

**ECO220Y1Y, Term Test #5, Prof. Murdock**

**April 5, 2018, 9:10 – 11:00 am**

**U of T E-MAIL:** \_\_\_\_\_ **@MAIL.UTORONTO.CA**

**SURNAME  
(LAST NAME):** \_\_\_\_\_

**GIVEN NAME  
(FIRST NAME):** \_\_\_\_\_

**UTORID:**  
(e.g. LIHAO118)

## **Instructions:**

- You have 110 minutes. Keep these test papers and the *Supplement* closed and face up on your desk until the start of the test is announced. You must stay for a minimum of 60 minutes.
  - You may use a non-programmable calculator.
  - There are 4 questions (all with multiple parts) with varying point values worth a total of 100 points.
  - This test includes these 8 pages plus the *Supplement*. The *Supplement* contains graphs, tables, and other materials required for the test questions and the aid sheets (formulas and Normal, *t* and *F* statistical tables). For each question referencing the *Supplement*, carefully review *all* materials, noticing continuations onto the next page.
    - The *Supplement* will *not* be graded. We will only collect these test papers, not the *Supplement*.
  - Write your answers clearly, completely and concisely in the designated space provided immediately after each question. An answer guide ends each question to let you know what is expected. For example, a quantitative analysis (which shows your work), a fully-labelled graph, and/or sentences.
    - Anything requested by the question and/or the answer guide is required.
    - Similarly, limit yourself to the answer guide. For example, if the answer guide does not request sentences, provide only what is requested (e.g. quantitative analysis). Leave yourself time to complete all questions rather than overdoing some questions and running out of time.
    - For questions with multiple parts (e.g. (a) – (d)), ***attempt each part***.
  - ***Your entire answer must fit in the designated space provided immediately after each question.*** No extra space/pages are possible. You *cannot* use blank space for other questions nor can you write answers on the *Supplement*. ***Write in PENCIL and use an ERASER as needed.*** This way you can make sure to fit your final answer (including work and reasoning) in the appropriate space. Most questions give more blank space than is needed to answer. ***Follow the answer guides and avoid excessively long answers.***

**(1)** See **Supplement for Question (1): Analyzing ON Public Sector Salary Disclosure Data for University of Waterloo.**

**(a)** [6 pts] For the 1,067 observations, what is the mean salary of full professors? *Explain.* (Note: You must decide which regression – **(1)** through **(6)** – is relevant to answer.) Answer with 1 – 2 sentences.

**(b)** [6 pts] Why is the coefficient on *Male* different in Regression **(3)** versus **(1)**? Explain. Answer with 2 – 3 sentences.

**(c)** [8 pts] For Regression **(6)**, compute the **99.9%** confidence interval estimate for the coefficient on *Male*. Interpret it. Answer with a quantitative analysis & 1 sentence that would be clear to someone who has read the supplement.

**(d)** [3 pts] If salary were measured in dollars, rather than \$1,000, what would the coefficient on *Male* be in Regression **(1)**? In Regression **(4)**? Answer with two numeric values.

**(e)** [8 pts] Comparing Regressions **(1)** and **(2)**, what does the difference in the  $R^2$  mean? Answer with 2 – 3 sentences.

**(f)** [5 pts] Consider Regression **(3)** in Table 1 and the diagnostic plot. How many Clinical Lecturers are female? Explain.  
Answer with a quantitative analysis & 1 – 2 sentences.

**(g)** [8 pts] In Regression **(7)**, find **1198.11625**. What does it measure? What is the meaning of its square root (giving its value and units)? Would it be bigger, smaller or the same for Regression **(3)**? Why? Answer with 2 – 3 sentences.

**(h)** [3 pts] What is the salary of the president of the University of Waterloo? Answer with a quantitative analysis.

**(i)** [5 pts] In Regression **(8)** what is the statistical test that the value **0.004** refers to? What is the practical meaning of that test? Answer with hypotheses using formal notation & 1 sentence.

**(2)** See **Supplement for Question (2): Mortality Inequality for Males Aged 0 Years: Currie and Schwandt (2016).**

**(a)** [8 pts] Which overall conclusions does the figure suggest for male infants? (Give interpretations that would be clear to readers of *The Economist*, who are *not* familiar with regression and who are not sure which conclusions to draw from the figure.) Answer with 2 – 3 sentences that would be clear to someone who has read the supplement.

**(b)** [4 pts] Write down a multiple regression *model* for the figure. Name your variables in a way that makes their meanings clear. (*Hint:* The correct model has  $k = 3$ .) Answer with a multiple regression model using formal notation.

**(c)** [6 pts] Using the figure and reasonable approximation, what are the multiple regression coefficient estimates? (*Hint:* Start with the line for each year by itself.) Answer with the estimated regression equation & your work.

**(3)** See **Supplement for Question (3): Currie and Schwandt (2016), Again (Males Aged 20-24 & 65+ Years).**

**(a)** [6 pts] Is the regression for males aged 20 to 24 years statistically significant? If so, at which significance levels?  
Answer with hypotheses using formal notation, a quantitative analysis & 1 sentence.

**(b)** [6 pts] Comparing the regressions for males aged 20 to 24 years with males aged 65+ years, which has a larger value of the  $s_e$ ? *Explain.* Answer with 1 – 2 sentences.

**(c)** [6 pts] Consider dividing males aged 65+ years into two groups: county groups ranked below 80 and county groups ranked 80 or higher. For 3-year mortality (per 1,000), the mean and standard deviation for Group 1 ( $n = 15$ ) are 121.25 and 6.42, respectively, and for Group 2 ( $n = 5$ ) are 140.87 and 4.64, respectively. In a simple regression with  $n = 20$  where the y-variable is 3-year mortality (per 1,000) and the x-variable is a dummy for Group 2, what is the standard error of the coefficient on the dummy? (*Hint:* Recall the homoscedasticity assumption of regression and the link between differences in means and a simple regression with a dummy.) Answer with a quantitative analysis.

**(4)** See **Supplement for Question (4): Waterloo's report "Salary Anomaly Working Group: Analysis & Findings."**

**(a)** [6 pts] What is the predicted annual salary for: a faculty member with the rank of Lecturer in 2015, a mean annual merit score of 1.44, no previous Outstanding Performance Awards, hired in 2013 as a Lecturer in the Faculty of Environment, and who earned a Master's degree in 2013? *Note:* This asks for a point prediction (not an interval) and you must decide which specification – **(1)** or **(2)** – is relevant to answer. Answer with a quantitative analysis.

**(b)** [6 pts] Consider the relationship between annual salary and years since hire (until 2015) in Specification **(1)**. For values to plug in for the other x variables, consider a person with a typical annual merit score of 1.6, no previous performance awards, no lag between earning highest degree and hire, with a PhD, in the current rank of Professor, in Applied Health Sciences, and hired at the rank of Assistant Professor. Which of the three graphs – Graph (A), Graph (B), or Graph (C) – is the correct one? Explain why the other two are definitely not correct. Answer 2 – 3 sentences.

This *Supplement* contains graphs, tables, and other materials required for the test questions and the aid sheets (formulas and Normal, *t* and *F* statistical tables). For each question referencing this *Supplement*, carefully review *all* materials, noticing continuations onto the next page.

**Supplement for Question (1):** Recall the 2017 Ontario disclosure of 2016 salaries of public sector employees making at least \$100K. Consider those employed at the University of Waterloo with any of these five job titles: Assistant Professor, Associate Professor, Professor (aka Full Professor), Lecturer, or Clinical Lecturer. *Male* is a dummy variable that equals 1 for males and 0 for females. Dummy variables also measure the five distinct faculty job titles at Waterloo, where the rank of Full Professor is the reference (omitted) category. See Table 1 below and the continued discussion right after it.

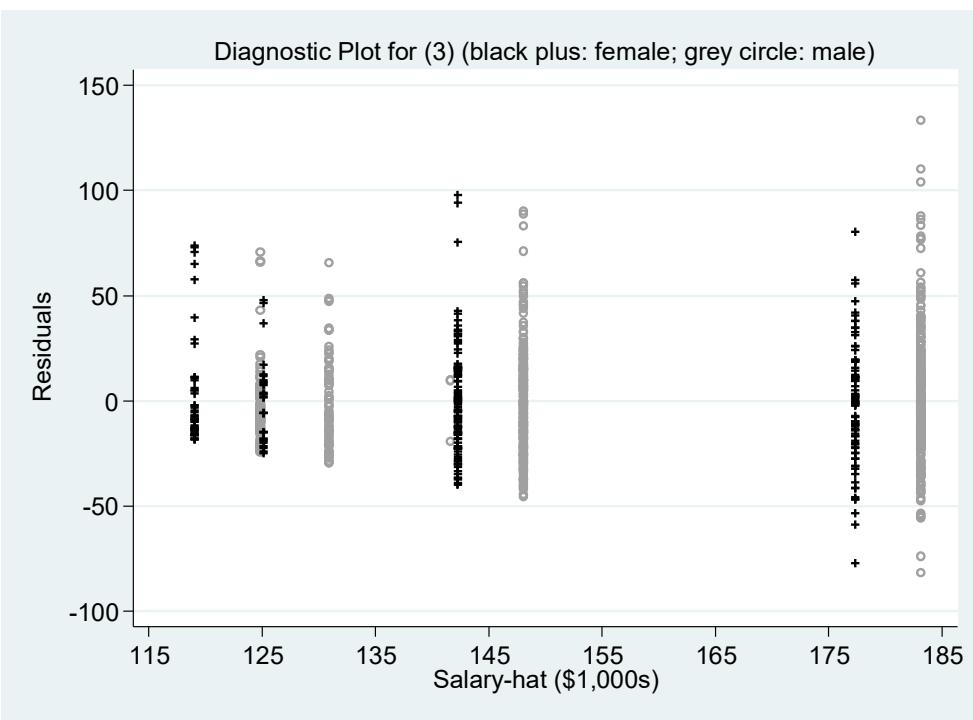
**Table 1.** Regressions to Explain Variation in Faculty Members' Salaries at the University of Waterloo

x-variables:	y-variable: <i>Salary (\$1,000s)</i>			y-variable: <i>Ln(Salary)</i>		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Male</i>	12.3352 (2.3155)	-	5.8317 (1.7493)	0.0822 (0.0146)	-	0.0403 (0.0108)
<i>Assistant Professor</i>	-	-59.2306 (2.4880)	-58.3145 (2.4915)	-	-0.3947 (0.0153)	-0.3883 (0.0153)
<i>Associate Professor</i>	-	-35.7378 (1.7439)	-35.0904 (1.7465)	-	-0.2200 (0.0108)	-0.2155 (0.0108)
<i>Full Professor</i>	-	-	-	-	-	-
<i>Lecturer</i>	-	-52.7632 (2.7732)	-52.2335 (2.7647)	-	-0.3434 (0.0171)	-0.3397 (0.0170)
<i>Clinical Lecturer</i>	-	-40.4039 (14.5397)	-41.5487 (14.4750)	-	-0.2446 (0.0897)	-0.2525 (0.0891)
<i>Constant</i>	147.1844 (1.984856)	182.0359 (1.2061)	177.3490 (1.8486)	4.9682 (0.0125)	5.1928 (0.0074)	5.1604 (0.0114)
<i>R</i> <sup>2</sup>	0.0260	0.4513	0.4570	0.0289	0.4774	0.4842
Observations	1,067	1,067	1,067	1,067	1,067	1,067

Notes: Reports six regressions. Standard errors in parentheses.

See the diagnostic plot for Regression (3) to the right.

Table 1 includes only regular faculty members without additional titles such as Dean. Regressions (7) and (8) (next page), include the president of Waterloo, Feridun Hamdullahpur, who is also a full professor of mechanical engineering. He is male and has the highest reported salary at Waterloo. Regression (7) shows STATA output for a regression like Regression (3) in Table 1 but including this outlier. Regression (8) adds a dummy variable (appropriately named president) for this outlier.



**Supplement for Question (1), cont'd:** Note: Some specific numbers are in boldface below for easy reference.

**Regression (7):**

Source	SS	df	MS	Number of obs	=	1068
Model	569538.696	5	113907.739	F( 5, 1062)	=	170.66
Residual	708851.34	1062	667.468305	Prob > F	=	0.0000
Total	1278390.04	1067	<b>1198.11625</b>	R-squared	=	0.4455
				Adj R-squared	=	0.4429
				Root MSE	=	25.835

salary	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
male	6.039971	1.80917	3.34	0.001	2.490016 9.589925
assis_prof	-58.78131	2.576379	-22.82	0.000	-63.83668 -53.72594
assoc_prof	-35.5668	1.805515	-19.70	0.000	-39.10958 -32.02402
lecturer	-52.71411	2.858988	-18.44	0.000	-58.32402 -47.10421
clin_lect	-42.08904	14.97174	-2.81	0.005	-71.46659 -12.71149
_cons	177.6811	1.911676	92.95	0.000	173.93 181.4322

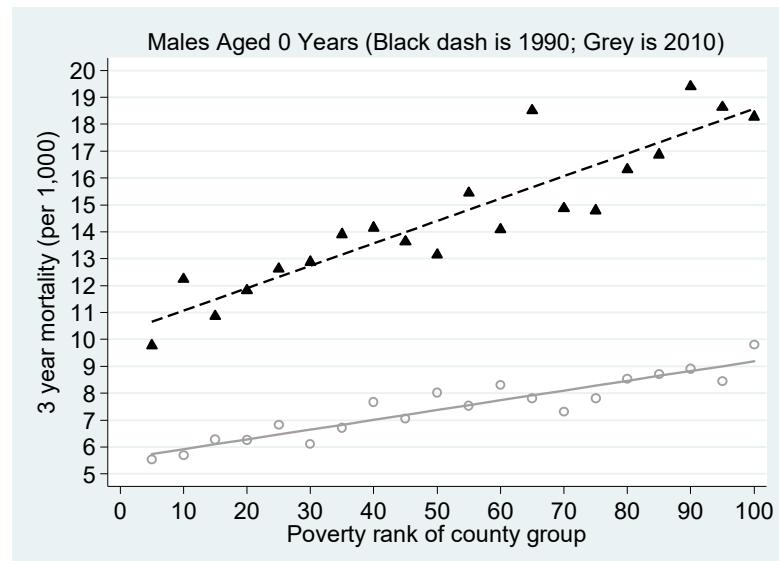
**Regression (8):**

Source	SS	df	MS	Number of obs	=	1068
Model	616432.124	6	102738.687	F( 6, 1061)	=	164.67
Residual	661957.913	1061	623.900012	Prob > F	=	0.0000
Total	1278390.04	1067	1198.11625	R-squared	=	0.4822
				Adj R-squared	=	0.4793
				Root MSE	=	24.978

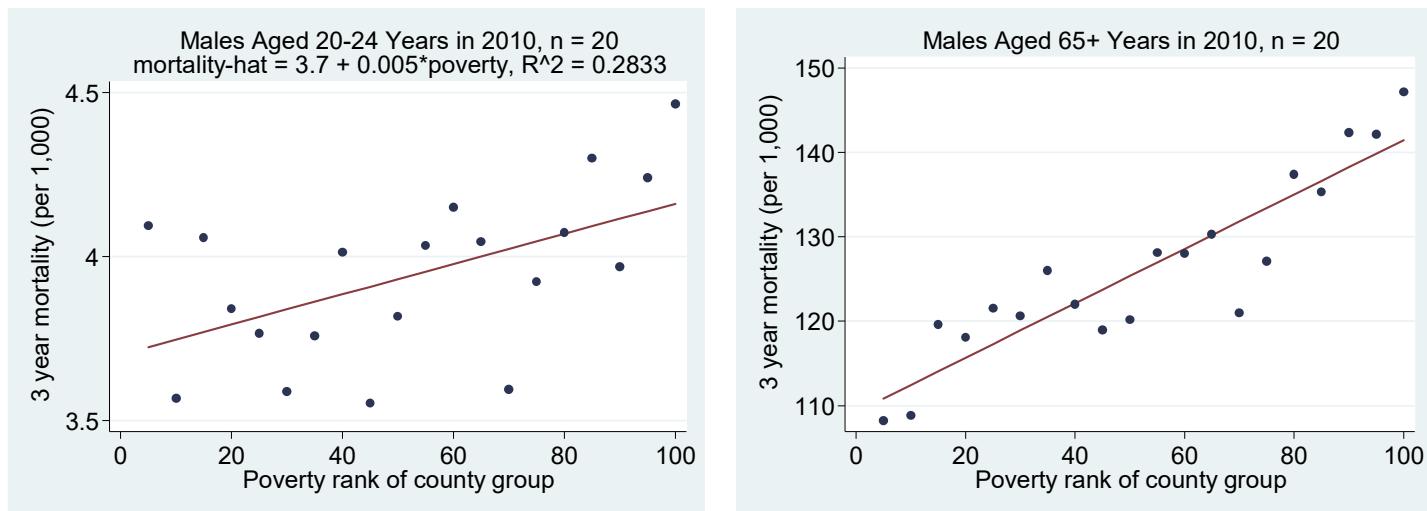
  

salary	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
male	5.831735	1.749293	3.33	0.001	2.399268 9.264202
assis_prof	-58.31453	2.491457	-23.41	0.000	-63.20327 -53.42579
assoc_prof	-35.09043	1.746458	-20.09	0.000	-38.51733 -31.66352
lecturer	-52.23354	2.76466	-18.89	0.000	-57.65836 -46.80871
clin_lect	-41.54867	14.47499	-2.87	<b>0.004</b>	-69.95154 -13.1458
president	216.8193	25.00917	8.67	0.000	167.7462 265.8923
_cons	177.349	1.848629	95.94	0.000	173.7216 180.9764

**Supplement for Question (2):** Recall Currie and Schwandt (2016) “Mortality Inequality: The Good News from a County-Level Approach.” The figure to the right shows males aged 0 years (infants) in the U.S. The y-axis is the three-year mortality rate per 1,000 population (which in 1990 is “the ratio of all deaths that occurred in a cohort [sex and age group] between April 1, 1990, and March 31, 1993, divided by the 1990 Census population count [for that sex and age group]” p. 37). The x-axis is the poverty rate rank of each county group, where counties are grouped into 20 bins (each is five percent of the U.S. population). The county groups ranked highest by poverty rate are the poorest.



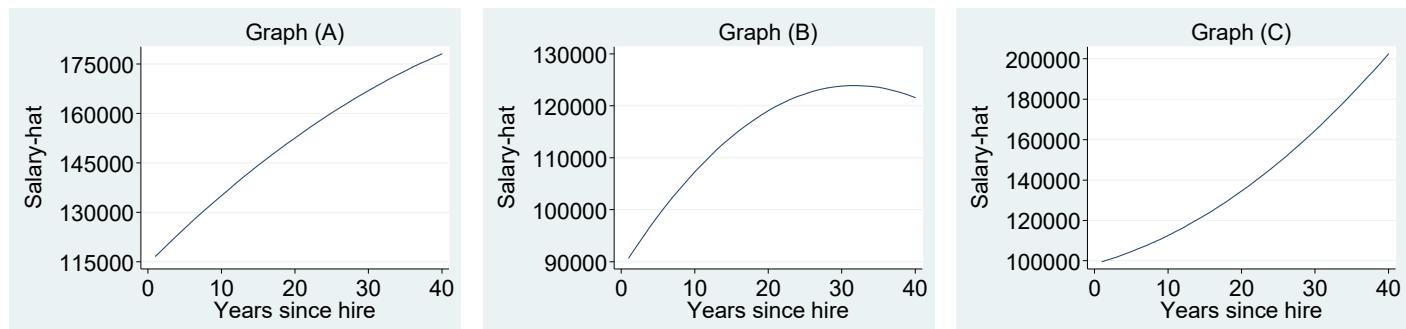
**Supplement for Question (3):** Again, recall Currie and Schwandt (2016), noting that the material already covered in the *Supplement for Question (2)* is not repeated here. The first figure below is for males aged 20 to 24 years in 2010. (Notice the estimates given in the title of the first figure.) The second figure below is for males aged 65+ years in 2010.



**Supplement for Question (4):** Recall the May 26, 2016 “Salary Anomaly Working Group: Analysis & Findings” report. It analyzes 1,171 Waterloo faculty members who completed an annual performance review on May 1, 2015, including 344 females and 827 males. The next page shows the regression results: Specifications **(1)** and **(2)**. Some variables are described below, with others described with the regression results. Categorical information is included with dummy variables: names clearly indicate when the value is 1. The omitted category for each set of dummies is specified below.

- *Annual salary*: Annual base salary as of May 1, 2015 excluding stipends (CAN dollars)
- *Mean annual merit score*: Average annual merit score for all available years for the faculty member from 2009 through 2014. Each year faculty members receive a merit score on a two-point scale (2.0 is the highest score) that scores overall productivity in research, teaching, and service. A typical score is 1.6.
- *LAG*: Number of years since earning degree and hire at Waterloo (i.e. year hired minus year earned degree)
- *Highest degree earned*: The omitted category is *Bachelor Degree (BA)*.
- *Current academic rank*: The omitted category is *Assistant Professor*.
- *UNIT: Academic unit of faculty member*: The omitted category is *Applied Health Sciences*.
- *RANK@HIRE: Academic rank when hired at Waterloo*: The omitted category is *Assistant Professor*.
- *PROFS\_RANKS*: Is a dummy variable equal to 1 for those with a rank of Assistant Professor, Associate Professor or Professor. The omitted category is Lecturers and Clinical Lecturers.

Also, one part of Question (4) asks you identify which of the graphs below are correct. (Note: Two are incorrect.)



**Supplement for Question (4) continues on next page >>>**

Multiple Regression Estimation Results for the Waterloo Case Study (Dependent variable: Annual salary)

	<i>Specification (1)</i>	<i>Specification (2)</i>		
	Estimate	Std. Error	Estimate	Std. Error
<i>Male</i>	-	-	2904.59	700.82
<i>Mean annual merit score, 2009 through 2014</i>	26097.98	1791.94	25821.67	1699.22
<i>Number of previous Outstanding Performance Awards</i>	3664.14	685.64	3587.37	651.23
<i>Years since hire until 2015 (i.e. 2015 minus the year hired)</i>	2226.66	140.71	2252.47	133.39
<i>Years since hire until 2015 squared</i>	-15.84	3.47	-16.60	3.29
<i>LAG: Lag of years between earning highest degree and hire at Waterloo</i>	829.88	147.10	858.08	138.79
<i>Highest degree earned:</i>				
<i>Doctoral (PhD)</i>	10084.27	3378.91	8564.38	3251.73
<i>Master's or equivalent</i>	1616.14	3519.85	146.71	3386.44
<i>Professional</i>	7339.61	4825.99	3533.10	4448.90
<i>Graduate Licence</i>	-7700.33	7201.54	-10290.55	6873.11
<i>Current academic rank:</i>				
<i>Professor</i>	15811.02	1598.82	15141.57	1525.42
<i>Associate Professor</i>	7666.75	1171.28	7325.93	1118.28
<i>Clinical Lecturer</i>	9038.29	11059.44	9675.22	10520.73
<i>Lecturer</i>	-7349.55	4462.49	-6553.49	4262.62
<i>UNIT: Academic unit of faculty member:</i>				
<i>School of Accounting &amp; Finance</i>	15244.60	5091.68	14049.27	4838.09
<i>Economics</i>	5084.50	5922.80	4167.66	5631.41
<i>Psychology</i>	-15413.75	6812.83	-16971.48	6468.86
<i>Chemical Engineering</i>	-7227.35	7405.35	-4803.26	7035.98
<i>Electrical &amp; Computer Engineering</i>	13516.82	6268.18	10629.27	5986.53
<i>School of Computer Science</i>	7520.51	5183.75	5114.92	4948.56
<i>School of Optometry</i>	10792.42	12693.56	11736.54	12002.76
<i>School of Pharmacy</i>	21043.85	2811.67	21177.10	2668.87
<i>Faculty of Environment</i>	-12225.70	8596.71	-13716.05	8169.56
<i>Arts (excluding units already listed above)</i>	-10388.70	4735.48	-11526.09	4515.81
<i>Engineering (excluding units already listed above)</i>	14752.95	4697.43	14007.87	4502.53
<i>Mathematics (excluding units already listed above)</i>	6532.34	4439.02	5383.34	4239.10
<i>Science (excluding units already listed above)</i>	-1135.99	5025.57	-1950.10	4779.45
<i>RANK@HIRE: Academic rank when hired at Waterloo:</i>				
<i>Professor</i>	10079.46	3681.30	10097.90	3495.13
<i>Associate Professor</i>	5319.26	2096.70	4952.56	1996.99
<i>Clinical Lecturer</i>	4242.50	11932.31	5743.84	11349.63
<i>Lecturer</i>	-2210.70	1641.43	-2101.91	1630.66
<i>Interaction terms between UNIT and PROFS_RANKS:</i>				
<i>UNIT=School of Accounting &amp; Finance * PROFS_RANKS</i>	18910.82	5572.08	19887.39	5284.41
<i>UNIT=Economics * PROFS_RANKS</i>	7780.52	6408.68	7679.05	6089.70
<i>UNIT=Psychology * PROFS_RANKS</i>	16166.87	7150.68	17518.53	6793.48
<i>UNIT=Chemical Engineering * PROFS_RANKS</i>	19640.16	7753.66	15814.88	7365.46
<i>UNIT=Electrical &amp; Computer Engineering * PROFS_RANKS</i>	3970.11	6512.00	5431.13	6200.93
<i>UNIT=School of Computer Science * PROFS_RANKS</i>	16368.08	5508.11	17615.55	5244.71
<i>UNIT=School of Optometry * PROFS_RANKS</i>	4233.38	12696.43	4146.50	12024.74
<i>UNIT=Faculty of Environment * PROFS_RANKS</i>	9908.77	8779.14	10683.69	8341.98
<i>UNIT=Arts * PROFS_RANKS</i>	5092.17	4996.96	5963.93	4769.95
<i>UNIT=Engineering * PROFS_RANKS</i>	779.42	4960.85	679.95	4740.64
<i>UNIT=Mathematics * PROFS_RANKS</i>	5397.36	4733.94	5372.86	4506.28
<i>UNIT=Science * PROFS_RANKS</i>	2055.99	5280.47	1765.49	5016.45
<i>Interaction terms between LAG and RANK@HIRE:</i>				
<i>LAG * RANK@HIRE=Professor</i>	889.20	228.67	889.46	216.39
<i>LAG * RANK@HIRE=Associate Professor</i>	410.74	237.82	444.89	225.81
<i>LAG * RANK@HIRE=Lecturer</i>	-335.24	181.71	-383.49	172.40
<i>Intercept</i>	46748.06	4592.89	47757.05	4401.09
<i>Number of observations</i>		1,171		1,171

The pages of this supplement will *not* be graded: write your answers on the test papers. **Supplement: Page 5 of 10**

**Sample mean:**  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$     **Sample variance:**  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{(\sum_{i=1}^n x_i)^2}{n(n-1)}$     **Sample s.d.:**  $s = \sqrt{s^2}$

**Sample coefficient of variation:**  $CV = \frac{s}{\bar{X}}$     **Sample covariance:**  $s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n-1} = \frac{\sum_{i=1}^n x_i y_i}{n-1} - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n(n-1)}$

**Sample interquartile range:**  $IQR = Q3 - Q1$     **Sample coefficient of correlation:**  $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n z_{x_i} z_{y_i}}{n-1}$

**Addition rule:**  $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$     **Conditional probability:**  $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

**Complement rules:**  $P(A^C) = P(A') = 1 - P(A)$      $P(A^C|B) = P(A'|B) = 1 - P(A|B)$

**Multiplication rule:**  $P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$

**Expected value:**  $E[X] = \mu = \sum_{all x} x p(x)$     **Variance:**  $V[X] = E[(X - \mu)^2] = \sigma^2 = \sum_{all x} (x - \mu)^2 p(x)$

**Covariance:**  $COV[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY} = \sum_{all x} \sum_{all y} (x - \mu_X)(y - \mu_Y)p(x, y)$

**Laws of expected value:**

$$\begin{aligned} E[c] &= c \\ E[X + c] &= E[X] + c \\ E[cX] &= cE[X] \\ E[a + bX + cY] &= a + bE[X] + cE[Y] \end{aligned}$$

**Laws of variance:**

$$\begin{aligned} V[c] &= 0 \\ V[X + c] &= V[X] \\ V[cX] &= c^2 V[X] \\ V[a + bX + cY] &= b^2 V[X] + c^2 V[Y] + 2bc * COV[X, Y] \\ V[a + bX + cY] &= b^2 V[X] + c^2 V[Y] + 2bc * SD(X) * SD(Y) * \rho \end{aligned}$$

where  $\rho = CORRELATION[X, Y]$

**Laws of covariance:**

$$COV[X, c] = 0$$

$$COV[a + bX, c + dY] = bd * COV[X, Y]$$

**Combinatorial formula:**  $C_x^n = \frac{n!}{x!(n-x)!}$     **Binomial probability:**  $p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$     for  $x = 0, 1, 2, \dots, n$

If  $X$  is Binomial ( $X \sim B(n, p)$ ) then  $E[X] = np$  and  $V[X] = np(1-p)$

If  $X$  is Uniform ( $X \sim U[a, b]$ ) then  $f(x) = \frac{1}{b-a}$  and  $E[X] = \frac{a+b}{2}$  and  $V[X] = \frac{(b-a)^2}{12}$

**Sampling distribution of  $\bar{X}$ :**

$$\begin{aligned} \mu_{\bar{X}} &= E[\bar{X}] = \mu \\ \sigma_{\bar{X}}^2 &= V[\bar{X}] = \frac{\sigma^2}{n} \\ \sigma_{\bar{X}} &= SD[\bar{X}] = \frac{\sigma}{\sqrt{n}} \end{aligned}$$

**Sampling distribution of  $\hat{P}$ :**

$$\begin{aligned} \mu_{\hat{P}} &= E[\hat{P}] = p \\ \sigma_{\hat{P}}^2 &= V[\hat{P}] = \frac{p(1-p)}{n} \\ \sigma_{\hat{P}} &= SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

**Sampling distribution of  $(\hat{P}_2 - \hat{P}_1)$ :**

$$\begin{aligned} \mu_{\hat{P}_2 - \hat{P}_1} &= E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1 \\ \sigma_{\hat{P}_2 - \hat{P}_1}^2 &= V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1} \\ \sigma_{\hat{P}_2 - \hat{P}_1} &= SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}} \end{aligned}$$

**Sampling distribution of  $(\bar{X}_1 - \bar{X}_2)$ , independent samples:**

$$\begin{aligned} \mu_{\bar{X}_1 - \bar{X}_2} &= E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2 \\ \sigma_{\bar{X}_1 - \bar{X}_2}^2 &= V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \\ \sigma_{\bar{X}_1 - \bar{X}_2} &= SD[\bar{X}_1 - \bar{X}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \end{aligned}$$

**Sampling distribution of  $(\bar{X}_d)$ , paired ( $d = X_1 - X_2$ ):**

$$\begin{aligned} \mu_{\bar{X}_d} &= E[\bar{X}_d] = \mu_1 - \mu_2 \\ \sigma_{\bar{X}_d}^2 &= V[\bar{X}_d] = \frac{\sigma_d^2}{n} = \frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n} \\ \sigma_{\bar{X}_d} &= SD[\bar{X}_d] = \frac{\sigma_d}{\sqrt{n}} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}} \end{aligned}$$

The pages of this supplement will *not* be graded: write your answers on the test papers.

**Supplement: Page 6 of 10**

### Inference about a population proportion:

$$\text{z test statistic: } z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad \text{CI estimator: } \hat{P} \pm z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

### Inference about comparing two population proportions:

$$\text{z test statistic under Null hypothesis of no difference: } z = \frac{\hat{P}_2 - \hat{P}_1}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n_1} + \frac{\hat{P}(1-\hat{P})}{n_2}}} \quad \text{Pooled proportion: } \bar{P} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\text{CI estimator: } (\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2} \sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$$

### Inference about the population mean:

$$\text{t test statistic: } t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \quad \text{CI estimator: } \bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad \text{Degrees of freedom: } v = n - 1$$

### Inference about a comparing two population means, independent samples, unequal variances:

$$\text{t test statistic: } t = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{CI estimator: } (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\text{Degrees of freedom: } v = \frac{\frac{1}{n_1-1} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2-1} \left( \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1-1} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2-1} \left( \frac{s_2^2}{n_2} \right)^2}$$

### Inference about a comparing two population means, independent samples, assuming equal variances:

$$\text{t test statistic: } t = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \quad \text{CI estimator: } (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad \text{Degrees of freedom: } v = n_1 + n_2 - 2$$

$$\text{Pooled variance: } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

### Inference about a comparing two population means, paired data: ( $n$ is number of pairs and $d = X_1 - X_2$ )

$$\text{t test statistic: } t = \frac{\bar{d} - \Delta_0}{s_d/\sqrt{n}} \quad \text{CI estimator: } \bar{X}_d \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} \quad \text{Degrees of freedom: } v = n - 1$$

### SIMPLE REGRESSION:

$$\text{Model: } y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{OLS line: } \hat{y}_i = b_0 + b_1 x_i \quad b_1 = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x} \quad b_0 = \bar{Y} - b_1 \bar{X}$$

$$\text{Coefficient of determination: } R^2 = (r)^2 \quad \text{Residuals: } e_i = y_i - \hat{y}_i$$

$$\text{Standard deviation of residuals: } s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (e_i - 0)^2}{n-2}} \quad \text{Standard error of slope: } s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$$

**Inference about the population slope:**

**t test statistic:**  $t = \frac{b_1 - \beta_{10}}{s.e.(b_1)}$    **CI estimator:**  $b_1 \pm t_{\alpha/2} s.e.(b_1)$    **Degrees of freedom:**  $v = n - 2$

**Standard error of slope:**  $s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$

**Prediction interval for  $y$  at given value of  $x (x_g)$ :**

$$\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}} \quad \text{or} \quad \hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{(s.e.(b_1))^2 (x_g - \bar{X})^2 + \frac{s_e^2}{n} + s_e^2}$$

**Degrees of freedom:**  $v = n - 2$

**Confidence interval for predicted mean at given value of  $x (x_g)$ :**

$$\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}} \quad \text{or} \quad \hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{(s.e.(b_1))^2 (x_g - \bar{X})^2 + \frac{s_e^2}{n}} \quad \text{Degrees of freedom: } v = n - 2$$

**SIMPLE & MULTIPLE REGRESSION:**

**Model:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$

$$SST = \sum_{i=1}^n (y_i - \bar{Y})^2 = SSR + SSE \quad SSR = \sum_{i=1}^n (\hat{y}_i - \bar{Y})^2 \quad SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$s_y^2 = \frac{SST}{n-1} \quad MSE = \frac{SSE}{n-k-1} \quad \text{Root MSE} = \sqrt{\frac{SSE}{n-k-1}} \quad MSR = \frac{SSR}{k}$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad \text{Adj. } R^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = \left( R^2 - \frac{k}{n-1} \right) \left( \frac{n-1}{n-k-1} \right)$$

$$\text{Residuals: } e_i = y_i - \hat{y}_i \quad \text{Standard deviation of residuals: } s_e = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^n (e_i - 0)^2}{n-k-1}}$$

**Inference about the overall statistical significance of the regression model:**

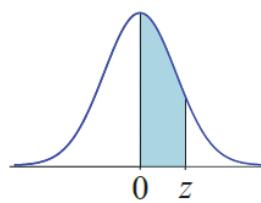
$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{(SST-SSE)/k}{SSE/(n-k-1)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

**Numerator degrees of freedom:**  $v_1 = k$    **Denominator degrees of freedom:**  $v_2 = n - k - 1$

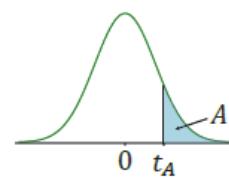
**Inference about the population slope for explanatory variable  $j$ :**

**t test statistic:**  $t = \frac{b_j - \beta_{j0}}{s_{b_j}}$    **CI estimator:**  $b_j \pm t_{\alpha/2} s_{b_j}$    **Degrees of freedom:**  $v = n - k - 1$

**Standard error of slope:**  $s.e.(b_j) = s_{b_j}$  (for multiple regression, must be obtained from technology)

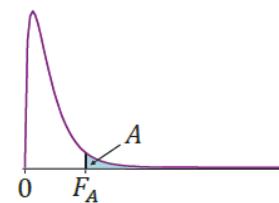


## The Standard Normal Distribution:

**Critical Values of Student  $t$  Distribution:**

$\nu$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$	$\nu$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$
1	3.078	6.314	12.71	31.82	63.66	318.3	636.6	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
2	1.886	2.920	4.303	6.965	9.925	22.33	31.60	39	1.304	1.685	2.023	2.426	2.708	3.313	3.558
3	1.638	2.353	3.182	4.541	5.841	10.21	12.92	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	41	1.303	1.683	2.020	2.421	2.701	3.301	3.544
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	43	1.302	1.681	2.017	2.416	2.695	3.291	3.532
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	45	1.301	1.679	2.014	2.412	2.690	3.281	3.520
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	47	1.300	1.678	2.012	2.408	2.685	3.273	3.510
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	49	1.299	1.677	2.010	2.405	2.680	3.265	3.500
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	51	1.298	1.675	2.008	2.402	2.676	3.258	3.492
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	52	1.298	1.675	2.007	2.400	2.674	3.255	3.488
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	53	1.298	1.674	2.006	2.399	2.672	3.251	3.484
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	54	1.297	1.674	2.005	2.397	2.670	3.248	3.480
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	75	1.293	1.665	1.992	2.377	2.643	3.202	3.425
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	140	1.288	1.656	1.977	2.353	2.611	3.149	3.361
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	160	1.287	1.654	1.975	2.350	2.607	3.142	3.352
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	180	1.286	1.653	1.973	2.347	2.603	3.136	3.345
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340
31	1.309	1.696	2.040	2.453	2.744	3.375	3.633	250	1.285	1.651	1.969	2.341	2.596	3.123	3.330
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	300	1.284	1.650	1.968	2.339	2.592	3.118	3.323
33	1.308	1.692	2.035	2.445	2.733	3.356	3.611	400	1.284	1.649	1.966	2.336	2.588	3.111	3.315
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	500	1.283	1.648	1.965	2.334	2.586	3.107	3.310
35	1.306	1.690	2.030	2.438	2.724	3.340	3.591	750	1.283	1.647	1.963	2.331	2.582	3.101	3.304
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	1000	1.282	1.646	1.962	2.330	2.581	3.098	3.300
37	1.305	1.687	2.026	2.431	2.715	3.326	3.574	$\infty$	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Degrees of freedom:  $\nu$

**The  $F$  Distribution:**

$\nu_1$	1	2	3	4	5	6	7	8	9	10	11	12	15	20	30	$\infty$
<b>Critical Values of <math>F</math> Distribution for <math>A = 0.10</math>:</b>																
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.24	3.21	3.17	3.10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.24	2.20	2.16	2.06
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	1.97	1.92	1.87	1.76
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.84	1.79	1.74	1.61
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.72	1.67	1.61	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.66	1.61	1.54	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.60	1.54	1.48	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.63	1.60	1.55	1.48	1.41	1.19
$\infty$	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55	1.49	1.42	1.34	1.00
<b>Critical Values of <math>F</math> Distribution for <math>A = 0.05</math>:</b>																
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.50	4.36
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.70	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.25	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.04	1.84
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.84	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.74	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.84	1.75	1.65	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.75	1.66	1.55	1.25
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.67	1.57	1.46	1.00
<b>Critical Values of <math>F</math> Distribution for <math>A = 0.01</math>:</b>																
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.72	9.55	9.38	9.02
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.25	3.91
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.21	2.87
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.78	2.42
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.39	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.20	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.35	2.20	2.03	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.19	2.03	1.86	1.38
$\infty$	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.04	1.88	1.70	1.00
<b>Critical Values of <math>F</math> Distribution for <math>A = 0.001</math>:</b>																
5	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	26.6	26.4	25.9	25.4	24.9	23.8
10	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.59	8.45	8.13	7.80	7.47	6.76
15	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.94	5.81	5.54	5.25	4.95	4.31
20	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.94	4.82	4.56	4.29	4.00	3.38
30	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.11	4.00	3.75	3.49	3.22	2.59
40	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.40	3.14	2.87	2.23
60	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.42	3.32	3.08	2.83	2.55	1.89
120	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.12	3.02	2.78	2.53	2.26	1.54
$\infty$	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.84	2.74	2.51	2.27	1.99	1.00

Numerator degrees of freedom:  $\nu_1$ ; Denominator degrees of freedom:  $\nu_2$