Biostats VHM 802 Course
Winter 2014, Atlantic Veterinary College, PEI
Henrik Stryhn

Final examination, 22 April 2014

All aids are allowed, except personal computer and personal assistance. The exam consists of 3 questions, which have equal weight (*10 points each*) and should all be answered; further detail about the points is given for specific parts of each question. The duration of the exam is 3 hours.

Generally, all statistical models used should be specified, and to such detail that it is clear which terms are present and in which form. Your answers should generally (unless specified otherwise) be based on the information provided. Nevertheless, if at some point you think it is necessary to carry out additional analysis in statistical software, explain carefully the purpose of your proposed analysis and how you would implement it in the statistical software.

Question 1.

An animal nutrition study was conducted to investigate the effects of protein in the diet on the level of leucine (an "essential amino acid") in the plasma of pigs. Pigs were randomly assigned to one of twelve protein diets: the combinations of protein source (fish meal, soybean meal, or dried skim milk) and protein concentration (9, 12, 15, or 18%). The response measured was the free plasma leucine level, in mcg/ml. The table shows mean leucine levels for all diet combinations, each obtained by averaging values of 3 pigs (except the 9% fish meal protein group where only 2 pigs completed the trial).

Mean leucine	Protein source			
Protein conc.	Fish meal	Dried skim milk	Soybean meal	Mean
9%	25.75	35.40	34.63	32.70
12%	30.93	43.47	39.63	38.01
15%	30.33	49.93	39.23	39.83
18%	32.33	66.50	45.43	48.09
Mean	30.21	48.82	39.73	39.86

A) (*4 points*) Describe the study type and its design in statistical terms, using standard descriptors such as factors, treatments, replication, blocks, balancedness, completeness, treatments, experimental units etc. Describe the statistical model for which results are presented in the Minitab and Stata listings, and draw general conclusions about the impact of protein source and concentration on leucine levels in pigs (postpone any specific comparisons to parts B)−C)).

B) (*2 points*) Use the information provided and any relevant additional calculations to compare the impact of different protein sources on leucine levels. As part of the discussion it should be addressed whether protein source appears to have the same impact in different concentrations; if not, any such differences should be described. (*Note:* To facilitate calculations you may carry out such comparisons as if also the 9% fish meal protein group had included 3 pigs.)

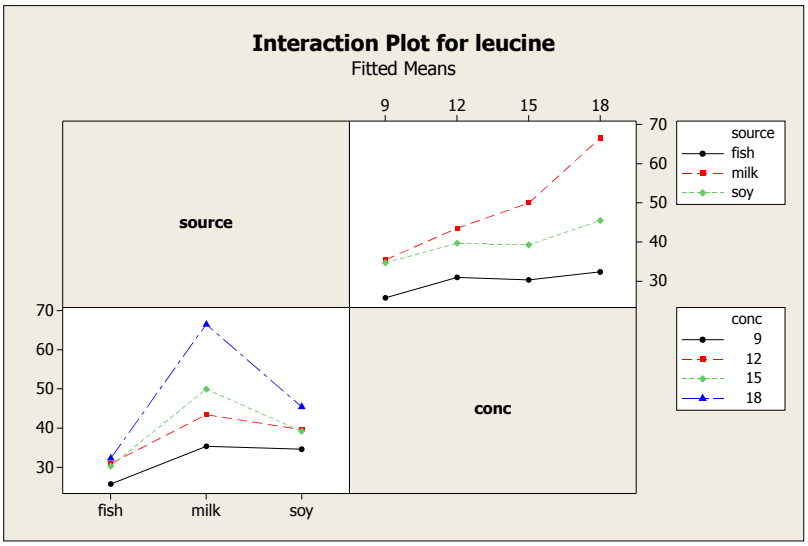
C) (*2 points*) Use the information provided and any relevant additional calculations to compare the impact of different protein concentrations on leucine levels. As part of the discussion it should be addressed whether protein concentration appears to have the same impact when originating from different protein sources; if not, any such differences should be described. In addition, explore whether the data seem to support a linear modelling of protein concentrations. (*Note:*

Depending on your conclusions up till this point, you may explore this either overall across all protein sources or for a single, purposefully selected, protein source. Also in this part, to facilitate calculations you may assume that the 9% fish meal protein group included 3 pigs.)

D) (*2 points*) In planning of a future experiment it was considered to use pigs from the same litter instead of pigs selected across a population. For the sake of the discussion, assume that it is feasible to select 4 pigs per litter from a number of litters for the study. Discuss how could this be used to refine the design of an experiment with the same factors and factor levels. In particular, describe the resulting data structure and the impact such a change in the design would be expected to have on comparisons of protein sources and concentrations.

Minitab listing and graph for Question 1 – note also the Stata listing below:

```
MTB > GLM 'leucine' = source | conc;
General Linear Model: leucine versus source, conc

Factor  Type   Levels  Values
source  fixed       3  fish, milk, soy
conc    fixed       4  9, 12, 15, 18

Analysis of Variance for leucine, using Adjusted SS for Tests

Source       DF   Seq SS   Adj SS   Adj MS      F      P
source        2  1989.19  2036.10  1018.05  48.64  0.000
conc          3  1196.93  1125.12   375.04  17.92  0.000
source*conc   6   602.00   602.00   100.33   4.79  0.003
Error        23   481.37   481.37    20.93
Total        34  4269.49

S = 4.57481   R-Sq = 88.73%   R-Sq(adj) = 83.33%

Unusual Observations for leucine

Obs  leucine      Fit  SE Fit  Residual  St Resid
 30  59.5000  49.9333  2.6413    9.5667      2.56 R
 32  41.4000  49.9333  2.6413   -8.5333     -2.28 R

R denotes an observation with a large standardized residual.
```



Stata listing for Question 1 – note also the Minitab listing and graph above:

```
. regress leucine source##conc

      Source |       SS       df       MS              Number of obs =      35
-------------+------------------------------           F( 11,    23) =   16.45
       Model |  3788.12065    11  344.374605            Prob > F      =  0.0000
    Residual |   481.36489    23  20.9289083            R-squared     =  0.8873
-------------+------------------------------           Adj R-squared =  0.8333
       Total |  4269.48554    34  125.573104            Root MSE      =  4.5748

------------------------------------------------------------------------------
     leucine |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      source |
       milk  |   9.649999   4.176213     2.31   0.030     1.010844    18.28915
        soy  |   8.883333   4.176213     2.13   0.044     .2441773    17.52249
             |
        conc |
         12  |   5.183333   4.176213     1.24   0.227    -3.455822    13.82249
         15  |   4.583333   4.176213     1.10   0.284    -4.055822    13.22249
         18  |   6.583334   4.176213     1.58   0.129    -2.055821    15.22249
             |
 source#conc |
   milk#12   |   2.883335   5.602978     0.51   0.612    -8.707308    14.47398
   milk#15   |   9.950002   5.602978     1.78   0.089    -1.640641    21.54065
   milk#18   |   24.51667   5.602978     4.38   0.000     12.92602    36.10731
    soy#12   |  -.1833331   5.602978    -0.03   0.974    -11.77398    11.40731
    soy#15   |   .0166677   5.602978     0.00   0.998    -11.57398    11.60731
    soy#18   |   4.216668   5.602978     0.75   0.459    -7.373975    15.80731
             |
       _cons |      25.75   3.234881     7.96   0.000     19.05814    32.44186
------------------------------------------------------------------------------

. testparm i.source
 ( 1)  2.source = 0
 ( 2)  3.source = 0

       F(  2,    23) =    3.10
            Prob > F =    0.0643

. testparm i.conc
 ( 1)  12.conc = 0
 ( 2)  15.conc = 0
 ( 3)  18.conc = 0

       F(  3,    23) =    0.87
            Prob > F =    0.4685

. testparm source#conc
 ( 1)  2.source#12.conc = 0
 ( 2)  2.source#15.conc = 0
 ( 3)  2.source#18.conc = 0
 ( 4)  3.source#12.conc = 0
 ( 5)  3.source#15.conc = 0
 ( 6)  3.source#18.conc = 0

       F(  6,    23) =    4.79
            Prob > F =    0.0026
```

Question 2.

A study was carried out to investigate the variability of aflatoxin concentrations in peanuts. The study was motivated as follows (in the article reporting it),

> The wide variation of aflatoxin results obtained on independent samples from a given lot of peanuts have caused concern regarding the adequacy of a 10 lb sample to characterize the true aflatoxin content of one lot of peanuts. Ten pounds has been used as the standard sample size for aflatoxin assay. This study was undertaken the determine the magnitude of the sampling variation [...]. It is important to know the sampling variation in order to properly characterize peanut lots for their suitability in making manufactured products.

According to one theory, the measured variability in aflatoxin concentrations is due to a relatively small number of contaminated kernels (peanuts), which however vary greatly in contamination levels. Based on this theory one would expect aflatoxin concentrations to show substantial small-sample variation and limited additional large-sample variation. The study was designed to explore this theory. In the study, the aflatoxin concentration was determined by two analytic methods, here denoted by the acronyms `BF` and `CB`, and a comparison between the two methods was also part of the study objective.

A large shipment of peanuts comprising 800 bags (100 000 lb) was divided into 16 sections consisting of 50 bags per section. Four handfuls of peanuts were removed from each bag in the section and composited to give a sample of about 20 lb. This sample was mixed well and split into two subsamples (denoted A and B) that were ground in a mill and prepared for biochemical analysis. Each subsample was analyzed by both the `BF` and `CB` methods. The concentrations of aflatoxin (in ppb, parts per billion) for the 32 subsamples analyzed by the two methods are shown in the table below.

Section/	Subsample A		Subsample B	
Method	`BF`	`CB`	`BF`	`CB`
1	22	36	9	26
2	20	35	336	286
3	65	62	30	31
4	8	16	11	26
5	10	11	20	89
6	24	36	20	89
7	27	36	42	61
8	32	47	10	0
9	49	78	18	15
10	16	42	10	8
11	19	11	31	26
12	42	52	39	52
13	57	66	10	19
14	20	11	30	56
15	10	21	45	103
16	15	37	40	76

A) (*2 points*) Describe the study type and its design in statistical terms, using standard descriptors such as factors, treatments, replication, blocks, balancedness, completeness, treatments, experimental units etc. As part of your description, address specifically whether the data have a hierarchical structure, and if so use a suitable diagram to illustrate the structure.

B) (*2 points*) Formulate a statistical model (with its assumptions) for the data that will allow estimation of the effect(s) and variation(s) described above. Make sure to identify which parts

of the model correspond to the different effect(s) and variation(s) of interest. Your model may coincide with the model behind the subsequent software listings, but is not necessarily required to do so.

C) (*1 point*) The analysis was carried out in two versions, one for the original aflatoxin concentrations and one for log-transformed concentrations. The value 0 in Section 8 poses a problem for the log-transformation. Indicate which of the following ways of dealing with this value you think would be valid in this context (no justification required, but false answers give deductions; more than one answer may be acceptable),

i) remove the observation,

ii) replace the observation by a value very close to 0 (such as 0.01),

iii) replace the observation by the lowest non-zero value in the dataset,

iv) replace the observation by the lower detection limit for the `CB` method (if that value is known),

v) replace the observation by the average concentration for Section 8,

vi) other suggestion(s).

D) (*2 points*) Use the attached output from analyses on either original or log-transformed scale in Minitab and Stata to answer this and the following question. Explain the statistical model used, briefly if it coincides with your model from part B), otherwise in full detail with all model assumptions. Discuss based on the information provided on which of the two scales you think it is preferable to try to answer the questions the study tries to address (as described above).

E) (*2 points*) Use the results (from your preferred scale) to draw conclusions about the effect(s) and variation(s) referred to in the study motivation and objective. Make sure you include both quantitative and significance statements about these effect(s) and variation(s). In your view, do the results support the theory put forward about the variability in aflatoxin concentrations?

F) (*1 point*) For the comparison between the two analytic methods, the researchers were also interested in whether the two methods had the same analytical precision (or variability). Can you answer this question from the information provided? Explain your answer.

Minitab listing and graphs for Question 2, parts D) and E) – note also the Stata listing below:

```
MTB > GLM 'aflatoxin' = section subsample(section) method;
SUBC>   Random 'section';
SUBC>   Brief 2 ;
SUBC>   EMS;
SUBC>   Means method;
SUBC>   GFourpack;
SUBC>   RType 2 .
General Linear Model: aflatoxin versus section, method, subsample

Factor              Type    Levels  Values
section             random      16  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
subsample(section)  random      32  A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B,
                                    A, B, A, B, A, B, A, B, A, B, A, B, A, B
method              fixed        2  BF, CB
```

Analysis of Variance for aflatoxin, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
section	15	75649.0	75649.0	5043.3	0.84	0.629
subsample(section)	16	95967.2	95967.2	5998.0	22.64	0.000
method	1	2795.8	2795.8	2795.8	10.55	0.003
Error	31	8211.7	8211.7	264.9		
Total	63	182623.7				

S = 16.2756 R-Sq = 95.50% R-Sq(adj) = 90.86%

Unusual Observations for aflatoxin

Obs	aflatoxin	Fit	SE Fit	Residual	St Resid
7	336.000	304.391	11.687	31.609	2.79 R
8	286.000	317.609	11.687	-31.609	-2.79 R
19	20.000	47.891	11.687	-27.891	-2.46 R
20	89.000	61.109	11.687	27.891	2.46 R
23	20.000	47.891	11.687	-27.891	-2.46 R
24	89.000	61.109	11.687	27.891	2.46 R

R denotes an observation with a large standardized residual.

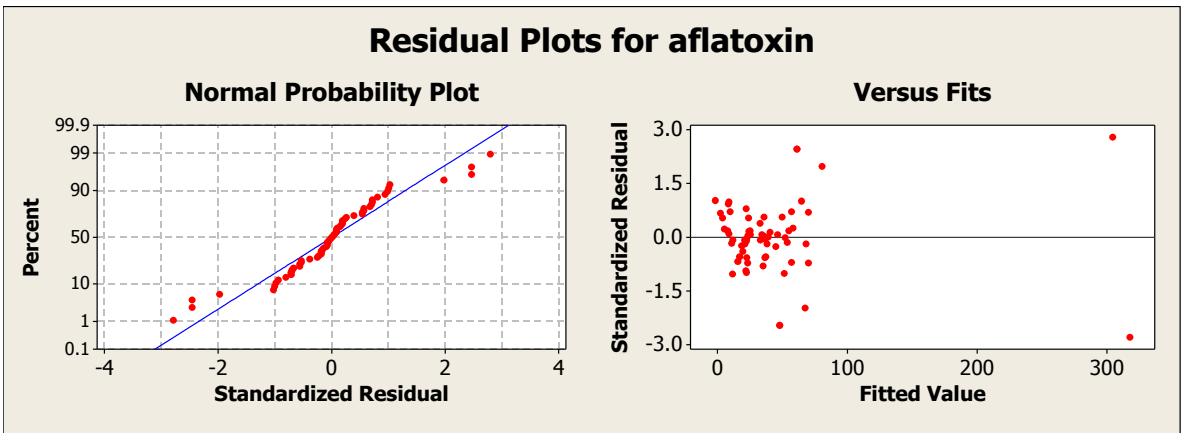
...

Variance Components, using Adjusted SS

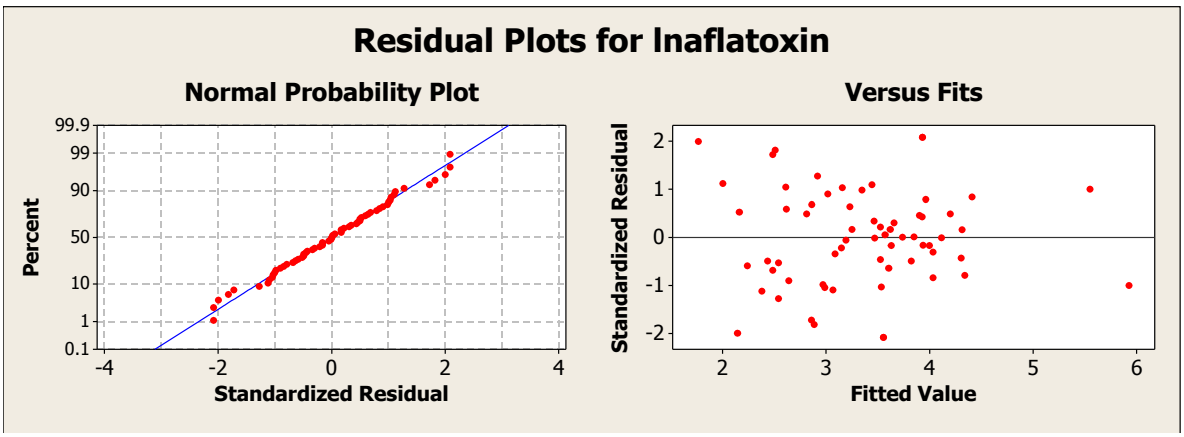
Source	Estimated Value
section	-238.7
subsample(section)	2866.5
Error	264.9

Least Squares Means for aflatoxin

method	Mean
BF	35.53
CB	48.75



```
MTB > GLM  'lnaflatoxin' = section subsample(section) method;
SUBC>   Random 'section';
SUBC>   Brief 2 ;
SUBC>   EMS;
SUBC>   Means method;
SUBC>   GFourpack;
SUBC>   RType 2 .

General Linear Model: lnaflatoxin versus section, method, subsample

Factor              Type    Levels  Values
section             random      16  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
subsample(section)  random      32  A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B, A, B,
                                    A, B, A, B, A, B, A, B, A, B, A, B, A, B
method              fixed        2  BF, CB

Analysis of Variance for lnaflatoxin, using Adjusted SS for Tests

Source              DF   Seq SS   Adj SS  Adj MS      F      P
section             15  14.0990  14.0990  0.9399   0.68  0.768
subsample(section)  16  22.0503  22.0503  1.3781   9.29  0.000
method               1   2.2734   2.2734  2.2734  15.32  0.000
Error               31   4.5988   4.5988  0.1483
Total               63  43.0215

S = 0.385160   R-Sq = 89.31%   R-Sq(adj) = 78.28%

Unusual Observations for lnaflatoxin

Obs  lnaflatoxin      Fit   SE Fit  Residual  St Resid
 19      2.99573  3.55371  0.27657  -0.55798     -2.08 R
 20      4.48864  3.93066  0.27657   0.55798      2.08 R
 23      2.99573  3.55371  0.27657  -0.55798     -2.08 R
 24      4.48864  3.93066  0.27657   0.55798      2.08 R

R denotes an observation with a large standardized residual.
...
Variance Components, using Adjusted SS

                    Estimated
Source                  Value
section               -0.1096
subsample(section)     0.6149
Error                  0.1483

Least Squares Means for lnaflatoxin

method   Mean
BF      3.166
CB      3.543
```



Stata listing for Question 2, parts D) and E) – note also the Minitab listing and graphs above:

```
. mixed aflatoxin i.method || section: || subsampsect:, reml
...

Mixed-effects REML regression                   Number of obs      =        64
-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
        section |       16          4        4.0          4
    subsampsect |       32          2        2.0          2
-----------------------------------------------------------

                                                Wald chi2(1)       =     10.55
Log restricted-likelihood = -311.51452          Prob > chi2        =    0.0012
------------------------------------------------------------------------------
   aflatoxin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      method |
         CB  |   13.21875   4.068894     3.25   0.001     5.243865    21.19364
       _cons |   35.53125   9.520462     3.73   0.000     16.87149    54.19101
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
section: Identity            |
                  var(_cons) |   2.12e-09   8.26e-08      1.65e-42    2.73e+24
-----------------------------+------------------------------------------------
subsampsect: Identity        |
                  var(_cons) |    2635.56   2061.231      569.0642    12206.31
-----------------------------+------------------------------------------------
               var(Residual) |   264.8943   67.37973      160.9004    436.1021
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =    54.15   Prob > chi2 = 0.0000
```

```
. mixed lnaflatoxin i.method || section: || subsampsect:, reml
...

Mixed-effects REML regression                   Number of obs      =        64
------------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+-------------------------------------------
        section |       16          4        4.0          4
    subsampsect |       32          2        2.0          2
------------------------------------------------------------

                                                Wald chi2(1)       =     15.32
Log restricted-likelihood = -64.244782          Prob > chi2        =    0.0001
------------------------------------------------------------------------------
 lnaflatoxin |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      method |
         CB  |   .3769466   .0962899     3.91   0.000     .1882218    .5656713
       _cons |   3.166054    .143312    22.09   0.000     2.885168     3.44694
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
section: Identity            |
                  var(_cons) |   2.71e-18          .             .           .
-----------------------------+------------------------------------------------
subsampsect: Identity        |
                  var(_cons) |   .5088785   .1492891      .2863507    .9043366
-----------------------------+------------------------------------------------
               var(Residual) |    .148348   .0376804      .0901729    .2440546
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =    28.37   Prob > chi2 = 0.0000
```

Question 3.

For several years in the 1970s, data were collected on the salinity of water during the spring in Pamlico Sound, North Carolina. Analysis of the data was part of a project for forecasting the shrimp harvest. The response variable of interest (here) is the salinity of the water, measured in ppt (parts per thousand). Our data contain average salinity measured over two-week periods from March to May during these years (presumably at the same monitoring station). Additional variables that may be useful to describe the salinity levels include: the water flow (i.e. river discharge into the sound; `water`), the year and period of the measurement, and the salinity value for the previous two-week period (`lag`). Note that periods are coded by the values $0 - 6$ within years, and that because only spring data were used, the lagged salinity value is not always the previous value of salinity. The full dataset is shown in the Stata listings below.

A) (*3 points*) Describe the study type, explain the statistical model and analysis shown in the Stata listing for A), and use all the information to draw (first) conclusions about the effects of water flow and the time of the measurement. Postpone all discussion about the model assumptions and diagnostics until part B).

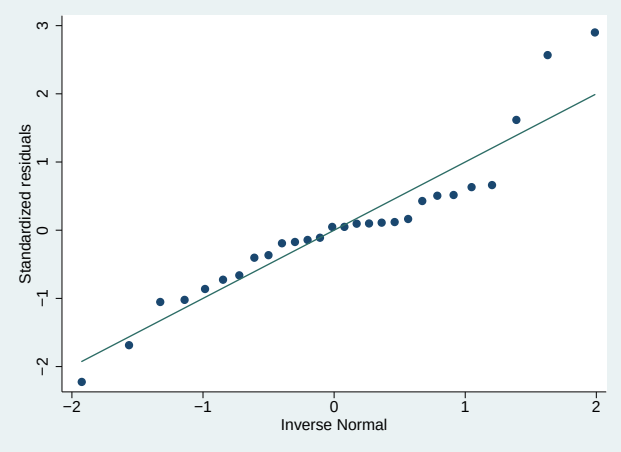
B) (*4 points*) Review briefly the assumptions behind the statistical analysis from A), and use the information provided to discuss how well each of these assumptions have been met and all other potential concerns one might have with the fitted model. Use statistical tests and decision rules/guidelines to support your discussion and conclusions. If you would want to carry out additional analyses related to model validation before embarking on the model-building, explain their purpose and implementation in statistical software.

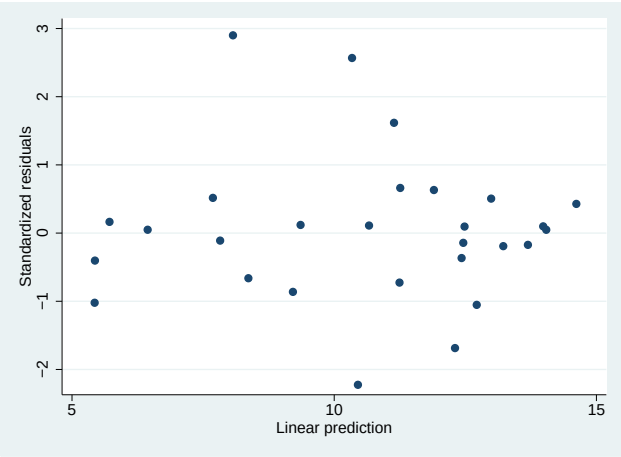
C) (*3 points*) Three additional Stata listings are presented for part C). Explain each of these analyses and describe what information they give you of relevance for the model-building (and perhaps as well for the model validation in B)). Based on the information provided, suggest the next steps of the development of the "best" model for salinity. For each analysis you propose, make sure to explain its purpose and the rationale behind it from the information you have already obtained; therefore, avoid suggestions of general model-building steps that do not utilize the information already obtained.

Stata listings and graphs for Question 3, parts A) and B):

```
. regress salinity lag water year period

      Source |       SS       df       MS              Number of obs =      28
-------------+------------------------------           F(  4,    23) =   32.44
       Model |  207.812606     4  51.9531516           Prob > F      =  0.0000
    Residual |  36.8370523    23  1.60161097           R-squared     =  0.8494
-------------+------------------------------           Adj R-squared =  0.8232
       Total |  244.649659    27  9.06109847           Root MSE      =  1.2655

------------------------------------------------------------------------------
    salinity |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lag |   .6163913   .1186028     5.20   0.000     .3710428    .8617399
       water |  -.2633923   .1029941    -2.56   0.018    -.4764519   -.0503328
        year |   .4264764   .2273268     1.88   0.073    -.0437849    .8967378
      period |   .0600201   .1598746     0.38   0.711    -.2707057     .390746
       _cons |  -21.57792   16.87763    -1.28   0.214    -56.49196    13.33611
------------------------------------------------------------------------------
```

```
. predict stdres, rstandard
. predict fit, xb
. scatter stdres fit
. qnorm stdres
```






```
. swilk stdres
                    Shapiro-Wilk W test for normal data

    Variable |    Obs       W           V         z       Prob>z
-------------+--------------------------------------------------
      stdres |     28    0.91997      2.417     1.817    0.03462

. predict delres, rstudent
. predict lev, leverage
. predict cookd, cooksd
. format stdres-cookd %5.3f
. list salinity-cookd, table
```

	salinity	lag	water	year	period	stdres	fit	delres	lev	cookd
1.	7.6	8.2	23.005	72	4	-0.662	8.364	-0.654	0.169	0.018
2.	7.7	7.6	23.873	72	5	-0.110	7.825	-0.108	0.192	0.001
3.	4.3	4.6	26.417	73	0	-1.022	5.432	-1.023	0.233	0.063
4.	5.9	4.3	24.868	73	1	0.165	5.715	0.161	0.215	0.001
5.	5	5.9	29.895	73	2	-0.403	5.437	-0.396	0.265	0.012
6.	6.5	5	24.2	73	3	0.049	6.443	0.048	0.156	0.000
7.	8.3	6.5	23.215	73	4	0.517	7.687	0.508	0.120	0.007
8.	8.2	8.3	21.862	73	5	-0.862	9.213	-0.857	0.138	0.024
9.	13.2	10.1	22.274	74	0	2.568	10.340	2.974	0.226	0.384
10.	12.6	13.2	23.83	74	1	0.632	11.901	0.623	0.236	0.025
11.	10.4	12.6	25.144	74	2	-0.726	11.245	-0.718	0.154	0.019
12.	10.8	10.4	22.43	74	3	0.111	10.664	0.109	0.062	0.000
13.	13.1	10.8	21.785	74	4	1.618	11.140	1.680	0.084	0.048
14.	12.3	13.1	22.38	74	5	-0.143	12.461	-0.140	0.202	0.001
15.	10.4	13.3	23.927	75	0	-1.686	12.304	-1.762	0.204	0.146

	salinity	lag	water	year	period	stdres	fit	delres	lev	cookd
16.	10.5	10.4	33.443	75	1	2.900	8.070	3.562	0.562	2.155
17.	7.7	10.5	24.859	75	2	-2.225	10.452	-2.456	0.044	0.046
18.	9.5	7.7	22.686	75	3	0.120	9.359	0.117	0.131	0.000
19.	12	10	21.789	76	0	0.662	11.259	0.654	0.218	0.025
20.	12.6	12	22.041	76	1	0.096	12.486	0.094	0.110	0.000
21.	13.6	12.1	21.033	76	4	0.506	12.993	0.497	0.100	0.006
22.	14.1	13.6	21.005	76	5	0.099	13.985	0.097	0.151	0.000
23.	13.5	15	25.865	77	0	-0.172	13.694	-0.169	0.209	0.002
24.	11.5	13.5	26.29	77	1	-1.053	12.717	-1.055	0.165	0.044
25.	12	11.5	22.932	77	2	-0.366	12.429	-0.359	0.141	0.004
26.	13	12	21.313	77	3	-0.192	13.224	-0.188	0.148	0.001
27.	14.1	13	20.769	77	4	0.049	14.044	0.048	0.159	0.000
28.	15.1	14.1	21.393	77	5	0.428	14.617	0.421	0.207	0.010

Stata listing for Question 3, part C):

```
. pwcorr salinity lag water year period, sig

             | salinity      lag    water     year   period
-------------+-------------------------------------------
    salinity |   1.0000
             |
             |
         lag |   0.8715   1.0000
             |   0.0000
             |
       water |  -0.4771  -0.2614   1.0000
             |   0.0103   0.1790
             |
        year |   0.7478   0.7362  -0.2121   1.0000
             |   0.0000   0.0000   0.2786
             |
      period |   0.1217   0.0156  -0.4524  -0.1498   1.0000
             |   0.5374   0.9370   0.0156   0.4466
             |

. regress salinity lag water i.year i.period

      Source |       SS       df       MS              Number of obs =      28
-------------+------------------------------           F( 12,    15) =   22.68
       Model |  231.871326    12  19.3226105           Prob > F      =  0.0000
    Residual |  12.7783321    15   .85188881           R-squared     =  0.9478
-------------+------------------------------           Adj R-squared =  0.9060
       Total |  244.649659    27  9.06109847           Root MSE      =  .92298
```

```
------------------------------------------------------------------------------
    salinity |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lag |    .149589    .162158     0.92   0.371    -.1960425    .4952205
       water |  -.0339915   .1006652    -0.34   0.740    -.2485544    .1805713
             |
        year |
          73 |   .2141379   .8273842     0.26   0.799     -1.54939    1.977666
          74 |   4.955115   1.136018     4.36   0.001     2.533749    7.376481
          75 |   3.268683   1.103408     2.96   0.010      .916826    5.620541
          76 |   5.404302   1.171997     4.61   0.000     2.906249    7.902355
          77 |   5.870632   1.310347     4.48   0.000     3.077693     8.66357
             |
      period |
           1 |  -.0026241   .6186005    -0.00   0.997     -1.32114    1.315892
           2 |  -1.412322    .640119    -2.21   0.043    -2.776703   -.0479405
           3 |  -.1390592   .7018849    -0.20   0.846    -1.635091    1.356973
           4 |   1.314395   .6457158     2.04   0.060    -.0619151    2.690706
           5 |   1.276696   .6900713     1.85   0.084    -.1941558    2.747548
             |
       _cons |   5.969429   2.835621     2.11   0.053     -.074554    12.01341
------------------------------------------------------------------------------
```

```
. regress salinity lag c.water##c.water year period

      Source |       SS       df       MS              Number of obs =      28
-------------+------------------------------           F(  5,    22) =   49.52
       Model |  224.685141     5  44.9370283           Prob > F      =  0.0000
    Residual |  19.9645173    22  .907478057           R-squared     =  0.9184
-------------+------------------------------           Adj R-squared =  0.8998
       Total |  244.649659    27  9.06109847           Root MSE      =  .95262

------------------------------------------------------------------------------
      salinity |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
           lag |   .6336046   .0893651     7.09   0.000     .4482726    .8189365
         water |  -4.586305   1.005539    -4.56   0.000    -6.671666   -2.500944
c.water#c.water |    .080217   .0186035     4.31   0.000     .0416357    .1187983
          year |   .1380723   .1837233     0.75   0.460    -.2429465    .5190911
        period |   -.213981   .1360892    -1.57   0.130    -.4962127    .0682507
         _cons |   57.28178   22.26828     2.57   0.017      11.1002    103.4634
------------------------------------------------------------------------------
```