STATISTICS 3A03

Fall 2015 Dr. Angelo Canty

TERM TEST 1 October 24, 2017.

IMPORTANT INFORMATION

Time and Location The test will start at 10:30am **SHARP** on Tuesday, October 24, 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 simple linear regression and multiple regression. That is the first three sets of my lecture notes, Chapters 1–3 of the textbook and the first 2 assignments.

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.

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

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

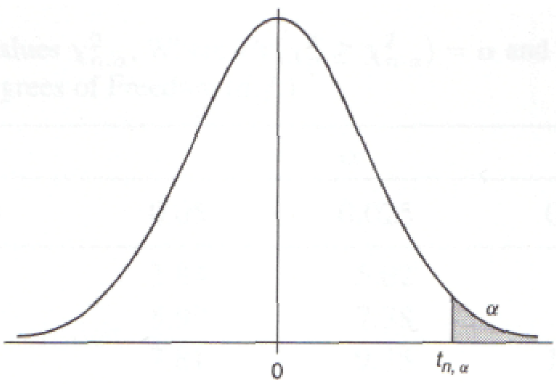
- For multiple regression with design matrix $\boldsymbol{X}$ the least squares estimators are $\hat{\boldsymbol{\beta}} \;=\; \left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{Y}$.
- The variance-covariance matrix of the estimators in a multiple regression is $\mathrm{Var}(\boldsymbol{\beta}) \;=\; \left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1}\sigma^2$
- 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.

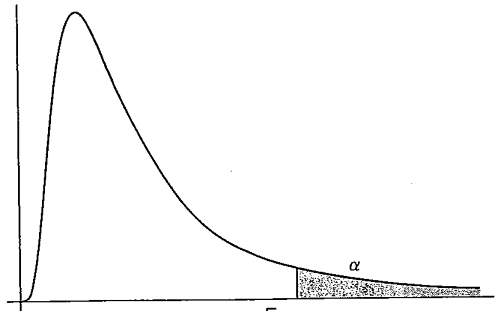


The probability density function of the Student's t-distribution with n degrees of freedom ($d.f.$).

Table A.2 Critical Values $t_{n,\alpha}$, Where $Pr(T_n \geq t_{n,\alpha}) = \alpha$ and T_n Is the Student's t-Distribution With n Degrees of Freedom ($d.f.$)

n ($d.f.$)	α: 0.10	0.05	0.025	0.010	0.005
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.97	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.42	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
12	1.36	1.78	2.18	2.68	3.06
14	1.34	1.76	2.14	2.62	2.98
16	1.34	1.75	2.12	2.58	2.92
18	1.33	1.73	2.10	2.55	2.88
20	1.32	1.72	2.09	2.53	2.84
30	1.31	1.70	2.04	2.46	2.75
40	1.30	1.68	2.02	2.42	2.70
60	1.30	1.67	2.00	2.39	2.66
120	1.29	1.66	1.98	2.36	2.62
∞	1.28	1.64	1.96	2.33	2.58

Source: Adapted from Table III of Fisher and Yates (1963), *Statistical Tables for Biological, Agricultural and Medical Research*, 6th Ed., published by Oliver and Boyd, Edinburgh, with kind permission of the authors and publishers.



The probability density function of the F-distribution with n_1 (numerator) and n_2 (denomenator) degrees of freedom.

Table A.4 The 5% Critical Values $f_{n_1,n_2;0.05}$, Where $\Pr(F_{n_1,n_2} \geq f_{n_1,n_2;0.05}) = 0.05$ and F_{n_1,n_2} is the F-Distribution with n_1 (numerator) and n_2 (denominator) (df)

	n_1								
n_2	1	2	4	6	8	10	12	24	∞
1	161.4	199.5	224.6	234.0	238.9	241.9	243.9	249.1	254.30
2	18.51	19.00	19.25	19.33	19.37	19.40	19.41	19.45	19.50
3	10.13	9.55	9.12	8.94	8.85	8.79	8.74	8.64	8.53
4	7.71	6.94	6.39	6.16	6.04	5.96	5.91	5.77	5.63
5	6.61	5.79	5.19	4.95	4.82	4.74	4.68	4.53	4.36
6	5.99	5.14	4.53	4.28	4.15	4.06	4.00	3.84	3.67
7	5.59	4.74	4.12	3.87	3.73	3.64	3.57	3.41	3.23
8	5.32	4.46	3.84	3.58	3.44	3.35	3.28	3.12	2.93
9	5.12	4.26	3.63	3.37	3.23	3.14	3.07	2.90	2.71
10	4.96	4.10	3.48	3.22	3.07	2.98	2.91	2.74	2.54
11	4.84	3.98	3.36	3.09	2.95	2.85	2.79	2.61	2.40
12	4.75	3.89	3.26	3.00	2.85	2.75	2.69	2.51	2.30
13	4.67	3.81	3.18	2.92	2.77	2.67	2.60	2.42	2.21
14	4.60	3.74	3.11	2.85	2.70	2.60	2.53	2.35	2.13
15	4.54	3.68	3.06	2.79	2.64	2.54	2.48	2.29	2.07
20	4.35	3.49	2.87	2.60	2.45	2.35	2.28	2.08	1.84
25	4.24	3.39	2.76	2.49	2.34	2.24	2.16	1.96	1.71
30	4.17	3.32	2.69	2.42	2.27	2.16	2.09	1.89	1.62
40	4.08	3.23	2.61	2.34	2.18	2.08	2.00	1.79	1.51
60	4.00	3.15	2.53	2.25	2.10	1.99	1.92	1.70	1.39
120	3.92	3.07	2.45	2.17	2.02	1.91	1.83	1.61	1.25
∞	3.84	3.00	2.37	2.10	1.94	1.83	1.75	1.52	1.00

Source: Abridged from Table 18 of Pearson and Hartley (1954), *Biometrika Tables for Statisticians, Volume I*, published by the Cambridge University Press for the *Biometrika* Trustees, with kind permission of the authors and publishers.