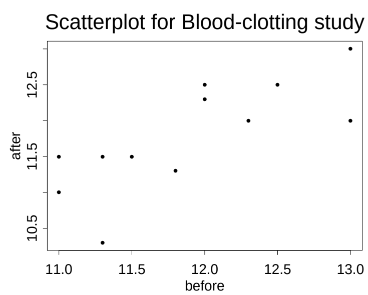
ST 512, Summer Session II, Dr. Jason A. Osborne
Homework assignment #4

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



(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) (For reading only, not to be graded.) Show that, in general, the paired t-test is equivalent to the F-test for the RCBD with block size equal to 2.

(d) (For reading only, not be graded.) 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. It can be shown that the expected block mean square is given by

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

Use this result to estimate the variance component for subject effects in a mixed model for the prothrombin data and to estimate the intra-subject correlation. Is the scatterplot above consistent with this correlation?

2. (Rao 15.11) A 2×3 experiment was conducted to study the effect of two factors, A- the height of the bed (low,high) and row spacing (0.25, .5 and $1m$) - on the moisture content of soils in which a plant variety was grown under a specific type of irrigation practice. The expt. was conducted in 18 plots, divided into blocks of six plots each. The six combinations of the levels of the two factors were assigned at random to the six plots in each block. The data (y is percentage soil water contents at a depth of 100 cm) are available online as "moist.dat". The `i` column is height (0 for low, 1 for high), the `j` column is row spacing, and `k` is the block.

   (a) Write a model for the data.

   (b) Construct an ANOVA table.

   (c) Construct a 95% confidence interval for the difference between mean moisture content at low and high bed heights.

   (d) Use Tukey's procedure to identify significant differences among the three spacings, while controlling the familywise error rate at $\alpha = .05$.

3. (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) Did you model driver and model effects as fixed or random? Why?

   (c) Using $\alpha = 0.05$, test the effects of blend. Is it plausible that different blends lead to the same fuel efficiency, on average?

   (d) Use Tukey's procedure, with familywise error rate $\alpha = .05$, to identify which differences among blends are significant.

4. (Rao 14.1 and 14.5) To determine the *variability* of the protein content of seeds produced in the F_3 generation of a soybean cross, a study was conducted in 30 experimental plots with similar soil and env. characteristics. Seeds from each of a random sample of 10 F_2 generation plants were planted in three randomly selected plots. Suppose that the % protein contents of the seeds produced by the F_3 plants are as follows:

Plant	Plot1	Plot2	Plot3
1	42.4	41.0	39.6
2	38.6	36.3	42.2
3	43.2	42.1	40.2
4	40.8	41.0	38.9
5	41.0	38.3	41.1
6	39.4	39.5	37.2
7	39.6	40.4	38.9
8	38.1	38.3	37.9
9	35.9	36.1	35.6
10	39.6	39.9	39.7

(a) Write a one-way ANOVA for this model.

(b) Using parameter(s) in the model, describe the null hypothesis H_0 that protein content of seeds is constant over the F_2 generation plants.

(c) Using parameter(s) in the model, describe the null hypothesis H_0 that the average protein content of seeds from F_3 generation is 40%

(d) Which parameter in the model quantifies variability in the measured protein contents of F_3 plants grown from the seeds of a single F_2 plant.

(e) Construct an ANOVA table, complete with a column for expected mean squares.

(f) Is the plant-to-plant variation in the protein content of the seeds produced by the F_3 generation plants significant at $\alpha = .05$?

(g) Estimate the variance components.

(h) Estimation the coefficient of variation.

(i) Obtain a 95% confidence interval for the average protein content of seeds from the F_3 generation.

(Two factor random and mixed effects models (Ch. 14) begin here.)

5. Rao 14.16 (p. 685)

(a) Write out an appropriate mixed effects model.

(b) Estimate the variance components associated with each random effect in the model from part a)

(c) Construct a 95% confidence interval for the mean weight gain of animals fed diet 1 in a randomly selected location. (See 14.14, p. 684 and use the fact that

$$
\begin{aligned}
\bar{Y}_{1++} &= \mu + \alpha_1 + \bar{B}_{+} + (\bar{\alpha B})_{1+} + \bar{E}_{1++} \\
\text{Var}(\bar{Y}_{1++}) &= \frac{\sigma^2_B}{b} + \frac{\sigma^2_{\alpha B}}{b} + \frac{\sigma^2}{nb} \\
\widehat{Var}(\bar{Y}_{1++}) &= \frac{\hat{\sigma}^2_B}{b} + \frac{\hat{\sigma}^2_{\alpha B}}{b} + \frac{\hat{\sigma}^2}{nb} \\
&= \frac{MS(AB)}{nb} + \frac{MS(B) - MS(AB)}{na}
\end{aligned}
$$

6. Rao 14.18 (p. 687)

   (a) Write out a model that takes lab effects as fixed. Complete the ANOVA table on p. 687 by adding columns for SS and df and a row for TOTAL.

   (b) Use Tukey's procedure to conduct all pairwise comparisons among labs.

7. Rao 14.20 p.695

   b Construct an ANOVA table for the mixed, nested model for this factorial experiment with subsampling.

   e Report an estimate, a standard error and a 95% confidence interval for the mean difference between the first treatment and the other three:

$$\theta_2 = \tau_1 - \frac{1}{3}(\tau_2 + \tau_3 + \tau_4)$$

8. A disease in wheat called "black point" is known to increase in its incidence as nitrogen (N) increases. An experiment is run to investigate whether this increase is the same when N is broadast as a fertilizer (`fertN`) as when it is delivered during irrigation (`irrN`). Three levels of N (0,60 or 120 kg/ha) are used for each method of delivery, in combination, for a total of $3 \times 3 = 9$ nine treatment combinations. The same level of irrigation N must be applied to entire plots, while levels of fertilizer N may applied to subplots. Both factors are assigned at random, in two separate randomizations, to whole plots and to split-plots. SAS code for fitting a fixed effects model for this experiment, along with output, is included on at the end of this question.

   (a) Using an appropriate mixed model, test the hypotheses listed below. Report F-ratios, associated degrees of freedom and critical values at level $\alpha = .05$.

      i. No interaction effect between fertilizer N and irrigation N
      ii. No main effect of fertilizer N
      iii. No main effect of irrigation N

   (b) Report standard errors for pairwise comparisons among

      i. marginal means for irrigation N, averaging over fertilizer N
      ii. marginal means for fertilizer N, averaging over irrigation N

   (c) Use a Bonferroni correction, along with the standard errors computed in part (b) to identify which of the 6 pairwise comparisons among marginal means may be declared significant, while controlling the familywise error rate at $\alpha = .05$.

   (d) In light of part (c) what conclusions may be drawn about the effects of increasing N on blackpoint incidence using the two methods of delivery. Is it the same?

```
proc glm data=blackpoint;
   class fertN irrN plot;
   model bppct=fertN|irrN plot(irrN);
   lsmeans fertN|irrN;
run;
```

The GLM Procedure
Class Level Information

Class	Levels	Values
fertN	3	0 60 120
irrN	3	0 60 120
plot	9	1 2 3 4 5 6 7 8 9

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	383.5891952	27.3992282	85.23	<.0001
Error	12	3.8578191	0.3214849		
Corrected Total	26	387.4470143			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
fertN	2	320.1546311	160.0773156	497.93	<.0001
irrN	2	46.5377983	23.2688992	72.38	<.0001
fertN*irrN	4	1.6155472	0.4038868	1.26	0.3397
plot(irrN)	6	15.2812185	2.5468698	7.92	0.0013

fert N	bppct LSMEAN
0	5.1235762
60	8.8640675
120	13.5410081

irr N	bppct LSMEAN
0	7.3282793
60	10.2559978
120	9.9443746

fert N	irr N	bppct LSMEAN
0	0	3.2033342
0	60	6.3698114
0	120	5.7975829
60	0	6.6561144
60	60	10.0234499
60	120	9.9126381
120	0	12.1253893
120	60	14.3747323
120	120	14.1229028