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

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 and the binomial cumulative tables. 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 5 questions (marks are indicated)

Do all 8 questions

1. **(30 marks, 5 each)** Suppose that sales (S) at a store are independently normally distributed with mean of 100 and a variance of 9.

   (a) What is the $P(97 < S < 102)$

$$
\begin{aligned}
P(97 &< S < 102) = P\left(\frac{97-100}{3} < \frac{S-\mu}{\sigma} < \frac{102-100}{3}\right) \\
&= P(-1 < Z < .67) = (F(.67) - .5) + (F(1) - .5) \\
&= (.7486 - .5) + (.8413 - .5) = .59
\end{aligned}
$$

   (b) What is the $P(97 < \bar{S}_{10} < 102)$ where $\bar{S}_{10}$ is the sample mean with 10 observations

$$
\begin{aligned}
P(97 &< \bar{S}_{10} < 102) = P\left(\frac{97-100}{\frac{3}{\sqrt{10}}} < \frac{\bar{S}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{102-100}{\frac{3}{\sqrt{10}}}\right) \\
&= P(-3.16 < Z < 2.1) = .5 + (.9818 - .5) = .9818
\end{aligned}
$$

(c) Suppose that two independent samples from this population are drawn S^1 and S^2 with samples sizes 10 and 20 respectively Calculate the relative efficiency of their sample mean estimation of the population mean.

$$RE = \frac{V[\bar{S}^1_{10}]}{V[\bar{S}^2_{20}]} = \frac{\frac{9}{10}}{\frac{9}{20}} = 2.0$$

(d) What is the minimum variance unbiased estimator of the population mean μ that you can construct if both sets of data are available to you

$$\bar{S}_{30} = \sum_{i=1}^{30} \frac{1}{30} S_i \text{ use all the observations}$$

(e) Suppose a researcher unaware of the mean but knows the variance obtains a sample mean for the first sample of 98. Construct a 99% confidence interval for the population mean and interpret

$$\begin{aligned} & \bar{S}_{10} \pm Z_{\frac{.01}{2}} \times \frac{\sigma}{\sqrt{n}} \\ & 98 \pm 2.576 \times \frac{3}{\sqrt{10}} \\ = \; & (95.6, \; 100.4) \\ & \text{If we do a large number of confidence intervals} \\ & \text{we expect 99\% of them to bracket the true value of } \mu \end{aligned}$$

(f) Suppose the researcher did not know the variance but obtained an estimate of 9 (which coincidently equals the actual variance). Construct a 99% confidence interval for the population mean and interpret

$$\begin{aligned} & \bar{S}_{10} \pm t_{n-1, \frac{.01}{2}} \times \frac{s}{\sqrt{n}} \\ & 98 \pm 3.250 \times \frac{3}{\sqrt{10}} \\ & (94.9, 101.08) \\ & \text{If we do a large number of confidence intervals} \\ & \text{we expect 99\% of them to bracket the true value o} \end{aligned}$$

2. (**20 , 5 each**) Newly built planes are given a thorough inspection and 10% of them are set up incorrectly.

(a) What is the probability that between 1 and 3 planes inclusively out of 20 planes are set up incorrectly

$$\begin{aligned} P(1 & \leq X \leq 3 \mid n = 20, p = .1) \\ & = F(3 \mid n = 20, p = .1) - F(0 \mid n = 20, p = .1) \\ & = .867 - .122 = .75 \end{aligned}$$

(b) What is the normal approximation that between 1 and 3 planes inclusively out of 20 planes are set up incorrectly? Do this with and with out continuity corrections.

$$\begin{aligned}
\text{Normal Approximation } X &\sim N(np, np(1-p)) \\
X &\sim N((2, 1.8) \text{ asymtotically} \\
P(1 &\leq X \leq 3) \approx P\left(\frac{1-2}{\sqrt{1.8}} < \frac{X-np}{\sqrt{, np(1-p)}} < \frac{3-2}{\sqrt{1.8}}\right) \\
&= P(-.75 < Z < .75) = 2 \times (F(.75) - .5) \\
&= 2 \times (.7734 - .5) = .55
\end{aligned}$$

$$\begin{aligned}
&\text{Continuity Corrections} \\
P(.5 &\leq X \leq 3.5) \approx P\left(\frac{.5-2}{\sqrt{1.8}} < \frac{X-np}{\sqrt{, np(1-p)}} < \frac{3.5-2}{\sqrt{1.8}}\right) \\
P(-1.12 &< Z < 1.12) = 2 \times (F(1.12) - .5) \\
&= 2 \times (.8686 - .5) = .74 \text{ much closer}
\end{aligned}$$

(c) Suppose a researcher unaware of the true probability of the defection rate obtains a sample of 20 planes and finds 2 are not set-up correctly. Construct the 95% confidence interval for the population proportion and interpret.

$$\hat{p} \pm Z_{\frac{.05}{2}} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$.1 - 1.96 \times \sqrt{\frac{.1 \times .2}{20}}$$

$$(-0.0315, 0.2315)$$

:

(d) Suppose now that a second researcher who also does not know the true defection rate and obtains an independent sample of 20 with 0 set up incorrectly. How could we incorporate the two samples information into a single estimate that is the most accurate. Explain and give as much detail as possible

$$\begin{aligned}
\hat{p} &= \frac{X_1 + X_2}{n_1 + n_2} \\
&\text{add the number of successes in two samples and} \\
&\text{divide by the number of observations in the combined sample}
\end{aligned}$$

3. **(15 marks, 5 each)** Taxi cabs in Toronto are equally likely to arrive at a house between 10 and 20 minutes after one has been ordered.

(a) What is the probability that a person will have to between 12 and 15 minutes?

$$\begin{aligned} T &\sim U(10,20) \\ P(12 &\leq T \leq 15) = \frac{15-12}{20-10} = .3 \end{aligned}$$

(b) Suppose 20 cabs are ordered what is the expected average waiting time and variance assuming that taxi cab arrivals are independent

$$\begin{aligned} E[T] &= 15, V[T] = \frac{(20-10)^2}{12} = 8.3 \\ \text{Since } T_i \text{ is } iid\ T &\sim iid(15, 8.3) \\ E[\bar{T}] &= 15 \quad V[\bar{T}] = \frac{8.3}{20} = .415 \end{aligned}$$

: 0.415

(c) What is the probability that the average wait time is between 12 and 15 minutes. What assumptions are you making to do this calculation?

$$\begin{aligned} \text{Central limit Theorem:} \quad & iid \\ \bar{T} &\sim N(15, .42) \text{ asymototically} \\ P(12 &< \bar{T} < 15) = (\frac{12-15}{\sqrt{.42}} < Z < \frac{15-15}{\sqrt{.42}}) \\ &= (-4.6 < Z < 0) = .5 \end{aligned}$$

4. (**15 marks, 5 each**) Everyone has heard of the expression warning against putting all of ones eggs in a single basket. This problem illustrates this. Suppose there are two assets X_1 and X_2 with the follwoing returns for a \$1 invested ..

(a) Suppose $X_1 \sim N(1,4)$ and $X_2 \sim N(1,4)$ and independent. what is

$$\begin{aligned} P(0 &< X_1 < 1) = P\left(\frac{0-1}{2} < \frac{X-\mu}{\sigma} < \frac{1-1}{2}\right) \\ &= P(-.5 < Z < 0) = F(.5) - .5 \\ &= (.6915 - .5) = ..1915 \\ P(0 &< \frac{1}{2}X_1 + \frac{1}{2}X_2 < 1) \Rightarrow \left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) \sim N(1,2) \\ P(0 &< \left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) < 1) = P\left(\frac{0-1}{\sqrt{2}} < Z < \frac{1-1}{\sqrt{2}}\right) \\ &= P(-.7071 < Z < 0) = F(.71) - .5 \\ &= .7611 - .5 = .2611 \text{ note probability is higher, independence increases} \\ &\quad \text{the chances of postive values} \end{aligned}$$

(b) Suppose $X_1 \sim N(1,4)$, $X_2 \sim N(1,4)$ and $\rho = .5$ what is the $P(0 < \frac{1}{2}X_1 + \frac{1}{2}X_2 < 1)$?

$$\begin{aligned}
V\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) &= \frac{1}{4} \times (V[X_1 + 2C[X_1, X_2] + V[X_2]) \\
C[X_1, X_2] &= \rho\sigma_1\sigma_2 = .5 \times 2 \times 2 = 2 \\
V\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) &= \frac{1}{4} \times (4 + 2 + 4) = 2.5 \text{ larger variance than independence} \\
P(0 &< \frac{1}{2}X_1 + \frac{1}{2}X_2 < 1) = P\left(\frac{0-1}{\sqrt{2.5}} < \frac{X-\mu}{\sigma} < \frac{1-1}{\sqrt{2.5}}\right) \\
&= P(-.632 < Z < 0) = F(.63) - .5 = .7357 - .5 = .2357
\end{aligned}$$

less chance of positeve returns than 2 independent assets

positive covariance raises risk!

(c) Suppose $X_1 \sim N(1,4)$ and $X_1 = -X_2$.

i. What is $E[\frac{1}{2}X_1 + \frac{1}{2}X_2]$?

$E[\frac{1}{2}X_1 + \frac{1}{2}X_2] = 1 - 1 = 0$ Perfect negative correlation yields asset with no gain

ii. What is $V[\frac{1}{2}X_1 + \frac{1}{2}X_2]$

$$\begin{aligned}
V[\frac{1}{2}X_1 + \frac{1}{2}X_2] &= \frac{1}{4} \times (V[X_1 + 2C[X_1, X_2] + V[X_2]) \\
&= \frac{1}{4} \times (4 - 2 \times 4 + 4]) = 0 \\
\text{Since } C[X_1, X_2] &= \rho\sigma_1\sigma_2 = -1 \times 2 \times 2 = -4
\end{aligned}$$

iii. What is $P(0 < \frac{1}{2}X_1 + \frac{1}{2}X_2 < 1)$. Can you explain what is going on and why this might be useful to an investor

$P(0 < \frac{1}{2}X_1 + \frac{1}{2}X_2 < 1) = 0$

This would act a s a perfect heging strategy

Suppose you had to buy asset A (say exchange contract becasue you are expo

Youmight want to buy another asset in forward market that guards

against movements on the exchange contract

If one that is perfectly negatively correlated was available

you would be perfectly safe! (usually there are none of these)

5. **(20 marks, 5 each)** Linear transformations are extremely common in economic data and handy in a variety of applications. Suppose that a teacher has a test score with a mean of 50 and variance of 25 for 40 students. Assume the scores comes from a normal population and is independent.

(a) What is the score that at least 80% of the students have?

$$
\begin{aligned}
P(X_i &< ?) = .8 \\
P(\left(\frac{X_i - \mu}{\sqrt{\sigma}} \leq \frac{? - 50}{\sqrt{25}}\right) &= .8 \\
P(Z &< \frac{? - 50}{\sqrt{25}}) = .8 \\
\frac{? - 50}{\sqrt{25}} &= .85 \text{ Since } F_X(.85) = .8 \\
? &= (.85 \times \sqrt{25}) + 50 = 54.3
\end{aligned}
$$

(b) If 5 is added to all the scores, what is the score now that at least 80% of the students have

$$
\begin{aligned}
Y_i &= (5 + X_i) \sim N(55, 25) \\
& P(\left(\frac{Y_i - \mu_y}{\sqrt{\sigma}} \leq \frac{? - 55}{\sqrt{25}}\right) = .8 \\
\frac{? - 55}{\sqrt{25}} &= .85 \Rightarrow ? = 59.3
\end{aligned}
$$

(c) If instead of adding 5 the teacher takes the grade and multiples by 1.1. What is the score now that at least 80% of the students have?

$$
\begin{aligned}
Y_i &= (1.1 \times X_i) \sim N(55, 1.1^2 \times 25 = 30.3) \\
P(\left(\frac{Y_i - \mu_y}{\sqrt{\sigma}} \leq \frac{? - 55}{\sqrt{30.3}}\right) &= .8 \\
\frac{? - 55}{\sqrt{30.3}} &= .85 \Rightarrow ? = 59.6
\end{aligned}
$$

(d) Discuss the differences in the two methods of adjusting the grades. *The two methods yield the same average for the class as a whole but different grades for the individual. Any student receiving a grade less than 50% perfers the first method to the second. the opposite is true for the student who started with a grade over 50. the student with 50% is indifferent. Also the second method adds variance so that there is more fluctuation around the mean, so that a higher grade (slightly) is required to get 80% coverage.*

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 :

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

$$E[X] = np$$

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

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 $-\infty < x < \infty$:

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

$$E[X] = \mu$$

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

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

Confidence Intervals:

(a) $100(1-\alpha)\%$ confidence interval for θ : $\hat{\theta} \pm Z_{\alpha/2} SD[\hat{\theta}]$, with known variance.

(b) If $V[\hat{\theta}]$ is unknown, then

$$t = \frac{\hat{\theta} - \theta}{SE[\hat{\theta}]} \sim t$$

with appropriate degrees of freedom. $100(1-\alpha)\%$ confidence interval for θ : $\hat{\theta} \pm t_{\alpha/2} SE[\hat{\theta}]$, where $SE[\hat{\theta}]$ is an estimator of $SD[\hat{\theta}]$.

Estimating Means and Proportions

$$\bar{X} = \frac{1}{n} \sum X_i; \quad s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

$$E[\bar{X}] = \mu; \quad V[\bar{X}] = \frac{\sigma^2}{n}$$

$$\hat{p} = \frac{X}{n}; \quad E[\hat{p}] = p; \quad V[\hat{p}] = \frac{p(1-p)}{n}$$