INDIAN STATISTICAL INSTITUTE

Linear Statistical Models

(B.Stat. 3rd year)

Semester 1, 2025-26

Final Examination

Date: November 14, 2025

Duration: 3 hours

Total score for this exam is 100 points. You are allowed to bring two-sheets (both sided) of notes. You must bring a calculator. All notations have meanings explained in the class.

1. Consider the two-way ANOVA model with balanced design:

$$\begin{aligned} Y_{ijk} &= \mu_{ij} + \varepsilon_{ijk} \\ &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \qquad i = 1, \ldots, a;\ j = 1, \ldots, b;\ k = 1, \ldots, m, \quad (1) \end{aligned}$$

where $\varepsilon_{ijk} \overset{i.i.d.}{\sim} N(0, \sigma^2)$. To ensure identifiability of the parameters, we impose the following *linear constraints*: $\sum_{i=1}^{a} \alpha_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$, $\sum_{i=1}^{a} \gamma_{ij} = 0$ for all j and $\sum_{j=1}^{b} \gamma_{ij} = 0$ for all i.

In the following, show detailed calculations with supporting arguments.

(a) Show that

$$\mathbb{E}(SSA) = \sigma^2(a-1) + mb\sum_{i=1}^{a} \alpha_i^2.$$

[5 points]

(b) Show that

$$\mathbb{E}(SSAB) = \sigma^2(a-1)(b-1) + m\sum_{i=1}^{a}\sum_{j=1}^{b} \gamma_{ij}^2.$$

[6 points]

(c) Suppose Factor A has $a = 3$ levels and Factor B has $b = 4$ levels. For each of the following models, find the expression for the 95% confidence interval for μ_{12} for each of the following submodels.

$$\begin{aligned}
Y_{ijk} &= \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \qquad \sum_{i=1}^{a} \alpha_i = 0,\ \sum_{j=1}^{b} \beta_j = 0 \\
Y_{ijk} &= \mu + \alpha_i + \gamma_{ij} + \varepsilon_{ijk}, \qquad \sum_{i=1}^{a} \alpha_i = 0,\ \sum_{i=1}^{a} \gamma_{ij} = 0 \text{ all } j, \sum_{j=1}^{b} \gamma_{ij} = 0 \text{ all } i \\
Y_{ijk} &= \mu + \gamma_{ij} + \varepsilon_{ijk}, \qquad \sum_{i=1}^{a} \gamma_{ij} = 0 \text{ all } j, \sum_{j=1}^{b} \gamma_{ij} = 0 \text{ all } i.
\end{aligned}$$

[9 points]

[Total: 20 points]

2. Consider the *one factor random effects ANOVA* model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad j = 1, \ldots, m;\ i = 1, \ldots, r,$$

where τ_i are i.i.d. $N(0, \sigma_\mu^2)$, ε_{ij} are i.i.d. $N(0, \sigma^2)$, and $\{\tau_i\}$ and $\{\varepsilon_{ij}\}$ are independent. Define $SSE = \sum_{i=1}^{r} \sum_{j=1}^{m} (Y_{ij} - \overline{Y}_{i\cdot})^2$ and $SSTR = m \sum_{i=1}^{r} (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$, where $\overline{Y}_{\cdot\cdot} = \frac{1}{mr} \sum_{i=1}^{r} \sum_{j=1}^{m} Y_{ij}$ and $\overline{Y}_{i\cdot} = \frac{1}{m} \sum_{j=1}^{m} Y_{ij}$.

Prove the following statements:

(a) $\overline{Y}_{\cdot\cdot}$ has $N(\mu, \vartheta^2)$ distribution, where $\vartheta^2 = (\sigma^2 + m\sigma_\mu^2)/mr$.
[6 points]

(b) $\overline{Y}_{\cdot\cdot}$, SSE and $SSTR$ are independently distributed.
[6 points]

(c) $SSE/\sigma^2 \sim \chi^2_{r(m-1)}$ and $SSTR/(\sigma^2 + m\sigma_\mu^2) \sim \chi^2_{(r-1)}$.
[8 points]

[Total: 20 points]

3. A plant contains a large number of coil winding machines. A production analyst studied a certain characteristic of the wound coils produced by these machines by selecting four machines at random and then choosing 10 coils at random from the day's output of each selected machine. The results follow.

	Coil number (j)									
Machine Number (i)	1	2	3	4	5	6	7	8	9	10
1	205	204	207	202	208	206	209	205	207	206
2	201	204	198	203	209	207	199	206	205	204
3	198	204	196	201	199	203	202	198	202	197
4	210	209	214	215	211	208	210	209	211	210

Assume that an one factor random effects model is appropriate.

(a) Prepare the ANOVA table for this data. [**Hint:** Computation can be reduced by first subtracting 200 from each observation.]

[10 points]

(b) Test whether or not the mean coil characteristic is the same for all machines in the plant; use $\alpha = 0.10$. State the alternatives, decision rule and conclusion.

[4 points]

(c) Estimate the mean coil characteristic for all coil winding machines in the plant; use a 90% confidence interval.

[6 points]

(d) Estimate $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}$ with a 90 percent confidence interval. Interpret your interval estimate.

[10 points]

(e) Test whether or not σ_μ^2 and σ^2 are equal; use $\alpha = 0.10$. State the alternatives, decision rule and conclusion.

[5 points]

[Total: 35 points]

4. The marketing director of a major softwood timber products company has developed a model to predict monthly lumber orders from her firm's domestic markets. Data were collected for the past 24 months for each of the following variables:

Y = lumber orders (in thousands of board feet), $X^{(1)}$ = building permits issued in market region; $X^{(2)}$ = interest rate on conventional first mortgages, $X^{(3)}$ = federal defense spending (in billions of dollars)

The fitted regression equation turned out to be

$$Y = 36.5460 + 0.8011X^{(1)} - 5.3088X^{(2)} + 2.2881X^{(3)}.$$

Here $SSTO = 215.214$, $SSE = 107.474$, $s(b_1) = 0.5256$, $s(b_2) = 2.5394$, $s(b_3) = 1.4301$.

(a) Set up the ANOVA table associated with this regression problem (no need to do any actual test).

[6 points]

(b) Calculate R^2, the coefficient of multiple determination.

[4 points]

(c) If you consider dropping only one variable from the full regression model (containing all three predictor variables), which one is the best candidate for deletion? Explain.

[5 points]

Now suppose, the marketing director also fitted a regression model for predicting lumber orders (Y) from variables $X^{(1)}$ and $X^{(2)}$. The error sum of squares for this multiple regression model turned out to be 121.23.

(d) Can we drop variable $X^{(3)}$? Write down the appropriate null and alternative hypotheses, and then carry out a test to decide this at 0.05 level of significance.

[6 points]

(e) Find the *coefficient of partial determination* $R^2_{Y3|12}$ and interpret it. Is the value of $R^2_{Y3|12}$ consistent with your decision in part (d) ? Explain.

[4 points]

[Total: 25 points]

Useful facts

The following quantiles of t and F distributions will be useful for solving the problems.

Table 1: F quantiles: $F(0.05; \nu_1, \nu_2)$

$\nu_1 \backslash \nu_2$	12	14	15	16	19	20	21	30	36	39
1	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
2	0.052	0.052	0.052	0.052	0.051	0.051	0.051	0.051	0.051	0.051
3	0.114	0.115	0.115	0.115	0.115	0.115	0.116	0.116	0.116	0.116
4	0.169	0.170	0.171	0.171	0.172	0.172	0.173	0.174	0.175	0.175
5	0.214	0.216	0.217	0.217	0.219	0.219	0.220	0.222	0.223	0.224
6	0.250	0.253	0.254	0.255	0.257	0.258	0.259	0.263	0.264	0.265

Table 2: F quantiles: $F(0.95; \nu_1, \nu_2)$

$\nu_1 \backslash \nu_2$	12	14	15	16	19	20	21	30	36	39
1	4.747	4.600	4.543	4.494	4.381	4.351	4.325	4.171	4.113	4.091
2	3.885	3.739	3.682	3.634	3.522	3.493	3.467	3.316	3.259	3.238
3	3.490	3.344	3.287	3.239	3.127	3.098	3.072	2.922	2.866	2.845
4	3.259	3.112	3.055	3.007	2.895	2.866	2.840	2.690	2.634	2.612
5	3.106	2.958	2.901	2.852	2.740	2.711	2.685	2.533	2.477	2.456
6	2.996	2.848	2.790	2.741	2.628	2.599	2.573	2.421	2.364	2.342

Table 3: t quantiles: t_ν^A

$A \backslash \nu$	3	4	5	6	10	12	14	16	20	21	30	36	39
0.90	1.638	1.533	1.476	1.383	1.372	1.356	1.345	1.337	1.325	1.323	1.310	1.306	1.303
0.95	2.156	2.132	2.0150	1.833	2.228	2.179	2.145	2.120	1.725	1.721	1.697	1.688	1.685
0.975	3.182	2.776	2.571	2.262	2.764	2.681	2.624	2.583	2.086	2.080	2.042	2.028	2.023