

**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:** 110 minutes. You MUST STAY for at least 50 minutes.

**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 **39** multiple choice questions with point values from 1.5 to 3 points each for a total of 87 points.

- Questions with 2 or 3 alternatives, (A) – (B)/(C), are worth: 1.5 point each correct answer
- Questions with 4 alternatives, (A) – (D), are worth: 2 points each correct answer
- Questions with 5 alternatives, (A) – (E), are worth: 3 points each correct answer

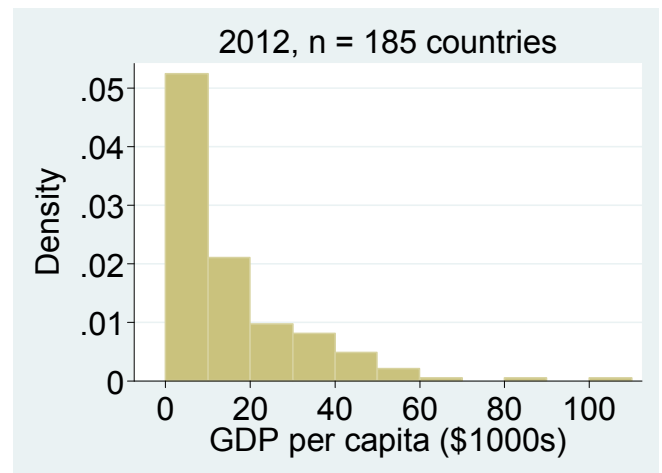
Complete the BUBBLE FORM before the end of the test, including entering your name.

**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

► **Questions (1) – (3):** This histogram shows GDP per capita in \$1,000s of USD on a purchasing power parity basis for 185 countries.

[https://www.cia.gov/library/publications/the-world-factbook/rankorder/rawdata\\_2004.txt](https://www.cia.gov/library/publications/the-world-factbook/rankorder/rawdata_2004.txt)



(1) Approximately what percent of countries have GDP per capita between \$20,000 and \$30,000?

- (A) 0.01%
- (B) 0.1%
- (C) 1%
- (D) 10%

(2) How would you describe the distribution of GDP per capita across countries?

- (A) positively skewed because most countries have fairly low GDP per capita
- (B) positively skewed because most countries have fairly high GDP per capita
- (C) negatively skewed because most countries have fairly low GDP per capita
- (D) negatively skewed because most countries have fairly high GDP per capita

(3) Which of these could be the median of these data?

- (A) \$9,100
- (B) \$12,300
- (C) \$22,800
- (D) \$54,900

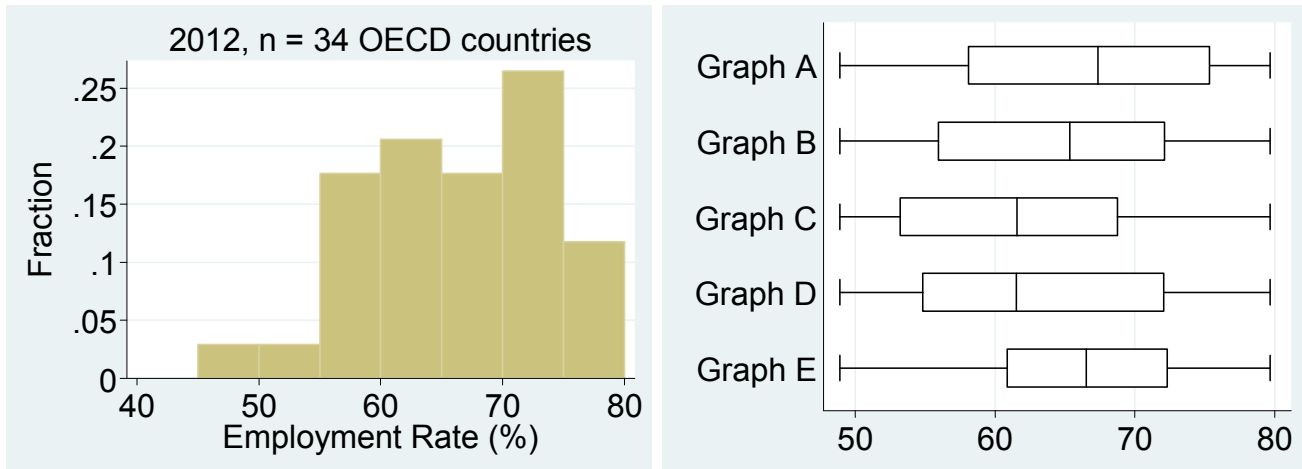
(4) Data on the total compensation for a sample of CEOs is highly skewed. The variable recording compensation is named X and it is measured in dollars. Consider creating a new variable Y, which transforms X. Which of the following would result in a more symmetric histogram?

- (A)  $Y = \frac{X}{1000}$
- (B)  $Y = \frac{X - \bar{X}}{s_X}$
- (C)  $Y = \ln(X)$
- (D) All of the above

(5) If you transform a variable using a linear transformation that adds a constant to all observations then which of the following would NOT change?

- (A) the sample median
- (B) the sample interquartile range
- (C) the sample coefficient of variation

► **Question (6):** Consider this histogram. (Source: OECD website.)



(6) Which box plot could summarize these 2012 employment rate data?

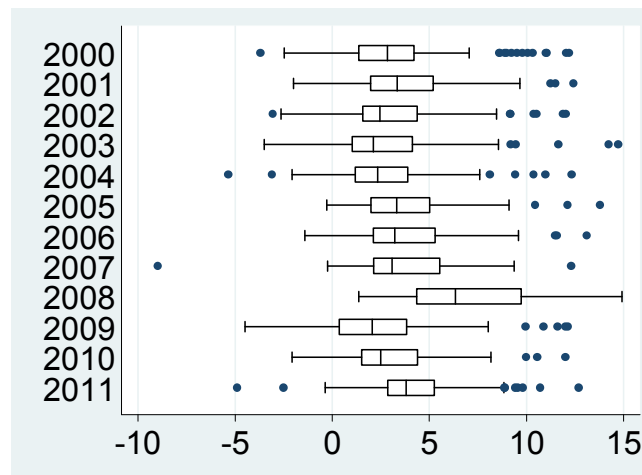
- (A) Graph A
- (B) Graph B
- (C) Graph C
- (D) Graph D
- (E) Graph E

(7) In Galton's original research, which has observations for over 1,000 father-son pairs, the heights of sons are regressed on the heights of the fathers. Why would the height of mothers be an unobserved variable (i.e. a lurking or confounding variable)?

- (A) because the height of mothers is as important as the height of fathers in determining a son's height
- (B) because a mother's height may affect the height of her son and her choice about how tall of a man to marry
- (C) because the genes of the mother are arguably more important than the father's in determining the height of the son
- (D) because short women may tend to give extra food to their growing sons to try to boost his height and offset some of the genetic effect
- (E) because the height of the mother is correlated with her nutrition and this affects the nutrition the fetus receives and hence the son's ultimate height

► **Questions (8) – (15):** Consider these data for inflation (%) for 107 countries (not all are shown) for each year from 2000 through 2011 from the World Bank website. Box plots of the data are also shown. <http://data.worldbank.org/indicator/FP.CPI.TOTL.ZG/countries>

country	country_code	inf00	inf01	inf02	inf03	inf04	inf05	inf06	inf07	inf08	inf09	inf10	inf11
Albania	ALB	0.050	3.108	7.771	0.484	2.280	2.367	2.371	2.933	3.359	2.281	3.552	3.450
Algeria	DZA	0.339	4.226	1.418	4.269	3.962	1.382	2.315	3.674	4.863	5.734	3.913	4.522
...													
United Kingdom	GBR	0.785	1.236	1.256	1.363	1.345	2.050	2.334	2.321	3.613	2.166	3.286	4.484
United States	USA	3.377	2.826	1.586	2.270	2.677	3.393	3.226	2.853	3.839	-0.356	1.640	3.157
Vanuatu	VUT	2.462	3.654	1.964	3.016	1.419	1.199	2.041	3.957	4.827	4.251	2.809	0.863



**(8)** What kind of data are these?

- (A) panel
- (B) time series
- (C) cross-sectional

**(9)** If all of the observations except for the first one (Albania) were dropped, what kind of data would the remaining data be?

- (A) panel
- (B) time series
- (C) cross-sectional

**(10)** If all of the variables except for “inf11” were dropped (not counting the country names), what kind of data would the remaining data be?

- (A) panel
- (B) time series
- (C) cross-sectional

(11) Using these data, why would a cross-tabulation be a poor way to assess how inflation in 2000 is related with inflation in 2010? It would be a poor choice because a cross-tabulation \_\_\_\_.

- (A) gives frequencies of every pair of values and there would be way too many
- (B) is a tool that is restricted to summarizing relationships amongst nominal variables
- (C) requires at least three variables to be assessed and we have only two in this case

(12) For which year is the interquartile range of inflation the largest?

- (A) 2007
- (B) 2008
- (C) 2009

(13) For which year is the range of inflation the largest?

- (A) 2007
- (B) 2008
- (C) 2009

(14) In 2005 roughly what fraction of countries in these data had inflation below 5%?

- (A) one quarter
- (B) one half
- (C) three quarters

(15) This STATA summary of inflation corresponds to which year?

-----				
	Percentiles	Smallest		
1%	-1.658031	-1.989279		
5%	.0559899	-1.658031		
10%	.9008585	-1.207572	Obs	107
25%	1.977099	-1.112516	Sum of Wgt.	107
50%	3.348123		Mean	3.695829
		Largest	Std. Dev.	2.719892
75%	5.187012	9.66575		
90%	7.329631	11.22698	Variance	7.397811
95%	8.825389	11.50209	Skewness	.7637981
99%	11.50209	12.43129	Kurtosis	3.917981

- (A) 2001
- (B) 2003
- (C) 2009
- (D) 2010
- (E) 2011

(16) For two interval variables the coefficient of correlation is 0.12 and the s.d. of the residuals is 1.4. Which is NOT a reason to check a scatter diagram before calculating these statistics?

- (A) to check for outliers
- (B) to check for heteroscedasticity
- (C) to check for an endogeneity bias
- (D) to check for a non-linear relationship

► **Questions (17) – (20):** In an academic journal article entitled “Sunk Costs, Firm Size and Firm Growth” Luis Cabral opens: “The conventional wisdom has been that expected firm growth rates are independent of firm size, a property known as Gibrat’s Law. However, recent empirical work has found a negative relation between firm growth and firm size.” In one of the cited papers entitled “Firm Size and Growth,” Bronwyn Hall runs a regression on a sample of 1,098 manufacturing firms where the dependent variable is the annual growth rate of the number of employees between 1976 and 1983 and the independent variable is the logarithm of the number of employees (a measure of firm size) in 1976. A negative slope is obtained.

(17) These data are \_\_\_\_.

- (A) experimental
- (B) observational
- (C) an observational experiment

(18) Hall’s regression analysis assumes a linear relationship between growth and \_\_\_\_.

- (A) firm size
- (B) the logarithm of firm size

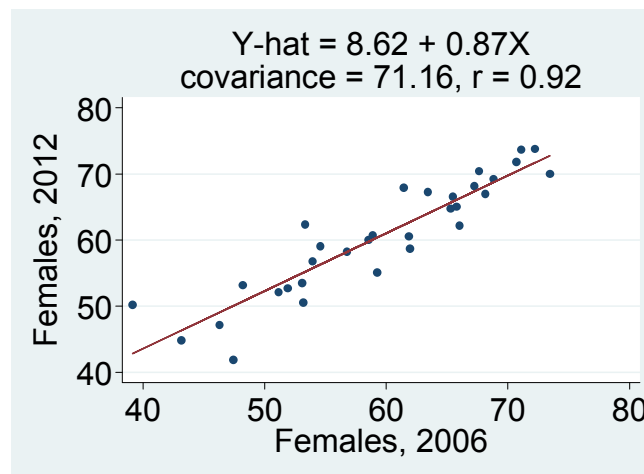
(19) If Gibrat’s Law were true we would expect \_\_\_\_ between firm size and growth.

- (A) little correlation
- (B) little association
- (C) a linear relationship
- (D) a non-linear relationship

(20) Hall finds heteroscedasticity. Which is a good explanation for why there is heteroscedasticity?

- (A) there are relatively few firms that are very small or large
- (B) growth is more variable across small firms than large firms
- (C) growth rates and firm sizes both vary substantially across firms
- (D) some firms grow faster than others for reasons that have nothing to do with firm size
- (E) the number of employees is a noisy measure of the size of a firm (e.g. market capitalization would be better)

► **Questions (21) – (23):** Excluding Iceland and Turkey, consider this scatter diagram and summary statistics for the percentage of females of working age employed in the 32 remaining OECD countries in 2012 versus 2006.



**(21)** About what percent of the variation across the OECD countries in the percent of women employed in 2012 is explained by variation in the percent of women employed in 2006?

- (A) 71%
- (B) 76%
- (C) 85%
- (D) 87%
- (E) 92%

**(22)** How do you interpret the slope of the OLS line?

- (A) On average 87 percent of the females employed in 2006 are still employed in 2012
- (B) When female employment is 0.87 percentage points higher in 2006, female employment is typically 1 percentage point higher in 2012
- (C) When female employment is 1 percentage point higher in 2006, female employment is typically 0.87 percentage points higher in 2012
- (D) On average a 0.87 standard deviation increase in female employment in 2006 is associated with a 1 standard deviation increase in female employment in 2012
- (E) On average a 1 standard deviation increase in female employment in 2006 is associated with a 0.87 standard deviation increase in female employment in 2012

**(23)** If instead of percentage points the employment rates of females in both 2006 and 2012 were recorded as the fraction employed (e.g. 0.61 instead of 61%) what would the covariance be?

- (A) 0.007116
- (B) 0.7116
- (C) 71.16
- (D) 7,116
- (E) 711,600

► **Questions (24) – (28):** A summary of undergraduate students from U of T's "Quick Facts":

- The University of Toronto has **65,612** undergraduate students enrolled in Fall 2012-2013
- Domestic: **56,380**
- International: **9,232**
- St. George campus has **42,347** undergraduate students
- Scarborough campus has **11,107** undergraduate students
- Mississauga campus has **12,158** undergraduate students

**(24)** If you randomly select an undergraduate student from U of T what is the probability that s/he is an international student?

- (A)** 0.024
- (B)** 0.141
- (C)** 0.164
- (D)** insufficient information provided to answer

**(25)** If you randomly select a student from the Scarborough campus what is the probability that s/he is an international student?

- (A)** 0.024
- (B)** 0.141
- (C)** 0.164
- (D)** insufficient information provided to answer

**(26)** Consider three independent surveys of undergraduate students where a random sample of 1,000 students is selected for each one. What is the probability that the same student (Albert Chung) is selected for two out of the three surveys?

- (A)** 0.0001
- (B)** 0.0002
- (C)** 0.0005
- (D)** 0.0007
- (E)** 0.0305

**(27)** If domestic undergraduate students pay \$8,000 of tuition and international undergraduate students pay \$27,000 in tuition then how much do undergraduate students pay on average?

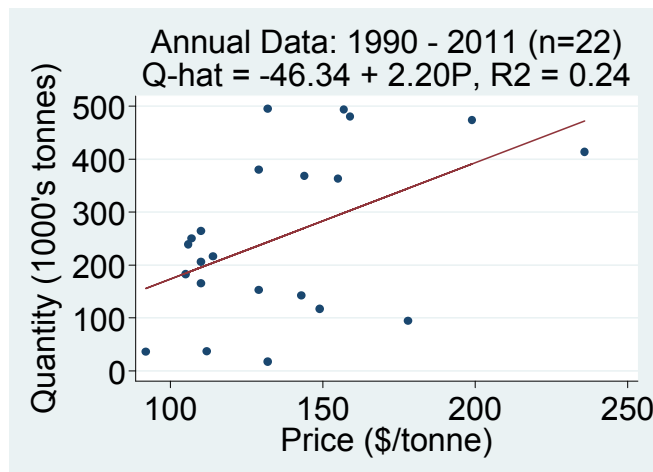
- (A)** \$10,244
- (B)** \$10,385
- (C)** \$10,457
- (D)** \$10,598
- (E)** \$10,673



**(28)** A survey of 1,000 undergraduates asks how much tuition the student pays. It yields an average that is less than the answer to the previous question. Which of these is a good explanation for why non-response bias may be the cause of part of the discrepancy?

- (A) some students chose not to respond because they were busy with their courses
- (B) the survey instrument is fairly long and some students got bored and did not finish it
- (C) the sample represents only a small fraction of U of T's undergraduate students (1,000 versus 65,612)
- (D) one third of invited students declined to participate and this disproportionately included international students who faced some language barriers
- (E) by only surveying 1,000 undergraduate students the survey design did not give any opportunity for all students to answer, which creates a non-response bias

► **Questions (29) – (30):** Consider this scatter diagram and summary statistics for corn prices and production for Manitoba corn farmers. [http://www.gov.mb.ca/agriculture/statistics/pdf/crop\\_grain\\_corn\\_sector.pdf](http://www.gov.mb.ca/agriculture/statistics/pdf/crop_grain_corn_sector.pdf)



**(29)** This regression line is not an estimate of the supply curve. Why not?

- (A) because of sampling error
- (B) because of endogeneity bias
- (C) because of heteroscedasticity
- (D) because a negative intercept makes no sense
- (E) because the slope should be negative, not positive

**(30)** When trying to estimate the supply curve using data on quantity produced and price, what would be a good example of an unobserved (i.e. a lurking or confounding) variable?

- (A) the increased demand for corn to make ethanol
- (B) the rapid creation of new uses for corn-based products
- (C) livestock producers (e.g. hogs and poultry) shifting away from corn-based feed
- (D) negative media coverage of high-fructose corn syrup as a sweetener in soft drinks
- (E) the development of corn plants with superior yield and a high tolerance for drought

► **Questions (31) – (32):** A May 30, 2013 New York Times article “Joblessness Shortens Lifespan of Least Educated White Women,” reports on new research. One thing the study finds is that less educated women are more likely to die than more educated women.

(31) What kind of data does this research use?

- (A) experimental
- (B) observational
- (C) an observational experiment

(32) The finding about the probability of dying is a comparison of \_\_\_\_ probabilities.

- (A) joint
- (B) marginal
- (C) conditional

► **Questions (33) – (35):** Consider this joint probability table for the Canadian population in 2012, which is derived from the Statistics Canada website.

(33) What percent of females are aged 65 to 69?

- (A) 1.2%
- (B) 2.4%
- (C) 4.8%
- (D) 51.1%

(34) Focusing on people aged 30 to 34, what can we say about the proportion of males versus females?

- (A) there are somewhat more females than males in this age bracket
- (B) there are somewhat more males than females in this age bracket
- (C) there are an equal number of males and females in this age bracket

(35) What is the 95<sup>th</sup> percentile of age?

- (A) it's within 75 to 79
- (B) it's within 80 to 84
- (C) it's within 85 to 89

Age	Male	Female	Total
0 to 4	0.0283	0.0270	0.0553
5 to 9	0.0274	0.0259	0.0532
10 to 14	0.0277	0.0262	0.0538
15 to 19	0.0318	0.0302	0.0620
20 to 24	0.0360	0.0340	0.0700
25 to 29	0.0357	0.0346	0.0703
30 to 34	0.0345	0.0345	0.0690
35 to 39	0.0331	0.0330	0.0661
40 to 44	0.0344	0.0340	0.0684
45 to 49	0.0387	0.0382	0.0769
50 to 54	0.0388	0.0387	0.0775
55 to 59	0.0344	0.0352	0.0696
60 to 64	0.0290	0.0302	0.0591
65 to 69	0.0229	0.0243	0.0472
70 to 74	0.0162	0.0180	0.0341
75 to 79	0.0120	0.0145	0.0265
80 to 84	0.0087	0.0119	0.0206
85 to 89	0.0047	0.0082	0.0129
90 and older	0.0021	0.0053	0.0074
Total	0.4962	0.5038	1.0000



**(36)** When is the slope of the regression line the same as the coefficient of correlation?

- (A)** when the data are experimental
- (B)** when the data are observational
- (C)** when the variables have been standardized
- (D)** when all of the underlying conditions of the regression analysis are met
- (E)** when the dependent and independent variables share the same units of measurement

► **Questions (37) – (39):** Summit Projects provides marketing services and website management. To understand consumer Web behavior, the company experiments with different offers and website designs. The results of such experiments can help to maximize the probability that visitors purchase products during a visit to a website. In one experiment visitors are randomly offered one of the following: an instant discount, free shipping, or nothing. Visitors make a purchase 6.20% of the time when offered the instant discount, 7.10% when offered the free shipping, and 1.40% when offered nothing. 30% of visitors are offered the discount and 30% are offered the free shipping.

**(37)** What percent of customers who visit the site make a purchase?

- (A)** 1.81%
- (B)** 2.47%
- (C)** 3.99%
- (D)** 4.55%
- (E)** 5.43%

**(38)** If you know that a visitor did NOT make a purchase, what is the chance that s/he had been offered nothing?

- (A)** 0.397
- (B)** 0.400
- (C)** 0.405
- (D)** 0.409
- (E)** 0.413

**(39)** Consider a repeat of this experiment that finds that 10 percent of visitors make a purchase. If purchasing behavior were completely independent of any offers of discounts and free shipping then what would be the probability that a visitor offered nothing would make a purchase?

- (A)** 0.04
- (B)** 0.10
- (C)** 0.15
- (D)** 0.25
- (E)** 0.40

**MAKE SURE TO FILL IN THE BUBBLE FORM COMPLETELY BEFORE THE END OF THE TEST IS ANNOUNCED. YOU MAY KEEP THESE QUESTION PAPERS AND THE AID SHEETS: DO NOT TURN THEM IN. YOUR FORM CODE IS A.**