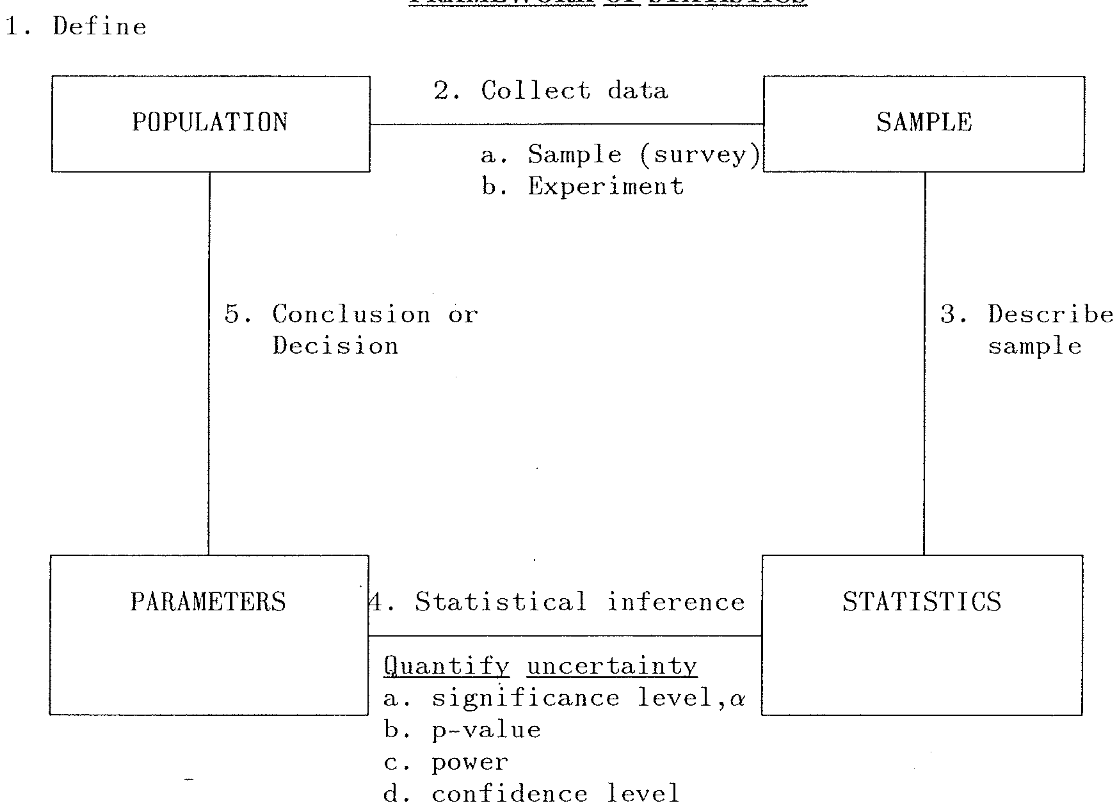
Describing and Reporting Statistics in Research Publications

Steve Stehman

- 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

```
            Statistical    Sampled     Subjective     Target
SAMPLE  ------------->  Population  ------------>  Population
             Inference                  Inference
```

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 <u>target</u> 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 <u>sampled</u> <u>population</u> 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	MSBl	
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")

<u>Guide to Reporting Cluster Analysis Studies</u> (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.

2194 CAN. J. FOR. RES. VOL. 24, 1994

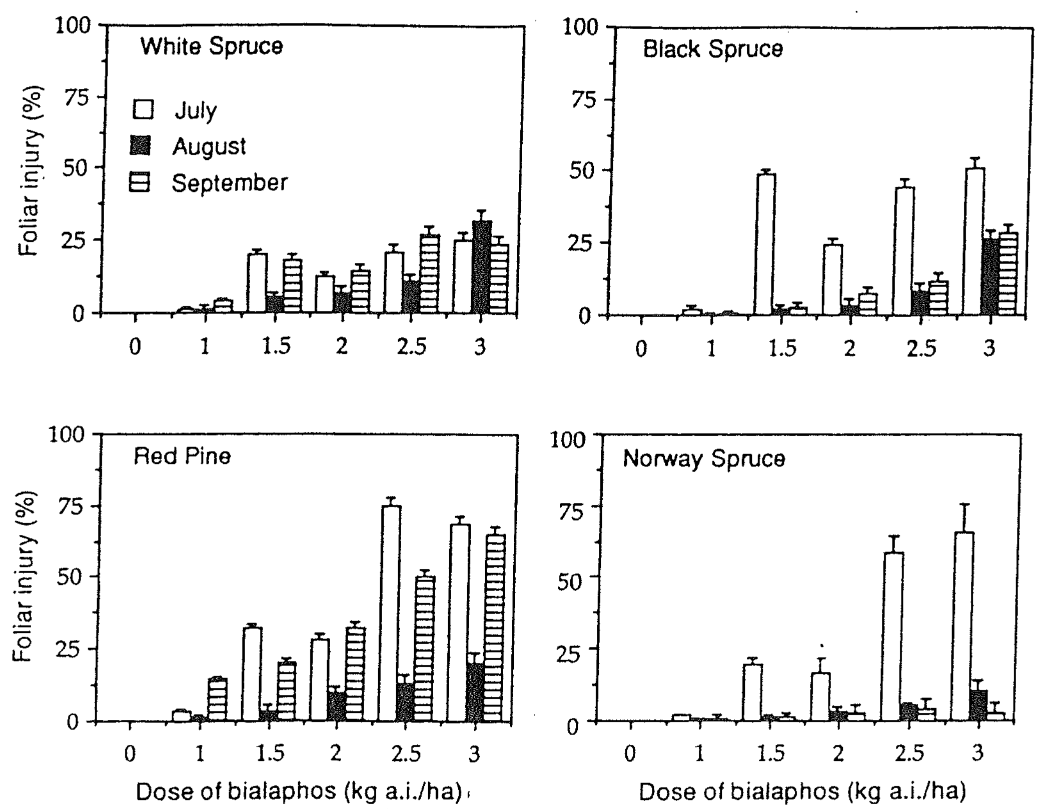


FIG. 2. Foliar injury as observed 2 months after bialaphos application as a function of application date, species, an
ANOVA showed a significant interaction of date, species, and dose (MS = 3.48, F = 2.80, P = 0.0001). SEM bars are show

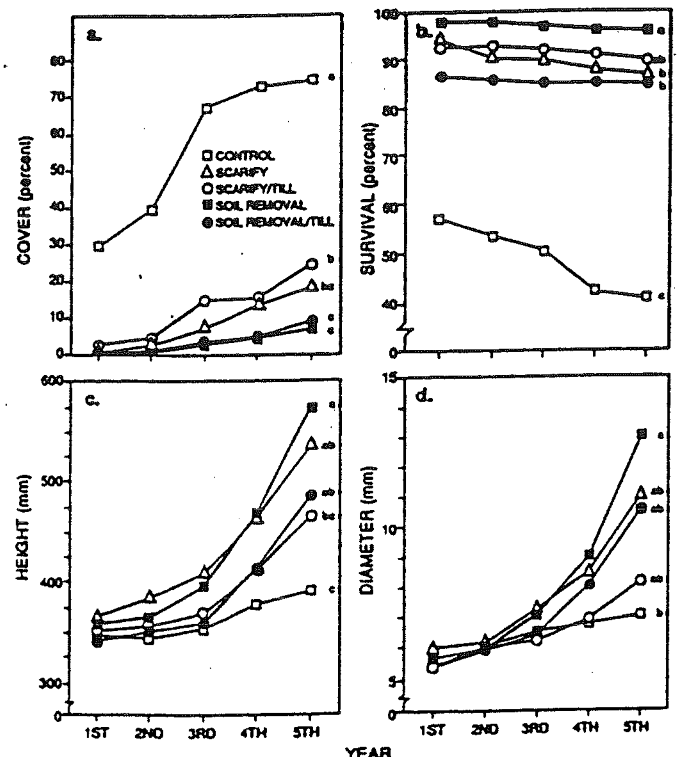


Figure 1. Mean percentage of vegetation cover (a), percentage of seedling survival (b), seedling heights (c), and seedling diameters (d), by year and by site preparation treatment. Fifth-year means with different letters are significantly different at the 0.05 level.

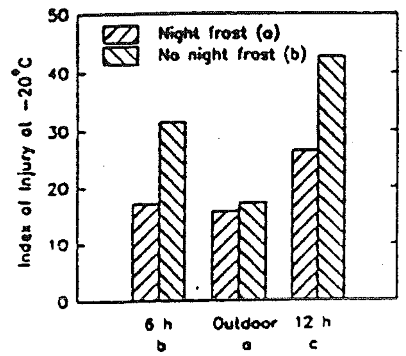


FIG. 4. Index of injury in yellow cypress seedlings exposed to three environments and two night-frost treatments on January 30. Index of injury at −20°C is interpolated from three freezing temperatures and averaged for three provenances. Note that environments or night-frost treatments with the same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

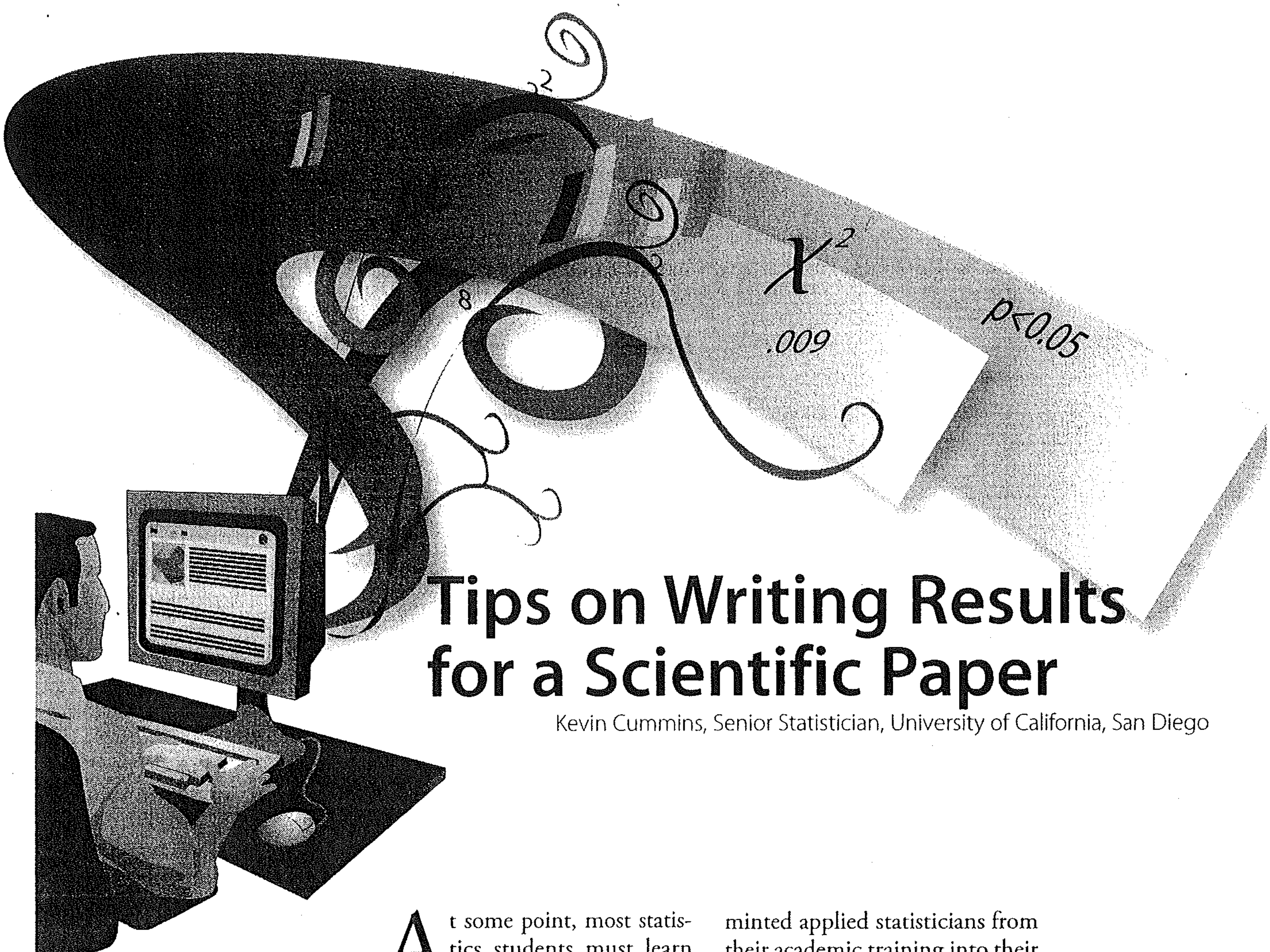


Illustration by Bret Alexander

Tips on Writing Results for a Scientific Paper

Kevin Cummins, Senior Statistician, University of California, San Diego

At some point, most statistics students must learn how to communicate scientific findings. It is hoped that they are already comfortable with the formulaic structure of a scientific article: introduction, methods, results, and conclusion. The scientific community dictates this gross structure because it allows readers to immediately begin understanding the content, without first having to understand the paper's organization. Little flexibility is allowed in the organization; however, there are fewer conventions below the structural outline, and students should recognize they have some latitude and adopt a style that improves the quality of their manuscripts.

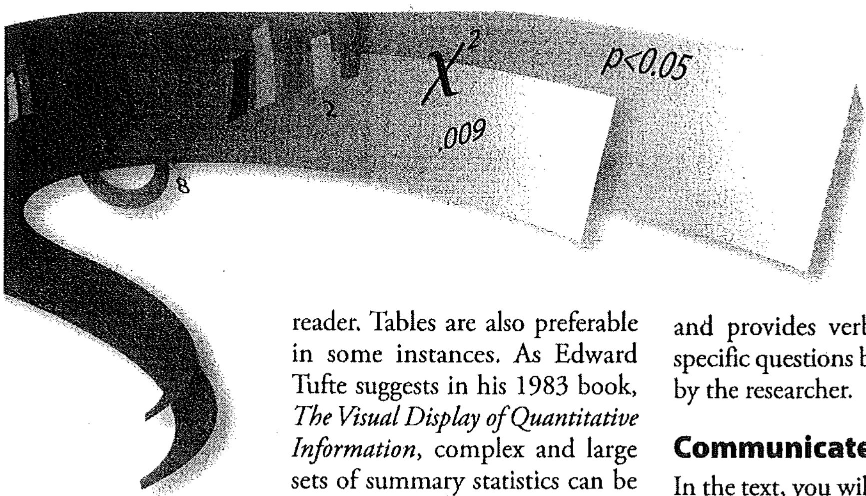
When collaborating on a manuscript (in an applied field), statisticians' contributions are most direct in the methods and results sections, and they are most likely to draft the results. I hope the following comments will ease the transition of newly minted applied statisticians from their academic training into their professional activities.

The results section is the core of a scientific paper. It should present new data that will be used to render a judgment about the issues raised in the introduction. The textual component of the results summarizes the general patterns and describes important statistical details, telling the reader what was found in the particular study. Implications of the findings are not presented in the results—those are saved for the conclusion.

Lean Heavily on Statistical Graphics

The use of statistical graphics often provides the strongest means for conveying the patterns in data. You should not be reluctant to lean heavily on statistical graphics. Indeed, this may be the only way to present more than a few parameter estimates without bogging down the

Kevin Cummins is a senior statistician at the Department of Psychiatry, School of Medicine, University of California, San Diego. Part of his job is to advise undergraduate and graduate students on how to analyze and report their findings. Over the years, he has put together a long list of helpful hints that he references when reviewing students' drafts.



reader. Tables are also preferable in some instances. As Edward Tufte suggests in his 1983 book, *The Visual Display of Quantitative Information*, complex and large sets of summary statistics can be recorded in well-organized tables. Tables can allow the reader to explore the data when there are many relevant combinations of comparisons among variables.

Integrate Nontextual Results and the Text

You should be able to read the text in the results section and know the answer to each question raised in the introduction without looking at tables or figures. This recommendation suggests you narrate the story in the graphs by parenthetically referencing your graphs when describing patterns in the data graphics. Consider the following examples:

Example 1

Watermelon production was 23% [95% CI: 20–26%] higher in plots inoculated with mycorrhizal fungi as compared to controls (Figure 2).

Example 1 is much more informative than writing, "The results of the mycorrhizal inoculation are presented in Figure 2."

Example 2

Subjects with low self-efficacy relapsed sooner than those with high self-efficacy (Figure 3). Relapse risk was 2.1 [95% CI: 1.8–2.3] times higher for those with low self-efficacy.

Example 2 is a better use of text than, "The relationship between self-efficacy and relapse is presented in Figure 3."

The examples demonstrate how the text highlights patterns that can be seen in the graphics and provides verbal answers to specific questions being addressed by the researcher.

Communicate It Once

In the text, you will tell the reader what can be found in the figures and tables. Don't list the same statistics in the text as are already plotted in a graph. Do describe the general patterns and details not directly plotted in the figure. Consider the following examples:

Example 3

Figure A illustrates that the mean TPS score was 21 (SE = 3.4) for new recruits and 45 (SE=5.5) for veteran employees.

If Figure A is a plot of the means, the text is redundant. If the text is sufficient to recreate the graphs, dump the graphs or use the text to present another perspective on the data or focus on the take-home message.

Example 4

TPS scores for veterans were 114% higher than for new recruits (Figure A).

This gets the reader thinking about relative differences and absolute scores, which are presented in an accompanying figure.

Respect Your Reader

Keep the inferential statistics in the background. Some authors use a repetitive pattern in which the first sentence declares an effect significant and the second sentence describes the direction of the effect. This couplet is repeated for each effect. Here is an example:

Example 5

The logistic model predicting relapse was significant (LR $\chi^2_{(3)} = 21$, $p<0.05$). Level of social supports was found to be significantly related to substance use relapse ($p<0.05$). Higher levels of abstaining social supports decreased the rate of relapse. Joint diagnosis with internalizing and externalizing psychiatric disorders was significantly related to

(See Example 4)

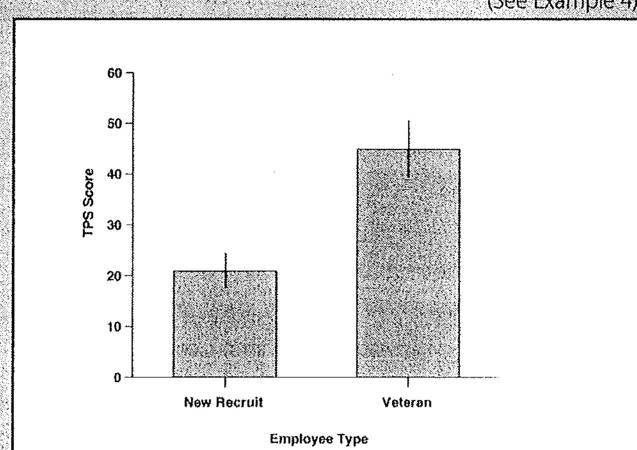


Figure A. Mean TPS scores by employee type. Error bars are 95% confidence intervals.

relapse ($p<0.05$). Subjects with both disorders displayed an increased rate of relapse. Age was also significantly related to relapse ($p<0.05$). Older subjects were more likely to relapse.

The style used in Example 5 is both a reflection and a perpetrator of the misuses of inferential statistics. The dominance of the *p*-value is at the root of concerns among analysts such as J. Scott Armstrong in the *International Journal of Forecasting* article "Significance Tests Harm Progress in Forecasting." He argues that the practice of conducting significance tests should be abandoned. In extreme cases, the *p*-values' dominance brings authors to forego describing the effect they are studying, because they emphasize the statistical significance to the exclusion of other statistics.

There is a style representing good statistical principles that is conciliatory between extreme views regarding significance testing and provides the most concise message without loss of information. This style, shown in Example 6, uses the text to describe the general patterns in the data and parenthetically provides either the *p*-value or a reference to a figure that contains the *p*-values.

Example 6

The hypothesized factors were found to jointly predict substance use relapse (LR x2(3) = 21, $p<.001$). Having a greater proportion of abstaining social supports increased the odds of being in the abstaining category three times (OR = 4.14, CI: 1.43–11.98, Wald = 2.62, p = .009). Conversely, the odds of remaining an abstainer were reduced by two-thirds if they had internalizing and externalizing psychiatric disorders (OR = 0.32, CI: 0.14– 0.74, Wald = −2.65, p = .008). Older subjects are less likely to relapse, such that the odds of relapse decrease by 25% for each one-year increase in age (OR = 0.75, CI: 0.58–0.99, Wald = −2.04, p = .042).

Use the Past Tense

The past tense is appropriate because the results section is supposed to be a description of what was observed in the particular study. The data are historical when published. Use of the present tense connotes an extrapolation beyond the data whereby the author is generalizing. Generalizations and the present tense are best left in the conclusions section.

Use CONSORT's Outline

CONSORT is a consortium organized for the purpose of standardizing research reports for randomized controlled trials. The use of the CONSORT guidelines ensures important information is not omitted from the research report. An example of one of the standards is the requirement for standardized patient flow charts, which describes what happens to the subject pool between randomization and final follow-up. Another important CONSORT recommendation is to report confidence intervals for all parameter estimates. Many of the guidelines can be applied or adapted to a variety of study designs. More information about CONSORT's recommendation can be found at *www.consort-statement.org*.

Sequence Consistently

Consistently order the topics presented in each section of the paper, as consistency makes reading easier. Consistency of order should be carried out whenever there is repetition. For instance, in a series of tables and figures that contain the same independent variables, all the tables and figures should have the variables presented in the same order. An important exception is when ordering is used to communicate different results.

Works on the Principles of Good Graphical Presentation

Written by Edward Tufte

The Visual Display of Quantitative Information. Graphics Press, 1983

Visual Explanations. Graphics Press, 1997

Beautiful Evidence. Graphics Press, 2006

Written by William Cleveland

The Elements of Graphing Data. Wadsworth Advanced Books, 1985

Eliminate Methods

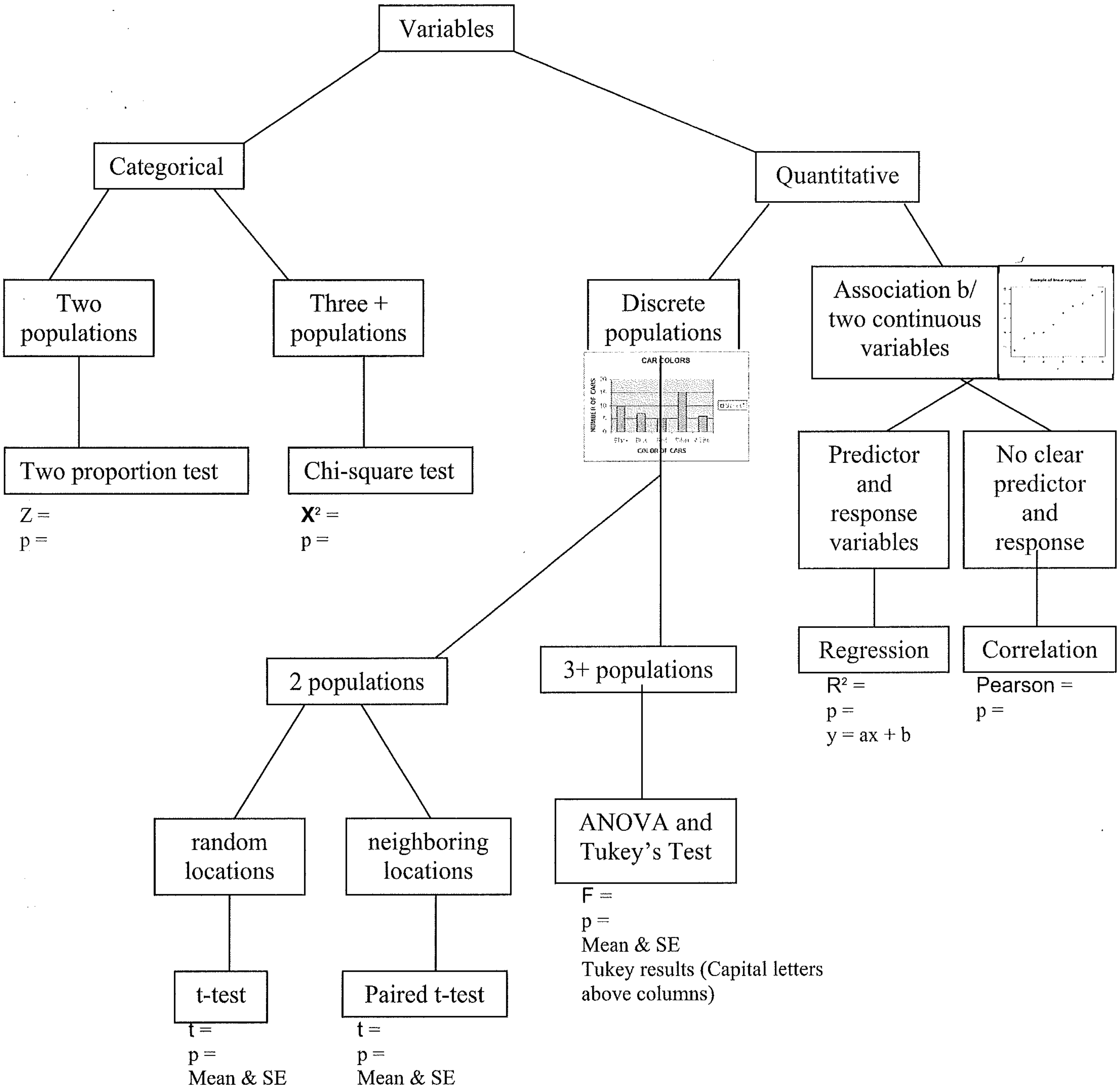
What you did to get the results belongs in the methods section. The details of the statistical modeling are considered methods in most disciplines. Avoid reiterating the methods in the results section. Reference to the methods as an aid to orient the reader can be acceptable when the study is complex. Example 7 is adapted from *The Scientist's Handbook for Writing Papers and Dissertations.*

Example 7

Salivary gland squashes probed with affinity-purified anit-75 antibodies showed the puffed regions of chromosomes had 52% higher fluorescence intensity (Figure 2).

Statisticians often know the data better than anyone else on the research team. The value of their contribution can be increased through a thoughtfully crafted communication style. ■

IV. Statistical Test Flow Chart



Statistical values you should provide for each test are indicated.

Compiled by S. Stehman with modifications by J. Tessier, G. McGee, J. Brunner, M. Fierke.