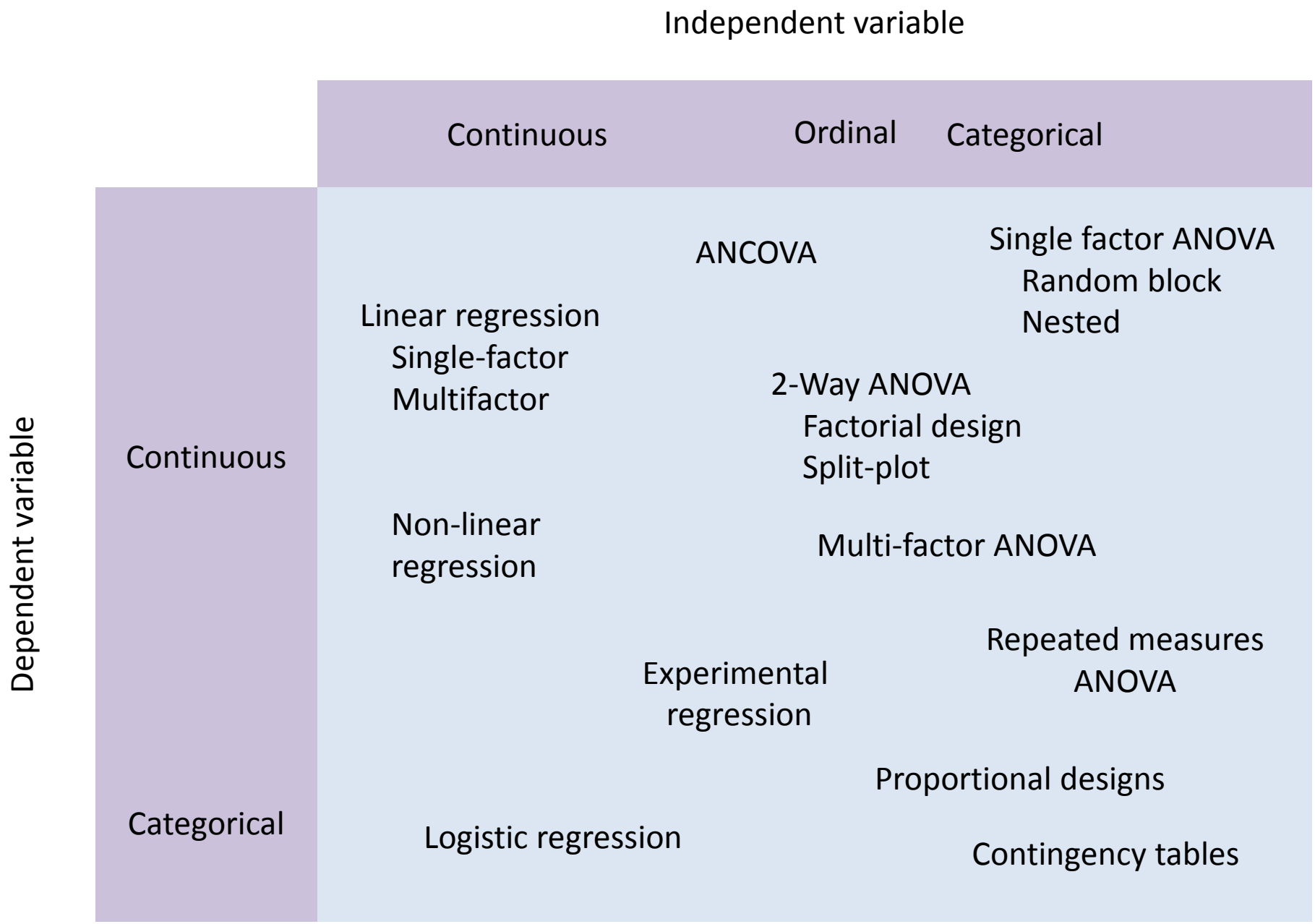
NRES_798_6_2015

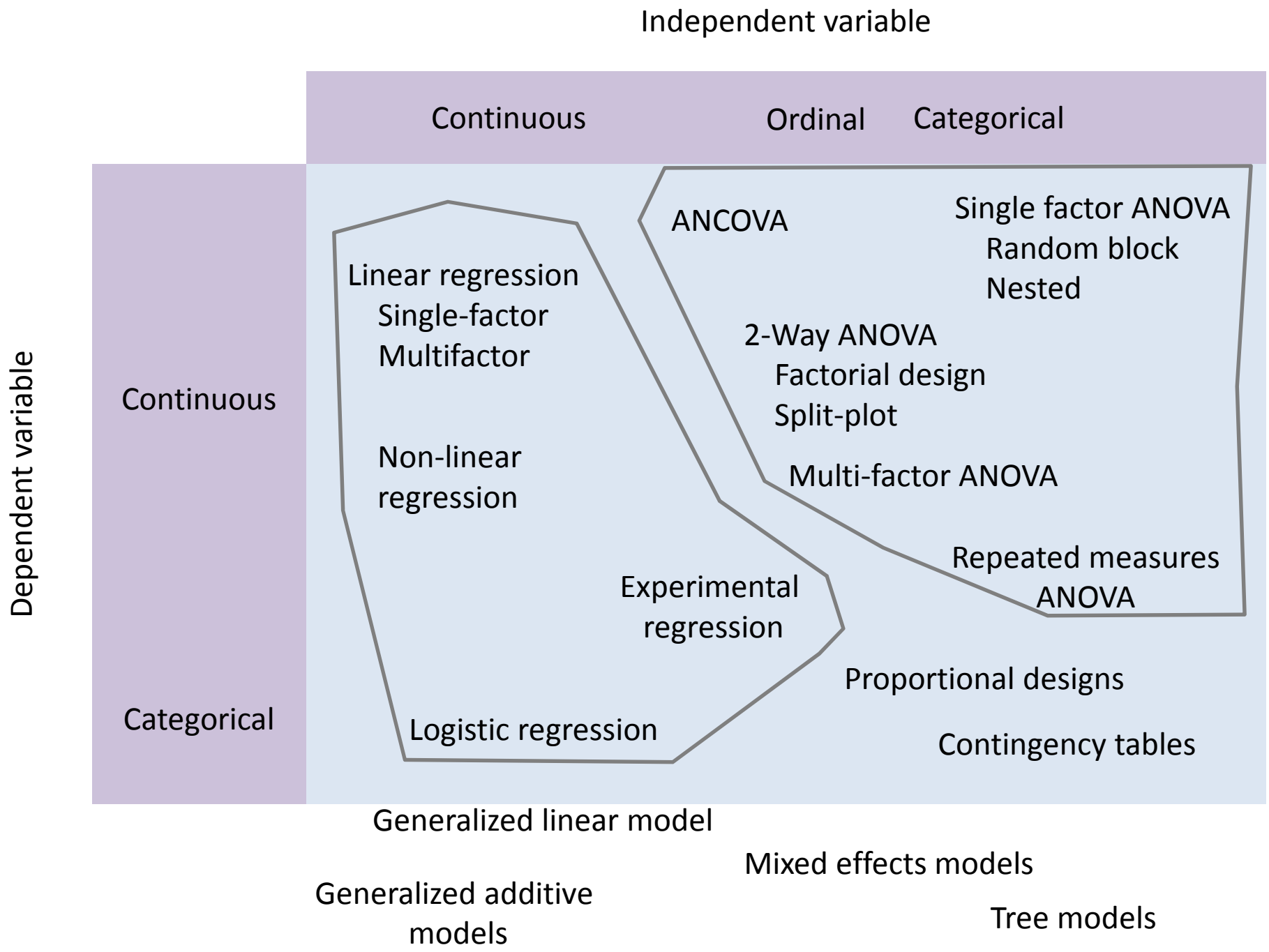
Sampling design and analysis frameworks

Landscape of experimental design

Field experiment	Natural experiment	Observational
Control	Replication	Randomization
Independence		Blocking

		Independent variable: Continuous	Independent variable: Categorical
Dependent variable	Continuous	Regression	ANOVA
	Categorical	Logistic regression	Tabular


Independent variable
Continuous
Ordinal
Categorical
Dependent variable
Continuous
Categorical
ANCOVA
Single factor ANOVA
Random block
Nested
Linear regression
Single-factor
Multifactor
2-Way ANOVA
Factorial design
Split-plot
Non-linear
regression
Multi-factor ANOVA
Repeated measures
ANOVA
Experimental
regression
Proportional designs
Logistic regression
Contingency tables


Independent variable
Continuous
Ordinal
Categorical
Dependent variable
Continuous
Categorical
ANCOVA
Single factor ANOVA
Random block
Nested
Linear regression
Single-factor
Multifactor
2-Way ANOVA
Factorial design
Split-plot
Non-linear
regression
Multi-factor ANOVA
Repeated measures
ANOVA
Experimental
regression
Proportional designs
Logistic regression
Contingency tables
Generalized linear model
Mixed effects models
Generalized additive
models
Tree models

Experiment design

- How replicates are physically arranged in space, and how those replicates are sampled through time

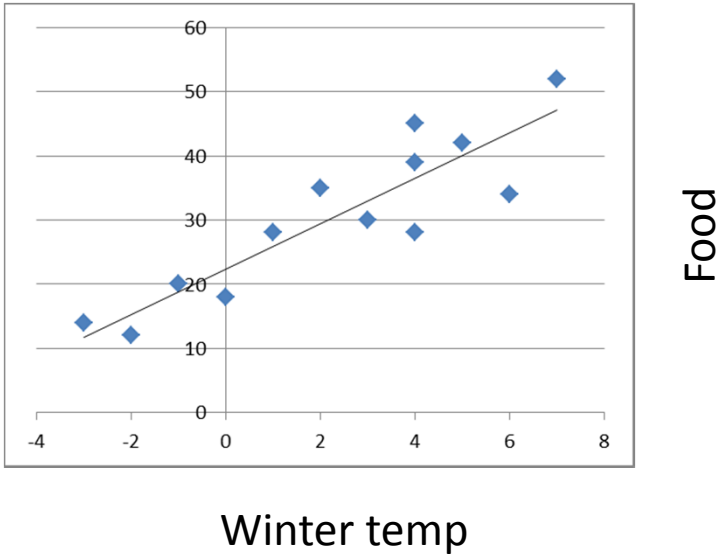
- Focus on the costs and benefits (advantages, disadvantages) of each of the design frameworks
- The most appropriate design will often depend on the nature of the system, the focus of the study, and constraints on sampling

- Underlying question that most experimental design and statistical analysis is how to best assess an partition various sources of variation (uncertainty)

Regression

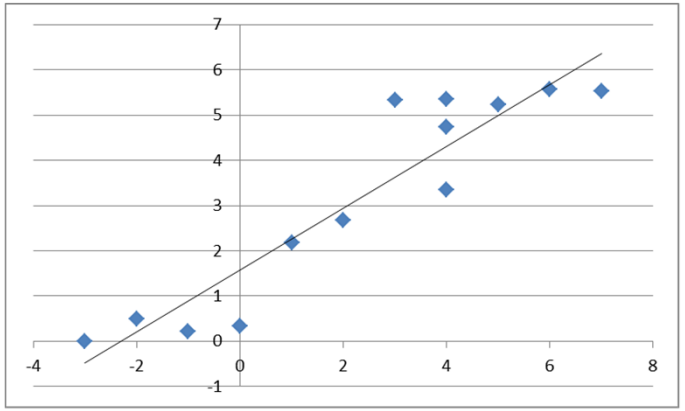
- Correlation vs. regression
  - Correlation
    - Evaluate the relationship between two variables
    - No causality assumed
    - Evaluation of correlations between all variable should be performed during the data exploration stage, prior to performing an analysis

Winter temperature	Food available during the previous summer	Offspring weight

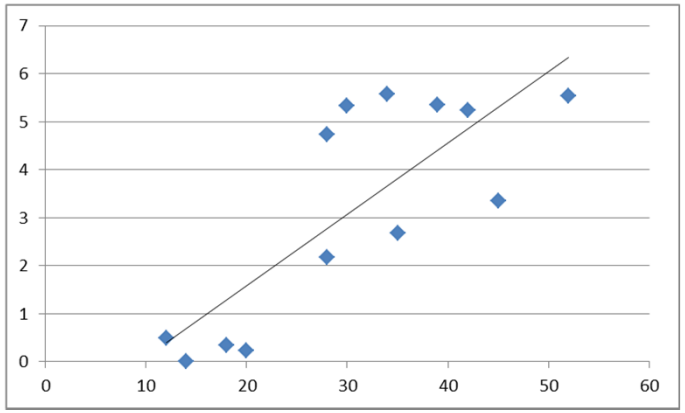


Regression

- Correlation vs. regression
  - Regression
    - Assumed asymmetric functional dependence between variables
    - X value influence Y value
    - Assumes that the value of the independent variable is known exactly
    - regression lines are fitted by only minimizing residual deviations only along the Y axis, while a correlation is fit by minimizing residual deviations along both X and Y



Winter temp.

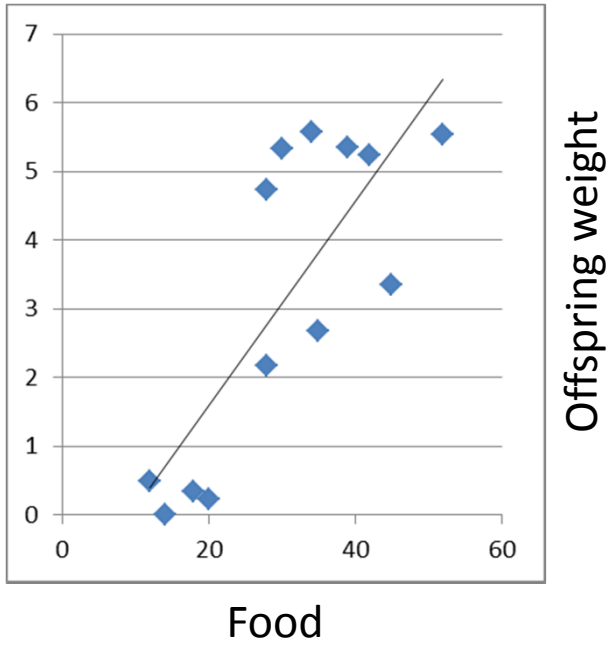


Food

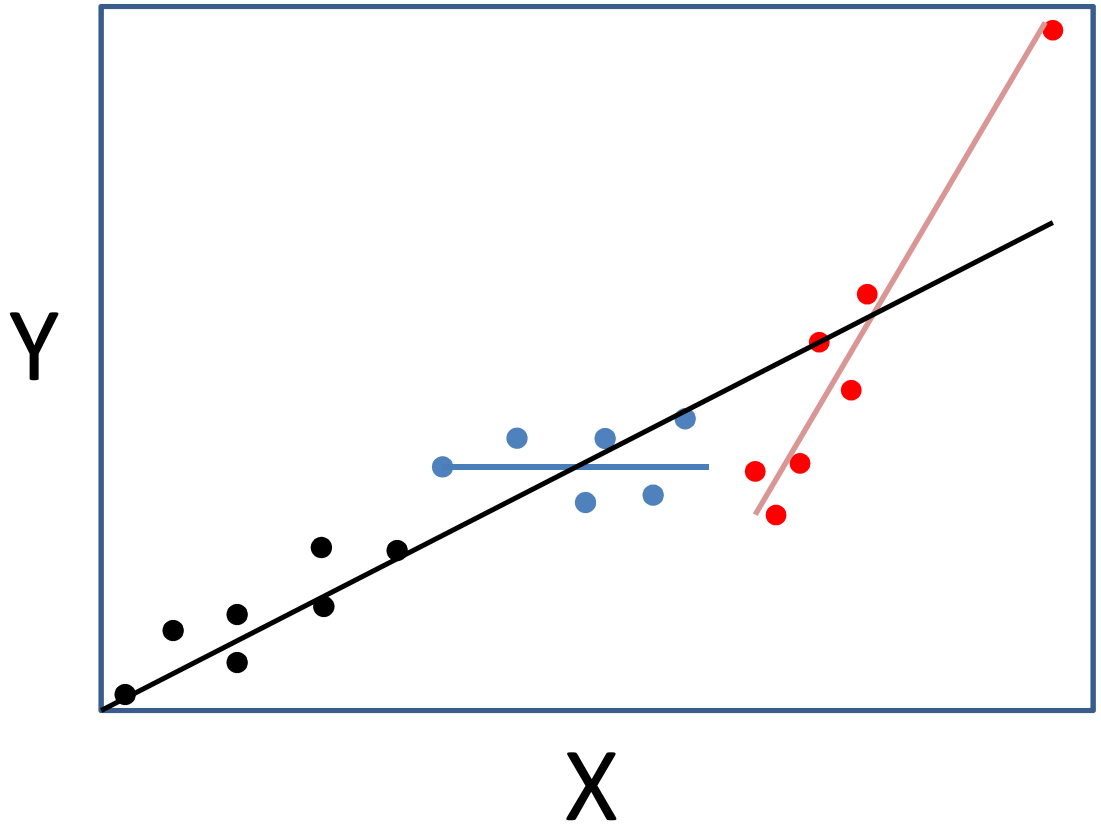
Offspring weight

Regression

- Single-factor regression
  - One continuous independent variable, one continuous dependent variable
    - Causality of independent variable state influencing the dependent variable is assumed
    - Observational study
      - neither variable is manipulated, natural variance in both is used
    - Experimental study
      - levels of the predictor variable are altered (e.g. food addition, predator exclusion), but the realized value of the variable is still measured.
  - Assume
    - Independence of data
    - Replication



Simple regression



Designing regression studies

- Ensure that the range of values sampled for the predictor variable is large enough to capture a full range of responses by the response variable.
- Ensure that the distribution of the predictor values is approximately uniform within the sampled range (What distribution would you expect?)
- Beware of extreme predictor values that can drive fit

Multiple regression

- Two or more continuous predictor variables with one continuous response variable
- Multiple predictor variables should be independent of one another (ideally)
- All predictor variables should be evaluated across the full range of possible values (usually not achievable)

Winter temperature	Food available during the previous summer	Offspring weight

Multiple regression

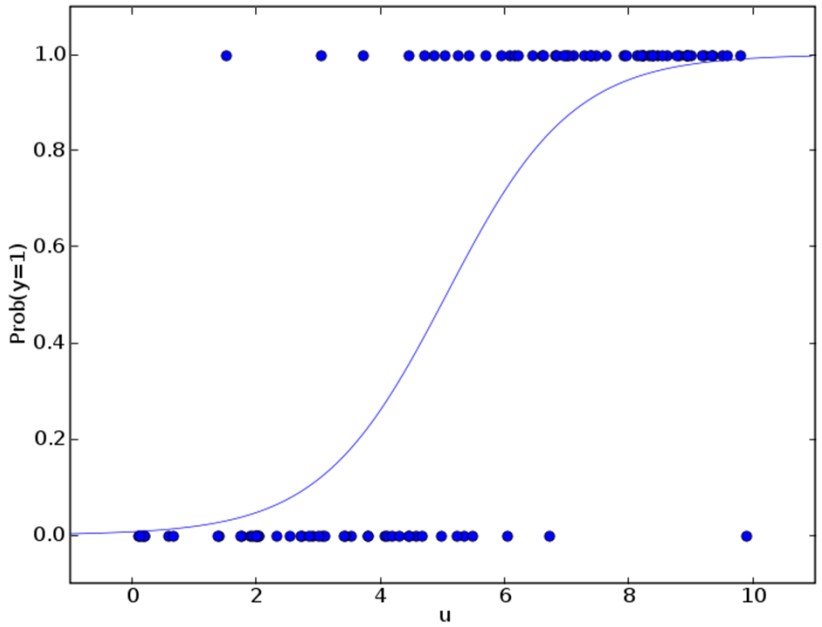
- Collinearity
  - Approximate linear relationship between two predictor variables
  - E.g. Rodents ~ Seeds + Veg Cover
    - In high Veg Cover, seed density is low
- Multicollinearity
  - Many of the predictor variables are correlated with one another

- **Designing multiple regression studies**
  - Replication important: at least 10 replicates for each variable in your study
  - Avoid the temptation to measure and include everything in the multiple regression model

Logistic regression

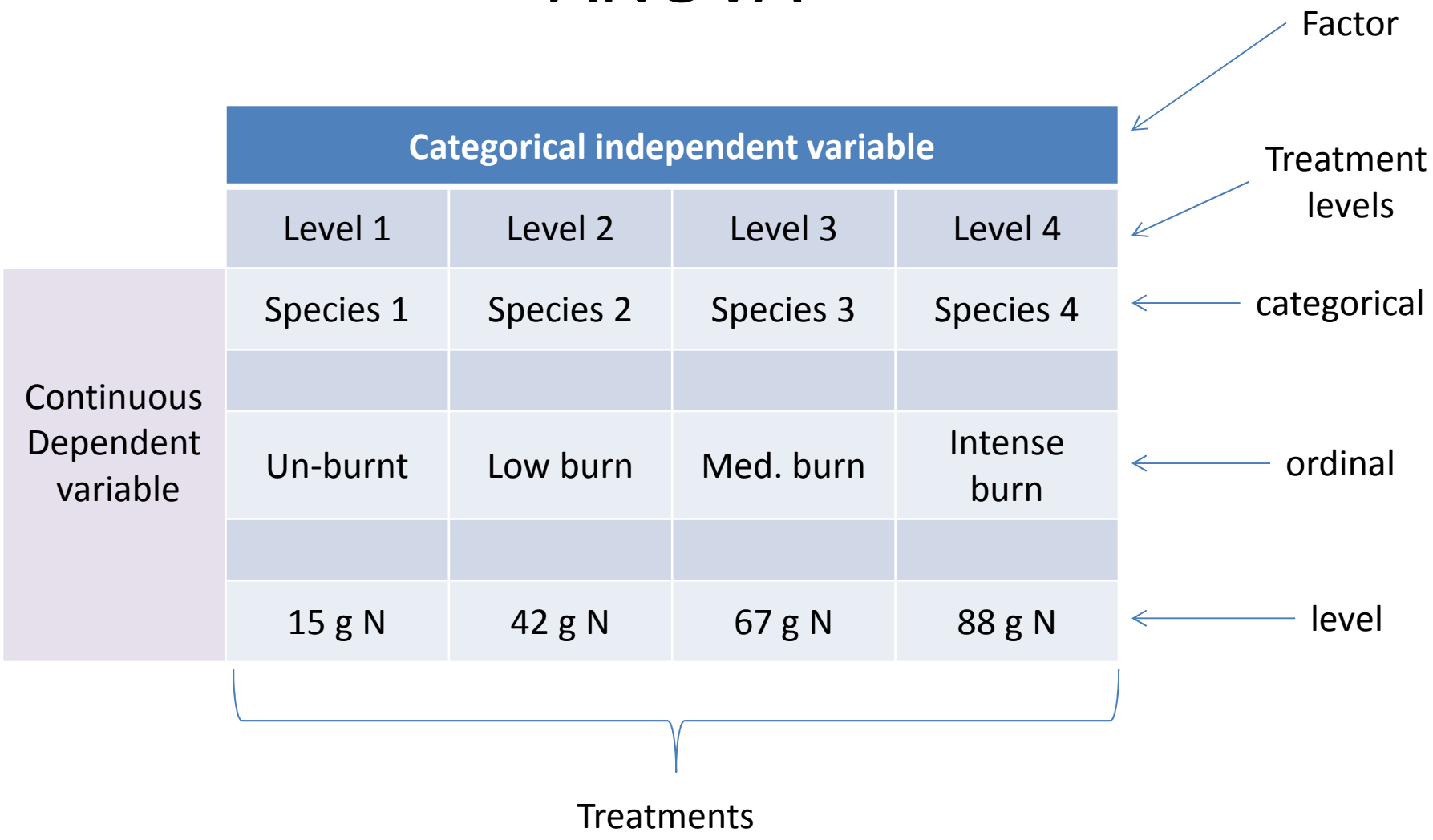
- Dependent variable categorical, independent variable continuous
- Type of generalized linear model
- Uses a “link” function (Logit) to allow the categorical variable to be evaluated within a linear model (regression) framework

Binary Response variable



Continuous predictor variable

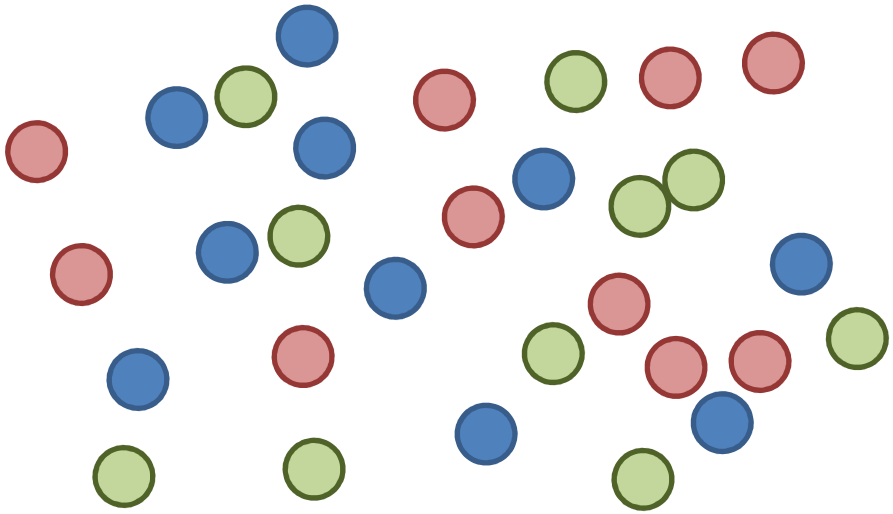
ANOVA


Factor
Categorical independent variable
Treatment levels
Level 1
Level 2
Level 3
Level 4
Continuous Dependent variable
Species 1
Species 2
Species 3
Species 4
categorical
Un-burnt
Low burn
Med. burn
Intense burn
ordinal
15 g N
42 g N
67 g N
88 g N
level
Treatments

ANOVA

- Single factor (one-way) ANOVA

	Treatment (Factor A)		
	0	25	75
Replicates	10	10	10

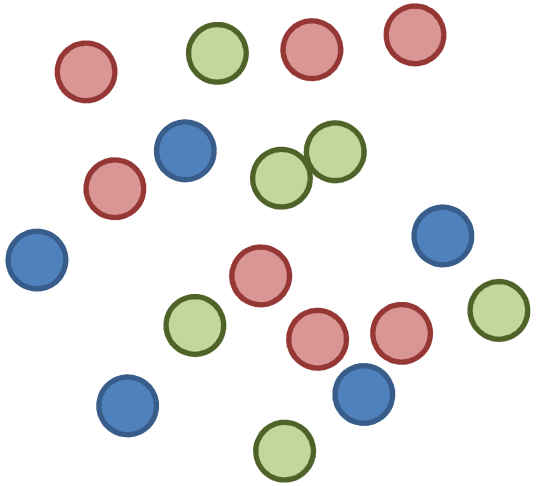


- Aim for equal replication across treatments
- Replicates should be independent
- Randomly allocated on environment

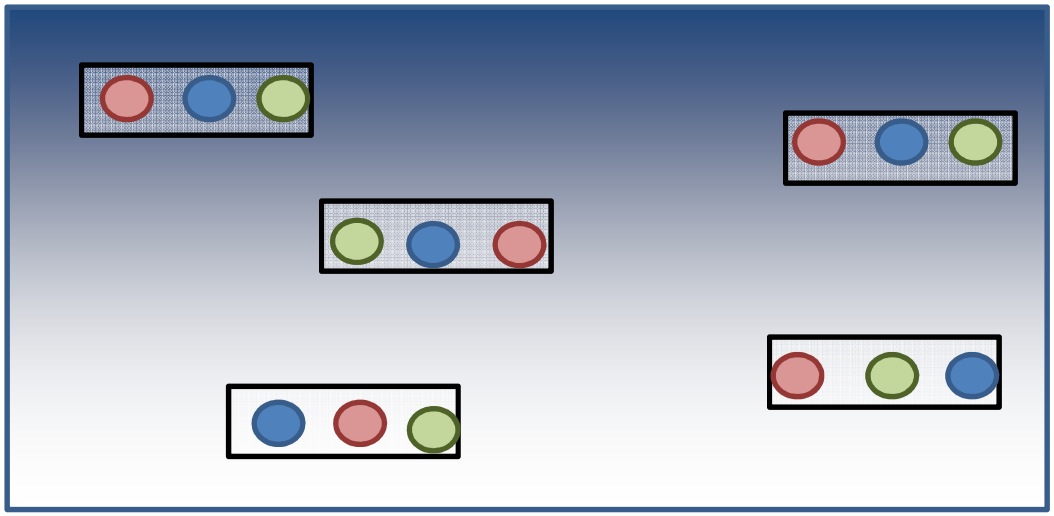
ANOVA

- Single factor (one-way) ANOVA

  - **Advantages of a one-way layout:**
    - Can accommodate unequal replicate number per treatment
    - Powerful test for differences among treatments
  - **Disadvantages of a one-way layout:**
    - Does not explicitly deal with environmental heterogeneity
      - Large unaccounted noise, can reduce power
      - But results can more likely be applied to other areas (i.e. stronger biological effect size, but weaker statistical effect size)
    - Explicitly implies that treatments are of only one factor (i.e. no colinearity with other factors)



ANOVA: Randomized block

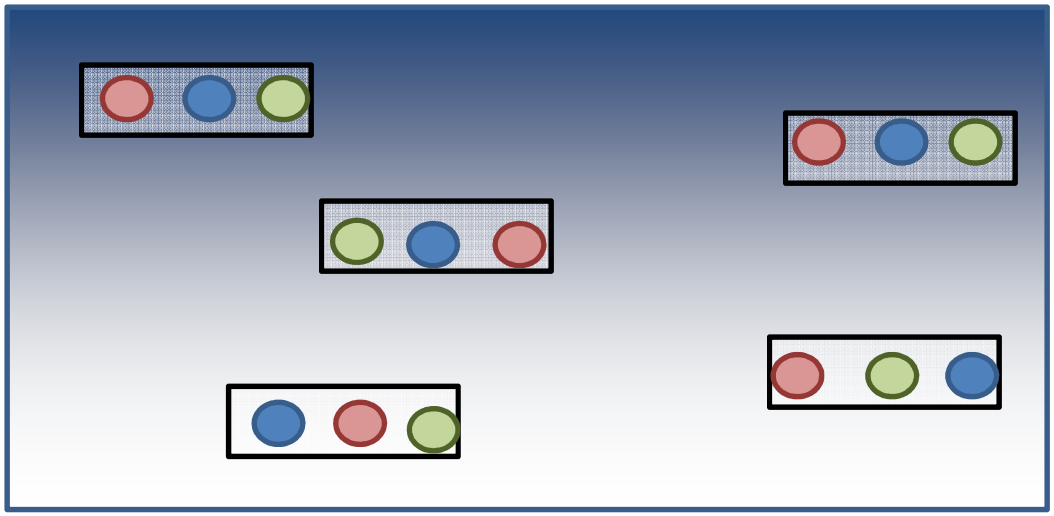


- Effective way of incorporating environmental homogeneity into a one-way ANOVA

- Block: a delineated are or time period within which the environmental conditions are relatively homogeneous.
- Blocks should be randomly distributed across the landscape.
- Environmental conditions should be more similar within blocks than between them
- One treatment replicate per block, random allocation of treatments within block

- **Within a block**
  - Block small enough that environmental conditions are homogeneous (i.e. gradients
  - Block large enough that one replicate of each treatment fits
  - Block large enough that each replicate is independent
  - Blocks randomly assigned is spatial environmental heterogeneity is unknown

ANOVA: Randomized block

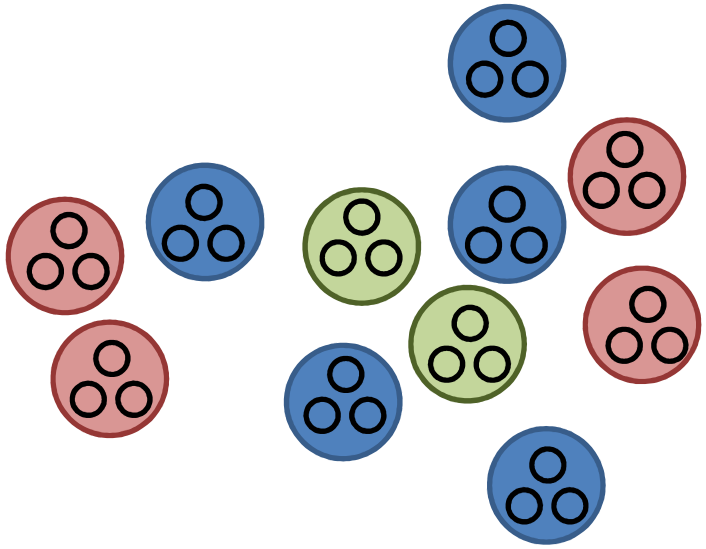


- **Advantages to a Random block design**
  - Effective means for dealing with spatial environmental heterogeneity, or temporal heterogeneity (i.e. necessity to run replicates through time)

- **Disadvantages to a Random block design**
  - If the “block” effect is weak, the statistical power of the test is less
  - Risk of non-independence if blocks are small
  - If data from one replicate is lost, the data from the whole block is compromised
  - The design assumes no interaction between the blocking factor and the treatment. (there is the risk that treatment does interact with environmental state)

ANOVA: Nested design

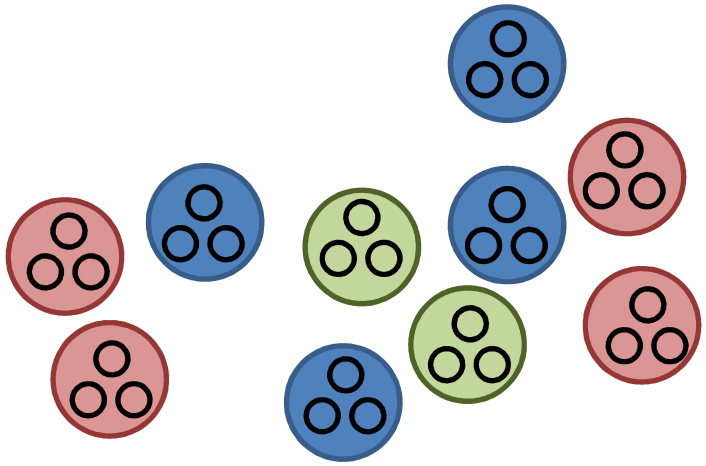
- Sub-sampling within replicates to increase precision
- Sub-samples not independent (nested)



ANOVA: Nested design

- **Advantages to a Nested design**
  - Increase in precision of the estimate for the replicate
  - Allows to test for variation within a treatment (within vs. between)
  - Design is beginning of a hierarchical design (design that allows partitioning of variance within and between hierarchical levels: e.g. within replicate, within river reach, within river, within watershed.

- **Disadvantages to a Nested design**
  - Risks not being analyzed properly (no nesting included in analysis framework, and pseudoreplication)
  - Sample sizes should be equal within each group.
  - Investment in obtaining more replicates is often better than increasing power within a replicate.



Multiple factor ANOVA

		Treatment (Factor A: 3 levels)		
		0	25	75
Treatment (Factor B: 4 levels	0	10	10	10
	0.2	10	10	10
	0.4	10	10	10
	0.6	10	10	10

- In a proper factorial design treatments are fully crossed
- If treatments are not fully crossed design is described as being confounded (i.e. can't properly partition the variance between the factors)

Multiple factor ANOVA

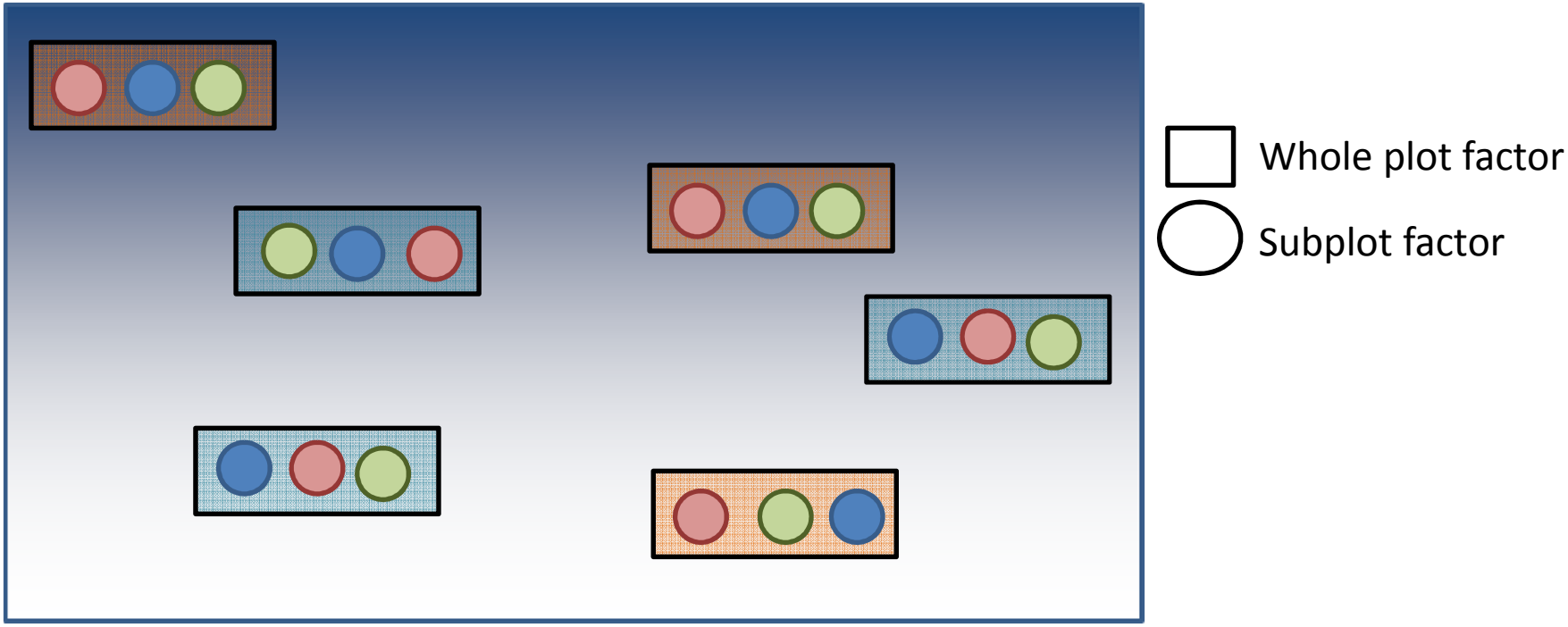
		Treatment (Factor A: 3 levels)		
		0	25	75
Treatment (Factor B: 4 levels	0	10	10	10
	0.2	10	10	10
	0.4	10	10	10
	0.6	10	10	10

- **Advantages of a two-way design**
  - Allows for the main effect and interaction effects to be examined
  - Interaction effect represent non-additive components of response
  - Interaction measures the extent to which treatment combinations act additively, synergistically, or antagonistically.

- **Disadvantages of a two way design**
  - Number of treatment combinations can quickly become too large (for adequate replication)
  - In some ecological circumstances all treatment combinations may not be feasible (all treatment combinations not possible or realistic)

ANOVA: split plot design

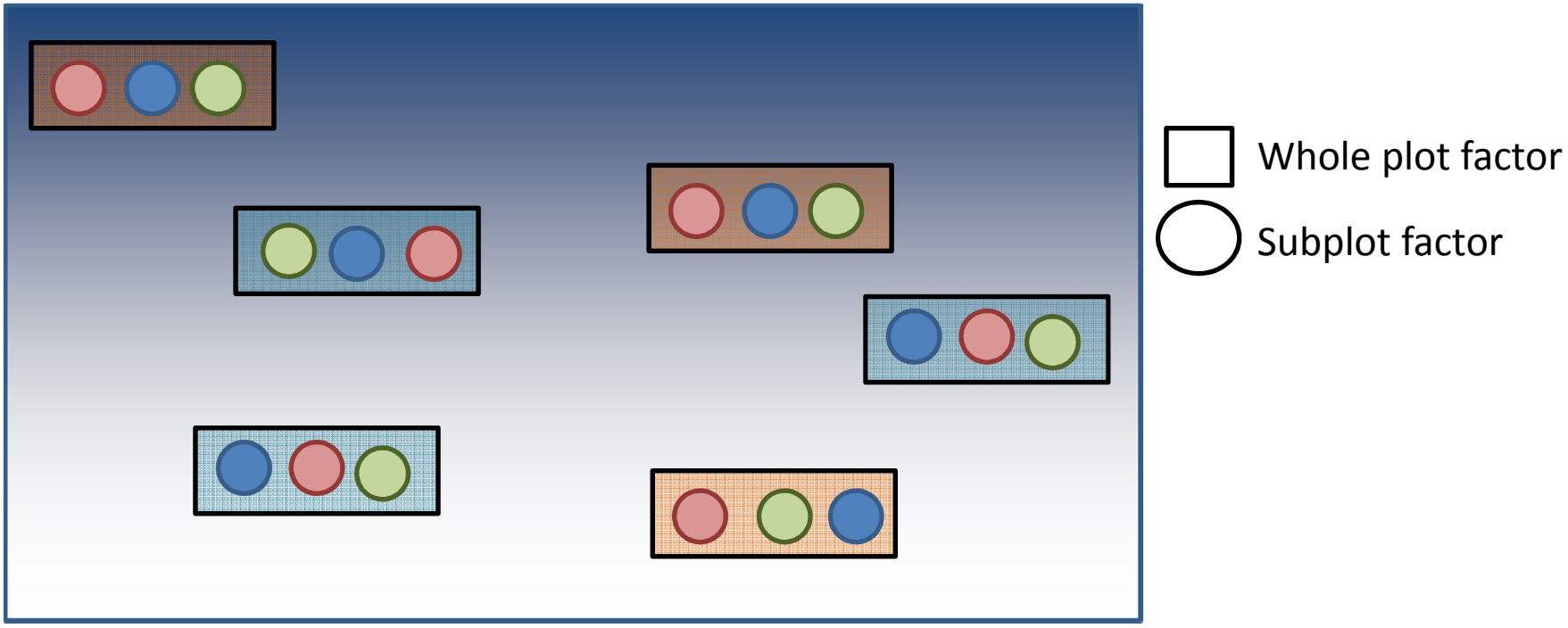
- Extension of the random block design to two experimental treatments.
- Second factor is applied at the level of a block.



ANOVA: split plot design

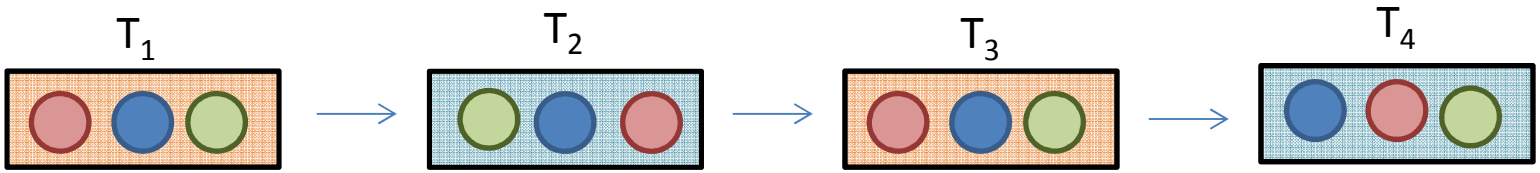
- **Advantages of a split-plot design**
  - Can control for environmental heterogeneity
  - Removes the additive effect of the blocks, and allows for test of the main effects and interactions between the two factors

- **Disadvantages of a split-plot design**
  - You can't test for interactions between blocks and subplot factors



ANOVA: repeated measures

- Repeated measure of replicate through time
- Measures of replicate not independent



- **Advantages of repeated measures**
  - Efficient use of sampling if getting high replicate number is problematic
  - Each replicate functions as its own "block" or control
    - i.e. controlling for size, age, life history etc.
  - Allows for testing of time by treatment interactions
    - i.e. test if differences between treatments change through time

- **Disadvantages of repeated measures**
  - Lack of independence

Environmental impact assessment
BACI design

- Type of repeated measures
- BACI: Before, After, Control, Impact

- Can include replication (spatial and temporal) before and after the perturbation, and blocking factors

Experimental Regression ~ ANOVA

- ANOVA good, but limited and can be divorced from ecological process of interest
- Is the predictor variable naturally categorical, or is just applied that way?

- Experimental regression, increase the number of treatment levels, at the expense of replication

- Range and spacing of treatment levels needs to be well designed
- Allows for a “response surface design”
- Design makes it easier to detect non-linear impacts and thresholds
- A regression framework allows model parameters (intercept, slope) to be estimated

		Treatment (Factor A)						
		0	1	2	3	4	5	6
Treatment (Factor B)	0.0	1	1	1	1	1	1	1
	0.1	1	1	1	1	1	1	1
	0.2	1	1	1	1	1	1	1
	0.3	1	1	1	1	1	1	1
	0.4	1	1	1	1	1	1	1
	0.5	1	1	1	1	1	1	1

Tabular designs

- Categorical independent variable, categorical response variable

- Contingency table analysis
  - Significance tested using chi-square, G-test, etc.

	Right-handed	Left-handed	Total
Males	43	9	52
Females	44	4	48
Totals	87	13	100