International Institute for
Technology and Management

March 11 ,2009 **Duration : 90 min**

Unit 04b: Statistics 2 – (Stats 2) **Assignment – 5**

Statistics 2 (half unit)

Candidates should answer all **FOUR** questions: **QUESTION 1** of Section A (40 marks) and all **THREE** questions from Section B (60 marks in total).

Candidates are strongly advised to divide their time accordingly.

A list of formulae is given after the final question on this paper.

New Cambridge Statistical Tables (second edition) are provided.

A calculator may be used when answering questions on this paper and it must comply in all respects with the specification given with your Admission Notice. The make and type of machine must be clearly stated on the front cover of the answer book.

PLEASE TURN OVER

SECTION A

Answer all five parts of question 1 (40 marks in total).

1. (a) For each one of the statements below say whether the statement is true or false explaining your answer.

i. If A and B are independent, then A^C and B^C are independent.

True

It is required to show that: $P(A^c \cap B^c) = P(A^c)\, P(B^c)$

We have $P(A^c \cap B^c) = P(A \cup B)^c$ using Demorgan's

$= 1 - P(A \cup B)$

$= 1 – P(A) - P(B) - P(A \cap B)$

$= 1- P(A) - P(B) - P(A)P(B)$ coz A & B are indep.

$=[1- P(A)] – P(B)[1- P(A)]$

$= [1- P(A)][1- P(B)] = P(A^c)\, P(B^c)$

ii. Let X be a random variable; The variance of $X + 1$ is larger than the variance of X.

False

$Var(X + 1) = Var(X) + Var(1) = Var(X) + 0 = Var(X)$

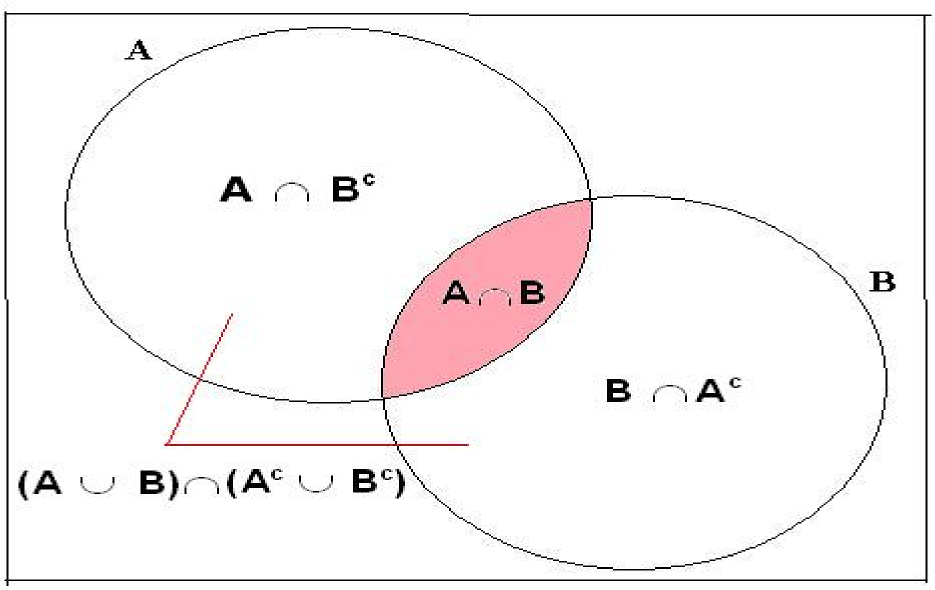
iii. $P((A \cup B) \cap (A^c \cup B^c)) = P(A) + P(B) - 2P(A \cap B)$.

True

$(A \cup B) \cap (A^c \cup B^c) = \{ s \mid s \in A \text{ or } B \textit{ and } s \notin (A \cap B) \}$

$= (A \cap B^c) \cup (B \cap A^c)$

$\Rightarrow P((A \cup B) \cap (A^c \cup B^c)) = P(A \cap B^c) + P(B \cap A^c)$



$A = (A \cap B^c) \cup (A \cap B)$, $(A \cap B^c)$ and $(A \cap B)$ are mutually exclusive

$P(A) = P(A \cap B^c) + P(A \cap B) \Rightarrow \boxed{P(A \cap B^c) = P(A) - P(A \cap B)}$

$B = (B \cap A^c) \cup (A \cap B)$, $(B \cap A^c)$ and $(A \cap B)$ are mutually exclusive

$P(B) = P(B \cap A^c) + P(A \cap B) \Rightarrow \boxed{P(B \cap A^c) = P(B) - P(A \cap B)}$

$P((A \cup B) \cap (A^c \cup B^c)) = P(A \cap B^c) + P(B \cap A^c)$

$= P(A) + P(B) - 2\,P(A \cap B)$

iv. Var (X + Y) = Var (X) + Var (Y) – 2Cov(X,Y)
False
Var (X + Y) = Var (X) + Var (Y) + 2Cov(X,Y)

Proof that $\mathsf{var}(X \pm Y) = \mathsf{var}(X) + \mathsf{var}(Y) \pm 2\mathsf{cov}(X,Y)$

This follows quickly from the algebraic identities $(a+b)^2 = a^2 + b^2 + 2ab$ and $(a-b)^2 = a^2 + b^2 - 2ab$.

From the definition of a variance we have for random variables X and Y:

$$\mathsf{var}(X+Y) = \mathsf{E}\left[(X+Y-\mathsf{E}(X+Y))^2\right] = \mathsf{E}\left[(X-\mathsf{E}(X)+Y-\mathsf{E}(Y))^2\right]$$

Expanding the squared term gives

$$(X-\mathsf{E}(X)+Y-\mathsf{E}(Y))^2 = (X-\mathsf{E}(X))^2+(Y-\mathsf{E}(Y))^2+2(X-\mathsf{E}(X))(Y-\mathsf{E}(Y))$$

Taking expectations of this expression gives the result, namely

$$\begin{aligned}\mathsf{var}(X+Y) &= \mathsf{E}(X-\mathsf{E}(X))^2 + \mathsf{E}(Y-\mathsf{E}(Y))^2 + 2\mathsf{E}(X-\mathsf{E}(X))(Y-\mathsf{E}(Y)) \\ &= \mathsf{var}(X) + \mathsf{var}(Y) + 2\mathsf{cov}(X,Y)\end{aligned}$$

as required. The result for $\mathsf{var}(X-Y)$ follows using the same kind of argument based on expanding $(a-b)^2$.

v. Consider the mutually exclusive and exhaustive events A, B and C. Is it possible to have
$P(A \cup B) = ½$, $P(B \cup C) = ½$ and $P(C \cup A) = 2/3$?
False
$P(A \cup B) = P(A) + P(B) = ½$
$P(B \cup C) = P(B) + P(C) = ½$
$P(C \cup A) = P(C) + P(A) = 2/3$
Adding these three : $2(P(A) + P(B) + P(C)) = 5/3$
$\Rightarrow P(A) + P(B) + P(C) = 5/6$, but A, B and C are mutually exclusive and exhaustive i.e.
$P(A) + P(B) + P(C) = 1$ which is not the case here.

(15 Marks)

(b) A random variable, X, is normally distributed with mean μ and variance σ^2.

Given that $E(X^2) = 5$ and $P(X > 5) = 0.02275$,
evaluate

μ and σ^2.

(5 Marks)

We have σ^2 = E(X^2) – E^2(X) = 5 - μ^2

and $P(X \succ 5) = P\left(\dfrac{X-\mu}{\sigma} \succ \dfrac{5-\mu}{\sigma}\right) = 0.02275 \Rightarrow \dfrac{5-\mu}{\sigma} = 2$

$\Rightarrow \mu = 5 - 2\sigma$ **solving this with σ^2 = 5 - μ^2 , we get**

$\mu = 1$ **and σ^2 = 4**

(c) For the binomial distribution with a probability of 0.25 of success in an individual trial, calculate the probability that, in 50 trials, there are at least 8 successes:

(a) from the normal approximation *without* continuity correction;
(b) from the normal approximation *with* continuity correction.

Compare these results with the exact probability 0.9547 and comment.

(5 Marks)

The probability required is $P(X \geq 8)$.
The Normal approximation uses Normal variable $X' \sim N(12.5, 9.375)$.

$$P(X' \geq 8) = P\left(Z \geq \frac{8-12.5}{\sqrt{9.375}}\right) = P(Z \geq -1.46969) = 0.9292$$

The required probability could have been expressed as $P(X > 7)$. Or any number in [7,8).

$$P(X' > 7) = P\left(Z \geq \frac{7-12.5}{\sqrt{9.375}}\right) = P(Z > -1.79627) = 0.9641$$

The continuity correction procedure would use

$$P(X' > 7{,}5) = P\left(Z \geq \frac{7.5-12.5}{\sqrt{9.375}}\right) = P(Z > -1.63299) = 0.9484$$

Obvious simple comments

(d) At one stage in the manufacture of an article a piston of circular cross-section has to fit into a similarly shaped cylinder. The distributions of diameters of pistons and cylinders are known to be normal with parameters:

Piston diameters:	mean 10.42 cm, standard deviation 0.03 cm.
Cylinder diameters:	mean 10.52 cm, standard deviation 0.04 cm.

If pairs of pistons and cylinders are selected at random for assembly, for what proportion will the piston not fit into the cylinder?

What is the probability that in 100 pairs, selected at random, no pistons will fail to fit?

Calculate an approximation for this probability, using a Poisson distribution, and discuss the appropriateness of using such an approximation.

[Hint : Assume D is the difference in diameter of a piston and a cylinder , then the piston will fit if D > 0]

(5 Marks)

Let D be the difference in diameter of a piston and cylinder. According to the information provided $D \sim N(0.01,(0.05)^2)$

The piston will fit if $\{D > 0\}$

$$P(D > 0) = P(Z > \frac{-0.01}{0.05}) = PZ > -2) = 0.02275$$

Let X denote the number of pairs where the piston does not fit.

$X \sim Bin(100,0.02275)$

The probability that all pistons fit is $P(X = 0) = 0.100131$

The Poisson approximation uses $X' \sim Poisson(2.275)$ which would be regarded as within the appropriate range for this approximation. The probability according to this is $P(X' = 0) = 0.102797$, which appears a reasonable approximation ·

(e) X and Y are discrete random variables that can assume values 0,1 and 2 only.

$p(X = x, Y = y) = A(x + y)$ for some constant A and $x, y \in \{0,1,2\}$

i) Draw up a table to describe the joint distribution of X and Y and find the value of the constant A.

		X		
		0	1	2
Y	0	0	A	2A
	1	A	2A	3A
	2	2A	3A	4A

Since A + 2A + A + 2A + 3A + 2A + 3A + 4A = 1 must have A = 1/18.

ii) Describe the marginal distributions of X and Y.

Marginal distribution for X and similarly for Y

	X	
0	1	2
3A = 1/6	6A = 1/3	9A = 1/2

iii) Give the conditional distribution of $X|Y = 1$ and find $E(X|Y = 1)$.

Distribution of $X|Y = 1$

	$X|Y = 1$	
0	1	2
A/6A = 1/6	2A/6A = 1/3	3A/6A = 1/2

$E(X|Y = 1) = 0x1/6 + 1x1/3 + 2x1/2 = 4/3$.

iv) Are X and Y independent? Give reasons for your answer.

iv) Even though the distributions of X and $X|Y = 1$ are the same, X and Y are not independent. $p(X = 0, Y = 0) = 0$ although $p(X = 0) \neq 0$, $p(Y = 0) \neq 0$.

(10 Marks)

SECTION B

Answer all the THREE questions in this section (60 marks in total).

2. (a) State and prove Baye's theorem. **(5 Marks)**

Refer to Tutoring Sheet 1b , question 6(i.)

(b) A man has two bags, bag A contains five keys and bag B seven. Only one of the twelve keys fits the lock that he is trying to open.

What is the probability that the key which would fit the lock is in the bag A?

The man selects a bag at random, picks out a key from the bag and tries that key in a lock.

What is the probability that the key he has chosen fits the lock?

Suppose the key first chosen does not fit the lock.

What is the probability that the bag chosen

i) is the bag A?

ii) contains the required key?

Should the man take a second key from the bag he first selected or should

$P(key\ in\ \ first\ bag) = \frac{5}{12}$

Define a partition $\{C_i\}$ s.t.

$C_1 =$ key in bag 1 and bag 1 chosen	$\frac{5}{12}\cdot\frac{1}{2} = \frac{5}{24}$
$C_2 =$ key in bag 2 and bag 1 chosen	$\frac{7}{12}\cdot\frac{1}{2} = \frac{7}{24}$
$C_3 =$ key in bag 1 and bag 2 chosen	$\frac{5}{12}\cdot\frac{1}{2} = \frac{5}{24}$
$C_4 =$ key in bag 2 and bag 2 chosen	$\frac{7}{12}\cdot\frac{1}{2} = \frac{7}{24}$

$P(key\ \ fits\,(=F)) = \frac{1}{5}P(C_1) + \frac{1}{7}P(C_4) = \frac{1}{12}$

$$P(bag\ 1 \,|\, F^c) = \frac{P(F^c \,|\, C_1).P(C_1) + P(F^c \,|\, C_2).P(C_2)}{\sum P(F^c \,|\, C_i).P(C_i)}$$

$$P(F^c \,|\, C_1) = \tfrac{4}{5} \quad P(F^c \,|\, C_2) = 1 \quad P(F^c \,|\, C_3) = 1 \quad P(F^c \,|\, C_4) = \tfrac{6}{7}$$

$$P(bag\ 1 \,|\, F^c) = \frac{\frac{4}{5}.\frac{5}{24}. + 1.\frac{7}{24}}{\frac{4}{5}.\frac{5}{24} + 1.\frac{7}{24} + 1.\frac{5}{24} + \frac{6}{7}.\frac{7}{24}} = \frac{1}{2}$$

$$P(right\ bag \,|\, F^c) = \frac{P(F^c \,|\, C_1).P(C_1) + P(F^c \,|\, C_4).P(C_4)}{\sum P(F^c \,|\, C_i).P(C_i)} = \frac{5}{11}$$

(c) Suppose that $P(A) = 2p$, $P(B) = p$, $P(B \mid A) = \frac{1}{2}p$ and $P(A \cup B) = 0.8$.

Evaluate p.

(2 Marks)

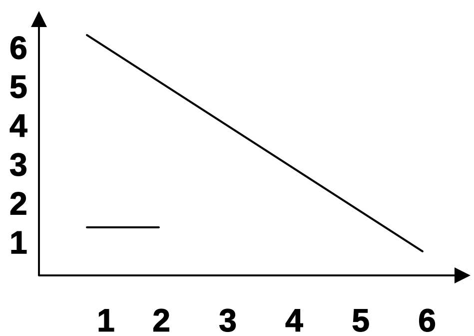
$$P(A \cup B) = P(A) + P(B) - P(B \mid A).P(A) = 2p + p - \tfrac{1}{2}p.2p$$

So $p^2 - 3p + 0.8 = 0 \quad \Rightarrow p = 0.295841$

3. (a) A fair die is thrown twice. Let A be the event that the first throw is less than 3 and B be the event that the sum of the two throws is 7 or 8.
i. Draw a diagram of the sample space and mark the events A and B.
ii. Evaluate P(A), P(B) and P(B|A).

(5 Marks)

i.



ii. P(A) = 2/6 = 1/3 , P(B) = 11/36

P(B|A) = $\dfrac{P(B \cap A)}{P(A)} = \dfrac{3/36}{1/3} = 0.25$

(b) The distribution of a random variable X is as follows

x	-1	0	1
P(X=x)	a	b	a

Calculate Cov(X , X^2) .Are X and X^2 correlated ? (5 Marks)
E(X) = -1 (a) + (0)(b) + (1)(a) = 0
E(X^2) = (1)(a) + (0)(b) + (1)(a) = 2a
E(X^3) = -1(a) + (0)(b) + (1)(a) = 0
Cov(X,X^2) = E(X.X^2) – E(X)E(X^2) = E(X^3) - E(X)E(X^2) = 0
X and X^2 are uncorrelated.

(c) The random variable X has a density function given by

$$f_X(x) = \frac{a}{(1+x)^5}$$

defined over the region x > 0. Find the value of the parameter a

Find also E (1 + X) , E (1 + X)^2 , Var(X) and

$$Cov\left(X+1, \frac{1}{X+1} \right)$$

(10 Marks)

$$\int_0^\infty \frac{a}{(1+x)^5} dx = 1 \Rightarrow \mathbf{a = 4}$$

E(1+X) = $\displaystyle\int_0^\infty \frac{4(1+x)}{(1+x)^5} dx = \frac{4}{3}$

E (1 + X)^2 = $\displaystyle\int_0^\infty \frac{4(1+x)^2}{(1+x)^5} dx = 2$

Var(X) = Var(X+1) = E (1 + X)^2 – E^2 (1 + X) = 2 – 16/9 = 2/9

$$Cov\left(X+1, \frac{1}{X+1} \right) = E((1+X) \times \frac{1}{1+X}) - E(1+X)E(\frac{1}{1+X})$$

$$= E(1) - \frac{4}{3}E(\frac{1}{1+X}) = 1 - \frac{4}{3}E(\frac{1}{1+X})$$

$$E(\frac{1}{1+X}) = \int_0^\infty \frac{4}{(1+x)^6} dx = \frac{4}{5}$$

$$Cov\left(X+1, \frac{1}{X+1} \right) = 1 - \frac{4}{3} \times \frac{4}{5} = \frac{-1}{15}$$

4. Consider two random variables X and Y. They both take the values 0, 1 and 2. The joint probabilities for each pair are given by the following table.

	$X=0$	$X=1$	$X=2$
$Y=0$	0.10	0.06	0.14
$Y=1$	0.08	0.06	0.16
$Y=2$	0.20	0.08	0.12

(a) Calculate the marginal distributions, and the expected values of X and Y.

(5 marks)

(b) Calculate $E(X|Y=1)$ and $E(X|X+Y=3)$.

(6 marks)

(c) Define $U = |X-1|$ and $V = Y$. Calculate the covariance of U and V.

(6 marks)

(d) Are U and V are independent variables? Explain your answer.

(5 marks)

a)

$\Pr(X=0) = 0.10 + 0.08 + 0.20 = 0.38$,
$\Pr(X=1) = 0.06 + 0.06 + 0.08 = 0.20$, $\Pr(X=2) = 0.14 + 0.16 + 0.12 = 0.42$,
$\Pr(Y=0) = 0.10 + 0.06 + 0.14 = 0.30$, $\Pr(Y=1) = 0.08 + 0.06 + 0.16 = 0.30$ and
$\Pr(Y=2) = 0.20 + 0.08 + 0.12 = 0.40$. So

$$E(X) = 0 \times 0.38 + 1 \times 0.20 + 2 \times 0.42 = 1.04$$

$$E(Y) = 0 \times 0.30 + 1 \times 0.30 + 2 \times 0.4 = 1.1.$$

(5 marks).

b)

We have that:

$$P(X=0|Y=1) = \tfrac{0.08}{0.3} = \tfrac{8}{30} = \tfrac{4}{15},$$
$$P(X=1|Y=1) = \tfrac{0.06}{0.3} = \tfrac{6}{30} = \tfrac{1}{5}$$

and

$$P(X=2|Y=1) = \tfrac{0.16}{0.3} = \tfrac{16}{30} = \tfrac{8}{15}$$

and therefore

$E(X|Y=1) = 0 \times \frac{4}{15} + 1 \times \frac{1}{5} + 2 \times \frac{8}{15} = \frac{19}{15} = 1.267.$

We also have

$$P(X+Y=3) = 0.16 + 0.08 = 0.24$$

and so

$$P(X=1|X+Y=3) = \tfrac{0.08}{0.24} = \tfrac{1}{3}$$

and

$$P(X=2|X+Y=3) = \tfrac{0.16}{0.24} = \tfrac{2}{3}.$$

Hence

$E(X|X+Y=3) = 1 \times \frac{1}{3} + 2 \times \frac{2}{3} = \frac{5}{3} = 1.667.$

(6 marks).

c)

Here is the table of probabilities:

	$U=0$	$U=1$
$V=0$	0.06	0.24
$V=1$	0.06	0.24
$V=2$	0.08	0.32

We then have $\Pr(U=0)=0.06+0.06+0.08=0.2, \Pr(U=1)=0.24+0.24+0.32=0.8$. Of course $\Pr(V=0)=0.3, \quad \Pr(V=1)=0.3$ and $\Pr(V=2)=0.4$. So

$$E(U)=0\times 0.2+1\times 0.8=0.8,$$

$$E(V)=E(Y)=1.1$$

and

$$E(UV)=1\times 0.24+2\times 0.32=0.88.$$

Hence $\mathrm{Cov}(U,V)=0.88-0.8\times 1.1=0$. (**6 marks**).

d)

The fact that their covariance is 0 is not enough to show that they are independent. However the table of probabilities can be rewritten as

	$U=0$	$U=1$
$V=0$	0.2×0.3	0.8×0.3
$V=1$	0.2×0.3	0.8×0.3
$V=2$	0.2×0.4	0.8×0.4

We observe that $\Pr(U=i,V=j)=\Pr(U=i)\Pr(V=j)$ for all possible pairs (i,j) and so they are independent.

Note that all pairs have to be checked. (**5 marks**).

END of ANSWERS

Formulae for Statistics

Discrete Distributions

Binomial Distribution

The probability of x successes in n trials is

$$\binom{n}{x} \pi^x (1-\pi)^{n-x}$$

for $x = 0, 1, \ldots, n$ The mean number of successes is $n\pi$ and the variance is $n\pi(1-\pi)$.

Poisson Distribution

The probability of x is

$$e^{-\mu}\frac{\mu^x}{x!}.$$

The mean number of successes is μ and the variance is μ.

Hypergeometric Distribution

The probability of x successes in a sample of size n from a population of size N with M successes is

$$\binom{M}{x}\binom{N-M}{n-x} / \binom{N}{n}.$$

The mean number of successes is nM/N and the variance is $n(M/N)(1-M/N)(N-n)/(N-1)$.