

Solutions to Computer Exercise 2 (R)

1. For males, the mean pupal weight is 152.02 ± 1.73 , and for females it is 144.72 ± 2.47 (males are heavier than females, which is unusual for butterflies, but has to do with nuptial gift giving). Confidence intervals (95%) are [148.57, 155.47] and [139.71, 149.73] for males and females.

2. The sample variances are 194.27 with $df=64$ and 226.37 with $df=36$ for males and females. The sample F is less than the critical ($p=0.584$), so we accept that the variances are the same. The t -test gives $t=2.470$ and $p=0.015$. Thus, we reject the null hypothesis of equal mean pupal weight for the sexes.

3. You should use a paired test, comparing the consumption of each deer from the two types of bowls. If we believe that the distribution (over individuals) of the consumption difference between the bowl types is acceptably close to normal, we can use a paired t test. The command for a paired t -test is

```
t.test(dat$AmntLow, dat$AmntHigh, paired=T)
```

This gives $t = 2.724$ with 9 df , with $p = 0.023$. However, with data from only 10 deer we know rather little about the shape of the distribution. If we are suspicious about the normality, we could use a Wilcoxon matched-pairs test instead. Using the command

```
wilcox.test(dat$AmntLow, dat$AmntHigh, paired=T)
```

the test gives a statistic $V = 51$ and $p = 0.014$. This is a so-called exact p -value. Calculating the exact p -value is very computer intensive, and is an iterative process. Because of this, the approximate p -value was/is often reported for this test (before computers were as clever as they are now). If you want the often used approximation that e.g. Statistica uses, you can use the command

```
wilcox.test(dat$AmntLow, dat$AmntHigh, paired=T, correct=F, exact=F)
```

This test gives $V = 51$ with $p = 0.017$ and the p -value is an approximation. Nevertheless, the conclusion is the same as for the parametric test: we reject the null hypothesis of equal expected consumption from the two types of bowls. Apparently, the deer preferred the less bitter food

4. Using the `cor.test` function, e.g. `cor.test(AmntLow, AmntHigh)`, we get the following (product moment) correlations and p -values ($n = 10$, $df = 8$ in all cases):

	AmntHigh	TimeLow	TimeHigh
AmntLow	$r=-0.41, p=0.24$	$r=0.88, p=0.001$	$r=-0.42, p=0.23$
AmntHigh		$r=-0.22, p=0.53$	$r=0.94, p<0.001$
TimeLow			$r=-0.17, p=0.64$

Since we have tested 6 correlations, we ought perhaps to adjust the significance levels. Using the simple, but maybe too conservative, Bonferroni procedure, the level α is adjusted to $\alpha/6$, which means that a p-value should be smaller than $0.05/6 = 0.0083$ to be regarded as significant. In this case, we get the same conclusion with or without the adjustment: AmntLow and TimeLow are positively correlated, and the same holds for AmntHigh and TimeHigh. Since it takes time to eat, this result is perhaps not so surprising.

If we worry about the distributional assumptions of the test, we could use Spearman rank correlations instead, e.g. `cor.test(AmntLow, AmntHigh, method="spearman")`.

	AmntHigh	TimeLow	TimeHigh
AmntLow	$r_s = -0.4, p = 0.91$	$r_s = 0.93, p < 0.001$	$r_s = -0.03, p = 0.93$
AmntHigh		$r_s = -0.14, p = 0.70$	$r_s = 0.90, p < 0.001$
TimeLow			$r_s = -0.06, p = 0.99$

Although some of these Spearman correlations seem to differ quite much from the product moment correlations (and this might be a reason to be suspicious of them), the conclusion is the same as before.

To get all correlations at once you can use the `rcorr` function in the `Hmisc` package (courtesy of Simon). You first need to remove the column with the individuals' names

```
dat1 <- dat[, -1]
install.packages("Hmisc")
library(Hmisc)
rcorr(dat1)
```

5. Looking at the plot of Grapple against AbsAsymm, one sees that the fit to a linear regression would be quite bad. However, log transforming Grapple gives a rather nice looking fit, with residuals that appear normally distributed with homogeneous variances. A test of the null hypothesis of zero regression slope gives $t = -6.50$ with 98 degrees of freedom, which is highly significant, $p < 0.000001$, so we reject the null hypothesis. The regression equation is:

$$\text{LogGrapp} = 4.8095 - 13.5435 * \text{AbsAsymm}$$

The conclusion is that grapple duration decreases with the absolute value of the size asymmetry.

6. The regressions of log adult weight on log pupal weight look quite fine both for males and for females. We get the following regression equations:

$$\text{males: log adult weight} = -2.051 + 0.905 * \text{log pupal weight}$$

$$\text{females: log adult weight} = -1.424 + 1.136 * \text{log pupal weight}$$

In both cases, the regression slopes are (highly) significantly different from zero.

For the regressions of log head width on log adult weight, there are some outliers (with “small heads”), and one might try removing these to see if they have a strong effect on the regression lines. However, if we skip that we get the following regression equations:

$$\text{males: log head width} = 1.670 + 0.161 * \text{log adult weight}$$

$$\text{females: log head width} = 1.543 + 0.158 * \text{log adult weight}$$

Again, the regression slopes are significantly different from zero.