While you wait for the start of this test you may fill in the FRONT AND BACK of the BUBBLE FORM and read this cover BUT please keep this test paper face up and flat on your desk.

Instructor: Prof. Murdock

Duration: 50 minutes. You MUST STAY for the entire 50 minutes: you cannot leave early.

Allowed aids: A non-programmable calculator; Aid sheets provided with this test.

Format: This test includes these question papers and a BUBBLE FORM. There are 24 multiple choice questions with point values from 1 to 3 points each for a total of 46 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

Once the test begins you may detach the aid sheets and statistical tables (Standard Normal, Student t and F) stapled to this test. Complete the BUBBLE FORM before the end of the test, including entering your name.

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

Instructions:

- 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 circle 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.
- **Your FORM CODE is A.**
- 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

(1) Which of these describe time series data? Data recording ___.

(A) GDP growth and income inequality for 120 different countries (n = 120)
(B) price per liter of gasoline and quantity sold at 40 different gas stations (n = 40)
(C) weight at age 18 and risk of diabetes at age 60 for 240 different people (n = 240)
(D) tuition and average class size at a university for 30 different academic years (n = 30)

(2) Each day an analyst records the posted online price (in 100s of dollars) of a flight from Toronto to Calgary and the number of tickets sold. Noting a non-linear relationship the natural log of both variables is taken. A regression analysis yields: ln(Q)-hat = $b_0 + b_1\ln(P)$. Why is it *incorrect* to interpret the coefficient b_1 as the elasticity of demand for flights from Toronto to Calgary?

(A) because the analyst has violated the linearity condition
(B) because variables such as day of week affect both the price and quantity
(C) because the analyst has measured price in 100s of dollars instead of dollars
(D) because b_1 reflects the slope: the average change in Q associated with a change in P
(E) because b_1 measures the coefficient of correlation between ln(Q) and ln(P), not elasticity

(3) An analyst performs a simple regression of the number of employees on the age (in years) of a firm for a random sample of firms. On average how many employees are seven-year-old firms predicted to have? This question can be addressed by computing ___.

(A) a prediction interval with $x_g = 7$
(B) a confidence interval with $x_g = 7$

► **<u>Questions (4) – (5):</u>** Presuming the underlying conditions hold, consider these regression results where X and Y are measured in dollars and the standard errors are reported in parentheses:

$$\text{Y-hat} = \underset{(1.231)}{23.240} + \underset{(0.088)}{0.931}{*}X$$

(4) The P-value for the test of statistical significance of the slope is ___.

(A) very small: will meet significance levels well below the usual 5%
(B) somewhat small: will meet significance levels around the usual 5%
(C) somewhat large: will not meet significance levels below 10%
(D) very large: will not meet any reasonable significance level

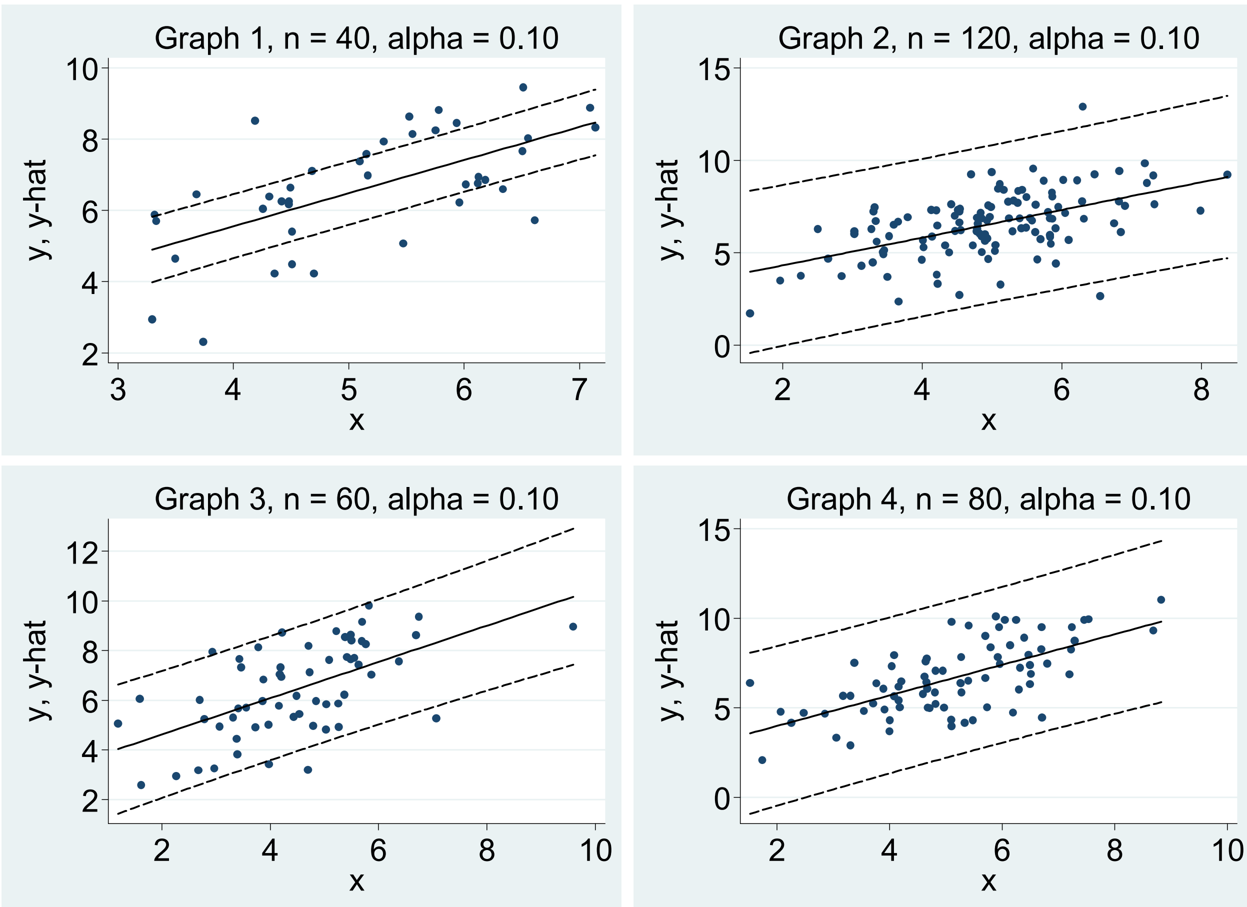
(5) The sample coefficient of correlation between Y and X is ___.

(A) weakly positive
(B) strongly positive
(C) positive but we cannot determine if it is weak or strong with the given information
(D) completely unknown: it may be either positive or negative and possibly weak or strong

(6) In a multiple regression analysis, which graphical technique can you use to check for autocorrelation of the error?

(A) a histogram of the residuals
(B) line graph of the residuals against time
(C) a series of line graphs with each explanatory variable against time
(D) a series of scatter diagrams with the dependent variable against each explanatory variable

► **Question (7):** Consider these graphs constructed (possibly incorrectly) by an analyst where the dashed bounds mark the 90% prediction intervals.



(7) Presuming that the regression lines themselves are correct in each graph, which of these graphs show correct prediction intervals?

(A) Only Graph 1
(B) Only Graph 2
(C) Only Graph 3
(D) Only Graph 4
(E) Both Graphs 2 and 4

► **Questions (8) – (14):** The 1991 publication "Economic Growth in a Cross Section of Countries" by Robert Barro in *The Quarterly Journal of Economics* reports many multiple regressions for a sample of 98 countries. Here are some results – Specifications (1), (3), and (4) – and the variable definitions. Standard errors of coefficient estimates appear in parentheses.

Table I: Regressions for per Capita Growth				**Definitions of Variables**
	(1)	**(3)**	**(4)**	
Dep. var.	GR6085	GR7085	GR7085	GR6085 (GR7085): Annual average growth rate of per capita real GDP from 1960 to 1985 (1970 to 1985)
No. obs.	98	98	98	
Const.	0.0302 (0.0066)	0.0287 (0.0080)	0.0294 (0.0082)	
GDP60	-0.0075 (0.0012)	-0.0089 (0.0016)	-0.0071 (0.0048)	GDP60: 1960 value of real per capita GDP in $1000's USD (1980 base year)
GDP70	-	-	-0.0015 (0.0037)	GDP70: 1970 value of real per capita GDP in $1000's USD (1980 base year)
SEC60	0.0305 (0.0079)	0.0331 (0.0137)	0.0350 (0.0128)	SEC60: 1960 secondary-school enrollment rate
PRIM60	0.0250 (0.0056)	0.0276 (0.0070)	0.0279 (0.0072)	PRIM60: 1960 primary-school enrollment rate
g^c/y	-0.119 (0.028)	-0.142 (0.034)	-0.147 (0.036)	g^c/y: Average from 1970 to 1985 of the ratio of real government consumption (exclusive of defense and education) to real GDP
REV	-0.0195 (0.0063)	-0.0236 (0.0071)	-0.0241 (0.0071)	REV: Number of revolutions or coups per year (1960 – 1985 or subsample)
ASSASS	-0.0333 (0.0155)	-0.0485 (0.0185)	-0.0490 (0.0188)	ASSASS: Number of assassinations per million population per year (1960 – 1985 or subsample)
PPI60DEV	-0.0143 (0.0053)	-0.0171 (0.0078)	-0.0174 (0.0079)	PPI60DEV: Deviation of the 1960 PPP value for the investment deflator (U.S. = 1) from the sample mean
R^2	0.56	0.49	0.50	
$\hat{\sigma}$	0.0128	0.0168	0.0169	

(8) These data are ___.

(A) time series and experimental
(B) time series and observational
(C) cross-sectional and experimental
(D) cross-sectional and observational

(9) In testing the joint significance of all of the coefficients what is the rejection region for Specification (4)? Rounding to the nearest integer, it is from ___.

(A) positive 2 to positive infinity
(B) positive 11 to positive infinity
(C) negative infinity to negative 2 and from positive 2 to positive infinity
(D) negative infinity to negative 11 and from positive 11 to positive infinity

(10) For Specification (1), which is a valid statement regarding the negative coefficient on GDP60?

(A) it is not statistically significant and hence it should not be interpreted
(B) GDP60 and GR6085 have a weak but statistically significant negative correlation
(C) countries with higher per capita GDP in 1960 grew more slowly from 1960 to 1985
(D) after controlling for some differences between countries, such as education and political stability, being richer is associated with slower growth
(E) on average countries with a high per capita GDP in 1960 saw their GDP grow more slowly during the period of these data compared to countries which started with lower per capita GDP

(11) For Specification (4) what is the P-value for the test of statistical significance for the coefficient on GDP60?

(A) less than 0.01
(B) between 0.01 and 0.05
(C) between 0.05 and 0.10
(D) between 0.10 and 0.20
(E) greater than 0.20

(12) GDP60 has a sample range from 0.208 to 7.380 (corresponding to $208 to $7,380 in 1980 USD). The sample range in per capita growth rates is -0.017 to 0.074, with a mean of 0.022. According to Specification (3), how would the predicted per capita growth rates compare for two countries with identical values for the other explanatory variables but where Country “Poor” has a per capita GDP of $208 in 1960 and Country “Rich” has a per capita GDP of $7,380 in 1960? Per capita GDP growth is predicted to be about ___ in Country Rich compared to Country Poor.

(A) 6.4 percentage points lower annually
(B) 6.4 percentage points higher annually
(C) 9.1 percentage points lower annually
(D) 9.1 percentage points higher annually

(13) In the previous question the requested prediction ___ an out of sample prediction.

(A) is
(B) is not

(14) The final row reports $\hat{\sigma}$, which in the notation of our course would be written s_e. What is the interpretation of this number in Specification (1)?

(A) the standard deviation GR6085 is 1.28 percentage points
(B) the standard deviation GR6085 is 0.0128 dollars (1980 USD)
(C) the standard deviation of the residuals is 1.28 percentage points
(D) the standard deviation of the residuals is 0.0128 dollars (1980 USD)

► **Questions (15) – (17):** Below is STATA output for a multiple regression analysis of the selling prices (in 1000s of dollars) of 519 homes based on the living area (square feet), number of bedrooms, number of bathrooms, number of fireplaces, age (in years), and size of the lot (square feet).

```
      Source |       SS       df       MS              Number of obs =     519
-------------+------------------------------           F(  6,   512) =  130.79
       Model |  2082015.68     6  347002.614           Prob > F      =  0.0000
    Residual |  1358400.86   512  2653.12668           R-squared     =  0.6052
-------------+------------------------------           Adj R-squared =  0.6005
       Total |  3440416.54   518  6641.73078           Root MSE      =  51.509

------------------------------------------------------------------------------
  price_1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  livingarea |   .0490309   .0073433     6.68   0.000     .0346042    .0634576
    bedrooms |  -10.12991   4.053272    -2.50   0.013      -18.093   -2.16682
   bathrooms |   22.87966   5.481045     4.17   0.000     12.11155    33.64776
  fireplaces |     10.383   4.762074     2.18   0.030     1.027388    19.73861
         age |   .0254905   .0795629     0.32   0.749    -.1308194    .1818004
     lotsize |   .0254734   .0043514     5.85   0.000     .0169245    .0340222
       _cons |  -20.26362   11.91661    -1.70   0.090     -43.6751    3.147856
------------------------------------------------------------------------------
```

(15) The sample mean house price is $168,387. What is the standard deviation of house prices?

(A) $63,398
(B) $63,769
(C) $81,497
(D) $81,973

(16) Because bigger houses tend to be on bigger properties the variables livingarea and lotsize have a significant positive correlation. How should this affect the interpretation of the coefficient on livingarea?

(A) it should not affect the interpretation
(B) we must be careful to say that we are holding everything else constant except lotsize
(C) we must be careful to say holding bedrooms, bathrooms, fireplaces, and age constant but not lotsize
(D) it means that the coefficient on livingarea is picking up both the effect of livingarea and lotsize on price

(17) What can we say about the coefficient on age?

(A) Age has zero effect on the selling price of a house
(B) Age has a weak positive association with the selling price of a house
(C) Age has no statistically significant relationship with the selling price of a house
(D) We have almost no evidence that age is related to price after we account for living area, lot size, bedrooms, bathrooms and fireplaces
(E) Once we control for all factors that affect the selling price of a house, age has almost no effect and certainly not a statistically significant effect

► **Questions (18) – (23):** Consider a multiple regression analysis with standardized variables. Assess whether each statement describes a situation that could arise versus one that could not arise. For all questions consider an analyst working with real data, using a fairly large sample size, and making reasonable modeling decisions with no significant violations of the underlying conditions.

(18) The coefficient on a predictor variable is large in magnitude but the coefficient is not statistically different from zero.

(A) this situation could arise
(B) this situation could not arise

(19) The adjusted R^2 is a little lower than the R^2.

(A) this situation could arise
(B) this situation could not arise

(20) When testing a positive coefficient the analyst can reject the null hypothesis $H_0\colon\ \beta_j = 0$ at a 1% significance level if the research hypothesis is $H_1\colon\ \beta_j \neq 0$ but not if it is $H_1\colon\ \beta_j > 0$.

(A) this situation could arise
(B) this situation could not arise

(21) At least one of the coefficients is statistically significant at the 1 percent level but the multiple regression model overall is not statistically significant at the 5 percent level.

(A) this situation could arise
(B) this situation could not arise

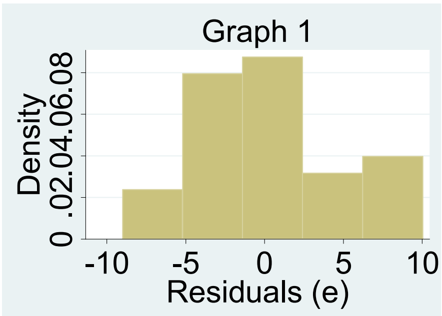
(22) A predictor variable is strongly correlated with the dependent variable—with a coefficient of correlation that is both large and statistically significant—but the multiple regression coefficient on this variable is near zero and statistically insignificant.

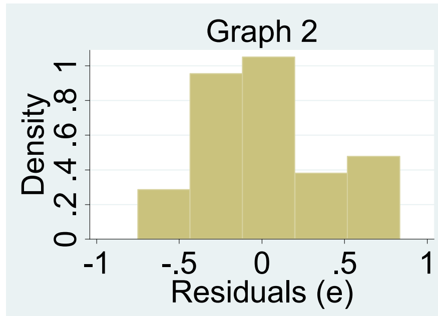
(A) this situation could arise
(B) this situation could not arise

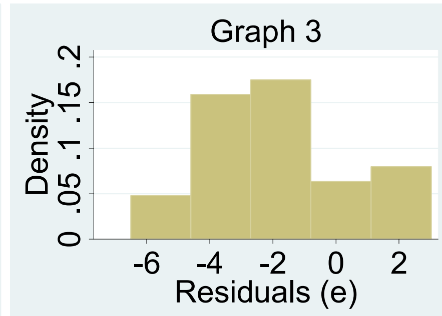
(23) A predictor variable has very little correlation with the dependent variable—with a coefficient of correlation that is both near zero and statistically insignificant—but the multiple regression coefficient on this variable is large and statistically significant.

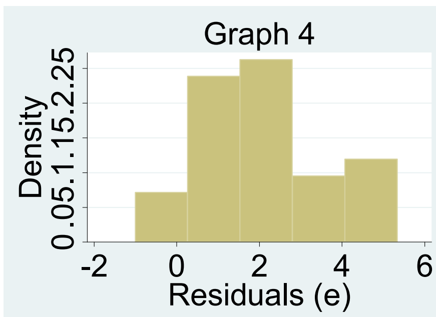
(A) this situation could arise
(B) this situation could not arise

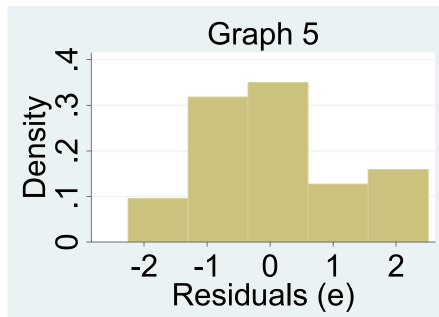
► **<u>Question (24):</u>** Recall the cross-country regression of income inequality (measured by the Gini index) on the growth rate of GDP (average annual % growth): $\widehat{GDP_{growth}} = 2.407 + 1.676\ inequality$ for the 33 OECD countries. The R^2 is 0.0069 and the standard deviation of residuals s_e is 1.2064.











(24) Which of the graphs above could be from the growth versus inequality analysis?

(A) Graph 1
(B) Graph 2
(C) Graph 3
(D) Graph 4
(E) Graph 5

MAKE SURE TO FILL IN THE BUBBLE FORM COMPLETELY BEFORE THE END OF THE TEST IS ANNOUNCED.

YOUR FORM CODE IS <u>A</u>.

YOU MAY KEEP THESE QUESTION PAPERS AND THE AID SHEETS: DO NOT TURN THEM IN. TURN IN ONLY THE BUBBLE FORM.