Consider an experiment in which 10 identical small boxes are placed side-by-side on a table. A crystal is placed, at random, inside one of the boxes. A self professed "psychic" is asked to pick the box that contains the crystal.

Answer the following questions on the separate and included paper.

Part (A): If the "psychic" simply guesses, what is the probability that she picks the box with the crystal?

Part (B): If the experiment is repeated seven times, what is the probability that the "psychic" guesses correctly at least once?

Part (C): If the psychic is guessing, what is the expected number of correct decisions in seven trials?

Part (D): If the psychic is guessing, what is the probability of no correct decisions in seven trials?

Part (E): Now suppose the psychic possesses extra-sensory perception (ESP) and $p = 0.5$. What is the probability that the psychic guesses incorrectly in all seven trials?

Part (F): A skeptic does not believe that the "psychic" is any better at picking the box than anyone else. The skeptic asked the "psychic" to pick the box with the crystal seven times. Each and every time the "psychic" failed to pick the correct box. What would you infer about this persons psychic ability?

There have been numerous studies conducted about the connection between smoking and lung cancer conducted by both academic sources and government agencies. It is generally accepted knowledge that tobacco smoke increases the risk of cancer; J. R. Tobacco Inc. believes that their new brand of cigarette, "light tar" reduces lung damage compared to "regular" cigarettes.

The effects of lung damage are measured by a lung capacity/strength measure that is generally accepted by the medical community. To study the possible differences in lung damages from the two types of cigarettes, the effects will be studied over a period of 10-weeks using 300 non-obese subjects who each smoke six (6) cigarettes per day. Each of these 300 participants will be assigned to one of two groups: the control group who will only smoke "regular" cigarettes and the experimental group who will only smoke the "light tar" cigarettes. The subjects are not aware throughout the study which cigarettes they are smoking since they are all unlabelled. At the end of the 10-week study, each participant will have their lung capacity/strength measured by a physician who is also unaware of which group the subject is in.

Answer the following questions on the separate and included paper.

Part (A): Is this an experiment or observational study? Explain why you chose your answer.

Part (B): Describe an appropriate method for assigning the subjects to the two groups? (Note: Your method must ensure that there are an equal number of subjects in each group.)

Part (C): For the situation, the researchers decided to include a group who smoked regular cigarettes as a control group. Explain why it is important to include the control group for this study despite the fact that the lung capacity/strength will be measured at the beginning and end of the study?

Part (D): What is the term applied to the fact that both the subjects and physicians in this test are unaware of which groups the subjects are assigned to? What difference does this make for the design?

Part (E): Why would the researchers chose to only include non-obese people in the study?

The 2004 summer Olympics in Athens, Greece will feature plenty of track-and-field events, including the four member relay. In this relay, each team member runs one lap around the track, then another member runs, and this repeats until the six members have each run one lap. Their individual and team (sum of six member) times are monitored in seconds. The following is based on data from the lap times for each member (and are normally distributed):

	Mean (μ)	Standard Deviation (σ)
Relay Team Member 1	37	6
Relay Team Member 2	39	3
Relay Team Member 3	41	4
Relay Team Member 4	40	5
Relay Team Member 5	37	2
Relay Team Member 6	38	6

Answer the following questions on the separate and included paper. Leave all your measures in seconds, even if it takes an unusual form such as 285 seconds. Do not convert to minutes/seconds.

Part (A): The distribution of team times is normal (since it is the sum of six normal variables). What are the mean and standard deviation of this distribution?

Part (B): Runner 5 thinks that he can improve the team score by running his lap in 29 seconds. Is it likely that his next lap will be in 25 seconds? Why or why not?

Part (C): Are the six individual lap times independent normal random variables? Why or why not?

Part (D): Is the overall team time an independent random variable? Why or why not?

Part (E): What is the probability that the team finished the race in less than 200 seconds?

The Minnesota Employer Opinion Survey gives the results of an extensive survey of employer opinions in Minnesota. Each employer was asked to rate his or her satisfaction with the preparation of employees by the public education system. Responses were 1, 1.5, or 2, representing very dissatisfied, neither satisfied nor dissatisfied, and very satisfied, respectively. A sample of 651 employers was selected. Assume that the mean for all employers in Minnesota is 1.50 (the "dividing line" between satisfied and dissatisfied) and the standard deviation is 0.45.

Answer the following questions on the separate and included paper.

Part (A): Which type of distribution describes the individual survey responses, continuous or discrete? Explain.

Part (B) Describe the distribution that best approximates the sample mean response of 651 employers. What are the mean and standard deviation of this distribution?

Part (C): What is the approximate probability that the sample mean will be less than 1.45?

Part (D): The mean of the sample of 651 employers surveyed in a recent year was 1.36. Given this result, do you think it is likely that all Minnesota employers' opinions were evenly divided on the effectiveness of public education? That is, do you think the assumption that population mean is 1.50 is correct? Why or why not?

Suppose an insurance company wants to relate the amount of fire damage in major residential fires to the distance between the burning house and the nearest fire station. The study is to be conducted in a large suburb of a metropolitan area. A sample of fifteen (15) recent fires in the suburb is selected. The amount of damage, y, and the distance between the fire and nearest fire station, x, are recorded for each fire.

The data is given below ($n = 15$):

Distance from the fire station, x, in miles	Fire damage, y, in thousands of dollars
3.4	26.2
1.8	17.8
4.6	31.3
2.3	23.1
3.1	27.5
5.5	36.0
0.7	14.1
3.0	22.3
2.6	19.6
4.3	31.3
2.1	24.0
1.1	17.3
6.1	43.2
4.8	36.4
3.8	26.1

Answer the following questions on the separate and included paper.

Part (A): Calculate a regression line for the given data. Regress, y, fire damage on x, distance from station. Write your regression equation on the answer sheet for part (A). Be sure to use the form $y = B_o + B_1 x + e$.

Part (B): Given an interpretation for your value of $\hat{\beta}_0$

Part (C): Given an interpretation for your value of $\hat{\beta}_1$

Part (D): Calculate the value of r, the correlation coefficient.

Part (E): What kind of association does the r value lead you to believe is present in the original data? Explain.

Part (F): What percentage of variation in fire damage can be explained by the variation in distance from the fire station?

Part (G): What would the value of the correlation coefficient be if you measured the fire damage variable in dollars instead of thousands of dollars? Explain.

Part (H): Based on the given data, is it reasonable to attempt to predict the fire damage for a structure that is 17.2 miles away from the station? Why or why not? If it is reasonable, perform the calculation and in addition to your explanation, provide the predicted value.

Part (I): Based on the given data, is it reasonable to attempt to predict the fire damage for a structure that is 4.2 miles away from the station? Why or why not? If it is reasonable, perform the calculation and in addition to your explanation, provide the predicted value.