Queen's University
Faculty of Arts and Sciences
Department of Economics
Fall 2005

Economics 250: 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. **Part A** has 12 questions worth 5 marks each for a total of 60 marks. **Part B** has 1 question worth 40 marks.

PART A: Do all 12 questions (5 Marks each)

A1. In a Binomial distribution model, there is a series of n independent trials and x successes. The probability of success in each trial is π. What are the assumptions using the Binomial distribution. What is the formula for the number of possible sequences of these x successes? What is the probability that there are x successes in n independent trials?

Answer: The assumptons for the binomial distribution are

1. *There are only two outcomes: labelled "success" or "failure"*
2. The probability of one trial to the next is independent
3. The probabilioty of success is π and it is the same for all n

The number of sequences of x successes in n trials is just the number of combinations of x objects chosen from n objects, $C^n_x = \frac{n!}{x!(n-x)!}$. The probability that there are x successes in n independent trials given parameter π is $P(x|n, \pi) = \frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{(n-x)}$.

A2. A factory has two production lines that produce the same product. 70% of the output comes from line A and the rest from line B. Typically the quality of 10% of line A output is above industry standard, and the quality of 20% of line B output is above industry standard. Suppose a batch of product in the warehouse is produced by one of the two lines, but we do not know which line produced it. A random sample of 20 items from this batch is examined and only 1 of them is found to have quality above the industry standard. What is the probability that this batch comes from line A? [**Hint:** use Binomial probability model and Bayes' Theorem.]

Answer: In Binomial distribution model, the probability of having 1 success in 20 trials when the success probability $\pi = 0.1$ *is* $P(x = 1|n = 20, \pi = 0.1) = C_1^{20} \times 0.1^1 \times (1 - 0.1)^{(20-1)} = 0.27017$, *and that of having 1 success in 20 trials when the success probability is* $\pi = 0.2$ *is* $P(x = 1|n = 20, \pi = 0.2) = C_1^{20} 0.2^1 (1 - 0.2)^{(20-1)} = 0.057646$.

Now we know $P(A) = 0.7$, $P(B) = 0.3$, $P(x = 1|A) = 0.27017$, *and* $P(x = 1|B) = 0.057646$. Using the Bayes' Theorem, we know that

$$P(A|x = 1) = \frac{P(A)P(x = 1|A)}{P(x = 1)} = \frac{P(A)P(x = 1|A)}{P(A)P(x = 1|A) + P(B)P(x = 1|B)}$$
$$= \frac{0.7 * 0.27017}{0.7 * 0.27017 + 0.3 * 0.057646} = 0.91622.$$

A3 The population sample proportion is 40% in favour of free trade. Two independent sample proportions are obtained. What is the chance that **one of two** (not both) samples of 100 exceed 50% in favour.

Answer:

$$p = \frac{X}{n} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right) \qquad asngetslarge$$

Therefore for any one sample we have

$$P(p > .50) = P\left(\frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} > \frac{.50 - .40}{\sqrt{\frac{.4(1-.4)}{100}}}\right)$$
$$= P(Z > 2.0412) = 1 - F_Z(2.0412) = 1 - .978 = .0219$$

Let F be the event that the first sample exceeds 50% which happens with probability :

$$P(F = 2) = 21.0219 \times (1 - .0219) = .0428$$

:

A4. Suppose we have wealth holdings from a trimodal population distribution with independently and identically distributed data with a mean of 50 and variance 36. The data is measured in 1000's of dollars. We take two independent samples (call them X and Y with 30 and 40 observations respectively) . What is the probability of the two samples wealth **averages** of X and Y differ by less 2? (Hint: Think what the range difference $X - Y$ can be) Be clear as to what assumptions if any you are making.

Invoking the central limit theorem we know that

$$\bar{X} - \bar{Y} \sim N\left(0, \frac{36}{30} + \frac{36}{40}\right) asthesamplesizesgetlarge$$

$$\begin{aligned} -2 &< P(X-Y) < 2 = P(\left(\frac{-2-0}{\sqrt{\frac{36}{30}+\frac{36}{40}}} < \frac{X-Y-(0)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{n}}} < \frac{2-0}{\sqrt{\frac{36}{30}+\frac{36}{40}}} \right) \\ &= P(-1.38 < Z < 1.38) \\ &= 2*(F_z(1.38) - .5) = 2 \times .416 = .832 \end{aligned}$$

A5. If random variable X is normally distributed with mean μ and variance σ^2, that is, $X \sim N(\mu,\ \sigma^2)$, what is the mean and variance of the random variable X/n, where n is a positive integer? Comment of the distribution as n gets larger

Answer: Applying the formula for the mean and variance of a linear function of a random variable, we know that

$$E(\frac{X}{n}) = \frac{\mu}{n}, \qquad Var(\frac{X}{n}) = \frac{\sigma^2}{n^2}.$$

as the sample size increases the mean is heading to zero but so is the variance. In fact the variance is getting there faster!

A6. The number of persons donating to a charity in any given month is found to be normally distributed with mean 50 and standard deviation 5.

1. What is the probability that the number of person donating to the charity in a given month is between 45 and 50?
2. The probability is 0.5 that the number of persons donating is less than how many?

3. Find the shortest range such that the probability is 0.95 that the number of persons donating will fall in this range

Answer:

1. *Let the CDF of standard normal distribution be denoted by* $F(z)$*. Then,* $P(45 \leq X \leq 50) = P(\frac{45-50}{5} \leq \frac{X-50}{5} \leq \frac{50-50}{5}) = P(-1 < z < 0) = F(0) - F(-1) \approx 0.5 - 0.1587 = 0.3413$.

2. *(2) 50.*

3. *(3) From the symmetric bell-like shape of the PDF of normal distribution, we know that shortest range is symmetric around the mean 50. Thus,* $[50 - 1.96 \times \sigma,\ 50 + 1.96 \times \sigma] = [40.2,\ 59.8] \approx [41,\ 59]$, *with the boundary values inclusive.*

A7. Answer the following:

1. Using the formula for covariance given in the formula sheet, interpret covariance in terms of measuring linear associations between random variables, in no more than 5 lines of words and expressions.

2. Suppose X and Y are two random variables and $Y = a + bX$. X is normally distributed with mean μ and variance σ^2. Use the same formula to find the covariance and correlation between X and Y. [**Hint:** the formula $Var(X) = E(X^2) - [E(X)]^2$ may also be useful at some point.]

Answer:

1. *Covariance between* X *and* Y*, say, is the expectation of the product of* $X - E(X)$ *and* $Y - E(Y)$*, which is linear in both* X *and* Y*. Thus, if high values of* X *tend to be associated with high values of* Y *and vice versa, we would expect this product to be positive, and the stronger the association, the larger the expectation of this product. Similarly for the negative association case.*

2. *First,* $E(X) = \mu$, $Var(X) = \sigma^2$, $E(Y) = a + b\mu$, *and* $Var(Y) = b^2\sigma^2$*. Then,*

$$
\begin{aligned}
& Cov(X, Y) = E(XY) - E(X)E(Y) = E[X(a + bX)] - E(X)E(Y) \\
= \ & aE(X) + bE(X^2) - E(X)E(Y) = aE(X) + b\{Var(X) + [E(X)]^2\} - E(X)E(Y) \\
= \ & a\mu + b(\sigma^2 + \mu^2) - \mu(a + b\mu) = b\sigma^2,
\end{aligned}
$$

and

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{b\sigma^2}{\sigma \mid b \mid \sigma} = sign(b)1.$$

which makes sense since X and Y are perfectly linear related and is either perfecctly positive or negative depending on the sign of b.

A8. When walking by a fund-raising event carried out by students in front of the library, one either donates or not. Suppose, somewhat unrealistically, that the probability of each passer-by donating is the same and is equal to 0.5. A random sample of 40 passers-by is drawn.

1. What kind of probability distribution does this situation correspond to? Which probability distribution can we use to approximate the true probability distribution? Write down the mean and variance of the approximating distribution.
2. Denoting the CDF of the approximating distribution by $F(x)$, write down an expression of the probability that more than 15 persons donated in this sample, in terms of $F(x)$. [**Note:** "in terms of $F(x)$"simply means that, for example, if you find the probability is equal to $2 * F(10)$, then that is your final answer, and you do not need to find out the specific value of $F(10)$ and thus of $2 * F(10)$.
3. In the same manner as in part (2), write down an expression of the probability that the number of persons who donated is between 18 and 22 (inclusive).

Answer:

1. *Binomial; normal; mean=* $n\pi = 40 \times 0.5 = 20$, *variance=* $n\pi(1-\pi) = 10$.
2. $1 - F(15)$.
3. $F(22) - F(17)$. *Some students may have different interpretations of boundary values being inclusive or not and thus different ways to deal with that. For example, in part (2), one may write* $1 - F(15.5)$. *As long as the answer is reasonable and consistent, marks are given.*

A9. The random variable X has probability density function

$$\begin{aligned} f(x) &= x \qquad \text{for } 0 < x < 1, \\ &= 2 - x \qquad \text{for } 1 < x < 2, \\ &= 0 \qquad \text{for all other values of } x. \end{aligned}$$

1. Draw the probability density function for this random variable, show that the area under the curve sums to 1 (and therefore is a proper probability)

2. Find the probability that this random variable takes on values between 0.5 and 1.7.

Answer:

1. *The shape is a triangle so the probability corresponds to the area under a triangle. Figure omitted here.*

$$\begin{aligned} AreaofaTriangle &= \frac{1}{2} \times base \times height \\ & \frac{1}{2} \times 2 \times 1 \\ &= 1 therefore it is a proper probability \end{aligned}$$

2. *This is the sum of two areas* $\frac{(0.5+1)\times 0.5}{2} + \frac{(0.3+1)\times 0.7}{2} = 0.83$.

A10. Suppose that the time it takes a kind of chemical process to complete is uniformly distributed over 2 hours to 4 hours. Suppose a random sample of 10 such processes is taken. Find the probability that at least 2 processes take more than 3.5 hours to complete.

Answer: Let $C \sim U(2,4)$ and therefore

$$P(C > 2.5) = \frac{1}{4-2} \times (4 - 3.5) = .25$$

This can now be characterized by a Binomial distribution, with $\pi = 0.25$ *and* $n = 10$. *Thus, the probability of interest is given by*

$$\begin{aligned} & 1 - P(x=0) - P(x=1) \\ = \; & 1 - \frac{10!}{0!10!}0.25^0(1-0.25)^{10} - \frac{10!}{1!9!}0.25^1(1-0.25)^9 = 0.75597 \; . \end{aligned}$$

A11. Use a simple numerical example to illustrate the meaning of "the sampling distribution of the sample mean". In particular, show the sampling distribution of the sample mean in your example and how it is derived.

Answer

Student Specific.

A12. The variance of a uniform distribution over [0,6] is 3. Suppose that a sample of 10 numbers are drawn each time from this uniform distribution and that such drawing is repeated for 100 times. What distribution can be used to approximate the distribution of the mean calculated from these samples and what are the parameters of this distribution?

1. Draw a rough figure that depicts the PDF's of the original uniform distribution and the approximating distribution.
2. What is the probability that the sample mean is greater than 3?
3. What is the probability that the sample mean is greater than 6?

Answer: $N(3, 0.3)$, *Figure omitted; 0.5; 0.*

PART B: Do all 7 parts, 40 marks total

Suppose that a population contains an infinite number of items and can be characterized by a random variable X that is normally distributed, with $X \sim N(30, 100)$.

1. What is the probability that a randomly drawn item from this distribution is less than 26?
2. What is the probability that X is less than 30?
3. Suppose we draw a random sample of size 100 and calculate the sample mean, $\bar{X}$. What is the probability that this sample mean is less than 26?
4. What is the answer to 3. if the sample size is instead 25?
5. If we draw 5 items from this normal distribution of X, $N(30, 100)$, what is the probability that more than 2 items is less than 30?
6. If we draw 5 items from this normal distribution of X, $N(30, 100)$, what is the probability that the sum of the 5 is less than 140?
7. If we simply take 1 draw of X and multiply it by 5 is

$$P(5X < 140)$$

the same as in 6? Why or why not?

Answer:

1. *If we denote the standard normal random variable by Z, then* $P(X < 26) = P(\frac{X-30}{10} < \frac{26-30}{10}) = P(Z < -0.4) = F(-0.4)$.
2. $P(X < 30) = 0.5$ *by the symmetry of the normal distribution.*
3. $\bar{X} \sim N(30, 1)$, *where* $Var(\bar{X}) = \frac{Var(X)}{n} = 1$ *implies that the standard deviation of the sampling distribution of $\bar{X}$ is 1. Thus, this probability is $F(-4)$, which is roughly 0.*
4. *If the sample size is 25, the sampling distribution is $N(30, 4)$, and* $P(\bar{X} < 26) = P(\frac{\bar{X}-30}{2} < \frac{26-30}{2}) = P(Z < -2) = F(-2) \approx 0.0228$.

5. *This situation can be modeled as a Binomial distribution with parameters* $\pi = 0.5$ *and* $n = 5$. *Thus,* $P(x > 2|\pi = 0.5, n = 5) = P(x = 3|\pi = 0.5, n = 5) + P(x = 4|\pi = 0.5, n = 5) + P(x = 5|\pi = 0.5, n = 5) = \frac{5!}{3!2!}0.5^3 0.5^2 + \frac{5!}{4!1!}0.5^4 0.5^1 + \frac{5!}{5!0!}0.5^5 0.5^0 = 16 \times 0.5^5 = 0.5$. *Normal approximation of the Binomial distribution cannot be used for this part, since* $n\pi(1-\pi) = 1.25 < 9$.

6. *Since the draws are independent we know that*

$$S =_{i=1}^{5} X_i \sim N(150, 500)$$

and we want to know

$$\begin{aligned} P(S &< 140) = \left(\frac{S - \mu_s}{\sigma_s} < \frac{140 - 150}{\sqrt{500}} \right) \\ &= P(Z < -.4472) = 1 - F_Z(.48) = .317 \end{aligned}$$

7. One draw multiplied by 5 implies

$$5X \sim (150, 25 \times 100)$$

and we want to know

$$\begin{aligned} P(5X &< 140) = P\left(\frac{5x - 5\mu_x}{\sqrt{25\sigma_x^2}} < \frac{140 - 150}{\sqrt{25 \times 100}} \right) \\ &= P(Z < -.2) = 1 - F_Z(.2) = .421 \end{aligned}$$

They are not the same since there is much more variation in $5X$ than $_{i=1}^{5} X_i$ so that we find a higher probability of being away from the mean (i.e. less than 140)

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

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}[x_i - \mu]^2$$
$$= \frac{1}{N}\sum_{i=1}^{N} {x_i}^2 - \mu^2$$

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)

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

Probability Theory

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (additive law)$$
$$P(A \cap B) = P(B)P(A \mid B) \quad (multiplicative law)$$

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

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

Counting Formulae

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

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

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)

$$CV = \frac{\sigma}{\mu} \times 100\% for population$$
$$CV = \frac{s}{\bar{X}} \times 100\% for sample$$

Univariate Probability Distributions

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

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

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 $-\infty < x < \infty$:

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

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

Selected Values of the CDF of Standard Normal Distribution:

z	-3	-2	-1.96	-1	0
F(z)	0.0014	0.0228	0.025	0.1587	0.5