Hypothesis testing: One-sample Inference

Antonio Falcó

Unit 7

Case to study

Familiar aggregation of cardiovascular risk-factors in general and lipid levels in particular.

Data

A group of men who have died from heart disease within the past year are identified, and the cholesterols levels of their offspring are measured.

Claim

Two hypothesis are considered:

1. *The average cholesterol level of these children is 175 mg/dL.*
2. *The average cholesterol level of these children is $>$ 175 mg/dL.*

Question

Why is hypothesis testing so important?

Answer

Hypothesis testing provides an objective framework for making decisions using probabilistic methods, rather than relying on subjective impressions.

Definition

The *null hypothesis*, denoted by H_0 is the hypothesis that is to be tested. The *alternative hypothesis*, denoted by H_1 is the hypothesis that in some sense contradicts the null hypothesis.

Example

Due to dental anxiety experienced by the patients, their pulse rate tends to increase while the patients are sitting in the dental chair. Suppose the average pulse rate of the dental patients during the treatment is 105 beats per minute. Research is being conducted to test what effect soothing music might have on the patients: decreased pulse rate or no changes at all. State the null and alternative hypotheses.

Answer

Letting μ be the average pulse rate of the patients

- $H_0 : \mu = 105$,
- $H_1 : \mu < 105$.

Possibilities

- Accept H_0 and it is true.
- Accept H_0 and it is not true, we have a *type II error*.
- Reject H_0 and it is true, we have a *type I error*.
- Reject H_0 and it is not true.

$$Pr(\text{ Reject } H_0 | H_1 \text{ false }) = \alpha \text{ (significance level) },$$

and

$$Pr(\text{ Accept } H_0 | H_1 \text{ true }) = \beta.$$

Definition

The *power of a test* is defined as

$$1-\beta = 1 - Pr(\text{ Accept } H_0 | H_1 \text{ true }),$$

that is,

$$1-\beta = Pr(\text{ Reject } H_0 | H_1 \text{ true }),$$

Example

Let

$$X \equiv \text{ Pulse rate of a patient that is sitting in a dental chair },$$

and assume that $X \sim N(\mu, \sigma^2)$. Then we would to test:

- $H_0 : \mu = 105$,
- $H_1 : \mu < 105$.

Disccusion (Part I)

Given a sample $X_1, \ldots, X_n$ from X we prevously known

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}.$$

Fix a siginificance level equal to α, say us 0.05. Recall

$$Pr(\text{ Reject } H_0 | H_1 \text{ false }) = \alpha,$$

thus

$$Pr(\text{ Accept } H_0 | H_1 \text{ false }) = 1 - \alpha (= 0.95).$$

Note that we are assume that

$$H_1 \text{ false } \equiv H_0 \text{ true } (\mu = 105).$$

Disccusion (Part II)

Since

$$H_1 \text{ false } \equiv H_0 \text{ true } (\mu = 105).$$

Then

$$\frac{\overline{X} - 105}{\frac{S}{\sqrt{n}}} \sim t_{n-1}.$$

We would to identify the events

$$\{ \text{ Accept } H_0 | H_1 \text{ false } \} \equiv \{\overline{X} \text{ belongs to some acceptance region}\}$$

To this end consider the expression

$$Pr\left(t_\alpha \leq \frac{\overline{X} - 105}{\frac{S}{\sqrt{n}}} \right) = 1 - \alpha.$$

Disccusion (Part III)

Take a sample of pulse rates from ten patiens:

```
98.13701  95.01450 102.99256  93.44211 122.86482 100.30452
88.03586 91.62925 111.84260 120.99088.
```

We obtain

```
> mean(x)
[1] 102.5254
> sd(x)
[1] 12.17227
```

Thus

$$\frac{\overline{X} - 105}{\frac{S}{\sqrt{n}}} = \frac{102.5254 - 105}{\frac{12.17227}{\sqrt{10}}} = -0.6428826.$$

Disccusion (Part VI)

For $\alpha = 0.05$ we have that for $n = 10$

$$Pr\left(t_\alpha \leq t_9\right) = 1 - 0.05 = 0.95.$$

t_α can be computed as

```
qt(c(0.95), df=9, lower.tail=FALSE)
[1] -1.833113
```

Since for our data and under the hypothesis that H_0 is true we have obtained

$$t_9 = \frac{102.5254 - 105}{\frac{12.17227}{\sqrt{10}}} = -0.6428826 > -1.833113 = t_\alpha,$$

we can't reject H_0.

Remark

Note that

$$Pr\left(t_\alpha \leq \frac{\overline{X}-105}{\frac{S}{\sqrt{n}}}\right) = Pr\left(105 + t_\alpha \frac{S}{\sqrt{n}} \leq \overline{X}\right) = 1-\alpha$$

Thus the interval

$$105 + t_\alpha \frac{S}{\sqrt{n}} < \overline{x} < \infty$$

is called the aceptance region, and

$$-\infty < \overline{x} \leq 105 + t_\alpha \frac{S}{\sqrt{n}}$$

rejection region.

One sample t-test for the Mean of a Normal distribution

To test

- $H_0 : \mu = \mu_0$,
- $H_1 : \mu < \mu_0$.

with a significance level α, we compute

$$t = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

and t_α by means

$$Pr(t_\alpha \leq t_{n-1}) = 1 - \alpha \equiv Pr(t_{n-1} < t_\alpha) = \alpha.$$

- If $t < t_\alpha$ then we reject H_0,
- If $t_\alpha < t$ then we accept H_0.