Solutions to Computer Exercise 5 (R)

1.

(a) The faces get on average 2 mm wider between 5 and 6 years old and this difference is highly statistically significant (t = -19.7203, df = 14, p < 0.000001). In fact, looking at the data one sees that each girl's face gets bigger, which is not so surprising.

(b) A two-way linear mixed model (with girl as random factor) again shows a highly significant effect of age on face width (X^2 = 388.9, df = 1, p < 2.2e-16).

(c) A one-way analysis of variance, which does not use girl as block, fails to demonstrate that the faces get bigger (F = 3.17, df = 1,28, p = 0.086). The reason is that the face width varies among girls and this variation swamps the small increase in width from 5 to 6 years old.

2.

(a) From the one-way analysis of variance, there is a highly significant difference between the two sites: F = 11.66 with 1 and 180 degrees of freedom, leading to p = 0.0008.

(b) From the nested analysis of variance, there is no evidence for a difference between sites: X^2 = 1.671 with 1 degree of freedom, leading to p = 0.196 if you used `Anova()` whereas the equivalent comparison of a reduced model (with `Site` dropped) to a model including `Site` using the `anova()` command gives p = 0.19. The p-values are somewhat different and are only approximate. The figure and variance component estimates indicate that there is a lot of variance between larval groups, the proper sampling unit for testing differences between locations.

(c) The density plot of the effect of site (the difference in pupal weight between the two sites) show that the distribution overlaps zero, indicating that the difference may not be statistically significant. The p-value from `summary()` can vary (as this is a Bayesian analysis) but your teacher got p = 0.224 in her first run.

3.

A nested Anova gives the conclusion that there are significant differences between localities in water fluoride content (p = 0.000009). The variance component estimates are

$$s^2 = 0.017778$$

and

$$s^2_{\beta(\alpha)} = \frac{1}{2}(0.008333 - 0.017778) < 0$$

Since the estimate of the variance among water samples within localities became negative (this can happen!), we revise the estimate to

$$s^2_{\beta(\alpha)} = 0$$

Which is also what the lmer estimate gave.

Running

```
library(lmerTest)

ranova(fm2)
```

leads to $p = 1$ for the random effect of water sample.

Assuming that it is fairly "cheap" to get additional determinations from the same water sample, one would gain from doing it (perhaps taking only one sample per locality). However, the estimate of the among samples within locality variation is not so reliable (only 6 degrees of freedom), so it is safest to get several water samples per locality. When the main cost lies in determining fluoride content, it is always best to get as many water samples as possible if one wants to detect differences between localities.

Note finally that one might be interested in getting an estimate of the error involved in determining fluoride content, in order to be able to report this kind of measurement error. In such a case, two determinations per water sample are perhaps ideal.

5.

First, you need to form the average fluoride content for the nine water samples (averaging over the two determinations per water sample). In this case it is quite easy to do the averaging manually, but you can also do as follows

```
Fluoride2 <- tapply(dat$Fluoride, dat$WtrSmpl, mean)

Loc2 <- factor(c(1,1,1,2,2,2,3,3,3))
```

To see if there are differences between localities, you can now use a Kruskal-Wallis test,

```
kruskal.test(Fluoride2, Loc2)
```

which gives *H* = 7.2605 with 2 degrees of freedom and a *p*-value of 0.027. The conclusion is that fluoride content varies between localities, which is the same conclusion as you got from the nested analysis of variance in the previous exercise. However, note that the p-value you got that time was a lot stronger than the Kruskal-Wallis *p*-value. Thus, either the Kruskal-Wallis test can be much less powerful than analysis of variance, or it was really not correct to do an analysis of variance on these data. It is hard to say which of these possibilities apply, but the Kruskal-Wallis analysis must be regarded as more safe. The anova on the same data gives p = 0.000009 and F = 142.2. This is the same as for the nested anova on the original data.