Indian Institute of Technology Madras

Department of Data Science and Artificial Intelligence

DA5000: Mathematical Foundations of Data Science

Tutorial VIII

Problem

1. Consider the following hypothesis test:

$$H_0 \geq 28$$

$$H_1 < 28$$

Assume that $\sigma^2 = 10000$ and sample size is 100.

   (a) Find the probability of type-2 error assuming $\mu = 25$.

   (b) Find the probability of type-2 error assuming $\mu = 30$.

   (c) Find the probability of type-1 error in the above case.

   Assume significance level of the test is 0.05

2. In a clinical trial, a new medication is tested for its effectiveness compared to a placebo. The null hypothesis (H_0) states that the medication has no effect, while the alternative hypothesis (H_1) states that the medication does have an effect.

   (a) The significance level (α) is set to 0.05. If the null hypothesis is rejected when it is actually true, what is the probability of committing a Type I error? Calculate this probability.

   (b) In this trial, the power of the test is determined to be 0.80. If the alternative hypothesis is true and the null hypothesis is not rejected, what is the probability of committing a Type II error (β)? Calculate this probability.

   (c) If the sample size is increased, discuss how this affects the probabilities of Type I and Type II errors.

   (d) If the significance level (α) is reduced, what effect does this have on the confidence interval (CI)? Will the CI increase or decrease?

   (e) Find the effect on Precision and Recall of the test when:

      i. α increases keeping beta constant

      ii. β increases keeping alpha constant

3. A machine fills milk bottles, the mean amount of milk in each bottle is supposed to be 32 Oz with a standard deviation of 0.06 Oz. Suppose the mean amount of milk is approximately normally distributed. To check if the machine is operating properly, 36 f illed bottles will be chosen at random and the mean amount will be determined.

   (a) If an $\alpha = .05$ test is used to decide whether the machine is working properly, what should the rejection criterion be?

   (b) Find the power of the test if the true mean takes on the following values: 31.97, 31.99, 32, 32.01, 32.03. Draw the power curve.

   (c) Find the probability of a type II error when the true mean is 32.03.

   Note : The power is the probability of rejecting H0 given that the true value of the parameter being tested is some specified value. power= α when H0 is true, power= 1 - β for a particular alternative.

4. A new diagnostic test is designed to detect a rare disease. A test result of positive indicates the presence of the disease, while a negative result suggests its absence. The hypotheses are as follows:

$$H_0 : \text{The patient does not have the disease.}$$

$$H_1 : \text{The patient has the disease.}$$

   (a) Define Type-I and Type-II errors in this context and explain the potential impacts of each error on the patient.

   (b) Given the rarity of the disease, why might the designers of the test prioritize minimizing the Type-I error? What implications does this have for the probability of Type-II error?

   (c) Suppose the probability of a Type-I error (false positive) is 0.01. If the probability of a Type-II error (false negative) is 0.2, what is the test's power?

   (d) If the prevalence of the disease is very low (e.g., 0.1%), discuss how Type-I and Type-II errors might impact the test's overall reliability. How would you interpret a positive result?

5. A standard insecticide is used to control a particular insect pest in soybeans. The probability that this insecticide eliminates the insect from an individual plant infested with the insect is 0.7. A new insecticide is developed which is known not to decrease the probability of insect elimination; it is desired to determine if the new insecticide performs better than the standard. An experiment will be conducted in which 500 randomly selected plants (known to be infested with the insect) are sprayed with the new insecticide. After a fixed period of time, the plant for which the new insecticide eliminates the insect will be counted. The experiment will reject the null hypothesis that the insecticide has no effect if the insecticide eliminates the insect in 375 or more plants.

   (a) State symbolically the null and alternative hypotheses, defining all symbols in words.

   (b) What is the α-level of the proposed test?

   (c) If in fact the success rate of the insecticide is 0.8, what is the power of the test?

   (d) When the experiment was actually carried out, it was found that the new insecticide eliminates the insect in 395 plants. Find the p-value of the test.