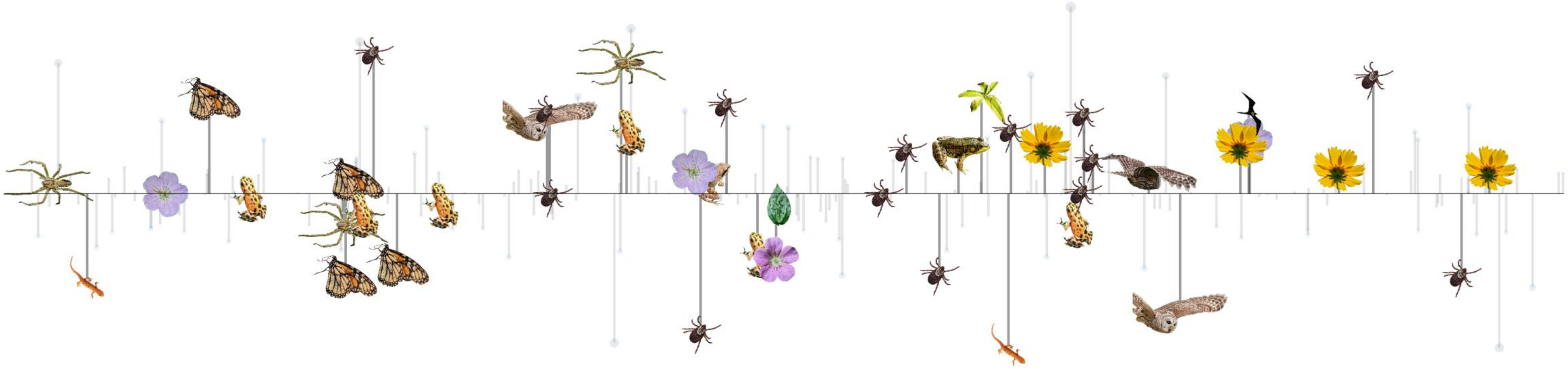
Intro to Quantitative Ecology
UMass Amherst – Michael France Nelson

Deck 7: Differences among more than two samples



Announcements

- The ANOVA material is more theoretically dense than what we've covered up until this point.
- We're going to have to work on our statistical intuition to master these inferential statistics concepts.
- What do I mean by *inferential?*

Descriptive and Inferential Statistics

Descriptive	Inferential
Describes a sample: a finite set of observations	Makes educated guesses about an unobservable population
Sample *statistics* – usually use Latin letters	Population *parameters* – usually use Greek letters
Sample mean: $\bar{x} = \frac{\sum x_i}{n}$	Population mean: μ is estimated by $\bar{x}$
Sample standard deviation: $s = \sqrt{\frac{\sum(x_i - \bar{x})}{n}}$	Population standard deviation: σ is estimated by s
No uncertainty: we can calculate quantities exactly	Uncertainty: the sample statistics are our best guess, but we don't expect the population parameters to be exactly the same as our sample statistics.

Tails and Samples

There is potential for confusion with the terms: two-sample, two-tailed, one-sample, one-tailed.

What do one-sample and two-sample mean?

- Do I want to compare two group means/medians?
  - Two-sample test
- Do I want to compare one group mean/median against a fixed value?
  - One-sample test

What do one-tailed and two-tailed mean?

- Am I proposing a one- or two-tailed alternative hypothesis?

Announcements

- The ANOVA material is more theoretically dense than what we've covered up until this point.
- We're going to have to work on our statistical intuition to master these inferential statistics concepts.
- What do I mean by *inferential?*

Descriptive and Inferential Statistics

Descriptive	**Inferential**
Describes a sample: a finite set of observations	Makes educated guesses about an unobservable population
Sample *statistics* – usually use Latin letters	Population *parameters* – usually use Greek letters
Sample mean: $\bar{x} = \frac{\sum x_i}{n}$	Population mean: μ is estimated by $\bar{x}$
Sample standard deviation: $s = \sqrt{\frac{\sum(x_i - \bar{x})}{n}}$	Population standard deviation: σ is estimated by s
No uncertainty: we can calculate quantities exactly	Uncertainty: the sample statistics are our best guess, but we don't expect the population parameters to be exactly the same as our sample statistics.

Tails and Samples

There is potential for confusion with the terms: two-sample, two-tailed, one-sample, one-tailed.

What do one-sample and two-sample mean?

- Do I want to compare two group means/medians?
  - Two-sample test
- Do I want to compare one group mean/median against a fixed value?
  - One-sample test

What do one-tailed and two-tailed mean?

- Am I proposing a one- or two-tailed alternative hypothesis?

Announcements

Probability Density Functions

A Probability Density Function (PDF) maps every element in a sample space to a measure of relative likelihood.

What is a sample space?

- The set of all possible outcomes.
- What is the sample space of a single coin flip?
- What is the sample space of two coin flips?
- What is the sample space of rolling a die one time?
- What is the sample space of measuring fish mass?

Sample Spaces

What is the sample space of a single coin flip?

- We could get a head or tail: {H, T}
- There are two elements.
- The size of the sample space is 2.

What is the sample space of two-coin flips?

- {HH, HT, TH, TT} if order matters.
- {HH, HT, TT} if order doesn’t matter.
- The size is 3 or 4, depending on whether or not order is important

What is the sample space of rolling a die one time?

- {1, 2, 3, 4, 5, 6}

Continuous and Discrete Sample Spaces

The probabilities of every event in the sample space must sum to 1.0.

- Rolling a head has probability 1/2 with a fair coin.

- Each number on a fair die has probabiltiy of 1/6.
  - The sum of numbers when you roll two dice is more complicated.
  - The sample space is {2, 3, 4, 5, 6, 7, 8. 9. 10, 11, 12}
  - Can you figure out the probabilities?

Continuous and Discrete Sample Spaces

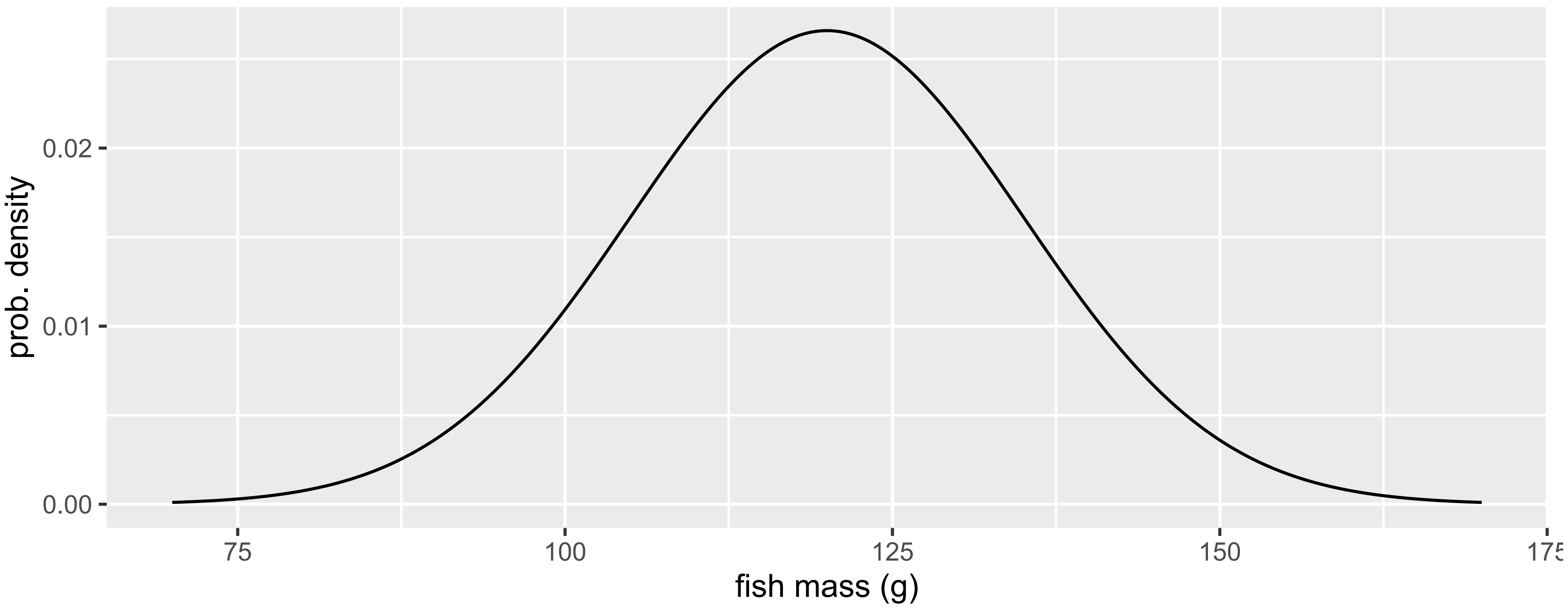
What is the sample space of measuring fish mass?

- This is a more subtle question.
- Continuous quantities can (potentially) take on any intermediate value.
- The probability of measuring any *particular* continuous quantity with infinite accuracy is zero.
  - This might feel weird and non-intuitive.
  - It has to do with the fact that a probability is an integral of the PDF, and the integral at a single point is always zero.
    - This is from calculus - don’t worry we won’t be solving any integrals!

Continuous Sample Spaces + Probabiltiy Density

Consider a population of fish with normally-distributed masses.

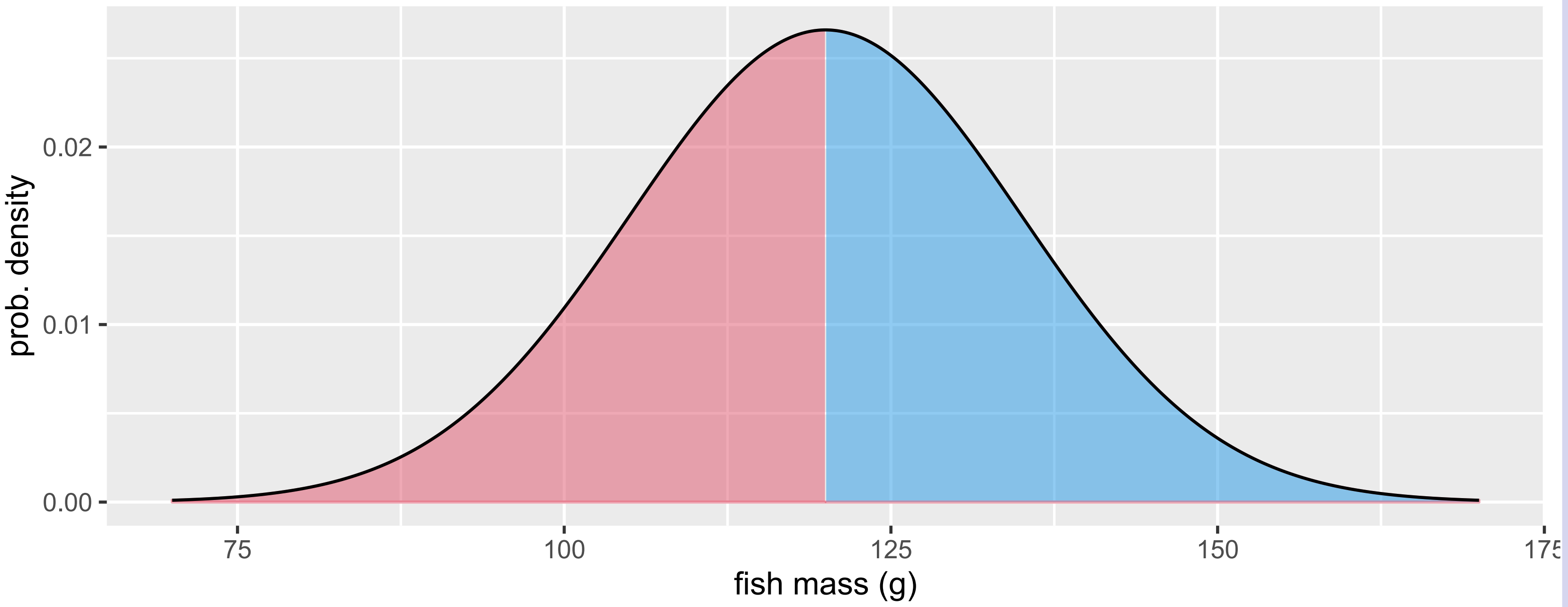
- Mean fish mass is 120 grams.
- Fish mass standard deviation is 15 grams.



Continuous Sample Spaces + Probabiltiy Density

Consider a population of fish with normally-distributed masses.

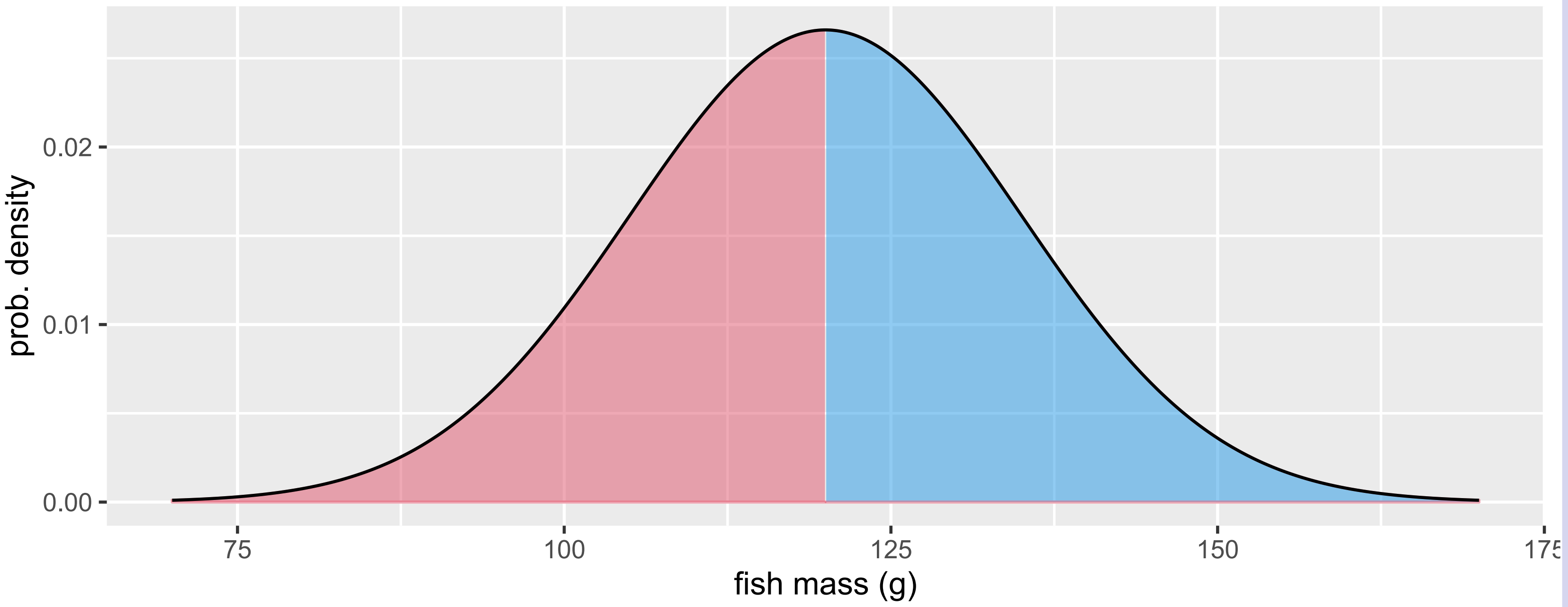
- What is the probability of catching a fish ***lighter*** than 120 grams?



Continuous Sample Spaces + Probabiltiy Density

Consider a population of fish with normally-distributed masses.

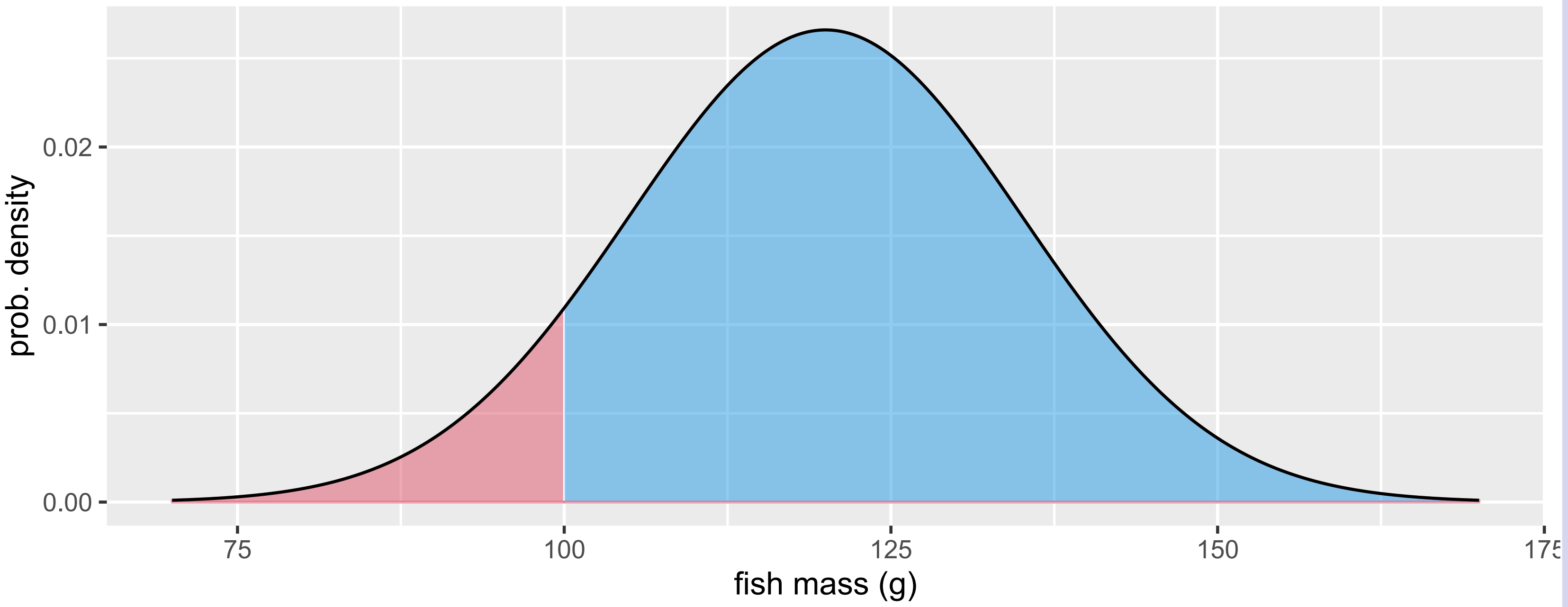
- What is the probability of catching a fish ***heavier*** than 120 grams?



Continuous Sample Spaces + Probabiltiy Density

Consider a population of fish with normally-distributed masses.

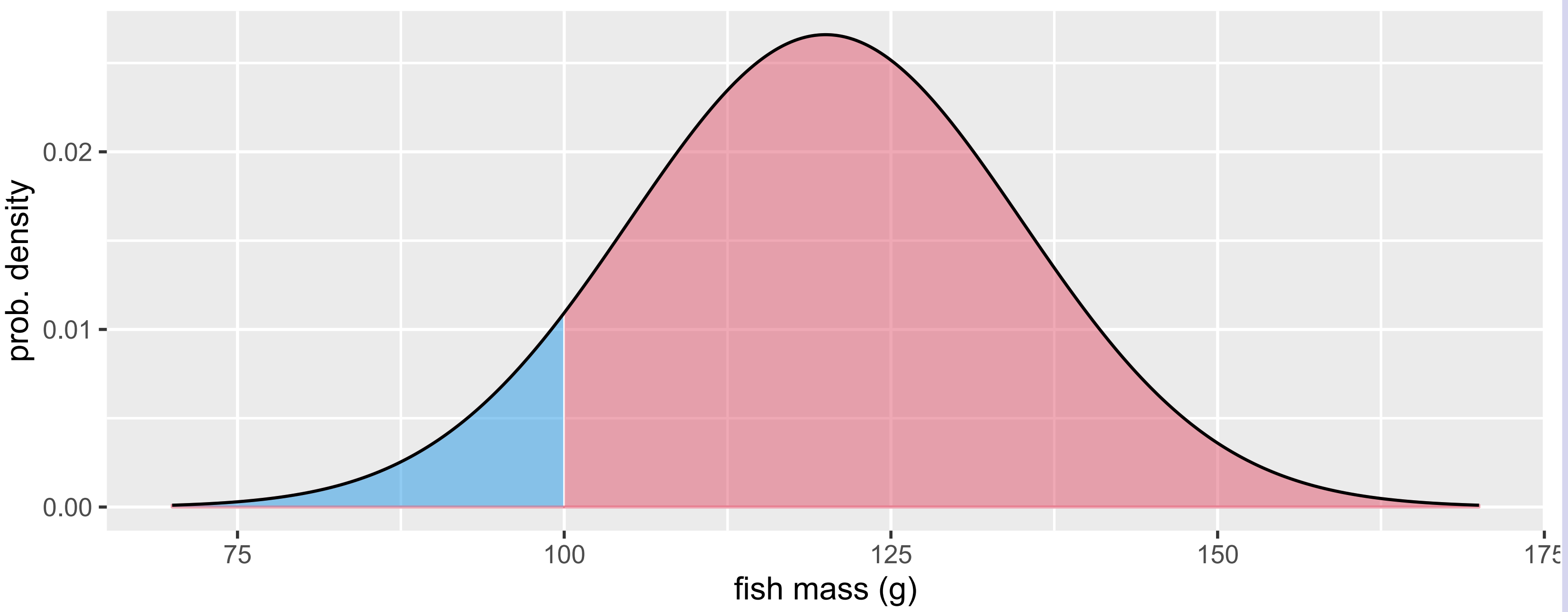
- What is the probability of catching a fish ***lighter*** than 100 grams?



Continuous Sample Spaces + Probabiltiy Density

Consider a population of fish with normally-distributed masses.

- What is the probability of catching a fish ***heavier*** than 100 grams?



Two-Sample Tests

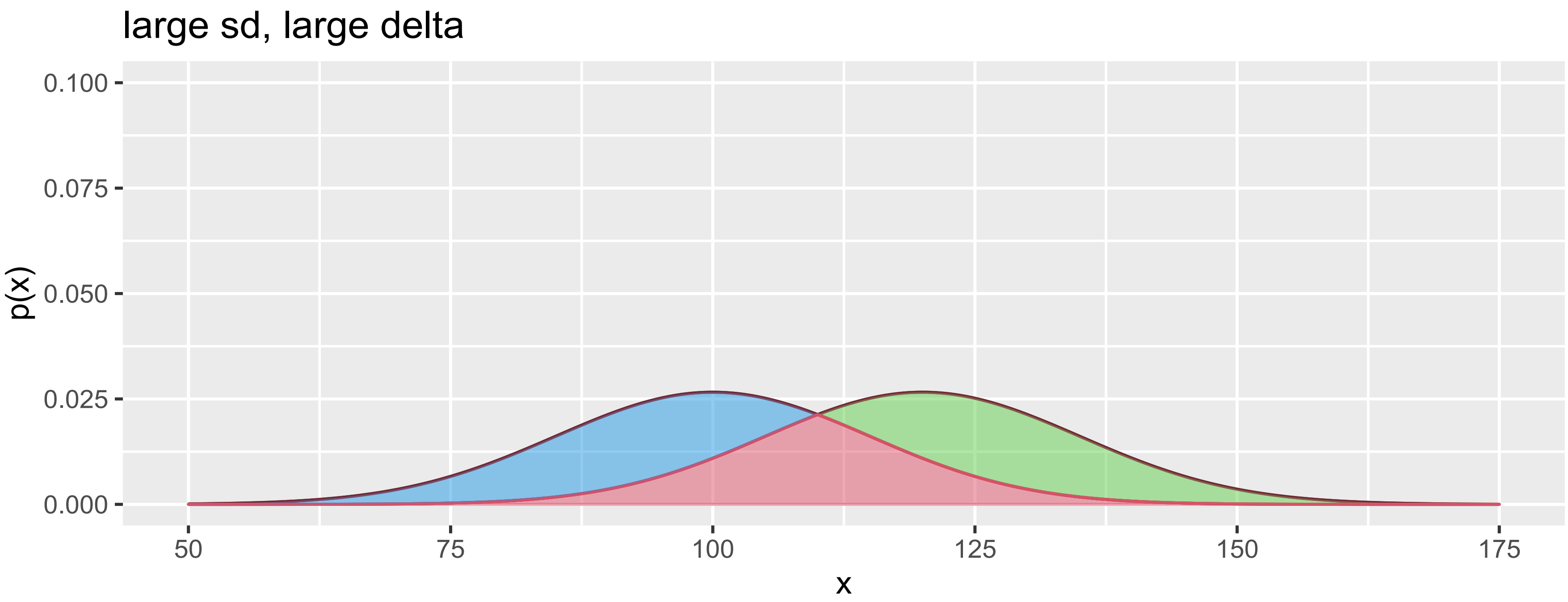
More Visual Intuition

Differences Between Two Groups

- What are we really asking with a t-test (or Wilcox test)?
  - Is there a *significant* difference between the group means (or medians)?
- How do we assess the significance?
  - We use a *test statistic* (t or U)
  - We calculate a p-value from the test statistic
- What is the two-group null hypothesis?
  - There is no difference between the group means.
  - The samples were all drawn from one (statistical) population.

Two Sample Tests: Visualizing Hypotheses

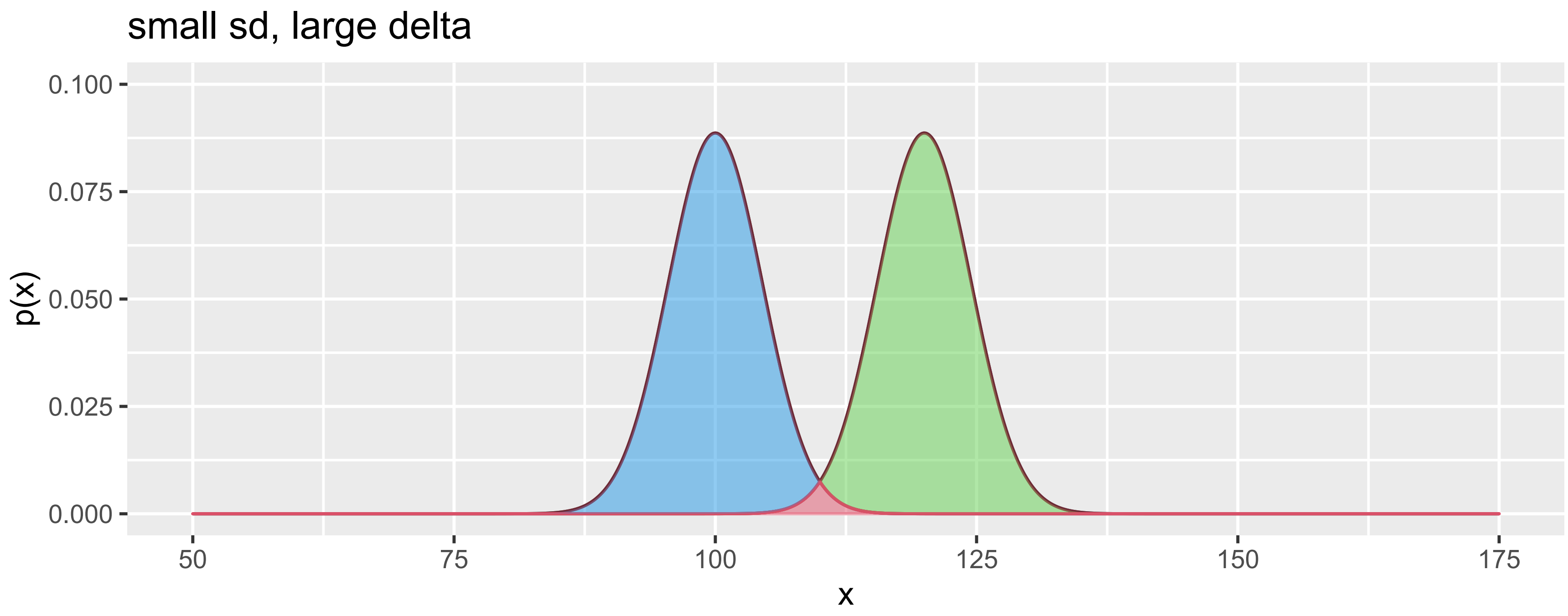
The two-sample test Alternative hypothesis:



These two groups have large difference in means, large dispersion. Lots of overlap, hard to tell apart.

Two Sample Tests: Visualizing Hypotheses

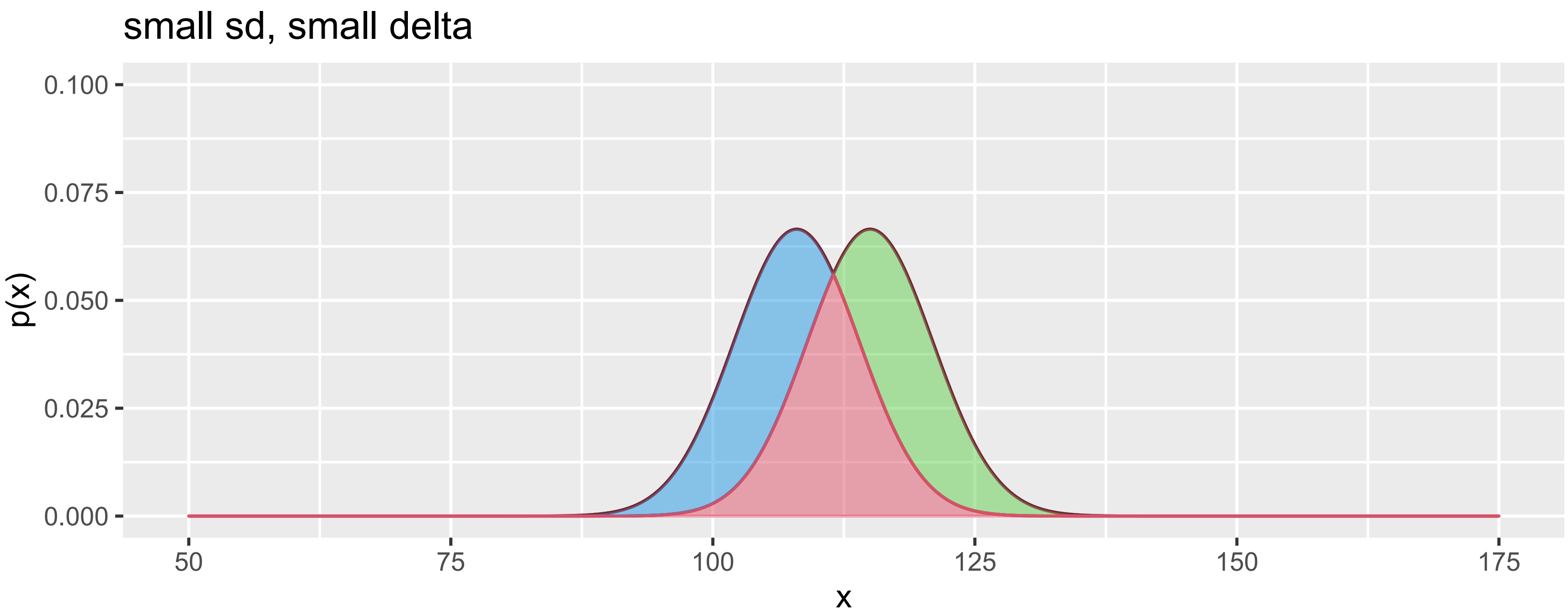
The two-sample test Alternative hypothesis:



These two groups have large difference in means, small dispersion. Small overlap, easy to tell apart

Two Sample Tests: Visualizing Hypotheses

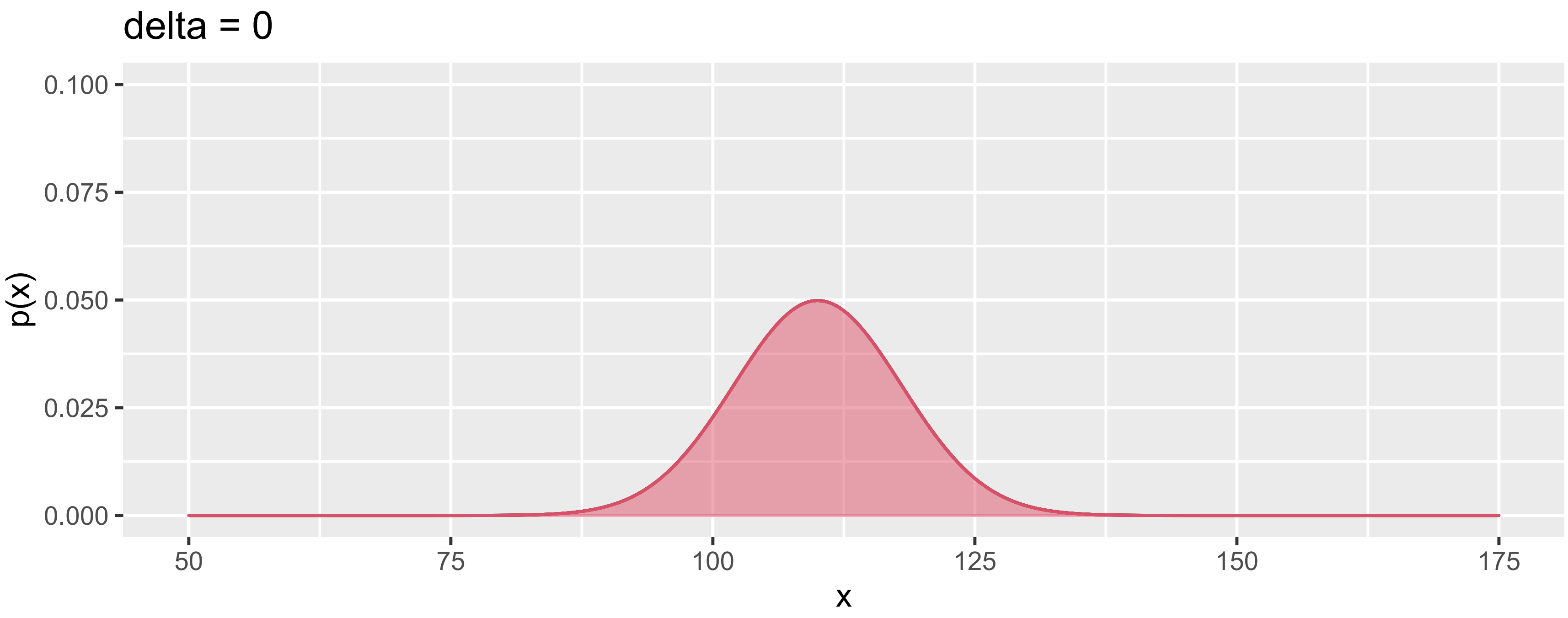
The two-sample test Alternative hypothesis:



These two groups have small difference in means, small dispersion. Lots of overlap, hard to tell apart.

Two Sample Tests: Visualizing Hypotheses

The two-sample test Null Hypothesis:



Complete overlap, impossible to tell apart.

Experimental Design and Sampling Units

What’s a Sampling Unit (SU)?

T-tests are great if we only have two groups, but what if we need to analyze more complicated scenarios? We need to know more about experiments and sampling units.

Scenario context: We’re interested in bluegill in Massachusetts lakes.

Sampling Units

What is a **sampling unit**?

- It's the thing that I measure.
  - A sampling unit has 1 or more attributes.
  - A sampling unit us *usually* the focus of my experiment or analysis.

Sampling units seem like a simple concept, but there is some subtlety.

- Keep an eye out for the different sampling units in the example scenarios.

Scenario 1

Thorsten is wants to know if the average mass of fishes differs among Massachusetts lakes.

- He sampled 30 fish from each of 16 lakes.
- What is the sampling unit and what were the attributes?
- He used an Analysis of Variance ANOVA to analyze his data. What were the predictor and response variables?

Scenario 1

Thorsten is wants to know if the average mass of fishes differs among Massachusetts lakes.

- He sampled 30 fish from each of 16 lakes.
- What is the sampling unit and what were the attributes?

SU = individual fishes, attributes = mass and lake

- He used an Analysis of Variance ANOVA to analyze his data. What were the predictor and response variables?

Lake = predictor, Fish mass = response

Scenario 1

He found a significant lake effect: i.e. the mean fish mass is not the same in all lakes.

- Did he observe a high, or a low p-value?

Scenario 1

Now he wants to know which lakes are different from each other

- What, exactly, is he asking?

- What does he need to compare?

- What test could he use?

Scenario 1

Now he wants to know which lakes are different from each other

- What, exactly, is he asking?

Is at least one lake different from the others?

- What does he need to compare?

Pairwise mean body masses for the lakes

- What test could he use?

Post-hoc test: Tukey Honest Significant Difference

Scenario 2

I am interested in testing whether there is a significant difference between the **population density** of fish in 30 **low salinity** lakes and 30 **high salinity** lakes:

1. What is the sampling unit

2. Which statistical test could I use?

3. Which is the test statistic for the test I chose?

Scenario 2

I am interested in testing whether there is a significant difference between the **population density** of fish in 30 **low salinity** lakes and 30 **high salinity** lakes:

1. What is the sampling unit

   Individual lakes – population density, a property of a lake

2. Which statistical test could I use?

   T-test: there are two groups of lakes (but beware of nested experiment design – lake is confounded with salinity)

3. Which is the test statistic for the test I chose?

   T-statistic

Scenario 3

I am interested in testing whether there is a significant difference between the **population density** of fish in 30 **low salinity** lakes and 30 **high salinity** lakes.

In fact, I actually sampled 10 **large**, 10 **medium**, and 10 **small** lakes in each at high and low salinity levels. I'm also interested in lake size

Scenario 3

I want to explore whether there are differences in population density based on lake **salinity** and lake **size**.

1. What is the sampling unit?
2. How has our question changed?
3. What are the predictor variables?
4. What test could I use
5. What is the test statistic?

Scenario 3

I want to explore whether there are differences in population density based on lake **salinity** and lake **size**.

1. What is the sampling unit?
   Individual lake
2. How has our question changed?
   Added another variable: lake size
3. What are the predictor variables?
   1. **Salinity**
   2. **Size**
4. What test could I use
   Two-way ANOVA
5. What is the test statistic?
   F statistic – one for each factor

Comparing differences - two smaples

T-Test	Significance
• tests whether group means differ significantly • H_0: there is no significant difference between the means • H_1: there is a significant difference between the means	• t-statistic: $t = \frac{\lvert \overline{x}_a - \overline{x}_b \rvert}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$ • Numerator is means • Denominator is standard deviations • degrees of freedom • *p*-value

Comparing differences - more than two samples

????

But what if there are more than two groups?

One-way ANOVA

One-Way Analysis of Variance (ANOVA):

- Statistical test for testing for differences among >2 groups
- ANOVA and t-test are *nearly* identical when there are 2 groups
- There is one factor/group/category (*One-way ANOVA*)

One-way ANOVA

What are some key ANOVA assumptions?

1. Data are normally distributed *within groups*.
2. Groups have the same amount of dispersion.
3. Observations are independent.

One-way ANOVA

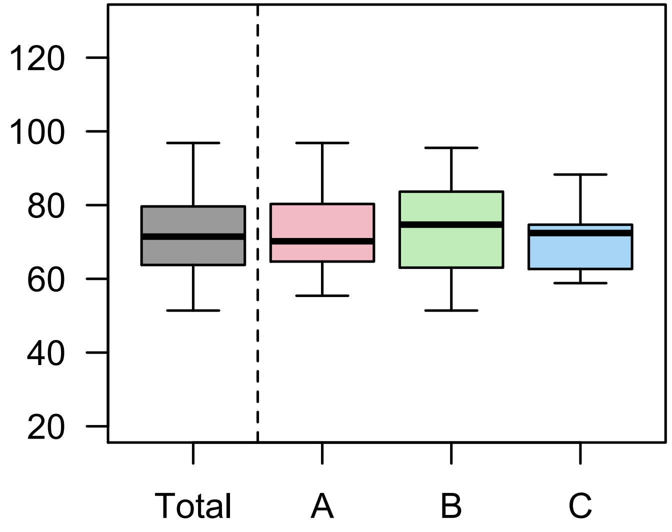
ANOVA hypotheses:

- H_0 - There is no difference in group means, the group means are all the same
- H_1 - Group means are not all the same

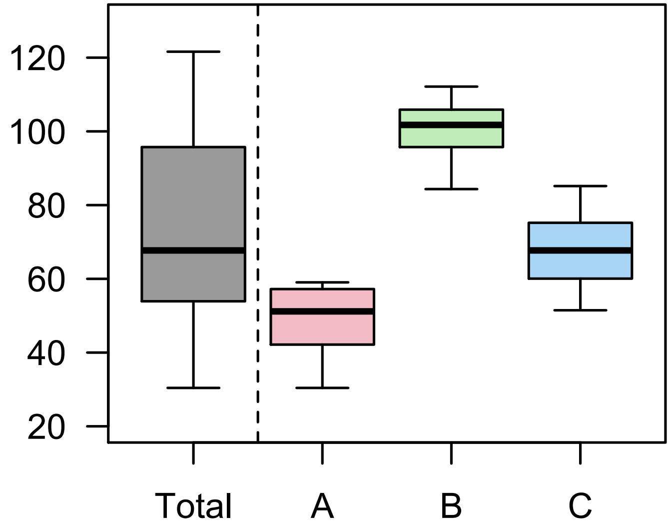
Why is there no directional option?

ANOVA Explained: graphical intuition

The ANOVA partitions the *total* variation into *within* sample (group) variation with *between* sample (group) variation to determine whether samples come from a single distribution or not.



Null hypothesis: All of the variation is within-groups.



Alternative Hypothesis: There is both within- and among-group variation

Detour: Sums of Squares (SS)

Sums of Squares are ubiquitous in ANOVA, but what is a sum of squares?

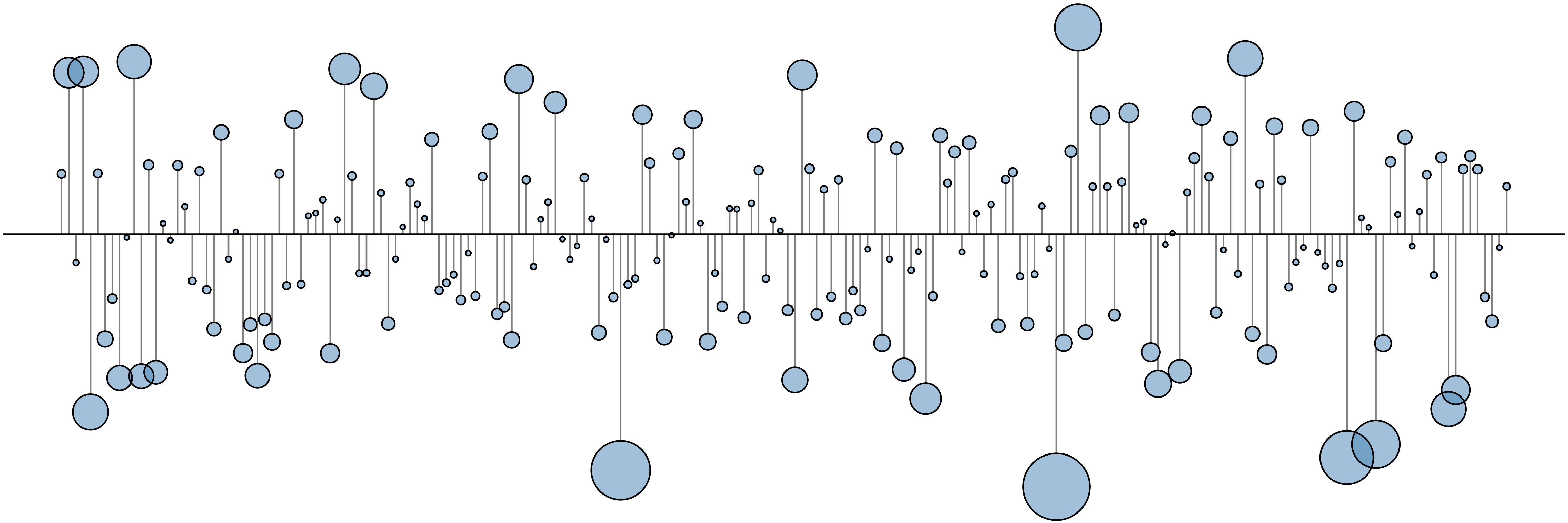
- “The sum of squared deviations from the mean”
- The difference between an observation and the mean…squared.
- A measure of the total variability in a collection of numbers.

Why do we use the sum of **squares**? Squaring has some desirable properties:

- Forces all the terms to be positive.
- An accelerating penalty for extreme values.

Squared Deviations

Circle area is proportional to the squared distance from mean value:



The standard normal distribution

What does standardized mean?

Recall that the *standard normal* has $\mu = 0$ and $\sigma = 1.$

Z-standardization:

- We can convert any value from any normal distribution into the *standard normal* by:

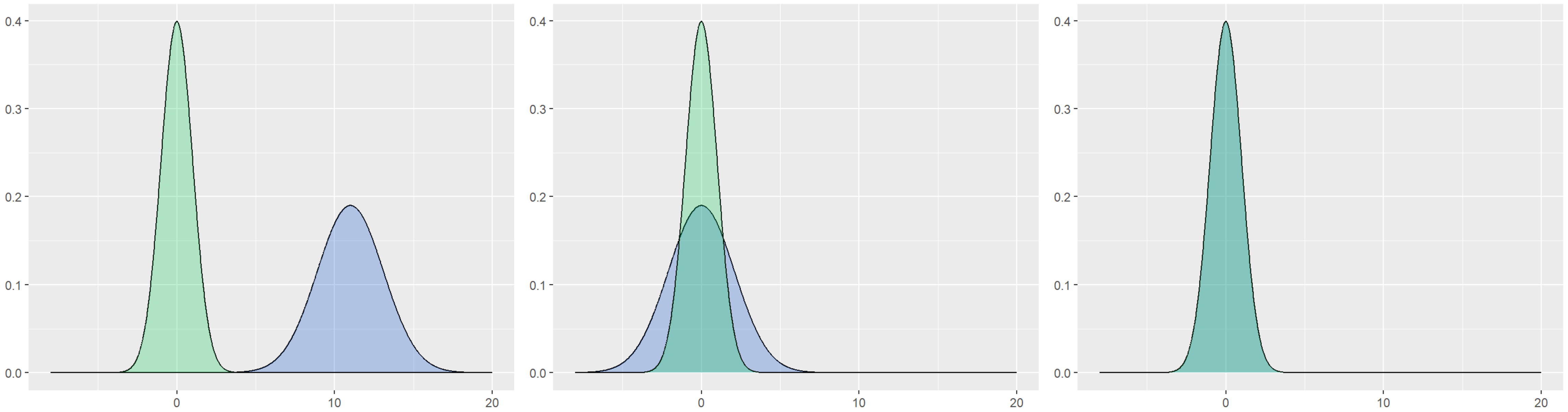
1. Subtracting the mean
2. Dividing by the standard deviation.

- We call a value standardized this way a *z-value*.
- It's very similar to how we calculate a *t-value*, more on that when we talk about t-tests and the t-distribution…

Z-standardization:

We can convert any value from any normal distribution into the *standard normal* by:

1. Subtracting the mean
2. Dividing by the standard deviation.



Independent Events

$$\mathbf{Pr}(\mathbf{A} \cap \boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{A}) \times \mathbf{Pr}(\boldsymbol{B})$$

- The probability that A and B both occur is equal to the product of the individual probabilities…
- We'll dissect this surprisingly important definition.

Independent events

Events are independent if knowing the value of one observation gives us no information about the value of another observation:

1. I measure the temperature in Neuquén, Argentina on November 23, 1823.
2. I measure the temperature in Amherst on July 4, 2020.

The Neuquén temperature in 1823 probably doesn’t tell me much about Amherst in 2020

- Likewise, the temperature here today probably won’t tell me much about what to expect there! [Other than knowing that it is fall here, and it was spring/summer there!]

Independent events

Non-Independent Temperatures

Compare the previous temperature example to:

1. I measure the temperature in Amherst on July 4, 2020 at 4:05PM (it is 20C)
2. I measure the temperature in Amherst on July 4, 2020 at 4:11PM (it is 21C)

The temperature at 4:05 gives me a lot of information about what the temperature will be in the same location six minutes later.

Independent events

Suppose we are equally likely to observe these temperatures:	Independent events: joint probability is product of individual probabilities
Temperatures on July 4^{th}: $Pr(temp = 19C) = 0.05$ $Pr(temp = 20C) = 0.05$ $Pr(temp = 21C) = 0.05$	If successive temperature measurements were independent: • $Pr(20){*}Pr(21)$=0.05∗0.05=0.0025 or about 0.25% Do you think observing a temperature of 20, followed by another temperature of 20 in the same location 6 minutes later is only 0.25%??? It's probably much higher than 5% (the *unconditional* probability of observing 20C.)

Independent events

If events are independent, the probability of observing a *specific* set of events (the joint probability) is the same as the product of the events of the individual events.

- I pick up an acorn in each hand simultaneously, from a *very large* collection of acorns of several species.
- Does knowing that the acorn in my left hand is from a Bur Oak tell me anything about the acorn in my right hand?

Independence and Maximum Likelihood

- This may not seem important now, but it is *crucial* to the likelihood concepts we'll examine later.
- It's also key to understanding Bayes' Rule.

Probability and Distributions: Probability Theory

Probability theory concerns the likelihood of events

Distributions are tools for describing the likelihood of observing specific events from the set of all possible events.

- They map ***events*** to ***likelihoods***

There are many named *parametric* distributions with well-understood, useful, and sometimes surprising properties.

Probability theory gets complicated and difficult *very* quickly!

- I’ll attempt to help you develop intuition about the most essential parts.
- This isn’t a course on probability theory – we’ll only cover the basics.

Probability Theory Essentials

Probabilities are non-negative	Sample space: the set of all possible events
• A probability can be any value between zero and 1.0, inclusive. • The probability of a specific event is usually less than 1.0 • **Law of total probability:** The sum of the probabilities of all possible events is 1.0	• Events: a possible outcome of a stochastic process • The definition of event is context-specific: • “What is the probability of catching a fish that weighs 405 grams?” • “What is the probability of catching a fish that weighs between 399 and 411 grams?” • “What is the probability of catching a fish that weighs less than 200 grams?” • “What is the probability that I observe 2 gray jays?”

Probability Notation Basics

Basic probability

- $\Pr(\mathrm{A}) = 0.05$
  - Read as: "The probability that event A occurs is 5%"

Joint probability

- $\Pr(\mathrm{A}\ and\ B)$ = $\Pr(\mathrm{A} \cap B) = 0.05$
  - Read as: "The probability that both events A and B occur is 5%"

Conditional Probability

- $\Pr(\mathrm{A}|B) = 0.05$
  - Read as: "The probability that event A occurs, given that B has already occurred is 5%"

Independent Events

$$\mathbf{Pr}(\mathbf{A} \cap \boldsymbol{B}) = \mathbf{Pr}(\boldsymbol{A}) \times \mathbf{Pr}(\boldsymbol{B})$$

- The probability that A and B both occur is equal to the product of the individual probabilities…
- We'll dissect this surprisingly important definition.

Independent events

Events are independent if knowing the value of one observation gives us no information about the value of another observation:

1. I measure the temperature in Neuquén, Argentina on November 23, 1823.
2. I measure the temperature in Amherst on July 4, 2020.

The Neuquén temperature in 1823 probably doesn't tell me much about Amherst in 2020

- Likewise, the temperature here today probably won't tell me much about what to expect there! [Other than knowing that it is fall here, and it was spring/summer there!]

Independent events

Non-Independent Temperatures

Compare the previous temperature example to:

1. I measure the temperature in Amherst on July 4, 2020 at 4:05PM (it is 20C)
2. I measure the temperature in Amherst on July 4, 2020 at 4:11PM (it is 21C)

The temperature at 4:05 gives me a lot of information about what the temperature will be in the same location six minutes later.

Independent events

Suppose we are equally likely to observe these temperatures:	Independent events: joint probability is product of individual probabilities
Temperatures on July 4^{th}: $Pr(temp = 19C) = 0.05$ $Pr(temp = 20C) = 0.05$ $Pr(temp = 21C) = 0.05$	If successive temperature measurements were independent: • $Pr(20) * Pr(21) = 0.05 * 0.05 = 0.0025$ or about 0.25% Do you think observing a temperature of 20, followed by another temperature of 20 in the same location 6 minutes later is only 0.25%??? It's probably much higher than 5% (the *unconditional* probability of observing 20C.)

Independent events

If events are independent, the probability of observing a *specific* set of events (the joint probability) is the same as the product of the events of the individual events.

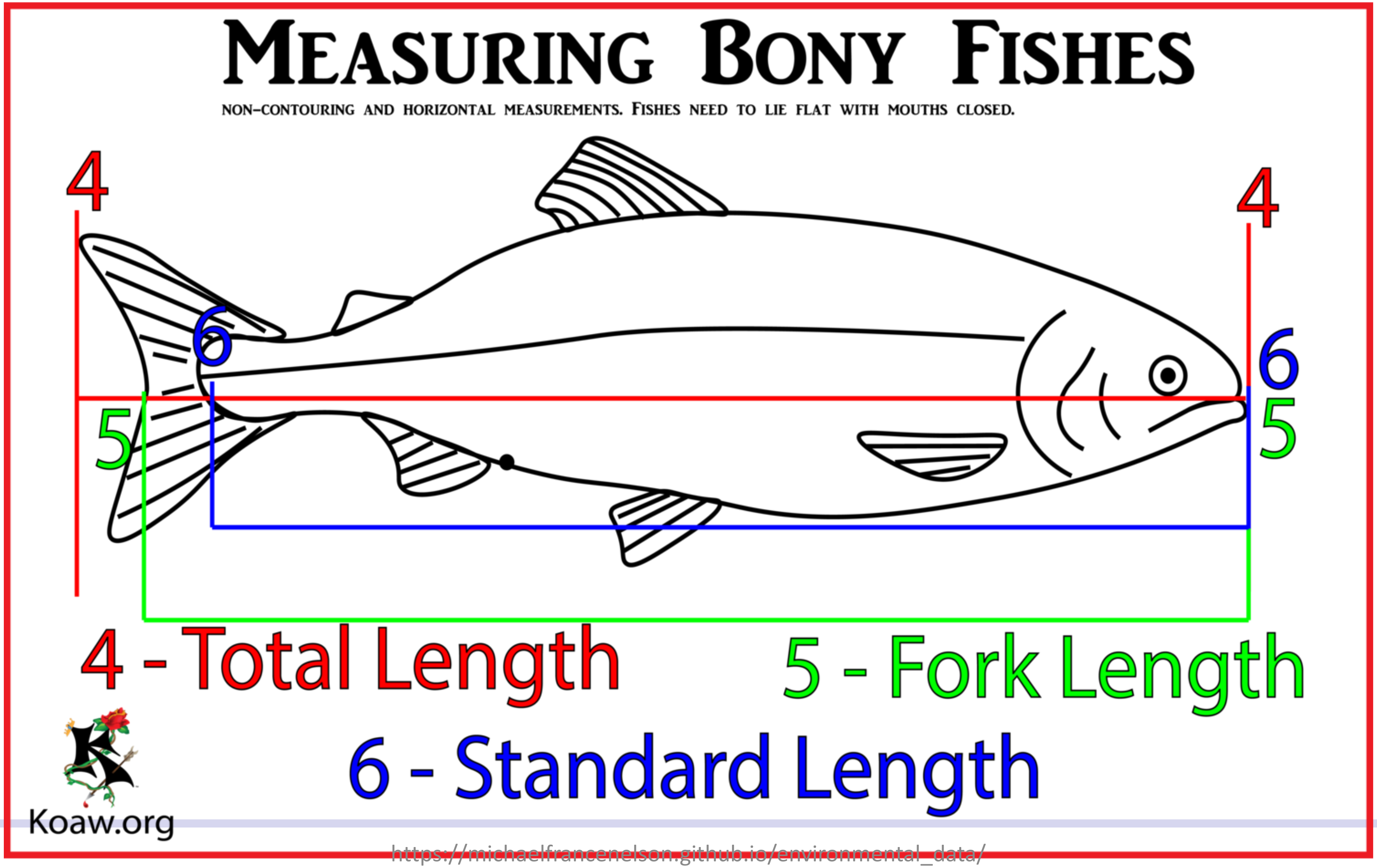
- I pick up an acorn in each hand simultaneously, from a *very large* collection of acorns of several species.
- Does knowing that the acorn in my left hand is from a Bur Oak tell me anything about the acorn in my right hand?

Independence and Maximum Likelihood

- This may not seem important now, but it is *crucial* to the likelihood concepts we'll examine later.
- It's also key to understanding Bayes' Rule.

Probability Distribution Functions

Probability Distribution Functions



What questions should a distribution function answer?

- Am I more likely to observe a fish that is 20cm, or a fish that is 11cm?
  - Probability Density Function: relative likelihood of x
- What is the probability that a fish is longer than 20cm?
  - Likelihood of x or smaller: cumulative Density Function
- How long is a fish in the 90th percentile?
  - Quantile Function

Probability Density/Mass Functions

Probability density or mass functions answer the questions:

- Am I more likely to catch a fish that measures 6cm or 14.5cm?
- What is the probability that I collect *exactly* two acorns of Red Oak out of a mixture of Red Oak and Bur Oak acorns?

They associate an event with a measure of likelihood

- This is the probability of the event for discrete distributions
- For continuous distributions it is more complicated, but you can think of it as relative likelihood.

Probability Density/Mass Functions

They are maps of events in the sample space to probabilities.

- **Probability Density Functions** for continuous distributions
- **Probability Mass Functions** for discrete distributions.

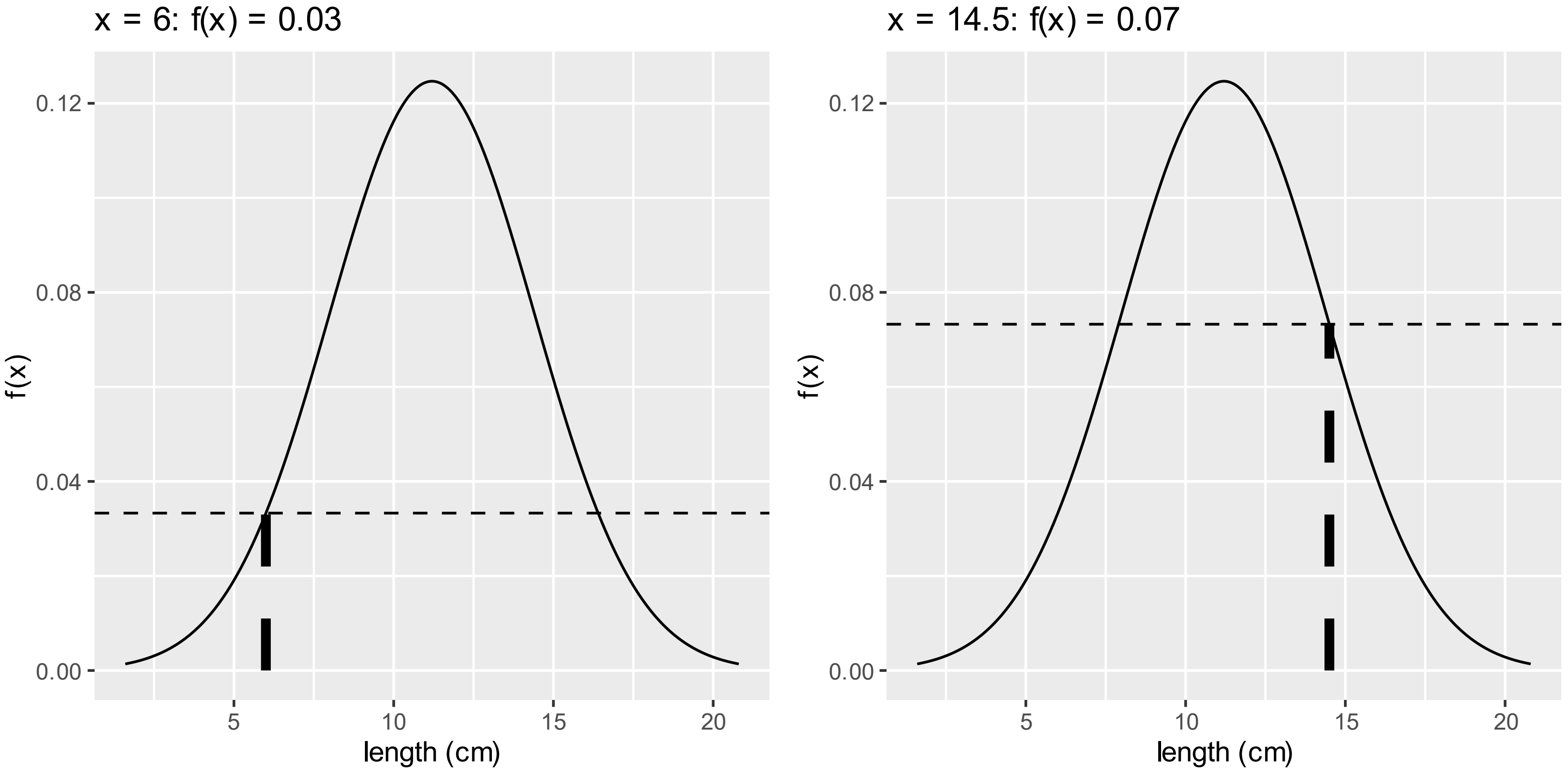
Probabilities are always between zero and one:

- The values of PDFs and PMFs are always non-negative.

Probability Density Function

PDFs tell us about relative likelihoods

Am I more likely to catch a fish that measures 6cm or 14.5cm?



Cumulative Probability Functions: CDFs & CMFs

The CDF/CMF answers:

- What is the probability that I catch a fish that weighs 153g or less?
- What is the probability that *at least* 3 of the acorns are Bur oak?

Cumulative density is the **accumulated area under the density curve** to the left of x.

- It's an integral (or a sum for discrete distributions).
- It is the probability of observing a value equal to or less than x.
- The nth percentile (quantile).

Cumulative Probability Functions: CDFs & CMFs

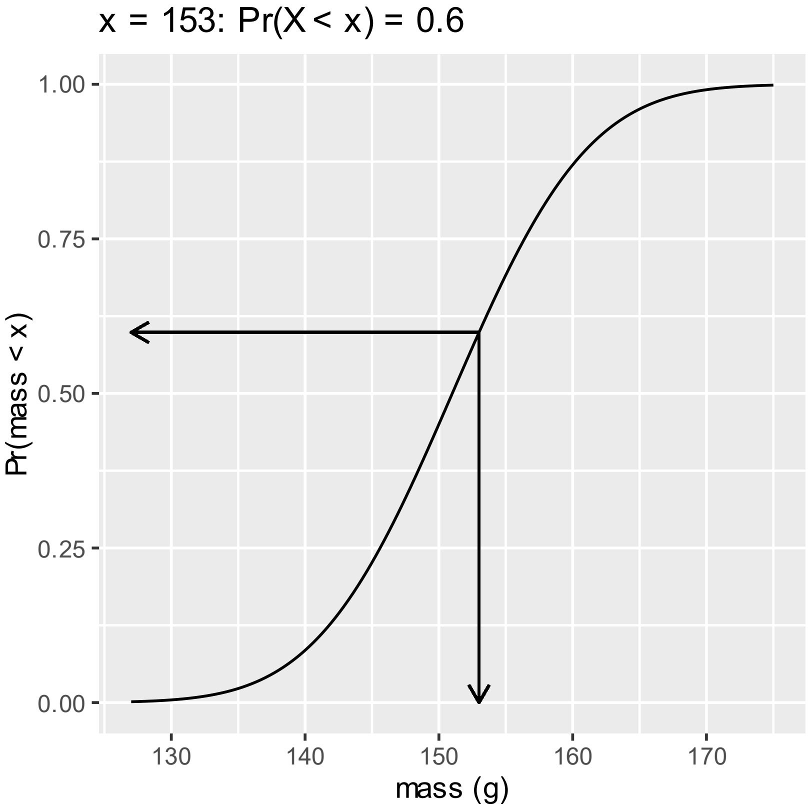
CDFs tell us the probability of an event:

What is the probability that I catch a fish that weighs 153g or less?

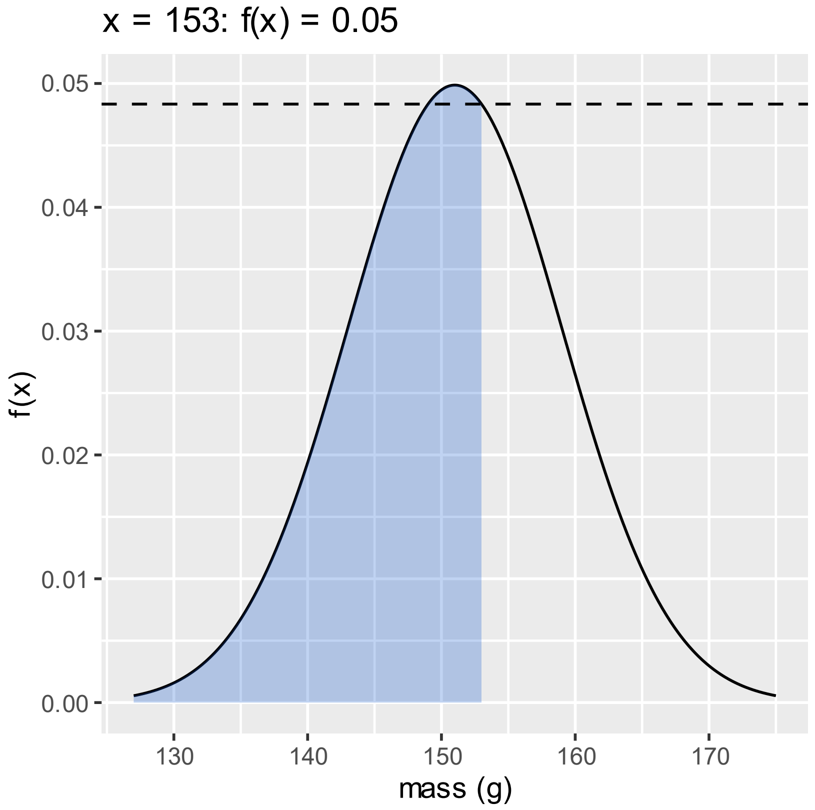
- **Read the mass on the x-axis: 153g.**

- **Read the corresponding probability on the left: 60%**

CDF



PDF

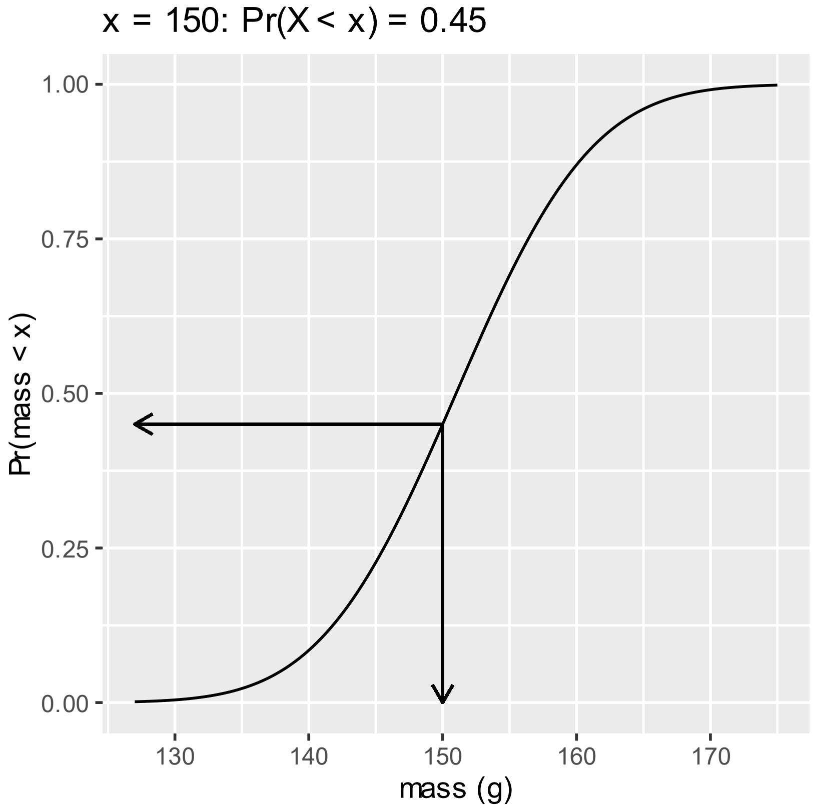


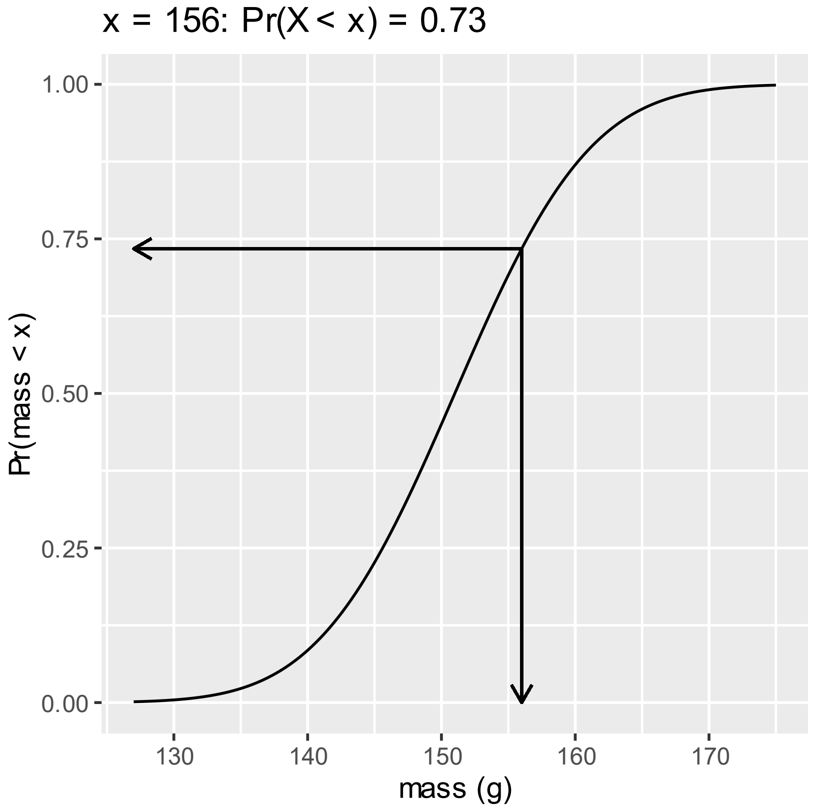
Cumulative Probability Functions: CDFs & CMFs

CDFs tell us the probability of an event:

What is the probability that I catch a fish that weighs between 150g and 156g?

- **Take the difference of probabilities from the CDF: 0.73 – 0.45 = 28%**



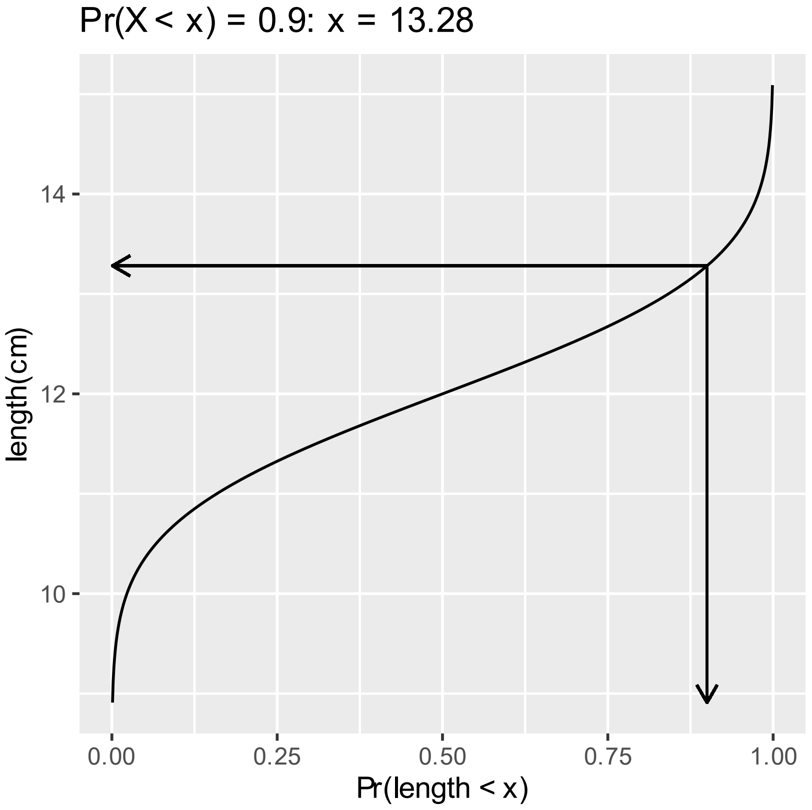


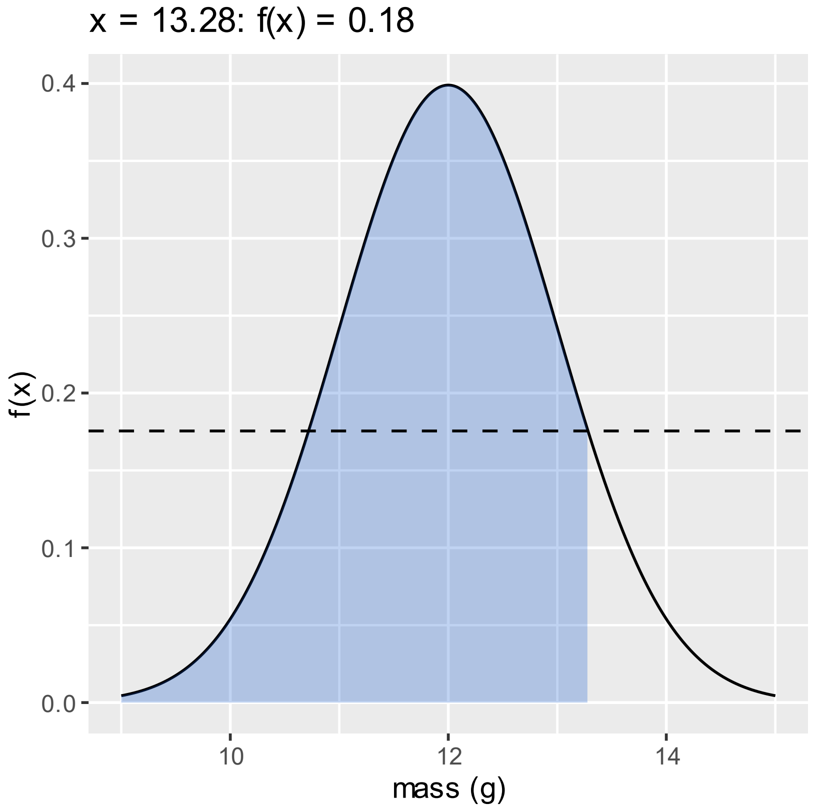
Quantile Functions

Quantile functions tell us about percentiles:

What length will 90% of all fishes will be shorter than?

- **Read the percentile on the x-axis.**
- **Read the size on the y-axis.**



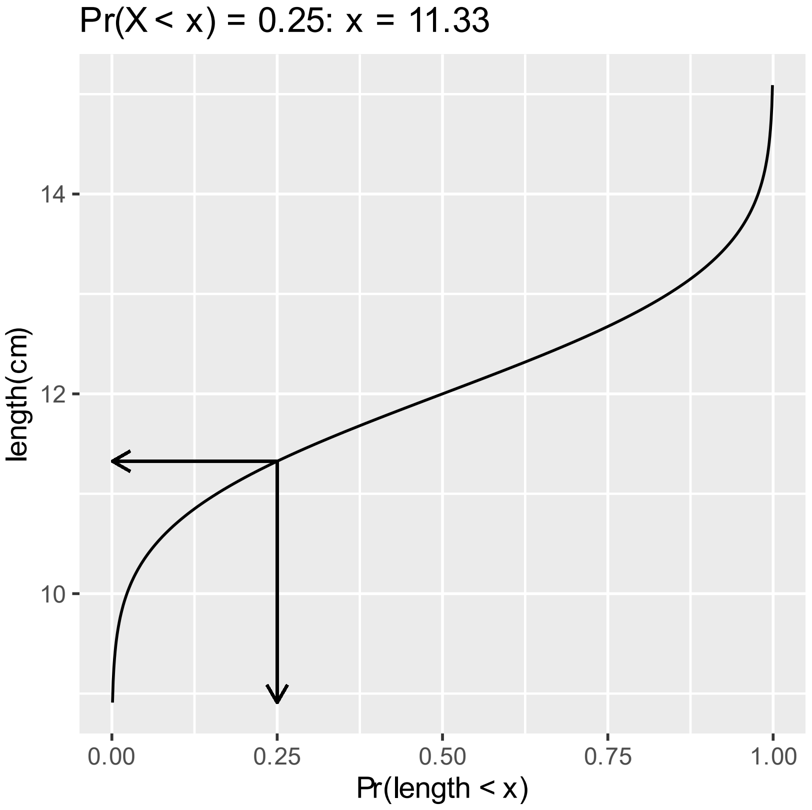


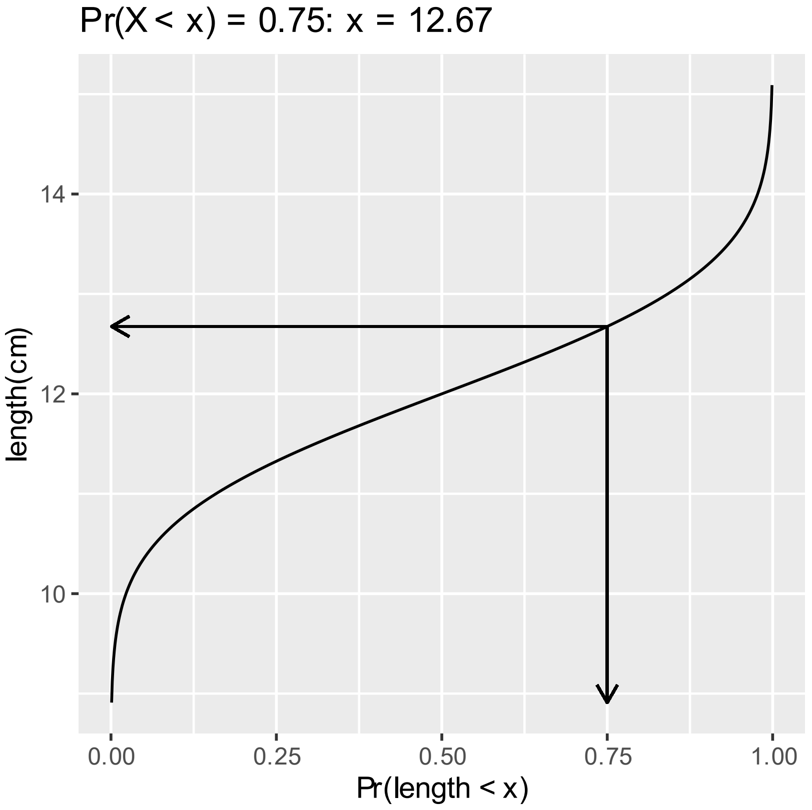
Quantile Functions

Quantile functions tell us about percentiles:

What lengths span the middle 50% of the range?

- **Read the percentiles on the x-axis.**
- **Read the sizes on the y-axis: 11.3cm – 12.7cm**





Parametric (Theoretical) Distributions

Parametric distributions are defined by mathematical *functions*

- The functions have one or more *parameters* that define how probabilities are allocated to events.
- We often want to estimate the parameters from samples.

The *binomial distribution* has two parameters: n, p.

The *Poisson* distribution has only one parameter: λ

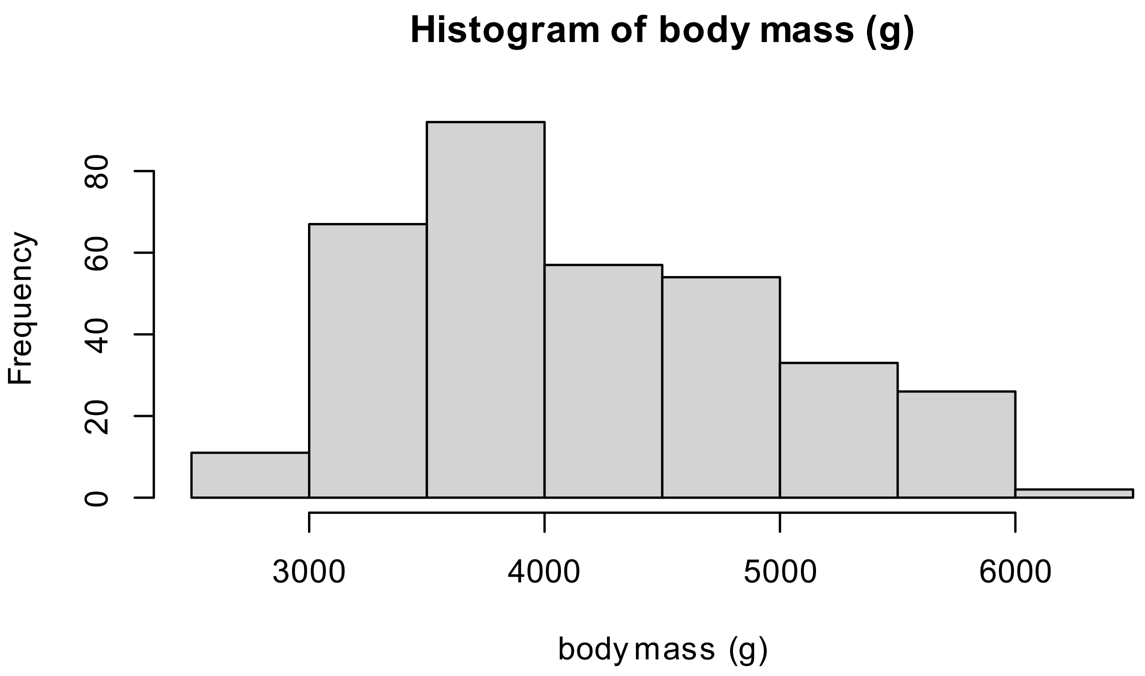
Empirical Distributions

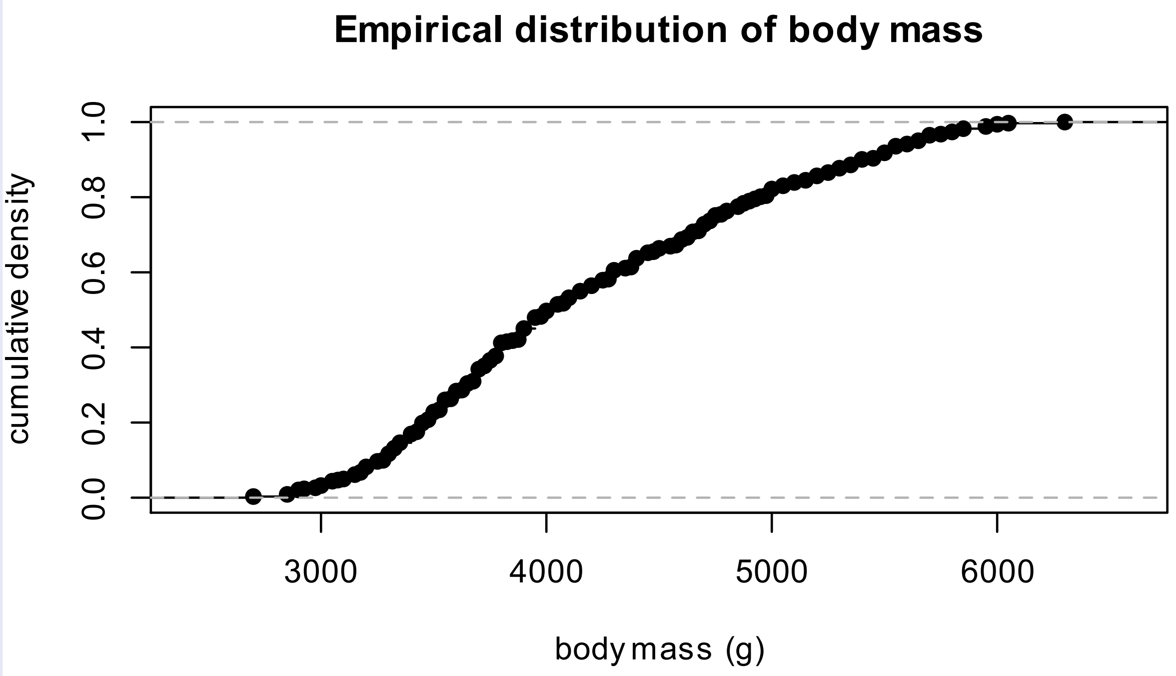
Empirical distributions are computed from *observations*.

- **There is no analytical function: the shape is computed from data.**
- **We can compare empirical distributions to parametric distributions.**

Histograms are analogous to a PDF/PMF

Empirical cumulative distribution function are analogous to the CDF/CMF





Recap

- Theoretical and empirical distributions
- Parameters
- Distribution functions
  - Probability Density/Mass
  - Cumulative Density/Mass
  - Quantile Functions

Announcements

- Assignments and lecture are out of sync…
  - Salamander ANOVA now due next Monday.
  - The next few assignments will also be pushed forward…details to follow.
- ANOVA degrees of freedom mistake fixed: correct values are:
  - Total: $df_T = n - 1$
  - Within: $df_W = \mathrm{n} - \mathrm{g}$
  - Between: $df_B = \mathrm{g}\ - 1$
  - (more details in a few slides!)

Squared Deviations

Sums of squares are used throughout statistics to quantify dispersion:

- Variance and standard deviation
  - These are standardized by the number of observations. That way we can compare different samples that may have different numbers of observations.
- Model error
  - What is the squared difference between our model prediction and the observed value?

Large deviations from the mean contribute a lot more because of the squaring.

Sums of Squares for ANOVA

In the ANOVA world, we are concerned with three main sums of squares:

1. **Total** sum of squares: the sum of all squared deviations form the grand mean.
2. **Within-group** sum of squares: squared deviations from group means.
3. **Between-group** sum of squares: squared deviations of the group means from the grand mean.
   - Each squared group mean is multiplied by the number of observations in the group.

Sums of Squares for ANOVA

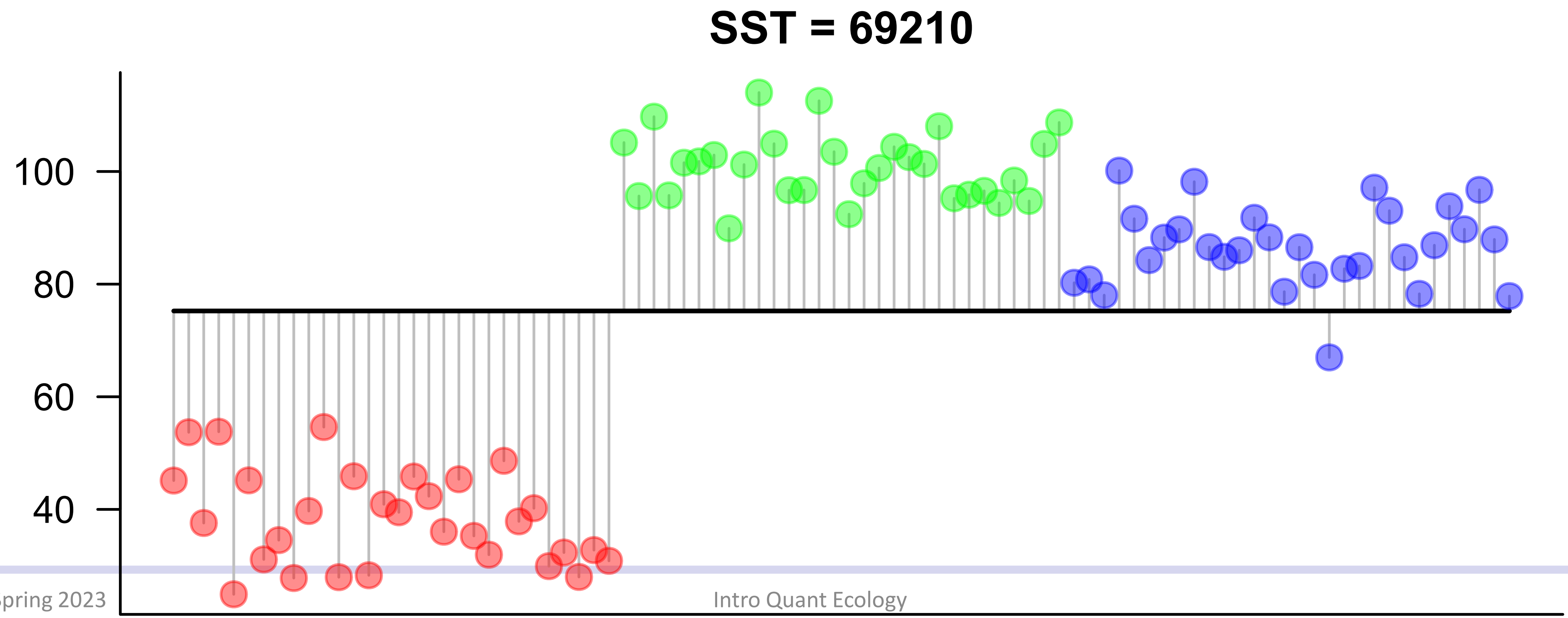
Procedure for calculating ANOVA sums of squares:

1. Calculate the grand mean of all observations.
2. Calculate the group means.
3. Sum the squared deviations of all observations from the grand mean. This is **SST**
4. Sum the squared deviations within groups. This is **SSW**
5. Sum the squared differences between the grand mean and each group mean. This is **SSB**

Total sum of squares: SS_T

Add up the squared differences from the overall mean

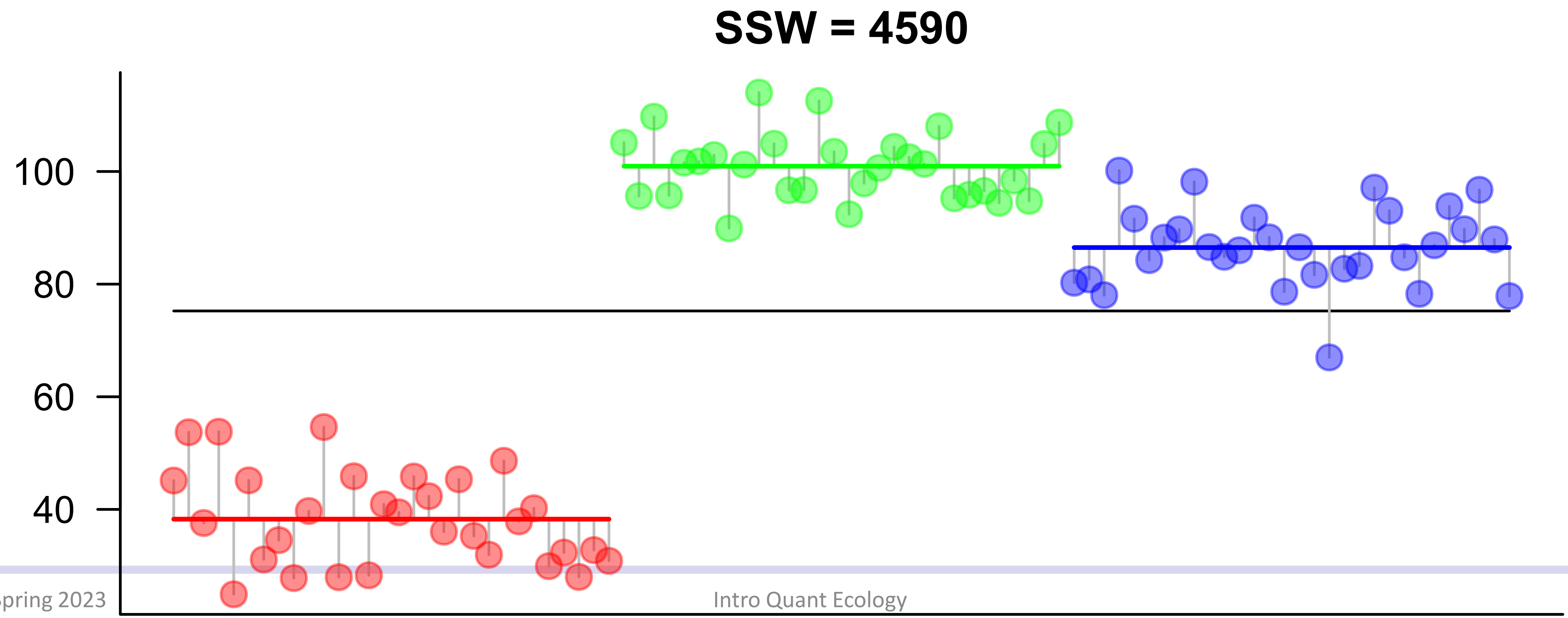
$$SS_T = \sum(x - \overline{x})^2$$



Within-group sum of squares: SS_W

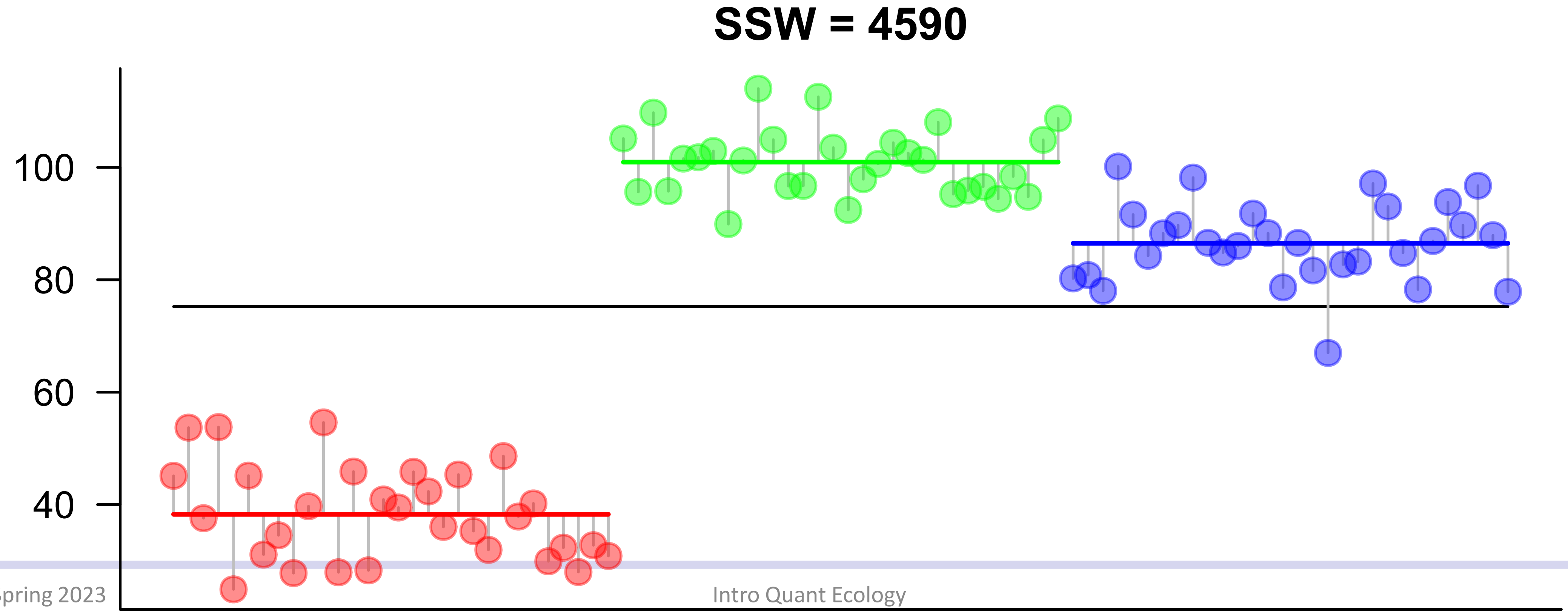
Add up the squared within-group differences from the group means:

$$SS_W = \sum(x_1 - \overline{x}_1)^2 + \sum(x_2 - \overline{x}_2)^2 + \sum(x_3 - \overline{x}_3)^2$$



Within-group sum of squares: SS_W

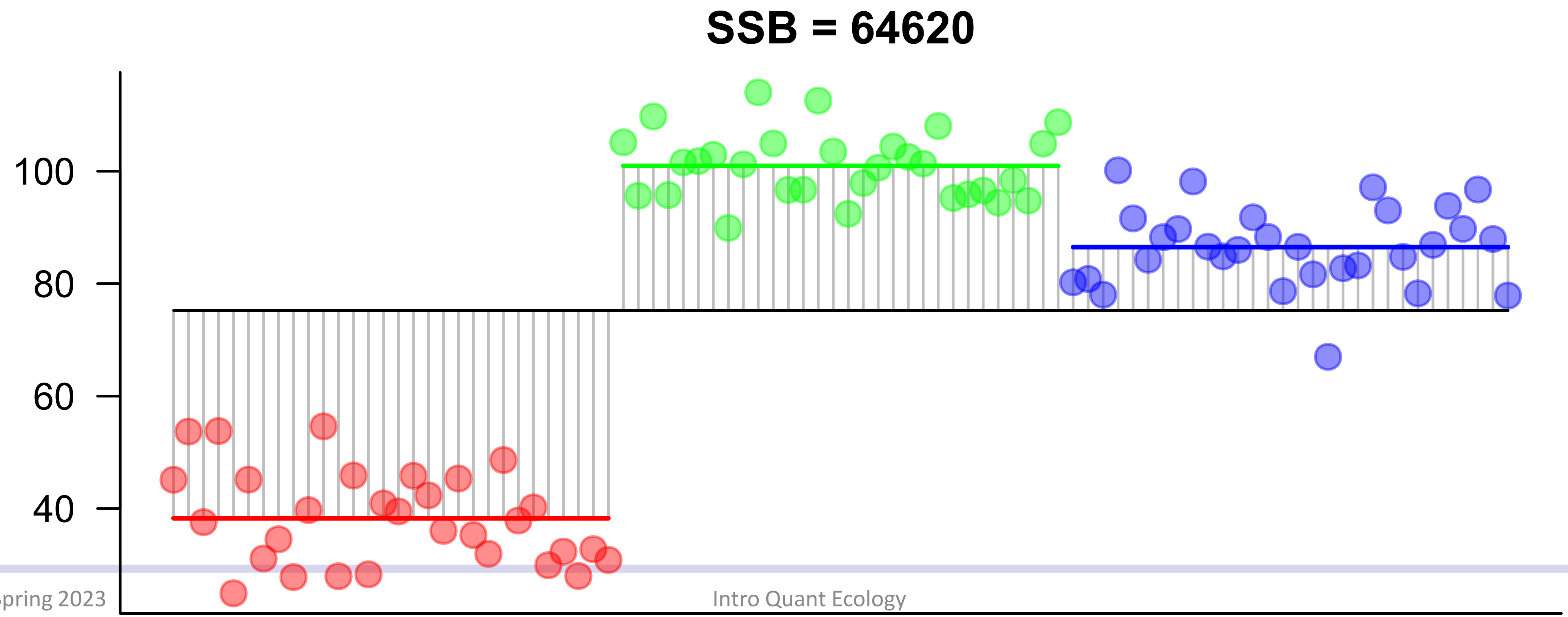
$$SS_W = \sum_g \sum_i (x_{ig} - \overline{x}_g)^2$$



Between-sample sum of squares: SS_B

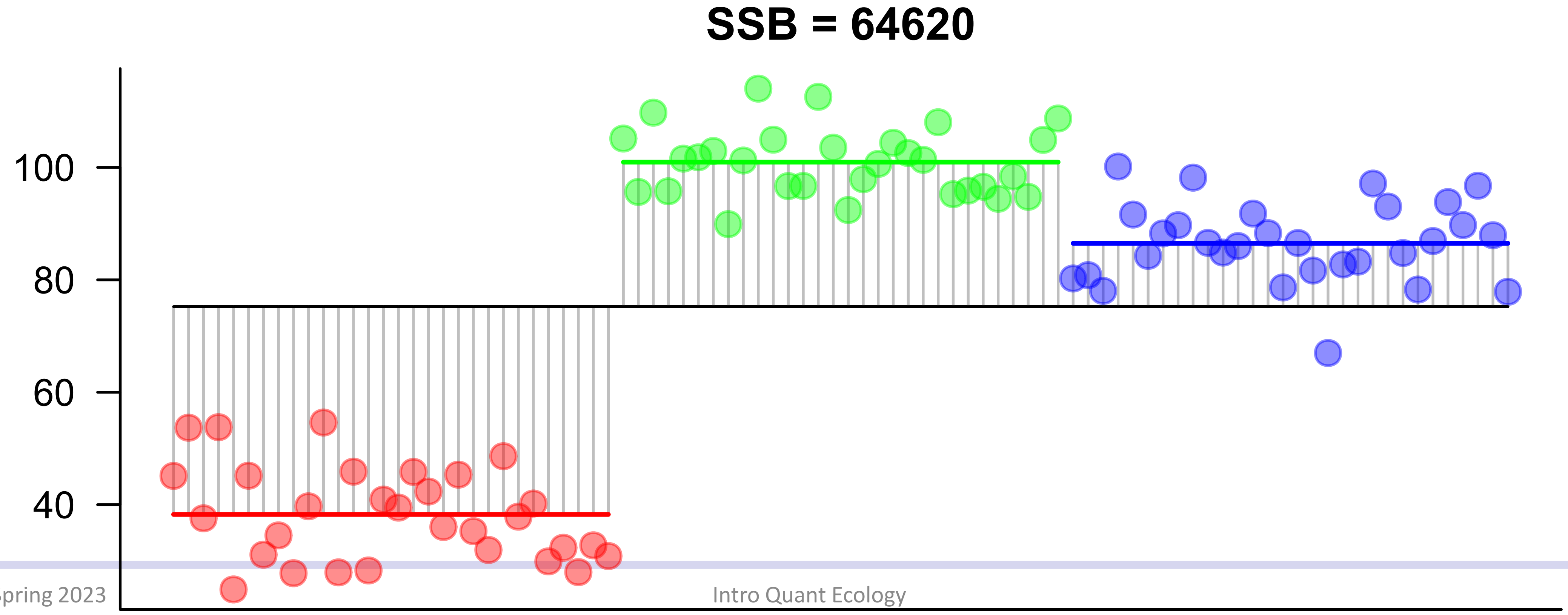
Add up squared differences of the group means $\bar{x}_g$ from the grand mean $\bar{x}$ and multiply by the number of observations in the group n_g:

$$n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 + n_3(\overline{x}_3 - \overline{x})^2$$



Between-sample sum of squares: SS_B

$$SS_B = \sum_g n_g \, (\bar{x}_g - \bar{x})^2$$

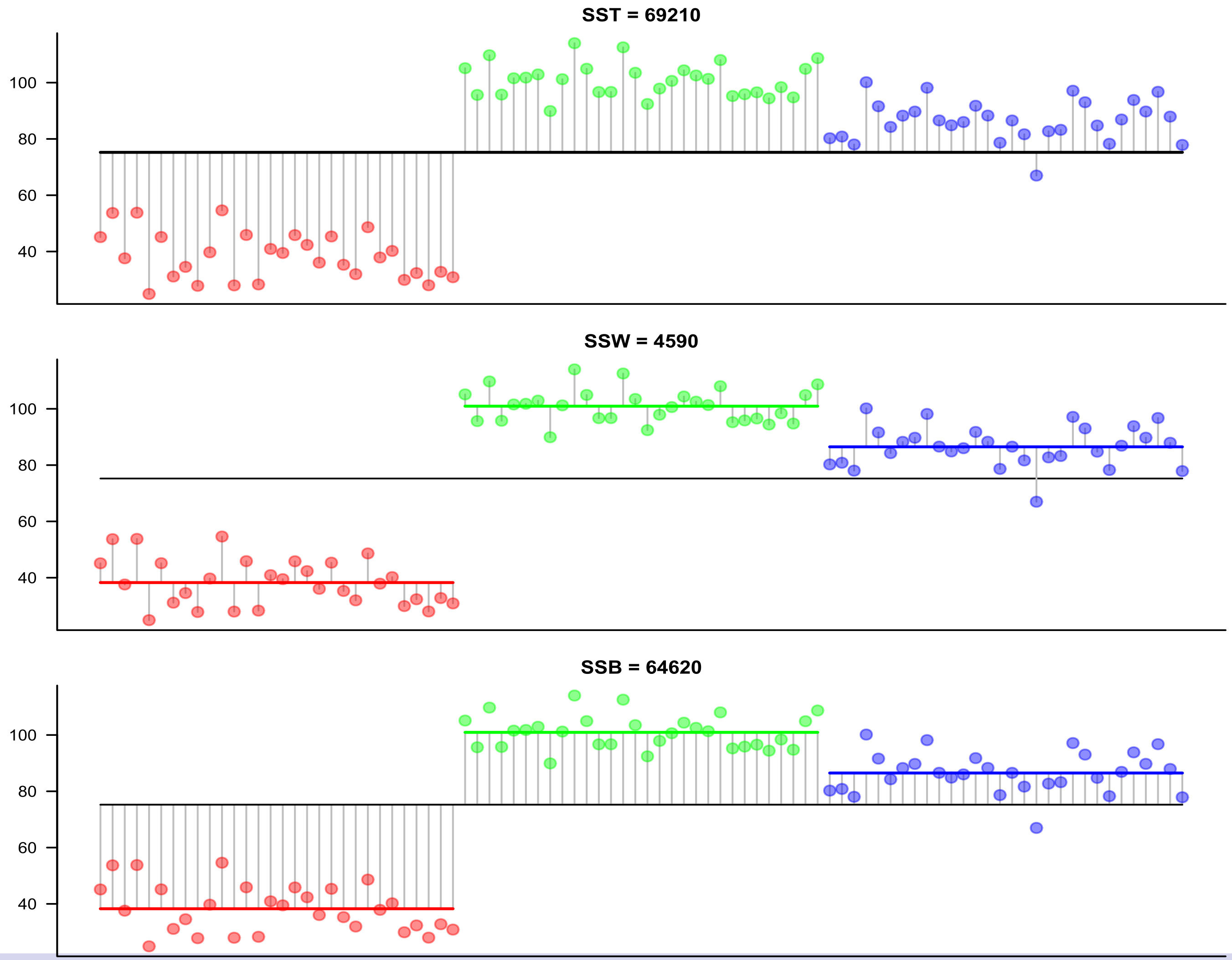


ANOVA Sums of Squares

$$SS_T = \sum (x - \overline{x})^2$$

$$SS_W = \sum_g \sum_i (x_{ig} - \overline{x}_g)^2$$

$$SS_B = \sum_g n_g \, (\overline{x}_g - \overline{x})^2$$



ANOVA degrees of freedom

Degrees of freedom are related to sample sizes.

- We generally subtract 1 from the sample size to get the degrees of freedom.

If we define the following:

- n is the total sample size (number of observations)
- g is the number of groups/samples

Then the degrees of freedom (df) are:

- Total: $df_T = n - 1$
- Within: $df_W = \mathrm{n} - \mathrm{g}$
- Between: $df_B = \mathrm{g} - 1$

ANOVA the *mean square*

The mean square (MS) is the sum of squares divided by the degrees of freedom:

$$MS = SS/df$$

So:

- Total: $MS_T = SS_T/df_T$
- Within: $MS_W = SS_W/df_W$
- Between: $MS_B = \frac{SS_B}{df_B}$

The Mean Squares are a *normalized* measure of the variability due to the grouping.

ANOVA: All the ingredients

Total

$SS_T = \sum(x - \overline{x})^2$

$df_T = n - 1$

Within-group

$$SS_W = \sum_g \sum_i (x_{ig} - \overline{x}_g)^2$$

$df_W = n - g$

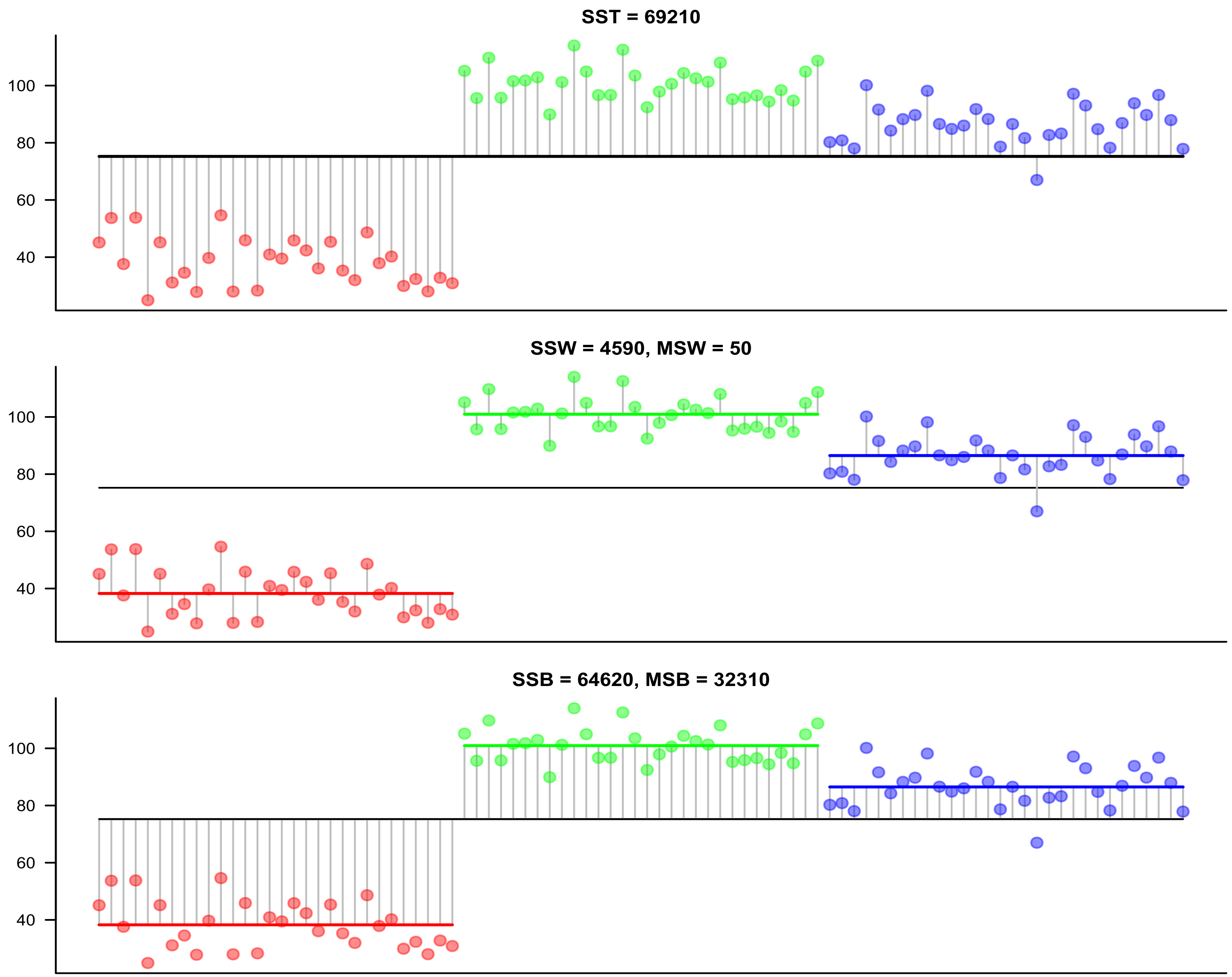
$MS_W = SS_W / df_W$

Between-group

$$SS_B = \sum_g n_g (\overline{x}_g - \overline{x})^2$$

$df_B = g - 1$

$MS_B = SS_B / df_B$



ANOVA Tables

ANOVA results are usually presented in a table

Source	SS	df	MS	F-value	p
Between-group	$\sum_g n_g (\overline{x}_g - \overline{x})^2$	$g - 1$	SS_B/df_B	MS_B/MS_W	
Within-group	$\sum_g \sum_i (x_{ig} - \overline{x}_g)^2$	$n - g$	SS_W/df_W		
Total	$\sum(x - \overline{x})^2$	$n - 1$	SS_T/df_T		

F is the test statistic for the ANOVA: $F = \frac{MS_B}{MS_W}$

ANOVA Tables

ANOVA results are usually presented in a table

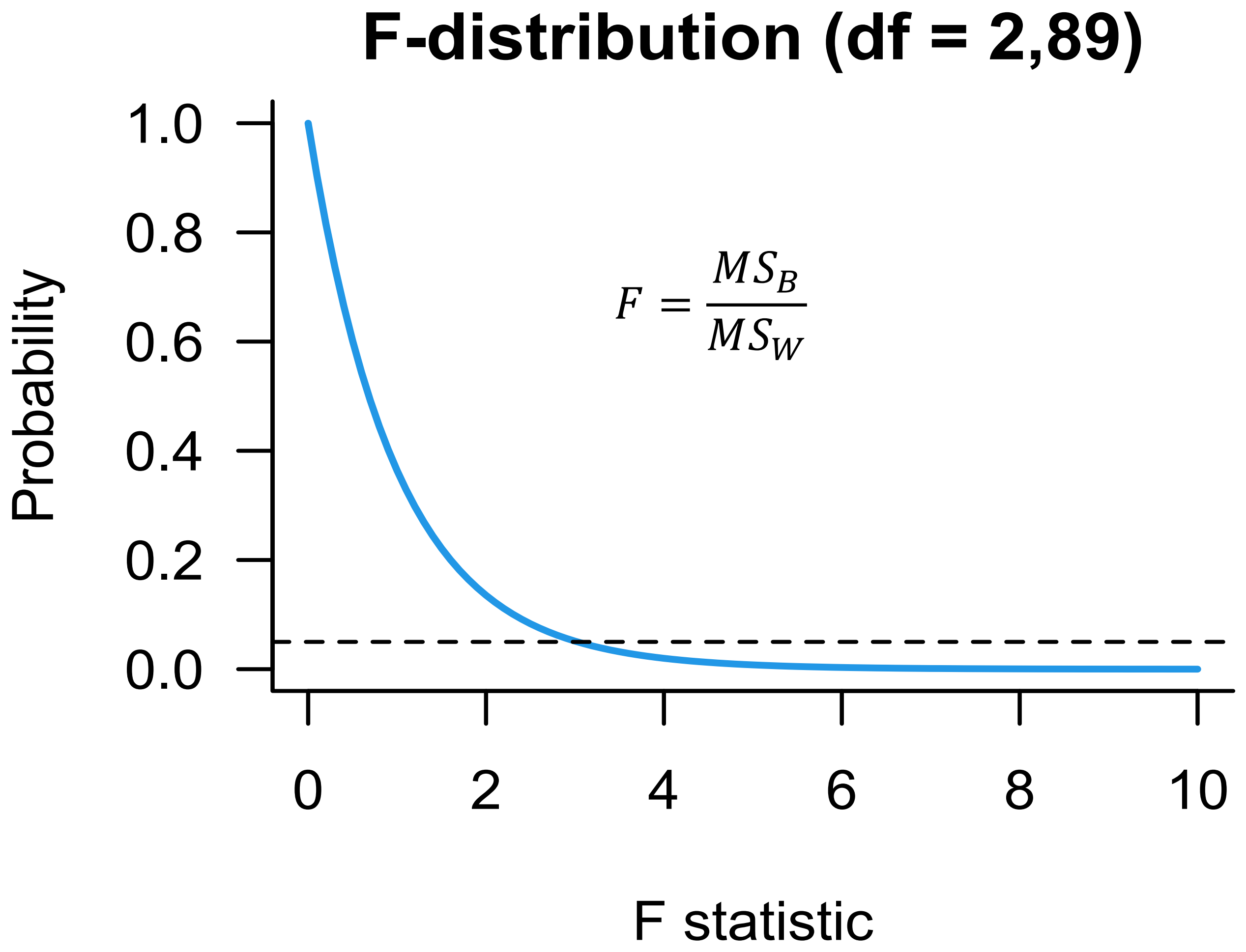
Source	SS	df	MS	F-value	p
Between-group	$\sum_g n_g (\overline{x}_g - \overline{x})^2$	$g - 1$	SS_B/df_B	MS_B/MS_W	
Within-group	$\sum_g \sum_i (x_{ig} - \overline{x}_g)^2$	$n - g$	SS_W/df_W		
Total	$\sum (x - \overline{x})^2$	$n - 1$	SS_T/df_T		

p is the probability of observing the F statistic with a given degrees of freedom if the null hypothesis were true: ‘no difference between the means’.
The calculation is of p is based on the F-distribution

ANOVA and the F distribution

The F distribution has 2 parameters:

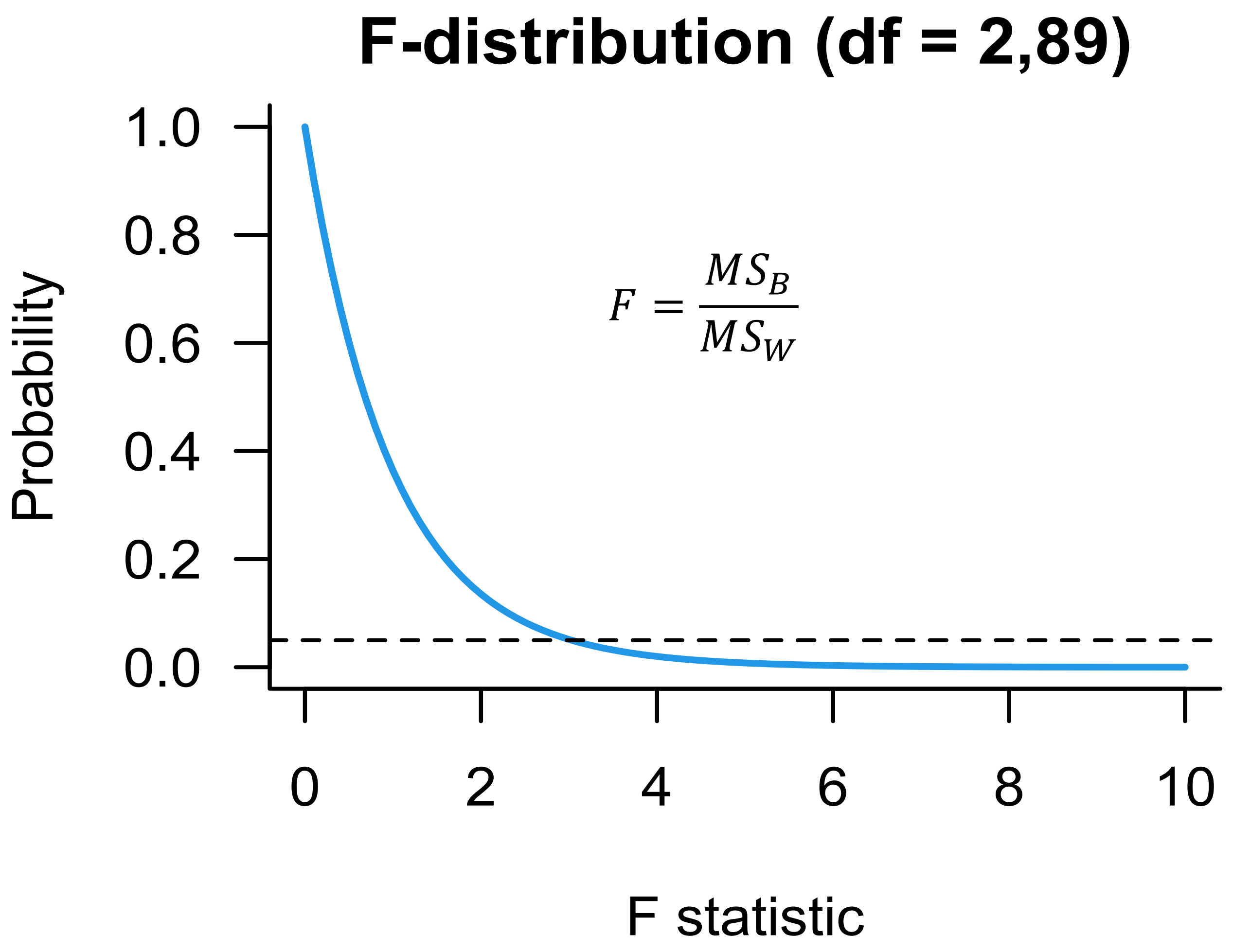
1. The numerator df
   Between group
2. The denominator df
   Within group



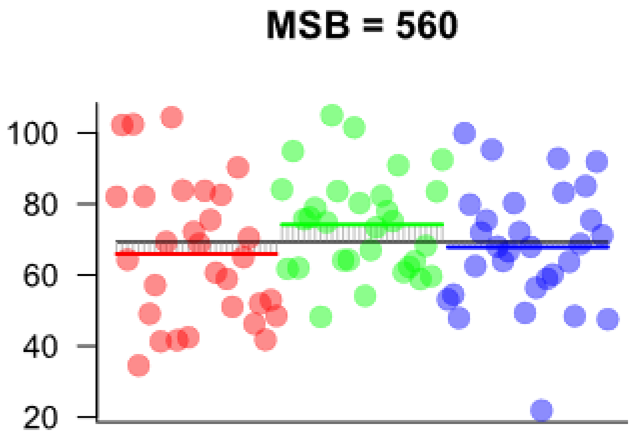
ANOVA and the F distribution

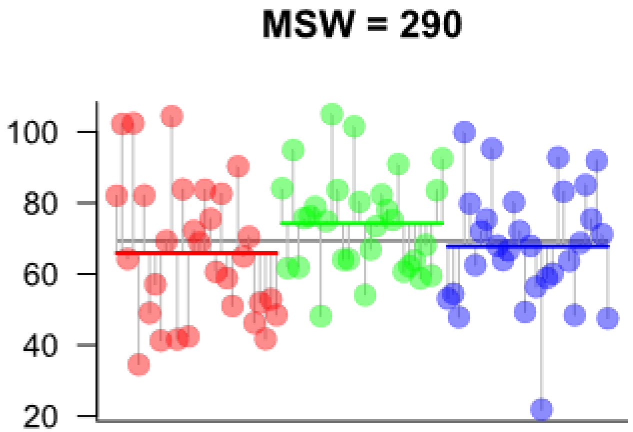
It's a ratio of the between-group noise and the leftover noise:

- How much of the variation is explained by the grouping factor?
- Did our grouping help explain the pattern in the data?



ANOVA and the F distribution - Null





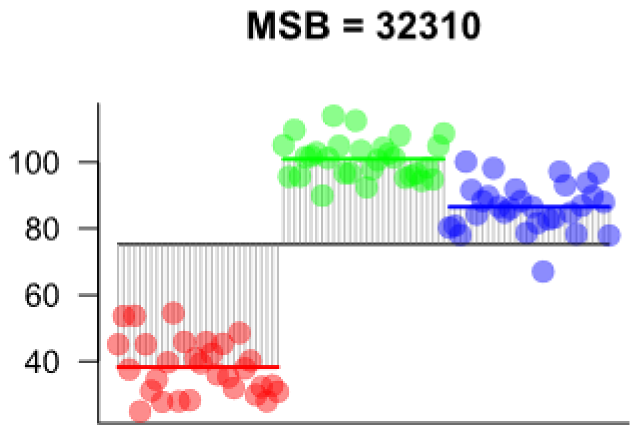
Small between-group differences:

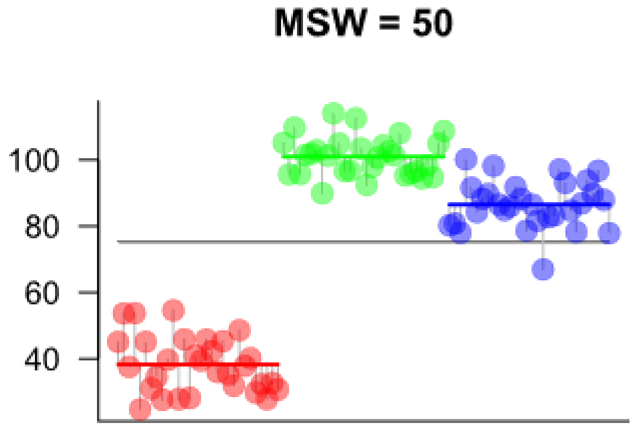
- Smaller F-value
- Higher p-value:
- Grouping did not explain much

$$F = \frac{MS_B}{MS_W} = \frac{560}{290} = 1.9$$

$$p = 0.15$$

ANOVA and the F distribution - Alternative





Large between-group differences:

- Big F-value
- Low p-value:
- Grouping explained a lot of the variation

$$F = \frac{MS_B}{MS_W} = \frac{32310}{50} = 646$$

$$p < [really\ low!]$$

Test statistics and p-values

- What’s a test statistic?
  - We’ve seen t, U, and now F
  - A value calculated from the data
  - Usually calculated using measures of center and spread
  - Allows us to quantify significance of differences.
  - Most of the time, large values for test statistics indicate something interesting.

Test statistics and p-values

- Test statistics and p-values
  - P-values are calculated using the value of a test statistic (and other information like degrees of freedom).
  - Most of the time low p-values are associated with high values for test statistics.
  - Low p-value means high significance.

Pairwise comparisons with ANOVA

The F statistic tells us whether there are differences, but *not* what the differences are.

We don't want to use many pairwise t-tests to make comparisons

- Some of the t-tests will return a significant result purely by chance
  - P values can be interpreted as the false-positive rate

Pairwise comparisons with ANOVA

The F statistic tells us whether there are group differences, but *not which groups are different.*

Instead, we conduct a *Post-hoc* test: Tukey Honest Significant Difference test (Tukey HSD)

- Not a turkey test!
- HSD helps us know which group means are different from one another.
- Significance groups

More than one factor with ANOVA

So far we have looked at multiple levels within a single factor

- factor: a single categorical predictor variable
- level: the categories within a factor

In some cases, we may be interested in >1 factor

- 2 factors: *two-way* ANOVA
- 3 factors: *three-way* ANOVA
- ⋯ multi-way ANOVA

ANOVA: Bird’s eye view

ANOVA and the F-test help us with:

- Global null and alternative hypotheses, global significance test
  - What does F tell us?
  - What can’t we learn from the F-test?
  - What does it mean to reject the null?

Notetaker Needed

Just a reminder that we're still looking for a notetaker:

"Disability Services is in need of a note taker for this class. If you are interested, please email notes@admin.umass.edu with your name, student ID and the course info (i.e. BIOLOGY 100, Section 2, Professor Jones). Disability Services staff will contact you to confirm and provide you instructions. You may earn 45 hours of community service for your efforts, or 1 pass/fail practicum credit."

ANOVA: Bird's eye view

One primary objective of an ANOVA is to quantify evidence that breaking up our observations into groups *improves* our description, relative to the null model that all data come from the same group.

- Null hypothesis is represented by the Total Sum of Squares: SS_T
- The alternative hypothesis is encapsulated by the relative magnitudes of the Within- and Between-group Sums of Squares: SS_W and SS_B
- What is our criterion for model improvement?

ANOVA Intuition

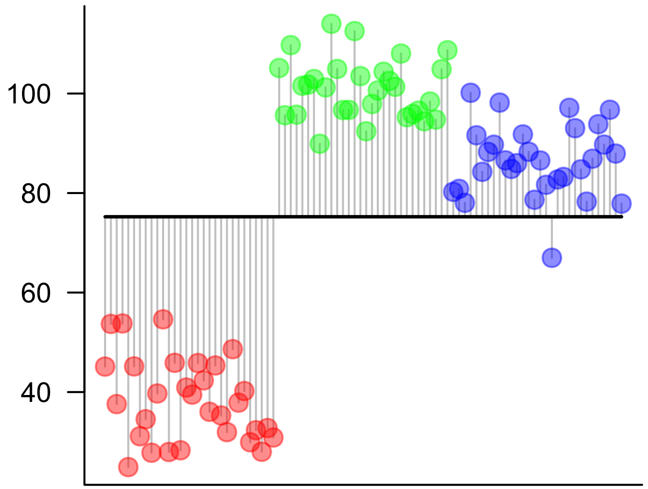
Sums of square terms are a way to quantify *variability.*

- Remember that the total SS is equal to the sum of the between- and within-group SS.
- The within- and between-group sums of square terms are calculated from different numbers of observations, so they are not directly comparable.
  - We ***normalize*** the sums of squares by their ***degrees of freedom***, which produces the mean squares.
  - This allows us to directly compare variability within- and between-groups.
  - What would we expect to observe if variability was similar within- and between-groups?

ANOVA Intuition

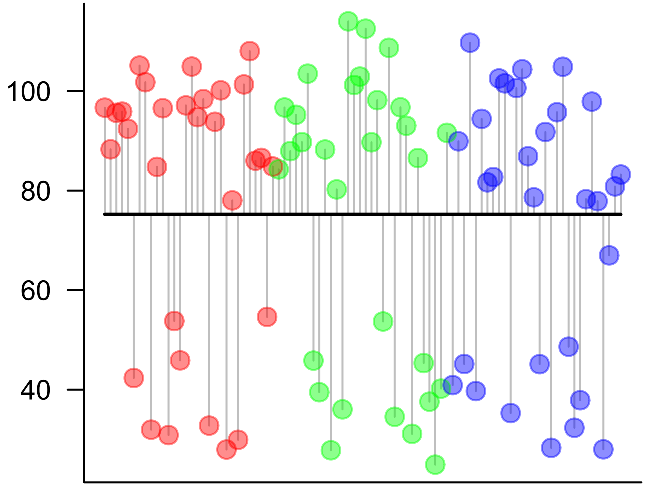
What if we randomly shuffled the group labels on our data?

- The new grouping probably *would not* improve our model.



This is another way to formulate the ANOVA null hypothesis:

"The group labels were selected at random."



Null Hypotheses: Model Thinking Perspective

A null hypothesis is…

- When we build a model, we like to consider *associations* among model components.
- We usually have some idea about the nature of the association:
- More water is associated with greater plant biomass, etc…

If the association were random…

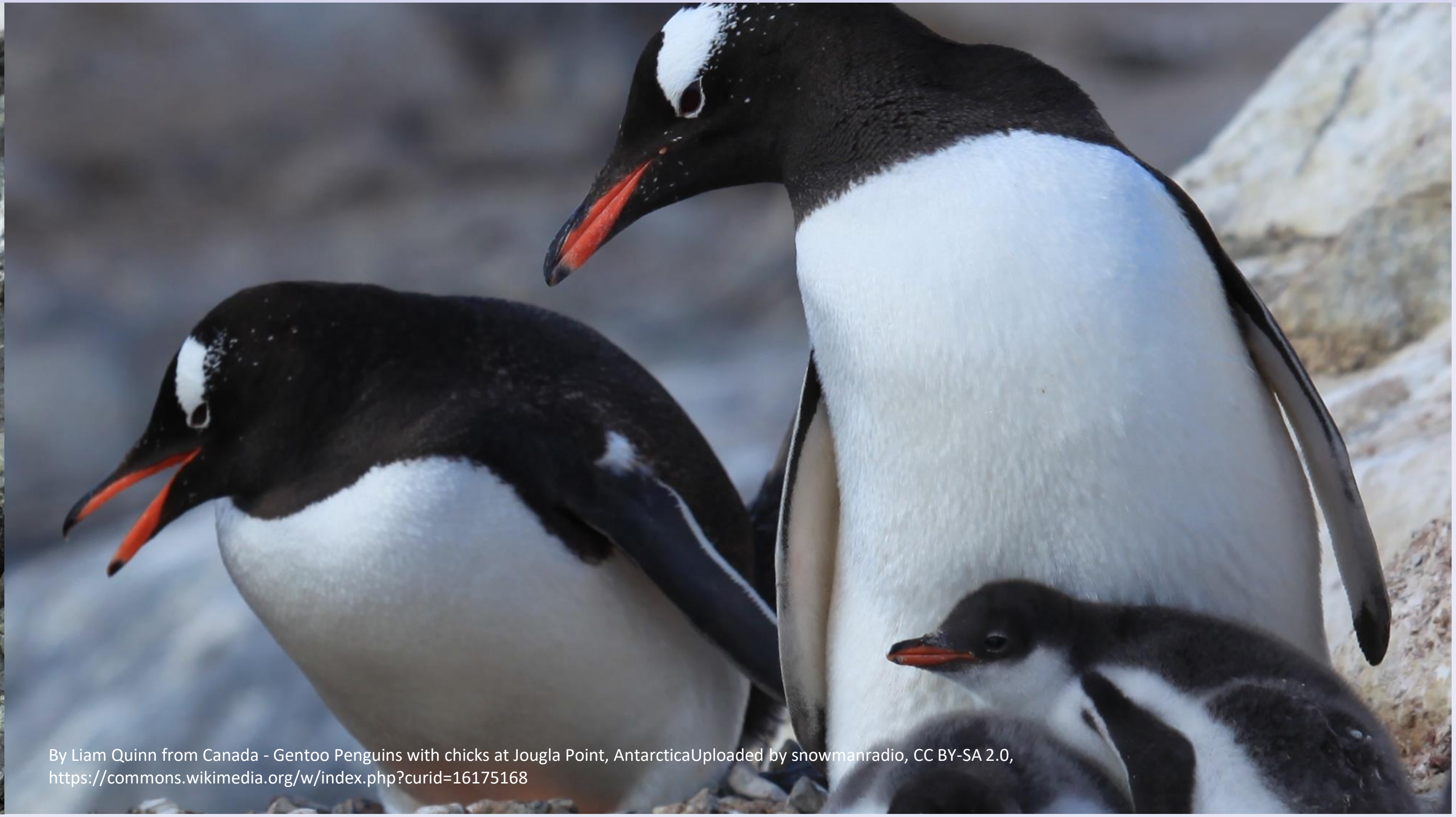
- In a null hypothesis, we usually that associations are totally random, i.e. quantities do not vary in a coordinated way.

Alternative Hypotheses

2-tailed hypotheses: non-directional	1-tailed hypotheses: directional
General: We don't propose a particular direction of the association. • Increased light might be associated with *greater* or *lower* biomass in an invasive plant. • We have no prior knowledge of its light response • We think that there will be *some* difference, but we don't' know the direction	More specific than 2-tailed hypotheses. • We have some previous knowledge that allows us to propose a *direction* to the association. • From our experience in the field and previous research we think that the plant will have *more* biomass in *higher* light. • We propose a *positive* association.

A Tale of Two Penguins

Adelie Penguins	Gentoo Penguins



A Null Hypothesis

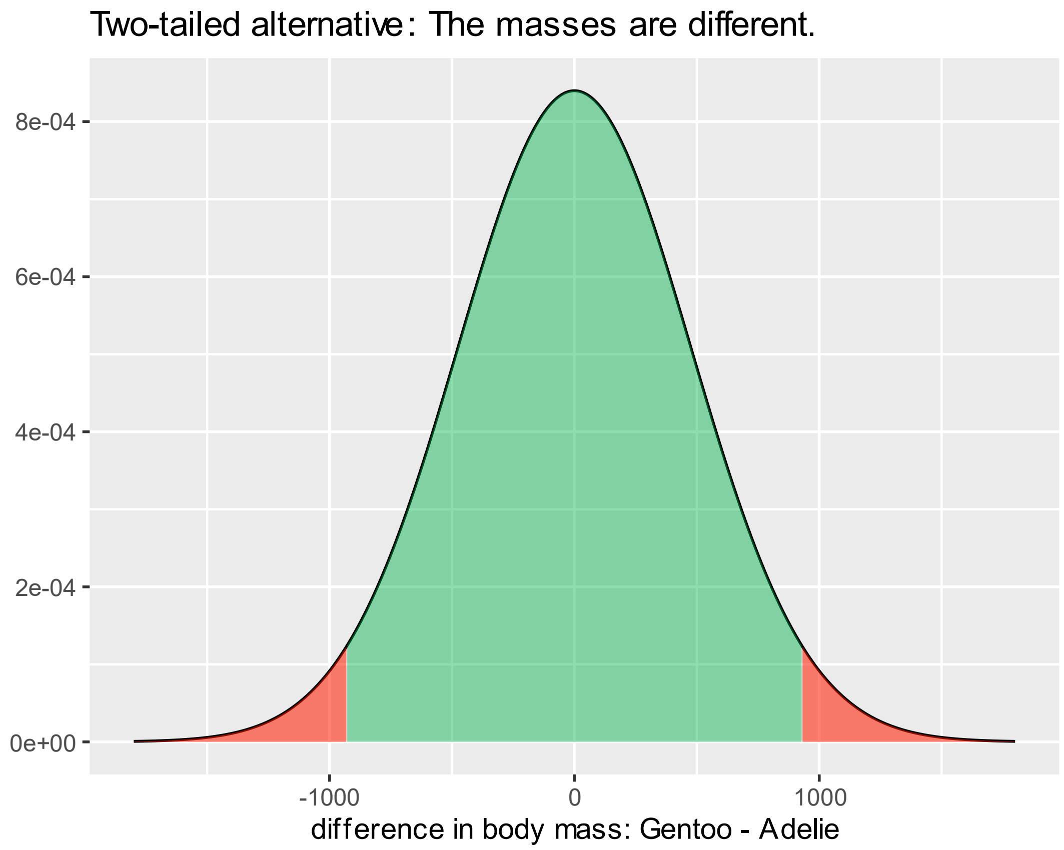
Question	Simple Null Hypothesis
• We want to know about differences between the two species.	Gentoo and Adelie penguins do not differ in body mass. A more technical, Frequentist formulation: • "The body masses of Adelie and Gentoo penguins are drawn from the same population of possible penguin body masses."

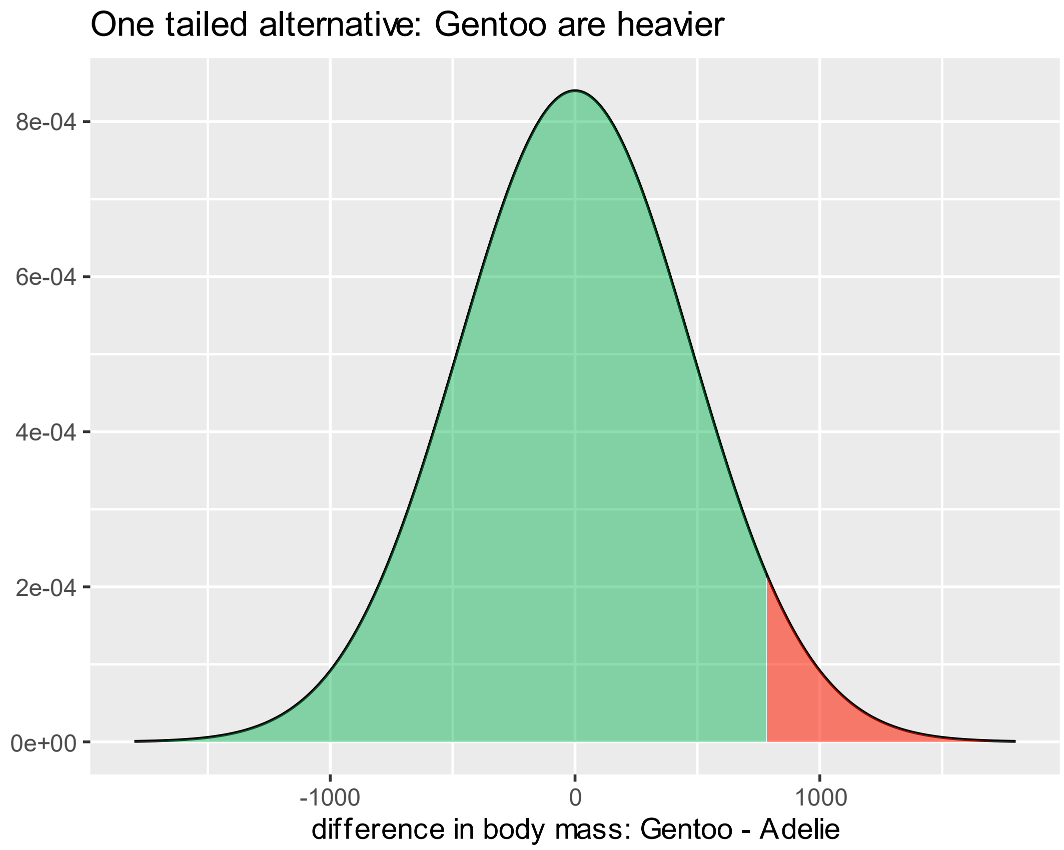
Alternative Hypotheses

2-tailed hypotheses: non-directional	**1-tailed hypotheses: directional**
General: We don’t propose a particular direction of the association. • We have no prior knowledge of the two species • We think that there will be *some* difference, but we don’t know which one is heavier • **“Gentoo and Adelie penguins have different body masses.”**	More specific than 2-tailed hypotheses. • We have some previous knowledge that allows us to propose a *direction* to the association. • From our experience in the field and previous research we think that Gentoo penguins are heavier • **“Gentoo penguins are heavier than Adelie penguins.”**

Alternative Hypotheses

2-tailed hypotheses: non-directional	1-tailed hypotheses: directional



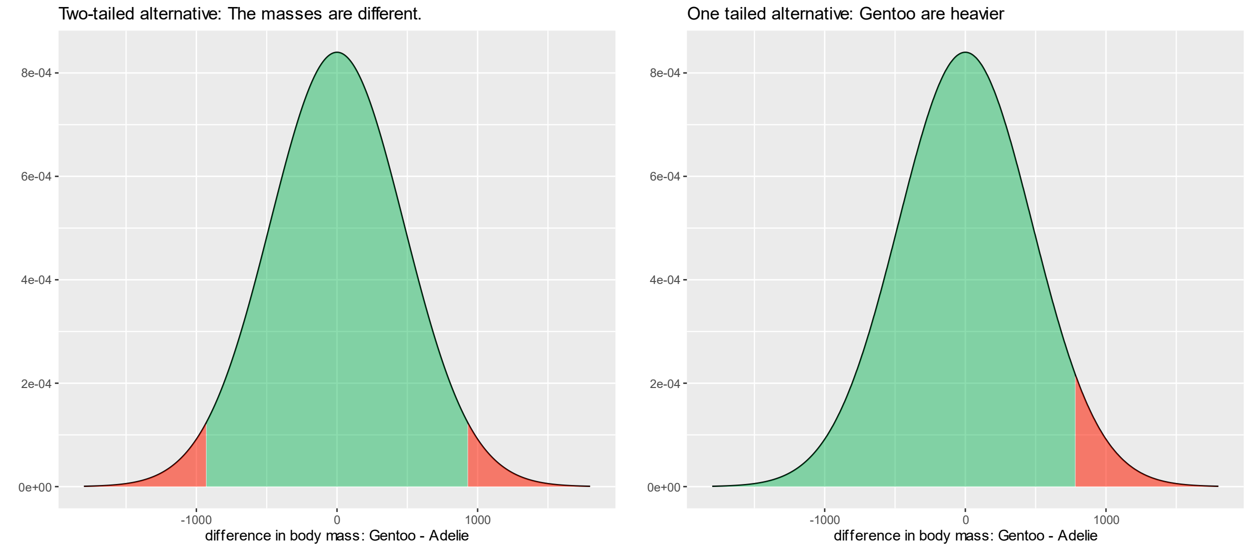


Decision criterion: tails and rejection regions

The area in red is α our Type I error rate, in this case 0.05.

1-tailed hypotheses can give us better *statistical power*.

- The tails define the *rejection region*, the region in which we say we have strong evidence to *reject the null hypothesis*.
- If the observed differences in masses falls in the rejection region, we can reject the null!



Some key points (not an exhaustive list)

- What's the difference between ANOVA and a t-test?
- What's a general hypothesis testing procedure?
  - Specify null and alternative hypotheses
  - Calculate test statistic (t, z, F, others)
  - Calculate the p-value using the test statistic and its probability distribution function
    - This is usually done for us in R, thankfully. We could do it by hand if we wanted to take a deeper dive into the inner workings of our tests.
  - Use p to evaluate the evidence against the null
- Squared deviations and sums of squares
- Normalization

Questions or demos?

- Remember to put your filenames in quotes!
- Remember your file extensions!
- read.csv produces a data.frame!

In-Class ANOVA

(Also finish in-class t-tests)