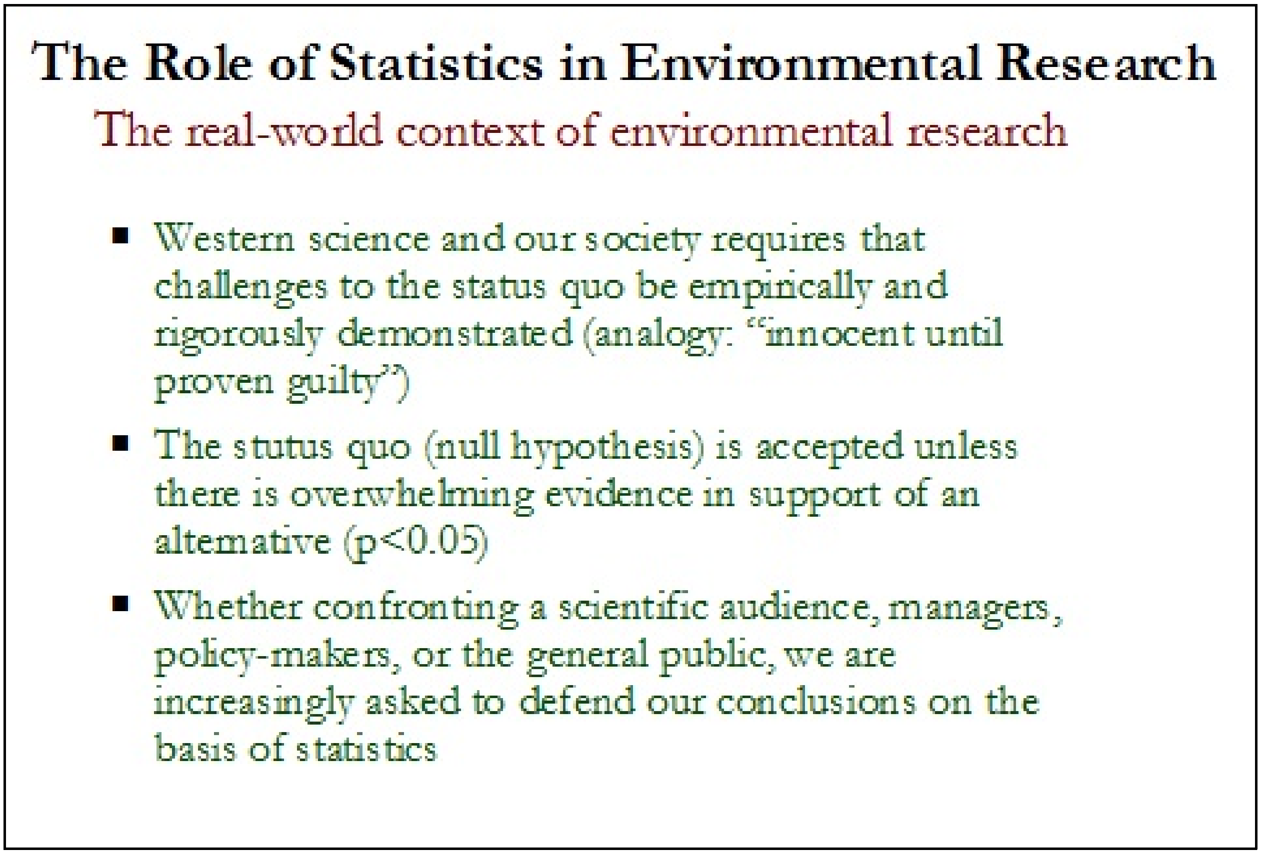
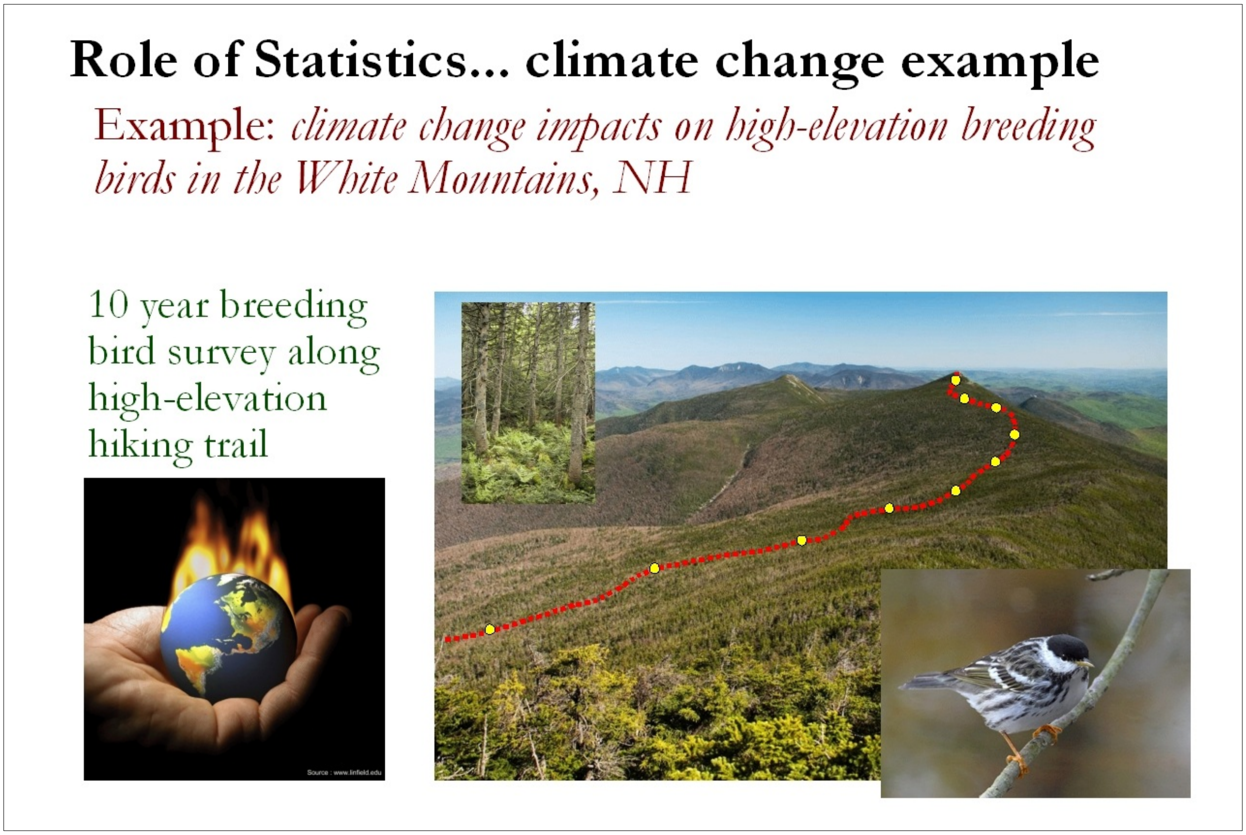
Analysis of Environmental Data
Chapter 1. Conceptual Foundations:
The Role of Statistics in Environmental Research



1. The real world context of environmental research

As scientists or managers we are faced with the challenges of defending our decisions every day. This is because we live in a world where challenges to the status quo are given little credence without solid evidence for the alternative (analogy: innocent until proven guilty). In other words, if we want to suggest a change to the current way of thinking, the burden of proof is on us. This idea lies at the core of the western scientific method. The classical approach to statistical inference, for example, involves testing hypotheses in which the null hypothesis (i.e., the status quo) is accepted unless there is overwhelming evidence in support of the alternative hypothesis – the one you are proposing. Indeed, we take great measures to ensure that we don't falsely reject the null hypothesis in favor of the alternative (i.e., a Type I error); this is why by conventional standard we set the *p*-value for rejecting the null hypothesis to be a very small number, usually 0.05, which means that we will continue to support the null hypothesis unless the data suggest that there is less than a 5% chance that we would have observed our data if in fact the null hypothesis were true. Moreover, it is not sufficient to simply claim strong support for the alternative hypothesis; support for the alternative must be demonstrated quantitatively and shown to be overwhelmingly more likely than the null hypothesis based on data. Whether we are presenting our findings to a scientific audience (e.g., in a scientific journal) or to managers, policy-makers, or the general public, we are increasingly asked to defend our conclusions on the basis of statistical evidence.



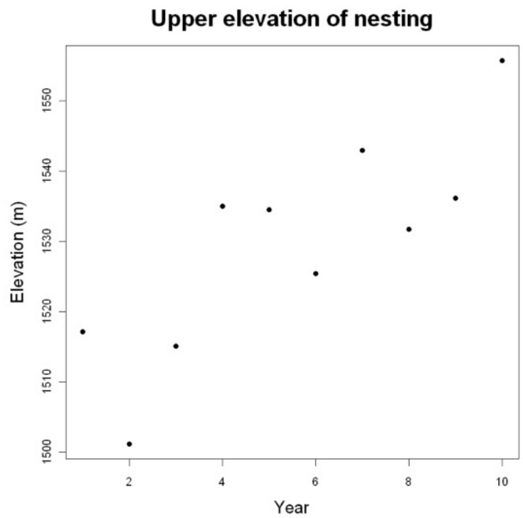
An example: The climate change debate

Consider the role of environmental statistics in the climate change debate. Imagine that you have been conducting a breeding bird survey of a single mountain top in the White Mountains of New Hampshire every year for the past 10 years and you have noticed that the upper elevational distribution of high-elevation bird species associated with spruce-fir forest appears to be getting higher; i.e., they seem to be shifting their distribution up in elevation, perhaps by as much as 50m over the past ten years.

Role of Statistics... climate change example

Testimony #1

- Observed increase in upper elevational limit of nesting along an elevational gradient

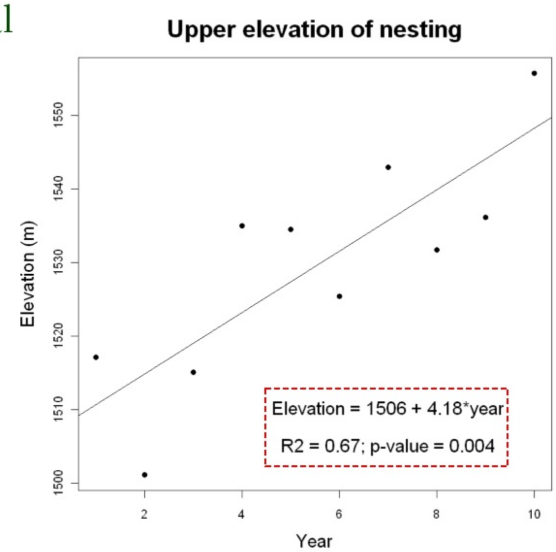


Hearing of your research and your expertise, you are asked to testify before a committee of scientists on climate change impacts. In your testimony (#1), you show a scatter plot (figure) of the relationship you have observed and make the claim that species are nesting higher in elevation in response to climate change. How do you think the committee will respond to your testimony?

Role of Statistics... climate change example

Testimony #2

- Statistically significant increase in upper elevational limit of nesting along an elevational gradient



Having a learned a painful lesson, you return to your lab and construct a statistical model that proposes a linear increase in the upper elevational distribution of several focal species, and then you confront the model with your data. Your results indicate a statistically significant linear trend. You return to testify (#2) at the next hearing and present your findings (figure), this time claiming that you have strong empirical support that the breeding bird community has been affected by climate change. As an example, you estimate that the upper elevational limit of nesting blackpoll warblers has been increasing by approximately 4 m/year. What do you think will be the response to your testimony this time?

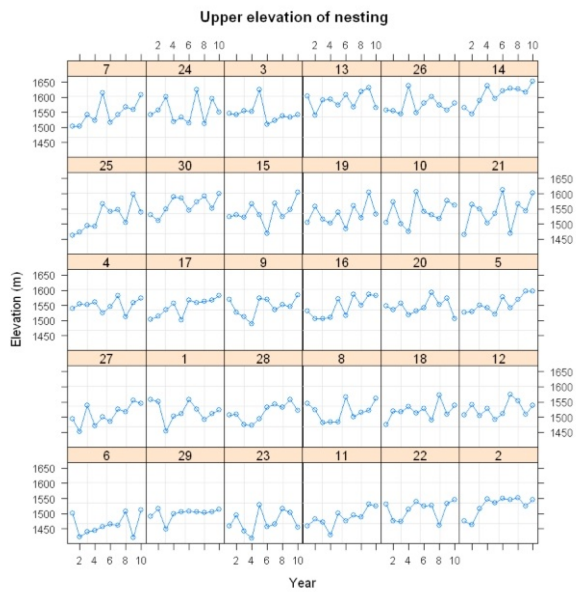
Role of Statistics… climate change example

Testimony #3

- Statistically significant increase in upper elevational limit based on 30 survey routes across the Forest along elevational gradients

```
Linear mixed model fit by maximum likelihood
Formula: elev ~ tsteps + (1 + tsteps | plot.id)
   Data: y.sim
  AIC  BIC logLik deviance REMLdev
 2970 2992  -1479     2958    2952
Random effects:
 Groups   Name        Variance   Std.Dev. Corr
 plot.id  (Intercept) 748.643730 27.36135
          tsteps        0.067454  0.25972 1.000
 Residual             887.300857 29.78760
Number of obs: 300, groups: plot.id, 30

Fixed effects:
             Estimate Std. Error t value
(Intercept) 1508.6996     6.2255  242.34
tsteps         4.4981     0.6006    7.49
```



Having learned another painful lesson, you return to your lab and decide to expand the geographic scope of the analysis in order to obtain replication of your observations; i.e., increase the sample size. You do some investigating and discover that the White Mountains National Forest has been conducting similar surveys throughout the Forest for the past 10 years, with approximately 30 survey routes on mountain tops distributed randomly across the Forest. Bonanza! The Forest agrees to share their data and you eagerly conduct a new statistical analysis, this time with replication of survey routes. You end up constructing a linear mixed effects model with varying intercept and slope, treating the survey route as a "random effect" and time as a "fixed effect", and allowing the intercept and slope of the trend to vary across routes. Sure enough, your earlier finding from a single survey route was nicely supported by the replicated data. Charged with these new findings, you return to testify (#3) at the next hearing, this time confident that you will be able to make a convincing case for climate impacts. What do you think the response will be this time?

Role of Statistics... climate change example

Testimony #4

- Weight of evidence strongly suggests climate as the major culprit, but habitat may play a role as well

	AIC	df	dAIC	weight
climate	2578.5	6	0.0	0.799
habitat	2581.3	6	2.8	0.201
recreation	2603.9	6	25.4	<0.001
deposition	2631.0	6	52.5	<0.001

Having suffered yet another humiliating defeat by the climate change skeptics, you are ready to give up your aspiring career as a scientist and take up farming. Fortunately, instead, your ego convinces you not to give up. You decide to sit in on the Analysis of Environmental Data stats class to see if there are any ideas that might help you with your case. You discover that your case can be substantially improved by considering a broad range of alternative or competing models that might also plausibly explain the observed patterns and let the data indicate the strength of evidence in support of each hypothesis. For example, you propose that vegetation changes caused by a recent spruce budworm outbreak might explain the shift to higher elevations in some species. After careful thought and review of the scientific literature, you end up with four alternative hypotheses to explain the elevational shift, all based on plausible mechanisms behind the observed shifts in distribution. You confront each of these hypotheses with the data and find that the climate change hypothesis has the strongest support based on a widely used Information Criterion known as AIC. Only one other hypothesis has weak support given the data, while the other two hypotheses can effectively be discarded as having no support in the data. You return to testify (#4) at the next hearing, where you now humbly present your findings. Based on the suite of plausible alternative hypotheses considered, you contend that there is very strong support for a real climate change impact on the elevational shift in the upper elevational distribution of breeding birds, but that you can't rule out the possibility of habitat-induced changes caused by the recent spruce budworm outbreak. In addition, although not shown, you also present confidence intervals for your estimate of the slope of the trend line. What do you think the response will be this time?

Role of Statistics... climate change example

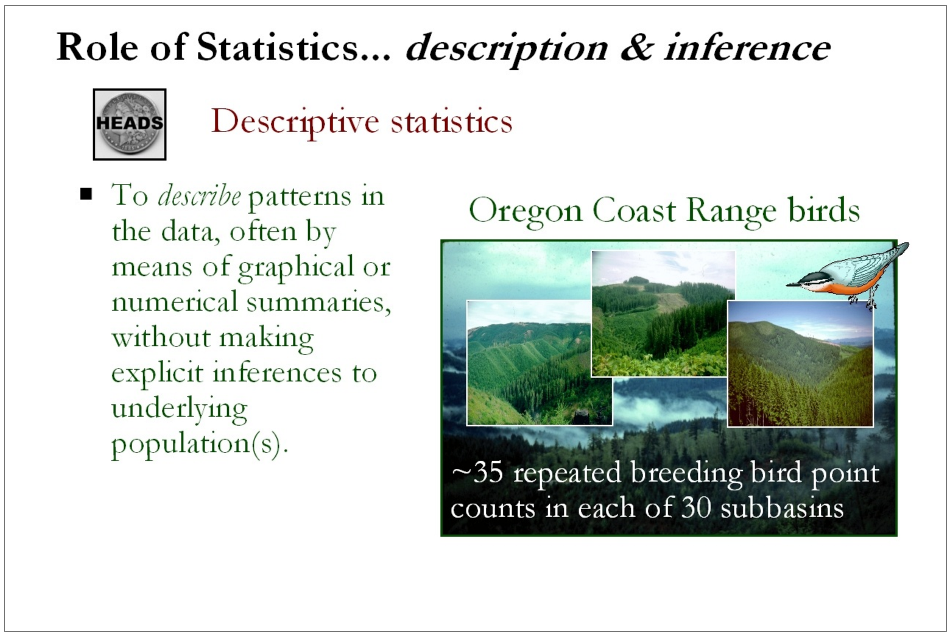
EUREKA! Committee funds $1,000,000 for an intensified research project for 10 years

Eureka! After careful deliberation, the task force (with one dissenting opinion - wonder who?) decides that your findings this time are credible and warrant careful consideration. However, one particularly belligerent committee member (can you guess which one?) notes that a 10-50 m shift in elevation, while perhaps real, doesn't seem like a very big deal biologically. You point out that while the 10-50 m elevational shift over the past 10 years may not have had a significant biological impact, if the trend were to continue, the shift would end up being much greater over the next 100 years and it could therefore have a significant biological impact. However, to understand the biological impact, it would be necessary to estimate population densities (instead of relative abundance) and fitness, and monitor the lower and upper elevational distribution of the focal species. To do this would require an intensified sampling effort that would allow you to estimate the detectability of each species, which would be necessary to estimate population densities and productivity. Specifically, instead of a single survey each year, you would need multiple surveys, and they would need to extend the survey routes along the full elevational gradient. After some discussion, the task force agrees to fund your research program at the tune of $100k per year for the next 10 years.

What are the lessons learned from this scenario?

2. The role of statistics

It should be clear even from the single scenario above that statistics can play a major role in environmental research and management. But what exactly are these roles? How do we use statistics in ecology and natural resources management? For purposes of organization, it is useful to consider statistics as having two major overarching purposes: (1) to describe patterns and (2) to make inferences. This dichotomy is based as much on the research and management context in which statistic are used as on the methods themselves. And while it is useful to distinguish descriptive from inferential statistics, this distinction is not always clear cut since the same statistical procedure can in one circumstance be used in a descriptive manner and in another circumstance be used as a means of conducting inference.



2.1 Description

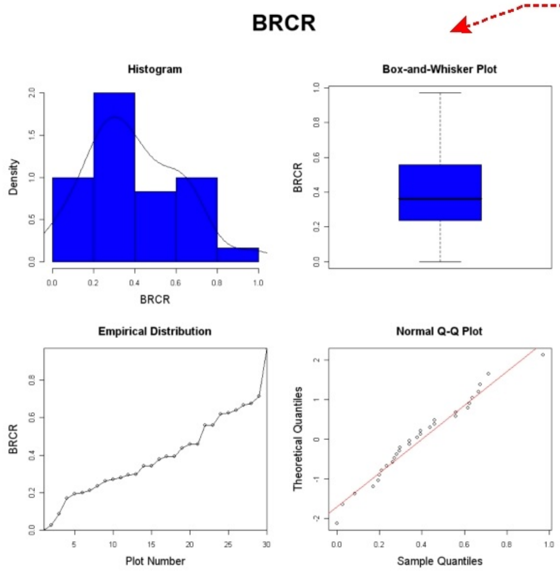
Definition.–To *describe* patterns in the data, often by means of graphical or numerical summaries, without making explicit inferences to underlying population(s).

The key distinction is that descriptive statistics involves describing patterns that are *intrinsic* to the data. No attempt is made, at least not explicitly, to use the patterns in the data to estimate attributes of the population from which the sample was drawn or to test hypotheses about the underlying environmental process. More often than not, descriptive statistics are used for exploratory data analysis to discover patterns and suggest underlying environmental processes that can be assessed more directly in subsequent studies.

Oregon breeding birds example.–To illustrate the use of descriptive statistics, we will draw on breeding bird data from the Oregon Coast Range. The data used here represent approximately 35 repeated breeding bird point counts in each of 30 subbasins, or small landscapes. The variables include the relative abundances of 97 different bird species and 52 habitat variables describing the composition and configuration of vegetative cover types of each subbasin.

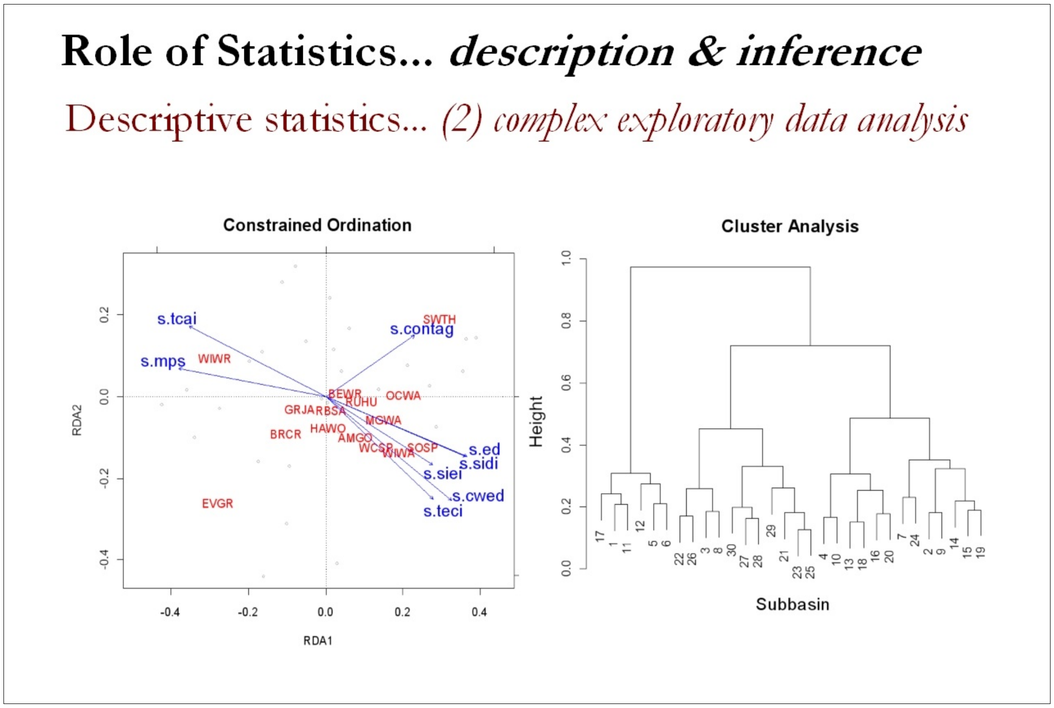
Role of Statistics… *description & inference*

Descriptive statistics… *(1) simple statistical summaries*



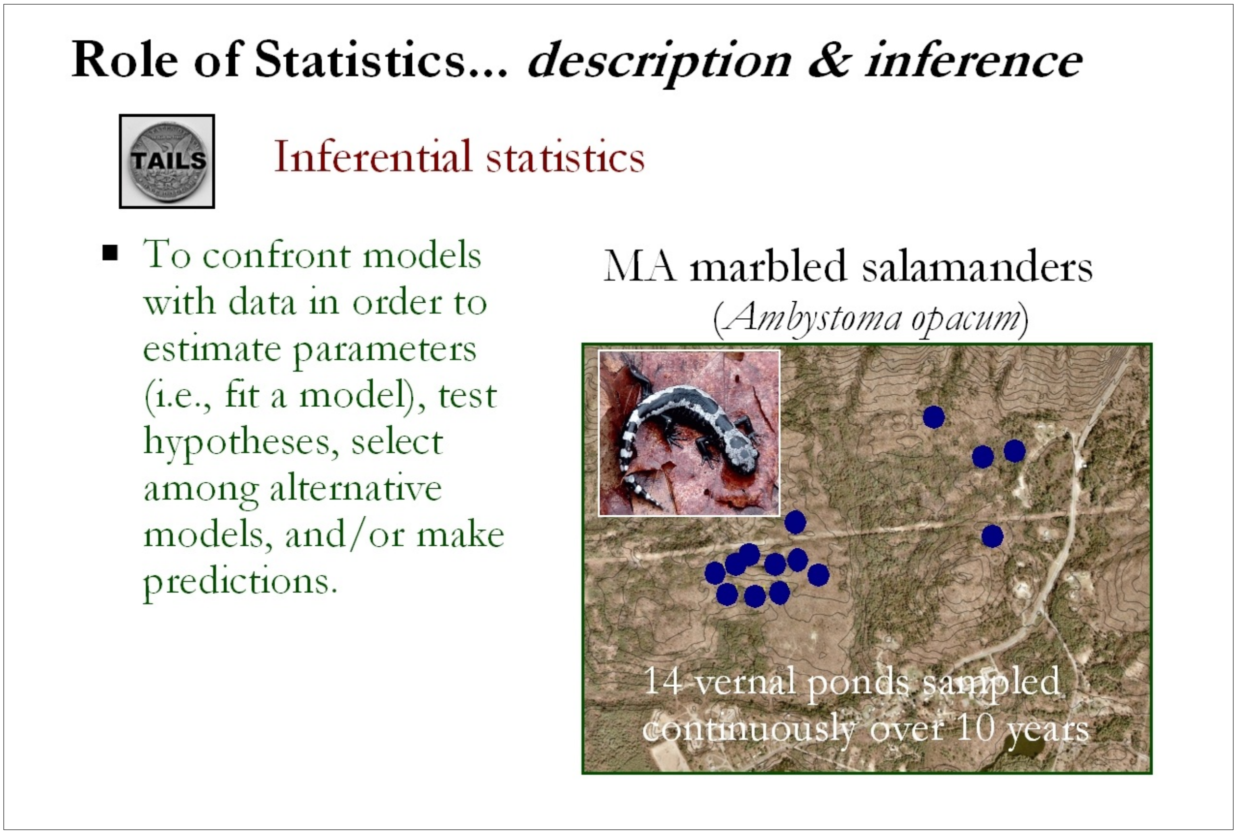
	BHGR	BRCR	BTPI	BUGR	BUSH
nobs	30.000	30.000	30.000	30.000	30.000
min	0.118	0.000	0.061	0.000	0.000
max	1.243	0.971	1.286	0.242	0.294
mean	0.580	0.392	0.532	0.013	0.057
median	0.572	0.361	0.500	0.000	0.054
sum	17.413	11.768	15.960	0.386	1.706
sd	0.254	0.224	0.321	0.047	0.071
cv	43.800	57.121	60.261	366.811	124.202
xeros	0.000	1.000	0.000	27.000	14.000
pct.xeros	0.000	3.333	0.000	90.000	46.667
nobs.missing	0.000	0.000	0.000	0.000	0.000
pct.missing	0.000	0.000	0.000	0.000	0.000
se	0.046	0.041	0.059	0.009	0.013
se.ratio	7.997	10.429	11.002	66.970	22.676
richness	30.000	29.000	30.000	3.000	16.000
sh.diversity	3.303	3.226	3.220	0.909	2.634
sh.evenness	0.971	0.958	0.947	0.827	0.950
si.diversity	0.960	0.956	0.955	0.533	0.917
si.evenness	0.994	0.990	0.988	0.800	0.978

Simple statistical summaries.–Perhaps the most common use of descriptive statistics is to provide a simple descriptive summary of the data. There are numerous tabular and graphical methods available for this purpose. For example, as shown here, there are numerous statistical measures that describe the distribution of a variable numerically (e.g., mean, variance, range, etc.) and graphically (e.g., histogram, box-and-whisker plot, empirical distribution, quantile-quantile plot, etc.). These descriptive statistics can be extremely useful for understanding the central tendency of a variable and/or its variability. Importantly, these descriptive statistics are used solely to describe the characteristics of the data; no attempt is made to infer characteristics of a larger underlying (statistical) population.



Complex exploratory data analysis.–There are many more complex methods for exploratory data analysis, and these can be especially useful for describing complex multivariate relationships for data sets involving multiple dependent or interdependent variables. Shown here are two examples: constrained ordination and cluster analysis. For the Oregon bird data, the constrained ordination (left figure) involves finding and displaying patterns in the bird species data that are explainable by the habitat variables. A set of axes is found that arranges the plots (points) such that their location in the ordination space indicates similarities in bird species assemblages that can be maximally explained by the habitat variables (constraints). The position of each species and the direction and length of the habitat vectors reveal relationships among species and habitat variables in a succinct manner. Importantly, the ordination plot displays relationships intrinsic to the data set. Typically, no inferences are made with respect to the underlying population from which the same was drawn and thus the method is purely descriptive. However, there are situations when constrained ordination is used to infer underlying relationships between the habitat variables and the species patterns, so the distinction between descriptive and inferential use of the procedure depends on the context in which it is used.

The cluster analysis (right figure) involves aggregating subbasins (the leaves of the tree) based on similarity in bird species assemblages into progressively fewer clusters until they are all aggregated into a single cluster. As the subbasins fuse together, they do so at increasingly lower levels of similarity, so that the final single cluster contains subbasins that are quite dissimilar to each other in terms of their species assemblages. The structure of the cluster tree-like diagram (known as a "dendrogram") reveals the nature of the similarity (or dissimilarity) of the subbasins in terms of their bird community. In this case, it appears that there are perhaps three relatively distinct groups or clusters of subbasins.

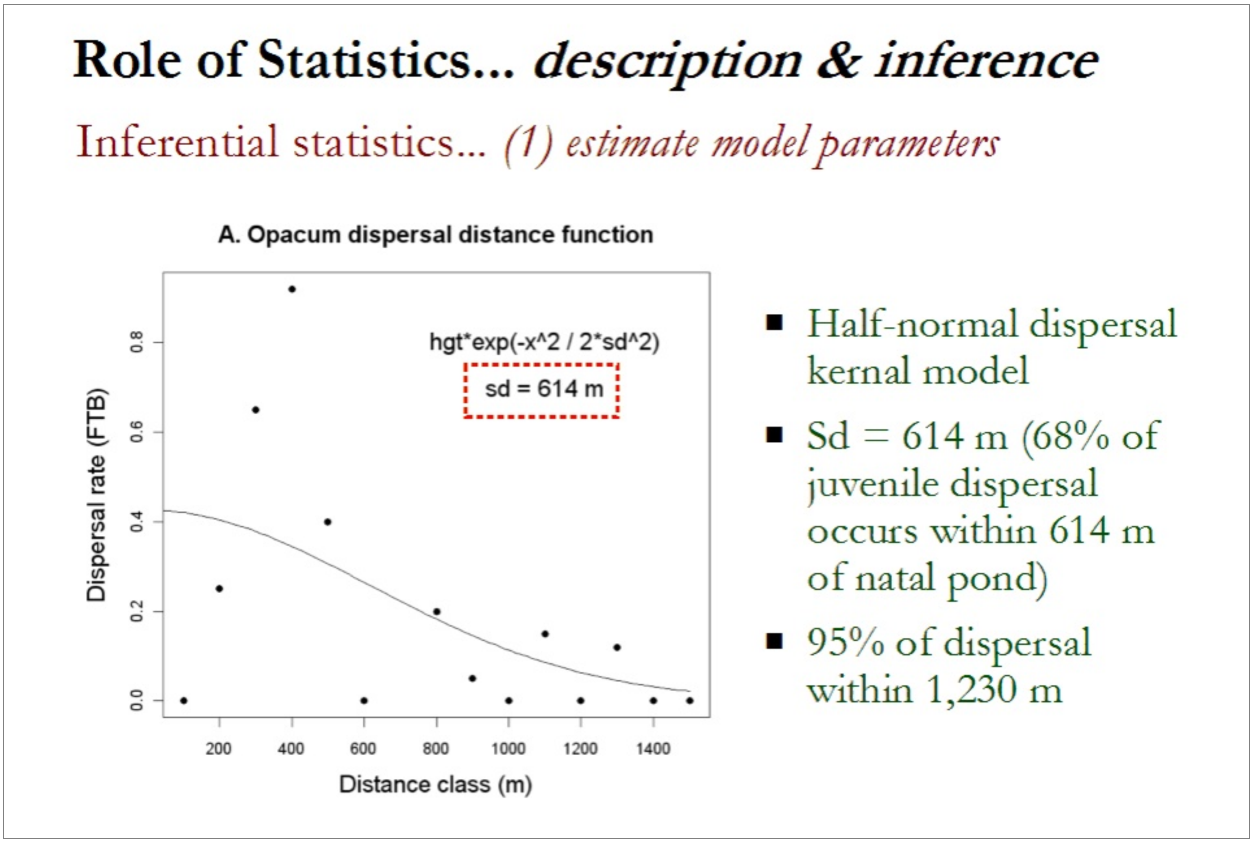


2.2 Inference

Definition.–To confront models with data in order to estimate parameters (i.e., fit a model), test hypotheses, select among alternative models, and/or make predictions.

The key distinction is that inferential statistics involves making statements about underlying processes *extrinsic* to the data. An attempt is made explicitly to describe the underlying population from which the sample was drawn, to chose among alternative explanations of the underlying environmental process (i.e., to weigh the evidence for each of the alternatives) or to make predictions for samples not yet collected. The scope of inferential statistics is exceedingly broad, however, as inferences can often be drawn from either comparative mensurative (observational) or manipulative experiments and from either field or laboratory studies. In addition, there are different conceptual and methodological frameworks (paradigms) for conducting statistical inference. The two major frameworks in use today are known as the classical (frequentist) framework and Bayesian framework, and each of these frameworks includes a variety of technical methods for conducting the inference (e.g., estimating parameters).

Massachusetts marbled salamander dispersal example.–To illustrate the use of inferential statistics, we will draw on a long-term study of marbled salamanders in western Massachusetts in which we sampled 14 seasonal ponds (or vernal pools) continuously over 10 years.

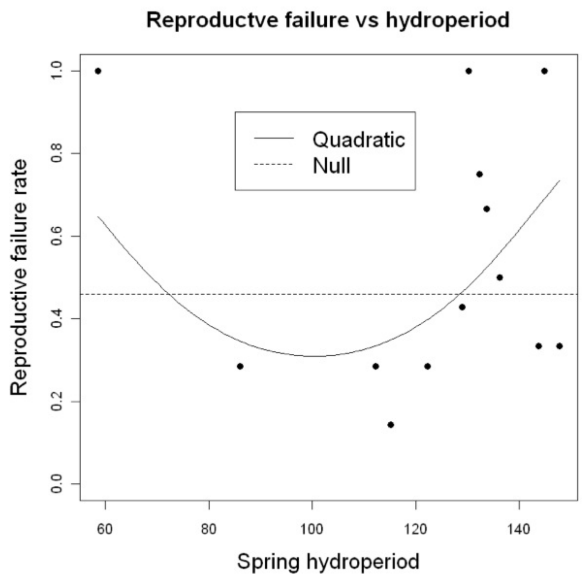


(1) *Estimate model parameters*.–The most basic and perhaps most important statistical inference is to estimate parameters of the proposed statistical model from the data. Often, parameter estimation is the sole purpose of statistical inference.

In the example here, the data represent the standardized dispersal rate (y-axis) for first-time breeders (FTB) as function of the distance from the natal pond (x-axis). Each point represents the standardized dispersal rate for the pairwise combination of ponds in each 100-m distance interval. I fit a half-normal curve (with normal errors) to the data and estimated the standard deviation to be 614 m. This means that 68% of the dispersing juveniles are expected to disperse less than 614 m from their natal ponds and 95% are expected to disperse within 1,228 m.

Role of Statistics... *description & inference*

Inferential statistics... *(2) test hypotheses*



H_a: Reproductive failure rate is associated with average spring hydroperiod

```
Analysis of Deviance Table

Model 1: y ~ 1
Model 2: y ~ spring.hydro + I(spring.hydro^2)
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1        12    22.2229
2        10    18.8422  2   3.3807    0.1845
```

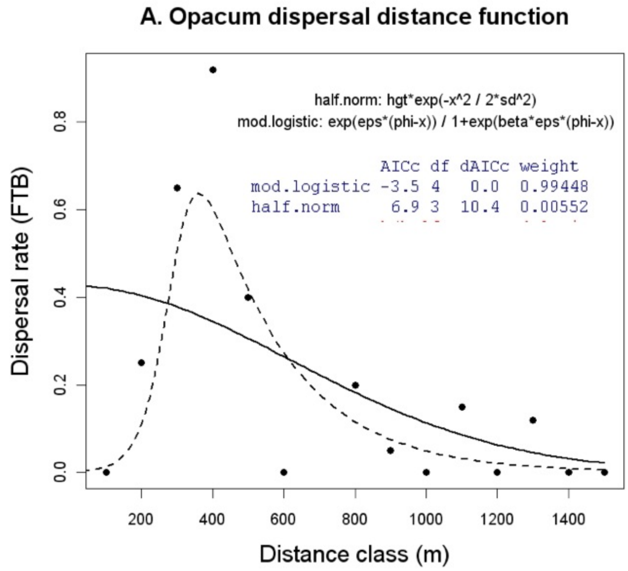
- Cannot accept on *statistical* basis

(2) *Test hypotheses*.–The classic use of statistical inference is to test hypotheses, which usually amounts to testing whether one model is significantly better than another. In some cases, this amounts to testing whether parameter estimates are significantly different from zero. Quite simply, we use the data to infer whether there is sufficient evidence to reject the null hypothesis in favor of the alternative, given the specified model.

In the example here, the data represent the reproductive failure rate (# years failed/# years attempted) (y-axis) as a function of average spring hydroperiod (#days from April 1 until pond drying) (x-axis). Each point represents a single pond. I tested whether a quadratic polynomial (with binomial errors) fits the data better than the null model of no relationship between reproductive failure rate and spring hydroperiod. The quadratic model only explained 15% of the deviance (something akin to the variation in the reproductive failure rate among ponds), which was not significantly greater than the deviance explained by chance under the null model (p=0.18), suggesting that the quadratic model cannot be accepted on a statistical basis.

Role of Statistics... *description & inference*

Inferential statistics... *(3) select among alternative models*



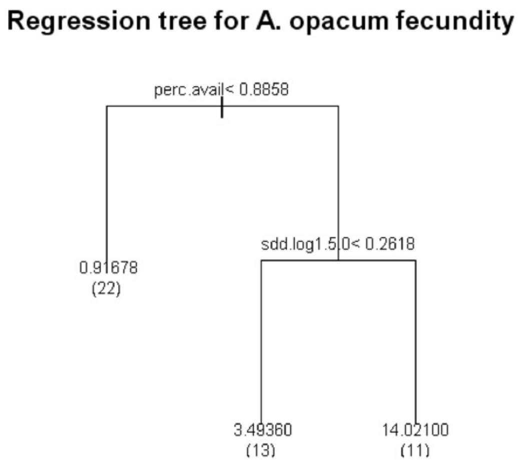
- Half-normal versus modified logistic model
- Information criterion (AIC) suggests modified logistic is vastly superior, despite being less parsimonious
- But what about biological relevance?

(3) *Select among alternative models*.–A closely related use of statistical inference is to chose among alternative models or competing hypotheses. This amounts to determining the relative strength of support in the data for one model over another and may or may not involve determining whether the most supported model is statistically significant as well.

In the example here, the data represent the dispersal data discussed above. However, here I also fit a modified logistic function to the data and used an Information Theoretic Criterion (AIC) to assess the strength of evidence for the two competing models. In this case, the AIC is considerably smaller for the modified logistic model, which is 99% more likely than the half-normal model. But what about the biological relevance of the modified logistic? It may be a much better fit, but does it make sense ecologically?

Role of Statistics... *description & inference*

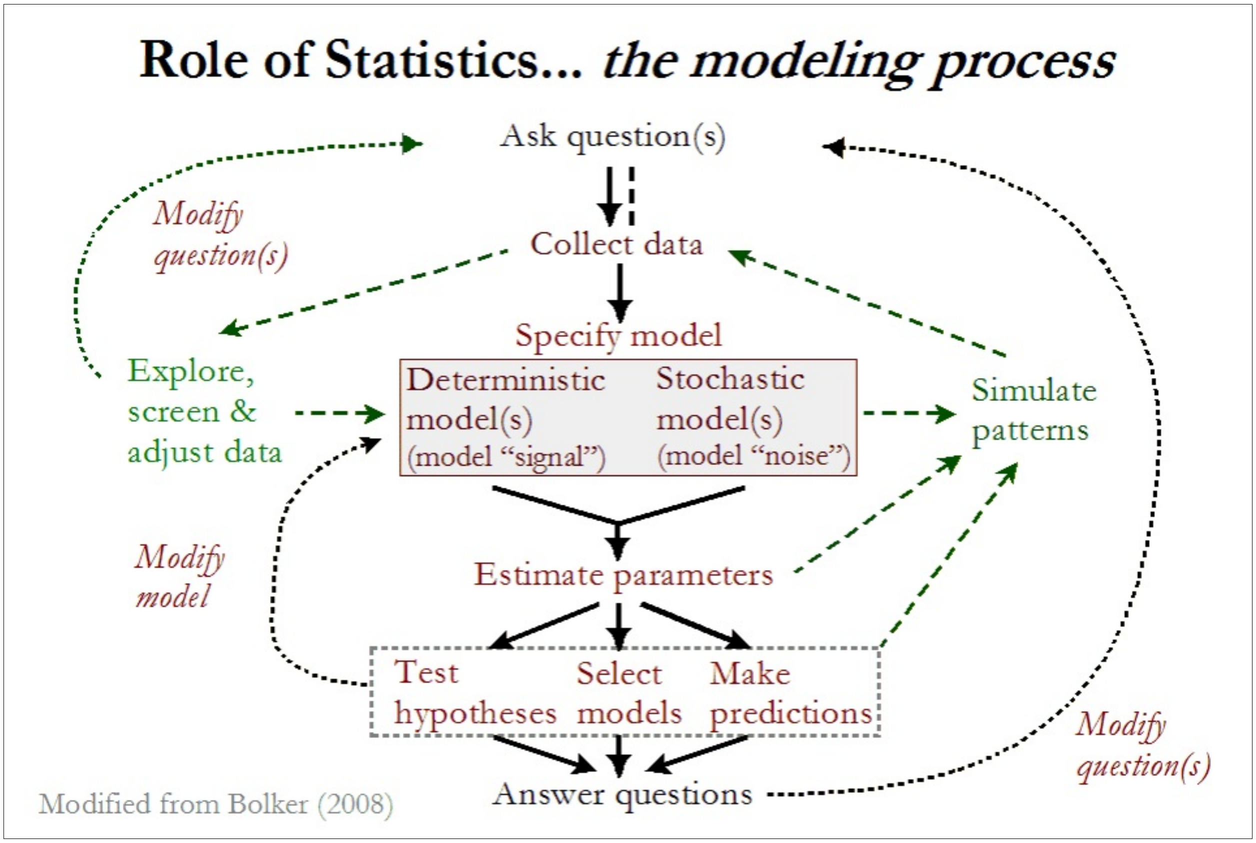
Inferential statistics... *(4) make predictions*



- Cross-validated regression tree explains 62% of the variation in female fecundity from Fall hydrological and meteorological conditions
- Predictions are made by dropping new observations (pond-year) down the tree

(4) *Make predictions*.–A final use of statistical inference is to make predictions for unobserved or unobservable situations. This amounts to using the statistical model to estimate the values for new observations. These predictions can be made either for observations within the existing domain of the data (i.e., the observed range of the predictor variable(s)) or for observations outside the data domain, but the latter are made with much less certainty since they involve making predictions in uncharted territory.

In the example here, the data represent fecundity (#female metamorphs/breeding female) plus a suite of hydrological and meteorological variables that describe conditions in the pond basins from adult entry to breed in the Fall through emergence of metamorphs (juveniles) the following summer for each pond-year combination (N=46). Here, we used Regression Tree Analysis to build a predictive model for fecundity. In the parsimonious model shown here, the cross-validated regression tree explains 62% of the variation in fecundity based on two independent variables: average percent of maximum nesting area available during the female period in the pond basin (representing nest site availability) and estimated egg survival based on the number of degree days (using a 0 degree threshold) of egg exposure prior to likely nest inundation. To predict fecundity for the current breeding year (i.e., how many emerging metamorphs to expect the following summer per breeding female this Fall), we calculate the percent nest area (perc.avail) and degree days of egg exposure (sdd.log1.5.0) and use the decision tree to predict the mean or expected fecundity.



3. The modeling process

Before jumping into the realm of statistical modeling further, we need an outline or road map of the modeling process (modified from Bolker, 2008) – which will serve as the organizational framework for this course.

1. Identify the question.–You have to know what you want to find out before you can start trying to model. In other words, you have to know the question before you can find the answer. Moreover, you should know what your question is both at a general, conceptual level (e.g., "how often do marbled salamander populations fail reproductively?") and at a specific level (e.g., "can the difference in reproductive failure rates between populations A and B be explained by differences in pond hydroperiod?). Unfortunately, no recipe can tell you how to ask good questions. Nevertheless, it is the first and most important step of any analysis and motivates all the other steps. Ideally, you began the study by asking good questions and now, with data in hand, you are ready to model.

2. Collect the data.–Ideally, once the question has been identified, the study is designed and the data is collected in an appropriate fashion. There are many considerations to the design of field and laboratory studies: identifying the desired scope of inference, choosing appropriate observational/experimental units, choosing the types of data to collect, establishing a robust sampling scheme (i.e., the number and spatial and/or temporal distribution of units and method(s) of collecting the data) to ensure accurate and precise inferences.

3. Optionally, explore the data for patterns and problems and return to step 1.–One of the basic tensions in all data analysis and modeling is how much you have all your questions framed before you begin to look at your data. In the classical statistical framework, you are suppose to have all your hypotheses laid out in advance and not stray from that course during the analysis. Allowing your data to suggest new statistical tests raises the risk of "fishing expeditions" or "data-dredging" – indiscriminate scanning of the data for patterns. But this philosophy may be too strict for environmental research. Unexpected patterns in the data can inspire you to ask new questions, and it is foolish not to explore your hard-earned data in this regard. In addition, exploratory analyses can reveal aspects of the data that may help you construct a more appropriate model to answer the original question. I see no particular harm in letting the data guide you to a better model, as long as you recognize the risk of detecting patterns that are not real and seek to confirm the findings with subsequent study.

4. Choose deterministic model(s).–Next, you need to choose a particular mathematical description of the pattern you are trying to describe. The *deterministic* model is the average, or expected pattern in the absence of any kind of randomness or measurement error. It's tempting to call this an "environmental" model, since traditional environmental models are described in deterministic terms, but environmental models can be either deterministic or stochastic (see below). The deterministic model can be phenomenological (i.e., a mathematical form that simply describes the observed pattern well, but is not environmentally based), mechanistic (e.g., a Type II functional response for predation rate, in which the parameters of the equation have a real-world ecological meaning), or even a complex individual-based simulation model. It is often useful to think of the deterministic model as defining the environmental process of interest.

5. Choose stochastic model(s).–To estimate the parameters of a model, you need to know not just the expected pattern but also something about the variation about the expected pattern. Typically, you describe the *stochastic* model by specifying a reasonable *probability distribution* for the variation. For example, we often assume that variation that comes from measurement error is normally distributed, while variation in the number of plants found in a quadrat of a specific size is Poisson distributed. In addition, sometimes the model includes more than one level of error (referred to variously as a multi-level or hierarchical model). For example, the samples may be clustered into groups (spatially) or taken from the same individuals over time (repeated measures), in which case there may be two (or more) levels to the model to distinguish the variability within each level from the variability among levels. Similarly, there may be a separate model for the observation or measurement error and the process error. Note, while most environmental models have a stochastic component that deals with the variation about the expected pattern, in some cases the variation cannot be described very well by any existing probability distribution. *Nonparametric* methods have been developed to deal with some of these special situations (e.g., quantile regression), in which case no particular error distribution is specified.

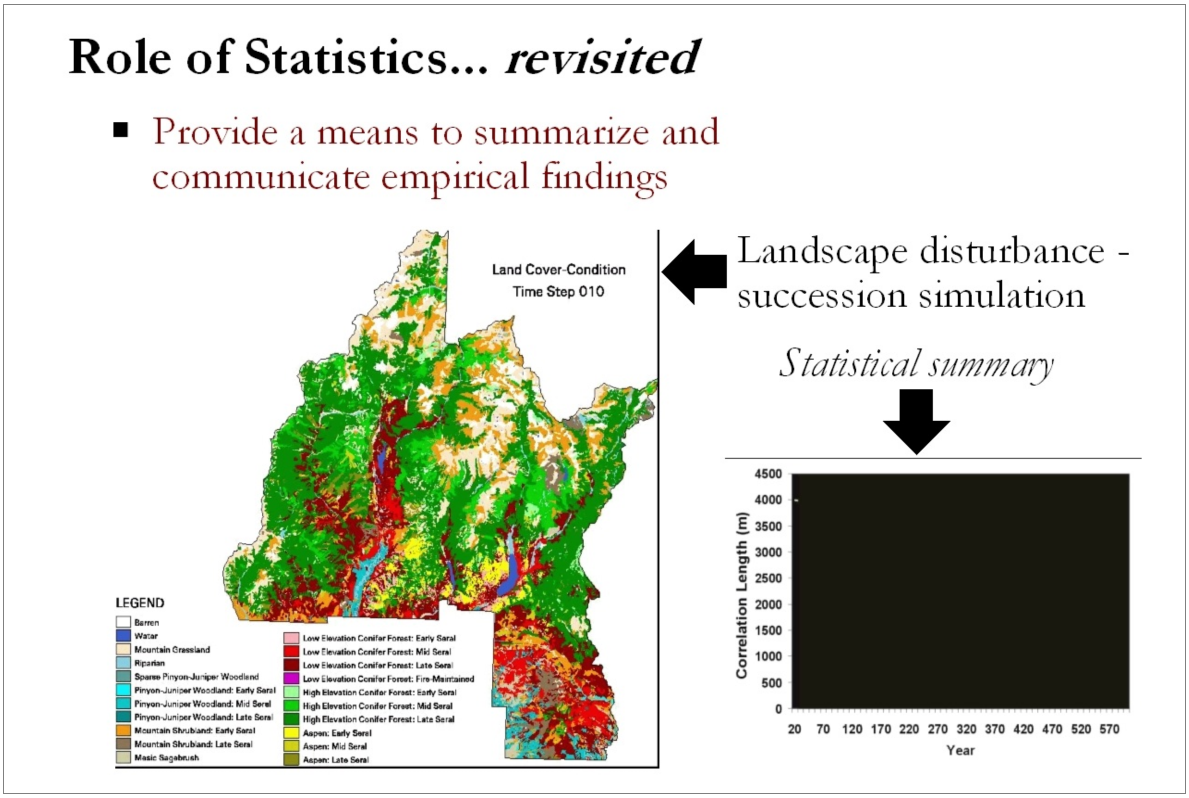
6. Fit parameters.–Once you have defined your model, you can estimate both the deterministic parameters (e.g., slope, attack rate, handling time, etc.) and stochastic parameters (e.g., the variance or parameters controlling the variance). This step is a largely technical exercise in figuring out how to get the computer to fit the model to the data. In some cases, an analytical solution is available; in other cases, a numerical solution must be found. Unlike the previous steps, it provides no particular insights into the basic environmental questions. Unfortunately, this step can require considerable

technical expertise, depending on the complexity of the model. Moreover, there are certain philosophical issues pertaining to the choice of an inference paradigm (e.g., classical frequentist versus Bayesian) that will determine the appropriate method for fitting the model.

7. Estimate confidence regions/test hypotheses/select model(s)/make predictions.–You need to know more that just the best-fit parameters of the model (the *point estimates*, in statistical jargon). Without some measurement of uncertainty, such estimates are meaningless. By quantifying the uncertainty in the fit of a model, you can estimate confidence limits for the parameters. You can also test environmental hypotheses, from both an environmental and a statistical point of view (e.g., can we tell the differences statistically between the reproductive failure rate in two different populations? And are these differences large enough to make any practical difference in the population dynamics?). You also need to quantify uncertainty in order to choose the best out of a set of competing models, or to decide how to weight the predictions of different models. And you need to quantify uncertainty in order to make honest predictions for future observations.

8. Optionally, use stochastic simulation to examine your environmental model.–Instead of immediately analyzing your data, you may want to start by simulating the system. You can use stochastic simulation to understand qualitative patterns that derive from your statistical model (i.e., the specific deterministic functions and probability distributions) and study design. You can also use simulated "data" from the system to test your estimation procedures. Since you never know the true answer to an environmental question – you only have imperfect measurements with which you're trying to get as close to the answer as possible – simulation is the only way to test whether you can correctly estimate the parameters of an environmental system. It is always good to test such a best-case scenario, where you know that the functions and distributions you're using are correct, before you proceed to real data. Lastly, *power analysis* is a specific kind of simulation testing where you explore how large a sample you would need to get a reasonably precise estimate of your parameters or more generally how variations in study design would change your ability to answer environmental questions. Ideally, stochastic simulation (and power analysis specifically) is done before the data are actually collected so that you can modify the statistical model and study design accordingly. However, stochastic simulation can also be useful after the data is collected. For example, if you can choose model parameters that make the simulated output from those functions and distributions look like your data, you can confirm that the models are reasonable – and simultaneously find rough estimates of the parameters.

9. Put the results together to answer questions and return to step 1.–Modeling is an iterative process. You may have answered your questions with a single pass through steps 1-8, but it is far more likely that estimating parameters and confidence limits will force you to redefine your models (changing their form or complexity or the environmental covariates they take into account) or even to redefine your original environmental questions. You may need to ask different questions, or collect another set of data, to further understand how your system works.

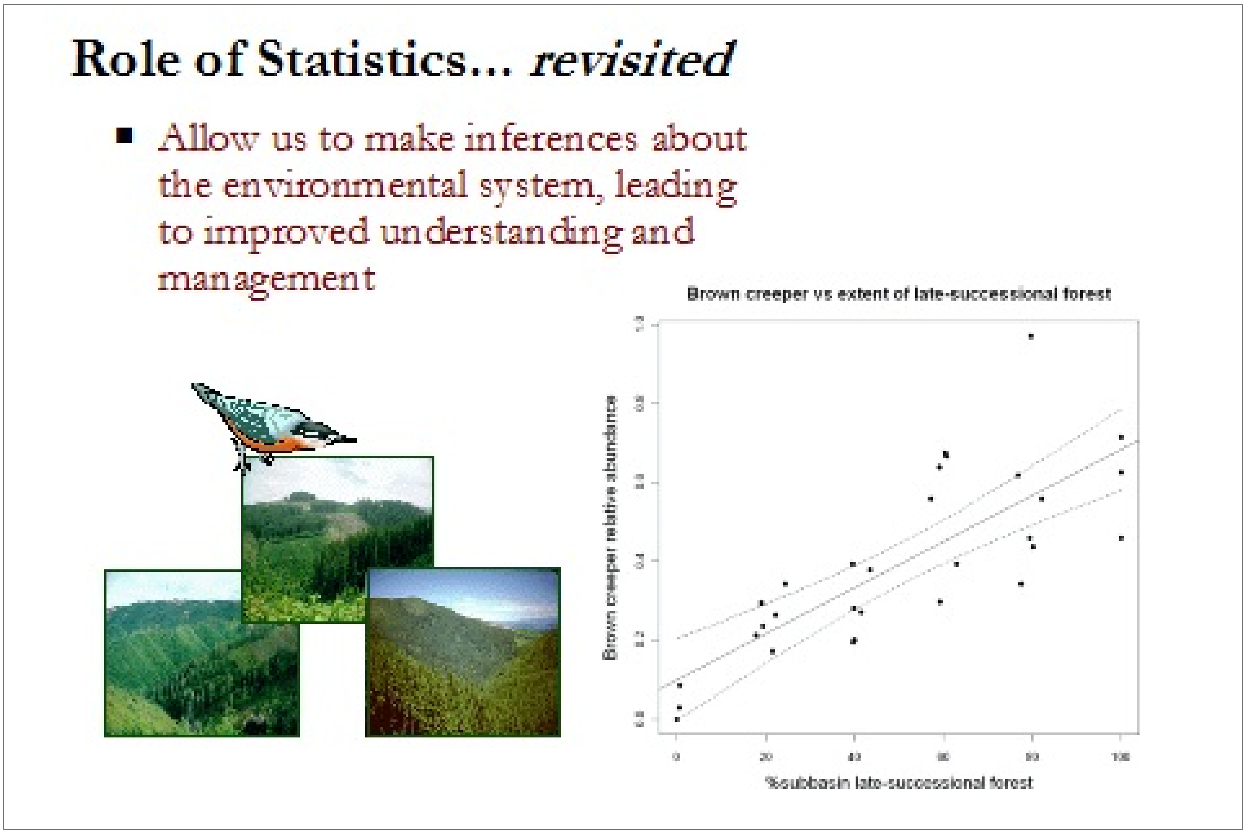


4. The role of statistics, revisited

Having briefly considered the role of statistics in ecology and the major elements of the modeling process, let's revisit the role of statistics in ecology. I contend that there are three major roles:

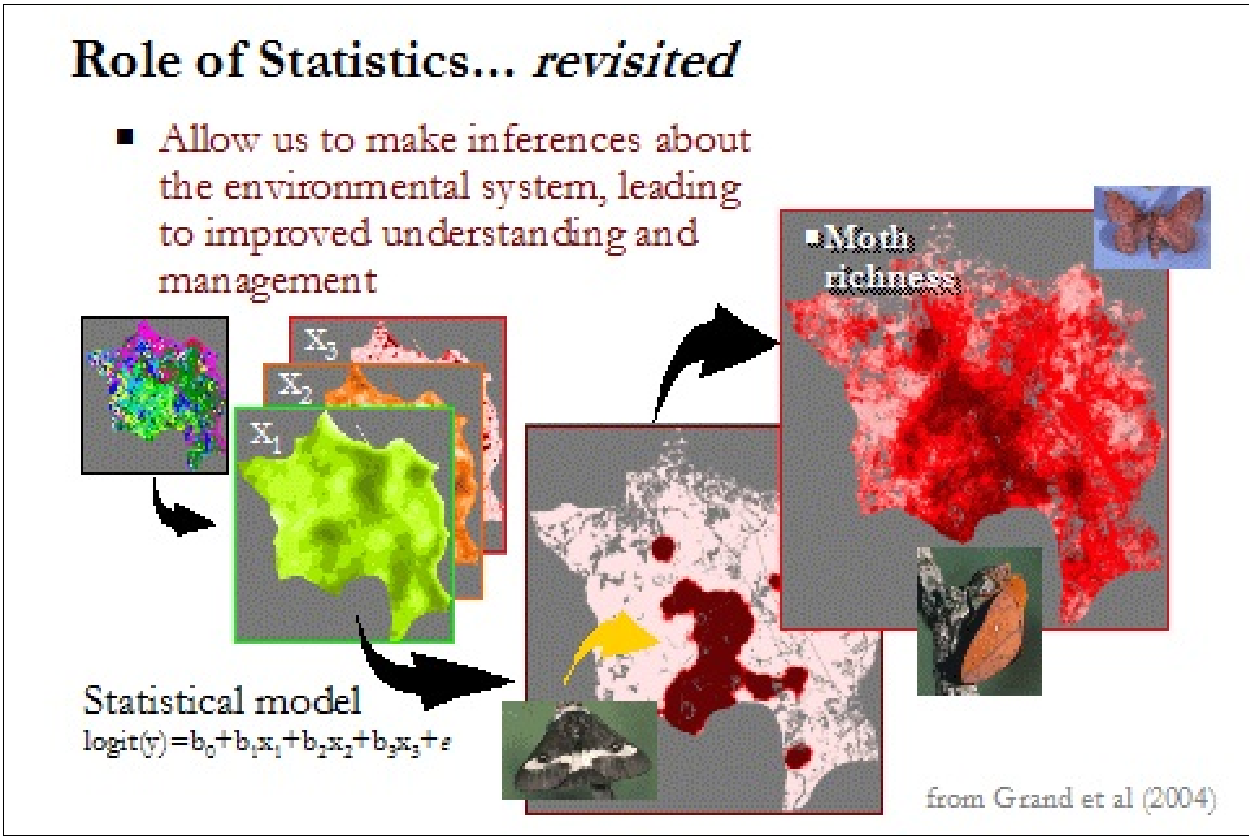
- Communication.–Perhaps the single most practical use of statistics is to provide a means to effectively summarize and communicate empirical findings. This is where the use of descriptive statistics largely resides of course, but summarizing and communicating what is learned through the use of inferential statistics is just as important. Finding ways to effectively communicate the results of complex statistical analyses should be of paramount concern to every practitioner of statistics.

  In the example here, a spatially explicit landscape disturbance-succession model (RMLands) was used to simulate the dynamics in vegetation patterns for a landscape in southwestern Colorado (left-side animation). The resulting vegetation patterns were quantified using FRAGSTATS, a spatial pattern analysis program for quantifying landscape patterns. The metric shown in the right-side animation is 'correlation length', a descriptive statistic that summarizes the overall physical connectedness of the landscape. This simple descriptive statistic summarizes complex landscape patterns and, as used here, effectively depicts the range of variation in landscape structure under this particular disturbance regime.

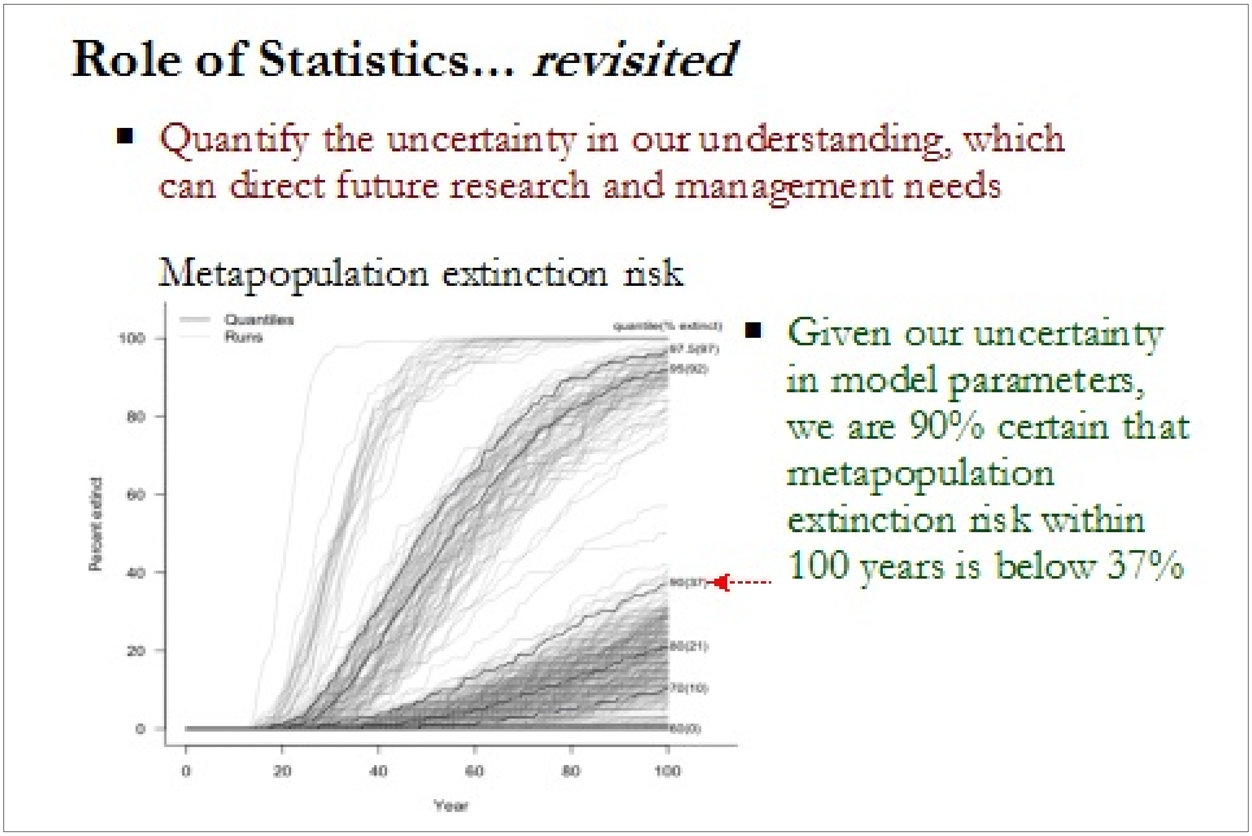


- Inference.–The goal of most statistical applications in ecology and conservation science is to make inferences (i.e., estimate parameters, test hypotheses, select models, make predictions) about the environmental system under consideration, leading to improved understanding and management. Inference is the essence of the scientific endeavor.

  In the example here, a linear regression model was used to estimate the relationship between percent of the landscape (small subbasins) comprised of late-successional forest (x-axis) and the relative abundance of brown creepers (y-axis) across 30 landscapes in the Oregon Coast Range. The fitted regression line (i.e., expected values) and a confidence envelope about the line (i.e., the uncertainty) is shown. The model is strongly supported by the data (results not shown) and allows us to infer that brown creepers increase in abundance as the percentage of late-successional forest increases. Therefore, forest management policies and practices that decrease late-successional forest are likely to negatively affect this species.



In the example here, a series of logistic regression models were used to estimate the probability of occurrence (i.e., presence) of several rare moth species based on a variety of habitat variables measured at different spatial scales in the pine barrens of southeastern Massachusetts (from Grand et al. 2004). These models, built on a sample data set, were then applied to the entire study area to predict the probability of occurrence of each species at every location (30 m resolution) and the results were combined across species to identify "hotspots" of moth richness which could serve as priorities for land conservation and/or habitat management.



- Uncertainty.–A less commonly recognized but increasingly important goal of statistical applications is to quantify the uncertainty in our understanding, which can direct future research and management needs. One of the great utilities of statistics is that it provides a means of summarizing not only what we think we know but how well we know what we think we know. That is, statistics provide us the means to quantify the uncertainty in our inference, which can be just as important as the inference itself.

  In the example here, the data from the marbled salamander field study were combined with published studies to parameterize a spatially realistic population viability model. The figure shown here depicts the risk of metapopulation extinction under 1,000 different model parameterizations in which two key model parameters, reproductive failure rate and adult survival rate, were allowed to vary within the bounds of our uncertainty in these parameters. Based on these simulation results, we can be 90% certain that metapopulation extinction risk within 100 years is below 37%. If we find this level of uncertainty unacceptable, we can direct additional research to improve our estimates of reproductive failure rate and adult survival rate.