Hypothesis Testing Examples

Tax Fraud Example

Example: Tax Fraud

Historically, IRS taxpayer compliance audits have revealed that about 5% of individuals do things on their tax returns that invite criminal prosecution.

A sample of $n = 1000$ tax returns produces $\hat{p} = 0.061$ as an estimate of the fraction of fraudulent returns.

Does this provide a clear signal of change in the tax payer behavior?

1. State the Hypotheses

Tax Fraud Example

2. The *test statistic* will be obtained from

$$Z = \frac{\hat{p} - \#}{\sqrt{\frac{\#(1-\#)}{n}}} = \frac{\hat{p} - 0.05}{\sqrt{\frac{0.05(0.95)}{n}}}$$

Under the null hypothesis, Z follows a N(0,1) distribution.

Plugging in our data values, we get the test statistic

$$z = \frac{0.061 - 0.05}{\sqrt{\frac{0.05(0.95)}{1000}}} = 1.59$$

Tax Fraud Cont.

3. Since we have a "$\neq$" in the H_A, the $p-$value is obtained from both the left-hand and right-hand area of the normal curve.

$$\begin{aligned} p - value &= P(|Z| \geq 1.59) \\ &= P(Z < -1.59) + P(Z > 1.59) \\ &= 2 \cdot P(Z < -1.59) \\ &= 2 * 0.0559 \\ &= 0.1118 \end{aligned}$$

This is not a very small $p-$value. We therefore only have very weak evidence against H_0. Thus, we *do not* reject the null hypothesis in favor of the alternative hypothesis.

There is not much evidence of change in tax payer behavior.

Disk Drive Example

Example: Disk Drive

$n_1 = 30$ and $n_2 = 40$ disk drives of 2 different designs were tested under conditions of "accelerated" stress and times to failure recorded:

Standard Design	New Design
$n_1 = 30$	$n_2 = 40$
$\bar{x}_1 = 1205$ hr	$\bar{x}_2 = 1400$ hr
$s_1 = 1000$ hr	$s_2 = 900$ hr

Does the new design have a larger mean time to failure under "accelerated" stress? In other word, is the new design better?

1. State the Hypotheses

Disk Drive Cont.

2. The *test statistic* will be obtained from

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Under the null hypothesis, Z follows a N(0,1) distribution.

Plugging in our data values, we get the test statistic

$$z = \frac{(1205 - 1400) - 0}{\sqrt{\frac{1000^2}{30} + \frac{900^2}{40}}} = -0.84$$

Disk Drive Cont.

3. Since we have a "$<$" in the H_A, the $p-$value is obtained from the left-hand area of the normal curve.

$$\begin{aligned} p - value &= P(Z < -0.84) \\ &= 0.2005 \end{aligned}$$

This is not a small $p-$value. We therefore only have very weak evidence against H_0. Thus, we *do not* reject the null hypothesis in favor of the alternative hypothesis.

There is not significant evidence that the new design is better.

Queuing System Example

Example: Queuing System

Suppose we have 2 queuing systems A and B. We'd like to know whether system A has a higher probability of having an available server in the long run than system B. The simulation data for the 2 servers is shown below:

System A	System B
$n_1 = 500$ runs	$n_2 = 1000$ runs
$\hat{p}_1 = \frac{303}{500}$	$\hat{p}_2 = \frac{551}{1000}$

where $\hat{p}$ is the proportion runs with available servers at $t = 2000$.

1. State the Hypotheses

Queuing System Cont.

2. The *test statistic* will be obtained from

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_{pool}(1 - \hat{p}_{pool})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Under the null hypothesis, Z follows a N(0,1) distribution.

Next, calculate $\hat{p}_{pool}$ to plug into the denominator of the test statistic.

$$\hat{p}_{pool} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{303 + 551}{500 + 1000} = 0.569$$

Plugging in our data values, we get the test statistic

$$z = \frac{(0.606 - 0.551) - 0}{\sqrt{0.569(1 - 0.569)}\sqrt{\frac{1}{500} + \frac{1}{1000}}} = 2.03$$

Queuing System Cont.

3. Since we have a "$>$" in the H_A, the $p-$value is obtained from the right-hand area of the normal curve.

$$\begin{aligned} p - value &= P(Z > 2.03) \\ &= 1 - 0.9788 \\ &= 0.0212 \end{aligned}$$

This is a small $p-$value. We therefore have strong evidence against H_0. Thus, we reject the null hypothesis in favor of the alternative hypothesis.

There is strong evidence that system A has a higher probability of having an available server than system B.