

STAT*3210 Experimental Design

G. Umphrey

Assignment #1

Fall 2020

Key solutions to this assignment will be submitted by Crowdmark, for this you will receive a template by email. Solutions to upload must be in either pdf or jpg format files. Further instructions will be forthcoming.

1. The “Weight of Chickens” data set handed out in class is available as a csv file on our course website. We have seen that this is an RCBD with one treatment factor (three treatments). Analyze the data with the assistance of the `aov()` function in R. Explicitly state the null and alternative hypotheses for testing for treatment effects. If appropriate, use Tukey’s HSD procedure at the 5% familywise error rate, and summarize the results with an underscore diagram.
2. Refer to Gary Oehlert’s book, “A First Course in Design and Analysis of Experiments”, which he kindly provides free as a pdf download. For each of the following data sets, concisely describe the design of the experiment (including the treatment design) and provide a model equation similar in format to those on the handout “Model Equations for Some Basic Experimental Designs”. For at least some of the experiments you can refer to an equation number, but specify what is represented by the model equation components. In each case choose a model that would make it possible to estimate experimental error variance. Write out the Source and Degrees of Freedom columns for all experiments, including a row for the Total SS and Total df. (A subset of these will be chosen for submission via Crowdmark. Also note that none of the experiments involve subsampling.)
 - (a) P. 61, Exercise 3.5.
 - (b) P. 62, Problem 3.2.
 - (c) P. 198, Exercise 8.2.
 - (d) P. 198, Problem 8.2.
 - (e) P. 200, Problem 8.4.
 - (f) P. 200-201, Problem 8.5.
 - (g) P. 201, Problem 8.6.
 - (h) P. 202, Problem 8.7.
 - (i) P. 346, Exercise 13.2.
 - (j) P. 350-351, Problem 13.7.
3. The following data set is taken from Bliss (1967), Statistics in Biology, Volume I. Bliss introduces the data: “The hemoglobin content (in grams %) of blood from niacin-depleted dogs was measured before (y_1) and after (y_2) treatment with 25mg niacin, with the following results.”

DOG NO.	1	2	3	4	5	6	7	8
y_1	12.6	12.6	13.7	11.1	11.3	12.2	10.0	11.4
y_2	10.4	11.5	13.6	12.0	10.7	9.3	8.8	9.4

In this exercise you will use R to analyze the data with a paired t -test and also as an RCBD. The data set has been posted in both “unstacked” and “stacked” data formats (make sure you know the difference).

- (a) In terms of means, state appropriate null and alternative hypotheses for this experiment for the most appropriate t -test. Assume that if there is a difference it is expected that the hemoglobin content would decline with the niacin treatment.
- (b). Perform a paired t -test on this data. Briefly summarize your conclusions; include the key evidence (value of t test statistic with associated p -value, and don’t forget the df).
- (c). Analyze the data as an RCBD (Randomized Complete Block Design). What represents the blocks in this experiment? What are appropriate null and alternative hypotheses for the

RCBD analysis?

- (d). What is the relationship between the value of the F test statistic you obtained for testing for treatment effects and the value of the paired t test statistic?
 - (e). What is the advantage of conducting a paired t test over an RCBD for this experiment?
 - (f). If the data was analyzed as a CRD (that is, the blocks were ignored), what changes would occur in the SS(residual) and MS(residual)? [You should be able to answer this without actually running a CRD analysis, but that is a good way to check if you are right!]
4. Refer to the description of the data set “Counting Beetles” (data originally due to Geoffrey Beall, description from Hand et al.’s *A Handbook of Small Data Sets*, to be provided on Monday Sept 21). The experimental design is a Latin Square Design of order 4, with the counters (the men) being the “treatments”.
- (a) We have discussed how to estimate all parameters explicit in the model equation; R provided the parameter estimates quite readily. There is another parameter implicit in the model assumptions, the error variance σ^2 . What quantity estimates the error variance?
 - (b) Show how the fitted (“predicted”) value and the residual are obtained from the parameter estimates for the observation in the lower left corner (observed value is 523).
 - (c) Use R to produce a boxplot of the residuals and briefly comment on what the boxplot tells us.
5. The following data set comes from Sokal & Rohlf (1995). “The oven-dry weights (in grams) of new growth in hybrid poplars grown in concrete soil frames and treated with lime (L), nitrogen (N), phosphorus (P), and potassium (K) are given below. The frames were laid out in three blocks.” Note that O represents a control (no supplement).

Treatments								
Blocks	O	P	PK	K	NK	N	NP	NPK
1	13.9	14.2	14.7	13.6	31.7	57.9	49.5	49.7
2	14.3	22.8	12.8	12.7	25.6	21.7	35.5	38.1
3	15.8	22.1	13.3	15.5	25.7	31.0	30.7	36.3

Treatments								
Blocks	L	LP	LPK	LK	LNK	LN	LNP	LNPK
1	15.3	11.8	17.8	16.6	41.2	43.0	63.8	53.4
2	19.4	23.2	21.4	20.1	59.3	62.5	59.7	53.5
3	15.9	22.7	20.6	15.1	32.0	37.1	41.3	58.5

- (a) How would you describe the design of this experiment? Write out an appropriate model equation for this experiment, if we assume blocks do not interact with treatment factors but all treatment factor interactions are possible.
- (b) How would you set up the data frame for analysis in R?
- (c) Write out the Source and df columns for the ANOVA table.
- (d) How many df would be associated with the error if all interactions involving three or more interactions were pooled with the error?
- (e) For your model, what are the values of $Y_{2\dots}$ and $Y_{21\dots}$?
- (f) How would the randomization have been conducted in this experiment?