BIOSTATS VHM 802 COURSE
WINTER 2015, ATLANTIC VETERINARY COLLEGE, PEI
HENRIK STRYHN

Final examination, 22 April 2015

All aids are allowed, except a computer-like device (including tablets and smartphones) 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.

A study explored how different compounds affected the concentration of a solution at which an individual first senses a specific odor; this is the called the (odor) threshold concentration. The study involved a total of 16 different solutions as well as 8 individuals, termed judges. The threshold-determining procedure was time consuming, and therefore each judge could only be used for four solutions. The 16 solutions were formed by choosing either low or high concentrations of four compounds, in all possible combinations. For simplicity we denote these four compounds as A−D, each taking either the value 1 (low) or 2 (high). After a suitable normalization (of no interest here), the threshold concentrations were as shown in the data listing, for the 8 judges and the different compounds determining the solutions.

	Solution compounds				Threshold	Natural
Judge	A	B	C	D	conc.	log conc.
1	1	1	1	1	8389	9.0347
	1	2	2	1	816	6.7044
	2	2	1	2	4	1.3863
	2	1	2	2	46	3.8286
2	2	1	1	1	4351	8.3782
	2	2	2	1	78	4.3567
	1	2	1	2	5941	8.6896
	1	1	2	2	27138	10.2087
3	2	2	1	1	6	1.7918
	2	1	2	1	262	5.5683
	1	1	1	2	1230	7.1148
	1	2	2	2	98	4.5850
4	1	2	1	1	375	5.9269
	1	1	2	1	33551	10.4208
	2	1	1	2	246	5.5053
	2	2	2	2	10	2.3026

(*Note:* the table continues on the next page)

(*Note:* the table is continued from the previous page)

Judge	Solution compounds A	B	C	D	Threshold conc.	Natural log conc.
5	1	1	1	1	56034	10.9337
	1	2	2	1	25046	10.1285
	2	2	1	2	19	2.9444
	2	1	2	2	490	6.1944
6	2	1	1	1	2346	7.7605
	2	2	2	1	35	3.5553
	1	2	1	2	228	5.4293
	1	1	2	2	6842	8.8308
7	2	2	1	1	67	4.2047
	2	1	2	1	3081	8.0330
	1	1	1	2	50991	10.8394
	1	2	2	2	784	6.6644
8	1	2	1	1	40581	10.6111
	1	1	2	1	90293	11.4108
	2	1	1	2	19103	9.8576
	2	2	2	2	61	4.1109

Use the description of the study and the information contained in the Minitab listings on the next two pages (equivalent Stata listings are available upon request) to answer the following questions. Note that the outcome is the (natural) log threshold concentration.

A) (*3 points*) Describe the experimental design in statistical terms, using standard descriptors such as factors, treatments, replication, blocks, balancedness, completeness, treatments, experimental units, hierarchical structure etc. Use your characterization of the design to motivate a statistical model for the data (possibly, but not necessarily, the model considered in the subsequent questions). Make sure to explain clearly the meaning of the different terms in your model.

B) (*4 points*) Describe the statistical model used in the Minitab listings, and use the results to assess and interpret, by means of estimates and statistical tests, the different effects of the model. Draw conclusions about whether each of the compounds affects the threshold concentration for odor? — if yes, quantify such effect(s). Comment also on the importance of the judges in the experimental design and analysis.

C) (*1 point*) The analysis was carried out on log scale. Do you think an analysis on original scale would have been likely to meet the model assumptions as well as on log scale? If not, explain how you would anticipate model assumptions on original scale to be violated, and why.

D) (*1 point*) The experimental design includes some features that distinguish it from the standard designs discussed in the course. One way this became noticeable in the analysis was that inclusion of the interaction `A*D` failed to produce any estimates for this term (both in Minitab and Stata; output not shown). Based on the listing of the experimental design, try to explain why this happened, and suggest at least one way to modify the model/analysis to enable estimation of the `A*D` interaction.

E) (*1 point*) For another run of the experiment involving three compounds and more judges (within reasonable numbers, say at most 20 judges), propose a "standard" design (among those covered in the course) that could be used for the new experiment. Describe the proposed design and the advantages you see of using such a design.

Minitab listing for Question 1:

```
MTB > GLM 'lnconc' = judge a|b a|c b|c b|d c|d;
SUBC>   Brief 2 ;
SUBC>   Means a b c d a*b a*c b*c b*d c*d.

General Linear Model: lnconc versus judge, a, b, c, d

Factor  Type   Levels  Values
judge   fixed       8  1, 2, 3, 4, 5, 6, 7, 8
a       fixed       2  1, 2
b       fixed       2  1, 2
c       fixed       2  1, 2
d       fixed       2  1, 2

Analysis of Variance for lnconc, using Adjusted SS for Tests

Source  DF   Seq SS   Adj SS   Adj MS       F      P
judge    7   57.390   57.390    8.199   11.63  0.000
a        1  104.236  104.236  104.236  147.87  0.000
b        1   79.783   79.783   79.783  113.18  0.000
a*b      1    3.392    3.392    3.392    4.81  0.044
c        1    0.384    0.384    0.384    0.54  0.472
a*c      1    0.565    0.565    0.565    0.80  0.385
b*c      1    1.261    1.261    1.261    1.79  0.201
d        1   12.912   12.912   12.912   18.32  0.001
b*d      1    0.126    0.126    0.126    0.18  0.679
c*d      1    1.352    1.352    1.352    1.92  0.186
Error   15   10.574   10.574    0.705
Total   31  271.975

S = 0.839592   R-Sq = 96.11%   R-Sq(adj) = 91.97%

Unusual Observations for lnconc

Obs   lnconc      Fit  SE Fit  Residual  St Resid
 30  11.4108  12.6588  0.6120   -1.2479     -2.17 R

R denotes an observation with a large standardized residual.


Least Squares Means for lnconc

a     Mean  SE Mean
1    8.596   0.2099
2    4.986   0.2099

b
1    8.370   0.2099
2    5.212   0.2099
```

```
c
1    6.901   0.2099
2    6.681   0.2099

d
1    7.426   0.2099
2    6.156   0.2099

a*b
1 1  9.849   0.2968
1 2  7.342   0.2968
2 1  6.891   0.2968
2 2  3.082   0.2968

a*c
1 1  8.572   0.2968
1 2  8.619   0.2968
2 1  5.229   0.2968
2 2  4.744   0.2968

b*c
1 1  8.678   0.2968
1 2  8.062   0.2968
2 1  5.123   0.2968
2 2  5.301   0.2968

b*d
1 1  8.943   0.2968
1 2  7.797   0.2968
2 1  5.910   0.2968
2 2  4.514   0.2968

c*d
1 1  7.330   0.2968
1 2  6.471   0.2968
2 1  7.522   0.2968
2 2  5.841   0.2968
```

Stata do-file for Question 1: (Stata output available upon request)

```
anova lnconc judge a##b a##c b##c b##d c##d
margins a##b a##c b##c b##d c##d, asbalanced
```

Question 2.

A study was carried out to detect differences in the percentage of red blood cell (RBC) survival at different ionic concentrations, among 3 different groups of mice, comprising 11, 8 and 11 animals, respectively. The table below shows the data obtained for some of the animals, and the (full) data are also displayed in 4 graphs on the next two pages.

RBC survival (%)		Ionic concentration					
Group	Mouse	0.30	0.35	0.40	0.45	0.50	0.55
1	1	94.1	97.2	94.6	87.3	33.1	5.6
	2	100	100	100	93.7	35.6	3.7
	3	100	100	98.3	90.3	29.1	0
	4	100	99.9	96.0	87.7	35.9	2.8
	5	100	100	97.7	82.4	31.8	0
	...		...		...		
	11	97.6	97.3	94.6	90.4	45.9	0
2	1	100	100	94.2	76.9	14.2	0
	2	97.5	98.1	92.3	71.3	10.2	0
	3	93.2	96.6	89.8	69.2	10.0	0
	...		...		...		
	8	97.9	99.5	94.8	76.4	19.5	0
3	1	100	96.4	94.4	82.1	19.4	1.3
	2	98.6	97.7	88.1	80.4	11.8	0
	3	100	100	100	100	17.1	0
	4	100	95.8	89.2	78.5	17.0	0
	5	100	100	97.3	17.8	17.7	0
	...		...		...		
	11	97.8	100	96.2	77.6	10.7	0

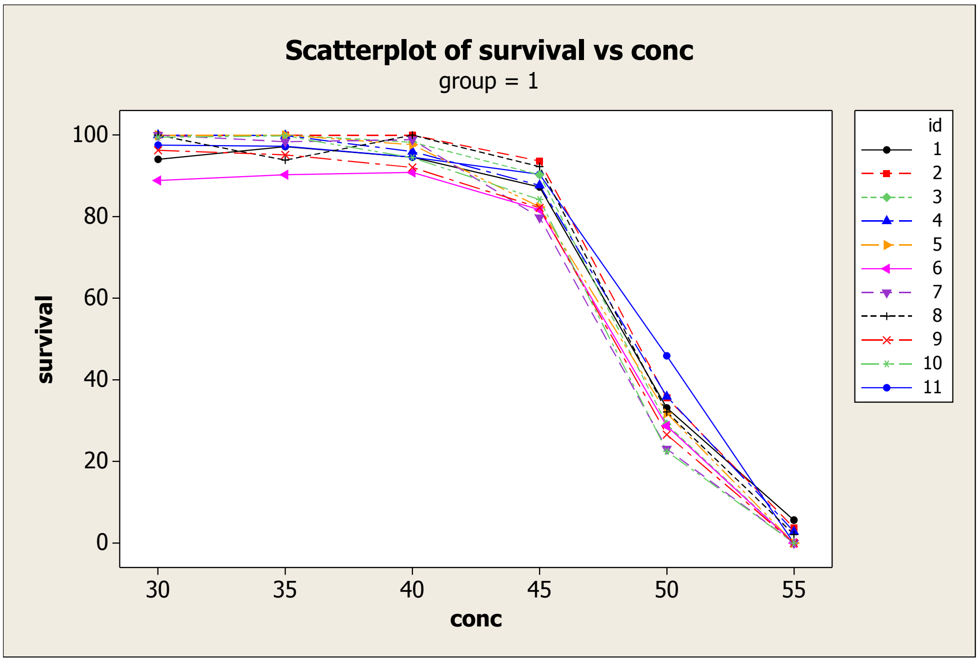
A) (*3 points*) Describe the data structure and experimental design; include any relevant diagrams to represent the data/design. Next, interpret the graphs; in particular, what do they tell us about the effects of groups, ionic concentrations and animals, on the outcome (RBC survival).

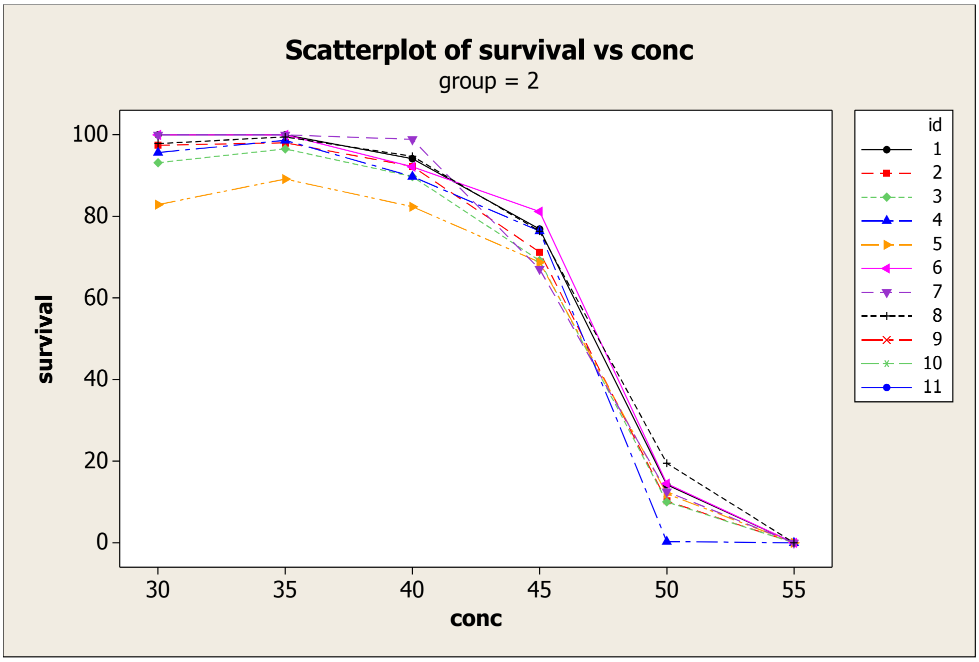
B) (*3 points*) Describe the statistical analysis, and the underlying statistical model, carried out for the Minitab and Stata listings below; note that the Minitab and Stata listings do not contain exactly the same information. Draw conclusions from the analysis about the effects of interest in the study; **defer any discussion of model assumptions** to C). Make sure to quantify any difference(s) between groups, or describe any additional analyses you would want to carry out to assist with this part (ideally you should use the information provided as much as you can).

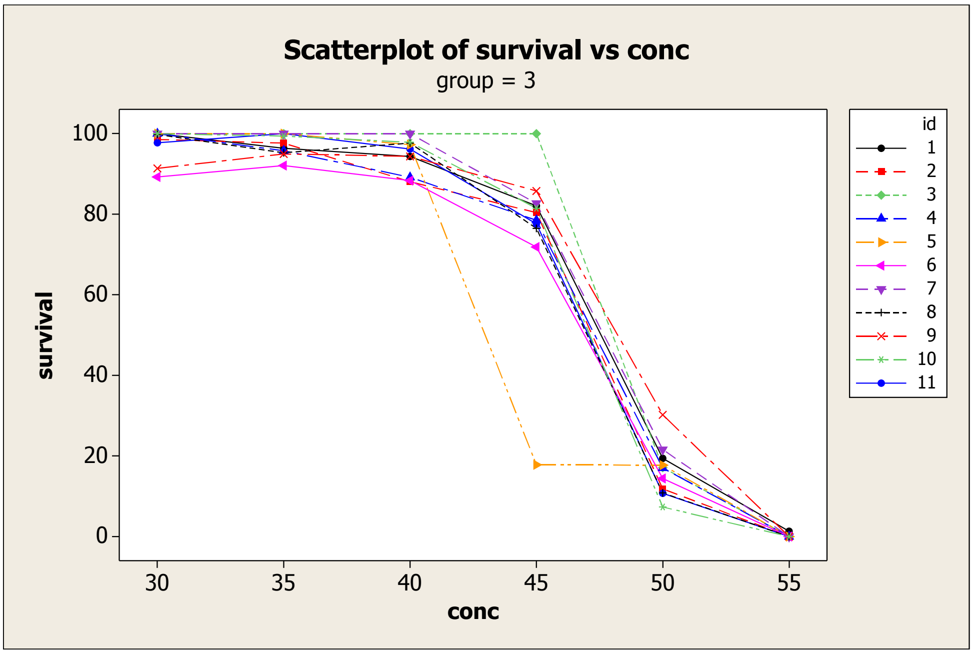
C) (*2 points*) Use the graphical displays, in addition to the other information provided, to critically discuss the suitability of the model/analysis for these data. If you note some concerns with the model assumptions, try to assess how they might affect your conclusions from B). Include any suggestions for improvements of the statistical analysis you might have, and then give also details for how you would implement them in statistical software.

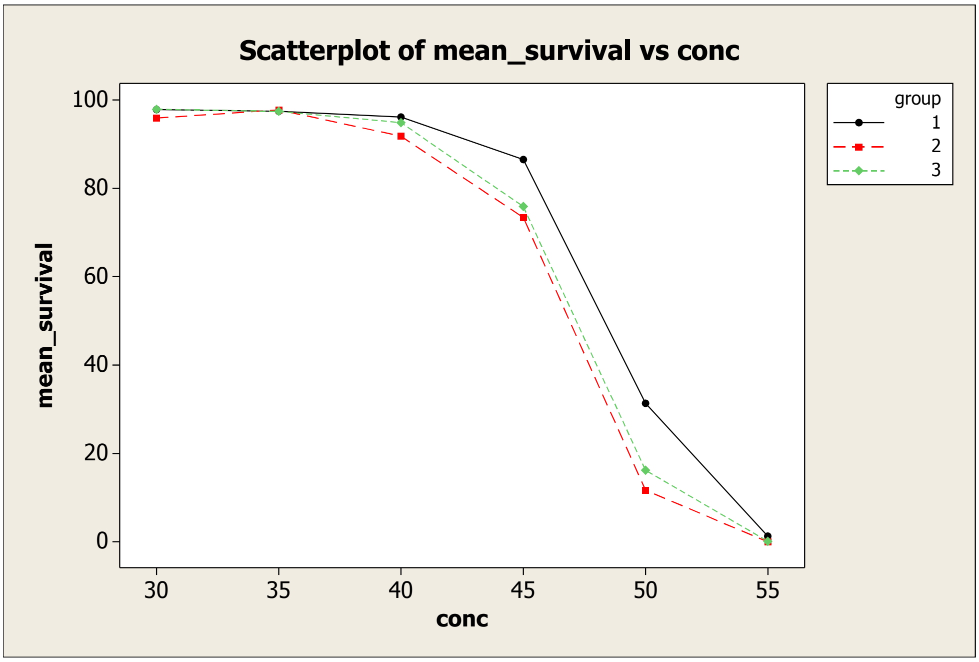
D) (*2 points*) Suggest some (at least two) alternative ways to analyse the data that in your view would be helpful to explore the effects of groups and ionic concentrations (without including the analysis discussed above, or variants of it). Make sure to motivate clearly your proposed analyses, and explain how you would carry out the analysis, both in terms of the statistical models used and the actual analysis in statistical software. Discuss, as well as the information provided allows you, whether there are concerns about the validity of the proposed analyses; ideally, you should not propose analyses that are not statistically valid.

Plots for Question 2, A):







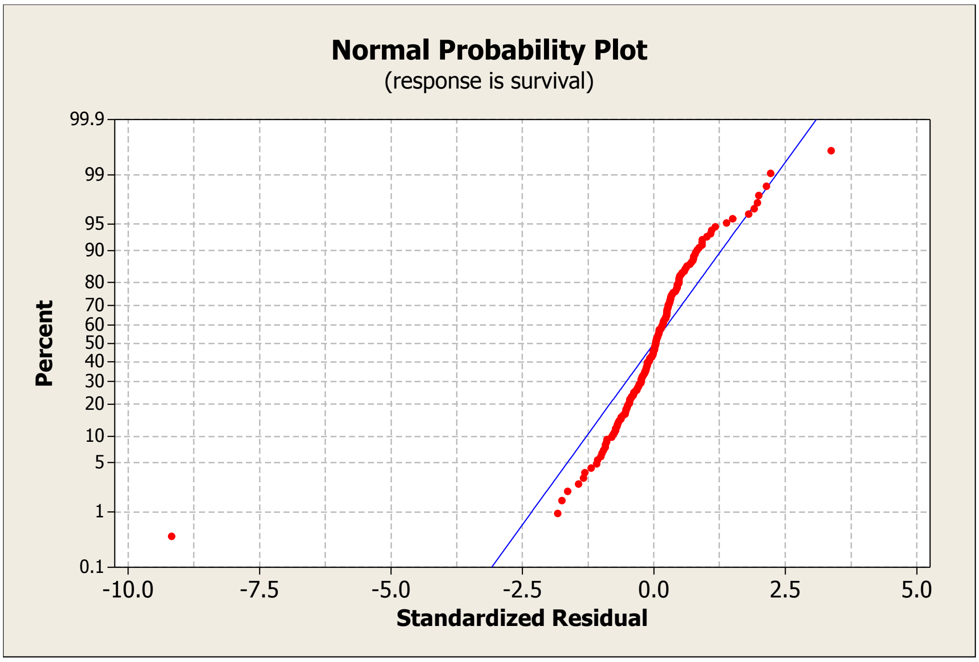


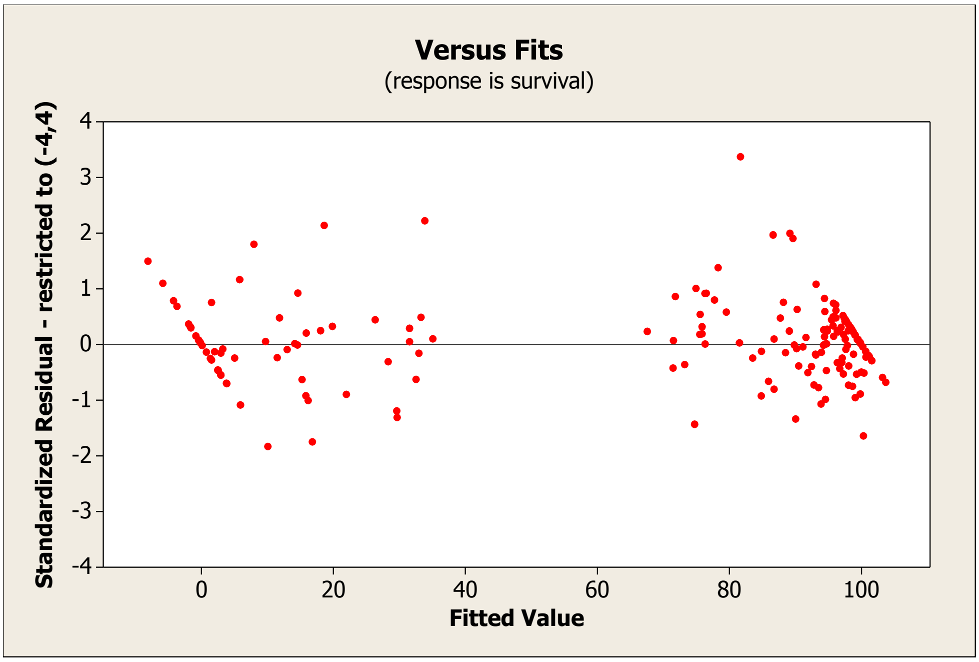
Minitab listing and graphs for Question 2:

```
MTB > GLM  'survival' = group| conc id(group);
SUBC>   Random 'id';
SUBC>   Means group|conc;

General Linear Model: survival versus group, conc, id

Factor     Type    Levels  Values
group      fixed        3  1, 2, 3
conc       fixed        6  30, 35, 40, 45, 50, 55
id(group)  random      30  1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 1, 2, 3, 4, 5, 6, 7, 8, 1,
                           2, 3, 4, 5, 6, 7, 8, 9, 10, 11
```

```
Analysis of Variance for survival, using Adjusted SS for Tests

Source       DF    Seq SS    Adj SS   Adj MS        F      P
group         2    1388.9    1388.9    694.5    11.19  0.000
conc          5  279369.7  274381.7  54876.3  1409.14  0.000
group*conc   10    1833.4    1833.4    183.3     4.71  0.000
id(group)    27    1676.1    1676.1     62.1     1.59  0.044
Error       135    5257.3    5257.3     38.9
Total       179  289525.4

S = 6.24045   R-Sq = 98.18%   R-Sq(adj) = 97.59%

Unusual Observations for survival

Obs  survival     Fit  SE Fit  Residual  St Resid
 65    45.900  33.842   3.073    12.058      2.22 R
130   100.000  81.676   3.073    18.324      3.37 R
142    17.800  67.626   3.073   -49.826     -9.17 R
167    30.200  18.583   3.073    11.617      2.14 R

R denotes an observation with a large standardized residual.

... (some lines deleted)

Variance Components, using Adjusted SS

           Estimated
Source         Value
id(group)      3.856
Error         38.943


Least Squares Means for survival

group          Mean
1           68.4333
2           61.7542
3           63.7318
conc
30          97.2182
35          97.5481
40          94.2727
45          78.6163
50          19.7136
55           0.4697
group*conc
1     30    97.8545
1     35    97.4545
1     40    96.1455
1     45    86.5455
1     50    31.3091
1     55     1.2909
2     30    95.9000
2     35    97.7625
```

```
2       40     91.8000
2       45     73.4125
2       50     11.6500
2       55     -0.0000
3       30     97.9000
3       35     97.4273
3       40     94.8727
3       45     75.8909
3       50     16.1818
3       55      0.1182
```





Stata do-file and listing for Question 2:

```
. generate uniqid=group*100+id
. mixed survival group##conc || uniqid:, reml

Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log restricted-likelihood = -553.42305
Iteration 1:   log restricted-likelihood = -553.41878
Iteration 2:   log restricted-likelihood = -553.41878
Computing standard errors:

Mixed-effects REML regression                   Number of obs     =        180
Group variable: uniqid                          Number of groups  =         30

                                                Obs per group: min =         6
                                                               avg =       6.0
                                                               max =         6

                                                Wald chi2(17)     =    7243.22
Log restricted-likelihood = -553.41878          Prob > chi2       =     0.0000
------------------------------------------------------------------------------
    survival |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |
          2  |  -1.954546   3.039843    -0.64   0.520    -7.912528    4.003436
          3  |   .0454552   2.789551     0.02   0.987    -5.421964    5.512874
             |
        conc |
         35  |  -.3999994   2.660937    -0.15   0.881    -5.615341    4.815342
         40  |  -1.709091   2.660937    -0.64   0.521    -6.924433     3.50625
         45  |  -11.30909   2.660937    -4.25   0.000    -16.52443    -6.09375
         50  |  -66.54545   2.660937   -25.01   0.000     -71.7608   -61.33011
         55  |  -96.56364   2.660937   -36.29   0.000     -101.779    -91.3483
             |
  group#conc |
       2 35  |   2.262499    4.10078     0.55   0.581    -5.774882    10.29988
       2 40  |  -2.390907    4.10078    -0.58   0.560    -10.42829    5.646474
       2 45  |  -11.17841    4.10078    -2.73   0.006    -19.21579   -3.141027
       2 50  |  -17.70454    4.10078    -4.32   0.000    -25.74193   -9.667164
       2 55  |   .6636369    4.10078     0.16   0.871    -7.373744    8.701018
       3 35  |  -.0727282   3.763134    -0.02   0.985    -7.448335    7.302878
       3 40  |  -1.318183   3.763134    -0.35   0.726    -8.693789    6.057424
       3 45  |      -10.7   3.763134    -2.84   0.004    -18.07561   -3.324394
       3 50  |  -15.17273   3.763134    -4.03   0.000    -22.54833   -7.797121
       3 55  |  -1.218182   3.763134    -0.32   0.746    -8.593789    6.157424
             |
       _cons |   97.85455    1.97251    49.61   0.000      93.9885    101.7206
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
uniqid: Identity             |
                  var(_cons) |   3.855533   2.924562      .8718006    17.05107
-----------------------------+------------------------------------------------
               var(Residual) |   38.94323   4.740017        30.678    49.43527
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =     2.70 Prob >= chibar2 = 0.0500
```

```
. testparm i.group
 ( 1)  [survival]2.group = 0
 ( 2)  [survival]3.group = 0

           chi2(  2) =    0.54
         Prob > chi2 =    0.7648

. testparm i.conc
 ( 1)  [survival]35.conc = 0
 ( 2)  [survival]40.conc = 0
 ( 3)  [survival]45.conc = 0
 ( 4)  [survival]50.conc = 0
 ( 5)  [survival]55.conc = 0

           chi2(  5) = 2454.64
         Prob > chi2 =    0.0000

. testparm group#conc
 ( 1)  [survival]2.group#35.conc = 0
 ( 2)  [survival]2.group#40.conc = 0
 ( 3)  [survival]2.group#45.conc = 0
 ( 4)  [survival]2.group#50.conc = 0
 ( 5)  [survival]2.group#55.conc = 0
 ( 6)  [survival]3.group#35.conc = 0
 ( 7)  [survival]3.group#40.conc = 0
 ( 8)  [survival]3.group#45.conc = 0
 ( 9)  [survival]3.group#50.conc = 0
 (10)  [survival]3.group#55.conc = 0

           chi2( 10) =   47.08
         Prob > chi2 =    0.0000

. predict stdres, rstandard
. predict fitted, fitted
. list if abs(stdres)>2

     +----------------------------------------------------------------+
     | group   id   conc   survival   uniqid      stdres     fitted |
     |----------------------------------------------------------------|
 65. |     1   11     50       45.9      111    2.186836   32.25315 |
130. |     3    3     45        100      303    3.517907   78.04667 |
142. |     3    5     45       17.8      305   -8.815203   72.81084 |
167. |     3    9     50       30.2      309    2.102931   17.07676 |
     +----------------------------------------------------------------+

. qnorm stdres
. scatter stdres fitted
```

Note: Stata graphs not included; use instead the Minitab graphs.

Question 3.

A frequently used literature dataset (Johnson & Raven (1973), *Science* **179**, 893–895) gives the number of tortoise species on 30 Galapagos islands. The interest is in the relationship between the diversity of tortoises and several geographic variables also included in the data (species diversity on these islands was first studied by Charles Darwin in the 1830's). As our measure of diversity, we will consider the number of endemic species (`ES` in the table below). Some geographical information was missing for three of the 30 islands, therefore only the 27 islands with complete data are included here. The table below explains the different variables. Furthermore, the Minitab output on the following pages includes a listing of the full dataset.

Variable	Description
island	name of island
NS	total number of (tortoise) species on islands
ES	number of endemic species on island, i.e., occuring only on this island
area	area of the island (km^2)
anear	area of the nearest island (km^2)
dist	distance to the nearest island (km)
distSC	distance to Santa Cruz, the largest island, (km)
elev	highest elevation of the island (m)

Use the information contained in the Minitab listing and graphs displayed below to answer to the following questions. A Stata do-file which would give equivalent output is shown as well (without repeating the output; Stata output is available upon request).

A) (*2 points*) Explain the statistical model on which the analysis is based, and interpret all the estimated parameters of the model (assuming that the model is valid and meaningful).

B) (*4 points*) Discuss as fully as the provided information allows, whether the assumptions of the model seem to be met. For any issue that you may consider as a (potential) model violation, assess its impact on the analysis and give specific directions for how you would deal with it in the analysis.

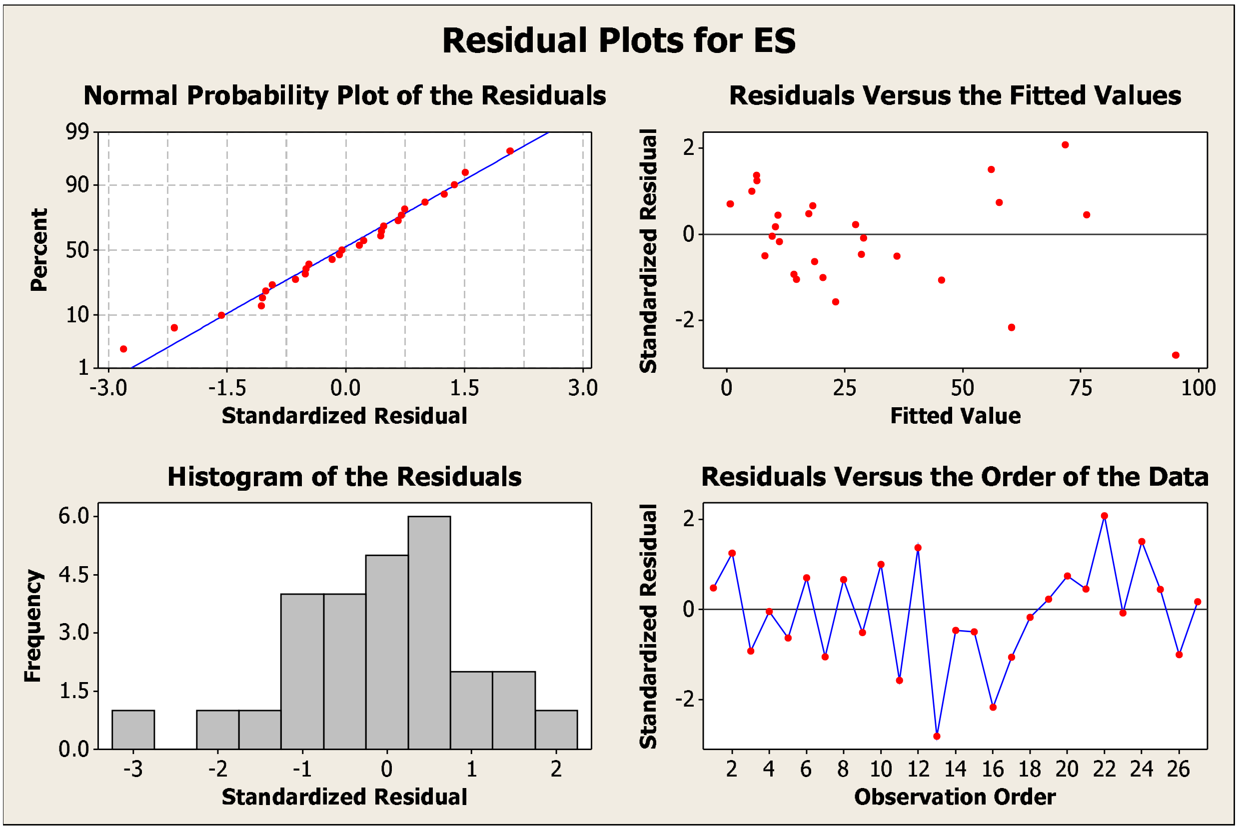
C) (*4 points*) Based on information provided, give a detailed outline of how you would develop a good (final) model for the number of endemic species (`ES`). Include in your outline any decisions you make about which predictors to include in the analysis leading to the final model; in particular, discuss whether the total number of species (`NS`) should be included as a predictor or not. Include also in your outline any analyses undertaken to examine in which form the predictors should be entered into the model.

Minitab listing for Question 3:

```
MTB > Correlation 'NS'-'elev'.
Correlations: NS, ES, area, anear, dist, distSC, elev

              NS      ES    area   anear    dist  distSC
ES         0.974
           0.000

area       0.614   0.618
           0.001   0.001

anear      0.034   0.095   0.185
           0.865   0.639   0.355

dist      -0.055  -0.049  -0.132  -0.114
           0.784   0.810   0.512   0.571

distSC    -0.243  -0.253  -0.133   0.061   0.599
           0.222   0.203   0.509   0.763   0.001

elev       0.725   0.778   0.751   0.562  -0.027  -0.047
           0.000   0.000   0.000   0.002   0.895   0.817

Cell Contents: Pearson correlation
               P-Value


MTB > Name c9 "SRES1" c10 "TRES1" c11 "HI1" c12 "COOK1"
MTB > Regress 'ES' 5 'area' 'anear' 'dist' 'distSC' 'elev';
SUBC>   SResiduals 'SRES1';
SUBC>   Tresiduals 'TRES1';
SUBC>   Hi 'HI1';
SUBC>   Cookd 'COOK1';
SUBC>   GFourpack;
SUBC>   RType 2;
SUBC>   Constant;
SUBC>   Brief 2.
Regression Analysis: ES versus area, anear, dist, distSC, elev

The regression equation is
ES = 9.54 - 0.00834 area - 0.0174 anear + 0.003 dist - 0.0774 distSC
     + 0.0807 elev

Predictor       Coef   SE Coef      T      P
Constant       9.539     4.279   2.23  0.037
area       -0.008339  0.004677  -1.78  0.089
anear      -0.017427  0.003823  -4.56  0.000
dist          0.0027    0.2174   0.01  0.990
distSC      -0.07737   0.04528  -1.71  0.102
elev         0.08069   0.01144   7.05  0.000

S = 12.4693   R-Sq = 83.4%   R-Sq(adj) = 79.4%
```

```
Analysis of Variance

Source          DF       SS      MS      F      P
Regression       5  16398.9  3279.8  21.09  0.000
Residual Error  21   3265.1   155.5
Total           26  19664.1

... (some lines deleted)

Unusual Observations

Obs  area     ES    Fit  SE Fit  Residual  St Resid
  9   634  35.00  36.07   12.29     -1.07     -0.51 X
 13  4669  89.00  95.11   12.28     -6.11     -2.81RX
 16    60  37.00  60.31    6.33    -23.31     -2.17R
 22   904  95.00  71.71    5.46     23.29      2.08R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.


MTB > Print 'island'-'COOK1'.
Data Display

Row  island          NS  ES     area    anear  dist  distSC  elev    SRES1    TRES1     HI1   COOK1
  1  Baltra          58  23    25.09     1.84   0.6     0.6   100   0.4774   0.4685  0.0904   0.003
  2  Bartolome       31  21     1.24   572.33   0.6    26.3   109   1.2479   1.2657  0.1097   0.032
  3  Caldwell         3   3     0.21     0.78   2.8    58.7   114  -0.9293  -0.9262  0.0679   0.010
  4  Champion        25   9     0.10     0.18   1.9    47.4    46  -0.0489  -0.0477  0.0807   0.000
  5  Daphne_Major    18  11     0.34     1.84   8.0     8.0   120  -0.6366  -0.6273  0.0858   0.006
  6  Darwin          10   7     2.33     2.85  34.1   290.2   168   0.7026   0.6939  0.4775   0.075
  7  Enderby          2   2     0.18     0.10   2.6    50.2   112  -1.0536  -1.0565  0.0660   0.013
  8  Espanola        97  26    58.27     0.57   1.1    88.3   198   0.6603   0.6512  0.1006   0.008
  9  Fernandina      93  35   634.49  4669.32   4.3    95.3  1494  -0.5135  -0.5043  0.9721   1.536
 10  Gardner1        58  17     0.57    58.27   1.1    93.1    49   1.0029   1.0031  0.1207   0.023
 11  Gardner2         5   4     0.78     0.21   4.6    62.2   227  -1.5730  -1.6344  0.0572   0.025
 12  Genovesa        40  19    17.35   129.49  47.4    92.2    76   1.3745   1.4062  0.4481   0.255
 13  Isabela        347  89  4669.32   634.49   0.7    28.1  1707  -2.8114  -3.4743  0.9696  42.043
 14  Marchena        51  23   129.49    59.56  29.1    85.9   343  -0.4690  -0.4601  0.1054   0.004
 15  Onslow           2   2     0.01     0.10   3.3    45.9    25  -0.5034  -0.4943  0.0828   0.003
 16  Pinta          104  37    59.56   129.49  29.1   119.6   777  -2.1695  -2.4036  0.2576   0.272
 17  Pinzon         108  33    17.95     0.03  10.7    10.7   458  -1.0650  -1.0686  0.1075   0.022
 18  Las_Plazas      12   9     0.23    25.09   0.5     0.6    25  -0.1761  -0.1720  0.1101   0.000
 19  Rabida          70  30     4.89   572.33   4.4    24.4   367   0.2269   0.2217  0.0636   0.000
 20  San_Cristobal  280  65   551.62     0.57  45.2    66.5   716   0.7428   0.7346  0.3759   0.055
 21  San_Salvador   237  81   572.33     4.89   0.2    19.8   906   0.4515   0.4428  0.2898   0.013
 22  Santa_Cruz     444  95   903.82     0.52   0.6     0.0   864   2.0776   2.2747  0.1918   0.170
 23  Santa_Fe        62  28    24.08     0.52  16.5    16.5   259  -0.0840  -0.0820  0.0955   0.000
 24  Santa_Maria    285  73   170.92     0.10   2.6    49.2   640   1.5117   1.5628  0.1823   0.085
 25  Seymour         44  16     1.84    25.09   0.6     9.6    30   0.4426   0.4339  0.1006   0.003
 26  Tortuga         16   8     1.24    17.95   6.8    50.9   186  -1.0114  -1.0120  0.0481   0.008
 27  Wolf            21  12     2.85     2.33  34.1   254.7   253   0.1703   0.1663  0.3417   0.002
```

Minitab plots for Question 3:



Stata do-file for Question 3: (Stata output available upon request)

```
pwcorr NS ES area anear dist distSC elev, sig
regress ES area anear dist distSC elev
predict fit, xb
predict stdres, rstandard
predict delres, rstudent
predict lev, hat
predict cook, cooksd
scatter stdres fit
qnorm stdres
list NS-cook, table
```