Inferential Statistics
L5 - Hypothesis Testing

Erlis Ruli (erlis.ruli@unipd.it)

Department of Statistics, University of Padova

Contents

Problem statement

Suppose that the average energy consumption of our population of WMs mounting a standard motor is μ_0.

It's claimed that NG1 family motors would lead to more efficient WMs, i.e. would lead to average consumption μ, with $\mu < \mu_0$.

There are two possibilities:

- the claim is false, so $\mu \geq \mu_0$; this is called Null Hypothesis ("null" because it adds nothing to the current state of art)
- the claim is true, so $\mu < \mu_0$; it's called Alternative Hypothesis.

↳ Either the claim is true, or it's false

Problem statement (cont'd)

Concretely, suppose that $\mu_0 = 20$.

We equip 10 WM's with the NG1 motor and measure their E consumption.

Let these energy values be
19.1, 20.6, 17.3, 21.1, 19.5, 19.5, 21.4, 19.1, 20.5, 19.5.

Their average is 19.76. But $19.76 < \mu_0 = 20$, so the NG1 motor seems to lead on average to more efficient WMs!

Is that really so? Couldn't this be due to a pure luck?

A Simple formulation

Suppose the sample above is a realisation of the iid random sample $Y_1, \ldots, Y_n$ with $Y_i \sim N(\mu, 5)$.

This is a reasonable assumption given that overall energy consumption of a WM is the sum of the consumptions due to the various components of a WM (motor, resistor, etc.).

For this specific example we have (from L4) that

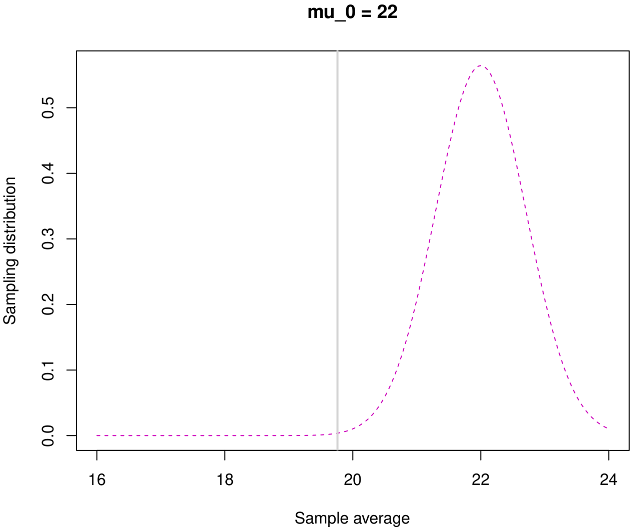
$$\frac{\sqrt{n}(\overline{Y}-\mu)}{\sqrt{5}} \sim N(0,1),$$

and thus

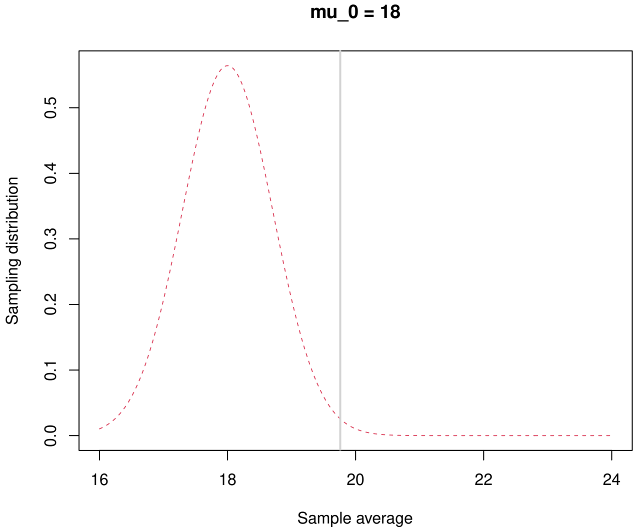
$$\overline{Y} \sim N(\mu, 0.5).$$

The following figure shows this distribution for several values of μ.

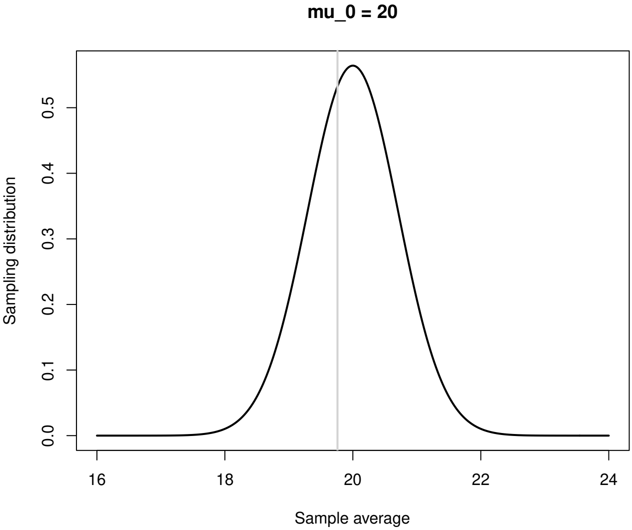
Sampling distribution of $\overline{Y}$



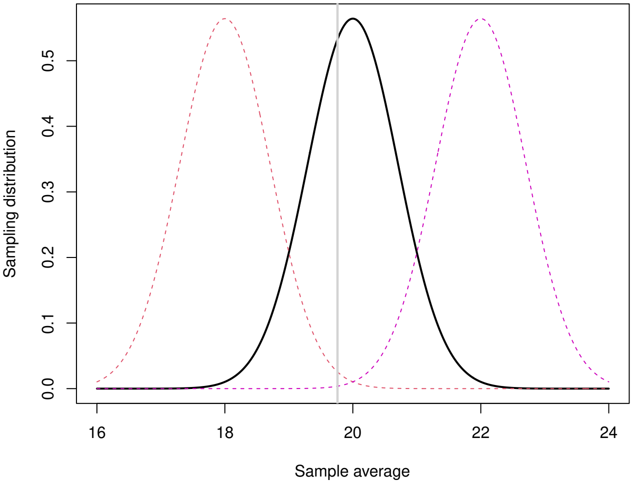
Sampling distribution of $\overline{Y}$



Sampling distribution of $\overline{Y}$



Sampling distributions of $\overline{Y}$



Test hypothesis and decision

Suppose that we judge surprising all values that under the sampling distribution are very unlikely to happen, say all those values with prob less than 0.01.

Under the sampling distribution with $\mu_0 = 20$, this value is 19.42, since $P_{\mu_0}(\overline{Y} \leq 19.42) = 0.01$.

↳ Find this through the quantile

We got a decision rule: a sample average < 19.42 is deemed surprising and should make us suspect about the worthiness of the null hypothesis, so we reject the null hypothesis. Otherwise we do not reject.

In the case above, the observed sample average was $\overline{y} = 19.76 > 19.42$, according to the rule, we should not be surprised and thus do not reject the null hypothesis.

Terminology

In the problem above we tested the null hypothesis $H_0 : \mu \geq \mu_0$ against the alternative $H_1 : \mu < \mu_0$.

In other situations we may be interested in testing

$$H_0 : \mu \leq \mu_0 \text{ against } H_1 : \mu > \mu_0 \tag{*}$$

or

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0 \tag{**}$$

Hypotheses s.t. (*) are called <u>one-tailed</u>, and (**) are called <u>two-tailed</u>.

That was too simple

In our motivating example we assumed population variance was known ($\sigma^2 = 5$). In practice, this assumption is not realistic and must be relaxed.

Furthermore, we assumed an iid random sample with Y_i following a normal distribution. But, in many problems, the normal distribution is not suitable.

The point is that, relaxing $\sigma^2 = 5$ or the distributional assumptions makes the above test useless; thus we have to look elsewhere. In other words, we need to know how to build a test for a given problem at hand.

In the rest of this lecture we will discuss two popular methods (both frequentist-parametric) and we will provide criteria for evaluating their performance.

Setting the scene

Let $Y_1, \ldots, Y_n$ be an iid random sample with $Y_i \sim F_\theta$, where $\theta \in \Theta$ is the unknown parameter with parameter space Θ.

We assume that F_θ has pdf f, indexed by the same parameter θ.

The iid assumption can be relaxed, but let's keep it simple for the moment.

We denote by $\mathcal{Y}$ the range of Y_i and by $\mathcal{Y}^n = \mathcal{Y} \times \mathcal{Y} \times \cdots \times \mathcal{Y}$, the Cartesian product of $\mathcal{Y}$ n times.

Setting the scene (cont'd)

Performing a statistical test essentially entails

building a decision rule from a sample to decide if reject

$$H_0 : \theta \in \Theta_0 \text{ in favour of } H_1 : \theta \in \Theta_0^c,$$

where $\Theta_0 \subset \Theta$.

Specifically, a test is a binary decision rule operating on a subset $R \subset \mathcal{Y}^n$ as follows:

reject H_0 if the observed sample $\mathbf{y} = (y_1, \ldots, y_n) \in R$, and accept H_0 otherwise.

Methods for computing statistical tests

A statistical test is typically defined on the basis of a test statistic $T(\mathbf{Y}) = T(Y_1, \ldots, Y_n)$ which is a function of the sample and closely related to a statistic seen in L4.

The test statistic used for computing the statistical test determines the nature and the name of the statistical test itself.

Likelihood Ratio Tests

The likelihood ratio test statistic for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{y}) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}.$$

A likelihood ratio test (LRT) is any test that has rejection region

$$R = \{\mathbf{y} : \lambda(\mathbf{y}) < c\},$$

where c is any number s.t. $0 \leq c \leq 1$.

Recalling that $\widehat{\theta}$ is the MLE of θ and denoting by $\widehat{\theta}_0$ the constrained MLE of θ when the parameter space is Θ_0, then the LRT statistic is

$$\lambda(\mathbf{y}) = \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}.$$

Example 1

Let $Y_i \sim N(\theta, 5)$, be an iid sample of size n and consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Here θ_0 is a number fixed prior to the experiment.

Under H_0 we have only one possible value for θ, thus the numerator of $\lambda(\mathbf{y})$ is $L(\theta_0)$. On the other hand the (unrestricted) MLE for θ is $\widehat{\theta} = \bar{y}$, the LRT statistic is

$$\begin{aligned}\lambda(\mathbf{y}) &= \frac{(10\pi)^{-n/2} \exp[-\sum_i (y_i - \theta_0)^2/10]}{(10\pi)^{-n/2} \exp[-\sum_i (y_i - \bar{y})^2/10]} \\ &= \exp[-n(\bar{y} - \theta_0)^2/10].\end{aligned}$$

The LRT is thus a test that rejects H_0 for small values of $\lambda(\mathbf{y})$, and the rejection region can be written as

$$\left\{\mathbf{y} : |\bar{y} - \theta_0| \geq \sqrt{-10(\log c)/n}, \right\}$$

for some $c \in (0, 1]$