Hypothesis Testing - Tutorial Sheet B

1. The strength of dosage of a plant growth enhancement chemical is often measured by the proportion of plants that grow faster. A particular dosage of the chemical is fed to 113 plants of these plants, 78 actually show faster growth.

   (a) Find a 95% confidence interval for the proportion of plants that would show faster growth.

   (b) The manufacturers of the product claim that the chemical is effective in more than 75% of cases. Test this claim using a 5% level of significance.

2. A press release for a broadband provider stated that 90% percent of customers are very satisfied with the standard of service. To test this claim, the local chamber of commerce surveyed 110 randomly selected customers. Among the sampled customers, 89 stated that they are very satisfied.

   (a) Compute the appropriate value for the standard error for a confidence interval.

   (b) Compute the 95% confidence interval for π, the true proportion.

   (c) What is your conclusion for this claim made by the press release? Justify your answer.

3. Six athletes run 400 metres both at sea level and 1000m above sea level. The times they take are given below.

Runner	1	2	3	4	5	6
Time at sea level	45.8	47.1	45.4	46.1	48.4	44.9
Time at altitude	44.8	45.1	45.4	45.1	45.4	45.9

   (a) Is there any evidence that their speed depends on altitude?

   (b) Calculate a 95% confidence interval for the mean difference between the time at sea level and the time at altitude. Using this confidence interval, test the null hypothesis that the average time at altitude is 3 seconds faster than the average time at sea level (state the appropriate significance level).

   (c) What assumption is used in both parts a) and b)?

4. A claim has been made that the mean body temperature of healthy adults is equal to 98.6 degrees Fahrenheit. A sample of 106 people who are taking medication for a chronic illness has produced a mean body temperature of 100.3 degrees Fahrenheit and a standard deviation of 0.65 degrees. Test the claim that this population of people taking medication has a different mean body temperature to the general population. Clearly state your null and alternative hypotheses and your conclusion. Use a 5% level of significance.

5. A geneticist claims that one in four of a certain strain of plant will have blue flowers. Out of 96 plants observed, 15 had blue flowers.

   (a) Use a 5% level of significance to test whether this result is consistent with the geneticists claim.

6. A well-known polling company estimates that more than 57% of Irish voters support a new constitutional amendment. 800 people were randomly surveyed and asked about their voting preferences. 488 of the 800 people responded positively to the amendment. You are required to:

(a) Obtain a point estimate of the proportion of people supporting the constitutional amendment.

(b) Construct a 95% confidence interval for the proportion of people in favour of the constitutional amendment.

(c) Test the polling companies claim about the outcome of the referendum, using a 5% level of significance.

7. Based on birth records for millions of babies, the percentage of newborn babies in Sweden that are female is 51%. A group of researchers in Sweden are interested in finding out if women who suffer from severe morning sickness are more likely to have a girl. The researchers looked at records for 1000 women admitted to hospital for severe morning sickness and determined that 560 of these women gave birth to a female baby. *[Source: Lancet 354: 2053, 1999].*

   (a) Calculate a 95% confidence interval for the percentage of women who suffer from severe morning sickness and give birth to a female baby. Interpret this interval.

   (b) Do these results provide evidence that women who suffer from severe morning sickness are more likely to produce female babies? Test this hypothesis using a 5% level of significance. Clearly state your null and alternative hypotheses and your conclusion.

8. An exercise physiologist wants to determine if several short bouts of exercise provide the same benefit for cardiovascular fitness as one long bout of exercise.
   50 volunteers are randomly assigned to group 1 and do standardised aerobic exercise on a stationary bicycle for 30 minutes once a day, 5 days a week. 40 volunteers are randomly assigned to group 2 and do the same exercise for 10 minutes, 3 times a day, 5 days a week. Cardiovascular fitness was measured by VO2 max (maximum oxygen consumption while exercising).

   Group 1 The mean change in VO2 after 12 weeks of exercise was 2.1 for group 1 with a standard deviation of 1.7.

   Group 2 The mean change in VO2 after 12 weeks of exercise was 0.7 for group 2 with a standard deviation of 1.

   Test the hypothesis that there is no significant difference between two groups are the same. You may assume a 5% level of significance.

   (a) Formally state your null and alternative hypotheses.

   (b) Compute the test statistic.

   (c) Discuss your conclusion to this test, supporting your statement with reference to appropriate values. You may assume a 5% level of significance.

9. A study was made of children who were hospitalized as a result of a car accident. 280 of the children were not wearing seat belts and 98 of these were seriously injured. 130 children wore seat belts and 26 were seriously injured.

   (a) Test the hypothesis that the rate of serious injury is the same for children who wear a seat belt or not. Clearly state your null and alternative hypotheses and your conclusion. Use a significance level of 5%.

10. It is generally assumed that older people are more likely to vote for the Conservatives than younger people. In a survey, 180 of 400 people over 40 and 120 of 400 people under 40 stated they would vote Conservative.

    (a) Do the data support this hypothesis at a significance level of 5%?

    (b) Calculate a 95% confidence interval for the difference between the proportion of people over 40 voting Conservative and the proportion of people below 40 voting Conservative.

11. A new process has been developed to reduce the level of corrosion of car bodies. Experiments were carried out on 11 cars using the new process and 11 cars using the old process. The average level of corrosion using the new process was 3.4 with a standard deviation of 0.5. The average level of corrosion using the old process was 4.2 with a standard deviation of 0.8.

    (a) Test the hypothesis that the mean of the level of corrosion does not depend on the process used. *For the time being, you may assume that assumption of equal variance is valid.*

    (b) Is there any evidence that the new process is better at a significance level of 5%?

    (c) Calculate a 95% confidence interval for the difference between the mean levels of corrosion under the two processes. Can it be stated that the mean level of corrosion is reduced by 1.5 at a significance level of 5%?

12. A researcher was investigating computer usage among students at a particular university. Three hundred undergraduates and one hundred postgraduates were chosen at random and asked if they owned a laptop. It was found that 150 of the undergraduates and 80 of the postgraduates owned a laptop.

    (a) Find a 95% confidence interval for the difference in the proportion of undergraduates and postgraduates who own a laptop.

    (b) On the basis of this interval, do you believe that postgraduates and undergraduates are equally likely to own a laptop?

13. The working lifetimes of 100 of both of two different types of batteries were observed. The mean lifetime for the sample of type 1 batteries was 25 hrs with a standard deviation of 4hrs. The mean lifetime for the sample of type 2 batteries was 23 hrs with a standard deviation of 3hrs.

    (a) Test the hypothesis that the mean working lifetimes of these batteries do not differ at a significance level of 5%.

    (b) Calculate a 95% confidence interval for the difference between the average working lifetimes of these batteries. Using this confidence interval, test the hypothesis that battery 1 on average works for 3 hours longer than battery 2.

14. A pharmaceutical Company wants to test Dozenol, a new cold medicine intended for night use. Tests for such products often include a treatment group of people who use the drug and a control group of people who dont use the drug. Fifty people with colds are given Dozenol, and 100 others are not. The systolic blood pressure is measured for each subject, and the sample statistics are given below.

Dozenol	Control Group
$n_1 = 50$	$n_2 = 100$
$\bar{x}_1 = 203.4$	$\bar{x}_2 = 189.4$
$s_1 = 39.4$	$s_2 = 39.0$

(i) Using a 0.05 level of significance, test the claim that Dozenol does not affect blood% pressure. Clearly state your null and alternative hypotheses and your conclusion.

15. An environmental group states that "*fewer than 60% of industrial plants comply with air pollution standards*". An independent researcher takes a sample of 120 plants and finds that 78 are complying with air pollution standards.

    (i) Carry out a hypothesis test to investigate the claim made by the environmental group. Clearly state your null and alternative hypotheses and your conclusion.

16. A test of a specific blood factor has been devised such that, for adults in Western Europe, the test score is normally distributed with mean 100 and standard deviation 10. A clinical research organization is carrying out research on the blood factor levels for sufferers of a particular disease.

    - A study has obtained the following test scores for 14 randomly selected patients suffering from the disease in Ireland

      $$\{115, 113, 111, 109, 115, 108, 121, 114, 104, 113, 122, 90, 103, 116\}$$

    - A similar study has obtained the following test scores for 15 randomly selected patients suffering from the disease in Germany .

      $$\{120, 137, 114, 120, 116, 118, 101, 110, 125, 113, 111, 128, 119, 117, 121\}$$

Sample	Mean	Std. Deviation	Sample Size
Ireland	111	8.143	14
Gemany	118	8.350	15

    You believe that there may be a difference in blood factor levels between the two countries. Test this hypothesis using a 5% level of significance. *For the time being, you may assume that assumption of equal variance is valid.*

    (a) Clearly state your null and alternative hypotheses.

    (b) Compute the pooled standard deviation and/or pooled variance.

    (c) Compute the test statistic.

    (d) Discuss your conclusion to this test, supporting your statement with reference to appropriate values.

17. The conventional treatment for a disease has been shown to be effective in 60% of all cases. A new drug is being promoted by a pharmaceutical company; the Department of Health wishes to test whether the new treatment is more effective than the conventional treatment.

    (a) Write down the null and alternative hypotheses for this problem.

    (b) A simple random sample of n. 400 patients suffering from the disease were given the new drug; the treatment was effective for 320 of them. What decision would you reach about the new drug? Give reasons for your decision.

A researcher was investigating computer usage among students at a particular university. Three hundred undergraduates and one hundred postgraduates were chosen at random and asked if they owned a laptop. It was found that 150 of the undergraduates and 80 of the postgraduates owned a laptop.

18. (a) Find a 95% confidence interval for the difference in the proportion of undergraduates and postgraduates who own a laptop.

    (b) On the basis of this interval, do you believe that postgraduates and undergraduates are equally likely to own a laptop?

19. The weight of 10 students was observed before commencement of their studies and after graduation (in kgs).

Student	1	2	3	4	5	6	7	8	9	10
Weight before	68	74	59	65	82	67	57	90	74	77
Weight after	71	73	61	67	85	66	61	89	77	83

    (i) Compute the mean and the standard deviation of the case-wise differences.

    (ii) Calculate a 95% confidence interval for the amount of weight that students put on during their studies.

    (iii) Test the hypothesis that the mean weight of students increases during their studies at a significance level of 5%.

20. A microbiologist measures the total growth in 24 hours of two strains of a germ culture in the same petri dish. Nine identical specimens are prepared. The growth rate for the eight samples for each strain are tabulated below:

Specimen	Strain 1	Strain 2
1	212	224
2	234	231
3	214	209
4	236	243
5	221	231
6	212	216
7	202	213
8	210	216
9	248	242

    At a significance level of 5%, is there sufficient evidence to state that there is any difference in growth rates between the two strains.

    (i) Formally state the null and alternative hypotheses.

    (ii) Compute the mean and standard deviation of the case-wise differences.

    (iii) Compute the test statistic.

    (iv) State the appropriate critical value for this hypothesis test.

(v) Discuss your conclusion to this test, supporting your statement with reference to appropriate values.

21. A simple random sample is conducted of 1486 college students who are near completion of a bachelor's degree in statistics. 802 students are female. Let π be the proportion of female students who are near completion of their bachelor degree in ' statistics.

    (i) Provide a point estimate of π.

    (ii) Calculate a 99% confidence interval for π.

    (iii) Use a 0.05 significance level to test the claim that the majority of college students who are near completion of a bachelors degree in statistics are female.
    Clearly state the null and alternative hypotheses and your conclusion.

22. A web-based software company claims that the average amount of time it takes for online queries to be dealt with is less than 2 hours. Out of a sample of 15 queries, the sample mean J? = 3.5 hours and the standard deviation is 30 minutes.

    (i) Construct the null and alternative hypothesis statements.

    (ii) Test this claim using a significance level of 0.05.

    (iii) Describe the two types of errors associated with hypothesis testing and how they relate to this question?

23. ABC Software has 125 programmers divided into two groups with 75 in Group A and 50 in Group B. In order to compare the efficiencies of the two groups, the programmers are observed for one day.

    - The 75 programmers of Group A averaged 76.21 lines of code with a standard deviation of 10.37.
    - The 50 programmers of Group B averaged 72.72 lines of code with a standard deviation of 10.07.

    (i) Using a significance level of 5%, test the hypothesis that there is no difference between the two groups versus the alternative that there is a difference. Clearly state your null and alternative hypotheses and your conclusion.

24. A manufacturer of a common cold cure claims that the product provides relief for 70% of people who use it. ln a test of 400 people, it was found that 300 people said the treatment provided relief.

    (i) Calculate a 95% confidence interval for the true proportion of people who would get relief from the product.

    (ii) Suppose the manufacturer wishes to be 95% confident that the prediction is correct to within 2% of the true proportion. What sample size is needed?

    (iii) Using a significance level of 5%, test the hypothesis that more than 70% of people who use the product find relief.
    Clearly state your null and alternative hypotheses and your conclusion.

25. A study was carried out to compare two treatments for the flu. A total of 500 newly diagnosed flu patients were randomly assigned to one of the two treatments.

    - Of the 280 assigned to the first treatment, 168 still had the flu after 2 days after diagnosis.

- Of the 220 assigned to the second treatment, 176 still had the flu after 2 days after diagnosis.

Let p_l denote the probability that a flu patient assigned to the first treatment will still have the flu after 2 days after diagnosis; let p_2 denote the corresponding probability for the second treatment.

26. A company organizes two evening courses, one on stock market trading and one on spread-betting. From a sample of 100 clients, 64 of them choose the stock market course.

    (i) Calculate a 95% confidence interval for the percentage of clients who prefer the stock market course. (5 marks)

    (ii) lt is known that in past years the percentage of clients who preferred the stock market trading classes was 75%. Should the company conclude that the interest in the stock market trading courses has decreased significantly in comparison to past years? Clearly state H_0 and H_1.

    (i) Provide an estimate of the difference between the population proportions (i.e. $\pi_l - \pi_2$).

    (ii) Calculate a 95% confidence interval for the difference between the population proportions.

    (iii) Use a 0.05 significance level to determine if there is a difference between the two proportions. [Clearly state the null and alternative hypotheses and your conclusion].

27. The height of 100 Americans and 50 Spaniards was observed. The mean and standard deviation of the height of the Americans was 172cm and 13cm, respectively. The mean and standard deviation of the height of the Spaniards was 167cm and 12cm, respectively.

    (i) Calculate a 95% confidence interval for the difference between the mean height of all Americans and the mean height of all Spaniards.

    (ii) Without doing any further calculations, test the hypothesis that the mean height of all Americans is equal to the mean height of all Spaniards. Give a brief justification of your conclusion. What is the significance level of this test'?

28. The mean and standard deviation of the salaries of 16 Irish full-time workers are 5000 and 3000 Euros, respectively.

    (i) Test the hypothesis that the mean salary of all Irish full-time workers is E4000 at a significance level of 5%.

    (ii) What assumption is made in this testing procedure? Is this assumption reasonable?

29. (Not using for MS4222 2018).
    A survey of 1000 Irish indicates that 750 have access to the Internet. A survey of 2000 Spaniards indicates that 1400 have access to the Internet.

    (i) By calculating the appropriate p-value, test the hypothesis that the proportion of all Irish having access to the Internet is equal to the proportion of all Spaniards having access to the internet at a significance level of 5

    (ii) Calculate a 99% confidence interval for the difference between the proportion of all Irish having access to the Internet and the proportion of all Spaniards having access to the internet.

30. (Not using for MS4222 2018).
Let π be the proportion of workers in Ireland who spend at least one hour per day in front of a computer terminal. Suppose that a researcher is going to take a sample of n workers and estimate π using $\hat{p}$, the proportion of workers in the sample who spend at least one hour per day in front of a computer terminal.

a. (1 mark) How large should n be if the researcher wants to be 90% certain that his error is less than 0.01?