**Queen's University**

**Faculty of Arts and Sciences**

**Department of Economics**

**Winter 2007**

**Economics 250 A and B: Introduction to Statistics**

**Term Test 1**

February 7th, 2007

**Instructions:**

- Read questions and instructions carefully.
- The last page of this exam paper displays several formulas.
- Nonprogrammable calculators are permitted.
- Answers are to be written in the examination booklet.
- Do not hand in the question sheet.
- Unless indicated otherwise, write enough intermediate steps to show how you got to your answers. Most of the marks will be awarded to intermediates steps showing how you got your answers.
- Number of questions: 9. **Answer all questions**.
- Number of points: 100.
- Time allotted: 120minutes (6pm-8pm).

**Question 1** 10 points
Here is a sample of 1-year returns for nine exchange-traded funds (ETFs):

| ETF name | Return (%) |
|---|---|
| S+P 500 Index Fund | 11 |
| European Monetary Union Index Fund | 24 |
| Canada Index Fund | 12 |
| Canadian Bond Index Fund | 4 |
| UK Index Fund | 18 |
| Emerging Markets Index Fund | 16 |
| Latin America Index Fund | 21 |
| China Index Fund | 30 |
| Japan Index Fund | 12 |

**(a)** The data above are of the continuous numerical type. What is their measurement type: nominal scale, ordinal scale, interval scale, or ratio scale? *No explanations necessary.*

**(b)** Calculate the sample mean, the median, and the mode (carry only two significant digits).

**(c)** Draw a box plot.

**(d)** Take an investor with a portfolio including only the Canadian Index Fund and the Canadian Bond Index Fund. Did her portfolio beat the median fund return over the last year?

**Answers**
**(a)** ratio scale.

**(b)** The sample mean is

$$\bar{X} = \frac{1}{9}\sum_{i=1}^{9} X_i = \frac{148}{9} = 16.44$$

The mode is 12 since 12 is the only value appearing more than once.
To calculate the median we first have to sort the data. The sorted data are

$$4, 11, 12, 12, 16, 18, 21, 24, 30$$

The median is the $0.5 \times (9+1) = 5$th ordered observation which here is equal to 16.

**(c)** To draw the box plot we need the five numbers summary: minimum, maximum, $Q_1$, $Q_2$, $Q_3$. Referring to the sorted data above we immediately get $Min = 4$, $Max = 30$. We know that $Q_2 = Median = 16$. $Q_1$ is the $0.25\times(9+1) = 2.5$th ordered observation. So we average the second and third ordered observation: $Q_1 = (11 + 12)/2 = 11.5$. $Q_3$ is the $0.75 \times (9 + 1) = 7.5$th ordered observation. So we average the seventh and eight ordered observation: $Q_3 = (21 + 24)/2 = 22.5$.

**(d)** The Canadian funds returned 4% and 12% while the median return is 16%. Given that $4 < 16$ and $12 < 16$, it is impossible for her portfolio to beat the median fund return.

**Marking Guidelines**

**(a) Out of 1 point.** 1 point for a correct answer. 0.5pt for "interval scale." No points otherwise.

**(b) Out of 4 points.**

1. 2pts for $\bar{X}$. Give only 1pt if the answer is simply 16.44 without intermediate steps. Give part marks when you deem necessary.
2. 1pt for the mode. The numerical answer on its own is sufficient to get full marks. If the answer is incorrect but there is some explanation of what the mode is, give some part marks.
3. 1pt for the median. Give only 0.5pt if the answer is simply 16 without intermediate steps. Give part marks when you deem necessary.

**(c) Out of 4 points.**

1. 0.5pt for the minimum. The numerical answer on its own is sufficient to get full marks.
2. 0.5pt for the maximum. The numerical answer on its own is sufficient to get full marks.
3. 1pt for $Q_1$. Give only 0.5pt if the answer is simply 11.5 without intermediate steps. Give part marks when you deem necessary.
4. 1pt for $Q_3$. Give only 0.5pt if the answer is simply 22.5 without intermediate steps. Give part marks when you deem necessary.
5. 1 point for the box-plot.

**(d) Out of 1 point.** Give 0.5pts for "no" without explanations. Give part marks when you deem necessary.

**Question 2** 10 points
A sample of 20 financial analysts was asked to provide forecasts of earnings per share of a corporation for the next year. The results are summarized in the frequency distribution below.

| **Class boundaries** | **Class Frequency** |
|---|---|
| greater or equal to 9.95, less than 10.45 | 2 |
| greater or equal to 10.45, less than 10.95 | 8 |
| greater or equal to 10.95, less than 11.45 | 6 |
| greater or equal to 11.45, less than 11.95 | 3 |
| greater or equal to 11.95, less than 12.45 | 1 |

**(a)** Represent the frequency distribution graphically using an histogram.

**(b)** Calculate the sample mean.

**(c)** Calculate the mode.

**Answers**
**(a)**

**(b)** Let $m_i$ denote the class mark (mid-point) and $f_i$ the class frequency for class $i$. Then

$$\bar{X} = \frac{1}{20}\sum_{i=1}^{5} m_i f_i = (10.2\times2+10.7\times8+11.2\times6+11.7\times3+12.2\times1)/20 = \frac{220.5}{20} = 11.025$$

**(c)** With grouped data, the mode is the mid-point of the class with highest frequency. Here it is $m_2 = 10.7$ (mid-point of the second class).

**Marking Guidelines**

**(a) Out of 5 points.**

1. Deduct 0.5pt when the horizontal axis has no title
2. Deduct 0.5pt when the vertical axis has no title
3. Deduct 1pt when they use relative frequencies rather then frequencies
4. Deduct 0.5pt if the vertical bars of the histogram do not touch
5. Deduct additional part marks for other errors

**(b) Out of 3 points.**

1. To get full marks it is sufficient to have the algebraic formula and the correct numerical answer **or** the application of the algebraic formula (even if the formula per se is not written down) and the correct numerical answer.

2. Give no marks to those who (attempt to) compute the median.

3. Give part marks when you deem necessary.

**(c) Out of 2 points.**

1. A correct numerical answer without intermediate steps or explanation is worth 1pt only.

2. When a student cannot compute the mode correctly but the answer gives some indication the student knows what the mode represents give 1pt.

**Question 3** 10 points
An investigator has a computer file showing family incomes for 1,000 subjects in a certain study. These range from $5,800 a year to $98,600 a year. By accident, the highest income in the file gets changed to $986,000.

**(a)** Does this affect the mean? If so, by how much?

**(b)** Does this affect the median? If so, by how much?

**Answers**
**(a)** Let $\bar{X}_{old}$ and $\bar{X}_{new}$ denote the mean before and after the change. Then, because the sum of the incomes of all families increases by 986000-98600,

$$\bar{X}_{new} * 1000 = \bar{X}_{old} * 1000 + (986000 - 98600).$$

Thus, the mean increases by $(986000 - 98600)/1000 = 887.4$. Or, simply, because there are 1000 subjects, the mean increases by $(986000 - 98600)/1000 = 887.4$.

**(b)** The median is the average of the 500th and 501th ordered observations. Since increasing the value of the 1st largest observation does not change the order of the other observations, the median is not affected.

**Marking Guidelines**

**(a) Out of 5 points.** Give 3 points if the answer is "yes" without a correct explanation and/or a correct answer (887.4).

**(b) Out of 5 points.** Give 3 points if the answer is "no" without a correct explanation. "The number of observations ($n$) did not change." does not constitute a correct explanation. Give 4 points if the answer is "no" with explanations such as "The median is the middle number." or "The median is not sensitive to outliers."

**Question 4** 5 points
Suppose that a farmer asks one of his helper to record the temperature (denoted by $F$) at noon each day during the month of June and to report on the mean and variance. The helper reports that the average temperature in June was $\mu_F = 77$ Fahrenheit with a variance $\sigma_F^2 = 100$. The farmer wished the helper used degrees Celsius rather than Fahrenheit. Note that the temperature measured in Celsius (denoted $C$) is related to the temperature measured in Fahrenheit as follows

$$F_i = 32 + \frac{9}{5}C_i$$

**(a)** Help the farmer by finding out $\mu_C$ and $\sigma_C^2$.

**(b)** Let $Z_i$ denote the temperature on day $i$ after it has been standardized. If on day 5 the temperature was below average, what is the sign of $Z_5$?

**(c)** If the median temperature in Celsius is 24, what is the median temperature when measured in Fahrenheit?

**Answers**
**(a)** Given the relationship between $F$ and $C$, we infer that

$$\mu_F = 32 + (9/5)\mu_C, \qquad \text{and that } \sigma_F^2 = \left(\frac{9}{5}\right)^2 \sigma_C^2.$$

We are told that $\mu_F = 77$ so we can easily solve for $\mu_C$ as follows

$$77 = 32 + (9/5)\mu_C \Rightarrow (77 - 32) \times \frac{5}{9} = \mu_C \Rightarrow \mu_C = 25$$

We are told that $\sigma_F^2 = 100$ so we can easily solve for $\sigma_C^2$ as follows

$$100 = \left(\frac{9}{5}\right)^2 \sigma_C^2 \Rightarrow \sigma_C^2 = 100 \times \left(\frac{5}{9}\right)^2 = 30.86$$

**(b)** Whether we work with temperature in Celsius or Fahrenheit has no bearing on the answer. For concreteness, suppose we work with Fahrenheits. Then

$$Z_5 = \frac{F_5 - \mu_F}{\sigma_F}.$$

The temperature below average on day 5 means that $F_5 < \mu_F$ which means $Z_5 < 0$ since $F_5 - \mu_F < 0$.

**(c)** Observe that the linear transformation we work with in this question does not affect the ordering of the sorted observations. That is, if 20 is the fifth sorted observation in Celsius, its equivalent on the Fahrenheit scale will also be the fifth sorted observation. Therefore, to calculate the median in Fahrenheit, simply plug the median in Celsius in the linear transformation equation. That is, the median temperature when measured in Fahrenheit is

$$Med_F = 32 + (9/5) \times 24 = 75.2.$$

**Marking Guidelines**

**(a) Out of 3 points.**

1. 1pt for the formula linking $\mu_C$ and $\mu_F$
2. 1pt for the formula linking $\sigma_C$ and $\sigma_F$
3. one point for the algebra.

**(b) Out of 1 point.**

1. Give 0.5pt for the answer "negative" and 0.5pt for an explanation. Give part marks if explanation is incomplete or unclear.
2. A numerical example is sufficient as a justification/explanation.

**(c) Out of 1 point.**

1. Give 0.8pt for plugging 24 in the equation.
2. Give 0.2pt for explaining that the transformation does not change the ordering

**Question 5** 5 points

A coin is tossed six times. Two possible sequences of results are

(i) H T T H T H (ii) H H H H H H

(The coin must land H or T in the order given; H = heads, T = tails.) Which of the following is correct? *No explanations necessary.*

**(a)** Sequence (i) is more likely.

**(b)** Sequence (ii) is more likely.

**(c)** Both sequences are equally likely.

**Answers**

(c) is correct. Both sequences occur with probability $(1/2)^6$ given the independence of the tosses.

**Question 6** 10 points

A mail-order firm considers three possible foul-ups in filling an order:

**A**. The wrong item is sent.
**B**. The item is lost in transit.
**C**. The item is damaged in transit.

Assume that event $A$ is independent of both $B$ and $C$ and that events $B$ and $C$ are mutually exclusive. The individual event probabilities are $P(A) = 0.02$, $P(B) = 0.01$, and $P(C) = 0.04$. Find the probability that at least one of these foul-ups occurs for a randomly chosen order. [Hint: $P(A \cup B \cup C) = P(A \cup (B \cup C))$].

**Answers**

We want to compute $P(A \cup B \cup C) = P(A \cup (B \cup C))$. First, note that

$$P(A \cup (B \cup C)) = P(A) + P(B \cup C) - P(A \cap (B \cup C))$$

Since $A$ is independent of both $B$ and $C$, $P(A \cap (B \cup C)) = P(A) \times P(B \cup C)$. Since $B$ and $C$ are mutually exclusive, $P(B \cup C) = P(B) + P(C)$. Therefore,

$$P(A \cup B \cup C) = 0.02 + (0.01 + 0.04) - 0.02 \times (0.01 + 0.04) = 0.069.$$

**Marking Guidelines**

1. 3 points for $P(B \cup C) = P(B) + P(C) = 0.01 + 0.04 = 0.05$.
2. 3 points for applying the addition rule of probabilities properly: $P(A \cup (B \cup C)) = P(A) + P(B \cup C) - P(A \cap (B \cup C))$.
3. 3 points for applying the multiplicative law of probability properly: $P(A \cap (B \cup C)) = P(A) \times P(B \cup C)$.
4. 1 point for computing the answer (0.069) correctly.

**Question 7** 20 points

A large corporation organized a ballot for all its workers on a new bonus plan. It was found that 60% of all night-shift workers favored the plan and that 45% of all female workers favored the plan. Also, 40% of all employees are night-shift workers and 35% of all employees are women. Finally, 15% of the night-shift workers are women.

**(a)** What is the probability that a randomly chosen employee is a woman in favor of the plan?

**(b)** What is the probability that a randomly chosen employee is either a woman or a night-shift worker (or both)?

**(c)** Is employee gender independent of whether the night-shift is worked?

**(d)** If 50% of all male employees favor the plan, what is the probability that a randomly chosen employee both does not work the night shift and does not favor the plan. [Hint: As an intermediate step in finding the answer, you have to calculate the probability of working the night shift or being in favor of the plan (or both).]

**Answers**

Let's first define some events.

$Fav$: the employee is in favor of the plan

$Ns$: the employee works night shifts

$W$: the employee is a woman

Correspondingly we have: $P(Fav|Ns) = 0.6$, $P(Fav|W) = 0.45$, $P(Ns) = 0.4$, $P(W) = 0.35$, and $P(W|Ns) = 0.15$.

**(a)**

$$P(W \cap Fav) = P(Fav|W)P(W) = 0.45 \times 0.35 = 0.1575$$

**(b)**

$$P(W \cup Ns) = P(W) + P(N) - P(W \cap Ns) = P(W) + P(N) - P(W|Ns)P(Ns)$$
$$= 0.34 + 0.4 - 0.15 \times 0.4 = 0.69.$$

**(c)** If they are independent, then $P(W|Ns) = P(W)$ which is not the case here since $P(W|Ns) = 0.15$ and $P(W) = 0.35$. Equivalently, independence requires that $P(W \cap Ns) = P(W)P(Ns)$ which is also violated since $P(W)P(Ns) = 0.35 * 0.4 = 0.14$ and $P(W \cap Ns) = P(W|Ns)P(Ns) = 0.15 \times 0.4 = 0.06$.

**(d)** We look for $P(\bar{N}s \cap F\bar{a}v)$. Note that events $Ns$ and $\bar{N}s$ are mutually exclusive and collectively exhaustive. The same can be said about events $Fav$ and $\bar{F}$. So there are only four joint outcomes for night/day shift and in favor/not in favor: $\{(Fav \cap Ns), (F\bar{a}v \cap Ns), (Fav \cap \bar{N}s), (F\bar{a}v \cap \bar{N}s)\}$. Or if you prefer, you can think of a joint probability table like

| | $Fav$ | $F\bar{a}v$ |
|---|---|---|
| $Ns$ | $P(Ns \cap Fav)$ | $P(Ns \cap F\bar{a}v)$ |
| $\bar{N}s$ | $P(\bar{N}s \cap Fav)$ | $P(\bar{N}s \cap F\bar{a}v)$ |

The discussion above makes clear that $P(\bar{N}s \cap F\bar{a}v) = 1 - P(Ns \cup Fav)$. So using the additive probability rule we get

$$P(\bar{N}s \cap F\bar{a}v) = 1 - P(Ns \cup Fav) = 1 - [P(Ns) + P(Fav) - P(NS \cap Fav)]$$

Using the fact that

$$P(Fav) = P(Fav \cap W) + P(Fav \cap \bar{W})$$
$$= P(Fav|W)P(W) + P(Fav|\bar{W})P(\bar{W})$$
$$= 0.45 \times 0.35 + 0.5 \times 0.65 = 0.4875.$$

Also

$$P(Ns \cap Fav) = P(F|Ns)P(Ns) = 0.6 \times 0.4 = 0.24$$

Therefore

$$P(\bar{N}s \cap F\bar{a}v) = 1 - [0.4 + 0.4875 - 0.24] = 0.3525$$

### Marking Guidelines

**(a)** 5 points

1. 2pts for clearly establishing the notation for the events "employee in favor of the plan," "employee works night shifts," and "employee is a woman." Give part marks if that is not done right. Alternatively, if a student write out the events explicitly (like $P(Woman|in\,favor\,of\,the\,plan)$ that's fine. Remove part marks when the notation is not clearly define though.

2. 1pt for a numerical answer without intermediate steps.

3. 2pts for the intermediate step(s) (give part marks).

**(b)** 5 points

1. 2pts for a numerical answer without intermediate steps.

2. 3pts for the intermediate steps.

3. Take off 1.5pt when they do not subtract the probability of the intersection.

4. Take off 1pt when they calculate the probability of the intersection as the product of marginal probabilities (implicitly assuming independence).

5. Take off part marks for other errors.

**(c)** 5 points

1. 1pt for stating that checking for independence entails verifying whether: $P(W|Ns) = P(W)$. Using condition $P(W \cap Ns) = P(W)P(Ns)$ is fine too.

2. 2pts for using a numerical example to show that one of these condition is violated.

3. 2pt for concluding that the events are not independent. Therefore, a student simply writing "not independent" gets only 2pts.

4. Give/take off part marks where appropriate.

**(d)** 5 points

1. Give 3pts to students who followed the Hint and correctly calculated $P(Ns \cup Fav)$ but could no go any further. Deduct no more than 1.5pts when they make mistakes in calculating $P(Ns \cup Fav)$.

2. When a student realizes that $P(\bar{Ns} \cap \bar{Fav}) = 1 - P(Ns \cup Fav)$, give 4pts even if the calculations are messed up.

3. This is a difficult questions so fell free to give part marks for a good effort.

**Question 8** 10 points
Subscribers to a newspaper were asked whether they regularly, occasionally, or never read the business section, and also whether they had traded common stocks (or shares in a mutual fund) over the last year. The associated joint probability table is below:

| | **Read the business section** | | |
|---|---|---|---|
| *Traded stocks* | **Regularly** | **Occasionally** | **Never** |
| *Yes* | 0.22 | 0.10 | 0.05 |
| *No* | 0.15 | 0.30 | 0.18 |

**(a)** What is the probability that a randomly chosen subscriber regularly reads the business section and has traded stocks (or shares of mutual funds) over the last year?

**(b)** What is the probability that a randomly chosen subscriber has traded stocks (or shares of mutual funds) over the last year?

**(c)** What is the probability that a subscriber who never reads the business section has traded stocks (or shares of mutual funds) over the last year?

**(d)** What is the probability that a subscriber who has traded stocks (or shares of mutual funds) over the last year reads the business section regularly or occasionally?

**Answers**
Let's first define some events.

R: subscriber reads the business section regularly
O: subscriber reads the business section occasionally
N: subscriber reads the business section occasionally
T: subscriber has traded stocks (or mutual funds) over the last year

**(a)** We get the answer directly from the joint probability table: $P(R \cap T) = 0.22$

**(b)** Using the fact that events $R, O, N$ are mutually exclusive and collectively exhaustive we have

$$P(T) = P(T \cap R) + P(T \cap O) + P(T \cap N) = 0.22 + 0.10 + .05 = 0.37$$

**(c)**

$$P(T|N) = \frac{P(T \cap N)}{P(N)} = \frac{0.05}{0.18 + 0.05} = 0.217$$

**(d)** Given that events $R$ and $O$ are mutually exclusive we have

$$P(R \cup O|T) = P(R|T) + P(O|T)$$

$$\frac{P(R \cap T)}{P(T)} + \frac{P(O \cap T)}{P(T)} = \frac{0.22}{0.37} + \frac{0.10}{0.37} = 0.86$$

**Marking Guidelines**

**(a)** 3 points

1. 1.5pts for clearly establishing the notation for the 4 events. Give part marks if that is not done right. Alternatively, if a student write out the events explicitly (like $P(\textit{trade stocks} \cap \textit{readregularly})$ that's fine. Remove part marks when the notation is not clearly define though.
2. It is sufficient to give only the numerical answer 0.22 to get the remaining 1.5 points. Give only 0.5pts when they compute a conditional probability.

**(b)** 2 points

1. Writing $P(T) = 0.22 + 0.10 + .05 = 0.37$ is sufficient to get full marks.
2. Give 1pt for 0.37 without intermediate steps.

**(c)** 3 points

1. 1.5pt for 0.217 without intermediate steps.
2. 1.5pt for applying the conditional probability rule properly.
3. give 1pt when they give the joint probability or never reading and trading stocks.

**(d)** 2 points

1. Give 2pts to students who get the right answer (with intermediate steps) but who do not indicate that mutual exclusiveness of $R$ and $O$ justifies simply adding the two conditional probabilities.
2. Give 1pt to those who get 0.86 without intermediate steps.
3. Give 0.5pt to those who simply sum the joint probabilities $P(T \cap R)$ and $P(T \cap O)$ and get 0.32.
4. As usual, give/take off part marks when you deem appropriate.

**Question 9** 20 points
Mr. Dupont is a professional wine taster. When given a French wine, he will identify it with probability 0.9 correctly as French, and will mistake it for a Californian wine with probability 0.1. When given a Californian wine, he will identify it with probability 0.8 correctly as Californian, and will mistake it for a French wine with probability 0.2. Suppose that Mr. Dupont is given ten unlabelled glasses of wine, four with French and six with Californian wines. He randomly picks a glass, tries the wine, and solemnly says: "French".

**(a)** What is the probability that the wine he tasted was Californian?

**(b)** Putting aside this glass, Mr. Dupont takes another glass and declares it "French". What is the probability he is wrong again? [Hint: there are two possibilities: (i) he picked the Californian first, or (ii) he picked the French first.]

**Answers**
**(a)** Let $F$ denote the event that the wine tasted was French, let $C$ be the event that the wine tasted was Californian, let $FD$ be the event that the wine tasted was declared French by Mr. Dupont, and let $CD$ be the event that the wine tasted was declared Californian by Mr. Dupont. We are given the following:

$$P(F) = 0.4, \quad P(C) = 0.6,$$
$$P(FD|F) = 0.9, \quad P(CD|F) = 0.1, \quad P(FD|C) = 0.2, \quad P(CD|C) = 0.8.$$

We want to calculate $P(C|FD) = P(C \cap FD)/P(FD)$. Since

$$\begin{aligned} P(FD) &= P(FD \cap F) + P(FD \cap C) \\ &= P(FD|F)P(F) + P(FD|C)P(C) = 0.9 \times 0.4 + 0.2 \times 0.6 = 0.48, \end{aligned}$$

it follows that $P(C|FD) = 0.12/0.48 = 0.25$.

**(b)** We first have to consider 2 cases that (i) he picked a Californian first or (ii) he picked a French first. The probability of interest is given by the union probability of (i) (pick a Californian wine first and declare the next pick a French when it is really a Californian) and (ii) (pick a French wine first and declare the next pick a French when it is really a Californian). Note that the intersection is zero (can't pick both a French and Californian first).

Let $P(C|FD)(i)$ denote the probability that a glass is Californian when it is declared to be French under scenario (i), and define $P(C|FD)(ii)$ similarly. Then the probability of interest is

$$P((i) \cup (ii)) = P(i) + P(ii) = P(C|FD)(i)P(C) + P(C|FD)(ii)P(F).$$

1. $P(C|FD)(i)$. Californian picked first, leaving 4 French and 5 Californian.

$$P(F) = 4/9, \quad P(C) = 5/9.$$

Since

$$P(FD) = P(FD|F)P(F)+P(FD|C)P(C) = 0.9\times 4/9+0.2\times 5/9 = 0.511,$$

it follows that

$$P(C|FD)(i) = P(FD|C)P(C)/P(FD) = 0.111/0.511 = 0.217.$$

2. $P(C|FD)(ii)$. French picked first, leaving 3 French and 6 Californian.

$$P(F) = 1/3, \quad P(C) = 2/3.$$

Since

$$P(FD) = P(FD|F)P(F)+P(FD|C)P(C) = 0.9\times 1/3+0.2\times 0.67 = 0.434,$$

it follows that

$$P(C|FD)(ii) = P(FD|C)P(C)/P(FD) = 0.134/0.434 = 0.309.$$

3.

$$\begin{aligned} P((i)\cup(ii)) &= P(C)P(C|FD)(i) + P(F)P(C|FD)(ii) \\ &= 0.6\times 0.217 + 0.4\times 0.309 = 0.254. \end{aligned}$$

**Marking Guidelines**

**(a) Out of 10 points.**

1. 5 points for the first step: $P(C|FD) = P(C\cap FD)/P(FD)$, or $P(C|FD) = P(FD|C)P(C)/P(FD)$. 4 points are given if a student defines the event $A$ as "Dupont is incorrect" or "Dupont is correct" and writes $P(C|A) = P(C\cap A)/P(A)$, or $P(C|A) = P(A|C)P(C)/P(A)$.

2. 3 points for the second step: $P(FD) = P(FD\cap F) + P(FD\cap C) = P(FD|F)P(F) + P(FD|C)P(C)$, conditional on that the first step is done correctly with events $FD$ and $C$.

3. 2 points for computing the answer (0.25) correctly.

**(b) Out of 10 points.**

1. 3 points for "$P(F) = 4/9, P(C) = 5/9$ if he picked a Californian first, and $P(F) = 3/9, P(C) = 6/9$ if he picked a French first."
2. 2 points for $P(C|FD)(i) = P(FD|C)P(C)/P(FD)$.
3. 2 points for $P(C|FD)(ii) = P(FD|C)P(C)/P(FD)$.
4. 2 points for $P((i)\cup(ii)) = P(i)+P(ii) = P(C|FD)(i)P(C)+P(C|FD)(ii)P(F)$.
5. 1 point for computing the answer (0.254) correctly.

# Statistics

**Means**

$$\text{Population} : \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{Sample} : \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

**Variances**

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

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

**Coefficient of variation (CV)**

$$\text{Population} : CV = \frac{\sigma}{\mu} \times 100\% \qquad \text{Sample} : CV = \frac{s}{\bar{X}} \times 100\%$$

# Probability Theory

**Probability Rules**

$$\text{Addition rule} : P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\text{Multiplication rule} : P(A \cap B) = P(B)P(A|B)$$

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

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i)P(A|E_i)$$

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