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: 90 minutes. You MUST STAY for at least 60 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 **34** multiple choice questions with point values from 1 to 3 points each for a total of 76 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

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

(1) To make an inference about young adults, a random sample of 200 Canadians aged 20 to 29 years is drawn. This is an example of a ___ random sample.

(A) simple
(B) cluster
(C) stratified
(D) systematic

► **Questions (2) – (6):** This STATA summary is for the number of employees at Fortune 500 companies in 2013. (Source: http://money.cnn.com/magazines/fortune/fortune500/2013/full_list/.)

	Percentiles	Smallest		
1%	1133	190		
5%	4455	233		
10%	6057	788	Obs	500
25%	11450	926		
50%	25000		Mean	52809.56
		Largest	Std. Dev.	117105.9
75%	54000	361000		
90%	128300	440000	Variance	1.37e+10
95%	182434	466995		
99%	341500	2200000		

(2) There is one very obvious outlier. Rounding to the nearest integer, what would the mean be if this outlier were excluded?

(A) 48,136
(B) 48,287
(C) 48,393
(D) 48,410
(E) 48,507

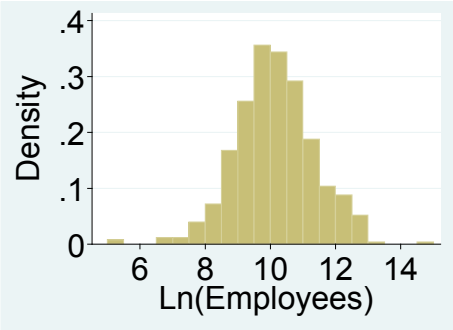
(3) Which is the best explanation for why the mean is bigger than the median?

(A) because of positive skew
(B) because of a large outlier
(C) because of negative skew
(D) because of sampling error

(4) What percent of Fortune 500 firms lie within one standard deviation of the mean?

(A) between 10% and 25%
(B) between 25% and 50%
(C) between 50% and 75%
(D) between 75% and 90%
(E) between 90% and 95%

(5) Considering also this histogram, how do the shapes of the distributions compare for the number of employees versus the natural logarithm of the number of employees?



(A) both are positively skewed
(B) both are negatively skewed
(C) both are approximately Normal
(D) one is positively skewed and one approximately Normal
(E) one is negatively skewed and one approximately Normal

(6) What is the standard deviation of the natural logarithm of employees?

(A) approximately 1.2
(B) approximately 2.5
(C) approximately 4.2
(D) approximately 6.5
(E) 11.67 = ln(117,105.9)

► **Questions (7) – (9):** For 185 countries the CIA measures GDP per capita in US$ on a purchasing power parity basis (GDP_PPP). The mean is $14,955.14 and the s.d. is $16,242.95. Consider a regression where the y variable is GDP_PPP and the x variable is GDP_PPP_1000, which is GDP_PPP but measured in $1,000's. (Source: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html.)

(7) What would the OLS slope be?

(A) 0
(B) 1/1000
(C) 1
(D) 1000

(8) What would be true about the SSE, SSR, and SST for this regression?

(A) SSE = SSR/SST
(B) SSE = SSR = SST
(C) SSE = 0 and SSR = SST
(D) SSR = 0 and SSE = SST

(9) What number would the SST be?

(A) 4.477×10^{10}
(B) 4.709×10^{10}
(C) 4.855×10^{10}
(D) 4.932×10^{10}

► **Questions (10) – (15):** The "Adult Literacy and Life Skills Survey" measures prose literacy, document literacy, numeracy, and problem solving. Rather than simply categorizing people as "literate" or "illiterate," it measures skills on a scale from level 1 (lowest) to level 4 or 5 (highest) for each dimension. (Source: http://www.statcan.gc.ca/pub/89-604-x/89-604-x2011001-eng.pdf, p. 61.)
For all questions presume the random sample is of Canadians aged 16 to 65.

	% of Canadian population aged 16 to 65 at each skill level, 2003, 2006 & 2008			
	Level 1	Level 2	Level 3	Level 4/5
prose literacy	14.6	27.3	38.6	19.5
document literacy	15.6	27.0	36.9	20.5
numeracy	19.5	30.3	33.4	16.9
problem solving	29.7	38.8	26.2	5.4

(10) What is the chance that one randomly selected person has Level 2 or 3 skill in numeracy?

(A) 0.101
(B) 0.637
(C) cannot be calculated with given information

(11) What is the chance that one randomly selected person has Level 1 skill in prose literacy and Level 1 skill in document literacy?

(A) 0.023
(B) 0.302
(C) cannot be calculated with given information

(12) What is the chance that one randomly selected person has Level 1 skill in prose literacy or Level 1 skill in document literacy?

(A) 0.023
(B) 0.302
(C) cannot be calculated with given information

(13) In a random sample of five people what is the chance two score Level 4/5 in numeracy?

(A) 0.0164
(B) 0.0249
(C) 0.1639
(D) 0.2492
(E) 0.2557

(14) In a random sample of 200 people what is the chance more than 50 score Level 4/5 in document literacy? (Round your answer to the nearest 0.05.)

(A) 0.05
(B) 0.10
(C) 0.15
(D) 0.35
(E) 0.45

(15) In a random sample of 50 people what is the chance more than two score Level 4/5 in problem solving?

(A) 0.51
(B) 0.55
(C) 0.59
(D) 0.63
(E) 0.67

► **Questions (16) – (18):** For the 34 OECD countries, an OLS regression of female employment (%) in 2012 on female employment (%) in 2006 yields: $\widehat{Fem\_emp}_{2012} = 9.12 + 0.86 Fem\_Emp_{2006}$.

(16) How do you interpret 0.86?

(A) in countries where 2012 female employment is one percentage point higher, on average 2006 female employment is 0.86 percentage points higher
(B) in countries where 2006 female employment is one percentage point higher, on average 2012 female employment is 0.86 percentage points higher
(C) in countries where 2012 female employment is one standard deviation higher, on average 2006 female employment is 0.86 standard deviations higher
(D) in countries where 2006 female employment is one standard deviation higher, on average 2012 female employment is 0.86 standard deviations higher

(17) T/F: On average female employment is 9.12 percentage points higher in 2012 compared to 2006.

(A) True
(B) False

(18) T/F: Countries with zero female employment in 2006 on average have female employment of 9.12% in 2012.

(A) True
(B) False

► **Questions (19) – (23):** Consider this graph from the *New England Journal of Medicine*, "Chocolate Consumption, Cognitive Function, and Nobel Laureates" by Franz H. Messerli, M.D., Oct. 18, 2012.

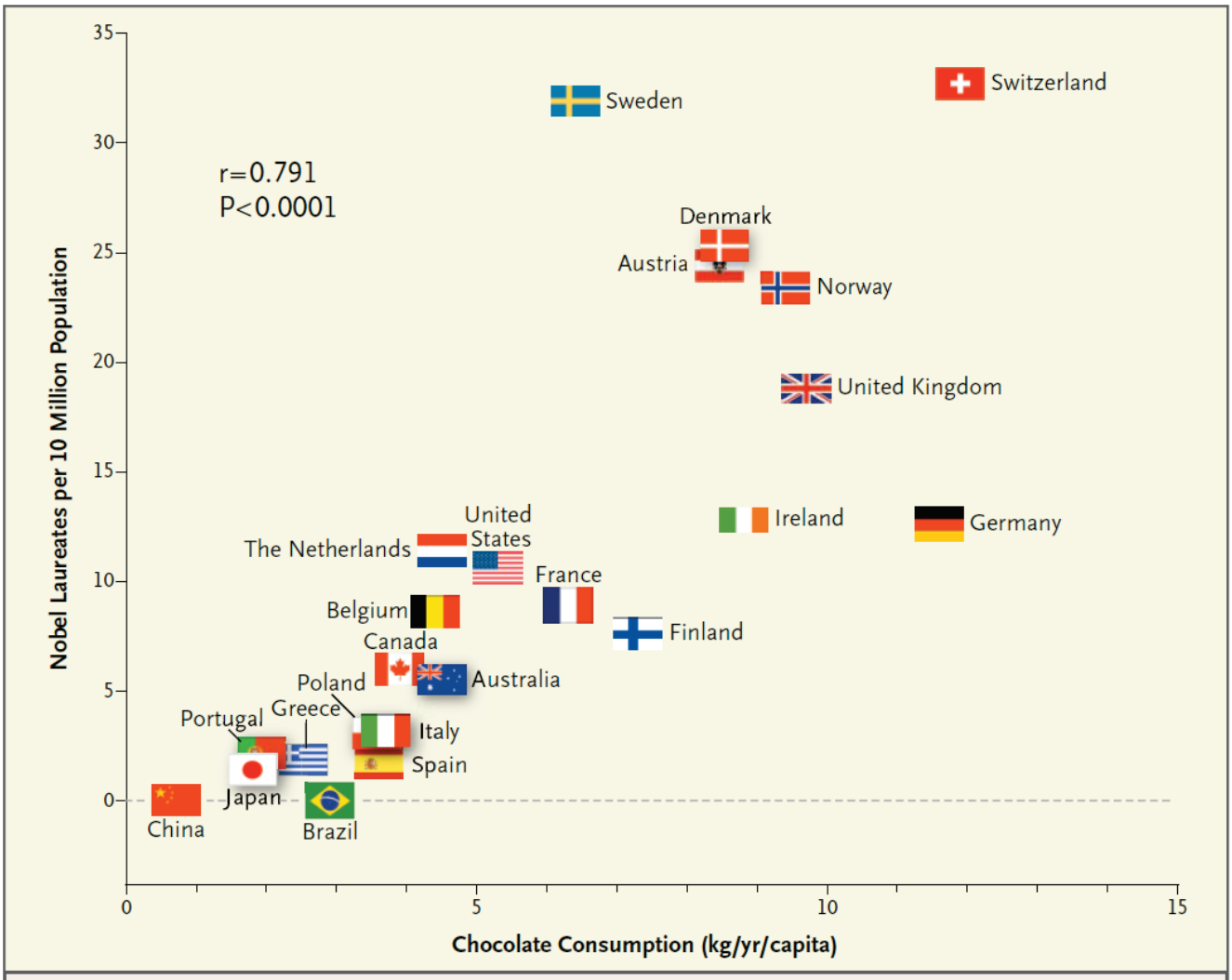


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

(19) Approximately, what would the slope of the OLS regression line be?

(A) between 0 and 0.25
(B) between 0.5 and 0.8
(C) between 2 and 3
(D) between 5 and 6

(20) The variation in chocolate consumption across countries explains what percent of the variation in the number of Nobel Laureates across countries?

(A) 0.626
(B) 0.791
(C) 0.889
(D) 0.902

(21) The Nobel laureates versus chocolate scatter diagram shows an example of ___.

(A) homoskedasticity
(B) heteroskedasticity

(22) Compared to an OLS line estimated with all of the data in the scatter diagram, if Sweden and Germany are removed and the OLS line re-estimated what should be expected?

(A) the R^2 would go down
(B) the mean of the residuals would go down
(C) the coefficient of correlation would go down
(D) the slope would better reflect the causal relationship
(E) the standard deviation of the residuals would go down

(23) Which is the best explanation for the positive correlation in the Nobel laureates versus chocolate scatter diagram?

(A) sampling error (chance)
(B) eating chocolate boosts mental performance in humans
(C) wealthy countries eat more chocolate and produce more Nobel Laureates
(D) countries with large populations eat more chocolate and produce more Nobel Laureates

► **Questions (24) – (26):** Checked bags are Normally distributed with mean 12.4 kg and s.d. 1.7 kg.

(24) What is the chance a randomly selected bag weighs within one standard deviation of the mean?

(A) 0.3174
(B) 0.3413
(C) 0.6587
(D) 0.6826

(25) For three randomly selected bags, what is the chance that all three weigh within two standard deviations of the mean?

(A) 0.8693
(B) 0.8919
(C) 0.9257
(D) 0.9544

(26) If a scale weighs to the nearest tenth of a kg, what is the chance a bag is weighed as 12.5 kg?

(A) approximately 0.01
(B) approximately 0.02
(C) approximately 0.04
(D) approximately 0.06
(E) approximately 0.08

► **Questions (27) – (29):** Recall the OECD data on the employment rates of males and females in 2006 and in 2012 all measured as percentages. Below are some summary statistics.

```
correlate fem_emp_2006 fem_emp_2012 male_emp_2006 male_emp_2012;
(obs=34)
             | fem~2006 fem~2012 mal~2006 mal~2012
-------------+------------------------------------
fem_emp_2006 |   1.0000
fem_emp_2012 |   0.9482   1.0000
male_em~2006 |   0.5904   0.4948   1.0000
male_em~2012 |   0.4758   0.6106   0.6498   1.0000

summarize fem_emp_2006 fem_emp_2012 male_emp_2006 male_emp_2012;
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
fem_emp_2006 |        34       58.92       11.45      22.75      80.88
fem_emp_2012 |        34       60.05       10.43      28.73      77.88
male_em~2006 |        34       74.43        6.31      60.88      88.05
male_em~2012 |        34       72.23        6.34      60.20      85.18
```

(27) What is the covariance between male employment in 2006 and male employment in 2012?

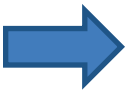
(A) 0.65 percent squared
(B) 26.00 percent squared
(C) 65.04 percent squared
(D) 1039.96 percent squared

(28) Supposing exactly half of the population is female and half is male, what is the mean employment for both sexes combined in 2012?

(A) 63.77
(B) 64.54
(C) 66.14
(D) 67.91

(29) Supposing exactly half of the population is female and half is male, what is the s.d. of employment for both sexes combined in 2012?

(A) 6.10
(B) 7.58
(C) 8.39
(D) 9.73
(E) 15.16



(30) Suppose $X_1 \sim N(0,1)$, $X_2 \sim N(0,1)$, and $X \sim N(0,1)$. Which statement is TRUE?

(A) $V[X_1 - X_2] > V[2X]$ if X_1 and X_2 are not correlated
(B) $V[X_1 - X_2] < V[2X]$ if X_1 and X_2 are not correlated
(C) $V[X_1 - X_2] = V[2X]$ if X_1 and X_2 are not correlated

► **Questions (31) – (33):** A random sample of 100 students is asked how many times they attended lecture for a one-term course. Respondents were offered these categories of replies: "0 – 2 times," "3 – 5 times," "6 – 8 times" "9 – 11 times" or "all 12 lectures." Below is a summary of the sample.

Reply	Percentage of sample
0 – 2 times	3%
3 – 5 times	4%
6 – 8 times	10%
9 – 11 times	36%
All 12 lectures	47%

(31) Assuming actual attendance is Uniformly distributed within each reply category, what is the mean number of lectures attended?

(A) 9.82
(B) 9.87
(C) 10.05
(D) 10.13

(32) Why is the assumption that actual attendance is Uniformly distributed within each reply category not plausible and what is the likely effect of this incorrect assumption?

(A) it is likely actual attendance within most categories is positively skewed: the answer to the previous question is a slight underestimate
(B) it is likely actual attendance within most categories is positively skewed: the answer to the previous question is a slight overestimate
(C) it is likely actual attendance within most categories is negatively skewed: the answer to the previous question is a slight underestimate
(D) it is likely actual attendance within most categories is negatively skewed: the answer to the previous question is a slight overestimate

(33) What is the name for the answer to Question 31?

(A) μ
(B) $\bar{X}$
(C) $E[X]$
(D) $V[X]$

(34) For 25 to 54 year olds in Canada in 2012, 81.4% are employed, 5.2% are unemployed and 13.4% are not in the labor force. For a random sample of 30, define W as the number employed, X as the number unemployed and Y as the number not in the labor force. How are W, X, and Y distributed?

(A) W, X, and Y are approximately Normal
(B) W is negatively skewed and X and Y are positively skewed
(C) W is positively skewed and X and Y are negatively skewed
(D) W is positively skewed and X and Y are approximately Normal
(E) W is approximately Normal, X is positively skewed and Y is negatively skewed

YOUR FORM CODE IS A. COMPLETE THE BUBBLE FORM BEFORE THE END OF THE TEST. YOU MAY KEEP THESE QUESTION PAPERS AND AID SHEETS: DO NOT TURN THEM IN.