SME. Final work

Inferential Statistics.

Sampling distributions and confidence intervals

Name: Jon Echevarria Sasia

January 10, 2024

Table of contents

Contents

Introduction

In this laboratory you will have a statement and you have to solve it using R, Rmarkdown and LaTeX.

All the necessary formulas will we written in LaTeX. And the answers will written using the R and Rmarkdown commands required. ($FORMULA = OUTPUT\ FROM\ CODE$)

You have to follow the index given and add the explanations required.

The statement

An industrial engineer has designed a machine to pack three-pound potato bags. However, due to various reasons, such as different weights of potatoes, filling problems, etc. is aware that the final weight of the potato bag won‘t be exactly three kilos, but there will be random variations in this quantity.

Read the data and describe them

1. The data of the sample comes in the following file: *potato_bags.csv*

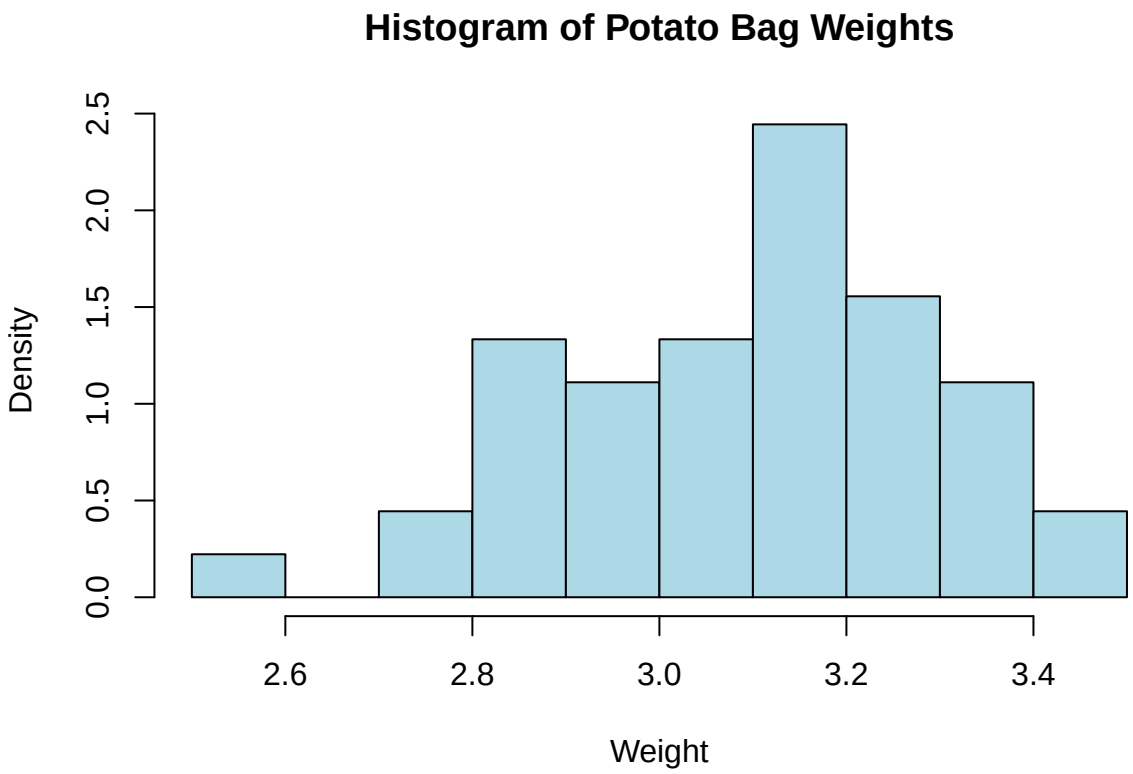
QUESTIONS

How many data are in the fille? 45 How many variables? 1

2. Find the mean, variance, standard deviation of the data in the sample

The mean of the sample is $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n} = 3.08$,
the variance is $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.037$,
and the standard deviation is $\sigma = \sqrt{Var(x)} = 0.18$.

3. Draw the histogram of the data showing the density not the frequencies.


Histogram of Potato Bag Weights
Density
0.0
0.5
1.0
1.5
2.0
2.5
2.6
2.8
3.0
3.2
3.4
Weight

Sampling distribution of the Sample Mean X_n

Find the expected value of the Sample Mean and its standard error.

$E[X_n] = \sum_{i=1}^{n} x_i * P(x_i) = 3.08$

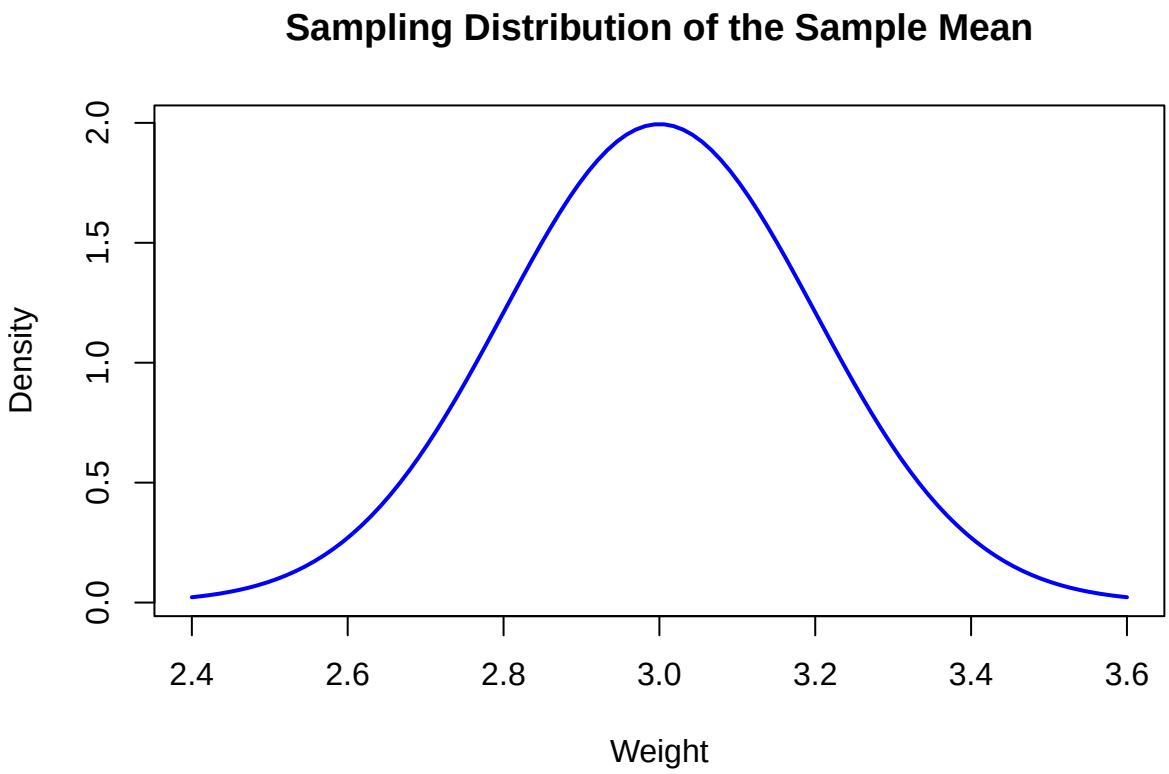
and standard error $SD(X_n) = \sqrt{Var(x)} = 0.03$

Which distribution follows the Sample Mean and why?

ANSWER:

>The sampling distribution of the sample mean follows a normal distribution. This is based on the Central Limit Theorem, which states that the distribution of the sample mean approaches a normal distribution as the sample size increases.

Draw the curve supposing the original population has mean $\mu = 3$ and $\sigma = 0.2$.


Sampling Distribution of the Sample Mean
Density
0.0
0.5
1.0
1.5
2.0
2.4
2.6
2.8
3.0
3.2
3.4
3.6
Weight

Confidence interval for the mean

1. Which distribution do you use for finding the probability of the confidence interval? Write the formula for the scores.

ANSWER:
We use the t-distribution to find the probability of the confidence interval. The formula for the t-scores is:

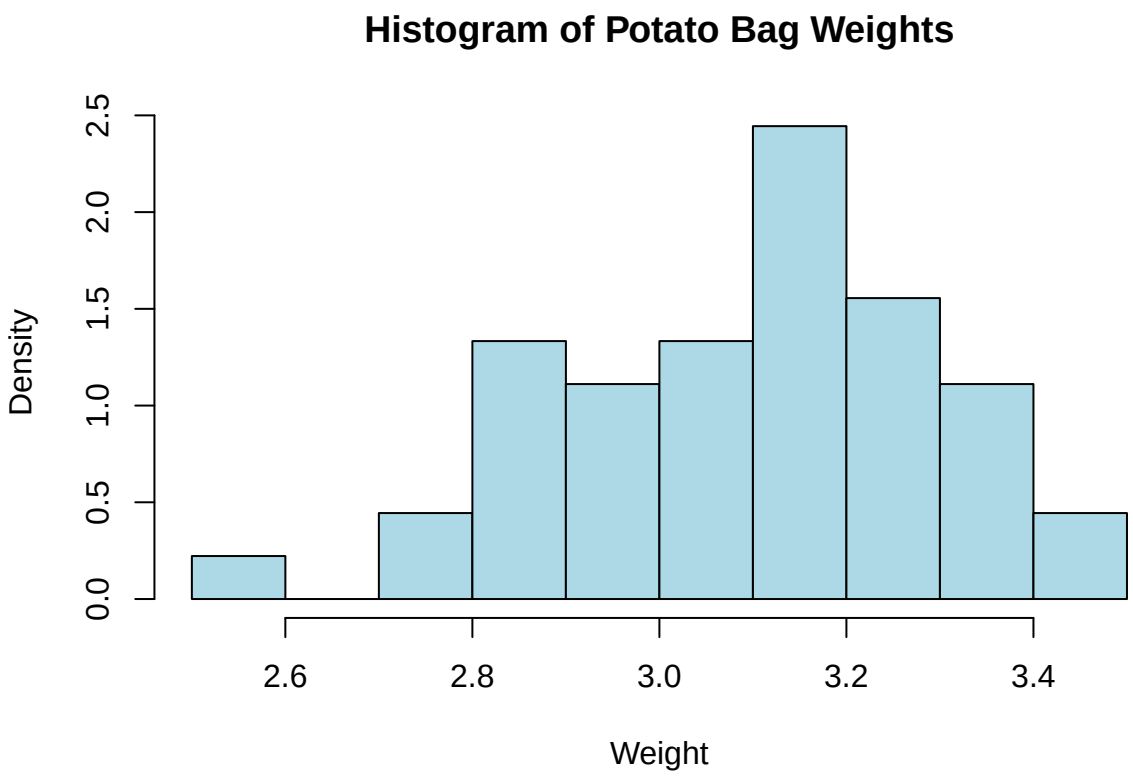
$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

2. Find the corresponding quantiles using that distribution.

ANSWER:

The value of the quantile corresponding to the significance level is $q = 2.015368$

3. And draw it in the graphic of the distribution.



3. Find the margin of error and the bounds of the interval.

ANSWER:

The margin of error is 0.058

The bounds of the interval for the mean: $(3, 3.2)$

Test the following hypothesis

To check if the machine is well calibrated, take a sample of 45 bags filled with potatoes and count their weight. With this information,

Does the engineer have reason to think the machine is poorly calibrated? (Please use 5% signication level).

Design the test

Define the null hypothesis and the alternative hypothesis

The null hypothesis is ... "The machine is well calibrated."

The alternative hypothesis is ... "The machine is poorly calibrated."

The significance level is ... 0.05

Find the corresponding statistic to the sample

The statistic follows the .... t-distribution

The value of the corresponding statistic is ..$FORMULA = 3.4$

Find the critical value corresponding to the significance level

The critical value is ..*2.015368*

Find the p-value

The p-value is ... $P(T >= t|H_0) = 0.00157508$

Make the decision

I reject the null hypothesis because the... p-value is less than the significance level.

Type of errors

The value of Type I error is ... $FORMULA = 0.05$

The value of Type II error is ... $FORMULA = 0.95$