ECO220Y, Term Test #4

March 31, 2017, 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 (some 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 the aid sheets (formula sheets and Normal, t and F statistical tables) as well as graphs, tables, and other information needed to answer some of the test questions.
    - Anything written on 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 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 or a one or two word answer).
    - For questions with multiple parts (e.g. (a) – (e)), ***attempt each part*** even if you have trouble with earlier parts. In other words, not being able to answer (a) does NOT imply you cannot answer (b).
    - Be careful with parts marked [No partial credit]: your answer and work must be completely correct.
- ***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)***: *Which is the Fair Sex? Gender Differences in Altruism*

(a) [3 pts] With available data, the standard deviation of total money passed by males is 16.44741, as reported in **Summary of total money passed, by sex**. Suppose Prof. Murdock had repeated this experiment in 2016/17 to obtain an even larger overall sample size. You should *expect* the standard deviation of total money passed by males to go up, go down, or be unchanged? Answer with two words. [No partial credit]

(b) [10 pts] Of the 534 females, 34 passed nothing to her partner in all eight decisions (`tot_mon_pass` equals 0), while, of the 334 males, 50 passed nothing. Is this difference statistically significant? If so, at which significance levels? Is the point estimate of the difference economically significant? Answer with formal hypotheses in standard notation, a quantitative analysis, and 2 – 3 sentences.

(c) [8 pts] Consider **Summary of total money passed, by sex** and also consider **Regression #1**. Show how to obtain the same t statistic as in Regression #1 $(t = 0.75)$ using the *appropriate **difference in means** approach* instead of a regression approach. Answer with formal hypotheses in standard notation and a quantitative analysis.

(d) [6 pts] Consider **Summary of total money passed, by sex** and also consider **Regression #2**. Show how to obtain the same t statistic as in Regression #2 $(t = 0.81)$ using the *appropriate **difference in means** approach* instead of a regression approach. Answer with a quantitative analysis.

(e) [9 pts] Consider **Summary of total money passed, by year**. Does student altruism differ among the three years? Write down a suitable regression model, which includes an intercept. Define your variables. Next, provide the numeric point estimates of all coefficients, including the intercept. Answer with variable definitions, a formal model, and the OLS point estimates written as an equation.

(f) [3 pts] An F test of the specification in Part (e) yields a P-value of 0.2079. Is there a statistically significant difference in altruism among the three years? Consider all usual α's. Answer with 1 sentence. [No partial credit]

(2) See the ***Supplement for Question (2)***: *Alternatives for addressing an outlier (South Korea)*

(a) [8 pts] What does the SST measure in this particular context? Why is the SST equal for Specifications (1) and (3)? Why is the SST much smaller for Specification (2)? Make sure to explain in plain English and to clearly discuss the practical meaning *in this particular context*. Answer with 3 sentences.

(b) [8 pts] What does the R-squared measure in this particular context? Why is the R-squared extremely small for Specification (2)? Why is the R-squared much larger for Specification (3)? Make sure to explain in plain English and to clearly discuss the practical meaning *in this particular context*. Answer with 3 sentences.

(3) [8 pts] Consider the model: $wage_i = \alpha + \beta educ_i + \delta female_i + \gamma female_i * educ_i + \varepsilon_i$. Suppose an R-squared of 0.0881 is obtained using data for 100 people. Are the results statistically significant overall? If so, at which significance levels? Answer with formal hypotheses in standard notation, a quantitative analysis, and 1 sentence.

(4) See the ***Supplement for Question (4)***: *Waterloo Salary Anomaly Working Group: Analysis & Findings*

(a) [8 pts] Suppose that, on average, male faculty members (n = 827) make $22,000 more than female faculty members (n = 344) at Waterloo and the standard error of the difference is $1,000. Explain *why* these results would *not* contradict the coefficient of $2,905 reported in ***Specification (2)***. Give TWO plausible and specific explanations for the difference between $22,000 and $2,905. Answer with 3 – 4 sentences.

(b) [5 pts] Consider the question of whether the coefficient on *Number of previous Outstanding Performance Awards* in ***Specification (2)*** is statistically greater than 2,000. What are the hypotheses? What is the P-value? Answer with formal hypotheses in standard notation and a quantitative analysis.

(c) [5 pts] Using ***Specification (2)***, what is the predicted annual salary for: a male with the rank of Assistant Professor in 2015, a mean annual merit score of 1.75, one Outstanding Performance Award, hired in 2010 as an Assistant Professor in the School of Computer Science, and who earned his PhD in 2010? Note that this question simply asks for a point prediction (not an interval). Answer with a quantitative analysis.

(d) [4 pts] Using ***Specification (2)***, find the point estimates of the slope between annual salary and years since hire for two different values: 5 years since hire and 35 years since hire. Answer with two numbers and a quantitative analysis.

(e) [10 pts] Using ***Specification (2)***, what is the 90% Confidence Interval Estimate of the coefficient on *Male*? How should the interval be interpreted? Is the interval wide or narrow? Offer a full interpretation in plain English and clearly discuss the meaning of the interval in practical terms. <u>Answer with a quantitative analysis and 2 – 3 sentences.</u>

(f) [5 pts] The authors also ran the regressions using the natural log of annual salary rather than annual salary (but do not report those results). What is a reasonable *rough estimate* of what the coefficient on *Male* would be if annual salary were logged? In other words, roughly what would b1 be in ln(salary)-hat = b0 + b1*Male + … ? Explain and support your estimate with relevant evidence from the *Supplement*. <u>Answer with a number and 1 – 2 sentences.</u>

This *Supplement* contains the aid sheets (formula sheets and Normal, t and F tables) as well as graphs, tables, and other information needed to answer some of the test questions. For each question directing you to this *Supplement*, make sure to carefully review all relevant materials. Remember, only your answers written on the test papers (in the designated space immediately after each question) will be graded. Any writing on this *Supplement* will *not* be graded.

THE FORMULA SHEETS AND NORMAL, t AND F TABLES ARE ON PAGES 5 THROUGH 10 OF THIS SUPPLEMENT.

Supplement for Question (1): Recall that on February 14, 2014, February 6, 2015, and on February 12, 2016 students in ECO220Y engaged in an experiment like the original participants from Andreoni and Vesterlund's "Which is the Fair Sex: Gender Differences in Altruism" published in 2001 and hereafter abbreviated as A&V (2001).

Each participant divided tokens between her/himself and another randomly selected participant in the room (whose identity would never be revealed). Each person made eight allocation decisions – budgets 1 through 8 shown below – where the number of tokens and the point values to each person (self and other) varied. Both A&V (2001) and ECO220Y (2014, 2015, 2016) *randomized* the order the eight budgets appeared to each participant on the decision sheet. Each point is worth $0.10 to all participants in all cases.

1. Divide 40 tokens: *Hold* _______ @ 1 point each, and *Pass* _______ @ 3 points each.
2. Divide 60 tokens: *Hold* _______ @ 1 point each, and *Pass* _______ @ 2 points each.
3. Divide 75 tokens: *Hold* _______ @ 1 point each, and *Pass* _______ @ 2 points each.
4. Divide 60 tokens: *Hold* _______ @ 1 point each, and *Pass* _______ @ 1 point each.
5. Divide 100 tokens: *Hold* _______ @ 1 point each, and *Pass* _______ @ 1 point each.
6. Divide 60 tokens: *Hold* _______ @ 2 points each, and *Pass* _______ @ 1 point each.
7. Divide 75 tokens: *Hold* _______ @ 2 points each, and *Pass* _______ @ 1 point each.
8. Divide 40 tokens: *Hold* _______ @ 3 points each, and *Pass* _______ @ 1 point each.

We attempted to replicate the original study. One difference is that rather than pay everyone for one randomly selected budget as A&V (2001) did using a research grant, ECO220Y (2014, 2015, 2016) paid randomly selected participants using money students donated right before the session and $20.00 per session (total of $220.00) donated by Prof. Murdock. Nearly all students voluntarily donated $2.00 to a collection jar as suggested by Prof. Murdock. ECO220Y (2014, 2015, 2016) used data from eleven sessions spread over February 14, 2014 (three sessions), February 6, 2015 (five sessions), and on February 12, 2016 (three sessions): 868 participated (334 males and 534 females). Because each participant made eight allocation decisions, 6,944 allocation decisions are observed. Consider the *total* money (CAN dollars) passed to the (anonymous) partner across all eight decisions: `tot_mon_pass`. Also, `male` is a dummy variable equal to one if the participant is male.

Summary of total money passed, by sex:

```
-> male = 0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
tot_mon_pass |        534    26.18446    12.23994          0       60.5


-> male = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
tot_mon_pass |        334    26.97126    16.44741          0       72.5
```

Supplement for Question (1) continues on next page >>>>>

Supplement for Question (1), cont'd:

Regression #1:

```
. regress tot_mon_pass male, robust;

Linear regression                                      Number of obs =     868
                                                       F(  1,   866) =    0.57
                                                       Prob > F      =  0.4513
                                                       R-squared     =  0.0007
                                                       Root MSE      =  14.008

------------------------------------------------------------------------------
             |               Robust
tot_mon_pass |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .7868006   1.044054     0.75   0.451    -1.262371    2.835972
       _cons |   26.18446   .5297886    49.42   0.000     25.14464    27.22428
------------------------------------------------------------------------------
```

Regression #2:

```
. regress tot_money_passed male;

      Source |       SS       df       MS              Number of obs =     868
-------------+------------------------------           F(  1,   866) =    0.65
       Model |  127.202988     1  127.202988           Prob > F      =  0.4210
    Residual |  169934.285   866  196.228967           R-squared     =  0.0007
-------------+------------------------------           Adj R-squared = -0.0004
       Total |  170061.488   867  196.149352           Root MSE      =  14.008

------------------------------------------------------------------------------
tot_mon_pass |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   .7868006   .9772317     0.81   0.421    -1.131219     2.70482
       _cons |   26.18446    .606193    43.19   0.000     24.99468    27.37424
------------------------------------------------------------------------------
```

Summary of total money passed, by year:

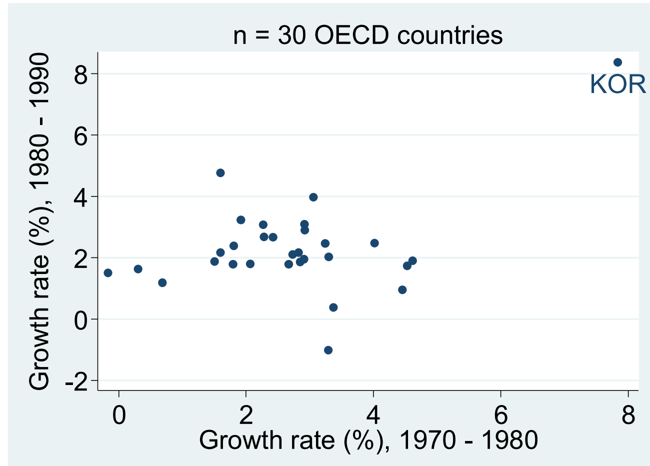
```
-> data_source = ECO220Y, Feb. 14, 2014

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
tot_mon_pass |       200      27.691    13.44333          0       72.5

-> data_source = ECO220Y, Feb. 6, 2015

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
tot_mon_pass |       461    26.52842    14.61084          0       72.5

-> data_source = ECO220Y, Feb. 12, 2016

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
tot_mon_pass |       207    25.23237    13.08705          0         53
```

Supplement for Question (2): Recall the 2014 *NBER* working paper “Asiaphoria Meets Regression to the Mean” (http://www.nber.org/papers/w20573.pdf). Test #1, Nov. 2016 provided a table “Table 1: Little persistence in cross-national growth rates across decades” and a scatter diagram. Recall the dot labeled “KOR,” which is South Korea. South Korea is an outlier and the table below shows three possibilities: (1) keep it, (2) drop it, or (3) put a dummy in for it.

Three Regression Specifications
(Dependent variable: *Growth rate (%), 1980 – 1990*)

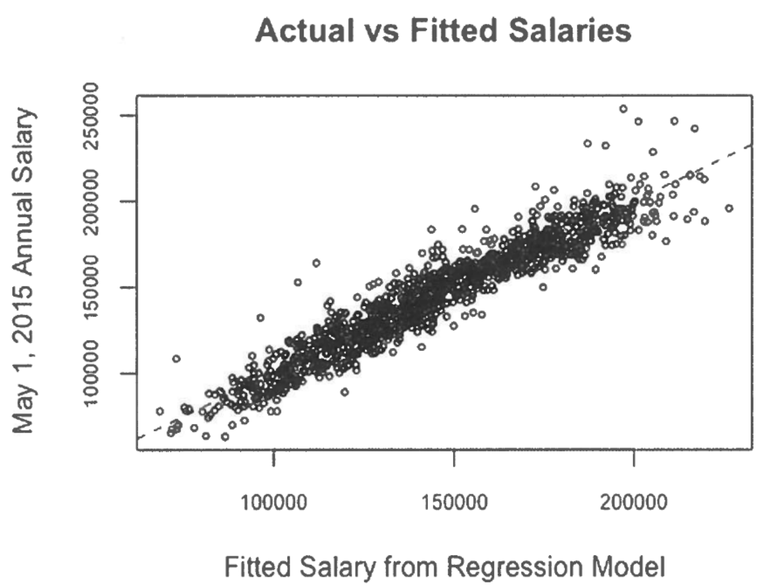
	(1)	(2)	(3)
Growth rate (%), 1970 – 1980	0.4326 (0.1772)	-0.1025 (0.1744)	-0.1025 (0.1744)
KOR	-	-	6.7889 (1.4287)
Intercept	1.1511 (0.5479)	2.3803 (0.4863)	2.3803 (0.4863)
Number of observations	30	29	30
R-squared	0.1755	0.0126	0.5510
SST	69.1815	31.4610	69.1815
SSR	12.1397	0.3975	38.1179
SSE	57.0418	31.0635	31.0635

Note: Coefficient estimates with standard errors in parentheses.



Supplement for Question (4): Recall the May 26, 2016 “Salary Anomaly Working Group: Analysis & Findings” report and appendix. It analyzes 1,171 Waterloo faculty members who completed an annual performance review on May 1, 2015. This includes 344 females and 827 males. The next page reproduces the regression results – Specifications (1) and (2) – with clarifications in places. The bullet list below describes some of the variables, with others described with the regression results. Categorical (nominal) information is included with dummy variables: the names clearly indicate when the value is 1. The omitted (aka reference) category for each set of dummies is specified in the bullet list. Also, the figure named “Actual vs Fitted Salaries” is reproduced from the original report and it corresponds to Specification (1).

- *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.
- *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.

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

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

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

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:** $s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(n-1)}$

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

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

Multiplication rule: $P(A\ and\ B) = P(A|B)P(B) = P(B|A)P(A)$

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>:

$V[c] = 0$

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

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

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

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

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>:

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

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

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

<u>Inference about a population proportion</u>:

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

<u>Inference about comparing two population proportions</u>:

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

<u>Inference about the population mean</u>:

***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:** $\nu = n - 1$

<u>Inference about a comparing two population means, independent samples, unequal variances</u>:

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

<u>Inference about a comparing two population means, independent samples, assuming equal variances</u>:

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

<u>Inference about a comparing two population means, paired data:</u> (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:** $\nu = n - 1$

<u>SIMPLE REGRESSION</u>:

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):

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

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

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

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

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:

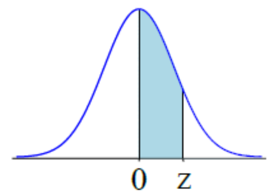
$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{(SST-SSE)/k}{SSE/(n-k-1)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

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:

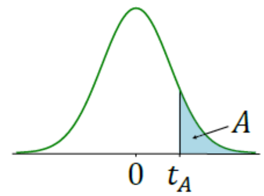
***t* test statistic:** $t = \frac{b_j - \beta_{j0}}{s_{b_j}}$ **CI estimator:** $b_j \pm t_{\alpha/2} s_{b_j}$ **Degrees of freedom:** $\nu = n - k - 1$

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



Normal Probabilities:

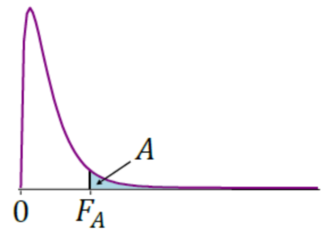
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	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 t:

ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657	38	1.304	1.686	2.024	2.429	2.712
2	1.886	2.920	4.303	6.965	9.925	39	1.304	1.685	2.023	2.426	2.708
3	1.638	2.353	3.182	4.541	5.841	40	1.303	1.684	2.021	2.423	2.704
4	1.533	2.132	2.776	3.747	4.604	41	1.303	1.683	2.020	2.421	2.701
5	1.476	2.015	2.571	3.365	4.032	42	1.302	1.682	2.018	2.418	2.698
6	1.440	1.943	2.447	3.143	3.707	43	1.302	1.681	2.017	2.416	2.695
7	1.415	1.895	2.365	2.998	3.499	44	1.301	1.680	2.015	2.414	2.692
8	1.397	1.860	2.306	2.896	3.355	45	1.301	1.679	2.014	2.412	2.690
9	1.383	1.833	2.262	2.821	3.250	46	1.300	1.679	2.013	2.410	2.687
10	1.372	1.812	2.228	2.764	3.169	47	1.300	1.678	2.012	2.408	2.685
11	1.363	1.796	2.201	2.718	3.106	48	1.299	1.677	2.011	2.407	2.682
12	1.356	1.782	2.179	2.681	3.055	49	1.299	1.677	2.010	2.405	2.680
13	1.350	1.771	2.160	2.650	3.012	50	1.299	1.676	2.009	2.403	2.678
14	1.345	1.761	2.145	2.624	2.977	51	1.298	1.675	2.008	2.402	2.676
15	1.341	1.753	2.131	2.602	2.947	52	1.298	1.675	2.007	2.400	2.674
16	1.337	1.746	2.120	2.583	2.921	53	1.298	1.674	2.006	2.399	2.672
17	1.333	1.740	2.110	2.567	2.898	54	1.297	1.674	2.005	2.397	2.670
18	1.330	1.734	2.101	2.552	2.878	55	1.297	1.673	2.004	2.396	2.668
19	1.328	1.729	2.093	2.539	2.861	60	1.296	1.671	2.000	2.390	2.660
20	1.325	1.725	2.086	2.528	2.845	65	1.295	1.669	1.997	2.385	2.654
21	1.323	1.721	2.080	2.518	2.831	70	1.294	1.667	1.994	2.381	2.648
22	1.321	1.717	2.074	2.508	2.819	75	1.293	1.665	1.992	2.377	2.643
23	1.319	1.714	2.069	2.500	2.807	80	1.292	1.664	1.990	2.374	2.639
24	1.318	1.711	2.064	2.492	2.797	90	1.291	1.662	1.987	2.368	2.632
25	1.316	1.708	2.060	2.485	2.787	100	1.290	1.660	1.984	2.364	2.626
26	1.315	1.706	2.056	2.479	2.779	120	1.289	1.658	1.980	2.358	2.617
27	1.314	1.703	2.052	2.473	2.771	140	1.288	1.656	1.977	2.353	2.611
28	1.313	1.701	2.048	2.467	2.763	160	1.287	1.654	1.975	2.350	2.607
29	1.311	1.699	2.045	2.462	2.756	180	1.286	1.653	1.973	2.347	2.603
30	1.310	1.697	2.042	2.457	2.750	200	1.286	1.653	1.972	2.345	2.601
31	1.309	1.696	2.040	2.453	2.744	250	1.285	1.651	1.969	2.341	2.596
32	1.309	1.694	2.037	2.449	2.738	300	1.284	1.650	1.968	2.339	2.592
33	1.308	1.692	2.035	2.445	2.733	400	1.284	1.649	1.966	2.336	2.588
34	1.307	1.691	2.032	2.441	2.728	500	1.283	1.648	1.965	2.334	2.586
35	1.306	1.690	2.030	2.438	2.724	750	1.283	1.647	1.963	2.331	2.582
36	1.306	1.688	2.028	2.434	2.719	1000	1.282	1.646	1.962	2.330	2.581
37	1.305	1.687	2.026	2.431	2.715	∞	1.282	1.645	1.960	2.326	2.576

Degrees of freedom: ν



Critical Values of F: $A = 0.10$

	ν_1											
ν_2	1	2	3	4	5	6	7	8	9	10	11	12
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
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
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
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
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
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
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
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
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55

Critical Values of F: $A = 0.05$

	ν_1											
ν_2	1	2	3	4	5	6	7	8	9	10	11	12
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
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
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
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
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
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
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
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
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75

Critical Values of F: $A = 0.01$

	ν_1											
ν_2	1	2	3	4	5	6	7	8	9	10	11	12
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71
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
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
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
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
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
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
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

Numerator degrees of freedom: ν_1; Denominator degrees of freedom: ν_2