Solution to Exercise 2.3 (Version 1, 16/09/14)

from Statistical Methods in Biology: Design & Analysis of Experiments and Regression (2014) S.J. Welham, S.A. Gezan, S.J. Clark & A. Mead. Chapman & Hall/CRC Press, Boca Raton, Florida. ISBN: 978-1-4398-0878-8

Exercise 2.3*

The one-sample t-test is rarely used in analysis of experimental data, except in the context of regression, but it can be useful for analysis of paired samples from a set of subjects. In this scenario, the two sample t-test is not valid because two samples from a single subject are not independent. However, if we analyse the differences between the samples from each subject, we can use a one-sample t-test to test the null hypothesis of no difference between samples.

An experiment made measurements of Rubisco protein (on a relative scale) in 12 grass plants before and after a drought stress period of five days. File PROTEIN.DAT contains the unit number (*DPlant*) and Rubisco measurements (variates *Before* and *After*) for each plant. Calculate the change in amount of Rubisco protein in each plant and analyse this change using a two-sided one-sample t-test. Write down the null and alternative hypotheses for this test and interpret them in the context of this experiment. Is there any evidence that the amount of Rubisco has changed after five days of drought stress?

Data 2.3 (PROTEIN.DAT)

Measurements of Rubsico protein (relative scale) in 12 grass plants before and after a drought stress was applied for five days:

Plant	1	2	3	4	5	6	7	8	9	10	11	12
Before	2.82	1.90	1.92	3.69	3.54	3.65	1.86	1.93	3.38	2.09	2.12	3.18
After	2.98	2.45	2.73	3.66	3.41	3.49	1.93	2.16	3.22	2.19	2.64	3.33

Solution 2.3

This data set consists of N = 12 paired observations. We start by calculating the change in pairs of protein measurements as *Change* = *After* – *Before* (see Table S2.3.1), to obtain the values to use in our one-sample t-test. Eight plants show increased levels of protein, whilst for four plants the level has decreased. If we denote the mean of our *Change* variate as μ then the null hypothesis of no change (in mean Rubisco protein levels) is H_0: $\mu = 0$, to be tested against the two-sided alternative hypothesis that there is a change, i.e. H_1: $\mu \neq 0$.

We will denote the i^{th} change as y_i, for $i = 1\ldots12$. The sample mean change, $\bar{y}$, is

$$\bar{y} \;=\; \frac{1}{N}\sum_{i=1}^{N} y_i \;=\; \frac{1}{12}\sum_{i=1}^{12} y_i \;=\; \frac{2.11}{12} \;=\; 0.176,$$

indicating that protein levels have increased on average.

Table S2.3.1 Calculation of change (after-before), deviation from the mean change (Deviation) and squared deviations.

Plant	Before	After	Change	Deviation	Deviation^2
1	2.82	2.98	0.16	-0.016	0.0003
2	1.9	2.45	0.55	0.374	0.1400
3	1.92	2.73	0.81	0.634	0.4022
4	3.69	3.66	−0.03	-0.206	0.0424
5	3.54	3.41	−0.13	-0.306	0.0935
6	3.65	3.49	−0.16	-0.336	0.1128
7	1.86	1.93	0.07	-0.106	0.0112
8	1.93	2.16	0.23	0.054	0.0029
9	3.38	3.22	−0.16	-0.336	0.1128
10	2.09	2.19	0.10	-0.076	0.0058
11	2.12	2.64	0.52	0.344	0.1185
12	3.18	3.33	0.15	-0.026	0.0007
Sum	-	-	2.11	0.000	1.0429

To calculate the variance, we take the deviations of the change values from the sample mean (see Table S2.3.1). We then take the sum of the squares of these values (1.0429, Table S2.3.1) and divide by $N-1 = 11$ to get the sample variance of the changes, s^2:

$$s^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - \bar{y})^2 = \frac{1.0429}{11} = 0.0948.$$

The t-statistic is then calculated as

$$t = \frac{\bar{y} - \mu}{\sqrt{s^2 / N}} = \frac{0.176 - 0}{\sqrt{0.0948 / 12}} = 1.978.$$

Note that the value of μ is that under the null hypothesis, so $\mu=0$ here. This statistic has $N - 1 = 12 - 1 = 11$ df. The 97.5th percentile value for the t-distribution with 11 df is $t_{11}^{[0.025]} = 2.201$ (see Table B.2). Because the observed test statistic is smaller than the critical value, we cannot reject the null hypothesis at the 5% ($\alpha = 0.05$) significance level, and conclude that there is not enough evidence to indicate that Rubisco protein levels have changed between the two samples. The observed significance level of this test is $P = 0.073$, which leads us to the same conclusion.