Stata destring with removing weird characters

STATA destring with only keeping numeric numbers

The newest solution to this problem is:
egen-method as described below followed by:

replace problemvar = subinstr(problemvar, ",", ".",.)
https://www.stata.com/statalist/archive/2009-05/msg00862.html

followed by:
destring problemvar, replace *without the 'dpcomma' option.

Dieter Van Der Westhuizen

Mon 2020-07-27 03:01 PM
Sent Items

To: Jody Rusch (jody.rusch@nhls.ac.za);
Follow-up from my previous email:

*This command works when one has either comma or period in the field:

destring kknew3, replace dpcomma

po4phos_new	nananew2	kknew2	kknew3
.			.
85	136	33	3.3
.			.
114	133	29	2.9
97	135	4	4
101	131	34	3.4
.			.
.			.
106	133	23	2.3
.			.
184	135	26	2.6
.			.
191	137	27	2.7
105	135	26	2.6
.			.
139			.
.			.
.			.
.			.
105	134	34	3.4
.			.
107	132	36	3.6
219	138	26	2.6
.			.
131	127	33	3.3
.			.

Dieter Van Der Westhuizen
Mon 2020-07-27 02:39 PM
Sent Items

How to strip special characters in STATA variables:
Type the following:
****

```
program define extrnum
version 7
syntax varlist(max=1) , gen(str)
local maxlen: type `varlist'
local maxlen=substr("`maxlen'",4,.)
tempvar work
qui gen str1 `work'=""
forvalues i=1/`maxlen' {
```

```
qui  replace  `work'=`work'+substr(`varlist',`i',1)  if
real(substr(`varlist',`i',1))<.
}
gen `gen'=real(`work')
end
```

****

How to use:

```
extrnum var1, gen(newvar1)
```

It works perfectly except it drops the commas too:

kk_new	clcl_new	hco3bicarb~w	hco3aniong~w	ureaurea_new	crtcreat_new	crtcomment~w	osmoosmola~w	caca_new	mgmg_new	po4phos_new
34	113	16	12	41	21	.	.	219	78	120
40	.	.	.	30	24	.	.	237	88	85
.	.	.	.	.	.	.	.	.	.	.
43	108	13	20	21	30	.	.	160	70	202
.	.	.	.	.	.	.	.	.	.	.
38	113	14	12	23	45	.	.	213	84	66
38	109	16	14	23	37	.	.	235	79	60
3	122	16	16	17	34	.	.	218	65	144
.	.	.	.	.	.	.	.	.	.	.
29	129	14	16	21	.	.	.	.	92	.
23	115	.	.	.	.	.	.	.	78	182
.	.	.	.	.	.	.	.	.	.	.
27	115	15	23	55	58	.	.	165	81	185
18	110	15	21	48	44	.	.	203	118	143
.	.	.	.	.	.	.	.	170	68	230
33	107	15	17	29	34	.	.	229	97	61
40	114	13	18	26	28	.	.	192	78	96
29	106	14	15	21	24	.	.	188	67	38
21	104	14	18	22	39	.	.	123	64	139
.	.	.	.	.	.	.	.	.	.	.
20	95	16	23	31	54	.	.	136	60	192
27	99	15	24	38	49	.	.	146	76	236
20	97	14	21	37	47	.	.	206	72	290
35	112	16	14	34	31	.	.	232	85	65
.	.	.	.	.	.	.	.	.	.	.
33	113	10	16	17	34	.	282	186	66	85
.	.	.	.	.	.	.	.	.	.	.
29	113	11	12	15	39	.	.	161	57	114
4	113	12	14	12	32	.	.	18	67	97
34	109	10	15	23	39	.	.	171	105	101

**Thus use:

```
findit egenmore
```

**and install the app “egenmore” by following the instructions.

**Then one can specify to keep the following characters:

```
egen kknew3 = sieve(kk), char(0123456789.,)
```

**Which lets you end up with this:

kknew2	kknew3
43	4,3
38	3,8
38	3,8
3	3
29	2,9
23	2,3
27	2,7
18	1,8
33	3,3
40	4.0
29	2,9
21	2,1
20	2.0
27	2.7

Destringing this doesn't work yet, as some characters have "." and others have ","

Dieter Van Der Westhuizen
Mon 2020-07-27 11:30 AM
Sent Items

https://www.statalist.org/forums/forum/general-stata-discussion/general/967675-removing-non-numeric-characters-from-strings

Dieter van der Westhuizen
Chemical Pathology RegistrarC17 NHLS Pathology Laboratory, Groote Schuur Hospital and Red Cross Children's Hospital LaboratoryTel: 021 404 4135 | Cell: +27 82 861 2093 | Fax-to-email: 0866 090 397dieter.vdwesthuizen@nhls.ac.za | www.nhls.ac.za