McGill University

Faculty of Science

MATH 423/533

REGRESSION AND ANALYSIS OF VARIANCE

Practice Examination

Date: – Time: –

Examiner: DAS
Associate Examiner: –

Please write your answers in the answer booklets provided. **Please write clearly.**

This paper contains five questions. Each question carries 20 marks. Credit will be given for all questions attempted. The total mark available is 100 but rescaling of the final mark may occur.

Calculators may be used. Dictionaries and Translation dictionaries are permitted.

Notation: Throughout this examination paper, the following standard notation from the course will be used: for $i = 1, \ldots, n$, y_i is the observed response; Y_i is the random variable version of the response; $\mathbf{y}$ and $\mathbf{Y}$ are the $n \times 1$ vector versions of the responses; $\mathbf{x}_i$ is the row vector of predictor values, $\mathbf{X}$ is the matrix of predictor values; $\widehat{y}_i$, $\widehat{Y}_i$, $\widehat{\mathbf{y}}$ and $\widehat{\mathbf{Y}}$ are the fitted or predicted response values or vectors arising from a given model; β is the vector of regression coefficients; $\widehat{\beta}$ is the vector of estimates or estimators.

1. (a) Show that for a linear regression model, the parameter estimates for the regression coefficients β derived under the least squares criterion are identical to the maximum likelihood estimates for the parameters under a specific distributional assumption made about the random residual errors, ϵ.

6 MARKS

(b) Demonstrate the difference between (i) the estimate of σ^2 commonly used under least squares methodology, and (ii) the estimate of σ^2 derived using maximum likelihood.

4 MARKS

(c) Suppose that, in a study where a single continuous predictor X is recorded, a bivariate Normal distribution assumption is made about the joint distribution of X and Y. The joint density is

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

where

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

σ_{XY} is the covariance parameter, ρ is the correlation parameter. Then it follows that $X \sim \text{Normal}(\mu_X, \sigma_X^2)$, and

$$Y|X = x \sim \text{Normal}\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2)\right).$$

By considering this factorization of this joint density of X and Y, show that there is an exact correspondence between the parameters in the joint model and parameters that appear in a standard simple linear regression model.

5 MARKS

(d) Suppose that a simple linear model with intercept β_0 set equal to zero is considered. For this model, derive the least squares estimate of β_1, and the variance of the corresponding estimator under standard assumptions.

5 MARKS

2. (a) Derive the decomposition (written in the standard notation of the course)

$$\mathbf{y}^\top(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y} = \mathbf{y}^\top(\mathbf{I}_n - \mathbf{H})\mathbf{y} + \mathbf{y}^\top(\mathbf{H} - \mathbf{H}_1)\mathbf{y}.$$

explaining carefully each term in the expressions.

8 MARKS

(b) Consider the terms in the decomposition from (a) written in their random variable versions. State the distributional results related to the decomposition that are used in the construction of the 'global' Fisher-F test.

4 MARKS

(c) In a simple linear regression analysis of a sample of size $n = 36$, the following sums of squares quantities were computed

$$\mathrm{SS}_{\mathrm{R}} = 678.90 \qquad \mathrm{SS}_{\mathrm{Res}} = 112.84.$$

Reconstruct, in standard format, the ANOVA table for carrying out the F-test for these data. 6 MARKS

Report the conclusion of the test if it is known that the 0.975 quantile of the Student-t distribution with 34 degrees of freedom lies at 2.032. 2 MARKS

3. (a) If $\mathbf{e} = (e_1, \ldots, e_n)^\top$ denote the residuals from the fit of a multiple regression model based on an $(n \times p)$ matrix $\mathbf{X}$ to data $\mathbf{y} = (y_1, \ldots, y_n)^\top$ using least squares, show that

$$\mathbf{X}^\top \mathbf{e} = \mathbf{0}_p$$

where the right hand side of this equation is a column vector of zeros of length p.

6 MARKS

(b) If the model in (a) includes an intercept, show that

$$\sum_{i=1}^{n} e_i = 0.$$

2 MARKS

(c) Suppose that true regression model relating a single predictor x_1 to response y takes the form

$$\mathbb{E}_{Y_i|X_i}[Y_i|\mathbf{x}_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2$$

but that a simple linear regression model in x_1 is fitted. Derive the expectation of the residuals from the fitted model, conditional on x_1, in this case.

6 MARKS

(d) Consider again data generated from the true model in (c), but where the residual errors $\epsilon_i, i = 1, \ldots, n$ are uncorrelated but have different variances, specifically

$$\mathrm{Var}_{Y_i|X_i}[Y_i|\mathbf{x}_i] = \sigma^2(1 + \exp(x_{i1})).$$

Derive the theoretical variance of the residuals from the fitted model in this case, conditional on x_1, if the correct conditional expectation model is fitted using (ordinary) least squares.

6 MARKS

4. A *one way analysis of variance* (ANOVA) is based on a linear regression model for a single factor predictor that takes, say, K levels, which identify subgroups indexed by different levels of the factor. In this question, $K = 4$, and the subgroup means are $\mu_1, \mu_2, \mu_3, \mu_4$. The analysis aims to test the null hypothesis that the subgroups have the same expected value,

$$\mathrm{H}_0 \; : \; \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

under the standard assumptions of linear regression modelling.

(a) If the total sample size is 20, and the study is balanced, construct the $\mathbf{X}$ matrix for fitting the one way ANOVA model, and write down the form of $\mathbf{X}^\top\mathbf{X}$, if the standard *contrast parameterization* is used.

6 MARKS

(b) Write down $\mathbf{X}$ if a parameterization in terms of $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)^\top$ is used, where

$$\gamma_0 = \overline{\mu} \qquad \gamma_1 = \mu_1 - \overline{\mu} \qquad \gamma_2 = \mu_2 - \overline{\mu} \qquad \gamma_3 = \mu_3 - \overline{\mu}$$

where

$$\overline{\mu} = \frac{1}{4}\sum_{j=1}^{4} \mu_j.$$

Hint: γ is a linear transformation of μ, and of the parameters in the contrast parameterization.

6 MARKS

(c) The following edited `R` output records the analysis of a data set under the conditions of the question: `x1` is the factor predictor. Do the results indicate that the null hypothesis above should be rejected ? Justify your answer.

```
> summary(lm(y ~ -1+x1))
Coefficients:
     Estimate Std. Error t value Pr(>|t|)
x11    2.0944     0.2348   8.920 1.31e-07 ***
x12    1.9157     0.2348   8.159 4.29e-07 ***
x13    2.6391     0.2348  11.240 5.27e-09 ***
x14    1.7602     0.2348   7.497 1.27e-06 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Residual standard error: 0.525 on 16 degrees of freedom
Multiple R-squared:  0.9536,    Adjusted R-squared:  0.942
F-statistic: 82.17 on 4 and 16 DF,  p-value: 1.86e-10
```

NB: Read the output carefully.

8 MARKS

5. (a) Explain the difference between a model selection procedure based on stepwise (forward and/or backward) selection and a selection procedure based on the information criteria AIC or BIC. Explain specifically the statistical summaries used and compared.

5 MARKS

(b) The following output in `R` demonstrates the analysis of a data set where three continuous predictors are used to explain the variation in an observed response:

```
> fit1<-lm(y ~ x1+x2+x3)
> fit2<-lm(y ~ (x1+x2+x3)^2)
> drop1(fit2,test='F')
Single term deletions
Model:
y ~  (x1 + x2 + x3)^2
       Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>              103.68 50.656
x1:x2   1    12.310 115.99 51.797  2.4932 0.1292864
x1:x3   1   109.648 213.33 68.858 22.2079 0.0001185 ***
x2:x3   1     1.897 105.58 49.164  0.3842 0.5420344
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
> fit3<-update(fit2, ~ .-x2:x3)
> drop1(fit3,test='F')
Single term deletions
Model:
y ~  x1 + x2 + x3 + x1:x2 + x1:x3
       Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>              105.58 49.164
x1:x2   1     11.23 116.81 49.994  2.3401    0.1403
x1:x3   1    120.03 225.61 68.425 25.0109 5.255e-05 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
> anova(fit3)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x1         1  35.330  35.330  7.3618   0.01269 *
x2         1   1.055   1.055  0.2198   0.64378
x3         1   0.641   0.641  0.1336   0.71820
x1:x2      1   0.777   0.777  0.1620   0.69124
x1:x3      1 120.030 120.030 25.0109 5.255e-05 ***
Residuals 22 105.581   4.799
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

```
> fit4<-lm(y ~ x1+x2+x3+x1:x3)
> drop1(fit4,test='F')
Single term deletions

Model:
y ~ x1 + x2 + x3 + x1:x3
       Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>              116.81 49.994
x2      1     3.007 119.82 48.706  0.5921 0.4494427
x1:x3   1   109.577 226.39 66.521 21.5756 0.0001128 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
> fit5<-lm(y ~ x1+x3+x1:x3)
> drop1(fit5,test='F')
Single term deletions

Model:
y ~ x1 + x3 + x1:x3
       Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>              119.82 48.706
x1:x3   1    106.71 226.53 64.538  21.374 0.0001083 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

In standard notation, list the five models fitted, and summarize the conclusions to be drawn from the analysis.

10 MARKS

(c) The formula for the BIC for a model with p regression parameters β as computed by R is

$$\text{BIC}(\beta) = n \log\left(\frac{\text{SS}_{\text{Res}}(\beta)}{n}\right) + (p+1)\log(n) + \text{constant}$$

where the constant depends on the sample size, and is identical for all models fitted.

Compute the BIC for the models fitted in the R output above, ignoring the constant term. Explain which model is selected using BIC here.

5 MARKS