Last updated on February 9, 2025

BM5163 Bayesian Inference in Bioengineering

Problem Set 0

Instructions

1. You are not expected to submit answers to these problems

Questions

1. Consider the following data obtained from 4 repetitions of an experiment measuring the rate of cell death due to an anti-cancer drug

Starting cell count	Time taken	Rate of cell death
10^6	2.0hrs	5.0×10^5 cells/hr
5×10^6	8.0hrs	6.3×10^5 cells/hr
2×10^6	4.0hrs	5.0×10^5 cells/hr
2×10^6	5.0hrs	4.0×10^5 cells/hr

2. The central limit theorem says that the sampling distribution of the mean will always be normally distributed, as long as the sample size is large enough (under certain conditions on sample size and number of repetitions etc.). The mean of this distribution is the same as that of the population and the standard deviation of this distribution is equal to $\sigma/\sqrt{n}$ (where σ is the population standard deviation, and n is the sample size) and is also called *standard deviation of the mean* or *standard error of the mean* or, sometimes, only *standard error*. Since in most situations, we do not know the population variance beforehand and have to estimate it from the sample itself. In that condition, the best estimate of the standard error is $s/\sqrt{n}$ where s is the standard deviation of the sample.

   Consider the dataset given below as a sample of measurements for systolic blood pressure (in mmHg)

   $$121, 125, 128, 134, 136, 138, 139, 141, 144, 145, 149, 151$$

   (a) Calculate the *standard error of the mean* from this sample.

   (b) What is the probability of obtaining a sample with a mean value > 146 for such a population? *Enumerate* all the assumptions you make for this calculation.

3. Suppose you are to perform an experiment to study the effect of nanoparticle based drug delivery system. For this you plan to compare the number of dead cells in cell culture system with and without the designed delivery method. Suppose the average (population) number of dead cells with and without the delivery method are μ_d and μ_c, respectively. Therefore, you aim to test for $H_0 : \mu_c = \mu_d$ at 5% significance level with a 95% probability of rejecting H_0 when $|\mu_c - \mu_d|$ is at least 100. The estimate of the population variance, σ^2, is $s^2 = 50$.

   (a) What minimum sample size should be used?

   (b) What sample size will be required if $\alpha = 0.01$?

   (c) What sample size will be needed if you desire $\alpha = 0.05$ and statistical power of 0.99.

   (d) If $n = 5$ and $\alpha = 0.05$, what is the smallest difference, $|\mu_c - \mu_d|$, that can be detected with 95% probability?

   (e) If $n = 5$ and $\alpha = 0.05$, what is the probability of detecting a difference, $|\mu_c - \mu_d|$, as small as 20?

   Note: You can use R to answer this question.

4. A species of marine arthropod lives in seawater that contains calcium in a concentration of 32 mmole/kg of water. Thirteen of the animals are collected and the calcium concentrations in their coelomic fluid are found to be: 28, 27, 29, 29, 30, 30, 31, 30, 33, 27, 30, 32, and 31 mmole/kg.

(a) State the null and alternate hypotheses if you are tasked to test whether members of this species maintain a coelomic calcium concentration less than that of their environment.

(b) Perform the test using R to conclude if H_0 can be rejected.

(c) Write down the null and alternate hypotheses if you seek to test the claim that this species maintains a coelomic calcium level less than that of their environment.

(d) Perform the test using R .

5. Using the data in the following table, you seek to test if male and female turtles have the same mean serum cholesterol concentrations.

(a) State the null and alternate hypotheses.

(b) Are data normally distributed.

(c) Which test are you going to perform. Justify your answer.

(d) Perform the test using R and report the outcome.

Male	Female
220.1	223.4
218.6	221.5
229.6	230.2
228.8	224.3
222.0	223.8
224.1	230.8
226.5	

Table 1: Serum Cholesterol (mg/100 ml)

6. In a hypothetical study to look at the relative effect of coffee and tea in exam performance of students, a group of students were given two sets of exams. One cup (150mL) of tea and coffee were consumed by each student before each exam. The marks obtained by the students after two exams are

Exam after coffee	**Exam after tea**
72	75
64	64
81	79
90	85
73	82
54	90
87	90

Table 2: Marks obtained by the students out of 100.

(a) Which statistical test will be best suited to settle the question?

(b) What are the assumptions you need to check before you perform the test ?

(c) Perform the test using R and report your results.