# York University

MATH 4939 – Final Exam

*Instructor: Georges Monette*

*April 5, 2019 – 9 am to 11 am (120 minutes)*

***WARNING***

**DO NOT OPEN THIS BOOKLET UNTIL YOU ARE INSTRUCTED TO DO SO**

**Student number:** _____

**Family name: (in BLOCK letters)**_____

**Given name: (in BLOCK letters)**_____

**Signature**_____

**Information:**

Be sure to read questions closely. Some may ask for multiple pieces of information. Make sure to respond completely. If you need more space to answer, write "**OVER**" and continue the answer on the back of the page.

The marks for each questions are shown at the end of the question. The sum of the marks is 175. The exam will be graded out of 160 so that you can potentially earn 15 bonus points.

**Aids allowed:** Non-programmable calculator, ruler, pencils, pens, erasers.

***WARNING***

**DO NOT OPEN THIS BOOKLET UNTIL YOU ARE INSTRUCTED TO DO SO**

1. Consider a mixed model of the form `lme( Y ~ X, data, random = ~ 1 + X| id)` in which there are two observations per cluster and the predictor, `X`, has the same two values, 0 and 1, in each cluster. Determine whether the variance parametrization of the model is identifiable. *(20 points)*

2. The output below uses the schizophrenia data in which patients were observed at years 1, 2, 3, 4, 5, and 6 taking one of three drugs: Atypical, Clozapine, Typical each year.

```
fitgl <- lme( gen ~ drug + cvar(drug,id) + year, dd, random = ~ 1 | id)
summary( fitgl )
```

```
|   Linear mixed-effects model fit by REML
|    Data: dd
|         AIC      BIC   logLik
|     2148.02 2177.964 -1066.01
|
|   Random effects:
|    Formula: ~1 | id
|           (Intercept) Residual
|   StdDev:    5.803759 6.097606
|
|   Fixed effects: gen ~ drug + cvar(drug, id) + year
|                               Value Std.Error  DF   t-value p-value
|   (Intercept)             33.64712  2.732709 262 12.312733  0.0000
|   drugClozapine           -1.55705  1.533425 262 -1.015409  0.3108
|   drugTypical              2.11299  1.244117 262  1.698387  0.0906
|   cvar(drug, id)Clozapine  8.58797  3.688862  50  2.328082  0.0240
|   cvar(drug, id)Typical   -1.29235  3.662873  50 -0.352825  0.7257
|   year                    -1.02063  0.249215 262 -4.095356  0.0001
```

Sketch the predicted response as a function of time (with years ranging from 1 to 6), for a patient who took the 'Typical' drug in years 1 and 2 and then switched to Clozapine for the remainder of the study. Label the graph clearly and show your work so it is clear that you know the numerical values that determine the lines you are plotting. *(15 points)*

3. Let Y and X be numerical variables and let Z represent a set of variables including a term that is a column of 1's. Consider the following models. All but one of these models must produce the same regression coefficient for X or Xr but they will produce different standard errors. Identify the model that may produce a different coefficient. Rank the others where you can according to the standard error of the estimated coefficient for X (or Xr) stating which would be equal if any (assume a very large n and ignore the effect of slight differences in degrees of freedom for the error term). Explain your reasoning briefly.

   (A) `Y ~ X + Z`

   (B) `Y ~ X`

   (C) `Yr ~ Xr` where `Yr` is the residual of `Y` regressed on `Z` and `Xr` is the same for `X`

   (D) `Y ~ Xr`

   (E) `Y ~ X + Xh` where `Xh` is the predictor of `X` in the regression of `X` on `Z`

   (F) `Y ~ X + Xh + Zg` where `Zg` consists of a subset of the columns in Z

   *(25 points)*

4. Explain clearly with a suitable sketch the relationship between the 'within effect', the 'between effect' and the 'contextual effect' when working with clustered data. *(20 points)*

5. Show the truth tables for logical OR and for logical AND. Note that R has 3 logical values. Write a brief but clear explanation of the logic behind the truth tables. *(10 points)*

6. Write an R function that takes a vector of character strings, factors or numeric values representing an amount of money that might be entered in a variety of formats for Canadian currency, e.g. "1,000.21", "$20", "C23.01", "22CDN", and returns a numeric value. *(10 points)*

7. Write an essay on variable selection strategies in statistical analysis. What major factors are relevant and how do they affect the choice of strategy. *(30 points)*

8. Consider the general linear model with the usual notation. Let $\eta_1 = L_1\beta$ and $\eta_2 = L_2\beta$. Suppose that within both $L_1$ and $L_2$ the rows are linearly independent and that the rows of $L_1$ can be expressed as linear combinations of the rows of $L_2$ and vice-versa. Show that the Wald tests for $\eta_1 = 0$ and for $\eta_2 = 0$ come to identical conclusions. *(10 points)*

9. What value will the following expressions return and why?
a) `"2" < 10`

b) `1 == TRUE`

c) `'1' == TRUE`

d) `f <- function(ab) {2*ab}; f(a = 5)`

*(10 points)*

10. Consider the following function in R:

```
f <- function(y = 2)  function(x = 3) x + y
```

State what the following expressions will return and explain briefly why:
a) `f()`

b) `f(20)`

c) `f()()`

d) `f(5)(20)`

e) `f(5,20)`

*(10 points)*

11. Consider a mixed model of the form `lme( Y ~ X, data, random = ~ 1 + X| id)` and let $G = \begin{pmatrix} g_{00} & g_{01} \\ g_{10} & g11 \end{pmatrix}$ be the variance matrix for the random effects. Prove that the value of $X$ at which the variance of predicted values is minimized is $-g_{01}/g_{11}$. *(15 points)*

**END**