

Mathematizing the biological world

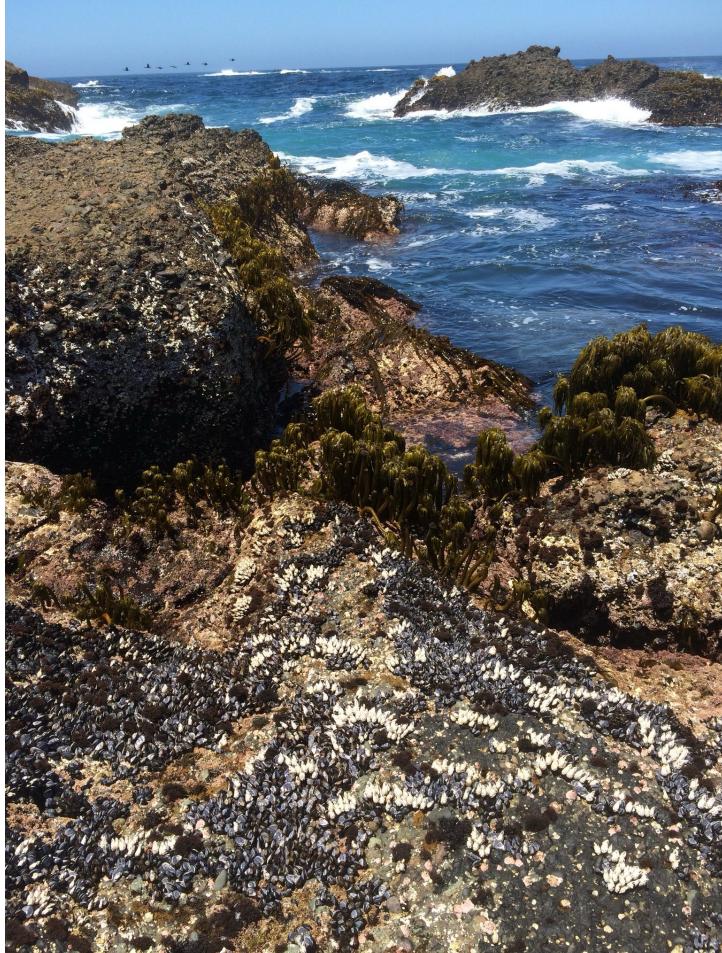


Quick VACC check

- Let's check connectivity



THE UNIVERSITY OF VERMONT
ADVANCED COMPUTING CENTER



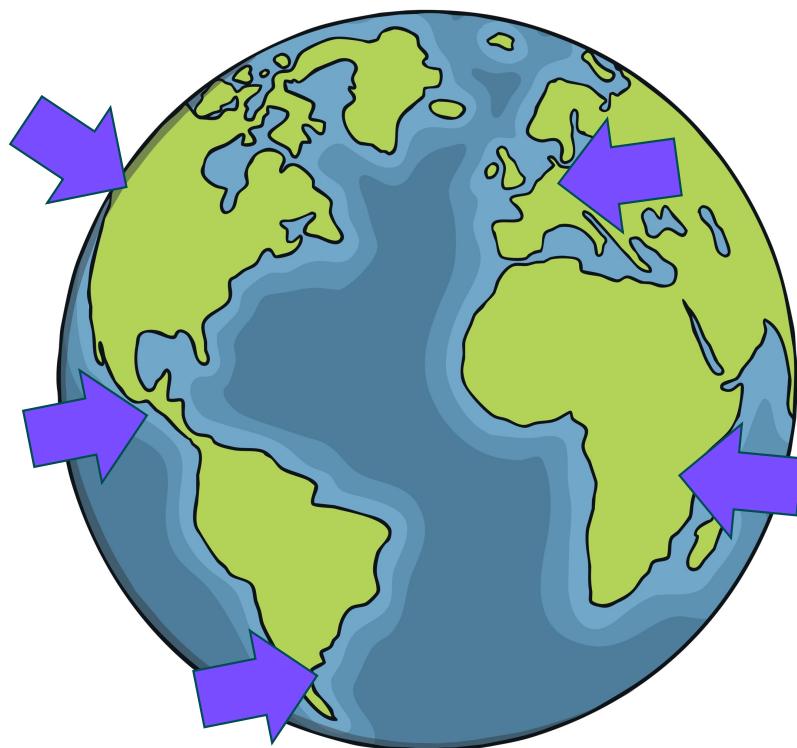
What exactly does it mean when new say

The critters from the upper intertidal zone have smaller shells than those in the lower intertidal zone (*t*-test; $P < 0.05$).

We make mathematical assumptions about the world

- We assume that "shell size" in the intertidal are generated by a process that can be generalized and mathematized in a formal way
 - Laws of physics, chemistry, biology

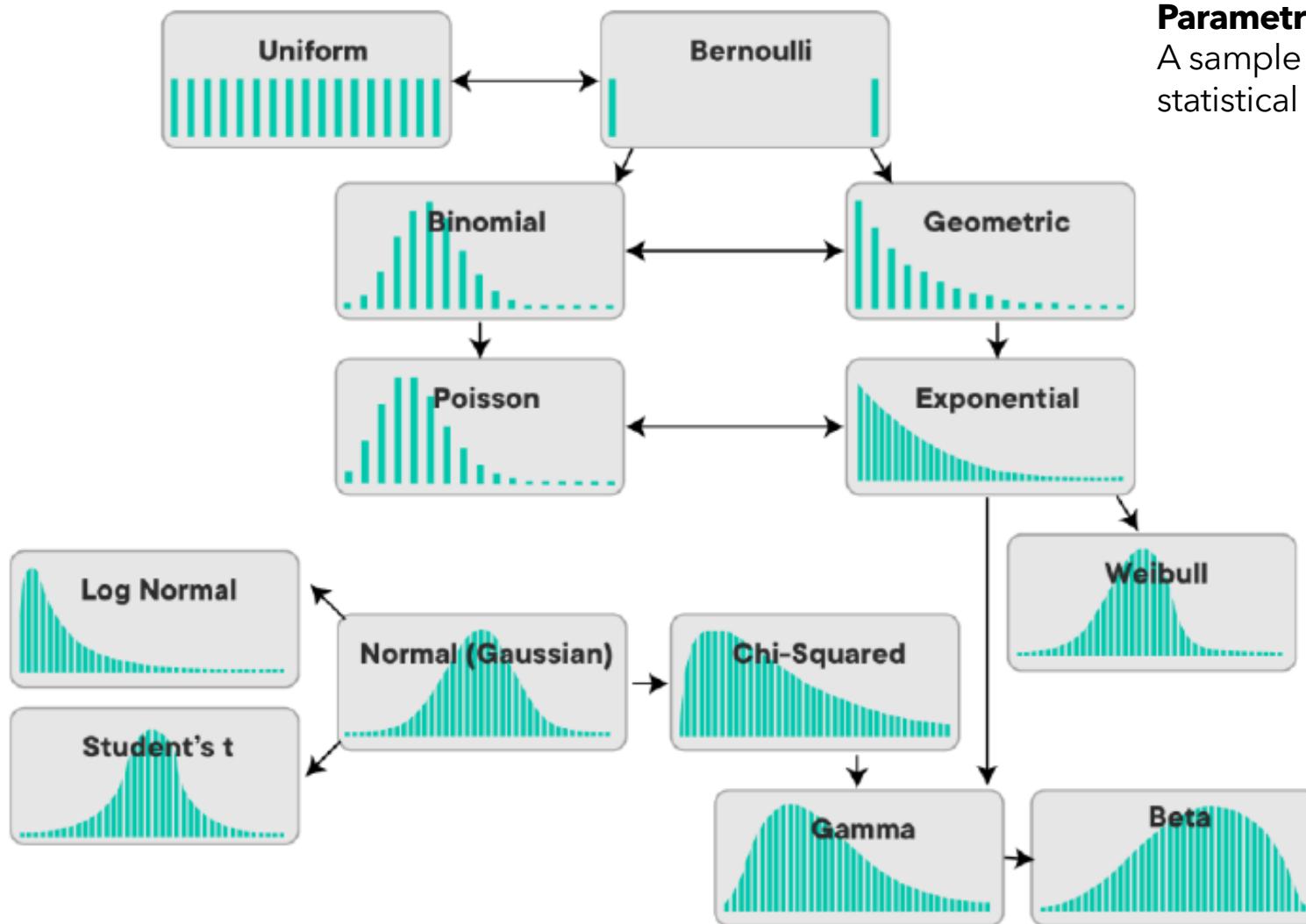
Generality of the laws of nature



Will the process of
DNA replication be
the same?

We make mathematical assumptions about the world

- We assume that "shell size" in the intertidal are generated by a process that can be generalized and mathematized in a formal way
 - Laws of physics, chemistry, biology
- We assume that "shell size" is a trait that statistical models can describe. For example, a statistical distribution (gaussian [normal], binomial, beta, chi - squared, ...)



Parametric Framework

A sample of commonly used statistical distributions

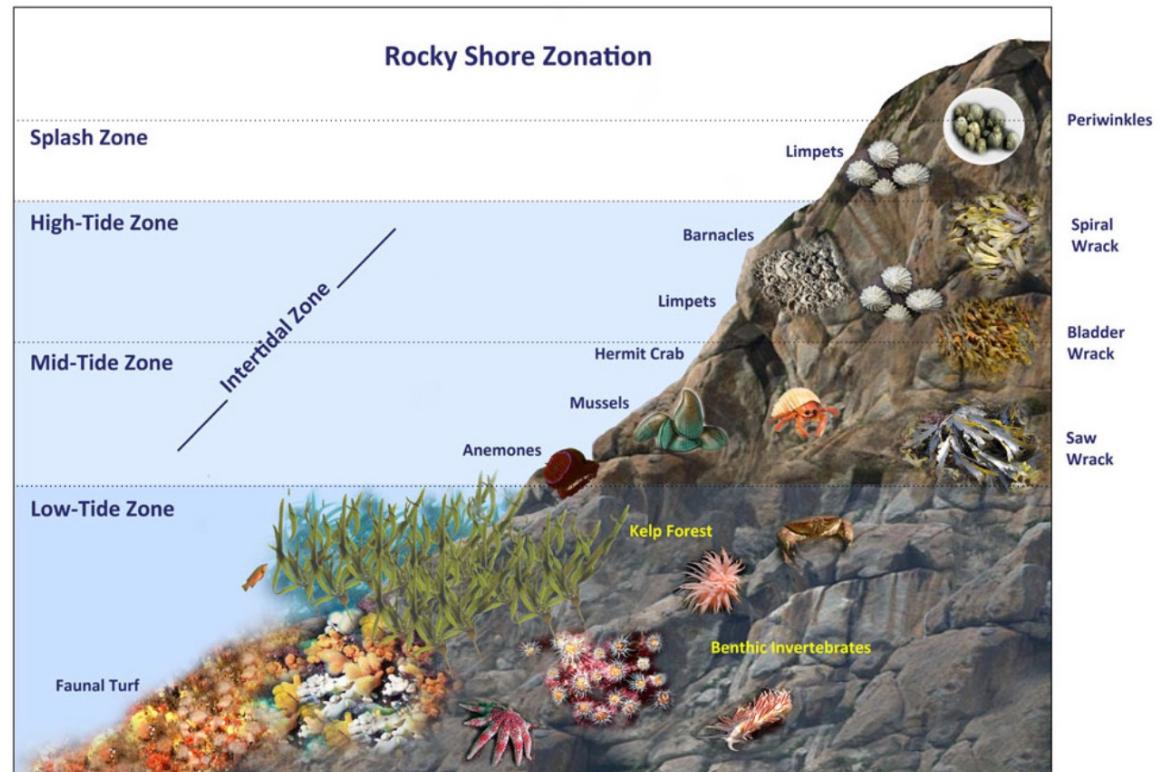
We make mathematical assumptions about the world

- We assume that "shell size" in the intertidal are generated by a process that can be generalized and mathematized in a formal way
 - Laws of physics, chemistry, biology
- We assume that "shell size" is a trait that statistical models can describe. For example, a statistical distribution (gaussian [normal], binomial, beta, chi - squared, ...)
- We assume that different physics, chemistry, and biology parameters, acting at different spaces across time and space (high/low intertidal) can modify the statistical properties of the "snail shells distribution."

What differs between high and low tide?

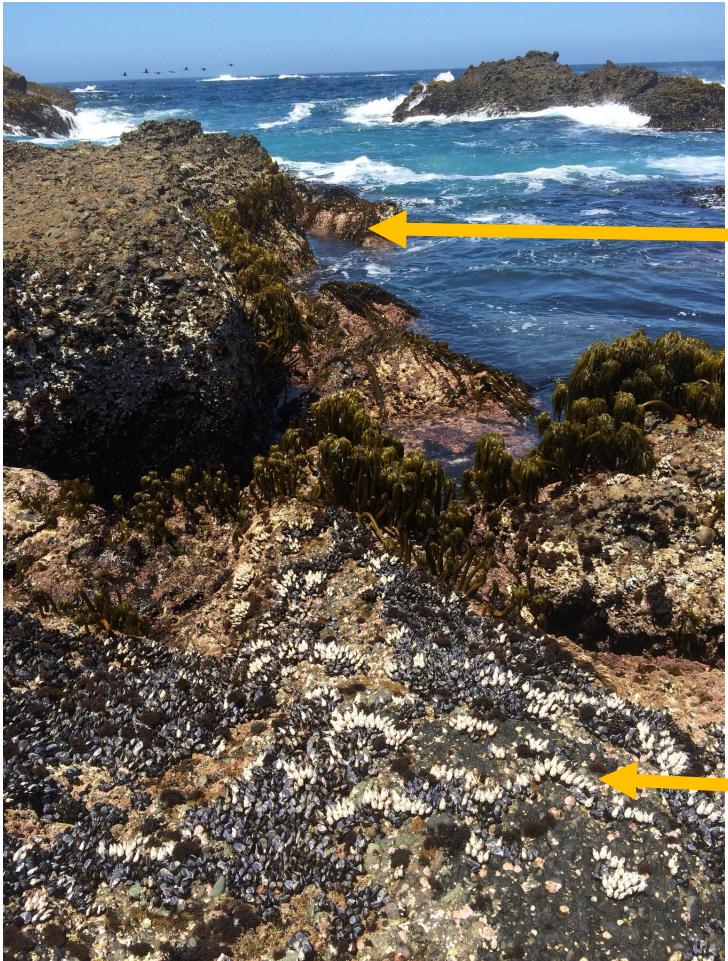
High abiotic stress
Low Predation

Low abiotic stress
High Predation



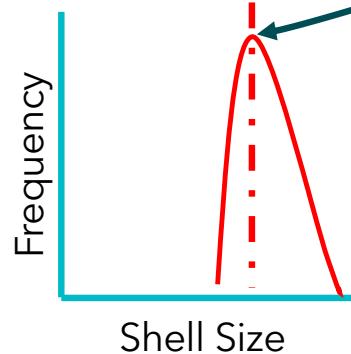
We make mathematical assumptions about the world

- We assume that "shell size" in the intertidal are generated by a process that can be generalized and mathematized in a formal way
 - Laws of physics, chemistry, biology
- We assume that "shell size" is a trait that statistical models can describe. For example, a statistical distribution (gaussian [normal], binomial, beta, chi - squared, ...)
- We assume that different physics, chemistry, and biology parameters, acting at different spaces across time and space (high/low intertidal) can modify the statistical properties of the "snail shells distribution."
- Thus, our statistical test asks whether samples from the upper and lower intertidal zones come from statistical distributions with different properties (e.g., mean).



Model assumption

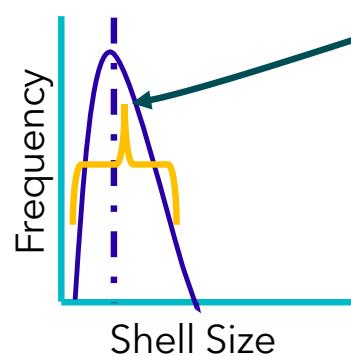
If all individuals could be measured



The **mean** of the distribution is a measure of "centrality."
And it is often the expected value of a sampling effort

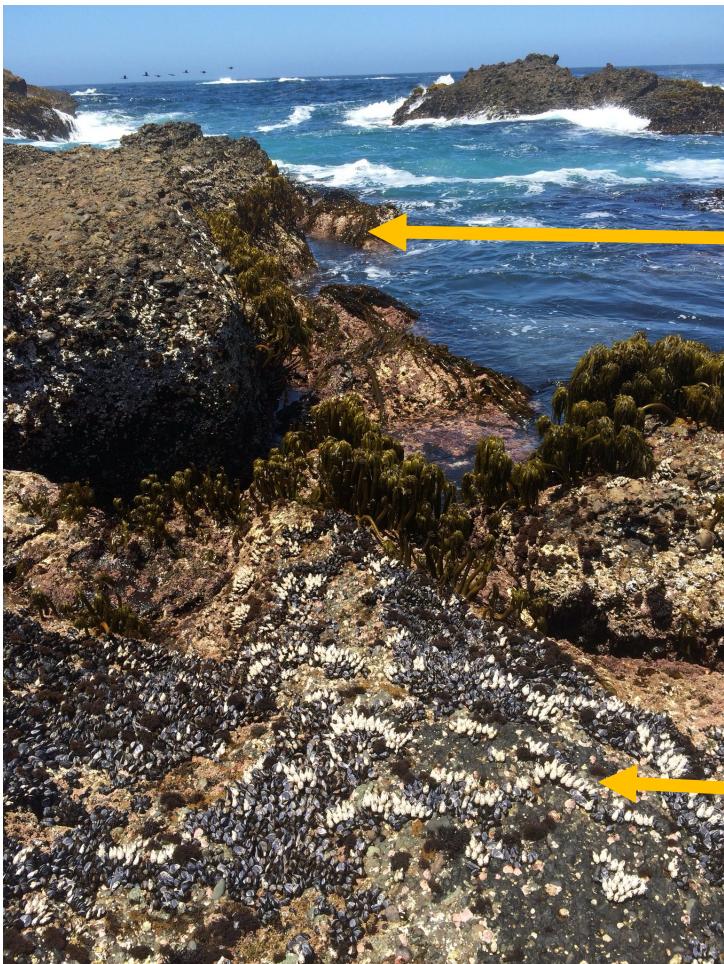
This means that if I were to sample just one critter, the "expected" shell size should be around the mean of the distribution.

If all individuals could be measured

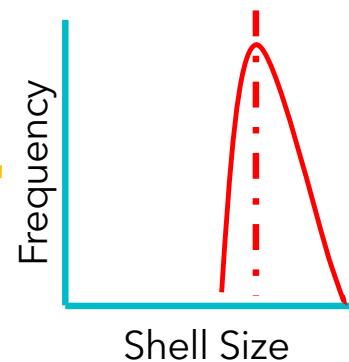


The **variance** of the distribution is a measure of "dispersion." *It is often the expected difference between the samples and the mean*

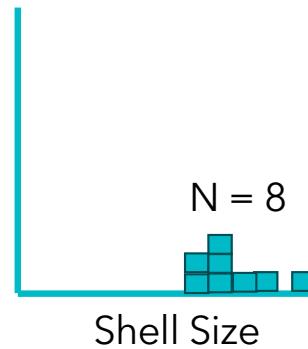
Models with low variance offer more predictability than those with high variance



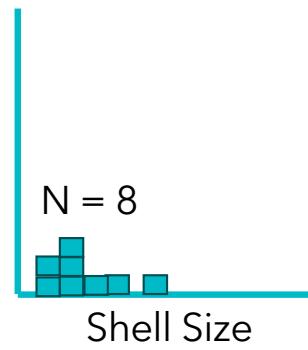
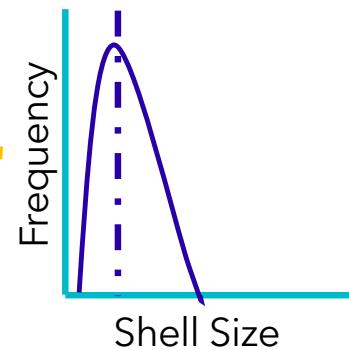
Model assumption

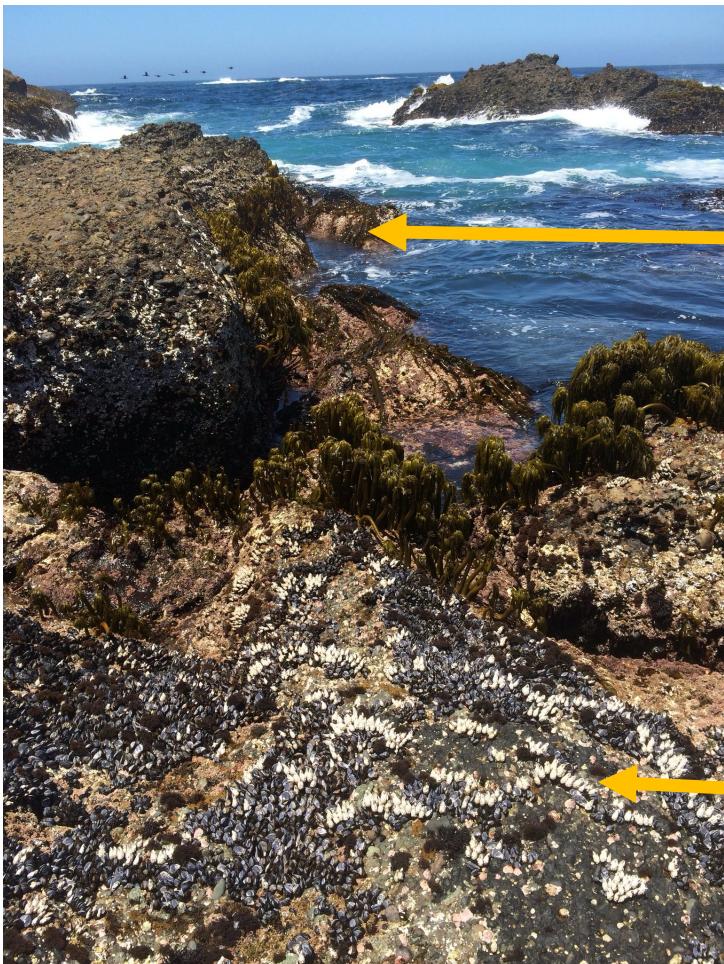


The process of sampling



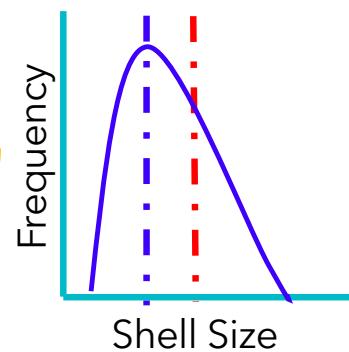
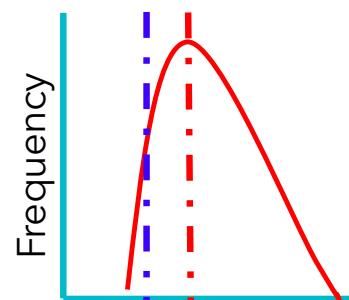
**Can all the individuals
really be sampled?**





Model assumption

If all individuals could be measured



The process of sampling

$N = 8$

Shell Size

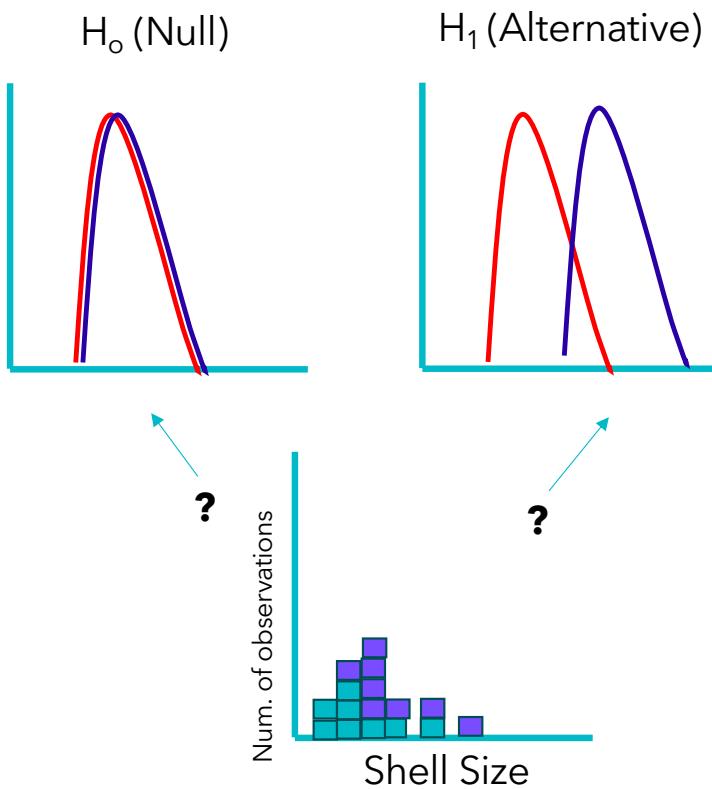
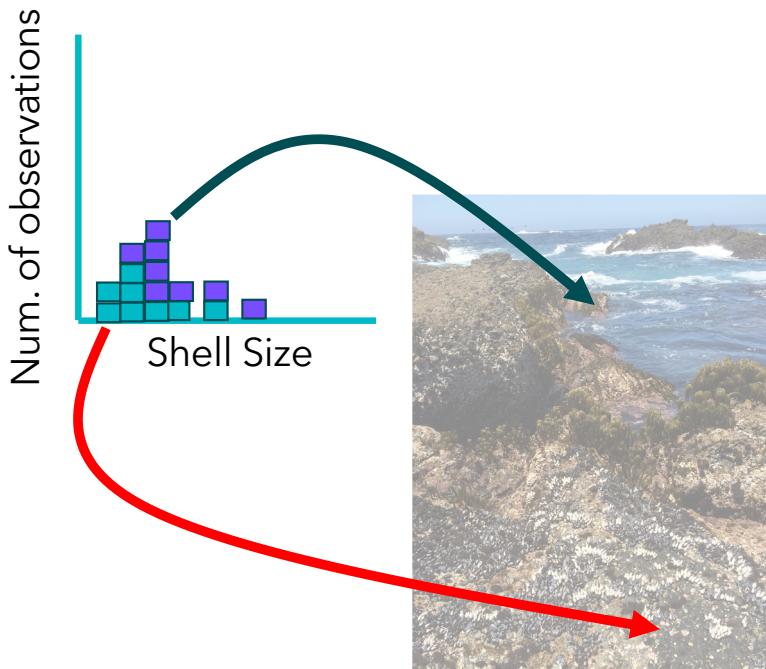
$N = 8$

Shell Size

Can we really say that these distributions are different, based on our sampling?

Two key concepts sample size and statistical power

- Based on our assumptions we seek to test the following



How many samples do we need? How much power do we have?

The chance that our test will tell us that the samples are different when they are not

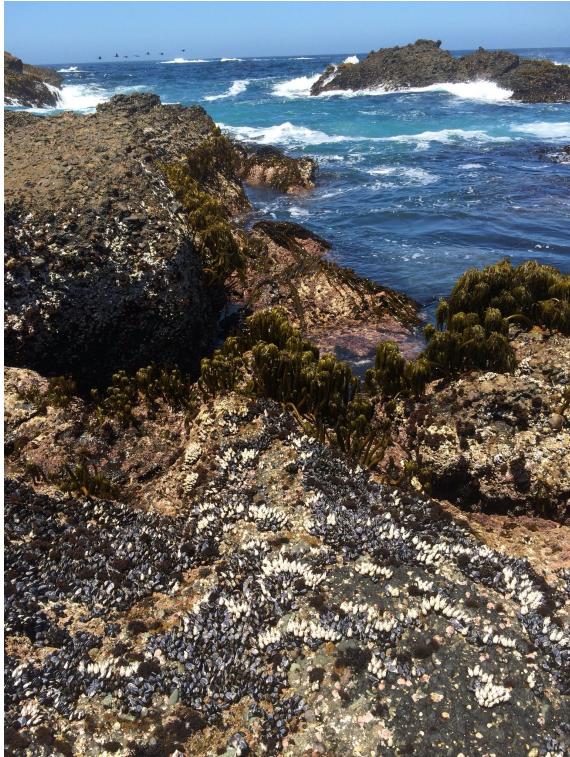
Probability of false positive	Probability of true positive
α	$1 - \beta$
$1 - \alpha$	β
Probability of true negative	Probability of false negative

We call this
power

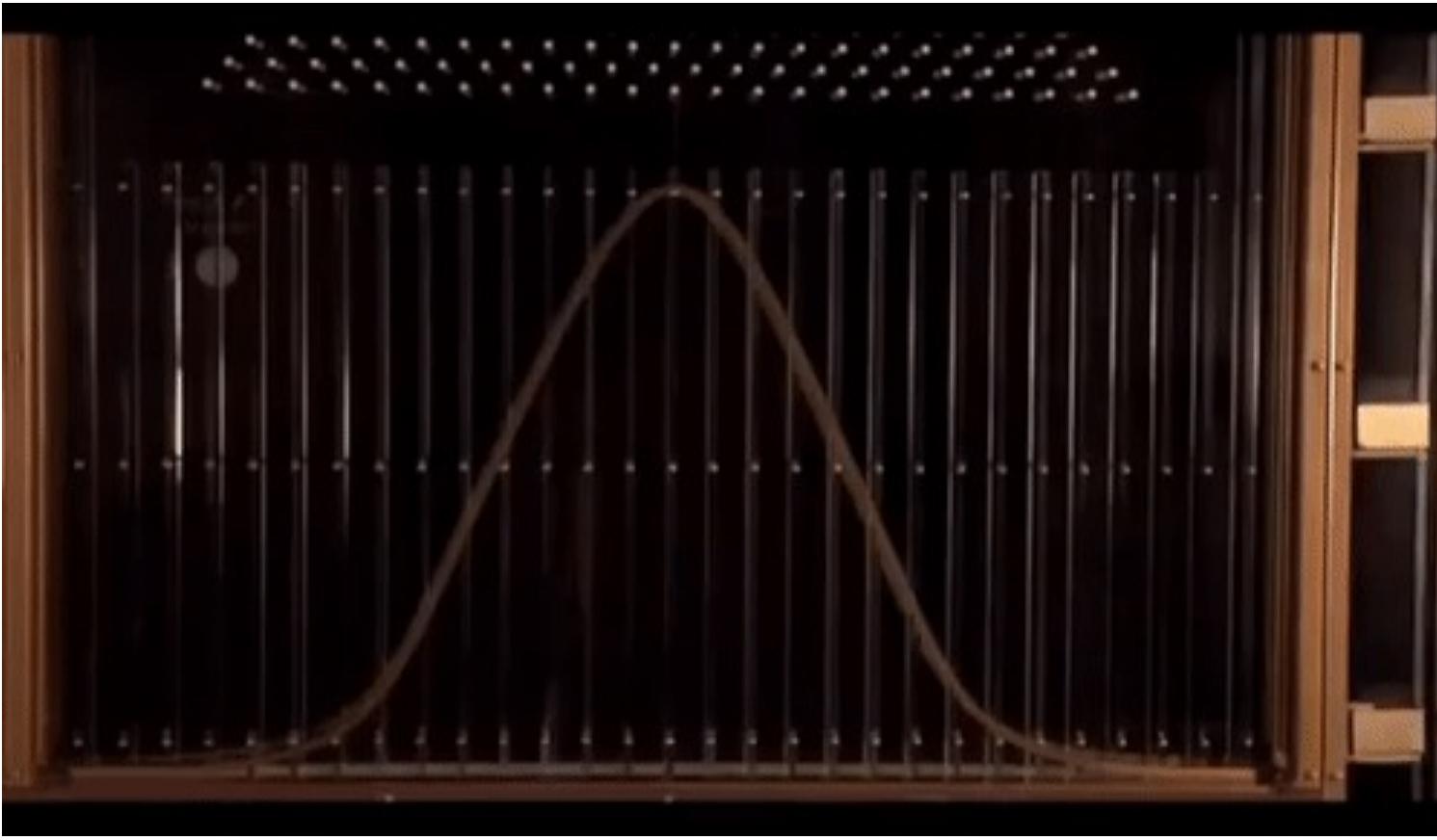
The chance that our test will tell us that the samples are NOT different when they are

**We want experimental design
and statistical tests that
maximize power!**

Simulations allow us to assess the power of our experimental design



We will assume that the trait (shell size) is
Gaussian distributed
(i.e., normally distributed*)



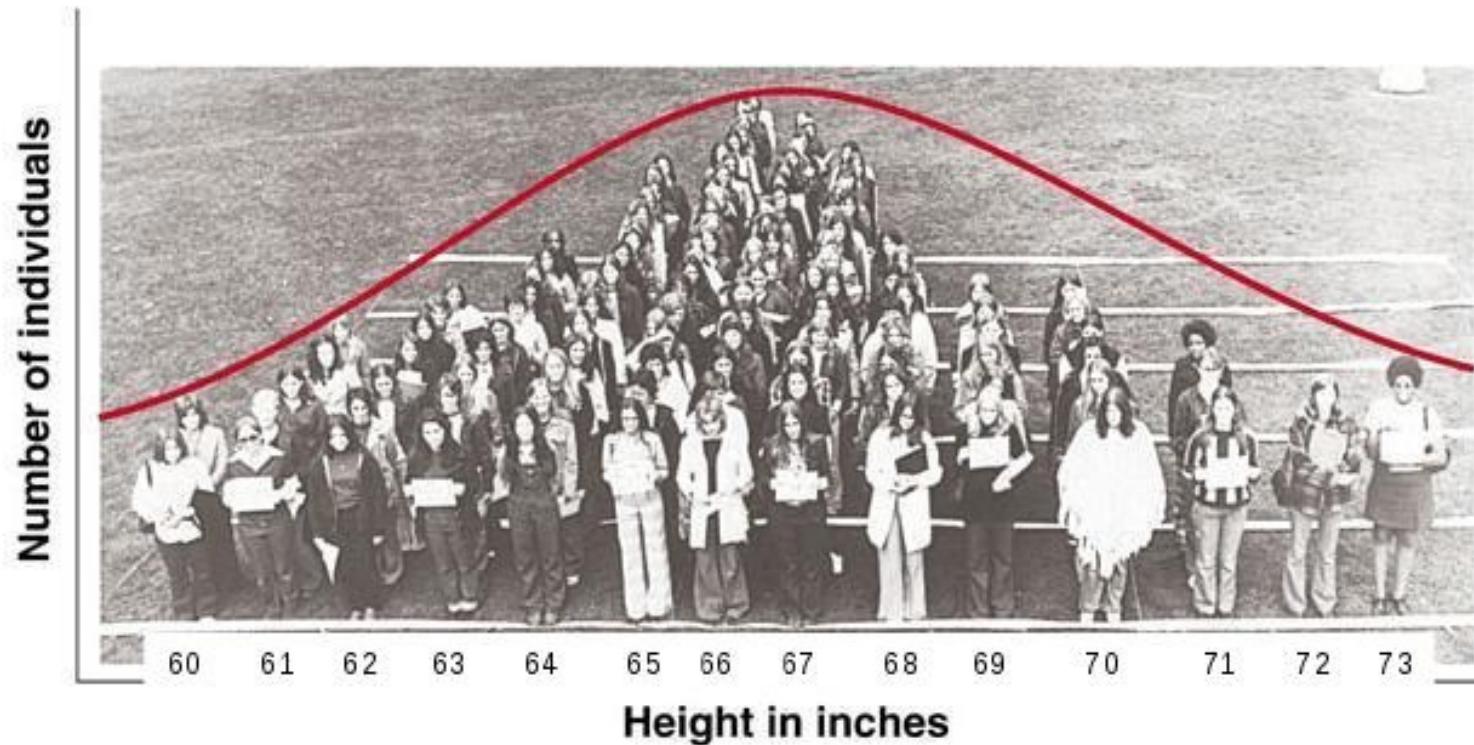
The Gaussian distribution naturally creates a “bell curve” where the mean and variance are the key parameters. **When the mean is 0 and the standard deviation is 1, we call it “Normal”**

Plotting Normal distribution in R

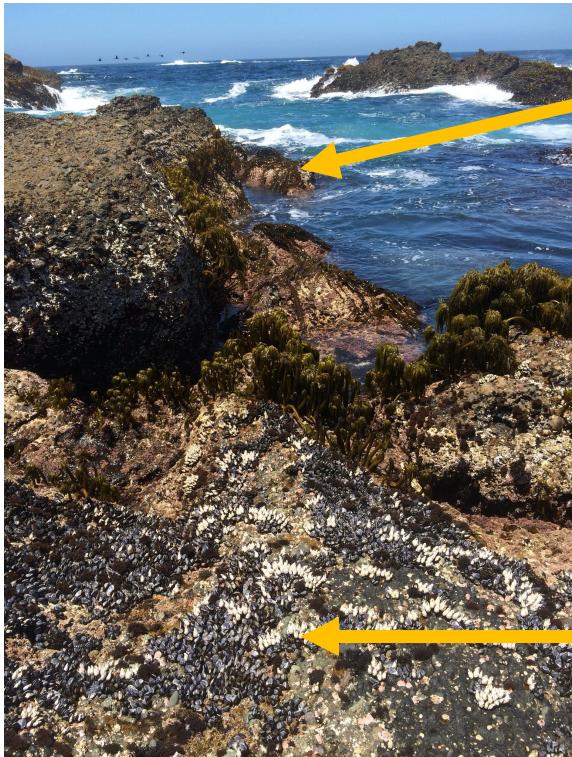
- Let's learn about the `rnorm()` function

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

Many things are normally distributed



Simulations allow us to assess the power of our experimental design



Lower ~ $N(1.2, 2.5)$

How many samples
do we need to detect
a true difference?

Higher ~ $N(1.0, 2.3)$

Day 2 Challenge

**What power do we get for a
“set” sample size?**