1. A person's blood-clotting ability is typcially expressed in terms of a "prothrombin time," which is defined to be the interval between the initiation of the prothrombin-thrombin (two proteins) reaction and the formation of the final clot. Does *aspirin* affect this function? Measurements made before administration of two tablets and three hours after.

	Prothrombin Times (seconds)		
Subject	Before Aspirin (Y_1)	After Aspirin (Y_2)	Difference (D)
1	12.3	12.0	0.3
2	12.0	12.3	-0.3
3	12.0	12.5	-0.5
4	13.0	12.0	1.0
5	13.0	13.0	0.0
6	12.5	12.5	0.0
7	11.3	10.3	1.0
8	11.8	11.3	0.5
9	11.5	11.5	0
10	11.0	11.5	-0.5
11	11.0	11.0	0
12	11.3	11.5	-0.2

Scatterplot for Blood-clotting study
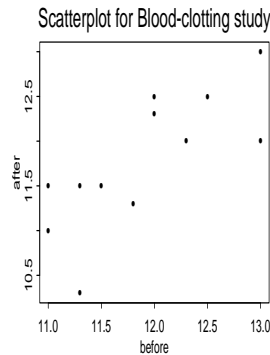
(a) Carry out a paired t-test of the hypothesis that prothrombin time is unaffected by aspirin.

(b) Carry out an F-test of the same hypothesis treating subjects as blocks in an analysis for a RCBD.

(c) (not graded) Show that, in general, the paired t-test is equivalent to the F-test for the RCBD with block size equal to 2. One approach is to write squared differences involving $Y_{ij}, \overline{Y}_{i\cdot}, \overline{Y}_{\cdot j}, \overline{Y}_{\cdot\cdot}$ in terms of differences, $D_j, \overline{D}$.

(d) Consider the mixed model
$$Y_{ij} = \mu + \tau_i + B_j + E_{ij}$$
where $B_j \overset{iid}{\sim} N(0, \sigma_B^2)$ and $E_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ with $B \perp E$ for $i = 1, \ldots, a$ and $j = 1, \ldots, b$.

 i. Show that $E[MS(block)] = \sigma^2 + a\sigma_B^2$

 ii. Use this result to estimate the variance component for subject effects in a mixed model for the prothrombin data.

 iii. Report an estimate of the intra-subject correlation. Is the scatterplot above consistent with this estimate?

2. (taken from Ott and Longnecker 15.10, p. 889) Fuel efficiency of four blends of gasoline is measured in MPG. There is considerable variability due to driver. Another source of variability is model of car. An experiment randomizes four models of car and gasoline blends (A,B,C,D) to drivers according to the design below

| | Model | | | |
Driver	1	2	3	4
1	15.5(A)	33.8(B)	13.7(C)	29.2(D)
2	16.3(B)	26.4(C)	19.1(D)	22.5(A)
3	10.5(C)	31.5(D)	17.5(A)	30.1(B)
4	14.0(D)	34.5(A)	19.7(B)	21.6(C)

(a) Assuming normally distributed data, propose a model in which the effects of model, driver and blend are additive on the mean.

(b) Obtain an ANOVA table for this model.

(c) Report the average fuel efficiency for each blend and the lowest level of significance α at which these averages can be said to differ significantly.

3. For each of several species, an Ecology researcher ran an exposure assay in which groups of $n = 25$ ants are measured for mortality after exposure to one of three different bacteria. For each species, ants are randomized to the three bacteria treatments within each of 15 colonies. That is, a randomized complete block design is used for each species, with colonies serving as complete blocks. Find the results for the DB species on moodle as "DBdat.mtx".

(a) Watch the video on the `hiddenf` package. What is the name of the function in R that will create a directory system containing all of the help files that a developer can complete to create documentation for a package? _____.skeleton.

(b) Obtain an interaction plot with proportion of ants dying out of $n = 50$ on the vertical axis, bacteria treatment on the horizontal axis and lines connecting mortality rates from the same colony.

(c) Either by using the `hiddenf` package in R, or by assigning colonies to groups, obtain an ANOVA table with the following sources of variability: treatments, groups, group-by-treatment interaction, colony within group.

In SAS use can use this code:

```
data DB;
   set DB;
   group=2-(colony in (1,2,4,6,7));
run;
```

```
proc glm;
   *by species; *there were other species but HW8 only uses species=DB;
   class treatment group colony;
   model y=treatment|group colony(group);
   lsmeans treatment|group;
run;
```

(d) Obtain the p-value for the test of group-by-treatment interaction after and report it after multiplying by $2^{15-1} - 1$. Is there evidence that the resistance to bacteria varies across colonies?

(e) Obtain another plot of survival versus bacteria with different lines for colonies. Color the colony lines according to group.

4. An experiment investigates the growth of oysters. Four bags with ten oysters each are randomly placed at four underwater stations next to a power plant:

- Trt1: At the bottom of a discharge canal
- Trt2: At the top of a discharge canal
- Trt3: At the bottom of an intake canal
- Trt4: At the top of an intake canal

(See "oysters.sas" and "oysters.dat" on moodle.) Average initial weight x and final weight y are measured for each of the 16 bags. (Bags serve as the experimental units.) Let $z = x - \bar{x}$ denote the difference from average of the initial weights. SAS code and output to fit an ANCOVA model appear at the end of the exam.

(a) Obtain the F-ratio for a test of equal final weights in a one-way ANOVA where initial weight z is ignored.

(b) Obtain the F-ratio for a test of equal final weights in a one-way ANOVA after controlling for initial weight z.

(c) Use the output to report the unadjusted means for treatments 1 and 4. (Use the 2nd column in the table below.)

Treatment	Unadjusted Mean	Adjusted Mean	Std. Error
1			
–	–	–	––
4			

(d) For bag i, let $x_{i1}, \ldots, x_{i4}$ denote indicator variables for treatments 1-4, respectively. Consider the analysis of covariance model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 z_i + E_i$$

  i. Use the fitted model to report the mean final weight, after adjustment to the average initial weight $\bar{x}$, for treatments 1 and 4. Fill in the appropriate column in the table above. Show work.

  ii. Report the standard error for the adjusted means for locations 1 and 4. (Fill in the table, writing "NP" if it is not possible to give a number based on the provided output.)

(e) Consider the difference between mean final weights under treatments 1 and 4. Estimate this difference after controlling for initial weight. Report a standard error and a p-value for a test of no difference.

```
proc glm;   /* These data taken from Freund, Littell and Spector */
   class trt;
   model final=trt z  /solution;
   means trt;
run;
```

output given next page

```
                              The SAS System                              1
                            The GLM Procedure

                    Class        Levels    Values
                    trt               4    1 2 3 4

                              Sum of
Source                    DF      Squares    Mean Square   F Value   Pr > F
Model                      4   144.7553608    36.1888402    113.59   <.0001
Error                     11     3.5046392     0.3186036
Corrected Total           15   148.2600000

            R-Square     Coeff Var     Root MSE    final Mean
            0.976362      1.747523     0.564450     32.30000

Source                    DF    Type I SS    Mean Square   F Value   Pr > F
trt                        3   29.0450000      9.6816667     30.39   <.0001
z                          1  115.7103608    115.7103608    363.18   <.0001

Source                    DF   Type III SS   Mean Square   F Value   Pr > F
trt                        3    7.6741036      2.5580345      8.03   0.0041
z                          1  115.7103608    115.7103608    363.18   <.0001

                                  Standard
  Parameter          Estimate        Error     t Value    Pr > |t|
  Intercept       32.82685402 B   0.28398639     115.59     <.0001
  trt      1       -1.23028630 B   0.43892225      -2.80     0.0172
  trt      2       -1.36002698 B   0.40124637      -3.39     0.0060
  trt      3        0.48289720 B   0.41086022       1.18     0.2647
  trt      4        0.00000000 B          .           .          .
  z                 1.04670265      0.05492404      19.06     <.0001

NOTE: The X'X matrix has been found to be singular, and a generalized
      inverse was used to solve the normal equations.  Terms whose
      estimates are followed by the letter 'B' are not uniquely estimable.

 Level of        -----------final-----------    ------------z-------------
 trt       N         Mean        Std Dev            Mean         Std Dev
 1         4     34.4750000     3.18891309      2.75000000      3.20572405
 2         4     31.6500000     1.53731367      0.17500000      0.96046864
 3         4     30.8500000     2.95578529     -2.35000000      2.75862284
 4         4     32.2250000     4.29757684     -0.57500000      4.04917687
```