# Lab 6: Hypothesis Testing – Type I Error

Researchers retain or reject hypothesis based on measurements of observed samples. The decision is often based on a statistical mechanism called **hypothesis testing**. A **type I error** is the mishap of falsely rejecting a null hypothesis when the null hypothesis is true. The probability of committing a type I error is called the **significance level** of the hypothesis testing, and is denoted by the Greek letter $\alpha$ .

The following exercises concern hypothesis testing in R with possible type I error.

## Exercise 1: 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 $\mu_0$ is a hypothesized lower bound of the true population mean $\mu$.

Let us define the test statistic z in terms of the sample mean, the sample size and the population standard deviation $\sigma$ :

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

Then the null hypothesis of the lower tail test is to be rejected if $z \leq -z_\alpha$ , where $z_\alpha$ is the $100(1 - \alpha)$ percentile of the standard normal distribution is the critical value and can be computed by the R function qnorm(1-alpha).

### *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?

## Exercise 2: 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 $\mu_0$ is a hypothesized upper bound of the true population mean $\mu$.

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 $z_\alpha$ 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?

## Exercise 3: 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 $\mu_0$ 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 $z \leq -z_{\alpha/2}$ or $z \geq z_{\alpha/2}$ , 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?

For the computation of critical value we use the qnorm(1−alpha/2). Why?

## Exercise 4: 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 $\mu_0$ is a hypothesized lower bound of the true population mean $\mu$.

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 $t_\alpha$ is the $100(1 - \alpha)$ percentile of the **Student t distribution** with $n - 1$ degrees of freedom. Use the R function qt.

### *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?

## Exercise 5: 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 $\mu_0$ is a hypothesized upper bound of the true population mean $\mu$.

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 $t_\alpha$ 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?

## Exercise 6: 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 $\mu_0$ is a hypothesized value of the true population mean $\mu$.

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?

## Exercise 7: 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 $z_\alpha$ is the $100(1 - \alpha)$ 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?

## Exercise 8: 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 $z_\alpha$ 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?

## Exercise 9: 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 $z \le -z_{\alpha/2}$ 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?