Homework #5, ST518

1. A researcher in Poultry Science runs a small experiment using a completely randomized design with $t = 4$ timepoints and $n = 3$ chickens per timepoint. One measurement of basal width is made for each chicken, for a total of 12 measurements. The means and standard deviations are given below:

```
The GLM Procedure

Level of              --------------bw-------------
dayclass      N               Mean           Std Dev

0             3         165.000000        27.4954542
2             3         165.000000        23.4307490
4             3         199.000000        22.5166605
6             3         250.000000        26.2106848
```

   (a) Construct an ANOVA table with rows for day $df = 3$, error and total ($df = 11$).

   (b) Report the F-ratio for a test of the hypothesis that there is no time effect. In other words, are these four observed averages significantly different? Use level of significance $\alpha = .05$.

   (c) Fit a model in which the dependence of mean basal width is a linear function of day. Report the ANOVA table with rows for day $df = 1$, error and total ($df = 11$).

   (d) Test the linear regression model for lack of fit.

   (e) Report the error mean square for both models. Which is smaller?

2. A researcher in Poultry Science is conducting sample size calculations for a completely randomized experimental design with 8 treatment combinations, listed below along with putative means for the response variable base width:

Theoretical treatment means

	Day			
Treatment	0	2	4	6
Control	150	150	150	150
Infected	150	170	200	250

   Pilot experiments suggest that the population standard deviation of base width among chickens under a given condition is $\sigma = 40$.

   (a) For a design that calls for $N = 24$ chickens to be randomized to the 8 combinations ($n = 3$ for each),

      i. report the degrees of freedom associated with the F-ratio for testing the hypothesis that all 8 population treatment means are equal.

      ii. report the non-centrality parameters associated with the F-ratio,

$$\gamma = \frac{n \sum \tau_i^2}{\sigma^2}$$

      iii. obtain the power of the F-test to reject $H_0 : \tau_i \equiv 0$ against the alternative above.

      iv. obtain a plot of the power versus samples size for $n = 2, 3, \ldots, 10$.

   (b) Add lines to the plot from part (b) that correspond to population standard deviations of $\sigma = 35$ and $\sigma = 45$.

3. Suppose that another response variable, apical width was not measured in the pilot study, but has theoretical means given below:

Theoretical treatment means

	Day			
Treatment	0	2	4	6
Control	70	70	70	70
Infected	70	90	120	180

For the first treatment combination, with hypothetical mean 70, the researcher expects all of the chickens to have apical width between 60 and 80, and for the last treatment between 170 and 190.

(a) Report the non-centrality parameter of the distribution of the F-ratio under the theoretical means given above.

(b) Since both base width and apical width will be measured on each chicken, will the power to detect the hypothetical treatment effect on base width be smaller or greater than the power computed in part (a)?

4. Consider all pairwise comparisons among antiobiotics binding fraction means from the CRD discussed in lecture.

(a) Obtain Tukey's HSD that can be used to compare any pair of means.

(b) List the means in order and provide letters next to the means such that any two means with the same letter do not differ significantly. Use software to check your work.

(c) Repeat (a) and (b) using Bonferroni's procedure.

5. This problem is challenging. It will not be graded, but should still be solved and solutions will still be provided. (Exercise 10.6.2 from Third Edition of Dowdy et al) The following table gives costs associated with driving a hybrid for 500 miles. $n = 6$ cars from each of three similar brands (D,E and F) of hybrid were sampled.

	Hybrid Brand		
	D	E	F
	20.3	24.5	21
	19.8	20.8	17.8
	21.1	22	18.1
	18.7	23.1	19.4
	20	23.5	17.5
	20.1	24.1	20.2
mean	20	23	19
st.dev.	.78	1.38	1.42

Consider the model $Y_{ij} = \mu + \tau_i + E_{ij}$ where Y_{ij} denotes cost for car j from brand i and $E_i \overset{iid}{\sim} N(0, \sigma^2)$.

(a) Report the F-ratio associated with a test of $H_0 : \tau_i \equiv 0$. Given that $F(.05, 2, 15) = 3.68$, make a declaration as to whether these observed averages may be said to differ "significantly."

(b) Fisher's Protected Least Significant Difference (LSD) is a protocol that gives weak control of the experimentwise error rate. It involves first checking to see whether the overall F-test of equality is rejected (part (a) above), and then, if rejected, proceeding to carry out pairwise comparisons without further adjustment for multiplicity. If the overall F-test is not significant, then no pairwise comparisons may be conducted. Use Fisher's protected LSD to carry out all three pairwise comparisons among costs for brands D,E and F. Which differences may be declared significant while (weakly) controlling the experimentwise type I error rate?

(c) See the simulation output from "hybrids.sas" on moodle. Give a Monte Carlo estimate of the experimentwise type I error rate when using Fisher's LSD and also when using Tukey's HSD.