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

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)** With the recent gas price surges people are concerned with their fuel mileage of their car. Unfortunately hybrid cars (H) which get on average 15 kilometers per liter and a standard deviation of 1 are very expensive relative to traditional cars which get 7 kilometers per liter and a standard deviation of 2. Suppose the following shows the relationship between gas mileage (k =kilometers per liter) and price of a car (C)

$$C = \$40,000 + 1000 * k$$

   (a) What is the average price and standard deviation of the two classes of cars?

$$\begin{aligned} \bar{C} &= 40000 + 1000 \times E[k] \\ \bar{C}_H &= 40000 + 1000 \times 15 = \$55,000 \\ \bar{C}_T &= 40000 + 1000 \times 7 = \$47,000 \end{aligned}$$

$$
\begin{aligned}
s_C^2 &= 1000^2 \times s_k^2 \\
s_H^2 &= 1000^2 \times 1^2 = 1.0 \times 10^6 \Rightarrow s_H = \sqrt{1.0 \times 10^6} = \$1000 \\
s_T^2 &= 1000^2 \times 2^2 = 4.0 \times 10^6 \Rightarrow s_T = \sqrt{4.0 \times 10^6} = \$2000
\end{aligned}
$$

(b) What are the coefficients of variation

$$
\begin{aligned}
CV_C &= \frac{s_c}{\bar{C}} \times 100\% \text{ for sample} \\
CV_H &= \frac{1000}{55000} \times 100 = 1.82\% \\
CV_T &= \frac{2000}{47000} \times 100 = 4.25\%
\end{aligned}
$$

Relative to its mean there is much more variation for the traditional cars than hybrids.

2. **(10 marks)** The worth of advertising is always questioned despite the fact that firms spend a large fraction of their revenues to promote their products. Assume for this question that advertising has **no** impact on loyalty to firm. Suppose you are told that two firms sell identical flowers (Eric the Flower Man and Heather's Flowers) with each having an even split of loyalty (L a yes/no variable)) of the available customers. Assume that the decision to advertise (A) is determined by the flip of a fair coin.

(a) What is the joint probability of loyalty and advertising decision for our two flower shops? Independence

Advertising Decision/Loyalty of firm shop	L_1	L_2
Y	.25	.25
N	.25	.25

(b) What would need to be true for advertising to have an impact on sales?

$$P(L_1|Y) > P(L_2|N)$$

This means that advertising can influence the loyalty of the customer in the positive direction. In the event that you do not advertise, you lose loyalty

(c) Discuss the key element in this question

$$P(Y, L)) = P(Y) \times P(L) \text{ independence}$$

3. **(10 marks)** In the recent stock market, the probability that the market goes up in any day is 0.1 and the probability of it going down is 0.9 with no chance of it staying the same. If we assume that stock market outcomes are independent from one day to the next. We want to calculate the probability of various events in the next 10 days:

(a) What is the probability of 5 days of declining stocks (I was not assuming in a row, give complete credi if they choose in a row. Of course this is a more complicated problem because there is a limited number of ways it could be done) ?

$$P(X = 5 \mid n = 10, p = .1) = \binom{10}{5} \times .1^5 \times .9^5 = 1.488 \times 10^{-3}$$

(b) What is the probability of at least 5 days of declining stocks ? Answer: Look at the probability of 4 or less declining stocks and subtract it from 1)

$$1 - P(X = 4 \mid n = 10, p = .1) = 1 - \sum_{k=0}^{4} \binom{10}{k} \times .1^k \times .9^{10-k} \approx 0$$

(c) Using the additive law of probability calculate the probability of having either 5 days of declining stocks or 5 days of increasing stocks or both. Does the additive law of probability formula work? Answer: There are two ways to have the outcome of 5 days of declining stock in a row and

$$\begin{aligned}
P(A &= \text{5 days of declining stocks} \cup B \\
&= \text{5 days of increasing stocks}) \\
&= P(A) + P(B) - P(A \cap B) \\
P(A \cap B) &= P(A) \text{ if there are 5 days of declining stock} \\
&\quad \text{there must also be 5 days of increasing stocks by default} \\
P(A) + P(B) - P(A \cap B) &= P(A) = 1.488 \times 10^{-3}
\end{aligned}$$

(d) What is the expected number of days the stock goes up and its variance?

$$\begin{aligned}
E[X] &= np = 10 \times .1 = 1 \\
V[X] &= np(1-p) = 10 \times .1 \times .9 = .9
\end{aligned}$$

4. **(10 marks)** Cumulative and joint probability distributions provide the same information and this problem illustrates this. Suppose you are given the following joint cumulative probability distribution $F(x, y)$ with $X = \{3, 4\}$ and $Y = \{1, 2\}$

$F(x,y)$	$F(x,1)$	$F(x,2)$
$F(3,y)$	.3	.4
$F(4,y)$	.3	1

(a) Interpret the table above and give the cumulative probability of $F(x)$ and $F(y)$

$$F(x, y) = P(X \leq x, Y \leq y)$$

(X)	$F(X)$	Y	$F(y)$
3	.4	1	.3
4	1	2	1

(b) What is joint probability distribution for X and Y.

$P(x,y)$	1	2
3	.3	.1
4	0	.6

(c) What is $P(X \mid Y = 2)$.. Are these variables independent?

$$\begin{aligned} P(X &= 3|Y=2) = \frac{P(X=3 \cap Y=2)}{P(Y=2)} = \frac{.1}{.7} = .14 \neq P(X=3) = .4 \quad \text{no} \\ P(X &= 4|Y=2) = \frac{P(X=4 \cap Y=2)}{P(Y=2)} = \frac{.6}{.7} = .86 \neq P(X=4) = .6 \quad \text{no} \end{aligned}$$

5. **(20 marks)** Averaging observations will smooth out fluctuations and is often used to display stock market data. Smoothing or averaging is defined as taking moving averages by adding one day's data to the calculation and dropping the earliest observation. For example suppose the two-day average index of the October 9 and October 10 is 1005 ($\bar{X}_{Oct10}$) with October 10th value being 1005. Below is the data for the next 2 days

Days	Oct. 10	Oct. 11	Oct. 12
Index	1005	1006	1008
$\bar{X}$	1005	1005.5	1007

(a) Calculate $\bar{X}_{Oct11}$ and $\bar{X}_{Oct12}$

$$\begin{aligned} \bar{X}_{Oct11} &= \frac{1}{2}(1005 + 1006) = 1005.5 \\ \bar{X}_{Oct12} &= \frac{1}{2}(1006 + 1008) = 1007 \end{aligned}$$

(b) Can you determine what the value of the index must have been for October 9th

$$\bar{X}_{Oct12} = 1005 = \frac{1}{2}(\text{Oct } 9 + 1005) = 1005$$

(c) Calculate the sample variance of the 3 days of the daily index and the two-day moving average. What sample variance is smaller and can you see why we say that averaging smooths data

$$\begin{aligned} \bar{X}_{index} &= \frac{1}{3}(1005 + 1006 + 1008) = 1006.3 \\ \bar{X}_{\bar{X}} &= \frac{1}{3}(1005 + 1005.5 + 1007) = 1005.8 \end{aligned}$$

$$\begin{aligned} s^2 &= 2.33 \\ s^2_{\bar{X}} &= 1.08 \end{aligned}$$

(d) In practice financial markets typically average over 30 days since that provides more smoothing and gives trend movements. It would tedious to redo the calculation of the mean by adding up all the previous 30 days of observations and dividing by 30. Fortunately there is a simple formula which can be worked out based on the same principal as a weighted average . Can you find b and the c in the following relationship between the mean of the previous day ($\bar{X}_t$) and the new price X_{t+1}

$$\bar{X}_{t+1} = \bar{X}_t + bX_t + cX_{t-30}$$
$$\bar{X}_{t+1} = \bar{X}_t + \frac{1}{30}X_t - \frac{1}{30}X_{t-30}$$

If the general calculation eludes you try using the specific numbers in the above formulation and work out:

$$\bar{X}_{Oct11} = \bar{X}_{Oct10} + bX_{Oct11} + cX_{Oct9}$$

6. (**10 marks**) People are often surprised by the fact that they share the same birthday as someone else and think this is a fantastic coincidence. In this example we will illustrate the probability of knowing someone in common. Suppose there is a one in ten chance that you share a friend in common (assume there are a lot of people out there). Determine how many people are needed to have a 88% probability at least two of them will share the a friend.

$$\begin{aligned} 1 - P(X &= 0 \mid n, p = .1) = .88 \\ \text{From tables we find:} \quad & n = 20 \text{ since } P(X = 0 \mid n, p = .1) = .12 \end{aligned}$$

7. (**10 marks**) Consider how knowing the probability outcome of some random event might assist us in determining something we want to know. Suppose Professor Gregory wants to find out how many students really hate Economics 250. He knows that he cannot just simply ask students that question in public but instead asks each student to flip a coin and answer one of the following questions:

$$\begin{aligned} Tails &: \text{ Do you hate Econ 250} \\ Head &: \text{ Did the coin flip land on heads?} \end{aligned}$$

Students answer truthfully according to their flip of the coin (A yes or no to either question). An independent observer records the number of Yes answers. Professor Gregory wants to calculate $P(E)$, where E is event of hating Econ 250 and records only the yes answers. Write out all of the probability information you have in this problem.

(a) What is the probability of a Head and what is the probability of a yes (Y) answer given the coin flip is a head?

$$P(H) = .5 \Rightarrow P(Y \mid H) = 1$$

(b) What would we conclude from a larger percentage of Yes answers? More than 50% yes would imply that people are hating Econ 250

(c) Write out in symbols an expression for the $P(Y)$ in terms of the intersection probabilities of the other events? Express this as conditional probabilities with the relevant conditons! What do we know already in this expression?

$$\begin{aligned} P(Y) &= P(Y \cap H) + P(Y \cap T) \\ &= P(Y \cap H) + P(Y|T) \times P(T) \\ P(H \cap Y) &= P(H) \times P(Y \mid H) = .5 \\ P(Y|T) &= \frac{P(Y) - 0.5}{P(T)} \end{aligned}$$

(d) Notice that this expression has only 1 unknown which is what we want to know. Suppose of the 70 students in Econ250, 45 answered yes. What is the best guess of the probability of students hating Econ250.

$$\begin{aligned} P(Y) &= \frac{45}{70} = .64 \\ P(Y|T) &= \frac{0.64 - 0.5}{0.5} \\ &= \frac{0.14}{0.5} \\ &= 0.28 \end{aligned}$$

8. **(20 marks)** One problem in the recent financial decline has been to sort out what assets are good and which are bad (likely to default) in mortgage backed securities. Mortgage back securities constitute about 25% of the market and these securities default 75% of the time. The remaining asset classes default only 5% of the time. Assume that mortgage and the other securities cannot default at the same time. Mortgage securities that do not default provide a 20% and a -95% rate of return when they default (i.e. you lose 95% of your invested money) . The other class of securities have a 5% rate of return when they do not default and 0 when they do.

(a) Wrtite out everything that you know in this problem.

(b) Suppose your rich aunt invests 1 million dollars in each of the mortgage back security and the other class of assets, what is your expected income?

Mortgage Gross Returns	P(X=x)	Other	P(Y=y)
ND=1.2	.25	ND=1.05	.95
D=.05	.75	D=1.0	.05

$$\begin{aligned} Y &= M + O \\ E[Y] &= E[M] + E[O] = 1.2 \times .25 + .05 \times .75 + 1.05 \times .95 + 1 \times .05 = 1.39 \end{aligned}$$

Your aunt is going to lose on average

(c) Suppose you hear that at least one of her securities has defaulted, what is the probability that it was the mortgage backed security? What is the chance it is the other class of securities?

$$\begin{aligned} P(MD \quad | \quad D) &= \frac{P(MD) \times P(D \mid MD)}{P(D)} \\ &= \frac{.25 \times .75}{P(MD) \times P(D \mid MD) + P(O) \times P(D \mid OD)} \\ &= \frac{.25 \times .75}{.25 \times .75 + .75 \times .05} = .83 \end{aligned}$$

$$\begin{aligned} P(OD \quad | \quad D) &= \frac{P(OD) \times P(D \mid OD)}{P(D)} \\ &= \frac{.75 \times .05}{0.225} = .17 \end{aligned}$$

(d) (**Bonus Question: Worth 5)** What is the expected value of her portfolio given this news that one of her investments has defaulted using the conditional probabilities above

Mortgage Gross Returns	P(X=x)	Other	P(Y=y)
ND=1.2	.17	ND=1.05	.83
D=.05	.83	D=1.0	.17

$$\begin{aligned} E[Y \quad | \quad D] &= E[M \mid D] + E[O \mid D] \\ &= 1.2 \times .17 + .05 \times .83 + 1.05 \times .83 + 1 \times .17 = 1.29 \end{aligned}$$

Your aunt is going to lose on average $2 - 1.29 = 0.71M$

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

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

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

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

Alternatively

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

Grouped Data (with k classes)

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

Probability Theory

$$\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{(additive law)} \\
P(A \cap B) &= P(B)P(A \mid B) \quad \text{(multiplicative law)}
\end{aligned}$$

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

$$\begin{aligned}
P(A) &= \sum_{i}^{n} P(A \cap E_i) = \sum_{i}^{n} P(E_i)P(A \mid E_i) \\
P(E_i \mid A) &= \frac{P(E_i)P(A \mid E_i)}{P(A)} \quad \text{(Bayes' Theorem)}
\end{aligned}$$

Counting Formulae

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

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

Random Variables

Let X be a discrete random variable, then:

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

The **covariance** of X and Y is

$$\begin{aligned}
Cov[X,Y] &= \sigma_{xy} = E\left[(X-\mu_x)(Y-\mu_y)\right] \\
&= E[XY] - E[X] \times E[Y]
\end{aligned}$$

The **correlation coefficient** of X and Y

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

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

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

If X and Y are correlated then

$$V[a+bX+cY] = b^2\sigma_x^2 + c^2\sigma_y^2 + 2bc\sigma_{xy}$$

Coefficient of Variation (CV)

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

Univariate Probability Distributions

Binomial Distribution: For $x = 0, 1, 2, ..., n$ and :

$$Pr[X = x] = \binom{n}{x} p^{x} (1 - p)^{n-x}$$

$$E[X] = np$$

$$V[X] = np(1 - p)$$