STAT 714 hw 8

Simultaneous confidence intervals, variance component estimation, mixed models

1. Let $\mathbf{A}$ be an $n \times n$ matrix. Prove each of the following results:
   (a) We have $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\tilde{\mathbf{A}}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, where $\tilde{\mathbf{A}} = (1/2)(\mathbf{A} + \mathbf{A}^T)$.
   (b) If $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$ then $\mathbf{A} = -\mathbf{A}^T$.
   (c) If $\mathbf{A}$ is symmetric, then $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$ for all $\mathbf{x} \in \mathbb{R}^n$ implies $\mathbf{A} = \mathbf{0}$.

2. If a random vector $\mathbf{z}$ has covariance matrix $\mathbf{\Sigma}$ and moment generating function $M_{\mathbf{z}}(\mathbf{t}) = e^{\mathbf{t}^T\boldsymbol{\mu} + \mathbf{t}^T\mathbf{\Sigma}\mathbf{t}/2}$, but $\mathbf{\Sigma}$ is singular, then $\mathbf{z}$ is said to have a singular multivariate Normal distribution. Come up with a way to generate a realization of $\mathbf{z}$ and describe it.

3. Let $Y_{ij} = \mu_i + \varepsilon_{ij}$, $\varepsilon_{ij} \sim \text{Normal}(0, \sigma^2)$ for $i = 1, \ldots, a$ and $j = 1, \ldots, n$. Suppose you are interested in building simultaneous confidence intervals for every contrast comparing a pair of means, that is for $\mu_i - \mu_j$ for all $i \neq j$.
   (a) Give the matrix representation $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ of the model.
   (b) Let $\mathbf{c}_1, \mathbf{c}_2, \ldots$ be the vectors defining the necessary contrasts and let $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \cdots]$. Give the values of the diagonal entries of $\mathbf{C}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{C}$.
   (c) Referring to Lecture 5, run a Monte Carlo simulation to obtain the value of $|t|^{\vee}_{N-a,\alpha}$ such that

   $$P\left(\mu_i - \mu_j \in \left[\bar{y}_{i.} - \bar{y}_{j.} \pm |t|^{\vee}_{N-a,\alpha}\hat{\sigma}\sqrt{2/n}\right] \text{ for all } i \neq j\right),$$

   where $N = na$. Use $\alpha = 0.05$, $n = 6$, and $a = 5$.
   (d) Tukey's HSD method for building simultaneous confidence intervals for all pairwise differences in a balanced one-way ANOVA design prescribes building the intervals

   $$\left[\bar{y}_{i.} - \bar{y}_{j.} \pm q_{a,N-a,\alpha}\hat{\sigma}\sqrt{1/n}\right],$$

   where the values of $q_{a,N-a,\alpha}$ appear in tables in the appendices of many textbooks. Use your Monte Carlo code to verify the numbers highlighted in the table attached to this homework (note that $q_{a,N-a,\alpha} = \sqrt{2}|t|^{\vee}_{N-a,\alpha}$). Each entry in the highlighted row of the table will correspond to a different (n, a) pair. For example, the value 3.96 corresponds to $n = 6$, $a = 4$. Note: Some of the Error df and Number of Groups combinations are not possible with a balanced design (e.g. 20 and 3). You may skip these, as these numbers are obtained with an adjusted method called the Tukey-Kramer method. Note that for the case of $a = 2$ one can run a much simpler simulation.

Table A.6 Critical Values of the Studentized Range, for Tukey's HSD.

Error df	Two-sided α	T = Number of Groups 2	3	4	5	6	7	8
5	0.05	3.64	4.6	5.22	5.67	6.03	6.33	6.58
5	0.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67
6	0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12
6	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61
7	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82
7	0.01	4.95	5.92	6.54	7.00	7.37	7.68	7.94
8	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60
8	0.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47
9	0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43
9	0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13
10	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30
10	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87
11	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20
11	0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67
12	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12
12	0.01	4.32	5.05	5.50	5.84	6.1	6.32	6.51
13	0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05
13	0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37
14	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99
14	0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26
15	0.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94
15	0.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16
16	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90
16	0.01	4.13	4.79	5.19	5.49	5.72	5.91	6.08
17	0.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86
17	0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01
18	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82
18	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94
19	0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79
19	0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89
20	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77
20	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84
25	0.05	2.91	3.52	3.89	4.15	4.36	4.53	4.67
25	0.01	3.94	4.53	4.88	5.14	5.35	5.51	5.65
30	0.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60
30	0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54
40	0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52
40	0.01	3.82	4.37	4.69	4.93	5.11	5.26	5.39
60	0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44
60	0.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25

Table produced using the SAS System using function PROBMC('SRANGE',,1 − α,df,T).

4. Obtain an expression for the REML estimator for σ^2 in the model $Y_i = \mu + \varepsilon_i$, $\varepsilon_i \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2)$, $i = 1, \ldots, n$.

5. Consider the model $Y_{ij} = \mu + \alpha_i + B_j + \varepsilon_{ij}$, where μ and α_i are fixed effects, $B_j \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_B^2)$, and $\varepsilon_{ij} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma_\varepsilon^2)$, $i = 1, \ldots, a$, $j = 1, \ldots, b$.

(a) Write the model in matrix form $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$.

(b) Give $\mathbf{V} = \text{Cov}\,\mathbf{y}$.

(c) Give the expected values of these sums of squares:

  i. $\text{SST} = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{..})^2$

  ii. $\text{SSA} = b \sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$

  iii. $\text{SSB} = a \sum_{j=1}^{b} (\bar{y}_{.j} - \bar{y}_{..})^2$

  iv. $\text{SSAB} = \sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$

(d) A randomized complete block design applied several pre-planting treatments to soybean seeds in different fields (blocking variable). The response is the number of plants, out of 100 planted seeds, which failed to emerge.

	Field			
Treatment	1	2	3	4
Control	8	11	12	13
Avasan	2	5	7	11
Spergon	4	10	9	8
Semaesan	3	6	9	10
Fermate	9	3	5	5

These data are taken from Dr. Michael Longnecker's course notes from 642 at TAMU in 2010.

  i. Obtain (numerically) the REML estimator of the variance of the Field effect.

  ii. Obtain (numerically) the REML estimator of the variance of the error term.

  iii. Obtain a p-value for testing the significance of the treatment effect using the test statistic

$$F_A = \frac{\text{SSA}\,/(a-1)}{\text{SSAB}\,/((a-1)(b-1))}$$

  iv. Obtain a p-value for testing $H_0\colon \sigma_B^2 = 0$ using the test statistic

$$F_B = \frac{\text{SSB}\,/(b-1)}{\text{SSAB}\,/((a-1)(b-1))}$$

  v. Complete an ANOVA table like the one below, providing F values and p values for Treatment and Field.

Source	df	SS	MS	F	p-value
Treatment		SSA			
Field		SSB			
Error		SSAB			
Total		SST			