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. However, please keep this test paper face up and flat on your desk.

Instructor: Prof. Murdock

Duration: 90 minutes

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

Format: This test has **35** multiple choice questions. Answer on the U of T bubble form. Total possible points: **80**.

- Questions with two possible answers, **(A) – (B)**, are worth: 1 point each
- Questions with three possible answers, **(A) – (C)**, are worth: 2 points each
- Questions with five possible answers, **(A) – (E)**, are worth: 3 points each

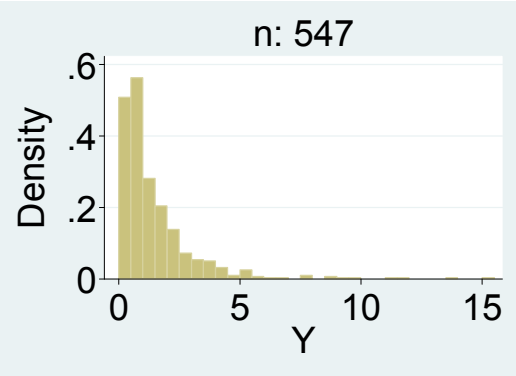
Instructions for completing your bubble form:

- Answers must be properly recorded on the U of T bubble form to earn marks
- **ON THE BACK OF THE FORM:** Print your **LAST NAME** and **INITIAL** in the boxes AND darken each letter in the corresponding bracket below each box
- **ON THE FRONT OF THE FORM:** Print your 9 (or 10) digit **STUDENT NUMBER** in the boxes AND darken each number in the corresponding bracket below each box. Your **FORM CODE** is **A.** Write in the other requested information in the upper right region of the form.
- ***NO EXTRA TIME once the end of the test is announced to fill in the form: everyone must immediately put down their writing instruments and promptly hand their forms to the TAs coming around to pick them up***
- Use a **pencil**
- Make dark solid marks that fill the bubble completely
- Erase completely any marks you want to change
- Crossing out a marked box is not acceptable and is incorrect
- If more than one answer is selected then that question earns 0 points

Instructions: Choose the best answer for each question. For questions with numeric answers that require rounding, round your final answer to be consistent with the choices offered. Use standard rounding rules.

DEFEND YOUR ACADEMIC INTEGRITY: Make sure to cover your answers and do not write your letter answers to each question in large font next to each question. Providing assistance to another student writing a test is as bad as receiving assistance. Protect yourself and your friends by not doing anything that could be interpreted as cheating and by guarding your own work.

► **Questions (1) – (6):** Consider this density histogram. The bin width is 0.5 and the first bin starts at 0.



(1) For these data the Empirical Rule ___ apply.

(A) would
(B) would not

(2) The median of these data is ___ 1.5.

(A) less than
(B) greater than

(3) Regarding the smallest values in these data we can say that ___.

(A) about 274 observations are less than 0.5
(B) about 274 observations are equal to zero
(C) about 25 percent of the observations fall in the range from 0 to 0.5

(4) The 90^{th} percentile of these data is ___ 10.

(A) less than
(B) greater than

(5) Standardizing Y would create a new variable where the majority of observations are ___.

(A) positive
(B) negative

(6) Standardizing Y would create a new variable that has a ___ distribution.

(A) positively skewed
(B) negatively skewed
(C) Bell shaped (Normal)

(7) Consider this excerpt from a 2011 academic research paper in economics "The Price Effects of a Larger Merger of Manufacturers: A Case Study of Maytag-Whirlpool."

> [Consider] eight major home appliance categories: dishwashers, clothes dryers, refrigerators, clothes washers, cooktops, ovens, and ranges. Within each appliance category there is substantial product differentiation. For example, the ratio of the price in the 75th percentile of the price distribution to the 25th percentile ranges from 1.7 for dishwashers to 2.8 for ovens.

The reported ratio of 2.8 in the last sentence measures ___.

(A) variability
(B) correlation
(C) association
(D) relative standing
(E) central tendency

(8) A November 6th, 2011 *New York Times* article ("Merger of Memphis and County School Districts Revives Race and Class Challenges") discusses a merger of two school districts: the Memphis district has mostly black students from poor families and the suburban district has mostly white students from upper-middle class families. Consider this excerpt: "Median family income in Memphis is $32,000 a year, compared with a suburban average of $92,000." This is an unethical use of statistics because ___.

(A) it uses inappropriate rounding
(B) it overstates the income difference by comparing a median and mean
(C) it understates the income difference by comparing a median and mean
(D) it implies causation when there is only a correlation between race and income
(E) it implies causation when there is only a correlation between income and student achievement

► **Questions (9) – (12):** Consider this cross-tabulation for two interval variables X and Y.

	Y				
X	0	1	2	3	Total
0	14	59	46	4	123
1	20	15	0	0	35
Total	34	74	46	4	158

(9) What is the median of X?

(A) 0
(B) 0.5
(C) 1

(10) What is the mean of Y?

(A) 0.9
(B) 1.1
(C) 1.3

(11) When Y equals 1 how often is X equal to 1?

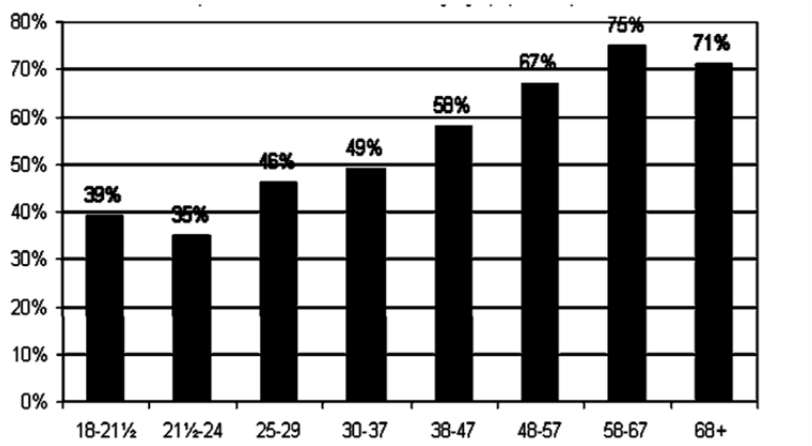
(A) 9.5% of the time
(B) 20.3% of the time
(C) 42.9% of the time

(12) There is a ___ correlation between X and Y.

(A) positive
(B) negative

► **Question (13):** Consider this figure from "Youth Voter Turnout in Canada: 1. Trends and Issues" April 7, 2010. (http://www.parl.gc.ca/Content/LOP/ResearchPublications/2010-19-e.pdf).

Figure 2 – Estimated Voter Turnout in Canada by Age Group, 2004 Federal General Election



(13) The heights of the bars relate to which kind of probability?

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

(14) In the 1971 article "Belief in the Law of Small Numbers" Tversky and Kahneman mention a researcher who tests two toys – Toy 1 and Toy 2 – and observes that the first 4 of 5 infants prefer Toy 1. The researcher feels confident that Toy 1 is superior given this evidence. This is an example of ___.

(A) the gambler's fallacy
(B) regression towards the mean
(C) how people underestimate the magnitude of chance variation in small samples
(D) how people misinterpret probabilities and expectations when faced with uncertainty
(E) All of the above

► **Questions (15) – (17):** Consider this graphic from an Oct. 8th, 2011 article "Business and the euro crisis" in *The Economist.*

(15) This graphic shows what is best described as ___.

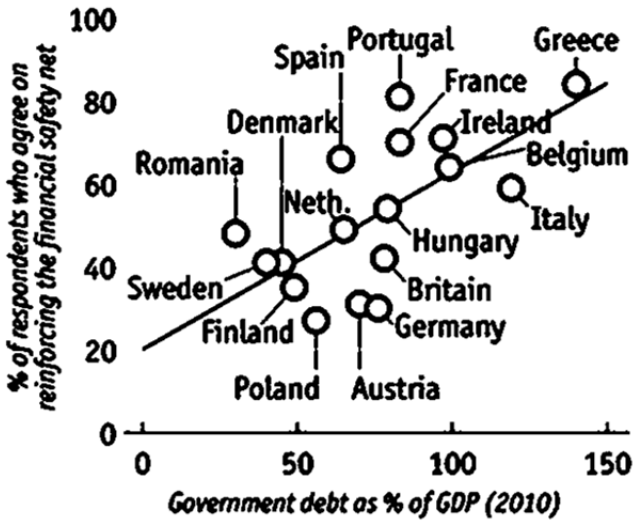
(A) a positive correlation
(B) a positive association

(16) This graphic illustrates ___.

(A) homoskedasticity
(B) heteroskedasticity

(17) For Hungary ___ is close to zero.

(A) the residual
(B) sampling error
(C) the standard error of estimate
(D) the standard deviation of the error term
(E) All of the above



Sources: Booz & Company; INSEAD; Eurostat; EEC

(18) By increasing the sample size for a linear regression you should expect which of the following?

(A) an increase in the SSE
(B) an increase in the SST
(C) no change in the standard error of estimate
(D) no change in the R^2 (coefficient of determination)
(E) All of the above

► **Questions (19) – (21):** Some top executives in Silicon Valley send their children to a Waldorf school: it does not use computers and technology. Instead they use pens, encyclopedias, and physical activity. Consider this excerpt from an October 23rd, 2011 *New York Times* article "A Silicon Valley School That Doesn't Compute."

> When asked for evidence of the schools' effectiveness, the Association of Waldorf Schools of North America points to research by an affiliated group showing that 94 percent of students graduating from Waldorf high schools in the United States between 1994 and 2004 attended college, with many heading to prestigious institutions like Oberlin, Berkeley, and Vassar. Of course, that figure may not be surprising, given that these are students from families that value education highly enough to seek out a selective private school, and usually have the means to pay for it. And it is difficult to separate the effects of the low-tech instructional methods from other factors. For example, parents of students say that it attracts great teachers who go through extensive training in the Waldorf approach, creating a strong sense of mission that can be lacking in other schools.

(19) Which best states the primary research question?

(A) What is the effect of a strong sense of mission on the quality of Waldorf teachers?
(B) What is the effect of Waldorf's low-tech instructional methods on students' learning and achievement?
(C) What is the effect of family means to pay (income) on families' choices amongst schooling alternatives such as selective private schools like the Waldorf?

(20) The passage describes observational data. Which would be necessary conditions if we imagine experimental data that could be used to measure the causal effect?

(A) Ensuring that students are randomly assigned families and incomes
(B) Ensuring that there are no random differences across students in skill levels
(C) Ensuring that instructional methods and teachers are randomly assigned to students
(D) Ensuring that the type of teaching method is the only factor to affect people's choices
(E) All of the above

(21) Which is NOT an unobserved (aka lurking or confounding) variable?

(A) teachers' skills
(B) families' incomes
(C) students' aptitude
(D) schools' effectiveness
(E) parents' values regarding importance of education

(22) Consider a statistical analysis of the linear relationship between the weight of laptops in kilograms (kg) and their retail price in hundreds of dollars ($100s). An interpretation that "A one kg increase in weight is on average associated with a price that is $67 lower" means that ___.

(A) the covariance is equal to -67
(B) the coefficient of correlation is equal to -67
(C) the coefficient of correlation is equal to -0.67
(D) the slope of the OLS (regression) line is equal to -67
(E) the slope of the OLS (regression) line is equal to -0.67

► **Questions (23) – (24):** Consider this STATA summary of an interval variable. Some of the output is deliberately erased.

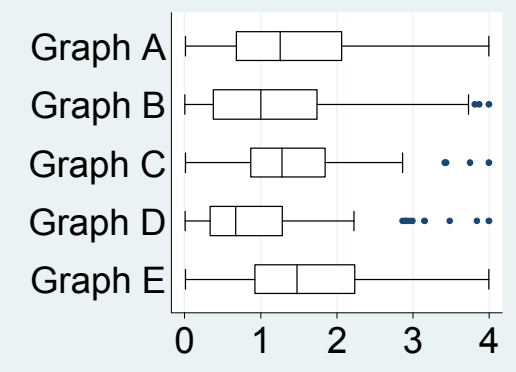
```
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .0865331          .0103
 5%     .3297926       .0865331
10%     .4564546       .0891625       Obs                 128
25%      .870467       .1029612       Sum of Wgt.         128

50%     1.281874                      Mean           1.383799
                        Largest       Std. Dev.
75%     1.847449        3.41924
90%     2.372096       3.435251       Variance       .5873525
95%     2.574215       3.750118       Skewness
99%     3.750118         3.9945       Kurtosis
```

(23) What percent of these data are more than two standard deviations below the mean?

(A) 0 percent
(B) about 2.3 percent
(C) about 4.6 percent
(D) at least 12.5 percent
(E) at least 25.0 percent

(24) Which of the graphs to the right corresponds to the same data as the STATA summary above?

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



► **Questions (25) – (26):** Consider this excerpt from an October 19th, 2011 *New York Times Magazine* article "Don't Blink! The Hazards of Confidence" by Daniel Kahneman.

> I asked for some data to prepare my presentation and was granted a small treasure: a spreadsheet summarizing the investment outcomes of some 25 anonymous wealth advisers, for eight consecutive years. … I computed the correlations between the rankings of advisers in different years, comparing Year 1 with Year 2, Year 1 with Year 3 and so on up through Year 7 with Year 8. That yielded 28 correlations, one for each pair of years. While I was prepared to find little year-to-year consistency, I was still surprised to find that the average of the 28 correlations was .01. In other words, zero. The stability that would indicate differences in skill was not to be found. The results resembled what you would expect from a dice-rolling contest, not a game of skill.

(25) The original data Kahneman received in a spreadsheet are ___ data.

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

(26) If skill played an even smaller role we would expect the correlation to ___.

(A) remain the same
(B) become negative
(C) get closer to zero

► **Questions (27) – (28):** A university has two campuses: Campus A and Campus B. It records the number of students enrolled at each campus each year for 20 years. Over that period the average enrolment at Campus A is 10,000 with a standard deviation of 500 and at Campus B it is 15,000 with a standard deviation of 700. Enrolments have a coefficient of correlation of -0.142 across campuses.

(27) What is the standard deviation of the total enrolment for both campuses combined?

(A) 540
(B) 600
(C) 800
(D) 860
(E) 1200

(28) If instead of negatively correlated the enrolments were positively across campuses then the mean total enrolment for both campuses combined would ___.

(A) increase
(B) decrease
(C) stay the same

(29) A politician is believed to have a 55 percent approval rating. What is the chance that a pollster randomly samples 500 people and finds that fewer than 250 approve of the politician?

(A) 0.01
(B) 0.02
(C) 0.04
(D) 0.48
(E) 0.49

► **Questions (30) – (31):** An economist is studying price dispersion: the exact same good being sold at different prices at various retail outlets. She records the retail price of LaraBars – a snack food (pictured at the right) – at retail outlets in the GTA. Prices are distributed Normally. Suppose the population mean is $2.05 and the population standard deviation is $0.23.

(30) What percent of retail outlets offer the LaraBar at $1.99 or less?

(A) 10%
(B) 20%
(C) 30%
(D) 40%
(E) 60%

(31) How much does the LaraBar cost at the most expensive retail outlets: the top 3 percent?

(A) at least $2.34
(B) at least $2.36
(C) at least $2.40
(D) at least $2.44
(E) at least $2.48

► **Questions (32) – (33):** Suppose the TTC routinely screens one percent of its workforce at random each day for drugs or alcohol.

(32) What is the chance an employee working eight consecutive days is screened more than once?

(A) 0.0009
(B) 0.0027
(C) 0.0052
(D) 0.0084
(E) 0.0101

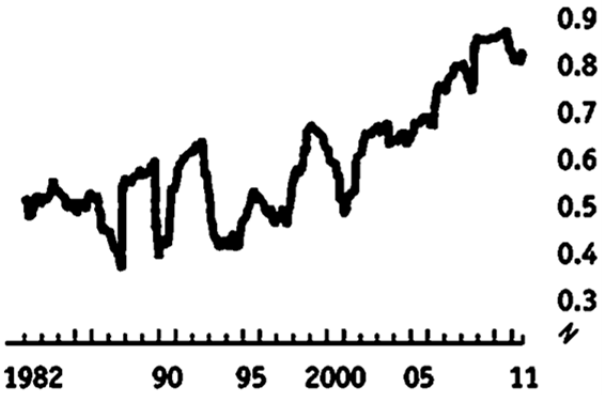
(33) Suppose ten members of the same extended family all work at the TTC. What is the chance that on a day when they are all working none of them is screened?

(A) 0.70
(B) 0.72
(C) 0.77
(D) 0.90
(E) Cannot answer because the independence assumption is violated by using family members

► **Questions (34) – (35):** Consider this excerpt and graphic from a September 10th, 2011 *Economist* article "All in the same boat: Why global stockmarkets have become more correlated."

> Diversification is supposed to be one of the rare free lunches in finance. Spread your assets geographically (or by asset class) and the chances are that your investments will not rise and fall together. Investors should be able to get the same reward with less risk. But, as the chart shows, global stockmarkets have steadily become more correlated over the past few decades. Wake to the financial headlines on any given morning and you will find that a sell-off in Asia has spread to Europe and that, all too often, both continents are reacting to a late plunge on Wall Street. It is rare for individual markets to go against the trend.

Correlation* in weekly price change of 23 developed stockmarkets against the MSCI World index, averaged



Sources: Société Générale; Thomson Reuters; *The Economist* *Two-year rolling

(34) The original data containing the price changes, which ultimately led to the production of the above chart, are ___ data.

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

(35) For it to be a valid summary, which are an underlying assumption behind this chart?

(A) it is rare for individual markets to go against the trend set by the MSCI World Index
(B) there are no unobserved variables that vary both over time and across these indices
(C) there are no unobserved variables the affect both the MSCI World Index and other indices
(D) weekly changes in each index are linearly related to weekly changes in the MSCI World Index
(E) All of the above

Your **FORM CODE** is **A**: make sure to darken **A** in the space labeled FORM CODE.
Marks will be posted online tonight (if possible): I will alert you with an e-mail.