

Keys and Searching Strings

March 7, 2016

Statisticool.com

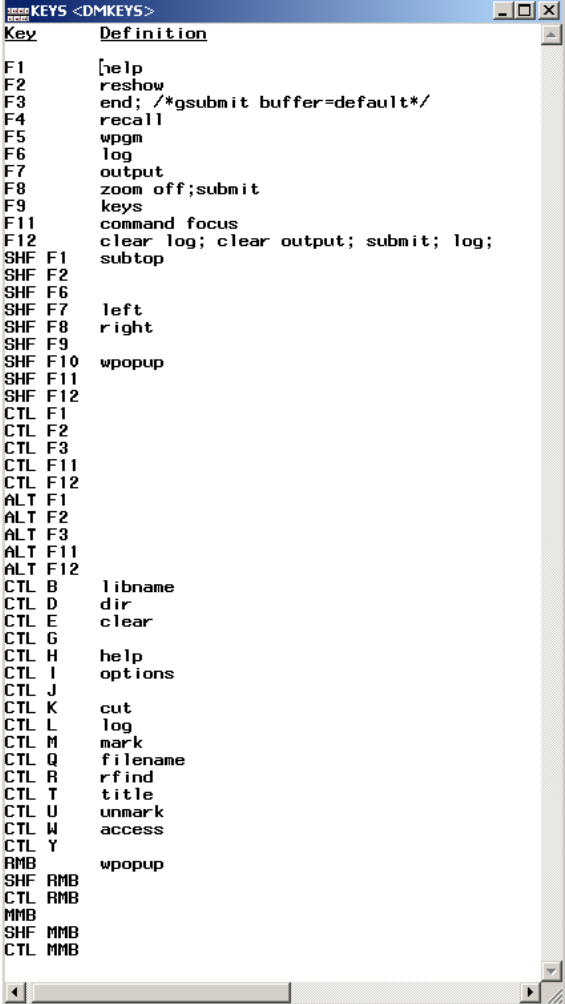
Keys

- Repetitive tasks: clearing log, submitting code, viewing the log
- To access shortcut keys
 - a. Tools → Options → Keys
 - b. Command line → "keys" → enter
 - c. F9 (sometimes this is set as default)
- Choose F key to assign (e.g. F12)

Keys (cont.)

- Enter in tasks
 - clear log; clear output; submit; log;

 - clears log, clears output, submits code, and brings log to the front



Key	Definition
F1	[help
F2	reshow
F3	end; /*gsubmit buffer=default*/
F4	recall
F5	wpgm
F6	log
F7	output
F8	zoom off;submit
F9	keys
F11	command focus
F12	clear log; clear output; submit; log;
SHF F1	subtop
SHF F2	
SHF F6	
SHF F7	left
SHF F8	right
SHF F9	
SHF F10	wpopup
SHF F11	
SHF F12	
CTL F1	
CTL F2	
CTL F3	
CTL F11	
CTL F12	
ALT F1	
ALT F2	
ALT F3	
ALT F11	
ALT F12	
CTL B	libname
CTL D	dir
CTL E	clear
CTL G	
CTL H	help
CTL I	options
CTL J	
CTL K	cut
CTL L	log
CTL M	mark
CTL Q	filename
CTL R	rfind
CTL T	title
CTL U	unmark
CTL W	access
CTL Y	
RMB	wpopup
SHF RMB	
CTL RMB	
MMB	
SHF MMB	
CTL MMB	

Searching Strings

- How to search a character variable for another character
- Real problem: searching a list of jobs for a specific job (“radiologist”)
 - List can have typos (“raddiologist”)
 - List can have variations (“I am a radiologist”)

Searching Strings (cont.)

```
data listtosearch;  
input person job $60.;  
datalines;  
1 radiologist  
2 radiologist  
3 radiologists  
4 xadiologist  
5 radioologist  
6 I am a radiologist  
7 radiology  
8 lawyer  
9 anthropologist  
10 radio operator  
11 dentist  
12 driver  
13 statistician  
14 mathematician  
15 teacher  
16 construction worker  
17 cleaner  
18 pilot  
19 coder  
20 banker  
;  
run;
```

- The first 7 jobs should be returned if we want to match on “radiologist”
- Number 10 may give us problems

Searching Strings (cont.)

- There are many ways to do this task
 - Search for exact match
 - Search for substrings
 - Search by sounds like (SOUNDEX)
 - Search by distance
 - Spelling distance (SPEDIS)
 - Levenshtein edit distance (COMPLEV)
 - Generalized edit distance (COMPGED)

Searching Strings (cont.)


- Tip: work with uppercase, job=upcase(job)
 - So “Radiologist” gets counted like “radiologist”

```
data method1;
```

```
set listtosearch;
```

```
where job='RADIOLOGIST';
```

```
run;
```

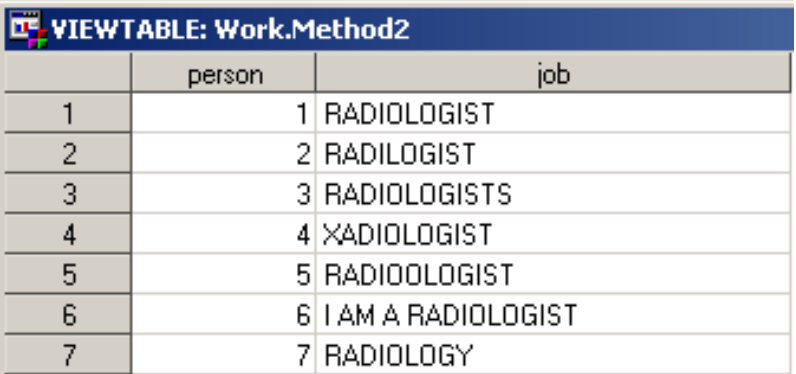


VIEWTABLE: Work.Method1

	person	job
1	1	RADIOLOGIST

Searching Strings (cont.)

```
data method2;  
set listtosearch;  
if index(job,'RADIOL') or  
index(job,'RADILO') or  
index(job,'RADIOO') or  
index(job,'XADIOL');  
/*etc*/  
run;
```



VIEWTABLE: Work.Method2		
	person	job
1	1	RADIOLOGIST
2	2	RADILOGIST
3	3	RADIOLOGISTS
4	4	XADIOLOGIST
5	5	RADIOOLOGIST
6	6	I AM A RADIOLOGIST
7	7	RADIOLOGY

Searching Strings (cont.)

- Algorithm for indexing names by sound, as pronounced in English
 - Create dataset of job I want to search for
 - $a = \text{soundex}(\text{job})$
 - Match merge by a in *listtosearch* dataset

VIEWTABLE: Work.Jobstosearchfor		
	job	a
1	RADIOLOGIST	R34223

VIEWTABLE: Work.Method3		
	job	person
1	RADIOLOGIST	1
2	RADILOGIST	2
3	RADIOOLOGIST	5

Searching Strings (cont.)

- Spelling distance
 - points = `spedis('RADIOLOGIST',job)`
 - `points=sum(costs)/length(query)`
 - choose points cutoff (I used 50)

VIEWTABLE: Work.Method4			
	person	job	points
1	7	RADIOLOGY	15
2	1	RADIOLOGIST	0
3	2	RADILOGIST	9
4	5	RADIOOLOGIST	2
5	3	RADIOLOGISTS	4
6	4	XADIOLOGIST	18

Searching Strings (cont.)

- Levenshtein edit distance
 - Number of insertions, deletions, or replacements of single characters required to convert one string to the other
 - points = `complev('RADIOLOGIST',job)`

VIEWTABLE: Work.Method5			
	person	job	points
1	7	RADIOLOGY	3
2	1	RADIOLOGIST	0
3	2	RADILOGIST	1
4	5	RADIOOLOGIST	1
5	3	RADIOLOGISTS	1
6	4	XADIOLOGIST	1