

Due: 5 pm, Friday April 12.

1. Fungal infections of crop plants are a common problem in the midwest where the summers are warm and humid. There is often some genetic control of which variety of fungus can infect which variety of crop plant. This is a study of 2 genetic varieties (A and B, commonly called isolates) of a particular fungal species and 3 genetic varieties of barley (1, 2, and 3). Barley varieties 1 and 3 are suspected to be sensitive to fungus B and resistant to fungus A while Barley variety 2 is sensitive to fungus A and resistant to fungus B. The quantity of interest is the log-transformed expression level of a gene thought to be involved in resistance. This is measured by sequencing RNA.

A recent experiment considered all 6 combinations of fungus genotype and barley genotype. Barley plants will be grown in a growth chamber. Because fungus spreads easily, you can't have two different fungus varieties in the same growth chamber. Barley plants don't move, so you can have more than one barley variety in the same growth chamber.

As a result of this difference in biological behaviour, growth chamber was randomly assigned to a fungus variety. Barley plants are planted in trays. More than one tray of Barley can be grown in a growth chamber. Trays were randomly assigned to one of the three barley varieties. The assigned barley variety was planted in each tray then inoculated with the assigned fungus. The trays were then placed in a growth chamber for six weeks for the barley to grow. There is one observation for each tray of plants. Again, fungal variety was assigned to the growth chamber and barley variety to the tray.

The researchers are interested in the differences in mean gene expression level between barley varieties, between fungal varieties, and especially in the interaction between barley variety and fungal variety. Hence, these three sources of variation will be considered fixed effects.

You have been asked to advise on the design of the study. You have lots of discussion about potential sources of variation with the researchers. You agree that:

- There is considerable random variability among growth chambers
- The variation between growth chambers is not consistent over time. That is two repetitions of the experiment in the same growth chamber are just as variable as running the experiment in different growth chambers at the same time.
- Repetitions of the study differ in many uncontrollable ways (e.g. characteristics of the master collection of each fungus variety).

As a consequence of these discussions, you and the researchers decide to block on repetitions of the study (if a study is repeated more than once) but not block on growth chamber. You consider four different ways to conduct this study.

For each of the following ways to conduct the study, write out the skeleton ANOVA table and indicate the appropriate denominator MS for the F test of each fixed effect (barley, fungus, and the interaction)

- (a) Randomly assign each combination of barley variety and fungus variety to a growth chamber. You have 18 growth chambers, so each of the 6 treatments (combination of barley variety and fungus variety) is replicated 3 times. Each growth chamber has 3 trays of the same barley and fungus variety.
 - (b) You only have 6 growth chambers, so you randomly assign each of the 6 combinations of barley variety and fungus variety to one of these growth chambers (1 chamber per treatment). Again there three trays of the same combination of barley variety and fungus variety (and hence three measurements) in each growth chamber. After you collect the data, you clean the growth chambers and repeat the study with a new randomization of treatment to chamber. You repeat again for a total of three repetitions. (Remember, your analysis should control for unwanted variability among repetitions when there is more than one repetition in the study).
 - (c) You only have 2 growth chambers, so you assign one chamber to fungus A and one to fungus B. Three trays, one of each barley variety, are grown in chamber 1 and three more are grown in chamber 2. After plants are harvested, the growth chambers are cleaned and the study repeated twice more. Again, a total of three repetitions of the study.
 - (d) One disadvantage of the design in part 1c is the time it requires. Your lab has large trays (used in the previous three designs) and small trays. Your professor proposes you make 18 small trays (6 for each barley variety). One growth chamber is randomly assigned to fungus A; the other chamber to fungus B. Nine small trays (3 of each barley variety) are placed in growth chamber 1 (fungus A); the other 9 small trays are placed in growth chamber 2 (fungus B).
 - (e) Without knowing about magnitudes of various sources of variability, it is hard to choose the best design, i.e. the one that provides the most precise estimates of differences. Even so, some proposed designs may have serious issues. Do any of the above designs have serious issues? If so, identify the design(s) that have issues and what those issues are.
2. The data in pigs.txt are birthweights (in pounds) of pigs born to 8 female Poland China pigs. Consider females to be a random effect in this problem. One quite plausible model is $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where Y_{ij} is the birth weight of piglet j in litter i , μ is the mean birthweight in this population, α_i is the deviation of female i from the mean birthweight, and ε_{ij} is the deviation of piglet ij from the female's mean birthweight. Assume that α_i and ε_{ij} are independent normal random variables.
- (a) Use type I (sequential SS) to estimate the variance component between females and the variance between piglets born to the same mom. Report your estimates of the two variance components.
 - (b) Use REML to estimate the variance component between females and the variance between piglets born to the same mom. Report your estimates of the two variance components.
 - (c) Are the REML estimates the same as the type I estimates? Explain why you would (or would not) expect them to be the same.

For all subsequent parts, use the REML estimates of the variance components.

- (d) Estimate the correlation in birthweights between two piglets born to the same female.
 - (e) Estimate the mean birthweight in this population of piglets (any female) and the mean birthweight of a piglet born to female 1.
 - (f) Predict the mean birthweight of new piglet born to female 1.
 - (g) Imagine that in the future you have accumulated data on 20 piglets born to female 1. The population mean, female 1 mean, and variance components are the same as in the current data set. Given this additional information, predict the mean birthweight of a new piglet born to female 1.
 - (h) Estimate the probability that a randomly chosen new piglet born to female 1 will weigh less than 2.5 pounds. You may ignore the additional uncertainty arising from estimating population parameters.
 - (i) Estimate the probability that a randomly chosen piglet born to a randomly chosen female in this population will weigh less than 2.5 pounds. You may ignore the additional uncertainty arising from estimating population parameters.
3. The data in `topsin.csv` are from a study of the sensitivity of different varieties of one fungus species to a particular fungicide. Varieties are called isolates in the data set. A large number of Petri dishes (called plates in the data set) are prepared containing one of 14 different isolates. There are 36 plates for each isolate. The 36 plates per isolate are randomly assigned to one of 6 fungicide concentrations. The response is the log-transformed growth rate of the fungus. In some cases, growth was so slow, it could not be measured accurately. These are indicated by missing values (NA) in the data set. (Yes, it would be better to treat these values as censored, i.e. growth $<$ some value, but that turns a HW problem into a MS creative component project). Log growth is commonly considered to be a linear function of fungicide concentration. Consider these 14 isolates to be randomly chosen from a population of all possible isolates of this species.
- (a) Estimate the mean intercept and slope for this species of fungus. Report your estimates and their standard error.
 - (b) Calculate a 95% confidence interval for the mean slope for this species of fungus.
 - (c) Estimate the mean log growth for this species of fungus at a concentration of 0.003 units of fungicide.
 - (d) If you prepared a new plate of the CUE2 isolate, what would you predict to be its log growth at a concentration of 0.003 units of fungicide?
 - (e) If you randomly sampled a new isolate of this species, what would you predict to be its log growth at a concentration of 0.003 units of fungicide?