Queen's University
Faculty of Arts and Sciences
Department of Economics
Fall 2007

Economics 250 A: Introduction to Statistics

Midterm Exam I:
Time allotted: 120 minutes.

Instructions:

READ CAREFULLY. Calculators are permitted. At the end of the exam are several formulae. Answers are to be written in the examination booklet. Do not hand in the question sheet. You are to answer **ALL** questions. **SHOW ALL YOUR WORK**. Most of the marks are awarded for showing how a calculation is done and not for the actual calculation itself!

There are a total of 100 possible marks to be obtained. Answer all 9 questions (marks are indicated)

Do all 9 questions

1. **(15 marks)** The age distribution of subscribers and purchase frequencies to two magazines are shown in the following table.

	10-19	20-29	30-39	40-49	50-59	60-69	70-79
Magazine As	9	533	827	208	59	13	2
Magazine Bs	29	533	825	1208	2000	2900	73

   (a) Set up the two relative frequency distributions.
   (b) Use them to make a recommendation to a client about placing an advertisement for baby food.
   (c) How about for denture cleaners?

   ANSWER: a) Relative frequency:

Age groups	A	B
10-19	0.005	0.004
20-29	0.323	0.07
30-39	0.501	0.109
40-49	0.126	0.16
50-59	0.036	0.264
60-69	0.008	0.383
70-79	0.001	0.01
	1	1

b) Baby food: Magazine A, because a large percentage of subscribers are found in the young groups (between 20 and 39, 82.4%) vs 17.9% for magazine B). Denture cleaners: magazine B, because a larger percentage of subscribers are in the older groups (between 50 and 79, 65.7% vs. 4.5%for magazine A).

2. **(10 marks)** A different summation is to be performed on some of the numbers that constitute the elements of the set

$$A = \{w : w = 17, 4, 6, 12, 8, 1, 8\}$$

Find the following sums.

(a) What is $\sum_{i=4}^{7}(w_i - \overline{w}$, where $\overline{w} = \frac{1}{n}\sum_{i=1}^{7} w_i$. Why does this not equal to zero?

(b) What is $\sum_{i=1}^{4} w_i^2$ and $(\sum_{i=1}^{4} w_i)^2$? Why are these not the same.

ANSWER:

a) $\overline{w} = (17 + 4 + 6 + 12 + 8 + 1 + 8)/7 = 8$, $\sum_{i=4}^{7}(w_i - \overline{w}) = -3$

b) $\sum_{i=1}^{4} w_i^2 = 17^2 + 4^2 + 6^2 + 12^2 = 485$, and $(\sum_{i=1}^{4} w_i)^2 = (17 + 4 + 6 + 12)^2 = 1521$

3. **(10 marks)** Considering the accompanying sample data on the R&D spending of five U.S. firms.

Firms	In Total Dollars	Firms	In percent of sales
General Motors	2250	Telesciences	22.1
Ford Motor	1718	Kulicke& Soffa	18.9
AT&T	1686	Computer Consoles	17.8
IBM	1612	Auto-Trol	17.2
Boeing	844	Amdahl	17

(a) Calculate the mean, median, and mode for each list.

(b) What is the interquartile range and mean absolute deviations for each list? Interpret your results.

ANSWER:

a) $mean_1 = 1622$, $median_1 = 1686$, and no mode. $mean_2 = 18.6\%$, $median_2 = 17.8\%$, and no mode.

b) For list 1, Q_1 is located at $0.25(n+1) = 1.5$ position, so $Q_1 \simeq 1228$, Q_3 is located at $0.75(n+1) = 4.5$ position, so $Q_3 \simeq 1984$, So $IRQ = Q_3 - Q_1 = 1984 - 1228 = 756$.

For list 2, $IRQ = Q_3 - Q_1 = 20.5 - 17.1 = 3.4$.

The interquartile range (IRQ) measures the spread in the middle 50% of the data.

$$MAD_1 = \frac{|2250 - 1622| + |1718 - 1622| + |1686 - 1622| + |1612 - 1622| + |844 - 1622|}{5} = 315.2$$

$$MAD_2 = \frac{7.6}{5} = 1.52$$

4. **(10 marks)** An oil company will drill two wells in a new area. The probability that the first well will be a success is 3/10. If the first well succeeds, the probability that the second well will be a success is 4/10. If the first well fails, the probability that the second well will be a success is 2/10. Using the designations S_1, and F_1 and S_2 and F_2 to succeed and fail on the first and second wells, respectively, find the following probabilities. (Hint: Do each intersection probability separately and then consider union.)

   (a) $P[(S_1 \cap F_2) \cup (F_1 \cap S_2) \cup (S_1 \cap S_2)]$
   (b) $P[S_1 \cup (F_1 \cap S_2)]$
   (c) $P(\overline{F1 \cap F2})$

ANSWER: Since $P(S_1) = 3/10$, $P(S_2|S_1) = 4/10$, $P(S_2|F1) = 2/10$.

a) $P[(S_1 \cap F_2) \cup (F_1 \cap S_2) \cup (S_1 \cap S_2)] = P(S_1)P(F_2 \mid S_1) + P(F_1)P(S_2 \mid F_1) + P(S_1)P(S_2 \mid S_1) = (3/10)(6/10) + (7/10)(2/10) + (3/10)(4/10) = 44/100$

b) $P[S_1 \cup (F_1 \cap S_2)] = P(S_1) + P(F_1)P(S_2 \mid F_1) = (3/10) + (7/10)(2/10) = 44/100$

c) $P(\overline{F1 \cap F2}) = 1 - (7/10)(8/10) = 44/100$

5. **(15 marks)** An analysis of a recent labor-union vote on a new contract shows this: In the Northeast, 5000 of 20000 members voted yes; in the Southeast 3000 of 7000 members voted no; in the Northwest, 4000 of 10000 members voted yes; and in the Southwest, 13000 of 21000 members voted yes. (There were no abstentions.)

   Calculate these probabilities and In each case, identify the nature of the probabilities involved.

   (a) write out information as a joint probability table and show all marginal probabilities.
   (b) A randomly chosen member of this union was from the Northeast and vote no.
   (c) A randomly chosen no-voter was from Southwest.
   (d) A randomly chosen member from the Northwest or Southwest voted yes.
   (e) A randomly chosen member voted yes.

ANSWER: a)

	Yes	NO	Total
NE	5000	15000	20000
SE	4000	3000	7000
NW	4000	6000	10000
SW	13000	8000	21000
Total	26000	32000	58000

Accordingly, the probabilities are the following:

b) $P(NE \text{ } and \text{ } N) = (15000/58000) = 0.26$; joint probability

c) $p(SW \mid N) = (8000/32000) = 0.25$; conditional probability

d) $P(Y \mid NW \text{ } or \text{ } SW) = (17000/31000) = 0.55$; conditional probability

e) $P(Y) = (26000/58000) = 0.45$; marginal probability

6. **(10 marks)** We have two independent portfolios called A and B with the following probability.

A	P(A=a)	B	P(B=b)
1%	0.4	1%	0.3
2%	0.6	2%	0.7

(a) What is the expected return of portfolio A and B?

(b) What is expected return of $0.5A + 0.5B$ portfolio?

(c) What is the variance of $0.5A + 0.5B$ portfolio?

(d) calculate coefficient variation of A, B, and $0.5A + 0.5B$?

ANSWER:

a) $E(A) = 0.016$, $E(B) = 0.017$.

b) $E(0.5A + 0.5B) = 0.0165$

c)

$$var(A) = (0.01 - 0.016)^2 * 0.4 + (0.02 - 0.016)^2 * 0.6 = 0.000024$$
$$var(B) = (0.01 - 0.017)^2 * 0.3 + (0.02 - 0.017)^2 * 0.7 = 0.000021$$

$$\begin{aligned} var(0.5A + o.5B) &= 0.25 * var(A + B) \\ &= 0.25 * (var(A) + var(B)) \\ &= 0.00001125 \end{aligned}$$

d) $cv(A) = \sqrt{var(A)}/0.016 = 30.6\%$

$cv(B) = \sqrt{var(B)}/0.017 = 27\%$

$cv(0.5A + 0.5B) = \sqrt{var(0.5A + 0.5B)}/0.0165 = 20.3\%$

7. **(10 marks)** Mr. Jones manages a home office department for an insurance company. The department reviews applications for insurance and issues policies to those who qualify. Ninety percent of the policies are issued without giving the applicant a chance to verify the information in the application (by mail). In this case, 60 percent of the applicants accept the policy as issued, 15 percent reject it, and the remainder return the policy for changes. Of those policies where applicants have the opportunity for verification, 80 percent accept the policy, 3 percent reject it, and 17 percent request changes. (Use the symbols V and NV for opportunity to verify and no opportunity to verify, respectively, and A, R, and C for accept, reject, and change.)

   (a) Write out all information given in the problem.

(b) Calculate the probability that each complete sequence will be pursued by an incoming policy application. (Hint: one of these sequence is $(V \cap A)$.)

(c) Mr. Jones has just been handed a policy that was rejected by the applicant. What is the probability that it was not verified prior to issuance.

ANSWER:

a) $P(V \cap A) = 0.1 * 0,8 = 0.08$, $P(V \cap R) = 0.1 * 0.03 = 0.003$, $P(V \cap C) = 0.1 * 0.17 = 0.017$, $P(NV \cap A) = 0.9 * 0.6 = 0.54$, $P(NV \cap R) = 0.9 * 0.15 = 0.135$, and $P(NV \cap C) = 0.9 * 0.25 = 0.225$

b) The condition is a rejected policy (R), and we want the probability of nonverification (NV) subject to the condition (R).

$$\begin{aligned} P(NV|R) &= (P(NV \cap R))/(P(R)) = P(NV \cap R)/(P(V \cap R) + P(NV \cap R)) \\ &= \frac{0.135}{0.003 + 0.135} \\ &= 0.978 \end{aligned}$$

8. **(10 marks)** Professor Gregory is mean and gives 5 random attendance checks in 12 classes.

(a) What are the features of a binomial distribution.

(b) What is the probability that all 5 attendance checks are done in the first 8 classes.

(c) Is this a binomial problem.

ANSWER:

a) To apply the binomial distribution, three conditions must hold.

1. There are a fixed number of trials of an experiment with only two possible outcomes for each trial: ‘success’ or ‘failure’.
2. The probability of success on any trial is constant.
3. The outcome of each trial is independent of every other trial.

b) $P(A \cap B) = P(A = checks\ in\ first\ 7\ class \cap B = checks\ in\ the\ last\ 8th\ class)$

$= \binom{12}{7} \frac{5}{12}^7 \frac{7}{12}^5 \times \frac{5}{12}$

c) Yes.

9. **(10 marks)** In Ontario, there are three major political parties, Liberal (L), Conservative (C), and NDP. The historical support of the Liberal and Conservative is 50% and 40% respectively. Suppose that 30% of NDP supports favor first past the post votary system (F). 80% of Conservative and 70% of Liberal face first past the pooh. If a voter is selected at random and faces proportional representative (the other system R), what is the probability that she is a Conservative.

ANSWER: Since $P(L) = 0.5$, $P(C) = 0.4$, $P(NDP) = 0.1$, $P(F|NDP) = 0.3$, $P(F|C) = 0.8$, and $P(F|L) = 0.7$, then $P(R|NDP) = 0.7$, $P(R|C) = 0.2$, and $P(R|L) = 0.3$

$$\begin{aligned} P(C|R) &= \frac{P(R|C)P(C)}{P(R)} \\ &= \frac{0.2 \times 0.4}{P(R|C)P(C) + P(R|L)P(L) + P(R|NDP)P(NDP)} \\ &= \frac{0.2 \times 0.4}{0.2 \times 0.4 + 0.3 \times 0.5 + 0.7 \times 0.1} \\ &= 0.2667 \end{aligned}$$

Formula Sheet

Statistics Formulas

Notation

- All summations are for $i = 1, \ldots n$ unless otherwise stated
- $\sim$ means 'distributed as'

Population Mean

$$\mu \;=\; \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sample Mean

$$\bar{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Mean Absolute Deviations (MAD)

$$MAD = \frac{1}{n}\sum_{i=1}^{n} \mid X_i - \overline{X} \mid$$

Population Variance

$$\begin{aligned}\sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}[x_i \;-\; \mu]^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} {x_i}^2 - \mu^2\end{aligned}$$

Sample Variance

$$s^2 \quad = \tfrac{1}{n-1}\sum_{i=1}^{n}[x_i \, - \, \bar{X}]^2$$

Alternatively

$$s^2 = \tfrac{1}{n-1}[\sum_{i=1}^{n} x_i^2 \, - \, n\bar{X}^2]$$
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

Grouped Data (with k classes)

$$\bar{X} \quad = \quad \frac{1}{n}\sum_{j=1}^{k}\nu_j f_j \quad \textit{where } \nu_j \textit{ is the class mark for class } j$$
$$s^2 \quad = \quad \frac{1}{n-1}\sum_{j=1}^{k} f_j(\nu_j - \bar{X})^2$$

Probability Theory

$$P(A \cup B) \quad = \quad P(A) + P(B) - P(A \cap B) \quad (\textit{additive law})$$
$$P(A \cap B) \quad = \quad P(B)P(A \mid B) \quad (\textit{multiplicative law})$$

If E_i is mutually exclusive and exhaustive events for $i = 1, \ldots n$, then

$$P(A) \quad = \quad \sum_{i}^{n} P(A \cap E_i) = \sum_{i}^{n} P(E_i)P(A \mid E_i)$$
$$P(E_i \quad \mid \quad A) = \frac{P(E_i)P(A \mid E_i)}{P(A)} \quad (\textit{Bayes's Theorem})$$

Counting Formulae

$$ {}_N P_R = \frac{N!}{(N-R)!} $$

$$ {}_N C_R = \binom{N}{R} = \frac{N!}{(N-R)!R!} $$

Random Variables

Let X be a discrete random variable, then:

$$
\begin{aligned}
\mu_x &= E[X] = \sum_x xP(X = x) \\
\sigma_x^2 &= V[X] = E[(x - \mu_x)^2] = \sum_x (x - \mu_x)^2 P(X = x) \\
\sigma_x^2 &= E[X^2] - (E[X])^2
\end{aligned}
$$

If X and Y are independent random variables and a, b, and c are constants, then:

$$
\begin{aligned}
E[a + bX + cY] &= a + b\mu_x + c\mu_y \\
V[a + bX + cY] &= b^2\sigma_x^2 + c^2\sigma_y^2
\end{aligned}
$$

Coefficient of Variation (CV)

$$
\begin{aligned}
CV &= \frac{\sigma}{\mu} \times 100\% \; for\ population \\
CV &= \frac{s}{\bar{X}} \times 100\% \; for\ sample
\end{aligned}
$$