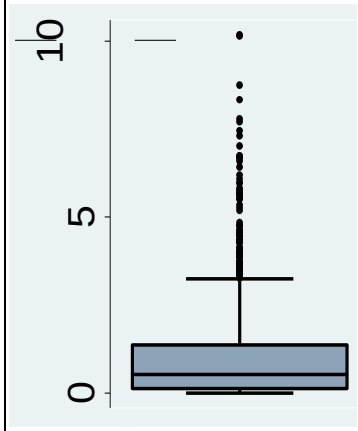
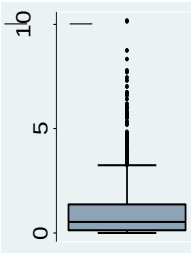
Repeated Measures, Part I

Charles E. McCulloch,
Division of Biostatistics,
Dept of Epidemiology and Biostatistics,
UCSF



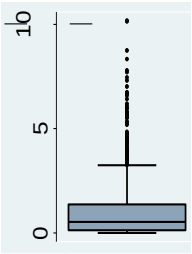
April, 2015

Outline



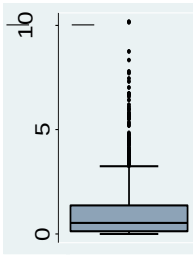
1. Motivating examples and introduction
2. Analysis of the fecal fat data
3. Accommodating correlated data
4. Correlation structures
5. Long and wide data formats
6. Descriptive methods for longitudinal data
7. Summary

Example: Fecal fat



Lack of digestive enzymes in the intestine can cause bowel absorption problems. This will be indicated by excess fat in the feces. Pancreatic enzyme supplements can be given to ameliorate the problem. Does the supplement form make a difference? (Graham, Enzyme replacement therapy of exocrine pancreatic insufficiency in man. *NEJM*, **296**: 1314-17, 1977 – But note: sex information made up for illustration.)

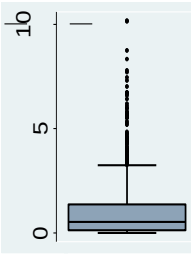
Example: Fecal fat



Considerable person to person variability

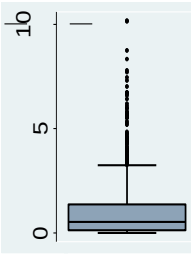
	Fecal Fat (g/day)				
	Pill type				
PatID / Sex	**None**	**Tablet**	**Capsule**	**Coated Capsule**	**Avg**
1 – M	44.5	7.3	3.4	12.4	16.900
2 – M	33.0	21.0	23.1	25.4	25.625
3 – M	19.1	5.0	11.8	22.0	14.475
4 – F	9.4	4.6	4.6	5.8	6.100
5 – F	71.3	23.3	25.6	68.2	47.100
6 – F	51.2	38.0	36.0	52.6	44.450
Avg	38.08	16.53	17.42	31.07	25.775

Example: Practice style and back pain (Korff, Barlow, Cherkin, and Deyo, 1994)



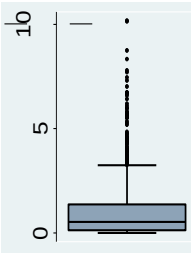
Forty-four primary care physicians in a large HMO were classified according to their practice style in treating back pain management (low, moderate or high frequency of prescription of pain medication and bed rest). An average of 24 patients per physician was followed for 2 years (1 month, 1 year and 2 year followups) after the indexed visit. Outcome measures included functional measures (pain intensity, activity limitation days, etc.), patient satisfaction (e.g., “After your visit with the doctor, you fully understood how to take care of your back problem”), and cost.

Example: Osteoarthritis Initiative (OAI): www.oai.ucsf.edu



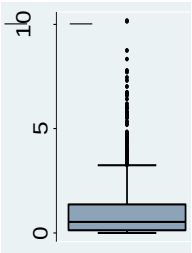
The OAI is a multi-center, longitudinal, prospective observational study of knee osteoarthritis (OA). 4,796 men and women ages 45-79 were enrolled between 2004 and 2006. Image (X-ray and MRI), demographic, and clinical data are being collected yearly. Some of the variables measured over time are: pain scores (one for each knee), presence of OA as judged from X-ray (one for each knee), functional limitation scores (one per person).

Example: Study of Osteoporotic Fractures (SOF): sof.ucsf.edu



The Study of Osteoporotic Fractures (SOF) is a longitudinal, prospective study of osteoporosis, breast cancer, stroke, and total and cause-specific mortality. In 1986, SOF enrolled 9,704 women and continues to track these women with clinical visits every 2 years. Data from the first seven visits are now available to the public. The data include measures of bone mineral density (BMD), sex and calcitropic hormones, tests of strength and function, cognitive exams, use of medication, health habits and much more.

Introduction: Hierarchical data

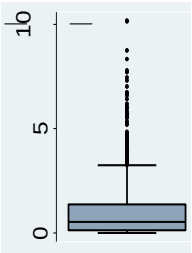


Hierarchical data (responses and/or predictors) are data collected from different levels within a study.

May be repeated measures data (e.g., fecal fat), clustered or multilevel data (e.g., back pain) or longitudinal data (over time, e.g., OAI or SOF).

A characteristic of hierarchical data is that predictors can be measured at any level in the hierarchy.

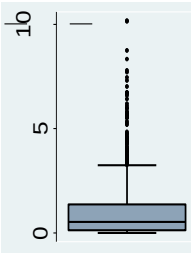
Some prototypical questions: Fecal fat example



Question 1: Does fecal fat depend on the repeated measures factor, pill type?

Question 2: Does fecal fat depend on the non-repeated measures factor, sex?

Some prototypical questions: Back pain example



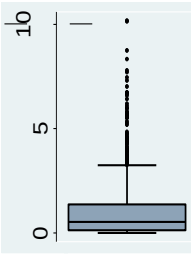
Question 1: Does log cost depend on the between physician factor, practice style?

Question 2: Does understanding of physician recommendation depend on practice style?

Question 3: Does log cost depend on the within physician, between patient factor, sex of the patient?

Question 4: Is there between physician variability in treatment of similar patients?

Some prototypical questions: SOF

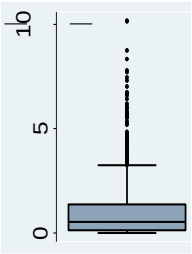


Question 1: Is change in BMD related to age at menopause? (time invariant predictor of change)

Question 2: Is change in BMD related to change in BMI? (time varying predictor of change)

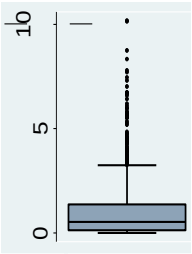
Question 3: Which participants are maintaining cognitive function into their 9^{th} and 10^{th} decades of life? (subject specific prediction)

Introduction



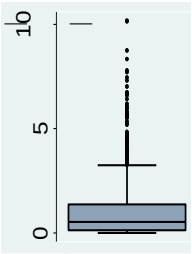
- Analysis technique depends on nature of the *outcome* variable and research question.
  - Binary: logistic regression (e.g., BMI>30)
    - Odds ratios, area under ROC curve
  - Numeric: linear regression (e.g., BMI, BMD)
  - Also – time to event (Cox model or pooled logistic regression), count outcomes (Poisson regression)
- Methods need to be modified for hierarchical data.

Accommodating hierarchical data



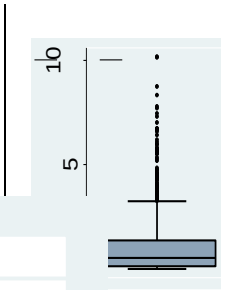
- Repeated measures/clustering in the *outcome* requires new methods of analysis. But not the predictor.
- SOF: Is visit 8 cognitive status (outcome) related to previous (repeatedly measured across multiple visits) physical activity? Does not have repeated measures on the outcome.
- This situation can be accommodated by including multiple values of physical activity as predictors or by calculating summary measure(s) (e.g., average physical activity).

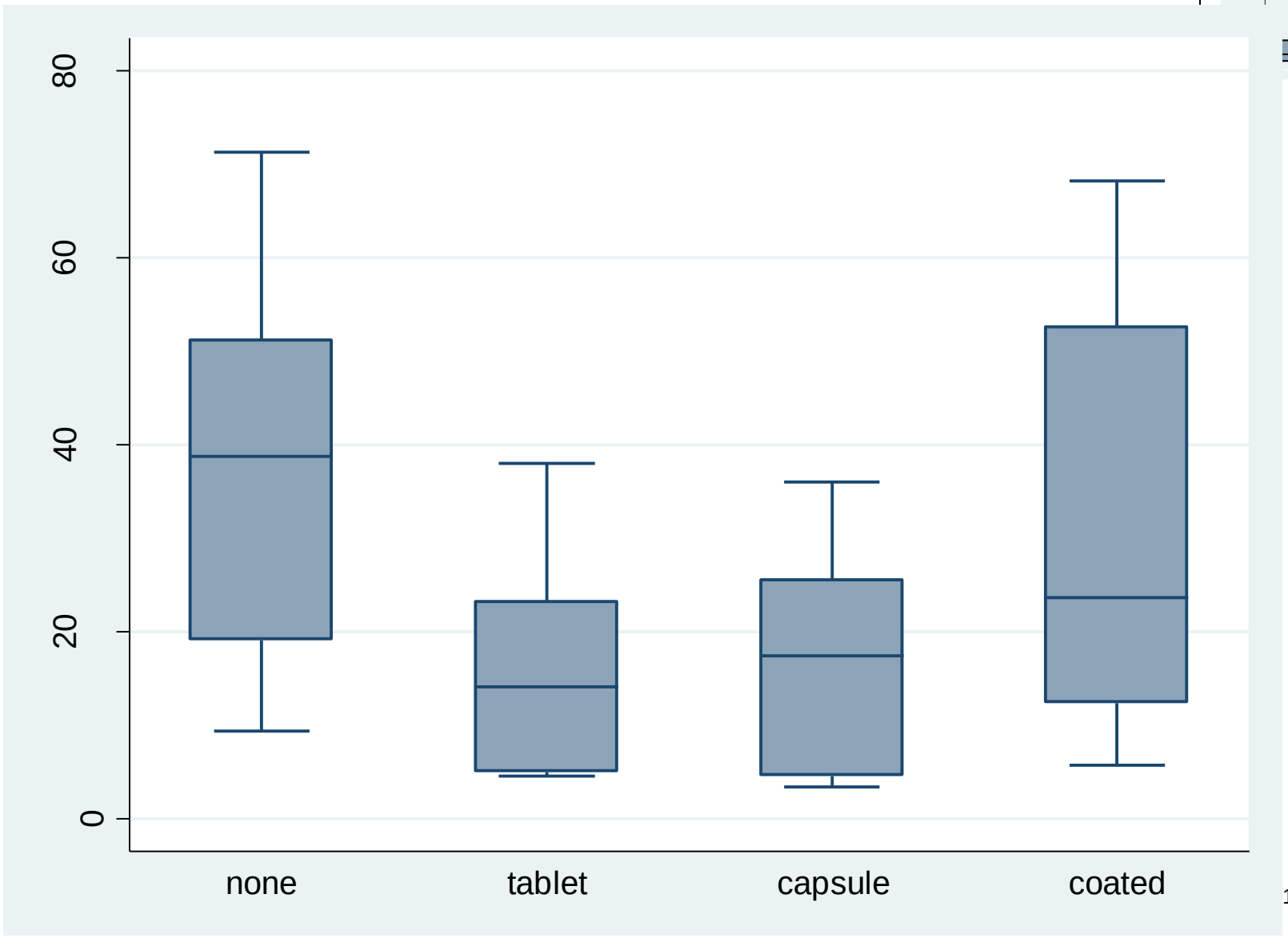
Fecal fat data analysis



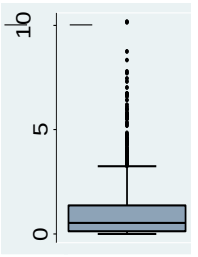

```
. bysort pilltype: summarize fecfat

-> pilltype=    none
Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+--------------------------------------------------------
  fecfat |       6    38.08333    22.47447         9.4       71.3
-> pilltype=  tablet
Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+--------------------------------------------------------
  fecfat |       6    16.53333    13.32091         4.6         38
-> pilltype= capsule
Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+--------------------------------------------------------
  fecfat |       6    17.41667    12.93745         3.4         36
-> pilltype=  coated
Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+--------------------------------------------------------
  fecfat |       6    31.06667     24.2641         5.8        68.2
```

Fecal fat data analysis





Regression/ANOVA ignoring sex effects (a wrong analysis)



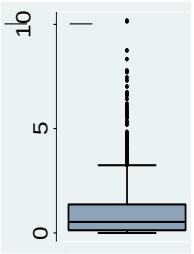

```
. regress fecfat i.pilltype

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  3,    20) =    1.86
       Model |   2008.6017     3  669.533901            Prob > F      =  0.1687
    Residual |  7193.36328    20  359.668164            R-squared     =  0.2183
-------------+------------------------------           Adj R-squared =  0.1010
       Total |  9201.96498    23  400.085434            Root MSE      =  18.965

------------------------------------------------------------------------------
      fecfat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pilltype |
          2  |     -21.55    10.9494    -1.97   0.063    -44.39005     1.29005
          3  |  -20.66667    10.9494    -1.89   0.074    -43.50672    2.173384
          4  |  -7.016668    10.9494    -0.64   0.529    -29.85672    15.82338
             |
       _cons |   38.08333   7.742396     4.92   0.000     21.93298    54.23369
------------------------------------------------------------------------------

. testparm i.pilltype
 ( 1)  2.pilltype = 0
 ( 2)  3.pilltype = 0
 ( 3)  4.pilltype = 0
       F(  3,    20) =    1.86
            Prob > F =    0.1687
```

Difference between average for Tablet – None

(16.53 – 38.08 = -21.55)

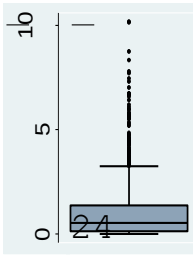
Overall test, p=0.17

A hierarchical analysis




```
. xtgee fecfat i.pilltype, i(patid)

GEE population-averaged model                   Number of obs      =        24
Group variable:                      patid      Number of groups   =         6
Link:                              identity     Obs per group: min =         4
Family:                            Gaussian                    avg =       4.0
Correlation:                   exchangeable                    max =         4
                                                Wald chi2(3)       =     22.53
Scale parameter:                   299.7235     Prob > chi2        =    0.0001
------------------------------------------------------------------------------
      fecfat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pilltype |
           2 |     -21.55   5.451781    -3.95   0.000    -32.23529   -10.86471
           3 |  -20.66667   5.451781    -3.79   0.000    -31.35196   -9.981373
           4 |  -7.016668   5.451781    -1.29   0.198    -17.70196    3.668626
             |
       _cons |   38.08333   7.067808     5.39   0.000     24.23068    51.93598
------------------------------------------------------------------------------
. testparm i.pilltype
 ( 1)  2.pilltype = 0
 ( 2)  3.pilltype = 0
 ( 3)  4.pilltype = 0
           chi2(  3) =   22.53
         Prob > chi2 =    0.0001
```

A hierarchical analysis (variation)

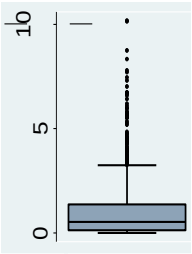



```
. xtgee fecfat i.pilltype, i(patid) robust

GEE population-averaged model                   Number of obs      =        24
Group variable:                      patid      Number of groups   =         6
Link:                             identity      Obs per group: min =         4
Family:                           Gaussian                     avg =       4.0
Correlation:                  exchangeable                     max =         4
                                                Wald chi2(3)       =     11.71
Scale parameter:                  299.7235      Prob > chi2        =    0.0084
                                                (Std. Err. adjusted for clustering on patid)
------------------------------------------------------------------------------
             |               Robust
      fecfat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pilltype |
           2 |     -21.55   6.931847    -3.11   0.002    -35.13617    -7.96383
           3 |  -20.66667   7.349407    -2.81   0.005    -35.07124   -6.262094
           4 |  -7.016668   5.246295    -1.34   0.181    -17.29922    3.265881
             |
       _cons |   38.08333   9.175163     4.15   0.000     20.10034    56.06632
------------------------------------------------------------------------------
```

Coefficient unchanged, but SE is slightly different (6.93 versus 5.45 without robust)

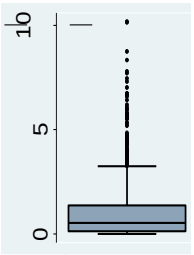
```
. testparm i.pilltype
 ( 1)  2.pilltype = 0
 ( 2)  3.pilltype = 0
 ( 3)  4.pilltype = 0
           chi2(  3) =   11.71
         Prob > chi2 =    0.0084
```

Accommodating hierarchical data



- The usual statistical methods (multiple regression, basic ANOVA, logistic regression, and many others) assume observations are independent.
- *Important idea*: observations taken within the same subgroup in a hierarchy are often more similar to one another than to observations in different subgroups, other things being equal. [correlated]
- Failure to accommodate the hierarchical nature of the data can lead to incorrect SEs, p-values and confidence intervals, sometimes grossly incorrect.

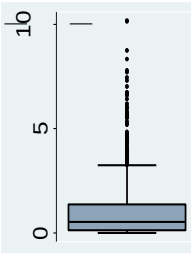
Regr/ANOVA with sex effects (incorrect analysis)




```
. regress fecfat i.pilltype i.sex

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  4,    19) =    2.43
       Model |  3110.21668     4  777.554169            Prob > F      =  0.0837
    Residual |   6091.7483    19  320.618332            R-squared     =  0.3380
-------------+------------------------------           Adj R-squared =  0.1986
       Total |  9201.96498    23  400.085434            Root MSE      =  17.906

------------------------------------------------------------------------------
      fecfat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pilltype |
           2 |     -21.55   10.33793    -2.08   0.051    -43.18753    .0875334
           3 |  -20.66667   10.33793    -2.00   0.060     -42.3042    .9708671
           4 |  -7.016668   10.33793    -0.68   0.505     -28.6542    14.62087
             |
       1.sex |       13.55    7.31002     1.85   0.079    -1.750047    28.85005
       _cons |   31.30833   8.172851     3.83   0.001     14.20236    48.41431
------------------------------------------------------------------------------
```

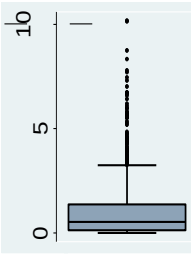
Sex effects are borderline statistically significant

A hierarchical analysis



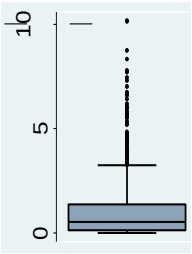

```
. xtgee fecfat i.pilltype i.sex, i(patid)

Iteration 1: tolerance = 9.971e-16

GEE population-averaged model                   Number of obs      =         24
Group variable:                      patid      Number of groups   =          6
Link:                              identity     Obs per group: min =          4
Family:                          Gaussian                   avg =        4.0
Correlation:                 exchangeable                   max =          4
                                                Wald chi2(4)       =      24.00
Scale parameter:                  253.8228      Prob > chi2        =     0.0001

------------------------------------------------------------------------------
      fecfat |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pilltype |
           2 |     -21.55   5.451781    -3.95   0.000    -32.23529   -10.86471
           3 |  -20.66667   5.451781    -3.79   0.000    -31.35196   -9.981373
           4 |  -7.016668   5.451781    -1.29   0.198    -17.70196    3.668626
             |
       1.sex |       13.55   11.16389     1.21   0.225    -8.330816    35.43082
       _cons |   31.30833   8.570992     3.65   0.000      14.5095    48.10717
------------------------------------------------------------------------------
```

Fecal fat data analysis - summary



- Failing to accommodate the hierarchical nature of the data led to grossly incorrect statistical inferences.
- Estimates were unchanged, but SEs were affected. (In general estimates tend to be little affected, but do change slightly).
- The proper hierarchical analysis can lead to smaller or larger SEs compared to a naïve analysis.

Not just an academic concern …



J Assist Reprod Genet
DOI 10.1007/s10815-010-9518-0

EMBRYO BIOLOGY

A prospective randomized sibling-oocyte study of two media systems for culturing cleavage-stage embryos—impact on fertilization rate

Received: 30 July 2010 / Accepted: 19 November 2010

Not just an academic concern …

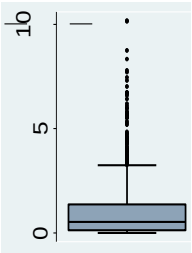
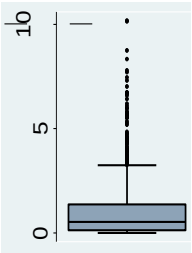


Table 2 Demographic data of patients included in the study is shown

Patients included (*n*)	110
Female age, mean±SD (years)	33.9±3.8
Infertility diagnosis, *n* (%)	
Tubal factor	8 (7.3)
Endometriosis	8 (7.3)
Unexplained	30 (27.3)
Anovulation	12 (10.9)
Male factor	37 (33.6)
Other	15 (13.6)
No. of oocytes retrieved per patient, mean±SD	10.9±4.8

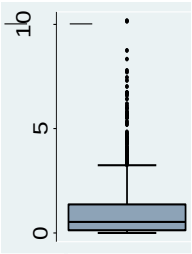
Not just an academic concern …



Statistical analysis

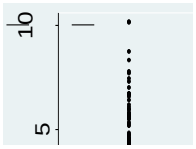
at least 269 oocytes in each group. Comparison of normal fertilisation rates, polyploid rates, embryo quality and embryo utilization rates was performed using the Chi-square test. Positive hCG rates, clinical pregnancy rates, implantation rates and delivery rates were compared by using the Fisher's exact test (two-tailed). A value of $p<0.05$ was considered statistically significant.

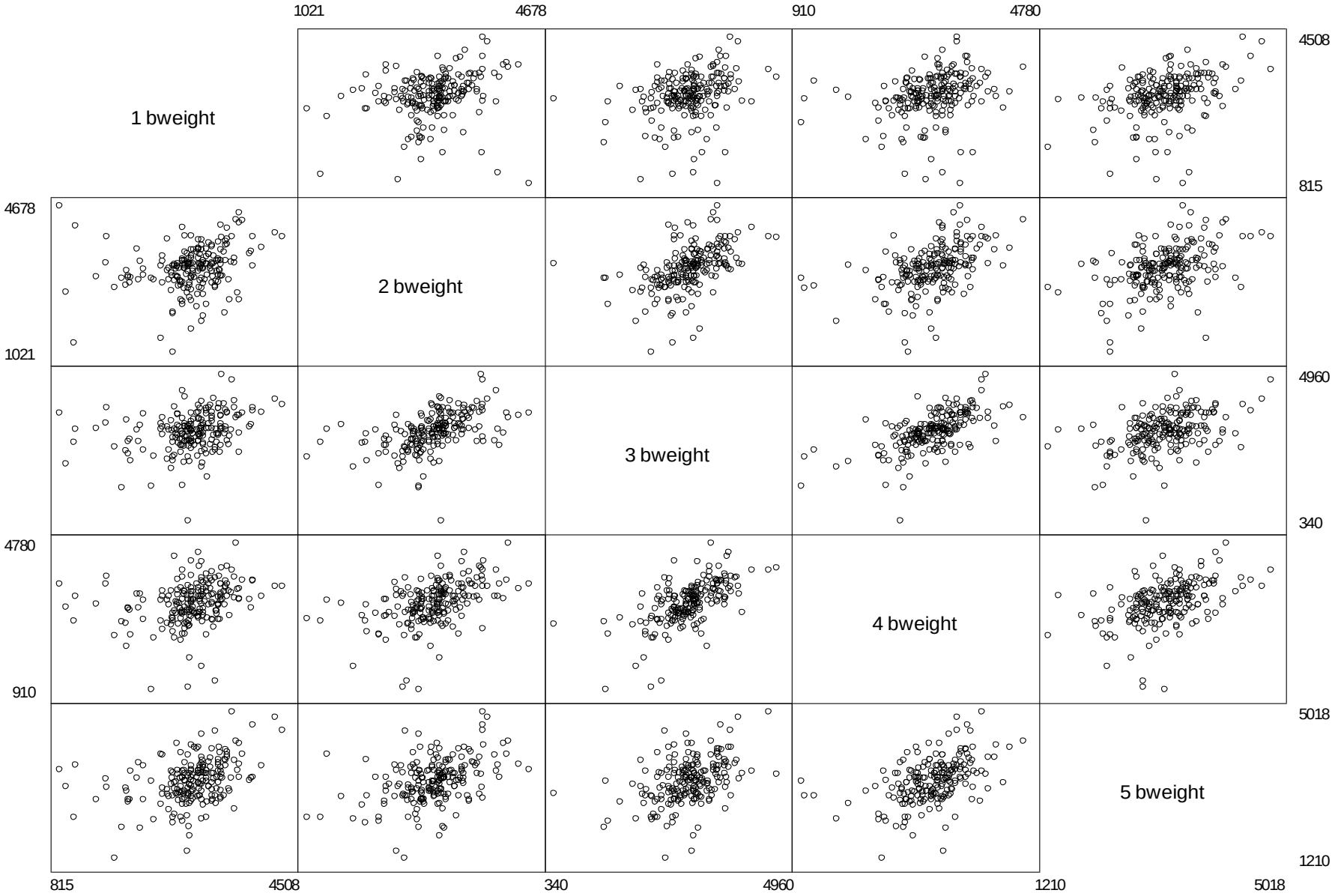
Correlation structures



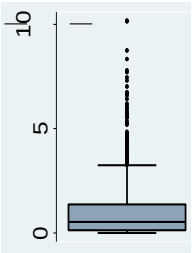
With continuous, balanced data we can plot the measurements that occur at different time points (or are repeated measurements of different types). The “Georgia babies” dataset follows successive birthweights of infants to mothers (each of whom had five children) from vital statistics in Georgia. We are interested in whether birthweight increases with birth order and mothers’ age. In lab we will generate the following plots.

Georgia Babies





Georgia Babies

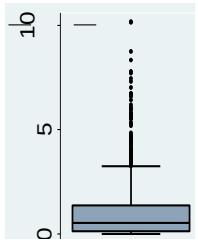


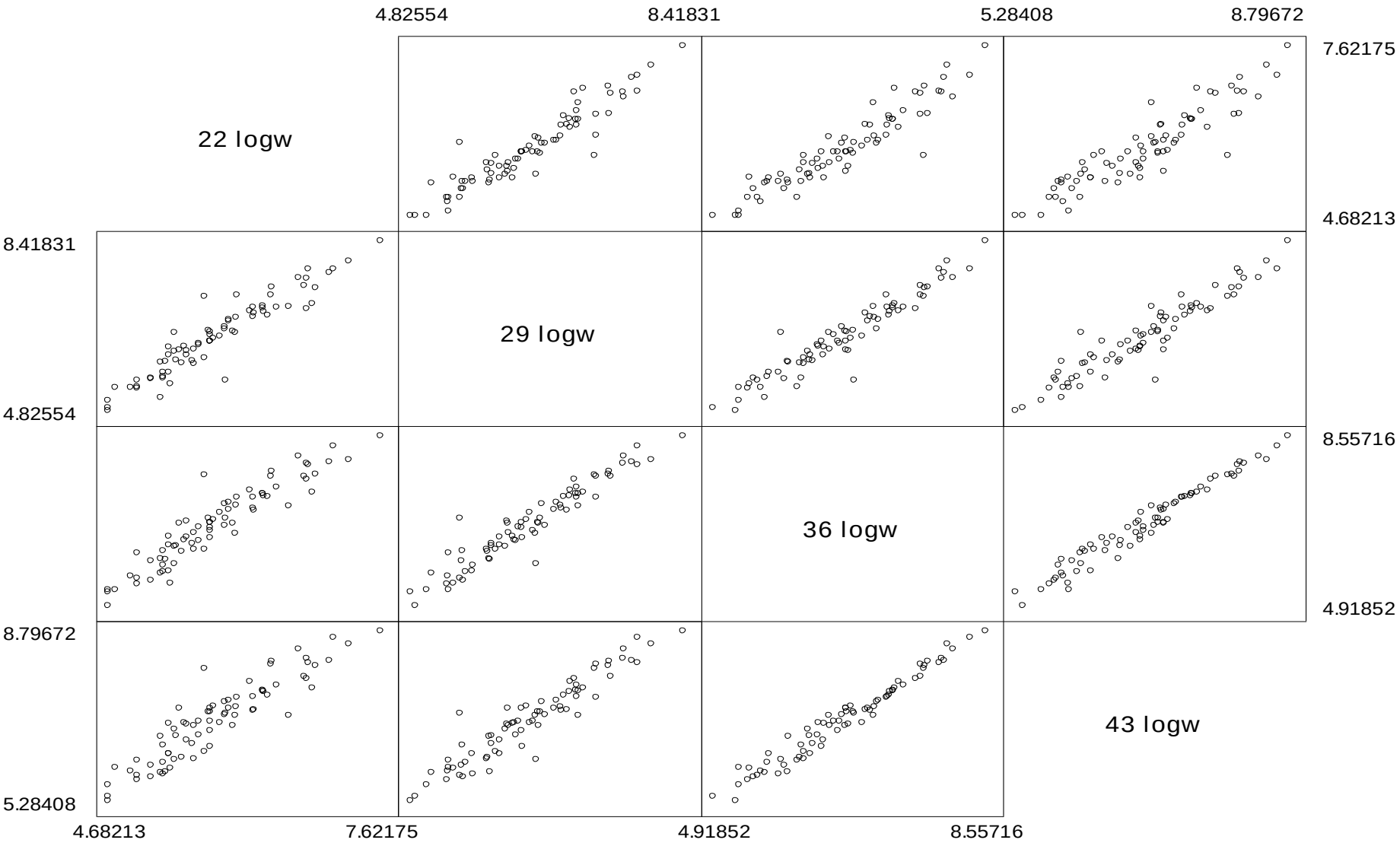
- Another common summary is the correlation matrix. Here is the correlation matrix for the Georgia babies data set:

```
. pwcorr  bweight1 bweight2 bweight3 bweight4 bweight5
             | bweight1 bweight2 bweight3 bweight4 bweight5
-------------+------------------------------------------
    bweight1 |   1.0000
    bweight2 |   0.2282   1.0000
    bweight3 |   0.2950   0.4833   1.0000
    bweight4 |   0.2578   0.4676   0.6185   1.0000
    bweight5 |   0.3810   0.4261   0.4233   0.4642   1.0000
```

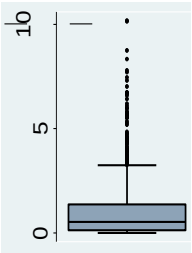
- How do we read this? Why isn't there anything in the top right hand corner?

Here is another example, giving the log weights (why log?) of mice for several weeks of measurement, mostly reflecting gain in tumor weight





Tumor/weight

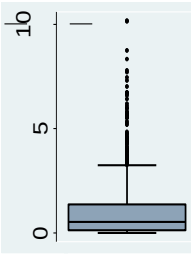


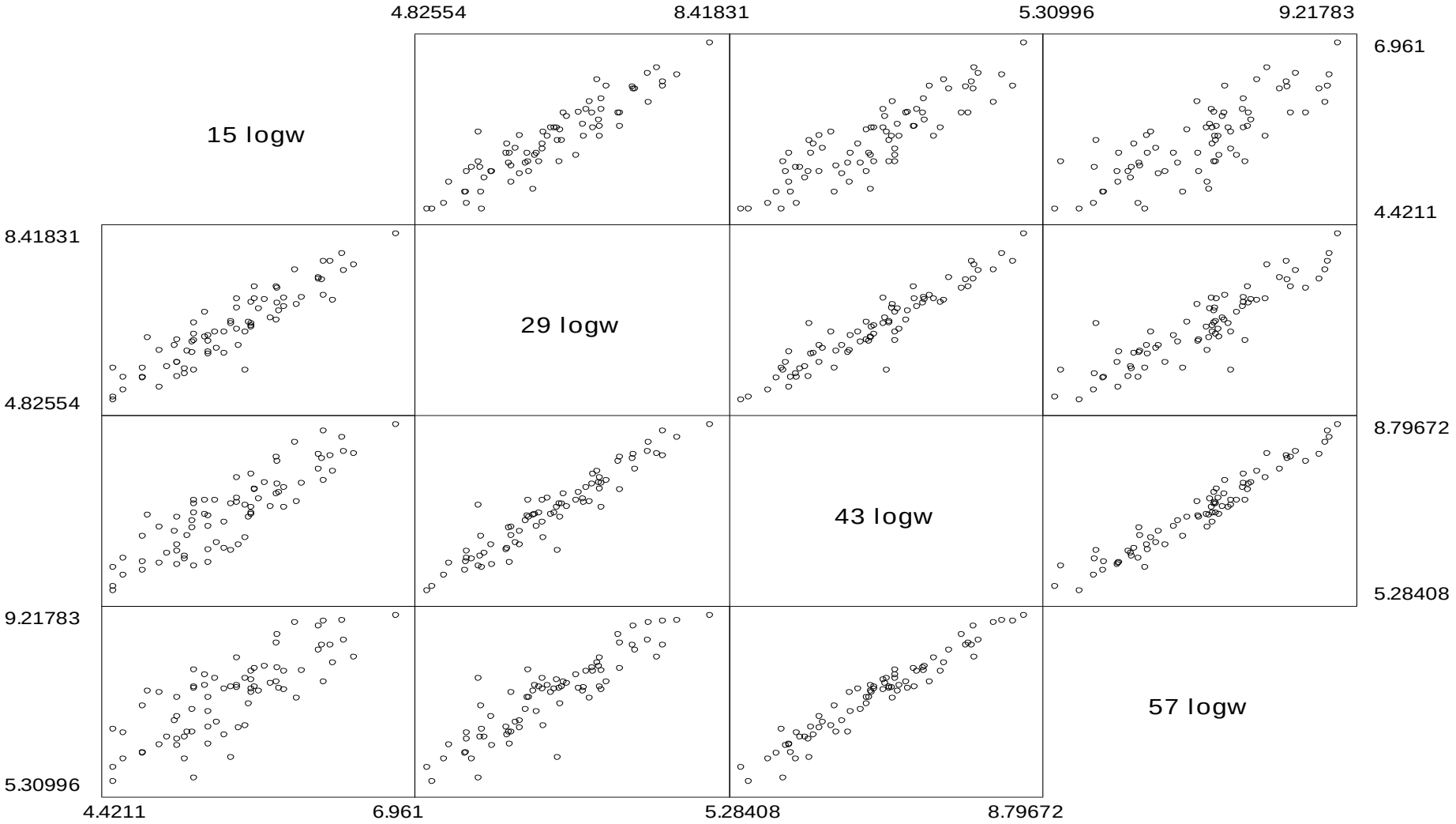
- And here is the corresponding correlation matrix:

```
pwcorr  logw22 logw29 logw36 logw43
             |   logw22   logw29   logw36   logw43
-------------+------------------------------------
      logw22 |   1.0000
      logw29 |   0.9414   1.0000
      logw36 |   0.9400   0.9568   1.0000
      logw43 |   0.9190   0.9466   0.9803   1.0000
```

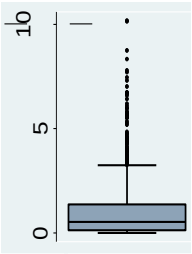
Tumor/weight

Here is a different collection of weeks for the same dataset. What does this suggest?





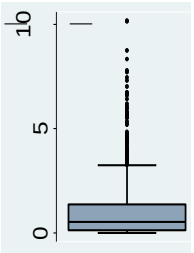
Tumor/weight



Here is the correlation matrix for that set of weeks:

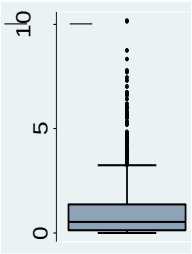
```
pwcorr  logw15 logw29 logw43 logw57

             |   logw15   logw29   logw43   logw57
-------------+------------------------------------
      logw15 |   1.0000
      logw29 |   0.9145   1.0000
      logw43 |   0.8713   0.9466   1.0000
      logw57 |   0.7937   0.8952   0.9692   1.0000
```

Correlation structures



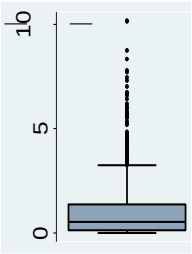
- The Georgia babies and tumor data sets are tidy because each “subject” has the same collection of observations (five observations for each mom and a tumor weight for each week). This is called “balanced” data.
- The Korff et al, back pain example is an example of unbalanced data, both because the sample sizes are unequal and because the “case-mix” is unequal. Because we don’t have a variable on which to reasonably order the observations (like parity for the Georgia babies data or time for the tumor data), there is not a reasonable plot we can make. But why are the data correlated in the back pain example?

Correlation structures



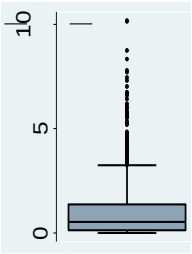
- Back to the tumor data: With 10 weeks of tumor data, there is the correlation of week 1 with week 2, week 1 with week 3, …, week 9 with week 10 for 45 unique correlations in all.
- Rather than having to estimate a separate correlation between each pair of times, we often use a simpler correlation "structure," both for ease of model specification and for statistical efficiency.

Correlation structures



- Common correlation structures used in STATA are:
  1. Exchangeable (all correlations equal).
  2. Autoregressive (correlations closer in time are more highly correlated, but drop off to zero as the difference in time increases).
  3. Unstructured (no assumptions made – estimate a separate correlation for each pair of time points).
  4. Independent (all correlations zero).
- In an AR(1) structure, if the correlation of adjacent time points is, say, 0.8, then the correlation of observations two time points apart is 0.8^2 = 0.64.

Correlation structures



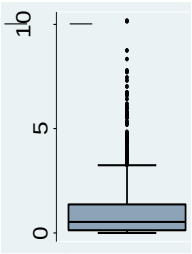
Which correlation structures do you feel best describe the examples we've considered?

Georgia babies

Tumor weights weeks 22 through 43

Tumor weights weeks 15 through 57

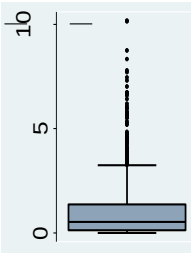
Back pain



Data layouts for longitudinal/clustered data

For longitudinal analyses: "long format"

id	visit	gender	bmi	totbmd
9000296	2	M	29.8	0.82
9000296	4	M	29.8	0.8
9000296	6	M	29.4	0.8
9000296	8	M	29.4	0.79
…				
9000798	2	F	32.4	0.4
9000798	4	F	32.3	0.41
9000798	6	F	32.3	0.39
9000798	8	F	32.5	0.39



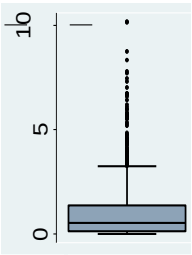
Data layouts for longitudinal/clustered data

For change score analyses: "wide format"

id	v02bmi	v04bmi	v06bmi
9000296	29.8	29.4	29.1
9000798	32.4	32.3	32.5

In lab: how to convert data back and forth between the two data formats.

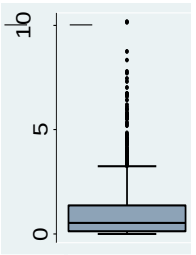
Example: HERS



Heart and Estrogen/Progestin Study (HERS - Hulley, *et al*, 1998, *JAMA*), was a randomized trial with long-term, yearly followup. pptid is the participant ID and nvisit is the visit number, with 0 being baseline:

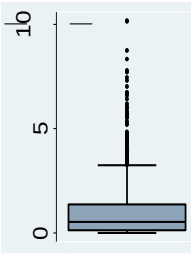
```
. xtset pptid nvisit
       panel variable:  pptid (unbalanced)
        time variable:  nvisit, 0 to 5, but with gaps
                delta:  1 unit
```

Descriptive methods: pattern of data



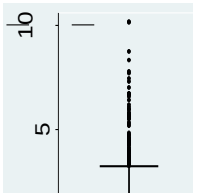

```
. xtdescribe
   pptid:  1, 2, ..., 2763                                 n =        2763
  nvisit:  0, 1, ..., 5                                    T =           6
           Delta(nvisit) = 1 unit
           Span(nvisit)  = 6 periods
           (pptid*nvisit uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       4         5         5       6       6

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
     1469     53.17   53.17 |  11111.
      531     19.22   72.39 |  1111..
      327     11.83   84.22 |  111111
      119      4.31   88.53 |  111...
      106      3.84   92.36 |  1.....
       87      3.15   95.51 |  11....
       40      1.45   96.96 |  111.1.
       22      0.80   97.76 |  1111.1
       11      0.40   98.15 |  11.11.
       51      1.85  100.00 | (other patterns)
 ---------------------------+---------
     2763    100.00         |  XXXXXX
```

Descriptive methods: variation between vs within individuals

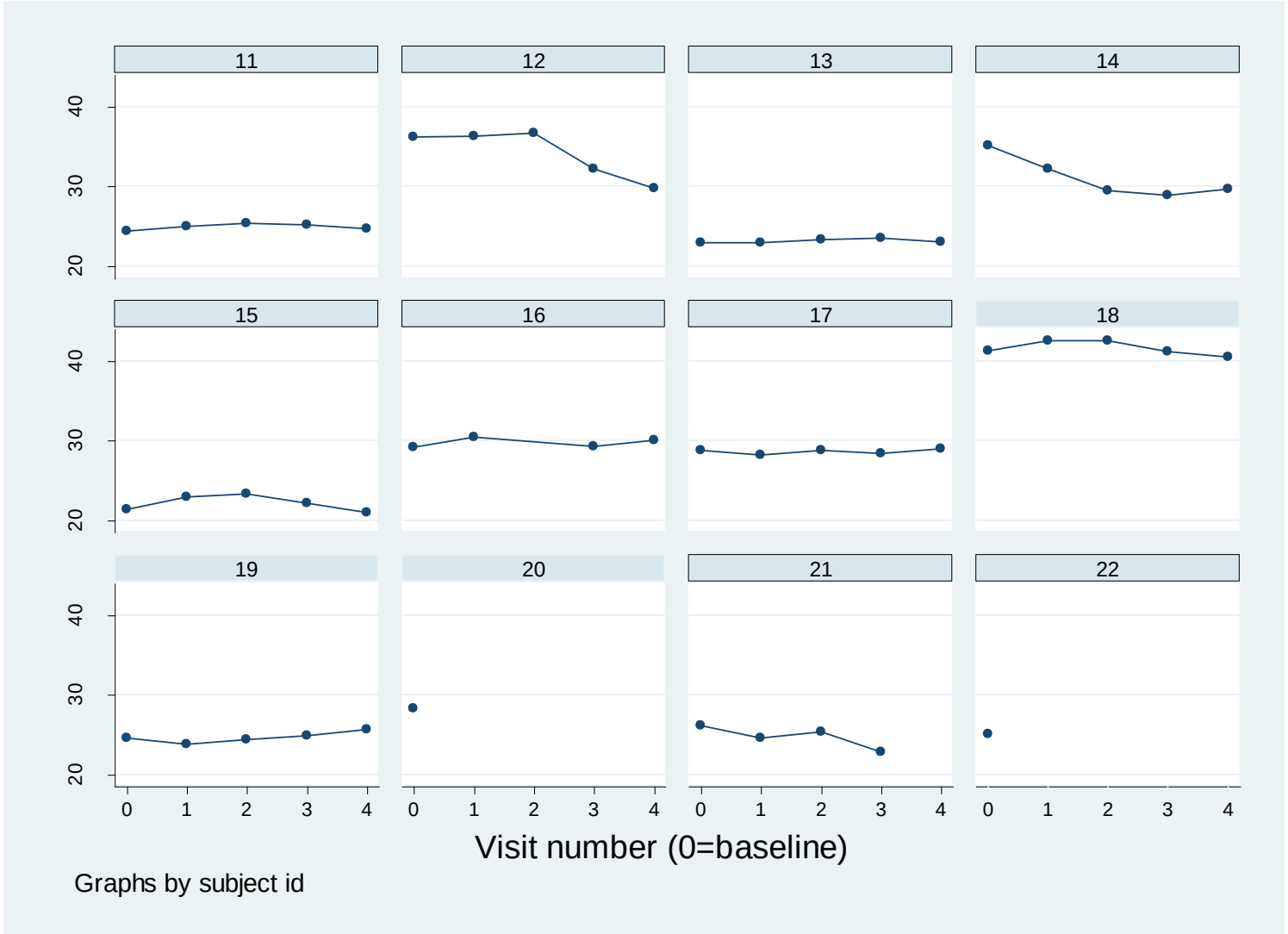



```
. xtsum bmi glucose white htnmeds

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
bmi        overall |  28.52184   5.565558      14.73      57.56 |     N =   12258
           between |             5.506242      15.21   54.89667 |     n =    2761
           within  |             1.174147   20.50784   38.41184 | T-bar =  4.4397
                   |                                          |
glucose    overall |  113.4589   41.18884          4        478 |     N =    7608
           between |             36.56199         53        311 |     n =    2763
           within  |             20.33938  -59.20781   322.7922 | T-bar = 2.75353
                   |                                          |
white      overall |  .8936408   .3083091          0          1 |     N =   12533
           between |             .3153701          0          1 |     n =    2760
           within  |                   0   .8936408   .8936408 | T-bar = 4.54094
                   |                                          |
htnmeds    overall |  .8341568   .3719546          0          1 |     N =   12548
           between |              .324773          0          1 |     n =    2763
           within  |             .1811389   .0008235    1.66749 | T-bar = 4.54144
```

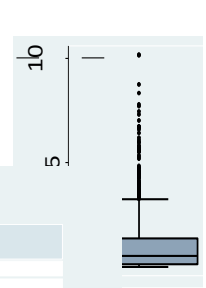
Descriptive methods: graphing variation over time

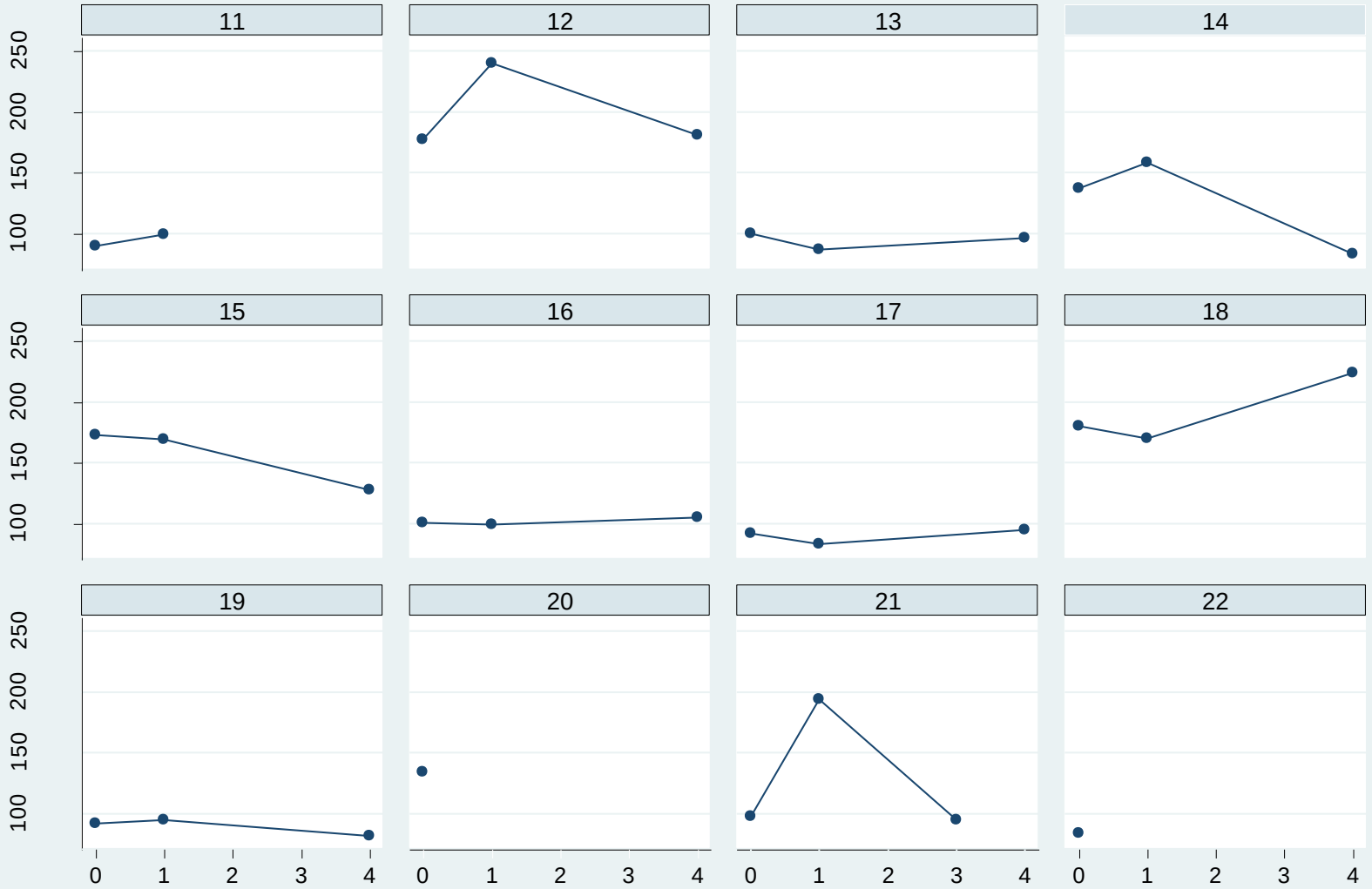



```
twoway (connected bmi nvisit) if pptid>10 & pptid<23, by(pptid)
```

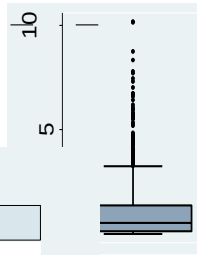


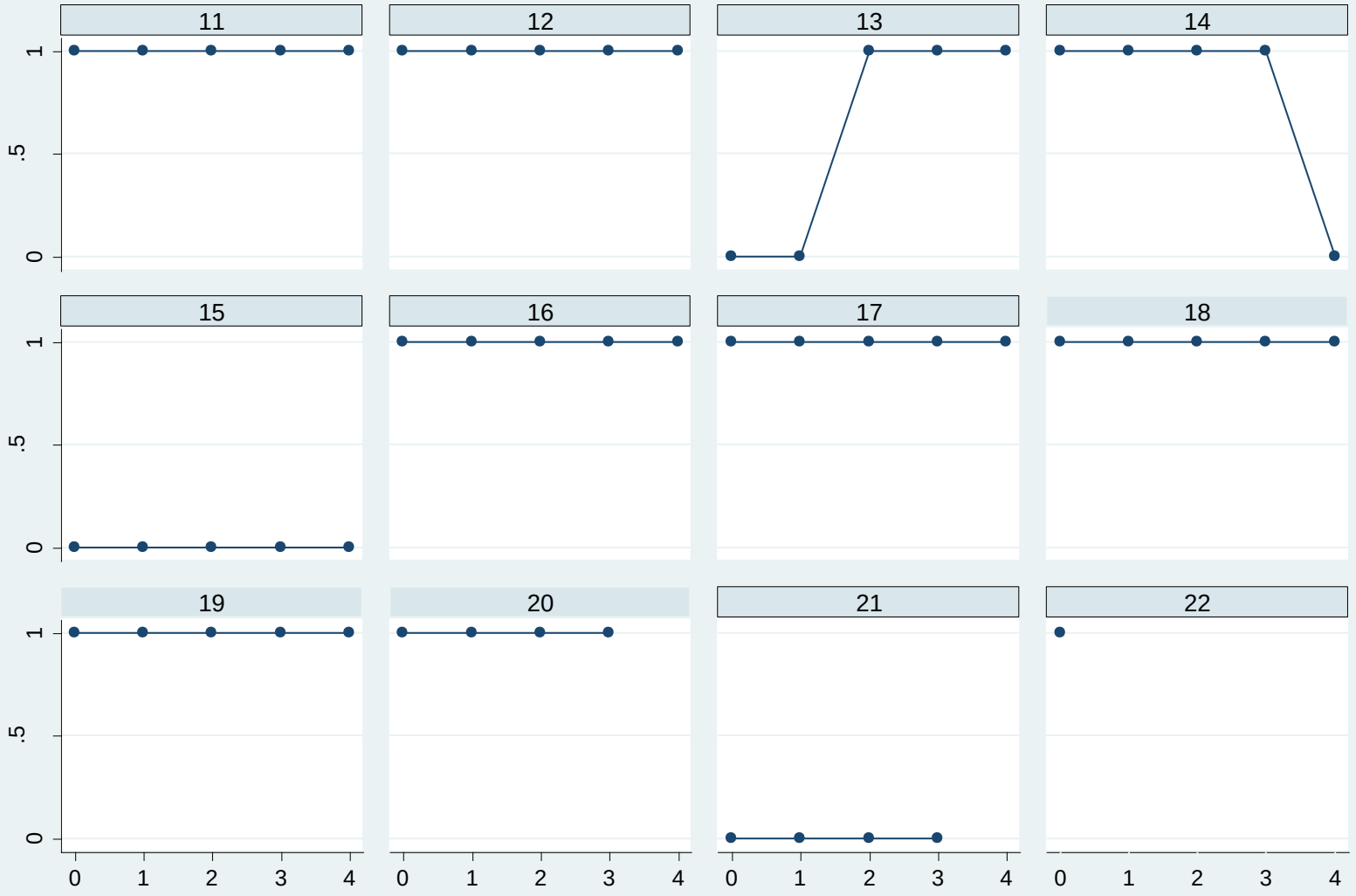
Descriptive methods: graphing



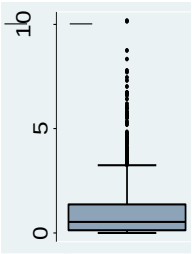


Descriptive methods: graphing



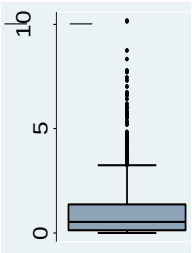


Summary – Part 1



- Hierarchical data structures are common.
- They lead to correlated data.
- Ignoring the correlation can be a serious error.
- Not easy to predict whether a proper, correlated data analysis will yield larger or smaller standard errors compared to an incorrect analysis that assumes all the data are independent.

Summary – Part 2



- We will often need to specify a correlation "structure" when using correlated data methods.
- Stata has convenient, built-in functions that generate descriptive statistics to better understand hierarchical data. Use them before embarking on formal statistical analyses.