

Open-book examination  
 Calculators : only authorized models  
 Duration : 3 hours

**14-IN-A7 Applied Probability and Statistics**

**Question 1 (10 points)**

If  $A$  and  $B$  are mutually exclusive events,  $P(A) = 0.25$  and  $P(B) = 0.45$ , find:

- a) (2.5 points)  $P(\bar{A})$ ;
- b) (2.5 points)  $P(A \cup B)$ ;
- c) (2.5 points)  $P(A \cap \bar{B})$ ;
- d) (2.5 points)  $P(\bar{A} \cap \bar{B})$ ;

**Question 2 (20 points)**

The following frequency table shows the classification of 58 landfills in the state according to their concentration of the three hazardous chemicals arsenic, barium and mercury.

		Barium			
		High		Low	
		Mercury		Mercury	
		High	Low	High	Low
Arsenic	High	1	3	5	9
	Low	4	8	10	18

If a landfill is selected at random, find the probability that it has:

- a) (5 points) a high concentration of mercury;
- b) (5 points) a high concentration of barium and low concentrations of arsenic and mercury;
- c) (5 points) a high concentration of any two of the chemicals and low concentration of the third;
- d) (5 points) a high concentration of any one of the chemicals and low concentrations of the other two.

### Question 3 (20 points)

You are responsible for the estimation of equipment **reliability**. The Marketing Department wants to offer a 2 years warranty (which corresponds to approximately **500 hours** of operation). We accept a return rate due to the guarantee of around **10%**. A test bench has done a cycling test of 12 equipment until they broke. The results of their lifetimes (in hours) are:

594	583	552	571	567	559
545	544	528	564	563	547

- (7 points) Estimate the average life (mean time) with 95% as confidence interval.
- (7 points) Can we certify if the average lifetime of the equipment meets the threshold of 500 hours at 90%?? Justify your answer with a statistical test.
- (6 points) For accuracy in predicting the life of  $\pm 50$  hours, what is the minimum sample size you should consider?

### Question 4 (20 points)

Levels of a basin  $Y$  (m) and the valve opening  $X$  (%) were measured. The results are shown in the table below.

$X$ (%)	10	20	30	40	50	60	70	80	90
$Y$ (m)	5,050	5,208	5,494	5,952	6,66	7,812	9,803	13,888	26,315

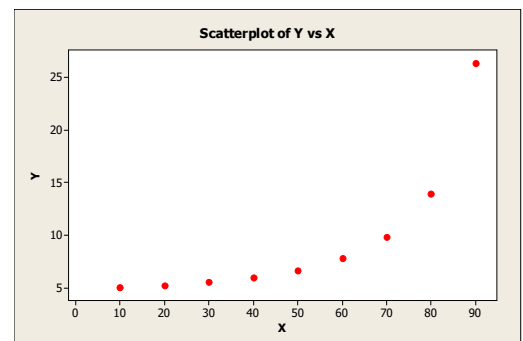
The chart is for reference only.

- (7 points) Considering the empirical model:

$$Y = a_0 + \frac{a_1}{1 - \left(\frac{X}{100}\right)^{1.5}} + \varepsilon$$

Where  $\varepsilon$  is a Gaussian noise with zero mean. You are asked to estimate the parameters  $a_0$  and  $a_1$  using the least squares method.

- (6 points) Is the value of  $X$  significant or can it be considered "virtually" zero? Justify your answer.
- (3 points) Calculate the linear correlation coefficient (*Pearson coef.*) between  $X$  and  $Y$ . Can it be considered "significant"? Justify your answer by considering the sample size  $n=9$  and an error  $\alpha=0.05$ .



- d) (4 points) Repeat the previous question, but considering the two variables  $X$  and  $Z$ , with :  $Z = \frac{1}{1 - \left(\frac{X}{100}\right)^{1.5}}$ .

### Question 5 (10 points)

As part of an industrial training program, some trainees are instructed by *Method A*, and some are instructed by *Method B*. If  $r$  random samples of size 10 are taken from large groups of trainees instructed by each of these two methods, and the scores which they obtained in an appropriate achievement test are:

Method A	68	74	71	68	66	73	69	65	75	71
Method B	72	77	84	78	69	70	77	73	65	75

Use the 0.05 level of significance to test the claim that Method B is more effective. Assume that the populations sampled can be approximated closely with normal distributions having the same variance.

### Question 6 (10 points)

Scientific research is conducted to demonstrate a relation between **smoking** and **cancer**. A statistical study was conducted on **13,420** people over five years. The results were classified into two categories: **Non-smoking** (2 months of cigarettes / week) and **Smoking**. The results are:

Profile	No Pathology	Pathology	Total
Non-smoking	10962	143	10905
Smoking	1874	541	2415
Total	12836	684	13320

- a) (5 points) With the assumption that lung cancer is related to smoking, formulate statistical hypotheses  $H_0$  and  $H_a$  for this contingency table.
- b) (5 points) Can we consider as probable the hypothesis that smoking is a risk for lung cancer? Justify your answer. Use error TYPE I = 5%.

### Question 7 (10 points)

Consider  $X$  as a random variable which can be considered normal with average equal to 1 and a unit variance  $x \sim N(\mu_x = 1, \sigma_x^2 = 1)$ .  $Y$  is a new variable defined as  $Y = 2X - 2$ . Calculate the mean and variance of the new variable  $Y$ .

---

**End**