

**Statistica Applicata**  
**Corso di Laurea in Scienze Naturali**  
**a. a. 2015/2016**

prof. Federico Plazzi

12 Settembre 2016

Nome: \_\_\_\_\_

Cognome: \_\_\_\_\_

Matricola: \_\_\_\_\_

**Alcune indicazioni:**

- La prova è costituita da cinque esercizi; dopo ogni esercizio c'è lo spazio in cui scrivere la risposta o le risposte. In caso questo spazio non sia sufficiente, si può continuare a rispondere sul retro del foglio, avendo cura di indicare il numero dell'esercizio a fianco della continuazione della risposta.
- Alcuni esercizi richiedono semplici calcoli, per i quali è consentito l'uso di una calcolatrice ed eventualmente la consultazione di una o più delle tabelle allegate.
- Altri esercizi richiedono invece la lettura dei dati: verrà valutata in questo caso l'argomentazione che giustifica l'interpretazione fornita.
- La durata massima della prova è di 60 minuti.
- Si prega di non scrivere nulla sulle tabelle allegate.

# 1 Dati

La tabella 1 mostra il numero di medaglie vinte dalle prime venti nazioni nel medagliere olimpico di Rio 2016 e le relative popolazioni (riferite al 2015).<sup>1</sup>

Tabella 1: Numero di medaglie d'oro ("G"), d'argento ("S"), di bronzo ("B") e totali ("T") e popolazione ("2015 Population") di ogni nazione.

			G	S	B	T	2015 Population
1	USA	Stati Uniti	46	37	38	121	321,418,820
2	GBR	Gran Bretagna	27	23	17	67	65,138,230
3	CHN	Cina	26	18	26	70	1,371,220,000
4	RUS	Federazione Russa	19	18	19	56	144,096,810
5	GER	Germania	17	10	15	42	81,413,150
6	JPN	Giappone	12	8	21	41	126,958,470
7	FRA	Francia	10	18	14	42	66,808,380
8	KOR	Repubblica di Corea	9	3	9	21	50,617,040
9	ITA	Italia	8	12	8	28	60,802,080
10	AUS	Australia	8	11	10	29	23,781,170
11	NED	Olanda	8	7	4	19	16,936,520
12	HUN	Ungheria	8	3	4	15	9,844,690
13	BRA	Brasile	7	6	6	19	207,847,530
14	ESP	Spagna	7	4	6	17	46,418,270
15	KEN	Kenya	6	6	1	13	46,050,300
16	JAM	Giamaica	6	3	2	11	2,725,940
17	CRO	Croazia	5	3	2	10	4,224,400
18	CUB	Cuba	5	2	4	11	11,389,560
19	NZL	Nuova Zelanda	4	9	5	18	4,595,700
20	CAN	Canada	4	3	15	22	35,851,770

## 2 Esercizi

### 2.1 Statistiche di base

Calcolare la media, la varianza e la deviazione standard del numero di medaglie d'oro vinte dalle nazioni europee<sup>2</sup>.

<sup>1</sup>I dati del medagliere olimpico sono tratti da <https://www.rio2016.com/en/medal-count-country>, mentre quelli demografici da <http://data.worldbank.org/indicator/SP.POP.TOTL>; entrambi i siti sono stati consultati l'8 Settembre 2016.

<sup>2</sup>La Federazione Russa va conteggiata tra le nazioni europee.

## 2.2 Distribuzioni

Eseguiamo il test di Shapiro e Wilk sui numeri di medaglie vinte dalle 20 nazioni considerate.

Shapiro-Wilk test

```
data: Gold
W = 0.72039, p-value = 6.985e-05
```

```
data: Silver
W = 0.81684, p-value = 0.001557
```

```
data: Bronze
W = 0.87752, p-value = 0.01597
```

```
data: Total
W = 0.78136, p-value = 0.0004615
```

Cosa possiamo concludere sulla normalità dei dati?

## 2.3 Confronto con le nazioni extraeuropee

Le nazioni europee hanno vinto significativamente più medaglie d'oro delle nazioni extraeuropee?

Di seguito è mostrato il risultato di quattro test statistici (rispettivamente, test  $t$ , test  $t$  a campioni appaiati, test di Mann e Whitney, test di Wilcoxon; tutti ad una coda): qual è il test più adeguato e cosa possiamo concluderne?

Two Sample t-test

```
data: Rio2016[europa, "Gold"] and Rio2016[non.europa, "Gold"]
t = -0.16671, df = 13.524, p-value = 0.565
alternative hypothesis: true difference in means is greater than 0
```

Paired t-test

```
data: Rio2016[europa, "Gold"] and Rio2016[non.europa, "Gold"]
t = -0.35479, df = 9, p-value = 0.6345
alternative hypothesis: true difference in means is greater than 0
```

Mann-Whitney test

```
data: Rio2016[europa, "Gold"] and Rio2016[non.europa, "Gold"]
W = 64, p-value = 0.1525
alternative hypothesis: true location shift is greater than 0
```

Wilcoxon signed rank test

data: Rio2016[europa, "Gold"] and Rio2016[non.europa, "Gold"]  
V = 36, p-value = 0.2065  
alternative hypothesis: true location shift is greater than 0

## 2.4 Correlazioni

É possibile che il numero di medaglie d'oro dipenda dal numero di medaglie totali? O dalla popolazione della nazione, che si riflette nel numero di atleti e quindi nelle possibilità di accedere alle medaglie?

Di seguito sono mostrati i dettagli dei modelli di correlazione lineare tra numero di medaglie d'oro e numero di medaglie totali (Tab. 2) o popolazione della nazione (Tab. 3).

Come si possono commentare questi dati?

Tabella 2: Correlazione con le medaglie totali

	Stima	p-value	$r$	$R^2$
Intercetta	-0,32637			
Pendenza	0,36983	$3,75 \times 10^{-13}$	0,9746454	0,9499

Tabella 3: Correlazione con la popolazione

	Stima	p-value	$r$	$R^2$
Intercetta	9,795			
Pendenza	$1,708 \times 10^{-8}$	0,02723	0,4929124	0,243

## 2.5 Proporzione di medaglie d'oro

Considerando la prime tre nazioni del medagliere (Stati Uniti, Gran Bretagna e Cina), possiamo affermare che abbiano più o meno tutte vinto la stessa quantità di medaglie d'oro rispetto al loro totale di medaglie o ci sono delle differenze significative? Qual è il test statistico più adeguato? Qual è il risultato?