**Queen's University**

**Faculty of Arts and Sciences**

**Department of Economics**

**Winter 2007**

# Economics 250 A/B: Introduction to Statistics

## Term Test 2

March 14th, 2007

## Solutions and marking guidelines

The total points are 100 points.

**Question 1** 10 points

The number of cars sold per day in Dan's Used Car Superstore is defined by the following probability distribution.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P(X = x)$ | 0.10 | 0.20 | 0.30 | 0.20 | 0.10 | 0.10 |

**(a)** Construct $F(x)$, the cumulative probability distribution function for $X$.

**(b)** Calculate $P(X > 3)$

**(c)** Calculate $P(X \leq 4)$

**Answers**

(a)

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $F(x)$ | 0.10 | 0.30 | 0.60 | 0.80 | 0.90 | 1.00 |

(b) $P(X > 3) = 0.1 + 0.1 = 0.2$

(c) $P(X \leq 4) = 0.10 + 0.20 + 0.30 + 0.20 + 0.10 = 0.9$ or
$P(X \leq 4) = 1 - P(X = 5) = 1 - 0.1 = 0.9$

**Marking guidelines**

(a) **Out of 4 points** Take off 1 point if one or two entries are incorrect because of calculation errors.

(b) **Out of 3 points** Take off 1 point for a calculation error.

(c) **Out of 3 points** Take off 1 point for a calculation error.

**Question 2** 14 points

Suppose $X$ is a discrete random variable that can take 3 different values. Its probability distribution is

| $x$ | $P(X = x)$ |
|---|---|
| 2 | 0.2 |
| 3 | 0.3 |
| 5 | 0.5 |

**(a)** Calculate $E[X]$.

**(b)** Calculate $V[X]$.

**(c)** Calculate $E[X^2]$.

**(d)** Calculate $E[2 + 3X]$.

**(e)** Calculate $V[2 + 3X]$.

**Answers**

(a) $E[X] = \sum_x xP(X = x) = 2 \times 0.2 + 3 \times 0.3 + 5 \times 0.5 = 3.8$

(b)
$$V[X] = \sum_x (x - \mu)^2 P(X = x)$$

$= (2-3.8)^2 \times 0.2 + (3-3.8)^2 \times 0.3 + (5-3.8)^2 \times 0.5 = 0.648 + 0.192 + 0.72 = 1.56$

or
$$V[X] = E[X^2] - (E[X])^2$$
$$= 2^2 \times 0.2 + 3^2 \times 0.3 + 5^2 \times 0.5 - 3.8^2 = 16 - 14.44 = 1.56$$

(c) $E[X^2] = \sum_x x^2 P(X = x) = 2^2 \times 0.2 + 3^2 \times 0.3 + 5^2 \times 0.5 - 3.8^2 = 16$

(d) $E[2 + 3X] = 2 + 3E[X] = 2 + 3 \times 3.8 = 13.4$

(e) $V[2 + 3X] = 3^2 V[X] = 9 \times 1.56 = 14.04$

**Marking guidelines**

(a) **Out of 4 points** Take off 2 points if the answer is incorrect because of a calculation error.

(b) **Out of 4 points** Take off 2 points if the answer is incorrect because of a calculation error.

(c) **Out of 2 points** Take off 1 point if the answer is incorrect because of a calculation error.

(d) **Out of 2 points** Take off 1 point if the answer is incorrect because of a calculation error.

(e) **Out of 2 points** Take off 1 point if the answer is incorrect because of a calculation error.

**Question 3** 6 points

**(a)** *Multiple choice question. Select the concept corresponding to the following definition:*

A variable that takes on numerical values determined by the outcome of a random experiment. *Remember to answer in your exam booklet.*

(A) Constant

(B) Random variable

(C) Expectation

(D) Standard deviation

(E) Covariance

**(b)** The probability density function of a continuous random variable is given by

$$f(x) = \begin{cases} 2x^3 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X = 0.5)$? Explain Briefly.

**Answers**

(a) (B), random variable.

(b) Zero. A valid explanation is either (i) for a continuous random variable, the probability of a specific value is zero since there is an infinite number of possible outcomes, or (ii) the probability is the area under the density curve, the area under a single point is zero.

**Marking guidelines**

(a) **Out of 2 points**. No part marks.

(b) **Out of 4 points.** 2 points (no part marks) for the numerical answer "zero." 2 points for the explanation. Give part marks for the explanation when you think it is appropriate.

**Question 4** 12 points

Suppose that you are in charge of marketing airline seats for a major carrier. Four days before the flight you have 16 seats remaining on the plane. You know from past experience that 80% of the people that purchase tickets in this time period will actually show up for the flight. If you sell 20 extra tickets,

(a) what is the probability that you will overbook the flight

(b) what is the probability that you will have at least one empty seat?

**Answers**

Let $X$ denote the number of people who show up. $X$ follows a binomial distribution with $n = 20$ and $\pi = 0.8$.

(a) The probability of interest is $P(X \geq 17) = 1 - P(X \leq 16) = 1 - 0.589 = 0.411$

(b) The probability of interest is $P(X \leq 15) = 0.370$

**Marking guidelines**

(a) **Out of 6 points.** Take off 2 points for a calculation error.

(b) **Out of 6 points.** Take off 2 points for a calculation error. Give 1 mark if the answer is $P(X \leq 15) = 1 - 0.370 = 0.63$.

**Question 5** 10 points

A campus food store sells hamburgers for \$3 each and donuts for \$1 each. Daily sales of hamburgers follow a normal distribution with mean 40 and standard deviation 10. Daily sales of donuts follow a normal distribution with mean 100 and standard deviation 30. The daily sales of hamburgers and the daily sales of donuts are independent.

**(a)** What is the distribution (uniform, binomial, *etc.*) of the daily total revenues of this campus food store? Briefly explain.

**(b)** Calculate the mean/expectation of the daily total revenues of this campus food store.

**(c)** Calculate the standard deviation of the daily total revenues of this campus food store.

**Answers**

Let $X$ denote the the sales of hamburgers, and let $Y$ denote the sales of donuts.

(a) Normal distribution, because linear combinations of normal variables are normally distributed.

(b) The daily total revenue is $3X + Y$. Thus its mean is $E[3X + Y] = 3E[X] + E[Y] = 3 \times 40 + 100 = 220$.

(c) Since $X$ and $Y$ are independent, the variance of the daily total revenue is $V[3X + Y] = 9V[X] + V[Y] = 9 \times 10 \times 10 + 30 \times 30 = 1800$. The standard deviation is $\sqrt{1800} = 42.43$.

**Marking guidelines**

(a) **Out of 4 points.** 2 points for the answer ("normal distribution"), 2 points for a correct explanation.

(b) **Out of 3 points.** Take off 1 point for a calculation error.

(c) **Out of 3 points.** Take off 1 point for a calculation error.

**Question 6** 10 points

Suppose that the time it takes a kind of chemical process to complete is uniformly distributed over 60 to 180 minutes.

(a) Calculate the probability that a randomly selected process takes more than 144 minutes to complete.

(b) Suppose a random sample of 6 (independent) processes is taken. Calculate the probability that at least 2 processes take more than 144 minutes to complete.

### Answers

(a) Let $C$ denote the time it takes for such a process to complete. Then $C \sim U(60, 180)$ and therefore

$$P(C > 144) = 1 - P(C \leq 144) = 1 - \frac{144 - 60}{180 - 60} = 1 - 0.7 = 0.3.$$

(b) Let $X$ denote the number of processes that take more than 144 minutes to complete. $X$ can be characterized by a binomial distribution, with $\pi = 0.3$ and $n = 6$. Thus, the probability of interest is given by $P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.420 = 0.580$.

### Marking guidelines

(a) **Out of 5 points.** Take off 1 point for a calculation error.

(b) **Out of 5 points.** Take off 2 points if an incorrect value of $\pi$ or $n$ is used but the rest of the calculation is correct. Take off 2 points for a calculation error. Take off 3 points if $1 - P(X \leq 2) = 1 - 0.744 = 0.256$ is computed as the answer.

**Question 7 (for Section A)** 10 points

A company receives a very large shipment of components. It is known that this shipment contains 10% defectives. A random sample of twenty components from this shipment is tested.

(a) Find the mean and variance of the number of defective components in this sample.

(b) What is the probability that this sample contains no more than two defective components?

(c) What is the probability that this sample contains exactly one defective component?

**Answer**

Let $X$ denote the number of defectives in this sample. Then $X$ follows the binomial distribution with $n = 20$ and $\pi = 0.1$.

(a) $E[X] = 20 \times 0.1 = 2$, $V[X] = 20 \times 0.1 \times 0.9 = 1.8$.

(b) The probability of interest is $P(X \leq 2) = 0.677$. One can also compute it as $C_0^{20}(0.1)^0(0.9)^{20} + C_1^{20}(0.1)^1(0.9)^{19} + C_2^{20}(0.1)^2(0.9)^{18}$.

(c) The probability of interest is $P(X = 1) = P(X \leq 1) - P(X \leq 0) = 0.392 - 0.122 = 0.270$. One can also compute it as $C_1^{20}(0.1)^1(0.9)^{19}$.

**Marking guideline**

(a) **Out of 4 points** 2 points for $E[X]$, 2 points for $V[X]$. 3 marks if an incorrect value of $n$ or $\pi$ is used but the rest of the calculation is correct.

(b) **Out of 3 points** Take off 1 point for a calculation error. 2 marks if an incorrect value of $n$ or $\pi$ is used but the rest of the calculation is correct.

(c) **Out of 3 points** Take off 1 point for a calculation error. 2 marks if an incorrect value of $n$ or $\pi$ is used but the rest of the calculation is correct.

**Question 7 (for Section B)** 10 points

Suppose that fifty-five percent of all students in campus dormitories are dissatisfied with the food service. A random sample of 100 students was taken.

**(a)** What is the probability that more than 45 students in the sample were dissatisfied with the food service? *Use the normal approximation to the binomial distribution <u>without</u> continuity correction.*

**(b)** What is the probability that the number of students who were dissatisfied with the food service was between 50 and 60 (inclusive)? *Use the normal approximation to the binomial distribution <u>with</u> continuity correction.*

### Answers

Let $X$ = the number of students dissatisfied with the food service. $X$ follows the binomial distribution with $n = 100$ and $\pi = 0.55$.

We have $E[X] = n\pi = 55$ and $V[X] = n\pi(1 - \pi) = 24.75$.

(a) The probability of interest is $P(X > 45)$. Using the normal approximation without continuity correction

$$P(X > 45) \cong P\left(Z > \frac{45 - 55}{\sqrt{24.75}}\right) = P(Z > -2.01) = P(Z < 2.01)$$

$$= F_z(2.01) = 0.9778$$

where the probability is taken from the normal distribution probability table.

*Aside: the exact probability computed using PHStat is 0.9716.*

(b) The probability of interest is $P(50 \leq X \leq 60)$. Using the normal approximation with continuity correction the probability of interest is replaced by $P(49.5 \leq X \leq 60.5)$. The we calculate the probability using the normal approximation

$$P(49.5 \leq X \leq 60.5) \cong P\left(\frac{49.5 - 55}{\sqrt{24.75}} < Z < \frac{60.5 - 55}{\sqrt{24.75}}\right)$$

$$= P(-1.11 < Z < 1.11) = P(Z < 1.11) - P(Z < -1.11) = F_z(1.11) - F_z(-1.11)$$

$$= F_z(1.11) - (1 - F_z(1.11)) = 0.8665 - (1 - 0.8665) = 0.733$$

*Aside: the exact probability computed using PHStat is 0.73117.*

**Marking guidelines** The 10 points are allocated as follows:

1. One point for realizing that $n = 100$. A student not explicitly writing out $n = 100$ but using $n = 100$ in his/her calculation get full marks (1pt).

2. One point for realizing that $\pi = 0.55$. A student not explicitly writing out $\pi = 0.55$ but using $\pi = 0.55$ in his/her calculation get full marks (1pt).

3. One point for calculating $E[X] = n\pi = 100 \times 0.55 = 55$. A student who do not write out $E[X] = n\pi$ but who uses $E[X] = 55$ in his/her calculations get full marks (1pt). Give 0.5pt to a student who gets the formula $E[X] = n\pi$ slightly wrong or who makes a calculation error.

4. One point for calculating $V[X] = n\pi(1-\pi) = 100 \times 0.55 \times 0.45 = 24.75$ A student who do not write out $V[X] = n\pi(1 - \pi)$ but who uses $V[X] = 24.75$ in his/her calculations get full marks (1pt). Give 0.5pt to a student who gets the formula $V[X] = n\pi(1 - \pi)$ slightly wrong or who makes a calculation error.

5. In part (a), one point for realizing that the probability asked is $P(X > 45)$ which is equivalent to $1 - P(X \leq 45)$.

6. In part (a), one point for the steps leading from $P(X > 45)$ to $P(Z > -2.01)$. Take off 0.5pt for a calculation error. A student who gets $P(Z > -2.01)$ without showing intermediate steps gets 0.5pt.

7. In part (a), 0.5 point for getting the numerical answer 0.9778 (or a value close to it given rounding). Give no marks (out of 0.5) to a student who gets the wrong numerical answer. *Note: if a student gets $P(Z > -2.01)$ wrong, you have to check whether is numerical answer is consistent with the probability statement derived.*

8. In part (b), one point for realizing that the probability asked is $P(50 \leq X \leq 60)$. *See note in the next item.*

9. In part (b), one point for approximating the probability as $P(49.5 \leq X \leq 60.5)$. *Note: A student who writes only $P(49.5 \leq X \leq 60.5)$ as the probability of interest gets 1pt for item 9 and 1pt for item 8.*

10. In part (b), one point for the steps leading from $P(49.5 \leq X \leq 60.5)$ to $P(-1.11 < Z < 1.11)$. Take off 0.5pt for a calculation error. A student who gets $P(-1.11 < Z < 1.11)$ without showing intermediate steps gets 0.5pt.

11. In part (b), 0.5 point for getting the numerical answer 0.733 (or a value close to it given rounding). Give no marks (out of 0.5) to a student who gets the wrong numerical answer. *Note: if a student gets* $P(-1.11 < Z < 1.11)$ *wrong, you have to check whether is numerical answer is consistent with the probability statement derived.*

**Question 8** 13 points

An instructor has found that times spent by students on a particular homework assignment follow a normal distribution with mean 150 minutes and standard deviation 40 minutes.

**(a)** What is the probability that a randomly chosen student spends less than 202 minutes on this assignment?

**(b)** What is the probability that a randomly chosen student spends more than 130 minutes on this assignment?

**(c)** The probability is 0.9 that a randomly chosen student spends less than how many minutes on this assignment?

**(d)** Two students are chosen at random. What is the probability that at least one of them spends more than 202 minutes on this assignment? *Assume that the time spent on the assignment by a student is independent of the time spent by other students.*

**Answers**

(a)

$$P(X < 202) = P\left(Z < \frac{202-150}{40}\right) = P(Z < 1.3) = 0.9032$$

(b)

$$P(X > 130) = P\left(Z < \frac{130-150}{40}\right) = P(Z > -0.5) = P(Z < 0.5) = 0.6915$$

(c) We first identify the value $z^*$ such that $P(Z < z^*) = 0.9$. From the normal cumulative probability table we get $P(Z < 1.28) = 0.9$ so $z^*$=1.28. Second we calculate $x^*$ using the fact that $X = \sigma Z + \mu$ which applied to the current problem yields $x^* = 40 \times 1.28 + 150 = 201.2$. Therefore, the probability is 0.9 that a randomly chosen student spends less than 201.2 minutes on this assignment.

(d) Let $Y$ = number of students spending more than 202 minutes on the assignment. $Y$ follows the binomial distribution with $n = 2$ and $\pi = 1 - 0.9032 \cong 0.1$ (0.9032 was calculated in (a)). The probability of interest is

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - 0.81 = 0.19$$

where $P(Y = 0) = 0.81$ comes from the binomial probability table.

**Marking guidelines**

*Note that for continuous random variables, probability statements can be written with weak ($\leq$, $\geq$) or strict ($<$, $>$) inequality signs without changes in the answer. So don't take off marks for that.*

(a) **Out of 4 points**

1. Give one point for realizing that the probability asked is $P(X < 202)$ which is equivalent to $1 - P(X > 202)$.
2. Give 2 points for the steps leading from $P(X < 202)$ to $P(Z < 1.3)$. Take off 0.5pt for a calculation error. A student who gets $P(Z < 1.3)$ without showing intermediate steps gets 1pt only.
3. Give 1 point for getting the numerical answer 0.9032 (or a value close to it given rounding). Give no marks (out of 1) to a student who gets the wrong numerical answer. *Note: if a student gets $P(Z < 1.3)$ wrong, you have to check whether is numerical answer is consistent with the probability statement derived.*

(b) **Out of 4 points**

1. Give one point for realizing that the probability asked is $P(X > 130)$ which is equivalent to $1 - P(X < 130)$.
2. Give 2 points for the steps leading from $P(X > 130)$ to $P(Z > -0.5)$. Take off 0.5pt for a calculation error. A student who gets $P(Z > -0.5)$ without showing intermediate steps gets 1pt only.
3. Give 1 point for getting the numerical answer 0.6915 (or a value close to it given rounding). Give no marks (out of 1) to a student who gets the wrong numerical answer. *Note: if a student gets $P(Z > -0.5)$ wrong, you have to check whether is numerical answer is consistent with the probability statement derived.*

(c) **Out of 3 points**

1. Two points for the steps leading to $z^* = 1.28$. Give only 1pt for a student stating $z^* = 1.28$ without intermediate steps. Give part marks when appropriate.
2. One point for going from the critical value $z^*$ to the critical value $x^* = 201.2$. Give only 1pt for a student stating $x^* = 201.2$ without intermediate steps. Give part marks when appropriate.

(d) **Out of 2 points**

1. Give 2 points if an incorrect value of $\pi$ is used but the answer is correct otherwise.

2. One point for using $n = 2$. Zero otherwise
3. One point for computing the correct probability. Give part marks.

**Question 9** 15 points

Suppose you are head coach of a coed softball team. After eliminating those who should not have showed up to training camp you are left with 16 equally good players. 10 of them are males and 6 of them are females.

(a) You are now faced with the task of selecting 10 players for the starting line-up of the next softball game. When selecting your starting line-up you have to comply to the rule stating that you must include at least 5 female players in the line-up. How many different starting line-up selections are possible?

(b) Julie (a female) is very motivated to beat the team they play next and would very much like to be part of the starting line-up. What is the probability that she will be part of the starting line-up? *(Recall that you must include at least 5 female players in the line-up.)*

**Answers**

(a) First, we figure out how many arrangements of 10 players there are when 6 females are included. This amounts to finding how many ways there are to select 4 male players from a set of 10 since all female players are included. The answer is $C_6^6 \times C_4^{10} = C_4^{10}$.

Second, we have to find the number of arrangements with 5 male and 5 female players. So we select 5 females from a set of 6 ($C_5^6$ possibilities) that we pair with five males from a set of 10 ($C_5^{10}$) for a total of $C_5^6 \times C_5^{10}$ arrangements with 5 males and 5 females.

Third, we add up the number of arrangements [4 males, 6 females] and the number of arrangements [5 males, 5 females]:

$$C_4^{10} + C_5^6 \times C_5^{10} = \frac{10!}{6!\ 4!} + \frac{6!}{5!\ 1!} \times \frac{10!}{5!\ 5!} = 210 + 6 \times 252 = 1,722$$

(b) We figure out the probability by using the classical definition of probabilities: number of favorable outcomes divided by the total number of possible outcomes. In part (a), we calculated the number of possible line-ups where at least 5 females player are selected. This represents the total number of possible outcomes. What is the number of favorable outcomes? That is, what is the number of line-ups including Julie? Again we have to consider arrangements of 10 players with either 5 or 6 female players.

First, let's count the number of arrangements including 5 female players (Julie plus 4 other) and 5 male players. Given that we are interested in line-ups including Julie, we have to select 4 females players from a set of 5 (Julie is already included) and select 5 male players form a set of 10. Algebraically this is $C_4^5 \times C_5^{10}$.

Second, let's count the number of arrangements including 6 female players (Julie plus 5 other) and 4 male players. Given that we are interested in line-ups including Julie, we have to select 5 female players from a set of 5 (Julie is already included) and select 4 male players form a set of 10. Algebraically this is $C_5^5 \times C_4^{10}$ or $C_6^6 \times C_4^{10}$ since all females are playing.

Third, we calculate the probability. The number of favorable outcomes is

$$C_4^5 \times C_5^{10} + C_5^5 \times C_4^{10} = \frac{5!}{4!\ 1!} \times \frac{10!}{5!\ 5!} + \frac{5!}{5!\ 0!} \times \frac{10!}{6!\ 4!} = 5 \times 252 + 210 = 1470.$$

Therefore

$$P(\text{Julie plays the next game}) = \frac{1,470}{1,722} = 0.8537.$$

## Marking guidelines

(a) **Out of 8 points**

1. Give only 2 points for "1,722" if no intermediate calculations are provided.
2. Give 5 points if a student correctly calculated the number of line-ups including 5 female players OR the number of line-ups including 6 female players but NOT both.
3. Take off a maximum of 2 points for calculation errors.
4. Take off 3 points if a student works with permutations rather than combinations.
5. Give or remove part marks as you deem appropriate.

(b) **Out of 7 points**

1. Two points for calculating the number of line-ups with 5 females (Julie plus 4 others) and 5 males. Give part marks.

2. Two points for calculating the number of line-ups with 6 females (Julie plus 5 others) and 4 males. Give part marks.
3. Three points for using the classical definition of probability to calculate the required probability. Do not take off marks if they plug in wrong values for the the number of favorable cases and/or the total number of possible cases.