Stat 131 (Mathematical Statistics III)

Lesson 3.1 Z Test for One Population Mean

Learning Outcomes

At the end of the lesson, students should be able to

1. derive the test statistic for testing hypothesis on one population mean; and
2. conduct the z test for testing hypothesis on one population mean.

Introduction

In this lesson we shall apply the concepts of hypothesis testing to testing problems involving the mean of a population. We shall revisit important elements of hypothesis testing such as stating the null and alternative hypothesis, computing the test statistic, and making a statistical decision either using the p-value approach or the critical value/rejection region approach.

The Z test statistic

Suppose that $Y_1, Y_2, \cdots, Y_n$ is a random sample from a distribution with mean μ and variance σ^2, and we are interested in testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ (or any other H_1). Recall that the Central Limit Theorem states that if the sample size, n, is sufficiently large then the sample means will have approximately normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. Thus, in this situation we have

$$\theta = \mu$$
$$\hat{\theta} = \overline{Y}$$
$$\sigma_{\hat{\theta}} = \frac{\sigma}{\sqrt{n}}$$
$$\hat{\sigma}_{\hat{\theta}} = \frac{S}{\sqrt{n}}$$

THus, in testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ (or any other H_1) assuming we have a sufficiently large $n \geq 30$, we shall use

$$Z = \frac{\overline{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

when σ^2 is known. Otherwise, if σ^2 is unknown, we estimate it using the sample standard deviation s and alternatively use the following test statistic.

$$Z = \frac{\overline{Y} - \mu_0}{\frac{S}{\sqrt{n}}}$$

Note that both test statistics have a standard normal distribution under H_0.

Example 3.1.1

Boys of a certain age are known to have a mean weight of $\mu = 85$ pounds. A complaint is made that the boys living in a municipal children's home are underfed. As one bit of evidence, n=25 boys (of the same age) are weighed and found to have a mean weight of $\bar{y} = 80.94$ pounds. It is known that the population standard deviation of the weights is 6.1 pounds. Is the data sufficient to conclude that indeed the boys living in this municipality are underfed?

SOLUTION

Let μ be the true mean weight of all children in this municipality. Then the null and alternative hypotheses for this situation are:

$$H_0 : \mu = 85$$
$$H_1 : \mu < 85$$

Test statistic:

The population standard deviation is known ($\sigma = 6.1$), though $n = 25 < 30$, so the appropriate test statistic is

$$Z = \frac{\overline{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

and the value of this test statistic is

$$\begin{aligned} Z &= \frac{\overline{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{80.94 - 85}{\frac{6.1}{\sqrt{25}}} \\ &= -3.33 \end{aligned}$$

Rejection Region:

Based on the form of the alternative hypothesis and $\alpha = 0.05$, the rejection region is given by $RR = \{z : z \leq -z_{0.05} = -1.645\}$. We can obtain the critical value either from a Z table or using the R command **qnorm(0.05)**.

The p-value associated with the value of the test statistic is given by

$$p - value = P(Z < -3.33) = 4 \times 10^{-4}$$

The p-value was obtained using the command **pnorm(-3.33)**.

Decision on H_0:

Using the rejection region approach, we reject H_0 since $z = -3.33$ is in the rejection region. In other words, $z = -3.33 < -1.645$. Alternatively, we also reject H_0 since $p - value < \alpha = 0.05$.

Conclusion:

Therefore, at the 5% level of significance, the data is sufficient to conclude that the boys living in the municipality are underfed.

Example 3.1.2

A brochure inviting subscriptions for a new diet program states that the participants are expected to lose over 10 kg in five weeks. Suppose that, from the data of the five-week weight losses of 56 participants, the sample mean and sample standard deviation are found to be 10.66 kg and 4.63 kg, respectively. Could the statement in the brochure be substantiated on the basis of these findings? Test at the $\alpha = 0.01$ level.

SOLUTION

Let μ be the true mean weight loss of all participants in the diet program. Then the null and alternative hypotheses are:

$$H_0 : \mu = 10$$
$$H_1 : \mu > 10$$

Test statistic:

We have a *large* sample of $n = 56$ observations from a normal population whose standard deviation is unknown. We therefore estimate σ using the sample standard deviation. Hence, the appropriate test statistic is

$$Z = \frac{\overline{Y} - \mu_0}{\frac{S}{\sqrt{n}}}$$

and the value of this test statistic is

$$
\begin{aligned}
Z &= \frac{\overline{Y} - \mu_0}{\frac{S}{\sqrt{n}}} \\
&= \frac{10.66 - 10}{\frac{4.63}{\sqrt{56}}} \\
&\approx 1.07
\end{aligned}
$$

Rejection Region:

For an $\alpha = 0.01$, the rejection region is given by $RR = \{z : z \geq z_{0.01} = 2.33\}$. We can obtain the critical value either from a Z table or using the R command **qnorm(0.01,lower.tail=FALSE)**.

The p-value associated with the value of the test statistic is given by

$$p - value = P(Z > 1.07) = 0.1423$$

The p-value was obtained using the command **pnorm(1.07,lower.tail=FALSE)**.

Decision on H_0:

Based on the rejection region approach, we fail reject H_0 since $z = 1.07$ does not fall in the rejection region. In other words, $z = 1.07 < 2.33$. Alternatively, we also reject H_0 since $p - value > \alpha = 0.01$.

Conclusion:

Therefore, at the 1% level of significance, the data cannot substantiate the brochure's claim that participants lose over 10 kg in five weeks.

Learning Tasks

Instruction: Answer the following as indicated.

1. The average normal body temperature is $37°C$. As part of the protocols for COVID-19, body temperatures of individuals passing through a check-point were obtained. The average body temperature for a random sample of 100 individuals passing through the checkpoint was computed to be $36.3°C$ with standard deviation of $2.15°C$. Is this data enough to conclude that the average body temperature of people passing the check point is different from the average normal body temperature?

2. The daily wages in a particular industry are normally distributed with mean PhP654.12 and standard deviation PhP123.89. A company in this industry employs 40 workers, paying them an average of PhP604.57 per day. Can this company be accused of paying substandard wages? Use $\alpha = 0.01$.