


Practical Stats  
Statistics, down to earth

# What's New in Statistics?

## 2. Free Statistics Software

Dennis R. Helsel  
PracticalStats.com

1



Practical Stats  
Statistics, down to earth

# What's New that My Employees Should Know About?

1. Permutation Tests. Never worry about a normal distribution again.
2. Free software for environmental statistics
3. Better methods for handling nondetect data
4. New and better methods for finding the best regression line.

2



## 2. Free software for environmental statistics

R is the world standard in statistical software today

It is completely free (open-source GNU license)

Most requested package when we teach courses because

1. Its free
2. Its free
3. It has everything

3



## What is R?

Free, open-source software

Modeled after S, a statistics language developed at Bell Laboratories in late 1980s (R is newer)

Written collaboratively by teams of volunteers

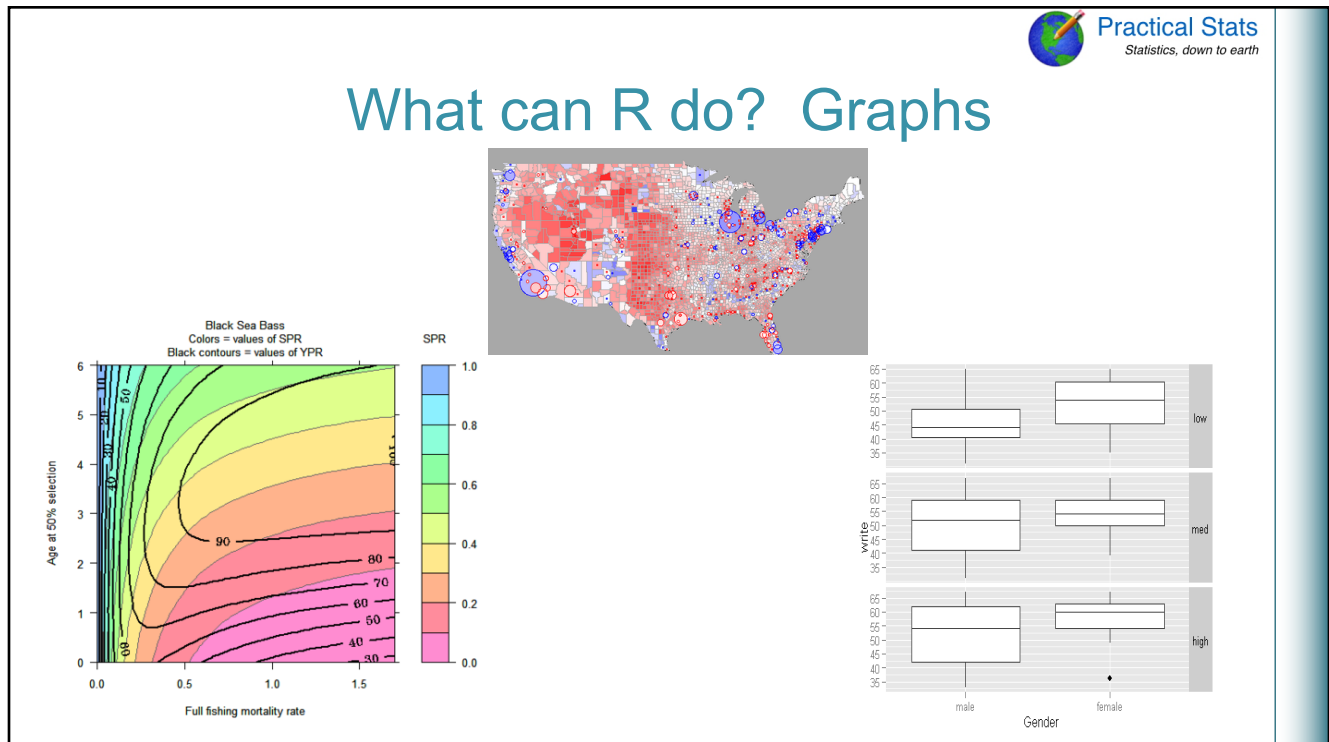
Broad suite of statistical methods

A programming language; you can develop your own routines (scripts)

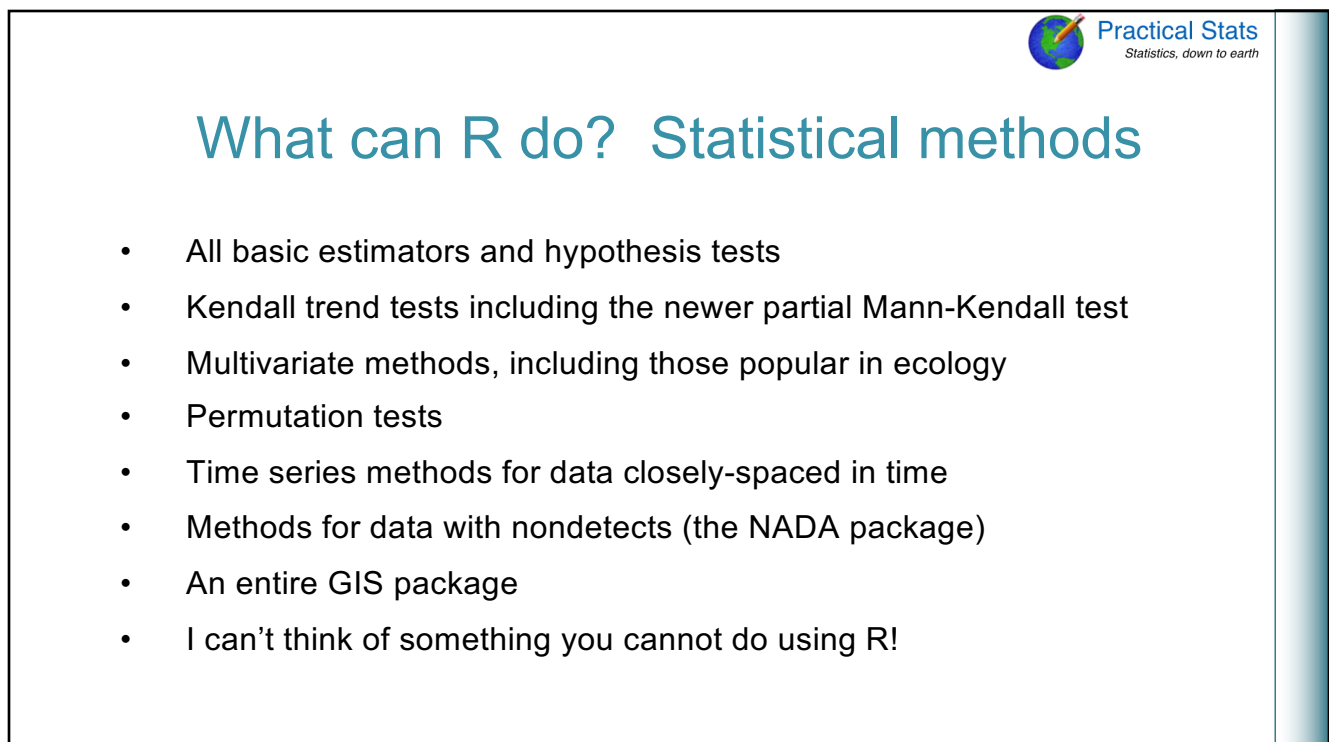
Difficulty much the same as SAS; some are put off by the default command line interface

However several GUI front ends are available that make it similar in style to commercial software

4



5



6



## Where can I get R software?

<http://cran.r-project.org/>

- For Linux, Windows and Mac OS
- Download the binary version for your OS
- Checked for viruses
- 'No one to call and yell at' but wikis and mailing lists provide much useful information and support
- Will also find free and low-cost books (pdf) onsite

7



## Importing Excel Datasets

Using the GUI interface you can directly read .xls and .xlsx datasets, as well as many other formats.

**Data > Import Data > Excel**

Read in the Green.xls file

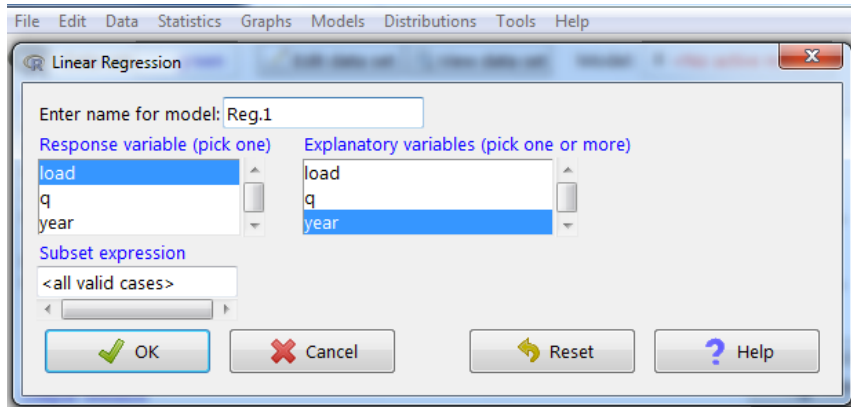
	year	load	q
1	1952	1070.89	1431.20
2	1953	357.52	639.81
3	1954	295.19	384.27
4	1955	744.41	975.47
5	1956	736.13	1055.61
6	1957	539.64	949.49
7	1958	771.53	1285.05
8	1959	303.81	630.79
9	1960	378.89	856.19
10	1961	429.81	948.60
11	1962	940.23	1243.40
12	1963	348.65	681.59
13	1964	426.99	633.86
14	1965	499.53	996.69
15	1966	272.62	566.71
16	1967	807.46	1483.09
17	1968	496.16	1049.41
18	1969	112.74	492.74
19	1970	411.82	856.47
20	1971	473.78	1146.52
21	1972	494.03	1120.90

8



# Stats with R: Regression

Statistics > Models > Linear Regression

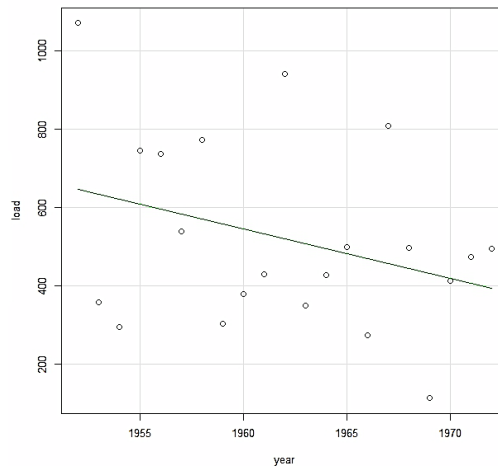


9



# Regression using Rcmdr

Graphs > Scatterplot



10

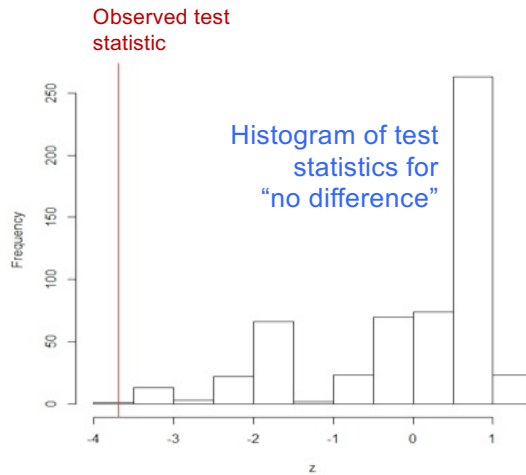


## Permutation Tests using R

Molybdenum concentrations before and after treatment in groundwater

t-test cannot see difference in the two means due to skewness (3.94 before, 0.25 after)

Perm test  $p=0.0017$ . Strong evidence of a difference between means



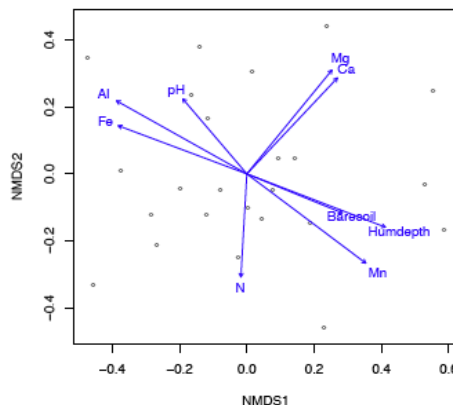
11



## R packages: Vegan

<http://cc.oulu.fi/~jarioksa/softhelp/vegan.html>

biplots, NMDS, correspondence analysis, PCA, factor rotation, permutation tests for correlation between groups of data.....



12



# R online: integration of processes

## CSIRO, Australia. Auto generate reports

**WQSAR**  
Water Quality Statistical Analysis and Reporting Tool

Start over (Home) | HELP | Glossary

Home Choose a Station Choose Data Inspect Data Final Data Analysis Get Results

Welcome to the Water Quality Statistical Analysis and Reporting Tool (WQSAR).  
You can use this tool to track water quality trends and generate [reports](#) on these for freshwater and estuarine systems all over Australia; as well as comparing results from water quality multiple monitoring stations (multi-site analysis).

Please read [how to use this tool](#) before you start.

Please select a service (source of water quality data), either:

- from the drop-down list below
- or by entering your own service URL.

(Learn more about [selecting or setting up a service](#)).

13



# R online: integration of processes

## Choose parameters for on-demand reports

**WQSAR**  
Water Quality Statistical Analysis and Reporting Tool

Start over (Home) | HELP | Glossary

Home Choose a Station Choose Data Inspect Data Final Data Analysis Get Results

View and Save Graphs and their Associated Data

The data exploration graphs below will help you to assess the suitability of raw data for analysis. Links to download the graphs and data/parameter files are available beneath the graphs.  
(Learn more about [inspecting the data](#)).

BACK NEXT

14



# R online: integration of processes

**WQSAR**  
Water Quality Statistical Analysis and Reporting Tool

Start over (Home) | HELP | Glossary

Home Choose a Station Choose Data Ingest Data Final Data Analysis Out Results

### Choose methods and transformations

(Learn more about [setting up the analysis parameters](#) or read a [report](#), discussing the uses of the trend methods below).  
NOTE that data run through analysis may be visible to other users.

**Trend methods - choose one or more:**

- Linear regression
- Seasonal Kendall's Tau / Mann-Kendall Test
- Generalised Additive Model

**Description of the methods:**

A parametric linear regression model is used to estimate the linear trend, after allowing for seasonal effects and possibly the effect of flow where available.

Seasonal Kendall's Tau is a nonparametric method for testing for monotonic trend based on the Kendall rank correlation. Seasonal Kendall's Tau is an extension that accounts for seasonal effects.

GAM is a semi-parametric regression model used to estimate a flexible non-linear trend, after allowing for seasonal effects and possibly the effect of other factors.

**Select Data Transformation:**

Water quality variable transform:

Replace zeros in water quality attribute data column with no-data values.

**(Generalised Additive Model Only)**

Select Flexibility in Non-Linear Trend:

Flexibility for non-linear time trend:

15



# R online: integration of processes

Get completed on-demand report

**CSIRO Water Quality Statistical Analysis and Reporting Tool**

**Water Quality Trend Analysis for Station 123456**  
Euston at Euston

Site ID: 123456  
Site Name: Euston at Euston  
Water Quality Variable: Conductivity  
Date Range: 1/1/2000 to 31/12/2010

**Table 1: Summary of station details.**

**Monthly Mean Data Plots**

**GAM Trend Analysis**

The fitted GAM trend is:  $14.43 \text{ mg/L per year}$ . The significance level is 0.05 for the trend  $t$ -test  $t = 2.002$  which means the likelihood of such a trend occurring by chance is approximately less than 1 in 1000. The significance level (p-value) of the regression test is 0.0499. The water flow data were not included in the analysis as there is no flow data available for the period of the analysis.

**Linear Trend Analysis**

The fitted linear trend is:  $14.43 \text{ mg/L per year}$ . The significance level is 0.05 for the trend  $t$ -test  $t = 2.002$  which means the likelihood of such a trend occurring by chance is approximately less than 1 in 1000. The significance level (p-value) of the regression test is 0.0499. The water flow data were not included in the analysis as there is no flow data available for the period of the analysis.


**Non-Parametric Trend Analysis**

The fitted non-parametric trend is:  $14.43 \text{ mg/L per year}$ . The significance level is 0.05 for the trend  $t$ -test  $t = 2.002$  which means the likelihood of such a trend occurring by chance is approximately less than 1 in 1000. The significance level (p-value) of the regression test is 0.0499. The water flow data were not included in the analysis as there is no flow data available for the period of the analysis.

Method	GAM coefficient	Linear	Linear-trend	Non-parametric	Non-parametric coefficient
Linear regression	14.43	14.43	14.43	14.43	14.43
Seasonal Kendall's Tau	14.43	14.43	14.43	14.43	14.43
Mann-Kendall Test	14.43	14.43	14.43	14.43	14.43
Generalised Additive Model	14.43	14.43	14.43	14.43	14.43

16



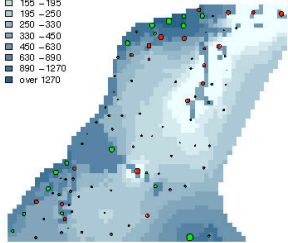

**Practical Stats**  
*Statistics, down to earth*

## R spatial analysis

### Universal kriging predictions

**Flood frequency UK model**

under 130
130 - 195
195 - 250
250 - 330
330 - 450
450 - 630
630 - 890
890 - 1270
over 1270




Of course, the resolution of the grid of prediction locations means that the shift from flood frequency class 1 to the others is too "chunky", but the effect of flood water "backin up" creeks seems to be captured:

```
> image(BMcD_grid, "UK_pred",
+       breaks = brks, col = cols)
```

Universal kriging adds a regression type of prediction relationship to the kriging process

17


**Practical Stats**  
*Statistics, down to earth*

## Summary

- R is a powerful and comprehensive package for statistics, for visualization, and for linking computer processes together to create information online
- Its free!
- Statisticians around the world use it. Recommended by USEPA's Research Division.
- With new interfaces, R is no more difficult to use than other stat software

18



## Summary: R Software

R can save corporations and agencies a great deal in software costs

Commonly used in universities. Employees with recent college experience in statistics and computer science are likely to know R

Scripts can be shared among offices, standardizing operations

19



## Resources – more info on these topics

Free Newsletters




<http://www.PracticalStats.com/news/>

Webinars and Training Courses


<https://PracticalStats.teachable.com>



our Applied Environmental Statistics course(s) are also an introduction to using R software.

 <p><b>Applied Environmental Statistics 1</b> Make Sense of Your Data</p> <p>Dennis Hiesel \$650</p>	 <p><b>Applied Environmental Statistics 2</b> Regression Models and Trend Analysis</p> <p>Dennis Hiesel \$650</p>
 <p><b>Bundle: AES 1 &amp; 2</b> Make sense of your data, from the basics through trend analysis.</p> <p>2 Course Bundle \$1,200</p>	<p><b>Intro to R</b></p> <p>Getting started with the free statistics software</p> <p>Dennis Hiesel FREE</p>

20




Practical Stats  
Statistics, down to earth

# Questions?

For answers to questions from the live broadcast --  
see the Q&A pdf file underneath this video







21



Practical Stats  
Statistics, down to earth

## More Details Are On Our Online Training Site

<https://practicalstats.teachable.com>  
where you'll find more videos and courses.

 <p><b>Applied Environmental Statistics 1</b> Make Sense of Your Data Dennis Helsel \$650</p>	 <p><b>Applied Environmental Statistics 2</b> Regression Models and Trend Analysis Dennis Helsel \$650</p>	 <p><b>Nondetects And Data Analysis (NADA)</b> Statistical analysis of censored data (data with nondetects) Dennis Helsel \$695</p>
 <p><b>Bundle: AES 1 &amp; 2</b> Make sense of your data, from the basics through trend analysis 2 Course Bundle \$1,200</p>	 <p><b>Intro to R</b> Getting started with the free statistics software Dennis Helsel FREE</p>	 <p><b>7 Perilous Errors in Environmental Statistics</b> Step Your Methods Into The 21st Century! Dennis Helsel FREE</p>

22