

STATA 001

**September 5,
2001**

```
log: C:\Stata\MyFiles\BiometryCourse\problems\kay.log
log type: text
opened on: 30 Aug 2001, 16:20:56
```

```
. rename dur_stay stay
. rename bact_cul bactcul
. label variable stay "Duration of Stay in Days"
. label variable age "Age in years"
. label variable sex "Gender"
. label variable temp "First Temperature Following Admission in degrees F"
. label variable wbc "First White Blood Count (x10^3) Following Admission"
. label variable antibio "Received Antibiotic - yes/no"
. label variable bactcul "Received Bacterial Culture - yes/no"
. label variable service "Hospital Service"
. label define sexmf 1 "Male" 2 "Female"
. label values sex sexmf
. label define yesno 1 "Yes" 2 "No"
. label values antibio yesno
. label values bactcul yesno
. label define serv 1 "Medical" 2 "Surgical"
. label values service serv
. sav,replace
file C:\Stata\MyFiles\BiometryCourse\problems\hospital.dta saved
```

. des

Contains data from C:\Stata\MyFiles\BiometryCourse\problems\hospital.dta

```

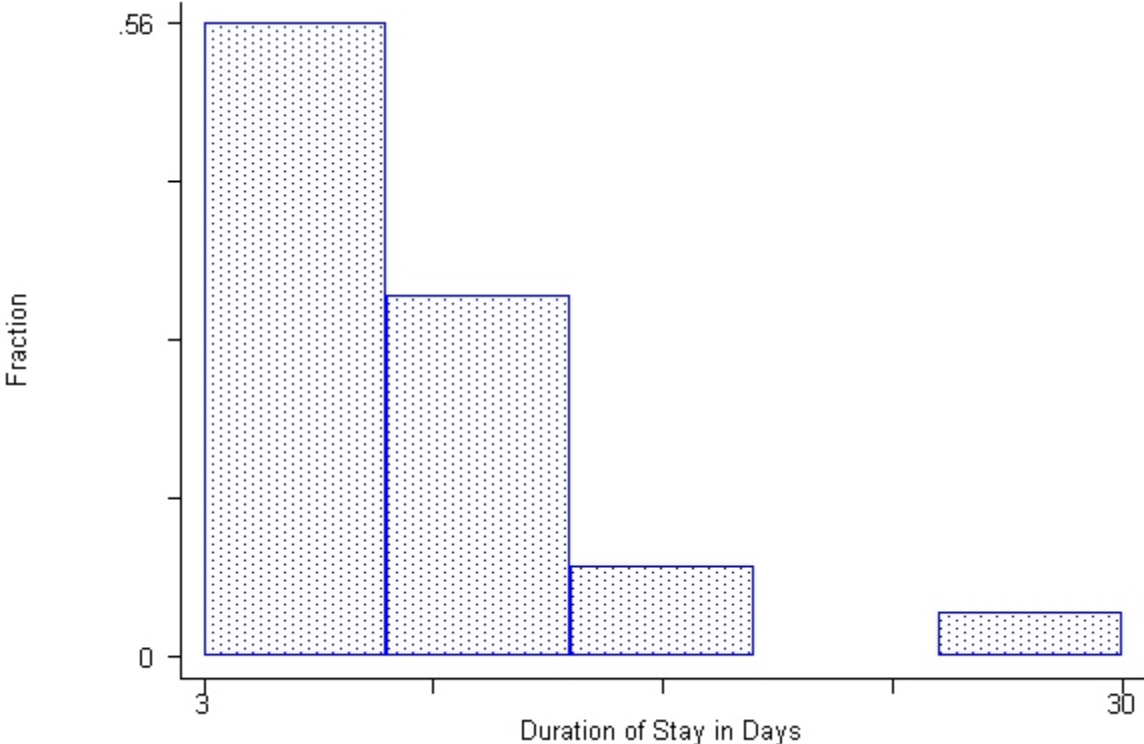
obs:      25
vars:     10
size:     1,100 (100.0% of memory free)
30 Aug 2001 17:28

```

variable name	storage type	display format	value label	variable label
id	float	%12.0g		
stay	float	%4.2f		Duration of Stay in Days
age	float	%12.0g		Age in years
sex	float	%12.0g	sexmf	Gender
temp	float	%12.0g		First Temperature Following Admission in degrees F
wbc	float	%12.0g		First White Blood Count (x10 ³) Following Admission
antibio	float	%12.0g	yesno	Received Antibiotic - yes/no
bactcul	float	%12.0g	yesno	Received Bacterial Culture - yes/no
service	float	%12.0g	serv	Hospital Service
catage	float	%9.0g		Age in Quartiles from lowest to highest

Sorted by:

```
. graph stay
```



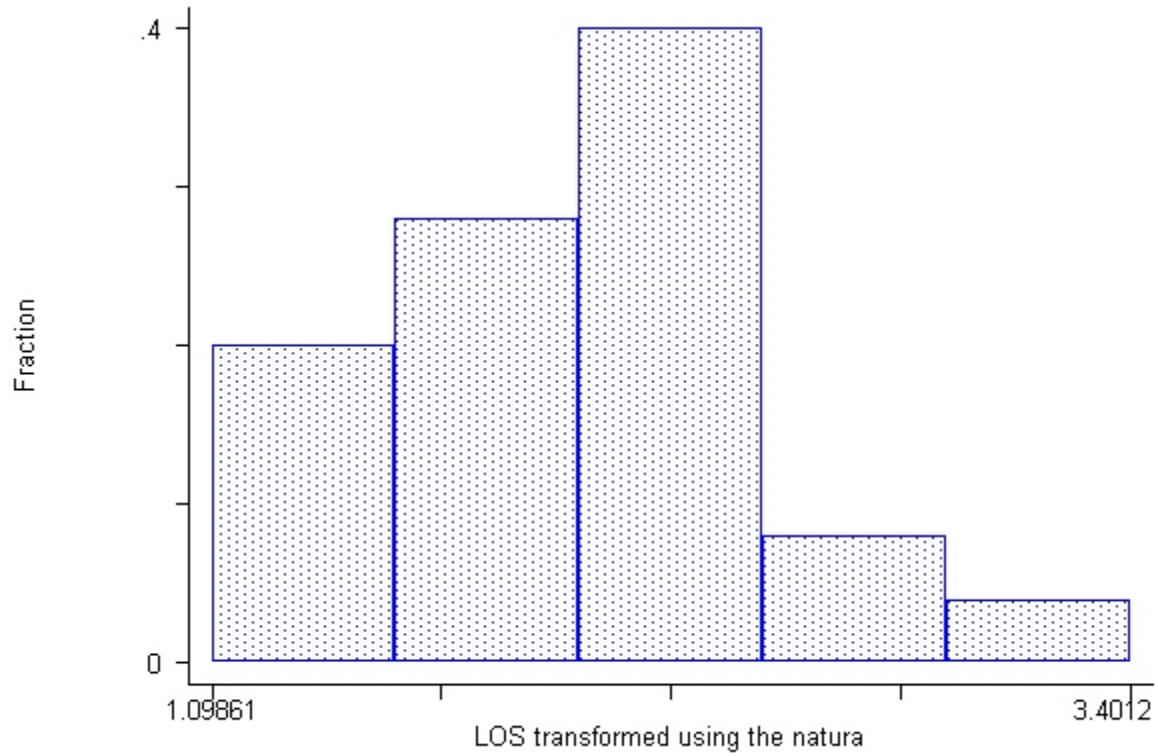
```
. sum stay,det
```

Duration of Stay in Days				
Percentiles		Smallest		
1%	3	3		
5%	3	3		
10%	3	3	Obs	25
25%	5	4	Sum of Wgt.	25
50%	8		Mean	8.6
		Largest	Std. Dev.	5.715476
75%	11	11		
90%	14	14	Variance	32.66667
95%	17	17	Skewness	2.203535
99%	30	30	Kurtosis	8.959067

```
. tab stay
```

Duration of Stay in Days	Freq.	Percent	Cum.
3	3	12.00	12.00
4	2	8.00	20.00
5	4	16.00	36.00
6	1	4.00	40.00
7	2	8.00	48.00
8	2	8.00	56.00
9	3	12.00	68.00
10	1	4.00	72.00
11	4	16.00	88.00
14	1	4.00	92.00
17	1	4.00	96.00
30	1	4.00	100.00
Total	25	100.00	

```
. gen logstay = log(stay)
. label variable logstay "LOS transformed using the natural log"
. graph logstay
```

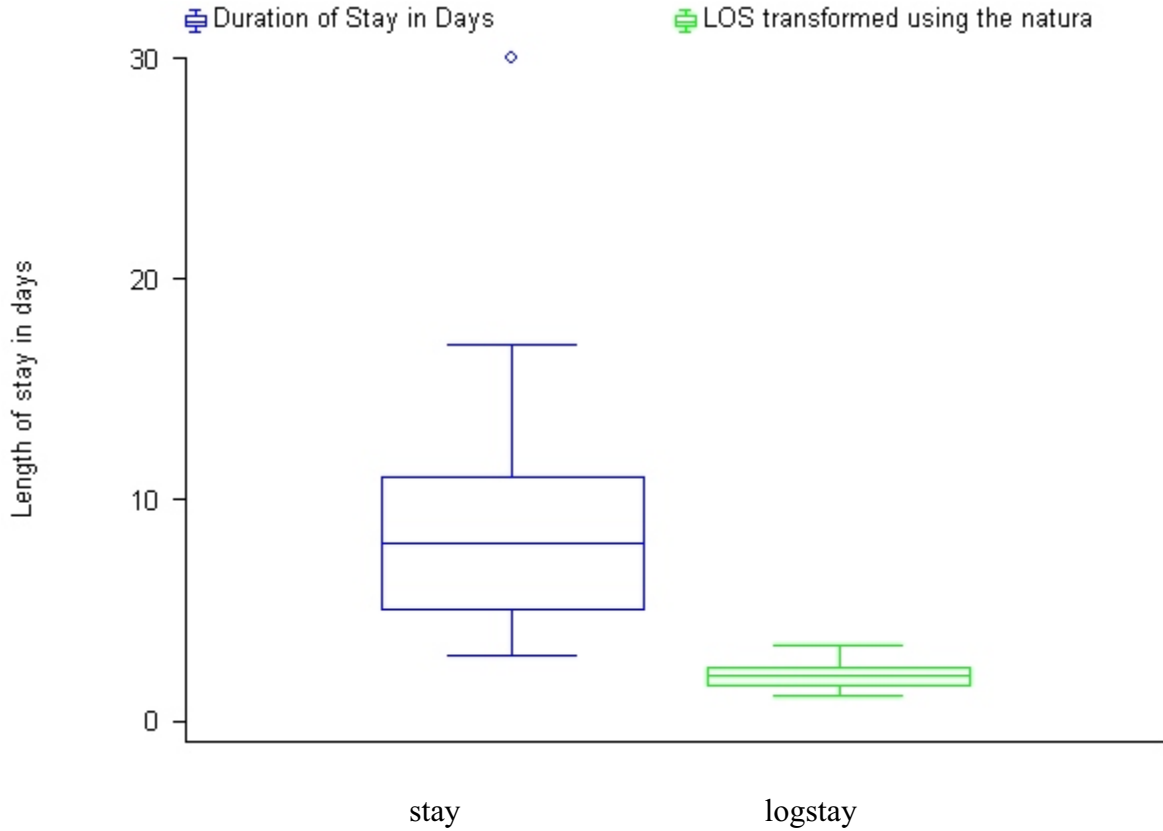


. sum

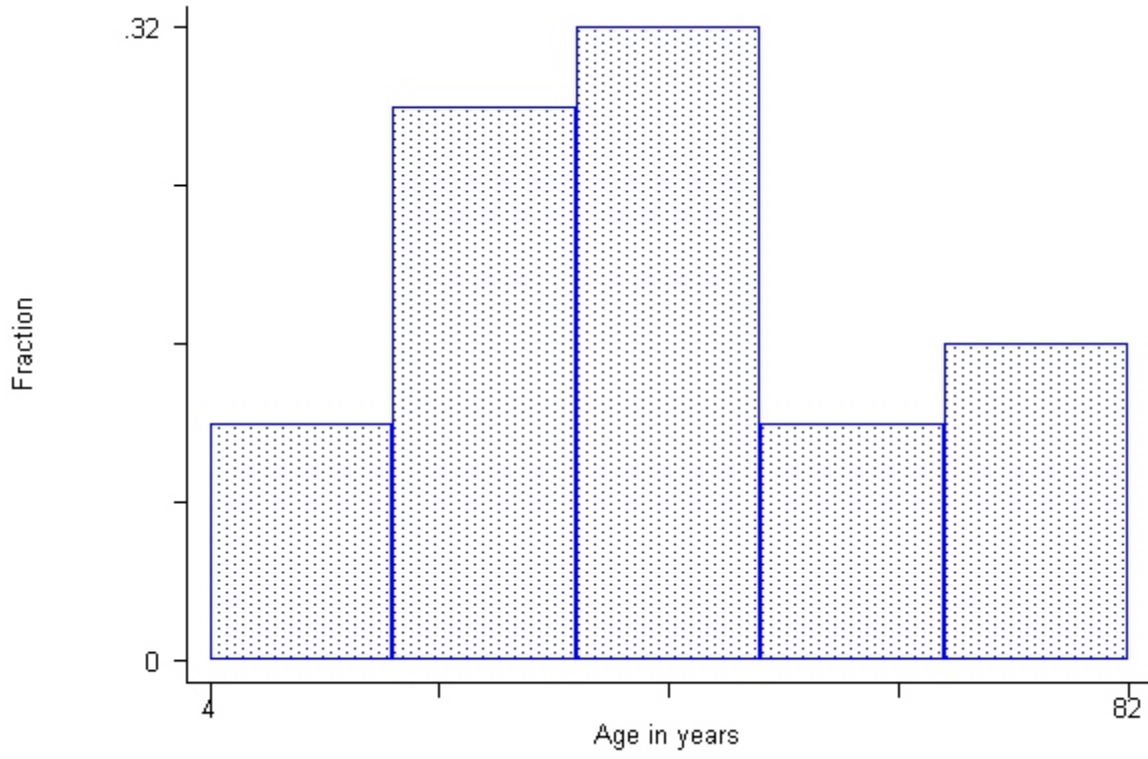
logstay, det

LOS transformed using the natural log

Percentiles		Smallest		
1%	1.098612	1.098612		
5%	1.098612	1.098612		
10%	1.098612	1.098612	Obs	25
25%	1.609438	1.386294	Sum of Wgt.	25
50%	2.079442		Mean	1.988318
			Std. Dev.	.5688759
75%	2.397895	2.397895		
90%	2.639057	2.639057	Variance	.3236198
95%	2.833213	2.833213	Skewness	.2909048
99%	3.401197	3.401197	Kurtosis	2.912761



. graph age



. sum age, det

Age in years					
Percentiles		Smallest			
1%	4	4			
5%	11	11			
10%	19	19	Obs	25	
25%	25	20	Sum of Wgt.	25	
50%	41		Mean	41.24	
		Largest	Std. Dev.	20.1024	
75%	56	67			
90%	69	69	Variance	404.1067	
95%	73	73	Skewness	.17486	
99%	82	82	Kurtosis	2.304504	

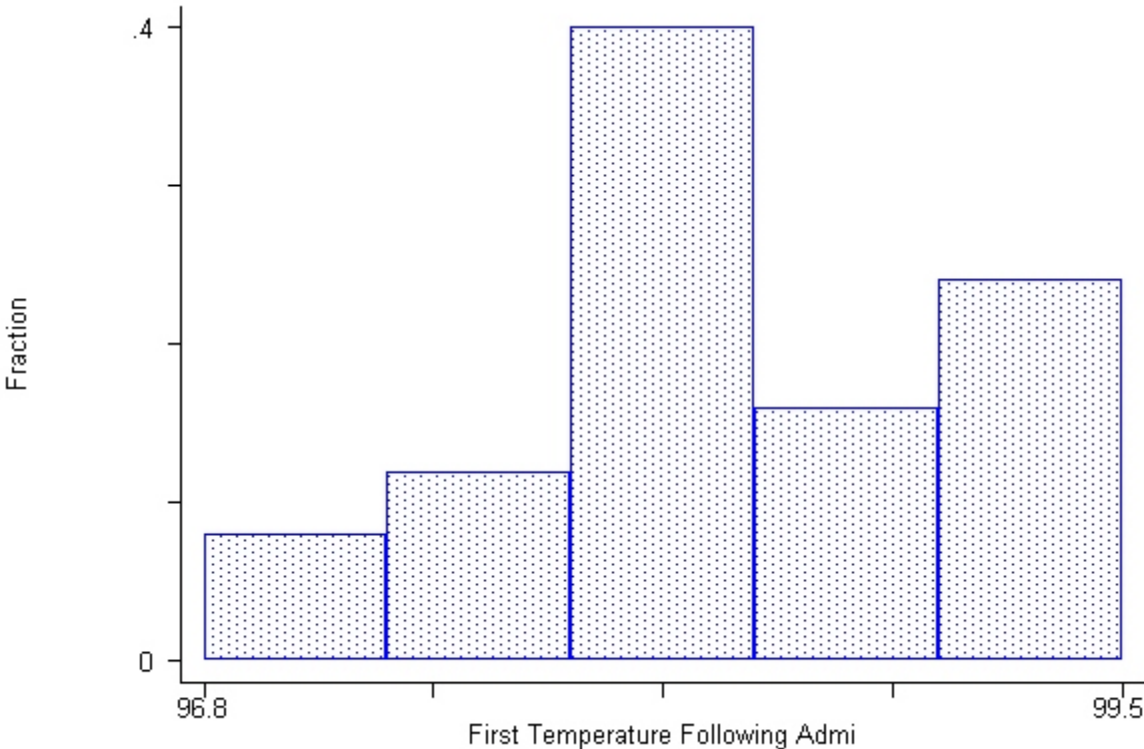
. tab age

Age in	Freq.	Percent	Cum.
years			
4	1	4.00	4.00
11	1	4.00	8.00
19	1	4.00	12.00
20	1	4.00	16.00
22	2	8.00	24.00
25	1	4.00	28.00
30	1	4.00	32.00
32	1	4.00	36.00
33	1	4.00	40.00
36	1	4.00	44.00
40	1	4.00	48.00
41	1	4.00	52.00
43	2	8.00	60.00
47	2	8.00	68.00
50	1	4.00	72.00
56	1	4.00	76.00
59	1	4.00	80.00
60	1	4.00	84.00
67	1	4.00	88.00
69	1	4.00	92.00
73	1	4.00	96.00
82	1	4.00	100.00
Total	25	100.00	

```
. tab sex
```

Gender	Freq.	Percent	Cum.
Male	11	44.00	44.00
Female	14	56.00	100.00
Total	25	100.00	

```
. graph temp
```



. sum temp,det

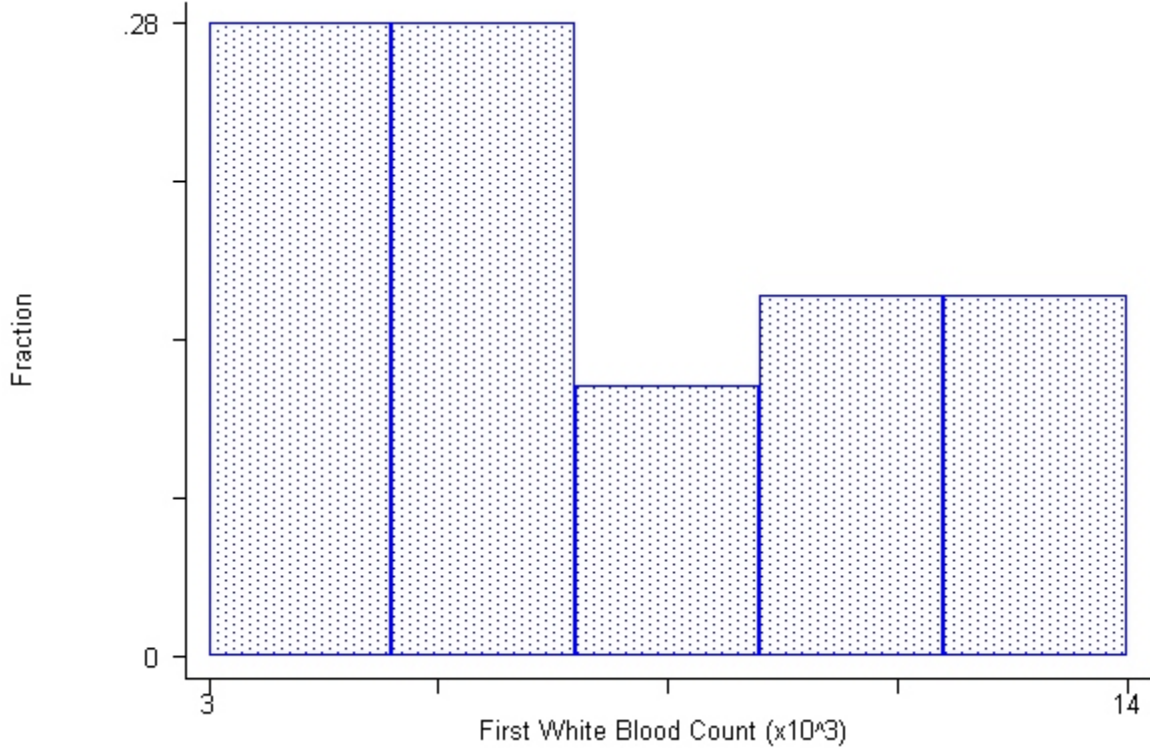
First Temperature Following Admission in degrees F

Percentiles		Smallest		
1%	96.8	96.8		
5%	97	97		
10%	97.6	97.6	Obs	25
25%	98	97.6	Sum of Wgt.	25
50%	98.2		Mean	98.308
		Largest	Std. Dev.	.6812484
75%	98.6	99		
90%	99.2	99.2	Variance	.4640994
95%	99.5	99.5	Skewness	-.1995118
99%	99.5	99.5	Kurtosis	2.840264

. tab temp

First Temperature Following Admission in degrees F	Freq.	Percent	Cum.
96.8	1	4.00	4.00
97	1	4.00	8.00
97.6	2	8.00	16.00
97.8	1	4.00	20.00
98	5	20.00	40.00
98.2	3	12.00	52.00
98.4	2	8.00	60.00
98.5	1	4.00	64.00
98.6	3	12.00	76.00
99	3	12.00	88.00
99.2	1	4.00	92.00
99.5	2	8.00	100.00
Total	25	100.00	

. graph wbc



. sum wbc,det

First White Blood Count (x10³) Following Admission

Percentiles		Smallest		
1%	3	3		
5%	4	4		
10%	4	4	Obs	25
25%	5	5	Sum of Wgt.	25
50%	7		Mean	7.84
		Largest	Std. Dev.	3.2104
75%	11	12		
90%	12	12	Variance	10.30667
95%	14	14	Skewness	.4791734
99%	14	14	Kurtosis	2.080755

. tab wbc

First White Blood Count (x10 ³) Following Admission	Freq.	Percent	Cum.
3	1	4.00	4.00
4	2	8.00	12.00
5	4	16.00	28.00
6	4	16.00	44.00
7	3	12.00	56.00
8	2	8.00	64.00
9	1	4.00	68.00
10	1	4.00	72.00
11	3	12.00	84.00
12	2	8.00	92.00
14	2	8.00	100.00
Total	25	100.00	

```
. tab antibio
```

Received Antibiotic - yes/no	Freq.	Percent	Cum.
Yes	7	28.00	28.00
No	18	72.00	100.00
Total	25	100.00	

```
. tab bactcul
```

Received Bacterial Culture - yes/no	Freq.	Percent	Cum.
Yes	6	24.00	24.00
No	19	76.00	100.00
Total	25	100.00	

```
. tab service
```

Hospital Service	Freq.	Percent	Cum.
Medical	9	36.00	36.00
Surgical	16	64.00	100.00
Total	25	100.00	

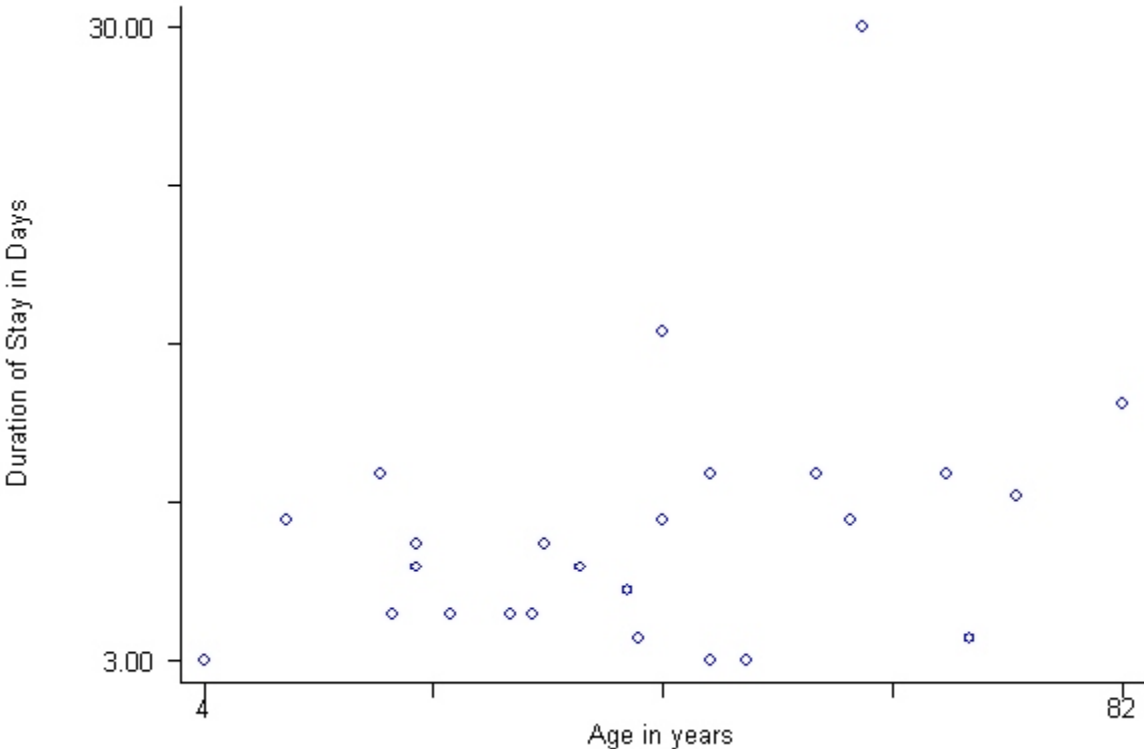
```
. log close
```

```
log: C:\Stata\MyFiles\BiometryCourse\problems\kay.log
```

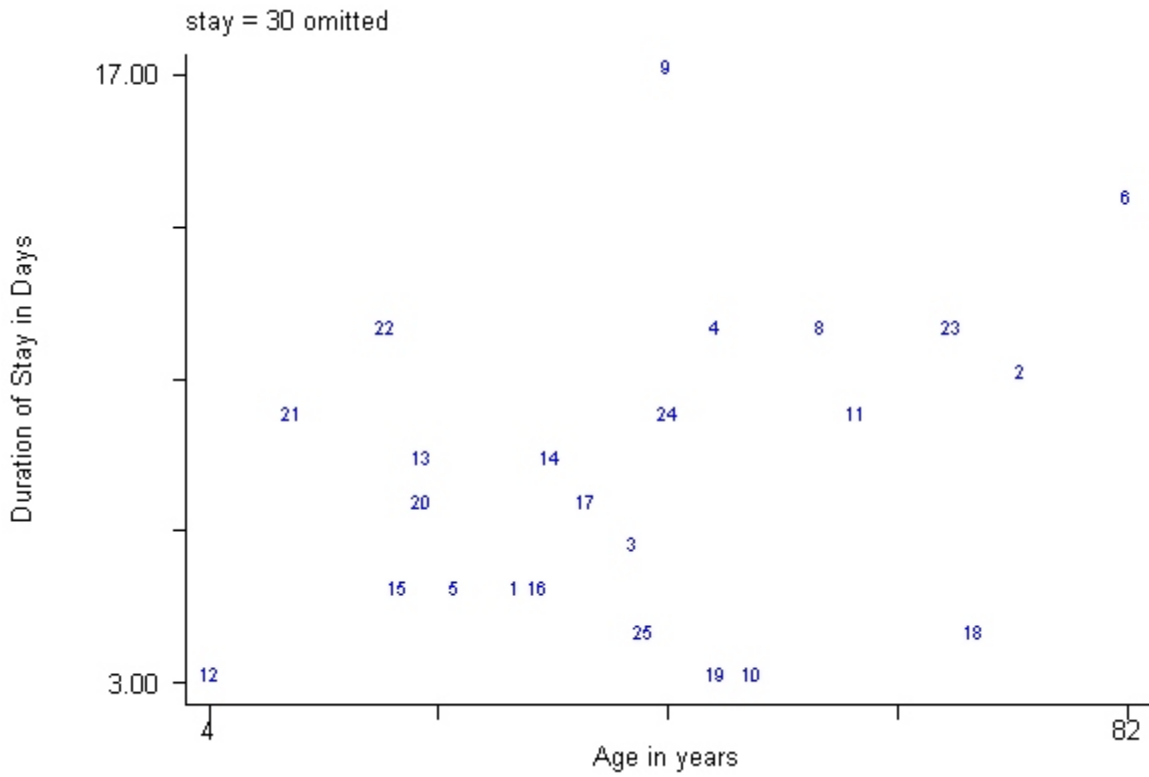
```
log type: text
```

```
closed on: 30 Aug 2001, 16:39:13
```

. graph stay age



```
. graph stay age if stay < 30, t1(stay = 30 omitted) s([id])
```



The commands below create a variable which divides age into quartiles. Since this version of age is a categorical variable, I called it catage.

```
. gen catage = 1 if age <= 25
(18 missing values generated)

. replace catage = 2 if age > 25 & age <= 41
(6 real changes made)

. replace catage = 3 if age > 41 & age <= 56
(6 real changes made)

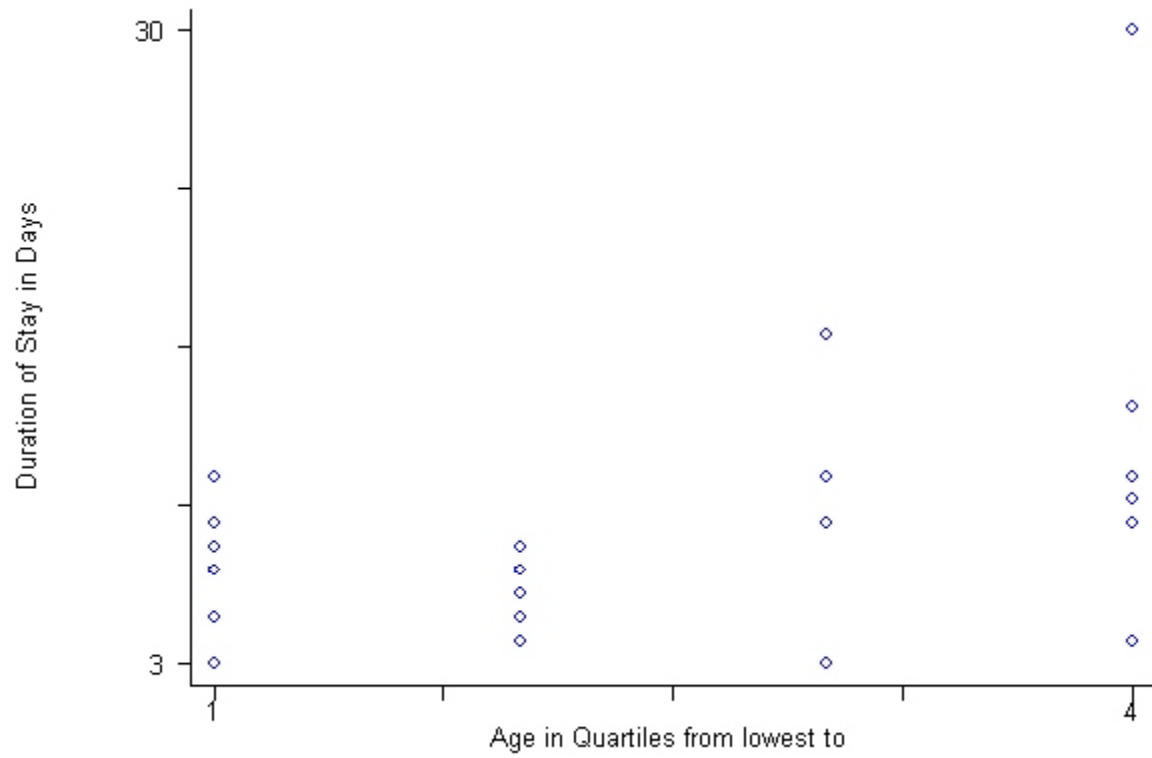
. replace catage = 4 if age > 56 & age ~= .
(6 real changes made)

. label variable catage "Age in Quartiles from lowest to highest"

. tab catage
```

Age in Quartiles from lowest to highest	Freq.	Percent	Cum.
1	7	28.00	28.00
2	6	24.00	52.00
3	6	24.00	76.00
4	6	24.00	100.00
Total	25	100.00	

. graph stay catage



```
. tab catage,sum(stay)
```

Age in Quartiles from lowest to highest	Summary of Duration of Stay in Days		
	Mean	Std. Dev.	Freq.
1	6.8571429	2.7342623	7
2	5.8333333	1.4719601	6
3	9	5.3665631	6
4	13	8.9442719	6
Total	8.6	5.7154761	25

```
. format stay %4.2f
```

```
. tab catage,sum(stay)
```

Age in Quartiles from lowest to highest	Summary of Duration of Stay in Days		
	Mean	Std. Dev.	Freq.
1	6.86	2.73	7
2	5.83	1.47	6
3	9.00	5.37	6
4	13.00	8.94	6
Total	8.60	5.72	25

```
. by catage:sum stay,det  
not sorted  
r(5);
```

```
. sort catage
```

. by catage:sum stay,det

-> catage = 1

Duration of Stay in Days

	Percentiles	Smallest		
1%	3	3		
5%	3	5		
10%	3	5	Obs	7
25%	5	7	Sum of Wgt.	7
50%	7		Mean	6.857143
		Largest	Std. Dev.	2.734262
75%	9	7		
90%	11	8	Variance	7.47619
95%	11	9	Skewness	.1078343
99%	11	11	Kurtosis	1.956854

-> catage = 2

Duration of Stay in Days

	Percentiles	Smallest		
1%	4	4		
5%	4	5		
10%	4	5	Obs	6
25%	5	6	Sum of Wgt.	6
50%	5.5		Mean	5.833333
		Largest	Std. Dev.	1.47196
75%	7	5		
90%	8	6	Variance	2.166667
95%	8	7	Skewness	.3053163
99%	8	8	Kurtosis	1.848284

-> catage = 3

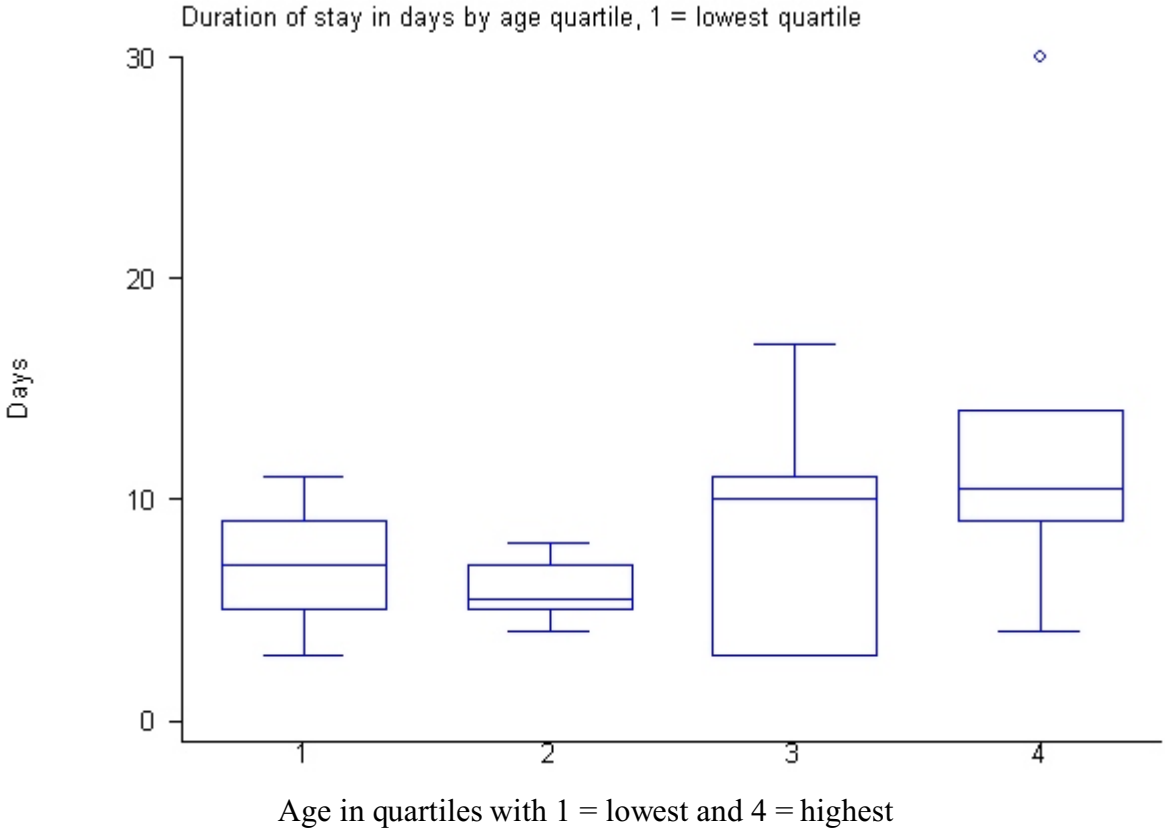
Duration of Stay in Days

	Percentiles	Smallest		
1%	3	3		
5%	3	3		
10%	3	9	Obs	6
25%	3	11	Sum of Wgt.	6
50%	10		Mean	9
		Largest	Std. Dev.	5.366563
75%	11	9		
90%	17	11	Variance	28.8
95%	17	11	Skewness	.1360828
99%	17	17	Kurtosis	1.944444

-> catage = 4

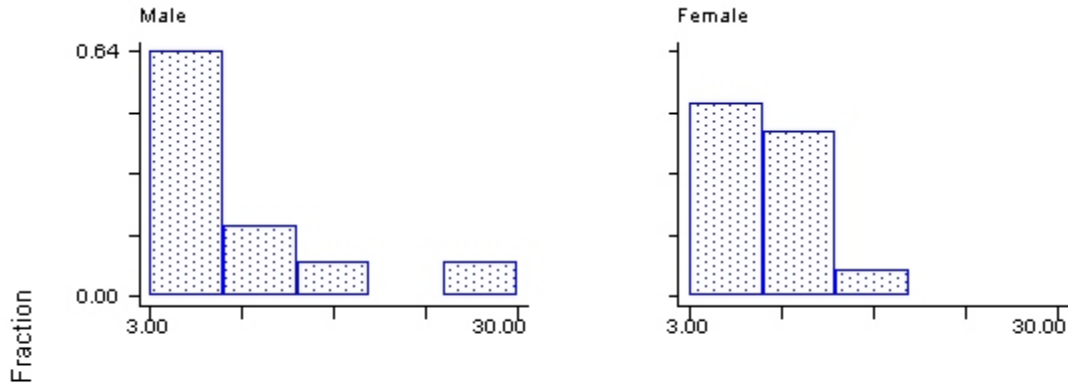
Duration of Stay in Days

	Percentiles	Smallest		
1%	4	4		
5%	4	9		
10%	4	10	Obs	6
25%	9	11	Sum of Wgt.	6
50%	10.5		Mean	13
		Largest	Std. Dev.	8.944272
75%	14	10		
90%	30	11	Variance	80
95%	30	14	Skewness	1.251077
99%	30	30	Kurtosis	3.39135



```
. graph stay,by(catage) box ylab t2(Duration of stay in days by age quartile, 1 = lowest quartile) l1(Days)
```

```
. sort sex  
. graph stay,by(sex)
```



Duration of Stay in Days
Histograms by Gender

```
. tab sex,sum(stay)
```

Summary of Duration of Stay in Days			
Gender	Mean	Std. Dev.	Freq.
Male	8.73	7.91	11
Female	8.50	3.48	14
Total	8.60	5.72	25

. by sex:sum stay,det

-> sex = Male

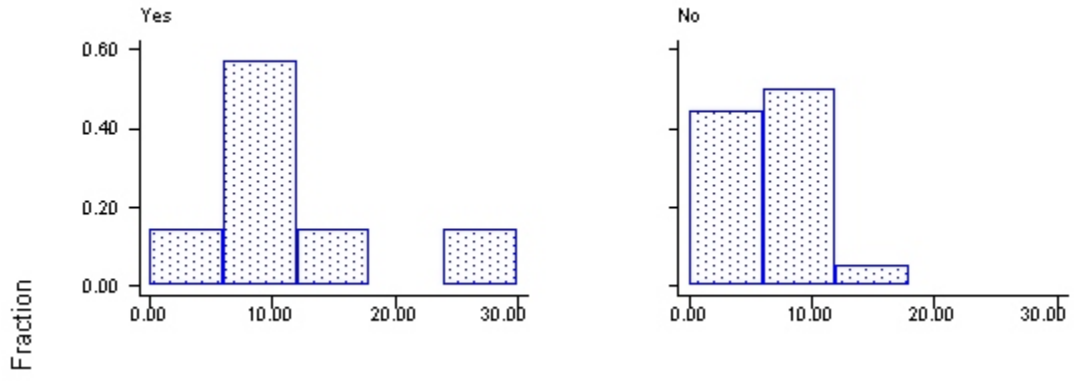
Duration of Stay in Days

	Percentiles	Smallest		
1%	3	3		
5%	3	3		
10%	3	3	Obs	11
25%	3	4	Sum of Wgt.	11
50%	7		Mean	8.727273
		Largest	Std. Dev.	7.913165
75%	11	9		
90%	14	11	Variance	62.61818
95%	30	14	Skewness	1.916305
99%	30	30	Kurtosis	5.877836

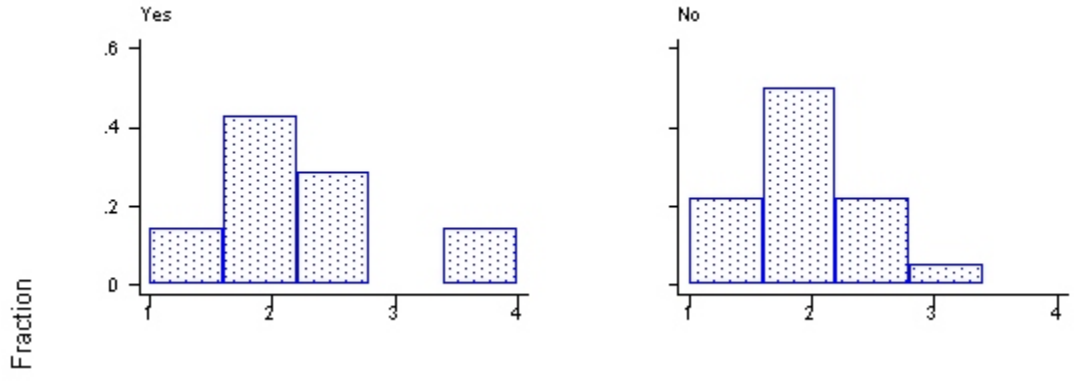
-> sex = Female

Duration of Stay in Days

	Percentiles	Smallest		
1%	4	4		
5%	4	5		
10%	5	5	Obs	14
25%	5	5	Sum of Wgt.	14
50%	8.5		Mean	8.5
		Largest	Std. Dev.	3.480716
75%	11	11		
90%	11	11	Variance	12.11538
95%	17	11	Skewness	.8120831
99%	17	17	Kurtosis	3.522751



Duration of Stay in Days
Histograms by Received Antibiotic - yes/no



LOS transformed using the natural logarithm
Histograms by Received Antibiotic - yes/no

```
. graph stay, by(antibio)
. graph logstay, by(antibio) b2(LOS transformed using the natural
logarithm)
```

```
. tab service antibio,row
```

Hospital Service	Received Antibiotic - yes/no		Total
	Yes	No	
Medical	2	7	9
	22.22	77.78	100.00
Surgical	5	11	16
	31.25	68.75	100.00
Total	7	18	25
	28.00	72.00	100.00

```
. tab service antibio,col
```

Hospital Service	Received Antibiotic - yes/no		Total
	Yes	No	
Medical	2	7	9
	28.57	38.89	36.00
Surgical	5	11	16
	71.43	61.11	64.00
Total	7	18	25
	100.00	100.00	100.00

```
. tab service antibio,cell
```

Hospital Service	Received Antibiotic - yes/no		Total
	Yes	No	
Medical	2	7	9
	8.00	28.00	36.00
Surgical	5	11	16
	20.00	44.00	64.00
Total	7	18	25
	28.00	72.00	100.00