Last Name:

First Name:

This test includes 7 written questions with varying point values. There is a total of 100 possible points. You will receive a separate handout with the formula sheet and Standard Normal table. You may use a non-programmable calculator. You have 3 hours.

Please keep these papers face-up and closed on your desk until the start of the test is announced. Put your writing instruments down immediately when the end of the test is announced.

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

- Write your answers clearly, concisely, and completely below each question.
- ***Make sure to show your work and reasoning***. ***Make sure your graphs are fully labeled***.
- 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 a quantitative analysis & 1 – 2 sentences]" To best demonstrate your understanding, follow the guides and focus on directly answering the questions asked.
    - Make sure to write actual sentences (not short-hand or bullet lists).
    - Apply your skills to the specific situation presented with the question.
    - Reproducing examples or discussion from our course materials is not an effective strategy because these will not address the specific situation asked about.
    - Extraneous analysis does not earn positive marks if it is correct and earns negative marks if incorrect: focus on the question that is asked.
- For questions with multiple parts, attempt each part even if you had trouble with earlier parts.

Student #:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Raw Total	Your Mark (out of 100%)
Point Value:	*8*	*14*	*24*	*10*	*12*	*16*	*16*	*100*	
Points Earned:									

(1) [8 pts] Statistics Canada reports unemployment rates of 25- to 29-year-olds, by educational attainment for 2012. An excerpt from that table is provided to the right. Treat the reported numbers as facts. (http://www.statcan.gc.ca/pub/81-582-x/2013001/tbl/tble3.2-eng.htm)

All levels	**7.4**
Less than high school	16.4
High school	8.8
College or trade	6.2
University	5.8

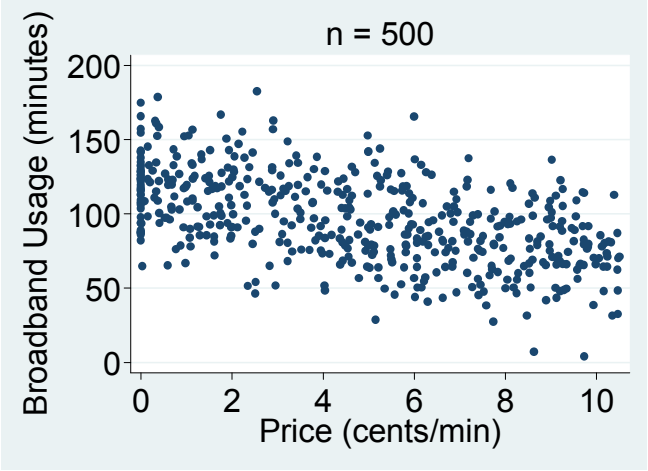
(a) [4 pts] Using notation from our course, replace the question mark in P(?) = 0.088 to show the meaning of 8.8 in the table above. What type of probability is it? [Answer with a formal probability statement and a phrase]

(b) [4 pts] If you randomly selected five 25- to 29-year-olds from Canada what is the probability that 1 of the 5 is unemployed? [Answer with a quantitative analysis]

(2) [14 pts] The two parts of this question are completely unrelated.

(a) [5 pts] Apartment complexes can have a single hot water system that supplies hot water to all of the units. Consider a complex with 50 units. Mean daily water usage is 280 liters per unit and the standard deviation is 160 liters. If we assume hot water usage is independent across units, what is the standard deviation of the total daily water usage for the entire apartment complex? Include the units of measurement. [Answer with a quantitative analysis]

(b) [9 pts] A hotel chain offers customers broadband access in their rooms. How much are its guests willing to pay for this service? Randomly selected guests receive a price ranging from 0 cents per minute up to 10 cents per minute at random. (Note: All guests had to use some minutes for check-out.) The hotel monitors how many minutes each guest purchases. For $n = 500$ an OLS line is obtained: $\widehat{usage}_i = 119.1 - 4.8price_i$ with $R^2 = 0.2813$. First, explain why you should review the scatter diagram to the right before interpreting these numbers. Second, *fully interpret* the OLS line and the R^2. [Answer with 4 – 6 sentences]



(3) [24 pts] Consider these excerpts from "The Wedding Industry's Pricey Little Secret" posted June 12, 2013 on Slate.

> "Average wedding cost $28,400 last year," reports CNN Money. "Average U.S. wedding costs $27,000!!" enthuses the New York Daily News. "Average cost of U.S. wedding hits $27,021," declares Reuters.
>
> One of the most extensive surveys, and perhaps the most widely cited, is the "Real Weddings Study" conducted each year by TheKnot.com and WeddingChannel.com. (It's the sole source for the Reuters and CNN Money stories, among others.) They survey some 20,000 brides per annum, an impressive figure. But all of them are drawn from the sites' own online membership, surely a more gung-ho group than the brides who don't sign up for wedding websites, let alone those who lack regular Internet access.
>
> In 2012 [for the "Real Weddings Study" data], when the average wedding cost was $27,427, the median was $18,086.

http://www.slate.com/articles/life/weddings/2013/06/average_wedding_cost_published_numbers_on_the_price_of_a_wedding_are_totally.single.html

(a) [4 pts] Which statistic is the better measure of what couples typically spend on a wedding? Explain using appropriate terminology from our course. [Answer with 2 – 3 sentences]

(b) [6 pts] Sketch a fully-labeled graph of the population distribution of wedding spending that is consistent with the last sentence in the excerpt above. [Answer with a graph]

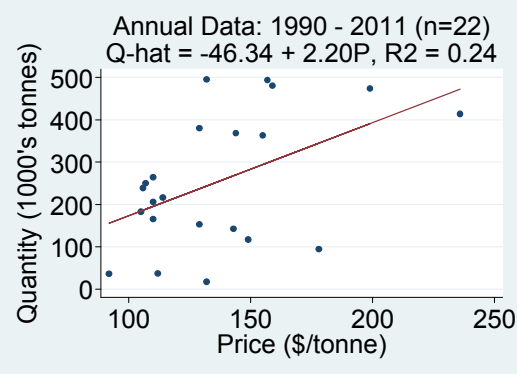
(c) [3 pts] If μ is \$28,000 and σ is \$18,000, what fraction of couples spends less than \$5,000 on their wedding?

(d) [4 pts] Is \$27,427 an unbiased estimator of the population mean where the population is all U.S. couples who had weddings in 2012? Explain. Make sure your answer is highly context-specific and indicates the direction of any bias. [Answer with 2 – 3 sentences]

(e) [7 pts] For this part only suppose that the population mean is \$28,000, the population standard deviation is \$18,000 and that the "Real Wedding Study" mentioned in the article is *only* affected by sampling error. Provide a *quantitative measure* of how much sampling error would affect the sample mean. *Fully interpret* the number you find and mention whether the comment in the article of "an impressive figure" is justified. [Answer with a quantitative analysis and 3 – 4 sentences]

(4) [10 pts] For the multiple-choice question below, explain why **(E)** is a correct answer. Also, give another *specific* example of a correct answer that could have appeared instead of **(E)**. What do **(A)** – **(D)** have in common that makes them all incorrect answers? Explain. [Answer with 4 – 8 sentences]

► 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



(MCQ) 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? **(E)**

(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

(5) [12 pts] A spam filtering firm claims that only 1 percent of valid e-mails are blocked as spam. A random sample of 200 blocked e-mails is evaluated. Of those e-mails, 2 percent are determined to have been valid. Is sampling error a plausible explanation for finding such a high fraction of valid e-mails in the sample? Illustrate your answer with a full-labeled graph. [Answer with a quantitative analysis, a graph, and 1 sentence]

(6) [16 pts] Consider these excerpts from “Mike Duffy tops Dalton McGuinty, Rob Ford in scandal poll” posted June 23, 2013 on the Toronto Sun.

> Pollster Forum Research asked more than 1,500 Canadians which of [three scandals – former Ontario premier Dalton McGuinty’s cancellation of gas plants, Senator Mike Duffy’s submission of ineligible reimbursements, and Toronto Mayor Rob Ford’s alleged crack cocaine usage—is] “the most embarrassing scandal to you as a Canadian.”
>
> Across the country, 33% picked Duffy. Because of his connection to the East Coast, 55% of Atlantic Canadians ranked his situation as “most embarrassing.”
>
> Just as Atlantic Canadians were hardest on Duffy, Ontario voters were hardest on McGuinty. Among Ontario respondents to the Forum survey, 41% said McGuinty's stink was the foulest, while the Duffy and Ford scandals were each picked by 21% of Ontario respondents as “most embarrassing.”
>
> The Forum poll of 1,525 was conducted via telephone on June 18. Forum says the results of this poll are accurate to within two percentage points, 19 times out of 20.
>
> http://www.torontosun.com/2013/06/23/mike-duffy-tops-dalton-mcguinty-rob-ford-in-scandal-poll

(a) [4 pts] Explain the meaning of the last sentence in the above excerpt using notation from our course. Show the underlying calculation. [Answer with a quantitative analysis and 1 sentence]

(b) [5 pts] Considering the table to the right from Statistics Canada, why is the last sentence of this article highly misleading for three of the four estimates reported in the excerpt? Illustrate your answer with a much better calculation for one of the three estimates. [Answer with a quantitative analysis and 2 – 3 sentences]

	2012 population (in thousands)
Atlantic provinces	2,363.5
Ontario	13,505.9
Rest of Canada	19,011.2

(c) [7 pts] Presuming the underlying conditions are met, compute and fully interpret the 95 percent confidence interval estimate based on the “33%” that appears above. [Answer with a quantitative analysis and 1 sentence]

(7) [16 pts] People vary in their opinions on tipping taxi drivers. Suppose an association of taxi drivers claims that on average people tip 11 percent with a standard deviation of 5. The distribution is positively skewed. An independent researcher randomly samples 1,000 taxi rides and obtains the tip amount (as a percent of the total bill) for each. The sample mean is 12.2.

(a) [10 pts] What is the chance that such a high sample mean is obtained if the original claim is true? Show your work AND illustrate your answer on a fully-labeled graph. [Answer with a quantitative analysis and graph]

(b) [6 pts] Assess the plausibility of each of the three possible explanations for the discrepancy between 12.2 and 11. [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.