.

Presentation of results

Environmental Sampling and Analysis

G. Quinn and M. Keough

Design by M. Logan

About

This presentation is intended to be used as a brief summary of basic statistical concepts. It should enable users to identify which concepts require revision.

Links…

Throughout the presentation, marks in the form of (qk2002, …) provide references to sections within the recommended statistical text;

Quinn, G. P. and Keough, M. J. (2002). Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge.

Words and phrases in blue type face provide links to popups that contain additional information and or definitions.

Navigation…

Navigation buttons on the right hand side of each page provide (from top to bottom) 'Previous Page', 'Next Page', 'First Page', 'Last Page', 'Go Back' and 'Quit' navigational shortcuts.

Presentation of results

The aim of presenting results is to report and validate your main findings. Therefore the following results could be presented;

- Statistical outcome of hypothesis testing (Presentation of analyses to validate results)
- Data summaries (Tabular or graphical)
- Trends in data (Tabular or graphical to accompany statistics and efficiently illustrate findings)

- Each result must be textually reported and validated in the results section.
- Validation involves referring to statistical analyses and/or a table or figure (graph).
- Tables and figures should be accompanied by an adequate caption so that they can largely stand alone from the main text and be meaningful.
- Tables and figures should appear embedded within the results section where practicable.

MONASH UNIVERSITY

- Your job to refer the audience to a table or figure and highlight exactly what point(s) they should get out of it

Presentation of analyses

Outcome of hypothesis testing

- Results and discussion are often centered around the outcome of some statistical analysis
- Each reported and discussed result **must** be backed up with an appropriate summary of the statistical output, that shows that your claims are based on statistics and not just opinion
- Note that a graph is not usually a statistical result

MONASH UNIVERSITY

Statistical summaries

Statistical packages often churn out copious amounts of results that are either redundant or not always relevant to all situations (particularly for simple tests). Following is a list of items that should always be reported:

- The degrees of freedom (df) gives an indication of sample size
- The test statistic (e.g. *t*-value, *r-value, F*-ratio)
- The actual *P*-values (when $P \geq 0.001$) allow readers to use any α for testing H_0

Example – Regression analysis

Below is a standard regression output from a SYSTAT. Non-bold items are redundant.

Dep var: LOGLIMP N: 40 Multiple R: 0.345 Squared multiple R: **0.119**

Adjusted squared multiple R: 0.096 Standard error of estimate: 0.373

Effect	Coef	SE	StdC	Tol	t	P
CONSTANT	**1.072**	0.083	0.0	.	12.994	0.000
ALGAE	**-0.006**	0.003	-0.3	1.0	-2.265	0.029

Analysis of Variance

Source	SS	DF	MS	F	P
Regression	0.713	**1**	0.713	**5.129**	**0.029**
Residual	5.282	**38**	0.139		

Example – Regression analysis

Much of the results in the tables on the previous slide are redundant or irrelevant. Greater, efficiency, clarity and use of space could be achieved by summarizing as follows:

The number of limpets fell as algae cover increased, although algal cover only explained 12% of the variation in limpet abundance (equation: log(limpets) = 1.076 – 0.006 X algal cover, $F_{1,38}$=5.129, P=0.029, r^2=0.119).

Example – Regression analysis

Further points of note

- Both *t*-value and *F*-ratio offer same outcome and thus only one is required
- Actual variables (c.f. shorthand variable names) are used in equations etc
- Any transformations should be reflected in the reported equations
- $F_{1,38}$ is shorthand for *F*-ratio with 1 and 38 degrees of freedom

Presenting data - Tables

Tabular summaries

Tables are used to summarise the following

- The results of complex analyses (e.g. complex linear models, correlation matrices)
- Qualitative data (e.g. descriptions)
- Comparing a large number of groups

Note that trends in data are usually more effectively illustrated in graphs than tables and that data should never be presented more than once

MONASH UNIVERSITY

Table captions

- Should appear above the table
- Should contain enough information so that the audience can easily determine
  - what information the table contains (including species names, number of replicates and measurement units, etc)
  - how the information was collected/compiled
  - and therefore (by examining the information in the table), what the information means (e.g. what the results are)

Example – Tabular summary

The peak milk-energy output recorded in this study for the koala is the lowest recorded for any mammal of equivalent mass, only 18%-28% of the predicted peak milk-energy output for eutherians (Table 4).

Table 4: Milk-energy output in nondairy species (in order of mass)

Species	Mass (kg)	Milk-Energy Output (Kj kg-0.75d-1)	Source
Marsupials:			
Ringtail possum	1	154	Monks 1990
Tammar wallaby	4.7	207	Cork and Dove 1989
Koala	6	99-122	This study
Eutherians:			
Black-tailed deer	49.8	494	Sadlier 1980
Sheep	52.6	586	Othtedal 1984
Red deer	85.3	320	Arman et al. 1974
Reindeer	107	332	McEwan and Whitehead 1971
Cattle	340	278	Yates et al. 1971
Horse	515	344	Oftedal et al. 1983

Example – Tabular summary

- Text in the results section reports a finding, and then invites the audience to a table for validation.
- When the audience is invited to examine the table it is also directed in what feature (milk-energy output of koala compared to eutherians) out of many possible features should be noted.
- This facilitates effective and efficient communication/conveying of results

Tabular formatting

The table below presents the results of testing for the effects of existing ascidians on settlement of marine invertebrates larvae. The table has only very basic formatting and does not facilitate quick or effective presentation of results.

Taxon	*P*	$\sqrt{MS_{Residual}}$	Power (ES=50%)
Serpulids	0.348	20.32	100
Spororbids	0.455	2.60	47
Elminius	0.531	24.89	71
Cryptosula	0.025	1.90	48
Scruparia	0.789	0.62	61
Tricellaria	0.017	4.72	98
Watersipora	0.525	3.45	94
Bugula neritina	0.118	10.36	69
Bugula stolonifera	0.042	18.60	100

Tabular formatting

The table below contains the same, yet more readable information as the previous slide.

Taxon	*P*	$\sqrt{MS_{Residual}}$	Power (ES=50%)
Serpulids	0.348	20.32	100
Spororbids	0.455	2.60	47
Elminius	0.531	24.89	71
Cryptosula	**0.025**	1.90	48
Scruparia	0.789	0.62	61
Tricellaria	**0.017**	4.72	98
Watersipora	0.525	3.45	94
Bugula neritina	0.118	10.36	69
Bugula stolonifera	**0.042**	18.60	100

- Bold font highlights significant results
- Spaces highlight the results for major taxonomic groupings

Presenting data - Figures

Graphical summaries

Graphs are used to summarise the following

- To illustrate trends in data (e.g. the means of categories or the polarity of relationships) to accompany statistical analyses
- Data exploration (e.g. boxplots, histograms)
- Analysis diagnostics (e.g. mean vs variance plot)

Note that usually graphs are **not** statistical hypothesis tests, and are therefore not adequate proof of a significant comparison or relationship.

Figure (graph) captions

- Should appear below the figure
- Should contain enough information so that the audience can easily determine
  - what information the figure contains (including species names, number of replicates and measurement units, etc)
  - how the information was collected/compliled
  - and therefore (by examining the information in the figure), what the information means (e.g what the results are)

Example – Graphical summary

The egg capsules of *Lepsiella vinosa* occurring in the mussel zone were found to contain significantly more eggs per capsule than those in the littorinid zone (t = -5.44, df = 77, P < 0.001,see Fig. 3).

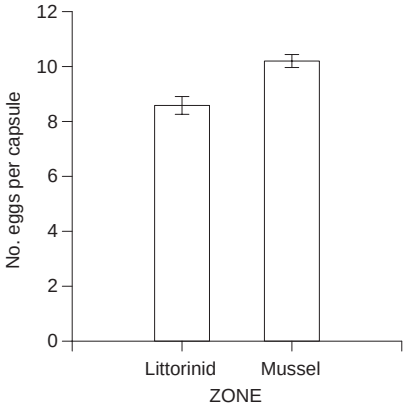


Figure 3. Mean (± SE) number of eggs per capsule of the predatory gastropod *Lepsiella vinosa* from two zones (Littorinid n=37, Mussel n=42) on a Victorian rocky intertidal shore.

Example – Graphical summary

- Text in the results section reports the finding, validates it statistically and then invites the audience to view a graph that illustrates the trend (mussel greater than littorinid).
- When the audience is invited to examine the graph it is also directed in what feature (greater number of eggs per capsule in mussel than littorinid zone) out of many possible features should be noted.
- This facilitates effective and efficient communication/conveying of results

Example – Graphical summary

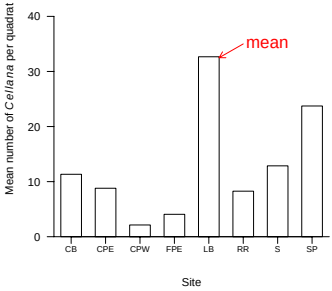
- Note that the graph is not a statistical validation of the results
- The analysis summary does not indicate the trend (which group is greater), hence the need for the graph.
- The graph helps the audience visualize the results

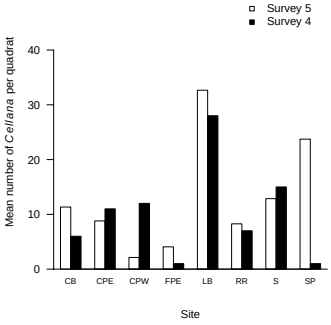
Graphical summaries

- When graphical summaries are used to illustrate trends in data, there are no assumptions imposed on data for graphical summaries
- Therefore, usually raw, untransformed data is graphed, even when analyses require transformations
- Raw data in graphical displays are easier to visualize and interpret than transformed data.

Bar graph

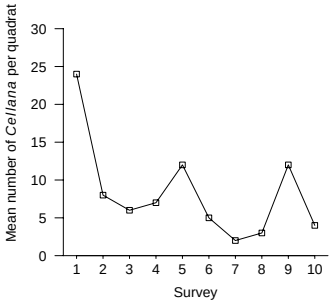
- Top of bar represents either a single observation or a group mean
- X-axis is categorical
- Different bars for different groups

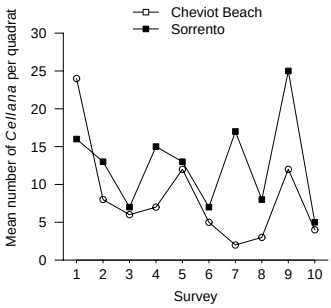




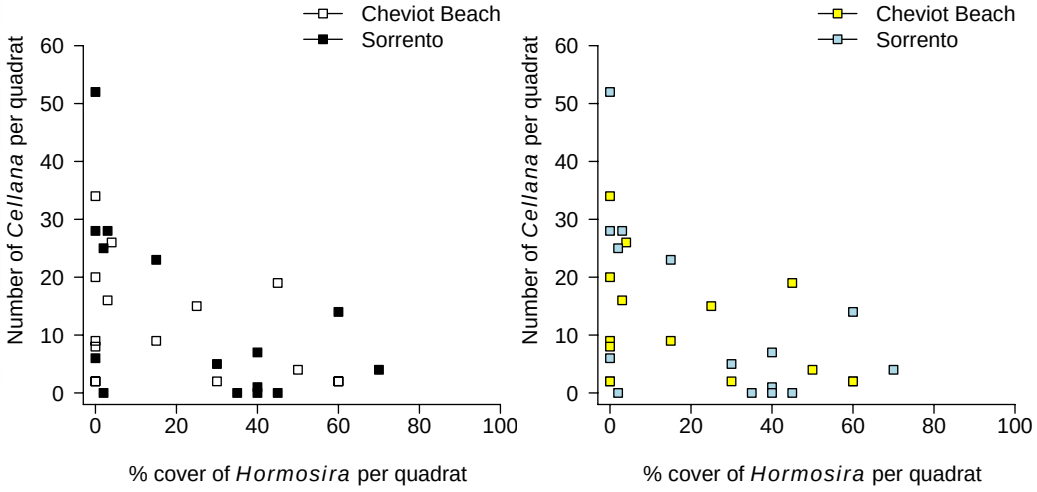
Category or line plot

- Each symbol represents either a single observation or group mean
- X-axis is categorical and yet quantitative
- Plotting symbols often joined by lines



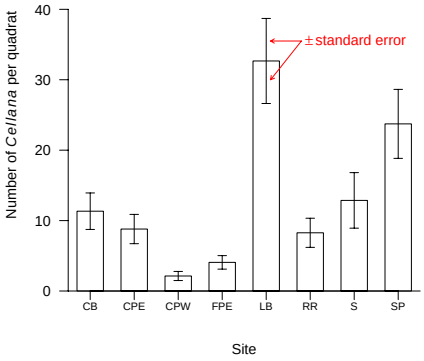


Multiple groups on scatterplots

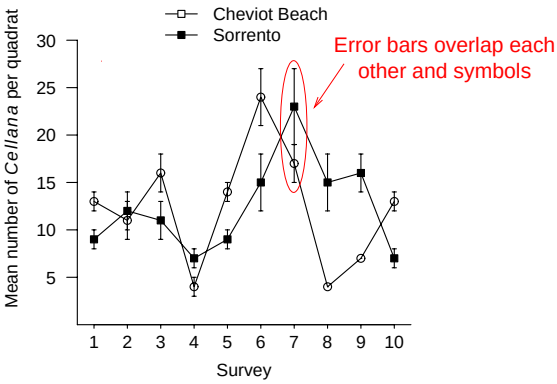


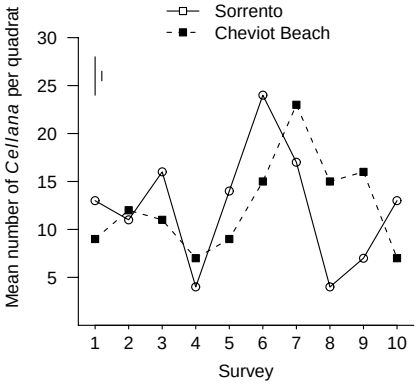
Depicting precision (error) or variability

- When plotting sample means, each group mean has an associated error bar
- Error bars can be:
  - standard deviation – indicating sample variability
  - standard error or confidence interval – indicating precision around mean
- Error bars are symmetrical and can be either one or two sided



Error bars on line plots



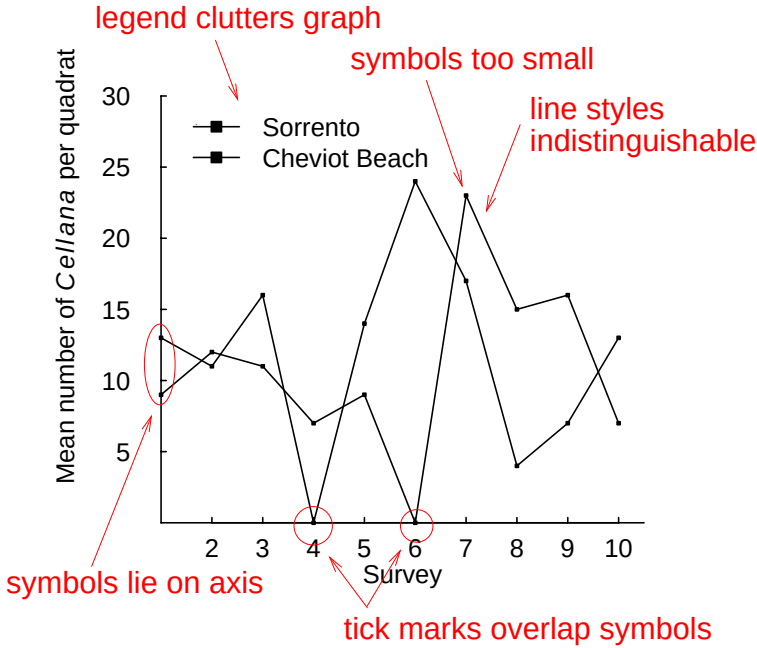


Rather than clutter the plot with error bars, indicate minimum and maximum error with two floating bars

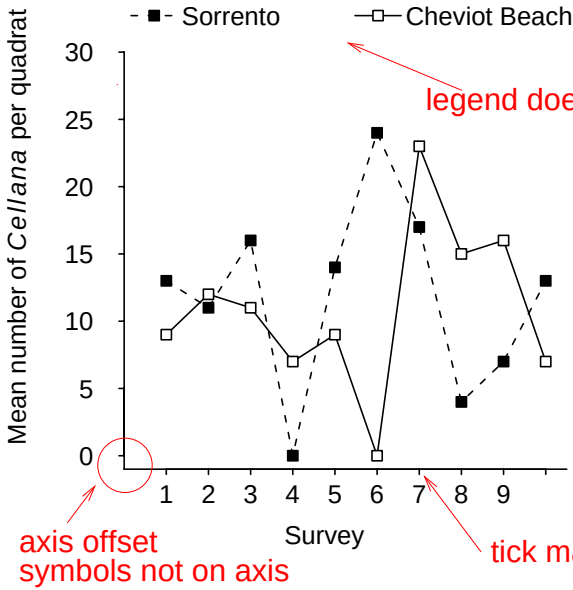
Recommendations – Graphs

- Emphasize data and nothing else!
  - avoid “chart-junk”
- Use visually prominent elements to show data
- Do not overlap graph elements
  - tick marks outside axis
  - use error bars carefully
- Avoid 3-D graphs for 2-D data

Line plots – bad points



Line plots – good points



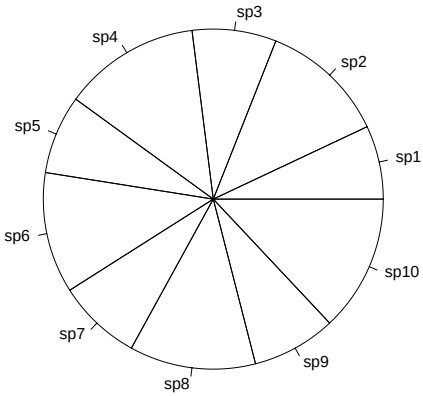
legend doesn't clutter graph

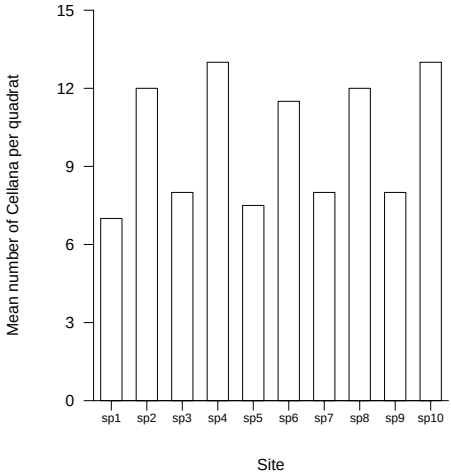
clearly visible and distinguishable symbols and lines

axis offset
symbols not on axis

tick marks outside

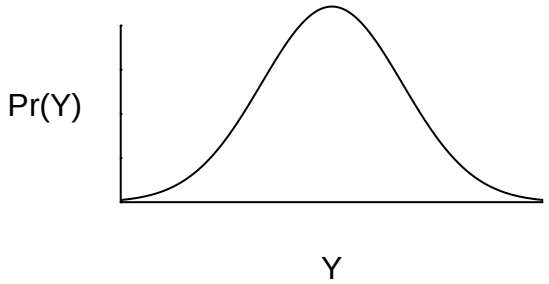
Pie charts





Normality

Normality refers to the state of a variable that is normally distributed. The normal (or Gaussian) distribution is a symmetrical probability distribution with a characteristic bell-shape. Statistical procedures that use sample means to characterize populations, assume that the observations that make up the sample (and thus the population) are normally distributed. Likewise, measures of the spread of data (often based on deviation from the center –mean) assume equal spread either side of the mean.



Click anywhere to close