STAT 741   SAMPLE PROBLEMS

Note: Some problems were taken from previous STAT 741 courses using different notation and textbooks.

1. Consider the following incomplete experimental design in three factors, each at two levels. In regression notation, the model is expressed as a regression model in variables $x_j = \pm 1$, $j = 1, 2, 3$. The data follow a main effects model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e.$$

(a) Suppose the design matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

This is a 2^{3-1}_{III} fractional factorial design. Its defining relation is $I = ABC$. Write the vector of observations as $\mathbf{Y} = [a,\ b,\ c,\ abc]^T$, following the notation that whenever a appears, Factor A is set at its high level, and whenever a is absent, Factor A is set at its low level, etc. For example, a represents the observation $\beta_0 + \beta_1 - \beta_2 - \beta_3 + e$.

Find the least squares estimates of the β_j and calculate $\operatorname{Var} \hat{\beta}_j$, $j = 0, \ldots, 3$.

(b) Suppose that the AB interaction term $\beta_{12} x_1 x_2$ is also present in the model. Which main effect terms, if any, can be estimated using this design?

(c) An alternative design (the *one-at-a-time* design) is given by the matrix

$$\mathbf{Z} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix} \quad \text{with} \ (\mathbf{Z}^T\mathbf{Z})^{-1} = \frac{1}{4} \begin{bmatrix} 4 & 2 & 2 & 2 \\ 2 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 2 \end{bmatrix}.$$

For this design, the observation vector is $[a,\ b,\ c,\ (1)]^T$. Assuming the main effects model, verify that the estimate of the coefficient β_j, $j = 1, 2, 3$, involves the difference of only two observations. Which design is preferable? Explain.

2. Consider the balanced two way ANOVA model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk},$$

$i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, K.$

(a) Assuming that the interaction parameters are all zero, show that $\bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}$ is the least squares estimator of the cell mean $\mu_{ij} = E[Y_{ijk}]$.

(b) In an actual data set, the following ANOVA table was obtained. The entries are Type I sums of squares.

Source	SS	d.f.	F
A	896.75	3	15.76
B	46.50	2	1.25
A*B	641.50	6	5.64
Error	455.04	24	

The F statistics for A and A*B are significant at the 0.001 level. Should these results be used to test for main effects? For interactions? Explain your answers.

(c) Assuming that the interaction terms are nonzero, is the contrast $\alpha_1 - \alpha_2$ estimable?

3. Consider the balanced nested model

$$Y_{ijkl} = \mu + \alpha_i^A + \alpha_{ij}^{B(A)} + \alpha_{ijk}^{C(AB)} + e_{ijkl},$$

where $i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K$ and $l = 1, \ldots, L$. All parameters represent fixed effects, B is nested within A, and C is nested within A and B. The e_{ijkl} are i.i.d $N(0, \sigma^2)$.

(a) Set up the ANOVA table, including formulas for all sums of squares, expected mean squares, and degrees of freedom.

(b) What is the F statistic for testing $H_0 : \alpha_i^A = 0$, $i = 1, \ldots, I$, against a general alternative? What is its distribution under the null hypothesis?

(c) What is the F statistic for testing $H_0 : \alpha_{ijk}^{C(AB)} = 0$, $k = 1, \ldots, K$, against a general alternative? What is its distribution under the null hypothesis?

4. Let $Y_{ij} = \mu + a_i + e_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, be data from a balanced one-way random effects ANOVA, where the a_i are i.i.d. $N(0, \sigma_a^2)$ and the e_{ij} are i.i.d. $N(0, \sigma_e^2)$.

In terms of sample averages and statistics calculated in the usual ANOVA table, find $1 - \alpha$ confidence intervals for μ, σ_e^2 and σ_a^2 / σ_e^2.

5. In an educational experiment, two teaching methods were compared. Four classrooms were selected at random from a population of first grade classrooms and assigned to method 1. An additional four classrooms were randomly selected and assigned to method 2. Each classroom teacher agreed to teach pre-reading skills to her students according to the method assigned to her class. The response was the child's reading readiness score after one month of instruction.

Let Y_{ijk}, $i = 1, 2$, $j = 1, \ldots, 4$, $k = 1, \ldots, 25$, denote the score of child k in classroom j receiving teaching method i.

(a) Write a model equation for Y_{ijk}, indicating clearly the side conditions on any fixed effect parameters and the assumed distributions for any random effects present.

(b) Write down the ANOVA table, including sums of squares, degrees of freedom and expected mean squares.

(c) How would you estimate the treatment means? What are the standard errors of your estimates?

2. An agricultural experiment was performed to compare $I = 5$ fertilizers on a certain variety of cotton. The experiment was run on $J = 6$ blocks, a block being a field. The blocks were thought to be a sample of fields on which the cotton might be grown.

(a) Formulate an appropriate model if the blocks are regarded as chosen at random. Write out the ANOVA table (source, sum of squares and degrees of freedom) and compute the expected mean squares under your model.

(b) Show how to test the hypotheses of no block effect and of no fertilizer effect. In each case find the test statistic and state its distribution under the null hypothesis.

(c) How would you create confidence limits for the fertilizer means and for a difference of fertilizer means?

3

(d) How, if at all, would your answers to (b) and (c) differ if the block effect was thought to be a fixed effect?

2. Consider the two way mixed model $Y_{ij} = \tau_i + b_j + e_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, where $b_j \sim N(0, \sigma_b^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, and the b_j and e_{ij} are mutually independent.

(a) Write out the usual ANOVA table for this problem, including the sums of squares, degrees of freedom and expected mean squares. With no assumptions on the parameters, what is the joint distribution of the sums of squares and the sample treatment means $\bar{Y}_{i.}$?

(b) How would you test the hypothesis of no treatment differences, that is, H_0: $\tau_1 = \cdots = \tau_I$? What is the distribution of your test statistic, under both H_0 and the alternative?

(c) How would you test H_0: $\tau_1 = \cdots = \tau_I = 0$? What is the distribution of your test statistic, under both H_0 and the alternative?

1. Let $Y_{ij} = \mu + a_i + e_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, be data from a one-way random effects ANOVA, where the a_i are i.i.d. $N(0, \sigma_a^2)$ and the e_{ij} are i.i.d. $N(0, \sigma_e^2)$.

(a) Write out the usual ANOVA table and compute the expected mean squares, $E(MS_A)$ and $E(MS_E)$.

(b) Find the distribution of the statistic $F = MS_A/MS_E$ under general conditions.

(c) Find a $1 - \alpha$ confidence interval for the intraclass correlation coefficient

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}.$$

4. In an agricultural study, the weight in pounds (Y) and age in weeks (x) were recorded for samples of turkeys selected from three different treatment groups. The following (full) model was fitted to the data:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_1 z_{ij} + \alpha_2 w_{ij} + e_{ij},$$

where $i = 1, 2, 3$ indexes treatment groups, $j = 1, \ldots, J_i$ indexes turkeys within group, $z_{ij} = I\{i = 1\}$, $w_{ij} = I\{i = 2\}$, and $I\{\cdot\}$ denotes the indicator function of an event. The sample sizes were $J_1 = 4$, $J_2 = 4$, and $J_3 = 5$. Least squares analysis of this model yielded $R^2 = 97.94\%$. By contrast, when the simple linear regression model (reduced model)

$$Y_{ij} = \beta_0^* + \beta_1^* x_{ij} + e_{ij}$$

was fitted to the data, it was found that $R^2 = 64.77\%$.

(a) The experimenters claimed that the large differences in R^2 showed that the treatment differences were significant. Can this statement be verified? If so, calculate an appropriate test statistic and give its distribution under the null hypothesis of no treatment differences. If not, explain why not.

(b) How would you test whether the mean difference between Groups 1 and 2 was nonzero, assuming this comparison had been planned in advance? Would the same testing procedure be used if this comparison was suggested by examination of the data?

4. Suppose we wish to compare the effects of three drugs on people by measuring some response Y. Let Y_{ij} be the response of the jth person taking the ith drug, $i = 1, 2, 3$, $j = 1, \ldots, J$. Assume all the error terms are independent $N(0, \sigma^2)$.

(a) Describe a one-way ANOVA model appropriate for this problem.

(b) Suppose we know that the effect of a drug depends quadratically on the age of the person. Explain how to model this problem.

(c) Suppose there is no interaction between the age and the type of drug. Explain how to model this problem.

(d) Summarize your models in (a), (b), (c) in matrix form.

4. Let Y_{ijk}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, satisfy the mixed effects model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + c_{ij} + e_{ijk}$$

where the c_{ij} are i.i.d. $N(0, \sigma_C^2)$, the e_{ijk} are i.i.d. $N(0, \sigma_e^2)$, and the c_{ij} and e_{ijk} are independent.

(a) Construct the ANOVA table for this model, showing sums of squares, degrees of freedom and expected mean squares.

(b) Show how to construct exact F tests for $H_A : \alpha_i \equiv 0$ and $H_C : \sigma_C^2 = 0$.

(c) Find the power of your test of $H_C : \sigma_C^2 = 0$.

3. In the nested random effects model

$$Y_{ijk} = \mu + a_i + b_{ij} + e_{ijk},$$

$i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, assume that the a_i, b_{ij} and e_{ijk} are mutually independent, that the a_i are i.i.d. $N(0, \sigma_a^2)$, that the b_{ij} are i.i.d. $N(0, \sigma_b^2)$ and that the e_{ijk} are i.i.d. $N(0, \sigma_e^2)$.

(a) Write out the ANOVA table for this model, including the expected mean squares.

(b) Find a $1 - \alpha$ confidence interval for μ.

(c) Suppose $J \to \infty$ while I and K are fixed. Which of μ, σ_a^2, σ_b^2, or σ_e^2 can be estimated consistently?

5. Suppose that $Y_{ij} = \mu + a_i + \beta x_{ij} + e_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$. Assume that the a_i and e_{ij} are mutually independent, that $a_i \sim N(0, \sigma_a^2)$ and that $e_{ij} \sim N(0, \sigma_e^2)$. Assume also that $\bar{x}_{i.} = 0$ for each i.

(a) Show that a unique least squares estimator of β exists. Is there a unique least squares estimator of μ?

(b) Find the joint distribution of $\bar{Y}_{..}$ and $\hat{\beta}$, the least squares estimator of the regression coefficient β.