

Question 1. [6 pts] Following data are the exam scores x of $n = 20$ students in a biology course.

88 75 80 77 77 80 84 84 94 90
87 62 93 80 85 65 70 72 66 66

$$\sum x = 1575, \sum x^2 = 125783$$

a) **[2 pts]** Fill in the following frequency table.

Score class	Frequency	Cumulative Frequency
60- < 70		
70- < 80		
80- < 90		
90- < 100		

b) **[1 pt]** What is relative frequency of the scores less than 80?

c) **[1 pt]** Use the raw data to find the mean and standard deviation of the scores.

d) **[2 pt]** Find Q_1 , Q_2 , Q_3 , and plot the box plot.

Question 2. [4 pts] A group of people are categorized according to gender and level of education. The results are given in the following table.

Level of Education	Gender	
	Male (M)	Female (F)
Elementary (E)	50	70
Intermediate (I)	80	70
Advanced (A)	30	20

A member of this group is selected randomly. Find the probability that the selected member is:

a) **[1 pt]** a male with advanced education.

b) **[1 pt]** a female or has elementary education.

c) **[1 pt]** a female given that she has intermediate education.

d) **[1 pt]** Are the two events M: male and A: advanced education independent?

Question 3. [4 pts] It is well-known that 45% of freshman students at a college have deficiency in the Physics subject. Find the probability that in a group of 25 students:

a) **[1 pt]** at most 12 have deficiency in Physics.

b) **[1.5 pts]** exactly 12 have deficiency in Physics.

c) **[1.5 pts]** between 5 and 10 (inclusive) have deficiency in Physics.

Question 4. [6 pts] Assume that the serum iron levels of a group of healthy men has normal distribution with mean 120 and standard deviation 15 micrograms per 100 ml. Find

a) **[2 pts]** the probability that the iron level of a healthy man is between 105 and 142.5 micrograms per 100 ml.

b) **[2 pts]** Find the top 5% serum iron level in healthy men [that is, find x that $P(X > x) = 0.05$].

c) **[2 pts]** Find the probability that a random sample of 9 healthy men has an average serum iron level of more than 125 micrograms per 100 ml.

Question 5. [7 pts] The following table shows the summary statistics on the hours of night sleep of males and females computed based on two random samples chosen from the populations of males and females in a society. Suppose that the hours of night sleep of both males and females have independent normal distributions.

Gender	Sample size	Mean	St. Dev.
Male	16	6 hours	1.5 hours
Female	20	7 hours	1 hour

a) **[2 pts]** Can we conclude, with $\alpha = 0.01$, that the population mean of the hours of male sleep is more than 5.5 hours?

b) **[2.5 pts]** Construct a 90% confidence interval on the difference between the population means of the hours of male and female sleep. (Assume that population variances are equal.)

c) **[2.5 pts]** Can we conclude, with 5% significance level, that the population mean of the hours of male sleep is less than female sleep? (Assume that population variances are equal.)

Question 6. [4 pts] The following table shows the number of smokers in two random samples chosen from two countries.

Country	Sample size	No. of smokers
1	200	80
2	250	95

a) **[2 pts]** Construct a 95% confidence interval on the difference between the proportions of smokers in Country 1 and Country 2.

b) **[2 pts]** Can we conclude, with 5% significance level, that the proportions of smokers in both countries are equal?

Question 7. [6 pts] The following data show the age x (in year) and the waist size y (in cm) of 10 middle-age men.

x	47	50	55	53	49	50	53	55	49	54
y	88	95	98	92	86	90	96	101	89	99

$$\sum x = 515 \quad SS_{xx} = \sum x^2 - n\bar{x}^2 = 72.5 \quad SS_{xy} = \sum xy - n\bar{x}\bar{y} = 115$$

$$\sum y = 934 \quad SS_{yy} = \sum y^2 - n\bar{y}^2 = 236.4$$

a) **[2 pts]** Compute the correlation coefficient between the age and the waist size. Comment on the direction and strength of the linear relationship between the two variables.

b) **[3 pts]** Write the equation of the simple linear regression line of y on x . Use this equation to predict the waist size of a 54 years old man. How much error is occurred in estimating the waist size of this man?

c) **[1 pt]** Find the coefficient of determination for the simple linear regression line of part (b).

Question 8. [3 pts] The following data on the frequencies of cardiovascular exercises and heart conditions of some random individuals have been collected.

	No exercise	Some exercise
No heart disease	10	40
Moderate heart disease	50	60
Sever heart disease	90	50

Can you conclude, with 0.05 significance level, that the heart condition is independent of the frequency of cardiovascular exercise?