

Introduzione alla statistica medica

La statistica

La **statistica** comprende un insieme di metodi per:

- la raccolta
- la descrizione
- l'analisi

di dati relativi a **fenomeni che hanno
attitudine a variare**

Se si rilevano in un gruppo di individui



i valori di altezza o il sesso, ad esempio,
1.67, 1.74, 1.94, 1.78
F, F, M, M

si può notare che

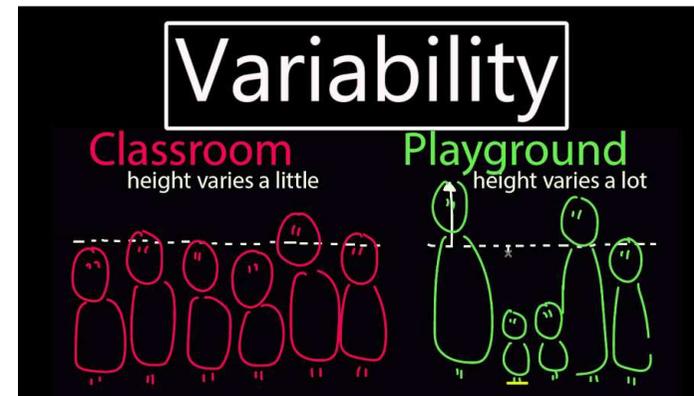
**i valori misurati variano da
individuo a individuo**

Fonti di variabilità

Strumentale



Biologica



Perché i valori cambiano da individuo a individuo?

Tra le possibili fonti di variabilità, quelle più rilevanti sono la:

- ✓ **variabilità strumentale** (legata alla procedura di misurazione, agli strumenti e a chi misura);
(ad es. la procedura operativa o lo strumento non sono ancora a punto, colui che misura non è sufficientemente esperto)
- ✓ **variabilità biologica** (intrinseca).

Perché i valori cambiano da individuo a individuo?

La variabilità strumentale può essere completamente controllata, agendo sulle modalità di misurazione.

(ad es. ottimizzando la procedura operativa, tarando lo strumento, facendo training al personale)

La variabilità biologica può essere solo parzialmente limitata, rendendo più omogeneo l'insieme di soggetti analizzati.

1.67, 1.74,
F, F

1.94, 1.78
M, M

Esempio



Pressione arteriosa del bambino in età scolare

La variabilità strumentale può essere completamente controllata addestrando il personale che effettua le misurazioni

(ad es. scelta del bracciale)

La variabilità biologica può essere parzialmente limitata identificando quei fattori che modificano la pressione arteriosa

(ad es. età, classe ponderale, familiarità)

Terminologia: universo

L'universo (o popolazione) consiste della totalità degli elementi (unità statistiche) che hanno certe caratteristiche



Es.:

☞ Studenti che seguono questa lezione di Statistica Medica

Terminologia: campione

Un **campione** è un sottoinsieme di elementi dell' universo che viene utilizzato per trarre conclusioni sulle caratteristiche dell'universo

campione  universo

Il campione non deve essere selezionato ma deve essere scelto in modo casuale



Es: Universo: Studenti che seguono questa lezione di Statistica Medica
Campione: 20 studenti presi a caso tra quelli che seguono questa lezione

Terminologia: campione

Un **campione** è un sottoinsieme di elementi dell' universo che viene utilizzato per trarre conclusioni sulle caratteristiche dell'universo

campione \Rightarrow universo

Il campione non deve essere selezionato ma deve essere scelto in modo casuale



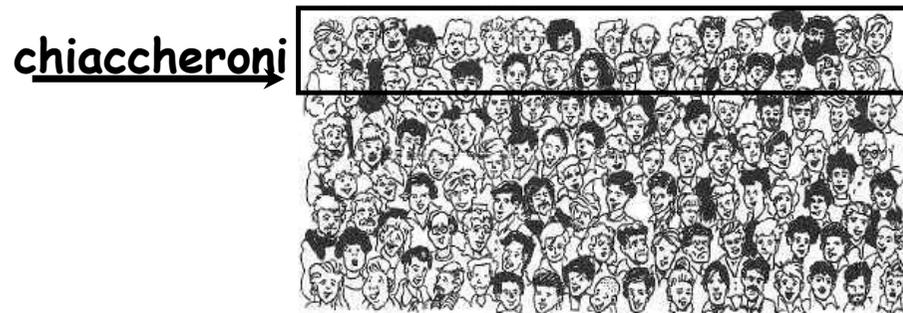
Es: Universo: Studenti che seguono questa lezione di Statistica Medica
Campione: 20 studenti presi a caso tra quelli che seguono questa lezione

Terminologia: campione

Un **campione** è un sottoinsieme di elementi dell' universo che viene utilizzato per trarre conclusioni sulle caratteristiche dell'universo

campione \Rightarrow universo

Il campione non deve essere selezionato ma deve essere scelto in modo casuale



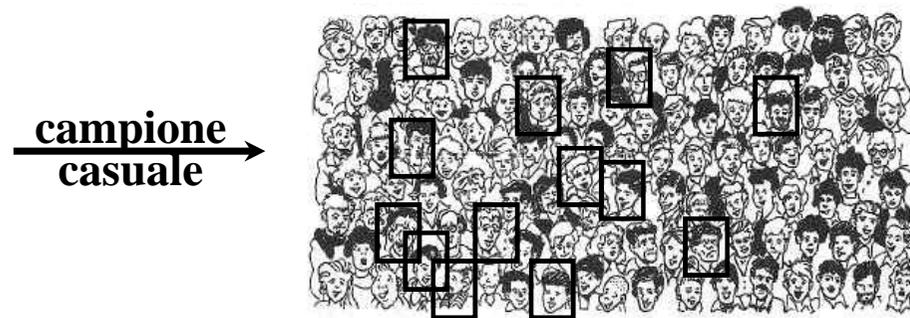
Es: Universo: Studenti che seguono questa lezione di Statistica Medica
Campione: 20 studenti presi a caso tra quelli che seguono questa lezione

Terminologia: campione

Un **campione** è un sottoinsieme di elementi dell' universo che viene utilizzato per trarre conclusioni sulle caratteristiche dell'universo

campione \Rightarrow universo

Il campione non deve essere selezionato ma deve essere scelto in modo casuale



Es: Universo: Studenti che seguono questa lezione di Statistica Medica
Campione: 20 studenti presi a caso tra quelli che seguono questa lezione

Il campione

Il campione casuale dovrebbe rappresentare una immagine in scala ridotta dell'universo.



campione come
miniatura
dell'universo



... ovvero dovrebbe essere **rappresentativo** dell'universo.

Questa è la condizione (non verificabile) di validità del processo di **generalizzazione dei risultati**.

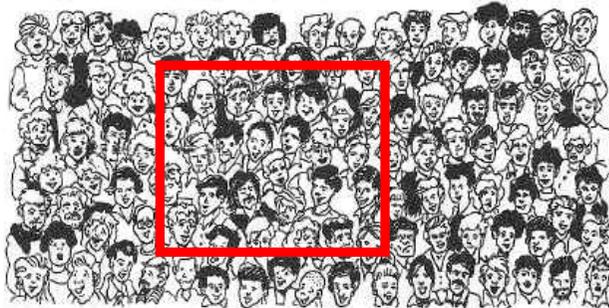
Il campione



$N=100$



1) $n=5$



2) $n=25$



3) $n=75$

Quale tra questi tre campioni contiene più informazioni sulla popolazione?

Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial



Yeming Wang*, Dingyu Zhang*, Guanhua Du*, Ronghui Du*, Jianping Zhao*, Yang Jin*, Shouzhi Fu*, Ling Gao*, Zhenshun Cheng*, Qiaofa Lu*, Yi Hu*, Guangwei Luo*, Ke Wang, Yang Lu, Huadong Li, Shuzhen Wang, Shunan Ruan, Chengqing Yang, Chunlin Mei, Yi Wang, Dan Ding, Feng Wu, Xin Tang, Xianzhi Ye, Yingshun Ye, Bing Liu, Jie Yang, Wen Yin, Aili Wang, Guohui Fan, Fei Zhou, Zhibo Liu, Xiaoying Gu, Jiuyang Xu, Lianhan Shang, Yizhang, Lianjun Cao, Tingting Guo, Yan Wan, Hong Qin, Yushen Jiang, Thomas Jaki, Frederick G Hayden, Peter W Horby, Bin Cao, Chen Wang

Summary

Background No specific antiviral drug has been proven effective for treatment of patients with severe coronavirus disease 2019 (COVID-19). Remdesivir (GS-5734), a nucleoside analogue prodrug, has inhibitory effects on pathogenic animal and human coronaviruses, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in vitro, and inhibits Middle East respiratory syndrome coronavirus, SARS-CoV-1, and SARS-CoV-2 replication in animal models.

Methods We did a randomised, double-blind, placebo-controlled, multicentre trial at ten hospitals in Hubei, China. Eligible patients were adults (aged ≥ 18 years) admitted to hospital with laboratory-confirmed SARS-CoV-2 infection, with an interval from symptom onset to enrolment of 12 days or less, oxygen saturation of 94% or less on room air or a ratio of arterial oxygen partial pressure to fractional inspired oxygen of 300 mm Hg or less, and radiologically confirmed pneumonia. Patients were randomly assigned in a 2:1 ratio to intravenous remdesivir (200 mg on day 1 followed by 100 mg on days 2–10 in single daily infusions) or the same volume of placebo infusions for 10 days. Patients were permitted concomitant use of lopinavir–ritonavir, interferons, and corticosteroids. The primary endpoint was time to clinical improvement up to day 28, defined as the time (in days) from randomisation to the point of a decline of two levels on a six-point ordinal scale of clinical status (from 1=discharged to 6=death) or discharged alive from hospital, whichever came first. Primary analysis was done in the intention-to-treat (ITT) population and safety analysis was done in all patients who started their assigned treatment. This trial is registered with ClinicalTrials.gov, NCT04257656.

Findings Between Feb 6, 2020, and March 12, 2020, 237 patients were enrolled and randomly assigned to a treatment group (158 to remdesivir and 79 to placebo); one patient in the placebo group who withdrew after randomisation was not included in the ITT population. Remdesivir use was not associated with a difference in time to clinical improvement (hazard ratio 1.23 [95% CI 0.87–1.75]). Although not statistically significant, patients receiving remdesivir had a numerically faster time to clinical improvement than those receiving placebo among patients with symptom duration of 10 days or less (hazard ratio 1.52 [0.95–2.43]). Adverse events were reported in 102 (66%) of 155 remdesivir recipients versus 50 (64%) of 78 placebo recipients. Remdesivir was stopped early because of adverse events in 18 (12%) patients versus four (5%) patients who stopped placebo early.

Interpretation In this study of adult patients admitted to hospital for severe COVID-19, remdesivir was not associated with statistically significant clinical benefits. However, the numerical reduction in time to clinical improvement in those treated earlier requires confirmation in larger studies.

Lancet 2020; 395: 1569–78

Published Online

April 29, 2020

[https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9)

This online publication has been corrected. The corrected version first appeared at [thelancet.com](https://www.thelancet.com) on May 28, 2020

See Comment page 1525

*Contributed equally

Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases (Ye Wang MD, F Zhou MD, Z Liu MD, L Shang MD, Y Zhang MD, Prof B Cao MD, Prof C Wang MD) and Institute of Clinical Medical Sciences (G Fan MS, X Gu PhD), China-

Japan Friendship Hospital, Beijing, China; Department of Respiratory Medicine, Capital Medical University, Beijing, China (Ye Wang, Prof B Cao);

Jin Yin-tan Hospital, Wuhan, Hubei, China (D Zhang MD, H Li MD, S Wang MS, S Ruan MS);

Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union

Medical College, Beijing, China (Prof G Du PhD);

Prof K Wang PhD, Prof Y Lu PhD);

Wuhan Lung Hospital, Wuhan, China (Prof R Du MD, C Yang MD,

Terminologia: variabili

Si dice **variabile** una **caratteristica** delle unità statistiche che può assumere una pluralità di valori al variare dell'unità su cui è rilevata

Es: Altezza, Sesso, Titolo di Studio, peso alla nascita

Le variabili possono essere:

- i) **quantitative**
- ii) **qualitative**

vengono indicate con lettere maiuscole scelte, in genere, tra le ultime lettere dell'alfabeto: Y, X, Z

Es: Y = Altezza X = Sesso

Terminologia: dati

I dati sono quei valori numerici o quelle modalità, assumibili da una variabile.

I dati sono rappresentati da lettere minuscole con un indice che distingue le diverse unità fra loro:

Es: $Y = \text{Altezza}$ $y_1 = 1.67$ $y_2 = 1.74$ $y_3 = 1.94$ $y_4 = 1.78$
 $X = \text{Sesso}$ $x_1 = F$ $x_2 = F$ $x_3 = M$ $x_4 = M$

Variabili quantitative discrete

Una variabile quantitativa è **discreta** se può assumere come valore un **qualsiasi numero naturale**

- Es.:*
- Numero automobili per famiglia
 - Voto esame di statistica
 - Durata dell'allattamento (in mesi)

Le variabili quantitative discrete derivano usualmente da **conteggi**

Variabili quantitative continue

Una variabile quantitativa è **continua**, se può assumere come valore un **qualsiasi numero reale**

Es.: Altezza, Peso, Concentrazione di glucosio nel sangue

I valori assunti da una variabile continua dipendono in realtà dal potere di risoluzione dello strumento di misura

Es.: Una altezza di 1.78324321.... m, potrebbe essere riportata al cm (1.78) o al mm (1.783) a seconda dell'uso

Le variabili quantitative continue derivano usualmente da **misurazioni**

Variabili qualitative nominali

Una variabile qualitativa è **nominale**, quando ogni possibile ordinamento delle modalità è arbitrario

Es: Sesso, Colore degli occhi, tipologia di parto

Etnia pazienti coinvolti in una sperimentazione clinica

caucasico - afroamericano - africano - indiano .. etc.

=

afroamericano - indiano - caucasico - africano .. etc.

Variabili qualitative ordinali

Una variabile qualitativa è **ordinale**, quando è possibile individuare un ordinamento naturale delle modalità.

Es.: Segno zodiacale, Titolo di studio

Misurazione dell'intensità del dolore

nulla < lieve < moderata < forte

forte > moderata > lieve > nulla

moderata - forte - nulla - lieve

NO!!

Variabili qualitative ordinali

Es: Misurazione dell'intensità del dolore

nulla < lieve < moderata < forte

Alle modalità si può associare un **codice numerico:**

(*Es.:* nulla=0, lieve=1, moderata=2, forte=3)

che però non ha significato quantitativo:

- ▶ 2 (dolore moderato) **non** è il doppio di 1 (dolore lieve),
3 (dolore forte) **non** è il triplo di 1
- ▶ la differenza tra 2 e 1 **non** è uguale a quella tra 3 e 2

Variabili qualitative a due livelli

Vengono chiamate anche dicotomiche (o binarie), segnalano la presenza (o l'assenza) di una caratteristica.

Es: Presenza di gravi complicazioni dopo un intervento chirurgico. Le uniche modalita' che questa variabile puo' assumere sono 'SI' , 'NO'

Dalla frequenza di 'SI' si ottiene la frequenza di 'NO' calcolando $1 - \text{la frequenza di 'SI'}$

Es: 0.12 (12%) di 'SI' implica 0.88 (88%) di 'NO'

La frequenza di 'SI' si chiama PROPORZIONE

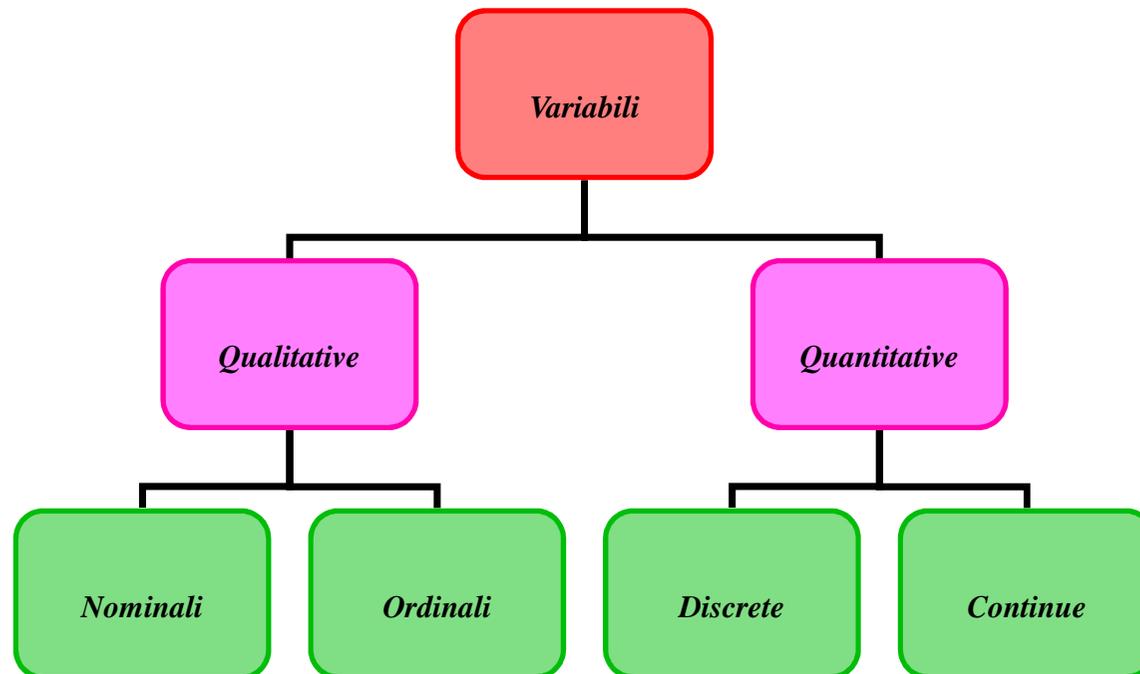
Cosa possiamo concludere dai dati relativi al campione di 20 studenti?

Soggetto	Altezza	Sesso	Soggetto	Altezza	Sesso
1	1.76	M	11	1.77	F
2	1.71	F	12	1.69	F
3	1.54	F	13	1.93	M
4	1.82	M	14	1.67	F
5	1.59	F	15	1.72	M
6	1.74	M	16	1.59	F
7	1.95	M	17	1.60	F
8	1.68	M	18	1.81	F
9	1.85	M	19	1.73	F
10	1.74	F	20	1.78	M

	Remdesivir group (n=158)	Placebo group (n=78)
Age, years	66.0 (57.0-73.0)	64.0 (53.0-70.0)
Sex		
Men	89 (56%)	51 (65%)
Women	69 (44%)	27 (35%)
Any comorbidities	112 (71%)	55 (71%)
Hypertension	72 (46%)	30 (38%)
Diabetes	40 (25%)	16 (21%)
Coronary heart disease	15 (9%)	2 (3%)
Body temperature, °C	36.8 (36.5-37.2)	36.8 (36.5-37.2)
Fever	56 (35%)	31 (40%)
Respiratory rate >24 breaths per min	36 (23%)	11 (14%)
White blood cell count, × 10 ⁹ per L		
Median	6.2 (4.4-8.3)	6.4 (4.5-8.3)
4-10	108/155 (70%)	58 (74%)
<4	27/155 (17%)	12 (15%)
>10	20/155 (13%)	8 (10%)
Lymphocyte count, × 10 ⁹ per L	0.8 (0.6-1.1)	0.7 (0.6-1.2)
≥1.0	49/155 (32%)	23 (29%)
<1.0	106/155 (68%)	55 (71%)
Platelet count, × 10 ⁹ per L	183.0 (144.0-235.0)	194.5 (141.0-266.0)
≥100	148/155 (95%)	75 (96%)
<100	7/155 (5%)	3 (4%)
Serum creatinine, μmol/L	68.0 (56.0-82.0)	71.3 (56.0-88.7)
≤133	151/154 (98%)	76 (97%)
>133	3/154 (2%)	2 (3%)
Aspartate aminotransferase, U/L	31.0 (22.0-44.0)	33.0 (24.0-48.0)
≤40	109/155 (70%)	49 (63%)
>40	46/155 (30%)	29 (37%)
Alanine aminotransferase, U/L	26.0 (18.0-42.0)	26.0 (20.0-43.0)
≤50	130/155 (84%)	66 (85%)
>50	25/155 (16%)	12 (15%)
Lactate dehydrogenase, U/L	339.0 (247.0-441.5)	329.0 (249.0-411.0)
≤245	36/148 (24%)	17/75 (23%)
>245	112/148 (76%)	58/75 (77%)
Creatine kinase, U/L	75.9 (47.0-131.1)	75.0 (47.0-158.0)
≤185	118/141 (84%)	54/67 (81%)
>185	23/141 (16%)	13/67 (19%)
National Early Warning Score 2 level at day 1	5.0 (3.0-7.0)	4.0 (3.0-6.0)
Six-category scale at day 1		
2—hospital admission, not requiring supplemental oxygen	0	3 (4%)
3—hospital admission, requiring supplemental oxygen	129 (82%)	65 (83%)
4—hospital admission, requiring high-flow nasal cannula or non-invasive mechanical ventilation	28 (18%)	9 (12%)
5—hospital admission, requiring extracorporeal membrane oxygenation or invasive mechanical ventilation	0	1 (1%)

(Table 1 continues on next page)

Per riassumere

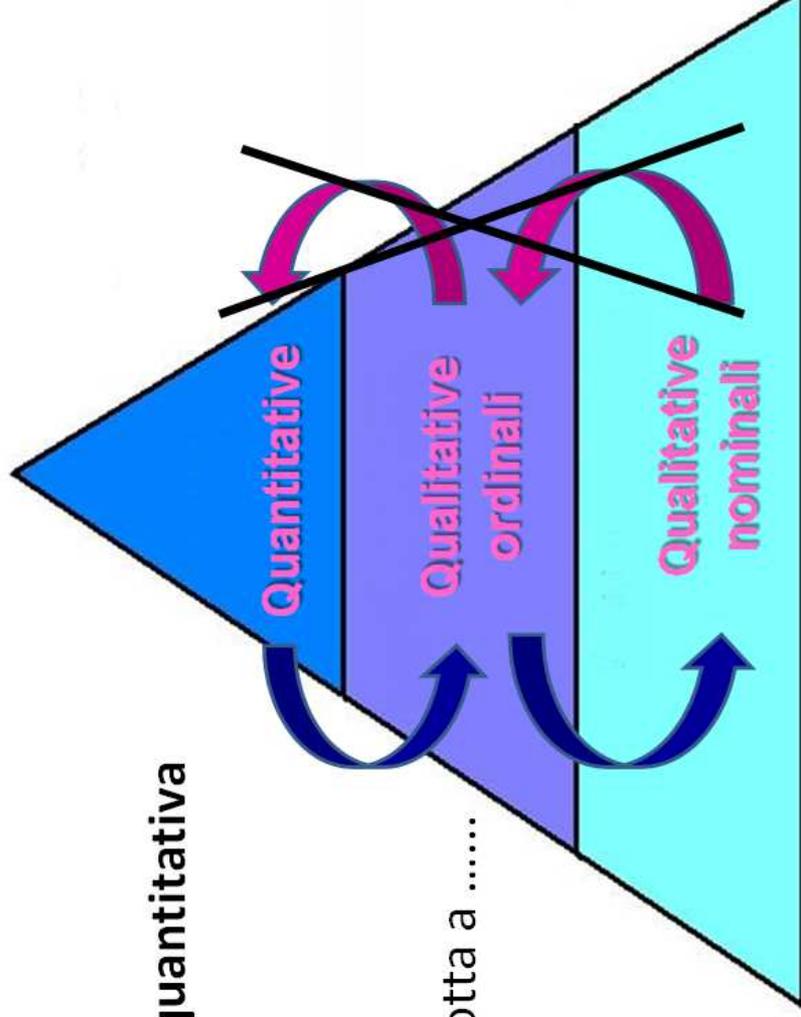


GERARCHIA DELLE VARIABILI

Una **variabile quantitativa**

può essere ridotta a

e ancora a



Esempi

Numero di carie presenti nell'arcata superiore	Quantitativa Discreta
Stato civile	Qualitativa Nominale
Consumo giornaliero di caffeina (mg)	Quantitativa Continua
Consumo giornaliero di caffè della macchinetta (bicchierini)	Quantitativa Discreta
Albumina sierica (g/l)	Quantitativa Continua
Tipologia Epatite	Qualitativa Nominale
Numero di linfonodi metastatici riscontrati alla TAC	Quantitativa Discreta