245W: Reading and writing about statistics[i]

Statistics are an essential component of scientific research, and in turn, scientific articles. In order to evaluate whether a research study has found sufficient support for its hypotheses, you as a reader will need to be able to understand basic statistical terminology.

Most statistical analyses are tests of specific hypotheses. These statistical hypotheses comprise null vs. alternative hypotheses. In a nutshell, the null hypothesis (H_0) expects no difference or no relationship between your variables of interest, whereas alternative hypotheses (H_A or H_1) expect differences or relationships. The H_A is usually a subset or prediction of the overall experimental hypothesis. The statistical hypotheses are usually not explicit; most articles assume the reader will know what the hypotheses were based on the types of statistical tests used. Thankfully, authors are typically required to briefly describe the statistical tests used as part of their Methods section. Therefore, it is possible to figure out the individual hypotheses from the authors' Methods and Results. For instance, let's take a look at the following excerpts from a recent article in **Evolution** discussing life history evolution in *Lestes* damselflies (De Block et al. 2008):

> (from Methods)
> Individual growth rates for each period were calculated as $[\ln(mass_{end}) - \ln(mass_{ini})]/(t_{end} - t_{ini})$ with $mass_{end}$ and $mass_{ini}$ being subsequent masses of an individual at the end (t_{end}) and start (t_{ini}) of the period. [...] Because we had repeated measurements on the same larva, repeated-measures analyses of variance (RM-ANOVAs) were conducted for development time and mass, using a compound symmetry covariance structure. In these analyses, starvation treatment, photoperiod, species, and sex were included as between-subject variables and period as within-subject variable. The uncorrected *P*-values did not differ from the Greenhouse and Geisser corrected *P*-values.
>
> (from Results)
> During the prestarvation period (i.e., the entire larval period excluding the final instar), *L. dryas* had considerably higher growth rates than *L. forcipatus*, which in turn had slightly higher growth rates than *L. congener* (species, $F_{2,1035} = 2449.2$, $P < 0.0001$; Fig. 2A, B). Females of all species had higher growth rates than males ($F_{1,1035} = 106.6$, $P < 0.0001$; all other $P > 0.13$).

From the first excerpt, we know that the authors used repeated-measures ANOVAs to look at the differences in two variables: development time and mass. (ANOVAs (for Analysis of Variance) are statistical tests that allow you to investigate the effects of multiple factors (e.g. starvation treatment, photoperiod, etc.) on a variable (e.g. development time, mass, etc.) you are interested in.) The growth rates they are talking about in the second excerpt were calculated using a function incorporating both development time and mass. Thus, it is safe to assume they used RM-ANOVAs to compare growth rates as well. Let's say that, based on the rest of the article and the first excerpt, you already know that the authors were interested in the affects of "starvation treatment, photoperiod, species, and sex" on individual growth rates. In the second excerpt, the authors are first comparing growth rates among individuals of the three different species they were looking at, then between females vs. males. The respective H_0s and corresponding H_As would have been something like this:

> $H_0(1)$: Growth rates of *Lestes dryas* = *L. forcipatus* = *L. congener*
> $H_A(1)$: Growth rates of *Lestes dryas* ≠ *L. forcipatus* ≠ *L. congener*
> $H_0(2)$: Growth rates of males = females
> $H_A(2)$: Growth rates of males ≠ females

Much like how citations are used to support statements in the Introduction and Discussion, in the Results section, statistics are used to support experimental hypotheses. The authors' statistical analyses refute the H_0s that species identity and sex do not affect growth rates, and we know this because of the *P*-values they report in parentheses. If the *P*-value of the statistical test is under a predetermined target probability (usually 0.05), the *P*-value is considered "significant", and H_0 is refuted, with H_A is supported instead. The *P*-value is the probability that you are observing the phenomenon of H_A when H_0 is true. That is, if the *P*-value is 0.05, 5% of the time you might observe a difference in growth rate despite there being no actual difference. What this also means is that, 95% of the time, the observed difference in growth rate is a reality. A "significant" *P*-value (i.e. less than 0.05) is considered reasonable support for H_A, although it does not disprove H_0.

Now, if you wanted to report these results in your critique or term paper, it is standard practice to gloss over the details of which statistical tests were used and what p-values were reported. Instead, you would emphasize what "significant" vs. "non-significant" effects (if any) the authors found. For instance, I might summarize the findings of De Block et al. (2008) as follows:

> De Block et al. (2008) found that individual growth rates of *L dryas* nymphs were higher compared to the other two species tested. Female nymphs grew more quickly than males in all three species.

Finally, while most experimental papers will have explicit statistical *P*-values for your consideration, others may not, depending on the type of analyses performed. For instance, phylogenetic analyses have different kinds of tests, hypotheses, and support values. With careful reading, however, you should be able to figure out, for example, which relationships were better supported compared to others.

References

De Block, M., M. A. McPeek, and R. Stoks. 2008. Life-history evolution when *Lestes* damselflies invaded vernal ponds. Evolution 62(2):485–493.
