

There may be more than one way to correctly answer a question, or several ways to describe the same answer. Not all of the possible correct answers may be listed here.

1. (a) (6 points) Since the statement of the problem implies that the conditional means of the responses are linear functions of the parameters $\beta = (\beta_1, \beta_2)^T$, to obtain a Gauss-Markov model we simply need to add that $\text{Var}(\mathbf{Y}) = \sigma^2 \mathbf{I}$. To obtain a normal theory Gauss-Markov model we must have $\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$.

(b) (4 points) $2\beta_2$ is the difference between the mean amount of escaped vapor when the device is installed in the gas tank and the mean amount of escaped vapor when no device is installed, and gasoline temperature kept the same for both cases.

(c) (6 points) $\frac{1}{\sigma^2} \text{SSE} = \frac{1}{\sigma^2} \mathbf{Y}^T (\mathbf{I} - \mathbf{P}_X) \mathbf{Y}$ has a central chi-squared distribution with 2 degrees of freedom, because $\text{rank}(\mathbf{I} - \mathbf{P}_X) = 4 - 2 = 2$. The non-centrality parameter is $\delta^2 = \beta^T \mathbf{X}^T (\mathbf{I} - \mathbf{P}_X) \mathbf{X} \beta = 0$.

(d) (6 points) Since the model matrix has full column rank, $\beta = (\beta_1, \beta_2)^T$ is estimable and the unique solution to the normal equations is $\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. For the normal theory Gauss-Markov model in part (a), \mathbf{b} is the unique best linear unbiased estimator for $\beta = (\beta_1, \beta_2)^T$. Also, $\mathbf{b} \sim \mathbf{N}(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$.

(e) (14 points) There are several ways to approach this problem. You could use the fact that $\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$ to argue that

$$\mathbf{c}^T \mathbf{Y} = (-1, -1, 1, 1) \mathbf{Y} = Y_3 + Y_4 - Y_1 - Y_2 \sim \mathbf{N}(\mathbf{c}^T \mathbf{X}\beta, \sigma^2 \mathbf{c}^T \mathbf{c}),$$

where $\mathbf{c}^T \mathbf{X}\beta = 4\beta_2$ and $\sigma^2 \mathbf{c}^T \mathbf{c} = 4\sigma^2$. Then, $\frac{(Y_3 + Y_4 - Y_1 - Y_2)}{2\sigma} \sim \mathbf{N}\left(\frac{2\beta_2}{\sigma}, 1\right)$ and by

the definition of the non-central chi-squared distribution $\frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{4\sigma^2} \sim \chi_1^2\left(\frac{4\beta_2^2}{\sigma^2}, 0\right)$.

Next show that $\frac{1}{\sigma^2} \text{SSE} = \frac{1}{\sigma^2} \mathbf{Y}^T (\mathbf{I} - \mathbf{P}_X) \mathbf{Y} = \frac{1}{\sigma^2} \mathbf{Y}^T (\mathbf{I} - \mathbf{P}_X)^T (\mathbf{I} - \mathbf{P}_X) \mathbf{Y}$ has a central chi-square distribution with 2 degrees of freedom. This follows from Result 4.7 in the course notes because $\mathbf{A} = \frac{1}{\sigma^2} (\mathbf{I} - \mathbf{P}_X)$ is a symmetric matrix with $\text{rank}(\mathbf{A}) = 2$ and

$\Sigma \mathbf{A} = \sigma^2 \mathbf{I} \frac{1}{\sigma^2} (\mathbf{I} - \mathbf{P}_X) = (\mathbf{I} - \mathbf{P}_X)$ is an idempotent matrix.

To show that $\frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{4\sigma^2}$ is stochastically independent of $\frac{1}{\sigma^2} \text{SSE}$, note that

$\frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{4\sigma^2}$ is a function only of $(-1, -1, 1, 1)\mathbf{Y} = Y_3 + Y_4 - Y_1 - Y_2$ and $\frac{1}{\sigma^2}\text{SSE} = \frac{1}{\sigma^2}\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_X)^T(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$ is a function only of $(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$. Then, $\frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{4\sigma^2}$ is stochastically independent of $\frac{1}{\sigma^2}\text{SSE}$ if $(-1, -1, 1, 1)\mathbf{Y}$ is independent of the residual vector $(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$. Since \mathbf{Y} has a multivariate normal distribution,

it follows that $\begin{pmatrix} \mathbf{c} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} -1 & -1 & 1 & 1 \\ \mathbf{I} - \mathbf{P}_X \end{pmatrix} \begin{pmatrix} \mathbf{Y} \\ \mathbf{u} \end{pmatrix}$ has a multivariate normal distribution. Then, to show that

$(-1, -1, 1, 1)\mathbf{Y}$ is independent of $(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$, we only need to show that the quantities are uncorrelated. Here,

$\text{Cov}((-1, -1, 1, 1)\mathbf{Y}, (\mathbf{I} - \mathbf{P}_X)\mathbf{Y}) = (-1, -1, 1, 1)(\sigma^2\mathbf{I})(\mathbf{I} - \mathbf{P}_X) = \sigma^2(-1, -1, 1, 1)(\mathbf{I} - \mathbf{P}_X) = (0, 0, 0, 0)$ because $(-1, -1, 1, 1)^T$ is a column of \mathbf{X} . Consequently,

$$F = \frac{\frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{4\sigma^2}}{\frac{\text{SSE}}{2\sigma^2}} = \frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{2\text{SSE}}$$

has an F-distribution with (1,2) degrees of freedom and non-centrality parameter $4\beta_2^2/\sigma^2$ because it is the ratio of a non-central chi-square random variable, divided by its degrees of freedom, over an independent central chi-square random variable divided by its degrees of freedom. This is the definition of a non-central F random variable.

Alternatively, you could note that

$$\frac{(Y_3 + Y_4 - Y_1 - Y_2)^2}{4\sigma^2} = \frac{(\mathbf{c}^T\mathbf{Y})(\mathbf{c}^T\mathbf{Y})}{(\mathbf{c}^T\mathbf{c})\sigma^2} = \frac{\mathbf{Y}^T\mathbf{c}(\mathbf{c}^T\mathbf{c})\mathbf{c}^T\mathbf{Y}}{\sigma^2} = \mathbf{Y}^T\mathbf{A}_1\mathbf{Y},$$

and apply Result 4.7 to this quadratic form and use Result 4.8 to show that this quadratic form is independent of SSE.

(f) (4 points) Since the non-centrality parameter for the F-statistic in part (e) is $4\beta_2^2/\sigma^2$, the null hypothesis is $H_0 : \beta_2 = 0$ and the alternative is $H_a : \beta_2 \neq 0$.

(g) (10 points) Since any linear function of a normally distributed random vector has a normal distribution, it follows that $\mathbf{a}^T\mathbf{Y} \sim N(\mathbf{a}^T\mathbf{X}\beta, \sigma^2\mathbf{a}^T\mathbf{a})$. Then,

$$\frac{\mathbf{a}^T\mathbf{Y}}{\sqrt{\sigma^2\mathbf{a}^T\mathbf{a}}} \sim N\left(\frac{\mathbf{a}^T\mathbf{X}\beta}{\sqrt{\sigma^2\mathbf{a}^T\mathbf{a}}}, 1\right) \text{ and it follows that } \frac{(\mathbf{a}^T\mathbf{Y})^2}{\sigma^2\mathbf{a}^T\mathbf{a}} \sim \chi_1^2\left(\frac{\mathbf{a}^T\mathbf{X}\beta}{\sqrt{\sigma^2\mathbf{a}^T\mathbf{a}}}\right)$$

In part (e) we also showed that $\frac{1}{\sigma^2}\text{SSE} = \frac{1}{\sigma^2}\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$ has a central chi-square distribution with 2 degrees of freedom. Then, the statistic shown in this part has an F-distribution if

$\frac{(\mathbf{a}^T\mathbf{Y})^2}{\sigma^2\mathbf{a}^T\mathbf{a}}$ is independent of $\frac{1}{\sigma^2}\text{SSE} = \frac{1}{\sigma^2}\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$. Using the argument from part (e), this

is true if $\text{Cov}(\mathbf{a}^T \mathbf{Y}, (\mathbf{I} - \mathbf{P}_X) \mathbf{Y}) = \mathbf{a}^T (\sigma^2 \mathbf{I}) (\mathbf{I} - \mathbf{P}_X) = \sigma^2 \mathbf{a}^T (\mathbf{I} - \mathbf{P}_X) = (0, 0, 0, 0)$. Hence, the statistic has an F-distribution if \mathbf{a} is in the space spanned by the columns of \mathbf{X} , that is, if

$$\mathbf{a} = a \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \\ \hat{e}_4 \end{pmatrix} + b \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \\ \hat{e}_4 \end{pmatrix} = \begin{pmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \hat{e}_3 \\ \hat{e}_4 \end{pmatrix} \begin{pmatrix} a-b \\ a-b \\ b \\ a+b \end{pmatrix}$$

for some constants a and b where either $a \neq 0$ or $b \neq 0$.

2. (a) (6 points) γ_1 is the mean corn yield for variety A grown with nitrogen applied at 150lb/a. $100 \delta_1$ is the difference between mean yields for nitrogen applied at 200lb/a and 100lb/a for either variety A or variety B.
- (b) (6 points) Since the model matrix has full column rank, any linear combination of the parameters is estimable. Hence, both (i) and (ii) are estimable.
- (c) (4 points) $\mu + \alpha_1$ is estimable because $E[(Y_1 + Y_2 + Y_3)/3] = \mu + \alpha_1$.
- (d) (6 points) Delete the third row and the third column of $\mathbf{X}^T \mathbf{X}$. Invert the remaining 4x4 matrix. Fill in zeros for the third row and third column of the generalized inverse. Alternatively you could use the spectral or singular value decomposition of $\mathbf{X}^T \mathbf{X}$ to obtain a generalized inverse.
- (e) (8 points) First show that $\alpha_1 - \alpha_2$ is estimable. Then, $(0, 1, -1, 0, 0) \mathbf{b}$ is the unique best linear unbiased estimator for $(0, 1, -1, 0, 0) \boldsymbol{\beta} = \alpha_1 - \alpha_2$. Furthermore, since the random errors are independent $N(0, \sigma^2)$ random variables, $\mathbf{c}^T \mathbf{b} \sim N(\alpha_1 - \alpha_2, \sigma^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c})$, where $\mathbf{c}^T = (0, 1, -1, 0, 0)$.
- (f) (6 points) Yes, because the columns of the model matrices span the same linear space. Let $\mathbf{W} = [\mathbf{w}_1 | \mathbf{w}_2 | \mathbf{w}_3 | \mathbf{w}_4]$ denote the model matrix for Model 1, and let $\mathbf{X} = [\mathbf{x}_1 | \mathbf{x}_2 | \mathbf{x}_3 | \mathbf{x}_4 | \mathbf{x}_5]$ denote the model matrix for Model 2. Then, $\mathbf{x}_2 = \mathbf{w}_1$, $\mathbf{x}_3 = \mathbf{w}_2$, $\mathbf{x}_1 = \mathbf{w}_1 + \mathbf{w}_2$, $\mathbf{x}_4 = \mathbf{w}_3/50$, and $\mathbf{x}_5 = (3\mathbf{w}_4/2500) - 2\mathbf{w}_1 - 2\mathbf{w}_2$. Also, $\mathbf{w}_1 = \mathbf{x}_2$, $\mathbf{w}_2 = \mathbf{x}_3$, $\mathbf{w}_3 = 50\mathbf{x}_4$, and $\mathbf{w}_4 = 2500(\mathbf{x}_5 + 2\mathbf{x}_2 + 2\mathbf{x}_3)/3$.
- (g) (6 points) Since the quantity is estimable, the unique b.l.u.e. for the mean yield of variety 1 using 120 lb/a of nitrogen is $\mathbf{c}^T \mathbf{b} = (1, 0, -30, 900) \mathbf{b} \sim N(\mathbf{c}^T \mathbf{g}, \mathbf{s}^2 \mathbf{c}^T (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{c})$, where $\mathbf{b} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y}$ is a solution to the normal equations. Using the MSE from the ANOVA to estimate \mathbf{s}^2 , a 95% prediction interval is

$$(1, 0, -30, 900) \mathbf{b} \pm t_{2,0.025} \sqrt{\text{MSE}(1 + \mathbf{c}^T (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{c})}$$

where $t_{2,0.025}$ is the upper 0.025 percentile of a central t-distribution with 2 degrees of freedom.

(h) (8 points) Write the null hypothesis as $H_0 : C\gamma = \begin{bmatrix} \hat{e}_1 & -1 & 0 & 0 \\ \hat{e}_2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} \hat{e}_0 \\ \hat{e}_0 \end{bmatrix}$ and let

$\mathbf{b} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y}$ be any solution to the normal equations. The null hypothesis would be rejected if

$$F = \frac{\mathbf{b}^T \mathbf{C}^T (\mathbf{C}^T (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{C})^{-1} \mathbf{C} \mathbf{b}}{2 \text{MSE}}$$

exceeds the upper α percentile of a central F-distribution with (2,2) degrees of freedom, where α is the type I error level of the test.

The exam scores (out of 100 points) are presented as a stem-leaf display:

```

90 | 8
90 | 0 1
80 | 7 8 9
80 | 0 3
70 | 5 5 5 5 7 8
70 | 0 1 2 3 3 4
60 | 7 7 8 9
60 | 0 0 1 1 3
50 | 6 7 7 7 7 8 8 8 8 9
50 | 0 1 4 4
40 | 5 6 8 9
40 | 0 2 3
30 | 9 9
30 | 2
20 | 5 9

```

A point value should be shown for the credit awarded for each part of your exam, corresponding to the point values listed above. If your answer failed to earn any credit, a zero should be shown. If no point value is shown for some part of your exam, show your exam to the instructor. Also, check if the total points recorded on the last page of your exam is correct.