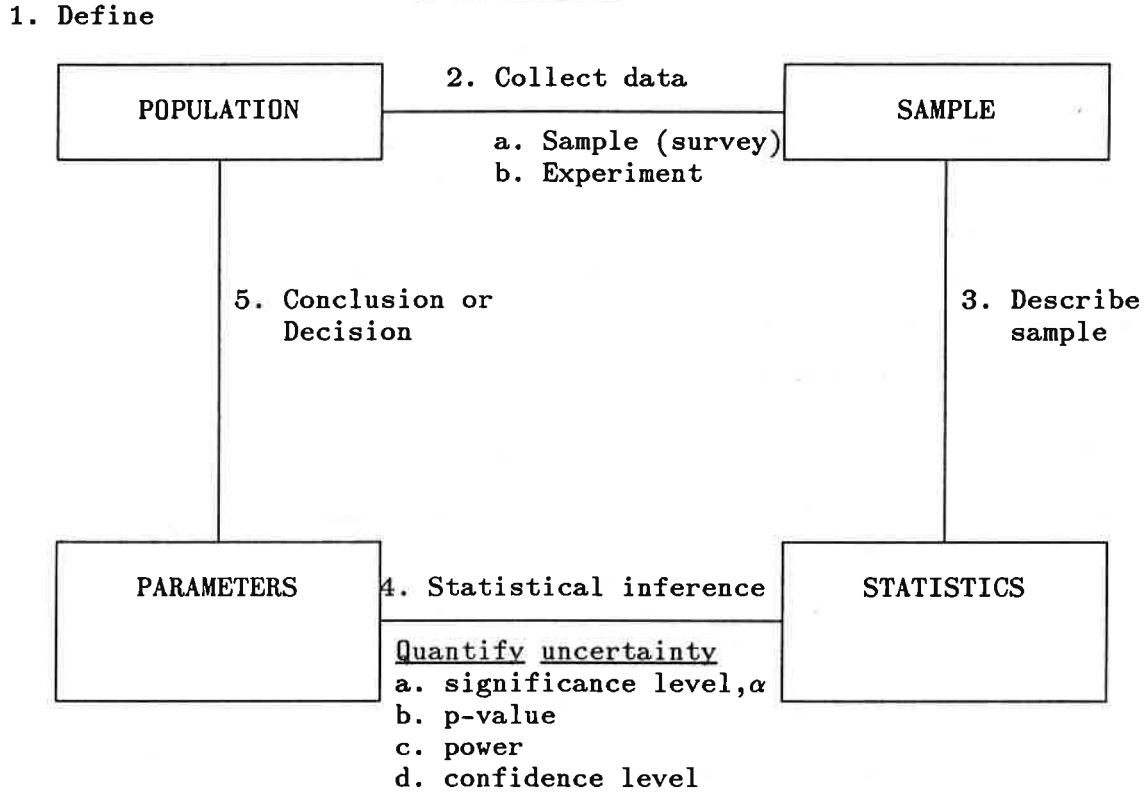
Describing and Reporting Statistics in Research Publications

Steve Stehman
September 27, 1999

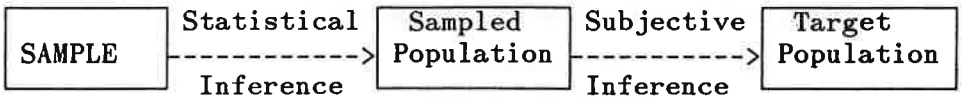
- To decide what should be reported in a research publication, it helps to return to the basics: WHAT IS STATISTICS USED FOR?
- Statistics focuses on collecting data, summarizing data, and drawing conclusions from data.
- STATISTICAL INFERENCE is the process of generalizing from a sample to a population.

FRAMEWORK OF STATISTICS



Definitions

1) *population* - the entire collection of individuals or measurements about which information is desired
2) *parameter* - numerical description of a population
3) *sample* - a subset of the population that has been selected for study
4) *statistic* - numerical description of a sample
5) *statistical inference* - methods or techniques for generalizing from a sample to the population from which the sample was selected



6) *target population* - the collection or set of measurements about which we want do draw conclusions

7) *sampled population* - the collection of objects from which we have a random sample

Example: Target population: all forest land in NY state.
Problems: denied access for field sampling, or non-response from survey

Five features that need to be addressed in the publication:

I. What did you do? (data collection and analysis)

II. What population do the inferences apply to?

III. What are the results: description and inference (connect to research objectives)?

IV. What is the uncertainty of the inferences (variability, α-level or p-value, power, confidence level)?

V. Conclusions should follow from your results

I. What did you do?

A. Methods

a. define target population: what population do you want your conclusions to apply to?

b. experiments and observational studies

1. define experimental units and how those were selected
2. experiment design: identify arrangement of experimental units (e.g., blocks, split-plots) (must be clear if and where **replication** is present!)
3. treatment design: identify arrangement of treatments (or groups)
4. state whether **randomization** was incorporated, how and where?
   - random or purposeful selection of experimental units?
   - random assignment of treatments to experimental units? (causality vs association)

c. surveys (sampling) studies

1. describe materials (e.g., frame) from which sample was selected
2. describe sampling design ("random" and "representative" are vague)
3. document non-response, other missing data problems

d. describe the variables you measured or observed, protocols ("response design")

B. Statistical Analysis

a. describe how you analyzed the data

b. provide a citation to methods that may be unfamiliar to users (include a page or section number if citing a book!)

c. do not cite software manuals for statistical methods -- i.e., do not cite the SAS manual for a paired t-test

d. state software and version used for analyses so reader knows you used a reputable package, or they know the problems the package has. SAS, SPSS, MINITAB, S-PLUS, SYSTAT, STATISTICA all ok, WalmartSTAT and EXCEL may have problems with non-standard analyses

II. What sampled population do the inferences apply to?

This should be covered by I.A. above. The description of how you selected the experimental units or how you selected your sample should clearly identify the sampled population. The reader may then judge the extent to which the sampled population differs from the target population.

III. What are the results?

A. Descriptive statistics: means, stnd. deviations or coefficients of variation, sample sizes, correlations

a. Why?

1. archiving information, documentation
2. information to assess practical importance of results
   - magnitude of treatment effects
   - differences in groups

   (Note: statistically significant differences may not be practically important, and "nearly" statistically significant differences may suggest important practical differences.)
3. record of variation, reliability of scales (useful to others for future planning)

b. Tables versus figures?

1. tables esthetically unappealing, but provide exact information
2. figures "look good", but often poorly done, waste a lot of ink and space, and convey potentially misleading information
3. figures should be used to show trends and relationships, not to document results
4. tables permit reader to conduct alternate analyses; order tables to show relationships

Table 8

***Mean Quality of Resolution* Scores Classified by Primary Reasons for Appealing**

	Quality of Resolution	SE				
Stop Implementation	2.24[a] (16)	.19				
Modify/Stop	2.56[a] (65)	.12				
Modify Implemenation	2.87[a] (51)	.16				
Neither	3.09[a] (10)	.40				
Outcome:						
	Effectiveness	**SE**	**Efficiency**	**SE**	**Equity**	**SE**
Stop Implementation	1.82[a] (18)	.27	3.12[ab] (17)	.25	1.74[a] (19)	.31
Modify/Stop	1.95[a] (77)	.17	2.88[b](66)	.17	1.82[a] (78)	.14
Modify Implementation	2.55[a] (65)	.19	3.60[a] (59)	.18	2.17[a] (62)	.19
Neither	2.95[a] (12)	.49	3.47[ab] (10)	.49	2.63[a] (11)	.50
Process:						
	Effectiveness	**SE**	**Efficiency**	**SE**	**Equity**	**SE**
Stop Implementation	2.05[b] (19)	.24	3.38[a] (18)	.34	1.75[a] (19)	.20
Modify/Stop	2.72[ab] (79)	.14	3.74[a] (74)	.13	2.08[a] (79)	.11
Modify Implementation	2.96[a] (65)	.17	3.38[a] (64)	.16	2.41[a] (63)	.16
Neither	3.18[ab] (11)	.47	2.95[a] (11)	.41	2.53[a] (12)	.44

Note: Means within a scale that do not share a superscript letter are significantly different @ .05 as determined by Tukey's pairwise comparison procedure.

** Number of valid cases per item*

B. Details of results for specific analyses (must go beyond just reporting very general results)

a. Analysis of Variance

1. table of means for each treatment (and variable)
2. analysis of variance table
   i. allows reader to reconstruct analysis
   ii. confirms what analysis and design were used
   iii. describes variability (Error mean squares) and provides information on power (df of error terms)

Example AVOVA table (for a split-plot, and two treatment factors, A and B)

Source	df	MS	p-value
Block	3	MSB/	
A	2	MSA	0.07
Error A	6	MSEA	
B	3	MSB	0.18
A*B	6	MSAB	0.44
Error A*B	27	MSEAB	
Total	47		

3. summarize contrasts of means, e.g. tests for main effects or interactions in a factorial (should address objectives: 1 df for treatment = 1 contrast = 1 research question)
4. multiple (pairwise) comparison procedures if no treatment structure (state which comparison procedure you used!)
5. state (briefly) how you evaluated assumptions and if this evaluation changed your analysis
6. minimize redundant information

b. Contingency tables

1. show the tables if you can (part of description, documentation)
2. examine and report standardized residuals to elaborate on significant findings

c. Regression

1. main features
   - graph (if possible)
   - estimated regression coefficients (prediction equation)
   - standard errors of estimated coefficients
   - measure of goodness of fit (R^2, lack of fit test, residual analysis, error mean square)
2. multiple regression
   - correlations among explanatory variables (so-called "independent" variables)
   - model development
     - model sequence if assessing importance of explanatory variables
     - type I, type II sums of squares, use of "extra sum of squares principle"
     - information measure
3. summary table

Source	Partial Reg. Coeff.	Std. Error	Type I SS	Type II SS
x_0 (int.)	4.50	2.02	1484	387
x_3	-0.20	0.04	287	245
x_1	1.05	0.21	125	97
x_2	0.77	0.23	80	22
x_4	0.03	0.01	5	5
Residual MS (32.5, with 41 df)				

d. Multivariate analyses

1. descriptive statistics
   - correlations
   - group means and variances (discriminant, cluster analysis)
   - variable means and variances (principal components, factor analysis)
2. figures such as canonical variable plots
3. information for interpreting canonical variables ("loadings")

Guide to Reporting Cluster Analysis Studies (pp. 80-81, Aldenderfer, M.S. and R.K. Blashfield, 1984, *Cluster Analysis*, Sage University)

1. Describe clustering method used.
2. State similarity measure (or statistical criterion if an iterative method is used).
3. State computer program used.
4. Describe procedure used to determine the number of clusters.
5. Provide adequate evidence of the validity of the cluster analysis solution - evidence should show that their results are independent of the clustering method.

IV. What is the uncertainty of the inferences (variability, α-level or p-value, power, confidence level)? If part III is done correctly, the reader has the necessary information to determine the uncertainty of the inferences. If sample sizes, degrees of freedom for different error terms, measures of variation, and α-levels, p-values, and confidence levels are provided, a reader should have a reasonably good idea of the power to detect differences. Formal power analyses may be difficult except for relatively simple experiments and analyses. Power analyses for simple experiments are probably unnecessary. An incorrect power analysis of a complex experiment is probably worse than no power analysis.

V. Conclusions and Decisions Should follow directly from results (research objectives→experiment or survey→analysis and results→conclusions and decisions). Statistics should contribute to conclusions and decisions, but it should not be the sole source of information on which conclusions and decisions are based.

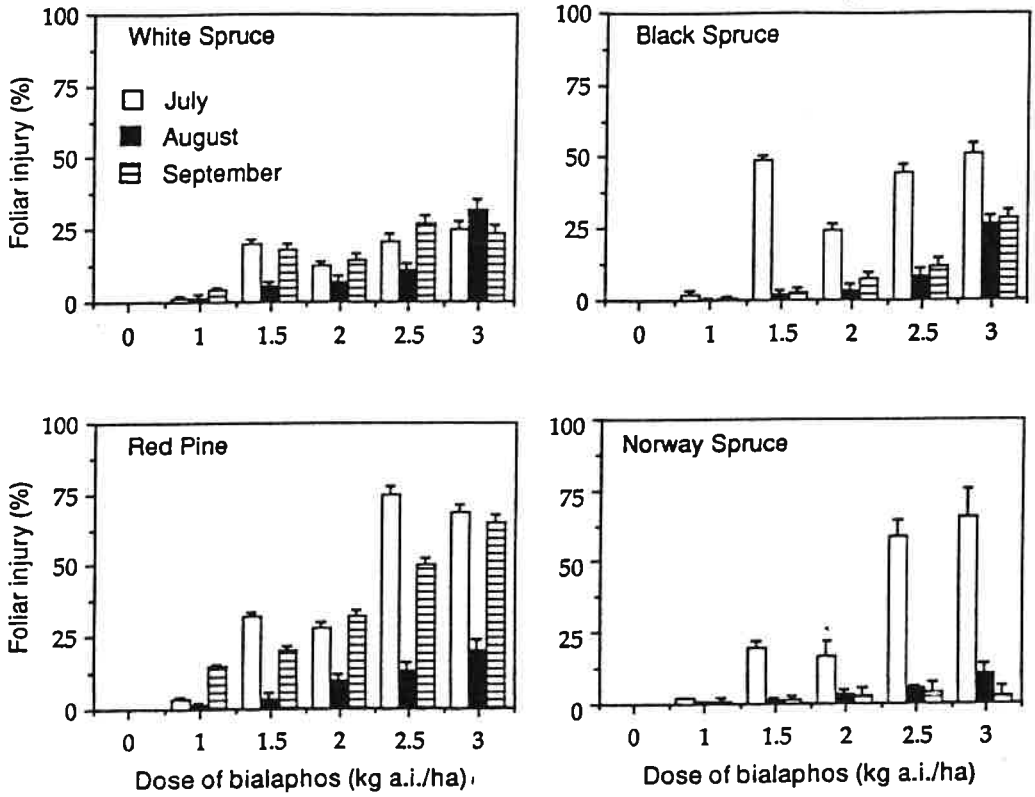


FIG. 2. Foliar injury as observed 2 months after bialaphos application as a function of application date, species, an ANOVA showed a significant interaction of date, species, and dose (MS = 3.48, F = 2.80, P = 0.0001). SEM bars are sho

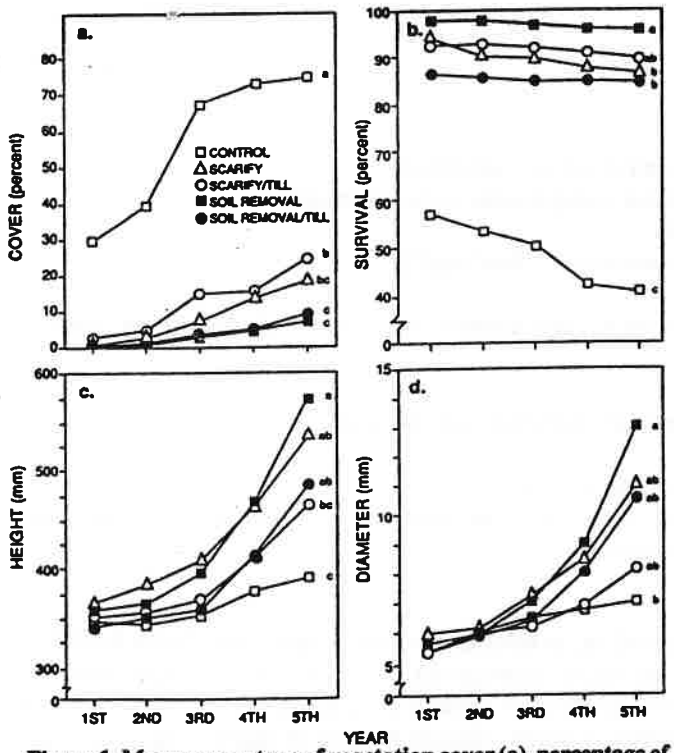


Figure 1. Mean percentage of vegetation cover (a), percentage of seedling survival (b), seedling heights (c), and seedling diameters (d), by year and by site preparation treatment. Fifth-year means with different letters are significantly different at the 0.05 level.

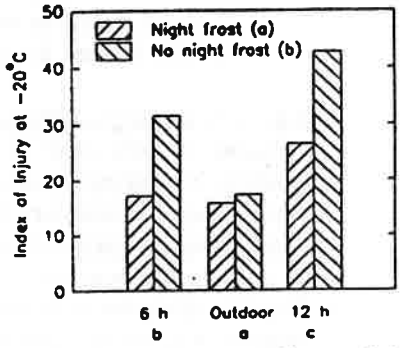


FIG. 4. Index of injury in yellow cypress seedlings exposed to three environments and two night-frost treatments on January 30. Index of injury at −20°C is interpolated from three freezing temperatures and averaged for three provenances. Note that environment or night-frost treatments with the same letter are not significantly different ($P < 0.05$, Scheffé's test).