Bootstrap Sampling in SAS, an example of Getting Means from Many Samples[1]

It is possible to use a macro in SAS to run procedures within an iterative loop. Following is an example of a program that repeatedly gets a data set, draws a 20% random sample, calculates the mean and standard deviation of the sample, and adds these into a file that has the means and standard deviations from the previous samples drawn. This approximates a bootstrap process for sampling. (Lines that begin with an asterisk document the file, and are not executed by SAS.)

```
* bootsas ; title1 'bootsas, example of bootstrap sampling with macros in sas' ;
options compress=yes ;
data one; input var1 @@ ;
cards ;
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
;
* Get descriptive statistics for the entire sample for comparison ;
proc means ; var var1;
* create a macro;
%macro bootst ;
** Change the number after "%to" to change N of iterations ;
%do i =1 %to 2000 ;
data sample ; set one ;
* Create a 20% random sample. ;
* Change ".2" to another number to change sample size ;
* The seed for the random number generator will vary with each iteration. ;
ransamp = ranuni(&i) ; if ransamp le .2 ;
* Run proc means on the sample, do not print ;
proc means noprint ;  var var1 ;
output out = meand&i mean = mean std = std n =n ;
** Append the data set created to proc means. Each proc means ;
** will create an observation with the mean, standard deviation ;
** and N of cases in the sample. ;
proc append base=meanall data=meand&i ;
** End the DO Loop ;
%end ;
** End the macro ;
%mend bootst ;
* execute the macro ;
%bootst ;
* Get descriptive statistics for the data set created by macro ;
proc means data = meanall ;  run ;
```

The mean for the population is 300.5. The mean of the mean of the 2,000 samples is 300.52, the range is from 255.0 to 338.3.

---

[1]Prepared by Patty Glynn,, University of Washington. May 18, 2001, updated 6/16/02 C:\all\help\helpnew\bootsas.wpd