UNIVERSITY OF TORONTO
Faculty of Arts and Science

APRIL 2013 EXAMINATIONS

ECO220Y1Y

PART 1 OF 2

Duration - 3 hours

Examination Aids: Non-programmable calculator

This exam has two parts and a BUBBLE FORM. You may detach the aid sheets and statistical tables (Standard Normal, Student t and F) stapled to PART 1. You are responsible for turning in the BUBBLE FORM and all 8 pages of PART 2. Complete both the BUBBLE FORM and PART 2 before the end of the exam, including entering your name. In total there are 120 possible points.

Unless otherwise specified, use the conventional 5 percent significance level.

PART 1: 31 multiple choice questions with point values from 1 to 3 points each for a total of 72 points

- Questions with 2 alternatives, (A) – (B), are worth: 1 point each correct answer
- Questions with 3 or 4 alternatives, (A) – (C)/(D), are worth: 2 points each correct answer
- Questions with 5 alternatives, (A) – (E), are worth: 3 points each correct answer
- Answers must be properly recorded on the U of T bubble form to earn marks
- On the FRONT of the BUBBLE FORM: Print your 9 (or 10) digit student number in the boxes AND darken each number in the corresponding bracket below each box. Print your last name and initial in the boxes AND darken each letter in the corresponding circle below each box. Write in the other requested information in the upper left region of the form.
- On the BACK of the BUBBLE FORM: Write in your name, sign, and record your answers.
- Use a pencil and make dark solid marks that fill the bubble completely
- Erase completely any marks you want to change; Crossing out a marked box is incorrect
- Choose the best answer for each question
- If more than one answer is selected then that question earns 0 points
- For questions with numeric answers that require rounding, round your final answer to be consistent with the choices offered. Use standard rounding rules

PART 2: 4 written questions with varying point values worth a total of 48 points

(1) Campaign volunteers periodically contact a random sample of the electorate to assess what fraction will vote for their candidate. Two weeks before the election the volunteers double the number of people contacted. What would be a good explanation for this?

(A) the volunteers wish to reduce potential non-response bias
(B) the volunteers wish to reduce the variability of peoples' replies
(C) the volunteers wish to reduce the standard deviation of the sample proportion
(D) All of the above

(2) When should we expect that a random sample will be Normally distributed?

(A) when the population is Normally distributed
(B) when the sample size is sufficiently large so that the Central Limit Theorem applies
(C) All of the above

(3) Variables measuring income (1000's of dollars) and education (years) are each *standardized*. A simple regression yields: $\widehat{std\_income}_i = 0.44 * std\_education_i$. What does the slope mean?

(A) it the coefficient of determination (R^2)
(B) it is the coefficient of correlation between income and education
(C) that 44 percent of the variation in income is explained by variation in education
(D) that on average an extra year of education is associated with 440 dollars of extra income

► **Questions (4) – (5):** A gambling casino may be built in Toronto. Toronto residents are divided over this issue. Consider this joint probability distribution.

	Support the casino	Against the casino
Unionized worker	0.21	0.07
Everyone else	0.26	0.46

(4) A randomly selected Toronto resident speaks to a pollster. If that person indicates being a unionized worker then what is the chance that that person supports the casino?

(A) 0.21
(B) 0.45
(C) 0.47
(D) 0.75

(5) Two Toronto residents are randomly selected. What is the chance that one supports the casino and one is against the casino?

(A) 0.2213
(B) 0.2491
(C) 0.4426
(D) 0.4982

(6) X is Normally distributed with mean 65 and standard deviation 10. P(50 < X < 60) = ___.

(A) 0.1915
(B) 0.2108
(C) 0.2417
(D) 0.2583
(E) 0.3085

► **Questions (7) – (8):** During winter the coefficient of correlation for monthly snowfall in the north versus south sides of a city is 0.8. The table to the right shows the other parameters. The city's snow removal costs are $120 per cm for the north side and $100 per cm for the south side and $5,000 per month regardless of snowfall.

	Monthly Snowfall (cm)	
	mean	s.d.
north side	80	25
south side	90	30

(7) What is the expected total cost of snow removal per month in the city?

(A) $5,220
(B) $18,800
(C) $23,600
(D) $41,900

(8) What is the standard deviation of the total cost of snow removal per month in the city?

(A) $3,978
(B) $4,243
(C) $5,692
(D) $6,000
(E) $6,050

► **Questions (9) – (10):** The table to the right is an excerpt from Statistics Canada's website (Table 202-01021), which is based on comprehensive data for all earners in Canada.

Earnings	2010
Average earnings, females (dollars)	31,700
Median earnings, females (dollars)	23,900

(9) Median earnings are lower than average earnings because the population is ___.

(A) skewed to the left (negatively skewed)
(B) skewed to the right (positively skewed)

(10) With a random sample of 500 females the sampling distribution of $\bar{X}$ would ___.

(A) be centered at $31,700
(B) follow the Normal distribution (be Bell shaped)
(C) have a s.d. equal to the population s.d. divided by the square-root of 500
(D) All of the above

(11) In testing H_0: p = 0.5 versus H_1: p < 0.5 what would a P-value of 0.02 imply?

(A) The power of the test is 0.02
(B) The sample proportion is below 0.5
(C) The probability that the result is statistically significant is 0.02
(D) There is very little evidence in favor of the research hypothesis
(E) The significance level (α) is 0.02 rather than the conventional 0.05

► **Questions (12) – (13):** For two firms in the same industry quarterly profits are related. Firm A's profits rise (Event A) with probability 0.8. Firm B's profits rise (Event B) with probability 0.6.

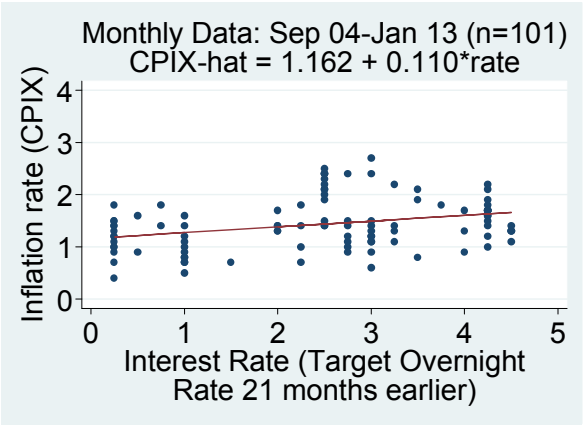
(12) Using only the given information which of these could you calculate?

(A) $P(A^C)$, which can also be written as P(A')
(B) P(A or B), which can also be written as $P(A \cup B)$
(C) P(A and B), which can also be written as $P(A \cap B)$
(D) All of the above

(13) Suppose that if Firm A's profits rise then the probability Firm B's profits rise is 0.7. If Firm B's profits do NOT rise then what is the probability that Firm A's profits do NOT rise?

(A) 0.2
(B) 0.3
(C) 0.4
(D) 0.5
(E) 0.6

► **Questions (14) – (19):** The Bank of Canada states: "At the heart of Canada's monetary policy framework is the inflation-control target. The target for inflation is the 2 per cent midpoint of a control range of 1 to 3 per cent. Inflation is measured as the year-over-year rate of increase in the total consumer price index (CPI). The Bank also monitors a set of "core" inflation measures, including the CPIX which strips out eight of the most volatile CPI components. The Bank carries out monetary policy through changes in its policy interest rate—the Target for the Overnight Rate. Monetary policy actions (changes in the policy rate) take time—usually between six and eight quarters—to work their way through the economy and to have their full effect on inflation." A student downloads the Bank's data obtaining the scatter diagram and OLS line shown to the right. http://www.bankofcanada.ca/wp-content/uploads/2010/11/monetary_policy.pdf



(14) What kind of data are these?

(A) time series
(B) cross-sectional

(15) In January 2013—the most recent observation—the interest rate (the Target Overnight Rate 21 months earlier) is 1 and the inflation rate is 0.5. What is the residual for this observation?

(A) -0.107
(B) -0.217
(C) -0.552
(D) -0.772

(16) Using visual approximation, what is standard deviation of the residuals?

(A) roughly 0.5
(B) roughly 2.0
(C) roughly 4.0
(D) roughly 6.0

(17) The standard error of the slope is 0.034. Is the slope statistically significant?

(A) no, not at any conventional significance level
(B) yes and the P-value is less than 0.01
(C) yes and the P-value is between 0.01 and 0.025
(D) yes and the P-value is between 0.025 and 0.05
(E) yes and the P-value is between 0.05 and 0.10

(18) Which is a correct interpretation of the slope of the regression line? (p.p. = percentage point)

(A) There is no statistically significant relationship between the interest rate and inflation
(B) Increasing the interest rate by 1 p.p. increases inflation 21 months later by 0.110 p.p.
(C) Increasing the interest rate by 0.110 p.p. increases inflation 21 months later by 1 p.p.
(D) Inflation is on average 1 p.p. higher 21 months after a 0.110 p.p. interest rate increase
(E) Inflation is on average 0.110 p.p. higher 21 months after a 1 p.p. interest rate increase

(19) If an observation with an inflation rate of 6 and interest rate of 0.25 (21 months earlier) were included in the regression analysis then the estimated slope of the regression line would be ___.

(A) unchanged
(B) a larger number
(C) a smaller number

► **Questions (20) – (26):** Many social scientists, including economists, now advocate evaluating economic performance using aggregate happiness in addition to GDP. In 2012 two Princeton sociologists published the article “The Pursuit of Happiness in China: Individualism, Collectivism, and Subjective Well-Being During China’s Economic and Social Transformation.” It uses multiple regression analysis to explore how individuals in China respond to survey questions regarding their own happiness and life satisfaction. “Happiness was measured via the question: ‘Taking all things together, would you say you are [very happy, quite happy, not very happy, or not at all happy]?’ Life

satisfaction was measured via the question: 'All things considered, how satisfied are you with your life as a whole these days,' measured on a scale from 1 to 10, with 1 representing 'dissatisfied' and 10 representing 'satisfied.' We combined these measures by standardizing them within each survey year and summing them [to form a new variable 'subjective well-being']." (p. 5) Pooled data combines the survey data from 1990, 2001 and 2007.

Table 1: Descriptive statistics for variables used in analysis (World Values Survey, 1990, 2001, 2007)

		1990 Data	2001 Data	2007 Data	Pooled Data
Variable	**Range**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Happiness	[1, 4]	2.94 (0.81)	2.91 (0.62)	3.05 (0.66)	2.97 (0.73)
Life satisfaction	[1, 10]	7.28 (2.10)	6.46 (2.49)	7.15 (2.17)	7.08 (2.22)
Subjective well-being	[-2.59, 1.39]	0.08 (0.76)	-0.11 (0.83)	0.13 (0.79)	0.06 (0.79)
Male (=1 if male, 0 otherwise)	[0, 1]	0.5995	0.5171	0.5213	0.5590
Age	[18, 65]	38.34 (12.82)	38.63 (11.48)	39.00 (12.55)	38.60 (12.48)
Married (=1 if married)	[0, 1]	0.7828	0.8229	0.7518	0.7809
Income	[1, 10]	3.22 (1.70)	6.98 (1.84)	4.49 (1.79)	4.35 (2.25)
Employed (=1 if employed)	[0, 1]	0.8756	0.6857	0.7216	0.7903
Self-rated health	[1, 4]	2.82 (0.94)	2.78 (0.94)	3.06 (0.90)	2.89 (0.93)
Freedom of choice/control	[1, 10]	7.07 (2.12)	6.84 (2.49)	7.46 (2.14)	7.15 (2.21)
National pride	[1, 4]	3.23 (0.78)	3.06 (0.75)	3.07 (0.71)	3.15 (0.76)
Support for collectivism	[1, 10]	3.57 (1.58)	4.44 (1.75)	4.67 (1.72)	4.08 (1.74)
Number of respondents		884	350	564	1,798

Table 2: OLS regression models of subjective well-being

	Coefficient Estimates (Standard Errors)			
	1990	**2001**	**2007**	**Pooled**
Married	0.028 (0.078)	0.280 (0.128)	0.305 (0.075)	0.182 (0.051)
Income	-0.013 (0.014)	0.057 (0.021)	0.111 (0.015)	0.042 (0.009)
Employed	0.019 (0.087)	0.078 (0.093)	0.127 (0.064)	0.119 (0.045)
Self-rated health	0.138 (0.026)	0.129 (0.043)	0.329 (0.031)	0.198 (0.018)
Freedom of choice/control	0.096 (0.011)	0.121 (0.017)	0.091 (0.013)	0.101 (0.008)
National pride	0.210 (0.031)	0.124 (0.053)	0.141 (0.036)	0.179 (0.022)
Support for collectivism	-0.039 (0.015)	-0.029 (0.023)	-0.014 (0.016)	-0.035 (0.010)
Male	-0.030 (0.048)	-0.131 (0.078)	-0.067 (0.053)	-0.077 (0.033)
Age	0.020 (0.016)	-0.053 (0.026)	-0.033 (0.015)	-0.018 (0.010)
Age-squared	-0.0001 (0.0002)	0.0012 (0.0004)	0.0004 (0.0001)	0.0003 (0.0001)
Year 2001 dummy				-0.240 (0.057)
Year 2007 dummy				-0.005 (0.041)
Constant	-1.949 (0.303)	-1.101 (0.543)	-2.162 (0.341)	-1.681 (0.207)
Observations	884	350	564	1,798
R^2	0.234	0.316	0.423	0.286

(20) Across individuals in these data the responses to the question on "National pride" are ___ variable than the responses to the question on "Self-rated health."

(A) less
(B) more

(21) For which year of data is there *insufficient* statistical evidence to infer that there is a non-linear association between a person's age and subjective well-being?

(A) 1990
(B) 2001
(C) 2007
(D) All of the above

(22) Which variable has a statistically significant coefficient in all four of the regressions (1990, 2001, 2007, and pooled data)?

(A) male
(B) income
(C) married
(D) employed
(E) national pride

(23) To test the overall statistical significance of the 1990 regression model the test statistic is ___.

(A) 22.8
(B) 23.4
(C) 24.2
(D) 25.9
(E) 26.7

(24) Based on the results in Table 2 what do we know about the coefficient of correlation between subjective well-being and "Freedom of choice/control" for 1990, 2001, and 2007?

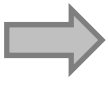
(A) it may be positive or negative
(B) it is positive for all three years
(C) it is negative for all three years

(25) Using the pooled regression results we ___ conclude that males generally have a lower subjective well-being than females in these data.

(A) can
(B) cannot

(26) Is the pooled regression model statistically significant overall?

(A) yes and the P-value is equal to 0.05
(B) yes and the P-value is less than 0.01
(C) yes and the P-value is between 0.01 and 0.05
(D) no, not at a 5% significance level but it is significant at the 10% level
(E) no, because the P-value is too large at any conventional significance level



(27) Using a random sample of 200 premium cable subscribers an analyst correctly calculates a 99 percent confidence interval estimate of the mean income in $1000s as (59, 68). How should this interval be interpreted? We are 99% confident that ___ between $59,000 to $68,000.

(A) the population mean income of all premium cable subscribers is
(B) a randomly selected premium cable subscriber would have an income
(C) the sample mean of a random sample of 200 premium cable subscribers will fall
(D) the mean income of the 200 premium cable subscribers will be captured by the interval
(E) All of the above are valid interpretations

(28) An analyst wants to show the mean change from last year in customer satisfaction is different from zero. With a random sample of 10 observations the test statistic is 2.07. What is the P-value?

(A) less than 0.005
(B) between 0.005 and 0.01
(C) between 0.01 and 0.025
(D) between 0.025 and 0.05
(E) between 0.05 and 0.10

(29) Consider a 5 percent significance level and a random sample of 18 observations. To infer a research hypothesis that the population mean is greater than 50 requires a ___.

(A) sample mean greater than 50
(B) sample mean greater than 58.3
(C) sample mean greater than 59.2
(D) standardized test statistic greater than 1.645
(E) standardized test statistic greater than 1.740

(30) Consider a 1 percent significance level and a random sample of 41 observations. To infer a research hypothesis that the population proportion is greater than 0.60 requires a ___.

(A) sample proportion greater than 0.600
(B) sample proportion greater than 0.778
(C) sample proportion greater than 0.791
(D) standardized test statistic greater than 1.960
(E) standardized test statistic greater than 2.423

(31) A random sample of female Toronto residents is proposed. The goal is to make an inference about the proportion that support a proposed casino. If amongst *all* female Toronto residents 37 percent support the casino, which hypothesis test would have the highest power?

(A) H_0: $p = 0.20$; H_1: $p > 0.20$
(B) H_0: $p = 0.30$; H_1: $p > 0.30$
(C) H_0: $p = 0.30$; H_1: $p \neq 0.30$
(D) H_0: $p = 0.38$; H_1: $p \neq 0.38$
(E) H_0: $p = 0.40$; H_1: $p \neq 0.40$

UNIVERSITY OF TORONTO
Faculty of Arts and Science

APRIL 2013 EXAMINATIONS

ECO220Y1Y

PART 2 OF 2

Duration - 3 hours

Examination Aids: Non-programmable calculator

Last Name:

First Name:

Student #:

This exam has two parts and a BUBBLE FORM. You may detach the aid sheets and statistical tables (Standard Normal, Student t and F) stapled to PART 1. You are responsible for turning in the BUBBLE FORM and all 8 pages of PART 2. Complete both the BUBBLE FORM and PART 2 before the end of the exam, including entering your name. In total there are 120 possible points.

Unless otherwise specified, use the conventional 5 percent significance level.

PART 1: 31 multiple choice questions with point values from 1 to 3 points each for a total of 72 points

PART 2: 4 written questions with varying point values worth a total of 48 points

- Write your answers clearly, concisely, and completely below each question.
- ***Show your work; Make the reasons justifying your work very clear***.
- ***Fully label all graphs***.
- A guide for your response is in brackets: it tells what is expected: a quantitative analysis, a graph, and/or sentences. For example, "What would the least squares intercept be and how should it be interpreted? [Answer with an analysis & 1 – 2 sentences]" To best demonstrate your understanding, follow the guides and focus on directly answering the questions asked.

	Q1	Q2	Q3	Q4	PART 2 Total	PART 1 Total	Raw Total	Percent Mark
Point Value:	*10*	*14*	*12*	*12*	*48*	*72*	*120*	
Points Earned:								

(1) [10 pts] An online store initiates a guarantee of quick delivery. From past experience it believes laptop delivery time is normally distributed with a mean of 4 days and a standard deviation of 1 day.

(a) [4 pts] Suppose it guarantees delivery within 5 days and advertises that a customer gets the laptop free if it is late. Each laptop cost the company $850 and it sells each for $1000. What would be the average profit per laptop under this guarantee policy? [Answer with a quantitative analysis]

(b) [6 pts] Draw a *fully-labeled* graph of the sampling distribution of the sample mean delivery time for a sample of 16 deliveries. Calculate the probability that the mean for a sample of 16 deliveries will exceed 4.5 days and illustrate it on the graph. [Answer with a graph and a quantitative analysis]

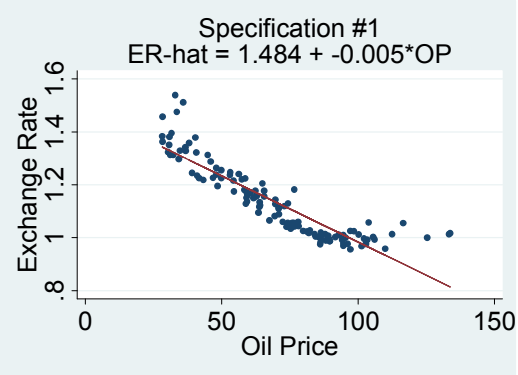
(2) [14 pts] For a pharmaceutical company to establish that its new drug is effective it must show that at least 66% of patients benefit. For a random sample of 142 patients, 97 are found to benefit.

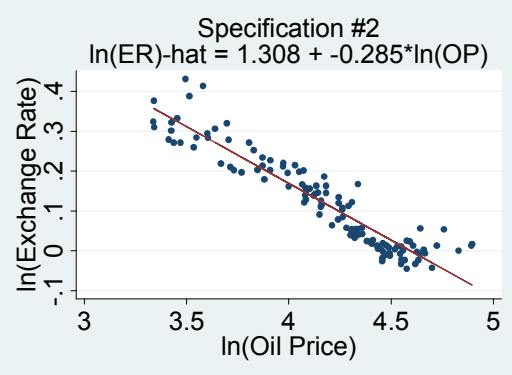
(a) [2 pts] Using notation from our course, write the null and research hypotheses. [Answer with two hypotheses]

(b) [6 pts] Is there sufficient statistical support to conclude that the drug is effective? Show your work and interpret your result with one sentence. [Answer with a quantitative analysis and 1 sentence]

(c) [6 pts] Supposing that the drug would benefit 75% of all eligible patients, compute the probability of a Type II error. [Answer with a quantitative analysis]

(3) [12 pts] Because oil production is an important part of the Canadian economy, some analysts believe that the value of the Canadian dollar is related to the price for oil. Consider data from the Bank of Canada and the St. Louis Fed: monthly observations from January 2003 through December 2012 of the US Canada exchange rate and the spot price of oil (US$/barrel). On the next page are two scatter diagrams and least square lines: ER=exchange rate and OP=oil price. In Specification #2 the variables have been transformed (re-expressed) by taking the natural logarithm of each.





(a) [4 pts] Does Specification #1 appear to violate the linearity condition? How about Specification #2? Explain and also state what this condition means. [Answer with 2 – 3 sentences]

(b) [4 pts] Does Specification #1 appear to violate the equal spread condition (also known as the equal variance condition or the homoscedasticity assumption)? How about Specification #2? Explain and also state what this condition means. [Answer with 2 – 3 sentences]

(c) [4 pts] For either specification is autocorrelation *likely* to be a concern? Explain and also state what autocorrelation means. In any case, how would you check for it? [Answer with 2 – 3 sentences]

(4) [12 pts] A shoe factory does an experiment to assess the relationship between hourly wages and productivity. 125 workers are each randomly assigned a wage and his/her productivity measured. A regression with productivity as the dependent variable yields the results below and an R^2 of 0.4828.

Variable	Definition
productivity	number of shoes a worker made per hour during the experiment period
wage	hourly wage ($'s) the worker received during the experiment period
female	dummy variable that equals 1 if the worker is female and 0 if the worker is male
education	worker's number of years of schooling (ranges from 10 to 14 for these workers)

	Coeff	SE(Coeff)	t-value	P-value
intercept	1.9941030	0.3128457	6.37	<0.001
wage	0.1258777	0.0153936	8.18	<0.001
wage*female	-0.0484462	0.0232552	-2.08	0.039
female	0.6550939	0.3274322	2.00	0.048
education	0.0944264	0.0185516	5.09	<0.001

(a) [4 pts] Is this regression statistically significant overall? Explain. [Answer with a quantitative analysis and 1 sentence]

(b) [4 pts] Fully *interpret* the coefficient on the variable `wage`. Include a comment on its statistical significance. [Answer with 2 – 3 sentences]

(c) [4 pts] Fully *interpret* the coefficient on the variable `wage*female`. Include a comment on its statistical significance. [Answer with 3 – 4 sentences]

EXTRA SPACE: If you use this space, *clearly indicate the question number and part and make a clear note in the original space directing the grader here*.