Applied Statistics Preliminary Examination
Theory of Linear Models
May 2024

Instructions:

- This preliminary examination consists of two parts: Linear Models and Design of Experiments.
- For this Linear Models portion, work all 3 problems. Neither calculators nor electronic devices of any kind are allowed. Show all your work and clearly state any theorem or fact that you use. Each of the 12 parts carries an equal weight of 10 points.
- Abbreviations/Acronyms:
  - IID – Independent and Identically Distributed
  - LSE – Least Squares Estimator.
  - OLS – Ordinary Least Squares estimator; synonymous with LSE.
  - GLS – Generalized Least Squares estimator.
- Notation:
  - $\boldsymbol{x}^T$ or $\boldsymbol{A}^T$: indicates the transpose of vector $\boldsymbol{x}$ or matrix $\boldsymbol{A}$.
  - $\text{tr}(\boldsymbol{A})$ and $|\boldsymbol{A}|$: denotes the trace and determinant, respectively, of $\boldsymbol{A}$.
  - $\boldsymbol{I}_n$: the $n \times n$ identity matrix.
  - $\boldsymbol{j}_n = (1, \dots, 1)^T$ is an n-vector of ones, and $\boldsymbol{J}_{m,n}$ is an $m \times n$ matrix of ones.
  - $\mathbb{E}(\boldsymbol{x})$ and $\mathbb{V}(\boldsymbol{x})$: expectation and variance(-covariance) of a random vector $\boldsymbol{x}$.
  - $x_i \sim IID(\mu, \sigma^2)$ indicates that the random variables $x_1, \dots, x_n$ are IID with mean μ and variance σ^2 with *no specific distribution assumed*.
  - $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: the m-dimensional random vector $\boldsymbol{x}$ has a multivariate normal distribution (or univariate normal if $m = 1$) with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Also: MVN_m.
  - $X \sim t(n, \lambda)$: a t distribution with n degrees of freedom and noncentrality parameter λ. If $\lambda = 0$, we write simply $X \sim t(n)$.
  - $X \sim F(n_1, n_2, \lambda)$: an F distribution with n_1 numerator and n_2 denominator degrees of freedom, and noncentrality parameter λ. If $\lambda = 0$, we write simply $X \sim F(n_1, n_2)$.
- Possibly useful results:
  - If $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given in partitioned form as:
  $$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \ (\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T),$$
  with $m_1 = \dim(\boldsymbol{x}_1)$, then the conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2$ is
  $$\boldsymbol{x}_1 | \boldsymbol{x}_2 \sim N_{m_1}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

1. Let $\boldsymbol{y} = (y_1, y_2, y_3)^T \sim N_3(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_3)$, where $\boldsymbol{\mu} = (3, -2, 1)^T$, and define the following matrices:
$$\boldsymbol{A} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$
In addition, let $z = y_1 + y_2 + y_3$, $q = \boldsymbol{y}^T \boldsymbol{A} \boldsymbol{y}$, and $w = (y_1 - 3)^2 + (y_2 + 2)^2 + (y_3 - 1)^2$.
   (a) Find the distribution of q/σ^2.
   (b) Are q and $\boldsymbol{B}\boldsymbol{y}$ independent? Justify your answer.
   (c) Find the distribution of z. Are z and q independent? Justify your answer.
   (d) True or False: If z and w are independent, then $z/\sqrt{w} \sim t(3, \lambda)$, where $\lambda = 2/\sqrt{3\sigma^2}$. Justify your answer. (Note: You don't need to determine if z and w are independent; assume that they are.)

2. Suppose that for $i = 1,2,3$ we fit the linear model
$$y_i = \beta_0^* + \beta_1^* x_i + \varepsilon_i^*, \quad \varepsilon_i^* \sim IID(0, \sigma^2),$$
when in fact the true model is
$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i, \quad \varepsilon_i \sim IID(0, \sigma^2),$$
and the design points are $(x_1, x_2, x_3) = (-1, 0, 1)$. Let $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ denote the LSEs of the fitted model.
   (a) Prove the following general result: If we fit the model $\boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1^* + \boldsymbol{\varepsilon}^*$ when in fact the true model is $\boldsymbol{y} = \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, then the expectation of the LSE of $\boldsymbol{\beta}_1^*$ is $\mathbb{E}(\widehat{\boldsymbol{\beta}}_1^*) = \boldsymbol{\beta}_1 + \boldsymbol{A}\boldsymbol{\beta}_2$, where $\boldsymbol{A} = (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{X}_2$.
   (b) For the setup posed in the introduction to this problem, find the expectation of the LSEs, i.e., find $\mathbb{E}(\hat{\beta}_0^*)$ and $\mathbb{E}(\hat{\beta}_1^*)$.
   (c) For $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ as described in the introduction to this problem, find $\mathbb{V}(\hat{\beta}_0^*)$ and $\mathbb{V}(\hat{\beta}_1^*)$.
   (d) For $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ as described in the introduction to this problem, find the correlation between the LSEs $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.

3. Consider the observations $\{y_{11}, \dots, y_{33}\}$ from the linear model
$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i, j = 1,2,3,$$
where the ε_{ij} are IID $N(0, \sigma^2)$, and $\{\mu, \tau_1, \tau_2, \tau_3\}$ are unknown parameters to be estimated. Of particular interest is the linear combination of parameters $\eta = 2\mu + \tau_1 + \tau_2$.
   (a) Determine the form of all *estimable* functions, and hence show that η is estimable.
   (b) Find the BLUE of η.
   (c) Show (with justification) that the following null hypothesis is *testable*:
$$H_0: \tau_1 = \tau_2 = \tau_3.$$
   (d) Propose a test statistic for the test in (c), and compute its distribution both under H_0 and under the alternative hypothesis.

Applied Statistics Preliminary Examination
Design of Experiments
May 2024

Instructions:

- For this Design of Experiments portion, work all 3 problems. Each of the 13 parts carries an equal weight of 10 points.

1. A study of the difference of 6 types of diets (factor A) on the weight gain of young rabbits is proposed. Because weight varies considerably among young rabbits, it is proposed to block the experiment based on ten breeds available to the researchers (factor B). The response variable is the weight gain of young rabbits. Assume Diet (A) is a fixed factor and that Breed (B) is a random factor. There are no interaction effects. Write down the statistical models and assumptions for each of the following designs.
   (a) There are 6 rabbits available for each of the 10 breeds. Each diet was applied to one rabbit at random for each breed. Weight gain for each rabbit was measured.
   (b) There are 12 rabbits available for each of the 10 breeds. Each diet was applied to two rabbits at random for each breed. Weight gain for each rabbit was measured.
   (c) There are 2 litters of 6 rabbits for each of the 10 breeds. Each diet was applied to one rabbit at random within each litter. Assume Litter (factor C) is a random factor. Weight gain for each rabbit was measured.

2. An investigator is interested in improving the number of rounds per minute fired from a Navy gun. He believes a new method of loading the gun will increase the number of rounds fired per minute. A team of people is needed to use this gun, so he divided the personnel into groups based on physique (slight, average, and heavy). He selected three teams from each of these groupings for the experiment. Each team was presented with both methods of loading and used each method twice in a random order. The number of rounds per minute fired from the gun was measured for each replication. The data structure is shown below:

Groups	Slight			Normal			Heavy		
Teams	1	2	3	4	5	6	7	8	9
Method 1	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x
Method 2	x	x	x	x	x	x	x	x	x
	x	x	x	x	x	x	x	x	x

Assume Method (factor A) is a fixed factor, Group (factor B) is a fixed factor, and Team (factor C) is a random factor. Suppose that the interaction between Method and Group and between Method and Team are to be included in the model. Denote y_{ijkl} as the l^{th} measure of the number of rounds per minute fired from the gun for the k^{th} team from the j^{th} group using the i^{th} method. The average number of rounds per minute fired by using the two methods are $\bar{y}_{1\ldots} = 23.6$ and $\bar{y}_{2\ldots} = 15.1$.

(a) Write down a model for this experiment. Clearly define all the terms and state all relevant assumptions. Additionally, explain (briefly) why an interaction between Group and Team cannot be included.

(b) Fill in the ANOVA table below. **Note:** Be sure to use the restricted (sum-to-zero) assumption for interactions between fixed and random factors.

Source	DF	MS	E(MS)
Method	______	652.0	______
Group	______	8.0	______
Team(Group)	______	6.5	______
Group*Method	______	0.6	______
Team(Group)*Method	______	1.8	______
Error	______	2.3	______
Total	______		

(c) State and conduct a statistical test for a Method effect at 0.05 significance level and interpret the test results.

(d) State and conduct a statistical test for a Group effect at 0.05 significance level and interpret the test results.

(e) Test whether the variance component for the Method × Team interaction is significant at 0.05 significance level and interpret the test results.

(f) Construct a 95% confidence interval for the effect difference of the two methods.

3. Consider a two-way design with fixed factors. Suppose that factor A has two levels while factor B has 4 levels. Further suppose that there are 2 replications. The means model for this design is then:

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}; i = 1,2, j = 1,2,3,4, k = 1,2$$

Denote $\mu_{i\cdot} = \frac{1}{4}\sum_{j=1}^{4}\mu_{ij}$, $\mu_{\cdot j} = \frac{1}{2}\sum_{i=1}^{2}\mu_{ij}$, and $\mu_{\cdot\cdot} = \frac{1}{8}\sum_{i=1}^{2}\sum_{j=1}^{4}\mu_{ij}$.

(a) State the following hypothesis in terms of the μ_{ij}'s (and possibly $\mu_{i\cdot}, \mu_{\cdot j}$, and/or $\mu_{\cdot\cdot}$):

H_0: There is no main effect for Factor A

(b) State the following hypothesis in terms of the μ_{ij}'s (and possibly $\mu_{i\cdot}, \mu_{\cdot j}$, and/or $\mu_{\cdot\cdot}$):

H_0: There is no interaction between Factor A and Factor B

(c) If there is an interaction between Factor A and Factor B, then what is the interpretation of the ANOVA F-test for the main effect of A? (That is, the test using $F = \frac{MS_A}{MS_E}$ as the test statistic, where MS_A is the mean-squares for A and MS_E is the mean-squares for error.)

(d) If there is an interaction between Factor A and Factor B, then how can the experimenter compare the two levels of Factor A?

F-table of Critical Values of α = 0.05 for F(df1, df2)

	DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00