COMPREHENSIVE WRITTEN EXAMINATION, PAPER III

FRIDAY AUGUST 21, 2009 , 9:00 A.M.–1:00 P.M.

STATISTICS 664 QUESTIONS

1. (70pts) Suppose we have a factor α occurring at $i = 1, ..., I$ levels, with J observations per level. We use the model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \tag{1}$$

where $\mu, \alpha_1, ..., \alpha_I$ are unknown parameters, and ϵ_{ij}'s are independent normally distributed random variables with zero means and constant variance σ^2. Let $Y_{i.} = \sum_{j=1}^{J} Y_{ij}/J$, and for $r \geq 2$ and $s = I - r \geq 2$, let $u = \sum_{i=1}^{I} Y_{i.}/I$, $v = \sum_{i=1}^{r} Y_{i.}/r$, and $w = \sum_{i=r+1}^{I} Y_{i.}/s$. We define

$$SS = \sum_{i=1}^{I}(Y_{i.} - u)^2, SS_1 = \sum_{i=1}^{r}(Y_{i.} - v)^2, SS_2 = \sum_{i=r+1}^{I}(Y_{i.} - w)^2, SS_3 = \sum_{i=1}^{I} rs(v - w)^2/I.$$

You may use matrix notation to answer the following questions, but no matrix should appear in your final answer except (a) and (e).

(a) (5 pts) Rewrite model (1) in matrix form.

(b) (10 pts) Let $\theta = c_0\mu + \sum_{i=1}^{I} c_i\alpha_i$ be an arbitrary linear function of the parameters. Derive necessary and sufficient conditions in terms of c_i for θ to be estimable.

(c) (10 pts) Derive the best linear unbiased estimator (BLUE) for estimable θ.

(d) (10 pts) Give confidence intervals for all of the differences $\alpha_i - \alpha_j$, $i, j \in \{1, 2, ..., I\}, i \neq j$, with simultaneous coverage probability P.

(e) (15 pts) Prove that $SS = SS_1 + SS_2 + SS_3$.

(f) (10 pts) Find the distribution of SS_1, SS_2 and SS_3, and prove that they are independent.

(g) (10 pts) Obtain an F-statistic based on SS_1 and residual sum of squares from Model (1), and give its distribution. What null hypothesis can be tested using this F-statistic?

2. (30pts) Let $Y_i, i = 1, ..., n$ be the dependent variable which are uncorrelated with constant variance, x_i the independent variable which is known, and z_i an indicator variable for fixed t, with $z_i = 0$ for $x_i \leq t$, and $z_i = 1$ otherwise. We consider the following two models:

Model I:

$$Y_i = \beta_0 + \beta_1 z_i + \beta_2 x_i + \beta_3 x_i z_i + \epsilon_i.$$

Model II: Model I with the constraint that $\beta_1 + \beta_3 t = 0$.

(a) (10 pts) For both Model I and Model II, describe in as much detail as possible the proposed relationship between Y and x. (A plot of Y versus x showing the shape of the fitted model can be used as part of the answer)

(b) (10 pts) For Model I, will the ordinary least squares estimator of β_2 depends only on those Y_i for which $x_i \leq t$? If your answer is yes, prove your claim. If your answer is no, provide a counter example.

(c) (10 pts) For Model II, will the restricted least squares estimator of β_2 depends only on those Y_i for which $x_i \leq t$? If your answer is yes, prove your claim. If your answer is no, provide a counter example.

Solutions, CWE 664, 2009

1(a)(b)(c) standard results. 1(d) use Tukey's method: $(Y_{i.} - Y_{j.}) \pm q^*_{1-P,I,IJ-I} s\sqrt{1/J}$, where q^* is the upper $1-P$ point of the studentized range distribution. 1(e) Write all SS in matrix notation and the proof is obvious 1(f) scaled non-central χ^2 distribution with $r-1, s-1$, and one degree of freedom. 1(g) F distribution with $r-1$ and $I(J-1)$ degree of freedom. Testing H_0: $\alpha_1 = \ldots = a_r$.

2(a) Model I two separate regression lines. Model II broken line. (b) Yes (c) No