

Queen's University
Faculty of Arts and Sciences
Department of Economics
Winter 2009

Economics 250 : Introduction to Statistics

Midterm Exam II :
Time allowed: 80 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 60 possible marks to be obtained. Answer all 6 questions (marks are indicated).

Do all 6 questions

1. (10 Marks) Consider the joint probability distribution of the returns on stocks X and Y

	X	
	10%	20%
0%	0.70	0.00
Y 10%	0.00	0.30

- (a) Compute the marginal probability distributions of the returns on X and Y .
- (b) Compute the covariance and correlation between the returns of X and Y .
- (c) Compute the mean and variance of the portfolio $W = \frac{1}{2}X + \frac{1}{2}Y$.
- (d) Assume that the portfolio is distributed normal with mean and variance as in c). What are the chances that it loses 50% of its value? (i.e. returns are at most half of the mean returns)

Note: as of today, the TSE has lost roughly half its value since the peak reached in 2008.

Solution:

		X		
		1	2	
(a)	Y	0	0.70	0.00
		1	0.00	0.30
			0.70	0.30

(b) $E[X] = 1 \times 0.7 + 2 \times 0.3 = 1.3$
 $E[Y] = 0 \times 0.7 + 1 \times 0.3 = 0.3$

$$V[X] = (1.3 - 1)^2 \times 0.7 + (1.3 - 2)^2 \times 0.3 = 0.21$$

$$V[Y] = (0 - 0.3)^2 \times 0.7 + (1 - 0.3)^2 \times 0.3 = 0.21$$

$$COV(X, Y) =$$

$$\sum_y \sum_x xyP(x, y) - E[X]E[Y] = 1 \times 2 \times 0.3 - 1.3 \times 0.3 = 0.21$$

$$Corr(X, Y) = \frac{COV(X, Y)}{\sigma_1 \sigma_2} = 0.21/0.21 = 1$$

(c) $E[W] = \frac{1}{2}E[X] + \frac{1}{2}E[Y] = \frac{1}{2} \times 1.3 + \frac{1}{2} \times 0.3 = 1.6$

$$V[W] = \frac{1}{4}V[X] + \frac{1}{4}V[Y] + 2(\frac{1}{2} \times \frac{1}{2}COV[X, Y])$$

$$= 0.21/4 + 0.21/4 + 0.21/2 = 0.21$$

(d) We are looking for

$$\begin{aligned} P(W < \frac{\mu}{2}) &= P(Z < \frac{\mu/2 - \mu}{\sigma_W}) \\ &= P(Z < \frac{-1.6/2}{\sqrt{0.21}}) \\ &= P(Z < \frac{-1.6/2}{\sqrt{0.21}}) \\ &= P(Z < -1.75) = 0.04 \end{aligned}$$

2. (10 Marks) The time at which a bus arrives at a station is uniform over an interval (a, b) with mean 2:00 PM and standard deviation $\sqrt{12}$ minutes. Determine the values of a and b.

Solution: Let t be the time the bus arrives. Then, we are given that

$$t \sim U(a, b)$$

From the properties of the uniform r.v. we know that $E[t] = \frac{a+b}{2}$ and $\frac{(b-a)^2}{12}$. Using the information in the question, we have that

$$\begin{aligned}\frac{a+b}{2} &= 2 : 00\text{PM and} \\ \frac{(b-a)^2}{12} &= (\sqrt{12})^2\end{aligned}$$

Simplifying we have,

$$\begin{aligned}a+b &= 4 : 00\text{PM and} \\ b-a &= 12\end{aligned}$$

Solving for a and b we have $a = 1 : 54$ and $b = 2.06$.

3. (10 Marks) A pizza delivery service delivers to a campus dorm. Delivery times follow a normal distribution with mean 20 minutes and standard deviation 4 minutes.
- (a) What is the probability that a delivery will take between 15 and 25 minutes?
 - (b) Pizza is free if the delivery takes more than 30 minutes. What is the probability of getting a free pizza from a single order?
 - (c) During final exams, a student plans to order pizza five consecutive evenings. Assume that these delivery times are independent of each other. What is the probability that the student will get at least one free pizza?

Solution: Let d be the delivery time for a pizza. Then, we know that $d \sim N(20, 4^2)$.

- (a) The probability that the delivery time is between 15 and 20 min-

utes is

$$\begin{aligned}P(15 \leq d \leq 25) &= P\left(\frac{15 - 20}{\sqrt{4^2}} \leq Z \leq \frac{25 - 20}{\sqrt{4^2}}\right) \\&= P(-1.25 \leq Z \leq 1.25) \\&= F_z(1.25) - F_z(-1.25) \\&= F_z(1.25) - (1 - F_z(1.25)) \\&= 0.79\end{aligned}$$

(b) The probability of getting a free pizza is

$$\begin{aligned}P(d > 30) &= P\left(\frac{d - 20}{4} > \frac{30 - 20}{4}\right) \\&= P(Z > 2.5) \\&= 1 - P(Z \leq 2.5) \\&= 1 - F_z(2.5) \\&= 0.006\end{aligned}$$

(c) Let X be the number of free pizzas the student gets in 5 deliveries. This is a binomial random variable with 5 trials and success probability given by $P(d > 30) = 0.01$. We are looking for the probability of at least one success:

$$\begin{aligned}P(X \geq 1) &= 1 - P(X \leq 0) \\&= 1 - P(X = 0) \\&= 1 - \binom{5}{0} 0.006^0 0.9938^5 \\&= 0.03\end{aligned}$$

4. (10 Marks) The breaking strength of a certain type of rope produced by a certain vendor is normal with mean 95 and standard deviation 9.5.

- (a) What is the probability that, in a random sample of size 10 from the stock of this vendor, the breaking strength of at least two are over 100?
- (b) Are we likely to get a good approximation using the normal distribution in this case?

- (c) Estimate this probability using the normal distribution.
- (d) Apply the continuity correction. Is there is a significant improvement?

Solution:

- (a) Let B be the breaking strength of a rope. According to the question, $B \sim N(95, 9.5^2)$. The probability that the breaking strength of one rope exceeds 100 is simply

$$\begin{aligned}
 P(B > 100) &= P\left(Z > \frac{100 - 95}{9.5}\right) \\
 &= P(Z > 0.53) \\
 &= 1 - P(Z \leq 0.53) \\
 &= 1 - F_z(0.53) = 0.30.
 \end{aligned}$$

Let X be the number of ropes with strength that exceeds 100 in the sample of 10. We have a binomial problem with success probability $P(B > 100) = 0.3$. Then the probability at least two ropes have breaking strength over 100 is given by:

$$\begin{aligned}
 P(X \geq 2) &= 1 - P(X \leq 7) \\
 &= 1 - 0.998 \\
 &= 0.002
 \end{aligned}$$

using the tables for the cumulative binomial distribution.

- (b) The approximation is good only if $n\pi(1 - \pi) > 9$. Here, $\pi = 0.3, 1 - \pi = 0.7, n = 10$ so $n\pi(1 - \pi) = 10 \times 0.3 \times 0.7 = 2.1 < 9$. So the approximation is not likely to be a good one.
- (c) Using the normal approximation to the binomial,

$$\begin{aligned}
 P(X \geq 2) &\approx P\left(\frac{X - E(X)}{\sqrt{V(X)}} \geq \frac{2 - E(X)}{\sqrt{V(X)}}\right) \\
 &= P\left(Z \geq \frac{2 - 10 \times 0.3}{\sqrt{10 \times 0.3 \times 0.7}}\right) \\
 &= P(Z \geq -0.69) \\
 &= 1 - P(Z \leq -0.69) \\
 &= 1 - (1 - P(Z \leq 0.69)) \\
 &= P(Z \leq 0.69) = 0.75
 \end{aligned}$$

which is a terrible approximation.

(d) Applying the continuity correction we have:

$$\begin{aligned}P(X \geq 1.5) &\approx P\left(\frac{X - E(X)}{\sqrt{V(X)}} \geq \frac{2 - E(X)}{\sqrt{V(X)}}\right) \\&= P\left(Z \geq \frac{1.5 - 10 \times 0.3}{\sqrt{10 \times 0.3 \times 0.7}}\right) \\&= P(Z \geq -1.04) \\&= P(Z \leq 1.04) \\&= 0.85\end{aligned}$$

which is even worse!

5. (10 Marks) A manufacturer of a liquid detergent claims that the mean weight of liquid in containers sold is at least 30 ounces. It is known that the population distribution of weights is normal with standard deviation 1.3 ounces. In order to check the manufacturer's claim, a random sample of 16 containers of detergent is examined. The claim will be questioned if the sample mean weight is less than 29.5 ounces. What is the probability that the claim will be questioned if, in fact, the population mean weight is 30 ounces?

Solution: Let W be the weight of a randomly chosen liquid container. Then we know that $W \sim N(30, 1.3^2)$. A sample size of 16 is taken and so the sampling distribution of the sample mean is

$$\bar{W} \sim N\left(30, \frac{1.3^2}{16}\right)$$

Therefore, the probability that the claim will be questioned is simply

$$\begin{aligned}P(\bar{w} < 29.5) &= P\left(\frac{\bar{w} - 30}{1.3/4} < \frac{29.5 - 30}{1.3/4}\right) \\&= P(Z < -1.43) \\&= 1 - F_z(1.43) \\&= 0.0764\end{aligned}$$

6. (10 Marks) A sample of 100 students is taken to determine which of two brands of beer is preferred in a blind taste test. Suppose that, in the whole population of students, 50% would prefer brand A.

- (a) What is the probability that more 60% of the sample members prefer brand A?
- (b) What is the probability that between 45% and 55% of the sample members prefer brand A?

Solution: Let p be the proportion of people in the general population that prefer brand A to B. Then, from the question we know that $p = 0.5$. Now, if we take a sample of $n=100$ students the sampling distribution of \hat{p} is

$$N\left(p, \frac{p(1-p)}{n}\right) = N\left(0.5, \frac{0.25}{100}\right) = N\left(0.5, \frac{1}{400}\right)$$

- (a) The probability that more than 60% of the sample members prefer brand A is

$$\begin{aligned} P(\hat{p} > 0.6) &= 1 - P(\hat{p} \leq 0.6) \\ &= 1 - P\left(Z \leq \frac{0.6 - p}{\sqrt{p(1-p)/n}}\right) \\ &= 1 - P\left(Z \leq \frac{0.6 - 0.5}{0.05}\right) \\ &= 1 - F_z(2) \\ &= 1 - 0.9772 = 0.0228 \end{aligned}$$

- (b) The probability that between 45% and 55% of the sample prefer brand A is

$$\begin{aligned} P(0.45 \leq \hat{p} \leq 0.55) &= P\left(\frac{0.45 - p}{\sqrt{p(1-p)/n}} \leq Z \leq \frac{0.55 - p}{\sqrt{p(1-p)/n}}\right) \\ &= P\left(\frac{0.45 - 0.5}{0.05} \leq Z \leq \frac{0.55 - 0.5}{0.05}\right) \\ &= P(-1 \leq Z \leq 1) \\ &= 2F_z(1) - 1 \\ &= 0.6826 \end{aligned}$$

Formula Sheet

Statistics Formulas

Notation

- All summations are for $i = 1, \dots, n$ 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

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

Alternatively

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

Grouped Data (with k classes)

$$\bar{X} = \frac{1}{n} \sum_{j=1}^k \nu_j f_j \quad \text{where } \nu_j \text{ is the class mark for class } j$$

$$s^2 = \frac{1}{n-1} \sum_{j=1}^k f_j (\nu_j - \bar{X})^2$$

Probability Theory

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (\text{additive law})$$

$$P(A \cap B) = P(B)P(A | B) \quad (\text{multiplicative law})$$

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

$$P(A) = \sum_i^n P(A \cap E_i) = \sum_i^n P(E_i)P(A | E_i)$$

$$P(E_i | A) = \frac{P(E_i)P(A | E_i)}{P(A)} \quad (\text{Bayes' Theorem})$$

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:

$$\mu_x = E[X] = \sum_x xP(X=x)$$
$$\sigma_x^2 = V[X] = E[(x - \mu_x)^2] = \sum_x (x - \mu_x)^2 P(X=x)$$
$$\sigma_x^2 = E[X^2] - (E[X])^2$$

The **covariance** of X and Y is

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

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)

$$CV = \frac{\sigma}{\mu} \times 100\% \text{ for population}$$

$$CV = \frac{s}{\bar{X}} \times 100\% \text{ for sample}$$

Univariate Probability Distributions

Binomial Distribution: For $x = 0, 1, 2, \dots, n$ 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 $-\infty < x < \infty$:

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

$$E[X] = \mu$$

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

$$X \sim N(\mu_X, \sigma_X^2)$$

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

Estimators in General

If $\hat{\theta} \sim N(\theta, V[\hat{\theta}])$, say, then under appropriate conditions:

$$Z = \frac{\hat{\theta} - \theta}{SD[\hat{\theta}]} \sim N(0, 1)$$

Estimating Means and Proportions

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

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

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

Differences of Means and Proportions for Independent Samples

$$s_{pool}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2; \quad V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If variances are assumed (i.e. $\sigma_1^2 = \sigma_2^2$) to be the same we may, estimate

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_{pool}^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\begin{aligned} p_{pool} &= \frac{X_1 + X_2}{n_1 + n_2}; \\ E[p_1 - p_2] &= \pi_1 - \pi_2; \\ V[f_1 - f_2] &= \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2} \end{aligned}$$