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

Economics 250 A: Introduction to Statistics

Midterm Exam II:
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 7 questions (marks are indicated)

Do all 10 questions

1. **(10 marks)** Find the probability of each of the following and draw a picture (Z is standard normal)

   (a) $P(Z > 1) = 1 - F_z = 1 - .8413 = 0.158\,7$

   (b) $P(-.5 < Z < .5) = 2 \times (F_z(.5) - .5) = 2 \times (.6915 - .5) = 0.383$

   (c) $P(2 \times Z > 1)$ this is a linear transformation $X = 2 \times Z$

   $$P(X > 1) = P\left(\frac{X-\mu}{2} \geq \frac{1-0}{2}\right) = P(Z \geq 0.5) = 1 - .6915 = .3085$$

   (d) $P(2 \times Z - 1 < -1)$ this is a linear transformation $X = 2 \times Z - 1$

   $$P(X \leq -1) = P\left(\frac{X-\mu}{2} \leq \frac{-1-(-1)}{2}\right) = P(Z \leq 0) = .5$$

   (e) $P(-d < Z < d) = .85$ find d

$$P(-d < Z < d) = .85 \Longrightarrow d \approx 1.44$$

2. **(10 marks)** Two variables are distributed as

$$\begin{aligned} X &\sim N(10,36) \text{ and } Y \sim N(-4,25) \text{ with } \rho_{xy} = -.1 \\ R &= 2X + 4Y \end{aligned}$$

(a) How is R distributed, work out its mean and variance

$$\begin{aligned} R &\sim N(4,496) \\ E[R] &= E[2X] + E[4Y] = 4 \\ V[R] &= 4 \times V[X] + 16 \times V[Y] + 2 \times 2 \times 4COV[X,Y] \\ &= 4 \times 36 + 16 \times 25 + 2 \times 2 \times 4 \times SD[X] \times SD[Y] \times \rho_{xy} \\ &= 4 \times 36 + 16 \times 25 + 2 \times 2 \times 4 \times 6 \times 5 \times (-.1) = 496 \end{aligned}$$

(b) What can we say about X and Y The are negatively correlated so that when one tends to be above its mean the other tends to be below. However the correlation is very low at -.1

(c)

$$P(10 < R < 20)?$$

$$\begin{aligned} &P\left(\frac{10-4}{\sqrt{496}} < \frac{R-\mu_R}{\sigma_R} < \frac{20-4}{\sqrt{496}}\right) \\ P(.269 &< Z < .718) = F_Z(.718) - F_Z(.269) = .7642 - .6064 = .1578 \end{aligned}$$

3. **(20 marks)** In a certain area of chronic unemployment, the probability of randomly finding an unemployed person is .25. Answer the following questions

   (a) Find the probability of exactly 5 to 10 people inclusively being unemployed when a sample of 20 is taken? What assumptions are you making (independence)

$$\begin{aligned} P(5 &\leq X \leq 10 \mid n = 20, \pi = .25) = F(10 \mid n = 20, \pi = .25) - F(4 \mid n = 20, \pi = .25) \\ &= .996 - .415 = .581 \end{aligned}$$

   :

   (b) Use the normal approximations with and without continuity corrections to calculate (a)

$$\begin{aligned} P(5 &\leq X \leq 10 \mid n\pi = 20 \times .25 = 5, \sigma^2 = n\pi(1-\pi) = 20 \times .25 \times .75 = 3.75 \\ &\approx P\left(\frac{5-5}{\sqrt{3.75}} \leq \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \leq \frac{10-5}{\sqrt{3.75}}\right) \\ &= P(0 \leq Z \leq 2.58) = F_Z(2.58) - .5 = .9951 - .5 = .4951 \end{aligned}$$

$$\begin{aligned} P(4.5 &\leq X \leq 10.5 \mid n\pi = 20 \times .25 = 5, \sigma^2 = n\pi(1-\pi) = 20 \times .25 \times .75 = 3.75 \\ &P\left(\frac{4.5-5}{\sqrt{3.75}} \leq \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \leq \frac{10.5-5}{\sqrt{3.75}}\right) \\ P(-.258 &\leq Z \leq 2.84) = F_Z(2.84) - .5 + F_Z(.258) - .5 = .9977 - .5 + .6026 - .5 = .600 \end{aligned}$$

   (c) The uniform distribution is not used to approximate this problem and will not provide a very accurate answer but use it to answer (a). To calculate the end points you will need to solve them from the mean and variance of the uniform (formula is at the back).

$$\begin{aligned} 5 &= \frac{a+b}{2} \\ 3.75 &= \frac{(b-a)^2}{12} \\ a &= 1.6 \text{ and } b = 8.4.\text{approximately} \end{aligned}$$

$$P(5 \leq X \leq 10) \approx P(5 < X < 8.4) = \frac{8.4-5}{8.4-1.6} = .5 \text{ not all that bad when considering the u}$$

   (d) Draw a picture to illustrate this problem

4. **(10 marks)** With the US. elections coming up, people are interested in polling. Suppose the true support for republicans is 40%, what is the probability that in 100

people, the support declared would be greater than 50?

$$
\begin{aligned}
f &= \frac{X}{n} \approx N(\pi, \frac{\pi(1-\pi)}{n} \\
P(f \geq .5) &\approx P\left(\frac{f-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \geq \frac{.5-.4}{\sqrt{\frac{.4\times.6}{100}}}\right) \\
&= P(Z \geq 2.04) = 1 - F_Z(2.04) = 1 - .9793 = .021
\end{aligned}
$$

5. **(15 marks)** Averaging tends to concentrate the distribution explain why. Answer the following questions where we assume each of the $X_{i's}$ are independent and normal with a mean of 3 and a variance of 100). Averaging keeps the mean the same but lowers the variance by a factor of $\frac{1}{n}$ where n is the sample size

   (a) $P(2.5 < X < 3.5) = P\left(\frac{2.5-3}{\sqrt{100}} < \frac{X-\mu}{\sigma} \leq \frac{3.5-3}{\sqrt{100}}\right) = P(-0.05 \leq Z \leq 0.05) = 2 \times (F_z(.05) - .5) = 2 \times (.5199 - .5) = 0.039\,8$

   (b) $P(2.5 < \bar{X}_2 < 3.5) = P\left(\frac{2.5-3}{\sqrt{\frac{100}{2}}} < \frac{X-\mu}{\sigma} \leq \frac{3.5-3}{\sqrt{\frac{100}{2}}}\right) = P(-.071 < Z < .071) = 2 \times (F_z(.07) - .5) = 2 \times (.5279 - .5) = 0.055\,8$

   (c) $P(2.5 < \bar{X}_5 < 3.5) = P\left(\frac{2.5-3}{\sqrt{\frac{100}{5}}} < \frac{X-\mu}{\sigma} \leq \frac{3.5-3}{\sqrt{\frac{100}{5}}}\right) = P(\ -0.111\,8 \leq Z \leq .1118) = 2 \times (F_z(..118) - .5) = 2 \times (.5478 - .5) = 0.095\,6$

   (d) What sample size would make $P(2.5 < \bar{X}_n < 3.5) = .99 = P\left(\frac{2.5-3}{\sqrt{\frac{100}{n}}} < \frac{X-\mu}{\sigma} \leq \frac{3.5-3}{\sqrt{\frac{100}{n}}}\right) \Longrightarrow$ $\frac{3.5-3}{\sqrt{\frac{100}{n}}} = 2.58$, Solution is: $\{n = 2662.\,6\}$

6. (**10 points)** Two schools are being examined for the quality of their teaching. School A has 50% of the students and school B has the remainder. Students have a 80% chance of passing in School A and a 75% passing in school B. Ten students are taken at random from one school and 9 pass, find the probability that they were taken from School A.

$$
\begin{aligned}
P(A) &= .5 = P(B) \\
P(Pa &\mid A) = .8 \qquad P(Pa \mid B) = .7 \\
P(X &= 9 \mid A) = \binom{10}{9}.8^9.2^1 = .27 \\
P(X &= 9 \mid B) = \binom{10}{9}.75^9.25^1 = .19 \\
P(A &\mid X = 9) = \frac{P(A) \times P(X = 9 \mid A)}{P(A) \times P(X = 9 \mid A) + P(B) \times P(X = 9 \mid B)} \\
&= \frac{.5 \times .27}{.5 \times .27 + .5 \times .19} \\
&= .587
\end{aligned}
$$

7. (**25 marks**) Two independent samples are drawn with the following properties

$$\begin{aligned} X_{1i} &\sim N(5,25) \quad i=1,\ldots,25 \\ X_{2j} &\sim N(5,25) \quad j=1,\ldots,50 \end{aligned}$$

(a) Consider

$$Y = \left(\frac{\bar{X}_1 + \bar{X}_2}{2}\right)$$

$$\begin{aligned} Y &\sim N(5, \\ E[Y] &= E\left(\frac{\bar{X}_1 + \bar{X}_2}{2}\right) = 5 \\ V[Y] &= \frac{1}{4}V[\bar{X}_1] + \frac{1}{4}V[\bar{X}_2] \\ &= .375 \end{aligned}$$

What is the sampling distribution Y?

(b) Calculate the probability of

$$P(4.9 < Y < 5.3)$$

$$\begin{aligned} & P\left(\frac{4.9-5}{\sqrt{.375}} < \frac{Y-\mu_y}{\sigma_y} \leq \frac{5.3-5}{\sqrt{.375}}\right) \\ = \;& P(-.16 < Z < .49) = F_z(..49) - .5 + F_z(.16) - .5 \\ = \;& .6879 - .5 + .5636 - .5 = .2515 \end{aligned}$$

(c) Consider

$$R = \frac{1}{75}\left(\sum_{i=1}^{25} X_{1i} + \sum_{j=1}^{50} X_{2j}\right)$$

What is sampling distribution of R?

$$R = \bar{X}_{75} \sim N(5, .33)$$

(d) Calculate the probability of

$$\begin{aligned} P(4.9 &< R < 5.3 \\ &= P\left(\frac{4.9-5}{\sqrt{.333}} < \frac{R-\mu_R}{\sigma_R} \leq \frac{5.3-5}{\sqrt{.333}}\right) \\ &= P(-.17 < Z < .52) = F_z(..52) - .5 + F_z(.17) - .5 \\ .6985 - .5 + .5675 - .5 &= .266 \end{aligned}$$

(e) Can you draw any conclusions on these results? The sample mean with 75 observations is more tightly distributed around the population mean than just the average of the two means.

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

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

Sample Variance

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

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 :

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

$$E[X] = n\pi$$

$$V[X] = n\pi(1 - \pi)$$

Uniform Distribution: For $a < x < b$:

$$f(x) = \frac{1}{b - a}$$

$$E[X] = \frac{a + b}{2}$$

$$V[X] = \frac{(b - a)^2}{12}$$

Normal Distribution:For $-\infty < x < \infty$:

$$f(x) = (2\pi\sigma^2)^{-1/2} exp[\frac{-(x - \mu)^2}{2\sigma^2}]$$

$$E[X] = \mu$$

$$V[X] = \sigma^2$$

$$\begin{aligned} X &\sim N(\mu_X, \sigma_X^2) \\ Z &= \frac{X - \mu_X}{\sigma_X} \sim N(0, 1) \end{aligned}$$