04b Mock 2009 Solution

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

i. Mutually exclusive events with positive probability can not be independent:**True**

If A and B are non-zero probability events , P(A)$\neq$0 , P(B)$\neq$0.

A and B are mutually exclusive $\Rightarrow$ Pr(A|B) = Pr(A ∩ B)/Pr(B)= 0/Pr(B)= 0 $\neq$ P(A)

Therefore, for two non-zero probability events, if they are mutually exclusive, then they are not ndependent.

If A and B are independent $\Rightarrow$ Pr(A|B) = P(A) $\Rightarrow$P(A ∩ B) = P(B)P(A|B)
= P(A)P(B) $\neq$0

Therefore, for two non-zero probability events, if they are independent, then they are not mutually exclusive.

Remark: A and B are both mutually exclusive and independent if and only if : P(A)= 0 or P(B)= 0 or both.

ii. If P (A) + P (B) > 1; A and B can not be independent.**False**

A and B are independent if Pr(A ∩ B) =P(A)P(B) > 0

Since Pr(A U B) $\leq$ 1 $\Rightarrow$ Pr(A)+Pr(B) − Pr(A ∩ B) $\leq$ 1

$\Rightarrow$ Pr(A)+Pr(B) − 1 $\leq$ Pr(A ∩ B)

$\Rightarrow$ Pr(A ∩ B) $\geq$ P(A) + P(B) – 1 $\Rightarrow$ Pr(A ∩ B) > 0 since P(A) + P(B) > 1

Therefore Pr(A ∩ B) > 0 and A , B are independent. Refer to ch2 summary p. 4

iii. If P (A) + P (B) > 1; A and B can not be mutually exclusive. **False**

mutually exclusive : Pr(A ∩ B) = 0

Pr(A □ B) = Pr(A)+Pr(B)− Pr(A ∩ B) = Pr(A)+Pr(B)

Since Pr(A □ B) $\leq$ 1 $\Rightarrow$ Pr(A)+Pr(B $\leq$ 1 which contradicts the given P (A) + P (B) > 1

iv. If T_1 and T_2 are unbiased estimators for a parameter θ; then T_1T_2 is an unbiased estimator for θ^2.**False**

We have $E(T_1) = \theta$, $E(T_2) = \theta$, for T_1T_2 to be an unbiased estimator of θ^2 , we should have : $E(T_1T_2) = E(T_1)\,E(T_2) = \theta^2$ which is not true unless T1 and T2 are independent.

v. If T_1 and T_2 are unbiased estimators for a parameter θ; then T_1T_2 is an unbiased estimator for θ^2 only if T_1 and T_2 are independent. **True**

We have $E(T_1) = \theta$, $E(T_2) = \theta$, since T_1 and T_2 are independent ,
$E(T_1T_2) = E(T_1)\,E(T_2) = \theta^2$

(10 Marks)

b) An amateur gardener is trying to predict his water consumption (measured in **litres**) in terms of the air temperature (measured in ^{o}**C**) and the time he spends mowing the lawn (measured in **hours**).
The following model was fitted: $y_i = a + bx_i + cz_i + \varepsilon_i$
where x_i **represents the maximum temperature** in day i, z_i **represents the time he spent mowing the lawn** in day i, y_i **is his water consumption** in day i and ε_i is the error which is normally distributed with mean 0 and unknown variance. The estimated values were -1.3 for a, 0.1 for b and 0.8 for c. Interpret the results.

The fitted model : $y_i = a + bx_i + cz_i = -1.3 + 0.1x_i + 0.8z_i$
Controlling the time spent mowing the lawn in day i , each additional ^{o}**C** of adds 0.1 **liters** to his water consumption.
Controlling the temperature , each additional **hour** spent mowing the lawn adds 0.8 **liters** to his water consumption.
If x = 0 , z = 0 i.e. no temperature and no time spent mowing the lawn then the water consumption would be -1.3 liters which doesn't make sense.

(4 Marks)

c) Suppose that you are given observations y_1, y_2, y_3 and y_4 that are such that

$y_1 = \alpha + \beta + \varepsilon_1$

$y_2 = \beta + \varepsilon_2$

$y_3 = \alpha + \varepsilon_3$

$y_4 = \alpha - \beta + \varepsilon_4$

The variables ε_i , i = 1; 2; 3; 4 are normally distributed with mean 0.
Find the least squares estimator for the parameter β and verify that it is unbiased.

We need to minimize the sum of errors squares :

$$S = (y_1 - \alpha - \beta)^2 + (y_2 - \beta)^2 + (y_3 - \alpha)^2 + (y_4 - \alpha + \beta)^2$$

$$\frac{\partial S}{\partial \alpha} = -2(y_1 - \alpha - \beta) + 0 - 2(y_3 - \alpha) - 2(y_4 - \alpha + \beta)$$

$$= 6\alpha - 2\beta - 2(y_1 + y_3 + y_4)$$

$$\frac{\partial S}{\partial \beta} = -2(y_1 - \alpha - \beta) - 2(y_2 - \beta) + 0 + 2(y_4 - \alpha + \beta)$$

$$= 6\beta - 2(y_1 + y_2 - y_4)$$

$\hat{\alpha}$ and $\hat{\beta}$ are the solutions of $\frac{\partial S}{\partial \alpha} = 0$, $\frac{\partial S}{\partial \beta} = 0$

Since the least squares estimator for the parameter **β is only required** , we use the second equation : $6\hat{\beta} - 2(y_1 + y_2 - y_4) = 0 \Rightarrow \hat{\beta} = \frac{y_1 + y_2 - y_4}{3}$

$$E(\hat{\beta}) = E\left(\frac{y_1 + y_2 - y_4}{3}\right) = E\left(\frac{\alpha + \beta + \varepsilon_1 + \beta + \varepsilon_2 - \alpha + \beta - \varepsilon_4}{3}\right)$$

Now E(α) = α , E(β) = β , E(ε_i) = 0 since ε_i are normally distributed with mean 0

= $E\left(\frac{3\beta}{3}\right) = E(\beta) = \beta$ hence $\hat{\beta}$ is unbiased.

(9 Marks)

d) Let U and V be two independent and normally distributed random variables with mean 0 and variance 1. Find k that satisfies
P (kU - V > 6.198) = 0.025

U ~ N(0,1) , V ~ N(0,1) , U and V are independent

kU - V ~ N(?,?)
E(kU – V) = E(kU) – E(V) = k(0) – 0 = 0

Var(kU – V) = Var(kU) + Var(V) - 2cov(kU,V)

but cov(kU,V)= 0 since U and V are independent.

Var(kU – V) = Var(kU) + Var(V)= k^2**Var(U) + Var(V)**
= $k^2 + 1$
$\Rightarrow$ **kU - V ~ N(0,**k^2 **+ 1) standardize : Let kU - V = W**
P(W > 6.198) = 0.025

$$P\left(\frac{W - 0}{\sqrt{k^2+1}} > \frac{6.198 - 0}{\sqrt{k^2+1}}\right) = P\left(Z > \frac{6.198}{\sqrt{k^2+1}}\right) = 0.025$$

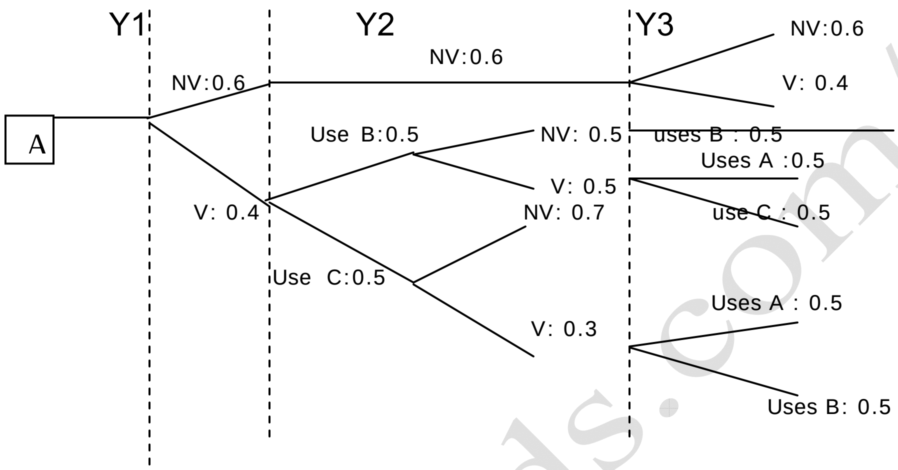
$$\Rightarrow 1 - P\left(Z \le \frac{6.198}{\sqrt{k^2+1}}\right) = 0.025 \Rightarrow 1 - \phi\left(\frac{6.198}{\sqrt{k^2+1}}\right) = 0.025 \Rightarrow \phi\left(\frac{6.198}{\sqrt{k^2+1}}\right) = 0.975$$

From Table 4 : $\phi(1.96) = 0.9750$

$$\frac{6.198}{\sqrt{k^2+1}} = 1.96 \Rightarrow 6.198 = 1.96\sqrt{k^2+1} \Rightarrow \sqrt{k^2+1} = 3.1622 \Rightarrow k^2 + 1 = 10$$

$\Rightarrow k^2 = 9 \Rightarrow k = \pm 3$ (6 Marks)

e) A man has a choice of three kinds of toothpaste. If he uses toothpaste A, there is a probability of 0.4 that he will develop a cavity in the next year. If he uses toothpaste B, the same probability is 0.5. If he uses toothpaste C, it is 0.3. If he gets no cavities during a particular year, he uses the same toothpaste the next year. Otherwise he chooses any of the other two toothpastes with equal probability. During year 1 he used toothpaste A.



V : cavity , NV : No Cavity

i. What is the probability that he used toothpaste A in year 3?

P(AV) = 0.4 , P($\overline{AV}$) = 0.6

P(BV) = 0.5 , P($\overline{BV}$) = 0.5

P(CV) = 0.3 , P($\overline{CV}$) = 0.7

P (used A in Y3) = (0.6)(0.6)(0.6) + (0.4)(0.5)(0.5) + (0.4)(0.3)(0.5) = 0.376

(4 marks)

ii. Given that he used toothpaste A in year 3, what is the probability he developed no cavities in years 1, 2 and 3?

(7 marks)

P(NV in Y1,2,3| A is used in Y3) = Pr(NV ∩ A3)/Pr(A3)

= (0.6)(0.6)(0.6) / 0.375 = 0.576

2. The random variable X has a density function given by
f (x) = ax² (x + 1) defined over the region 0 ≤ x ≤ 1.

Find a:

$$\int_{all} f(x)dx = 1 \Rightarrow \int_0^1 ax^2(x+1)dx = 1 \Rightarrow \int_0^1 (ax^3 + ax^2)dx = 1$$

$$\Rightarrow a\left(\frac{x^4}{4} + \frac{x^3}{3}\right)\Big|_0^1 = 1 \Rightarrow a(1/4 + 1/3) = 1 \Rightarrow a = \frac{12}{7}$$

P(X < 0.5 | X > 0.25) =

$$\frac{P(X \prec 0.5 \cap X \succ 0.25)}{P(X > 0.25)} = \frac{P(0.25 \prec X \prec 0.5)}{1 - P(X \le 0.25)} = \frac{\int_{0.25}^{0.5} \frac{12}{7}(x^3 + x^2)dx}{1 - \int_0^{0.25} \frac{12}{7}(x^3 + x^2)dx} = 0.089$$

$$E\left(\frac{1}{X}\right) = \int_0^1 \frac{1}{x} f(x)dx = \frac{12}{7}\int_0^1 \frac{x^2(x+1)}{x}dx = \frac{12}{7}\int_0^1 x(x+1)dx = \frac{12}{7}\int_0^1 (x^2+1)dx = \frac{10}{7} = 1.43$$

$$E\left(\frac{1}{X^2}\right) = \int_0^1 \frac{1}{x^2} f(x)dx = \frac{12}{7}\int_0^1 \frac{x^2(x+1)}{x^2}dx = \frac{12}{7}\int_0^1 (x+1)dx = \frac{18}{7}$$

Var (X) = $E\left(\frac{1}{X^2}\right) - \left(E\left(\frac{1}{X}\right)\right)^2 = \frac{18}{7} - \left(\frac{10}{7}\right)^2 = \frac{26}{49} = 0.53$

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

$$E\left(\frac{1}{X} \times \frac{1}{X+1}\right) = \int_0^1 \frac{1}{x(x+1)} f(x)dx = \frac{12}{7}\int_0^1 \frac{x^2(x+1)}{x(x+1)}dx = \frac{12}{7}\int_0^1 xdx = \frac{6}{7}$$

$$E\left(\frac{1}{X+1}\right) = \int_0^1 \frac{1}{(x+1)} f(x)dx = \frac{12}{7}\int_0^1 \frac{x^2(x+1)}{(x+1)}dx = \frac{12}{7}\int_0^1 x^2dx = \frac{4}{7}$$

$$Cov\left(\frac{1}{X}, \frac{1}{X+1}\right) = \frac{6}{7} - \frac{10}{7} \times \frac{4}{7} = \frac{2}{49} = 0.041$$

.

(19 marks)

3. A durian stall sells four varieties of durian and is open six days a week. The takings due to each variety of durian during each one of the six days of the same week were recorded. The total takings over the week measured in Singapore dollars for D99 durians were 844, for D24 durians 866, for D10 durians 839 and for D2 durians 849. The following is the calculated ANOVA table based on daily takings with some entries missing.

Source	degrees of freedom	sum of squares	mean square	F - value
Durian				
Day			27.54	
Error				
Total		345.96		

a) Complete the table using the information given above. (7 Marks)
We have r = 4 , c = 6
We kmow the Mean error for columns : 27.54
Sum of squares for columns = (r-1)MSE = (4-1)(27.54)= 82.62
There are four varieties of durian : D99,D24,D10 and D2
and the row sums are given

Row means are : $\bar{x}_{1.}$ = 844/6 = 140.7

$\bar{x}_{2.}$ = 866/6 = 144.3333

$\bar{x}_{3.}$ = 839/6 = 139.7

$\bar{x}_{4.}$ = 849/6 = 141.5

Overall mean $\bar{x}_{..}$ = 141.56

Sample variance of row means : S_R^2 = 3.9646

Mean sum squares between rows = c S_R^2 = 6(3.9646)=23.7876

sum squares between rows = (r-1) c S_R^2 =3(23.7876)=71.3628

Residual sum squares
= Total SS – SS between rows – SS between columns
= 345.96 - 71.3628 - 82.62 = 191.9772

Mean Residual SS = S² = $\frac{ResidualSS}{(r-1)(c-1)} = \frac{191.9772}{15} = 12.7985$

$F_R = c S_R^2 / S^2$ = 23.7876 / 12.7985 = 1.85

$F_C = r S_C^2 / S^2$ = 27.54 / 12.7985 = 2.15

Source	D.F.	Sum of squares	Mean Square	F-value
Durian(row)	r – 1 = 3	71.3628	23.7876	1.85
Day(column)	c – 1 = 5	82.62	27.54	2.15
Error	(r-1)(c-1)= 15	191.9772	12.7985	
Total	rc – 1 = 23	345.96		

b) Is there a significant difference between the daily takings due to each variety of durian? (4 Marks)

Durian is represented by rows :

H_0 : $\alpha_i = 0 \ \ \forall i$

H_1 : non zero row effect

Under H_0 : F_R = 1.85 ~ $F_{r-1,(r-1)(c-1)}$

The criterion : Reject H_0 if $F_R \geq F_{\alpha,r-1,(r-1)(c-1)} = F_{0.05,3,15} = 3.29$

Since 1.85 < 3.29 , we do not reject H_0 and therefore there is no significant difference between the daily takings due to each variety of durian.

c) Construct a 90% confidence interval for the diference in daily takings between D24 and D10 durians? Would you say there is a diference? (5 Marks)

For $\alpha_i - \alpha_j$: $\overline{X}_i - \overline{X}_j \pm t_{\frac{\alpha}{2},(r-1)(c-1)} S\sqrt{2/c}$

$\alpha = 0.1 \Rightarrow \alpha/2 = 0.05$, $t_{0.05,15}$ = 1.753

144.3333 - 139.7 $\pm t_{0.05,15}\sqrt{12.7985}\sqrt{2/6}$

= 4.6333 ± **(1.753)(3.5775)(0.5774)**

= 4.6333 ± **3.621 = (1.1023 , 8.2540)**

Since 0 does not belong to the interval , then there is a significant difference in daily takings between D24 and D10 durians.

d) Construct simultaneous 90% confidence intervals for the difference in daily takings between D24 and D10 and the difference between D99 and D2. (5 marks)

For $\alpha_i - \alpha_j$: $\overline{X}_i - \overline{X}_j \pm S\sqrt{(r-1)F_{\alpha,r-1,(r-1)(c-1)}(2/c)}$

$\alpha = 0.1 \Rightarrow \alpha/2 = 0.05$, $F_{0.05,3,15}$ = 8.70

D24 and D10 = **144.3333 - 139.7** $\pm\sqrt{12.7985}\sqrt{3(8.70)(2/6)}$

= 4.6333 ± **(3.5775)(5.023)**

= 4.6333 ± **17.9696 = (-13.3363, 22.60)**

D99 and D2 : = **141.5 – 140.7** $\pm\sqrt{12.7985}\sqrt{3(8.70)(2/6)}$

= 0.8 ± **17.9696 = (-17.9696, 18.7696)**

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

	X = -1	**X = 0**	**X = 1**
Y = -1	**0**	**0.2**	**0.1**
Y = 0	**0.1**	**0.2**	**0**
Y = 1	**0.1**	**0.05**	**0.25**

a) Calculate the marginal distributions and expected values of X and Y:

(6 marks)

P(X= -1) = 0 + 0.1 + 0.1 = 0.2
P(X= 0) = 0.2 + 0.2 + 0.05 = 0.45
P(X= 1) = 0.1 + 0 + 0.25 = 0.35
E(X) = -1 (0.2) + (0)(0.45) + (1)(0.35) = 0.15

P(Y= -1) = 0 + 0.2 + 0.1 = 0.3
P(Y= 0) = 0.1 + 0.2 + 0 = 0.3
P(Y= 1) = 0.1 + 0.05 + 0.25 = 0.4
E(Y) = -1 (0.3) + (0)(0.3) + (1)(0.4) = 0.1

b) Calculate E (X |Y = 1) and E (X | X + Y = 0).

(8 marks)

P(X= -1 | Y=1) = $\dfrac{P(X=-1 \cap Y=1)}{P(Y=1)} = \dfrac{0.1}{0.4} = 0.25$

P(X= 0 | Y=1) = $\dfrac{P(X=0 \cap Y=1)}{P(Y=1)} = \dfrac{0.05}{0.4} = 0.125$

P(X= 1 | Y=1) = $\dfrac{P(X=1 \cap Y=1)}{P(Y=1)} = \dfrac{0.25}{0.4} = 0.625$

E (X |Y = 1) = -1(0.25) + (0)(0.125) + (1)(0.625) = 0.375
X+ Y = 0 : (-1,1) , (0,0) , (1,-1)
P(X+Y = 0) = 0.1 + 0.2 + 0.1 0.4

P(X= -1 | X+Y=0) = $= \dfrac{0.1}{0.4} = 0.25$

P(X= 0 | X+Y=0) = $= \dfrac{0.2}{0.4} = 0.5$

P(X= 1 | X+Y=0) = $= \dfrac{0.1}{0.4} = 0.25$

E (X | X + Y = 0). = -1(0.25) + (0)(0.5) + (1)(0.25) = 0

c) Let W = min (X, Y) , Calculate Cov(X, W) :

(6 marks)

From (a) : E(X) = 0.15

Cov (X,W) = E(XW) – E(X)E(W)

E(W) = -1(0.1) + (-1)(0.2) + (-1)(0.1) +(-1)(0.1) + 0 + 0 + (-1)(0.1) + 0 + 0.25 = - 0.25

E(XW) = - 0.1 – 0.1 – 0.1 + 0.25 = - 0.05

Cov (X,W) = - 0.05 – (0.15)(- 0.25) = - 0.0125