VHM 812/802: Clustered Data Analysis Exercise

Data

These data come from Dr. Håkan Vigre's PhD research project. The original data file is called -ap2.dta- (on the VER web site), but you are going to use a simplified version of the data called -ap2_intro.dta- for this exercise.

Background:
Respiratory diseases are important causes of morbidity (sick pigs), mortality (dead pigs) and subclinical production losses (reduced daily weight gains) in intensive pig production units. The main production phases in pig production are:

Phase	Description
nursing	from birth to about 21 days of age when the piglets are removed from the sow
weaner period	from weaning to about 60 - 70 days of age
finishing period	from about 70 days of age to slaughter (at about 150 days of age)

In this assignment, we are going to focus on the first two periods combined and look at factors that influence growth rates of pigs from birth to time of transfer to the finishing barn. The data come from 416 litters (pigs born to one sow at a parturition are called a "litter") from 6 farms with up to 3 pigs per litter being bled at the time of transfer to the finishing barn and serologically tested for 4 important respiratory pathogens:

- ***Actinobacillus pleuropneumonia Type 2*** - this is a bacterial disease that causes inflammation of the pleura that lines the chest cavity and cover the lungs as well as inflammation within the lung tissue. There are several serotypes of Ap but we will just focus on the one (Type 2) which is most common (in Danish herds) and likely has the greatest effect.
- ***Mycoplasma hyopneumonia*** - this is a mycoplasmal agent which causes a disease called "Enzootic pneumonia" in pigs. It results in consolidation of the lung tissue, particularly in the antero-ventral lobe(s).
- ***Influenza*** - this is a highly contagious viral agent that causes respiratory distress. It can range from subclinical to an acute clinical condition with high fever that spreads rapidly through the barn.
- ***Porcine Respiratory and Reproductive Syndrome (PRRS)*** - this is a virus that causes both reproductive problems in sows and respiratory disease in growing pigs. The respiratory problems are characterized by interstitial pneumonia and can be highly variable in severity (due in part to strain variations in the virus).

All diseases except influenza were tested for using ELISA tests which give results in terms of "optical density", but these have already been converted for you into a simple 0/1 variable indicating whether the pig was negative (0) or positive (1). Influenza was tested for using a hemagglutination inhibition test with the results also given to you as negative (0) or positive (1).

Dataset:
The specific variables in the reduced dataset -ap2_intro- are shown in the table on the next page.

Variable	Description
farm_id	farm identification code (1-6)
litt_id	litter identification number
pig_id	pig identification number
dwg_wean	daily weight gain from birth to time of transfer to finishing barn (age_t)
parity	the parity of the sow (i.e. the dam of the pig)
ap2_t	serological reaction against A. pleuropneumoniae serotype 2 at age_t
mp_t	serological reaction against M. hyopneumoniae at age_t
infl_t	serological reaction against Influenza virus at age_t
prrs_t	serological reaction against PRRS virus at age_t

Questions

Use these data to answer the following questions. Note: In all of the following modelling questions, we will include the predictors: -parity-, -ap2_t-, -mp_t-, -infl_t-, and -prrs_t-, regardless of their statistical significance. We will also keep -parity- as a continuous variable without regard to the linearity of the assumption inherent in this. Finally, -farm_id- will be included in most models as a fixed effect (set of dummy variables).

1. Generate some descriptive statistics to get a feel for the data.
2. Build a multiple linear regression model with -dwg_wean- as the outcome but which ignores any clustering of pigs within a litter or farm.
3. Add -farm_id- as a fixed effect.
   (a) What effect does this have on the estimated coefficients and their statistical significance?
   (b) Is farm a significant predictor?
   (c) Did adding farm account for much of the unexplained variance from the previous model? (What proportion of the variance?)
   (d) How variable are average weight gains across farms?
4. Build a linear mixed model with -farm_id- as a fixed effect and -litt_id- as a random effect.
   (a) Does adding litter as a random effect change either the coefficients or the significance of the predictors in the model?
   (b) What proportion of variance resides at the litter level? How strongly correlated are weight gains of piglets from the same litter?
   (c) How variable are weight gains across litters?
5. Refit the linear mixed model but change farm from being a fixed effect to a random effect.
   (a) What are the arguments in favour of, and against, treating -farm_id- as a random effect?
   (b) What proportion of variance resides at each of the three levels?
   (c) What are the ICC's for: i) two pigs within a herd?, and ii) two pigs within a litter?