Interval Scales can have negative numbers

The zero can be moved, can put 0 at the boiling point of water instead of freezing point

Ratio Scales are (usually) non-negative

They have a meaningful zero, can only be all positive or all negative, never both

Circular scales wrap a maximum value back to zero

Circular scales are not as common, but they occur when thinking about angles  
 Exambles include wind direction and aspect  
 Both measured in degrees or radians

Modulo operator

The wrapping behavior of circular scales is just like addition in *modular algebra*

Sometimes it is convenient, or necessary, to convert a variable to a different scale

E.g. count data consisting mostly of 0 and 1, with only a few non-zero, non-one values, it may be useful to convert to binary data (presence/absence)

Aggregating into categories

Age, size, weight class

Destructive conversions: some information is lost in the conversion process. If you switch from exact ages to age classes, you can’t go back to exact ages from classes. Always keep your original data, always strive to acquire exact data, even if working with a categorical scale.

Theoretical (parametric) distributions as models

Remember that we fit models to data, not the other way around  
 Sometimes theoretical distributions are only *approximately* good models for our data  
 E.g. the Normal distribution often fits data, such as weights, very well, but its sample space includes negative numbers

The Row Data Paradigm

|  |  |
| --- | --- |
| Rows | Columns |
| Rows are observations | Columns represent attributes |
| Rows are samples | Columns are variables |
| Rows are sampling units (sometimes) | Columns are properties |
| A row is a collection of observations on a single entity | Columns are fields |
| Rows are |  |

Sample and Population

Populations are **large**

We typically cannot observe all individuals in a population (cornerstone of Frequentist thinking)

Samples are **subset** of the population

We can observe all individuals in a sample and completely characterize the properties of a sample

We use the sample to make informed guesses about the population

Populations have *parameters*, intrinsic characteristics of the entire population, which cannot be calculated directly

We can calculate *statistics* from samples

A Sampling Unit (SU) is the unit/entity/thing of interest for the research question

A variable is an attribute of the SU

Ecological population = the entire possible population of the organism  
A statistical population is (usually) a subset of an ecological population  
A sample is a subset of the statistical population  
A sampling unit is a subset of a sample  
A variable is a quantity measured on a single SU

Suppose we were studying bullhead in a single lake:  
• ecological population: entire species range  
• statistical population: the lake

What about bullhead in Massachusetts?  
• ecological population: entire species range  
• statistical population: all bullhead within MA

Note that the ecological population did not change.

Inferential Framework

Most widely used framework  
 Requires assumptions  
 Many tools

Key assumptions

Population exists, is infinite  
 Population parameters are true, but unknowable quantities  
 When we specify a model, there exist **true** model parameters (but they are unknowable)

Frequentism is based upon hypothetical infinite resampling

Frequentist assumptions are often asymptotically true  
 Source of misconceptions about terminology

Hypothesis testing: allows for quantification of *confidence* and *significance*

CI = Confidence Interval

Confidence does not mean “I’m 95% sure my CI contains the true value”

What it does mean is “If I were to repeat the experiment *ad infinitum*, approx. 95% of the CI’s I construct would contain the true population parameter

Null model: This is what we should see if there is no relationship between x and y  
Alternative model: This is what we want to observe if there is a relationship between x and y  
False positive: Sometimes the null is true, but by chance we observe a pattern akin to an alternative model  
False negative: Sometimes there is a true relationship, but by random chance we observe a patternakin to a null model