

Due: Thursday, December 6

1. Consider a matrix of the form $a\mathbf{I} + b\mathbf{1}\mathbf{1}'$, where $a, b \in \mathbb{R}$.

(a) Fill in the blank in the statement below with a condition potentially involving a, b , and n , and prove that your answer is correct.

$a\mathbf{I} + b\mathbf{1}\mathbf{1}'$ is positive definite if and only if _____.

(b) Suppose $a\mathbf{I} + b\mathbf{1}\mathbf{1}'$ is nonsingular. Show that the inverse of $a\mathbf{I} + b\mathbf{1}\mathbf{1}'$ is of the form $c\mathbf{I} - d\mathbf{1}\mathbf{1}'$ and provide expressions for c and d in terms of a, b , and n .

2. Prove the result on ranks given at the top of slide 24 of slide set 30. That is, show

$$\text{rank}(\mathbf{X}(\mathbf{I} - \mathbf{C}^{-1}\mathbf{C})) = \text{rank}(\mathbf{X}) - \text{rank}(\mathbf{C})$$

when $\mathbf{C}\beta$ is estimable.

3. Suppose $\mathbf{y} = \mathbf{X}\beta + \epsilon$, where $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{V})$ and \mathbf{V} is a known and nonsingular matrix. Suppose $\mathbf{C}\beta$ is a vector of q linearly independent estimable functions and that $\hat{\beta}$ is any solution to the Aitken equations.

(a) Provide a $100(1 - \alpha)\%$ confidence region for $\mathbf{C}\beta$.

(b) Prove that your confidence region in part (a) has exact coverage probability $1 - \alpha$. As usual, you may use any results that have been established in our course notes to simplify your proof.

(c) Suppose an experiment has been conducted to assess the effects of three feed types (1, 2, and 3) on weight gain in pigs. A total of 6 pens of pigs were used in the experiment. The three feed types were randomly assigned to the 6 pens with two pens per feed type. At the conclusion of the study, the average weight gained by pigs was computed separately for each pen. Due to illness that occurred at the end of the study and was unrelated to the feed type, not all pens ended up with the same number of pigs. The average weight gains and number of pigs per pen are provided in the table below.

Pen	Feed Type	Average Weight Gain (pounds)	Number of Pigs
A	1	79	6
B	1	69	2
C	2	44	6
D	2	52	6
E	3	91	1
F	3	75	6

Let y_{ijk} denote the weight gained by the k^{th} pig in the j^{th} pen for the i^{th} treatment, and let n_{ij} denote the number of pigs in the j^{th} pen for the i^{th} treatment. Suppose

$$y_{ijk} = \mu + \tau_i + P_{ij} + e_{ijk} \text{ for } i = 1, 2, 3; j = 1, 2; \text{ and } k = 1, \dots, n_{ij},$$

where the P_{ij} are i.i.d. $N(0, \sigma_P^2)$ independent of the e_{ijk} which are i.i.d. $N(0, \sigma_e^2)$; $\mu, \tau_1, \tau_2, \tau_3$ are unknown parameters, and σ_P^2 and σ_e^2 are unknown positive variance components with $\sigma_P^2/\sigma_e^2 = 2$. Provide a 95% confidence region for $(\tau_1 - \tau_2, \tau_1 - \tau_3)'$ and sketch the region in the (x, y) -plane.

(d) Is there evidence of some difference among the three treatment means? Explain.

4. Consider the model

$$y_i = \beta + bx_i + e_i \text{ for } i = 1, \dots, n \quad (1)$$

where x_1, \dots, x_n are known fixed values, β is an unknown parameter, $b \sim N(0, \sigma_b^2)$ independently of $e_1, \dots, e_n \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. This is a simple linear regression model with a random slope coefficient. The parameter space for the model is $\{(\beta, \sigma_b^2, \sigma^2) : \beta \in \mathbb{R}, \sigma_b^2 > 0, \sigma^2 > 0\}$. Let $\mathbf{y} = (y_1, \dots, y_n)'$, $\mathbf{x} = (x_1, \dots, x_n)'$, and $\mathbf{X} = [\mathbf{1}, \mathbf{x}]$. Assume that $\text{rank}(\mathbf{X}) = 2$.

(a) Determine $\text{Var}(\mathbf{y})$.

(b) Consider the quadratic forms $\mathbf{y}'(\mathbf{P}_X - \mathbf{P}_x)\mathbf{y}/\sigma^2$ and $\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}/\sigma^2$.

i. Prove that $\mathbf{y}'(\mathbf{P}_X - \mathbf{P}_x)\mathbf{y}/\sigma^2$ has a χ^2 distribution under model (1) and provide fully simplified, non-matrix expressions for its degrees of freedom and non-centrality parameter.

ii. Prove that $\mathbf{y}'(\mathbf{I} - \mathbf{P}_X)\mathbf{y}/\sigma^2$ has a χ^2 distribution under model (1) and provide fully simplified, non-matrix expressions for its degrees of freedom and non-centrality parameter.

iii. Are the quadratic forms independent of one another?

(c) Explain how to use the quadratic forms in part (b) to test $H_0 : \beta = 0$ vs. $H_A : \beta \neq 0$. Specifically,

i. provide a test statistic,

ii. state its distribution under H_0 ,

iii. state its distribution under H_A , and

iv. explain how you would use the test statistic to obtain a p -value for the test of H_0 vs. H_A .

5. Suppose $y_1, \dots, y_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$.

(a) Provide a vector of LIN error contrasts that could be used to obtain a REML estimate of σ^2 .

(b) Specify the joint distribution of the error contrasts provided in part (a).

(c) Write down the likelihood function for the error contrasts provided in part (a).

- (d) Find a simplified expression for the REML estimate of σ^2 from the error contrasts provided in part (a) by finding the value of σ^2 that maximizes the likelihood provided in part (c).
- (e) How does your answer to part (c) compare to the usual unbiased estimate of σ^2 ?
- (f) Would your answer to part (d) change if you were to choose a different set of LIN error contrasts in part (a)?
6. Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Consider partitioning \mathbf{y} into its first element and the rest as $\mathbf{y} = [y_1, \mathbf{y}'_2]'$. Similarly partition \mathbf{X} as $\mathbf{X} = [\mathbf{x}_1, \mathbf{X}'_2]'$ so that \mathbf{x}'_1 represents the first row of the \mathbf{X} matrix. Suppose that \mathbf{X} and \mathbf{X}_2 are both of full column rank. The *deleted residual* for the first observation is defined as $d_1 = y_1 - \hat{y}_{1(1)}$, where $\hat{y}_{1(1)} = \mathbf{x}'_1 \hat{\boldsymbol{\beta}}_{(1)} = \mathbf{x}'_1 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{y}_2$.

- (a) Show that $\mathbf{x}'_1 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{y}_2$ is the BLUP of y_1 if we were to delete y_1 from the data set.
- (b) Prove that $d_1 = \frac{y_1 - \hat{y}_1}{1 - \mathbf{x}'_1 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_1}$, where $\hat{y}_1 = \mathbf{x}'_1 \hat{\boldsymbol{\beta}} = \mathbf{x}'_1 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$. This equation shows that the deleted residual is the ordinary residual divided by one minus the observation's *leverage* (recall the diagonal elements of $\mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$ are known as the leverage values). Note that this expression for the deleted residual allows the computation of all deleted residuals from the fit of a single multiple regression, a fact that you most likely learned about in STAT 500.
- (c) Determine the distribution of d_1 .
- (d) Let $\hat{\sigma}_{(1)}^2 = (\mathbf{y}_2 - \mathbf{X}_2 \hat{\boldsymbol{\beta}}_{(1)})' (\mathbf{y}_2 - \mathbf{X}_2 \hat{\boldsymbol{\beta}}_{(1)}) / (n - p - 1)$. Determine the distribution of

$$\frac{d_1}{\hat{\sigma}_{(1)} \sqrt{1 + \mathbf{x}'_1 (\mathbf{X}'_2 \mathbf{X}_2)^{-1} \mathbf{x}_1}}.$$

- (e) Now suppose $\text{Var}(\varepsilon_1) = \sigma^2 + \sigma_\delta^2$, where $\sigma_\delta^2 \geq 0$. Explain how the statistic in part (d) could be used to test the hypotheses

$$H_0 : \sigma_\delta^2 = 0 \text{ vs. } H_A : \sigma_\delta^2 > 0.$$