Estimating the discrete probability distribution of the age characteristic of Veteran populations using SAS®, SAS/OR® and SAS Simulation Studio® for use in population projection models
SA-05

by
Michael C. Grierson

MWSUG 23
Minneapolis
Sept 16-18, 2012

MW
Minneap
Sept 16-18, 2012

Three Concepts

- American Community Survey / Social Security Administration Period Life Table
  - used to produce discrete probability distributions

- Discrete Event Simulation
  - SAS/Simulation Studio®
  - Predictive Modeling

- Forecast future service delivery obligations
  - Budgeting and Business futures

American Community Survey

- 'Stratified', so more Effective and Efficient
- Gathered periodically
  - Yearly is smallest period (1, 3, and 5 year periods supported)
    - 1 year results more current than 3 and 5 year
    - 3 and 5 year results have larger samples
  - Meant to provide more timely data than the Decennial Census (which is every 10 years).
- Supports Community Planning
  - Sampling to support Local Governments
    - Consequently, National issues get VERY well represented

ACS 2010 and National Data (age)

```
let acsds=pus.pus2010;

proc surveyfreq data=&acsds;
   table agep;
   ods output OneWay=pus.freqout_pus;
   weight pwgtp;
run;
```

The SURVEYFREQ Procedure

Data Summary

Number of Observations   3097724
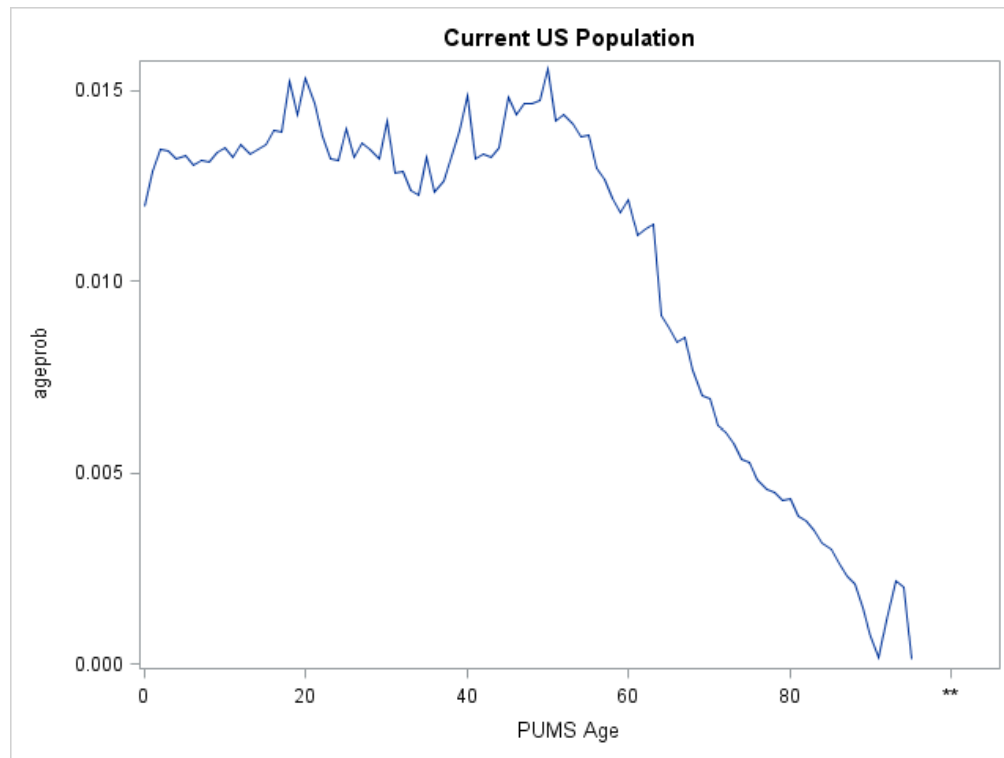Sum of Weights   313071822

The SURVEYFREQ Procedure

PUMS Age

Total Population of US (and Puerto Rico) 2010 by ACS is the 313,071,822 number.

AGEP	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
00	33448	3745300	24668	1.1963	0.0079
01	35589	4028867	25741	1.2869	0.0082
02	37431	4208986	26247	1.3444	0.0084
03	37738	4207428	26118	1.3439	0.0083
04	37555	4128056	25698	1.3186	0.0082
05	37944	4154459	25745	1.3270	0.0082
92	4310	356368	6367	0.1138	
93	8235	700237	9041	0.2237	
94	7024	625495	8834	0.1998	
95	533	43502	2136	0.0139	
Total	3097724	313071822	122364	100.000	

ACS 2010 and National Data (age)

I should call these Discrete Probability Distributions (but
sometimes I have called them Probability Density Distributions)

Veteran relevant questions in ACS

- Military Service?

Veteran relevant questions in ACS

- Disability Status and Health Coverage?

28 a. **Does this person have a VA service-connected disability rating?**

☐ Yes (such as 0%, 10%, 20%, ... , 100%)

☐ No ➔ *SKIP to question 29a*

b. **What is this person's service-connected disability rating?**

☐ 0 percent

☐ 10 or 20 percent

☐ 30 or 40 percent

☐ 50 or 60 percent

☐ 70 percent or higher

16 **Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans?** *Mark "Yes" or "No" for EACH type of coverage in items a – h.*

	Yes	No
a. Insurance through a current or former employer or union (of this person or another family member)	☐	☐
b. Insurance purchased directly from an insurance company (by this person or another family member)	☐	☐
c. Medicare, for people 65 and older, or people with certain disabilities	☐	☐
d. Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability	☐	☐
e. TRICARE or other military health care	☐	☐
f. VA (including those who have ever used or enrolled for VA health care)	☐	☐
g. Indian Health Service	☐	☐
h. Any other type of health insurance or health coverage plan – *Specify* ↙		

MWSUG 23
Minneapolis
Sept 16-18, 2012

Minneapolis
Sept 16-18, 2012

ACS 2010 and Veterans (age)

```
%let acsds=pus.pus2010;

data pus.veterans;
*  format lyms $lyms. drat $drat. dratx $dratx.;
   format lyms $lyms. ;
   set &acsds;
   if mil < 2 or mil > 3 then delete;
   if mlpa = 1 then lyms = 2009;
   else if mlpb = 1 then lyms = 2001;
   else if mlpc = 1 then lyms = 1990;
   else if mlpd = 1 then lyms = 1980;
   else if mlpe = 1 then lyms = 1974;
   else if mlpf = 1 then lyms = 1964;
   else if mlpg = 1 then lyms = 1960;
   else if mlph = 1 then lyms = 1954;
   else if mlpi = 1 then lyms = 1950;
   else if mlpj = 1 then lyms = 1946;
   else if mlpk = 1 then lyms = 1941;
   else lyms = . ;
   vetage = agep-(2010-lyms);
   if vetage < 17 then vetage = 17;
run;

proc surveyfreq data=pus.veterans;
   table agep;
   ods output OneWay=pus.freqout;
   weight pwgtp;
run;
```

← Veteran Status here
← Last possible year of Service starts here

← Latest possible age that Veteran created
(age when exited active duty)

again the use of a 'survey'
procedure

Veterans and the ACS

26 Has this person ever served on active duty in the U.S. Armed Forces, military Reserves, or National Guard? *Active duty does not include training for the Reserves or National Guard, but DOES include activation, for example, for the Persian Gulf War.*

☐ Yes, now on active duty
☐ Yes, on active duty during the last 12 months, but not now
☐ Yes, on active duty in the past, but not during the last 12 months
☐ No, training for Reserves or National Guard only → SKIP to question 28a
☐ No, never served in the military → SKIP to question 29a

```
/* the dataset with ACS data */
%let acsds=pus.pus2010;

proc surveyfreq data = &acsds;
    weight pwgtp;
    ods output OneWay=pus.mildist;
    tables mil / clwt alpha=0.1;
run;
```

Note the use of a 'survey' procedure

The SURVEYFREQ Procedure

Data Summary

Number of Observations   3097724
Sum of Weights   313071822

Served in Armed Forces

MIL	Frequency	Weighted Frequency	Std Dev of Wgt Freq	90% Confidence Limits for Wgt Freq		Percent	Std Err of Percent
1	9628	1048236	14046	1025132	1071340	0.4324	0.0058
2	4819	607755	10684	590182	625328	0.2507	0.0044
3	241316	21282767	49991	21200539	21364995	8.7797	0.0206
4	30303	2550498	17068	2522424	2578572	1.0522	0.0070
5	2154292	216918400	113709	216731366	217105434	89.4850	0.0226
Total	2440358	242407656	106878	242231858	242583454	100.000	

Frequency Missing = 657366

ACS 2010 and Veterans (age)

The SURVEYFREQ Procedure

Data Summary

Number of Observations   246135
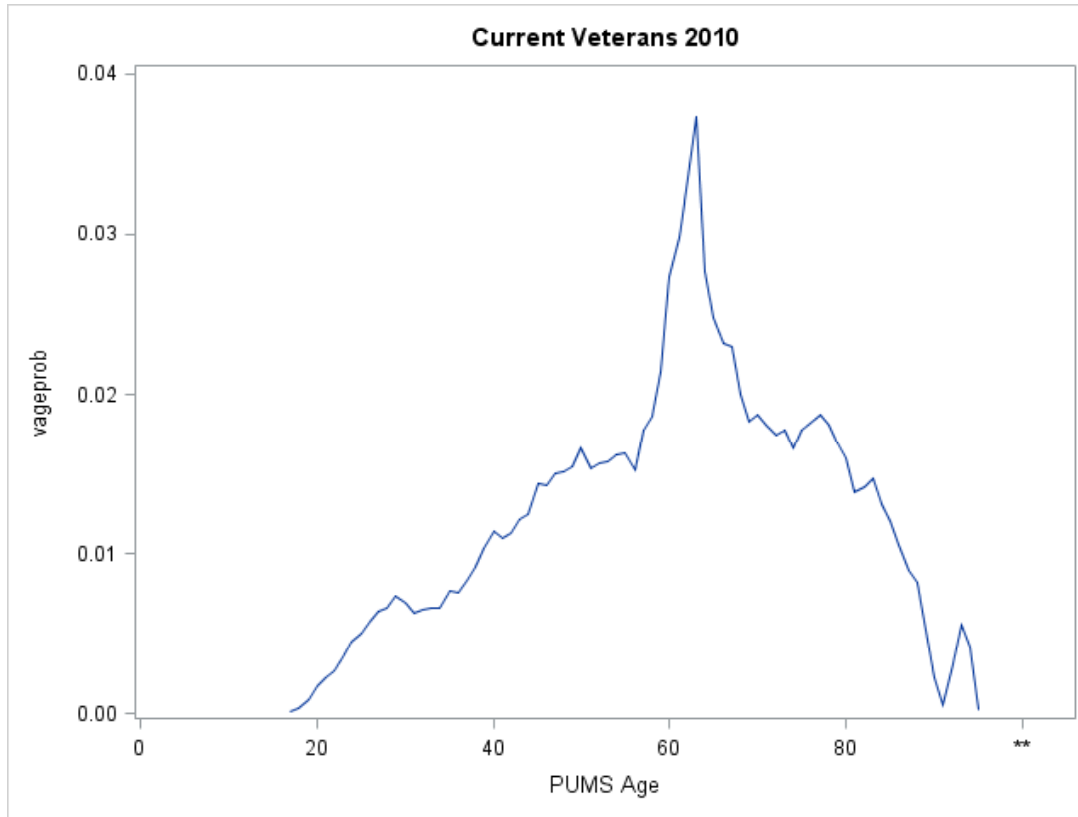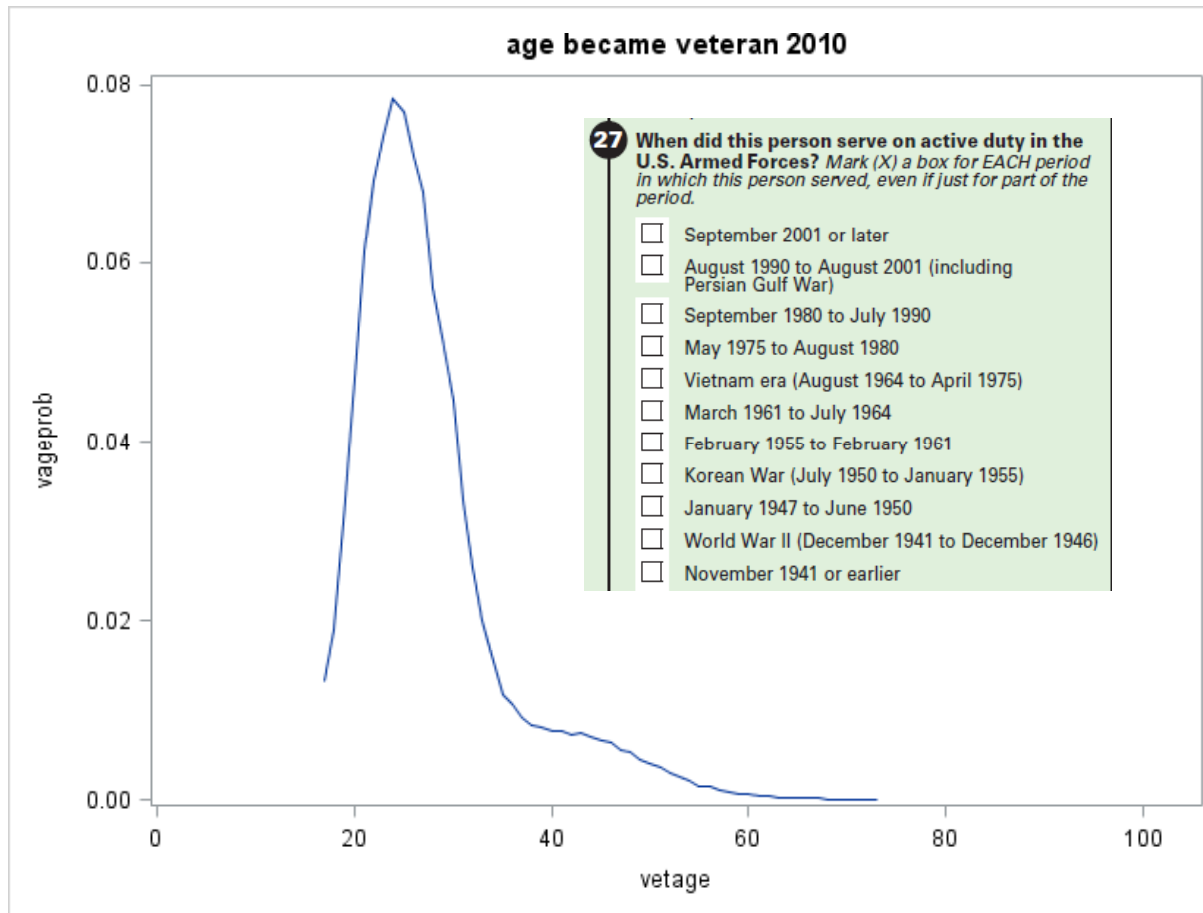Sum of Weights   21890522

The SURVEYFREQ Procedure

Table of vetage

vetage	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
17	3213	289717	6118	1.3235	0.0279
18	4869	414540	6950	1.8937	0.0317
19	8442	731215	9381	3.3403	0.0427
20	11934	1031681	11147	4.7129	0.0506
21	15658	1339253	12436	6.1180	0.0564
22	17695	1520050	13317	6.9439	0.0603
23	18645	1627191	13909	7.4333	0.0629
24	19337	1716637	14484	7.8419	0.0653
25	19182	1686654	14212	7.7050	0.0642
26	17850	1576705	13886	7.2027	0.0627
27	16546	1490579	13601	6.8092	0.0614
28	13773	1252373	12552	5.7211	0.0567
29	12380	1130161	12001	5.1628	0.0543
66	44	3665	631.34305	0.0167	0.0029
67	37	3185	569.12486	0.0145	0.0026
68	14	1385	405.29480	0.0063	0.0019
69	19	2273	665.36893	0.0104	0.0030
70	12	825.00000	243.14702	0.0038	0.0011
71	9	732.00000	280.95577	0.0033	0.0013
72	5	296.00000	144.34586	0.0014	0.0007
73	3	192.00000	111.02207	0.0009	0.0005
Total	246135	21890522	29191	100.000	

Total Population of US Veterans 2010 by ACS is the 21,890,522 number
AND
a stratified sample of 246,135 observations exist with much good info about this subset of the data (Veterans)

MWSUG 23
Minneapolis
Sept 16-18, 2012

MW
Minneapolis
Sept 16-18, 2012

ACS 2010 and Veterans (age)

Another Discrete Probability Distributions, useful for models in this form

ACS 2010, age, and 'New' Veterans

age became veteran 2010

From the Last Year of Military Service, the current year (2010), and the age of the Veteran at the time of the survey.

We can estimate the age of the Veteran at their last year of military service (lyms)

This gives us a Discrete Probability Distribution for the age of 'New' Veterans

ACS 2010, age, and 'New' Veterans

Stars and Stripes article:
"As separation rates plummet, military getting more selective"
By ERIK SLAVIN
Stars and Stripes
Published: November 22, 2009

but

DMDC is the usual ultimate source of this data.

The numbers in depth ...

FY 2005

	Retired	Separated	Total
Army	10,989	74,296	85,285
Navy	10,248	71,572	81,820
Marine Corps	2,799	29,464	32,263
Air Force	12,845	34,551	47,396
Coast Guard	908	3,347	4,255
Total	37,789	213,230	251,019

FY 2006

	Retired	Separated	Total
Army	10,676	60,357	71,033
Navy	10,106	43,531	53,637
Marine Corps	2,797	31,611	34,408
Air Force	11,304	29,498	40,802
Coast Guard	991	3,497	4,488
Total	35,874	168,494	204,368

FY 2007

	Retired	Separated	Total
Army	9,940	62,874	72,814
Navy	9,338	44,743	54,081
Marine Corps	2,280	31,503	33,783
Air Force	12,911	35,444	48,355
Coast Guard	1,035	3,023	4,058
Total	35,504	177,587	213,091

FY 2008

	Retired	Separated	Total
Army	9,662	61,921	71,583
Navy	8,985	39,248	48,233
Marine Corps	2,271	27,292	29,563
Air Force	9,257	26,942	36,199
Coast Guard	957	2,581	3,538
Total	31,132	157,984	189,116

FY 2009

	Retired	Separated	Total
Army	10,697	61,131	71,828
Navy	8,471	34,908	43,379
Marine Corps	2,223	28,407	30,630
Air Force	8,924	22,040	30,964
Coast Guard	967	2,619	3,586
Total	31,282	149,105	180,387

Figurative Model
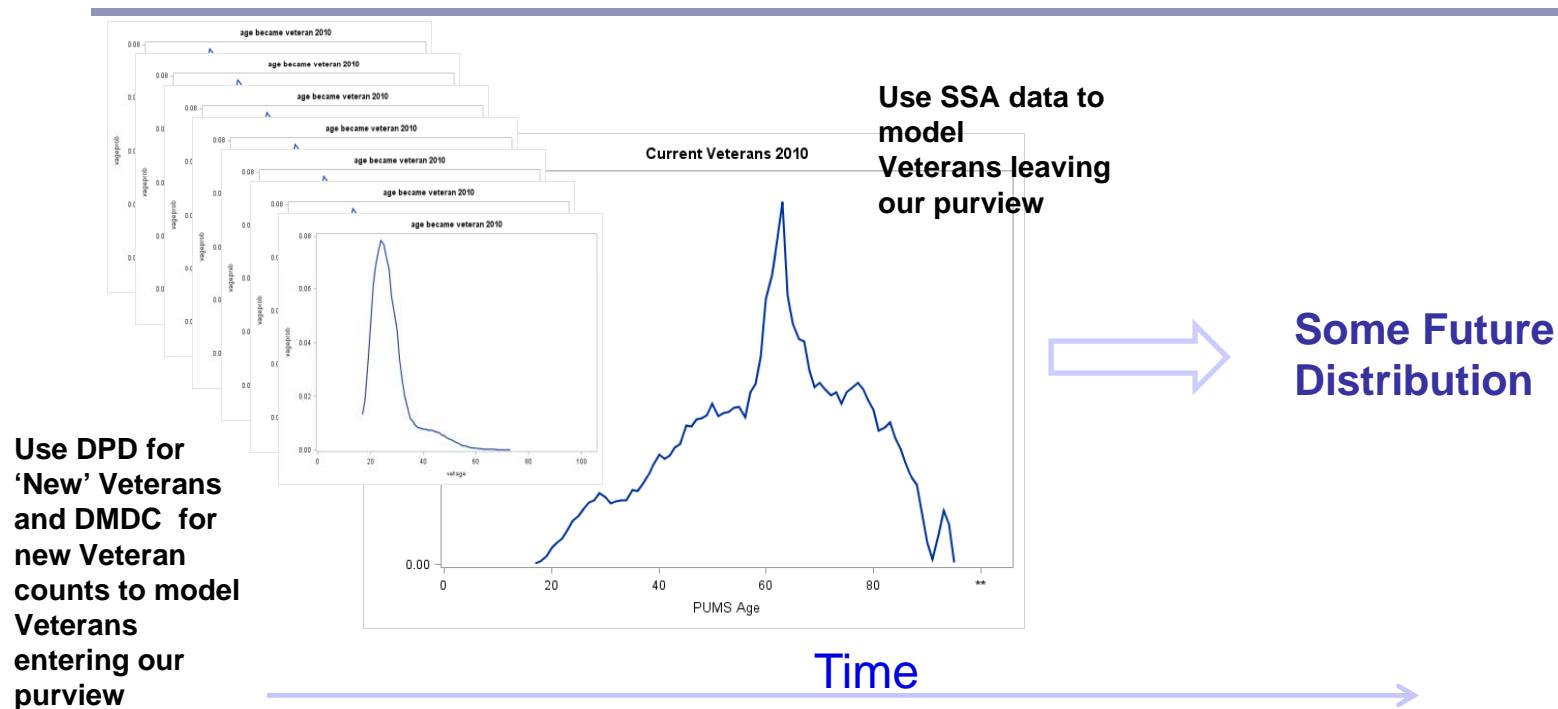
Use SSA data to model Veterans leaving our purview

Some Future Distribution

Use DPD for 'New' Veterans and DMDC for new Veteran counts to model Veterans entering our purview

Time

From the Last Year of Military Service, the current year (2010), and the age of the Veteran at the time of the survey, we can estimate the age of the Veteran at their last year of military service (lyms), this gives us a Discrete Probability Distribution for the age of 'New' and 'Current' Veterans

Expected Life Data

The Department of Veterans Affairs obligation to a Veteran ends when they die (excepting children and spouses)

Expected Life Data use

Male		
Age	Number surviving	Expected remaining life
0	100 000	75.7
5	99 247	71.3
10	99 168	66.4
15	99 059	61.4
20	98 642	56.7
25	98 098	52.0
30	97 458	47.3
35	96 777	42.6
40	96 044	38.0
45	95 138	33.2
50	93 799	28.6
55	91 750	24.2
60	88 421	20.1
65	82 846	16.3
70	74 065	12.8
75	61 792	9.9
80	46 114	7.3
85	27 898	5.4
90	12 522	4.2
95	4006	3.5
100	975	3.0
105	206	2.6

If alive at 35, the probability of still alive at 60 = 88421/96777 = 91.3657

If I have 100 entities that entered at 35 and are 60 at the year of interest, I count 91 entities.

The model will contain 91 at the year of interest.

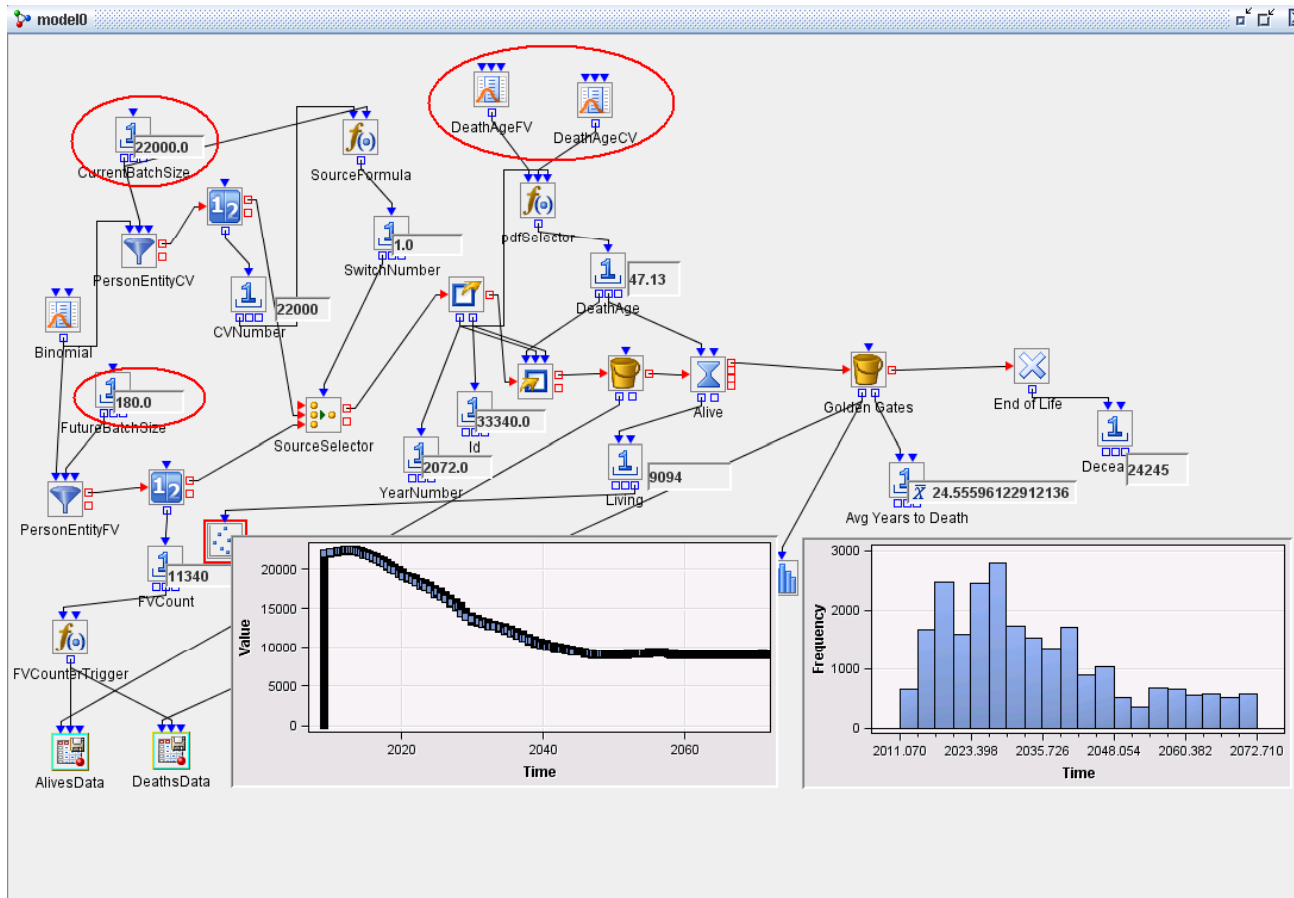# Expected Life Data (Current and New)
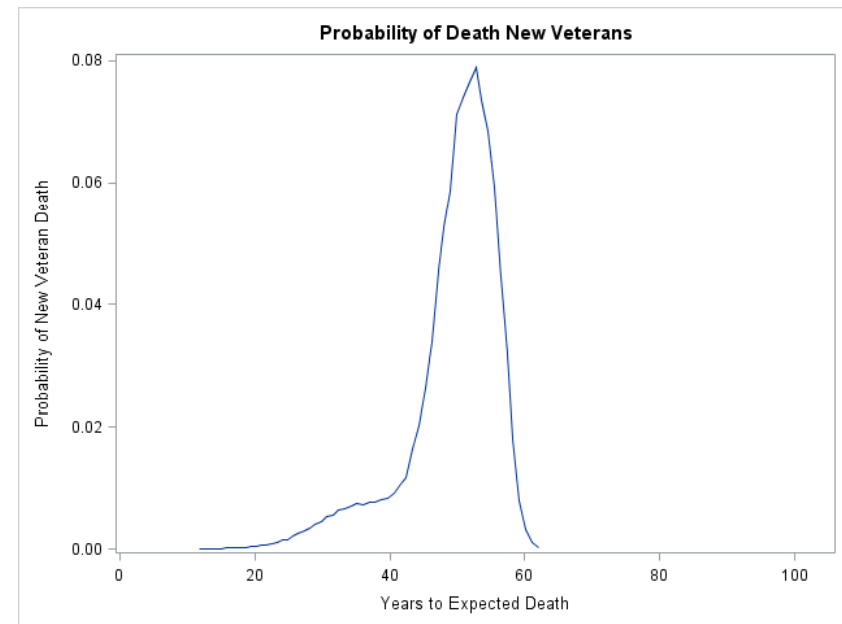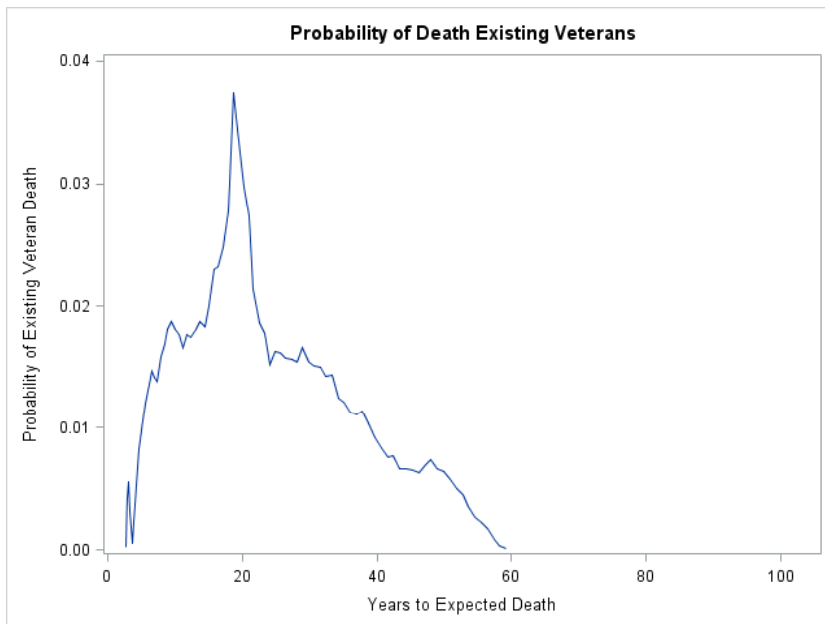
Here used Expected Life data in the Simulation to get population size

Can use Conditional (Bayes Rule) Probabilities in SAS to get accurate population shape.

SAS Simulation Studio Model
Discrete Event Simulation

Discrete Probability Distributions

Scaled down by 1/1000

Data output to SAS Datasets

Expected Life PROC FCMP

```
%macro getnol;
%global RC;
%let RC=;
%let inreturned=;
proc sql noprint;
  select nol into :inret
    from lt.LifeTable where age = &a ;
  select 1-nol/&inret into :p
    from lt.LifeTable where age = &b ;
quit;
%mend getnol;
proc fcmp outlib = sasuser.ds.functions;
  function getnol_macro(a, b);
    rc = run_macro('getnol', a, b, p);
    if rc eq 0 then return(p);
    else return(.);
  endsub;
run;
data lt.alive2020;
  set lt.goldengates2;
  inc=2020-ClockYear;
  pd=getnol_macro(eage,eage+inc);
  if ClockYear> 2020 then delete;
run;
```

Small sample of data

DeathAge	Time	ClockYear	PersonID	Exact Age	Male Life Expentancy	inc	pd
58.27	2070.3	2012	22472	18	58.27	8	0.0107
49	2061	2012	22473	28	49	8	0.0119
45.27	2057.3	2012	22474	32	45.27	8	0.0139
50.86	2062.9	2012	22475	26	50.86	8	0.0114
51.78	2063.8	2012	22476	25	51.78	8	0.0114
43.4	2055.4	2012	22477	34	43.4	8	0.0157
51.78	2063.8	2012	22478	25	51.78	8	0.0114
42.47	2054.5	2012	22479	35	42.47	8	0.0168
57.33	2069.3	2012	22480	19	57.33	8	0.0111
54.54	2066.5	2012	22481	22	54.54	8	0.0115
52.71	2064.7	2012	22482	24	52.71	8	0.0114
49.93	2061.9	2012	22483	27	49.93	8	0.0116
48.07	2060.1	2012	22484	29	48.07	8	0.0122
36.93	2048.9	2012	22485	41	36.93	8	0.0274
50.86	2062.9	2012	22486	26	50.86	8	0.0114
52.71	2064.7	2012	22487	24	52.71	8	0.0114

*
*
*

MWSUG 23
Minneapolis
Sept 16-18, 2012

MW
Minneapol
Sept 16-18, 2012

The Social Security Administration maintains Expected Life Data and one example is at http://www.ssa.gov/oact/STATS/table4c6.html

SAS Simulation Studio Model Results

Veteran age distribution 2020, by Conditional Probability
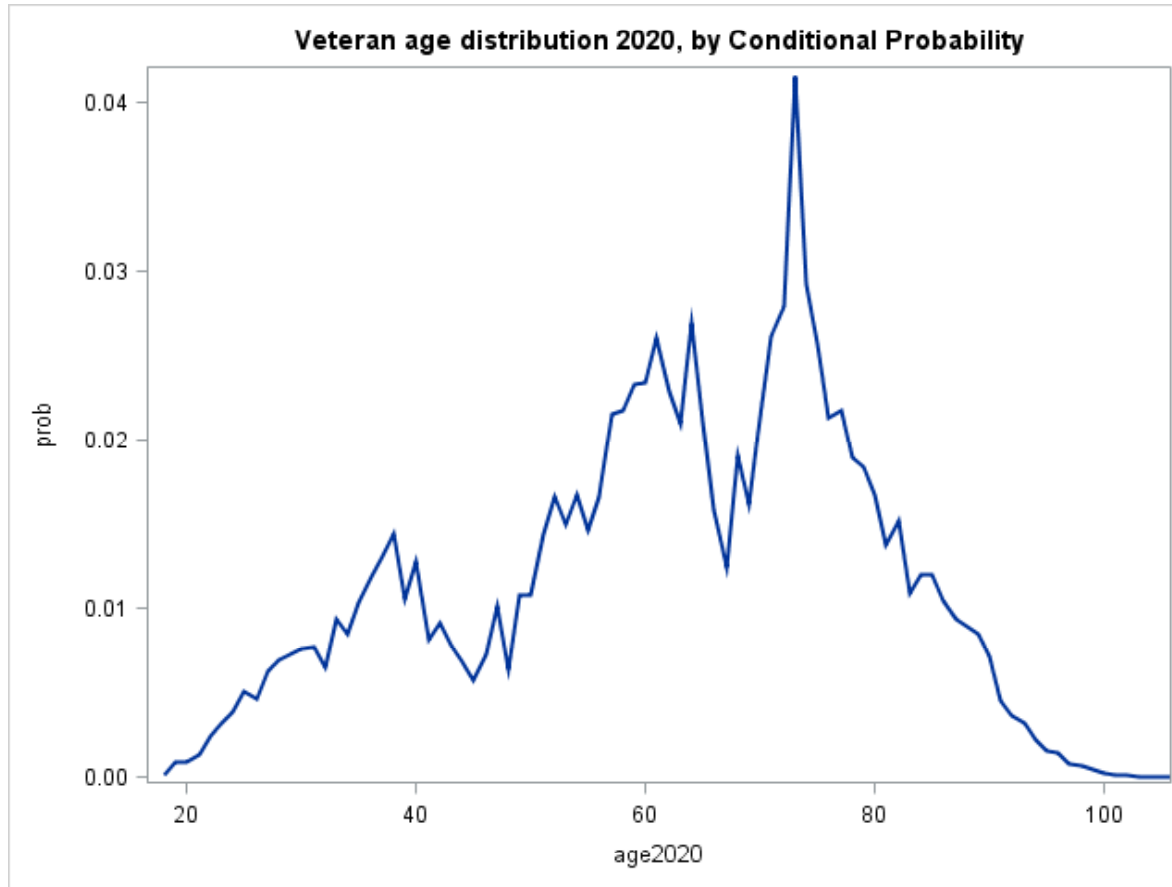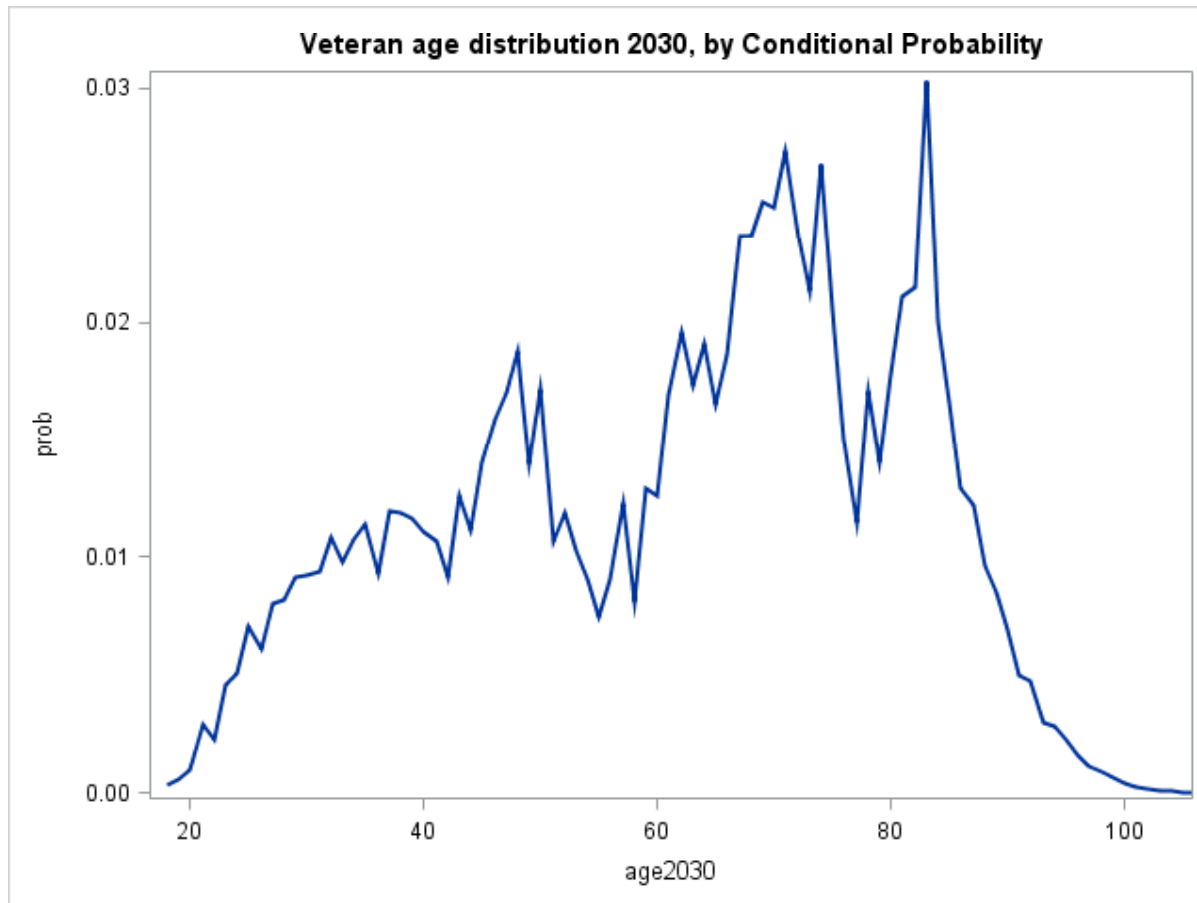
Use Conditional Probability in SAS to get more accurate population shape.

SAS Simulation Studio Model Results

Veteran age distribution 2030, by Conditional Probability

Forecast Future Service Delivery Obligations

- ## Three services examples

- Disability Compensation and Pension

- Veterans Health Administration services

- GI Bill education benefits

Data about Veterans in ACS

- Disability Status?

28 a. Does this person have a VA service-connected disability rating?

☐ Yes (such as 0%, 10%, 20%, ... , 100%)
☐ No → SKIP to question 29a

b. What is this person's service-connected disability rating?

☐ 0 percent
☐ 10 or 20 percent
☐ 30 or 40 percent
☐ 50 or 60 percent
☐ 70 percent or higher

The SURVEYFREQ Procedure

Data Summary

Number of Observations	246135
Sum of Weights	21890522

Disability rating checkbox

DRATX	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
Yes	38780	3427108	19599	15.6557	0.0873
No	207355	18463414	31340	84.3443	0.0873
Total	246135	21890522	29191	100.000	
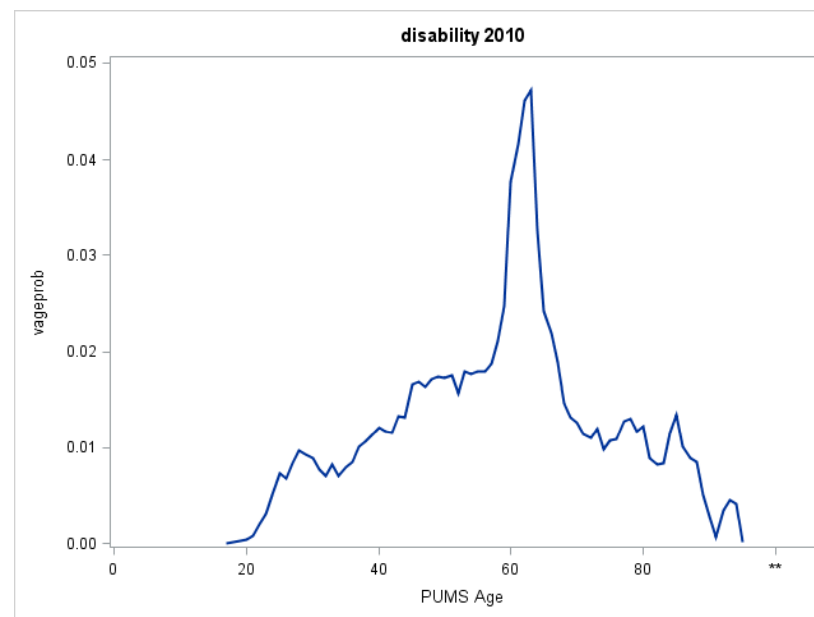
disability 2010

Population versus Utilization

- Disability distribution minus Population distribution
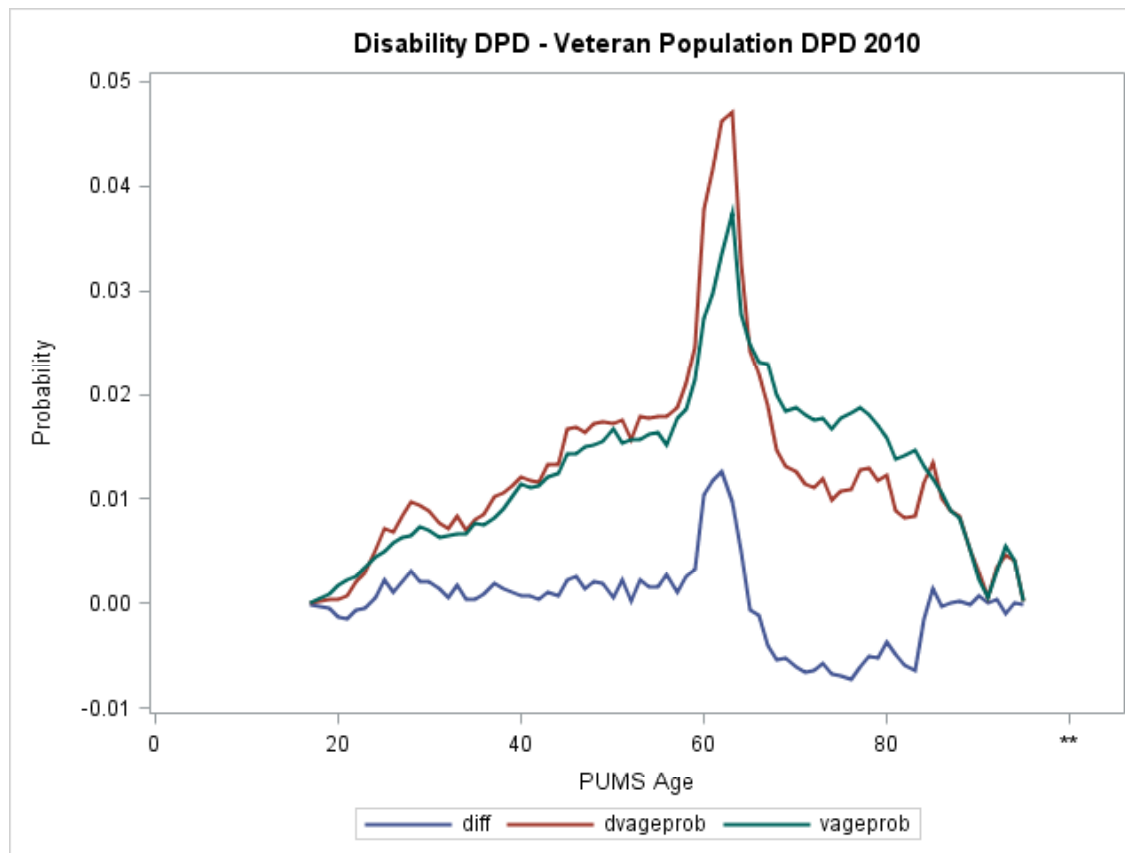
Compensation and Pension Claim Activity
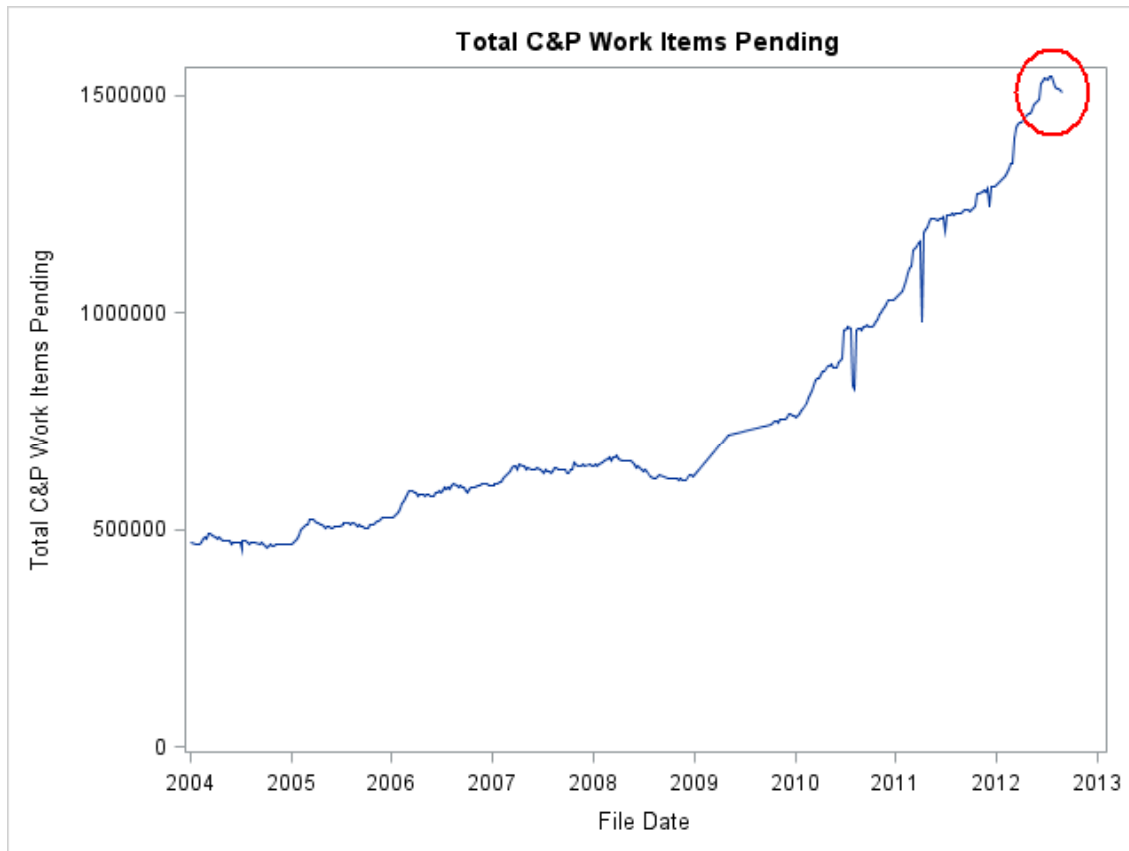
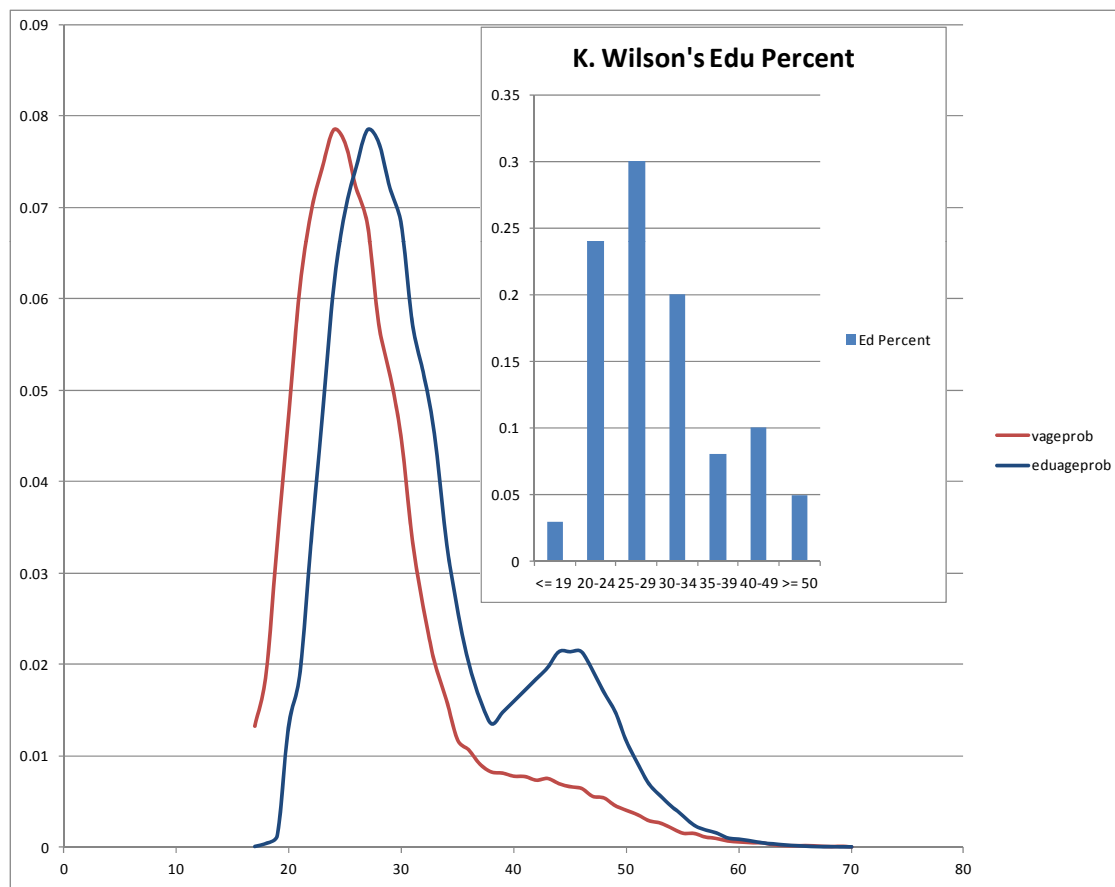# Compensation and Pension activity
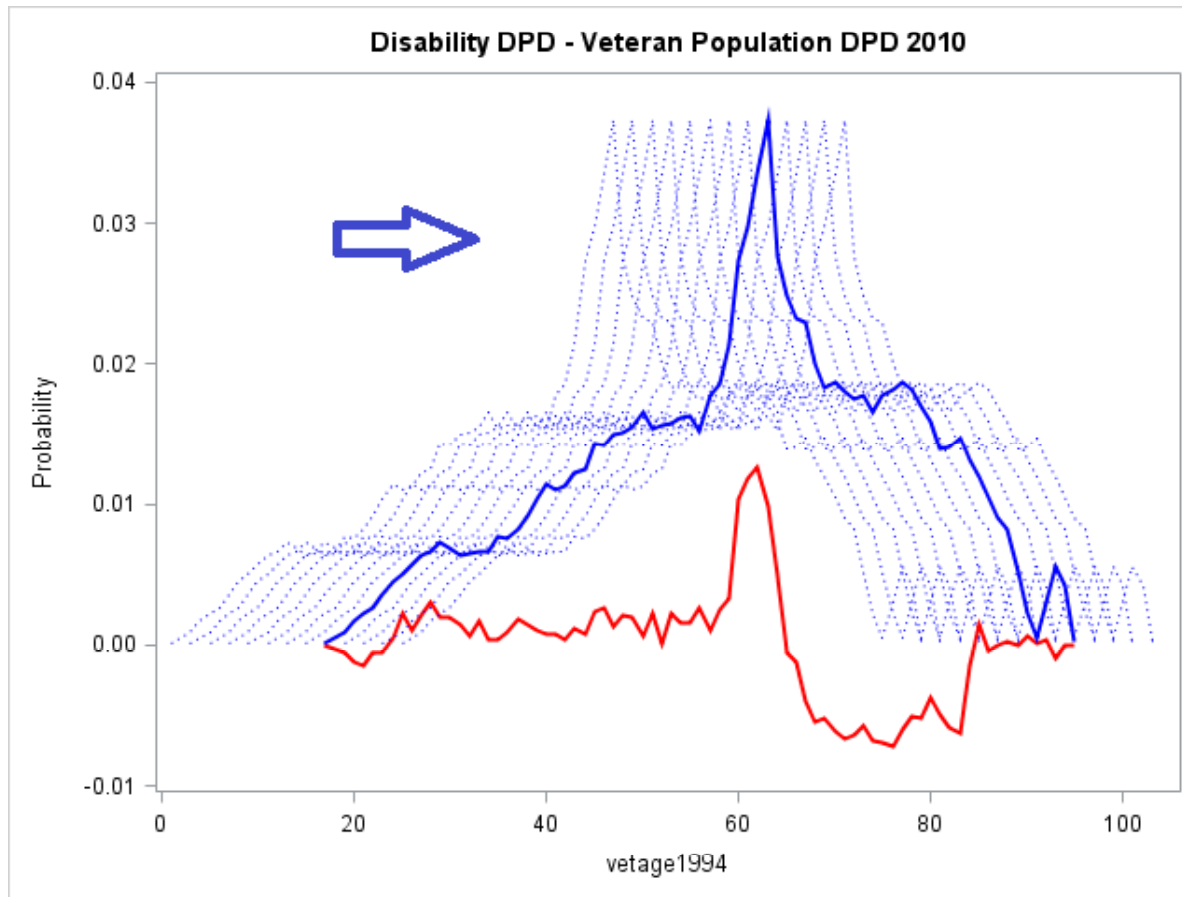
Ever used or Enrolled in VA Health?

Used or Enrolled DPD - Veteran Population DPD 2010
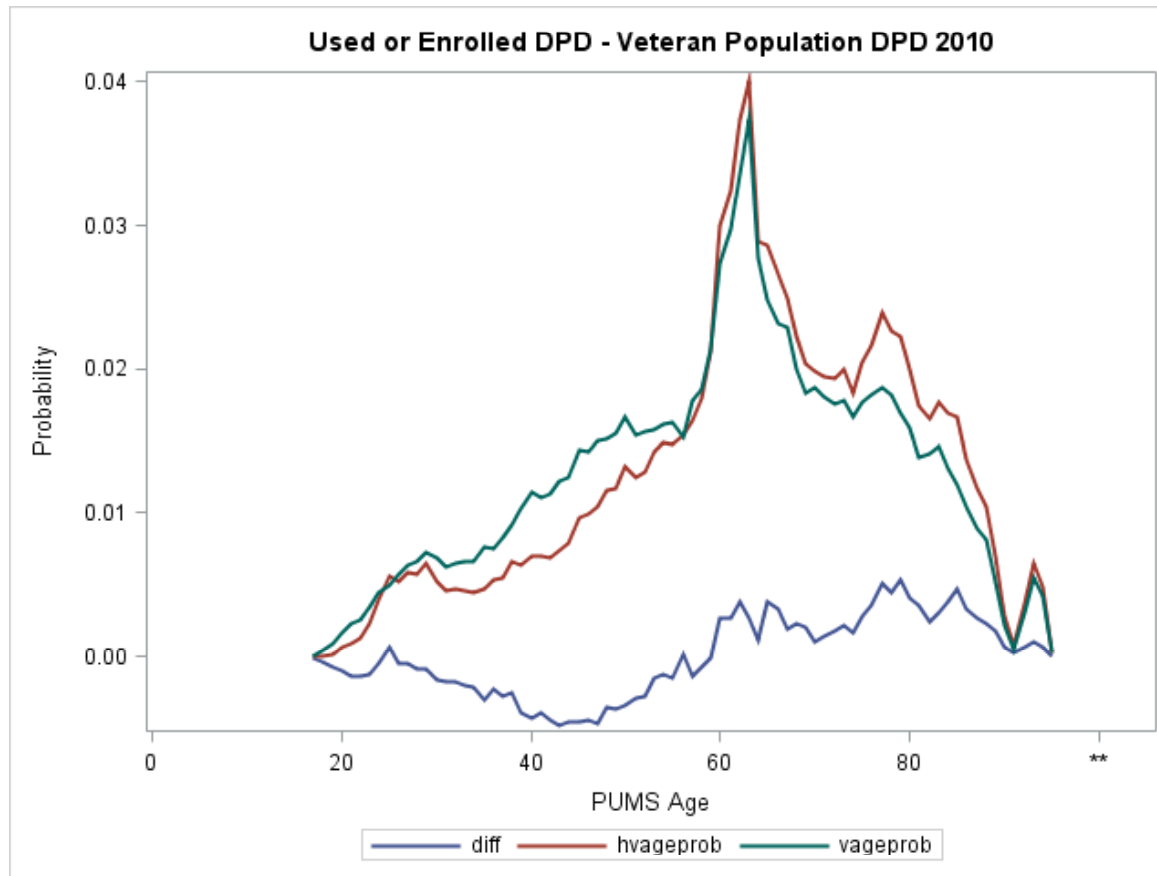
GI Bill educational Benefit

Approximated

Combination of Keith Wilson Presentation Data (from the American Council on Education presentation workbook) and the ACS data showing new Veteran age distribution

Conclusion

- **SAS®, SAS/OR® and SAS Simulation Studio® provide an unmatched capability**
  - Stratified surveys
  - Discrete Event Simulation
  - General Statistics and Analytics

Acknowlegements

Contact Information

Name: Michael C. Grierson
Enterprise: The Department of Veterans Affairs
Work Phone: 202 462-1195
E-mail: mcg@mcg-ct.com
Web: http://www.mcg-ct.com/
Find me at: www.sascommunity.org

The views expressed in this paper are those of the author and do not reflect the
official policy or position of the Department of the Veterans Affairs, or the U.S. Government.

MWSUG 23
Minneapolis
Sept 16-18, 2012
MW
Minneapo
Sept 16-18, 2012

BACKUP SLIDES

ACS 2010 and accuracy

The standard error of X can be approximated after the replicate estimates X_1 through X_{80} are computed. The standard error is estimated using the sum of squared differences between each replicate estimate X_r and the full sample estimate X. The standard error formula is:

$$SE(X) = \sqrt{\frac{4}{80} \sum_{r=1}^{80} (X_r - X)^2}$$

If X is zero, then use the generalized variance method for zero estimates given in section 6.2, Standard Errors for Totals and Percentages, to approximate the standard error.

As we mentioned earlier, the standard error can be used to form a 90% confidence interval around the estimate (X) as follows:
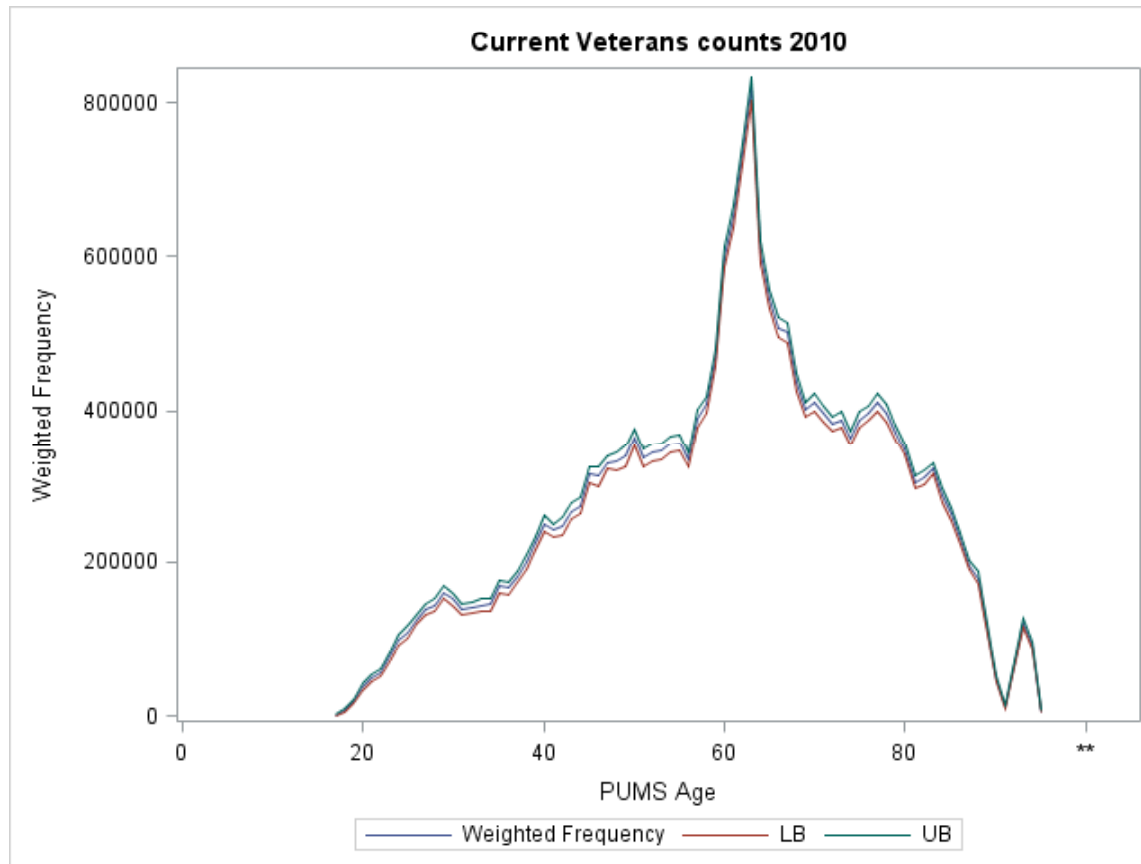
LB=Lower bound = X - 1.645*SE(X)
UB=Upper bound = X + 1.645*SE(X)

The 90% confidence interval is the interval (LB, UB).

ACS 2010 and accuracy

An actual distribution, not as useful for models in this form. Note that ACS uses a technique called Successive Difference Replication method, which must be manually coded in SAS.

ACS 2010 disability odds ratios

Odds Ratio Estimates		
Effect	**Point Estimate**	**95% Wald Confidence Limits**
agecat ages 25 to 35 vs ages 15 to 25	1.900	1.631	2.213
agecat ages 35 to 45 vs ages 15 to 25	1.594	1.373	1.850
agecat ages 45 to 55 vs ages 15 to 25	1.626	1.405	1.883
agecat ages 55 to 60 vs ages 15 to 25	1.652	1.424	1.917
agecat ages 60 to 65 vs ages 15 to 25	2.010	1.738	2.325
agecat ages 65 to 70 vs ages 15 to 25	1.186	1.023	1.375
agecat ages 70 to 75 vs ages 15 to 25	0.865	0.744	1.005
agecat ages 75 to 85 vs ages 15 to 25	0.909	0.785	1.053
agecat ages 85 to 95 vs ages 15 to 25	1.474	1.268	1.714
agecat ages 95+ vs ages 15 to 25	1.027	0.427	2.469

Note the odds ratios decrease after 65 just as the diff DPD graph shows
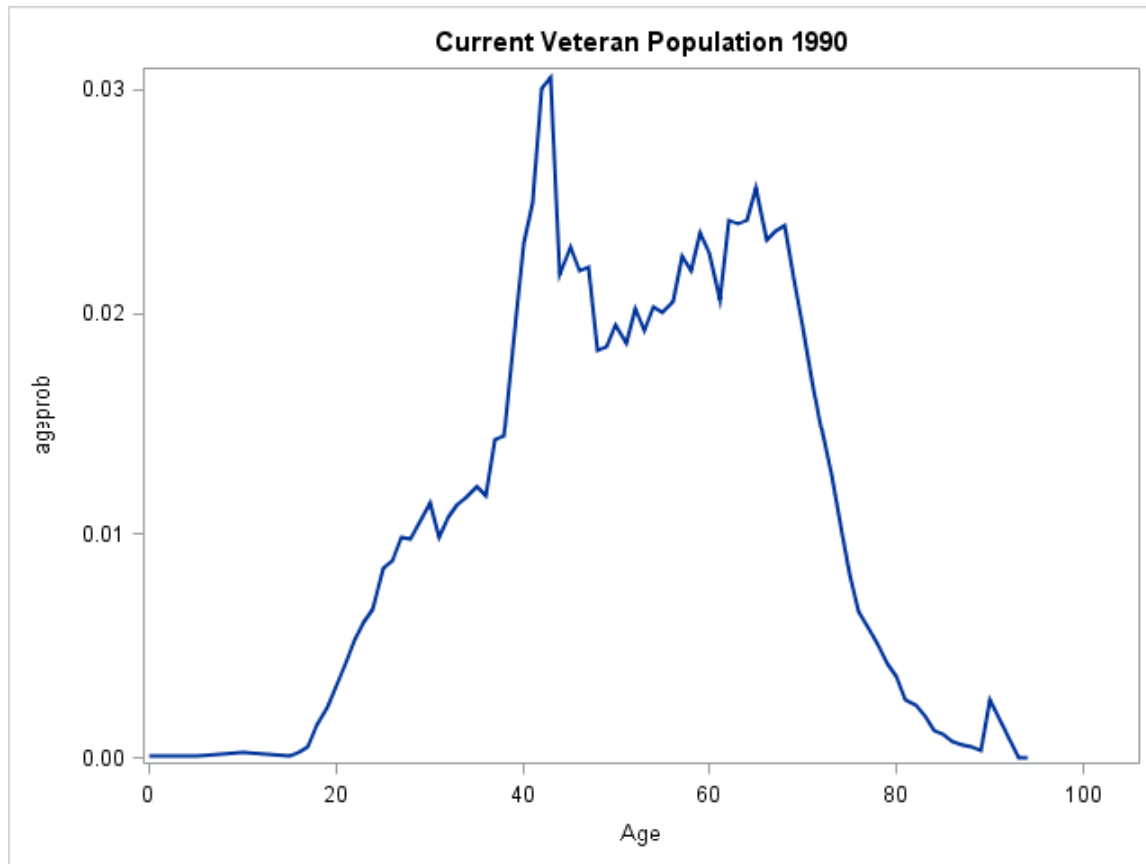
Veteran Population Distribution 2000

Current Veteran Population 2000
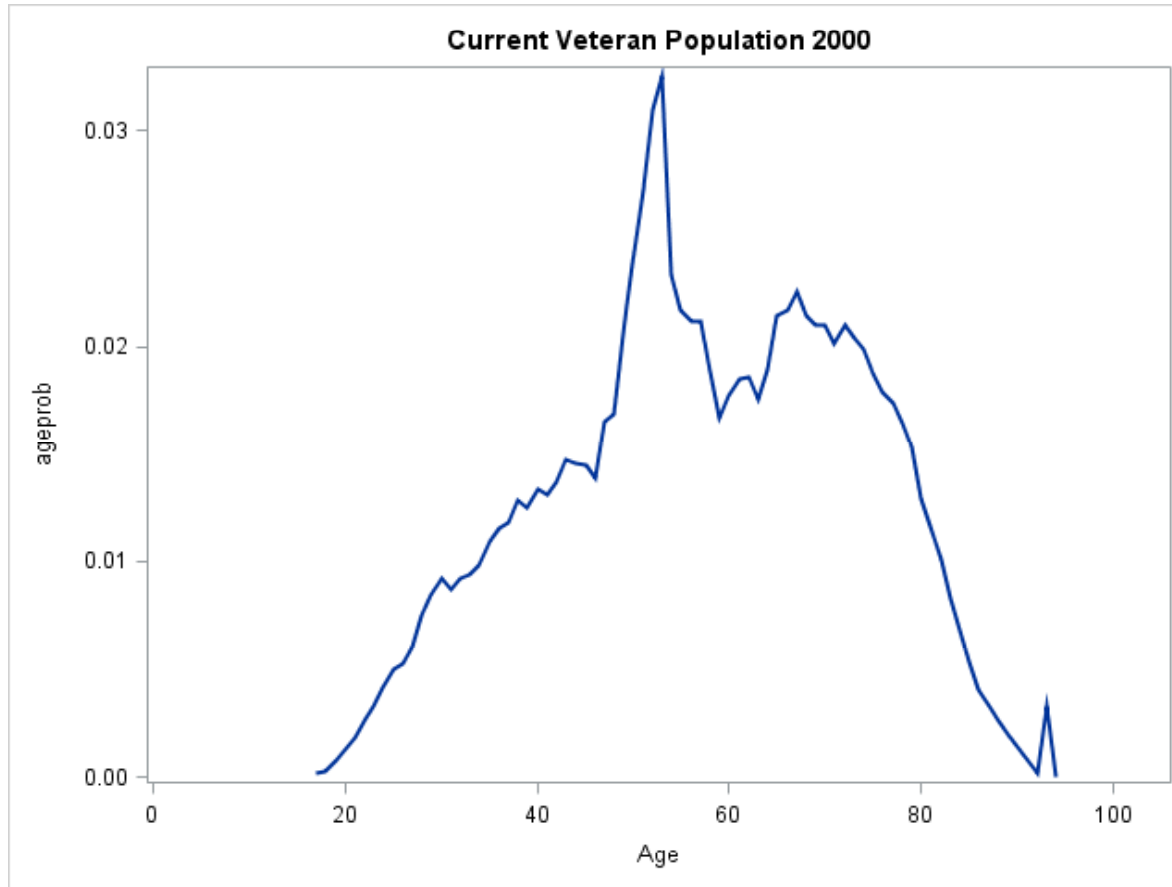
Of interest primarily to show
the WWII Veteran Population
Relative to the Vietnam
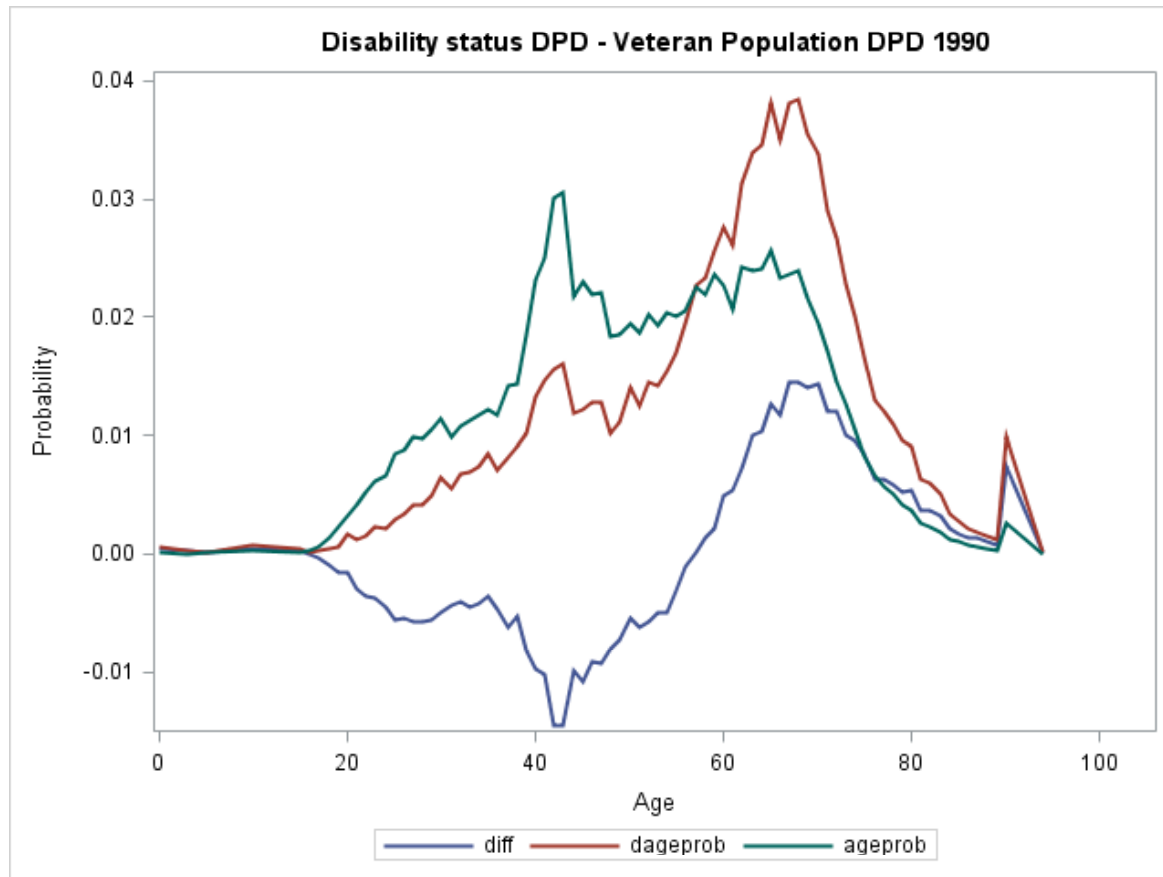Veteran population

Veteran Population Distribution 1990

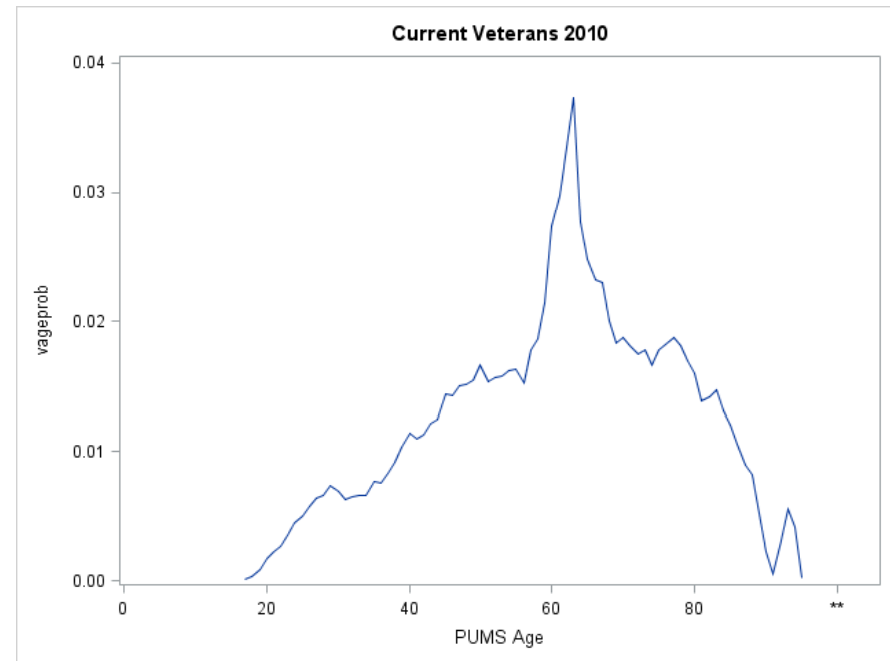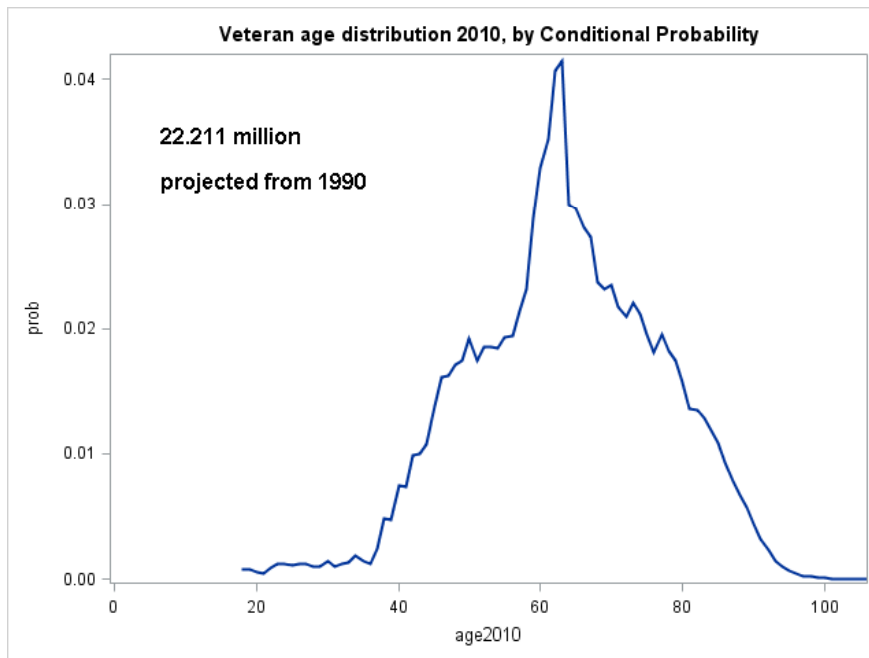**Current Veteran Population 1990**

Of interest primarily because WWII veterans are passing through the 65 year old utilization peak.

Veteran Population Distribution 1990

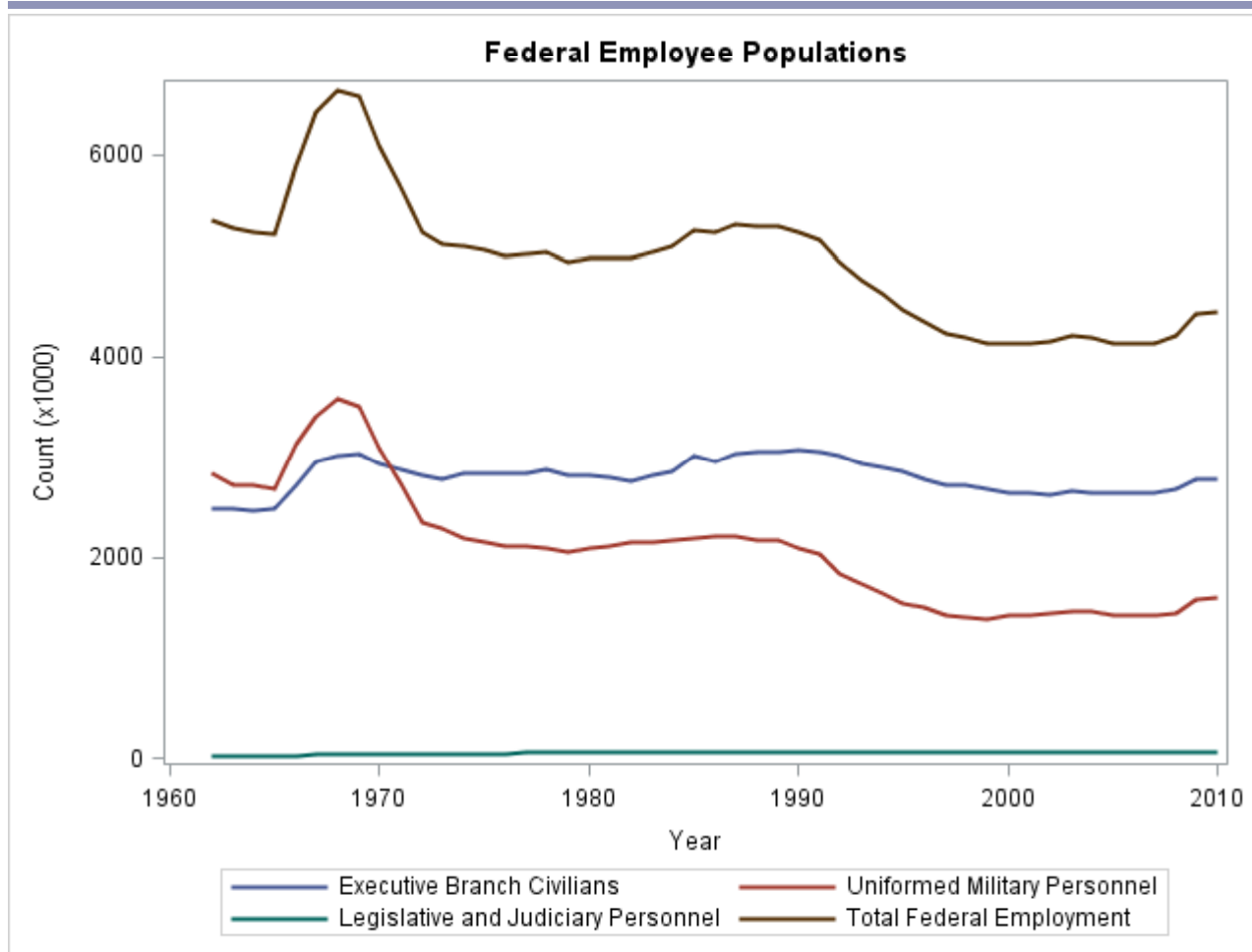**Disability status DPD - Veteran Population DPD 1990**

Of interest primarily because WWII veterans are passing through the 65 year old utilization peak.

Veteran Population Distribution
Project from 1990 to 2010 comparison

Veteran age distribution 2010, by Conditional Probability

22.211 million

projected from 1990

Current Veterans 2010

Federal Employee population history

Federal Employee Populations

From the opm.gov website "**Total Government Employment Since 1962**" at http://www.opm.gov/feddata/historicaltables/totalgovernmentsince1962.asp

Veteran Disability Odd Ratios 1990

| Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
agecat ages 25 to 35 vs ages 15 to 25	1.443	1.281	1.625
agecat ages 35 to 45 vs ages 15 to 25	1.624	1.450	1.820
agecat ages 45 to 55 vs ages 15 to 25	1.872	1.672	2.095
agecat ages 55 to 60 vs ages 15 to 25	3.141	2.803	3.519
agecat ages 60 to 65 vs ages 15 to 25	4.557	4.070	5.103
agecat ages 65 to 70 vs ages 15 to 25	5.759	5.146	6.446
agecat ages 70 to 75 vs ages 15 to 25	6.961	6.212	7.801
agecat ages 75 to 85 vs ages 15 to 25	9.947	8.843	11.189
agecat ages 85 to 95 vs ages 15 to 25	24.990	21.183	29.483

Of interest primarily because WWII veterans are passing through the 65 year old utilization peak.

Veterans by State ACS 2010

State Veteran Population
from the ACS 2010 survey

Of interest primarily because you can show population counts on a map with proc gmap.
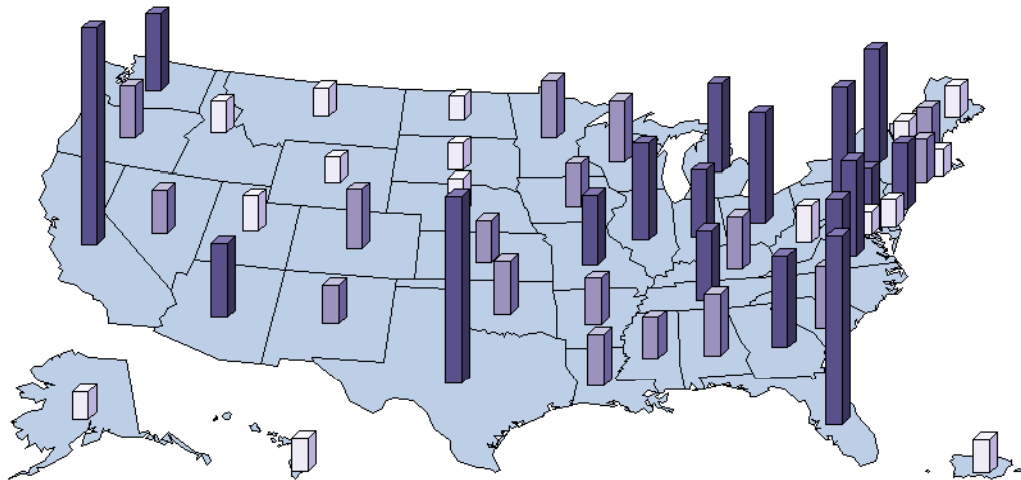
cnt   ▢ 30176 - 162582   ▨ 176349 - 423459   ▥ 427069 - 1944749

US Population by ACS 2000