# STATISTICS 3A03

Fall 2017 Dr. Angelo Canty

TERM TEST 2 November 28, 2017.

DAY CLASS
DURATION OF EXAMINATION: 50 Minutes

THIS EXAMINATION PAPER INCLUDES 6 PAGES AND 3 QUESTIONS. YOU ARE RESPONSIBLE FOR ENSURING THAT YOUR COPY OF THE PAPER IS COMPLETE. BRING ANY DISCREPANCY TO THE ATTENTION OF YOUR INVIGILATOR.

Instuctions:

1. Use of the Casio FX-991 calculator only is allowed.
2. Each question is worth 20 marks. Marks for parts of questions are given in the question.
3. Show **all of your work** for full marks!
4. Some useful results and formulae are given on Page 4.
5. A Table of critical values for the Student's $t$ distribution is given on Page 5.
6. A Table of 5% critical values for the $F$ distribution is given on Page 6.

---

**Q. 1** **a)** What are the three main assumptions of the linear model. For each state how you would assess violations of the assumption. **[9 marks]**

**b)** In the context of multiple regression explain what is meant by

**(i)** an outlier;

**(ii)** a highly influential point.

and describe how you would spot such points in a regression. **[6 marks]**

**c)** Suppose that we know that Suppose that a response variable $Y$ is related to a covariate $X$ with the following relationship.

$$\mathrm{E}[Y \mid X] = \beta_1 X \qquad \mathrm{Var}(Y \mid X) = \sigma^2 X$$

Show that taking the square roots of both variables results, approximately, in a linear model with constant variance. **[5 marks]**

**Q. 2** An experiment was done to examine the effect of diet on blood coagulation time. 24 animals were randomly assigned one of 4 diets (A, B, C and D) and their blood coagulation time was measured. Dummy variables were constructed for each of the diets and the linear model

$$Y = \beta_0 + \beta_1 \mathrm{DietB} + \beta_2 \mathrm{DietC} + \beta_3 \mathrm{DietD} + \varepsilon$$

was fitted in SAS producing the following output.

**Continued on Page 2**

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 3 | 228.00000 | 76.00000 | 13.57 | <.0001 |
| Error | 20 | 112.00000 | 5.60000 | | |
| Corrected Total | 23 | 340.00000 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 2.36643 | R-Square | 0.6706 |
| Dependent Mean | 64.00000 | Adj R-Sq | 0.6212 |
| Coeff Var | 3.69755 | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | 1 | 61.00000 | 1.18322 | 51.55 | <.0001 |
| DietB | 1 | 5.00000 | 1.52753 | 3.27 | 0.0038 |
| DietC | 1 | 7.00000 | 1.52753 | 4.58 | 0.0002 |
| DietD | 1 | -1.3323E-15 | 1.44914 | -0.00 | 1.0000 |

**a)** What is the interpretation of the parameters $\beta_0$, $\beta_1$ and $\sigma^2$ and give their estimates from the fitted model. **[9 marks]**

**b)** State the null hypotheses for the $F$-test in the Analysis of Variance table, its interpretation and your conclusion from that test. **[5 marks]**

**c)** The following SAS code was used to fit another model to this data and the Analysis of Variance table from that model is given below. Explain the model used here and why it is a reduced model relative to the model above. Test if this model is sufficient to explain the data. What is your conclusion from the test?

```
Data Blood;
  Set Blood;
  DietBC=DietB+DietC;
run;
PROC REG Data=Blood;
  Model coag=DietBC;
run;
```

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 1 | 216.00000 | 216.00000 | 38.32 | <.0001 |
| Error | 22 | 124.00000 | 5.63636 | | |
| Corrected Total | 23 | 340.00000 | | | |

**[6 marks]**

**Continued on Page 3**

**Q. 3** The price of a car is expected to depend on the horsepower of the car. A dataset was collected to see if that relationship differed across countries of origin. In this dataset the variable Country had 4 categories: Germany, Japan, USA and Other. Here is one model fitted to this dataset.

```
PROC GLM Data=S3A3.cars plots=none;
   Class Country(ref='Other');
   Model Price=Horsepower Country Horsepower*Country /solution SS3;
run;
```

**The GLM Procedure**

**Dependent Variable: Price**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 7 | 4889.298973 | 698.471282 | 43.39 | <.0001 |
| Error | 82 | 1319.845965 | 16.095683 | | |
| Corrected Total | 89 | 6209.144939 | | | |

| R-Square | Coeff Var | Root MSE | Price Mean |
|---|---|---|---|
| 0.787435 | 24.43431 | 4.011942 | 16.41930 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Horsepower | 1 | 1718.483598 | 1718.483598 | 106.77 | <.0001 |
| Country | 3 | 38.590633 | 12.863544 | 0.80 | 0.4978 |
| Horsepower*Country | 3 | 70.461585 | 23.487195 | 1.46 | 0.2317 |

| Parameter | Estimate | | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | -10.88176083 | B | 4.21576463 | -2.58 | 0.0116 |
| Horsepower | 0.23732073 | B | 0.03819472 | 6.21 | <.0001 |
| Country Germany | 11.77444778 | B | 9.23469248 | 1.28 | 0.2059 |
| Country Japan | 4.75521321 | B | 4.68515488 | 1.01 | 0.3131 |
| Country USA | 2.07573611 | B | 4.91644477 | 0.42 | 0.6740 |
| Country Other | 0.00000000 | B | . | . | . |
| Horsepower*Country Germany | -0.09507366 | B | 0.06639974 | -1.43 | 0.1560 |
| Horsepower*Country Japan | -0.07656473 | B | 0.04064078 | -1.88 | 0.0631 |
| Horsepower*Country USA | -0.05218890 | B | 0.04226560 | -1.23 | 0.2204 |
| Horsepower*Country Other | 0.00000000 | B | . | . | . |

**Note:** The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

**a)** Give the fitted linear relationships between Price and Horsepower for each of the country categories in this model. **[8 marks]**

**b)** Test the hypothesis that the four parallel lines is sufficient for this dataset. Ensure you give the null and alternative hypotheses, the value of the test statistic, its distribution if the null hypothesis is true and the results of your test. **[6 marks]**

**c)** Explain carefully how you would test the null hypothesis that the relationship between price and horsepower is identical for different countries of origin. Ensure you give the distribution of the test statistic when the null hypothesis is true. **[6 marks]**

**Continued on Page 4**