QGIS & DATA ANALYSIS

Dr Mark O'Connell

Director, ERT Conservation

Bristol, January 2016

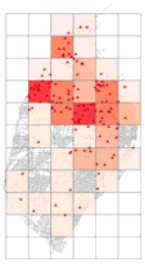
January 2016

ERT Conservation

Common functionality required from GIS (1)

Researchers often want to create….

- Random sample points
- Regular sampling grid
- Random sampling points within regular grid



January 2016

ERT Conservation

Common functionality required from GIS (2)

Researchers also want to measure….

- Number (count of individuals)
- Length
- Area: planar or topological
- Perimeter
- Density
- Environment: landscape features at particular point
- Proximity (distance to): between individuals & landscape features

January 2016

ERT Conservation

Why…

- Describe ***and*** understand
- Show differences, relationships, proportions

January 2016

ERT Conservation

Next part of course….

Exercises

- Operations already done
- Some new operations
- *Using* GIS = whole process + stats !!

January 2016

ERT Conservation

DATA & ANALYSIS

a quick word….

January 2016

ERT Conservation

Statistics (describe & understand)

Your new mantra….

- Statistics are **good**
- Statistics are **fun**
- Statistics are **easy**

January 2016

ERT Conservation

Three core test statistics

1. Tests to show ***differences***
2. Tests to identify ***relationships***
3. Tests to compare ***proportions***

January 2016

ERT Conservation

Types of data: scales of measurement

Four key types

1. Nominal scale data
2. Ordinal scale data
3. Interval scale data
4. Ratio scale data

January 2016

ERT Conservation

Nominal scale measurements

Classifying samples into mutually exclusive ***categories***

Examples

- Taxonomic groups such as *Coleoptera*, *Collembola* & *Diptera*
- Habitat groups such as *forest, grassland & urban*
- Male or female
- Colour groups

January 2016

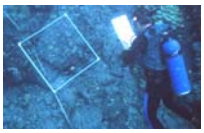
ERT Conservation

Ordinal scale measurements

Placing mutually exclusive categories into *rank order*

Example

The DAFOR scale of abundance



- *Dominant*
- *Abundant*
- *Frequent*
- *Occasional*
- *Rare*

January 2016

ERT Conservation

Interval & Ratio scale measurements

Placing observations (measurements) into rank order **and** saying *how far apart* they are

Example without absolute zero (interval)

- Temperature (Celsius)

Examples with absolute zero (ratio)

- Length
- Weight

January 2016

ERT Conservation

Scales of measurement

Why this is important for GIS studies…….

The *scale of measurement* you use limits the ***type*** of statistical analyses that are possible for your data

DO THIS *BEFORE* COLLECTING DATA !!

January 2016

ERT Conservation

Variables

Response variables

- The measure that responds to '*conditions*'
- Dependent variable
- Y-axis (vertical)

Explanatory variables

- The environmental *condition* to be measured (they drive and *explain* the response)
- Independent variable
- X-axis (horizontal)

January 2016

ERT Conservation

Which is which – how to remember

Response variables

- The measure that responds to '*conditions*'
- Dependent variable
- Y-axis

Explanatory variables

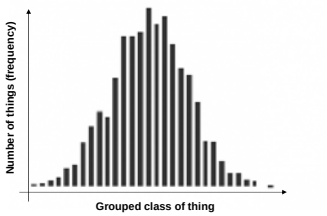
- The environmental *condition* to be measured
- Independent variable
- X-axis

January 2016

ERT Conservation

Frequency diagram

A simple but *VERY* important data representation in statistics



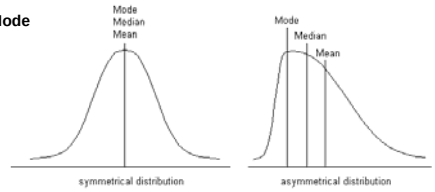
January 2016

ERT Conservation

Central tendency

Three commonly used measures of central tendency

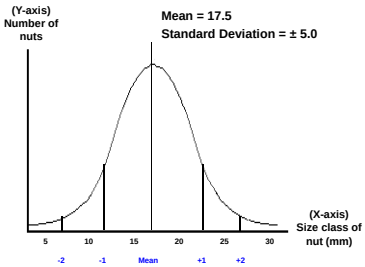
- **Mean**
- **Median**
- **Mode**



January 2016

ERT Conservation

The normal distribution



January 2016

ERT Conservation

Residuals

Residual = the mean value minus any particular value

$(\bar{X} - X)$

Also known as the ***deviation from the mean***

January 2016

ERT Conservation

Standard deviation (s)

We often need to know how much variability there is about the central tendency i.e. *how spread out is our frequency diagram* ?

We need to know ***average*** deviation from mean *(residual)*

January 2016

ERT Conservation

Standard deviation (s)

How do we calculate a mean (average) ?

Mean = $\frac{\text{sum of all measurements}}{\text{number of measurements}}$

Mean = $\frac{\sum X}{n}$

Average deviation from mean *(residual)*

Mean deviation = $\frac{\sum(X-\bar{X})}{n}$

$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$

Degrees of freedom

January 2016

ERT Conservation

Standard deviation (s)

Always report as…..

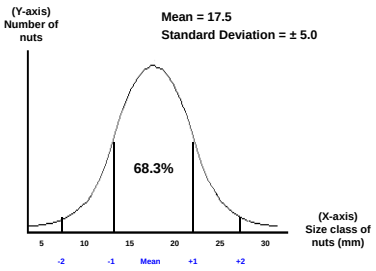
Mean ± standard deviation

Avoid the term <u>average</u>

January 2016

ERT Conservation

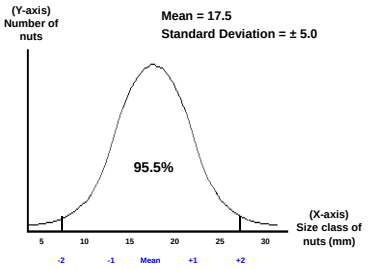
The normal distribution



January 2016

ERT Conservation

The normal distribution



January 2016

ERT Conservation

The basis of all statistical tests

- Compare an observation against a 'Test Statistic'
- Nowadays, stats software does this for you !!!
- Many Test Statistic types (but all work on same general principles)

THAT *P* VALUE… why so important?

January 2016

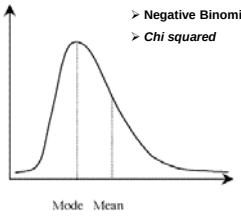
ERT Conservation

Also note……

There are other 'distributions' with their own *properties*

Examples

- Poisson distribution
- Negative Binomial
- *Chi squared*



January 2016

ERT Conservation

Parametric & non-Parametric statistics

- Some statistical tests require data to be normally distributed. They have this (and other) requirements.
- These requirements are called the parameters of the mathematical models used in the test.
- Tests that have these requirements are called Parametric tests.
- Tests that do not have these requirements are called Non-Parametric tests.

January 2016

ERT Conservation

Also note……

- Tests for normality
- Data transformation

January 2016

ERT Conservation

EXAMPLES: *types* of statistical tests in GIS

1. Tests to show ***differences***
2. Tests to identify ***relationships***
3. Tests to compare ***proportions***

January 2016

ERT Conservation

Scenario for examples of generic statistical test types

- ***Gloucester Wildlife Trust*** have been using pit-fall traps to monitor ***beetle*** numbers at ten forest sites in local nature reserves
- Trapping occurs monthly, with ten years of data available
- Halfway through the ten years of monitoring five sites received management aimed specifically at enhancing habitats for beetles

January 2016

ERT Conservation

1. Tests to show *differences*

Monitoring question

Does forest management increase *beetle* numbers

Statistical question

Are there significant differences in the number of beetles caught before and after management

Statistical method

Test for differences in mean (or median) *beetle* numbers caught before and after management

Explanation on flip chart……
Mention ANOVA….

January 2016

ERT Conservation

2. Tests to identify *relationships*

Monitoring question

Have *beetle* numbers in the forest changed with time

Statistical question

Is there a significant relationship (trend) between the number of *beetles* caught and year

Statistical method

Use regression to test for significant (non-zero) slope in graph of *beetle* numbers (y-axis) and year (x-axis)

Explanation on flip chart……

January 2016

ERT Conservation

3. Tests to compare *proportions*

Monitoring question

Are *beetle* numbers at the forests seasonal

Statistical question

Are the proportions of *beetle* in traps significantly different for either spring, summer, autumn or winter

Statistical method

Use Chi squared contingency table (or G test) to test for data homogeneity between seasons

Explanation on flip chart……

January 2016

ERT Conservation

Take home message

- Variables all derived *directly* from a GIS
- We are going to do some worked examples

January 2016

ERT Conservation

Multivariate analyses (models)

Many techniques…

- Generalised Linear [mixed] Models (GLMs)
- General Estimating Equations (GEEs)
- Multiple Regression
- General Additive Models
- *Etc*

Variables also derived *directly* from a GIS….

January 2016

ERT Conservation