

Informatica

Massimo Borelli, Ph.D.

17 marzo 2016

Gestione ed analisi dei dati

Master di Primo Livello *Metodologia della Ricerca applicata all'Infermieristica*



Dipartimento di
Scienze Mediche,
Chirurgiche e della Salute



SOCIETÀ DEI MATEMATICI
E NATURALISTI DI MODENA
www.socnatmatmo.unimore.it

Insegnamento 5 - Gestione ed analisi dei dati

L'insegnamento ha l'obiettivo di fornire allo studente le conoscenze necessarie all'esecuzione ed all'interpretazione dei principali test per la verifica delle ipotesi statistiche, nonché dell'analisi multivariata.

- reperire efficacemente la letteratura in rete
- strutturare un dataset per l'analisi dei dati
- conoscere ed utilizzare software per l'analisi statistica
- **eseguire i principali test di verifica delle ipotesi statistiche**
- produrre grafici consistenti con gli esiti misurati

il dataset monfalcone

	A	B	C	D	E	F	G	H	I	J	K
1	esito	markerNew	markerOld	eta	fumo	menopausa	etnia	statura	peso	imc	terapia
2	BENIGNO	3576	7014	34	si	PRE	africana	162	65	24,7	none
3	BENIGNO	3046	23216	21	no	PRE	africana	168	63	22,3	none
4	MALIGNO	29376	11226	64	si	POST	africana	183	65	19,4	high
5	MALIGNO	6296	5224	58	si	POST	africana	169	52	18,2	low
6	BENIGNO	3546	2069	74	no	POST	africana	171	87	29,7	low
7	BENIGNO	4056	6084	40	no	PRE	africana	170	71	24,5	low
8	MALIGNO	130456	196336	51	si	PRE	africana	144	72	34,7	low
207	BENIGNO	3776	6259	51	no	PRE	caucasica	154	55	23,1	low
208	BENIGNO	3276	1896	42	no	PRE	caucasica	186	80	23,1	low
209	BENIGNO	4116	6156	36	no	PRE	caucasica	176	69	22,2	none
210	BENIGNO	5816	903	55	no	POST	caucasica	159	46	18,1	high
211	MALIGNO	5226	5616	63	si	POST	caucasica	166	66	23,9	high

domande di ricerca

l'età delle donne e l'esito della patologia: vi è differenza?

- differenza (in media) tra due gruppi
 - > variabile categorica esito \rightarrow variabile continua eta
 - test t , modello lineare

```
t.test(eta ~ esito, var.equal = TRUE)
```

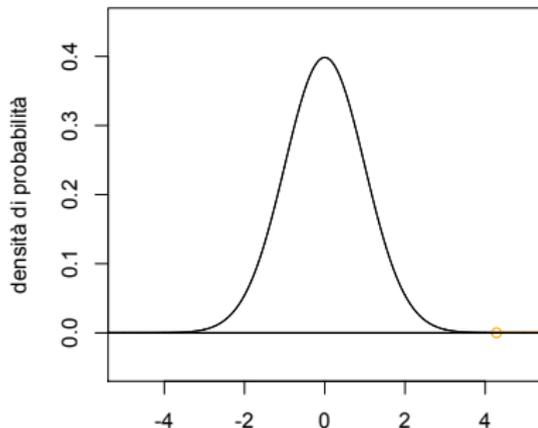
```
modello = lm(eta ~ esito)
```

```
modello = glm(eta ~ esito, family = gaussian)
```

domande di ricerca

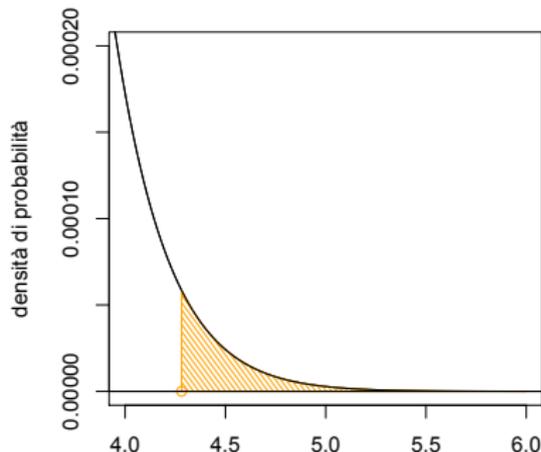
l'età delle donne e l'esito della patologia: vi è differenza?

$$58.538 - 47.233 = 11.305$$



quantili della distribuzione t di Student

$$11.305/2.640 = 4.282$$



quantili della distribuzione t di Student

domande di ricerca

l'età delle donne e l'esito della patologia: vi è differenza?

- differenza (in media) tra due gruppi
 - > variabile categorica esito \rightarrow variabile continua età
 - test t , modello lineare

domande di ricerca

la statura delle donne e l'esito della patologia: vi è differenza?

domande di ricerca

il peso delle donne e l'esito della patologia: vi è differenza?

domande di ricerca

il fumo e l'esito della patologia: vi è associazione?

- indipendenza di eventi
 - > variabile categorica fumo \rightsquigarrow variabile categorica esito
 - test di Fisher; modello lineare generalizzato binomiale

```
tabella = table(esito, fumo)
```

```
tabella
```

```
fisher.test(tabella)
```

```
modello = glm(esito ~ fumo, family = binomial)
```

```
summary(modello)
```

domande di ricerca

il fumo e l'esito della patologia: vi è associazione?

- indipendenza di eventi
 - > variabile categorica fumo \mapsto variabile categorica esito
 - test di Fisher; modello lineare generalizzato binomiale

domande di ricerca

la menopausa e l'esito della patologia: vi è associazione?

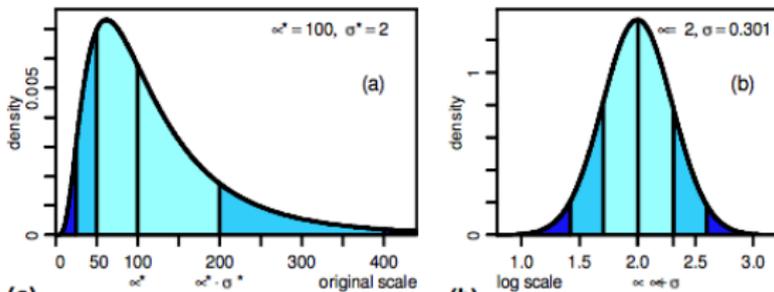
domande di ricerca

l'etnia e l'esito della patologia: vi è associazione?

domande di ricerca

l'indice di massa corporea e l'esito della patologia: vi è associazione?

ATTENZIONE! Sappiamo che il BMI non è distribuito in maniera gaussiana, ma è log-normale



hist(imc)

domande di ricerca

l'indice di massa corporea e l'esito della patologia: vi è associazione?

- indipendenza di eventi
 - > variabile continua $\log(\text{imc}) \rightarrow$ variabile categorica esito
 - test di Fisher; modello lineare generalizzato binomiale

`t.test(log(imc) ~ esito, var.equal = TRUE)`

`modello = glm(esito ~ log(imc), family = gaussian)`

domande di ricerca

Quale dei due marker è un migliore predittore dell'esito?

- metodo 'visuale'
 - > variabile continua MarkerOld, MarkerNew \rightarrow variabile categorica esito
 - curva R.O.C.

```
library(ROCR)
```

```
predizione = prediction( markerOld, esito )
```

```
risultato = performance( predizione , sens, spec)
```

```
plot(risultato , colorize=TRUE)
```