Exercise Set 1

1. A local bank reports that 80 percent of its customers maintain a checking account, 60 percent have a savings account, and 50 percent have both. A customer is chosen at random.
    a. What is the probability the customer has either a checking or a saving account?
    b. What is the probability the customer has neither a checking nor a saving account?

2. A Cork Society of Investors survey found that 30 of their 100 individual investors have used a discount broker. In a random sample of 9 individuals, what is the probability that:
    a. Exactly 2 of the sampled individuals have used a discount broker?
    b. Exactly 4 of them have used a discount broker?
    c. None of them have used a discount broker?
    d. State 2 conditions required for the use of the distribution used to solve parts (a) to (c).

3. A Normal population has a mean of 20.0 and a standard deviation of 4.0
    a. Compute the z value associated with 25.0.
    b. What proportion of the population is between 20.0 and 25.0?
    c. What proportion of the population is less than 18.0?
    d. Name 2 characteristics of the standard Normal distribution

4. The O'Leary family has twins, Majella and Elaine. Both Majella and Elaine graduated from college 2 years ago, and each is now earning €50,000 per year. Elaine works in the retail industry where the mean salary for executives with less than 5 years' experience is €35,000 with a standard deviation of €8,000. Majella is an engineer. The mean salary for engineers with less than 5 years' experience is €60,000 with a standard deviation of €5,000. Compute the z-values for Majella and Elaine and comment on your findings.

5. The mean weight of a large group of people is 180 lb and the standard deviation Is 15 lb. If the weights are normally distributed, find the probability that a person picked at random from the group will weigh:
    a. Between 160 and 180 lb
    b. Above 200 lb
    c. Below 150 lb

Exercises Set 2

1. A manufacturer of bulbs took a sample of 13 bulbs from a day's production and used them continuously until they were drained. The numbers of hours they were used until failure were:

342	**426**	**317**	**545**	**264**	**451**	**1049**	**631**	**512**	**266**	**492**	**562**	**298**

   i. Compute the mean, median and mode. Looking at the distribution of times of failure, which measures of location do you think are most appropriate and which least appropriate to use for these data? Why?

   ii. Calculate the range, variance and standard deviation

   iii. What would you advise if the manufacturer wanted to be able to say in the advertisement that these bulbs "should last 400 hours"? (Note: There is no right answer to this question. The point is to consider how to make such a statement precise).

   iv. Suppose that the first value was 1342 instead of 342. Repeat (i) through (iii), using this value. Comment on the difference in the results.

   v. Explain the difference between a sample and a population.

2. Describe the major characteristics of the standard deviation.

3. The daily water usage per person in Waterford follows a normal distribution with a mean of 20 gallons and a standard deviation of 5 gallons.

   i. What is the probability that a person from Waterford selected at random will use between 20 and 24 gallons per day?

   ii. What percent of the population use between 18 and 26 gallons per day? After calculating the percentage, use a diagram to show your answer.

4. The Wood County sheriff classifies crimes by age (in years) of the criminal and whether the crime is violent or nonviolent. As shown below, a total of 150 crimes were reported by the sheriff last year.

	Age (in years)			
Type of crime	**Under 20**	**20 to 40**	**Over 40**	**Total**
Violent	27	41	14	82
Non Violent	12	34	22	68

Total	39	75	36	150

i. What is the probability of selecting a case to analyse and finding it involved a violent crime?

ii. What is the probability of selecting a case to analyse and finding the crime was committed by someone less than 40 years old?

iii. What is the probability of selecting a case that involved a violent crime or an offender less than 20 years old? Which rule of addition did you apply?

iv. Two crimes are selected for review by Judge Smith. What is the probability that both are violent crimes?

5. If you roll a d20 die (a die with 20 faces) 4 times, how many different outcomes are possible?

Exercises Set 3

1. The director of a manufacturing at a clothing factory needs to determine whether a new machine is producing a particular type of cloth according to the manufacturer's specification, which indicate that the cloth should have a mean breaking strength of 70 pounds and a standard deviation of 3.5 pounds. A sample of 49 pieces of cloth reveals a sample mean breaking strength of 69.1 pounds.

    a. Is there evidence that the machine is not meeting the manufacturer's specifications for mean breaking strength? (Use a 0.05 level of significance)
    b. Compute the p-value and interpret its meaning.
    c. What is your answer in (a) if the standard deviation is 1.75 pounds?
    d. What is our answer in (a) if the sample mean is 69 pounds and the standard deviation is 3.5 pounds?

2. Children in the U.S.A. account directly for 36$ billion in sales annually, when their indirect influence over product decisions from stereos to vacations is considered, the total economic spending impacted by children is the U.S.S. is 290$ billion. It is estimated that by age 10, a child makes an average of over five trips a week to a store (Goldberg et al., "Understanding Materialism Among Youth, Journal of Consumer Psychology, 2003 (13(3):278-288). Suppose that you want to prove that children in Cork average more than five trips a week to a store. Let μ represent the population mean number of times children in your city make trips to a store.

    a. State the null and the alternative hypothesis
    b. Explain in the context of the above scenario the meaning of the Type I and Type II errors.
    c. Suppose that you carry out a study in Cork. Based on past studies, you assume that the standard deviation of the number of trips to the store is 1.6. You take a sample of 100 children and find that the mean number of trips to the store is 5.47. At the 0.01 level of significance, is there evidence that the population mean number of trips to the store is greater than 5 per week?
    d. Interpret the meaning of the p-value in (c).

3. The director of admissions at a large university advises parents of incoming students about the cost of text books during a typical semester. He selected a sample of 100 students and recorded their textbook expenses for the semester. He then computed a sample mean cost of 351.40$ and a sample standard deviation of 43.20.

    a. Using the 0.01 level of significance, is there evidence that the population mean is above 300$?
    b. What is your answer in (a) if the standard deviation is 75$ and the 0.05 level pf significance is used?
    c. What is your answer in (a) if the sample mean is $305.11 and the sample standard deviation is $43.20?

Exercises Set 4

Hypothesis Testing

1. A researcher suggests to you that the average square footage of a semidetached house in Ireland is 1200 square feet. To test this hypothesis, you collect a random sample of data for 200 semidetached houses and find the average square footage from this sample to be equal to 1180 square feet with a standard deviation equal to 72 square feet. Does your sample provide sufficient evidence statistically to suggest that the average square footage of 3-bed semidetached houses is lower than 1200 square feet? (Use a 5% level of significance for your hypothesis test)

2. A business travel magazine wants to classify international airports according to mean rating for a population of business travellers. A rating scale with a low score of 0 and a high score of 10 will be used, and airports with a population mean rating greater than 7 will be designated as superior service airports. Dublin airport got an average rating of 7.5 from 20 business travellers. The standard deviation of this sample of rating was equal to 1.1. Can we say from the data that Dublin should be designated as a superior service airport? (Use a 5% level of significance for your hypothesis test)

Regression Analysis

3. Find the regression equation for the following consumption schedule and interpret the results.

Year	n	Y_i	X_i
1988	1	102	114
1989	2	106	118
1990	3	108	126
1991	4	110	130
1992	5	122	136
1993	6	124	140
1994	7	128	148
1995	8	130	156
1996	9	142	160
1997	10	148	164
1998	11	150	170
1999	12	154	178

Where:
Y = Consumption in €
X = Income in €

Exercise Set 5

1. The owner of Brittens Egg Farm wants to estimate the mean number of eggs laid per chicken. A sample of 20 chickens shows they laid an average of 20 eggs per month with a standard deviation of 2 eggs per month

a) What is the value of the population mean?

b) Explain why we need to use the t distribution. What are the characteristics of the t-distribution? What assumption are these characteristics dependent on?

c) For a 95 per cent confidence interval, what is the value of t

d) Develop the 95 percent confidence interval for the population mean

e) Would it be reasonable to conclude that the population mean is 21 eggs? What about 25 eggs?

2. Describe the relationship between Confidence Interval and Significance Level in hypothesis testing

3. Explain what are Type I and Type II Errors and their relationship with what discussed in exercise 1 above.

4. A general sales manager plans to air a commercial for a digital camera on selected local TV stations prior to a sale starting on Saturday and ending on Sunday. To find out whether there is any relationship between the number of times the advertisement was aired and digital camera sales she collects the following information.

Location of TV station	**Number of Airings**	**Sat-Sun Sales ($ '000)**
Providence	4	15
Springfield	2	8
New Haven	5	21
Boston	6	24
Hartford	3	17

a) Which variable is the dependent variable? Which variable is the independent variable?

b) Draw a scatter diagram and use the scatter diagram to infer about the degree of correlation between the two variables

c) What factors do you need to keep in mind when interpreting the correlation coefficient?

d) Given that r = 0.9295, determine the coefficient of determination and interpret it.