# Test d'ipotesi
# Introduzione:seconda parte


Alessandra Nardi

alenardi@mat.uniroma2.it

Cosa cambia se la nostra variabile risposta è continua ...

Obiettivo del nostro studio è valutare il livello di espressione di un gene in pazienti affetti da leucemia mieloide acuta

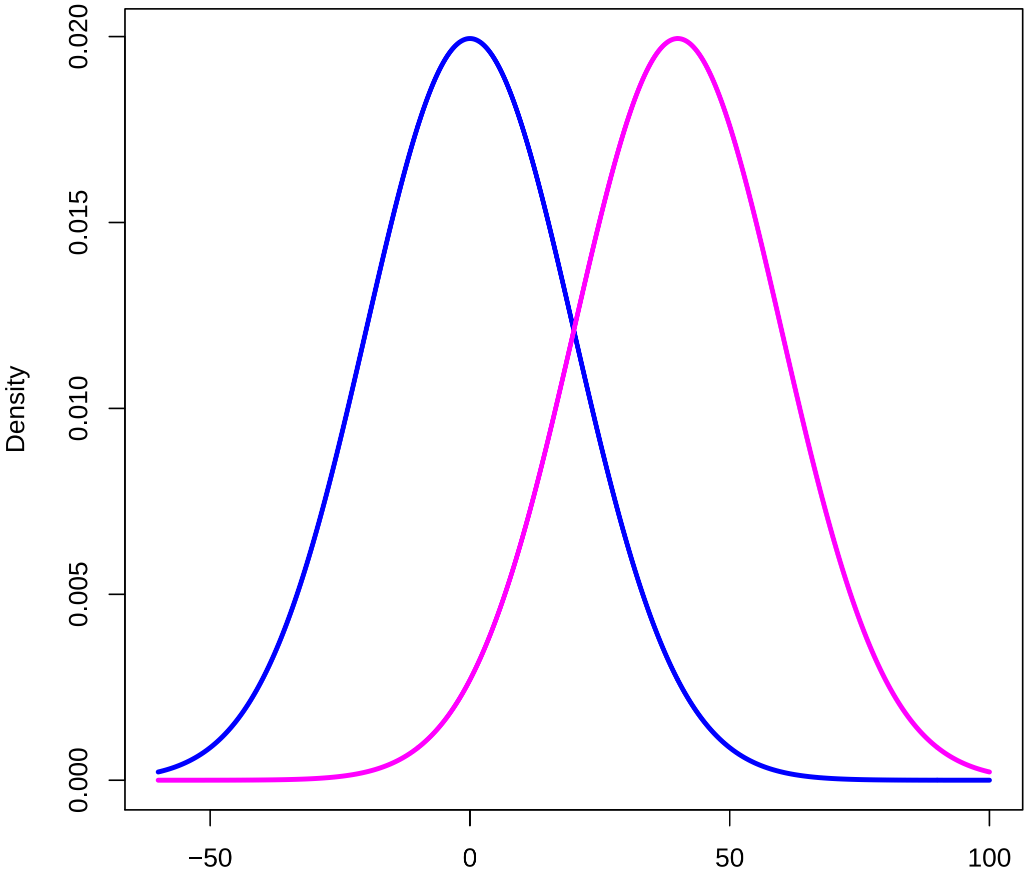
Assumiamo che in questi pazienti il livello di espressione del gene in esame segua un modello Normale di valore atteso $\mu$ e varianza $\sigma^2$

Vogliamo valutare il sistema d'ipotesi

$$
\begin{array}{lcll}
H_0 & : & \mu = 0 & \text{Ipotesi nulla: gene non espresso (curve di Andrews)} \\
H_1 & : & \mu = 40 & \text{Ipotesi alternativa: gene sovraespresso}
\end{array}
$$

Ipotizziamo di conoscere il valore di $\sigma = 20$ a priori

L'unico elemento incognito resta il valore atteso dell'espressione genica (ma la nostra incertezza e' limitata alle due possibili ipotesi)


Density
0.000
0.005
0.010
0.015
0.020
−50
0
50
100

Immaginiamo di aver estratto casualmente un campione di 10 pazienti su ciascuno dei quali e' stato osservato il livello di espressione del gene

In simboli

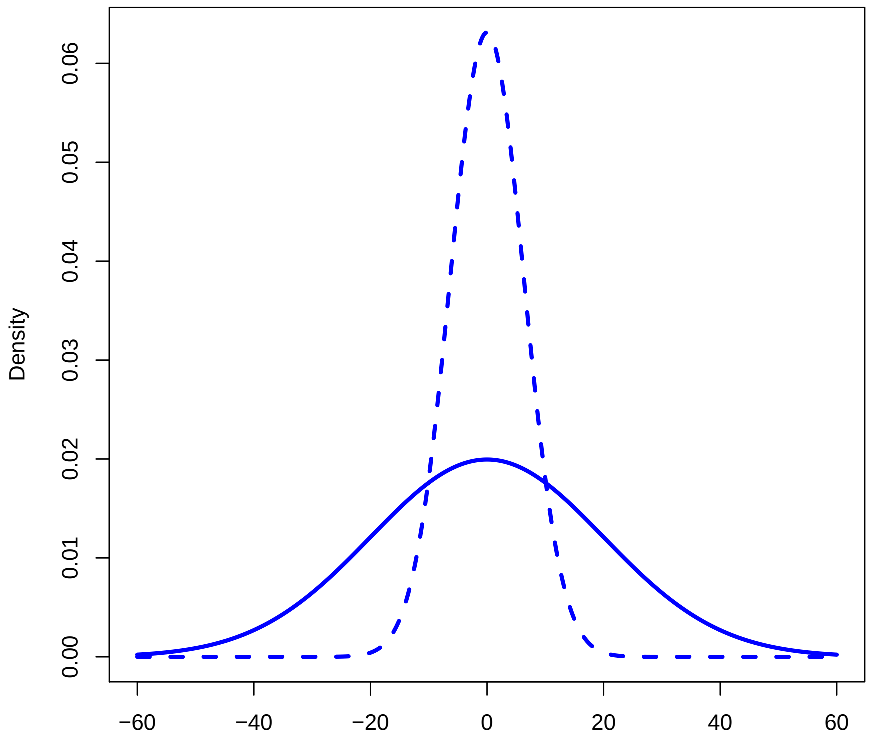
$(X_1, \ldots, X_{10})$ dove $X_i$ i.i.d. con $X_i \sim N(\mu, \sigma^2)$

Scegliamo come statistica test la media campionaria, stimatore naturale di $\mu$

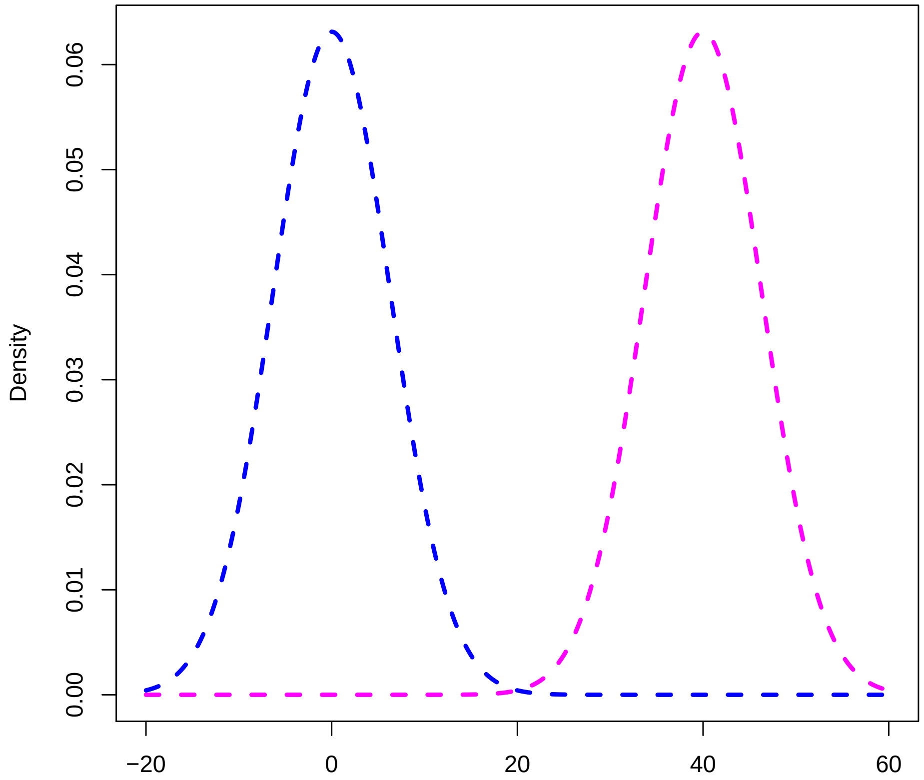
Dall'assunzione di un modello normale segue

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

Confrontiamo la distribuzione della singola osservazione (linea continua) e quella della media campionaria delle nostre 10 osservazioni (linea tratteggiata) sotto l'ipotesi nulla



Le distribuzioni della statistica test sotto l'ipotesi nulla e l'ipotesi alternativa sono adesso ben separate ...



Nella realtà la situazione si complica in primo luogo perche' l'ipotesi alternativa è spesso composta

$$\left\{ \begin{array}{lcll} H_0 & : & \mu = 0 & \text{Ipotesi nulla: gene non espresso} \\ H_1 & : & \mu > 0 & \text{Ipotesi alternativa: gene espresso} \end{array} \right\}$$

Questo implica che avremo un insieme di distribuzioni campionarie per la nostra statistica test sotto l'ipotesi alternativa, una per ogni possibile valore di $\mu > 0$ e la potenza del test (e la probablità di errore di seconda specie) diventera' una funzione di $\mu$

Inoltre generalmente $\sigma^2$ non e' noto e sara' necessario ricorrere al suo stimatore

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$$

La costruzione della nostra statistica test partirà dalla media campionaria che andremo a standardizzare assumendo vera l'ipotesi nulla

$$Z = \frac{\overline{X} - 0}{\sigma/\sqrt{n}}$$

per poi sotituire $\sigma$ con il suo stimatore

$$T = \frac{\overline{X} - 0}{S/\sqrt{n}}$$

T dipende adesso da due quantità aleatorie e la sua distribuzione campionaria sarà la distribuzione t di Student con n-1 gradi di libertà

Il test ottimale sarà caratterizzato dalla regione di rifiuto

$R = \{t\ tali\ che\ t > t_\alpha\}$ dove $t_\alpha$ si ricava dall'uguaglianza

$\int_{t_\alpha}^{+\infty} f_T(t)dt = \alpha$