Queen's University
Faculty of Arts and Sciences
Department of Economics
Fall 2007

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 standard normal table. 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 9 questions (marks are indicated)

Do all 9 questions

1. **(10 marks)** Why do we not require additional distribution tables (other than the standard normal) for the sample mean when the number of observations is large? Explain what assumptions are required in your answer.

   Solution If the observations are identically and independently distributed random variables from a common distribution then their mean, $\bar{X}$ is distributed $N(\mu, \sigma_{\bar{X}})$ in the limit as n goes to infinity. That is, we don't need the distribution tables other than the normal because we can apply the CLT.

2. **(15 marks)** The mean test score for 5th graders is 32 while the mean test score for 3rd graders is 28. The variances of the two test scores are 60 and 50 respectively, and the test scores for both groups are normally (and independently) distributed. You randomly sample 14 scores from the 5th graders and 12 scores from the 3rd graders. What is the probability that the mean of the 12 3rd graders will exceed the mean of the 14 5th graders by 2 or more ?

   Solution The mean of the difference is simply the difference of the means. That is, the mean of the difference is 28-32 = -4. The variance of the difference is the sum of the variances: $60/14 + 50/12 \approx 8.452$. The standard deviation is then $\sqrt{8.452} \approx= 2.91$. The difference is normally distributed as grades for both groups are normally distributed. Then $P($ difference $\geq 2) = P(Z \geq 2.062) \approx 0.0196$.

3. **(10 marks)** Let X be a uniform random variable over the interval $(0, 1 + \theta)$, where $0 < \theta < 1$ is a given parameter. Suppose the $P(X \leq 0.4) = 0.8$. Find θ and show all work.

   Solution If $P(X \leq 0.4) = 0.8$ then $0.4 \times 1/(1 + \theta) = 0.8$. Therefore, solving for θ yields $\theta = -0.5$.

4. **(10 marks)** Suppose that, for a discrete random variable X, $E[X] = 2$ and $E[X(X - 4)] = 5$. Find the variance and the standard deviation of $-4X + 12$.

   Solution $E[X(X - 4)] = 5$ implies that $E(X^2)\ -4E(X) = 5$. Using $E(X) = 2$ we have $E(X^2) = 13$. Therefore,

   $$Var(X) = E(X^2) - [E(X)]^2 = 13 - 4 = 9.$$

   Now $Var(-4X + 12) = 16 \times Var(X) = 16 \times 9 = 144$ and the standard deviation is just $\sqrt{144} = 12$.

5. **(10 marks)** Karen is interested in two games, Keno and Bolita. To play Bolita, she buys a ticket for \$1, draws a ball at random from a box of 100 balls numbered 1 to 100. If the ball drawn matches the number on her ticket, she wins \$75; otherwise, she loses. To play Keno, Karen bets \$1 on a single number that has a 25% chance to win. If she wins, they will return her dollar plus two dollars more; otherwise, they keep the dollar.

   a. What is the expected value of playing Keno or Bolita ? In terms of the expected value, does Karen favour either game?
   b. Which one is riskier ? (compare the variance of her gains between the two games). Comment on the games in terms of risk.

   Solution Let B and K be the amounts that Karen gains in one play of Bolita and Keno, respectively. Then

   $$E(B) = 74 \times 0.01 + (-1) \times 0.99 = -0.25$$
   $$E(K) = 2 \times 0.25 + (-1) \times 0.75 = -0.25$$

   Thus, in the long run both games have roughly the same payoff so Karen should be indifferent between the games. Now, computing the variances we find that

   $$Var(B) = E[(B - \mu)^2] = (74 + 0.25)^2 \times 0.01 + (-1 + 0.25)^2 \times 0.99 = 55.69 \text{ and}$$
   $$Var(K) = E[(B - \mu)^2] = (2 + 0.25)^2 \times 0.25 + (-1 + 0.25)^2 \times 0.75 = 1.6875$$

   Therefore, Keno involves much less risk that Bolita. In Keno players wins in smaller amounts but then win more often, in Bolita its the reverse.

6. **(10 marks)** Starting at 5:00 A.M., every half hour there is a flight from Toronto to Montreal. Suppose that none of these planes is completely sold out and that they always have room for passengers. A person who wants to fly to Montreal arrives at the airport at a random time (uniformly) between 8:45 A.M. and 9:45 A.M. Draw a diagram illustrating the relevant probabilities and find the probability that she waits:

   a. at most 10 minutes

   b. at least 15 minutes

   Solution Let the passenger arrive at the airport X minutes past 8:45. Then X is a uniform random variable over the interval (0, 60). Hence the density function of X is given by

$$f(x) = 1/60$$

   for $0 < x < 60$ and 0 otherwise. Now, the passenger waits at most 10 minutes if she arrives between 8:50 and 9:00 or 9:20 and 9:30. That is, if $5 < X < 15$ or $35 < X < 45$. So the answer to (a) is:

$$P(5 < X < 15) + P(35 < X < 45) = 10/60 + 10/60 = 1/3$$

   The passenger waits at least 15 minutes, if she arrives between 9:00 and 9:15 or 9:30 and 9:45. That is, if $15 < X < 30$ or $45 < X < 60$. Thus, the answer to (b) is:

$$P(15 < X < 30) + P(45 < X < 60) = 15/60 + 15/60 = 1/2$$

7. **(10 marks)** Mensa (from the Latin word for "mind") is an international society devoted to intellectual pursuits. Any person who has an IQ in the upper 2% of the general population is eligible to join. What is the lowest IQ that will qualify a person for membership? Assume that IQs are normally distributed with $\mu = 100$ and $\sigma = 16$.

   Solution Let the random variable Y denote a person's IQ, and let the constant y_L be the lowest IQ that qualifies someone to be a card-carrying Mensan. The two are related by a probability equation:

$$P(Y \geq y_L) = 0.02$$

   or equivalently,

$$P(Y < y_L) = 1 - 0.02 = 0.98$$

   Applying the Z transformation we have:

$$P(Y < y_L) = P\left(\frac{Y - 100}{16} < \frac{y_L - 100}{16}\right) = P\left(Z < \frac{y_L - 100}{16}\right) = 0.98$$

Now, using the table for the normal distribution we have that $\frac{y_L - 100}{16} \approx 2.05$. Therefore, solving for y_L, we have $y_L = 100 + 16(2.05) \approx 133$. That is, 133 is the lowest acceptable IQ for Mensa.

8. **(10 marks)** The elevator in the athletic dorm at Swampwater Tech has a maximum capacity of 2400 lbs. Suppose that 10 football players get on at the twentieth floor. If the weights of Tech's players are normally distributed with a mean of 220 lbs and a standard deviation of 20 lbs, what is the probability that the elevator will crash?

   Solution Let the random variables $Y_1, Y_2, ..., Y_{10}$ denote the weights of the 10 players. At issue is the probability that $Y = \sum_{i=1}^{10} Y_i$ exceeds 2400 lb. But

$$P\left(\sum_{i=1}^{10} Y_i > 2400\right) = P\left(\frac{1}{10}\sum_{i=1}^{10} Y_i > \frac{1}{10}2400\right) = P(\bar{Y} > 2400)$$

   Now applying a Z transformation we have:

$$P(\bar{Y} > 2400) = P\left(\frac{\bar{Y} - 220}{20/\sqrt{10}} > \frac{240.0 - 220}{20/\sqrt{10}}\right) = P(Z > 3.16) \approx 0.0008.$$

   Therefore, the chances of a crash are very small indeed.

9. **(15 marks)** The City of Kingston has 74,806 registered cars. A city by-law requires each owner to display a bumper sticker showing that the owner has paid an annual car tax of \$50. The law requires that the sticker be purchased during the month of the owner's birthday. This year's budget assumes that at least \$306,000 in revenue will be collected in November. Assume that probability of a birth in Novemeber is $\frac{1}{12}$.

   a. What is the probability that taxes reported in November will be less than anticipated \$306,000 ? Just write the expression for the probability, you do not need to calculate it as it is beyond the computational power of most calculators!

   b. Use the normal approximation with continuity corrections to compute the probability in *a*.

   Solution

   a. Let the binomial random variable X denote the number of cars (out of 74,806) whose fees will be paid in November. Since we can reasonably assume that roughly 1/12 of all birthdays will come in November,

$$\begin{aligned} P(\text{car tax revenue} < \$306,000) &= P(50X < \$306,000) \\ &= P(X < 6120) \\ &= P(X \leq 6119) \\ &= \sum_{k=0}^{6119} \binom{74,806}{k} \left(\frac{1}{12}\right)^k \left(\frac{11}{12}\right)^{74,806-k} \end{aligned}$$

b. Notice first that the continuity correction modifies the target event to be $X \leq 6119.5$ rather than $X \leq 6119$. Also,

$$np = 74,806(1/12) = 6233.8$$

and

$$np(1-p) = 74,806(1/12)(11/12) = 5714.3$$

Therefore,

$$\begin{aligned}
P(\text{budget deficit}) = P(X \leq 6119.5) = \\
& P\left(\frac{X - 6233.8}{\sqrt{5714.3}} \leq \frac{6119.5 - 6233.8}{\sqrt{5714.3}}\right) \\
&= P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq -1.51\right) \\
&\approx P(Z \leq -1.51) \\
&= 0.0655
\end{aligned}$$

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

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

Sample Variance

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

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 $-\infty < x < \infty$:

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

$$E[X] = \mu$$

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

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