STOR 664 CWE 2020

Name:________________

ID#:________________

I pledge that I have neither given nor received unauthorized aid on this exam.

Signature:________________

(1) Let x be the predictor representing the potassium/carbon atomic ratio (K/C, in $\%$) and y be the response representing the amount of absorbed carbon monoxide (CO, in mole/mole C $\%$). The data is shown in the following Table 1.

observation number	x	y
1	0.05	0.05
2	0.05	0.10
3	0.25	0.25
4	0.25	0.35
5	0.50	0.75
6	0.50	0.85
7	0.50	0.95
8	1.25	1.42
9	1.25	1.75
10	1.25	1.82
11	1.25	1.95
12	1.25	2.45
13	2.10	3.05
14	2.10	3.19
15	2.10	3.25
16	2.10	3.43
17	2.10	3.50
18	2.10	3.93
19	2.50	3.75
20	2.50	3.93
21	2.50	3.99
22	2.50	4.07

Based on the data table, consider the regression model $y_{ij} = x_i\theta_i + \epsilon_{ij}$ with $j = 1, ..., n_i$, $i = 1, ..., p$ and $n_1 + \cdots + n_p = n$. The model can be expressed as, in a matrix form, $Y = X\theta + \epsilon$ where X is a $n \times p$ matrix whose entries are known constants, $\theta_1, ..., \theta_p$ are unknown parameters, and the components ϵ_{ij}'s of ϵ are iid random variables with mean 0 and variance σ^2 (also an unknown parameter). Let

$$\overline{y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, \quad \overline{y}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{p} n_i\,\overline{y}_{i\cdot}.$$

(1a) (6 points) Write down the form of X, and specify the numerical values of p and $n_1, ..., n_p$.

(1b) (8 points) Express $(X^tX)^{-1}$ by their entries.

(1c) (6 points) Is this model a one-way ANOVA model? Explain why.

(1d) (8 points) Find the LSE $\widehat{\theta}_i$, $i = 1, ..., p$ numerically.

(1e) (6 points) Let $\mu_i = x_i\theta_i$, $i = 1, ..., p$ and assume ϵ_{ij}'s follow the common distribution $N(0, \sigma^2)$ [**Note: This normality is assumed in Parts (1e) — (1i)**]. Consider a F-test for H_0: $\mu_1 = \cdots = \mu_p$ (reduced model 1) versus the full model H_1. Complete the following ANOVA table by filling each empty cell with correct degree of freedom and corresponding mean square.

$Source$	$Sum\ of\ Squares$	$D.F.$	$Mean\ Square$
$Full\ Model$	SSR_1		
$Reduced\ Model$	SSR_0		
$Difference$	$SSE_0 - SSE_1$		
$Residual$	SSE_1		
$Total$	$SSTO$		

Table 2: ANOVA table for problem (1e) with two nested models

(1f) (6 points) Following (1e), spell out the following sums of squares in terms of y_{ij}'s:

$$SSTO =$$

$$SSR_1 =$$

$$SSE_1 =$$

$$SSR_0 =$$

$$SSE_0 =$$

(1g) (8 points) Under the reduced model 1, find the numerical values of LSE $\widehat{\theta}_i$, $i = 1, ..., p$.

(1h) (8 point) Under the full model H_1, write down SSE_1 numerically and use it to give the value of an unbiased estimate for σ^2.

(1i) (8 points) To construct 95% simultaneous confidence intervals for θ_1 and θ_4 using Scheffé method, denoted by $M_1 \pm c_1 G$ and $M_4 \pm c_4 G$ respectively, where $G^2 = q\ F_{q,n-p;1-\alpha}$, give the numerical values of midpoints M_1, M_4, the coefficients c_1, c_4, and specify q.

(1j) (8 points) Now let us consider the case of iid random errors ϵ_{ij}'s with mean 0 and variance σ^2 but **without** assuming the normal distribution for them. If we bootstrap the residuals e_{ij}'s under the full model H_1, what are the numerical values for the *little population* in this case?

(1k) (6 points) Define $\beta_i = \frac{x_i\theta_i - x_{i-1}\theta_{i-1}}{x_i - x_{i-1}}$, $i = 2, 3, ..., p$. Under the normality assumption for ϵ_{ij}'s given in (1e), consider testing H_0^*: $\beta_2 = \cdots = \beta_p$ (reduced model 2) vs the full model H_1, and complete the following ANOVA table by filling each empty cell with correct degree of freedom and corresponding mean square.

$Source$	$Sum\ of\ Squares$	$D.F.$	$Mean\ Square$
$Full\ Model$	SSR_1		
$Reduced\ Model$	SSR_0		
$Difference$	$SSE_0 - SSE_1$		
$Residual$	SSE_1		
$Total$	$SSTO$		

Table 3: ANOVA table for problem (1k) with two nested models

(1l) (8 points) Now let us consider the case of iid random errors ϵ_{ij}'s with mean 0 and variance σ^2 but **without** assuming the normal distribution for them. If we bootstrap the residuals e_{ij}'s under the reduced model 2 H_0^*, what are the numerical values for the *little population* in this case?

(1m) (8 points) If we use bootstrap to test H_0^* in (1l) and reject H_0^* when $\frac{(SSE_0 - SSE_1)/q}{SSE_1/(n-p)} > r$. What is the value for r based on the given data? Describe the step-by-step bootstrap procedure for calculating the p-value represented by $P_{H_0^*}\left(\frac{(SSE_0 - SSE_1)/q}{SSE_1/(n-p)} > r\right)$.

(2) (6 points) Interpret the two reduced models 1 and 2 given in (1e) and (1k) respectively, i.e. what do they imply in the context for Problem (1)? Is one of them nested in another? If so, in what order?