

## Problem set 1 solutions

### A: t-test practice

1.

(a)  $H_0: \mu_x \geq 30$

$H_A: \mu_x < 30$

(b) 1-tailed 1-sample t-test

(c)  $t = -0.4657$ ,  $df = 12$ ,  $P = 0.3249$

(d) At all levels of alpha,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that this species does not maintain a coelomic calcium concentration less than that of their environment.

2.

(a)  $H_0: \mu_x = 24.3$

$H_A: \mu_x \neq 24.3$

(b) 2-tailed, 1-sample t-test

(c)  $t = 2.7128$ ,  $df = 24$ ,  $P = 0.01215$

(d) For  $\alpha = 0.1$  and  $0.05$ ,  $p < \alpha$  so we reject the null hypothesis and conclude that the crabs maintain a body temperature different than air. For  $\alpha = 0.01$  and  $0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that the crabs maintain a body temperature the same as the air.

3.

(a)  $H_0: \mu_x = 29.5$

$H_A: \mu_x \neq 29.5$

(b) 2-tailed, 1-sample t-test

(c)  $t = -2.2859$ ,  $df = 13$ ,  $P = 0.03968$

(d) For  $\alpha = 0.1$  and  $0.05$ ,  $p < \alpha$  so we reject the null hypothesis and conclude that average human menstrual cycle does not equal a lunar month. For  $\alpha = 0.01$  and  $0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that the average human menstrual cycle equals a lunar month.

4.

(a)  $H_0: \mu_x \leq 45$

$H_A: \mu_x > 45$

(b) 1-tailed, 1-sample t-test

(c)  $t = 0.3665$ ,  $df = 7$ ,  $P = 0.3624$

(d) For  $\alpha = 0.1$ ,  $0.05$ ,  $0.01$ , and  $0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that the mean dissolving time is less than or equal to 45 seconds.

5.

(a)  $H_0: \mu_x \geq 45$

$H_A: \mu_x < 45$

(b) 1-tailed, 1-sample t-test

(c)  $t = 0.3665$ ,  $df = 7$ ,  $P = 0.6376$

(d) For  $\alpha = 0.1, 0.05, 0.01$ , and  $0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that the mean dissolving time greater than or equal to 45 seconds.

[Combining questions 4 and 5, if we use a 2-tailed, 1-sample t-test with  $H_0: \mu_x = 45$  and  $H_A: \mu_x \neq 45$ , ( $t = 0.3665$ ,  $df = 7$ ,  $P = 0.7248$ ).  $p > \alpha$  so we fail to reject the null hypothesis and conclude that the mean dissolving time is 45 seconds.]

6.

(a)  $H_0: \mu_B = \mu_G$

$H_A: \mu_B \neq \mu_G$

(b) 2-tailed, two-sample t-test

(c)  $t = -2.54$ ,  $df = 10.7$ ,  $P = 0.0277$

(d) For  $\alpha = 0.1$  and  $0.05$ ,  $p < \alpha$  so we reject the null hypothesis and conclude that the drugs differ in their effectiveness (mean blood-clotting times are not the same). For  $\alpha = 0.01$  and  $0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that there is no difference in the mean effectiveness of the drugs (the mean blood-clotting times are the same).

7.

(a)  $H_0: \mu_{\text{new}} \leq \mu_{\text{current}}$ , so  $\mu_{\text{new}} - \mu_{\text{current}} \leq 0$

$H_A: \mu_{\text{new}} > \mu_{\text{current}}$ , so  $\mu_{\text{new}} - \mu_{\text{current}} > 0$

(b) two sample t-test (one-tailed)

(c)  $t = -3.013$ ,  $df = 15.6$ ,  $P = 0.0042$

(d) For  $\alpha = 0.1, 0.05$  and  $0.01$ ,  $p < \alpha$  so we reject the null hypothesis and conclude that the mean height of plants grown with the new fertilizer is larger than the height of plants grown with the current fertilizer. For  $\alpha = 0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that the heights of the plants using the current fertilizer are at least as high or higher than using the new fertilizer (i.e. the new fertilizer does no better than the current fertilizer).

8.

(a)  $H_0: \mu_{\text{male}} = \mu_{\text{female}}$   
 $H_A: \mu_{\text{male}} \neq \mu_{\text{female}}$

(b) two-sample t-test (2-tailed)

(c)  $t = 0.632$ ,  $df = 10.9$ ,  $P = 0.5405$

(d) For  $\alpha = 0.1, 0.05, 0.01$ , and  $0.001$ ,  $p > \alpha$  so we fail to reject the null hypothesis and conclude that male and female turtles have the same mean serum cholesterol concentrations.

9.

(a)  $H_0: \mu_A = \mu_C$   
 $H_A: \mu_A \neq \mu_C$

(b) 2-tailed, paired t-test

(c)  $t = 5.3057$ ,  $df = 56$ ,  $P = 1.984e-06$

(d) For  $\alpha = 0.1, 0.05, 0.01$ , and  $0.001$ ,  $p < \alpha$  so we reject the null hypothesis and conclude that there is a difference in the mean peak streamflow of these two creeks.

Notice that if you had done a two-sample t-test then you would have reached a very different conclusion ( $P = 0.229$ ). This is because the test does not control for the large year-to-year fluctuations that are due to regional climate.

## B: Assessing normality

You can begin by glancing at the distributions (histograms) of each sample and you'll notice that most look roughly normal. The Shapiro-Wilk test provides a formal way to assess normality. For questions 1-5, you can run the test in JMP by going to Analyze -> Distribution -> Fit Distribution -> Normal -> Fitted Normal -> Goodness of Fit.

It turns out that the only strongly non-normal data are in problem 9, so you might want to jump to that one. The other problems came from a textbook; they are so strongly normal that I wonder if they came from a random number generator rather than real data!

1a) These data seem reasonably normal. Shapiro-Wilk test:  $P = 0.696$ ; fail to reject normality

2a) These data are clearly normal. S-W:  $P = 0.801$ ; fail to reject normality

3a) Again, normal as normal can be: S-W:  $P = 0.994$ ; fail to reject normality

4&5a) These are normal. S-W:  $P = 0.964$ .

6) The tests for normality should be evaluated for each group independently. You can run the test in JMP by going to Analyze -> Distribution but make sure to put the type of drug

in the “By” box. You will end up with distributions for Drug B and Drug G. Proceed as before for each distribution by clicking -> Fit Distribution -> Normal -> Fitted Normal -> Goodness of Fit.

We end up with the following S-W test statistics, so we fail to reject normality.

a)	Drug B	Drug G
S-W:	P = 0.997	P = 0.913

7a)	Current fertilizer	New fertilizer
S-W:	P = 0.683	P = 0.902

In both cases the data look reasonably normal, and we fail to reject normality.

8a)	Male	Female
S-W:	P = 0.674	P = 0.135

The S-W test supports the notion that the male distribution is normal. There is a hint of a non-normal pattern in the Female data, but with so few values it is hard to tell. This illustrates the challenge: if the population really were non-normal, then 6 data points probably aren't enough to satisfy the assumptions of the test; but with so few points it is hard to detect non-normality.

9) If we were doing a two-sample t-test (not the paired test) we would look at the individual creeks:

a)	Atascadero	Carpenteria
S-W:	P << 0.0001	P << 0.0001

In both cases the data are clearly non-normal, with strong positive skew. However, because n is ‘large’, our t-tests are likely valid due to the Central Limit Theorem. However, if n were smaller we would want to transform the data so that the resulting data were more normal looking in order to conduct a valid t-test. Alternatively, we could use a non-parametric test.

b) I first tried to take the square root of the creek flow but this does not transform the distributions to normal. I next tried logging the creekflow variable to see if that forces a normal-looking distribution.

To log a variable in JMP, go to the relevant datasheet (in this case the ‘creekflow’ sheet). Right click on the column headers to create a new column. Name the new column (here I choose “Inflow” to denote the natural log of the flow variable). Click Column Properties -> Formula -> Transcendental -> Log->flow. The transformed data now look reasonably normal (at least they look more normal).

a)	Atascadero	Carpenteria
S-W:	P = 0.378	P = 0.168

HOWEVER, since what we did was a paired t-test, the critical thing that we need to look at is the distribution of the *differences*. So create a new variable which is the difference between Atascadero and Carpenteria using the new column procedure as before. Go to the creekflow data sheet with paired data (rather than the stacked data). Click Column Properties -> Formula ->Atascadero - Carpenteria. This will generate a new variable that is the flow difference between the two creeks. Now go to Analyze and look at the distribution of the difference variable you just created.

a) The test results suggest the differences are not normally distributed, S-W: P = 0.0001.

The deviation is primarily through positive skew (the upward curve of the majority of the points) but the last few points on the left show evidence of kurtosis.

c) The problem with a log transformation here is that it won't work with negative values: the transformations are undefined. We could add a constant before we do the transformation, to make them all positive. However, let's instead perform a non-parametric test of the flow difference variable. Non-parametric tests are less sensitive to distributions. Click Analyze -> flowdiff->Y, Columns -> OK ->flowdiff-> test mean. Now put in '0' as the hypothesized mean and click the box to do a Wilcoxon Signed Rank test. The Signed-Rank test statistic is huge at 681.5 so we can reject the hypothesis that the mean and median of the differences are equal to zero. (If you are interested you can read more about the Wilcoxon Signed Rank test on your own).

## C: Integrative problems

Problem 1

- Determine the appropriate null and alternate hypotheses to make the decision. Explain why you chose these.**

$$H_0: \mu_x \geq 6 \text{ ppm}$$

$$H_A: \mu_x < 6 \text{ ppm}$$

The null hypothesis is that the mean TcCB concentration in the soil is greater than or equal to 6 ppm, and the alternative hypothesis is that the mean TcCB concentration in the soil is less than or equal to 6 ppm. This was selected because the null hypothesis, which generally represents the "no effect" scenario, suggests that the clean up has had no impact on soil concentrations. The alternative hypothesis suggests that an effect has been achieved as a result of the clean up. Notice that this setup puts the burden of proof, so to speak, on the company responsible for the cleanup.

**2. If you make a type I error, what will be the consequence? If you make a type II error, what will be the consequence?**

A Type I error occurs when the null hypothesis is rejected when it is actually true. In this case, we would conclude that the site is clean when in fact it is contaminated. Consequences of this error include threats to human and wildlife health and well being, as well as the continued degradation of soil and habitat.

A Type II occurs when the null hypothesis is false, though we fail to reject it. Here, we would conclude that the site is contaminated above EPA standards, when it is not. Consequences of this error include additional and unnecessary clean up costs.

**3. What statistical procedure will you use to test the null hypothesis? Justify your choice in words**

A one sample, one-sided t-test of the clean-up sites would be used in this situation. We are assuming normality of the non-transformed data, even though they are highly skewed. We select a t-test and not a z-test because we do not know the standard deviation of the population. We are using a one-sided t-test because we are interested in (depending on how your null and alternative hypotheses were set up) whether the data exceed the EPA standard in a particular direction. If we only wanted to know if the data are different from the EPA standard, we would select a two sided t-test.

**4. Run the test, report the results, and state your conclusions.**

With an alpha of 0.05, we fail to reject the null hypothesis ( $t \text{ stat} = -0.9140$ ,  $df = 76$ ,  $p = 0.1804$ ), and cannot conclude that the soil concentration of TcCB is less than the EPA standard.

**5. Examine distributions of the data. Are you concerned that the assumptions of the test might be violated? Why or why not? (If there is a problem, you are not required to fix it.)**

As mentioned above, the data have a strong positive skew and appear to have strong deviations from a normal distribution. This violates one of the main assumptions of the t-test, that our data are normal. The skewness is an issue here because it greatly increases the length of the tail of the distribution, making it highly leptokurtotic. This leads to the sample means not being normally distributed, and thus our t-test assumptions violated.

Log transforming helps to normalize our data, although the transformed data are still skewed. Given the large sample size, however, t-tests on the transformed data are likely to be reliable.

Problem 2

a)  $H_0: \mu_A = \mu_{SR}$

$H_A: \mu_A \neq \mu_{SR}$

b) Because we have measurements at both sites each year, we can do a paired t-test. In fact, if we did a two-sample t-test, we would be violating the assumptions of independence, as the values are correlated between the two sites:

We are doing a two-tailed test, so we can restate the hypotheses:

$H_0: \mu_A - \mu_{SR} = 0$

$H_A: \mu_A - \mu_{SR} \neq 0$

c)  $t = 9.47$ ,  $df = 13$ ,  $P \ll 0.0001$ . We reject the null hypothesis and conclude that kelp densities are different at the two sites.

d) The differences (Anacapa – Santa Rosa) are normally distributed (Shapiro-Wilks test:  $P = 0.207$ ). Thus the assumptions of the test are not likely to be violated.

However, there is a conceptual challenge. Our observations span just over a year, and indicate a consistent difference (Anacapa is almost always higher) over that time. But it would be risky to extend this conclusion across years, especially if large-scale oceanographic shifts occur.