Midterm Exam 2

Practice Problems

Scenario 1. Create an RCBD

Reconsider the car additive study. An experiment is to be carried out to study the carbon monoxide emission of gasoline used in cars which has been fortified with a polyether amine dispersant. The researchers wish to test 5 different levels of the additive (2%, 3%, 4%, 5%, 6%) on the CO emission in non-laboratory conditions. As such, the 16 ounces of the assigned additive will be added to a tank of gas. The additive is included to enhance engine performance, but the researchers are concerned it may have a detrimental effect on the carbon monoxide emissions. A sensor will be attached to each vehicle to measure the CO emission in the car exhaust every 5-seconds as it is driven 300-miles. The primary endpoints of interest are the highest amount of CO emitted and the area under the curve (AUC) of the CO emission profile for each vehicle.

a. What might be a limitation of using a strict inclusion criteria for the type of car? e.g., all compact 4-cylinder cars or all small SUVs?

b. Below, sketch a randomized complete block design using 5 blocks based on car type (4-cylinder sedans, 6-cylinder sedans, small SUVs, midsize SUVs, large SUVs) where the experimental unit is a car.

c. Write out the skeleton ANOVA for the study.

d. Write out the statistical effects model.

e. What is the the main goal of blocking on type of car rather than conducting a CRD with all available cars? Think about how your unexplained variablity is treated and what issues might arise if for example, all 4-cylinder sedans get the 2% additive.

Scenario 2. Pistachios `pistachio-data.csv`

A study was carried out to study the effects of a 'dust' treatment versus control on yields of pistachio trees. A dust-like substance called Surround, was sprayed on some trees, while other trees were left untreated (control). The goal of the Surround treatment was to trap moisture in the leaves of the trees, i.e., to keep the moisture from evaporating into the air.

The pistachio grove used in the study had the layout shown below. The grove was divided into 4 plots, which had similar "potential", vigor measurements, and surface temperature measurements based on proximity to the river. Although not randomized, you can see that each treatment (Surround, Control) was applied to a set of trees within each plot. (Note: Treatment should have been randomized within each plot, but issues with the grower, forced the researchers to apply the treatment in the same order within each plot.) Thus, the design is an RCBD.

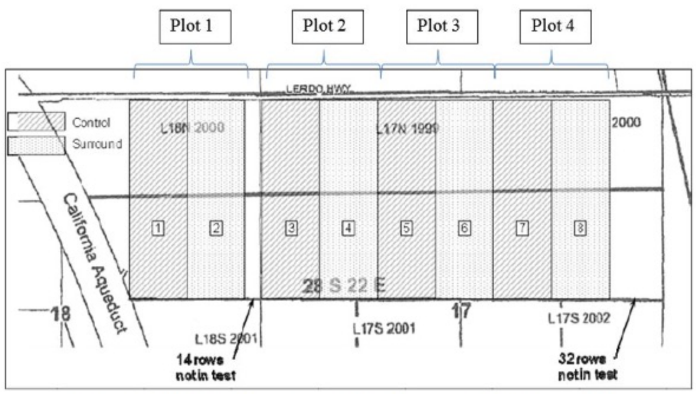

Figure 3. Eight test plots – four control and four treated with Surround™

The response variable of interest for this problem is the "split inshell edible unstained yield" measured in bushels/acre for each plot. Note: Growers get a better price for pistachios that are already split, but still in the shell, at harvest. However, they will harvest and sell even unsplit pistachios.

> ⚠ Warning
>
> Make sure plot is categorical when you read your data into JMP/R.

a. Describe the treatment structure:

b. Describe the design structure:

c. Provide the skeleton of the ANOVA table (SOURCE and DF).

d. **Linear Model:** Analyze the split inshell edible unstained yield in JMP/R, treating the plot as a **fixed** blocking factor.

- Make sure your analysis results are consistent with the skeleton ANOVA table in (c).
- Copy/paste the Effects Test (anova() in R). Identify the estimated block effects ($\hat{\rho}_j$) and residual variance ($\hat{\sigma}^2$).
- Provide a conclusion for the ANOVA F-test for the Surround factor using a 0.10 level of significance. (Note: Be clear about scope of inference. Hint: It is not "all plots similar to those in this study under similar conditions.")

e. Using a residual analysis, confirm the model assumptions are valid. Address both normality and constant variance.

f. Copy/paste the LSMeans table and pairwise comparisons (note with only 2 treatments, a multiplicity adjustment is not needed, thus "Student t's". Interpret results.

g. **Linear Mixed Model:** Refit the model using a **random** block effect.

- Make sure your analysis results are consistent with the skeleton ANOVA table.
- Copy/paste the Effects Test (anova() in R) and REML output (summary() in R). Identify the estimated block ($\hat{\sigma}^2_{blk}$) and residual variance ($\hat{\sigma}^2_{\epsilon}$).
- Provide a conclusion for the ANOVA F-test for the Surround factor using a 0.10 level of significance.

h. Copy/paste the LSMeans table and pairwise comparisons. How do the SE of the means and the SE of the difference in means compare to (f)?

i. What is the scope of inference of your study? How does the scope of inference compare between using a random block effect vs a fixed block effect?

Scenario 3. Coronary Heart Disease. `heart_disease.csv`

A researcher studied the effects of three experimental diets with varying fat contents on the total lipid (fat) level in plasma (Extremely Low, Fairly Low, and Moderately Low). Total lipid level is a widely used predictor of coronary heart disease. Fifteen male subjects were grouped into five blocks according to age. Within each block, the three experimental diets were randomly assigned to the three subjects. Data on reduction in lipid level (grams per liter) after the subjects were on the diet for a fixed period of time was recorded.

a. Identify the treatment design:

b. Identify the experimental design:

c. Sketch out the Skeleton ANOVA table for this study.

d. Write the statistical model for this experiment, making sure to fully explain each component.

e. Check the appropriate model assumptions.

f. Is there a significant difference in the reduction lipid level across the 3 diets? Cite all evidence.

g. Does this mean all the diets differ from one another? Explain.

h. Conduct the Tukey letter groupings and all pairwise comparisons. Which diet results in the largest reduction in lipid level?

Scenario 4. Remotivation. `remotivation_data.csv`

A remotivation team in a psychiatric hospital conducted an experiment to compare five methods for remotivating patients. Patients were grouped according to their level of initial motivation. Patients in each group were randomly assigned to the five remotivation methods. At the end of the experiment patients were evaluated by a team composed of a psychiatrist, a psychologist, a nurse, and a social worker, none of whom was aware of the method to which patients had been assigned. The team assigned each patient a composite score as a measure of their level of motivation.

a. Identify the treatment design.

b. Identify the experimental design.

c. Sketch out the Skeleton ANOVA table for this study.

d. Write the statistical model for this experiment.

e. Determine whether any of the model assumptions have been violated.

f. Do the data provide significant evidence of a difference in method of remotivation? If so, where?

Scenario 5. Nitrogen. `nitrogen_data.csv`

A field plot experiment was conducted to evaluate the interaction between timing of nitrogen application to the soil (early, optimum, late) and two levels of nitrification inhibitor (none, 0.5 lb/acre). The inhibitor delays conversion of ammonium forms of nitrogen into a more mobile nitrate form to reduce leaching loses of fertilizer-derived nitrates. The nitrogen was applied as pulse-labeled 15N through a drip irrigation system at an early, optimum, and late date of application. The data are percent of 15N taken up by sweet corn plants grown on the plots.

a. Identify the treatment design.

b. Identify the experimental design.

c. Sketch the skeleton ANOVA.

d. Write the statistical model for this experiment, making sure to fully explain each component.

e. Check the appropriate model assumptions.

f. Conduct the appropriate analysis and summarize your conclusions below.

Scenario 6. Bond Data `06-bond-data.csv`

An evaluation of diffusion bonding of zircaloy components is performed. The main objective is to determine which of three elements (nickel, iron, or copper) is the best bonding agent. Two components from each ingot are bonded together using each of three agents, and the amount of pressure required to separate the bonded components is measured. This experiment consisted of seven replications.

a. Treatment Structure:

b. Experimental Structure:

c. Sketch the skeleton ANOVA for this study (treat block as a random effect).

d. Write the statistical effects model (treat block as a random effect).

e. What error term is in the denominator of the F-test for element (treatment)?

f. Check model assumptions – residual vs predicted

g. What are $\hat{\sigma}^2_{blk}$ and $\hat{\sigma}^2_{\epsilon}$?

h. Based on the F-test, what conclusion would you make? Carry out the rest of the analysis accordingly (e.g., LSMEANS plot, Tukey letter groupings).

Scenario 7: Fertilizer

Consider an experiment to study the effects of two fertilizer applications. Each fertilizer will be randomly assigned to six pots of plants. Clearly, each plant in the pot will receive the same treatment. At the end of the study, four plants are randomly selected and measured from each pot. The researchers are interested in understanding not only the differences between fertilizers but also the variability between plants.

a. Sketch a study blueprint.

b. Identify the treatment structure.

c. Identify the experimental design structure.

d. Sketch the skeleton ANOVA.

e. Write out the statistical effects model.

Scenario 8: Example 5.3 from Design of Experiments: Statistical Principles of Research Design and Analysis, 2nd Edition by Robert O. Kuehl.
`06-pesticides-data.csv`

One of the concerns following the application of pesticides is the concentration of the pesticide residue that remains on the plants in the field after a certain period of time. An experiment is conducted to compare the ability of two standard chemistry methods (A and B) to recover pesticide residue on cotton plant leaves. Each method is randomly applied to a cotton plant and the concentration of pesticide residue from two leaves from each cotton plant is measured.

a. Identify the treatment design structure:

b. Identify the experimental design structure:

c. Sketch out the skeleton ANOVA (mixed model approach, maintaining subsamples).

d. Write the statistical effects model and assumptions (mixed model approach).

e. Fit the mixed model with subsamples in JMP/R.

f. Is there evidence of a treatment effect? Cite all relevant evidence.

g. What are $\hat{\sigma}^2_\epsilon$ and $\hat{\sigma}^2_s$?

h. Carry out the rest of the analysis according to your findings (e.g., LSMEANS plot, Tukey letter groupings, etc.)