NM317/318 Exam Paper 09/10
Three Questions for the Statistical Part

Three questions in this part are compulsory and each carries 25 marks. Battery-powered pocket calculators may be used. However, answers without any working are not acceptable, so full marks will not be awarded, even if the answer is correct, unless an outline of the calculations carried out is clearly shown.

1.(a) Two parts of a machine are manufactured independently. These two parts have probabilities p_1 and p_2 respectively of failing. Find the probability that (i) neither part fails, (ii) at least one part fails, (iii) exactly one part fails.

(8 marks)

(b) Assume that a random variable X has the probability density function

$$f(x) = \begin{cases} 0 & \text{if } x < -1; \\ x+1 & \text{if } -1 \le x \le 0; \\ 1-x & \text{if } 0 < x \le 1; \\ 0 & \text{if } x > 1. \end{cases}$$

Find the mean and variance of X.

(8 marks)

(c) Bags of sugar of nominal weight 1kg, have weights which are normally distributed with mean 1.005kg and standard deviation 6g. What is the probability that a randomly selected bag of sugar is underweight? If the firm wishes to ensure that only 1% of bags are underweight, what should the mean weight of bags be increased to? (Assume that the standard deviation remains unchange.)

(9 marks)

2.(a) A random sample of $n = 30$ observation from a $N(\mu, \sigma^2)$ population produced a sample mean $\bar{x} = 2.4$ and a sample standard deviation $s = 0.29$. Does the sample provide sufficient evidence to indicate that the population mean $\mu > 2.3$? (Test using $\alpha = 0.05$). Moreover, determine the 95% confidence interval for μ.

(8 marks)

(b) The following are the measurements of the height and weight of ten men.

(b) Given the following data:

x	1	2	3	4	5	6
y	7	5	5	3	2	0

find (i) the least square linear regression equation of y on x, and (ii) the least square linear regression equation of x on y.

(8 marks)

(c) There are three types of relay (electronic switch) to be used in a telephone exchange. From random samples we have the following measurements on the response times (milli-seconds) of these relays:

Type A	Type B	Type C
2.1	3.4	3.0
2.6	3.8	3.6
1.9	3.6	3.2
3.2	4.1	3.9
2.7	3.9	3.9

Construct an ANOVA table for the data. Is there any evidence to suggest that mean response times differ between types? Test using $\alpha = 0.05$.

(9 marks)

3.(a) An electronic component is mass-produced and then tested unit by unit on an automatic testing machine which classifies the unit as 'good' or 'defective'. But there is a probability 0.1 that the machine will mis-classify the unit, so that each component is in fact tested five times independently and regarded as good if so classified three or more times. What now is the probability of a miss-classification?

(9 marks)

(b) It is known that the number of patients arriving at the emergency room in a hospital between 6pm and 7pm has a Poisson distribution with parameter $\lambda = 6.9$. Determine the probability that, on a given day, the number of patients arriving the emergency room between 6pm and 7pm will be (i) exactly 4; (ii) at most 4; (iii) betweem 5 and 8, inclusive.

(7 marks)

(c) A gambler thinks that a die may be loaded; that is, he thinks the six numbers may not be equally likely. To test his suspicion, he rolls the die 150 times and obtains the results shown in the following table:

Number	1	2	3	4	5	6
Frequency	23	26	23	21	31	26

Do the date provide sufficient evidence to conclude that the die is loaded? Perform the hypothesis test using $\alpha = 0.05$.

(9 marks)

Solutions

1(a) Define $E_1 = \{\text{part 1 fails}\}$
$E_2 = \{\text{part 2 fails}\}$

(1) $P(\text{neither part fails})$
$= P(\bar{E}_1 \bar{E}_2) = P(\bar{E}_1) P(\bar{E}_2)$
$= (1 - P(E_1))(1 - P(E_2))$
$= (1-p_1)(1-p_2)$

(2) $P(\text{at least one part fails})$
$= 1 - P(\text{neither part fails})$
$= 1 - (1-p_1)(1-p_2)$
$= p_1 + p_2 - p_1 p_2$

(3) $P(\text{exactly one part fails})$
$= P(E_1 \bar{E}_2) + P(\bar{E}_1 E_2)$
$= P(E_1) P(\bar{E}_2) + P(\bar{E}_1) P(E_2)$
$= p_1(1-p_2) + (1-p_1)p_2$
$= p_1 + p_2 - 2p_1 p_2$

1(b) $\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{-1}^{1} x f(x) dx$
$= \int_{-1}^{0} x(x+1) dx + \int_{0}^{1} x(1-x) dx$
$= \left[\frac{1}{3}x^3 + \frac{1}{2}x^2\right]_{x=-1}^{0} + \left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_{x=0}^{1}$
$= 0 - \left(-\frac{1}{3} + \frac{1}{2}\right) + \left[\frac{1}{2} - \frac{1}{3}\right] - 0$
$= 0$

$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2 = \int_{-1}^{1} x^2 f(x) dx$
$= \int_{-1}^{0} x^2(x+1) dx + \int_{0}^{1} x^2(1-x) dx$
$= \left[\frac{1}{4}x^4 + \frac{1}{3}x^3\right]_{x=-1}^{0} + \left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_{x=0}^{1}$
$= 0 - \left[\frac{1}{4} - \frac{1}{3}\right] + \left[+\frac{1}{3} - \frac{1}{4}\right] - 0$
$= \frac{1}{6}$

1(c) Let X be the weight of a bag of sugar. Then $X \sim N(1.005,\ 0.006^2)$.

P(a bag of sugar is underweight)

$$= P(X < 1) = P\left(\frac{X-1.005}{0.006} < \frac{1-1.005}{0.006}\right)$$

$$= P(Z < -0.833) \qquad (Z \sim N(0,1))$$

$$= P(Z > 0.883) = 1 - P(Z \le 0.833)$$

$$= 1 - 0.7976 = 0.2024$$

If we require

$$P(X < 1) = 0.01$$

where $X \sim N(\mu,\ 0.006^2)$, then

$$P(X<1) = P\left(\frac{X-\mu}{0.006} < \frac{1-\mu}{0.006}\right)$$

$$= P\left(Z < \frac{1-\mu}{0.006}\right) = 0.01$$

Hence $P\left(Z > -\frac{1-\mu}{0.006}\right) = 0.01$

$$P\left(Z \le -\frac{1-\mu}{0.006}\right) = 0.99$$

But, by the Normal distribution table, we have

$$P(Z \le 2.3263) = 0.99$$

So
$$-\frac{1-\mu}{0.006} = 2.3263$$

$$\mu = 2.3263 \times 0.006 + 1$$

$$= 1.01396 \text{ (kg)}$$

That is, the mean weight of bags should be increased to 1.01396 (kg).

2(a) $H_0: \mu = 2.3$

$H_1: \mu > 2.3$ (1-tailed test)

Compute

$$t = \frac{\bar{x} - 2.3}{s/\sqrt{n}} = \frac{2.4 - 2.3}{0.29/\sqrt{30}} = 1.89$$

Given $\alpha = 0.05$, $d.f. = n-1 = 29$.

find $t_{\alpha, n-1} = t_{0.05, 29} = 1.699$

Since $t > t_{\alpha, n-1}$, we reject H_0 and conclude that $\mu > 2.3$ significantly.

To find the 95% C.I., we note

$$(1-\alpha)100\% = 1-\alpha = 0.95 \Rightarrow \alpha = 0.05$$

$$t_{\alpha/2, n-1} = t_{0.025, 29} = 2.045$$

$\therefore$ the 95% C.I. is

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} = 2.4 \pm 2.045 \times \frac{0.29}{\sqrt{30}}$$

$$= 2.4 \pm 0.11 = (2.29, 2.51)$$

2(b) $\bar{x} = \frac{1}{6}(1+2+\cdots+6) = 3.5$

$\bar{y} = \frac{1}{6}(7+5+\cdots+0) = 3.67$

$S_{xx} = 1^2+2^2+\cdots+6^2 - 6\times 3.5^2 = 17.5$

$S_{xy} = 1\times 7 + 2\times 5 + \cdots + 6\times 0 - 6\times 3.5\times 3.67 = -23$

$S_{yy} = 7^2+5^2+\cdots+0^2 - 6\times 3.67^2 = 31.33$

(i) To get the l.s.r. eq of y on x, compute

$$\hat{a}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-23}{17.5} = -1.314$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1\bar{x} = 3.67 - (-1.314)\times 3.5$$
$$= 8.269$$

$$\therefore \hat{y} = 8.269 - 1.314x$$

(ii) To get the l.s.r. eq of x on y, compute

$$\hat{a}_1 = \frac{S_{xy}}{S_{yy}} = \frac{-23}{31.33} = -0.734$$

$$\hat{a}_0 = \bar{x} - \hat{a}_1\bar{y} = 3.5 - (-0.734)\times 3.67$$
$$= 6.194$$

$$\therefore \hat{x} = 6.194 - 0.734y$$

2(c) ANOVA.

Treatment totals.

$T_A = 12.5, \quad T_B = 18.8, \quad T_C = 17.6$

Grand total.

$T = 48.9$

$$SS_{total} = 2.1^2 + 2.6^2 + \cdots + 3.9^2 - \frac{48.9^2}{3\times 5}$$

$$= 6.496$$

$$SS_{treat} = \frac{1}{5}\left[12.5^2 + 18.8^2 + 17.6^2\right] - \frac{48.9^2}{3\times 5}$$

$$= 4.476$$

$$SS_{error} = 6.496 - 4.476 = 2.02$$

$$MS_{treat} = \frac{4.476}{3-1} = 2.238$$

$$MS_{error} = \frac{2.02}{3\times(5-1)} = \frac{2.02}{12} = 0.168$$

$$F = \frac{2.238}{0.168} = 13.32$$

ANOVA Table

Source	SS	d.f.	MS	F
Treat.	4.476	2	2.238	13.32
Error	2.02	12	0.168	
Total	6.496	14		

Hypothesis test.

H_0: $\mu_A = \mu_B = \mu_C$ (mean response times are the same)

H_1: mean response times differ

Given $\alpha = 0.05$,

d.f. $v_1 = 2, \quad v_2 = 14$

find $F_{0.05, 2, 12} = 3.89$

$\because$ $F = 13.32 > F_{0.05, 2, 12}$

$\therefore$ reject H_0 and accept H_1, that is the mean response times differ.

3(a) Case 1. Assume that the unit tested is in fact good. Let X = the times when it was classified as good during the 5 independent tests. So

$X \sim B(5, 0.9)$

Then P(miss-classification)

$= P(X \le 2)$

$= \binom{5}{0} 0.9^0 \times 0.1^5 + \binom{5}{1} 0.9^1 \times 0.1^4 + \binom{5}{2} 0.9^2 \times 0.1^3$

$= 0.00856$

Case 2. Assume that the unit tested is in fact defective. Let Y = the times when it was classified as good during the 5 independent test. Then

$X \sim B(5, 0.1)$

and

P(miss-classification)

$= P(X \ge 3)$

$= \binom{5}{3} 0.1^3 \times 0.9^2 + \binom{5}{4} 0.1^4 \times 0.9 + \binom{5}{5} 0.1^5 \times 0.9^0$

$= 0.00856$

We therefore see that prob of miss-classification is independent of the unit tested being good or defective. So the prob = 0.00856

3(b) Let X = the number of patients arriving at the emergency room between 6–7pm.

(i) $P(X=4) = e^{-6.9} \times \frac{6.9^4}{4!} = 0.095$

(ii) $P(X \le 4) = e^{-6.9}\left[\frac{6.9^0}{0!} + \frac{6.9^1}{1!} + \frac{6.9^2}{2!} + \frac{6.9^3}{3!} + \frac{6.9^4}{4!}\right]$

$= 0.18825$

(iii) $P(5 \le X \le 8) = e^{-6.9}\left[\frac{6.9^5}{5!} + \frac{6.9^6}{6!} + \frac{6.9^7}{7!} + \frac{6.9^8}{8!}\right]$

$= 0.55972$

3(c) H_0: the die is fair

H_1: The die is loaded

Under the H_0, we should have

Number	1	2	3	4	5	6
Prob	1/6	1/6	1/6	1/6	1/6	1/6

Compute

$$X^2 = \sum_{i=1}^{6} \frac{(n_i - np_i)^2}{np_i}$$

where $n = 150$, $p_i = \frac{1}{6}$, $np_i = 25$, n_i's are the frequencies given. Hence

$$X^2 = \frac{(23-25)^2}{25} + \frac{(26-25)^2}{25} + \frac{(23-25)^2}{25} + \frac{(21-25)^2}{25} + \frac{(31-25)^2}{25} + \frac{(26-25)^2}{25}$$

$$= \frac{1}{25}(4+1+4+16+36+1)$$

$$= 2.48$$

Given $\alpha = 0.05$, d.f. $= 6-1 = 5$, we find

$$\chi_5^2(0.05) = 11.07$$

As $X^2 < \chi_5^2(0.05)$, we accept H_0 and conclude that the die is fair.