

Example Analysis with STATA

- Exploratory Data Analysis
 - ▷ Means and Variance by Time and Group
 - ▷ Correlation
 - ▷ Individual Series
- Derived Variable Analysis
 - ▷ Fitting a Line to Each Subject
 - ▷ Summarizing Slopes by Group
- Regression Analysis
 - ▷ Using Linear Mixed Model (`xtmixed`; `gllamm`)
 - ▷ Using GEE

CF Data – Summary

- Cystic Fibrosis Data
 - ▷ **Subjects:** A total of N=200 subjects enrolled in a cohort study (patient registry).
 - ▷ **Observation-level Measures:**
 - * FEV1 – measure of pulmonary function
 - * AGE – age in years
 - ▷ **Person-level Measures:**
 - * GENDER – male / female
 - * F508 – genotype (number of alleles)
 - * AGE0 – age-at-entry to study
- **Q:** Is level and/or rate of decline in pulmonary function associated with gender and/or genotype?

Example: CF Data

* cfkids.do *

* *

* PURPOSE: analysis of FEV1 among CF kids *

* *

* AUTHOR: P. Heagerty *

* *

* DATE: 31 March 2005 *

* 04 March 2006 *

infile id fev1 age gender pseudoA f508 pancreat using NewCFkids.raw

*

```
* ID          = patient id
* FEV1        = percent-predicted forced expiratory volume in 1 second
* AGE         = age (years)
* GENDER      = sex (1=male, 2=female)
* PSEUDOEA    = infection with Pseudo Aeruginosa (0=no, 3=yes)
* F508        = genotype (1=homozygous, 2=heterozygous, 3=none)
* PANCREAT    = pancreatic enzyme supplementation (0,1=no, 2=yes)
*
```

```
label variable age "Age (years)"
```

```
recode gender 1=0 2=1
```

```
label variable gender "female"
```

```
recode pseudoA 3=1
```

```
recode pancreat 1=0 2=1
```

```
recode f508 1=2 2=1 3=0

save NewCFkids, replace

***
*** some exploratory data analysis -- observation level
***

summarize fev1

gen y8 = fev1
recode y8 (min/max=.) if age > 8.75
recode y8 (min/max=.) if age < 7.25

gen y10 = fev1
recode y10 (min/max=.) if age > 10.75
recode y10 (min/max=.) if age < 9.25
```

```
gen y12 = fev1
recode y12 (min/max=.) if age > 12.75
recode y12 (min/max=.) if age < 11.25
```

```
gen y14 = fev1
recode y14 (min/max=.) if age > 14.75
recode y14 (min/max=.) if age < 13.25
```

```
gen y16 = fev1
recode y16 (min/max=.) if age > 16.75
recode y16 (min/max=.) if age < 15.25
```

```
gen y18 = fev1
recode y18 (min/max=.) if age > 18.75
recode y18 (min/max=.) if age < 17.25
```

```
gen y20 = fev1
recode y20 (min/max=.) if age > 20.75
```

```
recode y20 (min/max=.) if age < 19.25
```

```
***** the following creates a single record per kid:
```

```
collapse (mean) f508 gender y8 y10 y12 y14 y16 y18 y20, by(id)
```

```
***** look at means over these ages:
```

```
summarize
```

```
sort f508
```

```
by f508: summarize
```

```
sort gender
```

```
by gender: summarize
```

Results: CF Data

```
. summarize fev1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
fev1	1513	70.36416	27.22294	11.03	159.67

Results: CF Data

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	200	109141.9	5701.099	100073	119028
f508	200	1.335	.6745741	0	2
gender	200	.49	.5011544	0	1
y8	63	84.18426	25.81721	24.33	136.5
y10	81	80.96802	24.22817	24.645	119.95
y12	104	76.45958	26.55931	17.36	136.97
y14	116	72.9857	25.0613	14.2	134.415
y16	103	68.17985	26.34962	18.05	148.22
y18	90	66.46433	25.04077	21.25	136.51
y20	78	61.86303	25.19371	14.63	118.04

Results: CF Data

-> f508 = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
f508	23	0	0	0	0
y8	7	77.16214	24.50721	42.29	108.135
y10	8	73.28375	31.3521	24.645	108
y12	9	73.58333	34.82344	17.36	136.97
y14	12	82.67667	18.84188	43.51	104.88
y16	14	77.37286	25.45999	34.87	115.63
y18	11	86.28864	23.54095	42.97	109.78
y20	10	81.47	27.00938	26.93	118.04

Results: CF Data

-> f508 = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
f508	87	1	0	1	1
y8	25	83.4214	25.86229	39.845	117.7
y10	29	83.15086	22.07282	27.69	118.14
y12	38	77.83184	23.68194	24.26	123.39
y14	42	71.37131	25.35112	22.88	134.415
y16	38	67.60211	25.62312	21.39	148.22
y18	38	61.38513	21.67322	27.22	114.78
y20	34	61.33976	24.50993	21.65	110.29

Results: CF Data

-> f508 = 2

Variable	Obs	Mean	Std. Dev.	Min	Max
f508	90	2	0	2	2
y8	31	86.38511	26.55727	24.33	136.5
y10	44	80.92648	24.50779	31	119.95
y12	57	75.99889	27.40096	23.45	123.47
y14	62	72.20366	25.82999	14.2	126.1
y16	51	66.08676	27.0853	18.05	125.655
y18	41	65.85317	26.25174	21.25	136.51
y20	34	56.61956	23.15649	14.63	109.68

Results: CF Data

-> gender = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
gender	102	0	0	0	0
y8	38	85.9443	23.94876	38.42	122.61
y10	44	84.60659	23.16013	24.645	119.95
y12	56	79.9042	25.34632	17.36	136.97
y14	60	74.02294	21.95634	27.8	134.415
y16	50	72.0945	24.3662	31.315	120.52
y18	42	67.6119	24.78879	26.09	112.74
y20	37	63.53477	28.72807	14.63	118.04

Results: CF Data

-> gender = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
gender	98	1	0	1	1
y8	25	81.509	28.7279	24.33	136.5
y10	37	76.64108	25.06672	27.69	111.97
y12	48	72.44087	27.63062	23.45	123.47
y14	56	71.87437	28.17201	14.2	126.1
y16	53	64.48679	27.81735	18.05	148.22
y18	48	65.46021	25.47799	21.25	136.51
y20	41	60.35439	21.77503	21.65	109.68

Example: CF Data

```
***** information on correlation
```

```
corr y8 y10 y12
```

```
corr y12 y14 y16
```

```
corr y16 y18 y20
```

```
graph twoway (scatter y14 y8)
```

```
graph twoway (scatter y12 y8)
```

```
graph twoway (scatter y10 y8)
```

```
graph twoway (scatter y16 y10)
```

```
graph twoway (scatter y14 y10)
```

```
graph twoway (scatter y12 y10)
```

Results: CF Data

```
. corr y8 y10 y12  
(obs=50)
```

		y8	y10	y12
-----+-----				
y8		1.0000		
y10		0.8682	1.0000	
y12		0.8032	0.8904	1.0000

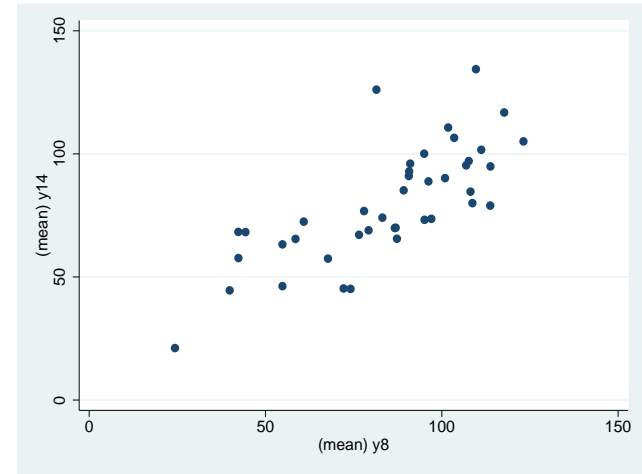
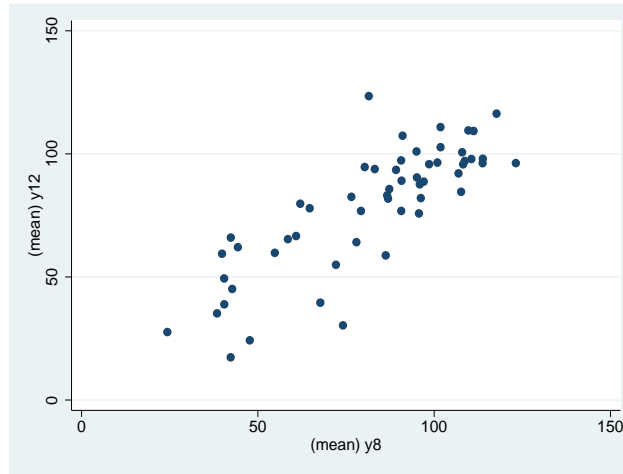
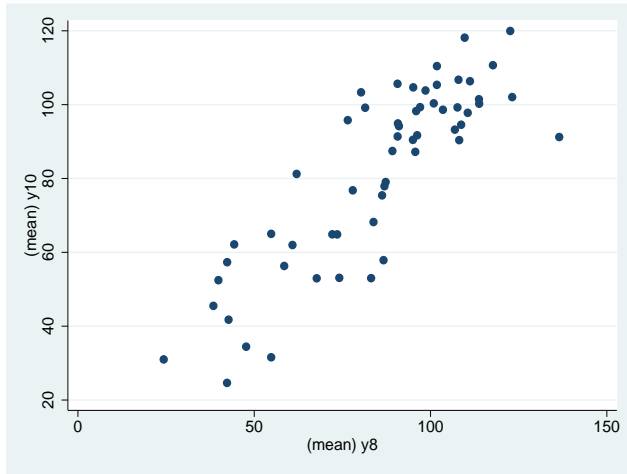
```
. corr y12 y14 y16  
(obs=53)
```

		y12	y14	y16
-----+-----				
y12		1.0000		
y14		0.7904	1.0000	
y16		0.6918	0.8830	1.0000

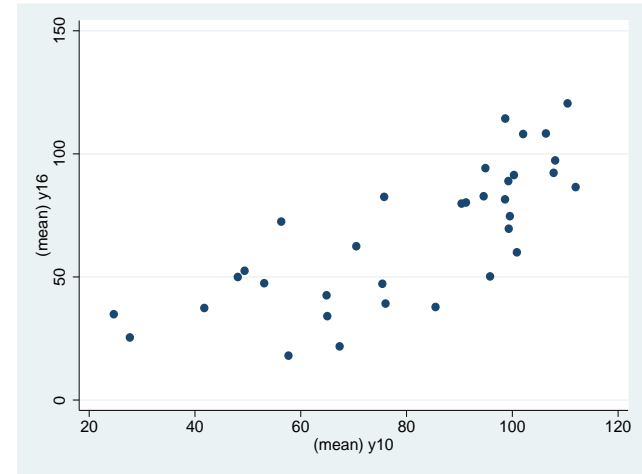
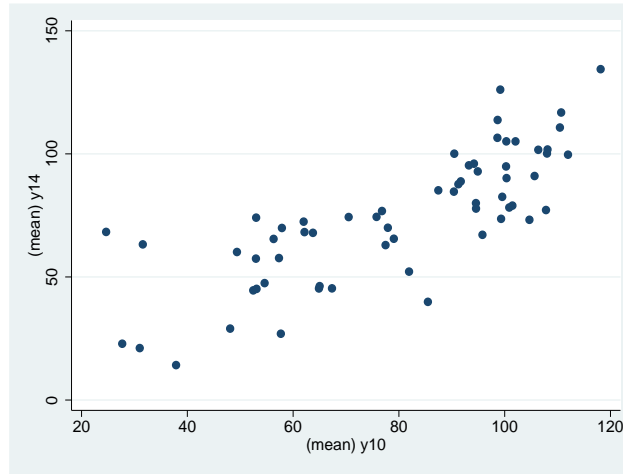
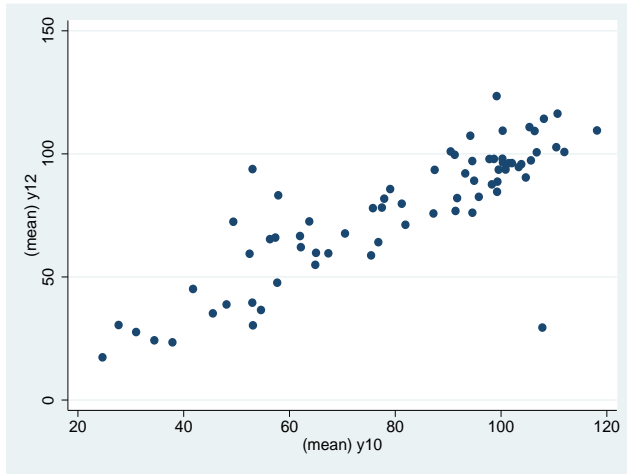

```
. corr y16 y18 y20  
(obs=47)
```

		y16	y18	y20
-----+-----				
y16		1.0000		
y18		0.8662	1.0000	
y20		0.8088	0.8638	1.0000

Results: CF Data – Pairwise Plots



Results: CF Data – Pairwise Plots



Example: CF Data

```
***  
*** some exploratory data analysis -- person level  
***  
  
clear  
  
use NewCFkids  
  
collapse (min) age f508 gender, by(id)  
  
tab f508  
tab gender  
tab f508 gender
```

Results: CF Data

f508	Freq.	Percent	Cum.
0	23	11.50	11.50
1	87	43.50	55.00
2	90	45.00	100.00
-----+			
Total	200	100.00	

gender	Freq.	Percent	Cum.
0	102	51.00	51.00
1	98	49.00	100.00
-----+			
Total	200	100.00	

Results: CF Data

```
. tab f508 gender
```

	(min) gender		
(min) f508	0	1	Total
0	13	10	23
1	46	41	87
2	43	47	90
Total	102	98	200

Example: CF Data

```
***
*** add the age-at-entry to the data
***

gen age0 = age

drop age f508 gender

***** save data with just ID and age0
save NewCFage0, replace

***** return to original data, and merge
clear

use NewCFkids
```

```
sort id
```

```
merge id using NewCFage0
```

```
list in 1/4
```

```
***** create ageL and save
```

```
gen ageL = age - age0
```

```
save NewCFkids, replace
```


Results: CF Data

```
. list in 1/4
```

```
+-----+
1. |      id |   fev1 |   age | gender | pseudoA | f508 | pancreat | age0 |
   | 100073 | 113.8 | 8.452 |      1 |         1 |    2 |         1 | 8.452 |
   |-----|
2. |      id |   fev1 |   age | gender | pseudoA | f508 | pancreat | age0 |
   | 100073 |  98.18 | 8.783 |      1 |         1 |    2 |         1 | 8.452 |
   |-----|
3. |      id |   fev1 |   age | gender | pseudoA | f508 | pancreat | age0 |
   | 100073 |  98.73 | 9.785 |      1 |         1 |    2 |         1 | 8.452 |
   |-----|
4. |      id |   fev1 |   age | gender | pseudoA | f508 | pancreat | age0 |
   | 100073 | 101.79 | 10.538 |      1 |         1 |    2 |         1 | 8.452 |
   |-----+
```

Example: CF Data

```
***
*** plot some individual series
***

graph twoway (lowess fev1 age) (scatter fev1 age) ///
    if (id<=101799), by(id)

graph twoway (lowess fev1 age) (scatter fev1 age) ///
    if (id>101799 & id<=103399), by(id)

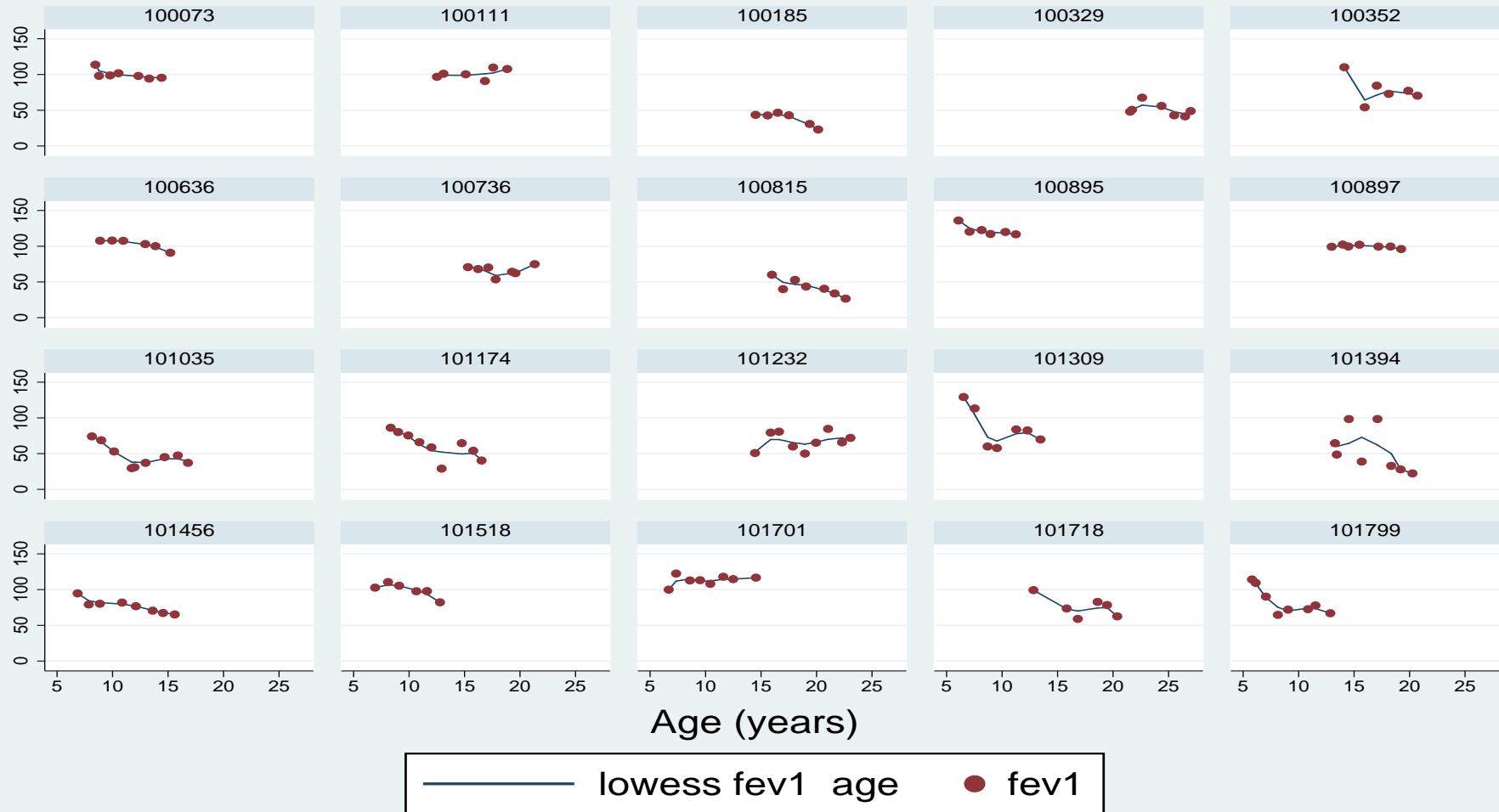
***
*** plot some fitted lines
***

graph twoway (lfit fev1 age) (scatter fev1 age) ///
```

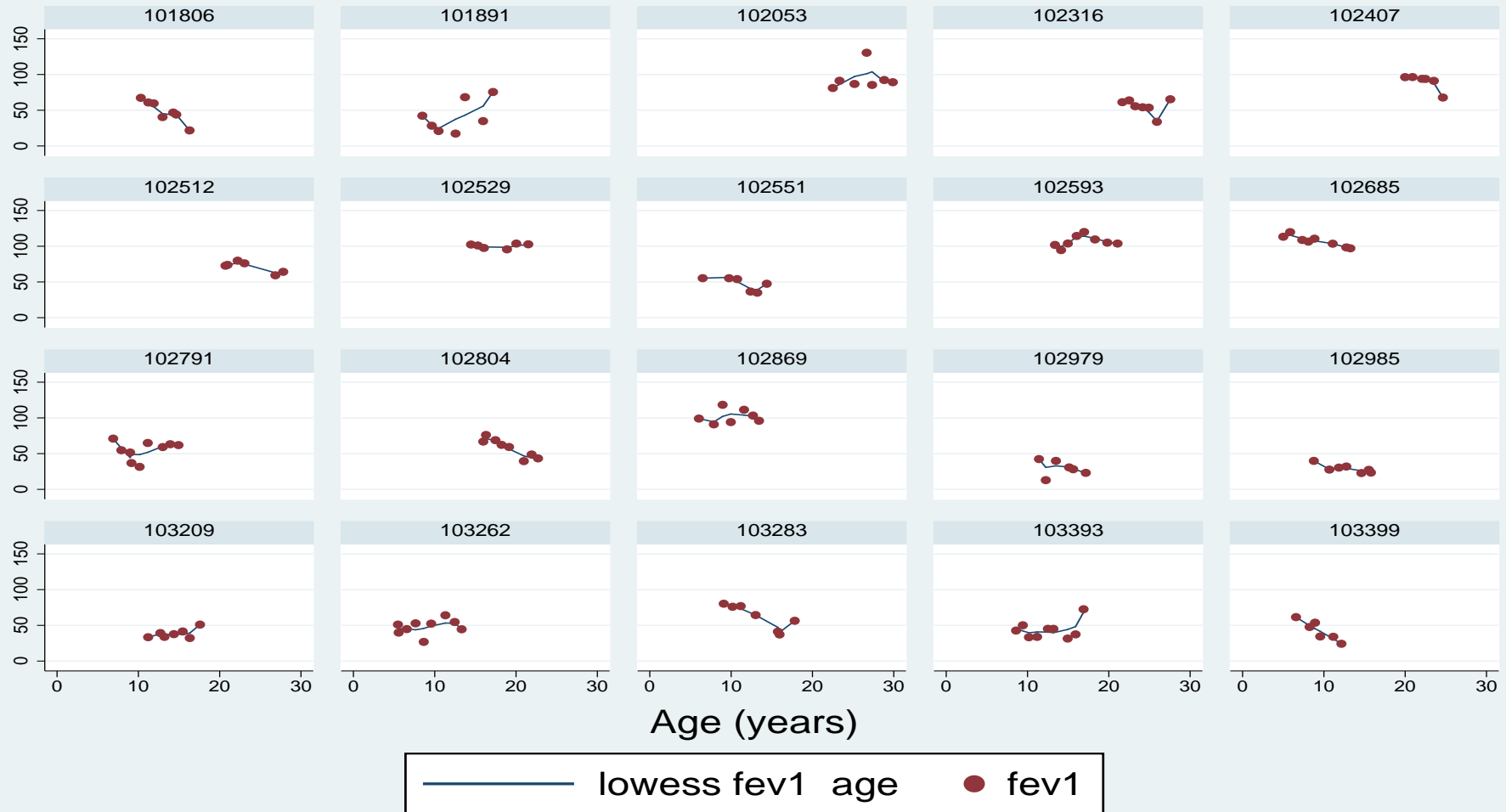
```
if (id<=101799), by(id)

graph twoway (lfit fev1 age) (scatter fev1 age) ///
if (id>101799 & id<=103399), by(id)
```

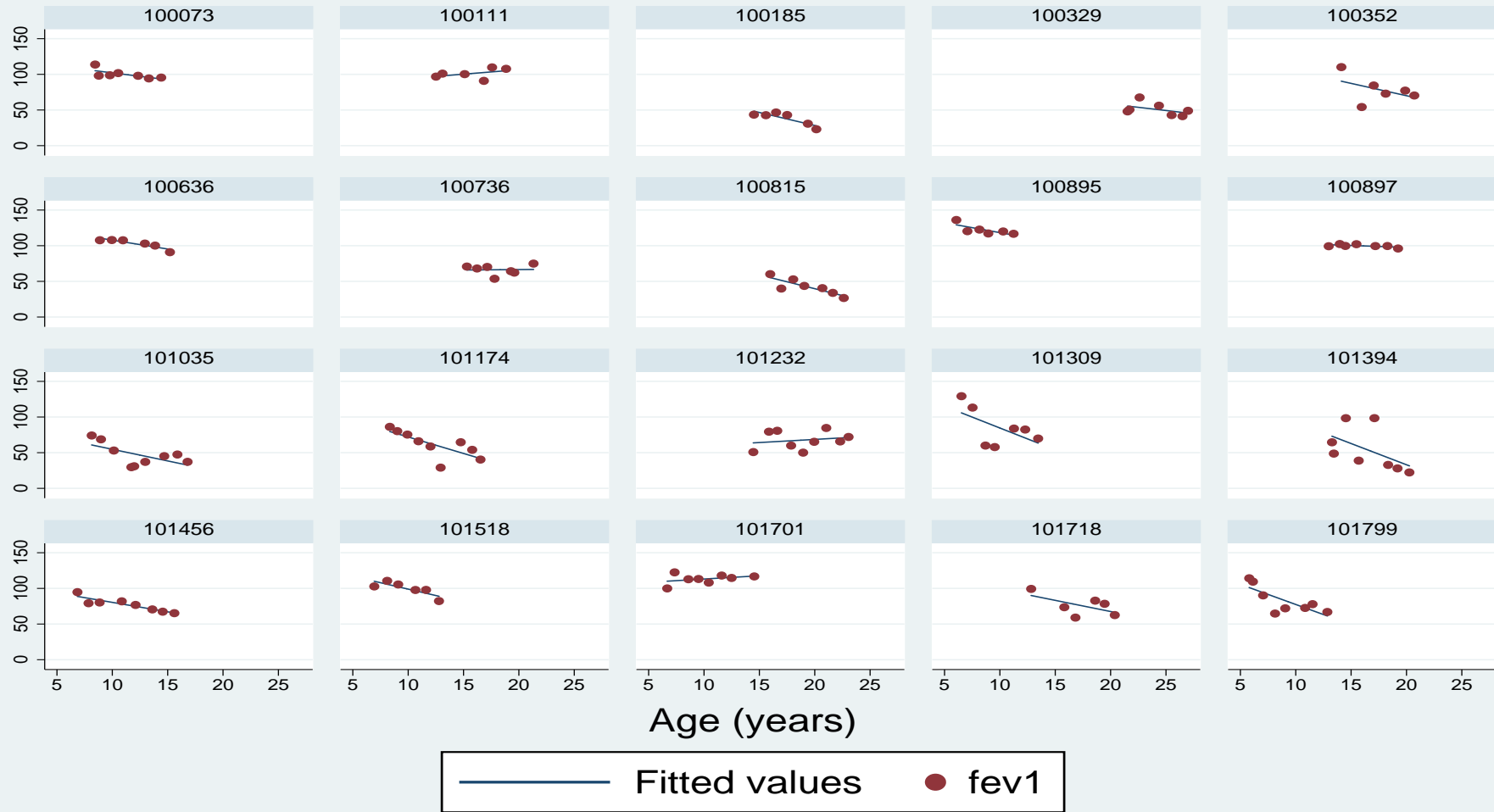
Results: CF Data – Individual Series



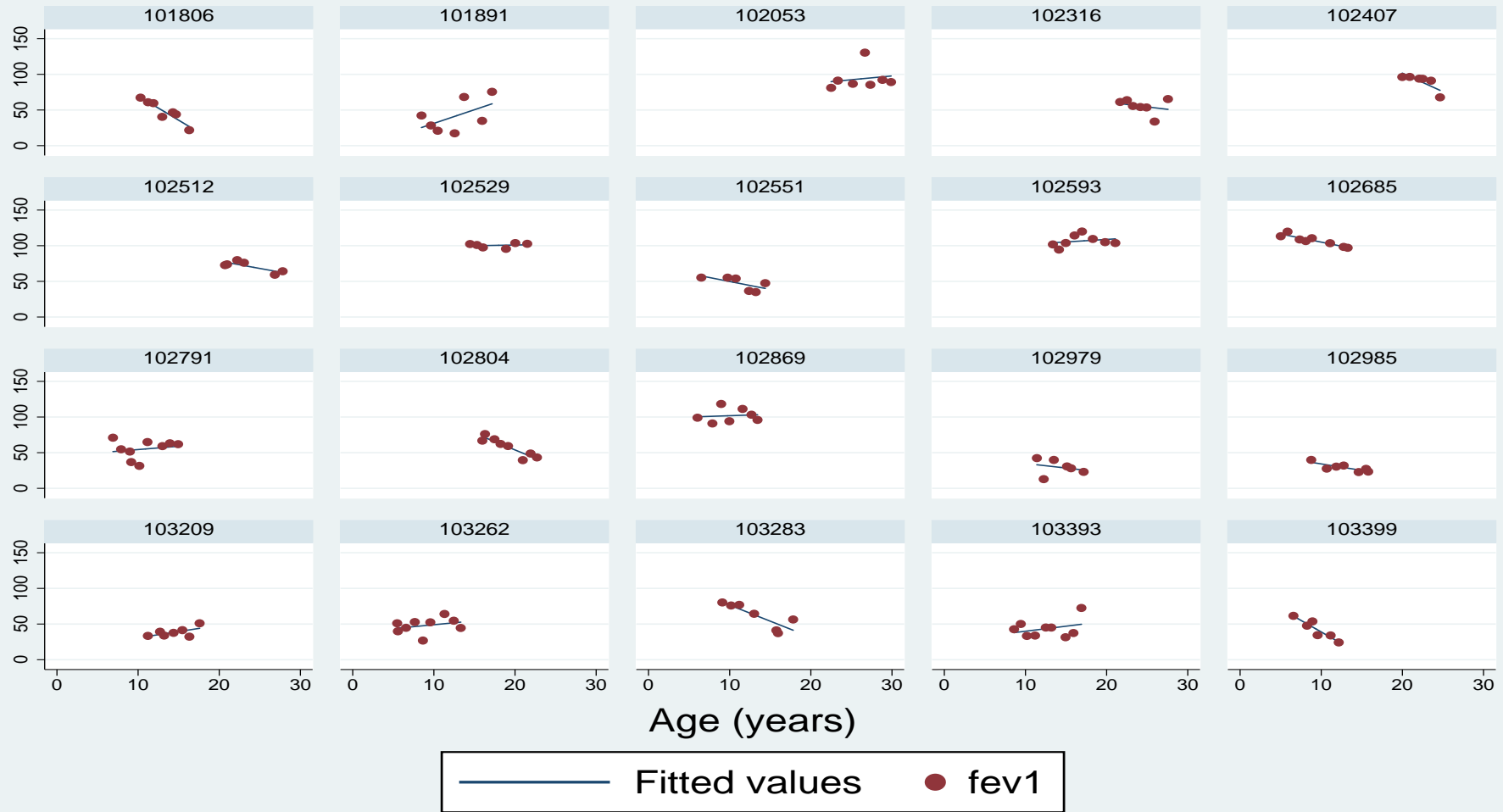
Results: CF Data – Individual Series



Results: CF Data – Fitted Lines



Results: CF Data – Fitted Lines



Graphs by id

Derived Variable Analysis

- Create Summary for Each Subject
 - ▷ Mean outcome
 - ▷ Fitted Slope
- Analysis of Summaries
 - ▷ t-test or ANOVA for summary (e.g. slope)
 - ▷ Regression for summary (e.g. slope)
- Limitations
 - ▷ Different number of observations?
 - ▷ Different times of observation?
 - ▷ Additional adjustment?

Example: CF Data

```
*****
* cfkids-derived.do *
*****
* *
* PURPOSE: analysis of FEV1 among CF kids *
* *
* AUTHOR: P. Heagerty *
* *
* DATE: 31 March 2005 *
*****

use NewCFkids

***
*** save kid-level data
```

```
***
```

```
collapse (mean) f508 gender age0, by(id)
```

```
save KidData, replace
```

```
clear
```

```
***
```

```
*** fit a line to each child's data
```

```
***
```

```
use NewCFkids
```

```
sort id
```

```
**** can't estimate line if less than 2 observations
```

```
by id: drop if _N < 2

statsby "reg fev1 ageL" _b, by(id)

list in 1/10

***** merge back with the KidData

merge id using KidData

*****
***** summarize b_ageL and b_cons by covariates
*****

sort gender
by gender: summarize b_ageL
ttest b_ageL, by(gender)
```

```
graph box b_ageL, over(gender)
```

```
sort f508
```

```
by f508: summarize b_ageL
```

```
xi: reg b_ageL i.f508
```

```
graph box b_ageL, over(f508)
```

Results: CF Data

```
. statsby "reg fev1 ageL" _b, by(id)
```

```
command:      reg fev1 ageL
statistics:   b_ageL      = _b[ageL]
              b_cons      = _b[_cons]
by:          id
```

```
.
```

```
. list in 1/10
```

```
+-----+
|      id      b_ageL      b_cons |
+-----+
1. | 100073    -1.936269    105.1512 |
2. | 100111     1.245856     97.19055 |
3. | 100185    -3.641983     48.351 |
4. | 100329    -1.794132     55.53651 |
```

Results: CF Data

-> gender = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
b_ageL	102	-1.212864	2.703113	-9.30644	4.765747

-> gender = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
b_ageL	98	-1.945086	3.063279	-14.32209	8.180941

Results: CF Data

```
. ttest b_ageL, by(gender)
```

```
Two-sample t test with equal variances
```

```
-----+-----  
Group |      Obs      Mean  Std. Err.   Std. Dev. [95% Conf. Int.]  
-----+-----  
      0 |      102  -1.2128   .2676     2.7031   -1.7438   -0.6819  
      1 |       98  -1.9450   .3094     3.0632   -2.5592   -1.3309  
-----+-----  
combined |      200  -1.5716   .2051     2.9012   -1.9761   -1.1671  
-----+-----  
diff |           .73222   .4081           -0.0725   1.5370  
-----+-----
```

```
Degrees of freedom: 198
```

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0

t = 1.7942

P < t = 0.9628

Ha: diff != 0

t = 1.7942

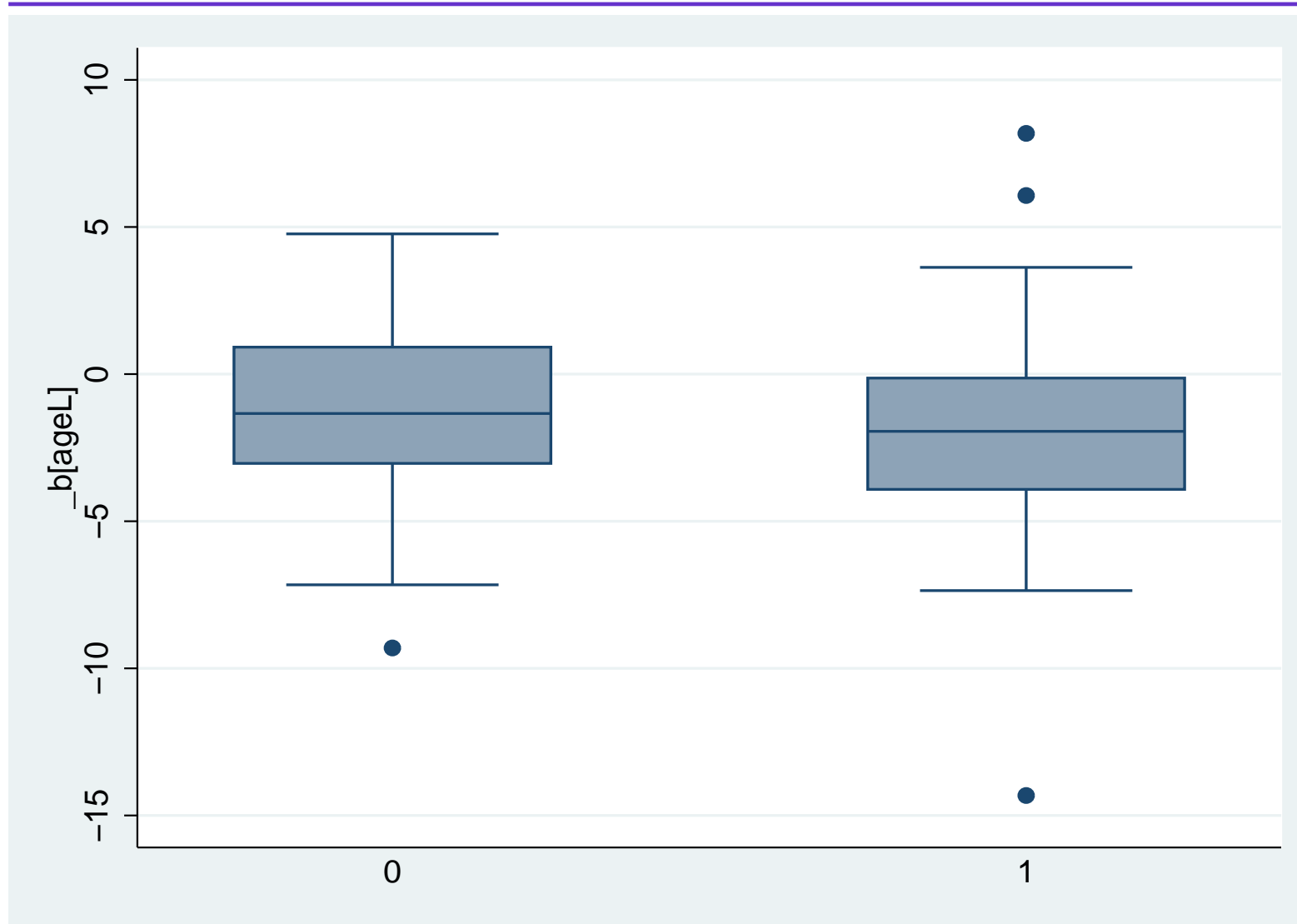
P > |t| = 0.0743

Ha: diff > 0

t = 1.7942

P > t = 0.0372

Results: CF Data – Slope Analysis



Results: CF Data

-> f508 = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
b_ageL	23	-.9898239	2.370343	-5.179185	3.831033

-> f508 = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
b_ageL	87	-1.482391	2.953235	-9.30644	6.069445

Results: CF Data

-> f508 = 2

Variable	Obs	Mean	Std. Dev.	Min	Max
b_ageL	90	-1.806628	2.975346	-14.32209	8.180941

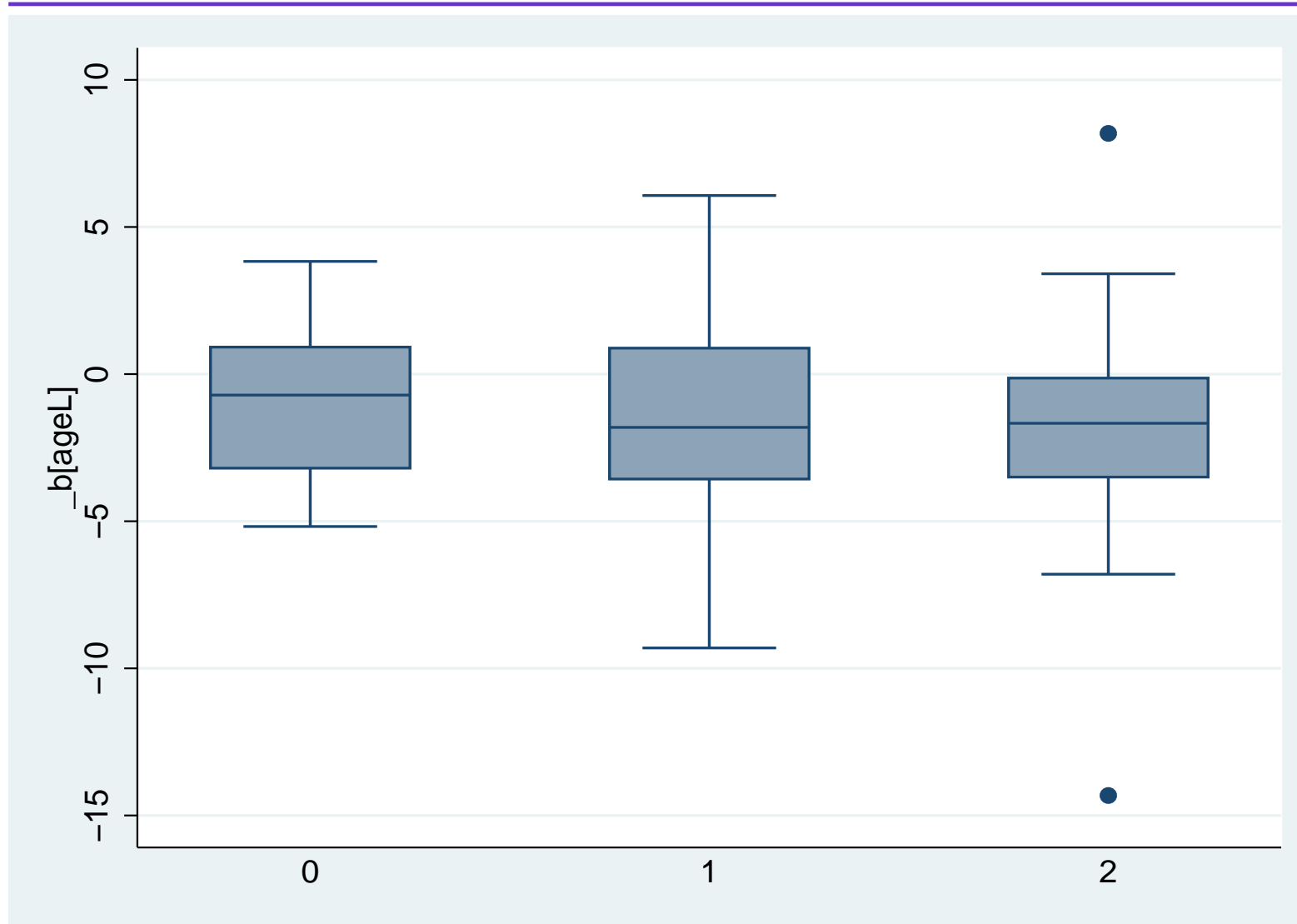
Results: CF Data

```
. xi: reg b_ageL i.f508
```

Source	SS	df	MS	Number of obs =	200
<hr style="border-top: 1px dashed black;"/>					
Model	13.4484747	2	6.72423734	F(2, 197) =	0.80
Residual	1661.55371	197	8.43428277	Prob > F =	0.4520
<hr style="border-top: 1px dashed black;"/>					
Total	1675.00218	199	8.41709639	R-squared =	0.0080
<hr style="border-top: 1px dashed black;"/>					
				Adj R-squared =	-0.0020
				Root MSE =	2.9042

b_ageL	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
<hr style="border-top: 1px dashed black;"/>						
_If508_1	-0.4925	.6809	-0.72	0.470	-1.835	.8502
_If508_2	-0.8168	.6785	-1.20	0.230	-2.154	.5213
_cons	-0.9898	.6055	-1.63	0.104	-2.184	.2043
<hr style="border-top: 1px dashed black;"/>						

Results: CF Data – Slope Analysis



Regression Analysis

- Linear Mixed Model
 - ▷ Random intercept
 - ▷ Random intercept and slope
- GEE
 - ▷ Independence correlation
 - ▷ Independence correlation with robust s.e.
 - ▷ Exchangeable correlation
 - ▷ Exchangeable correlation with robust s.e.

Regression Results: CF Data

```
*****
* cfkids-lmm-gee.do *
*****
* *
* PURPOSE: regression analysis of FEV1 among CF kids *
* *
* AUTHOR: P. Heagerty *
* *
* DATE: 05 April 2005 *
* 04 April 2006 *
*****
```

```
use NewCFkids
```

```
***
```

```
*** create the interactions and dummy variables
```

```
***
```

```
gen f508_1 = (f508==1)
```

```
gen f508_2 = (f508==2)
```

```
gen ageXf508_1 = f508_1 * ageL
```

```
gen ageXf508_2 = f508_2 * ageL
```

```
gen ageXgender = gender * ageL
```

```
*****
```

```
***** Linear Mixed Model Analysis
```

```
*****
```

```
*****
```

```
*** analysis using GLLAMM (old; slow; general)
```

```
***
```



```
*****
```

```
gen cons=1
```

```
eq int: cons
```

```
eq slope: ageL
```

```
set more off
```

```
*** Model 1: Random Intercepts
```

```
gllamm fev1 age0 ageL gender f508_1 f508_2 ///
```

```
    ageXgender ageXf508_1 ageXf508_2, ///
```

```
    i(id) eqs(int) nrf(1) nip(10) adapt
```

```
*** Model 2: Random Intercepts and Slopes
```

```
gllamm fev1 age0 ageL gender f508_1 f508_2 ///
```

```

ageXgender ageXf508_1 ageXf508_2,                                     ///
                    i(id) eqs(int slope) nrf(2) nip(10) adapt

*****
***  analysis using XTMIXED                                     ***
*****

*** Model 1:  Random Intercepts

xtmixed fev1 age0 ageL gender f508_1 f508_2                       ///
        ageXgender ageXf508_1 ageXf508_2      || id:

*** Model 2:  Random Intercepts and Slopes

xtmixed fev1 age0 ageL gender f508_1 f508_2                       ///
        ageXgender ageXf508_1 ageXf508_2      || id: ageL, cov(un)

```

***** Generalized Estimating Equations Analysis

```
xtgee fev1 age0 ageL gender f508_1 f508_2          ///
      ageXgender ageXf508_1 ageXf508_2,          ///
      corr(independent) i(id)
```

xtcorr

```
xtgee fev1 age0 ageL gender f508_1 f508_2          ///
      ageXgender ageXf508_1 ageXf508_2,          ///
      corr(independent) i(id) robust
```

```
xtgee fev1 age0 ageL gender f508_1 f508_2          ///
      ageXgender ageXf508_1 ageXf508_2,          ///
```

```
corr(exchangeable) i(id)
```

```
xtcorr
```

```
xtgee fev1 age0 ageL gender f508_1 f508_2          ///  
      ageXgender ageXf508_1 ageXf508_2,          ///  
      corr(exchangeable) i(id) robust
```

Regression Results: CF Data – Random Intercepts

```
. gllamm fev1 age0 ageL gender f508_1 f508_2 ageXgender ageXf508_1 ageXf508_2  
      i(id) eqs(int) nrf(1) nip(10) adapt
```

number of level 1 units = 1513 number of level 2 units = 200

log likelihood = -6255.0053

fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.8553	0.3300	-5.62	0.000	-2.5023	-1.2083
ageL	-0.5881	0.3924	-1.50	0.134	-1.3573	0.1810
gender	-1.1620	3.3575	-0.35	0.729	-7.7428	5.4187
f508_1	-4.2822	5.5447	-0.77	0.440	-15.1498	6.5852
f508_2	-6.7428	5.5759	-1.21	0.227	-17.6714	4.1857
ageXgender	-0.8257	0.2490	-3.32	0.001	-1.3138	-0.3376

ageXf508_1		-0.4873	0.4218	-1.16	0.248	-1.3140	0.3394
ageXf508_2		-0.6568	0.4207	-1.56	0.118	-1.4814	0.1677
_cons		103.8077	6.6248	15.67	0.000	90.8231	116.7922

Variance at level 1

148.32341 (5.7896206)

Variances and covariances of random effects

***level 2 (id)

var(1): 497.48033 (51.806823)

Regression Results: CF Data – Random Int & Slopes

```
. gllamm fev1 age0 ageL gender f508_1 f508_2 ageXgender ageXf508_1 ageXf508_2
      i(id) eqs(int slope) nrf(2) nip(10) adapt
```

number of level 1 units = 1513 number of level 2 units = 200

log likelihood = -6202.1683

fev1		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age0		-1.910407	.3283248	-5.82	0.000	-2.553912	-1.26690
ageL		-.6018747	.5839174	-1.03	0.303	-1.746332	.542582
gender		-1.296872	3.328747	-0.39	0.697	-7.821096	5.22735
f508_1		-4.238208	5.495178	-0.77	0.441	-15.00856	6.53214
f508_2		-6.655038	5.525834	-1.20	0.228	-17.48547	4.17539
ageXgender		-.7636477	.3770428	-2.03	0.043	-1.502638	-.024657

ageXf508_1		-.5002841	.6288493	-0.80	0.426	-1.732806	.732237
ageXf508_2		-.7450708	.6276375	-1.19	0.235	-1.975218	.48507
_cons		104.5162	6.572765	15.90	0.000	91.63378	117.398

Variance at level 1

118.05247 (5.0252714)

Variances and covariances of random effects

***level 2 (id)

var(1): 498.65795 (54.793715)

cov(2,1): -7.2774328 (4.6057063) cor(2,1): -.15612915

var(2): 4.3569809 (.73837017)

Regression Results: CF Data – Random Intercepts

```
. xtmixed fev1 age0 ageL gender f508_1 f508_2 ///
      ageXgender ageXf508_1 ageXf508_2      || id:
```

```
Mixed-effects REML regression      Number of obs      =      1513
Group variable: id                  Number of groups    =       200
```

Log restricted-likelihood = -6249.041

fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.855	0.3343	-5.55	0.000	-2.510	-1.200
ageL	-0.587	0.3930	-1.50	0.135	-1.358	0.182
gender	-1.162	3.3978	-0.34	0.732	-7.821	5.497
f508_1	-4.280	5.6110	-0.76	0.445	-15.27	6.716
f508_2	-6.740	5.6425	-1.19	0.232	-17.79	4.318

ageXgender		-0.825	0.2494	-3.31	0.001	-1.314	-0.336
ageXf508_1		-0.487	0.4224	-1.15	0.248	-1.315	0.340
ageXf508_2		-0.657	0.4213	-1.56	0.119	-1.483	0.168
_cons		103.806	6.7060	15.48	0.000	90.662	116.949

Regression Results: CF Data – Random Intercepts

```
-----  
Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]  
-----+-----  
id: Identity |  
      sd(_cons) | 22.599 1.1906 20.382 25.057  
-----+-----  
      sd(Residual) | 12.197 0.2384 11.738 12.673  
-----  
LR test vs. linear regression: chibar2(01) = 1554.17 Prob >= chibar2 = 0.0
```

Regression Results: CF Data – Random Int & Slopes

```
. xtmixed fev1 age0 ageL gender f508_1 f508_2 ///  
      ageXgender ageXf508_1 ageXf508_2      || id: ageL, cov(un)
```

```
Mixed-effects REML regression      Number of obs      =      1513  
Group variable: id                 Number of groups   =      200
```

Log restricted-likelihood = -6194.5865

fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.910	0.3310	-5.77	0.000	-2.559	-1.261
ageL	-0.602	0.5902	-1.02	0.307	-1.759	0.554
gender	-1.300	3.3701	-0.39	0.700	-7.905	5.304
f508_1	-4.238	5.5636	-0.76	0.446	-15.14	6.666
f508_2	-6.652	5.5944	-1.19	0.234	-17.61	4.312

ageXgender		-0.762	0.3812	-2.00	0.045	-1.509	-0.015
ageXf508_1		-0.500	0.6357	-0.79	0.431	-1.746	0.745
ageXf508_2		-0.745	0.6344	-1.18	0.240	-1.989	0.497
_cons		104.519	6.6443	15.73	0.000	91.496	117.541

Regression Results: CF Data – Random Int & Slopes

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
-----+				
id: Unstructured				
sd(ageL)	2.124	0.1794	1.800	2.506
sd(_cons)	22.636	1.2560	20.303	25.237
corr(ageL, _cons)	-0.158	0.0921	-0.332	0.025
-----+				
sd(Residual)	10.863	0.2312	10.419	11.326

LR test vs. linear regression: $\chi^2(3) = 1663.08$ Prob > $\chi^2 = 0.00$

Note: LR test is conservative and provided only for reference

Regression Results: CF Data – GEE Independence

```
. xtgee fev1 age0 ageL gender f508_1 f508_2 ageXgender ageXf508_1 ageXf508_2,  
      corr(independent) i(id)
```

GEE population-averaged model		Number of obs	=	1513
Group variable:	id	Number of groups	=	200
Link:	identity	Obs per group: min	=	6
Family:	Gaussian	avg	=	7.6
Correlation:	independent	max	=	9
		Wald chi2(8)	=	248.46
Scale parameter:	636.1352	Prob > chi2	=	0.0000
Pearson chi2(1513):	962472.50	Deviance	=	962472.50
Dispersion (Pearson):	636.1352	Dispersion	=	636.1352

fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.8614	0.1358	-13.71	0.000	-2.1276	-1.5952
ageL	-0.8910	0.7907	-1.13	0.260	-2.4409	0.6589
gender	-1.1329	2.2403	-0.51	0.613	-5.5239	3.2580
f508_1	-5.2795	3.7594	-1.40	0.160	-12.6478	2.0887
f508_2	-8.7756	3.7699	-2.33	0.020	-16.1646	-1.3865
ageXgender	-0.8187	0.5029	-1.63	0.104	-1.8045	0.1670
ageXf508_1	-0.1623	0.8505	-0.19	0.849	-1.8295	1.5047
ageXf508_2	-0.0393	0.8478	-0.05	0.963	-1.7009	1.6223
_cons	104.8027	3.9223	26.72	0.000	97.1151	112.4903


```
. xtcorr
```

Estimated within-id correlation matrix R:

	c1	c2	c3	c4	c5	c6	c7	c8	c9
r1	1.0000								
r2	0.0000	1.0000							
r3	0.0000	0.0000	1.0000						
r4	0.0000	0.0000	0.0000	1.0000					
r5	0.0000	0.0000	0.0000	0.0000	1.0000				
r6	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000			
r7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000		
r8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	
r9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Regression Results: CF Data – GEE Independence

```
. xtgee fev1 age0 ageL gender f508_1 f508_2 ageXgender ageXf508_1 ageXf508_2,  
      corr(independent) i(id) robust
```

GEE population-averaged model		Number of obs	=	1513
Group variable:	id	Number of groups	=	200
Link:	identity	Obs per group: min	=	6
Family:	Gaussian	avg	=	7.6
Correlation:	independent	max	=	9
		Wald chi2(8)	=	70.80
Scale parameter:	636.1352	Prob > chi2	=	0.0000
Pearson chi2(1513):	962472.50	Deviance	=	962472.50
Dispersion (Pearson):	636.1352	Dispersion	=	636.1352

(standard errors adjusted for clustering on id)

	Semi-robust					
fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.8614	0.3188	-5.84	0.000	-2.4864	-1.2364
ageL	-0.8910	0.6249	-1.43	0.154	-2.1159	0.3339
gender	-1.1329	3.4447	-0.33	0.742	-7.8845	5.6185
f508_1	-5.2795	6.0582	-0.87	0.384	-17.1536	6.5944
f508_2	-8.7756	6.1837	-1.42	0.156	-20.8955	3.3442
ageXgender	-0.8187	0.4545	-1.80	0.072	-1.7096	0.0721
ageXf508_1	-0.1623	0.6758	-0.24	0.810	-1.4870	1.1622
ageXf508_2	-0.0393	0.7010	-0.06	0.955	-1.4133	1.3347
_cons	104.8027	7.3461	14.27	0.000	90.4045	119.2009

Regression Results: CF Data – GEE Exchangeable

```
. xtgee fev1 age0 ageL gender f508_1 f508_2 ageXgender ageXf508_1 ageXf508_2,  
      corr(exchangeable) i(id)
```

GEE population-averaged model		Number of obs	=	1513
Group variable:	id	Number of groups	=	200
Link:	identity	Obs per group: min	=	6
Family:	Gaussian	avg	=	7.6
Correlation:	exchangeable	max	=	9
		Wald chi2(8)	=	185.56
Scale parameter:	636.5125	Prob > chi2	=	0.0000

fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.8553	0.3243	-5.72	0.000	-2.4911	-1.2195
ageL	-0.5896	0.4052	-1.46	0.146	-1.3838	0.2045
gender	-1.1619	3.3113	-0.35	0.726	-7.6519	5.3281
f508_1	-4.2875	5.4695	-0.78	0.433	-15.0076	6.4325
f508_2	-6.7531	5.5001	-1.23	0.220	-17.5332	4.0269
ageXgender	-0.8257	0.2571	-3.21	0.001	-1.3297	-0.3217
ageXf508_1	-0.4857	0.4355	-1.12	0.265	-1.3393	0.3679
ageXf508_2	-0.6537	0.4344	-1.50	0.132	-1.5051	0.1976
_cons	103.8126	6.5258	15.91	0.000	91.0221	116.6031

```
.  
. xtcorr
```

Estimated within-id correlation matrix R:

	c1	c2	c3	c4	c5	c6	c7	c8	c9
r1	1.0000								
r2	0.7515	1.0000							
r3	0.7515	0.7515	1.0000						
r4	0.7515	0.7515	0.7515	1.0000					
r5	0.7515	0.7515	0.7515	0.7515	1.0000				
r6	0.7515	0.7515	0.7515	0.7515	0.7515	1.0000			
r7	0.7515	0.7515	0.7515	0.7515	0.7515	0.7515	1.0000		
r8	0.7515	0.7515	0.7515	0.7515	0.7515	0.7515	0.7515	1.0000	
r9	0.7515	0.7515	0.7515	0.7515	0.7515	0.7515	0.7515	0.7515	1.0000

Regression Results: CF Data – GEE Exchangeable

```
. xtgee fev1 age0 ageL gender f508_1 f508_2 ageXgender ageXf508_1 ageXf508_2,  
      corr(exchangeable) i(id) robust
```

GEE population-averaged model		Number of obs	=	1513
Group variable:	id	Number of groups	=	200
Link:	identity	Obs per group: min	=	6
Family:	Gaussian	avg	=	7.6
Correlation:	exchangeable	max	=	9
		Wald chi2(8)	=	97.25
Scale parameter:	636.5125	Prob > chi2	=	0.0000

(standard errors adjusted for clustering on id)

	Semi-robust					
fev1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age0	-1.8553	0.3209	-5.78	0.000	-2.4844	-1.2261
ageL	-0.5896	0.5245	-1.12	0.261	-1.6176	0.4383
gender	-1.1619	3.3405	-0.35	0.728	-7.7093	5.3855
f508_1	-4.2875	5.9070	-0.73	0.468	-15.8651	7.2900
f508_2	-6.7531	5.9982	-1.13	0.260	-18.5095	5.0032
ageXgender	-0.8257	0.3715	-2.22	0.026	-1.5539	-0.0975
ageXf508_1	-0.4857	0.5798	-0.84	0.402	-1.6221	0.6507
ageXf508_2	-0.6537	0.5693	-1.15	0.251	-1.7697	0.4621
_cons	103.8126	7.2583	14.30	0.000	89.5865	118.0387

Summary – Regression Analysis

- Linear Mixed Models
 - ▷ Random effects used to model the subject-to-subject heterogeneity in trajectories.
 - ▷ Random effects model implies specific covariance forms (correlation).
 - ▷ Chosen model for random effects (between-subject) and the within-subject errors is used to:
 - * Weight observations in order to estimate the regression parameters.
 - * Provide standard errors for the regression parameters based on the implied correlation and variance model.

Summary – Regression Analysis

- GEE Analysis
 - ▷ The user directly chooses the correlation model.
 - ▷ Chosen model for correlation is used to:
 - * Weight observations in order to estimate the regression parameters.
 - * Provide standard errors for the regression parameters based on the assumed correlation and variance model.
 - ▷ **In addition**, a standard error estimate is calculated (if requested in STATA using `robust`) that is valid (correct) even if a poor correlation model was selected.