Experiment no 5 Title of Experiment: Hypothesis Testing

Abstract: Inferential Analysis can further made with the help of estimation of confidence region and hypothesis and can be represented using the three analysis methodologies.

- Numerical Analysis
- Graphical Analysis
- Inferential Analysis

Hypothesis testing is important for determining if there are statistically significant effects. Summary statistics, plots, histograms and practical considerations should be used for analyzing the result.

Introduction: Hypothesis testing is used to draw conclusions on the data and finally to arrive at a decision to test null and alternate hypothesis.A statistical hypothesis is an assumption made by the researcher about the data of the population collected for any experiment which can have either a positive or a negative value. In order to test hypothesis, it will consider the entire population into account. To test or validate a hypothesis, random samples from the population are collected. On the basis of the result from testing over the sample data, it either selects or rejects the hypothesis. The types of Hypothesis are the Null and Alternate Hypothesis.**Null Hypothesis** implies that there is no correlation between two variables.**Alternate Hypothesis** This hypothesis describes a correlation between two variables.

Essentials of Hypothesis Testing are as follows: p-value and the decision rule. Commonly an *alpha* value of 0.05 is used.

- If the *p*-value for the test is less than *alpha*, we reject the null hypothesis.
- If the *p*-value is greater than or equal to *alpha*, we fail to reject the null hypothesis.
- A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject it.
- A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject it.

Statistical Mean and Standard Deviation

The mean of a statistical distribution with a continuous random variable, also called the expected value, is obtained by integrating the product of the variable with its probability as defined by the distribution.

The expected value is denoted by the lowercase Greek letter mu (μ)An arithmetic mean is calculated using the formula $\mu = \frac{1}{n}\sum_{i=1}^{n} a_i$

The Standard Deviation is a measure of how spread out samples or numbers. It measures variability of the data. Its symbol is σ (the greek letter sigma).The formula is the square root of the Variance

Formula for Standard Deviation: $$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Experiment

Suppose the mean weight of crop Sugarcane found in an Konkan region in last year (2019) was 1.5 kg . In a sample of 22 sugarcane crop same time for current year (2020) in the same Konkan region, the mean sugarcane weight is 1.3 kg. Assume the population Standard Deviation (SD) is 1.7 kg. At .05 significance level, can we accept or reject the null hypothesis that the mean sugarcane weight does not differ from last year (2019).

Solution

State the Hypotheses

Ho : The mean weight of crop Sugarcane found in an Konkan region for last year 2019 was : 1.5 kg

H1 : Current year(2020) in the same Konkan region, the mean sugarcane weight for 22 sugarcane crop was: 1.3 kg.

Computing the test statistic

$\mu = \frac{1}{n}\sum_{i=1}^{n} a_i = 1.5$ .... Step-1 - last year mean (2019)

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} a_i = 1.3$ ---------------Step-2 - current year sample mean(2020)

$\mu_0 = 1.5$ -------------- Step-3 - hypothesized value

$\sigma = 1.7$-------------- Step-4 - standard deviation of population

n = 22 --------------Step-5 - sample size

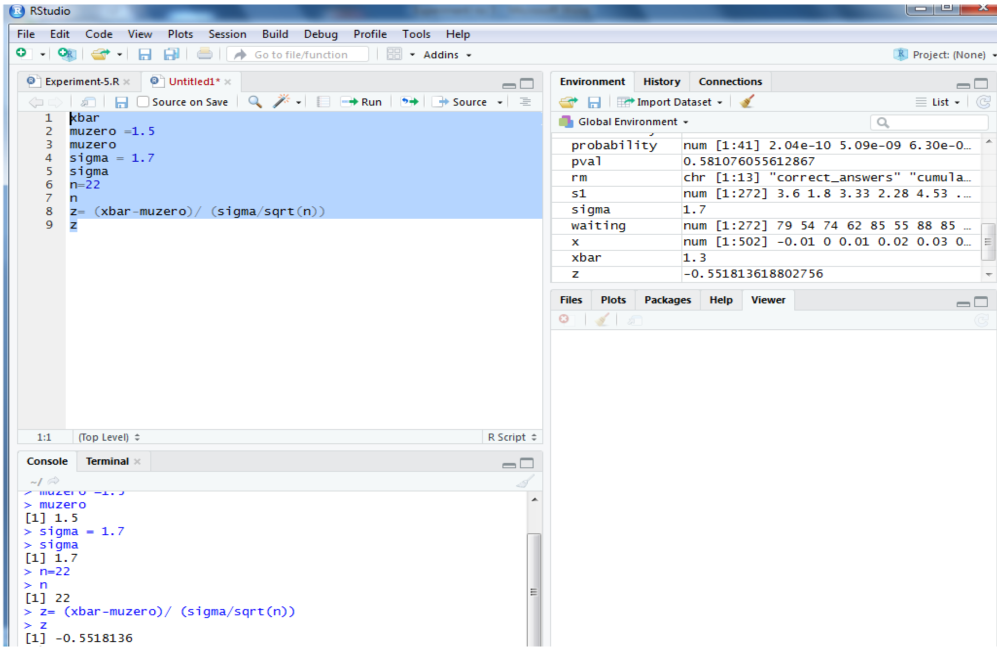
$z = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$ **---------------**Step-6 test statistic value

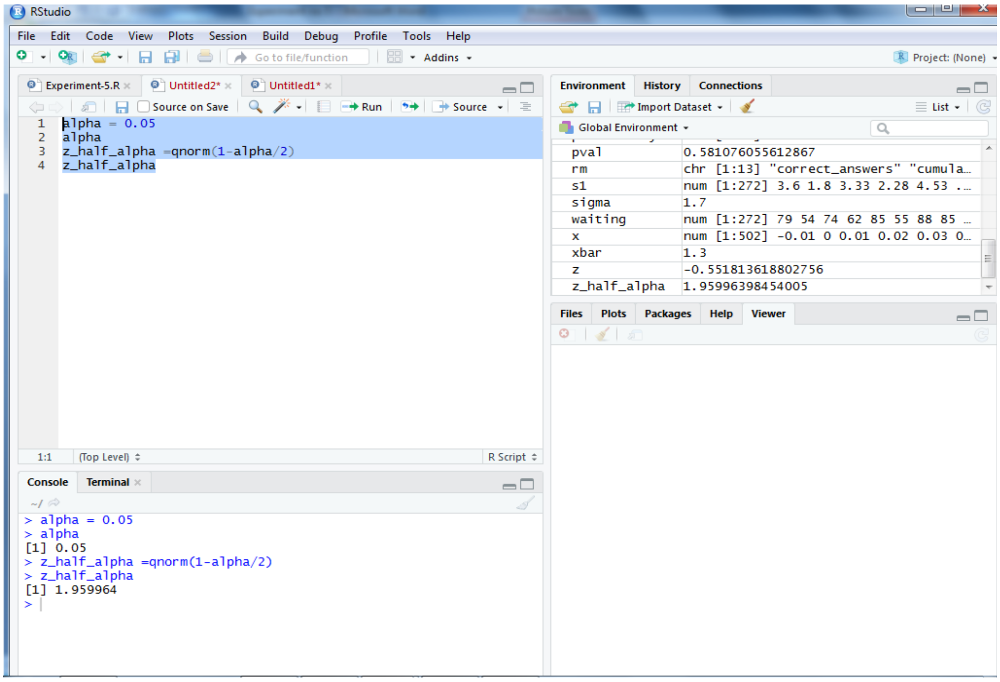
z = -0.5518136 ……….. Computed the critical values at .05 significance level

α = 0.05, z.1/α = qnorm(1−α/2) c (-z.1/α , z.1/α) **-1.959964 1.959964**

Numerical Analysis with R

(#windows 7 32 bit R version 3.6.1 and RStudio-1.1.463)





Inferential Analysis with R

(#windows 7 32 bit R version 3.6.1 and RStudio-1.1.463)

The test statistic -0.5518136 lies between the critical values -1.959964 and 1.959964. Hence at .05 significance level.

Interpretation

We accept the null hypothesis that the mean sugarcane weight does not differ from last year (2019) to current year (2020).

Alternative Solution

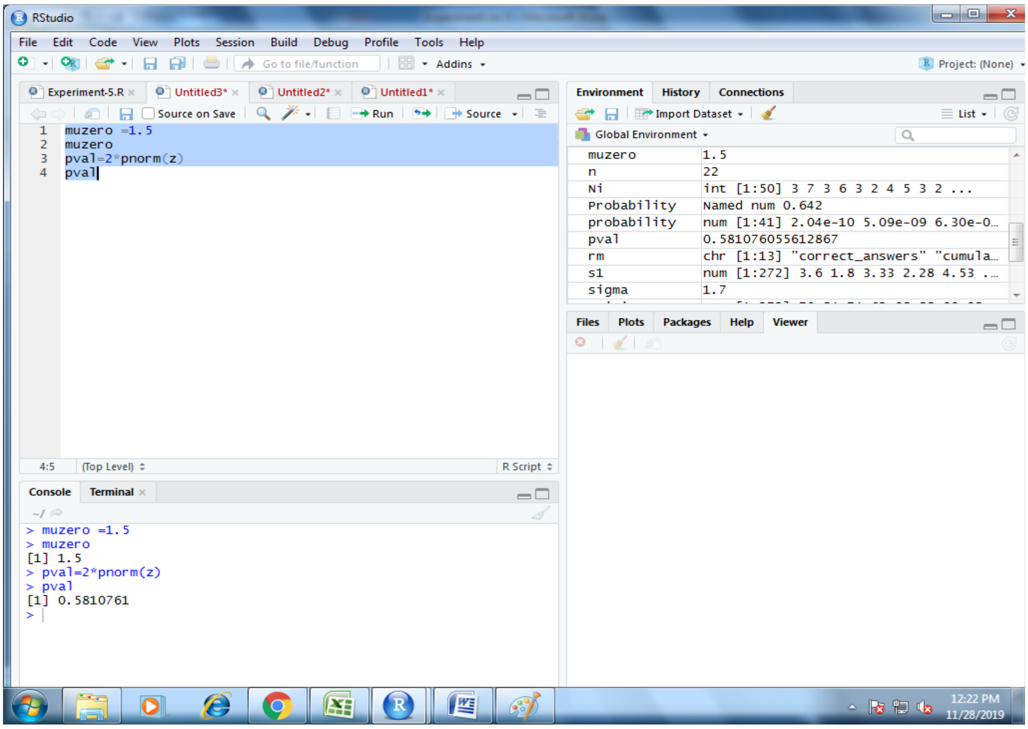
Instead of using the critical value, we apply the pnorm function to compute the two-tailed p-value of the test statistic. **It doubles the lower tail p-value as the sample mean is less than the hypothesized value. Since it turns out to be greater than the .05 significance level, we do not reject the null hypothesis**

μ = 1.5 ----step-1

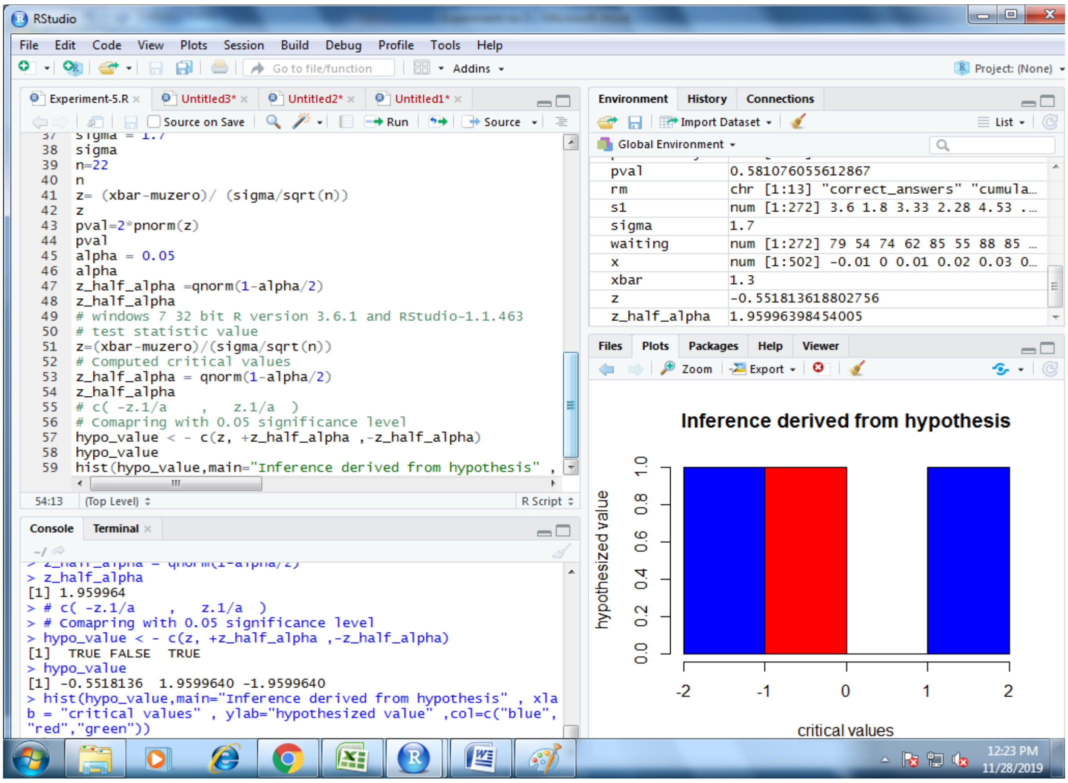
pval = 2 ∗ pnorm(z) ----step-2 --lower tail

>pval ----step-3 --two−tailed p−value

[1] **0.5810761**



Graphical Analysis using R



Here we have compared statistical value with computed critical (hypothesized) value and histogram seems to be same for both the values. So statistical Inferential derives that the estimation of confidence region for crop sugarcane mean weight found in a Konkan region does not differ from last year (2019) to current year (2020).So, the Null hypothesis is accepted.