AI1110: Probability and Random Variables

Assignment 9: Papoulis-Pillai Ex 8-27

Rishit D (cs21btech11053)

IIT Hyderabad

June 13, 2022

Outline

Problem

The weights of cereal boxes are the values of a random variable x with mean η. We measure 64 boxes and find that $\bar{x} = 7.7oz$ and $s = 1.5oz$. Test the hypothesis $H_0 : \eta = 8oz$ against $H_1 : \eta \neq 8oz$ with $\alpha = 0.1$ and $\alpha = 0.01$.

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:

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

We see that the random variable q obtains a Student t-distribution with $n-1$ degrees of freedom.

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:

$$R_c = (-\infty, t_{\alpha/2}(n-1)) \cup (t_{1-\alpha/2}(n-1), \infty) \tag{4}$$

where $t_k(n-1)$ represents the k^{th} percentile of the standard Student t-distribution. (As explained in (1))
We consider the $\alpha/2^{th}$ and its complementary percentile as the given hypothesis is double ended, i.e., it allows us accept values both slightly greater or less than the hypothesised mean value.

Student t-distribution

The Student t-distribution is used to estimate the mean of a normal distribution when it's variance is unknown. Given n observations in a sample, the t-distribution (with $n-1$ degrees of freedom) represents the sample mean with respect to the total mean.
The density of the Student t-distribution (with n degrees of freedom) is given by

$$p_x(x) = \frac{1}{\sqrt{n\pi}\beta(n/2, 1/2)}(1 + \frac{x^2}{n})^{\frac{-(n+1)}{2}} \tag{5}$$

Stating the Hypothesis

We state the null Hypothesis as

$$H_0 : \eta = 8 \tag{6}$$

and the alternate hypothesis as

$$H_1 : \eta \neq 8 \tag{7}$$

We are required to test the above hypotheses for significance values $\alpha_1 = 0.1$ and $\alpha_2 = 0.01$.

Calculate Test Statistic q

Given sample mean $\bar{x} = 7.7$ and sample variance $s = 1.5$, we get our test statistic q from (3) as

$$q = \frac{7.7 - 8}{1.5/\sqrt{64}} = -1.6 \tag{8}$$

This represents the value of random variable q with a Student t-distribution with mean 0 and $n - 1 = 63$ degrees of freedom.

Making Decision

We shall determine the critical regions for given significance values α_1 and α_2 using (4)
For $\alpha_1 = 0.01$, we find that $t_{0.005}(63) = -2.656$ and $t_{0.995}(63) = 2.656$. Hence critical region R_{c1} is

$$R_{c1} = (-\infty, -2.656) \cup (2.656, \infty) \tag{9}$$

Similarly for $\alpha_2 = 0.1$, we find that $t_{0.05}(63) = -1.669$ and $t_{0.95}(63) = 1.699$. Hence critical region R_{c2} is

$$R_{c2} = (-\infty, -1.699) \cup (1.699, \infty) \tag{10}$$

Conclusion

Note that using (8), (9), (10) we find that $q \notin R_{c1}$ and $q \notin R_{c2}$.
As a result, we can conclude that the evidence gathered does not reject the null hypothesis for both given significance values.