General Advice for Data Collection and Analysis (adapted from Krebs 1998)

1. Identify a problem and ask a question.
2. Not everything that can be measured should be.
3. In some situations, the questions may be unanswerable at the present time.
4. Use your non-statistical knowledge of the problem throughout the planning process.
5. Collect data that will answer your question and meet statistical assumptions.
6. Keep the design and analysis as simple as possible.
7. With continuous data, save time and money by deciding on the number of significant figures needed in the data before you start the experiment.
8. Never report an estimate without some measure of its possible error.
9. Always be skeptical of the results of statistical tests of significance.
10. Recognize the difference between statistical significance and biological significance.
11. Make backup copies of data sheets and database files.
12. The quality of planning and data collection will be apparent in the final report—garbage in, garbage out.

$$P(\text{data} \mid \text{model})$$

Null hypothesis tests

- An objective, uniform framework for making decisions
- Generation of competing statistical hypotheses (H_0 & at least one H_a)
- Probability of the data given the null hypothesis
- Frequentist characteristics:
  - Probability of an event = its <u>relative frequency</u>
  - All prior/collateral information is ignored (focused on the data)
  - Tests of significance assume the null hypothesis is true
    - Tests compute the probability of obtaining a result at least as extreme as the one actually observed by chance
    - **Shortcoming**: P –value is closely related to sample size – importance of determining the magnitude of effect (i.e., effect size) that is biologically relevant
    - May be more often applicable to experimental studies where true replication is possible

Null hypothesis testing

- Testing to determine if the null hypothesis is true
  - *Alternative hypothesis (samples derived from populations with different means): ($\mu 1 \neq \mu 2$)*
  - *Null hypothesis (no difference): (the mean, $\mu 1 = \mu 2$)*
- If data contradicts H_0: either the null hypothesis is false or a low probability event has occurred
  - *Apriori significance level (α) – usually 0.05 or 0.1*
- **P-value**: calculated probability of finding the observed, or more extreme, results when the null hypothesis is true
- Decision to reject or fail to reject H_0 is based on the test statistic and whether or not it falls in the critical region (rejection) of the statistical distribution

Power analysis

- **Definition:** The ability of a statistical test to detect an effect, given that the effect actually exists
- **Calculation:** 1 – the probability of a type II error ($1-\beta$)
- If the power is too low, then we have little chance of detecting a significant difference (i.e., analysis will yield a nonsignificant result)
  - *Even if there may be real differences*
  - *Example: Stream water chemistry*
- Ideally, we could use statistical tests that minimize α and β. Why can't we?
  - *As an alternative we fix α at a specified significance level and apply statistical tests that maximizes the power*

Power analysis

Calculation of sample size based on the level of significance, power, effect size, and variability (standard deviation):

1. Select a significance level (α)
2. Establish the power ($1 - \beta$) required to detect an effect (common = 0.80)
3. The effect size (biological change or significance) that needs to be detected is determined; based on the actual units of the response
   - *Effect size and the ability to detect it are indirectly related*
4. The extent of variation associated with the response variable in the population is determined (pilot study, literature)

- Talk about the relationship among sample size – variance – ability to detect significant differences.
- Suppose we know the maximum sample size. Why would we do this?

Parametric vs Nonparametric

- **Parametric:** assume sampled data are from populations that follow a certain distribution (usually normal)
  - *Assumes multiple samples have similar variability*
  - *Most commonly used…works well when the distribution is only approximately normal*
- **Nonparametric:** does not assume that data follow a normal distribution
  - *Fewer assumptions about the data*
  - *Conservation (less robust) – probability values tend to be higher, making it harder to detect real differences*
  - *Reduced power to detect differences with small sample sizes*
- Typically, we use nonparametric tests when:
  - *Variables are ranks or scores*
  - *Extremely small sample sizes*
  - *Data follow normal distribution but contains extreme outliers*

Data transformation

- Testing to determine whether data were sampled from a normal distribution is helpful – often less useful than we would like
  - *Shapiro-Wilk (SW)*
  - *Kolmogorov-Smirnov (KS)*
  - *Both tests quantify the difference between the data distribution and an ideal normal distribution*
    - Larger values represent larger discrepancies

- If we have a large sample size, a transformation may cause data to approximate a normal distribution

Data transformation

- **Definition:** a mathematical operation that changes the measurement scale of a variable
- Why?
  - *Approximate normality*
  - *Stabilize variance (assumption of many tests is that the variance is independent of the mean)*
  - *Reduce the effect of outliers*
  - *Linearize a relationship*
- Common transformations:
  - *Logarithmic (base 10 or natural log) – x' = log(x +1)*
  - *Square root transformation*
  - *Arcsine transformations*
- **Back-transformation:** Even though you've done a statistical test on a transformed variable (i.e. log of fish abundance) we don't report means on this transformed data.
  - *Why?*

$$P(\text{model} \mid \text{data})$$

Bayesian inference

- Alternative to null hypothesis testing
  - *Thought to provide more meaningful conclusions for field studies (observational)*
- Probability represents the degree of belief that a particular event will occur
  - *Probability of a model given the data*
  - *Incorporates uncertainty*
- Inferences are conditioned on prior (sometimes) information
  - *Useful when there isn't prior information for a given problem but there is a strong understanding of the mechanisms that effect the problem*
  - *Easy integration of new information, a necessary component of adaptive management*
- **Based on conditional probabilities, the objective is to incorporate prior knowledge in combination with new data to make statistical inferences**

Model selection

- Inference approach that is based on information theory (see Burnham and Anderson 2002)
- Relies on the existence of a knowledge base from which a suite of realistic competing models can be derived
  - *Deemphasizes the use of probability values and emphasizes model utility*
  - *Best model: candidate model best supported by the data (weighted by evidence)*
  - *Selection goal is to develop the simplest model that encompasses the suspected cause and effect relations (Occam's Razor)*
  - *Likelihood criterion is used to compare the competing models – assumes the "full truth" is known and weights are based on deviations from the "full truth"*
    - Closest model to "full truth" has the least information loss
- Thought to provide more meaningful conclusions for field studies (observational) than null hypothesis testing

Model selection

- Closest model to “full truth” has the least information loss
  - ***IMPORTANT:*** *Doesn’t necessarily mean it is a good model – not a measure of model performance*
- **AIC:** quantifies the goodness or lack of fit of a set of models, given the observed data (small sample sizes, AICc)
  - *Sensitive to sample size: models with numerous parameters may have good fit but are <u>overfitted</u> and suffer from low precision*
  - *Can’t interpret individual criterion, instead compare differences by ranking candidate models*
  - *Models with smallest Δ value is the best-fit among the candidate models*
    - $\Delta < 2$ have good support
    - $\Delta > 10$ have no support

Study design in Fisheries Management

- **Management:** the process of maintain fish populations at some target level, commensurate with the capacity of the environment and in accordance with established objectives set with the consideration of user or constituent needs
  - ***Misconception:*** *fisheries scientists need to be data rich to conduct effective management*
- Figure 2.1 (scope differs)
  - *Management assessments are likely to be site- or resource-specific, with results aimed at local rather than broad applications; lack of replication and controls*
- Four types of management investigations

Feasibility studies

- Precursory to more advanced studies, exploratory in nature, and entirely dependent on existing information and experiences to determine if a problem can be resolved
- At the end, it should be possible to estimate reasonably the time and personnel required to conduct a given amount of sampling
  - *Weigh those costs against the likely benefits of doing the study*
- Standard methods book

Pilot studies

- Opportunity to evaluate assessment alternatives identified in the feasibility study
  - *Small-scale studies*
- Determine what sampling is likely to be effective in achieving the specific objectives under the expected conditions
- Estimate the sampling variability and allow calculation of sample sizes needed to attain the desired level of precision
  - *Establishing statistical reliability (power analysis)*
  - *Example: weighing relative precision of data from large numbers of small samples or small numbers of large samples*
- Will the sampling procedures produce representative data?

Measurement of treatment effects

- Typically most intensive study done by a manager
- Before and after study on a single system (without replication)
  - *Example: size structure of crappie before and after altered bag limit or minimum size limit legislation*
- Cannot provide information about the probability of similar responses in another system
  - *Limited inference capabilities, but broader inferences can be facilitated by replicated treatments (e.g., more lakes)*

Monitoring

- Crucial to managing fish populations
- Understanding how management practices change fish communities, stocks, or populations to achieve specific goals and objectives
  - *Agencies have vast files of data collected for monitoring to establish the status of a resource*
- Long-term monitoring studies are designed to provide periodic updates on the status of the resource, detect major changes, and establish trends
- Name some major federal examples? Mississippi?

Design of the study & statistical efficiency

- **IMPORTANT: NO GEAR IS PERFECT!**
- Effectiveness in the distribution of samples through time and space affects sample size. Why?
- First and foremost, we need to be confident that the study design will produce data that will answer pertinent questions
  - *Representativeness of the data & precision of the estimate*
  - *Allocating of samples through space and time to achieve adequate accuracy and representation (in the face of many sources of bias and error)*
  - *Representativeness is most likely achieved through random sampling, either simple (total) random, stratified random, or cluster sampling (chapter 3)*
  - *Systematic sample can be representative but will likely produce inflated variability*
- Scope of inference is based on a compromise between (1) effective sampling to produce representative data and (2) adjustment of the sample design to accommodate real field conditions
- Sample variability is inflated by any inconsistency in sampling process
  - *Training and protocols (SOPs): established in advance to establish acceptable deviations and alleviate many of the inconsistency issues that can arise*

Sampling units

- Sampling units generally vary naturally
  - *For example, not all fish of a species are the same size or color, and not all sites have equal number of fish or species*
  - *This variability is innate to a species or to a habitat, and the only way to control it is by changing the species genetics or the habitat characteristics*
  - *The variance (or the Standard deviation) of a population of sampling units cannot be changed*

- The variance of the mean (not variance among sampling units), standard deviation of the mean (i.e., standard error), can be reduced by increasing sample size.

Conclusion – Chapter Two

- Knowledge of the inference space provided by the study design must be clear so the data can be properly interpreted in terms of whether the data are directly applicable to the parameters of interest **or only the samples themselves (peer-review example)**

A couple stats quotes

An old one…“There are three types of lies -- lies, damned lies, and statistics.”

-- B. Disraeli (before S. Clemens)

A newer one…“You can’t normalize ignorance.”

-- D. Sivia 1998

“All models are wrong, but some are helpful.”

-- George Box, 1976

Scales of measurement

- **Nominal**: male/female, large/med/small; stock/quality/preferred

- **Ordinal**: 1/2/3, large/med/small; stock/quality/preferred; can be ranked
  - *Differences between rankings not necessarily equal*

- **Interval**: differences between measurements have meaning. Take temperature. The difference between 20 and 30 is the same as the difference between 70 and 80 = 10 degrees. Cannot say that 80 degrees is four times as hot as 20.

- **Ratio**: has the additional property that ratios are meaningful. Length of fish.

Fixed sites

- Decisions about whether to establish sites that are judged to represent the environment, extremes in the environment, to be similar to one another, or to produce the highest catches
- Each will have a different effect of variance among samples for the index
- Random vs. subjective decisions of fixed sites – statistical advantages vs convenience/efficiency
- Random sampling conducted in conjunction with fixed sites can provide confidence in the degree of representativeness of the data
- **Data must be recognized as indices reflecting the sites rather than being a population parameter estimate**

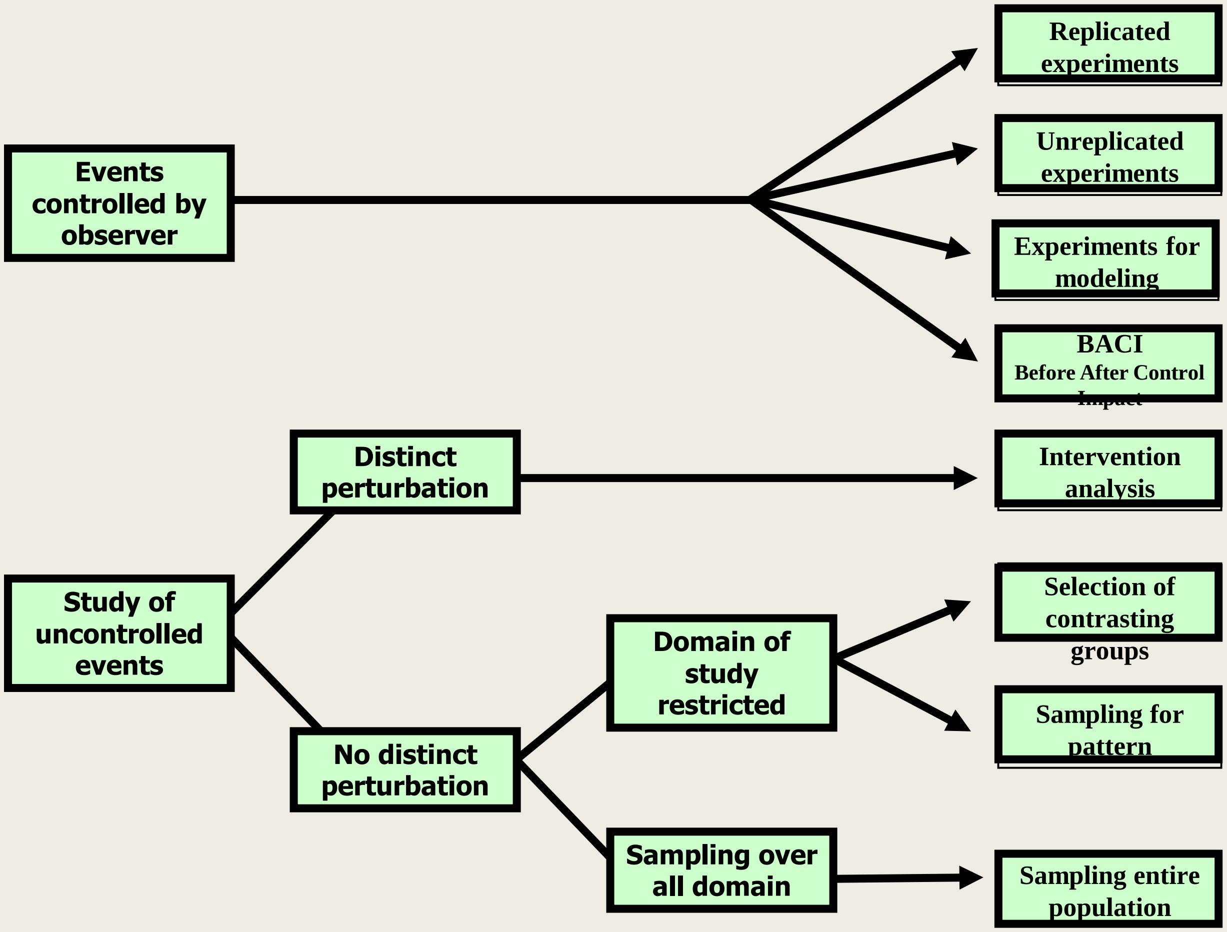
Accuracy and precision

- **Accuracy (bias)** - accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's actual (true) value
  - *How to increase accuracy?*

- **Precision** - precision of a measurement system, also called repeatability, is the degree to which repeated measurements show the same results.
  - *How to increase precision?*

- A measurement procedure can be accurate but not precise, precise but not accurate, neither, or both.

Experimental designs vs sampling design



Stratified random sampling

- The attribute of interest may follow a general trend or have similarity among subunits of space (e.g., watershed) or time (e.g., season)
- Stratification can be employed to increase precision (or more likely) to decrease the number of samples required to attain the desired precision
- Standardized technique in fisheries management and research studies
  - *Based on the assumption that sample data are more similar within subunits than for the resource component as a whole*