SSA 200
Strategic Use of Data

Lecture slides, activities, and additional supplementary materials are available online at:
ssa200.auburn.edu

What is a model?

The purpose of modeling

- Statistical analysis of data
- Use statistical analysis to predict the future
- Explaining variation
- Using data analysis to understand ecological processes
- Predict patterns in the future
- Evaluate competing hypothesis about how the system works

Conor P. McGowan
Anna M. Tucker
Nicole F. Angeli
Kylee Dunham

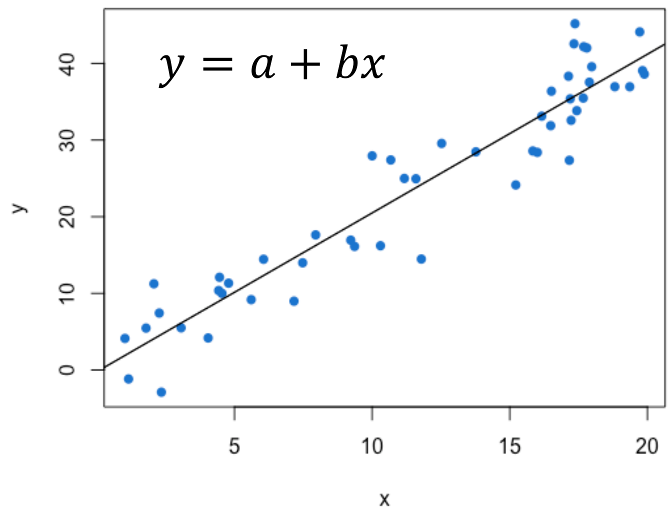
Statistical distributions

Name	Continuous or Discrete	Bounds	Common applications	Shape	Notes
Normal	Continuous	-∞, ∞	Linear regression		
Binomial	Discrete	0 or 1	Occupancy Survival		
Multinomial	Discrete	0, ∞	State transitions		
Poisson	Discrete	0, ∞	Count data		
Negative Binomial	Discrete	0, ∞	Counts with many zeros		
Log-normal	Continuous	0, ∞	Population-level productivity (projections)		
Beta	Continuous	0, 1	Population rates (projections)		
Uniform	Continuous	User-defined	Variety of applications (projections)		

Linear regression and AIC

General linear model – response variable (y) has a Normal distribution

Generalized linear model – response variable (y) has some other distribution
- **Logistic** regression – Binomial distribution
- **Poisson** regression – Poisson distribution



Parameter	Estimate	SE	*t*	p-value
Intercept	-0.22	1.29	-0.169	0.866
b	2.07	0.0978	21.16	< 0.0001

Model	AIC	ΔAIC	Np	w_i
Int + Covariate 1 + Covariate 3	345.8	0	3	0.82
Int + Covariate 1 + Covariate 2 + Covariate 3	349.1	3.1	4	0.18
Int	359.8	14.0	2	0
Int + Covariate 2	361.1	15.3	1	0

Types of uncertainty

Partial controllability – We are unable to control the exact management actions taken in a system.

Examples:
- Setting management goals – we may intend to fully restore a habitat, but may not be able to implement the exact management goals due to other logistical constraints

Observational uncertainty – We are unable to perfectly observe the state of natural systems.

Examples:
- Count data – in almost all cases, we cannot count every individual present at a specific location, but instead assume there is some probability of detecting individuals

Environmental variation – Stochastic environmental fluctuations mean that conditions typically vary randomly from year to year.

Examples:
- Predicting effects of temperature – we may estimate a relationship between temperature and survival probability that we can use to predict survival under future temperature conditions, but temperature will likely vary in a stochastic way from year to year.

Ecological uncertainty – We have an imperfect understanding of how ecological systems work.

Examples:
- Metapopulation dynamics – we think a set of populations function as a metapopulation, but have not conducted studies to explicitly estimate immigration among sites, and therefore we are unsure to what extent immigration plays a role in measured population growth rate at each site.

Some key terms

Response/dependent variable – in a statistical model, the variable that you are interested in better understanding or predicting (the "y" variable)

Predictor/independent variable – in a statistical model, the variable(s) that explain some of the observed variation in the response variable (the "x" variables)

Covariate – an environmental or ecological quantity that usually represents a stressor or species need and is included in a model as a predictor variable

Parameter – statistical quantities that are estimated to explain the relationship between predictor and response variables. Can also be used to refer to demographic vital rates of interest

Collinearity – occurs when two predictor variables in the same model are correlated with each other

Overfitting – occurs when too many predictor variables are included in the model, resulting in a model that is not very useful for prediction

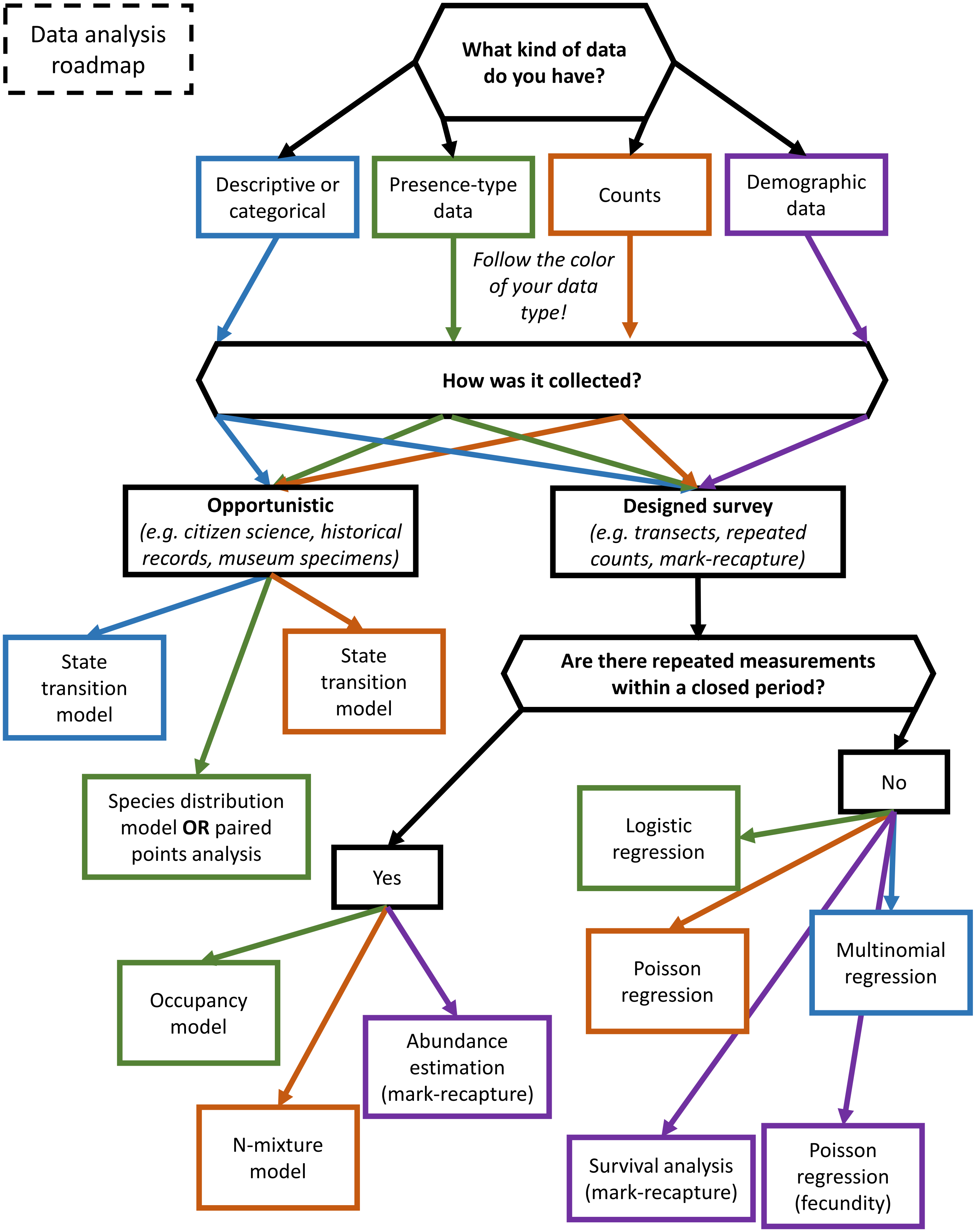
AIC – stands for Aikaike's Information Criterion – a metric used to rank models based on how well they fit the data with a penalty for the number of covariates in the model (to avoid overfitting)

Intercept – the theoretical value of the response variable if all predictors were equal to zero

Null model – the "intercept-only" model that does not include any covariates

Global model – the most complex model in the model set that includes all covariates

Population closure – an important concept for occupancy and abundance estimation, a population is considered "closed" when there are no births, deaths, immigration, or emigration



This roadmap is to serve as a general guide and is not an exhaustive list of all analysis options. Also, ***always check the specific assumptions of your planned modeling approach!***