PubH 5483 – Fall 2004

EXAM

This exam runs from 11:15 am until 12:30 pm. There are 12 questions (5 Application and 7 Concepts) plus 2 extra credit problems.

ANSWER ONLY 10 OF THE 12 QUESTIONS.

Each question is worth 10 points. Please write all your answers in two blue exam books: Application answers in one blue book, and Concepts and Extra Credit answers in another. In all questions, be sure to define any notation you introduce. All questions relate to the following data set.

Data: Oregano, an herb commonly used in cooking, has some antimicrobial, antifungal, and antioxidant properties and has been linked in animal studies to blood pressure. A randomized trial is conducted in rats to test whether supplemental oregano (in the form of oregano oil in their food) improves systolic blood pressure compared to a placebo (an inactive oil in their food). 56 rats are randomly allocated to either oregano or placebo and their systolic blood pressure (SBP) is measured at the same time daily for one week.

The researchers would like to know:

- Do rats on oregano have, on average, lower SBP than rats on placebo? If so, by how much?
- Do rats on oregano have on average a significantly faster decline in SBP over the one week than rats on placebo? If so, by how much?

ANSWER ONLY 10 OF THE 12 QUESTIONS.

If you answer more than 10 questions, we will grade the first 10 we find in your blue books.

APPLICATION

A subset of the data are printed on page 6 so you can see how the data are structured. On pages 7–14 you will find some exploratory data analysis plots. On pages 15–18 you will find model output and chi-square critical values. On pages 19–23 you will find some diagnostic plots.

1. Using the plots on pages 7–8, briefly describe what the data indicate about:
   (a) an effect of treatment vs. placebo, and
   (b) an interaction between treatment and time.

2. Using the plots on pages 8–14, briefly describe what the data indicate about:
   (a) what General Linear Mixed Models (GLMMs) might be reasonable, and
   (b) what General Linear Models (GLMs) might be reasonable.

   You do NOT need to describe which predictors are reasonable to include. Just describe which covariance structures are reasonable to consider.

3. On pages 15–18 you will find the `PROC MIXED` output from two GLMMs for these data. The models are labeled `COVARIANCE MODEL` #1 and `COVARIANCE MODEL` #2. Decide which of the two covariance structures is more appropriate for these data as follows:
   (a) write down a null hypothesis (in words or in symbols) for comparing the two models to each other,
   (b) write down the corresponding test statistic (if a test is possible) or comparison of AIC or BIC (if a test is not possible), and
   (c) based on your answer to (b), write down your conclusion on which model is preferred.

4. Suppose we proceed with `COVARIANCE MODEL` #1. Provide answers to the two questions posed by the researchers on page 1. (Your answers need to contain more of the statistical results than just significance or lack thereof.)

5. Again suppose we proceed with `COVARIANCE MODEL` #1. Using the output on pages 15–16 and the plots on pages 19–23, briefly describe what they indicate about:
   (a) violations of normality,
   (b) outlying clusters or observations, and
   (c) goodness of fit for the covariance structure.

CONCEPTS

6. Consider `COVARIANCE MODEL` #1.

   (a) What does the parameter `UN(2,1)` correspond to?

   (b) When fitting a GLMM, why is it often reasonable to allow for such a parameter?

7. GLMMs allow for the computation of both population-averaged (marginal) fitted values and cluster-specific (conditional) fitted values.

   (a) Write down the formula for how each is computed.

   (b) How do their interpretations differ?

8. For `COVARIANCE MODEL` #1, compute:

   (a) $Var[Y_{ij}]$

   (b) $Cov[Y_{ij}, Y_{ij'}]$

   where i corresponds to rat and j corresponds to day.

9. In `COVARIANCE MODEL` #2, there are two variance components, one for the random intercept (σ_0^2) and one for the measurement error (σ_e^2).

   (a) Indicate which variance component corresponds to between-person variability and which corresponds to within-person variability.

   (b) Suppose $\sigma_0^2 > \sigma_e^2$. What do we gain by fitting a model with random intercepts, compared to fitting a model without random intercepts? Your answer must be very specific to get full credit.

10. Suppose we fit two GLMs to these data instead: one with a compound symmetry structure, and one with a compound symmmetry structure and `GROUP=TRT`. These two models could be compared to each other with a likelihood ratio test.

    (a) What (in words or in symbols) is the null hypothesis being tested?

    (b) How are the degrees of freedom for this test computed?

    Your answers must be very specific to get full credit.

11. Consider the two scientific questions of interest shown on p. 1. Suppose you were to fit derived variables models to these data (instead of GLMMs or GLMs).

    (a) What derived variable would you use to answer the first question?

    (b) What derived variable would you use to answer the second question?

    (c) What is one disadvantage of this derived variables model approach?

    Briefly justify your answers.

12. Name two conditions which must be met in order to carry out a standard likelihood ratio test (i.e., NOT one of the Special Case likelihood ratio tests). Which estimation technique is used (ML or REML) should NOT be one of your conditions.

EXTRA CREDIT

1. (2 points) What is the following LSMEANS output testing and in what way is it a relevant test for the rat study?

Least Squares Means

Effect	trt	day	Estimate	Standard Error
trt	OIL	0.00	160.83	0.4411
trt	PLACEBO	0.00	160.42	0.5283

Differences of Least Squares Means

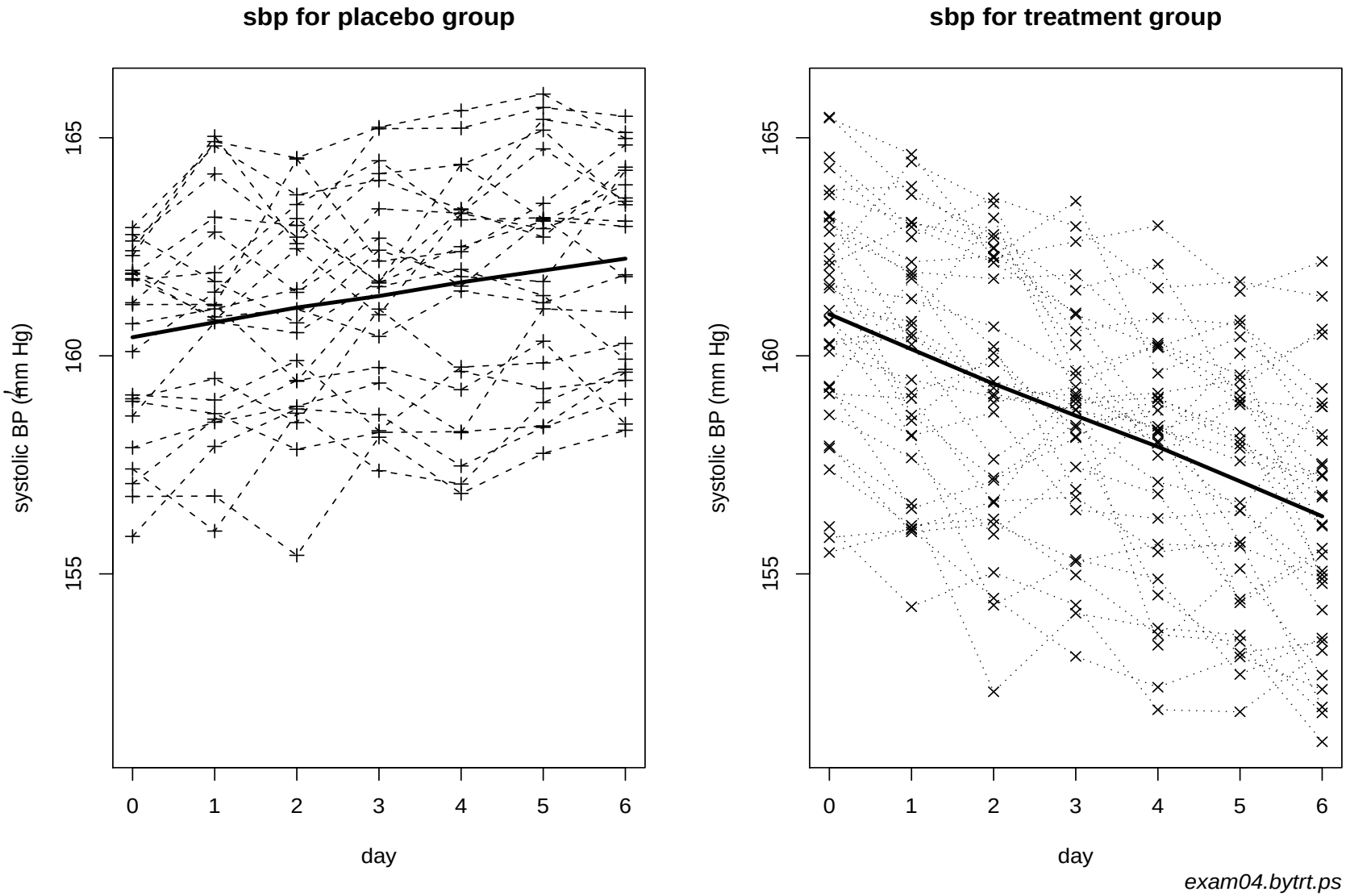
Effect	trt	_trt	day	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	OIL	PLACEBO	0.00	0.4075	0.6882	280	0.59	0.5542

2. (2 points) Write down the SAS code which generated `COVARIANCE MODEL` #1.

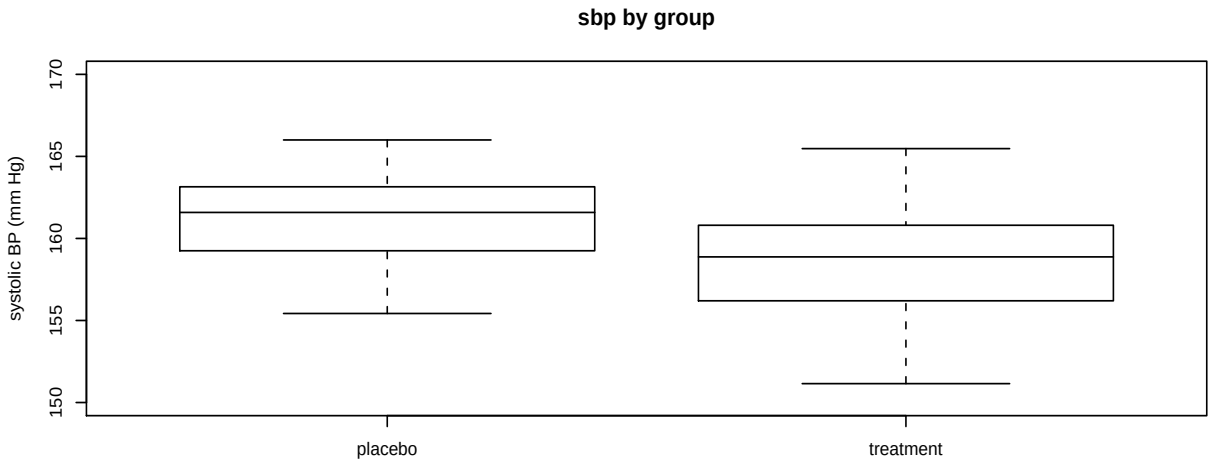
This page intentionally left blank.

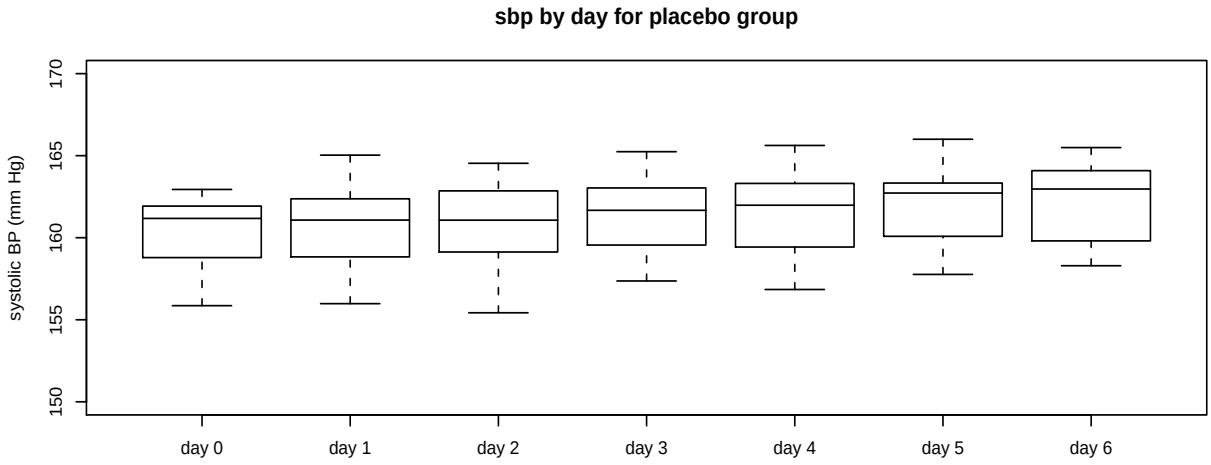
Partial printout of the rat data 1

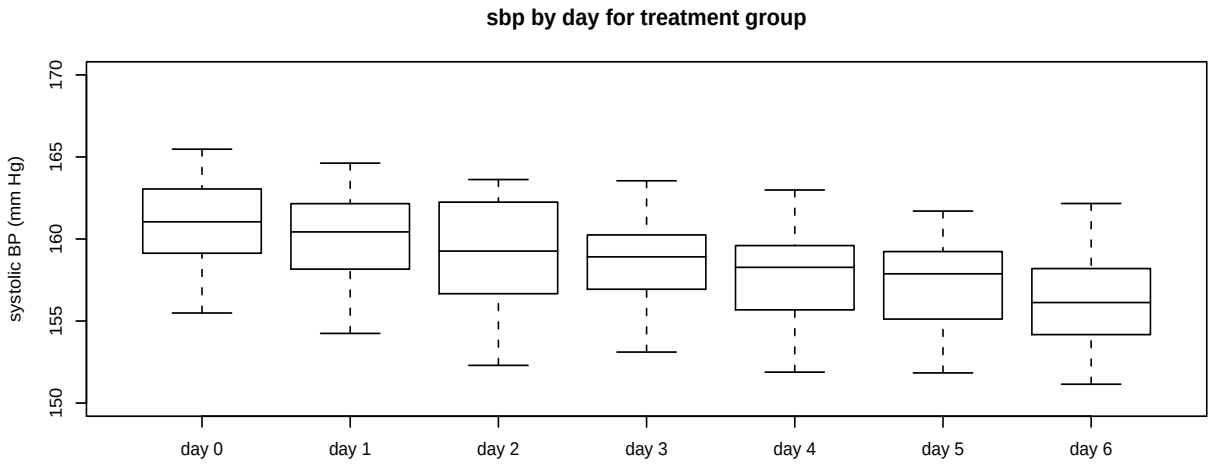
Obs	ratid	day	sbp	trt
1	1	0	164.559	OIL
2	1	1	162.955	OIL
3	1	2	162.141	OIL
4	1	3	160.985	OIL
5	1	4	158.233	OIL
6	1	5	159.062	OIL
7	1	6	156.127	OIL
8	2	0	160.736	PLACEBO
9	2	1	161.075	PLACEBO
10	2	2	161.530	PLACEBO
11	2	3	163.370	PLACEBO
12	2	4	163.272	PLACEBO
13	2	5	164.745	PLACEBO
14	2	6	163.544	PLACEBO
15	3	0	156.776	PLACEBO
16	3	1	156.785	PLACEBO
17	3	2	155.426	PLACEBO
18	3	3	158.133	PLACEBO
19	3	4	156.844	PLACEBO
20	3	5	157.764	PLACEBO
21	3	6	158.293	PLACEBO
22	4	0	158.957	PLACEBO
23	4	1	158.675	PLACEBO
24	4	2	157.856	PLACEBO
25	4	3	158.244	PLACEBO
26	4	4	158.262	PLACEBO
27	4	5	158.393	PLACEBO
28	4	6	159.625	PLACEBO
29	5	0	162.941	PLACEBO
30	5	1	164.812	PLACEBO
31	5	2	163.692	PLACEBO
32	5	3	164.023	PLACEBO
33	5	4	163.368	PLACEBO
34	5	5	165.421	PLACEBO
35	5	6	165.121	PLACEBO
36	6	0	161.624	OIL
37	6	1	161.305	OIL
38	6	2	160.668	OIL
39	6	3	159.542	OIL
40	6	4	157.998	OIL
41	6	5	156.452	OIL
42	6	6	158.921	OIL
43	7	0	161.554	OIL
44	7	1	160.508	OIL
45	7	2	159.264	OIL
46	7	3	159.078	OIL
47	7	4	158.273	OIL
48	7	5	157.877	OIL
49	7	6	157.268	OIL
50	8	0	157.936	OIL
51	8	1	155.963	OIL
52	8	2	156.137	OIL
53	8	3	155.330	OIL
54	8	4	154.885	OIL
55	8	5	152.693	OIL
56	8	6	153.530	OIL
57	9	0	160.256	OIL
58	9	1	160.431	OIL
59	9	2	159.122	OIL
60	9	3	158.771	OIL
61	9	4	158.043	OIL
62	9	5	160.065	OIL
63	9	6	154.776	OIL
64	10	0	161.869	PLACEBO
65	10	1	163.178	PLACEBO
66	10	2	162.988	PLACEBO
67	10	3	161.678	PLACEBO
68	10	4	164.389	PLACEBO
69	10	5	163.144	PLACEBO
70	10	6	161.814	PLACEBO



exam04.bytrt.ps

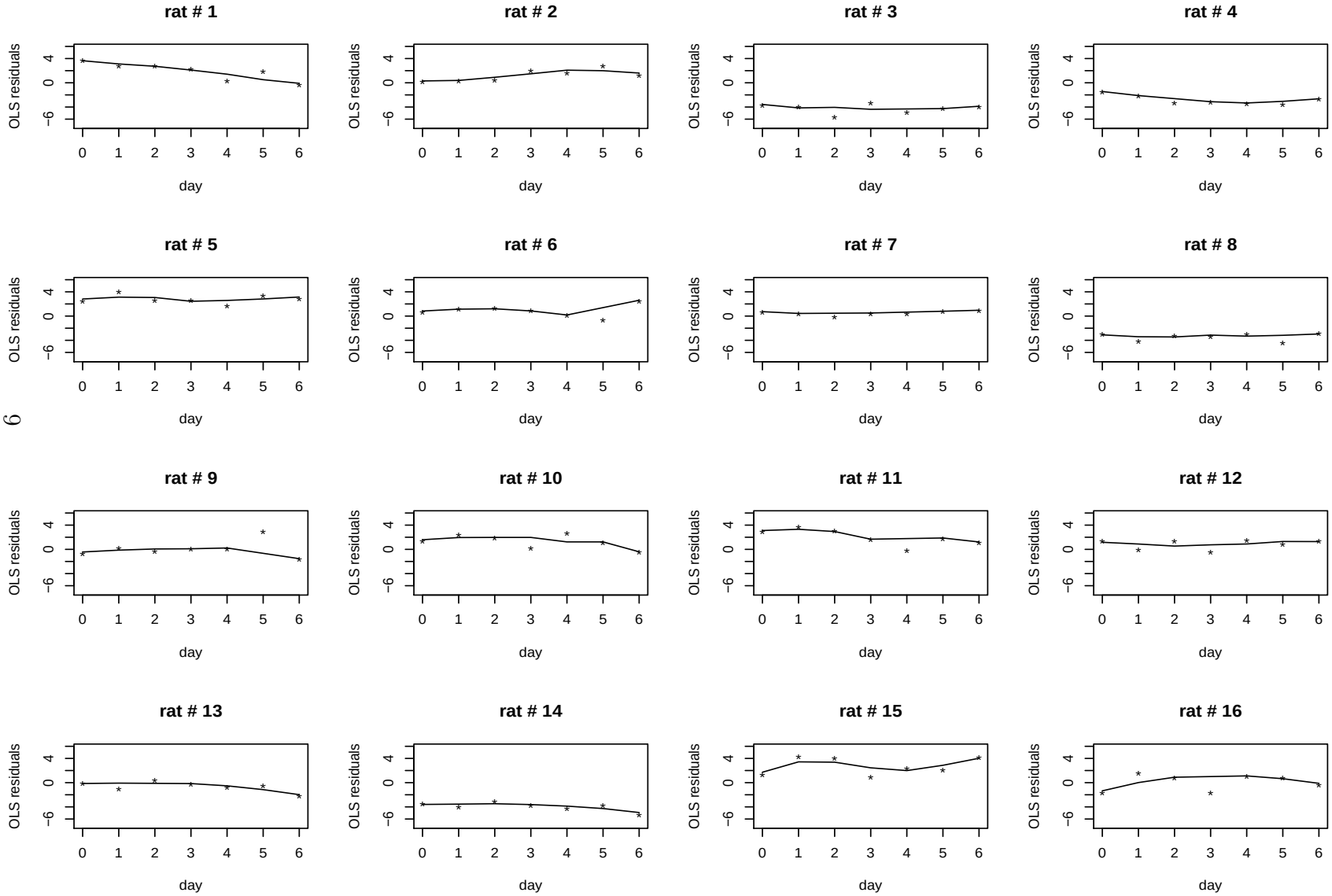




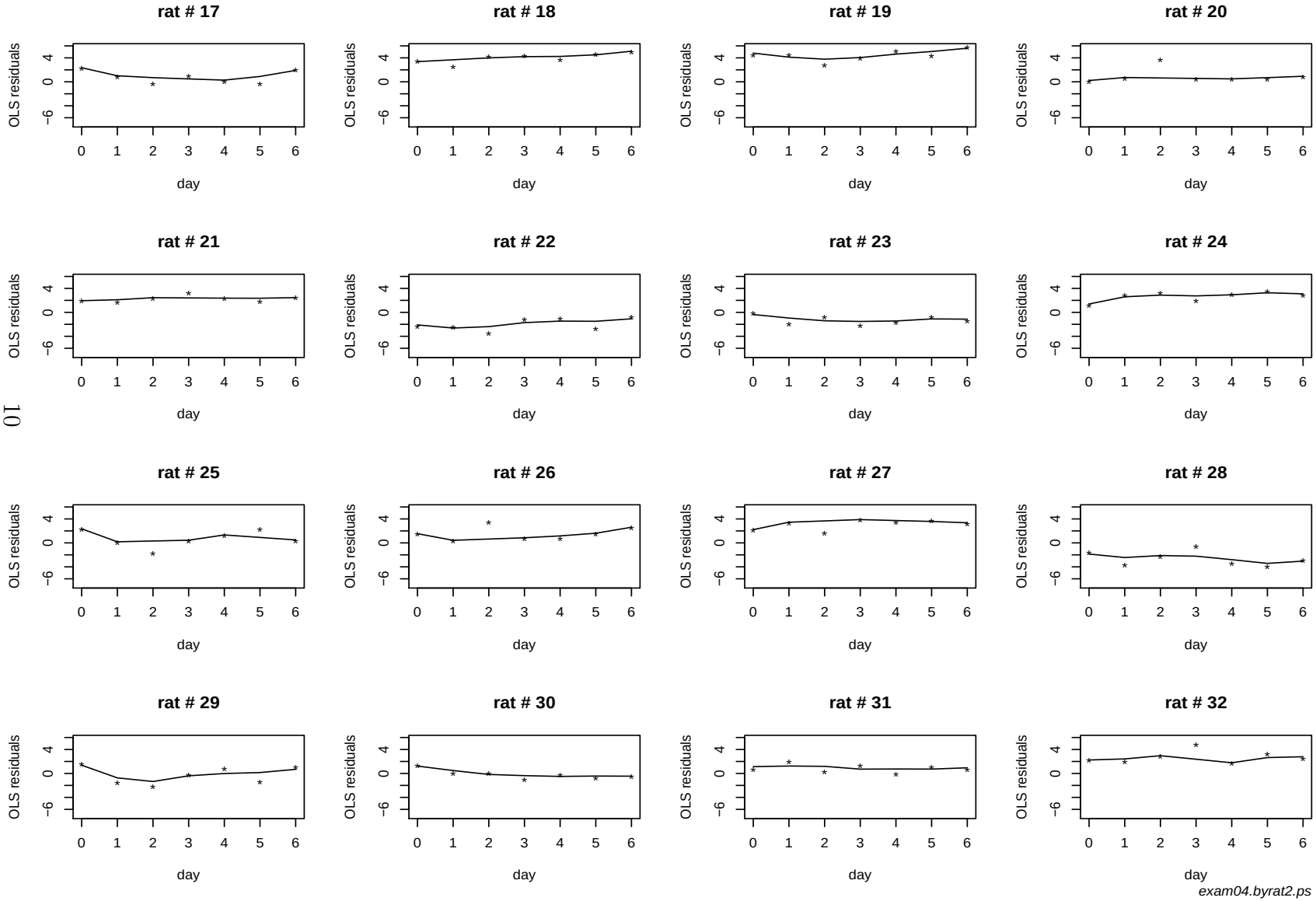


exam04.boxbytrt.ps

OLS residuals vs day by rat

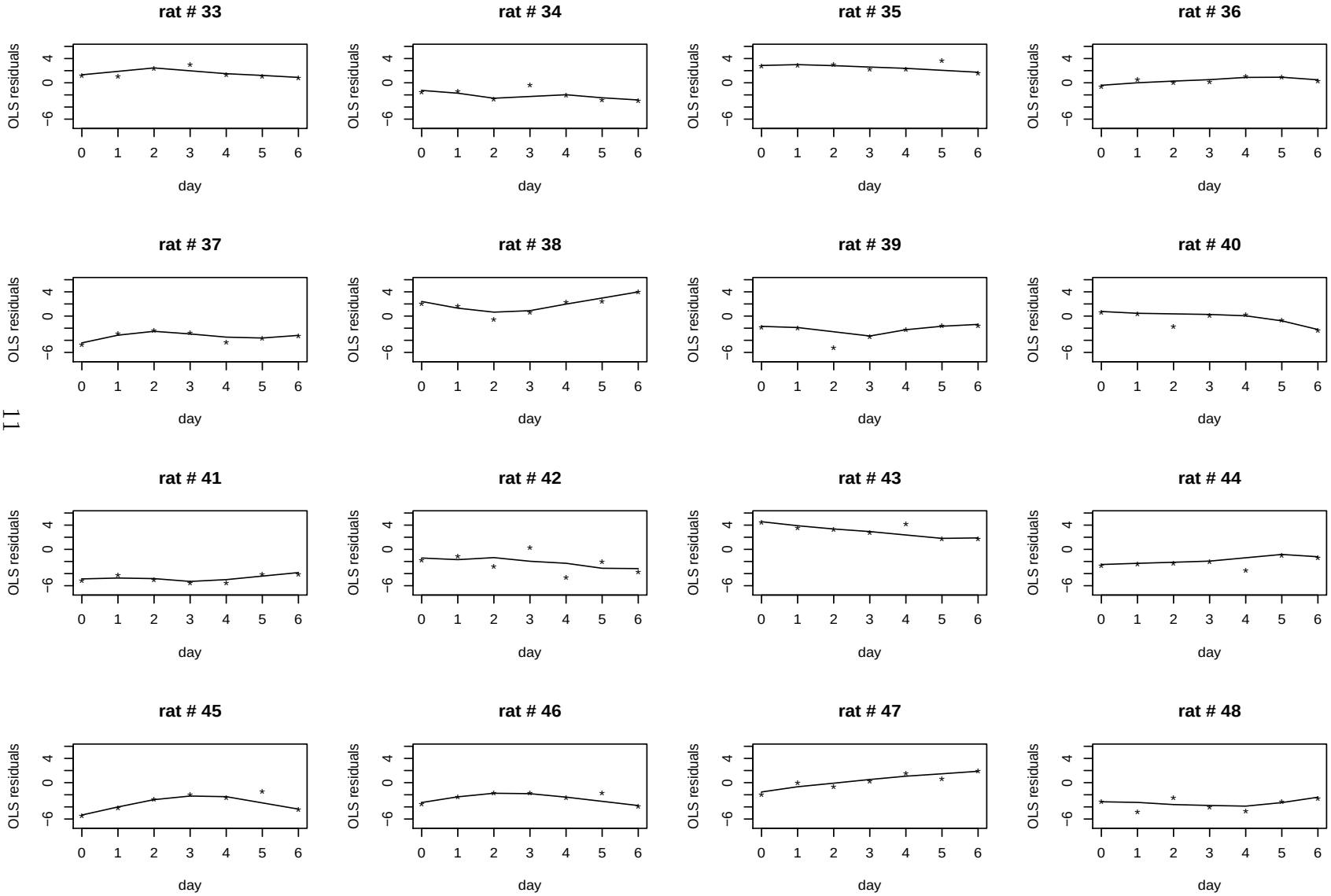


exam04.byrat1.ps

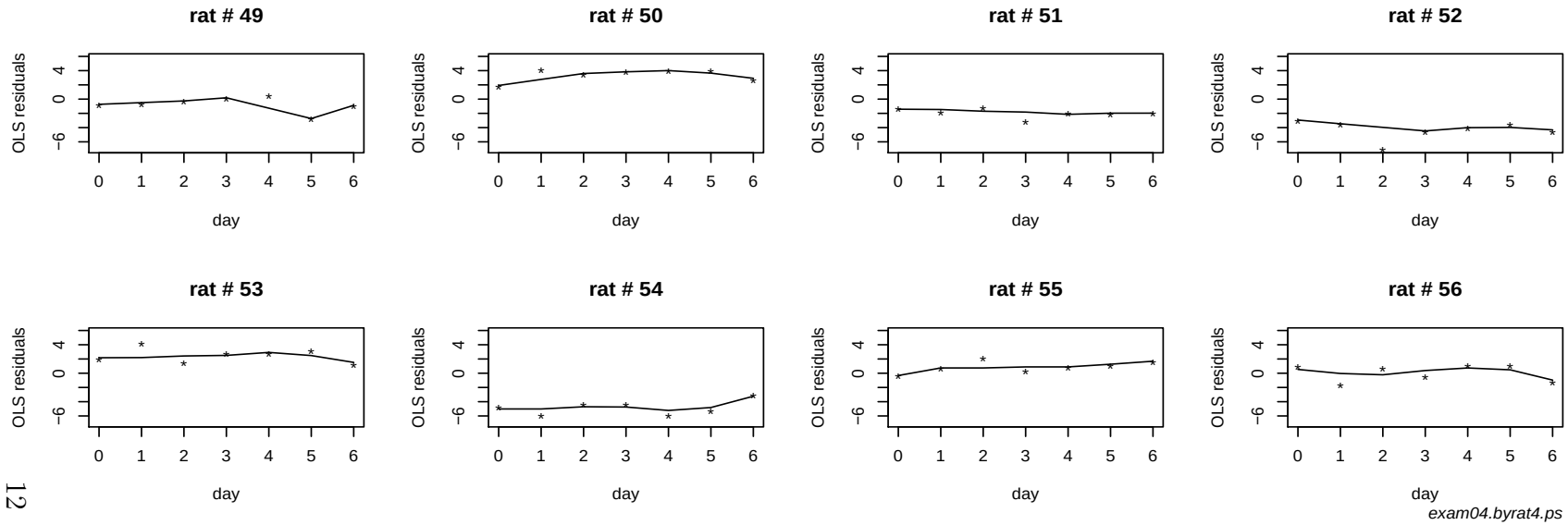


exam04.byrat2.ps

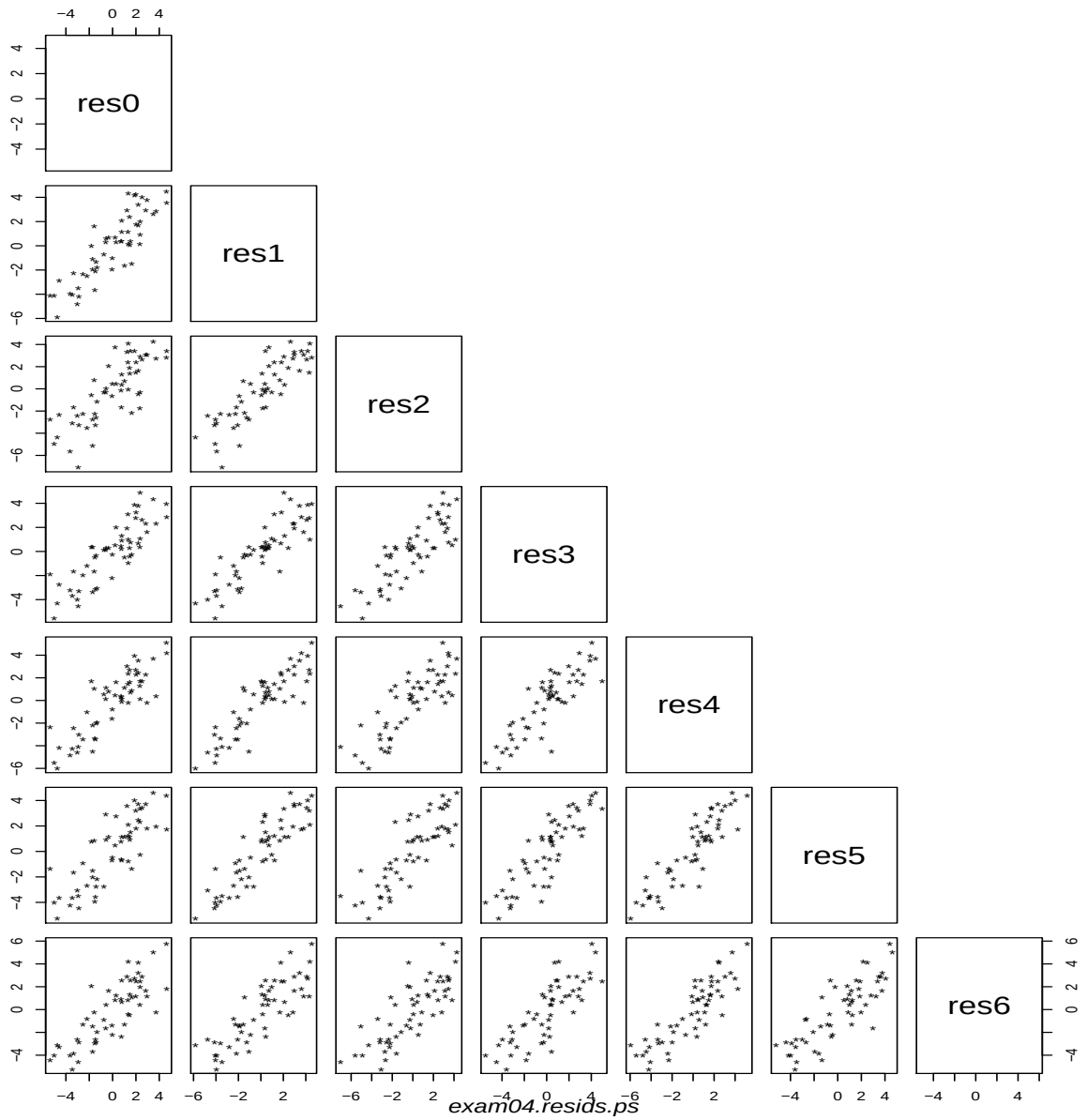
OLS residuals vs day by rat



exam04.byrat3.ps



exam04.byrat4.ps

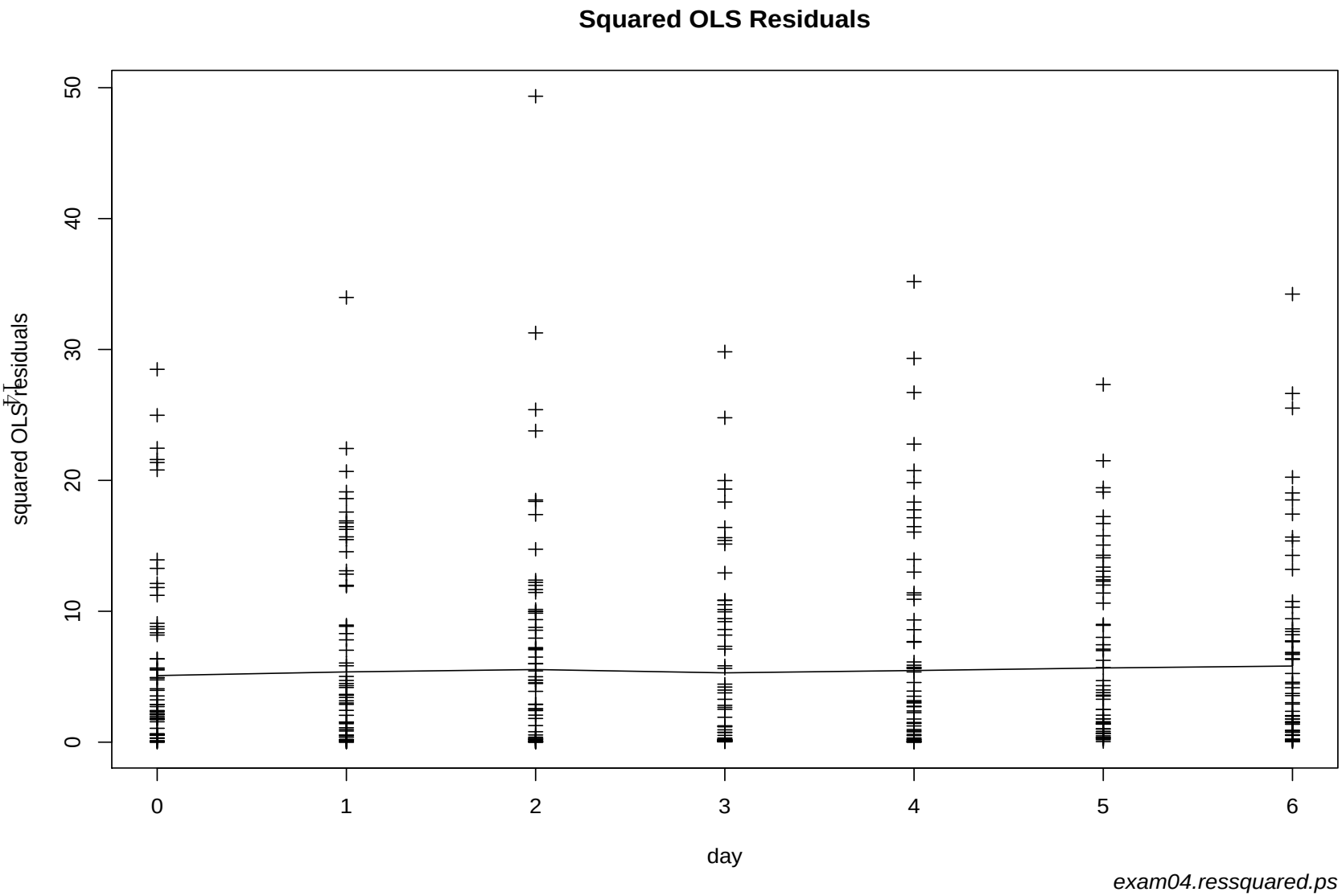


exam04.resids.ps

Correlation matrix of residuals:

	res0	res1	res2	res3	res4	res5	res6
res0	1.000	0.867	0.792	0.845	0.856	0.816	0.818
res1	0.867	1.000	0.856	0.872	0.887	0.893	0.836
res2	0.792	0.856	1.000	0.834	0.829	0.835	0.801
res3	0.845	0.872	0.834	1.000	0.858	0.862	0.808
res4	0.856	0.887	0.829	0.858	1.000	0.898	0.873
res5	0.816	0.893	0.835	0.862	0.898	1.000	0.833
res6	0.818	0.836	0.801	0.808	0.873	0.833	1.000

Variance overall by day: 6.21 7.348 7.763 6.174 7.459 7.069 7.136



exam04.ressquared.ps

The Mixed Procedure

Model Information

Data Set	WORK.REPDAT
Dependent Variable	sbp
Covariance Structure	Unstructured
Subject Effect	ratid
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
trt	2	OIL PLACEBO
ratid	56	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56

Dimensions

Covariance Parameters	4
Columns in X	6
Columns in Z Per Subject	2
Subjects	56
Max Obs Per Subject	7
Observations Used	392
Observations Not Used	0
Total Observations	392

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	1878.50343570	
1	1	1354.11793676	0.00000000

Convergence criteria met.

Estimated V Matrix for ratid 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	6.9770	5.9411	5.9459	5.9507	5.9554	5.9602	5.9650
2	5.9411	6.9963	5.9700	5.9844	5.9989	6.0133	6.0278
3	5.9459	5.9700	7.0348	6.0182	6.0423	6.0664	6.0905
4	5.9507	5.9844	6.0182	7.0927	6.0857	6.1195	6.1532
5	5.9554	5.9989	6.0423	6.0857	7.1698	6.1725	6.2159
6	5.9602	6.0133	6.0664	6.1195	6.1725	7.2663	6.2787
7	5.9650	6.0278	6.0905	6.1532	6.2159	6.2787	7.3821

Estimated V Correlation Matrix for ratid 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.8504	0.8487	0.8459	0.8420	0.8371	0.8312
2	0.8504	1.0000	0.8510	0.8495	0.8470	0.8434	0.8387
3	0.8487	0.8510	1.0000	0.8520	0.8508	0.8485	0.8452
4	0.8459	0.8495	0.8520	1.0000	0.8534	0.8524	0.8504
5	0.8420	0.8470	0.8508	0.8534	1.0000	0.8552	0.8544
6	0.8371	0.8434	0.8485	0.8524	0.8552	1.0000	0.8573

```
   7    0.8312    0.8387    0.8452    0.8504    0.8544    0.8573    1.0000
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	ratid	5.9363
UN(2,1)	ratid	0.004785
UN(2,2)	ratid	0.009658
Residual		1.0407

Fit Statistics

-2 Res Log Likelihood	1354.1
AIC (smaller is better)	1362.1
AICC (smaller is better)	1362.2
BIC (smaller is better)	1370.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	524.39	<.0001

Solution for Fixed Effects

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		160.42	0.5283	54	303.65	<.0001
trt	OIL	0.4075	0.6882	280	0.59	0.5542
trt	PLACEBO	0	.	.	.	.
day		0.2991	0.04512	54	6.63	<.0001
day*trt	OIL	-1.0516	0.05878	280	-17.89	<.0001
day*trt	PLACEBO	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	1	280	0.35	0.5542
day	1	54	59.50	<.0001
day*trt	1	280	320.09	<.0001

Least Squares Means

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	OIL	158.57	0.4335	280	365.81	<.0001
trt	PLACEBO	161.32	0.5192	280	310.69	<.0001

Differences of Least Squares Means

Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	OIL	PLACEBO	-2.7473	0.6764	280	-4.06	<.0001

```
================================================================================
```

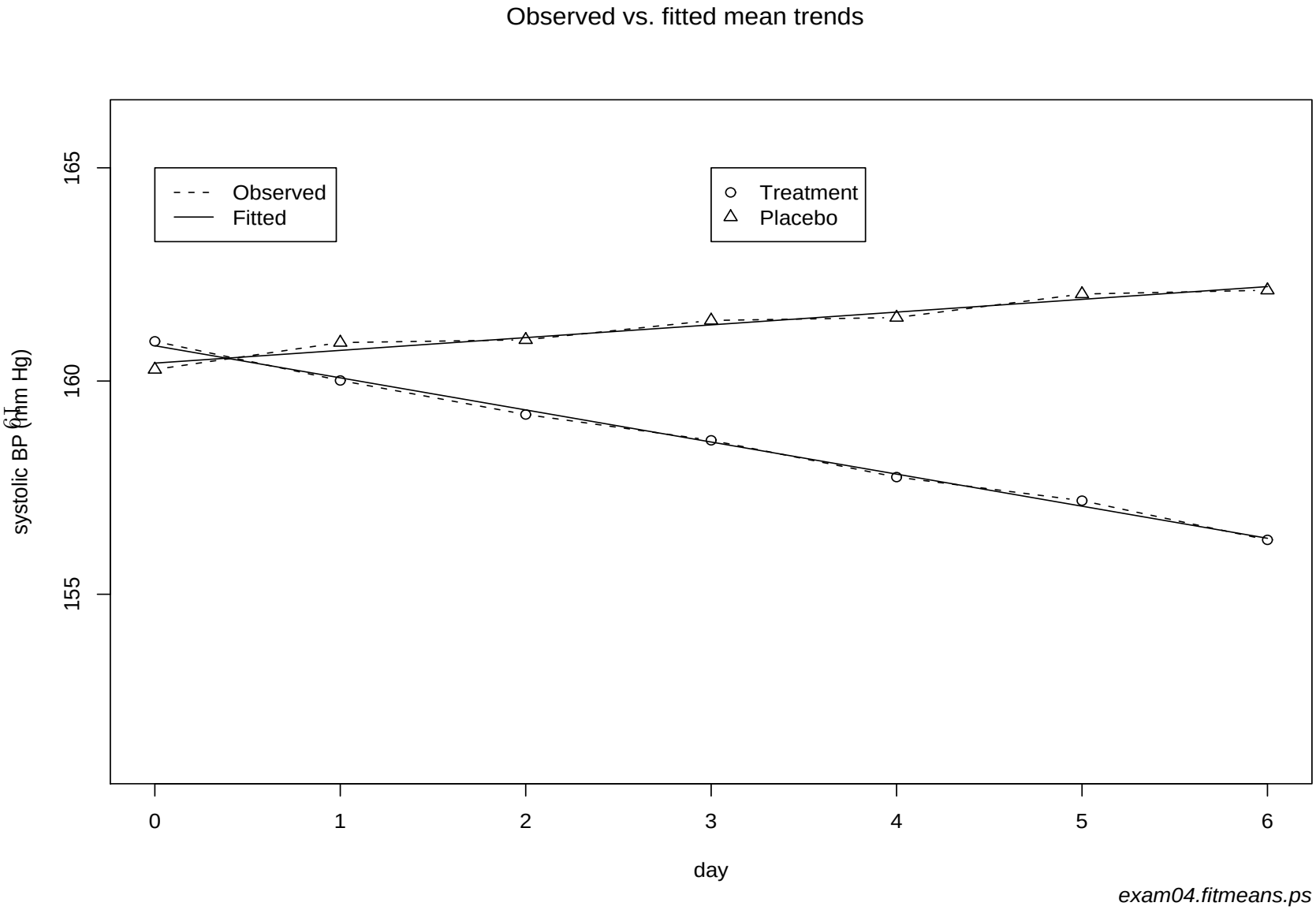
```
================================================================================

COVARIANCE MODEL #2                                                          2

The Mixed Procedure

                  Model Information

Data Set                     WORK.REPDAT
Dependent Variable           sbp
Covariance Structure         Variance Components
Subject Effect               ratid
Estimation Method            REML
Residual Variance Method     Profile
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Containment


            Class Level Information

Class    Levels    Values

trt           2    OIL PLACEBO
ratid        56    1 2 3 4 5 6 7 8 9 10 11 12 13
                   14 15 16 17 18 19 20 21 22 23
                   24 25 26 27 28 29 30 31 32 33
                   34 35 36 37 38 39 40 41 42 43
                   44 45 46 47 48 49 50 51 52 53
                   54 55 56


           Dimensions

Covariance Parameters             2
Columns in X                      6
Columns in Z Per Subject          1
Subjects                         56
Max Obs Per Subject               7
Observations Used               392
Observations Not Used             0
Total Observations              392


                    Iteration History

Iteration    Evaluations    -2 Res Log Like       Criterion

        0              1      1878.50343570
        1              1      1355.60175402      0.00000000


                  Convergence criteria met.


 Covariance Parameter Estimates

Cov Parm      Subject    Estimate

Intercept     ratid        6.0457
Residual                   1.0844


          Fit Statistics

-2 Res Log Likelihood          1355.6
AIC (smaller is better)        1359.6
AICC (smaller is better)       1359.6
BIC (smaller is better)        1363.7


                        Solution for Fixed Effects

                                    Standard
Effect       trt        Estimate       Error      DF    t Value    Pr > |t|

Intercept                160.42      0.5336      54     300.63      <.0001
trt          OIL          0.4075      0.6951     334       0.59      0.5581
```

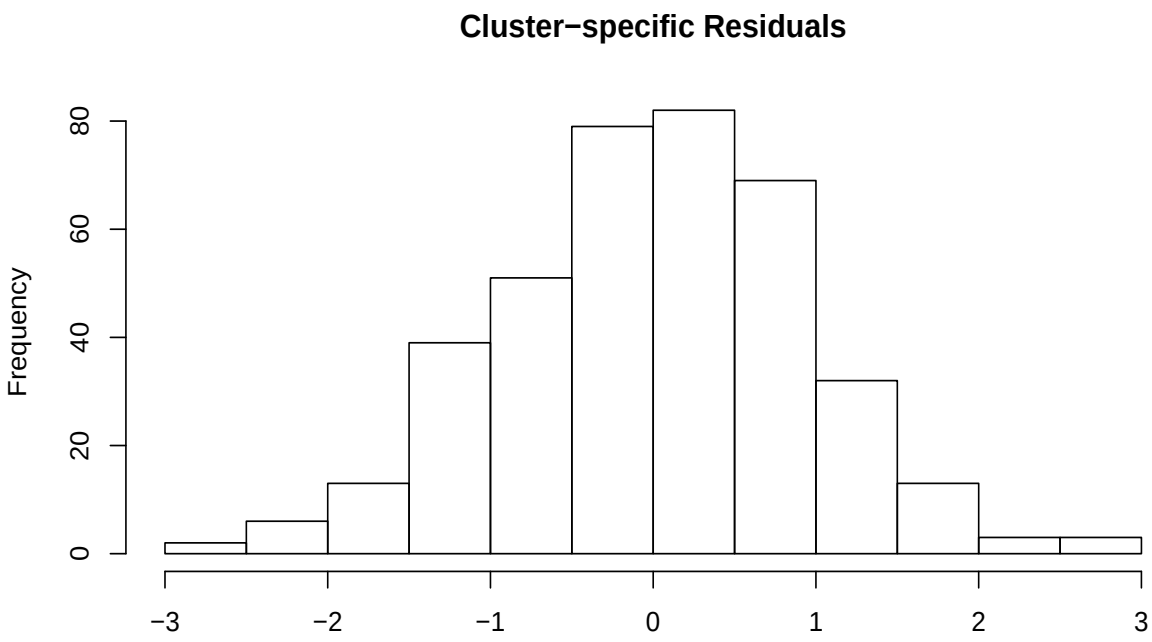
```
trt          PLACEBO           0          .       .        .        .
day                         0.2991     0.04104     334     7.29     <.0001
day*trt      OIL           -1.0516     0.05346     334   -19.67     <.0001
day*trt      PLACEBO           0          .       .        .        .


        Type 3 Tests of Fixed Effects

               Num     Den
Effect          DF      DF    F Value    Pr > F

trt              1     334       0.34    0.5581
day              1     334      71.94    <.0001
day*trt          1     334     387.01    <.0001


                           Least Squares Means

                                    Standard
Effect    trt        Estimate       Error      DF    t Value    Pr > |t|

trt       OIL          158.57      0.4335     334     365.81      <.0001
trt       PLACEBO      161.32      0.5192     334     310.69      <.0001


                    Differences of Least Squares Means

                                          Standard
Effect   trt       _trt      Estimate      Error     DF   t Value   Pr > |t|

trt      OIL       PLACEBO    -2.7473     0.6764    334     -4.06     <.0001

================================================================================
```

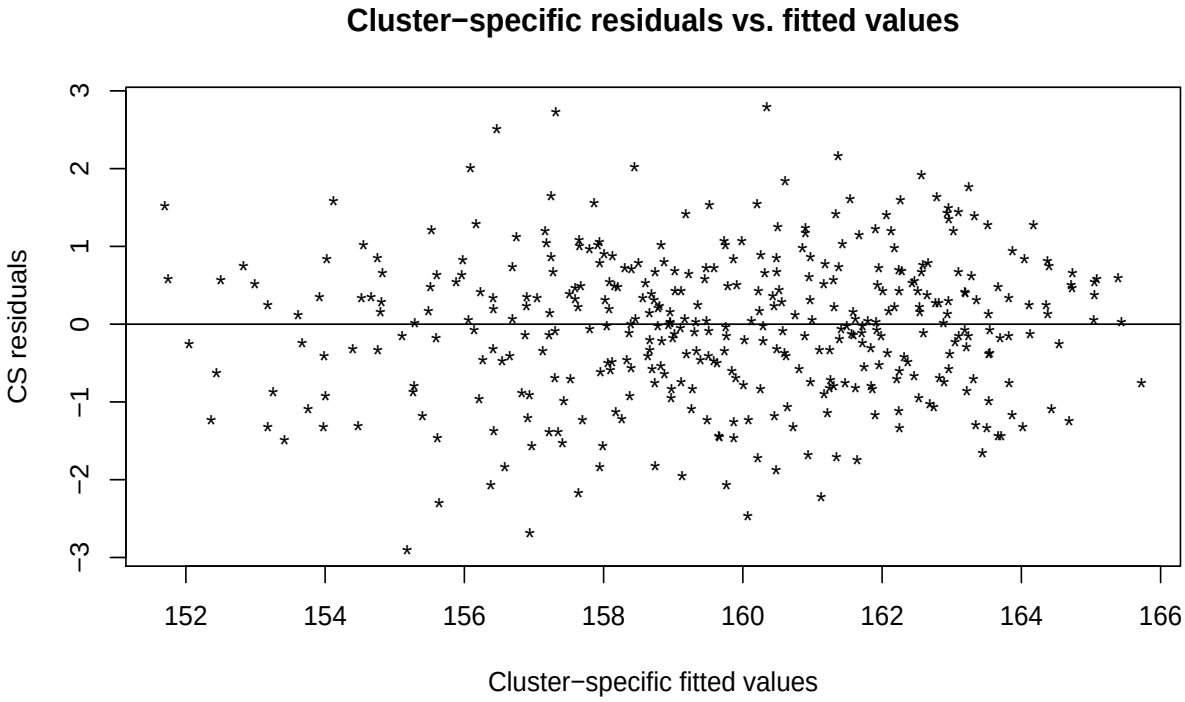
0.05 quantile cut-points for χ^2 distributions:

Distribution	Quantile cut-point
$\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$	2.72
χ^2_1	3.84
$\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$	5.14
χ^2_2	5.99

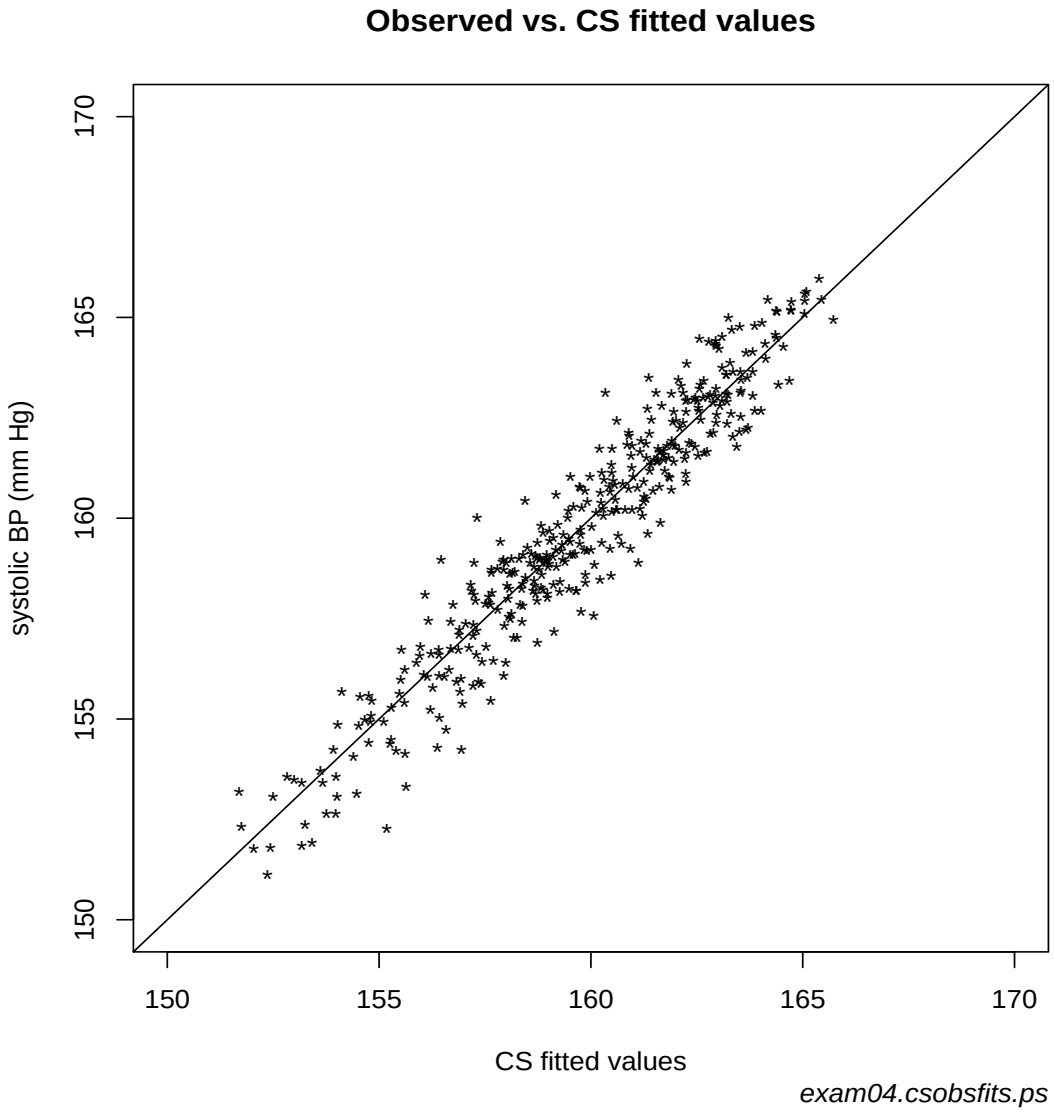
You do not need to compute p-values for any likelihood ratio tests you report. I have not given you sufficient tables to do that.


Observed vs. fitted mean trends
Observed
Fitted
Treatment
Placebo
systolic BP (mm Hg)
165
160
155
0
1
2
3
4
5
6
day
exam04.fitmeans.ps

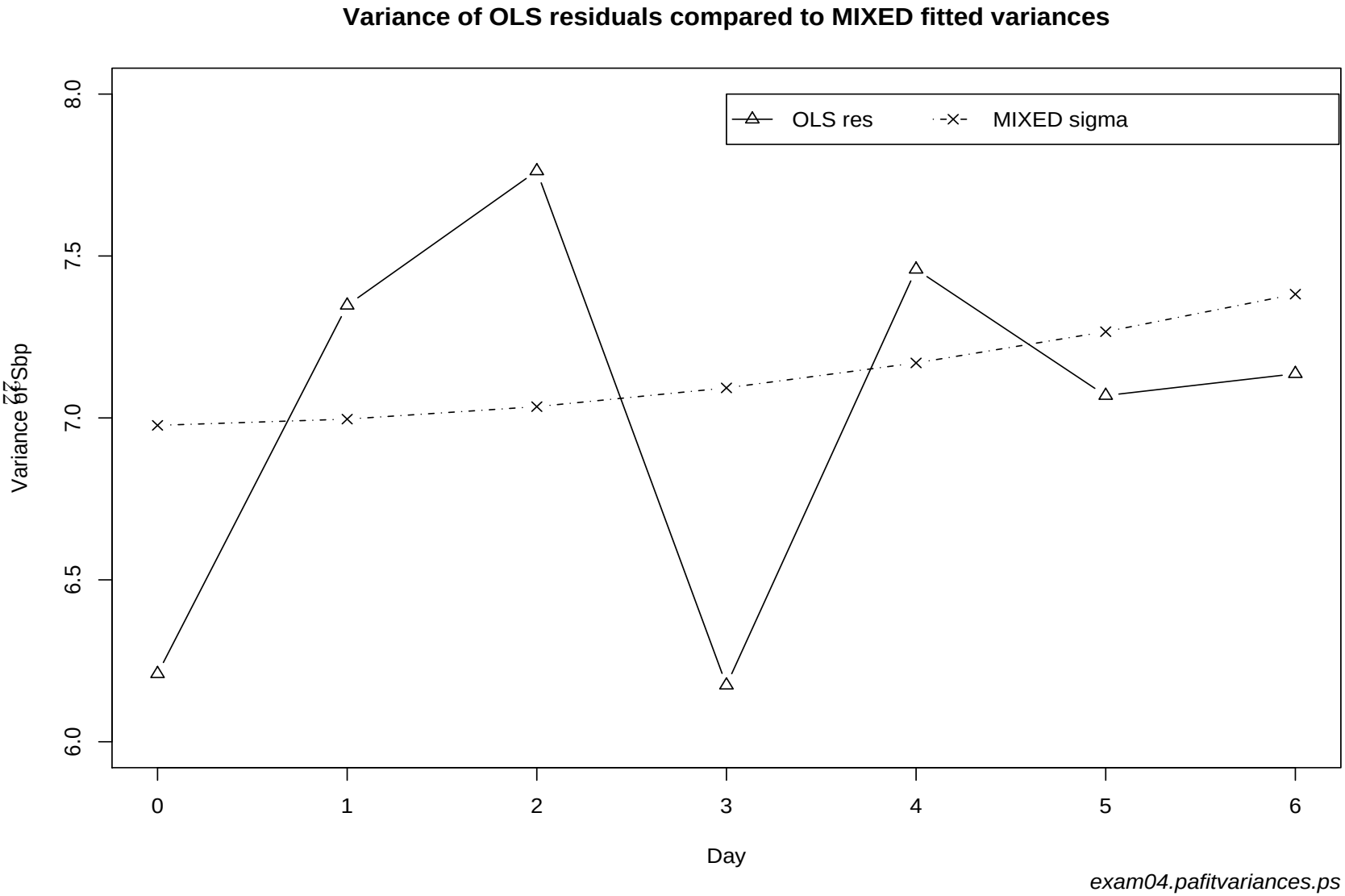


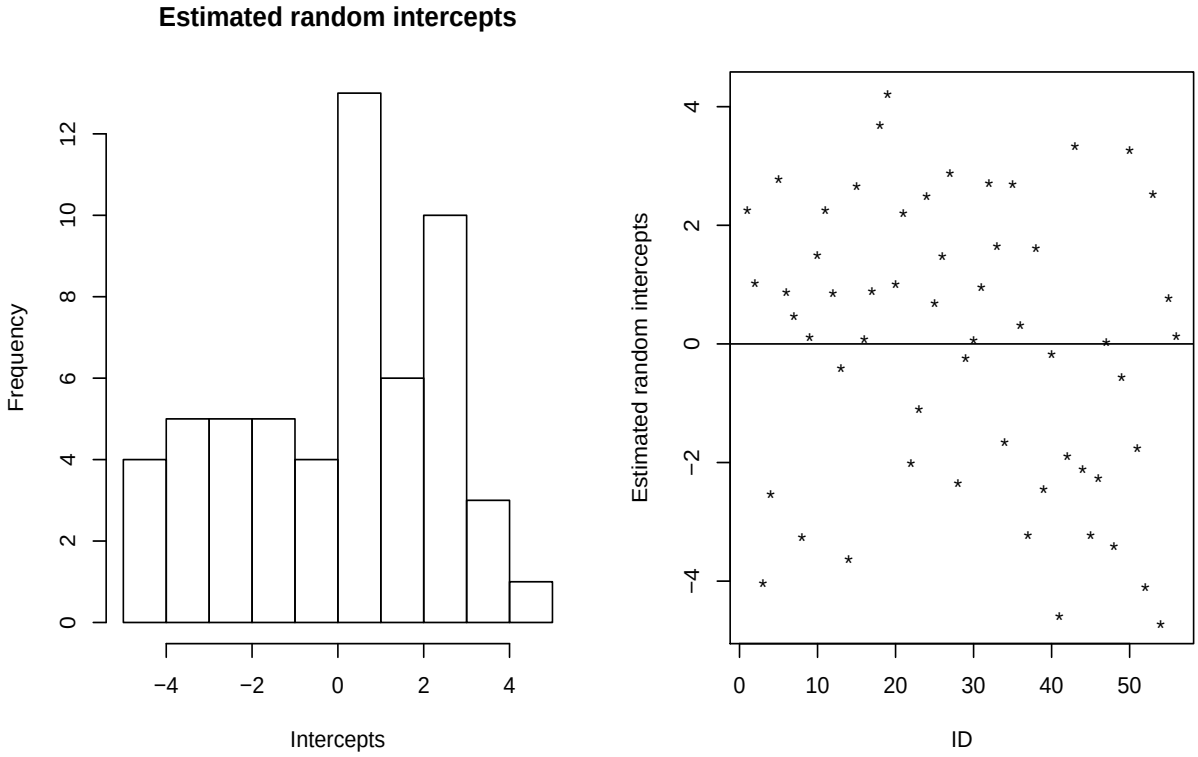


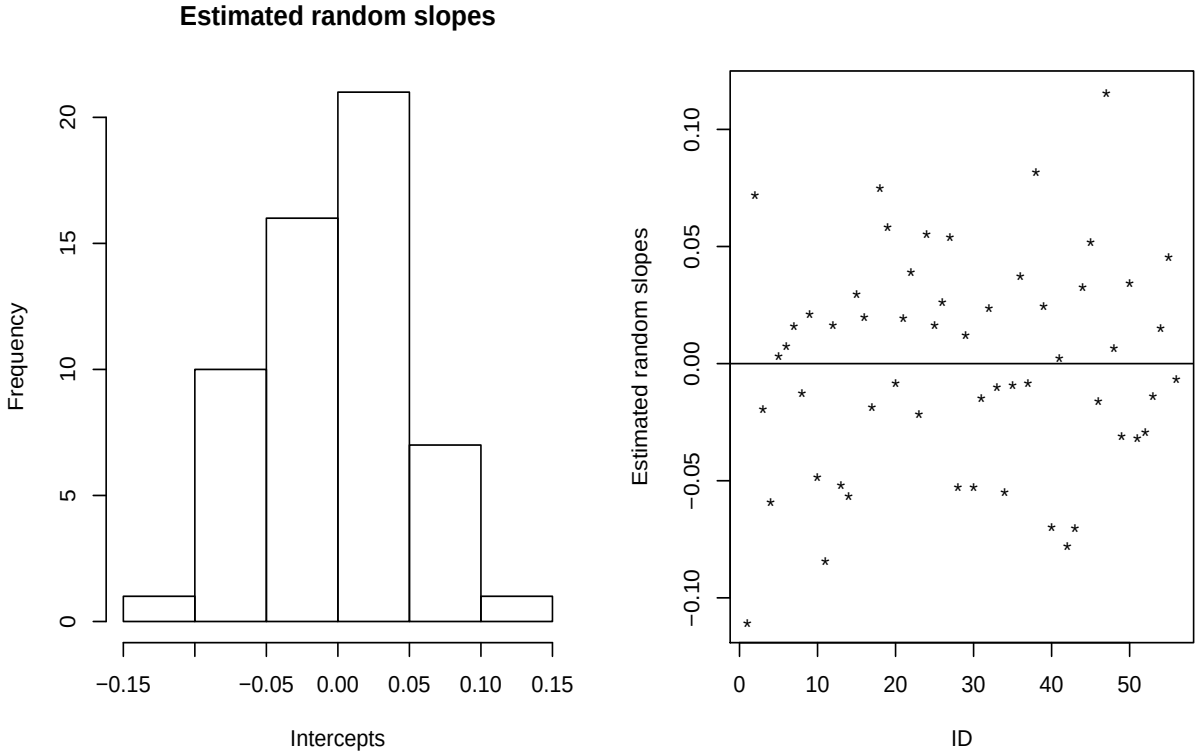
exam04.csresfits.ps



exam04.csobsfits.ps


Variance of OLS residuals compared to MIXED fitted variances
OLS res
MIXED sigma
Variance of Sbp
8.0
7.5
7.0
6.5
6.0
0
1
2
3
4
5
6
Day
exam04.pafitvariances.ps





exam04.rehist.ps