

Biost 518: Applied Biostatistics II
 #### Emerson, Winter 2007

Homework #3 Key
 #### Annotated Stata Log File
 #### January 31, 2007

The following output was used to generate the numbers that I wanted to present
 #### in tables, as well as the plots I wanted to present as figures in the paper.
 #### I note that Stata does not present its output in a form suitable for presentation.
 #### Numbers need to be rounded to an interesting number of significant digits, and
 #### the columns and rows need to facilitate comparison of relevant measures.
 #### I used Excel to manipulate this output into the form I wanted, then copied the
 #### resulting tables to the MS-Word document.

Comments edited into the log file produced by Stata are
 #### on the lines that start with the four '#' signs and are
 #### printed in italics.

The Stata commands are put in **bold face**.

Stata output is displayed in **regular typeface in blue**.

Create age categories. Note the two equivalent ways to do this.

```
. summ age
  Variable |      Obs      Mean   Std. Dev.   Min   Max
-----+-----
      age |    5000   72.8304   5.596418    65   100
```

```
. g agectg=age
```

```
. recode agectg 65/69.9=65 70/74.9=70 75/79.9=75 80/84.9=80 85/89.9=85 90/94.9=90 95/
> 100=95
```

```
(agectg: 583 changes made)
```

```
. g agectg2= trunc(age/5) * 5
```

```
. table agectg agectg2
```

```
-----
  agectg |      65      70      75      80      85      90      95
-----+-----
      65 |    117
      70 |         305
      75 |             187
      80 |                 81
      85 |                     35
      90 |                         8
      95 |                             2
-----
```

Descriptive statistics by diabetes and age.

```
. format crt %7.2f
```

```
. bysort diabetes: tabstat crt, col(stat) stat(n mean sd min p25 p50 p75 max)
by(agectg) format
```

```
-> diabetes = 0
```

Summary for variables: crt
by categories of: agectg

agectg	N	mean	sd	min	p25	p50	p75	max
65	99.00	1.05	0.26	0.50	0.90	1.00	1.20	2.40
70	275.00	1.04	0.25	0.60	0.80	1.00	1.20	2.10
75	168.00	1.05	0.30	0.50	0.80	1.00	1.20	2.60
80	71.00	1.04	0.29	0.50	0.80	1.00	1.20	1.80
85	32.00	1.25	0.61	0.60	0.95	1.10	1.45	4.00
90	7.00	1.23	0.17	1.00	1.10	1.20	1.40	1.50
95	2.00	1.10	0.28	0.90	0.90	1.10	1.30	1.30
Total	654.00	1.06	0.30	0.50	0.90	1.00	1.20	4.00

-> diabetes = 1

Summary for variables: crt
by categories of: agectg

agectg	N	mean	sd	min	p25	p50	p75	max
65	16.00	1.07	0.19	0.80	0.95	1.00	1.25	1.40
70	30.00	1.13	0.47	0.60	0.90	1.00	1.30	3.20
75	19.00	1.11	0.23	0.70	1.00	1.20	1.20	1.50
80	10.00	1.20	0.29	0.80	1.00	1.20	1.40	1.70
85	3.00	1.30	0.26	1.10	1.10	1.20	1.60	1.60
90	1.00	1.40	.	1.40	1.40	1.40	1.40	1.40
Total	79.00	1.13	0.34	0.60	0.90	1.10	1.30	3.20

```
. twoway (scatter crt age if diabetes==0, col(blue) msym(d) jitter(2)) (lowess crt
age if diabetes==0, col(blue)) (scatter crt age if diabetes==1, col(red) msym(T)
jitter(2)) (lowess crt age if diabetes==1, col(red) lpat(dash))
```

T test and regression for 1b. (note the similarity of answers—either one would do)

```
. ttest crt, by(diabetes) unequal
```

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	654	1.055963	.0116208	.2971837	1.033145	1.078782
1	79	1.13038	.0384862	.342073	1.05376	1.207
combined	733	1.063984	.0111903	.3029661	1.042015	1.085953
diff		-.0744164	.0402024		-.1542529	.00542

```
diff = mean(0) - mean(1) t = -1.8510
Ho: diff = 0 Satterthwaite's degrees of freedom = 92.7791
```

```
Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0337 Pr(|T| > |t|) = 0.0673 Pr(T > t) = 0.9663
```

```
. regress crt diabetes, robust
```

Linear regression

Number of obs = 733
 F(1, 731) = 3.46
 Prob > F = 0.0634
 R-squared = 0.0058
 Root MSE = .30229

crt	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
diabetes	.0744164	.0400206	1.86	0.063	-.0041525	.1529854
_cons	1.055963	.0116278	90.81	0.000	1.033135	1.078791

Regression for 1c.

. regress crt age, robust

Linear regression

Number of obs = 733
 F(1, 731) = 4.64
 Prob > F = 0.0315
 R-squared = 0.0100
 Root MSE = .30165

crt	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0055707	.0025859	2.15	0.032	.0004941	.0106473
_cons	.64849	.1897312	3.42	0.001	.2760069	1.020973

Regression for 1d and 1e.

. regress crt diabetes age, robust

Linear regression

Number of obs = 733
 F(2, 730) = 4.97
 Prob > F = 0.0072
 R-squared = 0.0162
 Root MSE = .30092

crt	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
diabetes	.0765244	.0396464	1.93	0.054	-.0013101	.1543589
age	.0056632	.0025746	2.20	0.028	.0006086	.0107178
_cons	.6333456	.1882908	3.36	0.001	.2636895	1.003002

Regression for 1f.

. regress crt age if diabetes==0, robust

Linear regression

Number of obs = 654
 F(1, 652) = 3.80
 Prob > F = 0.0517
 R-squared = 0.0100

Root MSE = .29592

crt	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0054463	.0027941	1.95	0.052	-.0000402	.0109328
_cons	.6495328	.2045285	3.18	0.002	.2479188	1.051147

Regression for 1g.

. regress crt age if diabetes==1, robust

Linear regression

Number of obs = 79
 F(1, 77) = 1.79
 Prob > F = 0.1849
 R-squared = 0.0139
 Root MSE = .34188

crt	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0075678	.0056567	1.34	0.185	-.0036962	.0188317
_cons	.5684487	.4324102	1.31	0.193	-.2925899	1.429487

Regression for 1h. Note that for comparison, I also compute a Z score from the results for 1f and 1g.

. g diabage= diabetes * age

. regress crt age diabetes diabage, robust

Linear regression

Number of obs = 733
 F(3, 729) = 4.04
 Prob > F = 0.0073
 R-squared = 0.0163
 Root MSE = .3011

crt	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0054463	.0027975	1.95	0.052	-.0000458	.0109383
diabetes	-.081084	.4745289	-0.17	0.864	-1.01269	.8505222
diabage	.0021215	.0062598	0.34	0.735	-.0101679	.0144109
_cons	.6495328	.204775	3.17	0.002	.2475137	1.051552

. display (.0075678 - .0054463)
 .0021215

. display sqrt(.0056567^2 + .0027941^2)
 .00630914

. display (.0075678 - .0054463) / sqrt(.0056567^2 + .0027941^2)
 .33625818


```

Log pseudolikelihood = -206.289
Wald chi2(2) = 15.06
Prob > chi2 = 0.0005
Pseudo R2 = 0.0290
    
```

h1crt	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	1.080451	.0216035	3.87	0.000	1.038928	1.123634
diabetes	1.300962	.5243976	0.65	0.514	.5904167	2.866624

Regression for Problem 3b. I chose to only use a quadratic term to look for nonlinearity.

```

. g agesqr= age^2
. logistic h1crt age agesqr diabetes, robust
    
```

```

Logistic regression
Log pseudolikelihood = -204.61896
Number of obs = 733
Wald chi2(3) = 14.35
Prob > chi2 = 0.0025
Pseudo R2 = 0.0368
    
```

h1crt	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	2.654518	1.215544	2.13	0.033	1.081951	6.512743
agesqr	.994297	.002896	-1.96	0.050	.9886372	.9999893
diabetes	1.293456	.5188437	0.64	0.521	.5892631	2.839186

Just for fun, I present alternative tests for nonlinearity based on a 3rd order polynomial. I have to test both the agesqr and agecub terms in this case.

```

. g agecub= age^3
. logistic h1crt age agesqr agecub diabetes, robust
    
```

```

Logistic regression
Log pseudolikelihood = -204.59883
Number of obs = 733
Wald chi2(4) = 15.00
Prob > chi2 = 0.0047
Pseudo R2 = 0.0369
    
```

h1crt	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	.5545643	3.791475	-0.09	0.931	8.40e-07	366009.3
agesqr	1.014204	.0865194	0.17	0.869	.8580476	1.19878
agecub	.9999167	.000353	-0.24	0.814	.999225	1.000609
diabetes	1.292792	.518724	0.64	0.522	.5888302	2.83836

```

. test agesqr agecub
    
```

- (1) agesqr = 0
- (2) agecub = 0

```
chi2( 2) = 5.01
Prob > chi2 = 0.0818
```

Yet another option that could have been used to test for nonlinearity. This is rather unusual.

```
. g logage= log(age)
```

```
. logistic hicrt age logage diabetes, robust
```

```
Logistic regression                               Number of obs   =          733
                                                  Wald chi2(3)    =          13.96
                                                  Prob > chi2     =          0.0030
Log pseudolikelihood = -204.63811                Pseudo R2      =          0.0367
```

hicrt	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
age	.4367581	.2057528	-1.76	0.079	.1734793	1.099599
logage	5.92e+30	2.18e+32	1.93	0.054	.2777408	1.26e+62
diabetes	1.293995	.5190559	0.64	0.521	.5895121	2.840352

Creating age strata for Problem 4. I have to make sure there are enough deaths in each stratum.

```
. table agectg death
```

agectg	death	
	0	1
65	102	15
70	261	44
75	155	32
80	59	22
85	19	16
90	4	4
95	2	

```
. g agestrata= agectg
```

```
. recode agestrata 65/74=70 75/84=80 85/99=90
(agestrata: 345 changes made)
```

Descriptive statistics for Problem 4a.

```
. stset obstime death
```

```
failure event: death != 0 & death < .
obs. time interval: (0, obstime]
exit on or before: failure
```

```
-----
735 total obs.
0 exclusions
-----
```

```

735 obs. remaining, representing
133 failures in single record/single failure data
1325995 total analysis time at risk, at risk from t = 0
        earliest observed entry t = 0
        last observed exit t = 2159
    
```

```

. sts graph, by(agestrata diabetes) col(black black blue blue yellow yellow)
lpat(solid dash solid dash solid dash)
    
```

```

        failure _d: death
        analysis time _t: obstime
    
```

```

. sts list, by(agestrata diabetes) at(730 1461)
    
```

```

        failure _d: death
        analysis time _t: obstime
    
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]	

agestrata=70 diabetes=0						
730	364	13	0.9654	0.0094	0.9412	0.9798
1461	352	12	0.9335	0.0128	0.9032	0.9546
agestrata=70 diabetes=1						
730	42	5	0.8913	0.0459	0.7584	0.9533
1461	39	4	0.8043	0.0585	0.6577	0.8930
agestrata=80 diabetes=0						
730	230	10	0.9582	0.0130	0.9236	0.9773
1461	211	19	0.8787	0.0211	0.8301	0.9141
agestrata=80 diabetes=1						
730	28	2	0.9310	0.0471	0.7514	0.9823
1461	24	4	0.7931	0.0752	0.5964	0.9013
agestrata=90 diabetes=0						
730	38	4	0.9024	0.0463	0.7606	0.9622
1461	31	7	0.7317	0.0692	0.5681	0.8415
agestrata=90 diabetes=1						
730	0	0	1.0000	.	.	.
1461	3	2	0.5000	0.2500	0.0578	0.8449

Note: Survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

```

. sts list, by(hicrt diabetes) at(730 1461)
    
```

```

        failure _d: death
        analysis time _t: obstime
    
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]	

hicrt=0 diabetes=0						
730	581	20	0.9667	0.0073	0.9488	0.9784
1461	551	30	0.9167	0.0113	0.8915	0.9362
hicrt=0 diabetes=1						
730	66	6	0.9155	0.0330	0.8215	0.9611
1461	57	9	0.7887	0.0484	0.6743	0.8668

hicrt=1 diabetes=0							
730	48	7	0.8704	0.0457	0.7472	0.9360	
1461	40	8	0.7222	0.0610	0.5822	0.8222	
hicrt=1 diabetes=1							
730	8	1	0.8750	0.1169	0.3870	0.9814	
1461	7	1	0.7500	0.1531	0.3148	0.9309	

Note: Survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

Regression for Problem 4b.

. **stcox crt**

failure _d: death
analysis time _t: obstime

Iteration 0: log likelihood = -854.83777
 Iteration 1: log likelihood = -850.41092
 Iteration 2: log likelihood = -838.57258
 Iteration 3: log likelihood = -837.47804
 Iteration 4: log likelihood = -837.45639
 Iteration 5: log likelihood = -837.45638
 Refining estimates:
 Iteration 0: log likelihood = -837.45638

Cox regression -- Breslow method for ties

No. of subjects =	733	Number of obs =	733
No. of failures =	133		
Time at risk =	1322206		
Log likelihood =	-837.45638	LR chi2(1) =	34.76
		Prob > chi2 =	0.0000

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
crt	3.996085	.7661438	7.23	0.000	2.744349 5.818756

Regression for Problem 4c.

. **stcox crt age diabetes**

failure _d: death
analysis time _t: obstime

Iteration 0: log likelihood = -854.83777
 Iteration 1: log likelihood = -840.34226
 Iteration 2: log likelihood = -825.15922
 Iteration 3: log likelihood = -823.8372
 Iteration 4: log likelihood = -823.8065
 Iteration 5: log likelihood = -823.80648
 Refining estimates:
 Iteration 0: log likelihood = -823.80648

Cox regression -- Breslow method for ties

No. of subjects =	733	Number of obs =	733
-------------------	-----	-----------------	-----

```
No. of failures =          133
Time at risk   =          1322206
Log likelihood  =          -823.80648
LR chi2(3)     =           62.06
Prob > chi2    =           0.0000
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
crt	3.398002	.6744821	6.16	0.000	2.302849 5.013971
age	1.065683	.0143224	4.73	0.000	1.037979 1.094128
diabetes	1.975062	.4368777	3.08	0.002	1.280258 3.04694

Problem 2.

Categories for cholesterol.

```
. tabstat chol, stat(me sd min q max)
```

variable	mean	sd	min	p25	p50	p75	max
cholest	211.6893	39.28814	73	186	210	236	430

```
. g cholctg=cholest
```

(47 missing values generated)

```
. recode cholctg 0/160=0 160/180=1 180/200=2 200/220=3 220/240=4 240/260=5 260/500=6
(4953 changes made)
```

Descriptive statistics by cholesterol category.

```
. tabstat crp, by(cholctg) stat(n me sd min q max)
```

Summary for variables: crp

by categories of: cholctg

cholctg	N	mean	sd	min	p25	p50	p75	max
0	404	5.324876	11.20052	.1	.75	1.9	3.635	99.71
1	586	4.311075	8.461507	.07	.97	1.94	3.53	107.97
2	960	3.439906	5.326833	.1	.885	1.805	3.235	53.96
3	1028	3.453745	5.23887	.1	.95	1.875	3.425	66.47
4	859	2.913923	3.833884	.07	.96	1.82	3.22	62.79
5	541	3.581128	5.233808	.16	1.1	2.11	3.51	56.83
6	552	3.229511	3.990233	.17	1.03	2.03	3.23	29.61
Total	4930	3.601103	6.143637	.07	.95	1.9	3.38	107.97

```
. twoway (scatter crp cholest) (lowess crp cholest)
```

Regression for 2b.

```
. regress crp cholest, robust
```

Regression with robust standard errors

```
Number of obs = 4930
F( 1, 4928) = 15.18
Prob > F = 0.0001
R-squared = 0.0060
Root MSE = 6.1258
```

crp	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
cholest	-.0121279	.0031129	-3.90	0.000	-.0182306 -.0060253
_cons	6.16838	.7063241	8.73	0.000	4.78367 7.55309

Regression for 2c.

. g cholsqr= cholest^2

(47 missing values generated)

. regress crp cholest cholsqr, robust

Regression with robust standard errors

Number of obs = 4930
 F(2, 4927) = 9.52
 Prob > F = 0.0001
 R-squared = 0.0150
 Root MSE = 6.0986

crp	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
cholest	-.1151435	.0291478	-3.95	0.000	-.1722861	-.0580009
cholsqr	.0002354	.0000623	3.78	0.000	.0001133	.0003574
_cons	17.06689	3.34314	5.11	0.000	10.51284	23.62093

Finding the minimum: If $Y = Ax^2 + Bx + C$, then the minimum (or maximum) occurs at

$x = -B / (2*A)$

(this is a minimum if A is positive, a maximum if A is negative)

. display - (-.1151435/.0002354/2)

244.56988

Problem 3.

Descriptive statistics.

. tabstat cholest, by(cholctg) stat(n me sd min q max)

Summary for variables: cholest

by categories of: cholctg

cholctg	N	mean	sd	min	p25	p50	p75	max
0	406	143.5936	14.20832	73	136	148	154	159
1	591	170.6447	5.632139	160	166	171	176	179
2	963	190.2388	5.694315	180	186	190	195	199
3	1033	209.8025	5.612979	200	205	210	215	219
4	862	229.0893	5.730709	220	224	229	234	239
5	543	248.5378	5.664836	240	244	248	253	259
6	555	282.8649	23.22892	260	266	276	292	430
Total	4953	211.6893	39.28814	73	186	210	236	430

. stset ttodth, fail(death)

failure event: death ~= 0 & death ~= .

obs. time interval: (0, ttodth]

exit on or before: failure

5000 total obs.

0 exclusions

5000 obs. remaining, representing

1121 failures in single record/single failure data

1.18e+07 total analysis time at risk, at risk from t = 0

earliest observed entry t = 0

last observed exit t = 2942

. sts graph, by(cholctg)

. sts list, by(cholctg) at(365 730 1826 2922)

failure _d: death

analysis time _t: ttodth

Time	Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]
------	-------	------	-------------------	------------	------------------

```
-----
```

cholctg							
cholctg=0							
365	394	13	0.9680	0.0087	0.9455	0.9813	
730	384	10	0.9433	0.0115	0.9160	0.9620	
1826	278	62	0.7845	0.0208	0.7404	0.8221	
2922	12	49	0.6060	0.0305	0.5434	0.6628	
cholctg=1							
365	577	15	0.9746	0.0065	0.9583	0.9846	
730	562	15	0.9492	0.0090	0.9282	0.9642	
1826	441	66	0.8343	0.0155	0.8014	0.8623	
2922	15	68	0.6889	0.0214	0.6448	0.7287	
cholctg=2							
365	951	13	0.9865	0.0037	0.9769	0.9921	
730	929	22	0.9637	0.0060	0.9497	0.9738	
1826	729	110	0.8453	0.0118	0.8205	0.8670	
2922	28	82	0.7378	0.0156	0.7058	0.7669	
cholctg=3							
365	1014	20	0.9806	0.0043	0.9702	0.9875	
730	990	24	0.9574	0.0063	0.9432	0.9681	
1826	802	82	0.8753	0.0104	0.8532	0.8942	
2922	25	103	0.7277	0.0256	0.6737	0.7743	
cholctg=4							
365	852	11	0.9872	0.0038	0.9771	0.9929	
730	837	15	0.9698	0.0058	0.9560	0.9794	
1826	693	55	0.9043	0.0101	0.8824	0.9223	
2922	27	71	0.8046	0.0146	0.7741	0.8314	
cholctg=5							
365	535	9	0.9834	0.0055	0.9684	0.9913	
730	522	13	0.9595	0.0085	0.9391	0.9731	
1826	437	43	0.8773	0.0143	0.8462	0.9025	
2922	16	30	0.7988	0.0202	0.7557	0.8352	
cholctg=6							
365	550	6	0.9892	0.0044	0.9761	0.9951	
730	538	12	0.9676	0.0075	0.9490	0.9794	
1826	448	42	0.8897	0.0135	0.8602	0.9133	
2922	14	50	0.7745	0.0206	0.7309	0.8119	

. #### Regression for 3b.

. **stcox** **cholest**

failure _d: death

analysis time _t: ttodth

```
No. of subjects =      4953          Number of obs   =      4953
No. of failures =      1111
Time at risk    =     11757827
Log likelihood  =    -9173.1039          LR chi2(1)      =      47.48
                                          Prob > chi2    =      0.0000
```

```
-----
```

	_t					
	_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
cholest		.9945688	.0007941	-6.82	0.000	.9930136 .9961265

```
-----
```

. **display** **.9945688^50**

.76162564

. **display** **.9930136^50**

.70430346

. **display** **.9961256^50**

.82357863

Regression for 3c.

```
. g cenchol= cholest - 200
(47 missing values generated)

. g cenchol2= cenchol^2
(47 missing values generated)

. stcox cenchol cenchol2, robust
      failure _d:  death
      analysis time _t:  ttodth
No. of subjects =      4953                Number of obs   =      4953
No. of failures =      1111
Time at risk   =      11757827
Log likelihood =    -9168.2642
Wald chi2(2)   =      63.20
Prob > chi2    =      0.0000
```

Variable	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
cenchol	.9937509	.0007865	-7.92	0.000	.9922106	.9952937
cenchol2	1.000037	.0000112	3.31	0.001	1.000015	1.000059

```
#### The following code will produce the estimated hazard ratios.
. g fitHR = exp(log(.9937509) * cenchol + log(1.000037) * cenchol2)
```

```
#### The following code will produce the estimated hazard ratios with more
#### precision (less roundoff error).
```

```
. predict fHR
(option hr assumed; relative hazard)

. summ fHR
      Variable |      Obs      Mean   Std. Dev.   Min      Max
-----+-----
      fHR      |    4953    1.014085   .2536373   .7680746  4.041481
```

```
. list fHR cholest if fHR<.76808
      fHR      cholest
-----+-----
    868.   .7680746      284
   2410.   .7680746      284
   4028.   .7680746      284
```

```
. scatter fHR cholest
```

```
. #### Problem 4.
. #### Descriptive statistics.
. centile crp, cen(10 20 30 40 50 60 70 80 90)
```

Variable	Obs	Percentile	Centile	-- Binom. Interp. -- [95% Conf. Interval]	
crp	4933	10	0.52	0.50	0.54
		20	0.78	0.75	0.82
		30	1.13	1.09	1.17
		40	1.50	1.43	1.56
		50	1.90	1.84	1.96
		60	2.40	2.33	2.47
		70	2.97	2.89	3.05
		80	4.13	3.84	4.55
		90	8.21	7.64	8.82

```
. g crpctg= crp
(67 missing values generated)
```

```
. recode crpctg 0/.5= 0 .5/1=1 1.001/2=2 2.001/4=4 4.001/8=8 8.001/110=16
(4906 changes made)
```

```
. tabstat crp, by(crpctg) stat(n me sd min q max)
Summary for variables: crp
by categories of: crpctg
```

crpctg	N	mean	sd	min	p25	p50	p75	max
0	428	.3663551	.0896874	.07	.31	.38	.44	.49
1	885	.7288475	.1495545	.5	.6	.71	.86	1
2	1265	1.486711	.2857354	1.01	1.23	1.47	1.73	2
4	1355	2.793129	.5251494	2.01	2.35	2.73	3.18	4
8	486	5.822922	1.068466	4.01	4.92	5.8	6.72	7.99
16	514	16.4849	12.41304	8.01	9.79	12.205	17.39	107.97
Total	4933	3.602349	6.143344	.07	.95	1.9	3.38	107.97

```
. sts graph, by(crpctg)
. sts list, by(crpctg) at(365 730 1826 2922)
failure _d: death
analysis time _t: ttodth
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]	
crpctg=0						
365	428	1	0.9977	0.0023	0.9835	0.9997
730	422	6	0.9836	0.0061	0.9660	0.9922
1826	359	32	0.9064	0.0143	0.8740	0.9307
2922	14	41	0.7892	0.0219	0.7424	0.8285
crpctg=1						
365	878	8	0.9910	0.0032	0.9820	0.9955
730	862	16	0.9729	0.0055	0.9598	0.9817
1826	742	61	0.9024	0.0101	0.8807	0.9203
2922	25	76	0.7727	0.0278	0.7125	0.8220
crpctg=2						
365	1253	13	0.9897	0.0028	0.9824	0.9940
730	1241	12	0.9802	0.0039	0.9709	0.9866
1826	1015	109	0.8908	0.0089	0.8720	0.9071
2922	40	102	0.7913	0.0125	0.7655	0.8147
crpctg=4						
365	1328	28	0.9793	0.0039	0.9702	0.9857
730	1293	35	0.9535	0.0057	0.9409	0.9635
1826	1042	130	0.8550	0.0097	0.8348	0.8728
2922	40	132	0.7315	0.0135	0.7040	0.7570
crpctg=8						
365	466	21	0.9568	0.0092	0.9345	0.9716
730	448	18	0.9198	0.0123	0.8918	0.9407
1826	326	61	0.7901	0.0187	0.7506	0.8241
2922	5	46	0.6436	0.0285	0.5848	0.6962
crpctg=16						
365	499	16	0.9689	0.0077	0.9497	0.9808
730	476	23	0.9241	0.0117	0.8976	0.9440
1826	330	66	0.7869	0.0186	0.7477	0.8207
2922	13	56	0.6196	0.0286	0.5608	0.6729

Note: Survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

```
. summ fib
```

Variable	Obs	Mean	Std. Dev.	Min	Max
----------	-----	------	-----------	-----	-----

```
-----+-----
      fib |      4915      322.978      67.28736      109      872
```

```
. g fibctg= fib
```

```
(85 missing values generated)
```

```
. recode fibctg 0/250=250 250/275=275 275/300=300 300/325=325 325/350=350 350/375=375
375/400=400 400/1000=450
```

```
(4915 changes made)
```

```
. tabstat fib, by(fibctg) stat(n me sd min q max)
```

```
Summary for variables: fib
```

```
by categories of: fibctg
```

fibctg	N	mean	sd	min	p25	p50	p75	max
250	479	223.8747	20.35781	109	217	230	238	249
275	692	262.1329	7.564271	250	253	266	270	274
300	673	286.2169	4.69459	275	281	285	289	299
325	797	308.6913	4.830035	300	307	307	311	324
350	776	333.8338	5.449067	325	331	334	337	349
375	631	361.8384	5.125645	350	358	361	367	374
400	298	390.4564	5.757131	375	387	393	395	398
800	569	450.6555	58.03154	400	407	436	464	872
Total	4915	322.978	67.28736	109	281	311	361	872

```
. sts graph, by(fibctg)
```

```
. sts list, by(fibctg) at(365 730 1826 2922)
```

```
failure _d: death
analysis time _t: ttodth
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]	
fibctg=250						
365	476	4	0.9916	0.0042	0.9779	0.9969
730	469	7	0.9770	0.0068	0.9589	0.9872
1826	400	36	0.8996	0.0139	0.8686	0.9236
2922	11	47	0.7611	0.0261	0.7051	0.8078
fibctg=275						
365	687	6	0.9913	0.0035	0.9808	0.9961
730	679	8	0.9798	0.0054	0.9661	0.9880
1826	597	48	0.9092	0.0110	0.8851	0.9285
2922	11	54	0.8000	0.0208	0.7555	0.8372
fibctg=300						
365	670	4	0.9941	0.0030	0.9842	0.9978
730	657	13	0.9747	0.0060	0.9597	0.9842
1826	540	59	0.8846	0.0125	0.8576	0.9068
2922	15	55	0.7483	0.0401	0.6594	0.8172
fibctg=325						
365	789	9	0.9887	0.0037	0.9784	0.9941
730	776	13	0.9724	0.0058	0.9584	0.9817
1826	635	66	0.8871	0.0114	0.8627	0.9074
2922	17	71	0.7746	0.0165	0.7403	0.8051
fibctg=350						
365	764	13	0.9832	0.0046	0.9713	0.9902
730	745	19	0.9588	0.0071	0.9422	0.9707
1826	605	76	0.8584	0.0126	0.8315	0.8813
2922	27	68	0.7540	0.0165	0.7199	0.7846
fibctg=375						
365	622	10	0.9842	0.0050	0.9707	0.9914
730	607	15	0.9604	0.0078	0.9419	0.9731

1826	476	68	0.8481	0.0146	0.8170	0.8743
2922	22	60	0.7242	0.0204	0.6818	0.7620
fibctg=400						
365	288	11	0.9631	0.0109	0.9343	0.9794
730	281	7	0.9396	0.0138	0.9058	0.9615
1826	203	31	0.8291	0.0224	0.7800	0.8682
2922	16	28	0.7074	0.0288	0.6467	0.7596
fibctg=450						
365	543	27	0.9525	0.0089	0.9316	0.9672
730	516	27	0.9051	0.0123	0.8779	0.9265
1826	352	72	0.7708	0.0181	0.7330	0.8039
2922	20	66	0.5978	0.0265	0.5439	0.6475

 Note: Survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

. #### Regression for 4b.

. **stcox crp age gender, robust**

```

failure _d: death
analysis time _t: ttodth
No. of subjects =          4933           Number of obs   =          4933
No. of failures =          1109
Time at risk    =       11713276
Log likelihood  =      -8888.3849           Wald chi2(3)     =        606.77
                                           Prob > chi2      =         0.0000
    
```

_t _d	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
crp	1.020546	.0038909	5.33	0.000	1.012948	1.0282
age	1.107258	.0052212	21.61	0.000	1.097072	1.117539
gender	1.761739	.1067056	9.35	0.000	1.564537	1.983798

. #### Regression for 4c.

. **stcox fib age gender, robust**

```

failure _d: death
analysis time _t: ttodth
No. of subjects =          4915           Number of obs   =          4915
No. of failures =          1098
Time at risk    =       11682465
Log likelihood  =      -8782.6313           Wald chi2(3)     =        676.08
                                           Prob > chi2      =         0.0000
    
```

_t _d	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
fib	1.0033	.000415	7.96	0.000	1.002487	1.004114
age	1.105707	.0051649	21.51	0.000	1.095631	1.115877
gender	1.806946	.1103815	9.69	0.000	1.603052	2.036774

. **g fib100= fib/100**

(85 missing values generated)

. **stcox fib100 gender age, robust**

```

failure _d: death
analysis time _t: ttodth
No. of subjects =          4915           Number of obs   =          4915
No. of failures =          1098
Time at risk    =       11682465
Log likelihood  =      -8782.6313           Wald chi2(3)     =        676.08
    
```

Log likelihood = -8782.6313 Prob > chi2 = 0.0000

_t _d	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
fib100	1.390201	.0575017	7.96	0.000	1.281947	1.507597
gender	1.806946	.1103815	9.69	0.000	1.603052	2.036774
age	1.105707	.0051649	21.51	0.000	1.095631	1.115877

. #### Regression for 4d.

. **stcox crp fib age gender, robust**

 failure _d: death

 analysis time _t: ttodth

No. of subjects = 4899 Number of obs = 4899

No. of failures = 1096

Time at risk = 11644078

Wald chi2(4) = 661.35

Log likelihood = -8759.4104 Prob > chi2 = 0.0000

_t _d	Haz. Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
crp	1.009063	.0044856	2.03	0.042	1.00031	1.017893
fib	1.002699	.0004932	5.48	0.000	1.001733	1.003666
age	1.106065	.0052128	21.39	0.000	1.095895	1.116329
gender	1.789276	.1096797	9.49	0.000	1.586719	2.017691

. **pwcorr crp fib**

	crp	fib
crp	1.0000	
fib	0.4809	1.0000