

## Reporting Statistical Results in Your Paper

### Overview

The results of your statistical analyses help you to understand the outcome of your study, e.g., whether or not some variable has an effect, whether variables are related, whether differences among groups of observations are the same or different, etc. Statistics are *tools* of science, *not* an end unto themselves. Statistics should be used to substantiate your findings and help you to say objectively when you have significant results. Therefore, when reporting the statistical outcomes relevant to your study, subordinate them to the actual *biological results*.

### Reporting Descriptive (Summary) Statistics

**Means:** Always report the **mean** (average value) along with a measure of **variability** (*standard deviation(s)* or *standard error of the mean*). Two common ways to express the mean and variability are shown below:

"Total length of brown trout (n=128) averaged 34.4 cm (s = 12.4 cm) in May, 1994, samples from Sebago Lake."

s = standard deviation (this format is preferred by [Huth and others \(1994\)](#))

"Total length of brown trout (n=128) averaged 34.4 ±12.4 cm in May, 1994, samples from Sebago Lake."

This style necessitates specifically saying in the Methods what measure of variability is reported with the mean.

If the summary statistics are presented in graphical form (a Figure), you can simply report the result in the text without verbalizing the summary values:

"Mean total length of brown trout in Sebago Lake increased by 3.8 cm between May and September, 1994 (Fig. 5)."

**Frequencies:** Frequency data should be summarized in the text with appropriate measures such as percents, proportions, or ratios.

"During the fall turnover period, an estimated 47% of brown trout and 24% of brook trout were concentrated in the deepest parts of the lake (Table 3)."

### Reporting Results of Inferential Tests

In this example, the **key result** is shown in blue and the **statistical result**, which *substantiates* the finding, is in red.

"Mean total length of brown trout in Sebago Lake increased significantly (3.8 cm) between May (34.4 cm ±12.4 cm, n=128) and September (38.2 cm ±11.7 cm, n=114) 1994 (twosample t-test, p < 0.001)."

**NOTE:** AVOID writing whole sentences which simply say what test you used to analyze a result followed by another giving the result. This wastes precious words (*economy*!!) and unnecessarily increases your paper's length.

## Summarizing Statistical Test Outcomes in Figures

If the results shown in a figure have been tested with an inferential test, it is appropriate to summarize the outcome of the test in the graph so that your reader can quickly grasp the significance of the findings. It is imperative that you include information in your Materials and Methods, or in the figure legend, to explain how to interpret whatever system of coding you use.

Several common methods for summarizing statistical outcomes are shown below.

Examples: Comparing groups (t-tests, ANOVA, etc)

Comparison of the means of 2 or more groups is usually depicted in a bar graph of the means and associated error bars.

For two groups, the larger mean may have 1-4 asterisks centered over the error bar to indicate the relative level of the p-value. In general, "\*" means  $p < 0.05$ , "\*\*\*" means  $p < 0.01$ , "\*\*\*\*" means  $p < 0.001$ , and "\*\*\*\*\*" means  $p < 0.0001$ . In all cases, the p-value should be reported as well in the figure legend.

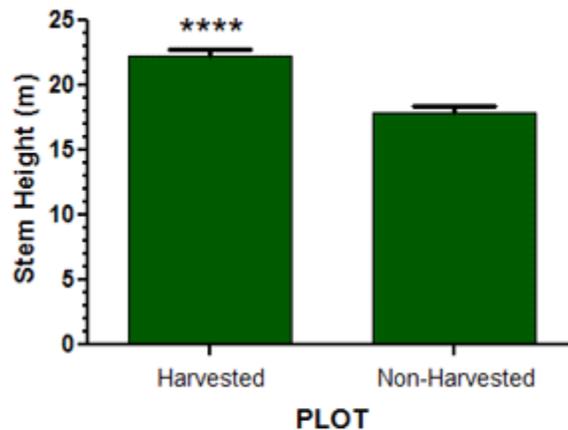


Figure 4. Stem height (mean  $\pm$  SEM) of mature eastern white pines in harvested and non-harvested plots five years after a one crown spacing, selective harvest in Range Pond State Park. Harvested plot  $n=20$  trees; Nonharvested plot  $n=22$  trees. Unpaired t-test, 40 df,  $p < 0.0001$ .

The asterisk may also be used with tabular results as shown below. Note how the author has used a footnote to define the p-values that correspond to the number of asterisks. (Courtesy of Shelley Ball)

Table 2. Log-likelihood tests of deviation from 1:1 sex ratios for nymphs collected from each population in 1997 and 1998. Values are ratios of female:male; sample sizes are in parentheses. Bonferroni corrected probabilities are shown with an asterisks.

Population	Year	
	1997	1998
Beaver Creek <sup>T</sup>	9.00:1(20)***	2.67:1 (22)*
Honey Creek <sup>T</sup>	9.00:1(56)***	2.27:1 (98)***
Rock Bridge <sup>T</sup>	3.33:1(26)**	2.09:1 (68)**
Cedar Creek <sup>P</sup>	2.05:1(119)***	1.87:1 (198)***
Grindstone Creek <sup>P</sup>	-	2.26:1 (140)***
Jacks Fork River <sup>P</sup>	2.89:1(35)**	5.17:1 (37)***
Meramec River <sup>P</sup>	2.80:1(38)**	2.41:1 (58)**
Little Dixie Lake <sup>L</sup>	2.45:1(494)***	2.46:1 (384)***
Little Prairie Lake <sup>L</sup>	2.38:1 (71)***	2.08:1 (157)***
Rocky Forks Lake <sup>L</sup>	2.55:1 (213)***	2.93:1 (299)***
Winegar Lake <sup>L</sup>	3.41:1 (207)***	2.34:1 (204)***
Whetstone Lake <sup>L</sup>	2.69:1 (381)***	2.01:1 (268)***

\* significant at  $p < 0.05$ ; \*\* significant at  $p < 0.005$ ; \*\*\* significant at  $p < 0.001$ .

<sup>T</sup> = temporary stream, <sup>P</sup> = permanent streams, <sup>L</sup> = lakes.

For three or more groups there are two systems typically used: lines or letters. The system you use depends on how complicated it is to summarize the result. The first example below shows a comparison of three means. The line spanning two adjacent bars indicates that they are not significantly different (based on a multiple comparisons test), and because the line does not include the pH 2 mean, it indicates that the pH 2 mean is significantly different from both the pH 5.3 (control) and the pH 3.5 group means. Note that information about how to interpret the coding system (line or letters) is included in the figure legend.

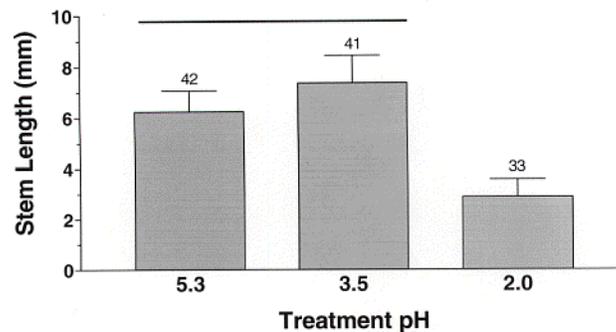


Figure 1. Mean stem length ( $\pm$  1 SD) of seedling clover watered to soil saturation daily for 2.5 weeks with simulated acid rain of varying pH. The control (pH 5.3) was normal city tapwater. The pH 3.5 and 2.0 water was acidified with 2 M sulfuric/ 1 M nitric acid solution. Line over bars indicates groups which were not significantly different (Kruskal-Wallis Test and Dunn's Multiple Comparison's Tests). Number over bar indicates sample size.

When lines cannot easily be drawn to summarize the result, the most common alternative is to use capital letters placed over the error bars. Letters shared in common between or among the groups would indicate no significant difference.

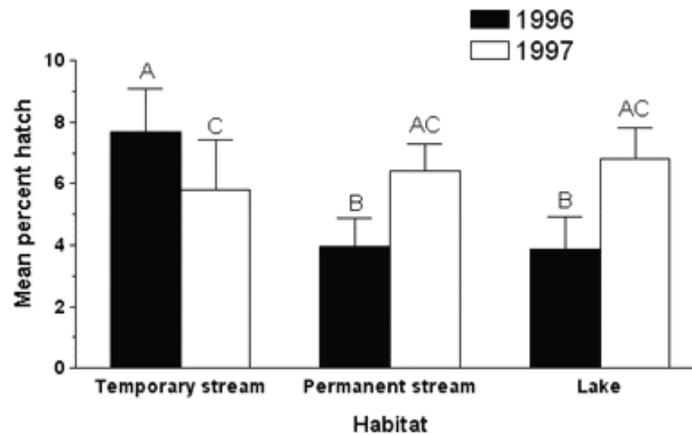


Figure 3. Effects of habitat and year on tychoparthenogenetic capacity (mean % hatching success  $\pm$  1 SD of unfertilized eggs) in mayflies. Means with different letters are significantly different (Tukey's HSD,  $p < 0.05$ ).

#### Example: Summarizing Correlation and Regression Analyses

For relationship data (X,Y plots) on which a correlation or regression analysis has been performed, it is customary to report the salient test statistics (e.g.,  $r$ ,  $r$ -square) and a  $p$ -value in the body of the graph in relatively small font so as to be unobtrusive. If a regression is done, the best-fit line should be plotted and the equation of the line also provided in the body of the graph.

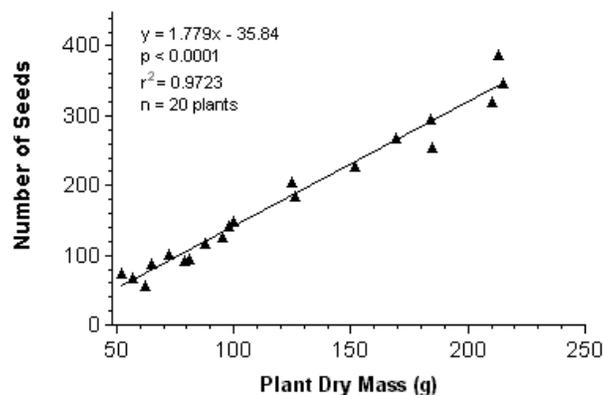


Figure 3. Seed production as a function of plant biomass in waterlilies (*Nuphar luteum*) harvested from Great Works Pond in Northern Maine in August, 2001.