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

SURNAME (Family name):

GIVEN NAME (First name):

UTORID: (e.g. LIHAO118)

UNIVERSITY OF TORONTO

Faculty of Arts & Science

APRIL 2023 EXAMINATIONS

ECO220Y1Y: Introduction to Data Analysis and Applied Econometrics

Duration: 3 hours

Aids Allowed: A non-programmable calculator

Exam Reminders:

- Above print your U of T e-mail address (the part *before* @mail.utoronto.ca), your name, and your UTORid.
- Leave these papers face up on your desk until the announcements end and the Exam Facilitator starts the exam.
- As a student, you help create a fair and inclusive writing environment. If you possess an unauthorized aid during an exam, you may be charged with an academic offence.
- Turn off and place all cell phones, smart watches, electronic devices, and study materials in your bag under your desk. If it is left in your pocket, it may be an academic offence.
- When you are done your exam, raise your hand for someone to come and collect your exam.
- If you are feeling ill and unable to finish your exam, please bring it to the attention of an Exam Facilitator so it can be recorded before leaving the exam hall.
- In the event of a fire alarm, do not check your cell phone when escorted outside.

Exam Format, Grading, and Special Instructions:

- This exam includes these ***10 pages*** plus the *Supplement*. There are ***5 questions*** (all with multiple parts) with varying point values worth a total of ***120 points***.

- The *Supplement* is 10 pages and contains graphs, tables, and other materials required for exam questions and the aid sheets. ***After the exam begins, carefully DETACH the Supplement.*** The *Supplement* is NOT graded.

Write your answers clearly, completely, and concisely. ***Your entire answer must fit in the designated space provided immediately after each question.*** No extra space/pages are possible. ***Write in PENCIL and use an ERASER as needed.*** Follow the answer guides that end each question and avoid excessively long answers.

Students must hand in all examination materials at the end

(1) See the ***Supplement for Question (1)****:* Title: *Persistent Overconfidence and Biased Memory: Evidence from Managers.*

(a) [3 pts] In **Panel A** of **Figure 1**, *interpret* the height of the tallest bar. (Recall the title above.) Answer with 1 sentence.

(b) [4 pts] In **Panel B** of **Figure 1**, locate the number 0.14. *Interpret* it. (Recall the title above.) Answer with 1 sentence.

(c) [4 pts] If managers predicted their own performance perfectly accurately, what would **Panel A** in **Figure 1** look like? Explain. What would **Panel B** look like? Explain. Answer with 2 sentences.

(d) [4 pts] If the correlation were below 0.47 (but still positive), how would **Panel B** in **Figure 1** look differently? Explain. Answer with 1 – 2 sentences.

(2) See the ***Supplement for Question (2)****: Locus of Control and Prosocial Behavior.*

(a) [6 pts] Using **Table A3**, draw a box plot summarizing the replies to Question **5)**. <u>Answer with a fully labelled graph.</u>

(b) [4 pts] In **Figure A1** the number of respondents in the tallest bar is *approximately* **_________** respondents. <u>Answer with a quantitative analysis & by filling in the blank.</u>

(c) [7 pts] In **Table A3**, for Question **8)** what does "(1.5)" measure in this context? How would "(1.5)" be expected to change if the sample size were $n = 400$ instead of $n = 18{,}405$? *Explain.* Answer with 2 – 3 sentences.

(d) [6 pts] In **Table A3** what is the standard deviation of a new variable that is the sum of replies to Questions **1)** and **5)**?

- *Answer #1:* 2.4 $(= 1.3 + 1.1)$
- *Answer #2:* 1.7 $\left(= \sqrt{1.3^2 + 1.1^2}\right)$

Which is correct: *Answer #1*, *Answer #2*, or neither? *Explain*. Answer with 2 – 3 sentences.

(e) [7 pts] In **Column (1)** of **Table A12**, *interpret* both 16.462 and 68.257. Answer with 2 – 3 sentences.

(f) [8 pts] Continuing, in **Column (2)** of **Table A12**, *interpret* 6.721. (Presume your reader sees your answer to Part **(e)** and do not be repetitive.) Next, see the **excerpt** immediately above Table A12. Importantly, explain how that excerpt relates to interpreting 16.462 and 6.721 in **Columns (1)** and **(2)**, respectively. Answer with 3 – 4 sentences.

(g) [4 pts] In **Column (6)** of **Table A12**, what is the interpretation of -96.647? Answer with 1 sentence.

(h) [4 pts] In **Column (8)** of **Table A12**, what is the interpretation of 0.222? Answer with 1 sentence.

(i) [5 pts] If the survey had asked about religion and gave respondents four choices – 1) Christian, 2) Muslim, 3) Other religion, and 4) Not religious (to match the four most common likely answers in Germany) – how would you add controls for religion in Columns (2), (4), (6), and (8) of **Table A12**? Explain. Answer with 2 sentences.

(3) See the ***Supplement for Question (3)****: Fighting Climate Change: International Attitudes Toward Climate Policies.*

(a) [14 pts] To test if opinions on **limiting flying** differ between **Canada** and **France**, what is the P-value? Is the difference statistically significant? If so, at which significance level? Next, is the difference *economically significant*? *Explain.* Answer with hypotheses in formal notation, a quantitative analysis & 2 – 3 sentences.

(b) [6 pts] The 95% confidence interval estimate of how **Italy** and **Japan** differ for **limiting beef/meat consumption** is [0.351980059, 0.408039387]. *Interpret* this interval. <u>Answer with 2 – 3 sentences.</u>

(c) [6 pts] Another table – <u>not</u> shown in the *Supplement* – reports that 42 percent of the population of China is over 50 years old, but just 27 percent of the sample from China is. What is the chance that sampling error could explain why the sample percent is so much lower than the population percent? <u>Answer with a quantitative analysis.</u>

(4) See the ***Supplement for Question (4)****: Mortality Inequality: The Good News from a County-Level Approach.*

(a) [8 pts] Following the figures in Currie and Schwandt (2016), draw one figure to *effectively communicate* the meaning of the regression coefficient estimates. Answer with one fully labelled graph & show your work.

(b) [7 pts] The coefficient on **quanX1990** is statistically significant. How does the Excel output support this and at which significance level? What does it *mean in this context* that it's statistically significant? *Interpret.* Answer with 3 sentences.

(5) For a multiple regression with three right-hand-side (explanatory) variables to explain the left-hand-side (dependent) variable, consider testing whether the model is statistically significant overall at the 1% level.

(a) [7 pts] For a random sample with 24 observations, how big does the R^2 need to be for the multiple regression to be statistically significant overall at the 1% level? Answer with hypotheses in formal notation & a quantitative analysis.

(b) [6 pts] If the sample had 2,400 observations, then the answer to Part **(a)** would be an R^2 of at least 0.0048. Explain *conceptually* why the minimum required R^2 is much smaller for a sample size of 2,400 versus a sample size of 24. (Do not discuss the mechanics of the calculations. A *conceptual* explanation is required.) Answer with 3 – 4 sentences.

CAREFULLY <u>DETACH</u> THIS SUPPLEMENT FROM YOUR EXAM PAPERS <u>NOW</u>

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

Supplement for Question (1): Consider the excerpts and Figure 1 from an academic journal article published in 2022 in the *American Economic Review* "Persistent Overconfidence and Biased Memory: Evidence from Managers." ***<u>NOTE:</u>*** You do **<u>NOT</u>** need the details in the *first two excerpts* for your interpretations: these simply help you understand the study.

> **Excerpt, Abstract:** A long-standing puzzle is how overconfidence can persist in settings characterized by repeated feedback. This paper studies managers who participate repeatedly in a high-powered tournament incentive system, learning relative performance each time.
>
> **Excerpt, pp. 3,145-3,146:** A manager's rank in the quarterly tournament is determined by relative performance on four dimensions: (i) a measure of store profits that is designed to isolate manager contributions independent of store characteristics and location; (ii) sales growth; (iii) a customer service rating by an undercover "mystery shopper"; (iv) an evaluation of the store manager by a regional manager against centrally set criteria.

A manager's rank affects their bonus. For the nationwide tournament in the upcoming fourth quarter, 212 managers predict their own performance quintile: either being in the first (worst), second, third, fourth, or fifth (best) quintile of the performance distribution. Below, Figure 1 summarizes the results, and the excerpt discusses Panel B of Figure 1.

> **Excerpt, pp. 3,153:** A comparison to realized outcomes in the fourth quarter, shown in Panel B of Figure 1, suggests that managers do have insights into predicting future performance: achieved outcome (realization) and prediction are significantly positively correlated, with a coefficient of correlation of 0.47.

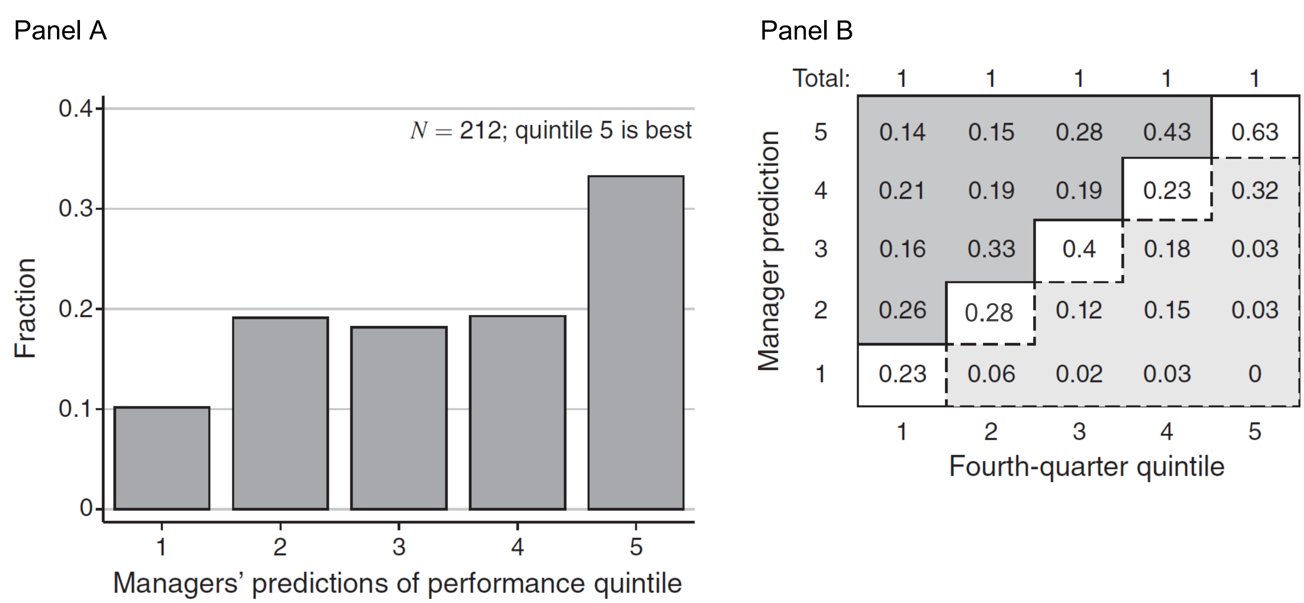


Figure 1: Distribution of managers' predictions for the upcoming fourth quarter and comparison to realizations

Supplement for Question (2): In a 2022 *NBER Working Paper* "Locus of Control and Prosocial Behavior," the abstract starts with: "We investigate how locus of control beliefs – the extent to which people attribute control over events in their life to themselves as opposed to outside factors – affect prosocial behavior." An example of prosocial behavior is donating to charity. They abbreviate "locus of control" as LOC and use survey questions to measure it. The German Socio-Economic Panel is a nationally representative survey of German households in 2010. Respondents assess the eight statements below on a Likert scale from 1 (strongly disagree) to 7 (strongly agree). In square brackets it indicates whether a higher numeric reply means a stronger or weaker locus of control (LOC). Table A3 summarizes the results.

1) [stronger LOC] *How my life goes depends on me.*
2) [weaker LOC] *Compared to other people, I have not achieved what I deserve.*
3) [weaker LOC] *What a person achieves in life is above all a question of fate or luck.*
4) [weaker LOC] *I frequently have the experience that other people have a controlling influence over my life.*
5) [stronger LOC] *One has to work hard in order to succeed.*
6) [weaker LOC] *If I run up against difficulties in life, I often doubt my own abilities.*
7) [weaker LOC] *The opportunities that I have in life are determined by the social conditions.*
8) [weaker LOC] *I have little control over the things that happen in my life.*

Table A3: Answers to the LOC-Module in the German Socio-Economic Panel, $n = 18{,}405$

	Percent of respondents selecting each value (Likert scale 1 to 7)							
Question	**1** *(Strongly disagree)*	**2**	**3**	**4**	**5**	**6**	**7** *(Strongly agree)*	Mean (S.D.)
1)	1.0	1.9	5.3	12.6	26.1	28.7	24.4	5.4 (1.3)
2)	20.2	21.2	14.7	15.9	14.8	9.0	4.3	3.3 (1.8)
3)	11.5	19.7	18.5	22.4	14.3	8.4	5.3	3.5 (1.7)
4)	20.6	23.9	17.1	15.4	12.5	7.3	3.3	3.1 (1.7)
5)	0.5	0.9	2.0	6.6	17.2	33.9	39.0	6.0 (1.1)
6)	17.4	24.1	18.2	16.9	13.7	6.8	2.8	3.2 (1.6)
7)	3.1	7.1	12.1	26.6	25.3	17.7	8.1	4.5 (1.5)
8)	24.1	30.4	17.7	14.2	7.7	4.1	1.8	2.7 (1.5)

They combine replies to the eight questions into an index where *higher index values* mean a *stronger* locus of control. (The method corrects for the fact that, in the raw data, higher Likert replies (1 to 7) for some questions mean a weaker locus of control.) Finally, they standardize the index. The histogram in Figure A1 summarizes the standardized index.

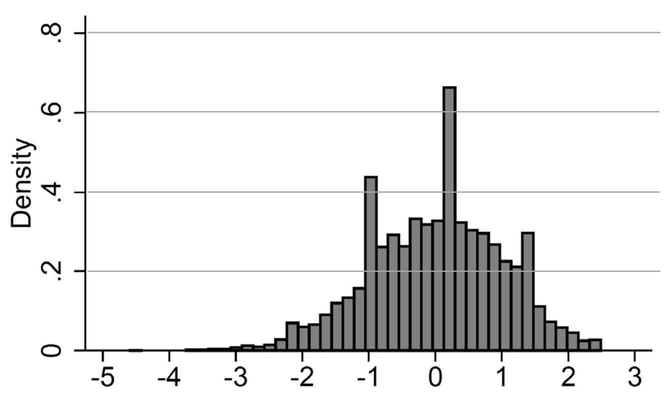


Figure A1: Distribution of the standardized LOC index: German Socio-Economic Panel, n = 18,405

Supplement for Question (2), continues the next page >>>>>

Supplement for Question (2), continued:

Respondents also answered the next questions. The researchers refer to them with the terms in square brackets.

- [Charitable donations in 2009] *We now have a question about donations. By donations we mean giving money for social, church, cultural, non-profit, and charitable purposes without receiving any direct consideration. These can be larger amounts, but also smaller ones, which can be put into a collection box. We also include the collection in the church. Did you donate money last year, in 2009 – membership fees not included?*
  - (If yes) *What was the total amount you donated last year?*
- [Hypothetical dictator game] *Imagine that you unexpectedly received a gift of 10,000 euros. How would you use this money? How much would you save, how much would you give away, and how much would you spend?*

The last question is called the "Hypothetical dictator game" because the person has complete control over any giving, and it is not subject to any bargaining or negotiations with other people.

Below, consider the excerpt and Table A12.

> **Excerpt, page 6:** Unconditional correlation estimates may capture not only the influence of LOC beliefs, but also confounding factors that are correlated with both LOC beliefs and the outcome of interest. For example, individuals with a high internal LOC tend to have better educational and labor market outcomes (Coleman and DeLeire, 2003; Cobb-Clark, 2015), both of which have been shown to affect prosociality (Bekkers and Wiepking, 2011; Wiepking and Bekkers, 2012).

Table A12: German Socio-Economic Panel – OLS estimates

	Charitable donations in 2009				Hypothetical dictator game			
Dependent variable:	*Amount donated, in euros*		*Amount donated, standardized*		*Amount given, in euros*		*Amount given, standardized*	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
LOC^{stand}	16.462 (1.346)	6.721 (1.293)	0.101 (0.008)	0.041 (0.008)	103.060 (25.667)	174.467 (23.001)	0.042 (0.010)	0.070 (0.009)
Ln(Income)	–	51.596 (2.383)	–	0.316 (0.015)	–	-96.647 (42.173)	–	-0.039 (0.017)
Female	–	-1.964 (2.533)	–	-0.012 (0.016)	–	551.261 (45.283)	–	0.222 (0.018)
Age in years	–	2.013 (0.076)	–	0.012 (0.000)	–	65.914 (1.292)	–	0.027 (0.001)
Years of education	–	12.222 (0.513)	–	0.075 (0.003)	–	-22.702 (8.877)	–	-0.009 (0.004)
Constant	68.257 (1.346)	-588.476 (18.602)	-0.000 (0.008)	-4.026 (0.114)	2163.221 (25.666)	-587.001 (328.090)	-0.000 (0.010)	-1.109 (0.132)
Number of obs.	14,539	14,539	14,539	14,539	9,327	9,327	9,327	9,327

Notes: Robust standard errors are reported in parentheses. LOC^{stand} is the standardized locus of control index based on eight questions where a higher value means a stronger locus of control. Ln(Income) is the natural logarithm of monthly household net income in euros. For each specification, it also reports a regression using a standardized outcome measure (dependent variable).

Supplement for Question (3): Consider an *NBER Working Paper* (2022) “Fighting Climate Change: International Attitudes Toward Climate Policies.” It uses an international survey run between March 2021 and March 2022.

Table 7. Share of people willing to adopt climate-friendly behaviors

	High-income	Australia	Canada	Denmark	France	Germany	Italy	Japan	Poland	South Korea	Spain	United Kingdom	United States	**Middle-income**	Brazil	China	India	Indonesia	Mexico	South Africa	Turkey	Ukraine
Have a fuel-efficient or electric vehicle	54	45	52	60	45	45	78	48	53	57	60	51	50	69	78	65	74	67	70	60	73	62
Limit flying	51	37	53	49	56	64	64	37	58	43	62	46	39	55	52	59	66	56	59	48	44	49
Limit beef/meat consumption	40	31	38	33	38	45	62	24	49	36	44	44	36	44	44	48	62	49	40	33	35	35
Limit driving	37	26	35	33	32	41	57	37	41	36	47	37	29	49	41	62	66	54	47	38	46	25
Limit heating or cooling your home	34	25	27	33	39	36	55	26	37	29	46	30	28	48	46	56	68	60	59	39	34	9
Sample size (number of people responding to the survey)	24,599	1,978	2,022	2,013	2,006	2,006	2,088	1,990	2,053	1,932	2,268	2,025	2,218	16,081	1,860	1,717	2,472	2,488	2,045	1,932	2,003	1,564

Notes: For “To what extent would you be willing to adopt the following behaviors” respondents answer on a 5-point scale: “Not at all,” “A little,” “Moderately,” “A lot,” and “A great deal.” The table reports the percent answering either “A lot” or “A great deal” for each climate-friendly behavior listed above.

Supplement for Question (4): Recall Currie and Schwandt (2016) “Mortality Inequality: The Good News from a County-Level Approach.” For women aged 65 years and over, the Excel regression output below uses 40 observations in total for 2010 and 1990. For each year, they create 20 county groups. The variable quantile is the poverty rate quantile of the county group: it takes values of 5, 10, 15, …, 100. A higher quantile means more poverty. The variable yr1990 is 1 for 1990 and 0 otherwise. The variable quanX1990 is the product of the two variables quantile and yr1990. The dependent variable is created with the following calculation: 1,000 times adjusted deaths divided by adjusted population.

Regression Statistics	
R Squared	0.888783818
Observations	40

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	3068.453397	1022.817799	95.89796736	3.13199E-17
Residual	36	383.9647677	10.66568799		
Total	39	3452.418164			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	107.2167507	1.517083781	70.67292661	3.20953E-40	104.1399622	110.2935392
quantile	0.154021206	0.025328741	6.080886677	5.41944E-07	0.102652138	0.205390275
yr1990	21.58067632	2.145480458	10.05866832	5.31233E-12	17.22944027	25.93191236
quanX1990	-0.102194588	0.03582025	-2.852983683	0.007135091	-0.174841421	-0.029547754

Supplement for Question (5): Not applicable because all information is given with the question.

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

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

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

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

<u>Laws of expected value</u>:

$E[c] = c$

$E[X+c] = E[X] + c$

$E[cX] = cE[X]$

$E[a+bX+cY] = a + bE[X] + cE[Y]$

<u>Laws of variance</u>:

$V[c] = 0$

$V[X+c] = V[X]$

$V[cX] = c^2V[X]$

$V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc * COV[X,Y]$

$V[a+bX+cY] = b^2V[X] + c^2V[Y] + 2bc * SD(X) * SD(Y) * \rho$

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

<u>Laws of covariance</u>:

$COV[X,c] = 0$

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

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 $x = 0,1,2,\ldots,n$

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** $V[X] = \frac{(b-a)^2}{12}$

<u>Sampling distribution of $\bar{X}$</u>:

$\mu_{\bar{X}} = E[\bar{X}] = \mu$

$\sigma_{\bar{X}}^2 = V[\bar{X}] = \frac{\sigma^2}{n}$

$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

<u>Sampling distribution of $\hat{P}$</u>:

$\mu_{\hat{P}} = E[\hat{P}] = p$

$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$

$\sigma_{\hat{P}} = SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}}$

<u>Sampling distribution of $(\hat{P}_2 - \hat{P}_1)$</u>:

$\mu_{\hat{P}_2-\hat{P}_1} = E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1$

$\sigma_{\hat{P}_2-\hat{P}_1}^2 = V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$

$\sigma_{\hat{P}_2-\hat{P}_1} = SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}}$

<u>Sampling distribution of $(\bar{X}_1 - \bar{X}_2)$, independent samples</u>:

$\mu_{\bar{X}_1-\bar{X}_2} = E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$

$\sigma_{\bar{X}_1-\bar{X}_2}^2 = V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$\sigma_{\bar{X}_1-\bar{X}_2} = SD[\bar{X}_1 - \bar{X}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

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

$\mu_{\bar{X}_d} = E[\bar{X}_d] = \mu_1 - \mu_2$

$\sigma_{\bar{X}_d}^2 = V[\bar{X}_d] = \frac{\sigma_d^2}{n} = \frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}$

$\sigma_{\bar{X}_d} = SD[\bar{X}_d] = \frac{\sigma_d}{\sqrt{n}} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}}$

Inference about a population proportion:

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

Inference about comparing two population proportions:

***z* test statistic under Null hypothesis of no difference:** $z = \frac{\hat{P}_2 - \hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}}$ **Pooled proportion:** $\bar{P} = \frac{X_1 + X_2}{n_1 + n_2}$

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:

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

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

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

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

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

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

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 $d = X_1 - X_2$)

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

SIMPLE REGRESSION:

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

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

Standard deviation of residuals: $s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - 0)^2}{n-2}}$ **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:** $\nu = n - 2$

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):

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

Degrees of freedom: $\nu = n - 2$

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

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

SIMPLE & MULTIPLE REGRESSION:

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

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

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

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

Residuals: $e_i = y_i - \hat{y}_i$ **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:

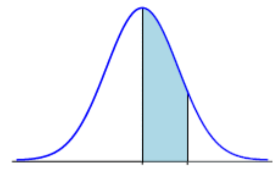
$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{(SST - SSE)/k}{SSE/(n-k-1)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$

Numerator degrees of freedom: $\nu_1 = k$ **Denominator degrees of freedom:** $\nu_2 = n - k - 1$

Inference about the population slope for explanatory variable j:

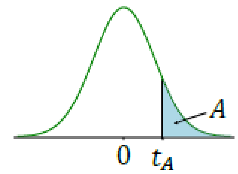
***t* test statistic:** $t = \frac{b_j - \beta_{j0}}{s_{b_j}}$ **CI estimator:** $b_j \pm t_{\alpha/2} s_{b_j}$ **Degrees of freedom:** $\nu = n - k - 1$

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



The Standard Normal Distribution:

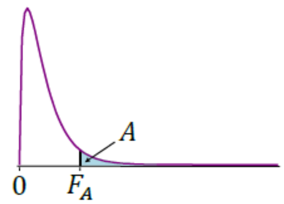
	Second decimal place in z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999



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