

- **read the CD4 data into STATA**

```
. infile time count age packs drugs sex cesd id using c:\data\cd4.data.txt,clear
(2376 observations read)
```

- **look at the variables**

```
. des
```

Contains data

```
obs:      2,376
vars:      8
size:      85,536 (99.8% of memory free)
```

variable name	storage type	display format	value label	variable label
time	float	%9.0g		
count	float	%9.0g		
age	float	%9.0g		
packs	float	%9.0g		
drugs	float	%9.0g		
sex	float	%9.0g		
cesd	float	%9.0g		
id	float	%9.0g		

Sorted by:

Note: dataset has changed since last saved

- **look at contents of the data**

```
. codebook
```

```
time ----- (unlabeled)
      type:  numeric (float)
      range:  [-2.989733,5.459274]          units:  1.000e-06
unique values:  1342                      coded missing:  0 / 2376

      mean:   .828425
      std. dev:  1.87823

      percentiles:      10%      25%      50%      75%      90%
                        -1.62081  -.392882  .729637  2.19439  3.47981
```

```
count ----- (unlabeled)
      type:  numeric (float)
      range:  [10,3184]                      units:  1
unique values:  1127                      coded missing:  0 / 2376

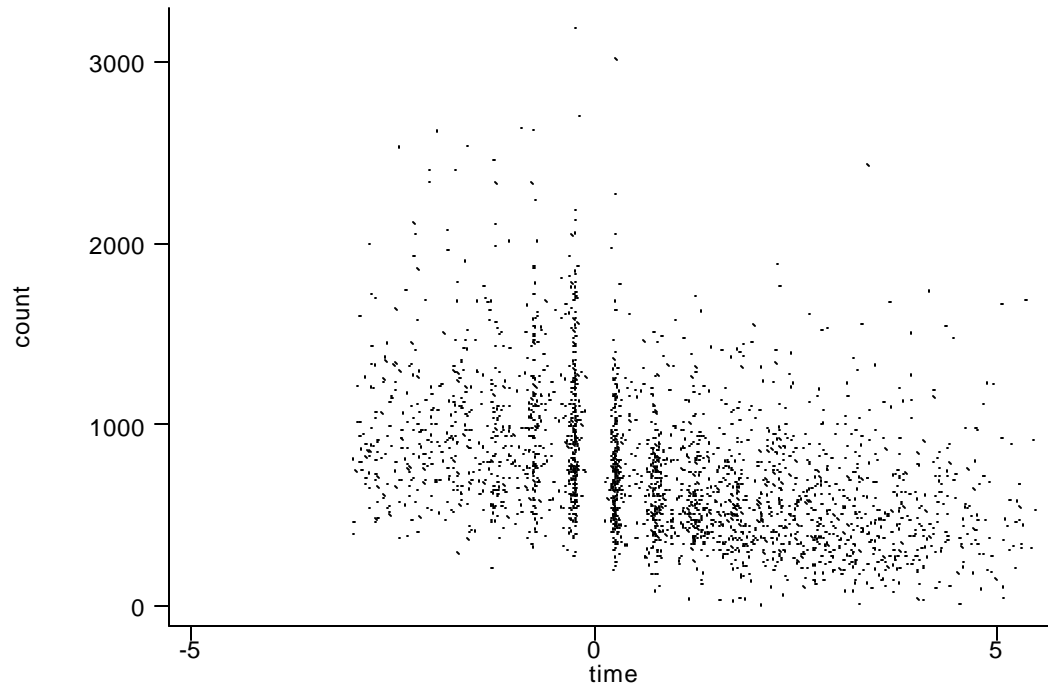
      mean:   765.131
      std. dev:  399.372

      percentiles:      10%      25%      50%      75%      90%
                        333      482.5      701.5      964      1267
```

```
..... (skip the details)
```

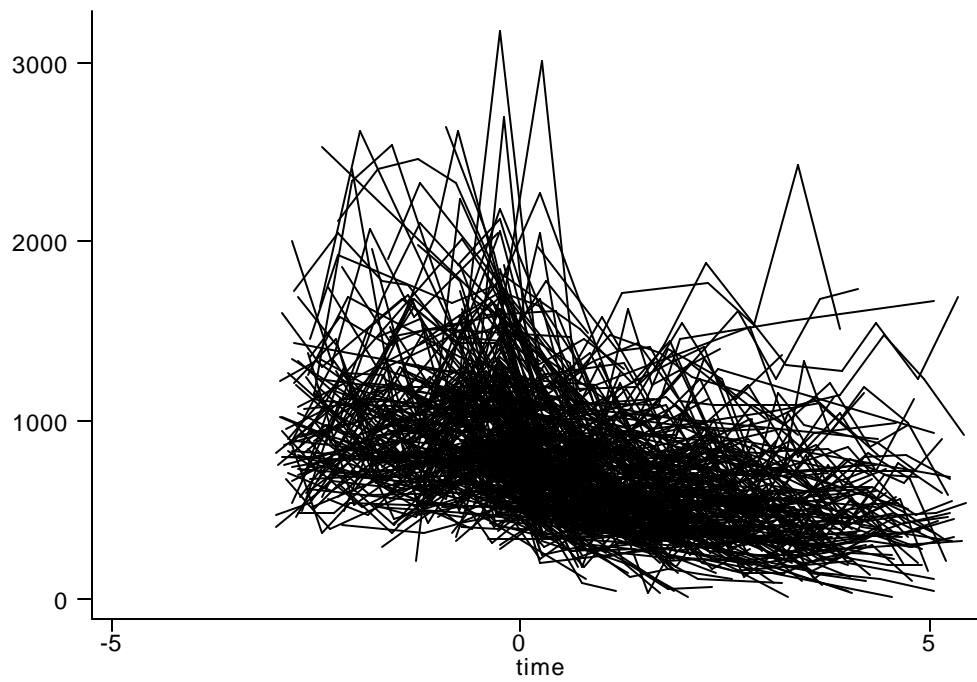
• **scatter plot**

```
. graph count time, s(.) xlab ylab
```



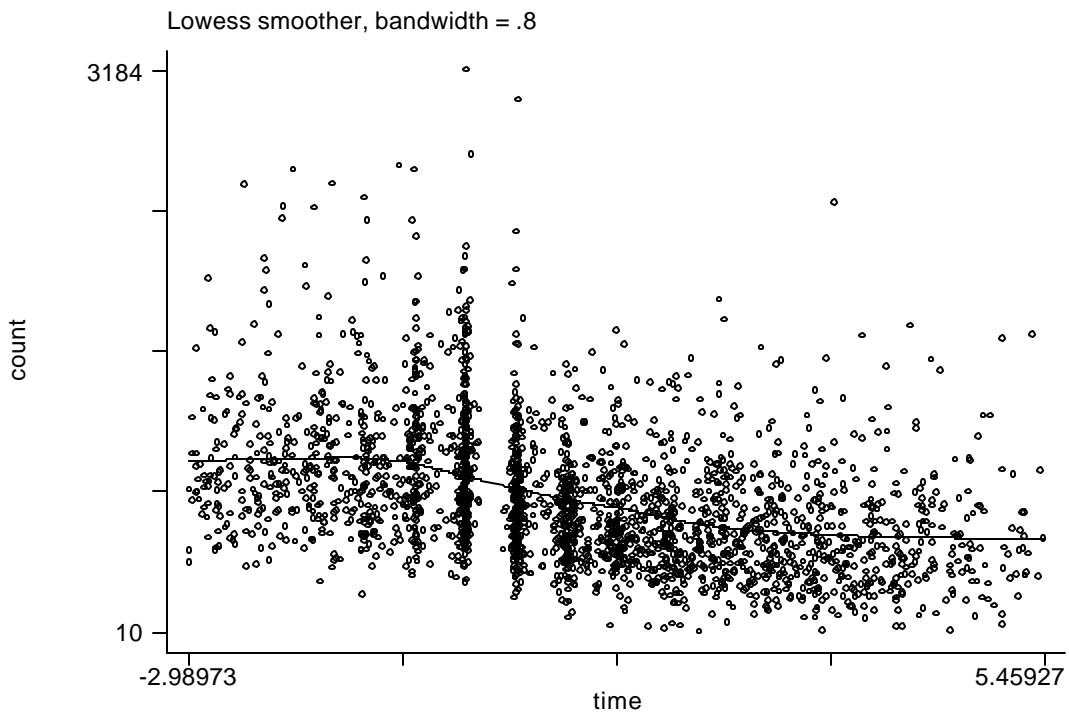
• **spaghetti plot**

```
. sort id time  
. graph count time, c(L) s(i) xlab ylab
```

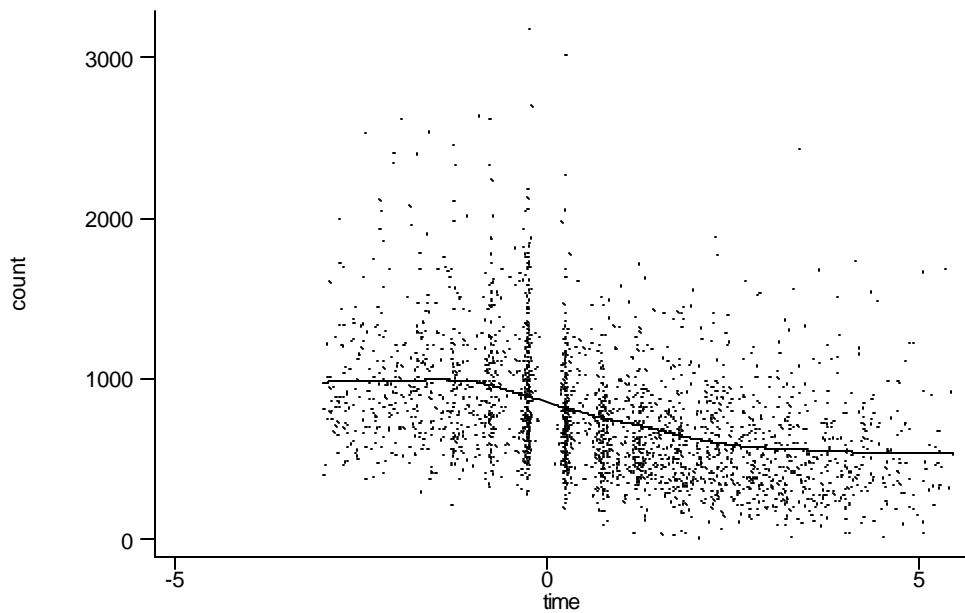


• **smoothing plot**

```
. ksm count time, lowess bw(.8)
```



```
. ksm count time, lowess bw(.8) s(.i) xlab ylab  
Lowess smoother, bandwidth = .8
```



• **set as time series data (show xt help)**

```
. iis id  
. tis time
```

• **within and between-subject variation**

. xtsum count

Variable		Mean	Std. Dev.	Min	Max	Observations
count	overall	765.1313	399.3716	10	3184	N = 2376
	between		290.4223	245.6	1979.5	n = 369
	within		280.5755	-304.8687	2762.131	T-bar = 6.43902

. xtsumcorr count

Variable		Mean	Std. Dev.	Min	Max	Observations
count	overall	765.1313	399.3716	10	3184	N = 2376
	between		290.4223	245.6	1979.5	n = 369
	within		280.5755	-304.8687	2762.131	T-bar = 6.43902
	corr. between		257.5667			
	corr. within		305.2164			
	rho		.4159	(betw. fract. of total)		

. *within-subject correlation is .42 (%58 variance come from within subjects)

. reg count time

Source	SS	df	MS	Number of obs =	2376
Model	66292987.4	1	66292987.4	F(1, 2374) =	503.59
Residual	312513916	2374	131640.234	Prob > F =	0.0000
Total	378806903	2375	159497.643	R-squared =	0.1750
				Adj R-squared =	0.1747
				Root MSE =	362.82

count	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
time	-88.95157	3.963822	-22.44	0.000	-96.72449	-81.17866
_cons	838.821	8.13554	103.11	0.000	822.8675	854.7745

. predict cres,resid

. xtsumcorr cres

Variable		Mean	Std. Dev.	Min	Max	Observations
cres	overall	-5.96e-08	362.7462	-733.8218	2323.504	N = 2376
	between		277.5963	-486.6496	1091.763	n = 369
	within		237.8347	-952.2429	1860.894	T-bar = 6.43902
	corr. between		254.2592			
	corr. within		258.722			
	rho		.4913	(betw. fract. of total)		

. *within variance decreases after adjusted by time

• **variogram of the residuals**

. reg count time age drugs sex

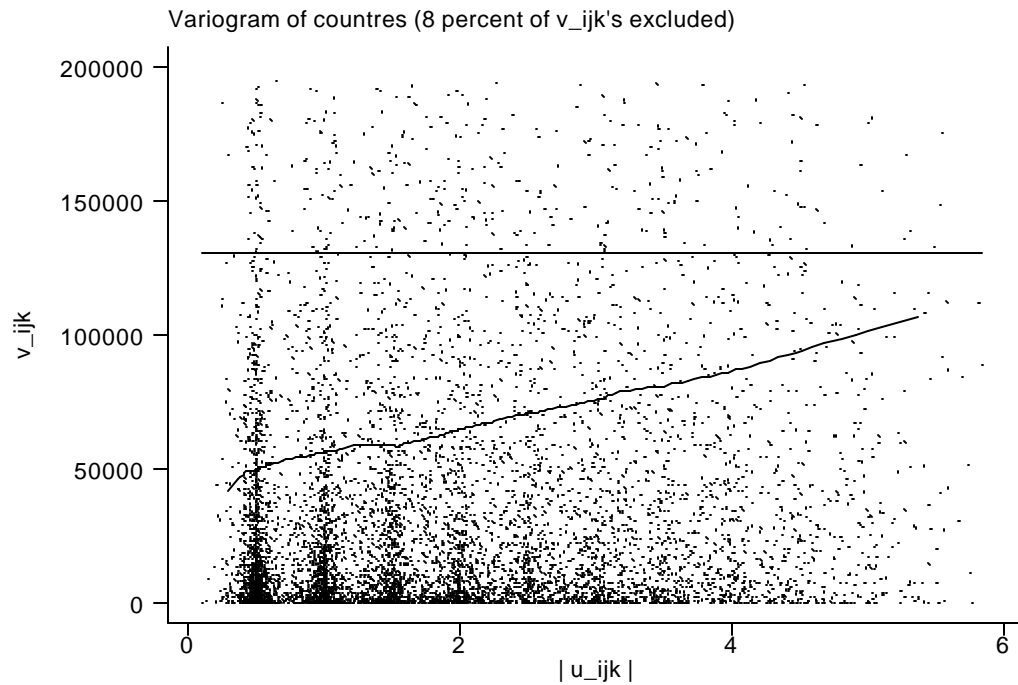
Source	SS	df	MS	Number of obs =	2376
Model	69154380.9	4	17288595.2	F(4, 2371) =	132.38
Residual	309652522	2371	130599.967	Prob > F =	0.0000
Total	378806903	2375	159497.643	R-squared =	0.1826
				Adj R-squared =	0.1812
				Root MSE =	361.39

count	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-------	-------	-----------	---	------	----------------------	--

time	-88.96601	4.130045	-21.54	0.000	-97.06489	-80.86714
age	2.147381	.9989652	2.15	0.032	.1884453	4.106317
drugs	78.15266	17.74669	4.40	0.000	43.35203	112.9533
sex	-2.104262	2.16046	-0.97	0.330	-6.340848	2.132324
_cons	774.0258	16.24769	47.64	0.000	742.1647	805.887

```

. predict countres, resid
. variogram countres
Computing smooth lowess model for v in ulag
    
```



• convert time into years and calculate the autocorrelation

```

. gen year=round(time,1)

. compress
count was float now int
packs was float now byte
drugs was float now byte
sex was float now byte
cesd was float now byte
year was float now byte

. tab year
    
```

year	Freq.	Percent	Cum.
-3	71	2.99	2.99
-2	198	8.33	11.32
-1	315	13.26	24.58
0	529	22.26	46.84
1	431	18.14	64.98
2	346	14.56	79.55
3	254	10.69	90.24
4	163	6.86	97.10
5	69	2.90	100.00
Total	2376	100.00	

```
. drop if year <-2 |year >4
(140 observations deleted)

. *keep one obs per year for each person
. sort id year

. by id year: drop if _n>1
(804 observations deleted)
```

• **set as time series data**

```
. tsset id year
      panel variable:  id, 10002 to 41844
      time variable:  year, -2 to 4, but with gaps

. sort year
. by year: sum count
```

-> year = -2

Variable	Obs	Mean	Std. Dev.	Min	Max
count	133	1004.594	421.6318	295	2534

-> year = -1

Variable	Obs	Mean	Std. Dev.	Min	Max
count	211	979.1659	415.7049	218	2640

-> year = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
count	307	975.4039	421.364	283	3184

-> year = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
count	279	666.8315	278.4096	89	1517

-> year = 2

Variable	Obs	Mean	Std. Dev.	Min	Max
count	226	593.5	307.2267	10	1771

-> year = 3

Variable	Obs	Mean	Std. Dev.	Min	Max
count	167	537.5689	295.7025	88	1615

-> year = 4

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

```
-----+-----
count |      109      519.0183      313.019          39      1681
```

. des

Contains data

```
obs:      1,432
vars:      9
size:      32,936 (99.9% of memory free)
```

```
-----+-----
```

variable name	storage type	display format	value label	variable label
time	float	%9.0g		
count	int	%9.0g		
age	float	%9.0g		
packs	byte	%9.0g		
drugs	byte	%9.0g		
sex	byte	%9.0g		
cesd	byte	%9.0g		
id	float	%9.0g		
year	byte	%9.0g		

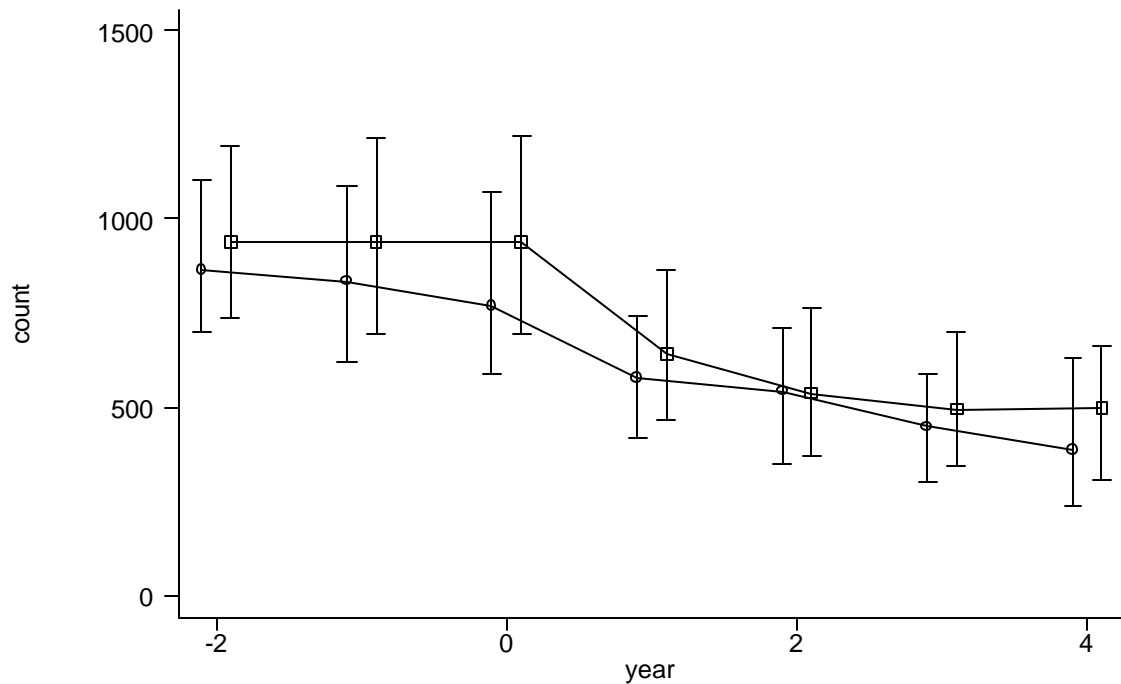
```
-----+-----
```

Sorted by: year

Note: dataset has changed since last saved

• **plot count over year by drug**

```
. xtgraph count, group(drug) av(median) bar(iqr) offset(.2) xlab ylab
      —○— 0      —□— 1
```



• **autocorrelation matrix and autocorrelation function**

. autocor count year id

	time1	time2	time3	time4	time5	time6	time7
time1	1.0000						
time2	0.6294	1.0000					
time3	0.5659	0.4856	1.0000				
time4	0.4377	0.5179	0.4981	1.0000			
time5	0.3220	0.4443	0.3889	0.6795	1.0000		
time6	0.4302	0.5336	0.3188	0.5680	0.7375	1.0000	
time7	0.8893	0.4241	0.4316	0.5241	0.7105	0.7781	1.0000

```

acf
1. .5357287
2. .4570088
3. .3958656
4. .4472889
5. .4302677
6. .8893483
    
```

