Queen's University
Faculty of Arts and Sciences
Department of Economics
Fall 2006

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 10 questions (marks are indicated)

Do all 10 questions

1. **(10 marks)** There are two classes of assets (technology and energy) and two markets for these assets (domestic and international). The following return results for each asset and market are recorded:

	Domestic		International	
	Tech	Energy	Tech	Energy
n	5	10	2	5
$\bar{X}$	8%	22%	10%	4%

   (a) Calculate the sample mean return for all tech stocks.

   (b) Calculate the sample mean return for all international stocks

   (c) Calculate the overall sample mean

Answers

$$\bar{X}_{tech} = \frac{5}{7} \times 8 + \frac{2}{7} \times 10 = 8.6$$
$$\bar{X}_{int} = \frac{2}{7} \times 10 + \frac{5}{7} \times 4 = 5.7$$
$$\bar{X}_{all} = \frac{5}{22} \times 8 + \frac{10}{22} \times 22 + \frac{2}{22} \times 10 + \frac{5}{22} \times 4 = 13.6$$

2. **(5 marks)** Suppose A is a subset of B and $P(A) = .3$ and the $P(B) = .5$
   (a) What is $P(A \cup B)$?
   (b) What is $P(A \cap B)$?
   (c) What is $P(A \mid B)$ and $P(B \mid A)$?

Answer

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A) = P(B) = .5$$
$$P(A \cap B) = P(A) = .3$$
$$P(A \quad | \quad B) = \frac{P(A \cap B)}{P(B)} = \frac{.3}{.5} = .6 \quad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

: 0.6

3. $\cup$ **(10 marks)** Suppose X is a discrete random variable that can take on 5 values with the following cumulative distribution

$X = x_i$	$F(X = x_i)$
1	.3
4	.4
9	.6
16	.6
25	

1. (a) Construct the marginal probabilities $P(X = x_i)$

(b) What is $E[X]$?

(c) What is $E[\sqrt{X}]$?

(d) Does $\sqrt{E[X]} = E[\sqrt{X}]$?

Answer

We obtain the marginal probabilities by subtracting the cumulatives

$X = x_i$	$P(X = x_i)$
1	.3
4	.1
9	.2
16	0
25	.4

$$\begin{aligned} E[X] &= 1 \times .3 + 4 \times .1 + 9 \times .2 + 16 \times 0 + 25 \times .4 = 12.5 \\ E[\sqrt{X}] &= 1 \times .3 + 2 \times .1 + 3 \times .2 + 4 \times 0 + 5 \times .4 = 3.1 \\ E[\sqrt{X}] &= 3.1 \neq \sqrt{E[X]} = \sqrt{12.5} = 3.16 = 3.5 \end{aligned}$$

: 3. 535 5 : 3. 1 : 12. 5

Since expectation is a linear operator, and the square root is a non-linear function

4. **(10 marks)** The following table records the sample of 10 domestic interest rates for various countries

3.4	5.8	7.2	3.5	3.5	6	4.2	5.3	7.3	2.8

1. (a) Calculate the sample mean, variance, median , mode (carry only 1 significant digit)

   (b) Draw a box plot

   (c) What is the coefficient of variation

Answer

$$\begin{aligned} Mean &= 4.9 \\ Variance &= 2.7 \\ Median &= 4.75 \\ Mode &= 3.5 \\ CV &= \frac{\sqrt{2.7}}{4.9} \times 100 = 33.5\% \end{aligned}$$

5. **(15 marks)** General Motors operates 3 plants (Oshawa, Hamilton and Burlington) with the following information regarding total car defects:

	Proportion of Defects	Proportion of Production
Oshawa	6%	40%
Hamilton	4%	25%
Burlington	3%	35%

1. (a) Define mutually exclusive and exhaustive events.
   (b) Draw a Venn diagram of this problem
   (c) Convert information in table to symbol form
   (d) What is the probability of a defect?
   (e) If a defective part is found, what is the probability it was made in Oshawa?
   (f) If a defective part is found, what is the probability it was made in Oshawa or Hamilton?

Mutally exclusive and exhasutive events are ones in which there is no intersection and the union of the events is the sample space

$$\begin{aligned} P(O) &= .4, \quad P(H) = .25, \quad P(B) = .35 \\ P(D &\mid O) = .06 \quad P(D \mid H) = .04 \quad P(D \mid B) = .03 \end{aligned}$$

$$
\begin{aligned}
P(D) &= P(O) \times P(D \mid O) + P(H) \times P(D \mid H) + P(B) \times P(D \mid B) \\
&= .4 \times .06 + .25 \times .04 + .35 \times .03 = ..0445
\end{aligned}
$$

$$
P(O \mid D) = \frac{P(O) \times P(D \mid O)}{P(D)} = \frac{.4 \times .06}{.0445} = .54
$$

$$
P(H \mid D) = \frac{P(H) \times P(H \mid O)}{P(D)} = \frac{.25 \times .04}{.0445} = .22
$$

$$
\begin{aligned}
P(O \cup H \quad \mid \quad D) &= P(O \mid D) + P(H \mid D) \\
&= .54 + .22 = .76
\end{aligned}
$$

6. . **(10 marks)** This problem illustrates the "randomized response technique". We use Bayes' Theorem here to estimate the response rate to sensitive questions. For example, suppose we are interested in the proportion of the population that has done illegal drugs and that we doubt that people will answer directly and truthfully in an open questionnaire. Imagine there are two questions that can be posed:

Question 1: Have you ever done illegal drugs?

Question 2: Is your mother born in July?

The interviewer tells the person to flip a coin and if it lands head they are to answer question 1, and if it is tails they are to answer question 2. The interviewer only records the yes answers and the number of people interviewed. Suppose that the people that are interviewed answer truthfully and that they conduct the coin flipping exercise properly. After all the interviews are done (say 100) there are 6 yes answers. Let Y be the event of yes. Let Q_1 and Q_2 be the event of answering that question 1 and questioned 2 were answered.

1. (a) If mothers are born evenly throughout the year, $P(Y \mid Q_2)$
   (b) What is $P(Q_1)$?

(c) In symbols show what probability we wish to estimate?

(d) What is the unconditional probability of Y?

(e) Answer (c)

Answers

$$P(Y \mid Q_2) = \frac{1}{12}$$
$$P(Q_1) = \frac{1}{2}$$
$$P(Y \mid Q_1) = ?$$

$$\begin{aligned} P(Y) &= .06 = P(Q_1) \times P(Y \mid Q_1) + P(Q_2) \times P(Y \mid Q_2) \\ P(Y \mid Q_1) &= \frac{P(Y) - P(Q_2) \times P(Y \mid Q_2)}{P(Q_1)} = \frac{.06 - .5 \times \frac{1}{12}}{.5} = .037 \end{aligned}$$

7. **(10 marks)** A professor who has a class average of 70 and a standard deviation of 20 with a class size of 100. Show what transformation yields the following results

   (a) What is a linear transformation?

   (b) An average of 80 with no change to the sample variance?

   (c) A sample standard deviation of 10 and no change to the average grade?

   (d) A sample average of 75 and a standard deviation of 10?

   (e) Will these transformations in general lead to a bell-shaped distribution of grades?

Answer

A linear transformation is a transformation that can be written as

$$Y_i = a + bX_i$$

$$\begin{aligned} a &= 10 \\ b &= \sqrt{.5} = .71 \quad a = 70 - \sqrt{.5} \times 70 = 20.5 \\ b &= \sqrt{.5} \quad a = 75 - \sqrt{.5} \times 70 = 25.5 \end{aligned}$$

Unless the $X_i's$ are normally distributed the $Y_i's$ will not be normallly distributed (bell shaped)

8. **(5 marks)** Certain phrases in the English language contain implicit conditional statements. In each of the following explain what is meant using probability symbols?
   (a) "Lightening never strikes in the same place twice"
   (b) "No two snowflakes are alike"
   (c) "A quitter never wins and a winner never quits"
   (d) "When it rains it pours"

Answers

Let L_1 be event that lightening strikes the location the first time and L_2

$$P(L_2 \mid L_1) = 0$$

Let S_i be the shape of snowflake i

$$P(S_i = S_j) = 0 \quad \text{for all } i \neq j$$

Let Q be the event a quitter and W be the event of a winner

$$P(W \mid Q) = P(Q \mid W) = 0$$

Let R be the event rain and P be the event pours. $R = P$ so that

$$P(P \mid R) = P(R)$$

The statement is nonsensical since we usually think of rain aas less than a (down)pour but the idea is to suggest it never just rains on a person.

9. **(15 marks)** An advertising firm is interested in determining whether an add campaign is effective. In one region the ad firms sends out flyers advertising their product, in another they do not advertise. They record the number of sales in a square kilometer in each region. Let X be the sales in the region with the advertising and Y be the sales in the region without advertising . Assume there were 2 measurements in region X and 3 in region Y and the probabilities measures the proportion sampled in the two regions for the two measurements.

x_i	$P(X = x_i)$	y_i	$P(Y = y_i)$
200	.3	150	.8
300	.7	200	.1
		400	.1

1. (a) Discuss the set-up for this experiment and state any potential problems
   (b) Construct the joint probability table making any necessary assumptions you require
   (c) What is $E[X]$, $E[Y]$, $E[X - Y]$?
   (d) What is $V[X]$, $V[Y]$, $V[X - Y]$?
   (e) Calculate coefficient of variation for X, Y and $X - Y$.
   (f) From the information calculated, what do you think of the advertising campaign

Answer

The experiment has several problems which include the independence assumption (people can straddle neighbourhoods and recieve information from flyers. Tthere are many other factors controlling purchase patterns besides advertising we are not controlling for any so neighbourhoods could differ in say their income, tastes and so on

If we can assume independence (if we have not got independence we have no way to construct the joint distribution)

x_i/y_i	150	200	400
200	.24	.03	.03
300	.56	.07	.07

$$
\begin{aligned}
E[X] &= 200 \times .3 + 300 \times .7 = 270 \\
E[Y] &= 150 \times .8 + 200 \times .1 + 400 \times .1 = 180 \\
E[X - Y] &= E[X] - E[Y] = 270 - 180 = 90 \\
V[X] &= (200 - 270)^2 \times .3 + (300 - 270)^2 \times .7 = 2100 \\
V[Y] &= (150 - 180)^2 \times .8 + (200 - 180)^2 \times .1 + (400 - 180)^2 \times .1 = 5600 \\
V[X - Y] &= V[X] + V[Y] = 2100 + 5600 = 7700 \\
CV[X] &= \frac{SD[X]}{E[X]} = \frac{\sqrt{2100}}{270} \times 100 = 17\% \quad CV[Y] = \frac{SD[Y]}{E[Y]} = \frac{\sqrt{5600}}{180} \times 100 = 42\% \\
CV[X - y] &= \frac{SD[X - Y]}{E[X - Y]} = \frac{\sqrt{7700}}{90} \times 100 = 98\%
\end{aligned}
$$

If if we are willing to make sufficent assumptions that this is a reliable experiment, we see that there is is almost as much variation ($SD[X - Y)$ around the difference as there is in the average number of extra sales ($E[X - Y]$)

10. **(10 marks)** The game rock, paper and scissors is played as follows. There are 3 symbols for rock paper and scissors. Two players compete by showing one of the symbols at the same time and the following decides the outcome: rock beats scissors, paper beats rock and scissors beats paper. Suppose there are two player's 1 and 2.

    (a) What is any player's best strategy in playing this game if both players play optimally?

    (b) If Player 1's strategy is to select her choices such that $P(R) = P(Pa)$ and $P(S) = 2 \times P(P_a)$, then what should Player 2 do?

    (c) Comment about the problem with selecting the optimal strategy

    (d) Can you fingure some mechanism to ensure that both parties play optimally

Answer

Without any prior information regarding someone's strategy the best thing to do is randomize among the possible outcomes

$$P(R) = P(Pa) = P(S) = \frac{1}{3}$$

If one player adopts a particular strategy, then the other player may take advantage of that strategy

$$\begin{aligned} P(R^1) &= P(Pa^1) = .25 \text{ and } P(S^1) = 2 \times P(P_a) = .5 \\ &\Longrightarrow P(P_a^2) = P(S^2) = .25 \text{ and } P(R^2) = 2 \times P(S) = .5 \end{aligned}$$

since player 1 plays S twice as often as the other two events, player 2 should counteract by playing R twice as frequent as the others.

Of course, calculating these probabilities and playing them is difficult (the human brain is unable to randomize and will develop patterns. This is why contests are really determining who can figure out the other's pattern and exploit it!). A mechansim for playing optimally for both players is to have say a die and roll it repeatedly (giving the random outcome) and choose

$$\begin{aligned} R &= 1,2 \\ P_a &= 3,4 \\ S &= 4,5 \end{aligned}$$