Biostats VHM 802 Course
Winter 2019, Atlantic Veterinary College, PEI
Henrik Stryhn

Final examination, 18 April 2019

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 paper by Koziol et al. (1981) describes an experiment in which 30 mice were injected with mouse colon carcinoma cells. The mice were divided into three treatment groups (A,B,C), and 5 days after injection the treatments — immunotherapy regimens — commenced. Group A received injections of tissue culture medium around the growing tumour, group B in addition received injections of normal spleen cells, and group C received injections of normal spleen cells, immune RNA and tumour antigens. The tumour volumes (in mm^2) in each mice were measured by a non-destructive procedure at days 7, 11, 12, 13, 14, 15 and 17 after injection.

The data are displayed in 4 graphs on the next two pages. Use the graphs and the information contained in the subsequent Minitab and Stata listings to answer to the following questions; note that the Minitab and Stata listings do not contain the same information.

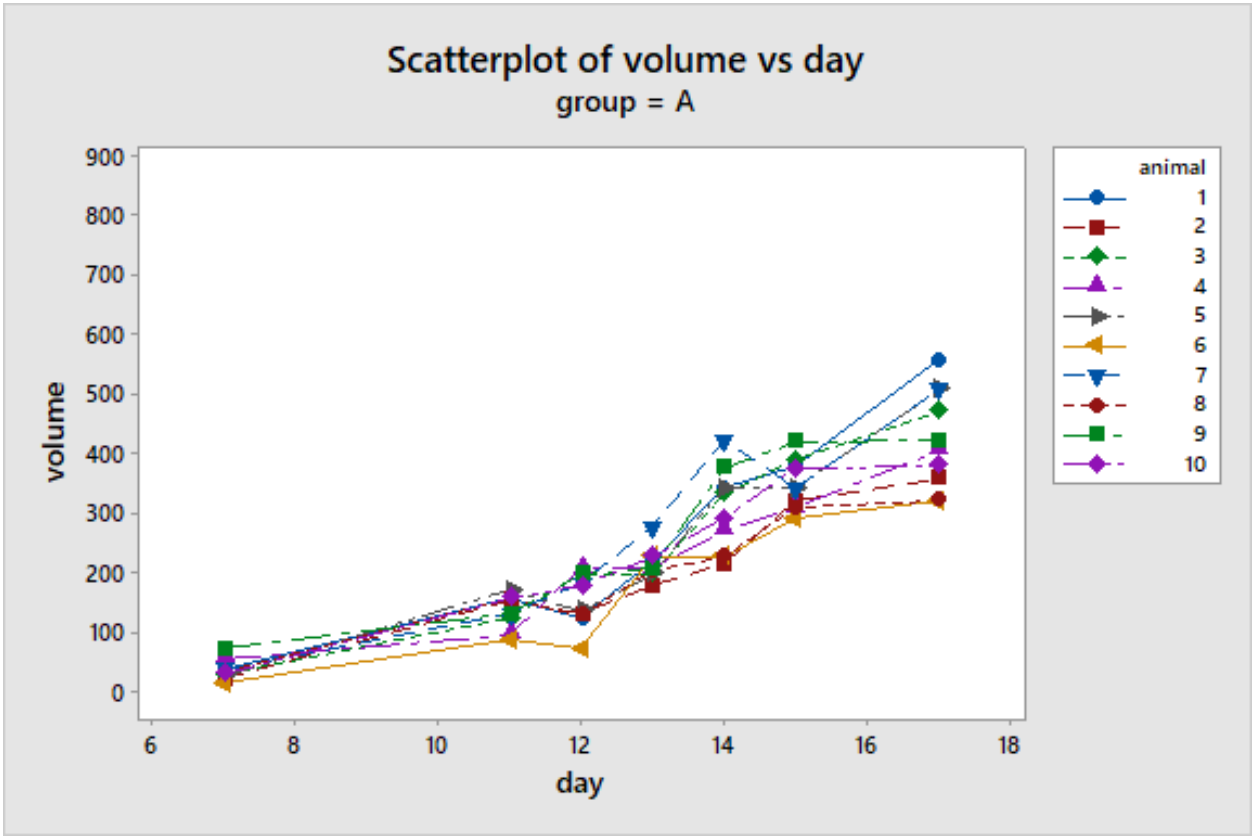
A) (*3 points*) Describe the data structure and experimental design. Interpret the graphs; in particular, what do they tell us about treatment, time and animal effects?

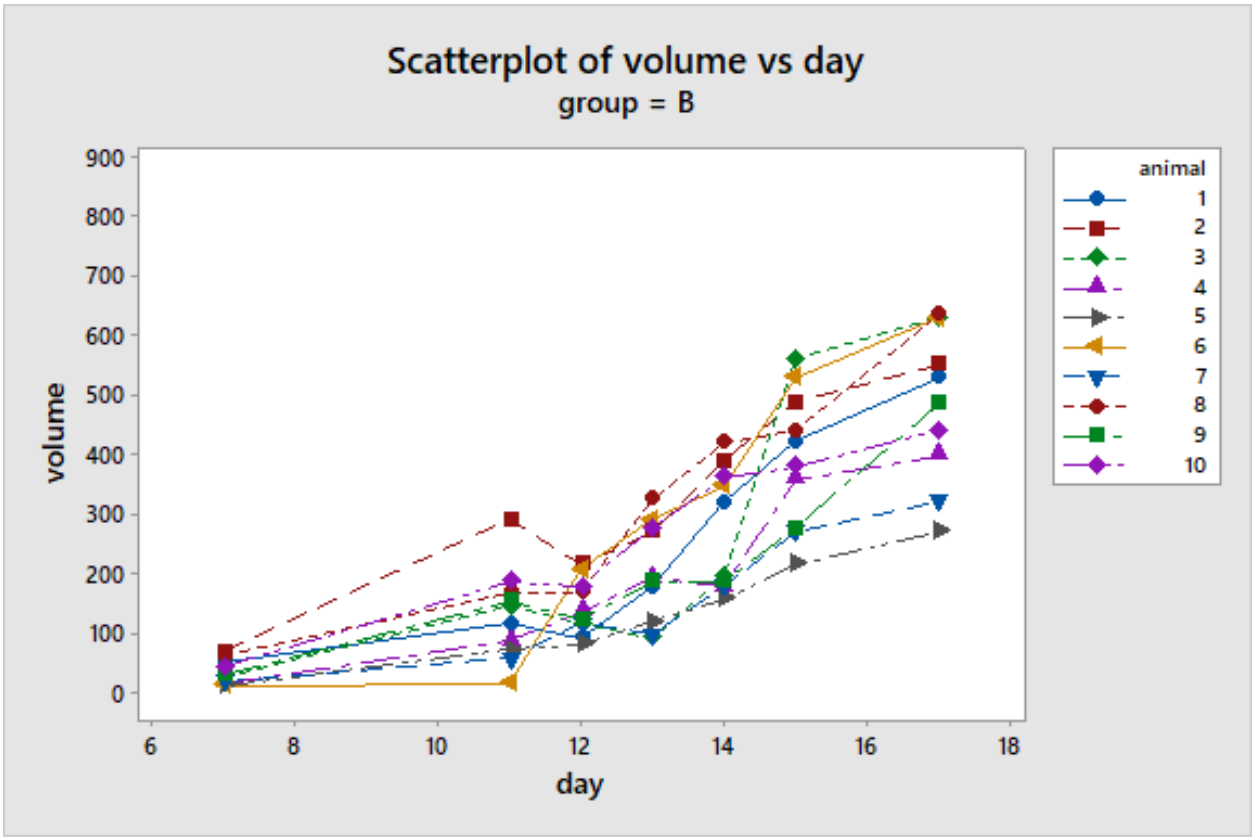
B) (*2 points*) Describe the statistical analyses, and the underlying statistical models, carried out for the listings presented for Question 1.B. Draw conclusions from the analyses, both with respect to the effects of interest in the study and the suitability of the models/analyses for these data. Include any suggestions for improvements of the statistical analysis you might have, and then give also details for how you would carry out the proposed analyses (in statistical sofware of your own choice).

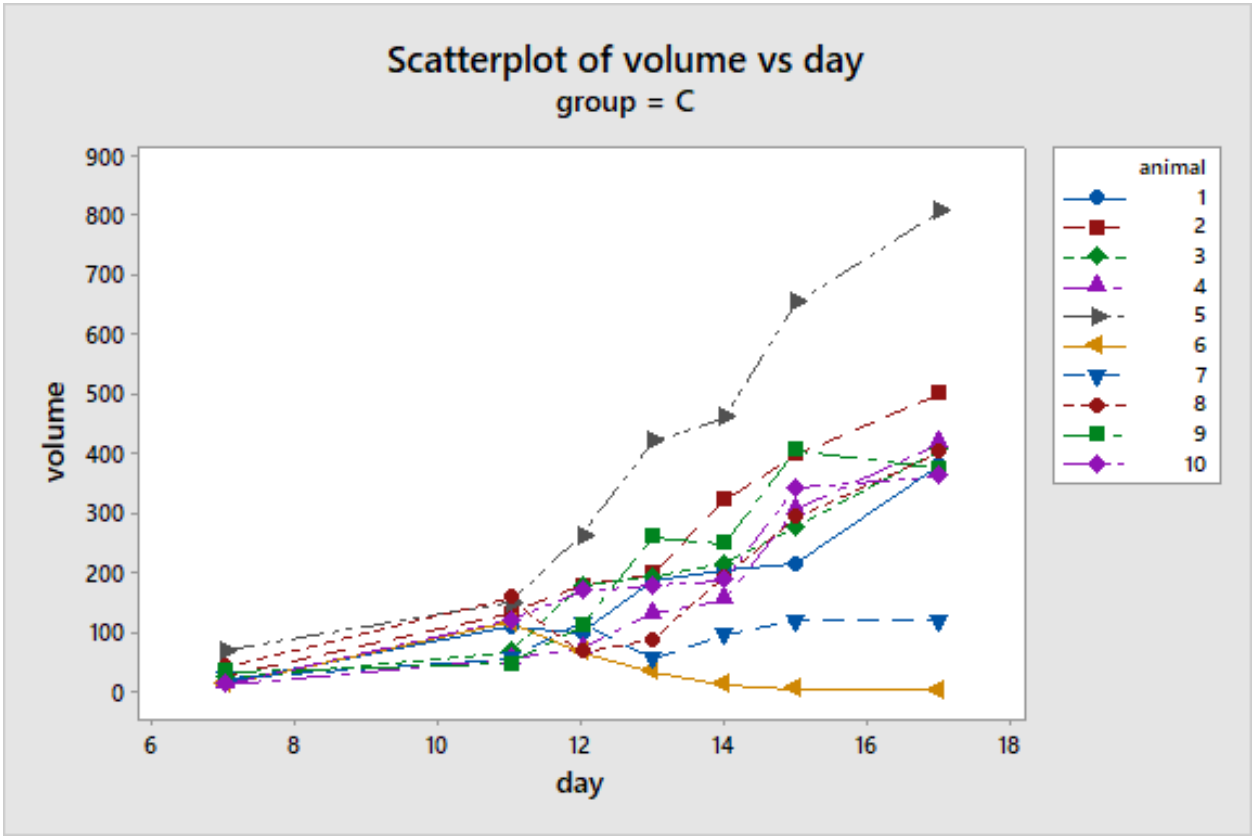
C) (*2 points*) Suggest some (at least two) response features (or summary statistics) to analyse the data. Make sure to motivate clearly your chosen features (features suggested without a motivation will not be credited). Explain how you would analyse your chosen features, both in terms of the statistical models used and the actual analysis in statistical sofware.

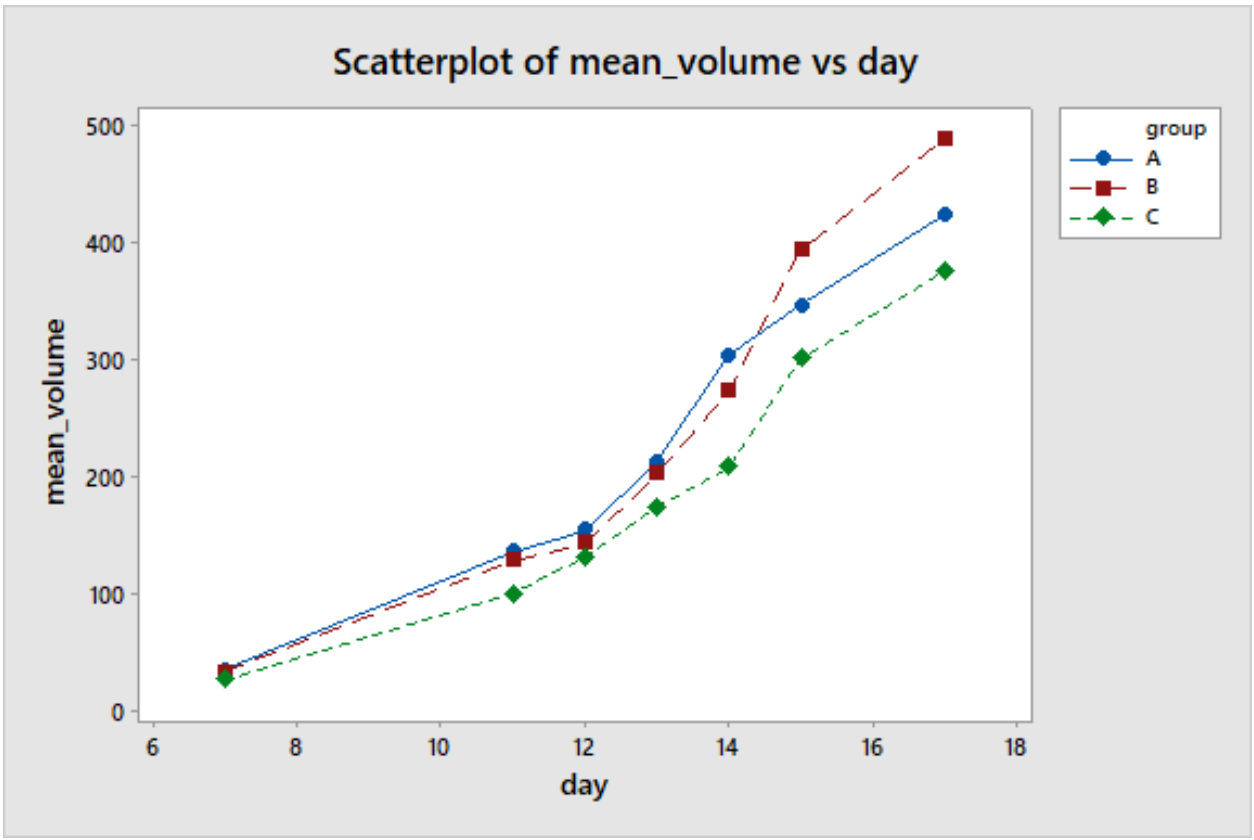
D) (*3 points*) Describe the statistical analysis, and the underlying statistical model, carried out for the listing presented for Question 1.D. Draw conclusions from the analysis, both with respect to the effects of interest in the study and the suitability of the models/analyses for these data. Include any suggestions for improvements of the statistical analysis you might have, and then give also details for how you would carry out the proposed analyses in statistical sofware.

Profile and mean plots for Question 1:






Scatterplot of volume vs day
group = C
volume
day
animal
1
2
3
4
5
6
7
8
9
10


Scatterplot of mean_volume vs day
mean_volume
day
group
A
B
C

Stata analyses for Question 1.B:

```
. oneway volume group if day==7, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |       33.86   16.932887          10
         B |   32.479999   20.523308          10
         C |       25.19   17.622299          10
-----------+------------------------------------
     Total |       30.51   18.194663          30

                      Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      434.057991      2   217.028996      0.64     0.5355
 Within groups      9166.26857     27   339.491429
------------------------------------------------------------------------
    Total           9600.32656     29   331.045744

Bartlett's test for equal variances:  chi2(2) =   0.3635  Prob>chi2 = 0.834


. oneway volume group if day==11, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |      135.39   28.289513          10
         B |      127.93   78.087075          10
         C |   99.240001   41.463672          10
-----------+------------------------------------
     Total |   120.85333   54.087865          30

                      Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      7285.30041      2   3642.65021      1.27     0.2976
 Within groups      77554.1152     27   2872.37464
------------------------------------------------------------------------
    Total           84839.4156     29   2925.49709

Bartlett's test for equal variances:  chi2(2) =   8.9059  Prob>chi2 = 0.012


. oneway volume group if day==12, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |       153.9   43.636834          10
         B |      142.34   46.703825          10
         C |      130.06   63.582183          10
-----------+------------------------------------
     Total |       142.1   51.191197          30
```

```
                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      2842.59204      2   1421.29602      0.52     0.5977
 Within groups        73153.03     27   2709.37148
------------------------------------------------------------------------
    Total            75995.622     29   2620.53869

Bartlett's test for equal variances:  chi2(2) =   1.4519  Prob>chi2 = 0.484


. oneway volume group if day==13, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |      212.53   26.306573          10
         B |      201.62   84.102025          10
         C |      173.06    112.0187          10
-----------+------------------------------------
     Total |   195.73667   81.182848          30

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      8308.61017      2   4154.30509      0.61     0.5488
 Within groups      182820.379     27   6771.12515
------------------------------------------------------------------------
    Total           191128.989     29    6590.6548

Bartlett's test for equal variances:  chi2(2) =  13.8959  Prob>chi2 = 0.001


. oneway volume group if day==14, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |      303.15   69.921757          10
         B |      272.58   103.21937          10
         C |      207.77   121.80545          10
-----------+------------------------------------
     Total |   261.16667   105.18611          30

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      47440.6864      2   23720.3432      2.34     0.1153
 Within groups       273418.72     27   10126.6193
------------------------------------------------------------------------
    Total           320859.407     29   11064.1175

Bartlett's test for equal variances:  chi2(2) =   2.5339  Prob>chi2 = 0.282
```

```
. oneway volume group if day==15, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |      346.17    42.27671          10
         B |       393.2   115.78009          10
         C |      300.12   175.12771          10
-----------+------------------------------------
     Total |   346.49667   125.40687          30

                       Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      43321.0284      2   21660.5142      1.42     0.2599
 Within groups       412758.57     27   15287.3545
------------------------------------------------------------------------
    Total           456079.599     29   15726.8827

Bartlett's test for equal variances:  chi2(2) =  13.5633  Prob>chi2 = 0.001


. oneway volume group if day==17, tabulate

           |          Summary of volume
     group |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         A |      424.03   83.326822          10
         B |      488.48   130.81083          10
         C |      375.88   213.57763          10
-----------+------------------------------------
     Total |   429.46333   154.34721          30

                       Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      63836.6111      2   31918.3056      1.37     0.2701
 Within groups      627032.144     27   23223.4128
------------------------------------------------------------------------
    Total           690868.756     29   23823.0605

Bartlett's test for equal variances:  chi2(2) =   7.1845  Prob>chi2 = 0.028
```

Minitab and Stata analyses for Question 1.D:

```
MTB > REML;
SUBC>   Response 'volume';
SUBC>   Random 'animal';
SUBC>   Categorical animal group day;
SUBC>   Nest animal(group);
SUBC>   Terms group day animal group*day;
SUBC>   Means group day;
...
```

(continues on the next page)

Mixed Effects Model: volume versus group, animal, day

Method

Variance estimation	Restricted maximum likelihood
DF for fixed effects	Kenward-Roger

Factor Information

Factor	Type	Levels	Values
group	Fixed	3	A, B, C
animal(group)	Random	30	1(A), 2(A), 3(A), 4(A), 5(A), 6(A), 7(A), 8(A), 9(A), 10(A), 1(B), 2(B), 3(B), 4(B), 5(B), 6(B), 7(B), 8(B), 9(B), 10(B), 1(C), 2(C), 3(C), 4(C), 5(C), 6(C), 7(C), 8(C), 9(C), 10(C)
day	Fixed	7	7, 11, 12, 13, 14, 15, 17

Variance Components

Source	Var	% of Total	SE Var	Z-Value	P-Value
animal(group)	4344.151608	49.58%	1355.888221	3.203916	0.001
Error	4417.241153	50.42%	490.804573	9.000000	0.000
Total	8761.392761				

-2 Log likelihood = 2253.848385

Tests of Fixed Effects

Term	DF Num	DF Den	F-Value	P-Value
group	2.00	27.00	1.45	0.253
day	6.00	162.00	128.97	0.000
group*day	12.00	162.00	1.37	0.185

Model Summary

S	R-sq	R-sq(adj)
66.4623	86.07%	84.60%

Conditional Fits and Diagnostics for Unusual Observations

Obs	volume	Fit	Resid	Std Resid	
25	66.599998	212.568429	-145.968431	-2.474532	R
26	11.400000	-108.556220	119.956220	2.033559	R
46	15.000000	173.792300	-158.792300	-2.691928	R
55	147.000000	286.618430	-139.618430	-2.366883	R
56	115.200000	-34.506219	149.706219	2.537897	R
103	90.699997	213.902914	-123.202917	-2.088599	R
163	559.599980	405.482914	154.117066	2.612671	R
175	653.400020	487.498433	165.901587	2.812449	R
176	3.200000	166.373784	-163.173784	-2.766206	R
193	629.299990	500.762914	128.537076	2.179026	R
195	268.799990	396.382441	-127.582451	-2.162843	R
205	806.400020	563.258433	243.141587	4.121861	R
206	1.400000	242.133784	-240.733784	-4.081042	R
207	118.300000	284.407379	-166.107379	-2.815937	R

R Large residual

Conditional Means

Term	Fitted Mean	SE Mean	DF	T-Value	P-Value
group					
A	229.861	22.3051	27.0000	10.31	0.000
B	236.947	22.3051	27.0000	10.62	0.000
C	187.331	22.3051	27.0000	8.40	0.000
day					
7	30.510	17.0894	76.3612	1.79	0.078
11	120.853	17.0894	76.3612	7.07	0.000
12	142.100	17.0894	76.3612	8.32	0.000
13	195.737	17.0894	76.3612	11.45	0.000
14	261.167	17.0894	76.3612	15.28	0.000
15	346.497	17.0894	76.3612	20.28	0.000
17	429.463	17.0894	76.3612	25.13	0.000

```
. encode group, gen(Group)

. mixed volume Group##day || animalid:, reml

Performing EM optimization:
...

Mixed-effects REML regression                   Number of obs     =        210
Group variable: animalid                        Number of groups  =         30

                                                Obs per group:
                                                              min =          7
                                                              avg =        7.0
                                                              max =          7

                                                Wald chi2(20)     =     793.16
Log restricted-likelihood = -1113.3962          Prob > chi2       =     0.0000
------------------------------------------------------------------------------
      volume |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Group |
           B |  -1.380001   41.86024    -0.03   0.974    -83.42455    80.66455
           C |      -8.67   41.86024    -0.21   0.836    -90.71455    73.37455
             |
         day |
          11 |     101.53   29.72285     3.42   0.001     43.27428    159.7857
          12 |     120.04   29.72285     4.04   0.000     61.78428    178.2957
          13 |     178.67   29.72285     6.01   0.000     120.4143    236.9257
          14 |     269.29   29.72285     9.06   0.000     211.0343    327.5457
          15 |     312.31   29.72285    10.51   0.000     254.0543    370.5657
          17 |     390.17   29.72285    13.13   0.000     331.9143    448.4257
             |
   Group#day |
       B#11  |  -6.079999   42.03446    -0.14   0.885    -88.46604    76.30604
       B#12  |     -10.18   42.03446    -0.24   0.809    -92.56603    72.20604
       B#13  |  -9.530003   42.03446    -0.23   0.821    -91.91604    72.85603
       B#14  |  -29.19001   42.03446    -0.69   0.487     -111.576    53.19603
       B#15  |      48.41   42.03446     1.15   0.249    -33.97603     130.796
       B#17  |      65.83   42.03446     1.57   0.117    -16.55604     148.216
       C#11  |     -27.48   42.03446    -0.65   0.513     -109.866    54.90604
       C#12  |     -15.17   42.03446    -0.36   0.718    -97.55604    67.21604
       C#13  |      -30.8   42.03446    -0.73   0.464     -113.186    51.58603
       C#14  |     -86.71   42.03446    -2.06   0.039     -169.096   -4.323966
       C#15  |  -37.37999   42.03446    -0.89   0.374     -119.766    45.00604
       C#17  |  -39.47999   42.03446    -0.94   0.348     -121.866    42.90604
             |
       _cons |      33.86   29.59966     1.14   0.253    -24.15426    91.87426
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
animalid: Identity           |
                  var(_cons) |   4344.156    1355.89      2356.297    8009.047
-----------------------------+------------------------------------------------
               var(Residual) |    4417.24   490.8044      3552.819     5491.98
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 73.68         Prob >= chibar2 = 0.0000
```

Question 2.

To study 3 diets and their impact on the cholesterol levels in humans, a study with 60 subjects (volunteers) is planned. The 60 subjects should be equally divided between women and men. Each subject will follow a diet for two weeks, and the blood cholesterol level is measured (in mg/dL) at the onset and end of this period. The intended outcome variable for data analysis is the decline (difference) between pre- and post-diet cholesterol levels. The data layout might be as follows,

Person	Sex	Diet	Pre-chol	Post-chol	Decline-chol
1	f	1	232	204	28
2	f	1	178	182	-4
...	...	...	...	...	...
60	m	3	205	192	13

A) (*3 points*) Describe the experimental design, and explain briefly how randomization would enter into the planning/execution of the study. Next, describe the statistical model you would use to analyze the data. Finally, outline the statistical analysis, either by giving an ANOVA table with all effects and their associated degrees of freedom, or by specifying how to perform the analysis in statistical sofware of your own choice. In the latter case, make sure to include all relevant specifications for the analysis (excluding model validation) in the statistical software.

B) (*2 points*) It is well-known that cholesterol levels may vary with (be affected by) age. Describe how you would take into account age in the planning and/or analysis of the data. Give as in A) the statistical model and an outline of the statistical analysis. (*Note:* Several possibilities exist, but a full explanation of one them is considered sufficient.)

C) (*3 points*) The researchers also considered an alternative layout of the study where each subject goes through all 3 diets, in different periods. Discuss briefly the general advantages and/or disadvantages such a layout of the study may have. Describe then again as in A) the experimental design, the statistical model you would use to analyze the data, and the statistical analysis. (*Hint:* It may be helpful to sketch the layout of the data in a similar way as above.)

D) (*2 points*) Prior to carrying out the study, the researchers considered a difference between two of the diets (in terms of their cholesterol declines) of 10 units to be of biological interest. Furthermore, the variation between choleresterol declines (in general) were expected to correspond to a standard deviation of 20 units, and the correlation between two cholesterol declines of the same subject was expected to be 0.3.

Explain how you would determine, based on one of the models from A)−C) of your own choice, whether the researchers were likely to find an observed difference of 10 units between the two diets to be statistically significant, and carry out as much of the calculation for the answer as you can from the information provided. If additional information is needed, explain how you would obtain it and how it would be used to come up with the desired answer.

Question 3.

A large epidemiological study was carried out to investigate risk factors for Coronary Heart Disease (CHD). The variables investigated included those listed in the table below. Except for CHD, these variables were recorded at study onset, and occurrence of CHD was monitored during a $8-9$ year follow-up period. The study population consisted of employed, male volunteers within a certain geographical area.

Variable	Description	Values
chd	coronary heart disease	0/1
age	age in years	$39-59$
height	height in inches	$60-78$
weight	weight in pounds	$78-320$
bmi	body-mass index	$11.2-38.9$
smoke	current smoker	0/1
sbp	systolic blood pressure	$98-230$
dbp	diastolic blood pressure	$58-150$
chol	blood cholesterol level	$103-645$
arcus	arcus sinus condition	0/1
behpat	behaviour pattern	1/2/3/4

The variable "behaviour pattern" classified subjects into four groups based on questionnaire information. The groups 1 and 2 may be considered as a subdivision of general behaviour type A, and similarly the groups 3 and 4 corresponded to two variants of a behaviour type B.

A) (*3 points*) Explain the statistical model used in the Stata listing presented for Question 3.A; you may either give a model formula or a verbal description of the model and the outcome/parameter being modelled. Interpret the effects of the variables included the model; your interpretation should for each effect include both a quantitative statement of the size of the effect and a statement about its statistical significance. If you think that additional analysis is needed to assess some effects of interest, outline how the additional analysis should be carried out (in Stata, or possibly another software of your choice) and indicate how you would use the information obtained from it.

B) (*3 points*) Review the assumptions behind the statistical model, and describe what checks in your view would be needed to validate the shown model. As part of your discussion, explain what the statistics already included in the listing tell you about the validity of the model.

C) (*4 points*) Use the information provided in the Stata listings (both for Questions 3.A and 3.C) to outline a strategy for developing a good predictive model for CHD. Explain and motivate the different steps you would go through in developing the model, while explaining how you use the information from the analyses already carried out in the process. Specifically, among the three models shown in the listings, which one do you think is preferable for prediction of CHD? And can you identify any shortcomings of that model or suggestions for improvement?

Stata listing for Question 3.A:

```
. logit chd age weight smoke i.behpat arcus

Iteration 0:   log likelihood =     -885.6
Iteration 1:   log likelihood = -831.19992
Iteration 2:   log likelihood = -825.84683
Iteration 3:   log likelihood = -825.83768
Iteration 4:   log likelihood = -825.83768

Logistic regression                             Number of obs     =      3,152
                                                LR chi2(7)        =     119.52
                                                Prob > chi2       =     0.0000
Log likelihood = -825.83768                     Pseudo R2         =     0.0675
------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0689622   .0118252     5.83   0.000     .0457853    .0921391
      weight |   .0128522   .0030188     4.26   0.000     .0069354     .018769
       smoke |   .6654465   .1382554     4.81   0.000      .394471    .9364221
             |
      behpat |
          2  |   .0541279   .2169931     0.25   0.803    -.3711707    .4794264
          3  |  -.7277124   .2393312    -3.04   0.002    -1.196793   -.2586318
          4  |  -.6018242   .3154834    -1.91   0.056     -1.22016    .0165119
             |
       arcus |   .2852151   .1398452     2.04   0.041     .0111236    .5593066
       _cons |  -8.102231   .8473147    -9.56   0.000    -9.762937   -6.441525
------------------------------------------------------------------------------

. estat gof

Logistic model for chd, goodness-of-fit test

       number of observations =      3152
 number of covariate patterns =      2474
           Pearson chi2(2466) =      2457.40
                 Prob > chi2 =         0.5450

. estat gof, group(8)

Logistic model for chd, goodness-of-fit test

  (Table collapsed on quantiles of estimated probabilities)

       number of observations =      3152
             number of groups =         8
     Hosmer-Lemeshow chi2(6) =         4.28
                 Prob > chi2 =         0.6386

. estat ic

Akaike's information criterion and Bayesian information criterion
------------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+----------------------------------------------------------------
           . |      3,152    -885.6  -825.8377       8    1667.675   1716.122
------------------------------------------------------------------------------
```

Stata listing for Question 3.C:

```
. logit chd age height weight bmi smoke sbp dbp chol arcus

Iteration 0:   log likelihood = -884.58565
Iteration 1:   log likelihood = -810.60763
Iteration 2:   log likelihood = -799.13846
Iteration 3:   log likelihood = -799.12406
Iteration 4:   log likelihood = -799.12406

Logistic regression                             Number of obs     =      3,140
                                                LR chi2(9)        =     170.92
                                                Prob > chi2       =     0.0000
Log likelihood = -799.12406                     Pseudo R2         =     0.0966
------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0646289   .0123198     5.25   0.000     .0404826    .0887752
      height |   .1466775   .2278245     0.64   0.520    -.2998504    .5932054
      weight |  -.0171515   .0449403    -0.38   0.703    -.1052329    .0709298
         bmi |   .1852966   .3148128     0.59   0.556    -.4317252    .8023184
       smoke |   .6282544   .1419017     4.43   0.000     .3501322    .9063766
         sbp |   .0192738   .0063903     3.02   0.003     .0067491    .0317986
         dbp |  -.0008293   .0108349    -0.08   0.939    -.0220652    .0204067
        chol |   .0108727   .0015148     7.18   0.000     .0079038    .0138415
       arcus |   .2062438   .1430328     1.44   0.149    -.0740953    .4865829
       _cons |  -22.81706   16.01255    -1.42   0.154    -54.20109    8.566964
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion
-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df          AIC         BIC
-------------+---------------------------------------------------------------
           . |      3,140 -884.5856  -799.1241      10    1618.248    1678.768
-----------------------------------------------------------------------------

. estat vce, corr

Correlation matrix of coefficients of logit model

             | chd
        e(V) |      age    height    weight       bmi     smoke       sbp       dbp      chol
-------------+------------------------------------------------------------------------------
chd          |
         age |   1.0000
      height |   0.0226    1.0000
      weight |  -0.0065   -0.9925    1.0000
         bmi |   0.0113    0.9895   -0.9963    1.0000
       smoke |   0.0712    0.0188   -0.0184    0.0289    1.0000
         sbp |  -0.0987   -0.0311    0.0302   -0.0316   -0.0842    1.0000
         dbp |  -0.0079    0.0209   -0.0240    0.0065    0.1110   -0.7623    1.0000
        chol |   0.0366    0.0493   -0.0366    0.0352   -0.0773   -0.0002   -0.0466    1.0000
       arcus |  -0.2089   -0.0201    0.0157   -0.0133   -0.0515    0.0062    0.0383   -0.0962
       _cons |  -0.0568   -0.9981    0.9896   -0.9894   -0.0317    0.0259   -0.0282   -0.0696
```

```
             | chd
        e(V) |    arcus     _cons
-------------+--------------------
chd          |
       arcus |   1.0000
       _cons |   0.0231    1.0000

. pwcorr age height weight bmi smoke sbp dbp chol arcus

             |      age   height   weight      bmi    smoke      sbp      dbp
-------------+---------------------------------------------------------------
         age |   1.0000
      height |  -0.0954   1.0000
      weight |  -0.0344   0.5329   1.0000
         bmi |   0.0266  -0.0658   0.8066   1.0000
       smoke |   0.0048  -0.0036  -0.1207  -0.1417   1.0000
         sbp |   0.1657   0.0184   0.2532   0.2878   0.0026   1.0000
         dbp |   0.1392   0.0103   0.2959   0.3424  -0.0830   0.7729   1.0000
        chol |   0.0892  -0.0889   0.0085   0.0706   0.0975   0.1231   0.1296
       arcus |   0.1867   0.0178  -0.0191  -0.0349   0.0734   0.0340   0.0031

             |     chol    arcus
-------------+------------------
        chol |   1.0000
       arcus |   0.1214   1.0000

. logit chd age smoke sbp chol arcus

Iteration 0:   log likelihood = -884.58565
Iteration 1:   log likelihood = -814.81323
Iteration 2:   log likelihood = -804.31135
Iteration 3:   log likelihood = -804.30009
Iteration 4:   log likelihood = -804.30009

Logistic regression                             Number of obs     =      3,140
                                                LR chi2(5)        =     160.57
                                                Prob > chi2       =     0.0000
Log likelihood = -804.30009                     Pseudo R2         =     0.0908
------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0598737    .012183     4.91   0.000     .0359956    .0837519
       smoke |   .5785868   .1390849     4.16   0.000     .3059855    .8511881
         sbp |   .0218924    .003934     5.56   0.000     .0141819    .0296029
        chol |   .0106913   .0014908     7.17   0.000     .0077693    .0136132
       arcus |   .2080544   .1424056     1.46   0.144    -.0710554    .4871643
       _cons |   -11.0517   .8079883   -13.68   0.000    -12.63533   -9.468073
------------------------------------------------------------------------------

. estat ic

Akaike's information criterion and Bayesian information criterion
------------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+----------------------------------------------------------------
           . |      3,140 -884.5856  -804.3001       6      1620.6   1656.912
------------------------------------------------------------------------------
```