8. ANSWERS TO PROBLEMS

1) Erythrocyte pH regulation in toads

a) Ho: $\mu_d = 0$
Ha: $\mu_d \neq 0$
$t=-1.9950$, $v=36$, $P=0.0537$ Do not reject Ho. In general, the paired t-test is the most powerful method available if its assumptions are met, and is therefore preferred if these assumptions are met.

b) The paired t-test assumes that the distribution of *differences* between paired measurements, d, follows a normal distribution. A visual inspection suggests that this assumption is not met. The sample distribution of d's looks reasonable except there are two outliers – one on either extreme – and kurtosis is high (2.9734). Use the hand tool to change the histogram interval sizes.

c) Ho: Differences between methods are normally distributed
Ha: Differences between methods are not normally distributed
The Shapiro-Wilk test did not reject Ho ($W=0.9456$, $n=37$, $P=0.0939$), therefore our assumption of normality *may* be justified. A sample size of 37, however, may not be sufficient to invoke the Central Limit Theorem in this case. In order to exercise caution, a non-parametric test (Wilcoxon signed-rank test) would be appropriate. A sign test (i.e., a binomial test) may be *too* cautious!

d) 95% CI: $-0.0605 \leq \mu \leq 0.00049$. The confidence interval may not be entirely accurate, since the data appear somewhat non-normal. However, we were unable to reject normality with the data, so perhaps we needn't be so cautious.

2) Effects of slash burning on spider diversity

a) A two-tailed test is most appropriate because we haven't got a compelling reason to predict that diversity will increase (or decrease) four years after burning.

b) First examine the data in a histogram, normal quantile plot, and outlier box plot to evaluate the assumption that the change in spider diversity has a normal distribution, which is an assumption of the paired t-test. Visually, the data do not appear to follow a normal distribution. This impression is upheld by a Shapiro-Wilk goodness of fit test:
Ho: Difference in diversity is normally distributed
Ha: Difference in diversity is not normally distributed

We reject Ho ($W=0.8983$, $n=27$, $P=0.0123$). Change in spider diversity is not normally distributed, therefore the paired t-test is inappropriate and we must turn to an alternative. The Wilcoxon signed-rank test gives the following results:
Ho: There is no difference in number of spider species before and after burning.
Ha: There is a difference in number of spider species before and after burning.
Wilcoxon signed rank statistic: 42.500, $P=0.199$. Do not reject Ho.

c) The problem with the design of the present study is that there are no controls. Perhaps spider diversity may change in time for reasons having nothing to do with burning per se. A better design would would be to split each plot in two and burn only one of the halves. The unburned halves would then serve as controls for the study. This design assumes that spider diversity in

the unburned half would not affect diversity in the burned half (e.g., it assumes that spiders don't move between half-plots). If spider movement really is a possibility, then the best approach is to use a two-sample design instead (burn some whole plots and not others).

d) In general, a paired design is an excellent approach because it minimizes the impacts of site-to-site variability in environmental conditions when testing for an influence of treatment. This is because environmental conditions are likely to be less variable between sides of the same plot than between plots. Variables such as soil type, moisture, aspect, elevation, may all confound results from plots in different sites.

e) 95% CI: $-2.2429 \leq \mu_d \leq 0.8355$. This is likely to be inaccurate because the data (change in spider diversity) are not normally distributed. The sample size is too small (27) to confidently invoke the Central Limit Theorem.

3) Species diversity of polyandrous vs. monandrous insect species clades

a) Visually, the data appear to strongly violate the assumption of normality, especially because of one large outlier. A Shapiro-Wilks goodness of fit test of the assumption of normality yielded the following:
Ho: Differences of species numbers are normally distributed
Ha: Differences of species numbers are not normally distributed
rejected Ho ($W=0.9592$, $n=25$, $P<0.0001$). Differences are not normally distributed, so we cannot use the paired t-test.

A signed-rank test gave the following results:
Ho: Number of species in polyandrous clades is not different from the number in monandrous clades.
Ha: Number of species in polyandrous clades is different from the number in monandrous clades
Wilcoxon signed rank statistic: 67.500, $P<0.068$. Do not reject Ho.

Note: I used a 2-tailed test here, as did the authors of the paper. It may be possible to justify a one-tailed test given that the theory predicted a greater number of species in the polyandrous clade.

b) Visually, the differences based on log(species number) conform to the normal distribution much better than did differences based on the un-logged data. A Shapiro-Wilks test of the assumption of normality for the logged data yielded the following:
Ho: The difference between the logs of species numbers in the two clades are normally distributed.
Ha: The difference between the logs of species numbers in the two clades are not normally distributed.
Do not reject Ho. ($W=0.9592$, $n=27$, $P=0.4120$)

The data do not allow us to reject our assumption of normality (although there may be one or two outliers, they are not too bad. Proceed cautiously with the paired t-test:
Ho: $\mu_d = 0$
Ha: $\mu_d \neq 0$
$t=2.4462$, $v=26$, $P=0.0221$. Reject Ho. There is a difference in species number between clades. The conclusion is the opposite of the non-logged data. This exercise suggests that it may be

rewarding to transform the data to more closely meet the assumptions of the paired t-test, which is more powerful than the Wilcoxon signed-rank test.

c) Examining the signs of the differences shows that 7 of the differences are negative and 18 are positive (you may get 7 positive differences and 18 negative ones if you calculated the differences by subtracting species numbers in the polyandrous clade from that in the monandrous clade). Under the null hypothesis, however, we expect equal numbers of positive and negative differences. A binomial test yielded the following:
Ho: $p=0.5$
Ha: $p \neq 0.5$
$P = 2 \times \text{Prob}(X \leq 7) = 0.0433$, Reject Ho.
(To calculate this P-value, use the JMP IN calculator function

Binomial Distribution(0.5, 25, 7)

and multiply the quantity by 2 to yield the 2-tailed probability.)

An alternative test is the χ^2 goodness of fit test, comparing the observed frequencies 7 and 18 with the expected frequencies 12.5 and 12.5, yielding:
Pearson's chi-squared: $\chi^2 = 4.8400$, df=1, $P=0.0278$
Likelihood: $G = 5.0097$, df=1, $P=0.0252$
(if you use the correction for continuity, as required when df=1, then you would get:
$\chi^2 = 4.00$, df=1, $P=0.0455$, which is closer to the exact binomial P-value).
Reject Ho, conclude that polyandrous clades exceed monandrous clades in species number.

d) Yes, I think so. Both the paired t-test and the sign test rejected Ho.

4) Mechanical efficiency of muscle between ACE genotypes

a) Ho: $\mu_{II} - \mu_{DD} = 0$
Ha: $\mu_{II} - \mu_{DD} \neq 0$
This experiment to compare change in muscle mechanical efficiency of ACE genotypes does not have a paired design. Rather, the design involves comparing means of two separate groups of individuals, "II" individuals and "DD" individuals. The test of different means therefore requires a 2-sample t-test rather than a paired t-test. You will need to reconfigure the data file somewhat to allow JMP IN to carry out the test. The results: $t=3.150$, $v=58$, $P=0.0026$. Reject Ho.

b) The method assumes that change in muscle mechanical efficiency has a normal distribution in both II and DD populations with equal variance.
Histograms revealed reasonable fits to the normal distribution, and Shapiro-Wilks goodness of fit tests do not the reject the null hypothesis of a normal distribution in either case. The sample variances of the two groups are similar, and a two-sample F-test (or Levene's test) does not reject the null hypothesis of equal variances.