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

Economics 250 : Introduction to Statistics

Midterm Exam II
Time allowed: 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. Do all 8 questions (marks are indicated).

1. **(10 marks)** Suppose that five people, including you and a friend, line up at random. Let the discrete random variable X denote the number of people standing between you and your friend. Please write down the probability distribution function of X, $P(X = x)$, and graph it. Also, write down the cumulative distribution function of X, $P(X \leq x)$.

   Solution:
   The pdf is:

x	$P(X = x)$
0	4/10
1	3/10
2	2/10
3	1/10

   The graph should have 4 bars, one each at $x = 0, 1, 2, 3$, of height $P(X = x)$ respectively.

   The cdf is:

x	$P(X \leq x)$
0	4/10
1	7/10
2	9/10
3	10/10

2. **(15 marks)** Consider the joint probability distribution

		X	
		1	2
Y	0	0.30	0.20
	1	0.25	0.25

(i) Compute the marginal probability distributions for X and Y.
(ii) Compute the covariance and correlation for X and Y.
(iii) Compute the mean and variance for the linear function $W = 2X + Y$.

Solution:
(i) $P(X = 1) = 0.55, P(X = 2) = 0.45,$
$P(Y = 0) = 0.5, P(Y = 1) = 0.5$
(ii) The means are:
$E[X] = 1 * 0.55 + 2 * 0.45 = 1.45$
$E[Y] = 0 * 0.5 + 1 * 0.5 = 0.5$
$Cov(X,Y) = \sum_y \sum_x xyP(x,y) - E[X]E[Y] = 0.25 * 1 + 0.25 * 2 - 0.5 * 1.45 = 0.025$
Now $\sigma_X = 0.497$, $\sigma_Y = 0.5$ (iii) The mean of W is $E[W] = 2 * E[X] + E[Y] = 2 * 1.45 + 0.5 = 2.4$
The variance of W is $V[W] = 2^2 V[X] + V[Y] + 2 * 1 * 2 * cov(X,Y) = 4 * 0.2475 + 0.25 + 4 * 0.025 = 1.34$

3. **(10 marks)** As the lawyer for a client accused of murder, you are looking for ways to establish "reasonable doubt" in the minds of the jurors. Central to the prosecutor's case is testimony from a forensics expert who claims that a blood sample taken from the scene of the crime matches the DNA of your client. One-tenth of 1% of the time, though, such tests are in error. Suppose, your client is actually guilty. If six other labs in the country are capable of doing this kind of DNA analysis (and you hire them all), what are the chances that at least one will make a mistake and conclude that your client is innocent?

Solution:
Each of the six analyses constitutes an independent trial, where $\pi = P(\text{lab takes makes mistake}) = 0.001$. Let X = number of labs that make mistake (and call client innocent). Then

$$\begin{aligned} P(\text{at least one lab says client is innocent}) &= P(X \geq 1) \\ &= 1 - P(X = 0) \\ &= 1 - \binom{6}{0}(0.001)^0(0.999)^6 \\ &= 0.006 \end{aligned}$$

So the chances of establishing "reasonable doubt" appear slim!

4. **(10 marks)** Suppose the length of time that you have to wait at a bank teller's window is uniformly distributed between 0 and 25 minutes.
(i) Draw the probability distribution function for the length of time you have to wait.
(ii) Find the probability that you wait less then 5 minutes.
(iii) Find the probability that you wait more than 15 minutes.
(iv) Find the probability that you wait between 12 and 18 minutes.

Solution:
(i) The pdf is $f(x) = 1/25$ for $0 \leq x \leq 25$. (ii) $5/25 = 0.2$ (iii) 1- 15/25 = 10/25 = 0.4 (iv) (18-12)/25 = 6/25

5. **(10 marks)** Records for the past several years show that the amount of money collected daily by a prominent televangelist is normally distributed with a mean of \$20,000 and a standard deviation of \$5,000.
(i) What are the chances that tomorrow's donation will exceed \$30,000?
(ii) What are the chances that it will at most \$10,000?
(iii) What are the chances that it will be between \$15,000 and \$25,000?

Solution:
(i) Let the random variable X denote tomorrow's donation. Then we

know that $X \sim N(20,000, 5,000^2)$. Therefore,

$$\begin{aligned} P(X > 30,000) &= 1 - P(X \leq 30,000) \\ &= 1 - P(Z \leq \frac{30,000 - 20,000}{5,000}) \\ &= 1 - P(Z \leq 2) \\ &= 0.0228 \end{aligned}$$

(ii)

$$\begin{aligned} P(X \leq 10,000) &= P(Z \leq \frac{10,000 - 20,000}{5,000}) \\ &= P(Z \leq -2) \\ &= 1 - P(Z \leq 2) \\ &= 0.0228 \end{aligned}$$

(iii)

$$\begin{aligned} P(15,000 \leq X \leq 25,000) &= P\frac{15,000 - 10,000}{5,000} \leq Z \leq \frac{25,000 - 20,000}{5,000}) \\ &= P(-1 \leq Z \leq 1) \\ &= F_z(1) - F_z(-1) \\ &= 2F_z(1) - 1 \\ &= 0.6826 \end{aligned}$$

6. **(15 marks)** Air Canada knows that on average only 90% of the ticket-holders for the Thursday night Toronto-Montreal flight will show up at the gate in time to board the plane. For that reason, the company routinely sells more tickets than their aircraft has seats. Suppose that they have just booked 140 passengers for next week's flight (and the plane has 130 seats).
(i) Write a formula for the exact probability that not all the ticket-holders who show up next Thursday can be accommodated.
(ii) Use the normal approximation to obtain the probability in part (i)
(iii) Use the continuity correction to obtain the probability in part (i)

Solution:
(i) This is just a binomial problem. Let X be the random variable that

denotes the number of people who bought a ticket and showed up for the flight. Then, we have a problem whenever $P(131 \leq X \leq 140)$. The probability of a passenger showing up is $\pi = 0.9$. Therefore,

$$P(131 \leq X \leq 140) = \sum_{k=131}^{140} \binom{140}{k} (0.9)^k (0.1)^{140-k}$$

(ii) The mean of X is $140 \times 0.9 = 126$. The variance of X is $\sqrt{140 \times 0.9 \times 0.1} = 3.55$. Then,

$$\begin{aligned} P(131 \leq X \leq 140) &\approx P(\frac{131 - 126}{3.55} \leq Z \leq \frac{140 - 126}{3.55}) \\ &= P(1.41 \leq Z \leq 3.94) \\ &= F_z(3.94) - F_z(1.41) \\ &\approx 1 - 0.9207 \\ &= 0.08 \end{aligned}$$

(iii) Applying the continuity correction we have

$$\begin{aligned} P(131 \leq X \leq 140) &\approx P(\frac{130.5 - 126}{3.55} \leq Z \leq \frac{140.5 - 126}{3.55}) \\ &= P(1.27 \leq Z \leq 4.08) \\ &= F_z(4.08) - F_z(1.27) \\ &\approx 1 - 0.8980 \\ &= 0.102 \end{aligned}$$

7. **(15 marks)** A manufacturing plant has 438 blue-collar employees. Of this group, 239 are concerned about future health care benefits. A random sample of 80 of these employees was questioned to estimate the population proportion concerned about future health care benefits.
(i) What is the standard error of the sample proportion who are concerned?
(ii) What is the probability that the sample proportion is less than 0.5?
(iii) What is the probability that the sample proportion is between 0.5 and 0.6?

Solution:

(i) The population proportion is $\pi = 239/438 \approx 0.54566$. The sample size is 9.

The variance of the sample proportion is

$$\begin{aligned} V[\hat{p}] &= Var[p]/n^2 = \pi(1-\pi)/n \\ &= \frac{(\frac{239}{438})(1-\frac{239}{438})}{9} \end{aligned}$$

Therefore the standard error is simply $s_{\hat{p}} = \sqrt{V[\hat{p}]} \approx 0.055668$.

(ii)

$$\begin{aligned} P(\hat{p} < 0.5) \approx P(Z < \frac{0.5-\pi}{s_{\hat{p}}}) &= F_z(-0.82) \\ &= 1 - F_z(0.82) \\ &= 0.2061 \end{aligned}$$

(iii)

$$\begin{aligned} P(0.5\hat{p} < 0.6) &\approx P(\frac{0.5-\pi}{s_{\hat{p}}} Z < \frac{0.6-\pi}{s_{\hat{p}}}) \\ &= F_z(0.98) - F_z(-0.82) \\ &= 0.8365 - 0.2061 \\ &= 0.6304 \end{aligned}$$

8. **(15 marks)** It has been found that times taken by people to complete a particular tax form follow a normal distribution with mean 100 minutes and standard deviation 30 minutes. A random sample of nine people who have completed this tax form was taken.
(i) What is the probability that the sample mean time taken is more than 120 minutes?
(ii) The probability is 0.2 that the sample mean time taken is less than how many minutes?

Solution:

(i) We know that the data is normally distributed, that is $X_i \sim N(100, 30^2)$ and that the sample size, $n = 9$. Thus,

$$P(\bar{X} > 120) = 1 - P(\bar{X} \leq 120) = 1 - P(Z \leq \frac{120 - 100}{\sqrt{\frac{30^2}{9}}})$$
$$= 1 - P(Z \leq 2)$$
$$= 0.0228$$

(ii) We need to find the x such that $P(\bar{X} < x) = 0.2$. We can rewrite this in terms of the standard normal as follows:

$$P(\bar{X} < x) = 0.2 \implies P(Z < \frac{x - 100}{\sqrt{\frac{30^2}{9}}}) = 0.2$$

Notice that we don't have tables for less than 0.5. That is, x must be to right of the mean: therefore, $\frac{x-100}{\sqrt{\frac{30^2}{9}}} < 0$. So,

$$P(Z < \frac{x - 100}{\sqrt{\frac{30^2}{9}}}) = 1 - P(Z < -1 \times (\frac{x - 100}{\sqrt{\frac{30^2}{9}}}))$$
$$= 1 - P(Z < \frac{100 - x}{\sqrt{\frac{30^2}{9}}})$$

Therefore, $P(Z < \frac{100-x}{\sqrt{\frac{30^2}{9}}}) = 0.8$. Now, reverse-looking up 0.8 in the tables we find that

$$\frac{100 - x}{\sqrt{\frac{30^2}{9}}} = 0.8416$$

Solving for x we have $x = 91.58$.

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)

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

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

Probability Theory

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

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

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

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:

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

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)

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

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 $-\infty < x < \infty$:

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

$$E[X] = \mu$$

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

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

Estimators in General

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

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

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

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

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

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

Differences of Means and Proportions for Independent Samples

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

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

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

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

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