Lab-8

One Sample Z-test

Aim:-To test the hypothesis for Large Samples by using one –sample Z-test

Procedure and R code:-

Test for significance of single mean:

Lower Tail Test of Population Mean with Known Variance:

The null hypothesis of the lower tail test of the population mean can be expressed as follows:

$$\mu \geq \mu_0$$

where μ_0 ***is a hypothesized lower bound of the true population mean*** μ***.***
Let us define the test statistic z in terms of the sample mean, the sample size and thepopulation standard deviation σ :

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Then the null hypothesis of the lower tail test is to be rejected if $z \leq -z_\alpha$ ***, where*** z_α ***is the 100(1− α) percentile of the standard normal distribution.***

Problem

Suppose the manufacturer claims that the mean lifetime of a light bulb is more than 10,000 hours. In a sample of 30 light bulbs, it was found that they only last 9,900 hours on average. Assume the population standard deviation is 120 hours. At .05 significance level, can we reject the claim by the manufacturer?

R code:

The null hypothesis is that $\mu \geq 10000$**. We begin with computing the test statistic.**

```
> xbar = 9900            # sample mean
> mu0 = 10000            # hypothesized value
> sigma = 120            # population standard deviation
> n = 30                 # sample size
> z = (xbar-mu0)/(sigma/sqrt(n))
> z                      # test statistic
[1] -4.564355
```

We then compute the critical value at .05 significance level.

```
> alpha = .05
> z.alpha = qnorm(1-alpha)
> -z.alpha               # critical value
[1] -1.644854
```

Interpretation:-

The test statistic -4.5644 is less than the critical value of -1.6449. Hence, at .05 significance level, we reject the claim that mean lifetime of a light bulb is above 10,000 hours.

Alternative Solution(compare with P value)

Instead of using the critical value, we apply the pnorm function to compute the lower tail p-value of the test statistic. As it turns out to be less than the .05 significance level, we reject the null hypothesis that $\mu \geq 10000$.

```
> pval = pnorm(z)
> pval                   # lower tail p-value
[1] 2.505166e-06
```

Upper Tail Test of Population Mean with Known Variance:

The null hypothesis of the upper tail test of the population mean can be expressed as follows:

$$\mu \geq \mu_0$$

where μ_0 is a hypothesized upper bound of the true population mean μ.
Let us define the test statistic z in terms of the sample mean, the sample size and thepopulation standard deviation σ :

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Then the null hypothesis of the upper tail test is to be rejected if $z \geq z_\alpha$ where z_α is the 100(1− α) percentile of the standard normal distribution

Problem

Suppose the food label on a cookie bag states that there is at most 2 grams of saturated fat in a single cookie. In a sample of 35 cookies, it is found that the mean amount of saturated fat per cookie is 2.1 grams. Assume that the population standard deviation is 0.25 grams. At .05 significance level, can we reject the claim on food label?

R code:-

The null hypothesis is that $\mu \leq 2$. We begin with computing the test statistic.

```
> xbar = 2.1                 # sample mean
> mu0 = 2                    # hypothesized value
> sigma = 0.25               # population standard deviation
> n = 35                     # sample size
> z = (xbar-mu0)/(sigma/sqrt(n))
> z                          # test statistic
[1] 2.366432
```

We then compute the critical value at .05 significance level.

```
> alpha = .05
> z.alpha = qnorm(1-alpha)
> z.alpha                    # critical value
[1] 1.644854
```

Interpretation:-

The test statistic 2.3664 is greater than the critical value of 1.6449. Hence, at .05 significance level, we reject the claim that there is at most 2 grams of saturated fat in a cookie.

Two-Tailed Test of Population Mean with Known Variance:-

The null hypothesis of the two-tailed test of the population mean can be expressed as follows:

$$\mu = \mu_0$$

where $\mu = \mu_0$ *is a hypothesized value of the true population mean* μ*.Let us define the test statistic z in terms of the sample mean, the sample size and the population standard deviation σ :*

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Then the null hypothesis of the two-tailed test is to be rejected if $z \le -z_{\alpha/2}$ *or* $z \ge z_{\alpha/2}$ *, where* $z_{\alpha/2}$ *is the 100(1 − α/2) percentile of the standard normal distribution.*

Problem

Suppose the mean weight of King Penguins found in an Antarctic colony last year was 15.4 kg. In a sample of 35 penguins same time this year in the same colony, the mean penguin weight is 14.6 kg. Assume the population standard deviation is 2.5 kg. At .05 significance level, can we reject the null hypothesis that the mean penguin weight does not differ from last year?

Solution

The null hypothesis is that $\mu = 15.4$**. We begin with computing the test statistic.**

```
> xbar = 14.6            # sample mean
> mu0 = 15.4             # hypothesized value
> sigma = 2.5            # population standard deviation
> n = 35                 # sample size
> z = (xbar-mu0)/(sigma/sqrt(n))
> z                      # test statistic
[1] -1.893146
```

We then compute the critical values at .05 significance level.

```
> alpha = .05
> z.half.alpha = qnorm(1-alpha/2)
> c(-z.half.alpha, z.half.alpha)
[1] -1.959964  1.959964
```

Interpretation :

The test statistic -1.8931 lies between the critical values -1.9600 and 1.9600. Hence, at .05 significance level, we do not reject the null hypothesis that the mean penguin weight does not differ from last year.

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 that μ = 15.4.

```
> pval = 2 * pnorm(z)    # lower tail
> pval                   # two-tailed p-value
[1] 0.05833852
```

Lower Tail Test of Population Proportion:

The null hypothesis of the lower tail test about population proportion can be expressed as follows:

$$p \geq p_0$$

where p_0 is a hypothesized lower bound of the true population proportion p. Let us define the test statistic z in terms of the sample proportion and the sample size:

$$z = \frac{p - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$$

Then the null hypothesis of the lower tail test is to be rejected if $z \leq -z_\alpha$, where z_α is the 100(1− α) percentile of the standard normal distribution.

Problem

Suppose 60% of citizens voted in last election. 85 out of 148 people in a telephone survey said that they voted in current election. At 0.5 significance level, can we reject the null hypothesis that the proportion of voters in the population is above 60% this year?

Solution

The null hypothesis is that $p \geq 0.6$. We begin with computing the test statistic.

```
> pbar = 85/148           # sample proportion
> p0 = .6                 # hypothesized value
> n = 148                 # sample size
> z = (pbar-p0)/sqrt(p0*(1-p0)/n)
> z                       # test statistic
[1] -0.6375983
```

We then compute the critical value at .05 significance level.

```
> alpha = .05
> z.alpha = qnorm(1-alpha)
> -z.alpha                # critical value
[1] -1.644854
```

Interpretation :

The test statistic -0.6376 is not less than the critical value of -1.6449. Hence, at .05 significance level, we do not reject the null hypothesis that the proportion of voters in the population is above 60% this year.

Alternative Solution 1

Instead of using the critical value, we apply the pnorm function to compute the lower tail p-value of the test statistic. As it turns out to be greater than the .05 significance level, we do not reject the null hypothesis that $p \geq 0.6$.

```
> pval = pnorm(z)
> pval
[1] 0.2618676
```

Alternative Solution 2

We apply the prop.test function to compute the p-value directly. The Yates continuity correction is disabled for pedagogical reasons.

Upper Tail Test of Population Proportion

The null hypothesis of the upper tail test about population proportion can be expressed as follows:

$$p \le p_0$$

where p_0 is a hypothesized upper bound of the true population proportion p. Let us define the test statistic z in terms of the sample proportion and the sample size:

$$z = \frac{p - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$$

Then the null hypothesis of the upper tail test is to be *rejected* if $z \ge z_\alpha$, where z_α is the $100(1-\alpha)$ percentile of the standard normal distribution.

Problem

Suppose that 12% of apples harvested in an orchard last year was rotten. 30 out of 214 apples in a harvest sample this year turns out to be rotten. At .05 significance level, can we reject the null hypothesis that the proportion of rotten apples in harvest stays below 12% this year?

Solution

The null hypothesis is that $p \le 0.12$. We begin with computing the test statistic.

```
> pbar = 30/214          # sample proportion
> p0 = .12               # hypothesized value
> n = 214                # sample size
> z = (pbar-p0)/sqrt(p0*(1-p0)/n)
> z                      # test statistic
[1] 0.908751
```

We then compute the critical value at .05 significance level.

```
> alpha = .05
> z.alpha = qnorm(1-alpha)
> z.alpha                # critical value
[1] 1.644854
```

Interpretation:-

The test statistic 0.90875 is not greater than the critical value of 1.6449. Hence, at .05 significance level, we do not reject the null hypothesis that the proportion of rotten apples in harvest stays below 12% this year.

Alternative Solution 1

Instead of using the critical value, we apply the pnorm function to compute the upper tail p-value of the test statistic. As it turns out to be greater than the .05 significance level, we do not reject the null hypothesis that $p \le 0.12$.

```
> pval = pnorm(z, lower.tail=FALSE)
> pval                    # upper tail p?value
[1] 0.1817408
```

Alternative Solution 2:

We apply the prop.test function to compute the p-value directly. The Yates continuity correction is disabled for pedagogical reasons.

```
> prop.test(30, 214, p=.12, alt="greater", correct=FALSE)

        1-sample proportions test without continuity correction

data:  30 out of 214, null probability 0.12
X-squared = 0.82583, df = 1, p-value = 0.1817
alternative hypothesis: true p is greater than 0.12
95 percent confidence interval:
 0.1056274 1.0000000
sample estimates:
        p
0.1401869
```

Two-Tailed Test of Population Proportion:

The null hypothesis of the two-tailed test about population proportion can be expressed as follows:

$$p = p_0$$

where p_0 is a hypothesized value of the true population proportion p.
Let us define the test statistic z in terms of the sample proportion and the sample size:

$$z = \frac{p - p_0}{\sqrt{p_0 q_0 / n}} \sim N(0,1)$$

Then the null hypothesis of the two-tailed test is to be *rejected* if $z \le -z_{\alpha/2}$ or $z \ge z_{\alpha/2}$, where $z_{\alpha/2}$ is the $100(1 - \alpha)$ percentile of the standard normal distribution.

Problem

Suppose a coin toss turns up 12 heads out of 20 trials. At .05 significance level, can one reject the null hypothesis that the coin toss is fair?

Solution

The null hypothesis is that p = 0.5. We begin with computing the test statistic.

```
> pbar = 12/20            # sample proportion
> p0 = .5                 # hypothesized value
> n = 20                  # sample size
> z = (pbar-p0)/sqrt(p0*(1-p0)/n)
> z                       # test statistic
[1] 0.8944272
```

We then compute the critical values at .05 significance level.

```
> alpha = .05
> z.half.alpha = qnorm(1-alpha/2)
> c(-z.half.alpha, z.half.alpha)
[1] -1.959964  1.959964
```

Interpretation:

The test statistic 0.89443 lies between the critical values -1.9600 and 1.9600. Hence, at .05 significance level, we do not reject the null hypothesis that the coin toss is fair.

Alternative Solution 1

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 upper tail p-value as the sample proportion is greaterthan the hypothesized value. Since it turns out to be greater than the .05 significance level, we do not reject the null hypothesis that p = 0.5.

```
>  pval = 2 * pnorm(z, lower.tail=FALSE)   # upper tail
>  pval
[1] 0.3710934
```

Alternative Solution 2

We apply the prop.test function to compute the p-value directly. The Yates continuity correction is disabled for pedagogical reasons.

```
> prop.test(12, 20, p=0.5, correct=FALSE)

        1-sample proportions test without continuity correction

data:  12 out of 20, null probability 0.5
X-squared = 0.8, df = 1, p-value = 0.3711
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3865815 0.7811935
sample estimates:
  p
0.6
```

Practice Problems:

1. **A sample of 100 tyres is taken from a lot. The mean life of tyres is found to be 39, 350 kilo meters with a standard deviation of 3, 260. Could the sample come from a population with mean life of 40, 000 kilometers?**

2. **the mean life time of a sample of 400 fluorescent light bulbs produced by a company is found to be 1, 570 hours with a standard deviation of 150 hours. Test the hypothesis that the mean life time of bulbs is 1600 hours against the alternative hypothesis that it is greater than 1, 600 hours at 1% and 5% level of significance.**
3. **In the sample of 1000 people in Maharashtra,540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance**
4. **A particular brand of tires claims that its deluxe tire averages at least 50,000 miles before it needs to be replaced. From past studies of this tire, the standard deviation is known to be 8000. A survey of owners of that tire design is conducted. From the 28 tires surveyed, the average lifespan was 46,500 miles with a standard deviation of 9800 miles. Do the data support the claim at the 5% level?**
5. **From generation to generation, the average age when smokers first start to smoke varies. However, the standard deviation of that age remains constant of around 2.1 years. A survey of 40 smokers of this generation was done to see if the average starting age is at least 19. The sample average was 18.1 with a sample standard deviation of 1.3. Do the data support the claim at the 5% level?**
6. **The cost of a daily newspaper varies from city to city. However, the variation among prices remains steady with a standard deviation of 6¢. A study was done to test the claim that the average cost of a daily newspaper is 35¢. Twelve costs yield an average cost of 30¢ with a standard deviation of 4¢. Do the data support the claim at the 1% level?**
7. **An article in the *San Jose Mercury News* stated that students in the California state university system take an average of 4.5 years to finish their undergraduate degrees. Suppose you believe that the average time is longer. You conduct a survey of 49 students and obtain a sample mean of 5.1 with a sample standard deviation of 1.2. Do the data support your claim at the 1% level?**
8. **The average number of sick days an employee takes per year is believed to be about 10. Members of a personnel department do not believe this figure. They randomly survey 8 employees. The number of sick days they took for the past year are as follows: 12; 4; 15; 3; 11; 8; 6; 8. Let X = the number of sick days they took for the past year. Should the personnel team believe that the average number is about 10?**
9. **In 1955, *Life Magazine* reported that the 25 year-old mother of three worked [on average] an 80 hour week. Recently, many groups have been studying whether or not the women's movement has, in fact, resulted in an increase in the average work week for women (combining employment and at-home work). Suppose a study was done to determine if the average work week has increased. 81 women were surveyed with the following results. The sample average was 83; the sample standard deviation was 10. Does it appear that the average work week has increased for women at the 5% level?**

10. Your statistics instructor claims that 60 percent of the students who take her Elementary Statistics class go through life feeling more enriched. For some reason that she can't quite figure out, most people don't believe her. You decide to check this out on your own. You randomly survey 64 of her past Elementary Statistics students and find that 34 feel more enriched as a result of her class. Now, what do you think?

11. According to an article in *Newsweek*, the natural ratio of girls to boys is 100:105. In China, the birth ratio is 100: 114 (46.7% girls). Suppose you don't believe the reported figures of the percent of girls born in China. You conduct a study. In this study, you count the number of girls and boys born in 150 randomly chosen recent births. There are 60 girls and 90 boys born of the 150. Based on your study, do you believe that the percent of girls born in China is 46.7?

12. A poll done for *Newsweek* found that 13% of Americans have seen or sensed the presence of an angel. A contingent doubts that the percent is really that high. It conducts its own survey. Out of 76 Americans surveyed, only 2 had seen or sensed the presence of an angel. As a result of the contingent's survey, would you agree with the *Newsweek* poll? In complete sentences, also give three reasons why the two polls might give different results.