GRIFFITH COLLEGE DUBLIN

Assignment Cover Sheet

Student name:	**Saujanya Bohara**
Student number:	**2892141**
Faculty:	**Computing Science**
Course:	**Bsch** **Stage/year:** **2**
Subject:	**Probability and statistics**
Study Mode:	Full time **yes** Part-time _____
Lecturer Name:	**Paddhy Fahy**
Assignment Title:	**Assignment 04**
No. of pages:	**6**
Disk included?	Yes _____ No _____
Additional Information:	(ie. number of pieces submitted, size of assignment, A2, A3 etc)
Date due:	
Date submitted:	

Plagiarism disclaimer:

I understand that plagiarism is a serious offence and have read and understood the college policy on plagiarism. I also understand that I may receive a mark of zero if I have not identified and properly attributed sources which have been used, referred to, or have in any way influenced the preparation of this assignment, or if I have knowingly allowed others to plagiarise my work in this way.

I hereby certify that this assignment is my own work, based on my personal study and/or research, and that I have acknowledged all material and sources used in its preparation. I also certify that the assignment has not previously been submitted for assessment and that I have not copied in part or whole or otherwise plagiarised the work of anyone else, including other students.

Signed: _____ **Date:** _____

Question 1

(a) Why is sampling necessary?

Answer:

Many things in this world would be impossible without sampling. For conducting large studies and implementing relevant factors we need sampling. For e.g. development activities in countries are carried out depending upon the population density and to know the density of population we need sampling. In one or other ways sampling is a vital factor in all sectors where a study is needed for large numbers. Sampling is a way of answering questions that deals with large numbers.

(b) What is random sampling? What are the characteristics of a random sample?

Answer:

It is the basic sampling technique where we select a group of subjects (a sample) for study from a larger group. Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample. There are two types of random sampling they are Sampling with Replacement and Without Replacement.
The characteristics of random sampling are discussed below:

- It is easy to conduct.
- High probability of achieving a representative sample.
- Meets assumptions of many statistical procedures.
- Identification of all members of the population can be difficult.
- Contacting all members of the sample can be difficult.

Question 2

What is:
(a) Systematic sampling?

Answer:

Systematic sampling is a type of probability sampling method in which sample members from a larger population are selected according to a random starting point and a fixes periodic interval. This interval, called the sampling interval, is calculated by dividing the population size by the desired sample size. Despite the sample population being selected

in advance, systematic sampling is still thought of as being random if the periodic interval is determined beforehand and starting point is random.

(c) Convenience sampling?

Answer:

It is a non-probability sampling technique where subjects are selected because of their convenient accessibility and proximity to the researcher.

Question 3

(a) What is a sampling frame? Explain, with the help of an example, how poor selection of a sampling frame can introduce bias.

Answer:

Sampling frame is a complete list of all the members of the population that we wish to study. The poor selection of sampling frame will lead to bias an example to support the statement.
Let's consider a sample is done to check if the people living at Terenure go to Aldi or Lidl for shopping. Since, there is no Lidl near Terenure but there is big store of Aldi. If we take sampling here, then it will be Aldi for sure because of the convenience.

We will get results that will show people prefer to go to Aldi than Lidl but the result in a large and fair scenario might be different. Hence, poor selection of sampling frame will lead to bias.

(b) Explain why a population mean might differ from a sample mean.

Answer:

Population mean
When the provided list represents a statistical population, then the mean is called the population mean. It is usually denoted by "μ."
Sample mean
When the provided list represents a statistical sample, then the mean is called sample mean. The sample mean is denoted by "X".

Population is the whole set of values whereas, sample is the subset of population and hence population and sample are two different things so the mean of them differ with each other.

Question 4

(a) Why is hypothesis testing necessary? What precisely is the objective of a hypothesis test?

Answer:

Hypothesis testing is necessary because it evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data. The main objective of hypothesis test is that it is concerned with how to infer population characteristics from samples and it allows us to tell the significance of a result.

(b) A machine fills packets of peanuts where it is stated on the packet that the mean weight is 40 grams. A random sample of 50 packets is taken and the mean weight is found to be 38.9 grams with a standard deviation of 2 grams. Carry out a significance test at the 5% level.

Answer:

This problem could be viewed from the point of view of the manufacturer or the customer. The manufacturer doesn't not want the packets overfilled because this would result in less profits. Neither does the manufacturer want under filled packets because this would result in customer complaints and possible damage to the reputation of the business. The customer however, is only concerned if the packet is underweight

If one takes manufacturer's view and decides on a two-tailed test, this determines the formulation of the null and alternative hypothesis

4 Steps to carry out two tailed test

1. First, state the null and alternative hypothesis:

H0 : $\mu = 40g$
HA: $\mu \neq 40g$

2. Find the standard error of the mean:

$Sx = s/\sqrt{n}$

$Sx = 2/\sqrt{50} = 0.282$

3. The acceptance range is:

$40 \pm 0.282 * 1.96$
40 ± 0.554

4. Conclusion:

The sample mean at 38.9g is outside the acceptance zone. Reject the null hypothesis and accept HA: $\mu \neq 40g$

Question 5

(a) In relation to significance testing, what is a one-tailed test? Give an example of where a one-tailed test might be used.

Answer:

A one tailed test is a statistical test in which the critical area of a distribution is one sided so that it is either greater than or less than a certain value but not both. If the sample that is being tested falls into the one-sided critical area, the alternative hypothesis will be accepted instead of the null hypothesis.

E.g.
School District A states that it's high schools have an 85% passage rate on the High School Exit Exam. A new school was recently opened in the district, and it was found that a sample of 150 students had a passage rate of 88%, with a standard deviation of 4%. Does this new school have a higher passage rate than the rest of School District A?

Here we are asked to check if the passage rate is higher or not in this scenario we use one tailed test but if the question was "Does this new school have a different passage rate then we would use two tailed-test". In simple words we can say that when we have a specific hypothesis about the direction of the relationship we use one tailed test.

(b) A car manufacturer claims that its cars are capable of travelling 9 kilometres per litre of fuel. Test runs with 90 cars result in an average distance travelled per litre of fuel of only 8.2 kilometres, with a standard deviation of 2.5 kilometres. Carry out a one tailed test at the 5% level to test the manufacturers claim.

Answer:

4 steps(1-tailed):

1: First, state the null and alternative hypothesis:

H0 : $\mu = 9km$
HA: $\mu < 9km$

2**: Find the standard error of the mean:**

$Sx = s/\sqrt{n}$

$Sx = 2.5/\sqrt{90} = 0.2635$

3: The acceptance range is:

= 9 – 0.2635 * 1.65

= 8.56

4. Conclusion:

The manufacturer is only concerned if his cars cannot travel 9 kilometres per litre of fuel. Our acceptance zone is from 8.56 to 9 and 8.2 does not lies on that so we accept the alternative hypothesis.

Reference:

https://moodle.griffith.ie/pluginfile.php/328112/mod_resource/content/3/significance_testing1.pdf

https://www.tutorialspoint.com

http://blog.minitab.com/blog/adventures-in-statistics/understanding-hypothesis-tests:-why-we-need-to-use-hypothesis-tests-in-statistics

http://www.differencebetween.net/science/difference-between-sample-mean-and-population-mean/