



Searching for Variable Values with CAT Functions

■ NEW FUNCTIONS IN V9.1

CAT concatenates *character* strings without removing leading or trailing blanks

CATS concatenates *character* strings and removes leading and trailing blanks

CATT concatenates *character* strings and removes trailing blanks only

CATX concatenates *character* strings, removes leading and trailing blanks, and inserts separators



Searching for Variable Values with CAT Functions

- NEW FUNCTION IN V9.2

CATQ concatenates *character or numeric* values by using a delimiter to separate items and by adding quotation marks to strings that contain the delimiter

the **CATX** function is similar to the **CATQ** function except that **CATX** does not add quotation marks

included "modifiers" that change the behavior of the function (no modifiers for other CAT functions)



Searching for Variable Values with CAT Functions

- results of the CAT, CATS, CATT, and CATX functions are usually equivalent to results that are produced by certain combinations of the concatenation operator (||) and the TRIM and LEFT functions



Searching for Variable Values with CAT Functions

Function	Equivalent Code (with conCATenation operator)
CAT(OF X1-X4)	X1 X2 X3 X4
CATS(OF X1-X4)	TRIM(LEFT(X1)) TRIM(LEFT(X2)) TRIM(LEFT(X3)) TRIM(LEFT(X4))
CATS(OF X1-X4)	STRIP(X1) STRIP(X2) STRIP(X3) STRIP(X4)
CATT(OF X1-X4)	TRIM(X1) TRIM(X2) TRIM(X3) TRIM(X4)
CATX(SP, X1-X4)	TRIM(LEFT(X1)) SP TRIM(LEFT(X2)) SP TRIM(LEFT(X3)) SP TRIM(LEFT(X4))
CATX(SP, X1-X4)	STRIP(X1) SP STRIP(X2) SP STRIP(X3) SP STRIP(X4)



Searching for Variable Values with CAT Functions

- default length for the CAT, CATS, CATT, and CATX functions is different from the length that is obtained when you use the concatenation operator

data names;

input last_name : \$25. first_name : \$15. @@;

* old way;

full_name_old_way = trim(last_name) || ', ' || first_name;

* new way;

full_name_new_way = catx(', ',last_name,first_name);

datalines;

Zdeb Mike Washington George Squirrel Rocket(J)

;



Searching for Variable Values with CAT Functions

last_name	first_name	full_name_old_way	full_name_new_way
Zdeb	Mike	Zdeb, Mike	Zdeb, Mike
Washington	George	Washington, George	Washington, George
Squirrel	Rocket(J)	Squirrel, Rocket(J)	Squirrel, Rocket(J)

Alphabetic List of Variables and Attributes			
#	Variable	Type	Len
2	first_name	Char	15
4	full_name_new_way	Char	200
3	full_name_old_way	Char	42
1	last_name	Char	25



Searching for Variable Values with CAT Functions

- NOT JUST FOR CONCATENATION
- ANOTHER USE ... SEARCH FOR VARIABLE VALUES

for example, given a data set named ANSWERS

create a new data set containing observations where at least one answer had a value of "Y"

DATA SET ANSWERS

id	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10
A1234	Y	Y	Y	Y	Y	Y	Y	Y	N	N
A2345	N	N	N	N	N	N	N	N	N	N
A3456	N	N	N	N	N	N	N	N	N	Y
A4567	N	N	N	N	Y	N	N	N	N	N
A5678	Y	N	N	Y	N	N	Y	N	N	Y



Searching for Variable Values with CAT Functions

- CONVENTIONAL
ARRAY AND A
LOOP

```
data yes;  
set answers;  
array a(10);  
do j=1 to 10 until (a(j) eq 'Y');  
end;  
if j lt 11;  
drop j;  
run;
```

- CAT AND FIND
FUNCTIONS

```
data yes;  
set answers;  
if find(cat(of a1-a10), 'Y');  
run;
```



Searching for Variable Values with CAT Functions

■ NEW TASK

create a new data set containing observations where at least TWO answers had a value of "Y"



Searching for Variable Values with CAT Functions

- CONVENTIONAL
ARRAY AND A LOOP

```
data yes;  
set answers;  
array a(10);  
do j=1 to 10 until (found_y eq 2);  
    found_y = sum(found_y, (a(j) eq 'Y'));  
end;  
if found_y eq 2;  
drop j found_y;  
run;
```

- CAT AND COUNTC
FUNCTIONS

```
data yes;  
set answers;  
if countc(cat(of a1-a10), 'Y') ge 2  
run;
```



Searching for Variable Values with CAT Functions

■ NEW TASK

add a new variable to each observation in data set ANSWERS

the variable will contain the number of answers that are "Y"



Searching for Variable Values with CAT Functions

```
* array and a loop;  
data yes;  
set answers;  
array a(10);  
* use _n_ as loop iterator (no DROP);  
do _n_=1 to 10;  
    found_y = sum(found_y, (a(_n_) eq 'Y'));  
end;  
run;  
  
* cat and countc;  
data yes;  
set answers;  
found_y = countc(cat(of a:), 'Y');  
run;
```



Searching for Variable Values with CAT Functions

- NEW DATA SET WITH NUMERIC ANSWERS
- SIMILAR TASK

create a new data set
containing observations where
at least one answer had a
value of 5

DATA SET ANSWERS

id	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10
A1234	5	5	3	3	2	1	1	1	1	1
A2345	0	0	0	0	0	0	0	5	5	5
A3456	1	1	2	2	4	0	0	0	0	0
A4567	5	5	5	5	5	5	5	5	5	5
A5678	1	0	1	0	1	0	5	5	5	3



Searching for Variable Values with CAT Functions

- USE CAT AND FIND
- NO PROBLEM WITH NUMERIC DATA (SEARCH FOR '5')

```
* cat and find;  
data five;  
set answers;  
if find(cat(of a:), '5');  
run;
```

SAS log ...

```
243 * cat and find;  
244 data five;  
245 set answers;  
246 if find(cat(of a:), '5');  
247 run;
```

NOTE: There were 5 observations read from the data set

WORK.ANSWERS.

NOTE: The data set WORK.FIVE has 4 observations and 11 variables.



Searching for Variable Values with CAT Functions

- YES ... YOU COULD HAVE USED ... **if max (of a:) eq 5;**

- A TEST ...

```
data five;  
set answers;  
answers = cat(of a:);  
if find(answers,'5');  
run;
```

- WHAT TYPE OF VARIABLE IS ANSWERS ... CHAR OR NUM (WHY?)



Searching for Variable Values with CAT Functions

■ HINT (THE LOG) ...

```
1497
1498 data five;
1499 set answers;
1500 answers = cat(of a:);
1501 if find(answers,'5');
1502 run;
```

NOTE: Character values have been converted to numeric values at the places given by: (Line):(Column).
1500:11

NOTE: Numeric values have been converted to character values at the places given by: (Line):(Column).
1501:9

NOTE: There were 5 observations read from the data set WORK.ANSWERS.

NOTE: The data set WORK.FIVE has 4 observations and 12 variables.

NOTE: DATA statement used (Total process time):
real time 0.00 seconds
cpu time 0.00 seconds



Searching for Variable Values with CAT Functions

Alphabetic List of Variables and Attributes

#	Variable	Type	Len
2	a1	Num	8
3	a2	Num	8
4	a3	Num	8
5	a4	Num	8
6	a5	Num	8
7	a6	Num	8
8	a7	Num	8
9	a8	Num	8
10	a9	Num	8
11	a10	Num	8
12	answers	Num	8
1	id	Char	1

id	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	answers
4	5	5	3	3	2	1	1	1	1	1	5533211111
5	0	0	0	0	0	0	0	5	5	5	555
7	5	5	5	5	5	5	5	5	5	5	5555555555
8	1	0	1	0	1	0	5	5	5	3	1010105553



Searching for Variable Values with CAT Functions

- DIGRESSION ... QUESTION FROM SAS-L

Is there anyway to check first three digits of numeric value?

- HOW ABOUT THE CAT FUNCTION

- THIS WORKS

```
data x;  
input abc @@;  
if cat(abc) eq : '111';  
datalines;  
111222 123456 11111111 987654 111  
;
```



Searching for Variable Values with CAT Functions

■ THE LOG ... NO ERRORS, NO NOTES

```
2    data x;  
3    input abc @@;  
4    if cat(abc) eq : '111';  
5    datalines;
```

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.

NOTE: The data set WORK.X has 3 observations and 1 variables.

NOTE: DATA statement used (Total process time):

real time	0.59 seconds
-----------	--------------

cpu time	0.04 seconds
----------	--------------



Searching for Variable Values with CAT Functions

- ANOTHER TEST ... THIS DOES NOT WORK (WHY?)

```
data x;  
input abc @@;  
if cat(abc) eq : '111';  
datalines;  
111222 123456 11111111 987654 111 11 1  
;
```

- HINT ... WHAT DOES eq : REALLY MEAN
- A FIX ... if cat(abc,'X') eq : '111';

abc
111222
11111111
111
11
1



Searching for Variable Values with CAT Functions

■ MORE COMPLEXITY ... SEARCH FOR DIAGNOSIS CODES

each observation has up to five,
three character diagnosis codes
and the code 250 indicates
diabetes ... find patient records
with diabetes

025	022			
682	401	244	493	
592	401	493		
428	493	780	V43	250
250				
414	V45	401	250	



Searching for Variable Values with CAT Functions

- INCORRECT ... WHY ?

```
data dia_incorrect;  
set diag3;  
if find(cat(of dx1-dx5), '250') ne 0;  
run;
```

- WHAT IS THE RESULT OF THE CAT FUNCTION ?
- LOOK AT THE FIRST RECORD ... THERE IS A 250 BUT NEITHER DIAGNOSIS IS 250 (A COMBINATION OF 025 AND 022)

```
025022  
682401244493  
592401493  
428493780V43250  
250  
414V45401250
```



Searching for Variable Values with CAT Functions

- CORRECT ... USE **CATX**

```
data dia_correct;  
set diag3;  
if find(catx('*', of dx1-dx5), '250') ne 0;  
run;
```

- WHAT IS THE RESULT OF THE CATX FUNCTION
- LOOK AT THE FIRST RECORD ...
THERE IS NO 250

```
025*022  
682*401*244*493  
592*401*493  
428*493*780*V43*250  
250  
414*V45*401*250
```



Searching for Variable Values with CAT Functions

- CATX MAKES IT EASY TO SEARCH FOR VARIOUS DIAGNOSES

250 indicates diabetes, 493 indicates asthma ... find patient records with diabetes and/or asthma

```
data dia_ast;  
set diag3;  
dia = (find(catx('*', of dx1-dx5), '250') ne 0);  
ast = (find(catx('*', of dx1-dx5), '493') ne 0);  
run;
```

- CODE ADDS DUMMY VARIABLES: 1, PRESENT; 0, NOT PRESENT



Searching for Variable Values with CAT Functions

■ DIGRESSION ... QUESTION FROM SAS COMMUNITY

find all observations where all non-missing values are 4

create a new variable with a value of 3 for above, otherwise 9

q1	q2	q3	q4	q5
4	4	4	4	4
.	2	3	1	4
4	.	.	.	4
.	.	.	.	.
1	2	3	4	5
.	4	4	4	.



Searching for Variable Values with CAT Functions

data test;

set test;

questcat = ifn(verify(catt(of q:),'.4') or sum(of q:) eq . , 9 , 3);

run;

q1	q2	q3	q4	q5	questcat
4	4	4	4	4	3
.	2	3	1	4	9
4	.	.	.	4	3
.	.	.	.	.	9
1	2	3	4	5	9
.	4	4	4	.	3



Searching for Variable Values with CAT Functions

- MORE COMPLEXITY ...
FIVE CHARACTER
DIAGNOSIS CODES,
ONLY THE FIRST THREE
CHARACTERS USED TO
FIND DISORDERS

25001	V180			
5680	78039	6250	49390	53081
4270	4111	4019	36250	
486	7990	25082	41400	
78659	25000	49320		
49392	4660	2449		
34839	2765	40493	4280	

- ONCE AGAIN ... FIND DIABETES (250) AND/OR ASTHMA (493)



Searching for Variable Values with CAT Functions

- COULD USE AN ARRAY PLUS A LOOP

```
data dia_ast;  
set diag5;  
array dx(5);  
dia = 0;  
ast = 0;  
do j=1 to 5;  
    if dx(j) eq : '250' then dia = 1;  
    if dx(j) eq : '493' then ast = 1;  
end;  
drop j;  
run;
```

- EQ :
LOOKS FOR VALUES THAT START WITH THE SPECIFIED STRING



Searching for Variable Values with CAT Functions

- OR, USE CATX FUNCTION ... INCORRECT ... WHY ?

```
data dia_ast_incorrect;  
set diag5;  
dia = (find(catx('*', of dx1-dx5), '*250') gt 0);  
ast = (find(catx('*', of dx1-dx5), '*493') gt 0);  
run;
```

- WHAT IS THE RESULT OF THE CATX FUNCTION ?
- LOOK AT THE FIRST RECORD ... STARTS WITH 250

```
25001*V180  
5680*78039*6250*49390*53081  
4270*4111*4019*36250  
486*7990*25082*41400  
78659*25000*49320  
49392*4660*2449  
34839*2765*40493*4280
```



Searching for Variable Values with CAT Functions

■ MODIFIED CATX FUNCTION

```
data dia_ast_correct;  
set diag5;  
dia = (find(catx('*', '*', of dx1-dx5), '*250') gt 0);  
ast = (find(catx('*', '*', of dx1-dx5), '*493') gt 0);  
run;
```

■ WHAT IS THE RESULT OF THE CATX FUNCTION

■ LOOK AT THE FIRST RECORD ... STARTS WITH **250

```
**25001*V180  
**5680*78039*6250*49390*53081  
**4270*4111*4019*36250  
**486*7990*25082*41400  
**78659*25000*49320  
**49392*4660*2449  
**34839*2765*40493*4280
```



Searching for Variable Values with CAT Functions

■ "REAL LIFE" EXAMPLE

data set with 2.5 million observations ... each observation has an admit, principal, and up to fourteen other diagnoses

identify observations with either: diabetes (250); asthma (493); heart attack (410); female breast cancer (174); influenza (487)



Searching for Variable Values with CAT Functions

■ DONE THREE DIFFERENT WAYS

```
* use an array and a loop;  
data array_loop;  
set temp;  
array dx(16) adx pdx odx1-odx14;  
dia = 0;  
ast = 0;  
ami = 0;  
brc = 0;  
flu = 0;  
do j=1 to 16 until (dx(j) eq ' ');  
    if dx(j) eq : '250' then dia = 1;  
    if dx(j) eq : '493' then ast = 1;  
    if dx(j) eq : '410' then ami = 1;  
    if dx(j) eq : '174' then brc = 1;  
    if dx(j) eq : '487' then flu = 1;  
end;  
drop j;
```



Searching for Variable Values with CAT Functions

*** CAT and FIND with a new variable
(and a LENGTH statement);**

```
data cat_find_new;  
set temp;  
length string $100;  
string = catx('*', '*', adx, pdx, of odx:);  
dia = (find(string, '*250') gt 0);  
ast = (find(string, '*493') gt 0);  
ami = (find(string, '*410') gt 0);  
brc = (find(string, '*174') gt 0);  
flu = (find(string, '*487') gt 0);  
drop string;  
run;
```



Searching for Variable Values with CAT Functions

*** CAT and FIND with repeated concatenation;**

```
data cat_find_repeat;  
set temp;  
dia = (find(catx('*', '*', adx, pdx, of odx:), '*250') gt 0);  
ast = (find(catx('*', '*', adx, pdx, of odx:), '*493') gt 0);  
ami = (find(catx('*', '*', adx, pdx, of odx:), '*410') gt 0);  
brc = (find(catx('*', '*', adx, pdx, of odx:), '*174') gt 0);  
flu = (find(catx('*', '*', adx, pdx, of odx:), '*487') gt 0);  
run;
```



Searching for Variable Values with CAT Functions

- EACH METHOD TRIED 25 TIMES
- DATA SET (NAMED TEMP) IN MEMORY USING SASFILE
`sasfile temp load;`
- TIMES ARE IN SECONDS

METHOD	ELAPSED TIME	CPU TIME
ARRAY AND LOOP	21.1	17.4
CAT AND FIND WITH A NEW VARIABLE	19.9	16.4
CAT AND FIND WITH A NEW VARIABLE (AND A LENGTH STATEMENT)	18.4	13.3
CAT AND FIND WITH REPEATED CONCATENATION	18.6	14.3



Searching for Variable Values with CAT Functions

■ WHAT IF SELECTION RULES ARE COMPLEX

Traumatic brain injury includes ICD-9 codes 800-801.9, 803-804.9, 850-854.1, 950.1-950.3, 959.01 and 995.55.

```
proc format;  
value $tbi  
'800'-'80199', '803'-'80499' , '850'-'85419',  
'9501'-'95039', '95901' , '99555' = '1'  
other = '0';  
run;
```



Searching for Variable Values with CAT Functions

■ CONCATENATE FORMATTED VALUES OF VARIABLES

```
filename nosee dummy;
```

```
data tbi;  
set temp;  
file nosee;  
put (adx pdx odx1-odx14) ($tbi.) @;  
if find(_file_, '1') then output;  
put;  
run;
```



Searching for Variable Values with CAT Functions

- SELECTION RULES NOT COMPLEX ... ANOTHER APPROACH FOR LOOKING AT THE FIRST THREE CHARACTERS IN FIVE CHARACTER DX CODES ... CONCATENATE FIRST THREE CHARACTERS

```
filename nosee dummy;
```

```
data tbi;  
set temp;  
file nosee;  
put (adx pdx odx1-odx14) ($3. +1) @;  
if find(_file_, '250') then output;  
put;  
run;
```



Searching for Variable Values with CAT Functions

■ LOOK FOR MULTIPLE DIAGNOSES ... SAVE PATTERN

```
proc format;  
value $dia_ast  
'250' - '25099' = 'D' '493' - '49399' = 'A' other = 'x';  
run;
```

```
data dia_ast;  
length pattern $16;  
set temp;  
file nosee;  
put (adx pdx odx1-odx14) ($dia_ast.) @;  
pattern = _file_;  
if findc(_file_,'DA') then output;  
put;  
run;
```



Searching for Variable Values with CAT Functions

pattern	adx	pdx	odx1	odx2	odx3	odx4	odx5	odx6	odx7	odx8	odx9	odx10	odx11	odx12	odx13	odx14
xxxxDxxxxxxxxxxx	V5789	V5789	5990	78079	25000	41400	2967	78057								
xxxxxxDxxxxxxxx	V5789	V5789	49121	41519	78079	7850	41400	25000								
xxxxxDxxxxxxxxxxx	V5789	V5789	82123	V4589	7140	25000	2727	2721								
xxxxxDxxxxxxxxxxx	V5789	V5789	496	4280	78079	7850	25000									
xxxxDxxxxxxxxxxx	V5789	V5789	40391	V451	25001	78079	2449									
xxxxxxxDxxxxxxx	V5789	V5789	486	5119	78079	41012	4019	2720	25000	4439	7804					
xxxxDxxxxxxxxxxx	V5789	V5789	4280	78079	25080	41400	5939	4019								
xxxDxxxxxxxxxxx	49121	49121	7990	25000												
xxxxxxxDxxxxxxx	V5789	V5789	78079	70710	4591	V1251	4019	43310	25000							
xxxxxDxxxxxxxxxxx	V5789	V5789	78079	V126	7806	25000	2727									
xxDxxxxxxxxxxx	V5789	V5789	25001	9986	78079	V1089	V1251	4019	2809	2449						
AAxxxxxxxxxxx	49320	49320	2761	2765	4580											
xxxxxDxxxxxxxxxxx	V5789	V5789	71536	V4365	41400	4019	25000	42789								
xxxDxxxxxxxxxxx	V5789	4409	8082	25001	V5789	V1582	73300	2720								
xxxxxDxxxxxxxxxxx	V5789	V5789	78079	V1089	V4589	25000	4019	2720	2449	53081	2809					
xxxxDxxxxxxxxxxx	V5789	V5789	78079	92300	25002	4019	41400	43310								
xxxxDxxxxxxxxxxx	V5789	V5789	71535	V4364	25001	436	42731	V1259	2720	53081						
xxxxxxxDxxxxxxx	V5789	V5789	78079	V1005	V4589	496	41400	25000	4111							
xxxxxxxDxxxxxxx	V5789	V5789	78079	V4972	V4975	V4971	40391	V451	25071	44381	74789					
xxxxxDxxxxxxxxxxx	V5789	V5789	78079	2900	41400	25060	25050	3371	3572	36201	5363					
xxxxxDxxxxxxxxxxx	V5789	V5789	78079	4139	4280	2851	25000									
xxxxDxxxxxxxxxxx	V5789	V5789	71535	V4364	25000	2859										
xxxxDAxxxxxxxxxxx	V5789	V5789	78079	4019	25000	49390	78057	4169	45981	4540						



Searching for Variable Values with CAT Functions

■ COMPARE

```
data cat_find_repeat;
set temp;
dia = (find(catx('*', '*', adx, pdx, of odx:), '*250') gt 0);
ast = (find(catx('*', '*', adx, pdx, of odx:), '*493') gt 0);
ami = (find(catx('*', '*', adx, pdx, of odx:), '*410') gt 0);
brc = (find(catx('*', '*', adx, pdx, of odx:), '*174') gt 0);
flu = (find(catx('*', '*', adx, pdx, of odx:), '*487') gt 0);
run;

data use_file_;
set temp;
file nose;
put (adx pdx odx1-odx14) ($3. +1) @;
dia = (find(_file_, '250') gt 0);
ast = (find(_file_, '493') gt 0);
ami = (find(_file_, '410') gt 0);
brc = (find(_file_, '174') gt 0);
flu = (find(_file_, '487') gt 0);
put;
run;
```



Searching for Variable Values with CAT Functions

■ LAST DIGRESSION ... FROM SAS COMMUNITY

if I concatenate several data sets, is there any way to find out what data set an observation came from



Searching for Variable Values with CAT Functions

■ VERSION 9.2+ ... CAN USE INDSNAME

```
data all;  
set one two three indsnames=temp;  
dsn = temp;  
run;
```

■ VERSION 9.13

```
data all;  
set one (in=_1) two (in=_2) three(in=_3);  
from = find(catt(of _:),'1');  
run;
```



Searching for Variable Values with CAT Functions

CONTACT INFORMATION ...

Mike Zdeb

msz03@albany.edu

518 402 6479