Analysis of repeated measurements (KLMED8008)

Eirik Skogvoll, MD PhD

Professor and Consultant

Institute of Circulation and Medical Imaging

Dept. of Anaesthesiology and Intensive Care

Day 2

- Practical issues …?
  - Lectures
  - Textbook
  - Software
  - Excercises
  - Exam
- Review of the dependence/ correlation example
- Taking baseline value into consideration
- Power and sample size determination
- Brief review of
  - One-way ANOVA (ANalysis Of VAriance)
  - ANOVA for repeated measurements
  - Friedman's non-parametric test

TAKING BASELINE VALUE INTO CONSIDERATION

Problem description

- To compare two independent groups (e.g. intervention/ control) with respect to one outcome variable
- Baseline (i.e. value before intervention) is known.
- Outcome is likely related to the baseline (subjects differ!!).
- Irrespective of intervention, extreme observations at baseline will move towards the average at follow-up (”regression to the mean”)

Vickers, A. J. and D. G. Altman (2001). Analysing controlled trials with baseline and follow up measurements." *BMJ* 323(7321): 1123-1124.

Example

Vickers (2001). *BMJ* 323(7321): 1123-1124.

```
. use "M:\Ntnu\KLMED8008\Dag2\Vickers.dta", clear

. describe

Contains data from M:\Ntnu\KLMED8008\Dag2\Vickers.dta
  obs:            52
 vars:             4                          4 Mar 2012 18:04
 size:           208
----------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
----------------------------------------------------------------------
pre             byte   %9.0g                  pretreatment score
post            byte   %9.0g                  post treatment score
group           byte   %11.0g      group      0= placeb, 1= acupunct
d               byte   %9.0g                  change score
```

	pre	post	group	d
1.	35	35	0	0
2.	54	37	0	-17
3.	35	40	0	5
4.	30	42	0	12
5.	44	45	0	1
6.	49	47	0	-2
7.	38	51	0	13
8.	52	52	0	0

	pre	post	group	d
28.	43	74	1	31
29.	75	99	1	24
30.	66	88	1	22
31.	65	85	1	20
32.	74	100	1	26
33.	58	58	1	0
34.	62	77	1	15
35.	64	100	1	36
36.	59	61	1	2

Taking baseline value into consideration

Possible analysis strategies:

- Compare group outcomes (”POST”) without considering (i.e. ignoring!) baseline (… may be misleading if there is baseline inequality, and is inefficient)
- Compare CHANGE within each group (…may be misleading due to regression to the mean)

Recommended strategy[1]

- Model the outcome using a linear model with baseline as covariate and group as factor (”ANCOVA”).
- Equivalently, change from baseline may be modeled with baseline as covariate and group as factor.
- Modelling the outcome using a linear mixed model (day 5)

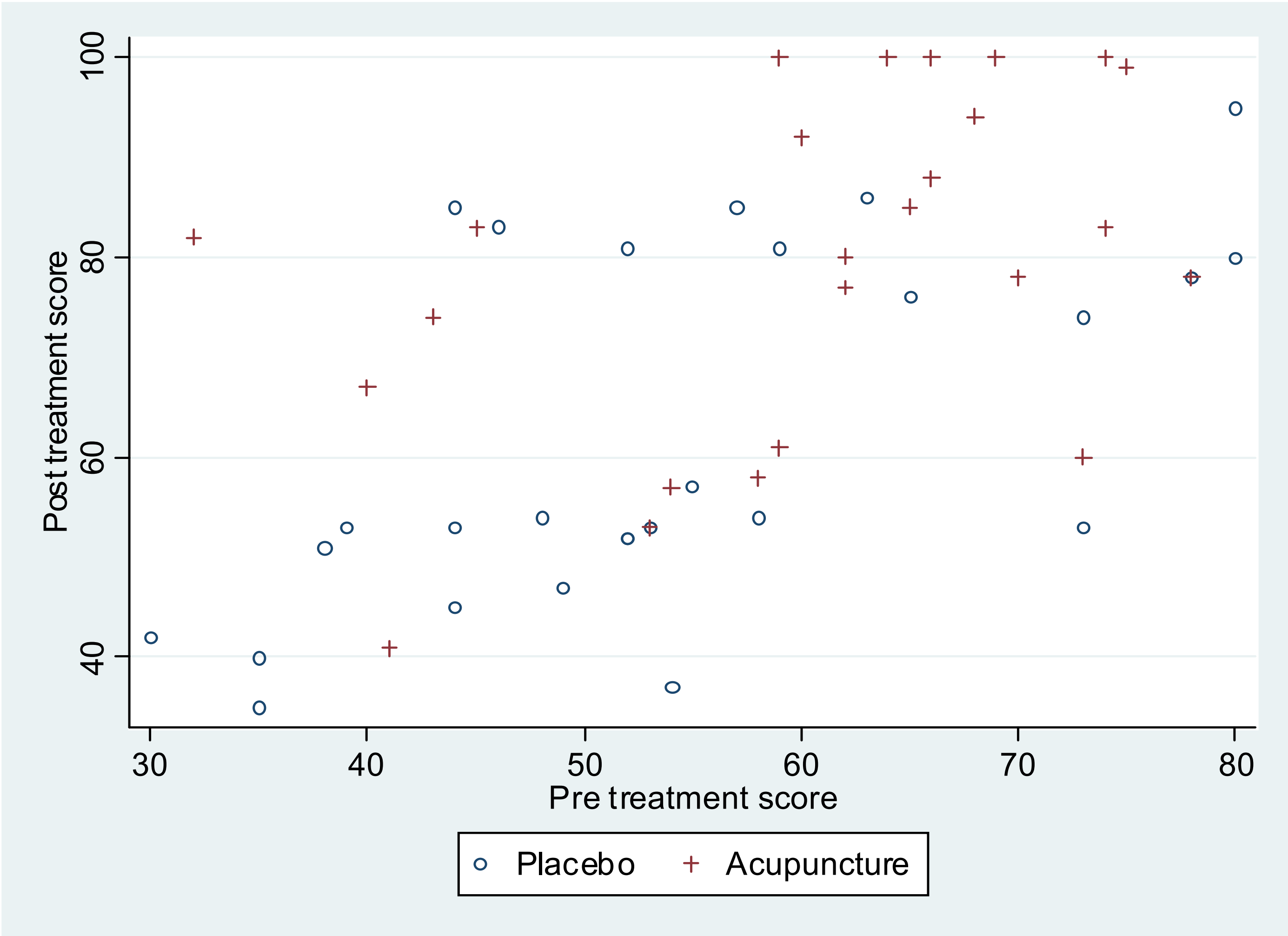
Vickers, A. J. and D. G. Altman (2001). Analysing controlled trials with baseline and follow up measurements." *BMJ* 323(7321): 1123-1124.

Example

Vickers (2001). *BMJ* 323(7321): 1123-1124.

```
. table group, c(mean pre mean post)

------------------------------------
0= placeb,   |
1= acupunct  |  mean(pre)  mean(post)
-------------+----------------------
     placebo |   53.92593     62.2963
 acupuncture |       60.4        79.6
------------------------------------


. ttest post, by (group) // bare outcome

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
 placebo |      27     62.2963    3.449602    17.92466    55.20554    69.38706
acupunct |      25        79.6      3.4288      17.144    72.52331    86.67669
---------+--------------------------------------------------------------------
combined |      52    70.61538    2.697291    19.45044    65.20035    76.03042
---------+--------------------------------------------------------------------
    diff |           -17.3037    4.872285               -27.08998   -7.517431
------------------------------------------------------------------------------
    diff = mean(placebo) - mean(acupunct)                         t =  -3.5515
Ho: diff = 0                                     degrees of freedom =       50

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0004         Pr(|T| > |t|) = 0.0008          Pr(T > t) = 0.9996
```

```
twoway (scatter post pre if group == 0, ms(oh)) (scatter post pre if group == 1, ///
ms(+)), ytitle (Post treatment score) xtitle ///
(Pre treatment score) legend (label(1 Placebo) label(2 Acupuncture) label ///
( 3 Placebo) label(4 Acupuncture) )
```



Example

Vickers (2001). *BMJ* 323(7321): 1123-1124.

```
. table group, c(mean pre mean post)
```

0= placeb, 1= acupunct	mean(pre)	mean(post)
placebo	53.92593	62.2963
acupuncture	60.4	79.6

```
. ttest pre, by (group) // bare pre-verdi
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
placebo	27	53.92593	2.700599	14.03273	48.37476	59.47709
acupunct	25	60.4	2.456963	12.28481	55.32908	65.47092
combined	52	57.03846	1.871224	13.49359	53.28182	60.7951
diff		-6.474074	3.670004		-13.84549	.8973453

```
    diff = mean(placebo) - mean(acupunct)                      t =  -1.7641
Ho: diff = 0                                     degrees of freedom =       50

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0419         Pr(|T| > |t|) = 0.0838          Pr(T > t) = 0.9581
```

Example

Vickers (2001). *BMJ* 323(7321): 1123-1124.

```
. table group, c(mean pre mean post)
```

0= placeb, 1= acupunct	mean(pre)	mean(post)
placebo	53.92593	62.2963
acupuncture	60.4	79.6

```
. ttest d, by(group)  // bare diff
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
placebo	27	8.37037	2.801174	14.55533	2.612475	14.12827
acupunct	25	19.2	3.220766	16.10383	12.55267	25.84733
combined	52	13.57692	2.23562	16.12129	9.088728	18.06512
diff		-10.82963	4.251638		-19.3693	-2.289963

```
    diff = mean(placebo) - mean(acupunct)                      t =  -2.5472
Ho: diff = 0                                  degrees of freedom =       50

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0070         Pr(|T| > |t|) = 0.0140          Pr(T > t) = 0.9930
```

Example

Vickers (2001). *BMJ* 323(7321): 1123-1124.

```
. table group, c(mean pre mean post)
```

0= placeb, 1= acupunct	mean(pre)	mean(post)
placebo	53.92593	62.2963
acupuncture	60.4	79.6

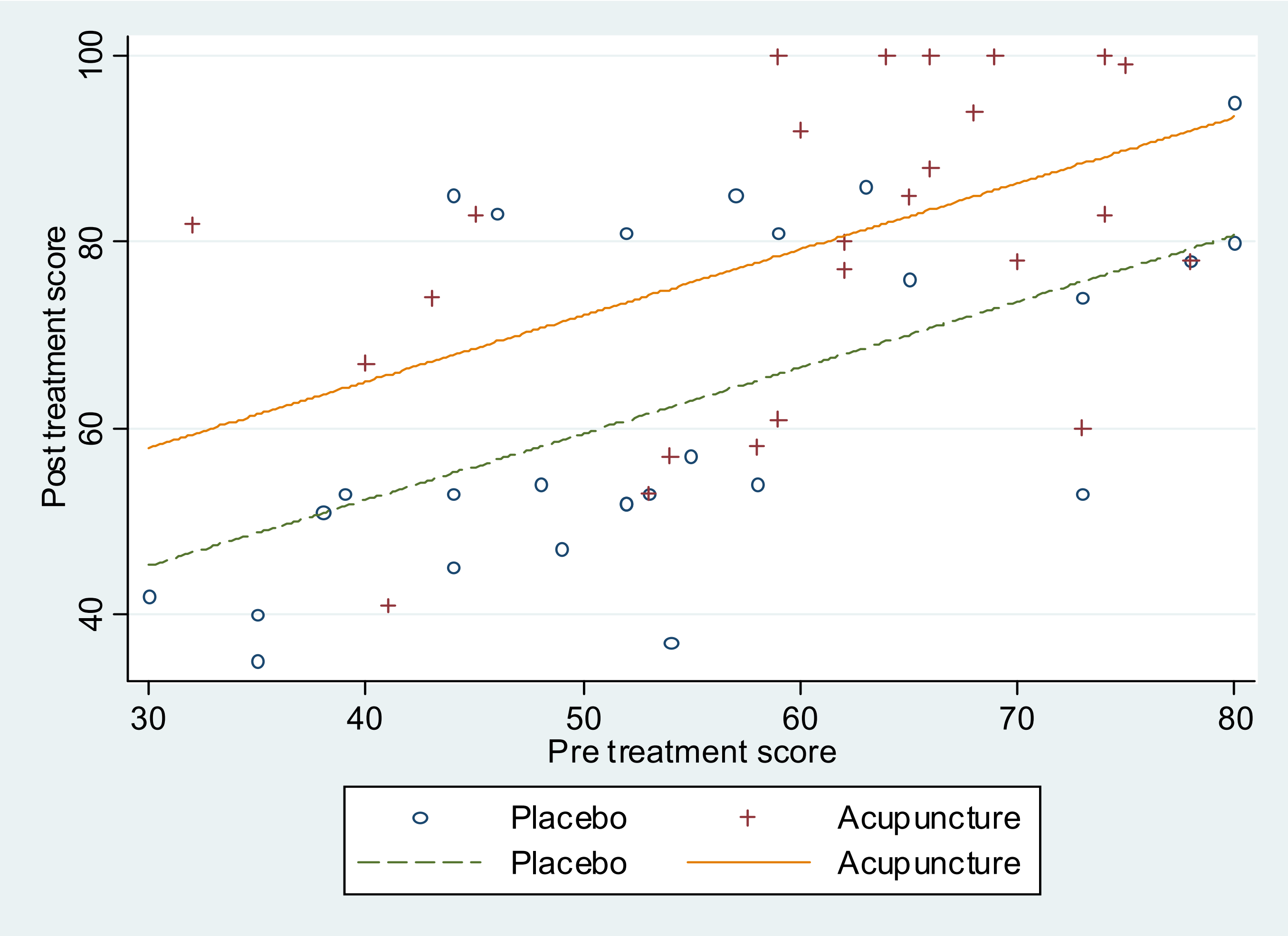
```
. regress post pre i.group // rett metode
```

Source	SS	df	MS
Model	8296.11243	2	4148.05621
Residual	10998.1953	49	224.452965
Total	19294.3077	51	378.319759

Number of obs =	52
F(2, 49) =	18.48
Prob > F =	0.0000
R-squared =	0.4300
Adj R-squared =	0.4067
Root MSE =	14.982

post	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pre	.7102148	.1602363	4.43	0.000	.3882079	1.032222
1.group	12.70572	4.285715	2.96	0.005	4.093254	21.31819
_cons	23.99731	9.109231	2.63	0.011	5.691621	42.30299

```
twoway (scatter post pre if group == 0, ms(oh)) (scatter post pre if group == 1, ///
ms(+)) (function y=23.9+ 0.71*x, range (pre) lpatt(dash)) (function y=23.9 ///
+12.7+ 0.71*x, range (pre) lpatt(solid)), ytitle (Post treatment score) xtitle ///
(Pre treatment score) legend (label(1 Placebo) label(2 Acupuncture) label ///
( 3 Placebo) label(4 Acupuncture) )
```



TAKING BASELINE VALUE INTO CONSIDERATION
– POWER AND SAMPLE SIZE

Problem description

- To compare two independent groups (e.g. intervention/ control) with respect to one outcome variable
- Baseline (i.e. value before intervention) is known.
- Outcome is likely related to the baseline.

- To calculate the necessary sample size, one needs to know:
  - The minimum difference to be detected (if small enough: it cannot be found!)
  - Variance of the baseline and outcome variables
  - Correlation between baseline and outcome
  - Required power (usu. 90 %) and significance level (usu. 5%, two-sided)

Example – Sample size

- Consider redoing the acupuncture experiment with:
  - A minimum difference of 5 (55 vs. 60) is to be detected
  - Variance ≈ 15^2
  - Correlation between baseline and outcome ≈ 0.55

```
. pwcorr pre post, sig

             |      pre     post
-------------+------------------
         pre |   1.0000
             |
             |
        post |   0.5725   1.0000
             |   0.0000
             |
```

  - Power 90 %, 5 % significance level, two-sided test

Example – Sample size

```
. sampsi 55 60, sd(15) pre(1) r01(0.55)

Estimated sample size for two samples with repeated measures
Assumptions:
                                      alpha =   0.0500  (two-sided)
                                      power =   0.9000
                                         m1 =       55
                                         m2 =       60
                                        sd1 =       15
                                        sd2 =       15
                                      n2/n1 =     1.00
            number of follow-up measurements =        1
             number of baseline measurements =        1
   correlation between baseline & follow-up =    0.550
```

Note: as for Stata 13, the command «power» has (officially) replaced «sampsi», but it does not seem to cover this situation (yet?)

Example
– Sample size

```
Method: POST
 relative efficiency =    1.000
    adjustment to sd =    1.000
        adjusted sd1 =   15.000

 Estimated required sample sizes:
                  n1 =      190
                  n2 =      190

Method: CHANGE
 relative efficiency =    1.111
    adjustment to sd =    0.949
        adjusted sd1 =   14.230

 Estimated required sample sizes:
                  n1 =      171
                  n2 =      171

Method: ANCOVA
 relative efficiency =    1.434
    adjustment to sd =    0.835
        adjusted sd1 =   12.527

 Estimated required sample sizes:
                  n1 =      132
                  n2 =      132
```

Lord’s paradox

- When comparing *non-randomized groups* (e.g. sex), adjustment for baseline may lead to contradictory[1] and unreliable[2] results.
- Similar problems may occur in observational studies when attempting to ”adjust” for various factors and covariates[2].

[1]Lord FM. A paradox in the interpretation of group comparisons. *Psychological Bulletin*. 1967;68(5):304-5.

[2]Tu Y-K et al. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon - the reversal paradox. Emerging Themes in Epidemiology. 2008;5(1):2.

The scandal of poor epidemiological research

Reporting guidelines are needed for observational epidemiology

Something surely must be wrong with epidemiology when the new editors of a leading journal in the field entitle their inaugural offering, "Epidemiology—is it time to call it a day?"[1] Observational epidemiology has not had a good press in recent years. Conflicting results from epidemiological studies of the risks of daily life, such as coffee, hair dye, or hormones, are frequently and eagerly reported in the popular press, providing a constant source of anxiety for the public.[2 3] In many cases deeply held beliefs, given credibility by numerous observational studies over long periods of time, are challenged only when contradicted by randomised trials. In the most recent example, a Cochrane review of randomised trials shows that antioxidant vitamins do not prevent gastrointestinal cancer and may even increase all cause mortality.[4 5]

Now Pocock et al describe the quality and the litany of problems of 73 epidemiological studies published in January 2001 in general medical and specialist journals (p 883).[6] Perhaps the most relevant findings relate to how investigators dealt with confounding, multiple comparisons, and subgroup analyses.

Papers p 883

BMJ 2004;329:868–9

Elm Ev, Egger M. The scandal of poor epidemiological research. BMJ. 2004 October 16, 2004;329(7471):868-9.

http://www.strobe-statement.org/

One way ANOVA

- The t-test extended to more than two groups
- Example:
  Reaction time (score) measured with four drugs:

```
. table drug, c(mean score sd score)
```

drug	mean(score)	sd(score)
drug1	26.4	8.763561
drug2	25.6	6.542171
drug3	15.6	3.847077
drug4	32	8

Obtained from
Winer 1991 in Stata manual: [R] anova
lwiner.dta (long format)

```
. use lwiner, clear
(T4.3 -- Winer, Brown, M

. list
```

	drug	score
1.	drug1	30
2.	drug2	28
3.	drug3	16
4.	drug4	34
5.	drug1	14
6.	drug2	18
7.	drug3	10
8.	drug4	22
9.	drug1	24
10.	drug2	20
11.	drug3	18
12.	drug4	30
13.	drug1	38
14.	drug2	34
15.	drug3	20
16.	drug4	44
17.	drug1	26
18.	drug2	28
19.	drug3	14
20.	drug4	30

One way ANOVA

- One *continuous* outcome variable
- ”One way” means a single ”axis of investigation”, e.g. different doses of the same drug
- One categorical predictor variable with two or more *levels;* thus defining two or more groups

  Example: Textbook section 1.4 illustrates ANOVA analysis with two groups (men/ women)

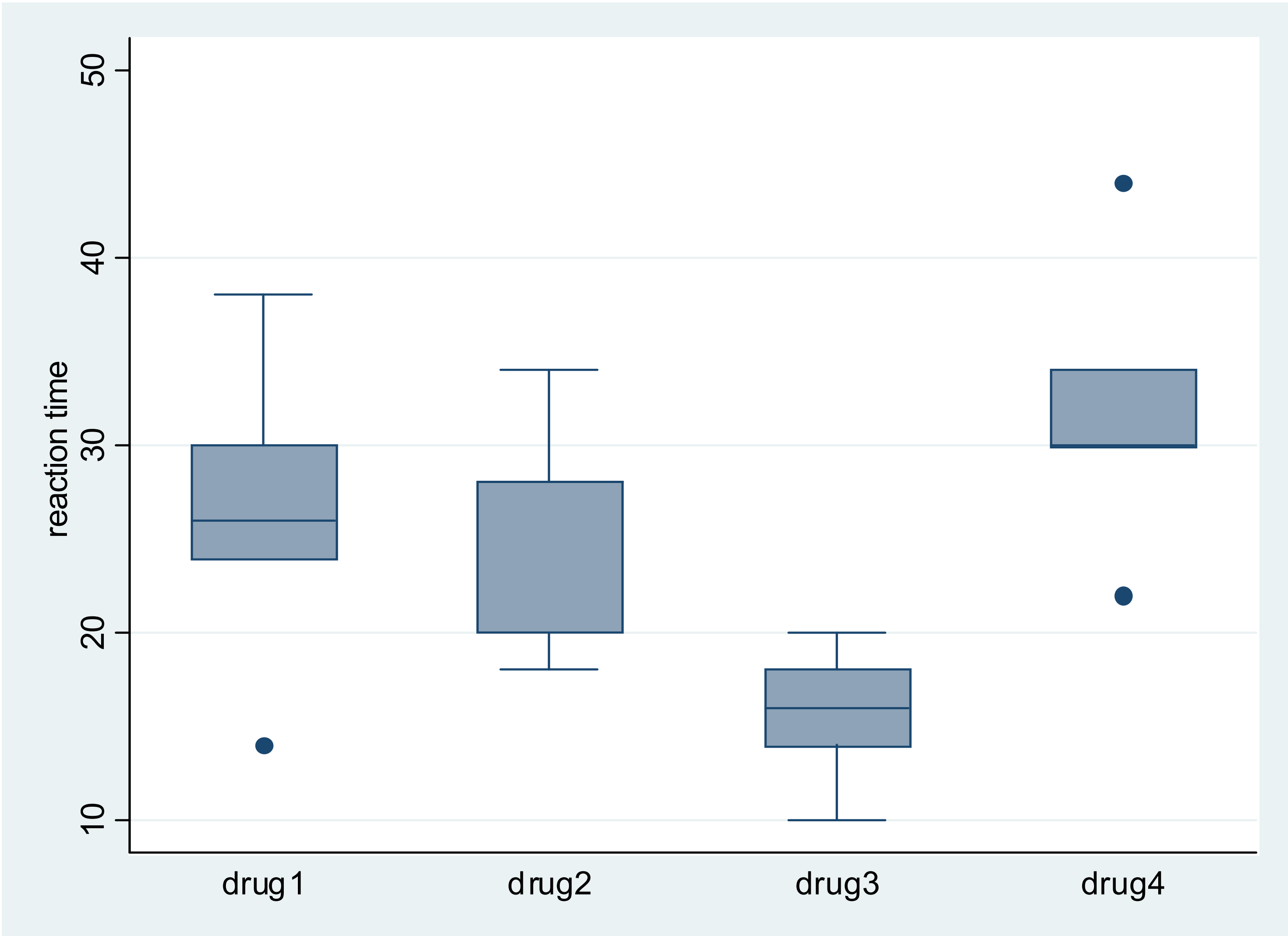
- H_0: The Expected (”mean”) values in all groups are equal
- The (global) F-test compares mean *model* sum-of-squares (SS) to mean *residual* sum-of-squares under H_0

Example

Winer 1991 in Stata manual: [R] anova

```
. anova score drug

                    Number of obs =        20     R-squared     =  0.4680
                    Root MSE      = 7.04273     Adj R-squared =  0.3683

          Source | Partial SS    df        MS           F     Prob > F
       ----------+----------------------------------------------------
           Model |      698.2     3   232.733333      4.69     0.0155
                 |
            drug |      698.2     3   232.733333      4.69     0.0155
                 |
        Residual |      793.6    16         49.6
       ----------+----------------------------------------------------
           Total |     1491.8    19   78.5157895

. margins drug

Adjusted predictions                              Number of obs   =         20

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |
          1  |       26.4   3.149603     8.38   0.000     20.22689    32.57311
          2  |       25.6   3.149603     8.13   0.000     19.42689    31.77311
          3  |       15.6   3.149603     4.95   0.000     9.426891    21.77311
          4  |         32   3.149603    10.16   0.000     25.82689    38.17311
------------------------------------------------------------------------------
```

```
. graph box score, over (drug)
```



ANOVA ”Post-hoc” tests

```
. pwcompare drug, bonferroni effects

Pairwise comparisons of marginal linear predictions

Margins      : asbalanced
```

	Number of Comparisons
drug	6

	Contrast	Std. Err.	Bonferroni t	Bonferroni P>|t|	Bonferroni [95% Conf.	Interval]
drug						
2 vs 1	-.8	4.454211	-0.18	1.000	-14.19976	12.59976
3 vs 1	-10.8	4.454211	-2.42	0.165	-24.19976	2.599755
4 vs 1	5.6	4.454211	1.26	1.000	-7.799755	18.99976
3 vs 2	-10	4.454211	-2.25	0.235	-23.39976	3.399755
4 vs 2	6.4	4.454211	1.44	1.000	-6.999755	19.79976
4 vs 3	16.4	4.454211	3.68	0.012	3.000245	29.79976

Repeated measurements ANOVA

- One *continuous* outcome variable
- Every ”subject” (e.g. patient, animal) is *repeatedly* observed
- Different treatments applied to the same individual. If possible, one should apply the interventions *in random order!*

  (Applying interventions as time evolves *may lead to "autocorrelation"* i.e. obs. no 1 is correlated with no 2, obs. no 2 with no. 3, no 3 with no 4 etc.)

Repeated measurements ANOVA

- May be viewed as a two-way analysis, with ”subject” as one experimental, fixed factor. However, the effect of this factor is generally without specific interest[1].
- The analysis requires “sphericity”: among other things equal correlation among all variables
  - Mauchly’s test: checks the requirement of sphericity
  - If not fullfilled, the denominator degrees of freedom of the global F-test may be adjusted, leading to stricter significance criteria. (Huyn-Feldt adjustment).
- Special problem: how to handle missing observations?
  - Usually, the entire row is deleted, wasting information.
  - Imputing missing observations is an option.

[1] In the remainder of the course, ”subject” will be considered a random factor.
Textbook section 3.8, p.124

Repeated measurements ANOVA

- Example:
  The reaction time (score) with four drugs was actually measured repeatedly in the same 5 persons:

	person	score_1	score_2	score_3	score_4
1.	1	30	28	16	34
2.	2	14	18	10	22
3.	3	24	20	18	30
4.	4	38	34	20	44
5.	5	26	28	14	30

Winer 1991 in Stata manual: [R] anova

```
. use winer, clear
(T4.3 -- Winer, Brown, Michels)

. list
```

	person	drug	score
1.	1	1	30
2.	1	2	28
3.	1	3	16
4.	1	4	34
5.	2	1	14
6.	2	2	18
7.	2	3	10
8.	2	4	22
9.	3	1	24
10.	3	2	20
11.	3	3	18
12.	3	4	30
13.	4	1	38
14.	4	2	34
15.	4	3	20
16.	4	4	44
17.	5	1	26
18.	5	2	28
19.	5	3	14
20.	5	4	30

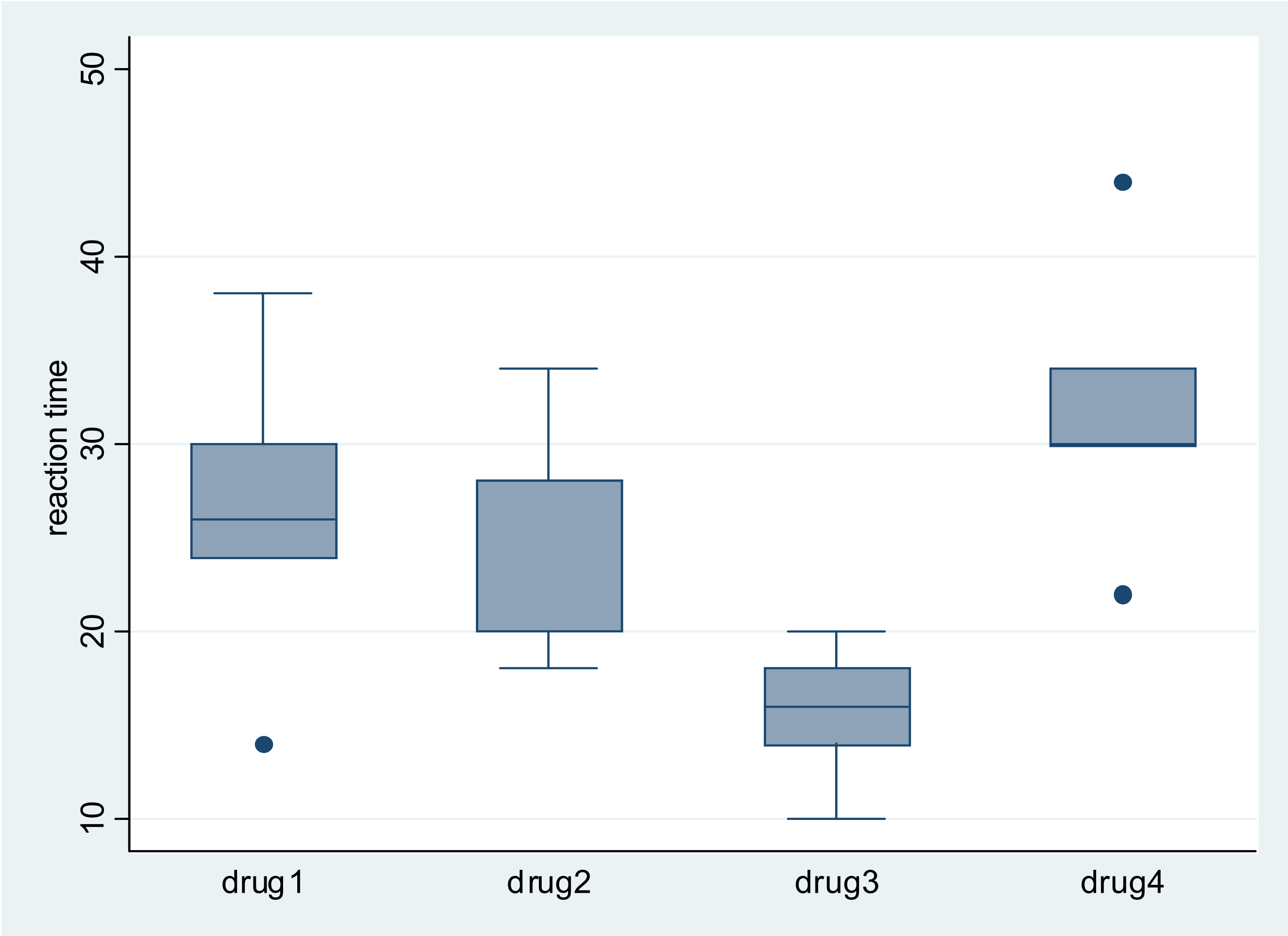
```
. anova score person drug, repeated(drug)

                          Number of obs =      20     R-squared     =  0.9244
                          Root MSE      = 3.06594     Adj R-squared =  0.8803

              Source |  Partial SS    df       MS           F     Prob > F
          -----------+----------------------------------------------------
               Model |        1379     7         197      20.96     0.0000
                     |
              person |       680.8     4       170.2      18.11     0.0001
                drug |       698.2     3  232.733333      24.76     0.0000
                     |
            Residual |       112.8    12         9.4
          -----------+----------------------------------------------------
               Total |      1491.8    19  78.5157895


Between-subjects error term:  person
                     Levels:  5         (4 df)
     Lowest b.s.e. variable:  person

Repeated variable: drug
                                          Huynh-Feldt epsilon        =  1.0789
                                          *Huynh-Feldt epsilon reset to 1.0000
                                          Greenhouse-Geisser epsilon =  0.6049
                                          Box's conservative epsilon =  0.3333

                                            ------------ Prob > F ------------
              Source |     df      F    Regular    H-F      G-G      Box
          -----------+----------------------------------------------------
                drug |      3    24.76   0.0000   0.0000   0.0006   0.0076
            Residual |     12
          -----------+----------------------------------------------------
```

Example

```
. margins drug

Predictive margins                              Number of obs     =         20

Expression   : Linear prediction, predict()
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
drug						
1	26.4	1.371131	19.25	0.000	23.71263	29.08737
2	25.6	1.371131	18.67	0.000	22.91263	28.28737
3	15.6	1.371131	11.38	0.000	12.91263	18.28737
4	32	1.371131	23.34	0.000	29.31263	34.68737

```
. graph box score, over (drug)
```



ANOVA ”Post-hoc” tests

```
. pwcompare drug, bonferroni effects

Pairwise comparisons of marginal linear predictions

Margins      : asbalanced
```

	Number of Comparisons
drug	6

	Contrast	Std. Err.	Bonferroni t	Bonferroni P>|t|	Bonferroni [95% Conf.	Interval]
drug						
2 vs 1	-.8	1.939072	-0.41	1.000	-6.913276	5.313276
3 vs 1	-10.8	1.939072	-5.57	0.001	-16.91328	-4.686724
4 vs 1	5.6	1.939072	2.89	0.082	-.5132759	11.71328
3 vs 2	-10	1.939072	-5.16	0.001	-16.11328	-3.886724
4 vs 2	6.4	1.939072	3.30	0.038	.2867241	12.51328
4 vs 3	16.4	1.939072	8.46	0.000	10.28672	22.51328

Friedman’s test: a non-parametric analogue to two-way ANOVA

- “Blocked” or “stratified” version of the Sign test/ Kruskal-Wallis’ test
- Conceptual: 2-way ANOVA without interaction, based on ranking the observations.
- Considers the individuals as “blocks” or “strata”
- Stata has no official procedure, but it is available from users. Type:

```
. findit Friedman
```

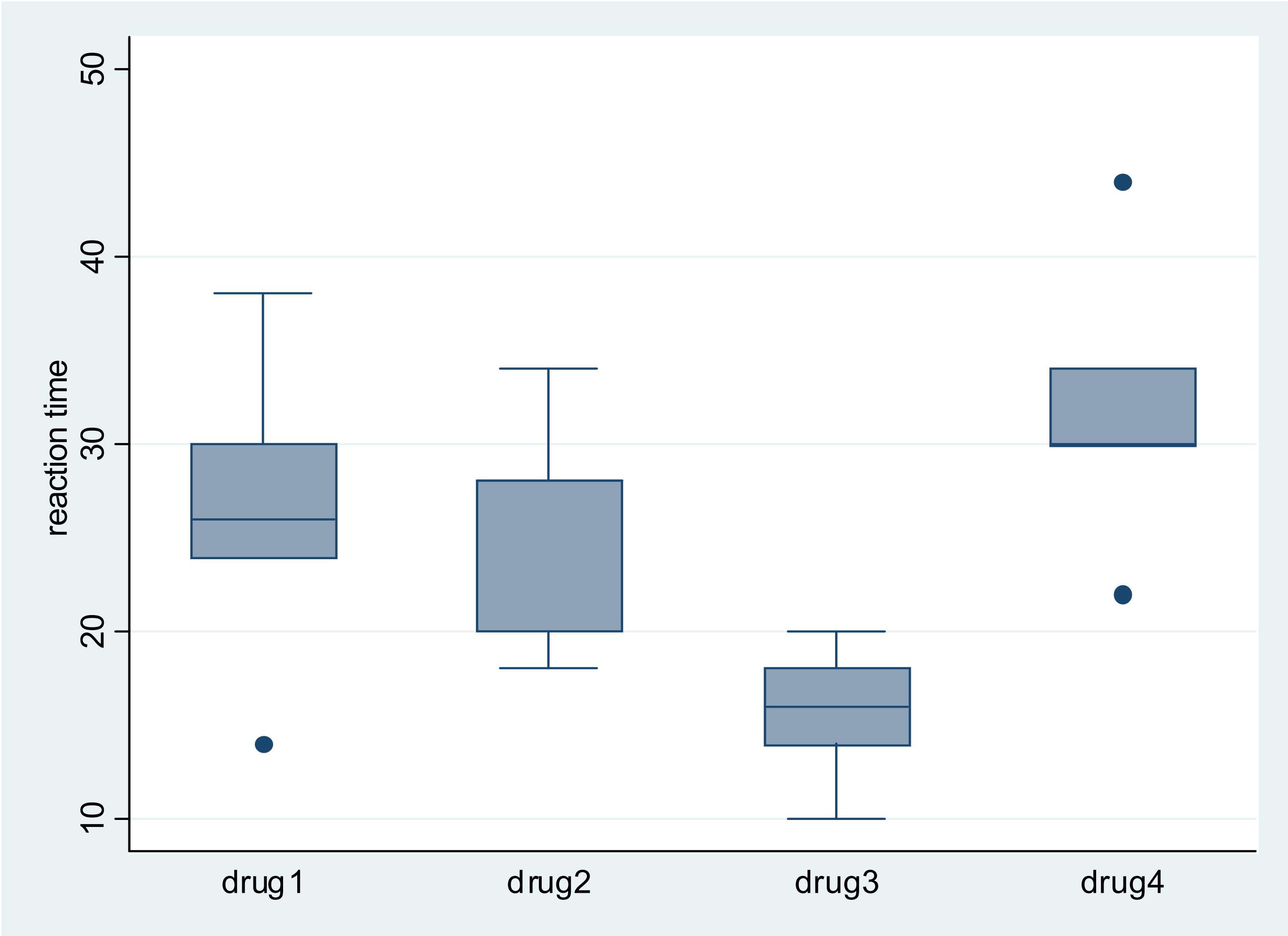
….and follow the instructions!

```
. use wwiner, clear
(Written by R.              )

. list
```

	person	score_1	score_2	score_3	score_4
1.	1	30	28	16	34
2.	2	14	18	10	22
3.	3	24	20	18	30
4.	4	38	34	20	44
5.	5	26	28	14	30

```
. friedman score*

Friedman =  13.7500
Kendall =    0.8594
P-value =    0.0081
```

```
. graph box score, over (drug)
```



Non-parametric ”post-hoc” tests

```
. use wwiner, clear //
(Written by R.                    )

. signrank score_1 = score_2

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        3         9.5         7.5
    negative |        2         5.5         7.5
        zero |        0           0           0
-------------+---------------------------------
         all |        5          15          15

unadjusted variance       13.75
adjustment for ties       -0.63
adjustment for zeros       0.00
                     ----------
adjusted variance         13.13

Ho: score_1 = score_2
             z =   0.552
    Prob > |z| =   0.5809
```

Non-parametric "post-hoc" tests

```
. signrank score_2 = score_3

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        5          15         7.5
    negative |        0           0         7.5
        zero |        0           0           0
-------------+---------------------------------
         all |        5          15          15

unadjusted variance        13.75
adjustment for ties        -0.13
adjustment for zeros        0.00
                       ---------
adjusted variance          13.63

Ho: score_2 = score_3
             z =   2.032
    Prob > |z| =   0.0422
```