Preliminary Examination - Multivariate Models

Tuesday, May 3, 2022, 1:00-3:30pm.

Answer all questions and show all your work. When appropriate, you may give your answers for confidence intervals and rejection regions in terms of percentiles of known distributions, e.g., $t_{0.05}(df = 10)$.

1. A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. A process engineer suspects that, in addition to the usual variation in strength within samples, σ^2, of fabric from the same loom, there may also be significant variation, σ_L^2, in strength between looms. To investigate this, she selects three looms at random and makes 3 strength determinations on the fabric manufactured on each loom.

   Let y_{ij} be the jth strength measurement on the ith loom.

   The data was used to calculate the following sum of squares, $\sum\sum(y_{ij} - \bar{y}_{i.})^2 = 24.0$ and $\sum\sum(y_{ij} - \bar{y}_{..})^2 = 68.0$.

   (a) Write down a suitable model for y_{ij} clearly describing all quantities and assumptions associated with the model.

   (b) Let $\boldsymbol{Y} = (y_{11}, y_{12}, y_{13}..., y_{33})$ be the vector of all nine measurements. Give the mean and variance of $\boldsymbol{Y}$ using the model in (a).

   (c) Giva a method of moment estimate and a 95% confidence interval for $\sigma_L^2/(\sigma_L^2 + \sigma^2)$.

   (d) Suppose that the company wants to test the hypotheses that the variance between looms, σ_L^2 is bigger than the variance within samples. Give an exact rejection region of 5% significance level.

2. Consider the model for three stage nested design, $Y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{ijkl}, i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, c; l = 1, \ldots, n$ where τ_i is the effect

of A, $\beta_{j(i)}$ is the effect of B within A, $\gamma_{k(ij)}$ is the effect of C with A and B. Assume that τ_i's, $\beta_{j(i)}$'s, are $\gamma_{k(ij)}$.s are all random, independent, and normally distributed. Construct the ANOVA table including the form of SS (e.g. $SST = \sum_i \sum_j \sum_k \sum_l (Y_{ijkl} - \bar{Y}_{....})^2$, df, MS, F-ratio, and Expected Mean Squares. Give the formulas for estimating the variance components.

3. Steel is normalized by heating above the critical temperature, soaking, and then air cooling. This process increases the strength of the steel, refines the grain, and homogenizes the structure. An experiment is performed to determine the effect of temperature and heat treatment time on the strength of normalized steel. Two temperatures and three heat treatment times are selected. The experiment is performed by heating the oven to a randomly selected temperature and inserting three specimens. After 10 minutes one specimen is removed, after 20 minutes the second specimen is removed, and after 30 minutes the final specimen is removed. Then, the temperature is changed to the other level and the process is repeated. Four shifts are required to collect the data, which are shown below.

		Temperature (F)	
Shift	Time(minutes)	1500	1600
1	10	63	89
	20	54	91
	30	61	62
2	10	50	80
	20	52	72
	30	59	69
3	10	48	73
	20	74	81
	30	71	69
4	10	54	88
	20	48	92
	30	59	64

(a) State the factors of interest in the experiment, and identify the design giving the necessary details to fully describe the design.

(b) Write down the model you would use to fit the data using the notation y_{ijk} for the measurement with Temperature i, Time k, in Shift j.

(c) Suppose that a student unfamiliar with the design calculated the following sum of squares (SS) using the typical SS formulas for a completely randomized design with all factors crossed: SS(Shift)= 145.6, SS(Temp)=2340.38, SS(Time)=159.25, SS(Shift*Temp)=240.46, SS(Shift*Time)=478.42, SS(Temp*Time)=795.25, SS(Total)=4403.63.

Construct a suitable ANOVA table including the the values of F-statistics for the design described at the top of this question.

(d) Find the 95% confidence interval for the difference between the mean strengths for the two temperatures.

4. A national retail chain wanted to study the effects of two advertising campaigns (factor A) on the volume of sales of athletic shoes over time. Ten similar test markets (factor B) were chosen at random to participate in this study. Five test markets were assigned to each campaign at random, and sales data were collected for three two-week time periods (factor C). Let y_{ijk} be the sales measurement on jth Test market using Campaign i during Time period k.

(a) Let $Y_{ij} = (y_{ijk})$ be the vector of sales measurements in Test market j with Campaign i. Write down two models for the vector Y_{ij}, with one having a compound symmetric, and the other having an unrestricted, correlation structure.

(b) Suppose you want to test whether the compound symmetry correlation structure is valid for Y_{ij}. Write down the hypotheses you would test at 5% significance level and state your conclusion using a part of the SAS output given below that was obtained running PROC GLM with repeated statement.

Sphericity Tests				
Variables	**DF**	**Mauchly's Criterion**	**Chi-Square**	**Pr > ChiSq**
Transformed Variates	2	0.5500635	4.1840508	0.1234
Orthogonal Components	2	0.5534131	4.1415532	0.1261

For questions below, assume a compound symmetric correlation structure for the sales measurements in the same test market, and use the following Type III SS obtained from the SAS output: SS(A)=168151, SS(C)=67073, SS(A*C)=391, SS(B(A))=1833681, SS(Total)=2075023. Sample means for the two campaigns are 739.40 and 589.67.

(c) Give the value of the test statistic to determine if there is significant interaction between Campaign and Time period, and state its distribution under the null hypothesis.

(d) Give an unbiased estimator for the difference between the mean sales of the two campaigns, find its standard deviation, and give a 95% confidence interval for the difference.

5. For this question, you may use the following result without proof.

Result: Let B be a $p \times p$ positive definite matrix and $b > 0$ is a scalar. Then

$$\frac{1}{|\Sigma|^b} e^{-tr(\Sigma^{-1}B)/2} \leq \frac{1}{|B|^b}(2b)^{pb} e^{-bp}$$

for all positive definite matrix Σ with equality holding when $\Sigma = (1/2b)B$.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be a random sample form a $p-$dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance matrix Σ.

Suppose we want to test the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ at 5% significance level, where $\boldsymbol{\mu}_0$ is a known $p-$ dimensional real vector.

(a) Derive the likelihood ratio tests statistic for testing this hypotheses.

(b) Give an exact rejection region that is valid for small sample, and an approximate rejection region that is valid for large sample, for the test given above based on this test statistic in part (a).