Reading and writing about statistical analyses

Why statistical analyses?

Found throughout the scientific literature
Allows one to draw conclusions from laboratory or field studies involving numerical data

Variability in data

Where does it come from?

Normal fact of natural world
- Bacterial activity, diversity, growth rates
- Geologic formations
- Seasonal variation in populations (nematodes, turtles, fish)

Imprecision of measurements

Number, size, mass, concentration.

Variability commonly referred to as **"error"**

Example: Abundance of fossil "X" in two Toledo limestone sites.

Site1: 72, 70, 68, 70 **Site 2**: 36, 125, 33, 86

Average is the same for each site: **70 fossils**

But, Site 1 counts are **much less variable** than Site 2 counts.

Therefore, you will often see some aspect of the variability in a collection of data reported, such as:

standard deviation,
standard error, or
confidence interval

Reading about statistics

Knowledge of how the statistical analyses were performed is not necessary to understand what the results mean.

$P < 0.001$

The difference indicated by the data would be expected by chance less than one time in 1000 repetitions of the experiment.

$P = 0.28$

The difference indicated by the data would be expected by chance as many as 28 times in 100 repetitions of the experiment.

Writing about statistics

Focus on the science, not the statistics

For the caterpillars reared on the mustard-flavored diet and subsequently given a choice of foods, the caterpillars showed a statistically significant preference for the mustard diet ($X^2 = 17.3$; d.f. = 1; $P < 0.05$). For 30 caterpillars reared on the quinine-flavored diet, however, there was no significant influence of previous experience on the choice of food ($X^2 = 0.12$; d.f. = 1; $P > 0.10$).

What does this mean?

The hermit crabs in our sample (N = 12) showed a significant relationship between their weights and the size of the shells they occupied in the field (r^2 = 0.477; test for zero slope: F = 9.104; d.f = 1.12; P = 0.013).

" r^2 " = The variation in hermit crab weight accounted for only 47.7% of the variation in the sizes of the shells occupied.

Focus should be placed on the **science** and not the statistics.

Statistics should only be used to support claims you wish to make about your results.

Resist the temptation to ramble on about how the statistics were calculated.

The chi-square value of 6.25, with a *P* of 0.0124 revealed that the physical condition of the shell had a significant influence on shell choice by the 15 hermit crabs used in our study.

What's wrong with this?

Corrected:

The physical condition of the shell had a significant influence on shell choice by the hermit crabs (X^2 = 6.25; d.f. = 1; N = 15; *P* = 0.0124).

A Student's *t*-test was used to determine the significance of the difference in mean interaction times. The data were not significant (i.e., we found no significant results in our experiment), and there was no difference in the mean contact time between hermit crabs in the presence or absence of predators.

What's wrong with this?

1. Eliminate the first sentence.
2. Statistical significance has been confused with scientific value of the data/study ("the data were not significant…).

The presence of a predator did not significantly affect the amount of time that hermit crabs spent interacting with each other (t = 1.012; d.f. = 12; P = 0.332).

The word "**significant**" is used carefully