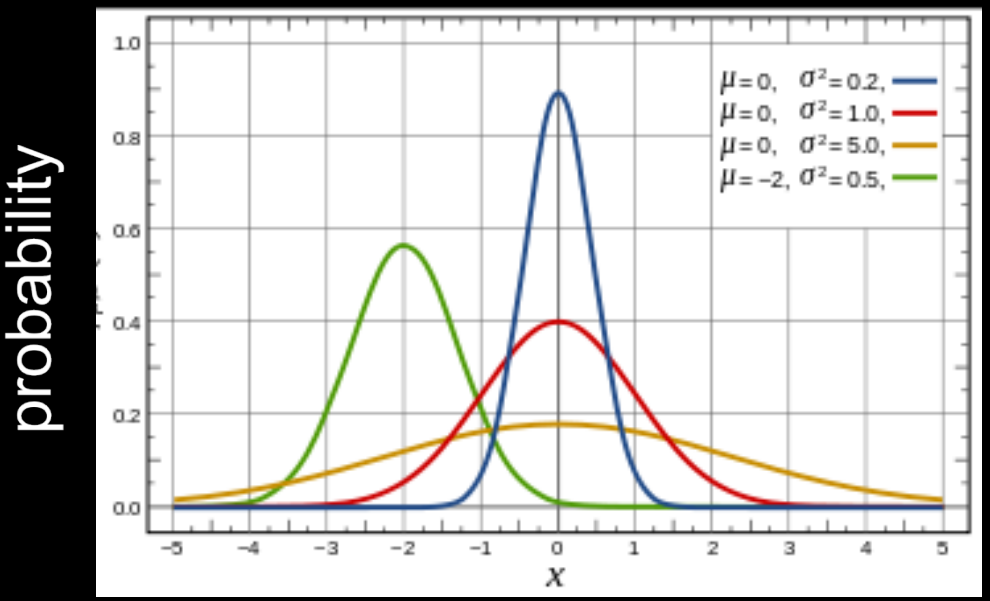
Distributions

Biological Statistics Course

Week 3

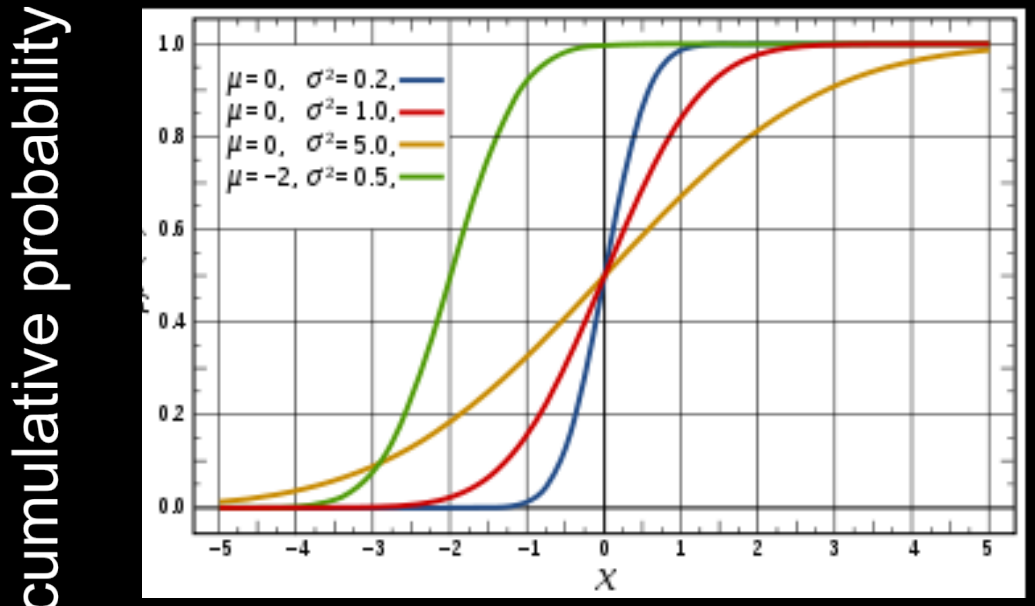
distribution functions

probability density function: the probability of getting that result



distribution functions

cumulative distribution function: the probability of getting that result or fewer

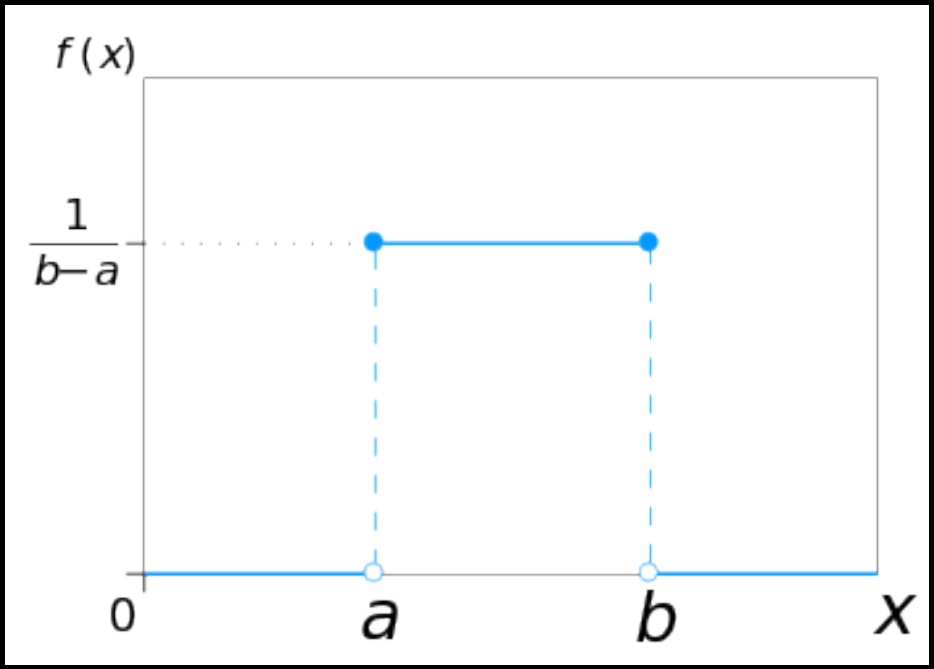


let’s look at some
common distributions

discrete probability distributions

uniform distribution

all values equally probable



binomial distribution

describes the outcome of n independent trials in an experiment

each trial has only two outcomes, either success or failure

the probability of a successful trial is p (between 0 and 1)

binomial distribution: real world examples

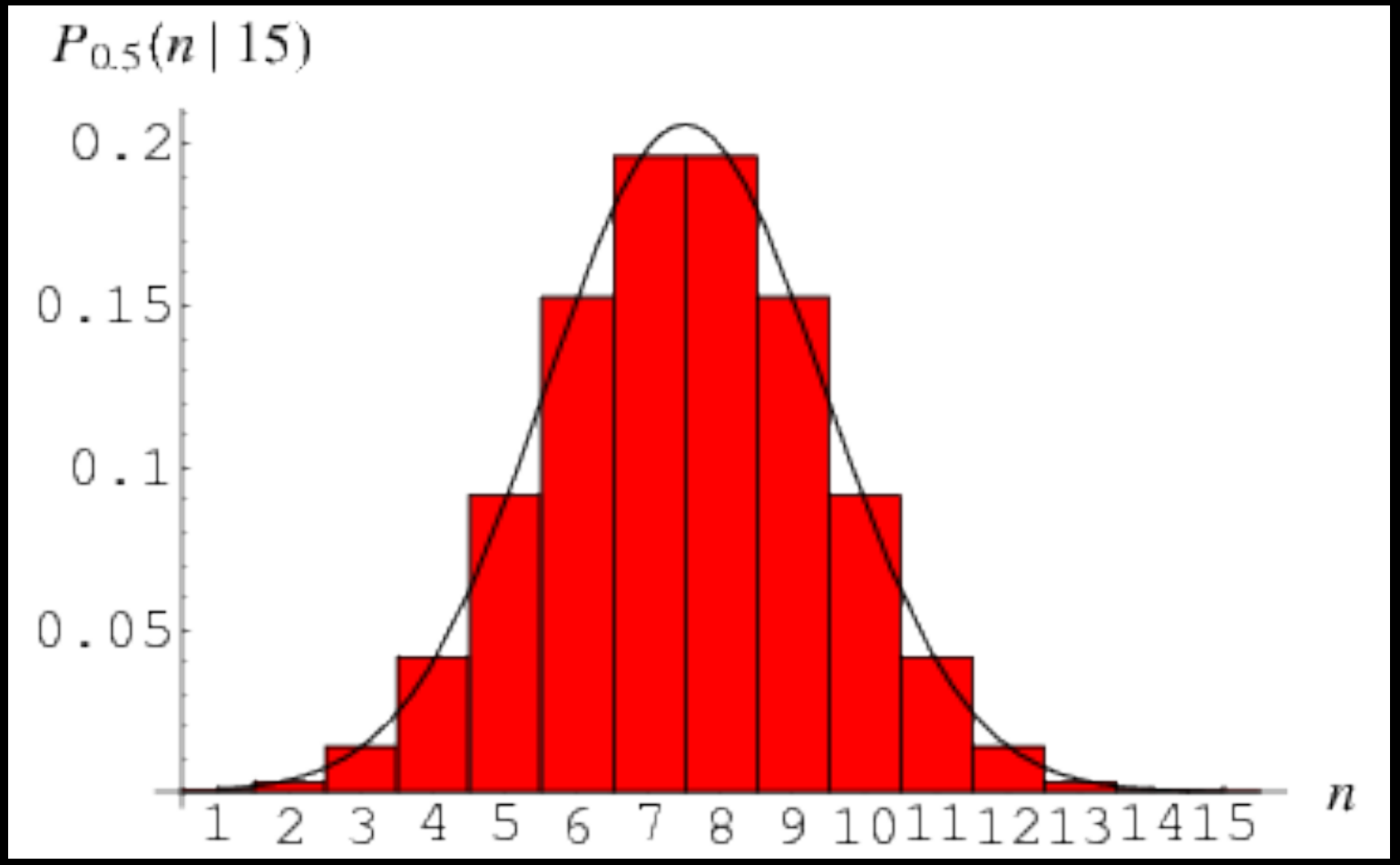
trials involving flipping a coin

drug trial success

winning the lottery

anything you can think of that has a "success" and a "failure"

binomial distribution



probability

value

distributions related to the binomial distribution

geometric distribution: number of failures before the first success

negative binomial: number of failures before the r^{th} success

multinomial: multivariate version of the binomial

Poisson distribution

used to model the counts of an event occurring randomly in space or time

λ is the mean number of events per time interval (rate)

Poisson distribution: real world examples

number of typing errors on a page

number of accidents occurring at an intersection in a given time

number of malfunctioning products made by manufacturing process

Poisson distribution: bioinformatics examples

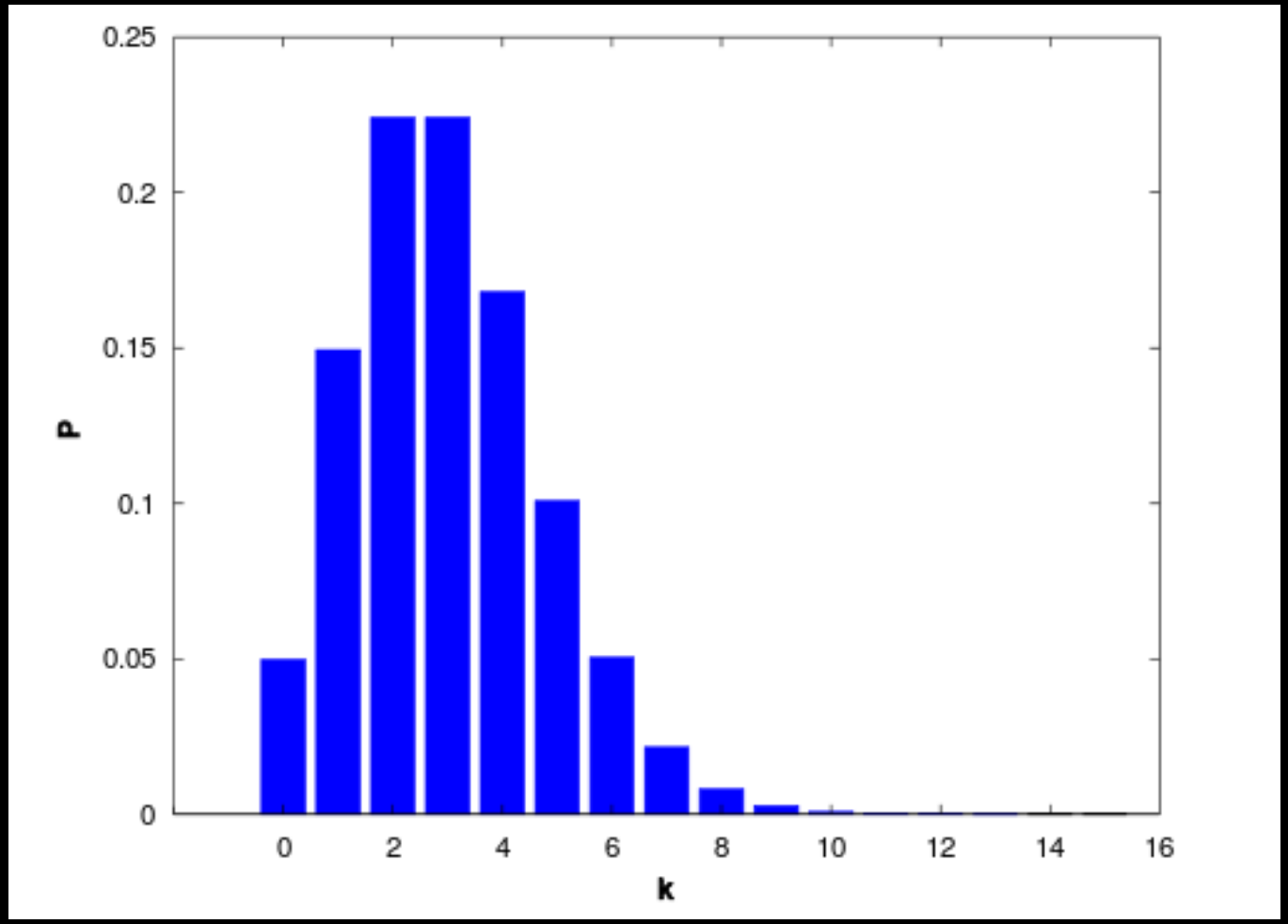
model instances of mutation or recombination in a genetic sequence

distribution of errors produced in sequencing process

probability of random sequence matches

counting occurrences of rare DNA patterns

Poisson distribution



continuous probability functions

normal distribution

typical “bell curve” distribution

also known as Gaussian distribution

μ is the mean, σ is the standard deviation

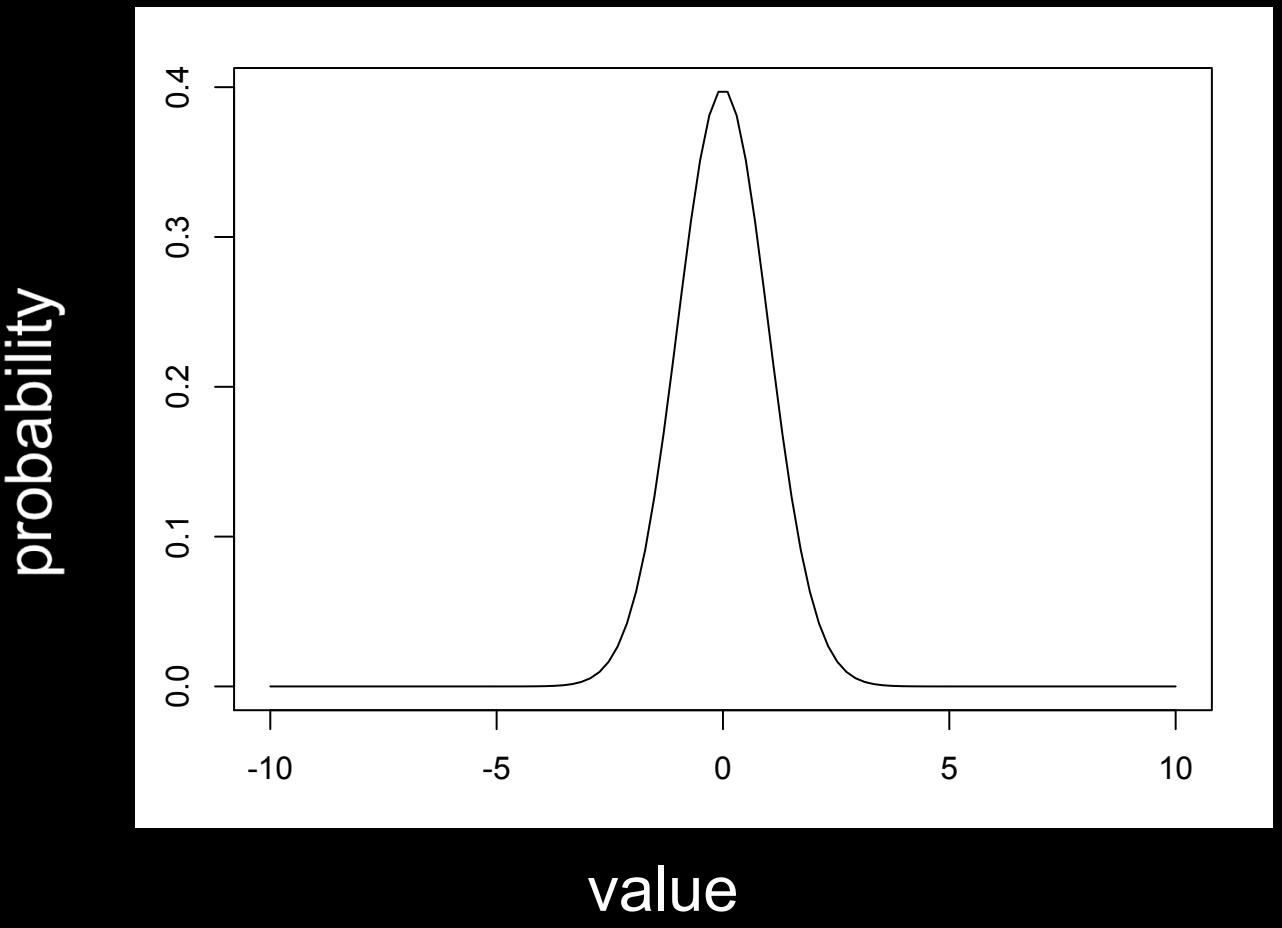
normal distribution: real world examples

height

weight

test scores

normal distribution



normal distribution: sampling

used to model data that is sampled

very important to inferential statistics

The Central Limit Theorem

standard normal distribution

mean is 0, standard deviation is 1

easy to convert any normal into a standard normal

known as Z distribution

$Z=(X-\mu)/\sigma$

but sometimes nature doesn't cooperate

sometimes your data won't be normally distributed

we'll look at two important nonnormal families of distributions, gamma and beta

Gamma Family

consists of a few related distributions

- gamma distribution
- exponential distribution
- Chi-square distribution

gamma distribution

base of the gamma family of distributions

only takes on positive real numbers

α is the shape parameter, β is the scale parameter

gamma distribution: real world examples

measurements of time until failure

concentrations of pollutants

gamma distribution

very flexible

useful for non-normally distributed data

mean=$\alpha*\beta$
variance=$\alpha*\beta^2$

exponential distribution

special case of the gamma distribution with $\alpha=1$

useful for modeling survival times

Chi-squared distribution

another gamma distribution variant

uses $\beta=2$ and $\alpha=k/2$

k=degrees of freedom

Chi-squared distribution

analyze count data in genetics

testing for independence

Beta Family

also have alpha and beta parameters

beta function is ratio of gamma functions

beta family

beta function used to model data measured as proportions

used in Bayesian statistics

interpretation of alpha and beta not as clear as in gamma, but called “shape1” and “shape 2”

like gamma, alpha and beta have a relationship with the mean and variance

mean $=\alpha/(\alpha+\beta)$

variance$=\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1)$

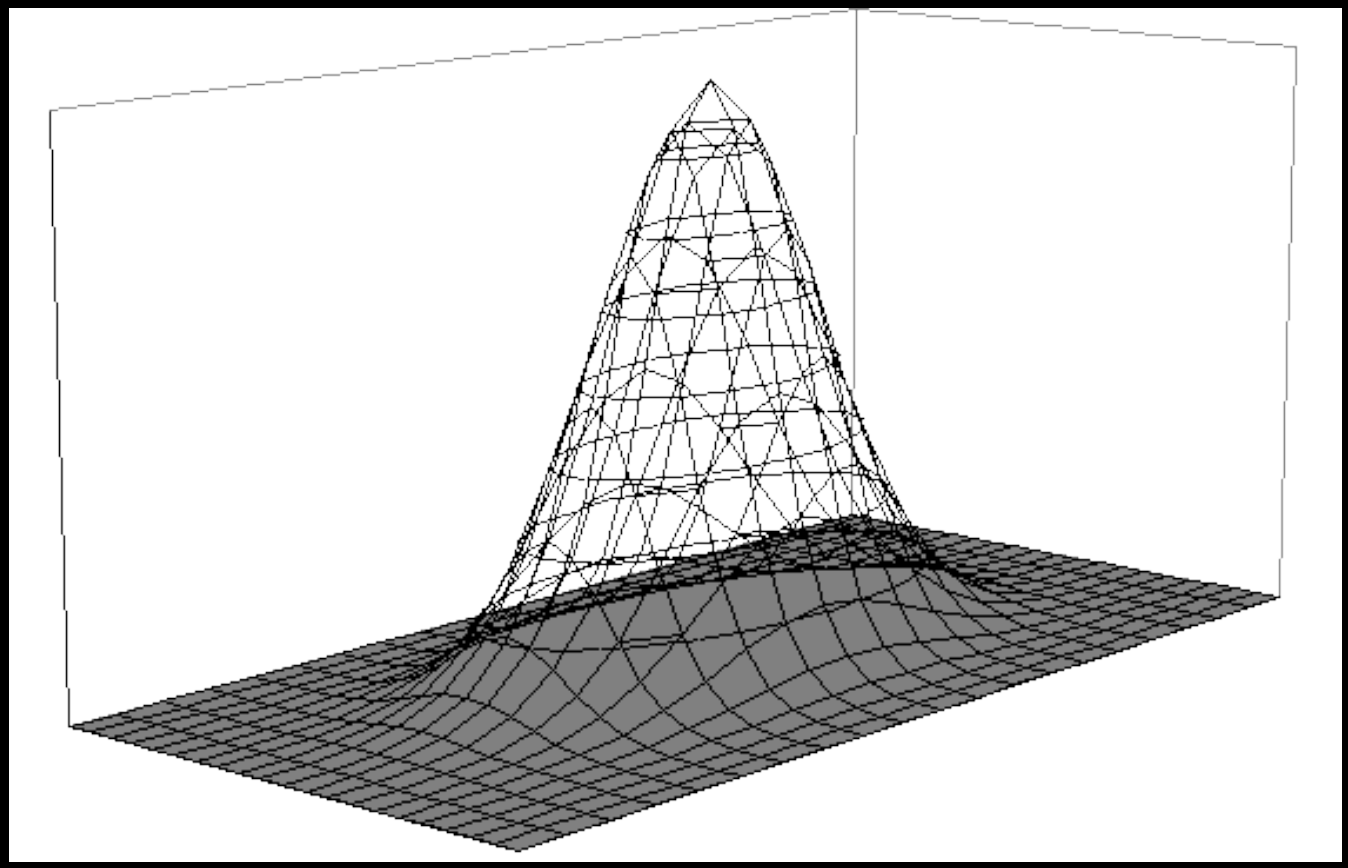
other important distributions

t distribution

F distribution

We'll cover these with hypothesis testing.

multivariate distributions



distributions in R

name	description
`dname()`	probability distribution
`pname()`	cumulative density function
`qname()`	quantile function
`r`*`name`*`()`	random samplins

R names for distributions

distribution	R	distribution	R
Beta	`beta`	Lognormal	`lnorm`
Binomial	`binom`	Neg. Binom.	`nbinom`
Cauchy	`cauchy`	normal	`norm`
Chi Square	`chisq`	Poisson	`pois`
Exponential	`exp`	Student's t	`t`
F	`f`	Uniform	`unif`
Gamma	`gamma`	Tukey	`tukey`
Geomentric	`geom`	Weibull	`weib`
Hypergeom.	`hyper`	Wilxocon	`wilcox`
Logistic	`logis`		