

UNIVERSITY OF TORONTO
Faculty of Arts and Science

APRIL 2018 EXAMINATIONS

ECO220Y1Y

Duration - 3 hours

Examination Aids: A non-programmable calculator

U of T E-MAIL: _____ @MAIL.UTORONTO.CA

**SURNAME
(LAST NAME):**

UTORID:
(e.g. LIHAO118)

This exam includes these **10 pages** plus the *Supplement*. There are **7 questions** (some with multiple parts) with varying point values worth a total of **120 points**. You are responsible for turning in all 10 pages of this exam and for writing your name and identifying information above *before* the end of the exam.

The *Supplement* is 10 pages and contains graphs, tables, and other materials required for some exam questions and the aid sheets. ***After the exam begins, carefully detach the Supplement.***

Write your answers clearly, completely and concisely in the designated space provided immediately after each question. An answer guide for your response ends each question: it is underlined so you do not miss it. It lets you know what is expected: for example, a quantitative analysis (which shows your work and reasoning), a fully-labelled graph, and/or sentences. Anything requested by the question and/or guide is required. If the answer guide does not request sentences, provide only what is requested (e.g. quantitative analysis). The guide sometimes specifies if your sentences should: **(1)** be clear for someone who *has not read* the *Supplement* and hence has no idea about the context for the numbers OR **(2)** presume your reader *has read* the *Supplement* and is familiar with the background information. While for **(2)** you can skip repeating basic background information (e.g. year data collected), you still have to explain what the results are and mean.

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 the ***Supplement for Question (1): Credit Card Choice, Carlin et al. (2017)***.

(a) [12 pts] Comparing people shown no taglines versus superfluous taglines, *regardless of which video they were shown*, what is the point estimate of the difference in the choice share of the dominant card? Interpret it. Is that difference statistically significant? If so, at which significance levels? Answer with a point estimate, hypotheses using formal notation, a quantitative analysis & 2 sentences (1 interpreting the point estimate and 1 assessing statistical significance) that would be clear to someone who has read the Supplement.

(b) [12 pts] For this part, consider *only those respondents shown superfluous taglines*. Comparing the baseline and implemental videos, what is the point estimate of the difference in the choice share of the dominant card? What is its margin of error for a 99% confidence level? Interpret the lower and upper confidence limits, measuring each to *the nearest percent*. Further, is the interval wide or narrow? Answer with a point estimate, quantitative analysis & 2 – 3 sentences that would be clear to someone who has read the Supplement.

(2) [5 pts] A survey asks about free trade agreements. Each respondent assesses the same list of statements and says how much s/he agrees with each (1 for strongly disagree to 7 for strongly agree). Use hypothesis testing to assess whether the average agreement with the statement “Canada benefits from NAFTA (North American Free Trade Agreement)” differs from the average agreement with the statement “Canada benefits from CETA (Canada-European Union: Comprehensive Economic and Trade Agreement).” What is the implied hypothesis test and test statistic formula? Answer with hypotheses in formal notation & test statistic formula.

(3) See the ***Supplement for Question (3): Histograms of Listing Prices on eBay.***

(a) [4 pts] Recall the common descriptive terms for the shape of a histogram: Normal (bell shaped); Uniform; symmetric; positively (right) skewed; negatively (left) skewed; unimodal; bimodal; multimodal. Which is the *single best* descriptive term to describe the shape of the distribution of listing prices in the following cases? Fill in the blanks with the *SINGLE BEST* descriptive for each case.

In Panel A, the shape of the distribution is _____.

In Panel B, the shape of the distribution from \$0 to \$50 is _____.

In Panel B, the shape of the distribution from \$900.01 to \$1,000 is _____.

In Panel B, the shape of the distribution from \$950.01 to \$1,000 is _____.

(b) [4 pts] Consider which percent of the subsample data lie in specific ranges of listing prices. Using approximation when needed, fill in the blanks with percent values. Fill in the blanks & show your work.

Observations with a listing price less than \$300.01 represent _____ percent of the subsample data.

Observations with a listing price from \$990.01 to \$1,000 represent _____ percent of the subsample data.

(4) See the ***Supplement for Question (4): Offshore Wealth as a Percentage of National GDP.***

(a) [6 pts] Among the following choices, which one is a reasonable estimate of the interquartile range (IQR): 4.6, 12.8, 17.4, 22.4, 33.3, 47, 72.1? What does it measure in this context? Is it large or small? Answer with 2 – 3 sentences that would be clear to someone who has read the Supplement.

(b) [6 pts] Draw a histogram of offshore wealth as a percentage of GDP. Make reasonable decisions where needed and use a bin width of 10 (to keep it simple). Answer with a fully-labelled graph.

(5) See the ***Supplement for Question (5): Australian wine producers' adaptation to climate change.***

(a) [5 pts] What is the regression line for a simple regression to predict adaptive practices using the natural log of firm age? Answer with a quantitative analysis & an OLS equation using standard notation.

(b) [5 pts] In a simple regression to predict adaptive practices using temperature change, what is the value of the s_e (also called: the Root MSE, standard deviation of residuals)? Answer with a quantitative analysis.

(c) [4 pts] Using stars, Table 2 reports if the correlation between each pair of variables is statistically significant at $\alpha = 0.05$ or $\alpha = 0.01$. Is the correlation between the natural log of firm age and rainfall change statistically significant at $\alpha = 0.10$? Answer with a quantitative analysis.

(6) See the ***Supplement for Question (6): Income Mobility in Canada.***

(a) [6 pts] Find the value **1.5** in the first row of results in Table 1. Is it a joint, marginal, or conditional probability? Write the probability using formal notation. Next, interpret 1.5. Answer with the type of probability, formal notation & 1 sentence that would be clear to someone who has not read the Supplement.

(b) [6 pts] In a random sample of eight taxfilers in the tenth income decile in 2007, what is the chance that seven or more of them are still in the tenth income decile in 2012? Answer with a quantitative analysis.

(c) [7 pts] In a random sample of 1,000 taxfilers in the fifth decile in 2007, what is the chance that more than 400 move to a higher income decile (upward) in 2012? Answer with a quantitative analysis.

(d) [8 pts] In hypothetical Country A, a taxfiler's income decile in 2007 is *independent* of her/his income decile in 2012. In other words, regardless of a taxfiler's 2007 decile, all 2012 deciles are equally likely. In Country A, for taxfilers in the *second decile in 2007*, what would be the values in the *last three columns* of Table 1? *Explain how Canada (values in boldface) differs from Country A. Answer with three numeric values & 2 – 3 sentences.*

(7) See the ***Supplement for Question (7): Is energy efficiency capitalized into the selling price of a home?***

(a) [5 pts] In Table 2, what is the P-value in testing whether the coefficient on the variable *New* is statistically significant? Answer with hypotheses using formal notation & a quantitative analysis.

(b) [8 pts] In Table 2, what is the interpretation of the coefficient on the number of bedrooms? What is the interpretation of the coefficient on the natural logarithm of the house size? Answer with 2 sentences that would be clear to someone who has read the Supplement.

(c) [10 pts] As you know, Table 2 reports *multiple regression* results. What would be the approximate value of the coefficient for the Energy Star variable in a *simple regression*? Why does the Energy Star coefficient in the multiple regression differ from the simple regression in the way it does? Make sure to support your arguments with the relevant information from the *Supplement*. Answer with a quantitative analysis & 2 – 3 sentences.

(d) [7 pts] The authors ran a multiple regression like Table 2 for Austin (a city in Texas) obtaining an Energy Star coefficient of 0.006 with a standard error of 0.009. Does this mean that in Austin the mean selling price of Energy Star homes is similar to the mean selling price of homes with no certification? Answer with 2 sentences.

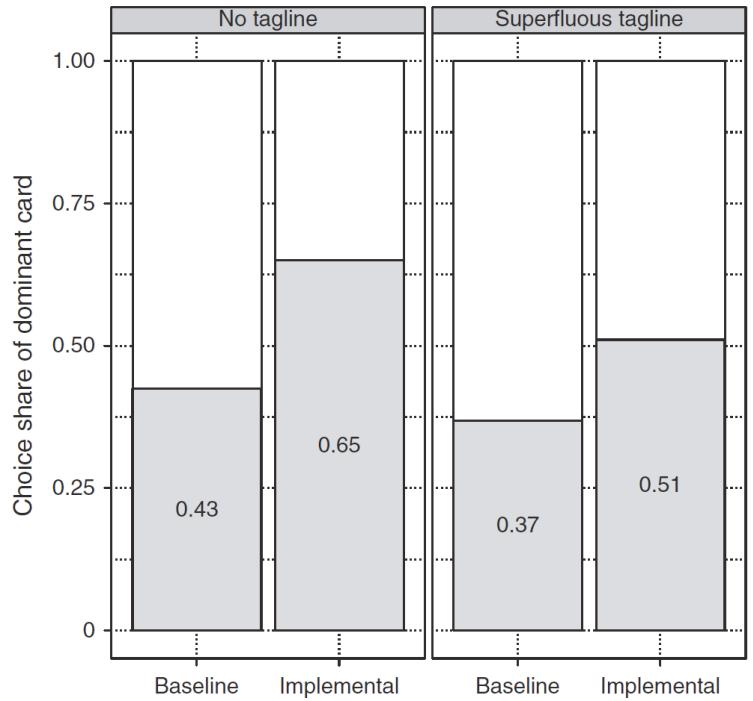
This *Supplement* contains graphs, tables, and other materials required for some exam questions and the aid sheets (formulas and Normal, *t* and *F* statistical tables). Review all relevant materials for each question.

Supplement for Question (1): Recall Carlin et al. (2017) “Millennial-Style Learning: Search Intensity, Decision Making, and Information Sharing” and Table A.1 (below) and Figure 6 (right). The researchers study peoples’ ability to choose the best credit card (*dominant card*) from among four offers. They show a short video (*baseline*) to some participants and a longer video (*implemental*) to others. Additionally, they show the four credit card offers either with misleading ads (*superfluous taglines*) or without misleading ads (*no taglines*). (<https://doi.org/10.1287/mnsc.2016.2689>)

Table A.1: Summary of Experimental Design:
Number of Respondents Receiving Each Treatment

	No tagline	Superfluous tagline	Total
Baseline video	407	394	801
Implemental video	397	405	802
Total	804	799	1,603

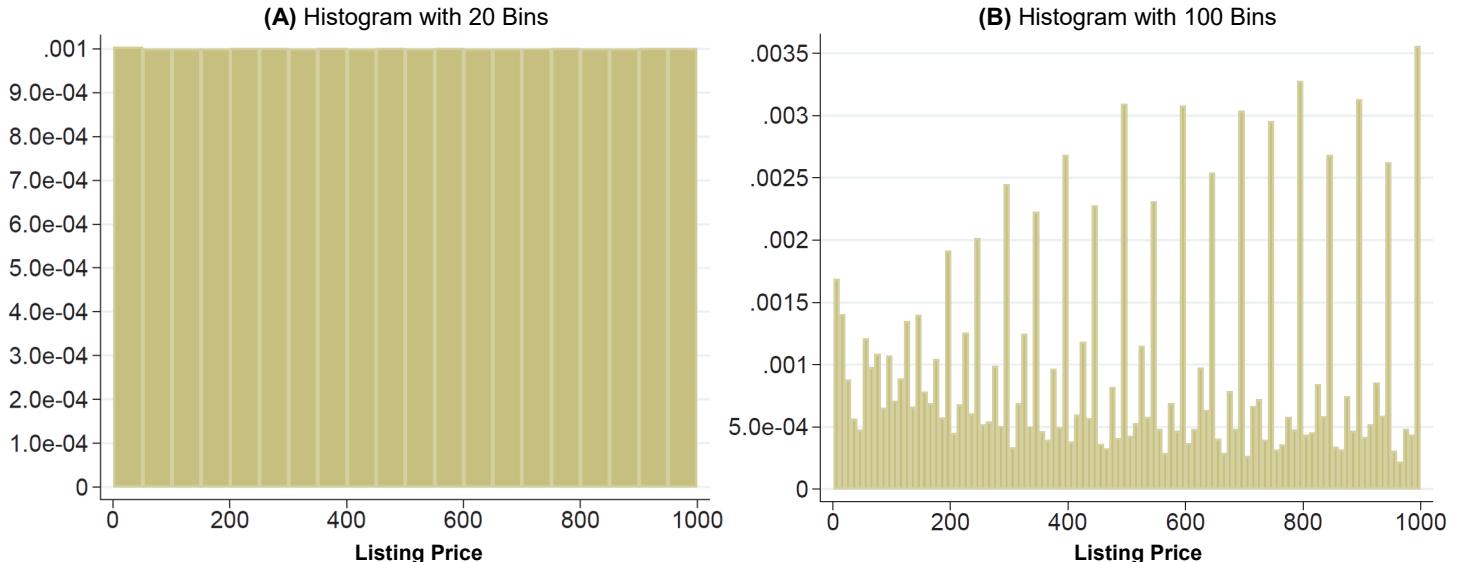
Figure 6. Choice Proportion of the Dominant Card in Each of the Four Experimental Treatments



Supplement for Question (3): A 2018 paper “Sequential Bargaining in the Field: Evidence from Millions of Online Bargaining Interactions” analyzes 88 million listings on eBay’s Best Offer platform (<http://www.nber.org/papers/w24306>). The authors use density histograms: the *area* of each bar tells the fraction of observations in that bin.

Excerpt, p. 3 of appendix: Given the large sample size of the dataset, we subsampled. We sampled 70,000 listings from 20 bins, each \$50 in length (inclusive on the upper extreme). Figure C1 presents two histograms: Panel A is a histogram with a bin width that reflects the stratification strategy, and Panel B has 100 bins of length \$10. The regular peaks in Panel B reflect the prevalence of round numbers (since our bins are constructed to be inclusive on the upper extreme).

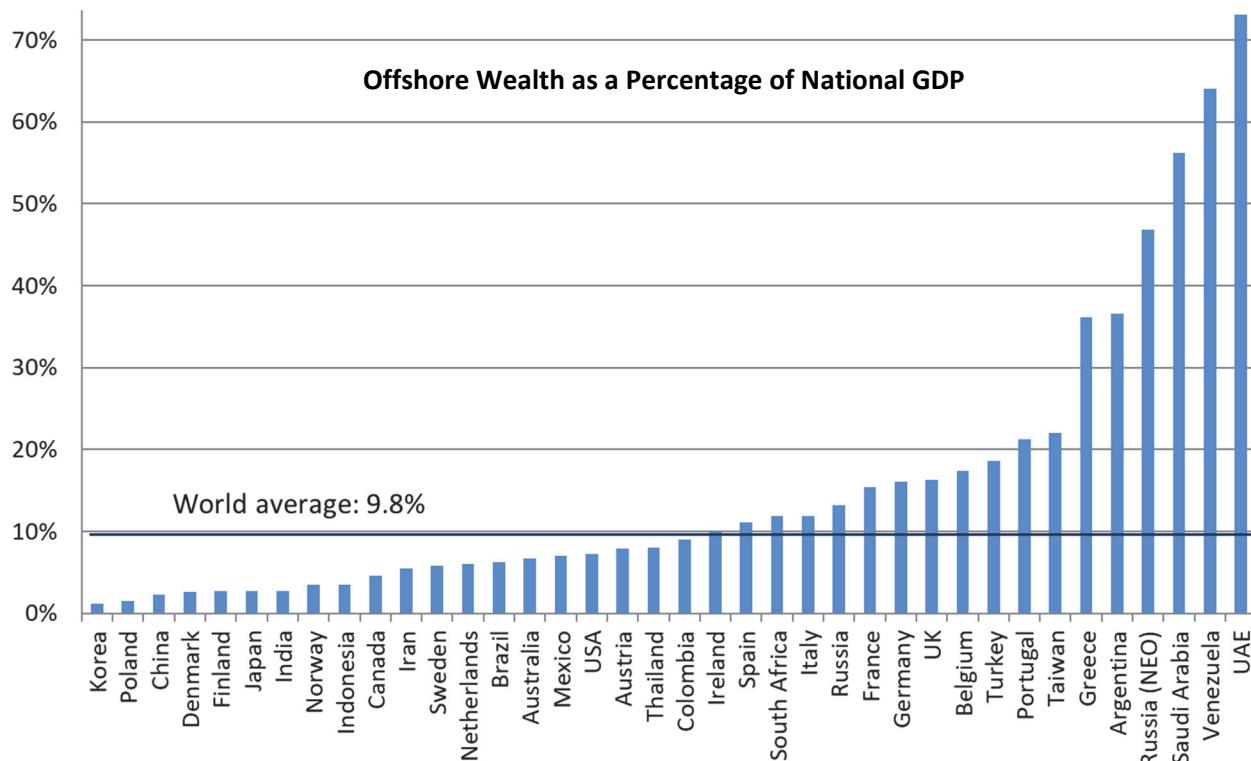
Figure C1: Histograms of the Subsample



The pages of this supplement will *not* be graded: write your answers on the exam papers. **Supplement: Page 2 of 10**

Supplement for Question (4): Consider “How Much Wealth is Stashed in Tax Havens?” in the Nov. 2017 *NBER Digest*.

Excerpt, p. 5: Offshore tax havens [places with low tax rates] in exotic locales [e.g. Cayman Islands] have been the subject of Hollywood movies and best-selling novels. But data on the importance of tax havens, and the amount of wealth held in them, is elusive. [A 2017 working paper] makes country-by-country estimates of tax haven wealth holdings. They estimate that wealth worth about 10 percent of global GDP is held offshore. There is substantial variation across nations in the share of their citizens’ wealth that is held in tax havens. [See figure below from the original research.] Scandinavians hold only a few percent of GDP as offshore wealth, but that number rises to about 15 percent for continental Europe as a whole, and to over 60 percent in some Persian Gulf and Latin American nations. (<http://www.nber.org/digest/nov17/nov17.pdf>)



Notes: This figure shows the amount of household wealth owned offshore as a percentage of GDP, in 2007. The sample includes all 38 of the world’s countries with more than \$200 billion in GDP in 2007. For Russia, we report an alternative estimate (“Russia (NEO)”) obtained by cumulating net errors and omissions in the balance of payment.

Supplement for Question (5): In a 2017 working paper “The impact of climate change on firm adaptation: A study of wine firms in South Australia,” Jeremy Galbreath uses survey replies from 207 wine firms. Table 2 (below) is an excerpt. To measure adaptive practices, he averages replies for seven statements on a seven-point Likert scale (1 for strongly disagree to 7 for strongly agree) such as “my firm is developing wine products based on hotter climate varieties.” The firm’s age is the natural log of years since the firm’s founding and the survey (2012). For each of 11 geographic areas the Australian Bureau of Meteorology gives temperature and rainfall data. He subtracts the 1982 annual mean from the 2012 annual mean to measure how much temperature and rainfall has changed from 1982 to 2012. He assigns each firm the change values for its geographic area. (<http://www.wine-economics.org/aawe-working-paper-no-220-economics/>)

Table 2: Descriptive Statistics and Correlations

Variable	Mean	SD	1.	2.	3.	4.
1. Adaptive practices	4.18	1.39	1.00			
2. Natural log of firm age (years)	1.26	0.40	0.15*	1.00		
3. Temperature change 2012 – 1982 (Celsius)	0.46	0.18	0.14*	0.23**	1.00	
4. Rainfall change 2012 – 1982 (mm)	95.49	39.63	0.04	-0.08	-0.34**	1.00

Notes: The first column explains the definition of each variable number (1 through 4). * P-value < 0.05; ** P-value < 0.01

The pages of this supplement will *not* be graded: write your answers on the exam papers. **Supplement: Page 3 of 10**

Supplement for Question (6): In 2016, Statistics Canada published “The evolution of income mobility in Canada: Evidence from the Longitudinal Administrative Databank, 1982 to 2012.” An excerpt and accompanying table are below. For ease of reference, Table 1 puts in boldface some numbers that the excerpt or the exam questions mention.

Excerpt, pp. 15 – 16: We employed several mobility indexes to measure the aggregate relative mobility among Canadian taxfilers. All of them were based on the income transition matrices. A transition matrix indicates how the income position of a person changes between two points of time. The transition matrix approach is quite flexible. Individuals can be classified into deciles, quintiles, quartiles, and so on over different lengths of transition period. We conducted our analysis using both the quintile and decile transition matrices and examined the transition probabilities over 5- and 10-year periods. The conclusions were essentially the same.

Table 1 gives an example of the transition matrix. It shows the estimated transition probabilities for the 2007-to-2012 panel of taxfilers. As an illustration, consider the third row of the matrix. The 10 elements in this row show the 5-year transition probabilities for taxfilers who were in the third decile in 2007. The third element from the left in the third row, **36.3%**, shows the proportion of taxfilers from the third decile in the 2007 distribution who stayed in the same decile in 2012. [This measures immobility.] The sum of the other nine elements from the third row tells the proportion of taxfilers who moved to a different decile five years later. [This] measures total relative mobility for taxfilers who were in the third decile in 2007.

Total relative mobility can be further decomposed into upward and downward mobility. Upward mobility shows the proportion of taxfilers who move up along the distribution ladder. For taxfilers who were in the third decile in 2007, upward mobility indicates how many of them moved to the fourth or higher deciles in 2012. Downward mobility shows the proportion of taxfilers who move down the distribution ladder. For taxfilers who were in the third decile in 2007, downward mobility indicates how many of them moved to the second or the first deciles in 2012. (<http://www.statcan.gc.ca/pub/75f0002m/75f0002m2016001-eng.htm>)

Table 1. Decile Income Transition Matrix for the 2007-to-2012 panel of Taxfilers

2007 decile	2012 decile										Total mobility statistics		
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Immobility	Upward	Downward
	Percentage										Percentage		
1st	39.7	22.9	11.2	7.6	5.4	4.0	3.1	2.5	2.0	1.5	39.7	60.3	0.0
2nd	13.5	39.4	18.5	10.0	6.2	4.4	3.1	2.2	1.6	1.0	39.4	47.0	13.5
3rd	6.4	14.9	36.3	16.9	9.7	6.1	4.1	2.8	1.8	1.0	36.3	42.4	21.4
4th	4.5	7.2	17.5	27.6	17.5	10.7	6.8	4.3	2.6	1.3	27.6	43.2	29.2
5th	3.1	4.4	8.2	17.0	25.6	17.6	11.3	6.9	4.0	1.8	25.6	41.6	32.7
6th	2.3	3.0	5.1	9.0	16.9	24.3	18.3	11.7	6.5	2.7	24.3	39.3	36.4
7th	1.8	2.1	3.4	5.9	9.5	16.9	24.3	19.6	11.8	4.6	24.3	36.0	39.7
8th	1.4	1.6	2.3	4.0	6.4	9.9	17.2	26.3	22.0	8.8	26.3	30.9	42.8
9th	1.2	1.2	1.6	2.7	4.1	6.4	10.1	18.1	32.5	22.1	32.5	22.1	45.4
10th	1.2	0.9	1.1	1.6	2.4	3.4	5.3	8.6	18.2	57.4	57.4	0.0	42.6

Source: Statistics Canada, Longitudinal Administrative Databank 2007 and 2012, authors' calculations.

The pages of this supplement will *not* be graded: write your answers on the exam papers. **Supplement: Page 4 of 10**

Supplement for Question (7): A 2017 article “Is energy efficiency capitalized into home prices? Evidence from three U.S. cities” uses transaction data to study how well a home’s selling price reflects its energy efficiency certification.

Excerpt, p. 104: One often given explanation for underinvestment in residential energy efficiency is that homeowners do not expect to occupy their homes long enough to realize energy savings benefits that offset up-front investment costs. If energy efficiency features of a home are capitalized in the selling price, then homeowners could recover their costs when they sell their homes, but asymmetric information [buyers not being able to directly observe the energy efficiency of the home, which is known to the sellers] is likely to be a barrier. The federal government’s Energy Star program was established, in part, to overcome some of these information problems. Homes certified under the program are 15 percent more efficient than homes that meet most current building codes. (<http://dx.doi.org/10.1016/j.jeem.2016.11.006>)

Consider the authors’ analysis of single-family homes sold in the Research Triangle area of North Carolina (N.C.) between 2009 and 2011. For each of the 16,038 transactions, the authors observe the selling price, characteristics of the home (e.g. number of bathrooms), and whether or not it has an Energy Star certification. See Tables 1 and 2 below.

Table 1. Descriptive Statistics for Selected House Characteristics by Certification Type, Research Triangle, N.C.

	No Certification		Energy Star	
	Mean	S.D.	Mean	S.D.
Home’s Selling Price (2011\$)	267,685	162,848	326,940	157,993
New (=1 if a brand new house, =0 otherwise)	0.14	0.34	0.65	0.48
Number of Bedrooms	3.57	0.79	3.86	0.77
Number of Full Bathrooms	2.35	0.77	2.69	0.76
Number of Half Bathrooms (toilet & sink only)	0.60	0.52	0.66	0.51
House Size (in square feet)	2,363	983	2,701	808
Garage (=1 if has a garage, =0 otherwise)	0.74	0.44	0.98	0.14
Number of Observations	14,068		1,970	

Table 2. Multiple Regression Results for Research Triangle, N.C.

Dependent Variable: Natural logarithm of the Home’s Selling Price	Coefficient Estimate (Standard Error)
Explanatory Variables:	
Energy Star (=1 if home is Energy Star Certified, =0 otherwise)	0.027 (0.008)
New	0.020 (0.007)
Age (years)	-0.010 (0.001)
Age squared	0.00007 (0.00001)
Number of Bedrooms	-0.054 (0.005)
Number of Full Bathrooms	0.091 (0.005)
Number of Half Bathrooms	0.044 (0.006)
Natural logarithm of the House Size	0.774 (0.014)
Fireplace (=1 if has a fireplace, =0 otherwise)	0.105 (0.009)
Garage	0.104 (0.008)
Number of Stories	-0.085 (0.008)
Natural logarithm of the Size of Property (in acres)	0.054 (0.005)
Private Pool (=1 if has pool, =0 otherwise)	0.058 (0.017)
Home Owners Assoc. Pool (=1 if has HOA pool, =0 otherwise)	0.038 (0.005)
Number of Observations	16,038
Adjusted R-squared	0.877

Notes: The regression model also includes Census tract-year/quarter fixed effects (coefficient estimates not reported), which control for variation in prices across neighborhoods and across time. The table reports robust standard errors.

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

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

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

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:

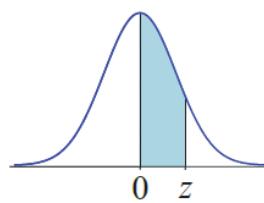
$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{(SST-SSE)/k}{SSE/(n-k-1)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

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

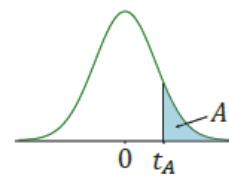
Inference about the population slope for explanatory variable j :

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)



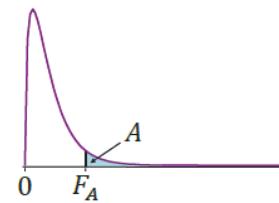
The Standard Normal Distribution:



Critical Values of Student t Distribution:

ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$	ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$
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	∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Degrees of freedom: ν



The F Distribution:

ν_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	30	∞
ν_2 Critical Values of F Distribution for $A = 0.10$:																
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
∞	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
ν_2 Critical Values of F Distribution for $A = 0.05$:																
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
∞	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
ν_2 Critical Values of F Distribution for $A = 0.01$:																
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
∞	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
ν_2 Critical Values of F Distribution for $A = 0.001$:																
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
∞	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: ν_1 ; Denominator degrees of freedom: ν_2