

Biost 517: Applied Biostatistics I
Emerson, Fall 2011

```
#### Annotated Stata Log File: Homework #3
#### October 25, 2011
```

```
#### In this file I give the Stata commands I used to produce
#### the key to Homework #3. In order to properly format
#### a table useful to casual readers, I cut and pasted some
#### of the output into Excel.
```

```
#### 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.

```
#### Begin logging commands and output
```

```

name: <unnamed>
log: \\.\psf\Home\Documents\teach\courses\b517\f11\hw3Stata.log
log type: text
opened on: 25 Oct 2011, 06:13:07

```

```
#### Read in data: The infile command was typed all on one line
```

```
> quietly: infile ptid str10 dbirth age str3 sex str10 race height weight bsa pctfat tri chol hdl ldl tpro
> alb str10 dstart str10 drand arm nvisit str10 dvisit tvisit bcarot ret retpalm vite dose str10 dlast
> tlast using teach/datasets/bcarot.txt
```

```
#### Create an indicator variable for sex
```

```
. generate female= .  
(700 missing values generated)  
. replace female= 1 if sex=="F"  
(355 real changes made)  
. replace female= 0 if sex=="M"  
(344 real changes made)
```

```
#### Verify that the first case should be dropped and drop it
. list in 1
```

[illegible]

tabulation:	Freq.	Value
	133	0
	64	15
	62	30
	50	45
	63	60

```
bcarot
(unlabeled)
```

```

      type:  numeric (float)

      range:  [34,1944]          units:  1
unique values: 270              missing .: 2/372

      mean:    324.57
      std. dev: 271.814

percentiles:      10%      25%      50%      75%      90%
                  100      156      251.5    386      646
```

```
ret
(unlabeled)
```

```

      type:  numeric (float)

      range:  [295,1042]        units:  1
unique values: 249              missing .: 2/372

      mean:    571.189
      std. dev: 131.99

percentiles:      10%      25%      50%      75%      90%
                  419      469      552      656      744.5
```

```
retpalm
(unlabeled)
```

```

      type:  numeric (float)

      range:  [0,308]           units:  1
unique values: 80              missing .: 2/372

      mean:    26.5243
      std. dev: 30.1081

percentiles:      10%      25%      50%      75%      90%
                  10       11       17       30       63
```

```
vite
(unlabeled)
```

```

      type:  numeric (float)

      range:  [0,12.82]          units:  .01
unique values: 274              missing .: 2/372

      mean:    7.35489
      std. dev: 1.84027

      percentiles:      10%      25%      50%      75%      90%
                        4.965      6.16      7.52      8.58      9.44
```

```
. list female arm nvisit bcarot ret retpalm vite if bcarot==.
```

	female	arm	nvisit	bcarot	ret	retpalm	vite
88.	0	45	0	.	.	.	.
109.	1	15	0	.	.	.	.

```
#### Problem 1
```

```
#### First find total number of cases by sex and arm
```

```
. tabulate female arm
```

female	arm					Total
	0	15	30	45	60	
0	82	35	21	25	31	194
1	51	29	41	25	32	178
Total	133	64	62	50	63	372

```
. tabstat bcarot ret retpalm vite, stat(n mean sd min q max) col(stat) by(female)
```

```
Summary for variables: bcarot ret retpalm vite
by categories of: female
```

female	N	mean	sd	min	p25	p50	p75	max
0	193	274.1503	248.2547	34	131	218	332	1944
	193	607	120.8177	340	518	608	695	975
	193	27.71503	36.23429	0	10	17	29	308
	193	7.202176	1.900485	0	6.06	7.36	8.35	12.82
1	177	379.548	286.0801	44	191	296	426	1680

	177	532.1412	132.8987	295	444	499	592	1042
	177	25.22599	21.56451	0	12	17	31	100
	177	7.521412	1.762578	3.09	6.37	7.87	8.88	12.27
Total	370	324.5703	271.8143	34	156	251.5	386	1944
	370	571.1892	131.9903	295	469	552	656	1042
	370	26.52432	30.10813	0	11	17	30	308
	370	7.354892	1.840274	0	6.16	7.52	8.58	12.82

Problem 2b

First find total number of cases by sex and arm

. tabulate female arm if nvisit<=3

female	arm						Total
	0	15	30	45	60		
0	21	20	12	16	19		88
1	16	20	24	16	20		96
Total	37	40	36	32	39		184

. tabstat bcarot ret retpalm vite if nvisit<=3, stat(n mean sd min q max) col(stat) by(female)

Summary for variables: bcarot ret retpalm vite
by categories of: female

female	N	mean	sd	min	p25	p50	p75	max
0	87	197	124.8194	34	105	169	254	600
	87	617.9195	124.7759	371	519	608	703	975
	87	32	47.29767	0	11	18	28	308
	87	7.793678	1.654061	0	6.87	8	8.66	12.02
1	95	256.5579	115.661	44	164	244	337	715
	95	530.7684	132.9246	295	444	488	583	990
	95	23.32632	19.08182	0	12	17	24	94
	95	8.155158	1.401497	5	7.38	8.19	9.15	12.27
Total	182	228.0879	123.4503	34	142	207.5	309	715
	182	572.4286	135.9382	295	469	552	669	990
	182	27.47253	35.64957	0	12	17	26	308
	182	7.982363	1.533884	0	7.13	8.135	9	12.27

Problem 2c

First find total number of cases by sex and arm

. tabulate female arm if dvisit==dstart

Problem 4a

. egen mncarot= mean(bcarot), by(ptid)

. tabstat mncarot, stat(n mean sd min q max) col(stat) by(female)

Summary for variables: mncarot
by categories of: female

female	N	mean	sd	min	p25	p50	p75	max
0	194	274.4802	172.2283	58.75	155.7222	235.5714	357.0625	933.1429
1	178	380.2669	152.827	154.7143	249	360.7143	494.4286	800.4286
Total	372	325.0986	171.3883	58.75	188.4286	305	421.6667	933.1429

. tabstat mncarot, stat(n mean sd min q max) col(stat) by(arm)

Summary for variables: mncarot
by categories of: arm

arm	N	mean	sd	min	p25	p50	p75	max
0	133	257.5639	104.7243	129.1176	155.7222	204	357.0625	405.4167
15	64	287.4766	136.3562	95.85714	188.4286	242.5714	402	513.1429
30	62	356.629	156.766	179	198.8333	305	494.4286	616.5714
45	50	422.0233	213.7885	93.25	228.25	498.7143	516.1429	800.4286
60	63	397.9365	219.1345	58.75	256.4286	360.7143	441.1429	933.1429
Total	372	325.0986	171.3883	58.75	188.4286	305	421.6667	933.1429

Problem 4c

. egen visits= count(bcarot), by(ptid)

. tabstat visits, stat(n mean sd min q max) col(stat) by(female)

Summary for variables: visits
by categories of: female

female	N	mean	sd	min	p25	p50	p75	max
0	194	10.80928	4.934778	4	7	7	16	18
1	178	8.657303	4.000067	4	7	7	7	18
Total	372	9.77957	4.632571	4	7	7	15	18

. tabstat visits, stat(n mean sd min q max) col(stat) by(arm)

Summary for variables: visits
by categories of: arm

arm	N	mean	sd	min	p25	p50	p75	max
0	133	15.51128	2.695722	7	15	16	18	18
15	64	6.515625	1.007782	4	7	7	7	7
30	62	6.903226	.2980636	6	7	7	7	7
45	50	6.38	1.104536	4	6	7	7	7
60	63	6.52381	1.013731	4	7	7	7	7
Total	372	9.77957	4.632571	4	7	7	15	18

Problem 4d

. codebook ptid if dvisit==dstart

ptid
(unlabeled)

type: numeric (float)

range: [210,9542]

units: 1

unique values: 46

missing .: 0/46

mean: 4771.43

std. dev: 2958.36

percentiles:	10%	25%	50%	75%	90%
	701	2548	4556	7927	8630

Problem 4e

. tabstat mncarot if dvisit==dstart, stat(n mean sd min q max) col(stat) by(female)

Summary for variables: mncarot
by categories of: female

female	N	mean	sd	min	p25	p50	p75	max
0	22	279.6992	192.1459	58.75	168.5714	238.2143	357.0625	933.1429
1	24	382.2937	158.8576	154.7143	244.5	381.3571	496.5714	800.4286
Total	46	333.2268	181.142	58.75	192.4286	288.5	460.1429	933.1429

. tabstat mncarot if dvisit==dstart, stat(n mean sd min q max) col(stat) by(arm)

Summary for variables: mncarot
by categories of: arm

arm	N	mean	sd	min	p25	p50	p75	max
0	9	256.0095	111.9356	129.1176	155.7222	204	357.0625	405.4167
15	10	287.6393	140.7173	95.85714	188.4286	242.5714	402	513.1429
30	9	354.1243	164.9467	179	198.8333	305	494.4286	616.5714
45	8	394.0298	231.7901	93.25	198.4107	418.4405	512.4286	800.4286
60	10	380.8595	229.4254	58.75	256.4286	345.7321	441.1429	933.1429
Total	46	333.2268	181.142	58.75	192.4286	288.5	460.1429	933.1429

Problem 5a

```
. egen grbg= count(bcarot) if nvisit<=3, by(ptid)
(188 missing values generated)
```

```
. egen nrunin= mean(grbg), by(ptid)
```

```
. replace nrunin=0 if nrunin==.
(0 real changes made)
```

```
. table nrunin
```

nrunin	Freq.
3	20
4	335
5	17

```
. codebook ptid if nrunin<4
```

```
ptid
(unlabeled)
```

```
type: numeric (float)
```

```
range: [3212,8236]
unique values: 3
```

```
units: 1
missing .: 0/20
```

```
tabulation: Freq. Value
              7 3212
              7 8115
              6 8236
```

```
. codebook ptid if nrunin >4
```

```
ptid
(unlabeled)
```

```
type: numeric (float)
```

```

        range:  [8630,8630]           units:  1
unique values:  1                     missing.: 0/17

```

```

tabulation:  Freq.  Value
              17   8630

```

```
. drop grbg
```

```
#### Problem 5b
```

```
. egen grbg= mean(bcarot) if nvisit<= 3, by(ptid)
(188 missing values generated)
```

```
. egen mncarotRI= mean(grbg), by(ptid)
```

```
. tabstat mncarotRI if dvisit==dstart, stat(n mean sd min q max) col(stat) by(female)
```

```
Summary for variables: mncarotRI
by categories of: female
```

female	N	mean	sd	min	p25	p50	p75	max
0	22	197.9121	117.3576	48.25	134.75	174.75	241.3333	476.25
1	24	259.0521	103.1455	98.25	177.125	232.625	346.75	496
Total	46	229.8112	113.2189	48.25	140	212	310.75	496

```
. tabstat mncarotRI if dvisit==dstart, stat(n mean sd min q max) col(stat) by(arm)
```

```
Summary for variables: mncarotRI
by categories of: arm
```

arm	N	mean	sd	min	p25	p50	p75	max
0	9	259.9333	131.1809	136.25	149.4	222.75	360	476.25
15	10	222.2	131.0684	64.75	136	185.625	237.75	496
30	9	224	87.55605	125.5	140	232.5	282.25	348.5
45	8	226.9792	105.538	93.25	147.9167	216.375	299.125	395.75
60	10	217.8083	122.3405	48.25	98.25	224.2917	310.75	407.5
Total	46	229.8112	113.2189	48.25	140	212	310.75	496

```
#### Problem 6
```

```
. egen mxcarotRI = max(bcarot) if nvisit <= 3, by(ptid)
(188 missing values generated)
```

```
. tabstat mxcarotRI if dvisit==dstart, stat(n mean sd min q max) col(stat) by(female)
```

Summary for variables: mxcarotRI
by categories of: female

female	N	mean	sd	min	p25	p50	p75	max
0	22	252.0455	135.0473	65	169	240	273	600
1	24	322.3333	143.8537	117	219	300.5	400.5	715
Total	46	288.7174	142.6464	65	190	257.5	385	715

```
. tabstat mxcarotRI if dvisit==dstart, stat(n mean sd min q max) col(stat) by(arm)
```

Summary for variables: mxcarotRI
by categories of: arm

arm	N	mean	sd	min	p25	p50	p75	max
0	9	305	155.5169	160	190	258	385	600
15	10	299.5	202.5527	87	169	240	279	715
30	9	277.6667	99.70707	154	199	255	326	428
45	8	289.875	106.8035	123	231	263	385.5	437
60	10	272.3	143.8333	65	117	290.5	388	488
Total	46	288.7174	142.6464	65	190	257.5	385	715

Problem 7

Verifying that all patients are off supplementation following the 12 month visit

```
. table dlast if nvisit > 12
```

dlast	Freq.
NA	111

Creating the variable that measures mean beta carotene level after supplementation stops

I have to drop grbg before I can use it again

Note also that it is crucial to go through the two step egen process here, because the

first call will have missing data when dstart==dvisit (those are run-in visits)

```
. drop grbg
```

```
. egen grbg = mean(bcarot) if nvisit > 12, by(ptid)
(261 missing values generated)
```

```
. egen mncarotFU = mean(grbg), by(ptid)
(31 missing values generated)
```

```
#### Now doing descriptive statistics using each patient at most once
```

```
. tabstat mncarotFU if dvisit==dstart, stat(n mean sd min q max) col(stat) by(female)
```

```
Summary for variables: mncarotFU
by categories of: female
```

female	N	mean	sd	min	p25	p50	p75	max
0	20	415.8917	357.0676	109.6667	238	290	527.5	1634
1	19	607.0702	297.6037	88.5	401.6667	565.3333	838	1340
Total	39	509.0299	339.2248	88.5	258.3333	450	695.6667	1634

```
. tabstat mncarotFU if dvisit==dstart, stat(n mean sd min q max) col(stat) by(arm)
```

```
Summary for variables: mncarotFU
by categories of: arm
```

arm	N	mean	sd	min	p25	p50	p75	max
0	8	224.2083	138.6056	88.5	118.8333	184.8333	306.0833	485.6667
15	8	395.7917	240.4956	137.3333	251.6667	314.6667	494	908.3333
30	9	533.0556	277.1484	231	332.3333	401.6667	792	974
45	6	760.2778	377.9054	246	491	785.1667	914.3333	1340
60	8	691.625	396.8891	363.3333	507	583.6667	677.1667	1634
Total	39	509.0299	339.2248	88.5	258.3333	450	695.6667	1634

```
. log close
```

```
name: <unnamed>
```

```
log: \\.\psf\Home\Documents\teach\courses\b517\f11\hw3Stata.log
```

```
log type: text
```

```
closed on: 25 Oct 2011, 07:21:04
```