

Uso di `dest_prov`

Nicola Tommasi
email: nicola.tommasi@gmail.com

31 dicembre 2017

Indice

Sommario

`dest_prov` converte una variabile stringa con i nomi delle provincie italiane in una variabile numerica con i corrispettivi codici ISTAT. Inoltre può creare anche la variabile numerica della regione e le variabili stringa dei codici NUTS3, NUTS2 e NUTS1 (versione 2010).

1 Introduzione

In questo documento mostrerò il funzionamento del comando `dest_prov` con le sue varie opzioni. Per fare questo userò i dati più recenti forniti da ISTAT a questo [indirizzo](#). I dati sono presenti in fondo alla pagina, io ho usato questi. Il presente documento è stato redatto con il pacchetto `texdoc` di Ben Jann; potete trovarlo [qui](#). Variabili stringa con i nomi delle provincie italiane andrebbero sempre converti nei rispettivi codici numerici ISTAT. Non sempre nei database esiste anche la variabile numerica e allora in questi casi cosa fare? Serve un tedioso lavoro di conversione; ma perché farlo? Sostanzialmente per tre motivi:

- perché le variabili numeriche sono più efficienti delle variabili stringa.
- perché semplifica certe operazioni
(prendere tutte le provincie venete: `keep if inrange(cod_prov,23,24,25,26,27,28,29)`).
- perché consente di fare dei merge che altrimenti con variabili stringa sarebbero alquanto problematici.
- perché avere una variabile numerica per le provincie serve per il comando `dest_com`.

2 Installazione

Dalla command bar di Stata digitate:

```
. net from https://raw.githubusercontent.com/NicolaTommasi8/dest_prov/master/

https://raw.githubusercontent.com/NicolaTommasi8/dest_prov/master/
(no title)

PACKAGES you could -net describe-:
    dest_prov      Comando per convertire i nomi delle provincie italiane nel rispettivo codice
                   ISTAT.

. **net from https://sites.google.com/site/nicolatommasi/stata
```

quindi cliccate su `dest_prov` e alla pagina successiva su (`click here to install`). Questa operazione installerà sul vostro PC quattro files:

- `dest_prov.ado`
- `dest_prov.do`
- `dest_prov_lab.ado`
- `dest_prov.sthlp`

Il comando richiede Stata dalla versione 13 o successive, ma è caldamente consigliata almeno la versione 14 dato che viene usata la funzione `ustrfrom()` per consentire l'uso di dataset in codifica non UTF-8. In caso di utilizzo di Stata 13 alcune provincie potrebbero non essere riconosciute (Forlì-Cesena per esempio) e per queste si renderà necessaria una codifica manuale.

3 I Dati

Ora carico i dati relativi ai nomi e ai codici delle provincie italiane presenti nel file scaricato dai ISTAT. Ho scelto di usare nomi maiuscoli per queste variabili in modo da distinguerle da quelle che creerò in minuscolo con il comando `dest_com`.

```
. import excel using Elenco-codici-statistici-e-denominazioni-al-01_07_2016.xls, ///
> firstrow case(upper) clear
```

Ecco che variabili sono state caricate

```
. desc, fullnames
```

Contains data

```
obs:      8,016
vars:      25
size:     1,683,360
```

variable name	storage type	display format	value label	variable label
CODICEREGIONE	str2	%9s		Codice Regione
CODICECITTÀMETROPOLITANA	str3	%9s		Codice Città Metropolitana
CODICEPROVINCIA1	str3	%9s		Codice Provincia (1)
PROGRESSIVODELCOMUNE2	str3	%9s		Progressivo del Comune (2)
CODICECOMUNEFORMATOALFANUMERI	str6	%9s		Codice Comune formato alfanumerico
DENOMINAZIONEINITALIANO	str34	%34s		Denominazione in italiano
DENOMINAZIONEINTEDESCO	str36	%36s		Denominazione in tedesco
CODICERIPARTIZIONEGEOGRAFICA	byte	%10.0g		Codice Ripartizione Geografica
RIPARTIZIONEGEOGRAFICA	str10	%10s		Ripartizione geografica
DENOMINAZIONEREGIONE	str29	%29s		Denominazione regione
DENOMINAZIONECITTÀMETROPOLITAN	str7	%9s		Denominazione Città metropolitana
DENOMINAZIONEPROVINCIA	str29	%29s		Denominazione provincia
FLAGCOMUNECAPOLUOGODIPROVINC	byte	%10.0g		Flag Comune capoluogo di provincia
SIGLAAUTOMOBILISTICA	str2	%9s		Sigla automobilistica
CODICECOMUNEFORMATONUMERICO	long	%10.0g		Codice Comune formato numerico
CODICECOMUNENUMERICOCON107P	long	%10.0g		Codice Comune numerico con 107 province (dal 2006 al 2009)
CODICECOMUNENUMERICOCON103P	long	%10.0g		Codice Comune numerico con 103 province (dal 1995 al 2005)
CODICECATASTALEDELCOMUNE	str4	%9s		Codice Catastale del comune
POPOLAZIONELEGALE201109102	long	%10.0g		Popolazione legale 2011 (09/10/2011)
CODICENUTS12010	str3	%9s		Codice NUTS1 2010
CODICENUTS220103	str4	%9s		Codice NUTS2 2010 (3)
CODICENUTS32010	str5	%9s		Codice NUTS3 2010
CODICENUTS12006	str3	%9s		Codice NUTS1 2006
CODICENUTS220063	str4	%9s		Codice NUTS2 2006 (3)
CODICENUTS32006	str5	%9s		Codice NUTS3 2006

Sorted by:

Note: Dataset has changed since last saved.

Ora cancello quelle superflue. Prima però aggiungo alla denominazione delle provincie quella delle città metropolitane dato che il valore nella variabile della provincia è uguale a "-" per le città metropolitane.

```
. replace DENOMINAZIONEPROVINCIA = DENOMINAZIONECITTÀMETROPOLITAN if DENOMINAZIONEPROVINCIA=="-"  
(911 real changes made)
```

```
. drop CODICECITTÀMETROPOLITANA PROGRESSIVODELCOMUNE2 CODICECOMUNEFORMATOALFANUMERI ///
```

```

> CODICERIPARTIZIONE GEOGRAFICA RIPARTIZIONE GEOGRAFICA DENOMINAZIONE REGIONE ///
> DENOMINAZIONE CITTÀ METROPOLITANA FLAG COMUNI CAPOLUOGO DI PROVINCIA ///
> CODICE CATASTALE DEL COMUNE POPOLAZIONE LEGALE 2011 09 10 2 CODICE COMUNE FORMATO NUMERICO ///
> CODICE COMUNE NUMERICO CON 107 P CODICE COMUNE NUMERICO CON 103 P ///
> CODICENUTS12006 CODICENUTS220063 CODICENUTS32006

```

e rinomino quelle di interesse

```

. rename (CODICEREGIONE CODICEPROVINCIA1 DENOMINAZIONE IN ITALIANO DENOMINAZIONE IN TEDESCO ///
> DENOMINAZIONE PROVINCIA SIGLA AUTOMOBILISTICA CODICENUTS12010 ///
> CODICENUTS220103 CODICENUTS32010) ///
> (COD_REGIO COD_PROV DENOM_ITA DENOM_TED DENOM_PROV TARGA ///
> NUTS1 NUTS2 NUTS3)

```

Ora pulisco il dataset dalle osservazioni missing caricate dal file excel e converto in numerica la variabile COD_PROV che viene letta come stringa

```

. drop if COD_PROV=="
(18 observations deleted)
. destring COD_PROV COD_REGIO, replace
COD_PROV: all characters numeric; replaced as int
COD_REGIO: all characters numeric; replaced as byte

```

4 Uso di dest_prov

Adesso supponiamo di avere una variabile tipo DENOM_PROV con la denominazione delle provincie italiane e di volerla trasformare in una variabile numerica con i codici numerici ufficiali ISTAT.

```

. fre DENOM_PROV
DENOM_PROV — Denominazione provincia

```

		Freq.	Percent	Valid	Cum.
Valid	Agrigento	43	0.54	0.54	0.54
	Alessandria	190	2.38	2.38	2.91
	Ancona	47	0.59	0.59	3.50
	Arezzo	37	0.46	0.46	3.96
	Ascoli Piceno	33	0.41	0.41	4.38
	Asti	118	1.48	1.48	5.85
	Avellino	118	1.48	1.48	7.33
	Bari	41	0.51	0.51	7.84
	Barletta-Andria-Trani	10	0.13	0.13	7.96
	Belluno	64	0.80	0.80	8.76
	Benevento	78	0.98	0.98	9.74
	Bergamo	242	3.03	3.03	12.77
	Biella	79	0.99	0.99	13.75
	Bologna	55	0.69	0.69	14.44
	Bolzano/Bozen	116	1.45	1.45	15.89
	Brescia	205	2.56	2.56	18.45
	Brindisi	20	0.25	0.25	18.70
	Cagliari	71	0.89	0.89	19.59
	Caltanissetta	22	0.28	0.28	19.87
	Campobasso	84	1.05	1.05	20.92
	:	:	:	:	:
	Siracusa	21	0.26	0.26	78.92
	Sondrio	77	0.96	0.96	79.88
	Taranto	29	0.36	0.36	80.25
	Teramo	47	0.59	0.59	80.83
	Terni	33	0.41	0.41	81.25
	Torino	315	3.94	3.94	85.18
	Trapani	24	0.30	0.30	85.48
	Trento	177	2.21	2.21	87.70
	Treviso	95	1.19	1.19	88.88
	Trieste	6	0.08	0.08	88.96
	Udine	135	1.69	1.69	90.65
	Valle d'Aosta/Vallée d'Aoste	74	0.93	0.93	91.57
	Varese	139	1.74	1.74	93.31
	Venezia	44	0.55	0.55	93.86

Verbano-Cusio-Ossola	76	0.95	0.95	94.81
Vercelli	86	1.08	1.08	95.89
Verona	98	1.23	1.23	97.11
Vibo Valentia	50	0.63	0.63	97.74
Vicenza	121	1.51	1.51	99.25
Viterbo	60	0.75	0.75	100.00
Total	7998	100.00	100.00	

```
. dest_prov DENOM_PROV, gen(cod_prov)
. fre cod_prov
cod_prov — Codice ISTAT provincia
```

			Freq.	Percent	Valid	Cum.
Valid	1	Torino	315	3.94	3.94	3.94
	2	Vercelli	86	1.08	1.08	5.01
	3	Novara	88	1.10	1.10	6.11
	4	Cuneo	250	3.13	3.13	9.24
	5	Asti	118	1.48	1.48	10.72
	6	Alessandria	190	2.38	2.38	13.09
	7	Aosta	74	0.93	0.93	14.02
	8	Imperia	67	0.84	0.84	14.85
	9	Savona	69	0.86	0.86	15.72
	10	Genova	67	0.84	0.84	16.55
	11	La Spezia	32	0.40	0.40	16.95
	12	Varese	139	1.74	1.74	18.69
	13	Como	154	1.93	1.93	20.62
	14	Sondrio	77	0.96	0.96	21.58
	15	Milano	134	1.68	1.68	23.26
	16	Bergamo	242	3.03	3.03	26.28
	17	Brescia	205	2.56	2.56	28.84
	18	Pavia	188	2.35	2.35	31.20
	19	Cremona	115	1.44	1.44	32.63
	20	Mantova	69	0.86	0.86	33.50
	:	:	:	:	:	:
	91	Nuoro	52	0.65	0.65	89.01
	92	Cagliari	71	0.89	0.89	89.90
	93	Pordenone	50	0.63	0.63	90.52
	94	Isernia	52	0.65	0.65	91.17
	95	Oristano	88	1.10	1.10	92.27
	96	Biella	79	0.99	0.99	93.26
	97	Lecco	88	1.10	1.10	94.36
	98	Lodi	61	0.76	0.76	95.12
	99	Rimini	25	0.31	0.31	95.44
	100	Prato	7	0.09	0.09	95.52
	101	Crotone	27	0.34	0.34	95.86
	102	Vibo Valentia	50	0.63	0.63	96.49
	103	Verbano-Cusio-Ossola	76	0.95	0.95	97.44
	104	Olbia-Tempio	26	0.33	0.33	97.76
	105	Ogliastra	23	0.29	0.29	98.05
	106	Medio Campidano	28	0.35	0.35	98.40
	107	Carbonia-Iglesias	23	0.29	0.29	98.69
	108	Monza e della Brianza	55	0.69	0.69	99.37
	109	Fermo	40	0.50	0.50	99.87
	110	Barletta-Andria-Trani	10	0.13	0.13	100.00
	Total		7998	100.00	100.00	

a questo punto verifico che la variabile `cod_prov` corrisponda alla variabile `COD_PROV`

```
. assert cod_prov==COD_PROV
```

Adesso aggiungo le opzioni per creare il codice numerico delle provincie delle regioni (`gregio`) e i codici alfanumerici `nuts3` (`gnuts3`), `nuts2` (`gnuts2`) e `nuts1` (`gnuts1`).

```
. drop cod_prov
.
. dest_prov DENOM_PROV, gen(cod_prov) gregio(cod_regio) ///
> gnuts3(nuts3) gnuts2(nuts2) gnuts1(nuts1)
```

e come fatto in precedenza verifico la correttezza della conversione

```

. compare cod_regio COD_REGIO

```

	count	minimum	difference average	maximum
cod_regio=COD_REGIO	7998			
jointly defined	7998	0	0	0
total	7998			

```

. compare nuts3 NUTS3

```

	count
nuts3=NUTS3	7998
jointly defined	7998
total	7998

```

. compare nuts2 NUTS2

```

	count
nuts2=NUTS2	7998
jointly defined	7998
total	7998

```

. compare nuts1 NUTS1

```

	count
nuts1=NUTS1	7998
jointly defined	7998
total	7998

5 Help del comando

```

. type C:\ado\plus\d\dest_prov.sthlp
help dest_prov

```

Description

`dest_prov` converte i nomi delle provincie italiane nel codice numerico ISTAT.

Syntax

```

dest_prov varname [if] [in] [, generate(varname) tl(full|sigla) ignore gregio(name)
macro3(name) macro5(name) gnuts3(name) gnuts2(name) gnuts1(name) ]

```

dove *varname* è la variabile stringa con i nomi delle provincie. I nomi possono essere in forma estesa (es. Verona) o come sigla (es. VR).

options	Description
<code>generate(varname)</code>	crea la variabile numerica <i>varname</i> con il codice ISTAT delle provincie italiane. Se non specificato verrà creata di default la variabile <code>cod_prov</code> .
<code>tl(full sigla)</code>	<i>varname</i> specifica se le labels dei valori della variabile creata saranno in forma estesa (es. Verona) o come sigla (es. VR). Il valore di default è <code>full</code> .
<code>ignore</code>	<code>dest_prov</code> esegue due controlli. Il primo controlla che tutte le stringhe di <i>varname</i> siano riconosciute e quindi convertite in codice numerico, il secondo controllo verifica che tutti i codici numerici generati abbiano una label. Se uno dei due controlli non è verificato, l'esecuzione del comando viene interrotta e nessuna conversione viene eseguita. Specificando l'opzione <code>ignore</code> , <code>dest_prov</code> viene eseguito anche se uno dei controlli non viene superato.
<code>gregio(varname)</code>	genera una variabile con i codici ISTAT delle regioni.
<code>macro3(varname)</code>	genera una variabile con l'appartenenza a 3 macro regioni (Nord, Centro, Sud e Isole).
<code>macro5(varname)</code>	genera una variabile con l'appartenenza a 5 macro regioni (Nord-Ovest,

	Nord-Est, Centro, Sud, Isole).
<code>gnuts3(varname)</code>	genera una variabile (stringa) con i codici NUTS3 2010.
<code>gnuts2(varname)</code>	genera una variabile (stringa) con i codici NUTS2 2010.
<code>gnuts1(varname)</code>	genera una variabile (stringa) con i codici NUTS1 2010.

Examples

codifica la variabile stringa provincia nella variabile numerica prov_num:

```
. dest_prov provincia, gen(prov_num)
```

come la precedente, ma la label dei valori della variabile prov_num sono in forma di sigla:

```
. dest_prov provincia, gen(prov_num) tl(sigla)
```

Saved results

Nessun risultato salvato

References

ISTAT [Pagina di riferimento](#)

Author

Nicola Tommasi
Centro Interdipartimentale di Documentazione Economica (C.I.D.E.)
University of Verona, Italy
nicola.tommasi@univr.it

Also see

Help: `dest_com`