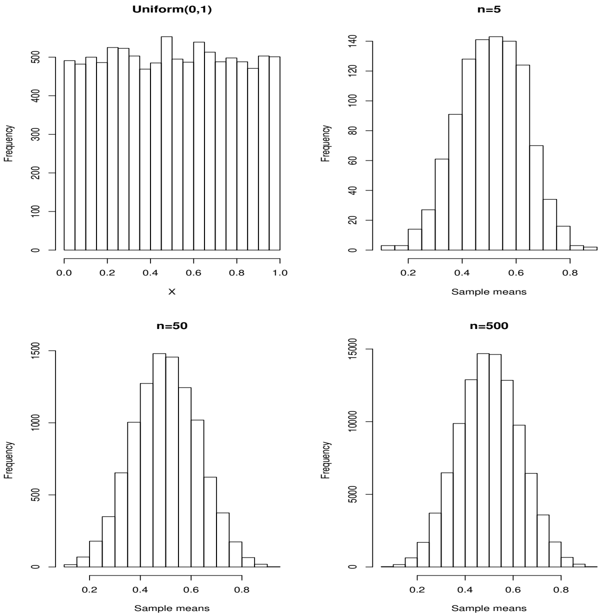
$$\bar{X} \sim \mathcal{N}(\mu, \sigma/\sqrt{n}).$$

Recall: When we are talking about the variability of a **statistic**, we use the term **standard error** (not standard deviation). The standard error of the sample mean is $\sigma/\sqrt{n}$.

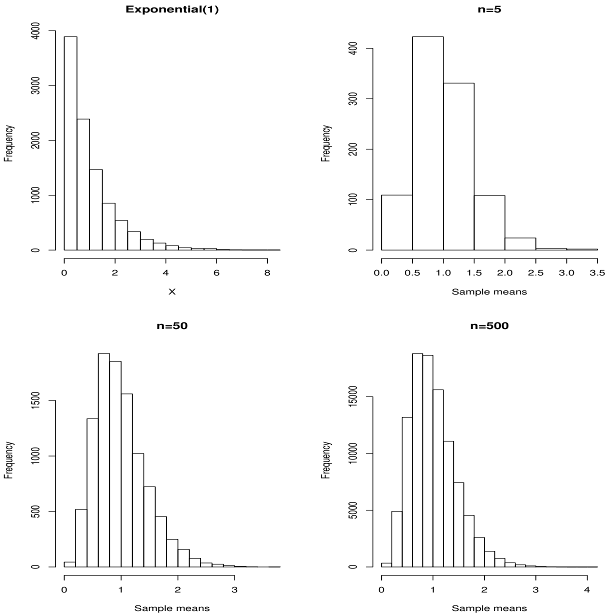
The Central Limit Theorem (CLT)



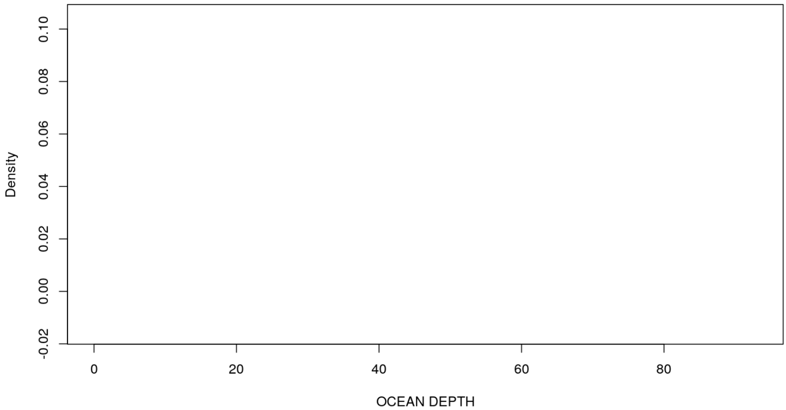
The Central Limit Theorem (CLT)



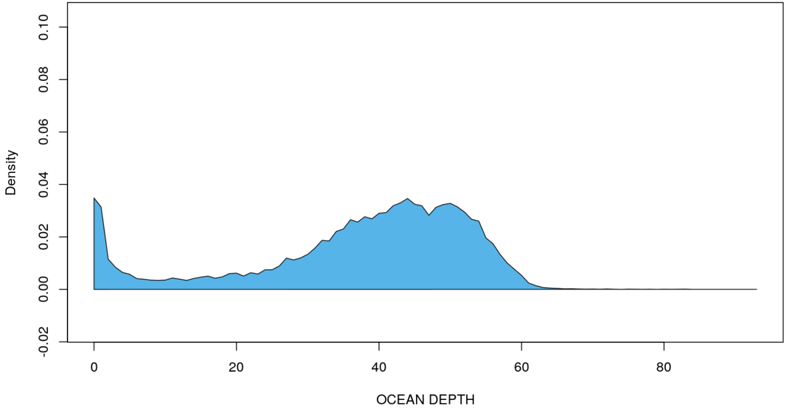
The Central Limit Theorem (CLT)



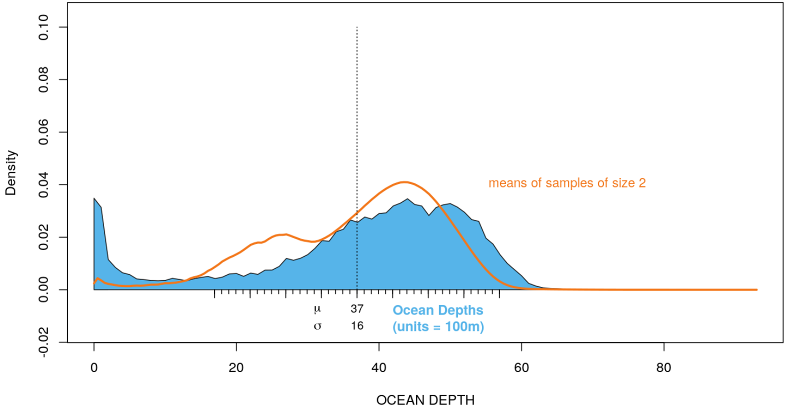
CLT in action: Depths of the ocean



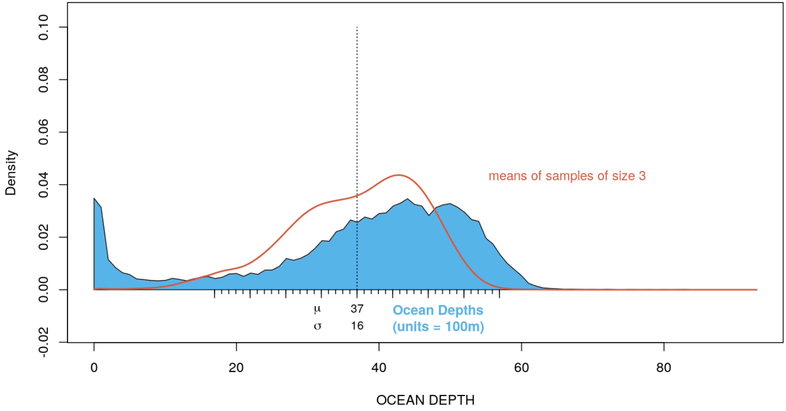
CLT in action: Depths of the ocean



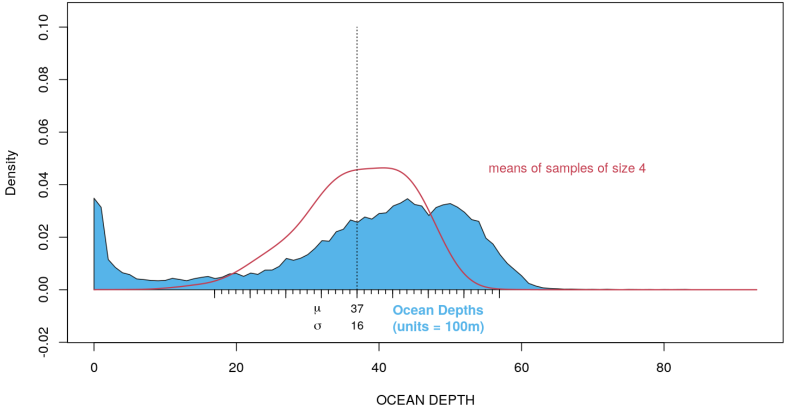
CLT in action: Depths of the ocean



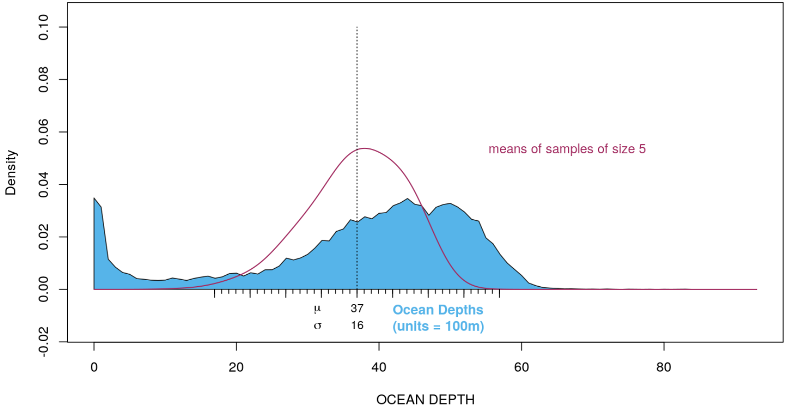
CLT in action: Depths of the ocean



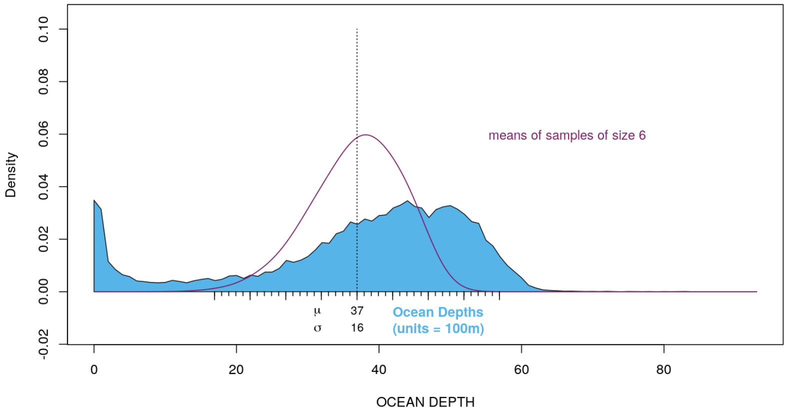
CLT in action: Depths of the ocean



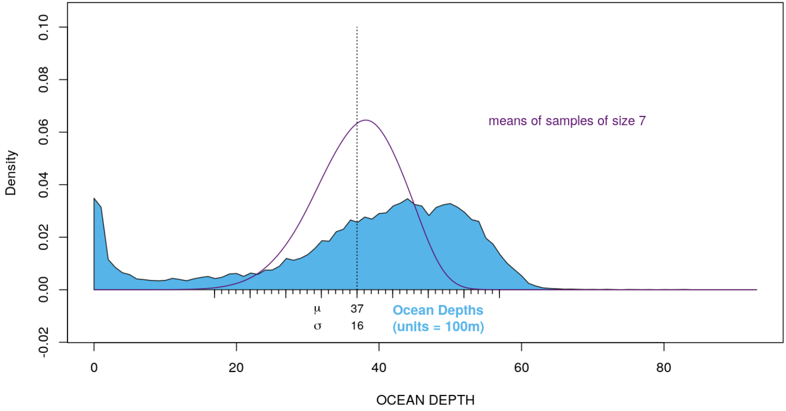
CLT in action: Depths of the ocean



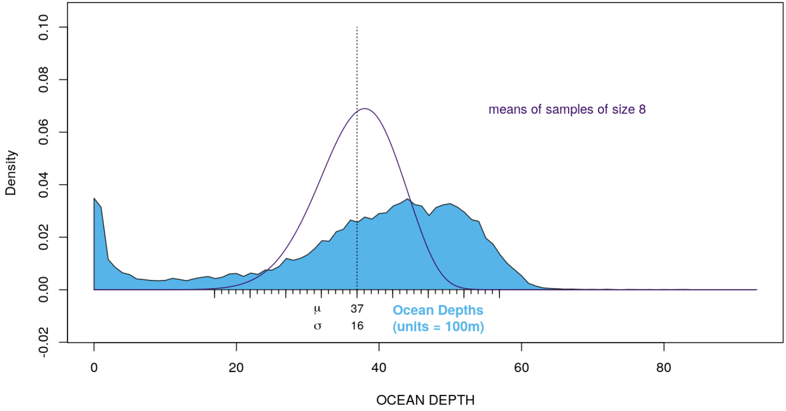
CLT in action: Depths of the ocean



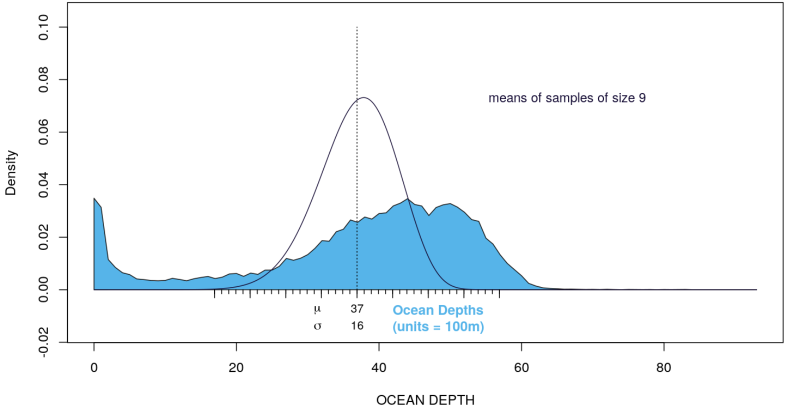
CLT in action: Depths of the ocean



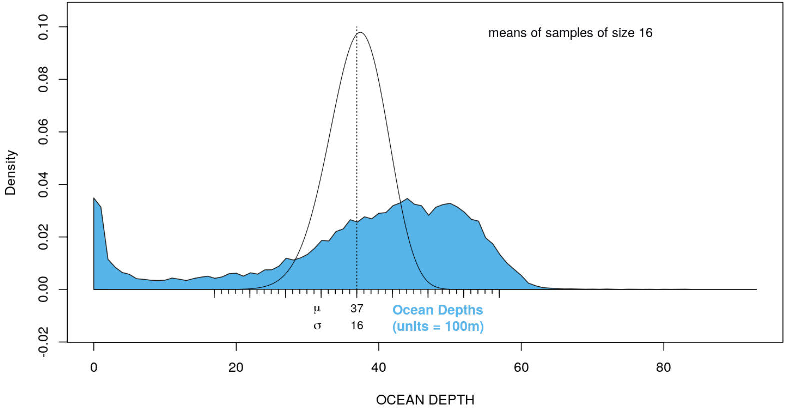
CLT in action: Depths of the ocean



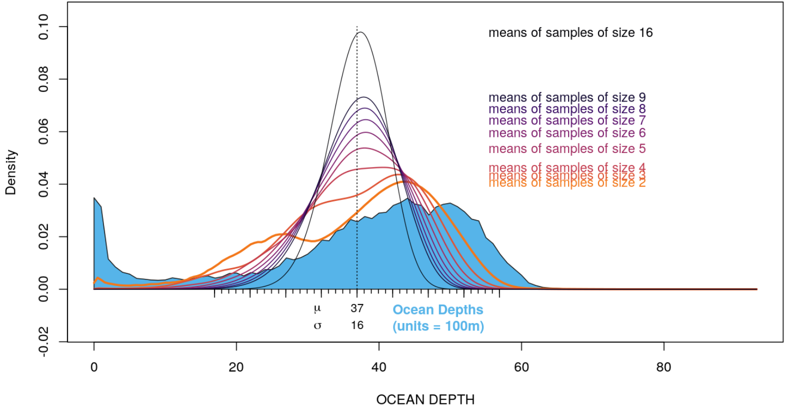
CLT in action: Depths of the ocean



CLT in action: Depths of the ocean



CLT in action: Depths of the ocean



Quadruple the work, half the benefit

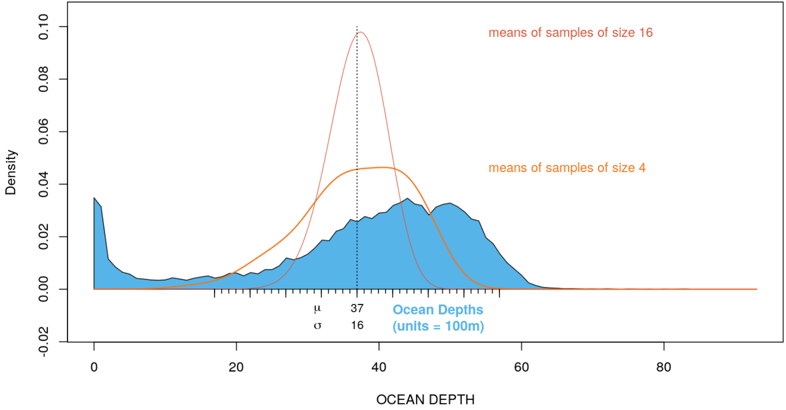


Fig.: When the sample size increases from 4 to 16, the spread of the sampling distribution for the mean is reduced by a half, i.e., the range is cut in half. This is known as the curse of the $\sqrt{n}$