Likelihood ratio tests

Jan Graffelman[1]

[1]Department of Statistics and Operations Research
Universitat Politècnica de Catalunya
Barcelona, Spain

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
UPC BARCELONATECH

jan.graffelman@upc.edu

October 2, 2022

Contents

Terminology

- A hypothesis is, in general, a statement about a population parameter.
- The value of θ is unknown, but we know $\theta \in \Omega$.
- Supose $\Omega = \Omega_0 \cup \Omega_1$ with $\Omega_0 \cap \Omega_1 = \emptyset$ and we wish to decide $\theta \in \Omega_0$ or $\theta \in \Omega_1$.
- We write $H_0 : \theta \in \Omega_0$ and $H_1 : \theta \in \Omega_1$. Because $\theta \in \Omega$, $\Omega = \Omega_0 \cup \Omega_1$ and $\Omega_0 \cap \Omega_1 = \emptyset$, one of the two is true.
- H_0 is called the null hypothesis, H_1 is called the alternative hypothesis.

Null and alternative hypothesis

- The null and alternative hypothesis are mutually exclusive. Only one of the two can be true.
- On the basis of sample data, we decide which hypothesis to reject.
- We can:

$$\text{reject } H_0 \Leftrightarrow \text{ accept } H_1$$

$$\text{accept } H_0 \Leftrightarrow \text{ reject } H_1$$

Hypothesis test

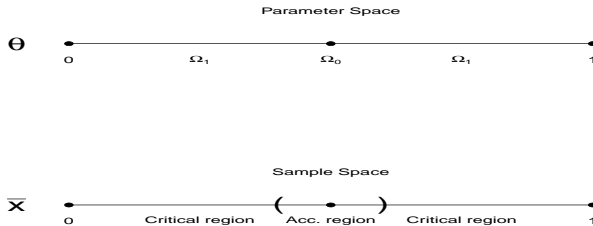A hypothesis test or test procedure is a rule specifying

- for which sample values we accept H_0 as true.
- for which sample values we reject H_0 and accept H_1.

The subset of the sample space for which we reject the null is called the critical region or rejection region. The subset of the sample space for which we accept the null is called the acceptance region.

Example

We wish to test if a coin is fair. We throw the coin 50 times and count the number of heads.

$$\left\{ \begin{array}{l} H_0 : \theta = \frac{1}{2} \\ H_1 : \theta \neq \frac{1}{2} \end{array} \right.$$

Parameter Space

Θ

```
•————————————•————————————•
0          Ω₁          Ω₀          Ω₁          1
```

Sample Space

$\overline{\mathbf{x}}$

```
•——————( •——— )——————•
0    Critical region  Acc. region  Critical region    1
```

Errors

- Sample information may lead us to believe that the coin is not fair, while in fact it is fair. We then reject a true null hypothesis. This is called a type I error.

- Sample information may lead us to believe that the coin is fair, while in fact it is not. We then accept a false null hypothesis. This is called a type II error.

Possibilities

		Test	
		Accept H_0	Reject H_0
Truth	H_0	correct decision	type I error
	H_1	type II error	correct decision

α and β

- $\alpha \equiv P\left(\text{Type I error}\right) = P\left(\text{Reject } H_0 | H_0 \text{ is true}\right)$
- $\beta \equiv P\left(\text{Type II error}\right) = P\left(\text{Accept } H_0 | H_0 \text{ is false}\right)$
- ideally, we would like both probabilities to be small.
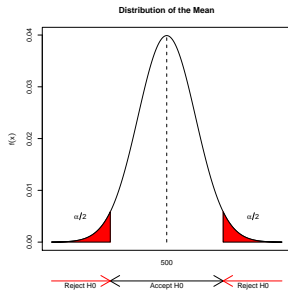
The test statistic and the *p*-value

- The test statistic is a quantity calculated on the basis of the sample, relevant for addressing the hypothesis test under consideration (e.g. $\overline{X}$ when testing $H_0 : \mu = \mu_0$).
- The **p**-value is the probability of observing a test-statistic as extreme or more extreme than the one we have observed.
- The *p*-value quantifies how much evidence there is against the null hypothesis
- The *p*-value is compared with the pre-set significance level α
- If the *p*-value $< \alpha$ then we reject H_0
- If the *p*-value $\geq \alpha$ then we do not reject H_0
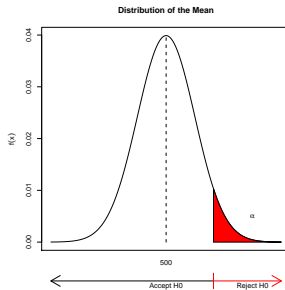
Procedure for a hypothesis test

- Choose a study design.
- Formulate the null hypothesis and the alternative hypothesis.
- Choose a test statistic.
- Determine the desired significance level α.
- Specify the distribution of the test statistic under the null and state assumptions made.
- Compute the test statistic and the p-value and state the decision.
- Optionally provide a confidence interval to supplement your test.

One-tailed and two-tailed tests

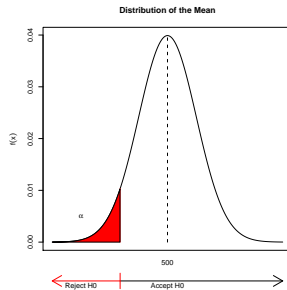$$\left\{ \begin{array}{l} H_0 : \mu = 500 \\ H_1 : \mu \neq 500 \end{array} \right.$$

$$\left\{ \begin{array}{l} H_0 : \mu = 500 \\ H_1 : \mu > 500 \end{array} \right.$$

$$\left\{ \begin{array}{l} H_0 : \mu = 500 \\ H_1 : \mu < 500 \end{array} \right.$$

How about β?

- We decided to control α
- How about β?
- If μ_1 is exactly specified, β can be calculated.

α and β

Distribution of the Mean
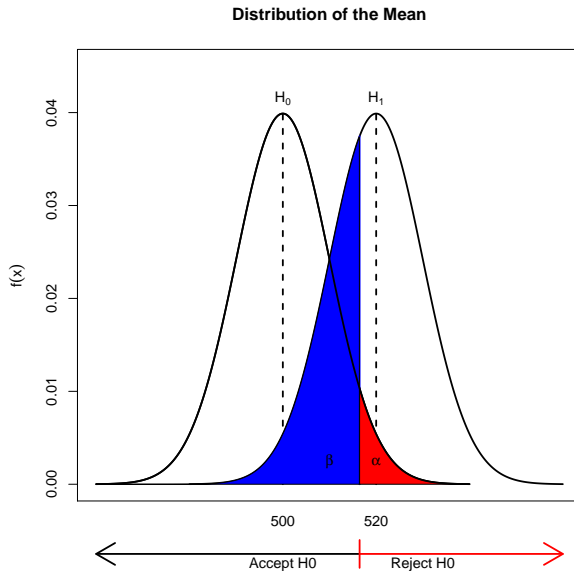
Power

- The probability of rejecting the null is called the power of a test.
- Power $= 1 - \beta$.
- Which factors do affect the power of a test?

Hypothesis test and confidence interval

- Confidence intervals and hypothesis tests are related.
- If $\mu_0 \in CI_{1-\alpha}(\mu)$ then we do not reject $H_0 : \mu = \mu_0$ at significance level α.
- If $\mu_0 \notin CI_{1-\alpha}(\mu)$ then we reject $H_0 : \mu = \mu_0$ at significance level α.

Power function

- A statistical test can be characterized by its power function
- The power function, $\pi(\theta)$, gives the probability of rejecting H_0 for each value of θ.
-
$$\Pi(\theta) = \left\{ \begin{array}{ccc} \alpha = P\left(\text{Type I error}\right) & \text{if} & \theta \in \Omega_0 \\ 1 - \beta = 1 - P\left(\text{Type II error}\right) & \text{if} & \theta \in \Omega_1 \end{array} \right.$$

- Power functions serve to compare test procedures
- The ideal power function is:
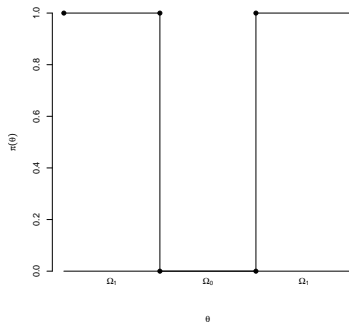
Most powerful test

A test that minimizes β, the probability of a type II error, for a given value of α is called the the most powerful test of size α.

Simple and composite hypothesis

If Ω_0 (or Ω_1) contains only one possible value of θ, then the correspoding hypothesis is called simple. If it contains more than one value, the hypothesis is called composite.

Neyman-Pearson Lemma

- Let $X_1, \ldots, X_n$ be a random sample of a random variable with $f(x \,;\, \theta)$.
- We wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.
- Let $L(\theta \,|\, \mathbf{x})$ be the likelihood function
- The most powerful test of size α has critical region

$$CR = \left\{ \mathbf{x} \in \mathcal{X}^n \,:\, \frac{L(\theta_1 \,|\, \mathbf{x})}{L(\theta_0 \,|\, \mathbf{x})} \geq C_\alpha \right\}$$

for some constant $C_\alpha > 0$.
- Note this lemma involves the ratio of two likelihoods.

Example: Normal distribution

Let $X \sim N(\mu, \sigma^2)$, σ^2 known. We wish to test

$$\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{array} \right.$$

where $\mu_1 > \mu_0$. Intuition: reject H_0 if $\overline{x}$ is "large", accept H_0 if $\overline{x}$ is "small"
Neyman-Pearson:

$$L_0 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu_0}{\sigma}\right)^2} \quad L_1 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu_1}{\sigma}\right)^2}$$

$$CR = \{\mathbf{x} : \frac{L_1}{L_0} > C_\alpha\} \rightarrow \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu_1}{\sigma}\right)^2}}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu_0}{\sigma}\right)^2}} > C_\alpha$$

With a bit algebra

$$CR = \{\mathbf{x} : \overline{x} > C'_\alpha\}$$

Example: Normal distribution

Let $\mu_0 = 0, \mu_1 = 1, \sigma^2 = 1, \alpha = 0.05$ and $n = 10$.
Then Under H_0

$$\overline{X} \sim N(\mu_0, \sigma = 1/\sqrt{10})$$

$$P\left(\text{Type I Error}\right) = P\left(\text{Reject } H_0 | H_0 \text{ is true}\right) = P\left(\overline{X} > k | \mu = 0\right) = \alpha = 0.05$$

$$P\left(\overline{X} > k | \mu = 0\right) = P\left(Z > \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) = 0.05$$

$$\frac{k - \mu_0}{\sigma/\sqrt{n}} = 1.65 \rightarrow k = \mu_0 + 1.65 \cdot \frac{\sigma}{\sqrt{n}} = \frac{1.65}{\sqrt{10}} = 0.52$$

$$CR = \{\mathbf{x} : \overline{X} > 0.52\}$$

Uniformly most powerful tests.

A size α test is uniformly most powerful (UMP) for the contrast $H_0 : \theta \in \Omega_0$ versus $H_1 : \theta \in \Omega_1$, with power function $\pi(\theta)$ and with $\sup_{\theta \in \Omega_0} \pi(\theta) = \alpha$ if, for any other power function $\pi^*(\theta)$, also with $\sup_{\theta \in \Omega_0} \pi^*(\theta) = \alpha$ $\pi(\theta)$ satisfies

$$\pi(\theta) \geq \pi^*(\theta), \qquad \text{for all } \theta \in \Omega_1$$

Graphically:
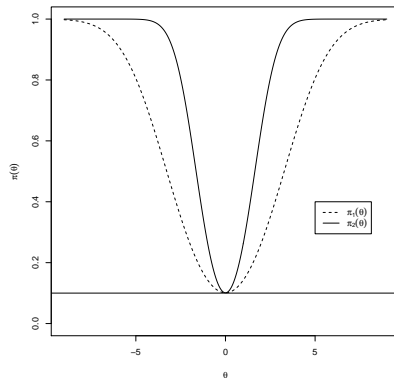
Generalized likelihood ratio test

Let $X_1, \ldots, X_n$ be a random sample of X with density $f(x \mid \theta)$ for some $\theta \in \Omega$.
We wish to test

$$\left\{ \begin{array}{l} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{array} \right.$$

where $\Theta = \Omega_0 \cup \Omega_1$ and $\Omega_0 \cap \Omega_1 = \emptyset$.
We define the likelihood ratio statistic as

$$\lambda = \lambda(\mathbf{x}) = \frac{\max_{\theta \in \Omega_0} L(\theta \mid \mathbf{x})}{\max_{\theta \in \Omega} L(\theta \mid \mathbf{x})}$$

The likelihood ratio test, also called the generalized likelihood ratio test in order to distinguish it from the test obtained by the lemma of Neyman-Pearson, has a critical region of the form

$$C = \left\{ \mathbf{x} : \lambda(\mathbf{x}) \leq A \right\}$$

for some constant A to be determined such that the size of the test is α.
Note that $0 \leq \lambda \leq 1$. The closer λ is to 1, the more likely it is that $\theta \in \Omega_0$, whereas if λ is far from 1 the more likely is the alternative hypothesis $\theta \in \Omega_1$.

Asymptotic distribution of the likelihood ratio statistic

Theorem:

Let $X_1, \ldots, X_n$ be a random sample, where X_i has density $f(x \mid \theta)$ for some $\theta \in \Omega$. We wish to test

$$\left\{ \begin{array}{l} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{array} \right.$$

where $\Omega = \Omega_0 \cup \Omega_1$ i $\Omega_0 \cap \Omega_1 = \emptyset$.
Assuming that the derivatives of the likelihood function exist and are continuous, and that the support of the distribution does not depend on the parameter. Given these conditions, the statistic

$$Q_n = -2 \log \lambda(x) \to \chi_d^2 \qquad \text{under } H_0 : \theta \in \Omega_0$$

with $d = dim(\Omega) - dim(\Omega_0)$.

Unifying framework

- Many classical statistical tests are actually likelihood ratio tests.
- Likelihood ratio theory provides a unifying framework for hypothesis testing.
- LRT tests are important in statistical modeling, and form the basis for comparing nested models.
- In the remainder of this module, we will give some examples of LR-tests.
- LR-tests underlie model comparison procedures in, among others, logistic regression, and Poisson regression, which are explained in the next sessions

A summary of classical statistical tests

Test	Hypothesis	Statistic	Distribution
One-sample Z	$H_o : \mu = \mu_0$	$Z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$	$N(0, 1)$
	$H_1 : \mu \neq \mu_0$		
One-sample Z	$H_o : p = p_0$	$Z = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}}$	$N(0, 1)$
(proportion)	$H_1 : p \neq p_0$		
One-sample T	$H_o : \mu = \mu_0$	$T = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$	t_{n-1}
	$H_1 : \mu \neq \mu_0$		
Two-sample T	$H_o : \mu_D = 0$	$T = \frac{\overline{D} - \mu_D}{s_{\overline{D}}}$	t_{n-1}
(paired)	$H_1 : \mu_D \neq 0$		
Two-sample T	$H_o : \mu_x = \mu_y$	$T = \frac{\overline{X}_m - \overline{Y}_n - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}}$	t_{m+n-2}
(independent)	$H_1 : \mu_x \neq \mu_y$		
Two-sample T	$H_o : \mu_x = \mu_y$	$T = \frac{\overline{X}_m - \overline{Y}_n - (\mu_1 - \mu_2)}{\sqrt{\frac{s_m^2}{m} + \frac{s_n^2}{n}}}$	$t_{\hat{\nu}}$
(independent)	$H_1 : \mu_x \neq \mu_y$		
Two-sample F	$H_o : \sigma_x^2 = \sigma_y^2$	$F = \frac{s_x^2}{s_y^2}$	$F(n_x - 1, n_y - 1)$
	$H_1 : \sigma_x^2 \neq \sigma_y^2$		

Example: Single Nucleotide Polymorphism (SNP)
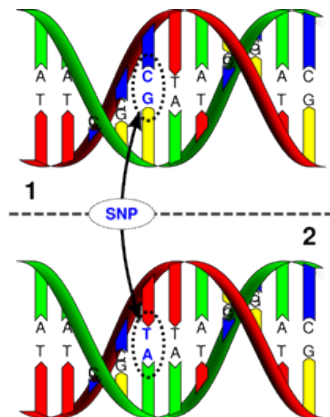
A glance at a SNP database

Id	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	⋯
NA18524	CC	CC	TT	TT	AT	AC	CC	AC	CT	GG	⋯
NA18526	CC	CC	CT	TT	AT	CC	CC	AC	CT	GG	⋯
NA18529	CC	CC	TT	TT	TT	AC	CG	AC	CT	GG	⋯
NA18532	CC	CC	TT	TT	TT	AC	CG	AC	CT	GG	⋯
NA18537	CC	CC	TT	TT	AT	CC	CC	AC	CT	GG	⋯
NA18540	CC	CC	CT	TT	AT	CC	CG	AC	CT	GG	⋯
NA18542	CC	CC	TT	TT	TT	CC	CG	AC	CT	GG	⋯
NA18545	CC	CC	CT	TT	AT	CC	CC	AC	CT	GG	⋯
NA18547	CC	CC	CT	TT	AT	CC	CC	AC	CT	GG	⋯
NA18550	CC	CC	CT	TT	AT	CC	CC	AC	CT	GG	⋯
NA18552	CC	CC	TT	TT	TT	CC	CG	AC	CT	GG	⋯
NA18555	CC	CC	TT	TT	TT	CC	CG	AC	CT	GG	⋯
NA18558	CC	NA	CC	TT	TT	CC	CG	CC	CT	GG	⋯
NA18561	CC	CC	TT	TT	TT	AC	CC	AC	CT	GG	⋯
NA18562	CC	CC	TT	TT	AT	AC	CG	AC	CT	GG	⋯
NA18563	CC	CC	CT	TT	AA	CC	CC	AA	CT	GG	⋯
NA18564	CC	CC	TT	TT	TT	AC	CC	AC	CT	GG	⋯
NA18566	CC	CC	TT	TT	TT	AC	CC	AC	CT	GG	⋯
NA18570	CC	CC	TT	TT	AT	AC	CC	AC	CT	GG	⋯
NA18571	CC	CC	TT	TT	AT	AC	CC	AC	CT	GG	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

What distribution does a SNP have?

Multinomial distribution

Binomial: $X_1 = \#Heads, X_2 = \#Tails$

$$P(X_1 = x_1 \cap X_2 = x_2) = \binom{n}{x_1} p^{x_1}(1-p)^{x_2}$$

Trinomial: $X_1 = \#AA, X_2 = \#AB, X_3 = \#BB$

$$P(X_1 = x_1 \cap X_2 = x_2 \cap X_3 = x_3) = \binom{n}{x_1 x_2 x_3} p_1{}^{x_1} p_2{}^{x_2} p_3{}^{x_3}$$

Multinomial: $X_1 = x_1, X_2 = x_2, \ldots, X_q = x_q$

$$P(X_1 = x_1 \cap X_2 = x_2 \cap \ldots \cap X_q = x_q) = \binom{n}{x_1 x_2 \ldots x_q} p_1{}^{x_1} p_2{}^{x_2} \ldots p_q{}^{x_q}$$

In all cases, the ML estimator is:

$$\hat{p}_i = \frac{x_i}{n}$$

Hardy-Weinberg equilibrium

- A biological population of n individuals.

- A bi-allelic genetic marker.

- One locus with alleles A and B, frequencies p and q.

- Three genotypes AA, AB, BB frequencies f_{AA}, f_{AB} and f_{BB}.

$$
\begin{array}{c}
& & & \female & \\
& & p & q \\
& & A & B \\
\male \quad p & A & \boxed{p^2 \quad pq} \\
q & B & \boxed{pq \quad q^2}
\end{array}
\qquad
\begin{array}{ccc}
f_{AA} & f_{AB} & f_{BB} \\
\hline
p^2 & 2pq & q^2
\end{array}
$$

- Equilibrium achieved in one generation.

- Note that the allele frequency of A in the new generation is
$p' = \frac{2p^2 + 2pq}{2} = p^2 + pq = p(p+q) = p.$

LR test for Hardy-Weinberg equilibrium

Let $X \sim MN(\boldsymbol{\theta})$. We wish to test

$$\begin{cases} H_0 : \mathcal{F}(x; \boldsymbol{\theta}) = \mathcal{F}_0(x; \boldsymbol{\theta} = (p^2, 2pq, q^2)) \\ H_1 : \mathcal{F}(x; \boldsymbol{\theta}) = \mathcal{F}_0(x; \boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)) \end{cases}$$

$$L_0 = \binom{n}{x_1 x_2 x_3} p^{2x_1} 2pq^{x_2} q^{2x_3} \qquad L_1 = \binom{n}{x_1 x_2 x_3} \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}$$

$$\max L_0 : \hat{p} = (2x_1 + x_2)/n, \hat{q} = 1 - \hat{p} \quad \text{(sample allele frequencies)}$$

$$\max L_1 : \hat{\theta}_1 = x_1/n, \hat{\theta}_2 = x_2/n, \hat{\theta}_3 = x_3/n \quad \text{(sample genotype frequencies)}$$

LR test:

$$-2 \log \frac{L_0}{L_1} \sim \chi_1^2$$

Example:

```
library(HardyWeinberg)
x <- c(AA=298,AT=489,TT=213)
HWLratio(x,verbose=TRUE)
Likelihood ratio test for Hardy-Weinberg equilibrium
G2 = 0.2214663 DF = 1 p-value = 0.637925
```

Example: testing goodness-of-fit with the multinomial distribution

Let $X \sim \mathcal{F}(x; \theta)$. We wish to test

$$\begin{cases} H_0 : \mathcal{F}(x; \theta) = \mathcal{F}_0(x; \theta) \\ H_1 : \mathcal{F}(x; \theta) \neq \mathcal{F}_0(x; \theta) \end{cases}$$

$\mathcal{F}_0$ maybe any distribution (Normal, Exponential, ...) We have values $X_1, \ldots, X_n$, and group these in exclusive classes $C_1, C_2, \ldots, C_m$, on $C_1 = (-\infty, c_1], C_2 = (c_1, c_2], \ldots, C_m = (c_{m-1}, c_m], C_{m+1} = (c_m, \infty)$
Define

$$p_j = P\left(X \in C_j\right) = p_j(\theta) \qquad \text{for } j = 1, \ldots, m+1$$

and

$$p_j^0 = P\left(X \in C_j \mid X \sim \mathcal{F}_0(x; \theta)\right) = p_j^0(\theta)$$

We can rewrite our test as

$$\begin{cases} H_0 : p_1 = p_1^0(\theta), \ldots, p_{m+1} = p_{m+1}^0(\theta) \\ H_1 : \text{some } p_j \neq p_j^0(\theta) \end{cases}$$

Let $Y_j = \sum_{i=1}^n 1_{X_i \in C_j}$. Then

$$(Y_1, \ldots, Y_{m+1}) \sim MN(n, p_1, \ldots, p_{m+1})$$

Example: testing goodness-of-fit with the multinomial distribution

The likelihood function is:

$$L(\theta) = \frac{n!}{y_1! \dots y_{m+1}!} \, p_1(\theta)^{y_1} \dots p_{m+1}(\theta)^{y_{m+1}}$$

and has a maximum in $\widehat{p}_j = \frac{y_j}{n}$. Let $\widehat{\theta}$ be the MLE of θ calculated using the sample $X_1, \dots, X_n$. Then the MLE under H_0 is given by $p_j(\widehat{\theta})$.
The likelihood ratio is

$$\Lambda(y_1, \dots, y_{m+1}) = \frac{p_1(\widehat{\theta})^{y_1} \dots p_{m+1}(\widehat{\theta})^{y_{m+1}}}{\widehat{p}_1^{\,y_1} \dots \widehat{p}_{m+1}^{\,y_{m+1}}} = \prod_{j=1}^{m+1} \left(\frac{p_j(\widehat{\theta})}{\widehat{p}_j} \right)^{y_j}$$

We use the asymptotic distribution of the likelihood ratio

$$-2 \, log\Lambda = -2 \sum_{j=1}^{m+1} y_j \, log\left(\frac{p_j(\widehat{\theta})}{\widehat{p}_j} \right) = -2 \, n \sum_{j=1}^{m+1} \widehat{p}_j \, log\left(\frac{p_j(\widehat{\theta})}{\widehat{p}_j} \right)$$
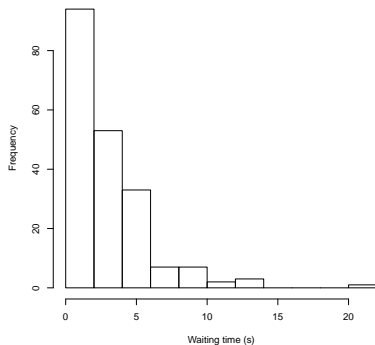
Thus, $-2 \, log\Lambda$ follows $\mathcal{X}^2_{m+1-k-1} = \mathcal{X}^2_{m-k}$ under H_0, with k the number of parameters estimated from the data. By using a Taylor expansion, tt can be shown that $-2 \, log\Lambda$ is asymptotically equivalent to

$$\sum_{j=1}^{m+1} \frac{(y_j - n \, p_j(\widehat{\theta}))^2}{n \, p_j(\widehat{\theta})}$$

and this is the usual chi-square statistic for goodness-of-fit. We will see examples later on.

The distribution of waiting times

Histogram of waiting times

	obs	pr	exp	chi2
(0,2]	94.00	0.48	96.80	0.10
(2,4]	53.00	0.25	50.00	0.20
(4,6]	33.00	0.13	25.80	2.00
(6,8]	7.00	0.07	13.20	2.90
(8,10]	7.00	0.03	6.80	0.00
(10,50]	6.00	0.04	7.40	0.30

$$X^2 = 0.1 + 0.2 + \cdots + 0.3 = 5.5$$

$$\chi^2_{4,0.95} = 9.49$$

$$P\left(\chi^2_4 \geq 5.5\right) = 0.24$$

```
     N  N*  Mean  Stdev   Med    Q1    Q3    Min    Max
X  200   0  3.022    2.9  2.316  0.954  4.274  0.013  21.633
```

$$\hat{\lambda} = 1/3.022 = 0.33$$

One-sample test for a proportion

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

The LR criterion leads to the test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1 - \hat{p})/n}}$$

Compare with a N(0,1) distribution.

Example

In Spain, in 1983, 475743 children were born, 246124 boys and 229619 girls.
Is it true that P ("newborn child is a boy") $= P$ ("newborn child is girl") $= 0.5$?

$$\begin{cases} H_0 : p_{boy} = p_{girl} = p = 0.5 \\ H_1 : p \neq 0.5 \end{cases}$$

$$\hat{p} = 0.5173466$$

$$Z = \frac{0.5173466 - 0.50}{\sqrt{0.5173466(1 - 0.5173466)/475743}} = 23.94368$$

$$\text{p-value} = 2 \cdot P(Z \geq 23.94368) = 1.075362e - 126$$

$$CI_{0.95} = (0.5159266, 0.5187665)$$

R output

```
> prop.test(boy,tot)

1-sample proportions test with continuity correction

data:  boy out of tot, null probability 0.5
X-squared = 572.5402, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5159254 0.5187674
sample estimates:
        p
0.5173466


> prop.test(boy,tot,correct=FALSE)

1-sample proportions test without continuity correction

data:  boy out of tot, null probability 0.5
X-squared = 572.6096, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5159265 0.5187663
sample estimates:
        p
0.5173466
```

Two-sample test for variances

$$\left\{ \begin{array}{l} H_0 : \sigma_x^2 = \sigma_y^2 \\ H_1 : \sigma_x^2 \neq \sigma_y^2 \end{array} \right.$$

$$\frac{s_x^2}{s_y^2} = \frac{\frac{(n_x-1)s_x^2}{\sigma_x^2} \frac{\sigma_x^2}{n_x-1}}{\frac{(n_y-1)s_y^2}{\sigma_y^2} \frac{\sigma_y^2}{n_y-1}} = \frac{\frac{U}{(n_x-1)}}{\frac{V}{(n_y-1)}} \frac{\sigma_x^2}{\sigma_y^2}$$
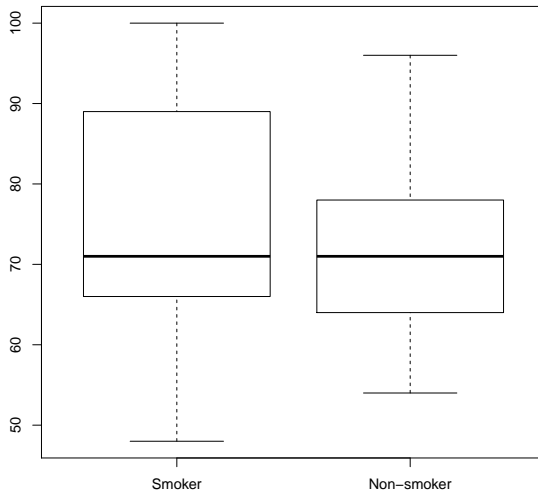
Where $U \sim \chi_{n_x-1}^2$ and $V \sim \chi_{n_y-1}^2$.
If H_0 is true then

$$F = \frac{s_x^2}{s_y^2} \sim F(n_x - 1, n_y - 1)$$

This test is also a likelihood ratio test.

Example: pulse of smokers and non-smokers

Example: pulse of smokers and non-smokers

```
   N N* Mean  Stdev Med Q1 Q3    Min Max
X 28  0 75    13.493 71 66 88.5  48 100
X 64  0 71.938 9.702 71 64 78     54  96
```

$$F = \frac{(13.493)^2}{(9.702)^2} = 1.9344$$

Critical value $F_{27,63,0.975} = 1.8334$

p-value: $2 \cdot P \left(F_{27,63} \geq 1.9344 \right) = 0.0328$

```
F test to compare two variances

data:  Pulse by Smoke
F = 1.9344, num df = 27, denom df = 63, p-value = 0.0328
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.055115 3.859449
sample estimates:
ratio of variances
         1.934427
```

Download pulse.csv

Comparing regression models

- We entertain two regression models for modeling our data:

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \qquad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

- The issue is wether x_2 is relevant for understanding y_i or not ($\beta_2 = 0$).
- Under assumption of normally distributed errors the LR criterion leads to comparing sums-of-squares of the two regressions, and this gives the F test for comparing two regression models:

$$F = \frac{RSS_{H_0} - RSS}{q s_R^2} \sim F_{q, n-p}$$

with RSS_{H_o} the residual-sum-of-squares under the reduced model, and RSS the residual-sum-of-squares of the full model, q the difference in number of parameters, and s_R^2 the variance of the residuals under the full model.

Example: student weight and height

	Alçada	Pes	Edat	Sexe	Germ	Esp	Alç.pare	Alç.mare	Pes.pare	Pes.mare	germ.m
1	169	59	22	0	2	1	168	165	70	58	5
2	178	75	20	0	2	1	170	160	68	60	2
3	185	66	21	0	2	1	173	168	83	60	1
4	175	73	21	0	2	1	173	161	73	56	1
5	178	81	21	0	2	1	170	-1	77	60	4
6	183	71	21	0	2	1	175	164	75	56	7
7	179	80	22	0	2	1	175	162	80	65	4
8	185	80	21	0	2	0	175	165	70	60	5
9	172	67	20	0	3	0	170	152	78	54	3
10	178	74	20	0	2	1	161	158	65	62	5
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.

Download PesAlcada.csv

Example: student weight and height
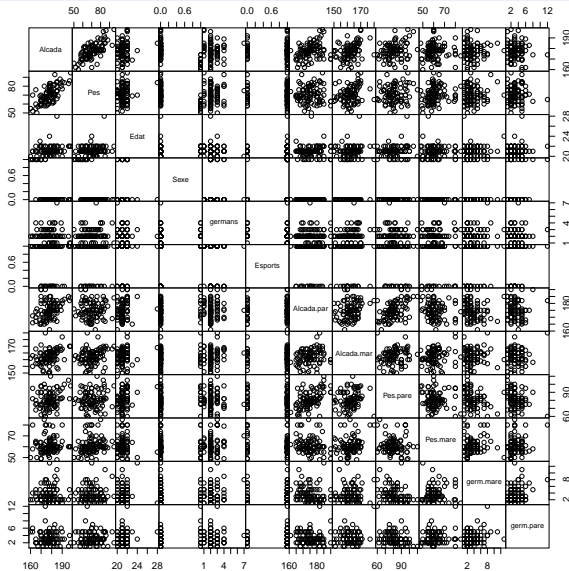
Example: student weight and height

```
 out1.lm <- lm(Pes~Alçada+Edat+Sexe+germans+Esports+Alç.pare+Alç.mare+Pes.pare+Pes.mare+germ.mare+germ.par
> summary(out1.lm)

Call:
lm(formula = Pes ~ Alçada + Edat + Sexe + germans + Esports +
    Alç.pare + Alç.mare + Pes.pare + Pes.mare + germ.mare + germ.pare,
    data = X)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -62.962540  29.861521  -2.108   0.0385 *
Alçada        0.840958   0.157498   5.339 1.04e-06 ***
Edat          0.768634   0.718386   1.070   0.2882
Sexe         -3.334807   3.094713  -1.078   0.2848
germans       0.964057   0.812694   1.186   0.2394
Esports       2.063933   1.771648   1.165   0.2479
Alç.pare     -0.315586   0.156447  -2.017   0.0474 *
Alç.mare      0.003795   0.148619   0.026   0.9797
Pes.pare      0.077042   0.090486   0.851   0.3974
Pes.mare      0.200909   0.103468   1.942   0.0561 .
germ.mare    -0.145142   0.332750  -0.436   0.6640
germ.pare     0.358032   0.463533   0.772   0.4424
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.314 on 72 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared: 0.5838,Adjusted R-squared: 0.5202
F-statistic: 9.181 on 11 and 72 DF,  p-value: 5.117e-10
```
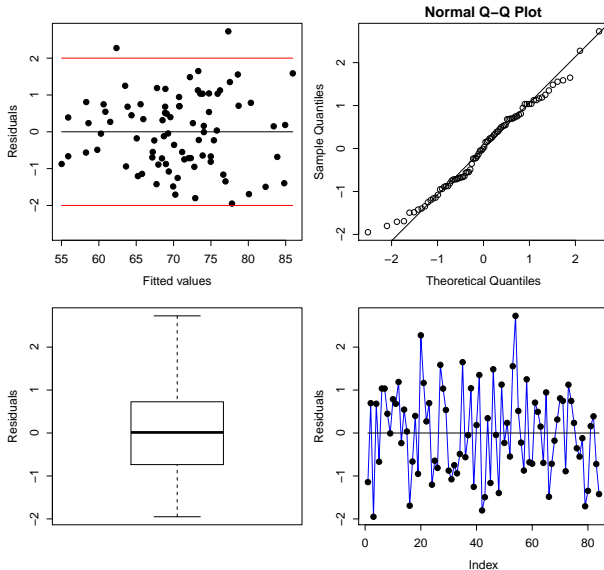
Example: student weight and height

Example: student weight and height

```
> out2.lm <- lm(Pes~Alçada+Alç.pare+Pes.mare,data=X)
> summary(out2.lm)

Call:
lm(formula = Pes ~ Alçada + Alç.pare + Pes.mare, data = X)

Residuals:
     Min      1Q   Median      3Q     Max
-12.1049  -4.7905  -0.0394  4.6255  14.5430

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -55.71624   23.29953  -2.391  0.01905 *
Alçada        0.95032    0.09577   9.923 9.35e-16 ***
Alç.pare     -0.31232    0.11401  -2.739  0.00754 **
Pes.mare      0.20070    0.09190   2.184  0.03180 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.19 on 83 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared: 0.5639,Adjusted R-squared: 0.5482
F-statistic: 35.78 on 3 and 83 DF,  p-value: 6.126e-15
```
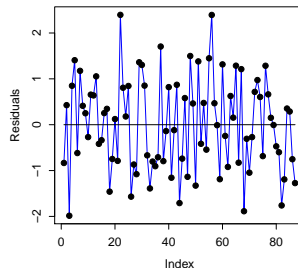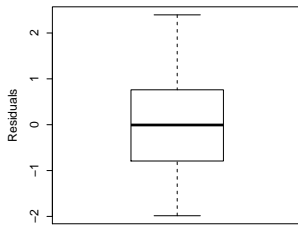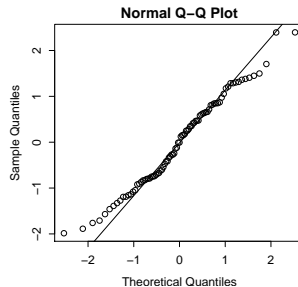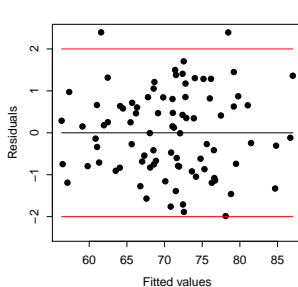
Example: student weight and height

Example model comparison

```
> anova(out2.lm)
Analysis of Variance Table

Response: Pes
           Df Sum Sq Mean Sq F value    Pr(>F)
Alcada      1 3506.0  3506.0 91.4989 4.832e-15 ***
Alcada.par  1  423.8   423.8 11.0610  0.001315 **
Pes.mare    1  182.7   182.7  4.7689  0.031800 *
Residuals  83 3180.4    38.3
```

$$\left\{ \begin{array}{l} H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_7 = \beta_8 = \beta_{10} = \beta_{11} = 0 \\ H_1 : \text{At least one different} \end{array} \right.$$

```
> anova(out1.lm)
Analysis of Variance Table

Response: Pes
           Df Sum Sq Mean Sq F value    Pr(>F)
Alcada      1 3248.9  3248.9 81.4906 1.859e-13 ***
Edat        1   18.8    18.8  0.4725   0.49406
Sexe        1  133.5   133.5  3.3495   0.07136 .
germans     1   66.9    66.9  1.6770   0.19946
Esports     1   33.5    33.5  0.8396   0.36258
Alcada.par  1  262.0   262.0  6.5718   0.01245 *
Alcada.mar  1   12.1    12.1  0.3041   0.58300
Pes.pare    1   30.9    30.9  0.7744   0.38178
Pes.mare    1  192.6   192.6  4.8299   0.03119 *
germ.mare   1    3.4     3.4  0.0858   0.77045
germ.pare   1   23.8    23.8  0.5966   0.44241
Residuals  72 2870.5    39.9
```

$RSS_{H_0} = 3180.4, \quad RSS = 2870.5, \quad q = 8, \quad (n-p) = 72.$

$$F = \frac{(3180.4 - 2870.5)/8}{2870/72} = 0.97$$

p-value $= 0.465$

References

- DeGroot, M. H. & Schervish, M. J. (2002) Probability and Statistics, Third edition, Addison-Wesley