

```
. *****;
. ***** Dental data set *****;
. ***** distance difference across boys and girls over time *****;
. *****;
```

```
. infile obs id age dist sex using "C:\Documents and Settings\Yijie\My Documents\ldata\dental.dat"
(108 observations read)
```

```
. drop obs
```

```
. summ
```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	108	14	7.825193	1	27
age	108	11	2.246493	8	14
dist	108	24.02315	2.928577	16.5	31.5
sex	108	.5925926	.4936425	0	1

```
. ** Reshape the data into "long" or "wide" format - depend on which one you want;
. ** We need long format for modelling;
```

```
. reshape wide dist, i(id) j(age);
(note: j = 8 10 12 14)
```

Data	long	->	wide
Number of obs.	108	->	27
Number of variables	4	->	6
j variable (4 values)	age	->	(dropped)
xij variables:	dist	->	dist8 dist10 ... dist14

```
. reshape long dist, i(id) j(age);
(note: j = 8 10 12 14)
```

Data	wide	->	long
Number of obs.	27	->	108
Number of variables	6	->	4
j variable (4 values)		->	age
xij variables:	dist8 dist10 ... dist14	->	dist

```
. tsset id age;
panel variable: id, 1 to 27
time variable: age, 8 to 14, but with gaps
```

```
. xtides;
```

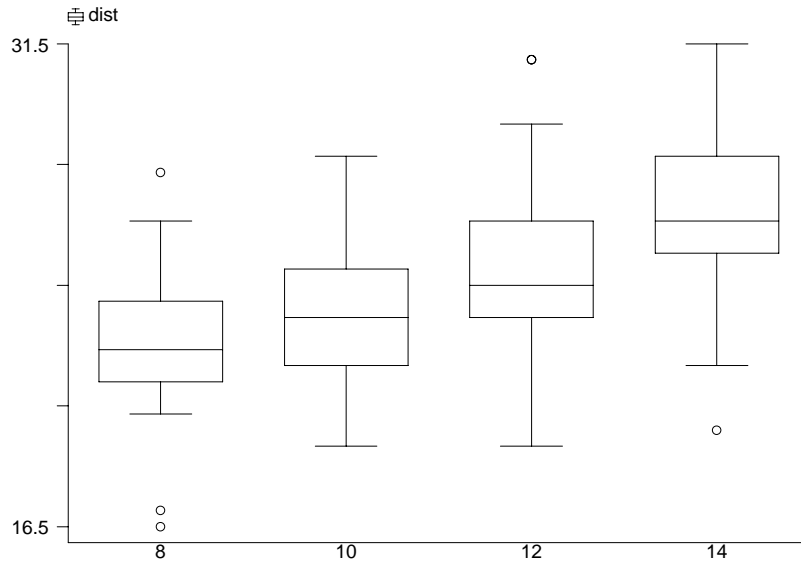
```
id: 1, 2, ..., 27          n =          27
age: 8, 10, ..., 14       T =          4
Delta(age) = 2; (14-8)/2 + 1 = 4
(id*age uniquely identifies each observation)
```

```
Distribution of T_i:  min    5%    25%    50%    75%    95%    max
                   4      4      4      4      4      4      4
```

Freq.	Percent	Cum.	Pattern
27	100.00	100.00	1111

```
27 100.00 | XXXX
```

```
. graph box dist, over(age)
```



```
. ** the time-varying variables are: age, dist;  
. ** the baseline variables are: id, sex;  
. ** to summarize: means, standard deviations, frequenct, for time-series (xt) data;  
. xttab sex;
```

sex	Overall		Between		Within
	Freq.	Percent	Freq.	Percent	Percent
0	44	40.74	11	40.74	100.00
1	64	59.26	16	59.26	100.00
Total	108	100.00	27	100.00	100.00

(n = 27)

```
. xtsum age;
```

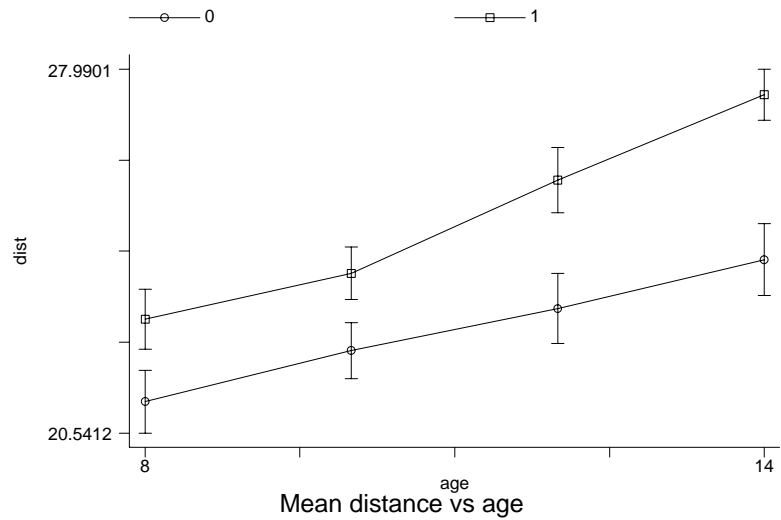
Variable		Mean	Std. Dev.	Min	Max	Observations
age	overall	11	2.246493	8	14	N = 108
	between		0	11	11	n = 27
	within		2.246493	8	14	T = 4

```
. xtsum sex;
```

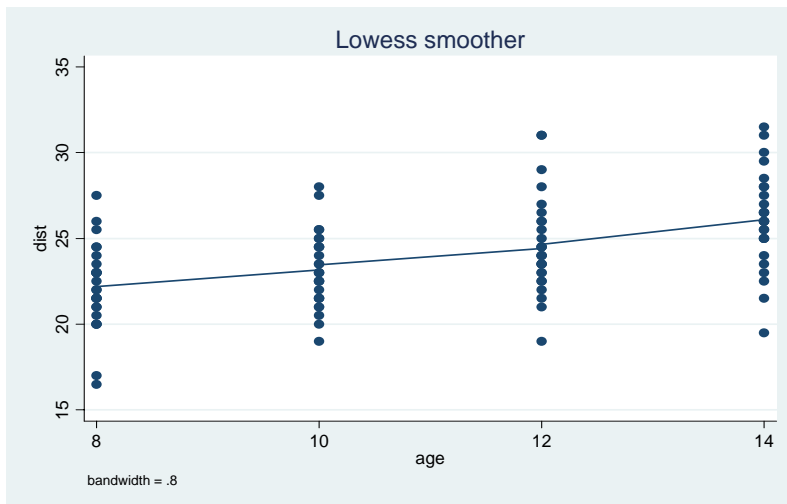
Variable		Mean	Std. Dev.	Min	Max	Observations
sex	overall	.5925926	.4936425	0	1	N = 108
	between		.5007117	0	1	n = 27
	within		0	.5925926	.5925926	T = 4

Sex is a time invariant covariates, therefore no within-subject variance. Age is a time variant covariate, and every individual has same age points, therefore no between-subject variance.

```
. ** Mean trend plot***;
. xtgraph dist, group(sex) ti("Mean distance vs age") bar(ci) i(id) t(age);
```



```
. ** kernel smooth **
. ksm dist age, lowess gen(distsmth)
```



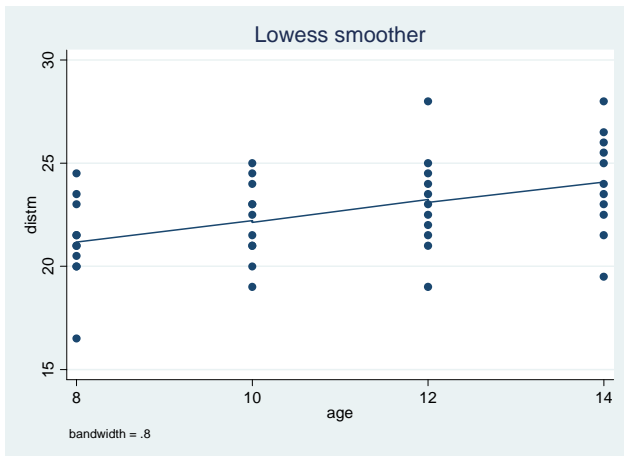
```
** separate kernel smooth for different gender group **
```

```
. gen distm = dist if sex == 0
(64 missing values generated)

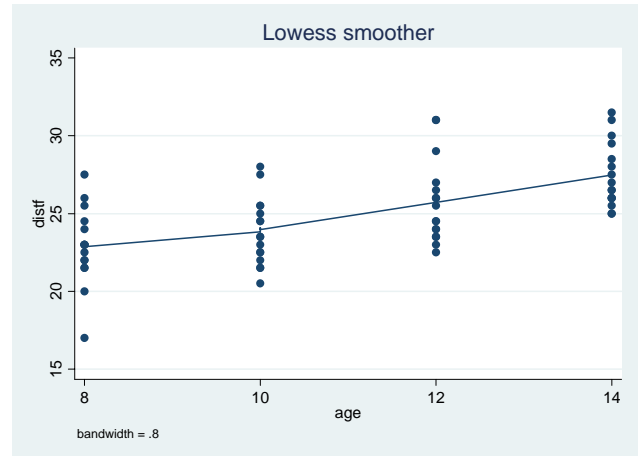
. gen distf = dist if sex == 1
(44 missing values generated)

. ksm distm age, lowess gen(distmsmth)
. ksm distf age, lowess gen(distfsmth)
```

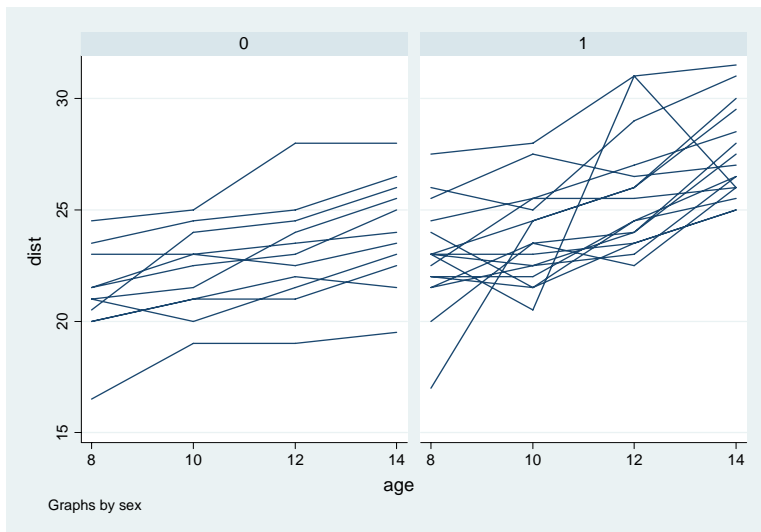
For Male:



For Female:



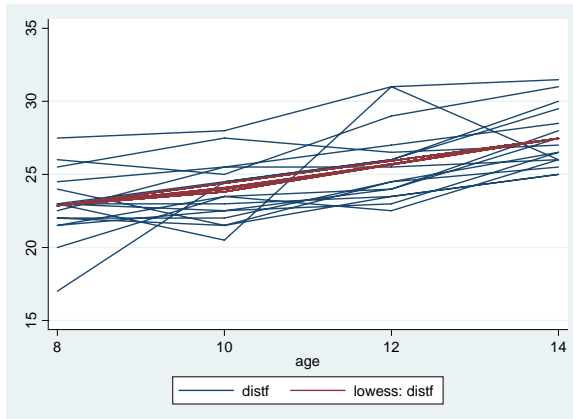
```
. ** ** Spaghetti plots ;
. sort sex id age;
. twoway line dist age, by(sex) connect(L);
. ** the option connect(L) is crucial here to guarantee that only observations from the same
. ** subject ordered by time is connected. Otherwise, STATA will connect the last observation from
. ** the 1st subject with the first observation from the 2nd subject.
```



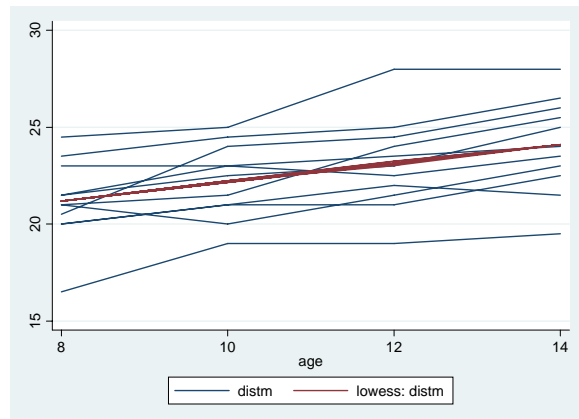
\*\* Spaghetti plots with lowess smooth;

```
. twoway line distm distmsmth age, connect(L)
. twoway line distf distfsmth age, connect(L)
```

For Female:



For Male:



ZAP plot:

```
. regress dist age sex
. predict rs, resid

. egen mrs=median(rs), by(id)

. reshape wide dist distm distf distsmth distfsmth rs, i(id) j(age)
. egen maxmrs=max(mrs)
. egen minmrs=min(mrs)
. egen medmrs=median(mrs)
. egen mrs75=pctile(mrs),p(75)
. egen mrs25=pctile(mrs),p(25)

. reshape long dist distm distf distsmth distfsmth rs, i(id) j(age)
. gen maxrs=rs if mrs==maxmrs
(104 missing values generated)
. gen minrs=rs if mrs==minmrs
(104 missing values generated)
. gen medrs=rs if mrs==medmrs
(104 missing values generated)
. gen rs25=rs if mrs==mrs25
(104 missing values generated)
. gen rs75=rs if mrs==mrs75
(104 missing values generated)

. twoway (scatter rs age) (line maxrs minrs medrs rs25 rs75 age)
```

