Introduction to Statistical Hypothesis Testing in R

A statistical hypothesis is an assumption made by the researcher about the data of the population collected for any experiment. It is not mandatory for this assumption to be true every time. Hypothesis testing, in a way, is a formal process of validating the hypothesis made by the researcher.

In order to validate a hypothesis, it will consider the entire population into account. However, this is not possible practically. Thus, to validate a hypothesis, it will use random samples from a population. On the basis of the result from testing over the sample data, it either selects or rejects the hypothesis.

Statistical Hypothesis Testing can be categorized into two types as below:

- **Null Hypothesis –** Hypothesis testing is carried out in order to test the validity of a claim or assumption that is made about the larger population. This claim that involves attributes to the trial is known as the Null Hypothesis. The null hypothesis testing is denoted by *H0*.
- **Alternative Hypothesis –** An alternative hypothesis would be considered valid if the null hypothesis is fallacious. The evidence that is present in the trial is basically the data and the statistical computations that accompany it. The alternative hypothesis testing is denoted by H_1 or H_a.

Let's take an example of the coin. We want to conclude that a coin is unbiased or not. Since null hypothesis refers to the natural state of an event, thus, according to the null hypothesis, there would an equal number of occurrences of heads and tails, if a coin is tossed several times. On the other hand, the alternative hypothesis negates the null hypothesis and refers that the occurrences of heads and tails would have significant differences in number.

Statisticians use hypothesis testing to formally check whether the hypothesis is accepted or rejected. Hypothesis testing is conducted in the following manner:

1. **State the Hypotheses –** Stating the null and alternative hypotheses.
2. **Formulate an Analysis Plan –** The formulation of an analysis plan is a crucial step in this stage.
3. **Analyze Sample Data –** Calculation and interpretation of the test statistic, as described in the analysis plan.
4. **Interpret Results –** Application of the decision rule described in the analysis plan.

Hypothesis testing ultimately uses a p-value to weigh the strength of the evidence or in other words what the data are about the population. The p-value ranges between 0 and 1. It can be interpreted in the following way:

- 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.

A p-value very close to the cutoff (0.05) is considered to be marginal and could go either way.

Decision Errors in R

The two types of error that can occur from the hypothesis testing:

- **Type I Error –** Type I error occurs when the researcher rejects a null hypothesis when it is true. The term significance level is used to express the probability of Type I error while testing the hypothesis. The significance level is represented by the symbol α *(alpha).*
- **Type II Error –** Accepting a false null hypothesis H_0 is referred to as the Type II error. The term power of the test is used to express the probability of Type II error while testing hypothesis. The power of the test is represented by the symbol β *(beta).*

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 the population standard deviation σ :

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Then the null hypothesis of the lower tail test is to be *rejected* if $z \leq -z_\alpha$, where z_α is the $100(1 - \alpha)$ 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?

Solution

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] −4.5644
```

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

```
> alpha = .05
> z.alpha = qnorm(1−alpha)
> −z.alpha             # critical value
[1] −1.6449
```

Answer

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

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] 2.5052e−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:

$$\mu \leq \mu_0$$

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 the population standard deviation σ :

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Then the null hypothesis of the upper tail test is to be *rejected* if $z \geq z_\alpha$, where z_α is the $100(1 - \alpha)$ 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?

Solution

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] 2.3664
```

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

```
> alpha = .05
> z.alpha = qnorm(1−alpha)
> z.alpha             # critical value
[1] 1.6449
```

Answer

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.

Alternative Solution

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 less than the .05 significance level, we reject the null hypothesis that $\mu \le 2$.

```
> pval = pnorm(z, lower.tail=FALSE)
> pval                   # upper tail p−value
[1] 0.0089802
```

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 μ_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 :

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

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 $100(1 - \alpha/2)$ 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] −1.8931
```

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] −1.9600  1.9600
```

Answer

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] 0.058339
```

Lower Tail Test of Population Mean with Unknown 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 t in terms of the sample mean, the sample size and the sample standard deviation s :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Then the null hypothesis of the lower tail test is to be *rejected* if $t \leq -t_\alpha$, where t_α is the $100(1 - \alpha)$ percentile of the Student t distribution with $n - 1$ degrees of freedom.

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 sample standard deviation is 125 hours. At .05 significance level, can we reject the claim by the manufacturer?

Solution

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

```
> xbar = 9900          # sample mean
> mu0 = 10000          # hypothesized value
> s = 125              # sample standard deviation
> n = 30               # sample size
> t = (xbar−mu0)/(s/sqrt(n))
> t                    # test statistic
[1] −4.3818
```

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

```
> alpha = .05
> t.alpha = qt(1−alpha, df=n−1)
> −t.alpha             # critical value
[1] −1.6991
```

Answer

The test statistic -4.3818 is less than the critical value of -1.6991. Hence, at .05 significance level, we can reject the claim that mean lifetime of a light bulb is above 10,000 hours.

Alternative Solution

Instead of using the critical value, we apply the pt 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 = pt(t, df=n−1)
> pval              # lower tail p−value
[1] 7.035e−05
```

Upper Tail Test of Population Mean with Unknown Variance

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

$$\mu \leq \mu_0$$

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

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Then the null hypothesis of the upper tail test is to be *rejected* if $t \geq t_\alpha$, where t_α is the $100(1 - \alpha)$ percentile of the Student t distribution with $n - 1$ degrees of freedom.

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 sample standard deviation is 0.3 gram. At .05 significance level, can we reject the claim on food label?

Solution

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

```
> xbar = 2.1          # sample mean
> mu0 = 2             # hypothesized value
> s = 0.3             # sample standard deviation
> n = 35              # sample size
> t = (xbar−mu0)/(s/sqrt(n))
> t                   # test statistic
[1] 1.9720
```

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

```
> alpha = .05
> t.alpha = qt(1−alpha, df=n−1)
> t.alpha             # critical value
[1] 1.6991
```

Answer

The test statistic 1.9720 is greater than the critical value of 1.6991. Hence, at .05 significance level, we can reject the claim that there is at most 2 grams of saturated fat in a cookie.

Alternative Solution

Instead of using the critical value, we apply the pt function to compute the upper 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 $\mu \leq 2$.

```
> pval = pt(t, df=n−1, lower.tail=FALSE)
> pval              # upper tail p−value
[1] 0.028393
```

Two-Tailed Test of Population Mean with Unknown Variance

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

$$\mu = \mu_0$$

where μ_0 is a hypothesized value of the true population mean μ.
Let us define the test statistic t in terms of the sample mean, the sample size and the sample standard deviation s :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Then the null hypothesis of the two-tailed test is to be *rejected* if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$, where $t_{\alpha/2}$ is the $100(1 - \alpha)$ percentile of the Student t distribution with $n - 1$ degrees of freedom.

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 sample 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
> s = 2.5              # sample standard deviation
> n = 35               # sample size
> t = (xbar−mu0)/(s/sqrt(n))
> t                    # test statistic
[1] −1.8931
```

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

```
> alpha = .05
> t.half.alpha = qt(1−alpha/2, df=n−1)
> c(−t.half.alpha, t.half.alpha)
[1] −2.0322  2.0322
```

Answer

The test statistic -1.8931 lies between the critical values -2.0322, and 2.0322. 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 pt 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 $\mu = 15.4$.

```
> pval = 2 ∗ pt(t, df=n−1)  # lower tail
> pval                      # two−tailed p−value
[1] 0.066876
```

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:

$$z = \frac{\bar{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

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] −0.6376
```

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

```
> alpha = .05
> z.alpha = qnorm(1−alpha)
> −z.alpha             # critical value
[1] −1.6449
```

Answer

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                 # lower tail p−value
[1] 0.26187
```

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(85, 148, p=.6, alt="less", correct=FALSE)

        1−sample proportions test without continuity
        correction

data:  85 out of 148, null probability 0.6
X−squared = 0.4065, df = 1, p−value = 0.2619
```

```
alternative hypothesis: true p is less than 0.6
95 percent confidence interval:
 0.0000 0.63925
sample estimates:
      p
0.57432
```

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:

$$z = \frac{\bar{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

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] 0.90875
```

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

```
> alpha = .05
> z.alpha = qnorm(1−alpha)
> z.alpha              # critical value
[1] 1.6449
```

Answer

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] 0.18174
```

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 = 0.8258, df = 1, p−value = 0.1817
alternative hypothesis: true p is greater than 0.12
95 percent confidence interval:
 0.10563 1.00000
sample estimates:
      p
0.14019
```

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:

$$z = \frac{\bar{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

Then the null hypothesis of the two-tailed test is to be *rejected* if $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$, where $z_{\alpha/2}$ is the 100(1 − α) 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] 0.89443
```

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] −1.9600  1.9600
```

Answer

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 *greater* than 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                 # two−tailed p−value
[1] 0.37109
```

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:
  0.38658 0.78119
sample estimates:
  p
0.6
```