Economics 250 — Midterm 1 **17 October 2013**

Instructions: You may use a hand calculator. Do not hand in the question and formula sheets. Answer all **three** questions in the answer booklet provided. Show your work: incorrect answers without any work shown cannot be given partial marks. Formulas and tables are provided at the end of the question pages; you may wish to detach these from the question pages for easier reference.

1. You obtain the following sample of 10 Canadian household incomes (in thousands of dollars): 67 67 66 64 68 73 69 77 90 69

a) Find the mean and median of the data.

b) Find the first and third quartiles of the data.

c) Find the sample standard deviation.

d) The data is actually *grouped data* of mean household incomes of the 10 Canadian provinces. Calculate the mean household income of the *last four* data points—the western provinces—by also using the following population counts (in millions):

	Manitoba	Saskatchewan	Alberta	British Columbia
Household income:	69	77	90	69
Population:	1.3	1.1	3.8	4.6

Answers:

Sorting the data first helps (for the median/quartiles):

64 66 67 67 68 69 69 73 77 90

a)

$$\overline{x} = \frac{1}{n}\sum x_i = \frac{710}{10} = 71$$
$$x_{\text{median}} = \frac{68+69}{2} = 68.5$$

b)

$$Q_1 = 67$$
$$Q_3 = 73$$

c)

$$
\begin{aligned}
s_x^2 &= \frac{1}{n-1}\sum(x_i-\overline{x})^2 \\
&= \frac{524}{9} \\
&= 58.22222 \\
s_x &= \sqrt{s_x^2} \\
&= 7.630349
\end{aligned}
$$

d) Using c_i to denote the population counts and m_i to denote the means:

$$
\begin{aligned}
\overline{x} &= \frac{\sum c_i m_i}{\sum c_i} \\
&= \frac{1.3\times 69 + 1.1\times 77 + 3.8\times 90 + 4.6\times 69}{1.3+1.1+3.8+4.6} \\
&= \frac{833.8}{10.8} \\
&= 77.2037
\end{aligned}
$$

Note that this is *not* the same as simple mean (which gives incomes equal weight) of 76.25.

2. Suppose that, in any given week, the price of gas per litre (in dollars), denoted X, is a random variable that follows a normal distribution with mean 1.2 and standard deviation 0.1.

a) Suppose that it costs you \$2 worth of your time, plus the cost of the gas, to fill up your car which takes 40 litres of gas. Denote the total cost of a fill-up as Y, and write down an expression for Y in terms of X.

b) Find the mean, variance, and standard deviation of Y.

c) Find the probability that, on a random week, the cost of a fill-up (i.e. the value of Y) will be:

(*i*) Below \$50.

(*ii*) Between \$42 and \$55.

(*iii*) Below \$42 or above \$55.

(*iv*) Exactly \$47 (assume no rounding of decimal prices).

d) A researcher obtains 25 observations of the gas price (i.e. observations of X, not of Y) and calculates the mean of those observations. Find the probability that the sample mean will be less than or equal to 1.23.

Answers:

a) $Y = 2 + 40X$

b)

$$\mu_Y = 2 + 40\mu_X = 2 + 40(1.2) = 50$$
$$\sigma_Y^2 = 0 + 40^2\sigma_X^2 = 1600(0.01) = 16$$
$$\sigma_Y = \sqrt{16} = 4$$

c) *(i)*

$$P(Y < 50) = P(z < \frac{50 - 50}{4}) = 0.5$$

(ii)

$$\begin{aligned} P(42 < Y < 55) &= P(Y < 55) - P(Y < 42) \\ &= P(z < \frac{55 - 50}{4}) - P(z < \frac{42 - 50}{4}) \\ &= P(z < 1.25) - P(z < -2.00) \\ &= 0.8944 - 0.0228 \\ &= 0.8716 \end{aligned}$$

(iii)

$$P(Y < 42 \text{ or } Y > 55) = 1 - P(42 < Y < 55) = 0.1384$$

This makes use of the fact that it is simply the complement of the previous answer. It could also be worked out explicitly, as follows:

$$\begin{aligned} P(Y < 42 \text{ or } Y > 55) &= P(Y < 42) + P(Y > 55) \\ &= P(Y < 42) + 1 - P(Y < 55) \\ &= P(z < -2.00) + 1 - P(z < 1.25) \\ &= 0.0228 + 1 - 0.8944 \\ &= 0.1284 \end{aligned}$$

(iv) $P(Y = 47) = 0$

d) According to the Central Limit Theorem learned in class,[1] the distribution of the mean of a sample of size 25 is:

$$\overline{x}_{25} \sim \text{N}(1.2, \frac{0.1}{\sqrt{25}}) = \text{N}(1.2, 0.02)$$

and so:

$$P(\overline{x}_{25} \leq 1.23) = P(z \leq \frac{0.03}{0.02}) = P(z \leq 1.5) = 0.9332$$

[1] On a technical note, though not something we learned in class, we don't actually need the Central Limit Theorem at all here: the sum of independent normals is always normal. Since the sample mean is just the sum of normals (divided by n), the "approximate" result of the CLT actually holds exactly, for all values of n, when the underlying distribution is normal.

3. A researcher studying the effect of education on income collects data on whether or not the employees of a large Canadian corporation, BlueBerry, earned a university degree, and classifies the worker as having either a high-skill or low-skill job. The researcher uses a simple random sample to select 50 of the corporation's employees to interview.

Of the 50 employees surveyed, 18 have university degrees. Of those 18, 14 have high-skill jobs. Of the remaining surveyed employees, 15 have high-skill jobs.

a) What is the population of the sample? Will the sample be appropriate for determining the relationship between university degrees and job skill levels among all Canadian workers?

b) Form a two-way table of counts summarizing the data. Include a "Total" column and row in your table.

c) Find the distribution of the skill variable conditional on having a university degree, and, separately, the distribution of the skill variable conditional on *not* having a university degree. Based on your results, is there a relationship between having a university degree and having a high-skill job for workers of this corporation?

d) The researcher conducts the same survey at another Canadian corporation, Tom Hurten's, and estimates the following probabilities, where H indicates a high-skill job, L indicates a low-skill job, U indicates having a university education, and N indicates not having a university education:

- $P(N \cap H) = 0.20$
- $P(N \cap L) = 0.60$
- $P(U \cap H) = 0.05$

(*i*) Find $P(U \cap L)$, $P(U)$, $P(H)$, and $P(U \cup H)$.

(*ii*) Find $P(H \mid U)$. Are having a university degree and having a high-skill job independent? Be sure to show a condition that is satisfied or not satisfied to support your answer.

Answers:

a) The population is workers of the corporation. It is not appropriate to use the results to infer details about the population at large because the workers of this corporation are not likely to be a random sample of Canadians.

b)

	High-skill	Low-skill	Total
Degree	14	4	18
No degree	15	17	32
Total	29	21	50

c) For workers with university degrees: $\frac{14}{18} = 77.8\%$ have high-skill jobs, 22.2% have low-skill jobs.

For workers without university degrees: $\frac{15}{32} = 46.9\%$ have high-skill jobs, 53.1% have low-skill jobs.

There appears to be a positive relationship: workers with university degrees are more likely than those without university degrees to have high-skill jobs.

d) (*i*) Since all the probabilities have to add up to 1:

$P(U \cap L) = 1 - P(N \cap H) - P(N \cap L) - P(U \cap H) = 0.15.$

$P(U)$ is the probability of all outcomes satisfying U, thus $P(U) = P(U \cap L) + P(U \cap H) = 0.2$.

Similarly, $P(H) = P(U \cap H) + P(N \cap H) = 0.25$.

From the adding rule for probabilities:

$P(U \cup H) = P(U) + P(H) - P(U \cap H) = 0.2 + 0.25 - 0.05 = 0.4.$

(*ii*) From the formula for conditional probability:

$$P(H|U) = \frac{P(H \cap U)}{P(U)} = \frac{0.05}{0.2} = 0.25$$

Because $P(H|U) = P(H)$, H and U are independent. You could also check this by checking that $P(H \cap U) = P(H)P(U)$, which is satisfied.

It isn't necessary,[2] but you could double-check independence by seeing whether $P(U|H) = P(U)$: $P(U|H) = \frac{0.05}{0.25} = 0.2 = P(U)$, verifying that H and U are independent.

[2] You can use Bayes Rule plus a couple of steps of algebra to show that $P(A|B) = P(A)$ and $P(B|A) = P(B)$ are equivalent statements.