Queen's University
Faculty of Arts and Sciences
Department of Economics
Winter 2008

Economics 250 : Introduction to Statistics

Midterm Exam I:
Time allowed: 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. **(10 marks)** There are two stocks A and B. Twelve samples for each stock are taken at the same time in a month.

   Stock A: 21 22 27 22 18 19 23 28 36 33 22 19

   Stock B: 40 55 62 50 43 55 58 60 65 42 45 55

   (a) What is the mean stock price for each sample?

   (b) Find the range of the stock price for each sample?

   (c) What is the standard deviation of the stock price for each sample?

   (d) Calculate the C.V for each stock and which stock should you purchase? Explain.

   (e) Supposed two stock A and B are independent discrete random variables. If the value X of the portfolio can be represented by the linear combination $X = 0.8A + 0.2B$, what is the expected value of X and the variance of X.

ANSWER:

a) $\overline{A} = 24.17$, $\overline{B} = 52.5$.

b) $range_A = 18$, $range_B = 25$

c) $SD_A = 5.7$, $SD_B = 8.372$.

d) $CV_A = 0.236$, $CV_B = 0.1595$. Stock B will be preferred to.

e) $\overline{X} = 0.8\overline{A}+0.2\overline{B} = 0.8*24.17+0.2*52.5 = 19.336$, $V(X) = 0.8^2*5.7^2+0.2^2*8.372^2 = 23.5972$.

2. **(10 marks)** Two variables X and Y assume the values

$X : \{2,\ -5,\ 4,\ -8\}$

$Y : \{-3,\ -8,\ 10,\ 6\}$

Calculate

(a) $\sum_2^4 (X_i + Y_i)(X_i - Y_i)$.

(b) $\sum_{i=1}^4 X_i \sum_{i=1}^4 Y_i$ and $\sum_{i=1}^4 X_i Y_i$, are those two the same? Why?

(c) $\sum_{i=2}^3 Y_i^2$ and $(\sum_{i=2}^3 Y_i)^2$, are those two the same?

ANSWER:

a) $\sum_2^4 (X_i + Y_i)(X_i - Y_i) = -151$

b) $\sum_{i=1}^4 X_i \sum_{i=1}^4 Y_i = -35$, $\sum_{i=1}^4 X_i Y_i = 26$.

c) $\sum_{i=2}^3 Y_i^2 = 164$, $(\sum_{i=2}^3 Y_i)^2 = 4$.

3. **(10 marks)** A corporation regularly takes deliveries of a particular sensitive part from three subcontractors. It found that the proportion of parts that are good or defective from the total received were as shown in the following table:

	Subcontractor		
Part	A	B	C
Good	0.27	0.30	0.33
Defective	0.02	0.05	0.03

(a) If a part is chosen randomly from all those received, what is the probability that it is defective?

(b) If a part is chosen randomly from all those received, what is the probability it is from subcontractor B?

(c) What is the probability that a part from subcontractor B is defective?

(d) What is the probability that a randomly chosen defective part is from subcontractor B?

(e) Is the quality of a part independent of the source of supply?

(f) In terms of quality, which of the three subtractors is most reliable?

ANSWER:

a) $P(D) = 0.02 + 0.05 + 0.03 = 0.10$

b) $P(B) = 0.3 + 0.05 = 0.35$

c) $P(D|B) = \frac{P(D \bigcap B)}{P(B)} = \frac{0.05}{0.35} = 0.1429$

d) $P(B|D) = \frac{P(D \bigcap B)}{P(D)} = \frac{0.05}{0.1} = 0.5$

e) No. For instance, events $defective$ and subcontractor B are not independent because $P(D|B) \neq P(D)$ or $P(B|D) \neq P(B)$.

f) Subtractor C because $P(C \bigcap Good) = 0.33 > P(B \bigcap Good) > P(A \bigcap Good)$.

4. **(15 marks)** Define the expectation of X, $E(X) = \mu_x$ and the variance of X, $V(X) = \sigma_x^2$. We can show relationship between the expectation and variance is

$$V(X) = E[X^2] - \mu_x^2 \quad (1)$$

Given the probability distribution function

x	0	1	2
Probability	0.2	0.7	0.1

(a) Draw the probability distribution function.

(b) Calculate and draw the cumulative probability function.

(c) Find the mean of the random variable X.

(d) Find the variance of X.

(e) Proof the equation (1).

(f) Evaluate your results of c) and d) by using the equation (1).

ANSWER:

c) d)

x	$P(X=x)$	$xP(X=x)$	$(x-\mu_x)^2 P(X=x)$	$x^2 P(X=x)$
0	0.2	0	0.162	0
1	0.7	0.7	0.007	0.7
2	0.1	0.2	0.121	0.4
		$E(X)=0.9$	$V(X)=0.29$	$E(X^2)=1.1$

e)

$$\begin{aligned} V(X) = E(X-E(X))^2 &= E(X^2 - 2XE(X) + (E(X))^2) \\ &= EX^2 - 2E(X)E(X) + (E(X))^2 \\ &= EX^2 - (E(X))^2 \\ &= EX^2 - \mu_x^2 \end{aligned}$$

f) By using the data in the previous table, we have

$$\begin{aligned} V(X) &= EX^2 - \mu_x^2 \\ 0.29 &= 1.1 - 0.9 \end{aligned}$$

5. **(10 marks)** An urn contains r red and b blue balls. We choose at random n balls without replacement. What is the probability that there will be exactly k red balls chosen?

ANSWER: The choice of n objects without replacement is $\binom{r+b}{n}$. We must have k red balls, which is choosing by $\binom{r}{k}$ ways. We must also have $n-k$ blue balls, which is in $\binom{b}{n-k}$ ways. So total number of choices is the product $\binom{r}{k}\binom{b}{n-k}$. Consequently, the probability in the question is

$$P(exactly\ \ k\ \ red\ \ balls) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}}$$

6. **(15 marks)** Suppose we have the set of values

$$X: \quad 1,\ 3,\ 5,\ 7,\ 9.$$

Perform the following:

(a) Calculate the mean and variance of X.

(b) Make the transformation $Y = 7 + 8X$.

(c) Calculate the mean and variance of Y.

(d) Note the relations between the respective means and variances.

ANSWER:

a) We have

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(25) = 5$$

and

$$\begin{aligned} s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \frac{1}{5}[(1-5)^2 + (3-5)^2 + \ldots + (9-5)^2] \\ &= \frac{40}{5} = 8. \end{aligned}$$

b) Under the transformation, we have

$$\begin{aligned} y_1 &= 7 + 8x_1 = 15, \\ y_2 &= 7 + 8x_2 = 31, \\ y_3 &= 7 + 8x_3 = 47, \\ y_4 &= 7 + 8x_4 = 63, \\ y_5 &= 7 + 8x_5 = 79. \end{aligned}$$

Thus, the values of Y become

$$Y: \; 15, \; 31, \; 47, \; 63, \; 79.$$

c) We have

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{5}(15 + 31 + \ldots + 79) = \frac{235}{5} = 47$$

and

$$\begin{aligned} s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{x})^2 &= \frac{1}{5}[(15-47)^2 + (31-47)^2 + \ldots + (79-47)^2] \\ &= \frac{1}{5}(1024 + 256 + \ldots + 1024) = \frac{2560}{5} = 512. \end{aligned}$$

d) The linear transformation involves an additive constant of 7 and a multiplicative constant of 8. Very careful comparison reveals that the mean of Y can be written in terms of the mean of X as

$$\overline{y} = 7 + 8\overline{x} = 7 + 8(5) = 47.$$

Also, the variance of Y can be written in terms of the variance of X as

$$s_y^2 = (8)^2 s_x^2 = 64(8) = 512.$$

Note that the mean is affected by both the additive and multiplicative constants, while the variance is affected only by (the square of) the multiplicative constant.

7. **(10 marks)** A survey of 500 students taking one or more courses in Algebra, Economics and Statistics during one semester revealed the following numbers of students in the indicated subjects.

Algebra 329	Algebra and Economics 83
Economics 186	Algebra and Statistics 217
Statistics 295	Economics and Statistics 63

Using Venn diagram to illustrate the relationship among those three subjects. How many students were taking

(a) all three subjects. What is the probability of a student taking all 3 subjects.

(b) Algebra but not Statistics.

(c) Economics but not Algebra.

(d) Algebra or Statistics but not Economics.

(e) Algebra but not Economics or Statistics.

ANSWER:

Let A denote the set of all students taking algebra, and denote by (A) the number of students belonging to this set. Similarly, let (B) denote the number taking Economics and (C) the number taking Statistics. Then (A+B+C) denotes the number taking either Algebra, Economics or Statistics or combinations, (AB) the number taking both Algebra and Economics, etc. a) We have

$$\begin{aligned}(A+B+C) &= (A)+(B)+(C)-(AB)-(BC)-(AC)+(ABC) \\ 500 &= 329+186+295-83-63-217+(ABC) \\ (ABC) &= 53\end{aligned}$$

The probability of a student taking all 3 subjects is 53/500.

b) $329 - 217 = 112$ or $82 + 30 = 112$.

c) $83 + 10 = 103$.

d) $82 + 164 + 68 = 314$.

e) 82.

8. **(10 marks)** Suppose that in a particular statistics course given by a large University, 60% of the students have good grades and 40% have poor grades. It is known that , on the average, the good students are late to class only 5% of the time while poor students are late about 15% of the time. If a student selected at random from the course comes in late on a particular day, what is the probability that he is a good student?

ANSWER:

We know $P(G) = 0.6$, $P(poor) = 0.4$, $P(L|G) = 0.05$, and $P(L|poor) = 0.15$, then

$$\begin{aligned} P(G|L) &= \frac{P(L|G)P(L)}{P(L)} \\ &= \frac{P(L|G)P(G)}{P(L|G)P(G) + P(L|poor)P(poor)} \\ &= \frac{0.05 * 0.6}{0.05 * 0.6 + 0.15 * 0.4} \\ &= 0.33 \end{aligned}$$

9. **(10 marks)** The following table illustrates federal personal income tax rates for different years. Read the table carefully
   (a) Compare the tax rate in 1974 and 1978. What conclusions you can draw from this observation.
   (b) Using weighted sample methods to confirm your conclusions.

Table 1: Total Income and Total Tax (in $), and Tax Rate for Taxable Income Tax Return, by Income Category and Year

		1974				
Adjusted Gross Income	Income	Tax	Tax Rate	Income	Tax	Tax Rate
under $5,000	41,651,643	2,244,467	.054	19,879,622	689,318	.035
5,000 to $9,999	146,400,740	13,646,348	.093	122,853,315	8,819,461	.072
10,000 to $14,999	192,688,922	21,449,597	.111	171,858,024	17,155,758	.100
15,000 to $99,999	470,010,790	75,038,230	.160	865,037,814	137,860,951	.159
100,000 or more	29,427,152	11,311,672	.384	62,806,159	24,051,698	.383
Total	880,179,247	123,690,314		1,242,433,934	188,577,186	
Overall Tax Rate			.141			.152

ANSWER: Between1974 and 1978, the tax rate decreased in each income category, yet the overall tax rate increased from 14.1 percent to 15.2 percent. Again, the overall rates are weighted averages, with the tax rate for each category weighted by that category's proportion of total income. Because of inflation, in 1978 there were relatively more persons and consequently relatively more taxable dollars assigned to the higher income (i.e., higher tax rate) brackets.

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

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

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

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)

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