Economics 250 — Final Exam (answers) **11 April 2015**

Instructions:

- The exam is 3 hours in length. There are 180 points in total. Allocate your time accordingly.
- Put your name and student number on each answer booklet used.
- You may use a hand calculator: either the standard Casio fx-991MS, or a university-approved calculator with a gold or blue sticker. No red-sticker calculators or other aids are allowed.
- Formulas and tables are printed at the end of the question papers.
- This midterm consists of 13 pages in all: this cover sheet, 5 question pages, 4 formula pages, and 3 statistic table pages. Please ensure you have all questions/sheets!
- This exam is divided into two sections:
  - Section A (page 2, worth 30 marks) consists of 10 very short questions requiring only a small calculation, value lookup, or a couple of words.
  - Section B (pages 4–12, worth 150 marks) consists of 9 longer questions with multiple parts. Show your work: part marks cannot be awarded for wrong answers without calculations.
- Answer all questions. The value of each question is shown in the exam.
- Proctors are unable to respond to queries about the interpretation of exam questions. Do your best to answer the exam questions as they are written.
- This material is copyrighted and is for the sole use of students registered in Economics 250 and writing this exam. This material shall not be distributed or disseminated. Failure to abide by these conditions is a breach of copyright and may also constitute a breach of academic integrity under the University Senate's Academic Integrity Policy Statement.
- Good luck!

Section A: Very short questions [30 points]

The following questions require only a couple of words, a lookup of a numerical value, or a small calculation. Each question is worth 3 marks.

1. Suppose X is a continuous, right-skewed distribution, and that $Y = \log(X)$ appears (approximately) normally distributed with a mean approximately equal to the median. Would you expect the mean of X to be *above*, *below*, or *approximately equal to* the median of X?

Answer: The mean will be $\boxed{\text{above}}$ the median.

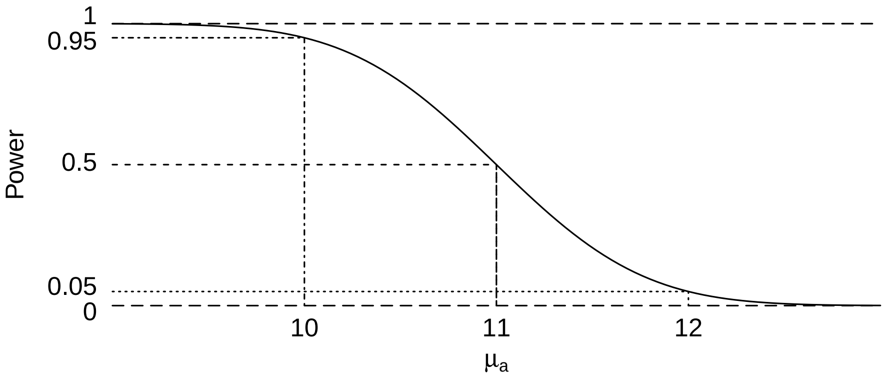
2. You have a sample of $n = 27$ observations and wish to test the hypothesis $\mu = 12$ against the alternative $\mu \neq 12$. The population's standard deviation is unknown. What will be the distribution of your test statistic?

Answer: The test statistic follows a $\boxed{t\text{-distribution with 26 degrees of freedom}}$.

3. What is the critical value of the appropriate distribution needed to calculate a 95% confidence interval for μ from a sample of $n = 27$ observations, without knowing σ?

Answer: $\boxed{2.056}$ *from Table D for p=.025, df=26*

4. The following graph shows the power of a test conducted at the $\alpha = 0.05$ level. Write down the null and alternative hypotheses of the test.



Answer: $\boxed{H_0 : \mu = 12,\ H_a : \mu < 12}$. *The one-sided less-than alternative is because the test rejects with increasing probability as we see negative values further from 12, but doesn't reject for positive values further from 12. The value is 12 because that's the point where the power equals α (the power at the null hypothesis is simply the probability of making a Type I error).*

5. Suppose A and B are disjoint events, and that $P(A) = 0.4$. What is the range of values that $P(B)$ could take?

Answer: $\boxed{[0, 0.6]}$. *Since A and B are disjoint, $P(A \cup B) = P(A) + P(B)$; since probabilities have to add up to a value between 0 and 1, $P(B)$ must be between 0 and 0.6.*

6. Suppose A and B are independent events, and that $P(A) = 0.4$. What is the range of values that $P(B)$ could take?

Answer: $\boxed{[0, 1]}$. *Independence only tells us that $P(B|A) = P(B)$, but doesn't otherwise put any restrictions on B.*

7. If 62% of Queen's students are female and 12% of female students own cars, and 14% of students own cars, what is the probability that a randomly selected student car belongs to a female student?

Answer: $P(female|car) = \frac{P(car|female)P(female)}{P(car)} = \frac{(.12)(.62)}{.14} = \boxed{0.53}$

8. What is the probability that a weighted coin that produces heads 75% of the time will produce exactly 9 heads in 12 flips?

Answer: $P(\text{9 heads in 12 flips}) = \binom{12}{9} 0.75^9 0.25^3 = (220)0.75^9 0.25^3 = \boxed{0.368}$

9. If X is distributed $\mathcal{U}(-10, 5)$, what is $P(X < 0 \ \cup \ X \geq 3)$?

Answer:

$$P(X < 0) = \frac{0 - (-10)}{5 - (-10)} = \frac{10}{15}$$
$$P(X \geq 3) = \frac{5 - 3}{5 - (-10)} = \frac{2}{15}$$

and since the two areas don't overlap, adding them together gives $P(X < 0 \ \cup \ X \geq 2) = \frac{12}{15} = \boxed{0.8}$.

10. Suppose that IQ scores in the general population follow the distribution $\mathcal{N}(100, 15)$. What is the distribution of the mean IQ score of a randomly selected group of 10 people?

Answer: $\overline{x} \sim \boxed{\mathcal{N}\left(100, \frac{15}{\sqrt{10}}\right)} = \boxed{\mathcal{N}(100, 4.743)}$

Section B: Longer questions [150 points]

1. **[15]** A random variable X takes one of the four values: 0, 5, 6, or 7. The value 0 occurs with probability 0.4; the remaining values all have equal probabilities of occurring.

a) Find the mean and standard deviation of X.

Answer:

$$\begin{aligned}\mu_x &= \sum_{i=1}^{4} P(X = x_i)x_i = 0.4(0) + 0.2(5 + 6 + 7)\\ &= \boxed{3.6}\\ \sigma_x^2 &= \sum_{i=1}^{4} P(X = x_i)(x_i - \mu_x)^2\\ &= 0.4(-3.6)^2 + 0.2(1.4)^2 + 0.2(2.4)^2 + 0.2(3.4)^2\\ &= 9.04\\ \sigma_x &= \sqrt{9.04} = \boxed{3.007}\end{aligned}$$

Another random variable Y follows the $\mathcal{U}(1, 3)$ distribution, which has mean 2 and standard deviation 0.577. X and Y are *not* independent: the correlation coefficient between them is -0.5. A third random variable, W, is defined by $W = 2X - 3Y$.

b) Find the mean of W.

Answer: $\mu_w = 2\mu_x - 3\mu_y = 2(3.6) - 3(2) = \boxed{1.2}$

c) Find the standard deviation of W.

Answer: $\sigma_w^2 = 2^2\sigma_x^2 + 3^2\sigma_y^2 - 2(2)(3)\sigma_{XY}$.

σ_{XY} isn't given, but can be calculated by rearranging the correlation formula $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$:

$-0.5\sigma_x\sigma_y = -0.5(3.007)(0.577) = -0.868$.

So: $\sigma_w^2 = 4(3.007)^2 + 9(0.577)^2 - 12(-0.868) = 49.58$ and thus

$\sigma_w = \sqrt{49.58} = \boxed{7.04}$

2. **[15]** The Canadian lottery "Lotto 6/49" is played by choosing any 6 different numbers from 1–49. The jackpot prize (of at least $5 million, and often much larger) is won if all 6 numbers on a ticket match the 6 numbers drawn by the Interprovincial Lottery

Corporation. For example, a ticket could contain the numbers 4, 7, 19, 34, 45, 49, and would win the jackpot if the same 6 numbers are drawn (in any order).

You may leave your answers to this question as a multiplication of fractions; it is not necessary to obtain a decimal value.

a) What is the probability that the *first* number drawn by the Lottery Corporation matches any of the 6 numbers on a Lotto 6/49 ticket?

Answer: Let M_i be the event that the ith number matches. Then: $P(M_1) = \boxed{\frac{6}{49}}$

b) If the first draw matches one of the ticket numbers, what is the probability that the second draw matches one of the remaining ticket numbers?

Answer: $P(M_2|M_1) = \boxed{\frac{5}{48}}$

c) Find the probability of winning the jackpot.

Answer: We want to find the probability that all 6 numbers match, in other words, $P_{jackpot} = P(M_6 \cap M_5 \cap M_4 \cap M_3 \cap M_2 \cap M_1)$. For ease of notation, let M_{1-j} denote the event that the first j numbers match. Applying the multiplication rule repeatedly gives:

$$\begin{aligned}
P_{jackpot} &= P(M_6|M_{1-5})P(M_{1-5}) \\
&= P(M_6|M_{1-5})P(M_5|M_{1-4})P(M_{1-4}) \\
&= \ldots \\
&= P(M_6|M_{1-5})P(M_5|M_{1-4})P(M_4|M_{1-3})P(M_3|M_{1-2})P(M_2|M_1)P(M_1) \\
&= \boxed{\left(\frac{1}{44}\right)\left(\frac{2}{45}\right)\left(\frac{3}{46}\right)\left(\frac{4}{47}\right)\left(\frac{5}{48}\right)\left(\frac{6}{49}\right)}
\end{aligned}$$

which can be simplified as 1/13983816 or calculated as 7.15×10^{-8} (though neither is required as stated by the question instructions.)

3. **[15]** A pollster conducts a poll of eligible, decided voters obtains information on the voting intentions of Canadians among the top three major national parties: Conservative, Liberal, and NDP. Among the 108 surveyed men, 53 intend to vote Conservative and 19 intend to vote NDP. Among surveyed women, 47 intend to vote Conversative and 38 intend to vote NDP. There are 93 Liberal supporters in the sample.

a) Create a two-way table using the above data. Use counts in the table rather than proportions.

Answer:

	Conservative	Liberal	NDP	Total
Male	53	36	19	108
Female	47	57	38	142
Total	100	93	57	250

The numer of male Liberal voters comes from $108 - 53 - 19 = 36$, the number of female Liberal voters then comes from $93 - 36 = 57$.

b) What is the marginal distribution of support for political parties in the sample?

Answer: The marginal distribution is: $100/250 = 0.400 = \boxed{40\%}$ vote Conservative, $93/250 = 0.372 = \boxed{37.2\%}$ vote Liberal, and $57/250 = 0.228 = \boxed{22.8\%}$ vote NDP.

c) Find the probability that a female voter responded with either Liberal or NDP support.

Answer: $P(\text{Liberal } \cup \text{ NDP}|\text{female}) = \frac{57+38}{142} = \boxed{0.669}$

d) Find the probability that a randomly selected voter in this sample is male and voted for either the Conservative or Liberal parties.

Answer: $\frac{53+36}{250} = \boxed{0.356}$

4. [10] Suppose that each student who takes Economics 250 has a 70% chance of passing the course.

a) What is the probability that at least 8 of a class of 10 students pass the course?

Answer: $\binom{10}{8}.7^8.3^2 + \binom{10}{9}.7^9.3^1 + \binom{10}{10}.7^{10} = 0.383$

b) What is the probability that at least 80 of a class of 100 students pass the course?

Answer: $P(X \geq 80) = P\left(z \geq \frac{80-70}{\sqrt{100(.7)(.3)}}\right) = P(z \geq 2.18) = 0.0146$

5. [20] A study involving 36 tenth grade students measures the students' IQ scores and measures a sample mean of 94.7. IQ scores are known to have a standard deviation of 15.

a) Test the hypothesis at the 95% confidence level that IQ scores equal 100 against the alternative the IQ scores are not equal to 100.

Answer: Use a z test:

$$z = \frac{94.7 - 100}{15/\sqrt{36}} = -2.12$$

This has a p-value of $2(.0170) = 0.034$, which is less than 0.05, so we reject the null hypothesis: there is evidence that the mean IQ score is different from 100.

b) What is the p-value of your test? Give an interpretation of this p-value.

Answer: See above: $p = 0.034$. This value tells us that, if the population mean IQ score actually equals 100, we have 3.4% chance of getting a sample as extreme as the one we found. In other words, there is a 3.4% chance that we are rejecting H_0 when it is true.

c) Calculate the power of this test when the population mean actually equals 98.

Answer: We reject when $|z| > 1.96$, which, from rearranging the z formula, corresponds to seeing a value of $\overline{\text{x}}$ more extreme than $\overline{\text{x}} = 100 \pm 1.96\frac{15}{\sqrt{36}}$, or in other words, a value of $\overline{\text{x}} < 95.1$ or $\overline{\text{x}} > 104.9$.

If the true value of $\mu = 98$,

$$P(\overline{\text{x}} < 95.1) = P\left(z < \frac{95.1 - 98}{2.5}\right) = P(z < -1.16) = 0.1230$$
$$P(\overline{\text{x}} > 104.9) = P\left(z > \frac{104.9 - 98}{2.5}\right) = P(z > 2.76) = 0.0029$$

and so the power of the test is $0.1230 + 0.0029 = 0.1259$.

d) Interpret your power value (in words).

Answer: We have about a 12.59% chance of rejecting the null hypothesis when the true population value actually equals 98 (which $\neq$ 100 and so the null hypothesis is actually false).

6. **[20]** You collect the following values of waiting times (in days) for ACL repair surgery in a major Canadian city.

107	106	117	162	81	127	156	107

a) Use this data to calculate a test statistic that will allow you to test the hypothesis that the mean waiting time equals 140 against the alternative that the mean waiting time is less than 140 days. What is the distribution of this test statistic?

Answer: We first need to calculate $\overline{\text{x}} = \frac{1}{8}(107 + 106 + 117 + \cdots + 107) = 120.375$ and $s = 27.1816$. Then we can get a t statistic:

$$t_7 = \frac{\overline{x} - 140}{s/\sqrt{8}} = -2.042113$$

This statistic follows a t distribution with 7 degrees of freedom.

b) Test the hypothesis of part *a)* at the $\alpha = 0.05$ level.

Answer: From Table D, our test statistic is between the critical values for $p = .025$ and $p = .05$ for the $df = 7$ row, so we conclude that our p-value is between these values: and since this range is less than 0.05, we reject the null hypothesis.

c) Construct a 95% confidence interval for the waiting time.

Answer: From Table D, we get a critical value of 2.365, and so our confidence interval is:

$$\mu \in [120.375 - 2.365(27.1816/\sqrt{8}), 120.375 - 2.365(27.1816/\sqrt{8})]\mu \in [97.647, 143.103]$$

Note that this includes 140, but that's because the confidence interval is always two-sided: we would fail to reject that the mean equals 140 in a two-sided test, even though we rejected it in a one-sided test.

7. **[20]** A survey is conducted using 47 six-year-old boys in a sub-Saharan African country. Participants are randomly assigned to one of two groups: the first group (of 21 boys) receives daily nutritional supplements over a four-year period, while the second group (26 boys) receives no treatment.

At the end of the four-year experiment, the height of the participants is measured. The first sample (the treatment group) has a mean height of 138cm with standard deviation of 15.7cm; the second sample has a mean height of 120cm and standard deviation of 13.6cm.

a) Test the hypothesis at the $\alpha = 0.05$ level that the nutritional supplement had no effect against the alternative that it increased heights.

Answer: Let A be the treatment group and B be the non-treatment group. Then we are testing:

$$H_0 : \mu_A - \mu_B = 0$$
$$H_a : \mu_A - \mu_B > 0$$

Use a t-test:

$$\begin{aligned} t &= \frac{(\overline{\mathrm{x}}_A - \overline{\mathrm{x}}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \\ &= \frac{18}{\sqrt{11.74 + 7.11}} \\ &= 4.146 \end{aligned}$$

Using even the most conservative rule for the degrees of freedom, 20, we get a p-value less than 0.0005 and so strongly reject H_0 in favour of H_a: the data contains strong evidence that the mean height was higher for the treatment group.

b) Construct a confidence interval at the 98% confidence level for the difference in mean heights.

Answer:

$$\mu_A - \mu_B \in \left[18 - t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}, 18 + t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}\right]$$
$$\mu_A - \mu_B \in \left[18 - t^* \sqrt{18.85}, 18 + t^* \sqrt{18.85}\right]$$

The value of t^* depends on which rule of thumb we use for the degrees of freedom; if using the smaller, we'll use $df = 20$ and get $t^* = 2.528$; using the Satterthwaite approximation, we get $df \approx 40$ and get $t^* = 2.423$.

This gives, for $df = 20$:

$$\mu_A - \mu_B \in [7.024, 28.976]$$

and for $df = 40$:

$$\mu_A - \mu_B \in [7.480, 28.520]$$

c) It is discovered that because of an incorrect formula in Excel, the data for the non-treatment group was actually divided by 1.1: the actual heights in the non-treatment sample need to be multiplied by 1.1.

Correct for this error and calculate the p-value for the test.

Answer: We can either write the hypotheses as:

$$H_0 : \mu_A - 1.1\mu_B = 0$$
$$H_a : \mu_A - 1.1\mu_B > 0$$

or we can define a new variable $C = 1.1B$ and use:

$$H_0 : \mu_A - \mu_C = 0$$
$$H_a : \mu_A - \mu_C > 0$$

As discussed in class, if we scale all of our observations by exactly the same factor, the mean and standard deviation will also be scaled by that factor, so $\overline{x}_C = 1.1\,\overline{x}_B = 132$ and $s_C = 1.1 s_B = 14.96$.

Thus our test becomes:

$$t = \frac{(\overline{x}_A - \overline{x}_C) - (\mu_A - \mu_C)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_C^2}{n_C}}}$$
$$t = \frac{6}{\sqrt{11.74 + 8.61}}$$
$$= 1.330$$

Whether we use $df = 20$ or $df = 40$ we get a p-value between 0.05 and 0.1.

8. **[20]** Sodexo, the company providing exclusive food service at Queen's, is considering opening a new Tim Hortons franchise at a newly-available location on campus. To help make the decision, they conduct a simple random sample of the campus population and ask the respondents whether people would visit the coffee shop at the proposed location at least once per week. 15 out of 109 responders indicate that they would use the new location regularly.

Since the proportions in this question are relatively close to 0, you should use Wilson's estimate (adding an appropriate number of fake positive and negative responses) throughout this question.

a) Sodexo management will only consider opening the new Tim Hortons if they are confident that more than 10% of the campus population will visit at least once per week. State and perform an appropriate test at the 95% confidence level to determine whether the location will attract the required number of customers.

Answer: Using Wilson's estimate's we get $\widehat{p} = \frac{15+2}{109+4} = 0.1504$. Our tests is:

$$H_0 : p = 0.1$$
$$H_a : p > 0.1$$

and so our test statistic is:

$$z = \frac{0.1504 - 0.1}{\sqrt{\frac{0.1(0.9)}{109}}} = 1.75$$

which has a p-value of 0.0401 (not doubled, since this is a one-sided test). Thus we reject H_0: there *is* significance evidence that the location will attract more than the required number of customers.

Sodexo performs a second simple random sample survey to determine whether another location is more suitable. 42 out of 170 participants in this survey indicate that they would visit the second location regularly.

b) Test the hypothesis at the $\alpha = 0.05$ level that the two locations will attract the same number of customers against the alternative that the second location will attract more customers.

Answer: Our hypotheses are:

$$H_0 : p_2 - p_1 = 0$$
$$H_a : p_2 - p_1 \neq 0$$

Since we're testing the null hypothesis of equal proportions, we need to pool to get a standard error. Note that we're still using Wilson's adjustments, but since we have two samples, we add one fake success and one fake failure to each sample.

$$\widehat{p}_{pool} = \frac{X_1 + X_2 + 2}{n_1 + n_2 + 4} = \frac{15 + 42 + 2}{109 + 170 + 4} = \frac{59}{283} = 0.2085$$

$$\begin{aligned} SE_{Dp} &= \sqrt{\widehat{p}_{pool}\,(1 - \widehat{p}_{pool})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \\ &= \sqrt{0.2085(0.7915)\left(\frac{1}{109} + \frac{1}{170}\right)} \\ &= 0.0498 \end{aligned}$$

and so the test statistic is:

$$\begin{aligned} z &= \frac{(\widehat{p}_2 - \widehat{p}_1) - 0}{0.0498} = \frac{\frac{43}{172} - \frac{16}{111}}{0.0498} \\ &= 2.13 \end{aligned}$$

which has a p-value of $0.0166 < \alpha = 0.05$, so we reject H_0 in favour of H_a: there is evidence that the second location will attract more customers.

c) Construct a 97% confidence interval for $p_2 - p_1$, the difference in customer proportions between the two locations.

Answer: We can't reuse the pooled standard error because we no longer have the null hypothesis that the samples are the same, so we need to use:

$$\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}} = \sqrt{\frac{(0.1441)(.8559)}{109} + \frac{(0.25)(0.75)}{170}}$$
$$= 0.0473$$

For a 97% confidence interval, we use $z^* = 2.17$, and so our confidence interval for the difference is:

$$\left[\left(\frac{43}{172} - \frac{16}{111}\right) - 2.17(0.0473), \left(\frac{43}{172} - \frac{16}{111}\right) \pm 2.17(0.0473)\right]$$
$$= [0.0032, 0.2085]$$

9. **[15]** A researcher is studying the performance of high school students at different schools using the following regression model:

$$math10 = \beta_1 + \beta_2 totcomp + \beta_3 enroll + u$$

where:

- $math10$ is the percentage (from 0 to 100) of a school's students who pass a standardized grade 10 math test
- *totcomp* is the average compension (salary plus benefits) of the school's teachers, measured in dollars
- *enroll* is the number of students enrolled at the school

Using a data set of $n = 408$ randomly selected schools, the researcher uses a linear regression program which outputs the following values:

	Coefficient	Std. Error
const	8.320	3.487
totcomp	0.0004244	0.00009630
enroll	-0.0001658	0.0002137

R^2	0.050671	Adjusted R^2	0.045983
$F(2, 405)$	10.80850	P-value(F)	0.000027

a) Provide an economic interpretation of the $\widehat{\beta}_2$ coefficient.

Answer: $\widehat{\beta}_2 = 0.0004244$ indicates that, holding the enrollment level constant, schools with higher average compensation paid to teachers increase average math test passing rates by 0.0004244% per extra dollar of compensation (or equivalently, 0.4255% higher for each extra thousands of dollars of compensation).

b) Perform a hypothesis test to determine whether the data provides evidence at the $\alpha = 0.1$ level that larger schools has a negative effect on the grade 10 math test performance of students. What is the distribution and p-value of your test statistic?

Answer: We want to test:

$$H_0 : \beta_3 = 0$$
$$H_a : \beta_3 < 0$$

and so we use a t-test:

$$t = \frac{-0.0001658 - 0}{0.0002137} = -0.775$$

This follows a t_{405} distribution. Using the table for $df = 100$, we find a p-value between 0.2 and 0.25. Since this is greater than 0.1, we fail to reject the null hypothesis: the data does not provide significant evidence that larger schools have lower passing rates.

c) What information does the $R^2 = 0.0507$ value provide?

Answer: It tells us that this model with this sample data is able to explain about 5.07% of the variation in the sample's *math*10.

d) One of the schools has $math10 = 51.1$, $totcomp = 39992$, and $enroll = 1116$. Calculate the residual for this observation. Does the fitted model predict a value for this observation that is too high or too low?

Answer: To find a residual, we first need the fitted value: $\widehat{math10} = 8.320 + 0.0004244(39992) - 0.0001658(1116) = 25.1$

The residual is $\widehat{u} = y - \widehat{y} = 51.1 - 25.1 = 26.0$.

The large, positive residual indicates that the model predicts a passing rate for this school that is much too low.