1. Consider again the data on the amount of physiologically active polyunsaturated fatty acids (PAPFUA, in percentages) for six different brands of diet/imitation margarine. Of these margarines, Mazola and Fleishman’s are corn-based whereas the others are soybean-based.

(a) Perform an F test for the hypothesis that mean PAPFUA does not vary within margarines of the same base (soy or corn) versus the hypothesis that it does vary by brand as well as by base. Describe the two model fits that you would compare in the F test. Does this F test correspond to a t-test on a coefficient? Describe why or why not.

(b) What is the p-value for this test? Describe briefly, as to a client, your conclusions from this test.

(c) Do any assumptions on the model appear to be violated? Provide diagnostic plots to support your conclusion.

2. (Based on a Statistics 572 assignment) A dairy scientist is interested in calibrating a “new” technique for determining mastitis cell counts by comparing it with a pre-existing (“old”) technique. Ten independent samples of milk are used. Each sample is randomly divided into two subsamples, one to be analyzed by each technique. The recorded data consist of a mastitis cell count for each technique (x is the old method, y is the new method) for each of the ten milk samples. You can read the data into R using

```r
> calib = read.table("http://www.stat.wisc.edu/~clayton/stat572/calib.dat")
> names(calib) = c("x", "y")
> str(calib)

'data.frame': 10 obs. of 2 variables:
$ x: int 607 517 833 583 1300 503 603 1390 549 1933
$ y: int 778 496 795 470 1624 551 526 1355 368 1731
```

(a) Plot the data. In what follows we will fit y versus x - do you think that is a good idea? Why or why not?

(b) Fit a simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Is it reasonable to conclude that $\beta_0$ is zero? Perform a formal hypothesis test.

(c) Fit the model $y_i = \beta_1 x_i + \epsilon_i$. Carry out the test $H_0 : \beta_1 = 1$ vs $H_A : \beta_1 \neq 1$. Do this in two ways, as an F test using the `anova` function and as a t-test. How to you fit the restricted model for the `anova` comparison?

(d) Based on the available data, how would you compare the “new” and the “old” technique?

(e) Plot the data. Add the fitted line from the model in part (c). (You will probably find it easier to use “old-style” graphics.) Add point-wise 95% confidence limits on the mean value across the range of x values. Add the line with unit slope and zero intercept (see the documentation for the `abline` function). Does this graphically confirm the results of the test in part (c)?

3. Consider the data on the yield of paddy (a grain grown in India) versus date of harvesting (measured as number of days after flowering) in the data set `xmp13.07` in the `Devore6` package.

(a) Fit the quadratic model to `yield` in two ways as done in the most recent assignment and check the `summary` of each fitted model. The t-tests on the quadratic term give the same p-value for the two fitted models. Explain what is being tested in each case and why the results should be the same.

(b) The t-tests on the linear term give very different p-values for the two fitted models. Explain what is being tested in each case and why the results should be different.