AI1110: Probability and Random Variables

Assignment 11: Papoulis-Pillai Ex 8-28

Suruaansh Jain (cs21btech11057)

June 16, 2022

Outline

Problem

Brand A batteries cost more than brand B batteries. Their life lengths are two normal and independent random variables x and y. We test 16 batteries of brand A and 26 batteries of brand B and find these values in hours: $\bar{x} = 4.6$ $s_x = 1.1$ $\bar{y} = 4.2$ $s_y = 0.9$ Test the hypothesis $\eta_x = \eta_y$, against $\eta_x > \eta_y$, with $\alpha = 0.05$

Percentile

Given a distribution on random variable x as $F_x(x)$, we define the k^{th} percentile of this distribution as

$$Percentile_k = F_x^{-1}(k) \tag{1}$$

In other words, the k^{th} percentile returns the value of random variable x_0 for which $F_x(x_0) = k$.

Hypothesis

Let us assume, we are given a random variable x whose distribution is $F(x, \theta)$ depending on some parameter θ (The parameter might be mean, variance etc.). We are required to use evidence that either supports or rejects a given prediction of the actual value of θ, which we will call θ_0.

Null Hypothesis

In the null hypothesis, we make the prediction that $\theta = \theta_0$. This is represented by $H_0 : \theta = \theta_0$.

Alternate Hypothesis

In the alternate hypothesis, we make the prediction that $\theta \neq \theta_0$. This is represented by $H_1 : \theta \neq \theta_0$. Note that the null hypothesis may be defined differently based on utility.

Testing Hypothesis

To test whether a given hypothesis is feasible based on evidence, we first define a random variable q whose density is convenient to plot and is a function of sample vector X as follows.

$$q = g(X) \tag{2}$$

We will call q as the test statistic.
The density of random variable q is given by $p_q(q, \theta)$ where θ is the parameter. Now consider the density $p_q(q, \theta_0)$ (based on H_0) and a region (Critical Region) R_c where $p_q(q, \theta_0)$ is negligible. If we find that the value of q lies in R_c, then we reject H_0.
One can decide the region R_c using the significance level α. α represents the probability that $q \in R_c$ when H_0 is true. Hence, when given a value of α one can determine R_c and thereby check the validity of the null hypothesis.

Mean as Parameter: Unknown Variance

Consider a random variable x, from which we have obtained a sample vector X. We are required to reject or support the hypothesis $H_0 : \eta = \eta_0$ against $H_1 : \eta \neq \eta_0$, where we check if the mean η equals a constant η_0.
In the case that the variance is unknown but the sample mean $\bar{x}$ and sample variance s^2 are given, we must use a Student t distribution. Note that the sample vector X has $n - 1$ degrees of freedom as we are constrained to ensure that the sum of the values of the vector $X - \bar{x}$ must be 0.
Assuming random variable $\bar{x}$ is represented by a normal distribution, we define test statistic q as follows:

$$q = \frac{\bar{x} - \eta_0}{s/\sqrt{n}} \tag{3}$$

Mean as Parameter: Unknown Variance

For an alternate hypothesis $H_1 : \eta \neq \eta_0$ and given significance value α, we note that the critical region R_c is given by:

$$R_c = (t_{1-\alpha}(n-1), \infty) \tag{4}$$

where $t_k(n-1)$ represents the k^{th} percentile (As explained in (1))
We consider the α^{th} and its complementary percentile as the given hypothesis is single ended, i.e., it allows us accept values less than the hypothesised mean value.

Stating the Transformation

Let,

$$w = \bar{x} - \bar{y} = 0.4 \tag{5}$$

$$\implies \sigma_w^2 = \frac{\sigma_x^2}{16} + \frac{\sigma_y^2}{26} = 0.32 \tag{6}$$

Stating the Hypothesis

We state the null Hypothesis as

$$H_0 : \eta = 0 \tag{7}$$

and the alternate hypothesis as

$$H_1 : \eta > 0 \tag{8}$$

We are required to test the above hypotheses for significance value $\alpha = 0.05$

Calculate Test Statistic q

Given sample mean $\bar{x} = 4.35$ and sample variance $s = 0.32$, we get our test statistic q from (3) as

$$q = \frac{0.4 - 0}{0.32} = 1.223 \tag{9}$$

Making Decision

We shall determine the critical regions for given significance values α_1 and α_2 using (4)
For $\alpha = 0.005$, we find that $t_{0.995} = 1.64$.As $q < 1.64$ we can accept this hypothesis.