**Assignment 1 (12 points)**

A longitudinal study investigated the effectiveness of a dietary intervention on body mass index (BMI). The intervention was at the population level, so everyone received the intervention. In the study, a measurement was taken several months before the intervention, at the time the intervention was started, and a measurement several months after the start of the intervention. In the dataset, two periods were distinguished: the period before the intervention is coded with 0 and the period after the intervention is coded with 1. Because the measurements were not taken at a fixed time, the variable time is a continuous variable in weeks. The aim of the study was to investigate whether the course over time in BMI was different in the period before the intervention than in the period during the intervention. Output 1.1 shows the result of the analysis.

Output 1.1

Mixed-effects ML regression Number of obs = 228

Group variable: id Number of groups = 57

 Obs per group:

 min = 4

 avg = 4.0

 max = 4

 Wald chi2(3) = 84.44

Log likelihood = -453.81584 Prob > chi2 = 0.0000

----------------------------------------------------------------------------------

 bmi | Coefficient Std. err. z P>|z| [95% conf. interval]

-----------------+----------------------------------------------------------------

 period | .1966332 .2077244 0.95 0.344 -.210499 .6037655

 time | 1.104516 .1489925 7.41 0.000 .8124963 1.396536

 |

 c.period#c.time | -1.145126 .5183836 -2.21 0.027 -2.16114 -.1291132

 |

 \_cons | 30.48301 .420871 72.43 0.000 29.65812 31.3079

----------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

id: Identity |

 var(\_cons) | 8.9643 1.744127 6.122143 13.12591

-----------------------------+------------------------------------------------

 var(Residual) | 1.374692 .1486748 1.112107 1.699276

------------------------------------------------------------------------------

LR test vs. linear model: chibar2(01) = 271.50 Prob >= chibar2 = 0.0000

1. What is the interpretation of the regression coefficient for period? (2 points)
2. What is the interpretation of the regression coefficient for time? (2 points)
3. Is the intervention effective? Explain the answer using Output 1.1. (2 points)

In a second analysis, a random slope for time was added to the model. Output 1.2 shows the result of this analysis.

Output 1.2

Mixed-effects ML regression Number of obs = 228

Group variable: id Number of groups = 57

 Obs per group:

 min = 4

 avg = 4.0

 max = 4

 Wald chi2(3) = 51.93

Log likelihood = -449.71104 Prob > chi2 = 0.0000

----------------------------------------------------------------------------------

 bmi | Coefficient Std. err. z P>|z| [95% conf. interval]

-----------------+----------------------------------------------------------------

 period | .1276168 .1810189 0.70 0.481 -.2271737 .4824074

 time | 1.34459 .2088948 6.44 0.000 .935164 1.754017

 |

 c.period#c.time | -1.446722 .4646707 -3.11 0.002 -2.35746 -.5359843

 |

 \_cons | 30.56066 .4225948 72.32 0.000 29.73239 31.38893

----------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

id: Unstructured |

 var(time) | .8704925 .4130137 .3434821 2.206104

 var(\_cons) | 9.27731 1.790935 6.354805 13.54384

 cov(tijd,\_cons) | .3536352 .4983901 -.6231914 1.330462

-----------------------------+------------------------------------------------

 var(Residual) | .9964192 .1440176 .7506086 1.322728

------------------------------------------------------------------------------

LR test vs. linear model: chi2(3) = 279.71 Prob > chi2 = 0.0000

1. Is it necessary to add a random slope for time to the model? Explain the answer. (2 points)

In a third analysis it was investigated whether the intervention has a different effect for males than for females. For this purpose, the variable male was added to the model. Male is 1 coded for males and 0 for females. Output 1.3 shows the result of this analysis.

Output 1.3

Mixed-effects ML regression Number of obs = 228

Group variable: id Number of groups = 57

 Obs per group:

 min = 4

 avg = 4.0

 max = 4

 Wald chi2(7) = 60.46

Log likelihood = -447.55183 Prob > chi2 = 0.0000

----------------------------------------------------------------------------------

 bmi | Coefficient Std. err. z P>|z| [95% conf. interval]

-----------------+----------------------------------------------------------------

 period | -.0273119 .2598321 -0.11 0.916 -.5365735 .4819498

 time | 1.07413 .2780572 3.86 0.000 .5291481 1.619112

 |

 c.period#c.time | -1.350457 .6627854 -2.04 0.042 -2.649492 -.0514213

 |

 male | .1178369 .408438 0.29 0.773 -.6826869 .9183607

 |

 c.period#c.male | .2940749 .3624115 0.81 0.417 -.4162386 1.004388

 |

 c.time#c.male | .5789084 .4248395 1.36 0.173 -.2537617 1.411578

 |

 c.period#c.time |

 #c.male | -.3113001 .9851513 -0.32 0.752 -2.242161 1.619561

 |

 \_cons | 30.50451 .4709802 64.77 0.000 29.5814 31.42761

----------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

id: Unstructured |

 var(time) | .7294677 .3801192 .2626954 2.025628

 var(\_cons) | 9.225256 1.784065 6.314882 13.47695

 cov(tijd,\_cons) | .0885125 .4937986 -.8793149 1.05634

-----------------------------+------------------------------------------------

 var(Residual) | .9986258 .1449973 .7512958 1.327378

------------------------------------------------------------------------------

LR test vs. linear model: chi2(3) = 277.19 Prob > chi2 = 0.0000

1. Is the effect of the intervention different for males and females? Explain the answer. (2 points)
2. What is the interpretation of the regression coefficient for the interaction between period and time (i.e. -1.350457)? (2 points)

**Assignment 2 (10 points)**

A longitudinal study investigated the relationship between physical activity and physical fitness. Both variables are continuous, have been measured multiple times and have been scaled in such a way that the values are between 10 and 50. Output 2.1 shows the result of the first analysis.

Output 2.1

Mixed-effects ML regression Number of obs = 113

Group variable: id Number of groups = 57

 Obs per group:

 min = 1

 avg = 2.0

 max = 2

 Wald chi2(1) = 55.29

Log likelihood = -274.35898 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

 fitness | Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

 activity | .6819912 .0917173 7.44 0.000 .5022287 .8617538

 \_cons | 8.494127 2.419501 3.51 0.000 3.751993 13.23626

------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

id: Identity |

 var(\_cons) | 8.26258 2.117439 5.00011 13.65375

-----------------------------+------------------------------------------------

 var(Residual) | 2.894333 .6113673 1.913154 4.378718

------------------------------------------------------------------------------

LR test vs. linear model: chibar2(01) = 31.87 Prob >= chibar2 = 0.0000

1. What is the interpretation of the regression coefficient for activity? (2 points)

In a second analysis, the analysis was adjusted for smoking. The variable smoking is coded 0 for no smoking and 1 for smoking. Output 2.2 shows the result of this analysis.

Output 2.2

Mixed-effects ML regression Number of obs = 113

Group variable: id Number of groups = 57

 Obs per group:

 min = 1

 avg = 2.0

 max = 2

 Wald chi2(2) = 60.74

Log likelihood = -272.08724 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

 fitness | Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

 activity | .6733816 .0895962 7.52 0.000 .4977763 .8489869

 smoking | -1.202401 .5565651 -2.16 0.031 -2.293249 -.1115538

 \_cons | 8.093983 2.385995 3.39 0.001 3.417518 12.77045

------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

id: Identity |

 var(\_cons) | 8.240875 2.070813 5.035918 13.48553

-----------------------------+------------------------------------------------

 var(Residual) | 2.702445 .5673798 1.790798 4.078187

------------------------------------------------------------------------------

LR test vs. linear model: chibar2(01) = 33.86 Prob >= chibar2 = 0.0000

1. Is it necessary to adjust for smoking in this analysis? Explain the answer. (2 points)

1. Based on output 2.2, calculate the relationship between activity and fitness for smokers? (2 points)

The same study also investigated the longitudinal relationship between physical activity and quality of life. Quality of life is a dichotomous variable (qol), where a low quality of life is coded by 0 and a high quality of life is coded by 1. Output 2.3 shows the result of the logistic GEE analysis.

Output 2.3

GEE population-averaged model Number of obs = 113

Group variable: id Number of groups = 57

Family: Binomial Obs per group:

Link: Logit min = 1

Correlation: exchangeable avg = 2.0

 max = 2

 Wald chi2(1) = 29.66

Scale parameter = 1 Prob > chi2 = 0.0000

 (Std. err. adjusted for clustering on id)

------------------------------------------------------------------------------

 | Robust

 qol | Odds ratio std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

 activity | 2.031304 .2643415 5.45 0.000 1.574 2.621471

 \_cons | 9.72e-09 3.23e-08 -5.54 0.000 1.43e-11 6.62e-06

------------------------------------------------------------------------------

1. What is the interpretation of the odds ratio for activity? (2 points)

The researchers could also have analyzed the relationship between quality of life and activity with a logistic mixed model analysis.

1. If they had done so, would the odds ratio have been lower, higher, or stayed the same? Explain the answer. (2 points)

**Assignment 3 (6 points)**

A large study in social medicine investigated which variables could explain health differences between neighbourhoods and regions. To do this, first a multilevel analysis was performed without independent variables. Output 3.1 shows the result of this analysis.

Output 3.1

Mixed-effects ML regression Number of obs = 6,285

 Grouping information

 -------------------------------------------------------------

 | No. of Observations per group

 Group variable | groups Minimum Average Maximum

 ----------------+--------------------------------------------

 region | 11 244 571.4 1,187

 neighbourhood | 396 4 15.9 25

 -------------------------------------------------------------

 Wald chi2(0) = .

Log likelihood = -22082.846 Prob > chi2 = .

------------------------------------------------------------------------------

 health | Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

 \_cons | 51.37577 .2291838 224.17 0.000 50.92657 51.82496

------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

region: Identity |

 var(\_cons) | .3836962 .2548758 .1043673 1.410622

-----------------------------+------------------------------------------------

neighbourhood: Identity |

 var(\_cons) | 2.121595 .4531857 1.395856 3.224663

-----------------------------+------------------------------------------------

 var(Residual) | 64.13078 1.181311 61.85675 66.48841

------------------------------------------------------------------------------

LR test vs. linear model: chi2(2) = 57.36 Prob > chi2 = 0.0000

1. What is the interpretation of the constant (i.e. 51.37577)? (2 points)
2. What is the interrpetation of the random intercept at the neighbourhood level (i.e. 2.121595)? (2 points)

In a second analysis, social class was added to the model. Output 3.2 shows the result of this analysis.

Output 3.2

Mixed-effects ML regression Number of obs = 6,285

 Grouping information

 -------------------------------------------------------------

 | No. of Observations per group

 Group variable | groups Minimum Average Maximum

 ----------------+--------------------------------------------

 region | 11 244 571.4 1,187

 neighbourhood | 396 4 15.9 25

 -------------------------------------------------------------

 Wald chi2(1) = 4904.03

Log likelihood = -20273.138 Prob > chi2 = 0.0000

------------------------------------------------------------------------------

 health | Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

 soc\_class | 10.78027 .1539406 70.03 0.000 10.47855 11.08199

 \_cons | 45.85419 .1750313 261.98 0.000 45.51113 46.19724

------------------------------------------------------------------------------

------------------------------------------------------------------------------

 Random-effects parameters | Estimate Std. err. [95% conf. interval]

-----------------------------+------------------------------------------------

region: Identity |

 var(\_cons) | .1760037 .1173949 .0476177 .6505421

-----------------------------+------------------------------------------------

neighbourhood: Identity |

 var(\_cons) | .6630208 .2192706 .3467547 1.267745

-----------------------------+------------------------------------------------

 var(Residual) | 36.43668 .6714406 35.14416 37.77673

------------------------------------------------------------------------------

LR test vs. linear model: chi2(2) = 26.90 Prob > chi2 = 0.0000

1. How much of the difference in health between neighbourhoods is explained by social status? Explain the answer. (2 points)

**Assignment 4 (3 points)**

Open the file HEL-8003 part 1. For each question provide the syntax and the relevant output of the analysis as answer. You don’t have to say something about the interpretation.

a) Analyze the relationship between smoking and health (outcome) and add all necessary random coefficients. (2 points)

b) Analyze to what extent the relationship between smoking and health is different for subjects with overweight and without overweight. (1 point)

**Assignment 5 (4 points)**

Open the file HEL-8003 part 2. For each question provide the syntax and the relevant output of the analysis as answer. You don’t have to say something about the interpretation.

a) Analyze the effect of the physical therapy + manual therapy on the continuous outcome variable pain on average over time. (2 points)

b) Analyze the effect of the physical therapy + manual therapy on the continuous outcome variable pain at the different time points and show at which time point the intervention is most effective. To do this, show an output with the effect, the 95% confidence interval and the corresponding p-value. (2 points)