STATISTICS 3A03

Fall 2017 Dr. Angelo Canty

TERM TEST 2 November 28, 2017.

IMPORTANT INFORMATION

Time and Location The test will start at 10:30 **SHARP** on Tuesday, November 28, 2017 in the **Canadian Martyrs Testing Center**. No extra time will be allowed for students who arrive late. Any student with an accommodation requiring extra time must arrange for an alternative test through the Student Accessibility Service well in advance of the test. Any such test **MUST** overlap with regular test unless I give prior approval to a different time.

Please note that the Canadian Martyrs Testing Centre **CANNOT** be accessed through the school. Access is from the west end of the school building as shown in this picture. Also note there is no parking allowed on Canadian Martyrs School property.

Test Duration You will be given 50 minutes to complete the test.

Format There will be three questions each consisting of 3 parts. Each question will be worth 20 marks and the marks for the parts of the questions will be stated on the test paper.

Absences There is no makeup test for this course so any absences must be reported through the usual McMaster Student Absence Form mechanism or your Dean of Students. For more information see the course outline and `https://www.mcmaster.ca/msaf/index.html`.

Coverage The test will cover the assumptions of the linear model and diagnostics, linear models with categorical variables including interactions between categorical variables and between categorical and continuous covariates and transformations. This corresponds to the fourth to sixth set of my lecture notes and Chapters 4–6 of the textbook. The earlier chapters will not be directly examined but the material may be needed as the examined chapters build on the earlier ones.

What to Bring The only things you may take to your desk are writing implements, the McMaster Standard calculator (Casio FX-991) and your student ID.

Tables and Formulae The following three pages will be attached to your test paper for use in the test. No other tables or formulae sheets are allowed. in room **T-29 101**. No extra time will be allowed for students who arrive late. Any student with an accommodation requiring extra time must arrange for an alternative test through the Student Accessibility Service well in advance of the test. Any such test **MUST** overlap with regular test unless I give prior approval to a different time.

Some Useful Results and Formulae

- $S_{xx} \; = \; \sum(x_i - \overline{x})^2 \; = \; \sum x_i^2 - n\overline{x}^2. \qquad S_{xy} \; = \; \sum(x_i - \overline{x})(y_i - \overline{y}) \; = \; \sum x_i y_i - n\overline{x}\,\overline{y}.$
- The population and sample correlation coefficients are

$$\rho \; = \; \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \qquad r \; = \; \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

- The least squares estimates for the simple linear model $Y = \beta_0 + \beta_1 x + \varepsilon$ are

$$\hat{\beta}_0 \; = \; \overline{y} - \hat{\beta}_1\overline{x} \qquad \hat{\beta}_1 \; = \; \frac{S_{xy}}{S_{xx}}$$

- An unbiased estimator of the error variance σ^2 in a simple linear model is

$$\hat{\sigma}^2 \; = \; \frac{\mathrm{SSE}}{n-2} \; = \; \frac{\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$$

- The standard error of the estimators in a simple linear model are

$$\text{s.e.}\left(\hat{\beta}_0\right) \; = \; \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}}\;\hat{\sigma} \qquad \text{s.e.}\left(\hat{\beta}_1\right) \; = \; \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

- For multiple regression with design matrix $\boldsymbol{X}$ the least squares estimators are $\hat{\boldsymbol{\beta}} \; = \; \left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{Y}$.
- The variance-covariance matrix of the estimators in a multiple regression is $\mathrm{Var}(\boldsymbol{\beta}) \; = \; \left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1}\sigma^2$
- An unbiased estimator of σ^2 is $\hat{\sigma}^2 \; = \; \mathrm{SSE}/(n-p-1)$.
- For any coefficient in a linear model with $p \geqslant 1$ covariates

$$\frac{\hat{\beta}_j - \beta_j}{\text{s.e.}(\hat{\beta}_j)} \; \sim \; t_{n-p-1} \qquad j = 0, 1, \ldots p$$

- The coefficient of determination is $R^2 \; = \; 1 - \mathrm{SSE}/S_{yy}$.
- To test if a reduced model is sufficient compared to a full model we use the test statistic

$$F \; = \; \frac{(\mathrm{SSE}_{\mathrm{red}} - \mathrm{SSE}_{\mathrm{full}})/(df_{\mathrm{red}} - df_{\mathrm{full}})}{\mathrm{SSE}_{\mathrm{full}}/df_{\mathrm{full}}}$$

  where the degrees of freedom are the error degrees of freedom in the models. This statistic has an F distribution with $df_{\mathrm{red}} - df_{\mathrm{full}}$ and df_{full} degrees of freedom if the reduced model is sufficient.
- The Delta method specifies that

$$\mathrm{E}\big(g(Y)\big) \approx g\big(E(Y)\big) \qquad \mathrm{Var}\big(g(Y)\big) \approx \big(g'\big(\mathrm{E}(Y)\big)\big)^2\,\mathrm{Var}(Y)$$

  where g' is the derivative of the function g.