Hypothesis tests

Aldo Solari
Statistical Inference II
PhD in Economics, Statistics and Data Science
University of Milano-Bicocca

XXXVII cycle

Outline

Simple null hypotheses

Composite null hypotheses

Main references

- Cox and Hinkley (1976) Theoretical Statistics. Chapman and Hall/CRC, §4, §5

The Neyman–Pearson formulation of hypothesis testing requires to fix the probability of rejecting H_0 when it is true, denoted by α, aiming to maximize the probability of rejecting H_1 when false.

This approach demands the explicit formulation of the *alternative hypothesis* H_1.

The decision procedure, i.e. rejecting or not H_0, is called the *test* of H_0 against H_1.

Suppose Y has distribution $f_Y(y;\theta)$ for $\theta \in \Theta$

Formulate a null hypothesis $H_0 : \theta \in \Theta_0$ and an alternative hypothesis $H_1 : \theta \in \Theta_1$ with $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$

A *test* or *critical function* $\phi = \phi(Y)$ assigns to each possible value y one of these two decisions

$$\phi : \mathcal{Y} \mapsto \{0, 1\}$$

where 1 denotes the decision of rejecting H_0 and 0 denotes the decision of not rejecting H_0, and thereby partition the sample space $\mathcal{Y}$ into two complementary regions $\mathcal{Y}_0$ and $\mathcal{Y}_1$

When performing a test one may arrive at the correct decision, or one may commit one of two errors: rejecting H_0 when it is true (*type I error*) or not rejecting it when it is false (*type II error*).

Critical region

Unfortunately, the probabilities of the two types of error cannot be controlled simultaneously

Choose the *level of significance* $\alpha \in (0, 1)$, and control the probability of type I error at α, i.e.

$$\text{pr}_\theta(Y \in \mathcal{Y}_1) \leq \alpha \quad \forall\, \theta \in \Theta_0$$

The *size* of the test is

$$\sup_{\theta \in \Theta_0} \text{pr}_\theta(Y \in \mathcal{Y}_1)$$

If, for all α, the size of the test is α, we call $\mathcal{Y}_1$ a *critical region of size* α, denoted by $\mathcal{Y}_\alpha$

Power function

Subject to

$$\sup_{\theta \in \Theta_0} \mathrm{pr}_\theta(Y \in \mathcal{Y}_\alpha) = \alpha$$

it is desired to maximize

$$\mathrm{pr}_\theta(Y \in \mathcal{Y}_\alpha) \quad \forall\ \theta \in \Theta_1$$

Considered as a function of θ, this probability is called the *power function* of the test

$$\mathrm{pow}(\theta; \alpha) = \mathrm{pr}_\theta(Y \in \mathcal{Y}_\alpha; \theta)$$

p-value

If we require that the rejection regions $\mathcal{Y}_\alpha$ and $\mathcal{Y}_{\tilde{\alpha}}$ are *nested* in the sense that

$$\mathcal{Y}_\alpha \subset \mathcal{Y}_{\tilde{\alpha}} \quad \text{if } \alpha < \tilde{\alpha}$$

the p-value is defined as the smallest significance level at which the null hypothesis would be rejected for the given observation:

$$p_{\text{obs}} = \inf\{\alpha : y \in \mathcal{Y}_\alpha\}$$

Table of Contents

In the present section we consider only the case where H_0 is a simple hypothesis

It is best to begin with a simple alternative hypothesis H_1

$$H_0 : Y \sim f_0(y) = f(y; \theta_0), \quad H_1 : Y \sim f_1(y) = f(y; \theta_1)$$

Let $\mathcal{Y}_\alpha$ and $\mathcal{Y}'_\alpha$ be two critical region of size α, i.e.

$$\mathrm{pr}_0(Y \in \mathcal{Y}_\alpha) = \mathrm{pr}_0(Y \in \mathcal{Y}'_\alpha) \tag{1}$$

We regard $\mathcal{Y}_\alpha$ as preferable to $\mathcal{Y}'_\alpha$ for the alternative H_1 if

$$\mathrm{pr}_1(Y \in \mathcal{Y}_\alpha) > \mathrm{pr}_1(Y \in \mathcal{Y}'_\alpha) \tag{2}$$

The region $\mathcal{Y}_\alpha$ is called the *best critical region* of size α if (2) is satisfied for all other $\mathcal{Y}'_\alpha$ satisfying the size condition (1).

We call $\mathrm{pr}_1(Y \in \mathcal{Y}_\alpha)$ the *size α power* of the test against H_1

Neyman-Pearson lemma

For simplicity, suppose that the likelihood ratio $\mathrm{lr}(Y) = f_1(Y)/f_0(Y)$ is, under H_0, a continuous random variable such that for all α, there exists a unique c_α such that

$$\mathrm{pr}_0(\mathrm{lr}(Y) \geq c_\alpha) = \alpha$$

We call the region defined by

$$\mathrm{lr}(y) \geq c_\alpha$$

the size α *likelihood ratio critical region*

A fundamental result, called the Neyman-Pearson lemma, is that,*for any size α, the likelihood ratio critical region is the best critical region.*

Let $\mathcal{Y}_\alpha$ be the likelihood ratio critical region and let $\mathcal{Y}_1$ be any other critical region, both being of size α. Then

$$\alpha = \int_{\mathcal{Y}_\alpha} f_0(y)dy = \int_{\mathcal{Y}_1} f_0(y)dy$$

so that

$$\int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_0(y)dy = \int_{\mathcal{Y}_1 \setminus \mathcal{Y}_\alpha} f_0(y)dy$$

since $\int_{\mathcal{Y}_\alpha} f_0(y)dy = \int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_0(y)dy + \int_{\mathcal{Y}_\alpha \cap \mathcal{Y}_1} f_0(y)dy$

Now, if $y \in \mathcal{Y}_\alpha \setminus \mathcal{Y}_1$, which is inside $\mathcal{Y}_\alpha$, $f_1(y) \geq c_\alpha f_0(y)$, while if $y \in \mathcal{Y}_1 \setminus \mathcal{Y}_\alpha$, which is outside $\mathcal{Y}_\alpha$, $c_\alpha f_0(y) > f_1(y)$.

We have that

$$\int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_1(y)dy \geq c_\alpha \int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_0(y)dy = c_\alpha \int_{\mathcal{Y}_1 \setminus \mathcal{Y}_\alpha} f_0(y)dy \geq \int_{\mathcal{Y}_1 \setminus \mathcal{Y}_\alpha} f_1(y)dy$$

with strict inequality unless the regions are equivalent

Then

$$\int_{\mathcal{Y}_\alpha} f_1(y)dy \geq \int_{\mathcal{Y}_1} f_1(y)dy$$

thus the power of $\mathcal{Y}_\alpha$ is at least that of $\mathcal{Y}_1$

Note that if $\mathcal{Y}_1$ had been of size less than α the final inequality holds

Normal mean with known variance

Let $Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$. Consider

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1$$

with $\mu_1 > \mu_0$.

$$\mathrm{lr}(y) = \exp\left\{n\bar{y}(\mu_1 - \mu_0) - \frac{1}{2}n\mu_1^2 + \frac{1}{2}n\mu_0^2\right\}$$

Because all quantities, except for $\bar{y}$, are fixed constants, and because $\mu_1 - \mu_0 > 0$, a critical region of the form $\mathrm{lr}(y) \geq c_\alpha$ is equivalent to one of the form $\bar{y} \geq d_\alpha$. Since $\bar{Y} \overset{H_0}{\sim} N(\mu_0, 1/n)$

$$d_\alpha = \mu_0 + \frac{z_\alpha}{\sqrt{n}}$$

where z_α is the $1 - \alpha$ quantile of $N(0, 1)$, and

$$\mathcal{Y}_\alpha^+ = \{y_1, \ldots, y_n : \sqrt{n}(\bar{y} - \mu_0) \geq z_\alpha\}$$

Suppose we have an observation from $N(\mu, 1)$ and that the hypotheses are $H_0 : \mu = 0$ and $H_1 : \mu = 10$.

We observe $y_{\text{obs}} = 3$. Then $p_{\text{obs}} = 1 - \Phi(y_{\text{obs}}) = 0.0013$ for testing H_0 against H_1.

On the other hand, $p_{\text{obs}} = \Phi(y_{\text{obs}} - 10) < 0.0001$ for testing H_1 against H_0.

Exponential family

Let $Y_1, \ldots, Y_n$ be i.i.d. in the single parameter exponential family

$$\exp\{a(\theta)b(y) + c(\theta) + d(y)\}$$

among them the normal, gamma, binomial and Poisson distribution, and that the hypotheses are $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$.

Then the likelihood ratio involves the data only through the sufficient statistic $S = \sum b(Y_j)$ and the best critical region has the form

$$\exp\{a(\theta_1) - a(\theta_0)\}s \geq e_\alpha$$

If $a(\theta_1) - a(\theta_0) > 0$, this is equivalent to $s \geq \tilde{e}_\alpha$, the critical region being the same for all such θ_1

Poisson mean

Let $Y \sim \text{Poisson}(\lambda)$. Consider

$$H_0 : \lambda = 1, \quad H_1 : \lambda = \lambda_1 > 1$$

The likelihood critical regions have the form $y \geq d_\alpha$

However, because Y is discrete, the only critical regions are of the form $y \geq r$, where r is an integer

r	0	1	2	3	4	5	6
$\text{pr}_0(Y \geq r)$	1	0.632	0.264	0.08	0.0189	0.0037	0.0006

If α is one of the values above, a likelihood ratio region of the required size does exist.

By a mathematical artifice, it is, however, possible to achieve likelihood ratio critical regions with other values of α

Suppose that $\alpha = 0.05$. The region $y \geq 4$ is too small, whereas the region $y \geq 3$ is too large. All values $y \geq 4$ are put in the critical region, whereas if $y = 3$ we regard the data as in the critical region with probability π such that

$$\mathrm{pr}_0(Y \geq 4) + \pi \cdot \mathrm{pr}_0(Y = 3) = 0.05$$

leading to $\pi = 0.51$. This is a *randomized critical region* of size 0.05.

The randomized definition of p_{obs} corresponding to $Y = y$ is

$$\mathrm{pr}_0(Y > y) + U \cdot \mathrm{pr}_0(Y = y)$$

where $U \sim \mathrm{Uniform}(0, 1)$, independently of Y. The corresponding random variable is, under H_0, $\mathrm{Uniform}(0, 1)$.

Observation with two possible precision

Suppose that a random variable Y is equally likely to be $N(\mu, \sigma_1^2)$ or $N(\mu, \sigma_2^2)$, where σ_1^2 and σ_2^2 are different and known.

A random variable C is observed, taking the value 1 or 2 according to whether Y has the first or second distribution. Thus it is known from which distribution y comes.

Then the likelihood of the data (c, y) is

$$f_{C,Y}(c, y) = \frac{1}{2}(2\pi\sigma_c^2)^{-\frac{1}{2}} \exp\{-(y - \mu)^2/(2\sigma_c^2)\}$$

so that $S = (C, Y)$ is sufficient for μ with σ_1^2 and σ_2^2 known.

Because $\mathrm{pr}(C = 1) = \mathrm{pr}(C = 2) = 1/2$ independent of μ, C is ancillary

Suppose $\sigma_1^2 = 1$ and $\sigma_2^2 = 10^6$ and consider $H_0 : \mu = 0$ and $H_1 : \mu = \mu_1 > 0$

If we work conditionally on c, the size α critical region is

$$\mathcal{Y}_\alpha = \begin{cases} y > z_\alpha & c = 1 \\ y > 10^3 z_\alpha & c = 2 \end{cases}$$

That is, we require

$$\mathrm{pr}(Y \in \mathcal{Y}_\alpha | C; H_0) = \alpha$$

and, subject to this, we require maximum power.

Conditional p-value

$$\mathrm{pr}(Y \geq y_{\mathrm{obs}} | C = c; H_0) = 1 - \Phi(y_{\mathrm{obs}}/\sigma_c)$$

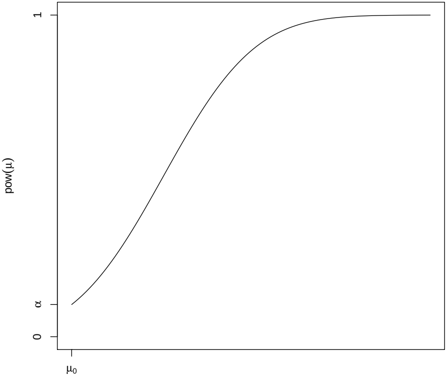
On the other hand, if we don't impose the conditional size condition, we apply the Neyman-Pearson lemma directly, we obtain a likelihood ratio critical region with more power.

The conclusion is that the requirement of using a conditional distribution cannot be deduced from that of maximum power and the two requirements may conflict.

Composite alternatives

Suppose $H_0 : \theta = \theta_0$ and $H_1 : \theta \in \Theta_1$. Two cases now arise.

- We get the same size α best critical region for all $\theta \in \Theta_1$. Then we say that the region is *uniformly most powerful* size α region. If this holds for each α, then the test itself is called uniformly most powerful (UMP).
- The best critical region depends on the particular $\theta \in \Theta_1$. Then no uniformly most powerful exists. One possibility is to take $\theta \in \Theta_1$ very close to θ_0, to maximize the power locally near the null hypothesis.



$Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$. Test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ with critical region $\mathcal{Y}_\alpha^+ = \{y_1, \ldots, y_n : \sqrt{n}(\bar{y} - \mu_0) \geq z_\alpha\}$

$$\text{pow}(\mu; \alpha) = \Phi(\sqrt{n}(\mu - \mu_0) - z_\alpha)$$

Two-sided tests

Let $Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$. Test $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$ with critical region $\mathcal{Y}_\alpha^- = \{y_1, \ldots, y_n : \sqrt{n}(\bar{y} - \mu_0) \leq -z_\alpha\}$

Suppose now that we wish to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

The critical region

$$\mathcal{Y}_\alpha = \mathcal{Y}_\alpha^- \cup \mathcal{Y}_\alpha^+$$

has size 2α, and no uniformly more powerful test exists for the two-sided alternative.

Unbiased tests

A critical region $\mathcal{Y}_\alpha$ of size α is called *unbiased* if

$$\text{pr}(Y \in \mathcal{Y}_\alpha; \theta) \geq \text{pr}(Y \in \mathcal{Y}_\alpha; \theta_0) = \alpha \quad \forall\, \theta \in \Theta_1$$

One may restrict attention to unbiased regions and among these look for the one with maximum power

A test which is uniformly most powerful amongst the class of all unbiased tests is *uniformly most powerful unbiased* (UMPU)

Type III error

Consider the problem of testing $H_0 : \theta = \theta_0$ versus $H_0 : \theta \neq \theta_0$.

If H_0 is rejected, then a decision is to be made as to whether $\theta > \theta_0$ or $\theta < \theta_0$.

We say that a Type III (or directional) error is made when it is declared that $\theta > \theta_0$ when in fact $\theta < \theta_0$ (or vice-versa).

Normal mean vector with known variance-covariance matrix

Suppose $Y = (Y_1, \ldots, Y_m)^t$ is multivariate normal with mean vector $\mu = (\mu_1, \ldots, \mu_m)^t \geq 0$ and known nonsingular covariance matrix Σ

For testing $H_0 : \mu = 0$ against $H_1 : \mu = \mu_1$, the most powerful test rejects for large values of

$$\mu_1^t \Sigma^{-1} Y$$

In particular, no UMP test exists

For testing $H_0 : \mu = 0$ against $H_1 : \mu = (k, \ldots, k)^t$ for $k > 0$, a UMP test exists and rejects for large values of the sum of the components of $\Sigma^{-1} Y$. If, in particular, Σ has diagonal elements 1 and off-diagonal elements ρ, then the test rejects when

$$\sum_i Y_i \geq z_\alpha (m + m(m-1)\rho)^{1/2}$$

Locally most powerful tests

Denote the pdf of the vector Y by $f_Y(y;\theta)$ and consider $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_0 + \epsilon$ for small $\epsilon > 0$

$$\begin{aligned}
\log \mathrm{lr}(y) &= \log f_Y(y;\theta_0+\epsilon) - \log f_Y(y;\theta_0) \\
&= \left[\log f_Y(y;\theta_0) + \epsilon \frac{\partial \log f_Y(y;\theta_0)}{\partial \theta_0} + \dots \right] - \log f_Y(y;\theta_0) \\
&= \epsilon \frac{\partial \log f_Y(y;\theta_0)}{\partial \theta_0} + \dots
\end{aligned}$$

Thus, for sufficiently small positive ϵ, we obtain the likelihood ratio critical region from large values of the *score* statistic

$$U = u(Y;\theta_0) = \frac{\partial \log f_Y(Y;\theta_0)}{\partial \theta_0}$$

In regular problems,

$$\begin{aligned}
\mathrm{E}(u(Y;\theta_0);\theta_0) &= 0 \\
\mathrm{Var}(u(Y;\theta_0);\theta_0) &= i(\theta_0) = \mathrm{E}\Big[-\frac{\partial^2 \log f_Y(Y;\theta_0)}{\partial\theta_0^2};\theta_0\Big]
\end{aligned}$$

where $i(\theta)$ the Fisher information about θ contained in Y

If $Y_1,\ldots,Y_n$ are independent, then

$U = \sum_{j=1}^n U_j$ with $U_j = \dfrac{\partial \log f_{Y_j}(Y_j;\theta_0)}{\partial\theta_0}$

$i(\theta_0) = \sum_{j=1}^n i_j(\theta_0)$ with $i_j(\theta_0) = \mathrm{Var}(u(Y_j;\theta_0);\theta_0)$

In large samples from regular models the null distribution of U is approximately normal with mean zero and variance equal to the Fisher information, so a locally most powerful critical region has form

$$\mathcal{Y}_\alpha = \{y_1, \ldots, y_n : u(y, \theta_0) \geq i(\theta_0)^{1/2} z_\alpha\}$$

Under the alternative hypothesis $H_1 : \theta = \theta_0 + \epsilon$

$$\begin{aligned} \mathrm{E}(U; \theta_0 + \epsilon) &\approx \epsilon i(\theta_0) \\ \mathrm{Var}(U; \theta_0 + \epsilon) &\approx i(\theta_0) \end{aligned}$$

hence the local power of the score test is

$$\mathrm{pr}_1\{u(y, \theta_0) \geq i(\theta_0)^{1/2} z_\alpha\} \approx \Phi(\epsilon i(\theta_0)^{1/2} - z_\alpha)$$

Exponential families

Suppose that Y_1 has the pdf in the single exponential family density

$$f_{Y_1}(y;\theta) = \exp\{a(\theta)b(y) + c(\theta) + d(y)\}$$

and that $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Then

$$\begin{aligned}
\frac{\partial \log f_{Y_1}(y;\theta_0)}{\partial \theta_0} &= a'(\theta_0)b(y) + c'(\theta_0) \\
-\frac{\partial^2 \log f_{Y_1}(y;\theta_0)}{\partial \theta_0^2} &= -a''(\theta_0)b(y) - c''(\theta_0)
\end{aligned}$$

It follows that for this single observation

$$\begin{aligned}
U_1 &= a'(\theta_0)b(y) + c'(\theta_0) \\
i_1(\theta_0) &= -a'(\theta_0)\frac{d}{d\theta_0}\left\{\frac{c'(\theta_0)}{a'(\theta_0)}\right\}
\end{aligned}$$

Location parameter of a Cauchy distribution

Let $Y_1, \ldots, Y_n$ be i.i.d. in the Cauchy distribution

$$\frac{1}{\pi[1 + (y - \theta)^2]}$$

For the null hypothesis $H_0 : \theta = \theta_0$ the score from Y_1 is

$$U_1(\theta_0) = \frac{2(Y_1 - \theta_0)}{1 + (Y_1 - \theta_0)^2}$$

and the information from a single observation is

$$i_1(\theta_0) = \frac{1}{2}$$

The test statistic is thus

$$U(\theta_0) = 2 \sum_{i=1}^{n} \frac{(Y_i - \theta_0)}{1 + (Y_i - \theta_0)^2}$$

Its null distribution has mean o and variance $n/2$

Table of Contents

A first type of composite null hypothesis is when we have a single parametric family of densities $f(y;\theta)$ with $\theta \in \Theta$ and

$$H_0 : \theta \in \Theta_0 \subset \Theta, \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

e.g. $Y_1, \dots, Y_n$ i.i.d. $N(\mu, 1)$, and

$H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$

$H_0 : \mu \in [-\Delta, \Delta]$ vs $H_1 : \mu \in (-\infty, -\Delta) \cup (\Delta, \infty)$ for some $\Delta > 0$

$H_0 : \mu \in (-\infty, -\Delta] \cup [\Delta, \infty)$ vs $H_1 : \mu \in (-\Delta, \Delta)$

A second type of composite null hypothesis is when we have a single parametric family of densities $f(y;\theta)$ where $\theta = (\psi, \lambda)$ and $\Theta = \Psi \times \Lambda$, and

$$H_0 : \psi = \psi_0, \quad H_1 : \psi \in \Psi \setminus \psi_0$$

e.g. $Y_1, \ldots, Y_n$ i.i.d. $N(\mu, \sigma^2)$ with σ^2 unknown, and

$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

UMP tests

In the one-parameter exponential family, a UMP test exists for testing

- the one-sided null hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$
- the interval null hypothesis $H_0 : \theta \leq \theta_1 \cup \theta \geq \theta_2$ against $H_1 : \theta_1 < \theta < \theta_2$

A UMP test does not exist for testing

- the two-sided null hypothesis $H_0 : \theta_1 \leq \theta \leq \theta_2$ against $H_1 : \theta < \theta_1 \cup \theta > \theta_2$

Let $Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$, and $H_0 : \mu \in (-\infty, -\Delta] \cup [\Delta, \infty)$ against $H_1 : \mu \in (-\Delta, \Delta)$ for some pre-specified $\Delta > 0$.

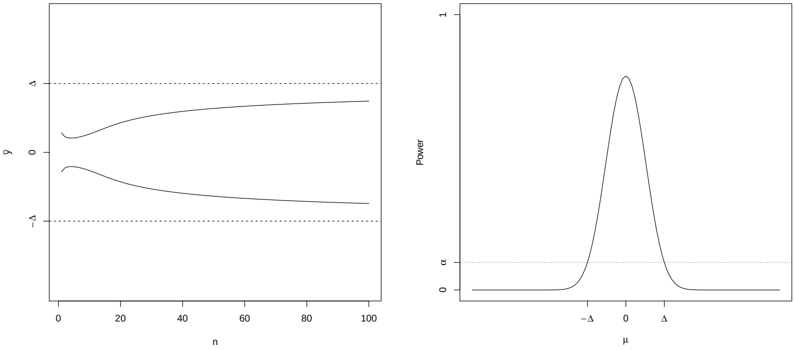
The best critical region of size α is given by

$$\mathcal{Y}_\alpha = \{(y_1, \ldots, y_n) : -\sqrt{c_\alpha/n} \leq \bar{y} \leq \sqrt{c_\alpha/n}\}$$

where c_α is the α quantile of $\chi_1^2(n\Delta^2)$

It satisfies

$$\mathrm{pr}_{-\Delta}(Y \in \mathcal{Y}_\alpha) = \mathrm{pr}_{\Delta}(Y \in \mathcal{Y}_\alpha) = \alpha$$



Δ
0
−Δ
ȳ
0
20
40
60
80
100
n
1
α
0
Power
−Δ
0
Δ
μ

Three-sided testing

Goeman, J.J., Stijnen, T. and Solari, A. (2010) Three-Sided Hypothesis Testing: Simultaneous Testing of Superiority, Equivalence and Inferiority. Statistics in Medicine, 29, 2117–2125.

$$
\begin{aligned}
H_0 &: -\Delta \leq \mu \leq \Delta && \text{(equivalence)} \\
H_+ &: \mu > \Delta && \text{(superiority)} \\
H_- &: \mu < -\Delta && \text{(inferiority).}
\end{aligned}
$$

Simultaneously test H_0, H_+ and H_- at level α using one-sided tests for H_+ and H_-, and a two-sided test for H_0.

Suppose that our parameter of interest is μ, and that we have an estimate $\hat{\mu}$ of μ, which has standard error s.

Suppose we have a t-test statistic

$$T_m = \frac{\hat{\mu} - m}{s},$$

for $H_{0,m} : \mu = m$.

Let t_α be the α quantile of the distribution of T_m under $H_{0,m}$.

1. reject H_+ if $T_\Delta \leq t_\alpha$
2. reject H_- if $T_{(-\Delta)} \geq t_{1-\alpha}$, and
3. reject H_0 if either $T_\Delta \geq t_{1-\alpha/2}$ or if $T_{(-\Delta)} \leq t_{\alpha/2}$.

Figure: Rejection regions for three-sided testing based on the t-test, as a function of Δ and $\hat{\mu}$, for fixed s. Legend: $a = st_{\alpha/2}$, $b = st_{\alpha}$, $c = st_{1-\alpha}$, $d = st_{1-\alpha/2}$.

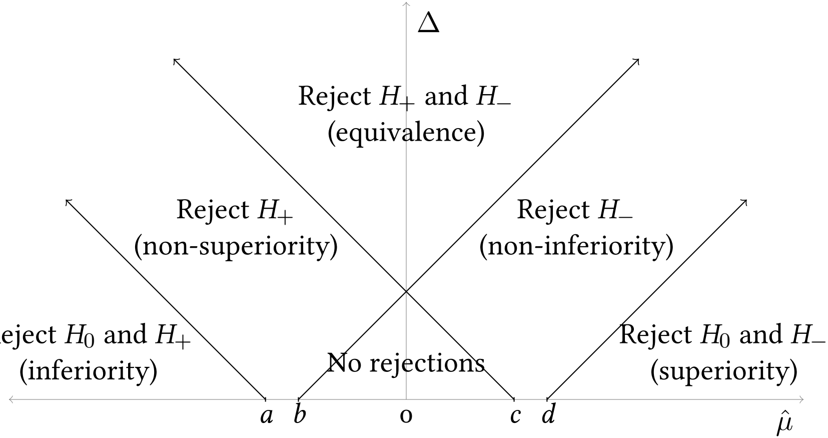
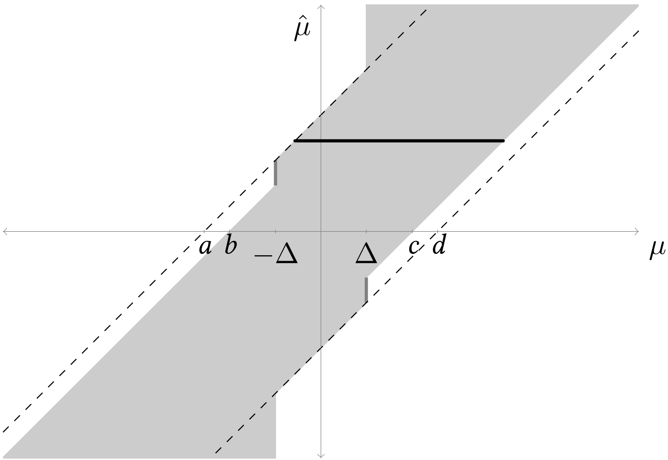


Figure: Confidence intervals consistent with three-sided testing (shading), compared with the classical confidence interval (dashed lines), based on the t-test. The thick gray lines indicate the "inclusive" part of the confidence interval. The thick black line illustrates how a confidence interval may be read off the graph for an observed value of $\hat{\mu}$. Legend: $a = st_{\alpha/2}$, $b = st_{\alpha}$, $c = st_{1-\alpha}$, $d = st_{1-\alpha/2}$.



For testing hypotheses about the parameter of interest ψ in the presence of nuisance parameters λ, a naive approach would be to fix λ at an arbitrary value, say λ^*

We can then test $H_0 : \psi = \psi_0$, obtaining a value of p_{obs} that will be a function of λ^*

We may hope that for $\lambda^* \in \Lambda$, this value of p_{obs} does not vary greatly

Suppose that $Y_1, \ldots, Y_n$ are i.i.d. $N(\mu, \sigma^2)$ and $H_0 : \mu = 0$ vs $H_1 : \mu > 0$. Then p_{obs} is

$$\text{pr}(\bar{Y} \geq \bar{y}; H_0, \sigma^*) = 1 - \Phi(\sqrt{n}\bar{y}/\sigma^*)$$

This probability varies between 1 and 1/2 if $\bar{y} < 0$, and between 0 and 1/2 if $\bar{y} > 0$; if $\bar{y} = 0$ it is 1/2

Similar regions

We require that for all $\lambda \in \Lambda$

$$\text{pr}(Y \in \mathcal{Y}_\alpha; \psi_0, \lambda) = \alpha$$

A region satisfying the above is called a *similar region of size* α

Suppose that, given $\psi = \psi_0$, S_λ is sufficient for the nuisance parameter λ. Then the conditional distribution of Y given $S_\lambda = s$ does not depend on λ when H_0 is true

If S_λ is boundedly complete, then any similar region of size α must be of size α conditionally on $S_\lambda = s$ for almost all s. We call a critical region $\mathcal{Y}_\alpha$ with this property

$$\text{pr}(Y \in \mathcal{Y}_\alpha | S_\lambda = s; \psi_0) = \alpha$$

for all s, a region of *Neyman structure*

UMPS tests

Suppose that S_λ is boundedly complete. By the Neyman-Pearson lemma, we can find the similar test with maximum power for a particular alternative hypothesis $\psi = \psi_1, \lambda = \lambda_1$, obtaining the best critical region

$$\left\{ y : \frac{f_{Y|S_\lambda}(y|s; \psi_1, \lambda_1)}{f_{Y|S_\lambda}(y|s; \psi_0)} \geq c_\alpha \right\}$$

If this same region applies to all ψ_1 and λ_1, then we call the region *uniformly most powerful similar*

Comparison of Poisson means

Suppose that Y_1 and Y_2 are independent Poisson random variables with means μ_1 and μ_2, and $H_0 : \mu_1 = \psi_0 \mu_2$ where ψ_0 is a given constant. Here we reparametrize so that $\psi = \mu_1/\mu_2$ and $\lambda = \mu_2$

Under $H_0 : \psi = \psi_0$, we have that $S_\lambda = Y_1 + Y_2$ is a complete sufficient statistic for λ

The conditional distribution of (Y_1, Y_2) given $S_\lambda = s$ is

$$f_{Y_1,Y_2|S_\lambda}(y_1, y_2|s; \psi, \lambda) = \binom{s}{y_1}(1+\psi)^{-s}\psi^{y_1}$$

If $H_1 : \psi > \psi_0$, the likelihood ratio test rejects H_0 for large y_1.

$$\text{pr}(Y_1 \geq r|S_\lambda = s; \psi_0) = \sum_{x=r}^{s} \binom{s}{y_1}(\frac{\psi_0}{1+\psi_0})^x(\frac{1}{1+\psi_0})^{s-x}$$

The test is uniformly most powerful similar

Normal mean with unknown variance

Let $Y_1, \ldots, Y_n$ be i.i.d. in $N(\mu, \sigma^2)$, both parameters unknown. Consider $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$.

Under H_0, $V(\mu_0) = \sum_{i=1}^n (Y_i - \mu_0)^2$ is a complete sufficient statistic for σ^2.

The likelihood ratio region for all alternatives $\mu_1 > \mu_0$ takes the form

$$\{y : \sum_{i=1}^n (Y_i - \mu_0) \geq c_\alpha \{v(\mu_0)\}^{1/2}\}$$

This is the one-sided Student t test, which is UMPS

If the alternatives are $\mu \neq 0$, then we are led to the two-sided Student t test

Invariant tests

Suppose that Y has probability density $f(y;\theta)$ with parameter space Θ, and $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$.

The hypothesis testing problem is then said to be invariant under a group $\mathcal{G}$ of transformations acting on the sample space if for any trasformation $g \in \mathcal{G}$, the distribution of gY is obtained from the distribution of Y by replacing θ by $g^*\theta$, such that the collection $\mathcal{G}^*$ of all such induced parameter trasformations g^* is a group on the parameter space preserving both Θ_0 and Θ_1, i.e.

for any $g \in \mathcal{G}$ and all sets $\mathcal{A}$ in the sample space

$$\text{pr}(gY \in \mathcal{A};\theta) = \text{pr}(Y \in \mathcal{A}; g^*\theta)$$

for some $g^* \in \mathcal{G}^*$ satisfying $g^*\Theta = \Theta$, $g^*\Theta_0 = \Theta_0$, $g^*\Theta_1 = \Theta_1$.

A test with critical region $\mathcal{Y}_\alpha$ is an *invariant test* if

$Y \in \mathcal{Y}_\alpha$ implies $gY \in \mathcal{Y}_\alpha$ for all $g \in \mathcal{G}$

Mean of multivariate normal distribution

Let $Y_1, \ldots, Y_n$ be a random sample from the m-variate normal distribution $N_m(\mu, \Sigma)$ with Σ unknown, and $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$.

Let $\mathcal{G}$ be the group of all non-singular $m \times m$ matrices A, so that

$$gY_i = AY_i, \quad i = 1, \ldots, n$$

The induced transformation on the parameter space is defined by

$$g^*(\mu, \Sigma) = (A\mu, A\Sigma A^t)$$

because AY_i has m-variate normal distribution $N_m(A\mu, A\Sigma A^t)$

Hotelling's test

If $n > m$, for testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, the Hotelling test statistic is

$$n(\overline{Y} - \mu_0)^t S^{-1} (\overline{Y} - \mu_0)$$

where $\overline{Y}$ is the sample mean vector and

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})(Y_i - \overline{Y})^t$$

is the sample variance/covariance matrix

Under H_0, the test statistic follows Hotelling$(m, n-1)$, i.e.

$$\frac{m(n-1)}{n-m} F_{m,n-m}$$

where $F_{m,n-m}$ is the F distribution with parameters m and $n-m$

Squared Student and Hotelling test statistics have a similar form:

$$
\begin{aligned}
(\mathrm{uvn})(\frac{\mathrm{Chisquared}}{\mathrm{df}})^{-1}(\mathrm{uvn}) &= \sqrt{n}(\bar{y}-\mu_0)[s^2]^{-1}\sqrt{n}(\bar{y}-\mu_0) \\
(\mathrm{mvn})^t(\frac{\mathrm{Wishart}}{\mathrm{df}})^{-1}(\mathrm{mvn}) &= \sqrt{n}(\bar{Y}-\mu_0)^t[S]^{-1}\sqrt{n}(\bar{Y}-\mu_0)
\end{aligned}
$$

where under H_0

$\sqrt{n}(\bar{y}-\mu_0)/\sigma \sim N(0,1)$

$s^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2/(n-1)$ with $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$,

$\sqrt{n}\Sigma^{-1/2}(\bar{Y}-\mu_0) \sim N_m(0, I_m)$

$(n-1)\Sigma^{-1/2}S\Sigma^{-1/2} \sim \mathrm{Wishart}(I_m, n-1)$

Hotelling's test is the most powerful test in the class of tests that are invariant to non-singular linear transformations

$$Y_i \mapsto AY_i + b$$

for a non-singular $m \times m$ matrix A and any $m \times 1$ vector b

Invariance here means that no direction away from μ_0 should receive special emphasis. Hotelling's test is equally powerful in all directions of the μ space, which is a strong condition.

A UMPS test will not exist, because any specific alternative μ_1 indicates a preferred direction in which the t test based on $\mu_1^t \overline{Y}$ is uniformly most powerful.

Pulmonary data

Changes in pulmonary function of 12 workers after 6 hours of exposure to cotton dust.

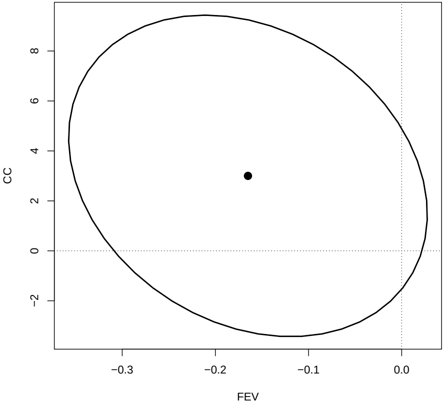
A data frame with 12 observations on the following 3 variables:

- `FVC` : change in FVC (forced vital capacity) after 6 hours.
- `FEV` : change in FEV_3 (forced expiratory volume) after 6 hours.
- `CC` : change in CC (closing capacity) after 6 hours.

Test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$: t_{obs} = 14.018, p_{obs} = 0.0512

i	1	2	3	4	5	6	7	8	9	10	11	12
FVC	-0.11	0.02	-0.02	0.07	-0.16	-0.42	-0.32	-0.35	-0.10	0.01	-0.10	-0.26
FEV	-0.12	0.08	0.03	0.19	-0.36	-0.49	-0.48	-0.30	-0.04	-0.02	-0.17	-0.30
CC	-4.30	4.40	7.50	-0.30	-5.80	14.50	-1.90	17.30	2.50	-5.60	2.20	5.50

Confidence region



The $(1-\alpha)$ confidence region is a hyperellipsoid centered at $\bar{Y}$

$$C_\alpha = \left\{ \mu : n(\bar{Y}-\mu)^t S^{-1}(\bar{Y}-\mu) \leq \frac{m(n-1)}{n-m} F_{m,n-m,\alpha} \right\}$$

Let $X_i = a^t Y_i$ for $i = 1, \ldots, n$ and $a = (a_1, \ldots, a_m)^t \in \mathbb{R}^m$. Then X_i is normal with $\mu_x = \mathrm{E}(X_i) = a^t\mu$ and $\sigma_x^2 = \mathrm{Var}(X_i) = a^t\Sigma a$

The squared Student t statistic is

$$\frac{(\bar{X} - \mu_x)^2}{s_x^2/n} = \frac{n(a^t\bar{Y} - a^t\mu)^2}{a^tSa}$$

The invariant Hotelling statistic is the largest of all such squared Student t statistics

$$\max_a \frac{n(a^t\bar{Y} - a^t\mu)^2}{a^tSa} = n(\bar{Y} - \mu)^t S^{-1}(\bar{Y} - \mu)$$

which occurs when $a \propto S^{-1}(\bar{Y} - \mu)$

Simultaneous confidence interval

A $(1-\alpha)$ confidence interval for $\mu_x = a^t\mu$ is

$$\bar{x} - \frac{s_x}{\sqrt{n}} c_\alpha \leq \mu_x \leq \bar{x} + \frac{s_x}{\sqrt{n}} c_\alpha$$

where $c_\alpha = t_{n-1;\alpha/2}$

A $(1-\alpha)$ simultaneous confidence interval for all $\mu_x = a^t\mu$ with $a \in \mathbb{R}^p$ is

$$\bar{x} - \frac{s_x}{\sqrt{n}} d_\alpha \leq \mu_x \leq \bar{x} + \frac{s_x}{\sqrt{n}} d_\alpha$$

where $d_\alpha^2 = \frac{m(n-1)}{n-m} f_{m,n-m;\alpha}$. It guarantees

$$\text{pr}(\tilde{L}_\alpha \leq \mu_x \leq \tilde{U}_\alpha, \forall\, a \in \mathbb{R}^p) \geq 1-\alpha$$

	$\bar{y}$	L_α	U_α	$\tilde{L}_\alpha$	$\tilde{U}_\alpha$
FVC	-0.14	-0.16	-0.13	-0.17	-0.12
FEV	-0.16	-0.20	-0.13	-0.22	-0.11
CC	3.00	-31.98	37.98	-56.82	62.82

Prediction region for a future observation

Suppose Y_i i.i.d. $N_m(\mu, \Sigma)$, and $\bar{Y}$ and S have been calculated from a sample of n observations

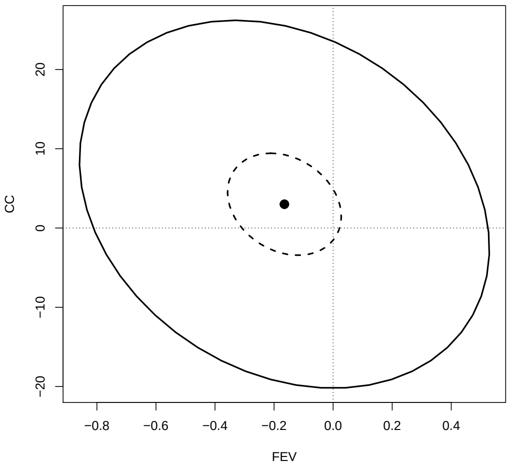
If Y_{n+1} is some new observation sampled from $N_m(\mu, \Sigma)$, then

$$\frac{n}{n+1}(Y_{n+1} - \bar{Y})^t S^{-1}(Y_{n+1} - \bar{Y}) \sim \frac{(n-1)m}{n-m} F_{m,n-m}$$

given that $\mathrm{Var}(Y_{n+1} - \bar{Y}) = \mathrm{Var}(Y_{n+1}) + \mathrm{Var}(\bar{Y}) = \frac{n+1}{n}\Sigma$

The $(1-\alpha)$ prediction ellipsoid is given by all y that satisfy

$$(y - \bar{Y})^t S^{-1}(y - \bar{Y}) \leq \frac{(n^2-1)m}{n(n-m)} f_{m,n-m;\alpha}$$


CC
FEV
−20
−10
0
10
20
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4