

STAT 510 Homework 8
Due Date: 11:00 A.M., Wednesday, March 27

As practice for your upcoming exam, treat these problems as if they were exam questions and try solving them without the use of a computer.

1. A plant scientist was interested in comparing two plant genotypes (denoted as A and B). An experiment was conducted in a greenhouse with one table, eight trays, and sixteen pots. The table in the greenhouse held the eight trays with two pots on each tray. For each of the eight trays, two genotype A seeds were planted in one pot, and two genotype B seeds were planted in the other pot. The assignment of genotypes A and B to the two pots within each tray was determined by flipping a fair coin. The response of interest is a quantitative measurement of overall plant health that was calculated for each plant 42 days after planting. These quantitative measurements of overall plant health are presented as integers in Table 1 to make hand calculations easier, but please answer all questions as if each measurement is a realization from a normal distribution.

Table 1. Measurements of overall plant health for each plant.

Tray	Genotype A Pot		Genotype B Pot	
	Plant 1	Plant 2	Plant 1	Plant 2
1	7	8	6	7
2	8	9	4	5
3	8	8	7	7
4	5	7	2	4
5	5	6	3	4
6	9	10	7	9
7	5	7	1	4
8	4	6	5	5

Let i index genotypes ($i = A, B$), j index trays ($j = 1, \dots, 8$), and k index plants within pots ($k = 1, 2$). Let y_{ijk} denote the response corresponding to genotype i , tray j , and plant k . Suppose

$$y_{ijk} = \mu_i + t_j + p_{ij} + e_{ijk} \quad \forall i, j, k, \quad (1)$$

where μ_A and μ_B are unknown real-valued parameters, $t_j \sim N(0, \sigma_t^2) \forall j$, $p_{ij} \sim N(0, \sigma_p^2) \forall i, j$, $e_{ijk} \sim N(0, \sigma_e^2) \forall i, j, k$, and all t_j , p_{ij} , and e_{ijk} terms are mutually independent.

- (a) Explain what the p_{ij} terms represent and provide one reason for including them in model (1).
- (b) Let $\bar{y}_{ij\cdot} = \frac{1}{2} \sum_{k=1}^2 y_{ijk} \forall i, j$. Determine the distribution of $\bar{y}_{A1\cdot} - \bar{y}_{B1\cdot}$.
- (c) Compute the value of an unbiased estimator of the variance of $\bar{y}_{A1\cdot} - \bar{y}_{B1\cdot}$.
- (d) Provide a 95% confidence interval for $\mu_A - \mu_B$.
- (e) Model (1) can be written in the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$. Provide \mathbf{X} , $\boldsymbol{\beta}$, \mathbf{Z} , and \mathbf{u} .
- (f) Suppose the researchers would like to repeat their experiment again, using the same basic resources: eight trays, two pots per tray, sixteen seeds of genotype A , and sixteen seeds of genotype B . Would you recommend any changes to their experimental design? Explain why or why not.

2. Now suppose the experiment described in problem 1 actually involved a second factor – *soil moisture* with levels *low* and *high* – in addition to the factor *genotype* discussed previously. Four of the trays (2, 4, 5, and 7) were randomly assigned to low soil moisture level, and the remaining four trays (1, 3, 6, and 8) were assigned to high soil moisture level. Throughout the experiment, each pot was watered to maintain the soil moisture level assigned to its tray.

Let ℓ index soil moisture levels with $\ell = 1$ for low soil moisture level and $\ell = 2$ for high soil moisture level. Let $\mu_{i\ell}$ denote the expected value of the response for genotype i and soil moisture level ℓ . Let $h(j) = 1$ for $j \in \{2, 4, 5, 7\}$ and $h(j) = 2$ for $j \in \{1, 3, 6, 8\}$. Note that the function h simply maps the tray index to the index for the soil moisture level assigned to the tray. Thus, $\mu_{ih(j)}$ is the mean response of plants of genotype i grown on tray j at soil moisture level $h(j) \forall i, j$. With y_{ijk} defined as in problem 1, assume that

$$y_{ijk} = \mu_{ih(j)} + t_j + p_{ij} + e_{ijk} \quad \forall i, j, k, \quad (2)$$

where, as in model (1), $t_j \sim N(0, \sigma_t^2) \forall j$, $p_{ij} \sim N(0, \sigma_p^2) \forall i, j$, $e_{ijk} \sim N(0, \sigma_e^2) \forall i, j, k$, and all t_j , p_{ij} , and e_{ijk} terms are mutually independent. Use the raw data in Table 1 (on page 1) and/or the R code and output below to complete part (a) through (f).

```
tray=factor(rep(1:8, each=4))
soilmois=factor(rep(c("H", "L", "H", "L", "L", "H", "L", "H"), each=4))
geno=factor(rep(c("A", "A", "B", "B"), 8))
y=c(7, 8, 6, 7, 8, 9, 4, 5, 8, 8, 7, 7, 5, 7, 2, 4, 5, 6, 3, 4, 9, 10, 7, 9, 5, 7, 1, 4, 4, 6, 5, 5)
anova(lm(y~soilmois+tray+geno+soilmois:geno+tray:geno))
```

Analysis of Variance Table

Response: y

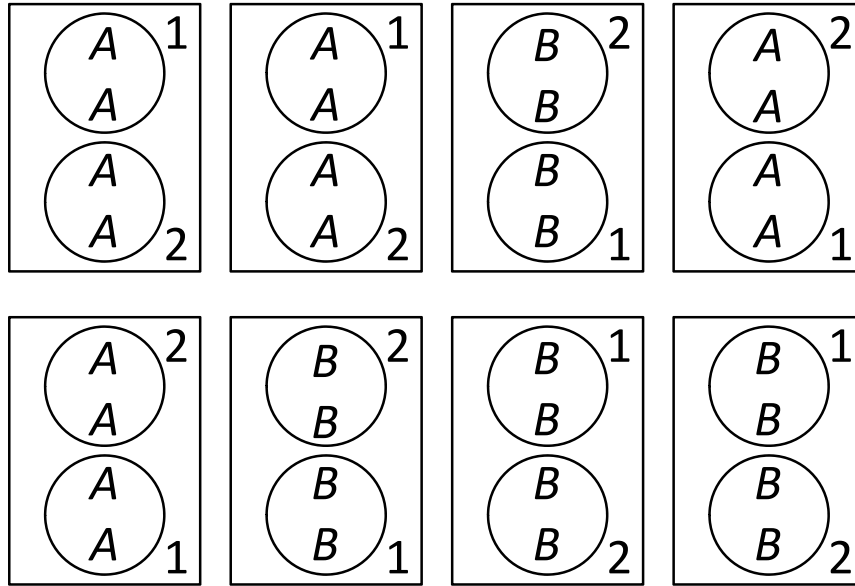
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
soilmois	1	36.125	36.125	32.1111	3.504e-05	***
tray	6	42.375	7.062	6.2778	0.001520	**
geno	1	32.000	32.000	28.4444	6.725e-05	***
soilmois:geno	1	10.125	10.125	9.0000	0.008479	**
tray:geno	6	3.375	0.562	0.5000	0.799299	
Residuals	16	18.000	1.125			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Let $\bar{\mu}_{\cdot\ell} = (\mu_{A\ell} + \mu_{B\ell})/2$ for $\ell = 1, 2$. Compute an F -statistic for testing $H_0 : \bar{\mu}_{\cdot 1} = \bar{\mu}_{\cdot 2}$.
- Provide the degrees of freedom associated with the F -statistic computed in problem 2(a).
- State the null hypothesis of no interaction between the factors genotype and soil moisture in terms of the $\mu_{i\ell}$ parameters.
- Compute a t -statistic for testing the null hypothesis of no interaction between the factors genotype and soil moisture.
- Provide the degrees of freedom associated with the t -statistic computed in problem 2(d).
- If there is interaction between the factors genotype and soil moisture, the test statistic computed in problem 2(d) has a noncentral t -distribution. Determine the noncentrality parameter of this noncentral t -distribution in terms of model (2) parameters.

3. Suppose a second version of the investigation described in problems 1 and 2 was independently conducted using a different experimental design. As before, eight trays were used with two pots per tray and two seeds per pot. However, in the second experiment, genotypes were randomly assigned to trays with four trays per genotype, and the two soil moisture levels were randomly assigned to the two pots within each tray. Figure 1 provides an overhead view of the second experiment in which rectangles correspond to trays, circles to pots, letters to plants (A for a genotype A plant and B for a genotype B plant), and numbers (1 and 2) to the moisture levels assigned to pots.

Figure 1. Overhead view of the second experiment.



Let $\bar{\mu}_i = (\mu_{i1} + \mu_{i2})/2$ for $i = A, B$. In the analysis of the first experiment (described in problems 1 and 2), the Best Linear Unbiased Estimator (BLUE) of $\bar{\mu}_A - \bar{\mu}_B$ took the value 2.0 with a standard error denoted by se_1 . An analysis of data from the second experiment yielded 1.5 as the BLUE of $\bar{\mu}_A - \bar{\mu}_B$ with standard error se_2 .

- Which standard error would you expect to be larger, se_1 or se_2 ? Explain.
- The researchers would like to report a single estimate of $\bar{\mu}_A - \bar{\mu}_B$ that combines the results of both experiments. Provide an expression for the estimate you would recommend. Your answer should be in terms of the estimate 2.0, the estimate 1.5, se_1 , and se_2 .