MATH 60604A
Statistical modelling
§ 6f - Prediction from linear mixed models

HEC Montréal
Department of Decision Sciences

Prediction

- The random effects b, are random variables and not parameters (parameters are fixed, but unknown quantities).
- We can always get predictions for these random variables.
- Once we have predicted values for the b terms and estimates for the fixed effect parameters, β, we can predict the outcome variables Y by its conditional mean.

Prediction: model **without** random effects

- If there are no random effects in the model (for example, if we had fitted a model that directly specified the covariance structure using `repeated`), then we make predictions in the same way as we did for ordinary linear regression.
- That is, the prediction for Y_{ij} is

$$\widehat{Y}_{ij} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{ij1} + \ldots + \widehat{\beta}_p X_{ijp}.$$

- This quantity is also the estimate of the mean (at the population level) of the response variable.

Prediction: model **with** random effect

- If there are random effects in the model, the estimation of the marginal mean (at the population level) of the response variable for an individual with the characteristics of individual j from group i is

$$\widehat{Y}_{ij} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{ij1} + \ldots + \widehat{\beta}_p X_{ijp}.$$

- But we can also get predictions of the values of the response variable for individual j in group i
- For example, in a model with a random intercept for group i, b_i,

$$\widehat{Y}_{ij} = \widehat{\beta}_0 + \widehat{b}_i + \widehat{\beta}_1 X_{ij1} + \ldots + \widehat{\beta}_p X_{ijp}.$$

- If, however, we want to get predictions for a new individual that was not included in the original dataset, then we have no choice but to use the mean prediction, because the random effect estimate of this group is not available.

Predictions for random effects

SAS code for the random intercept model

```
proc mixed data=statmod.motivation;
class idunit;
model motiv = sex yrserv agemanager nunit / solution;
random intercept / subject=idunit type=vc solution;
ods output Mixed.SolutionR=re;
run;
```

The option `solution` in the command `random` is used to get predictions of the random effects. The command `ods output` saves these in order to make diagnostic plots for the random effects.

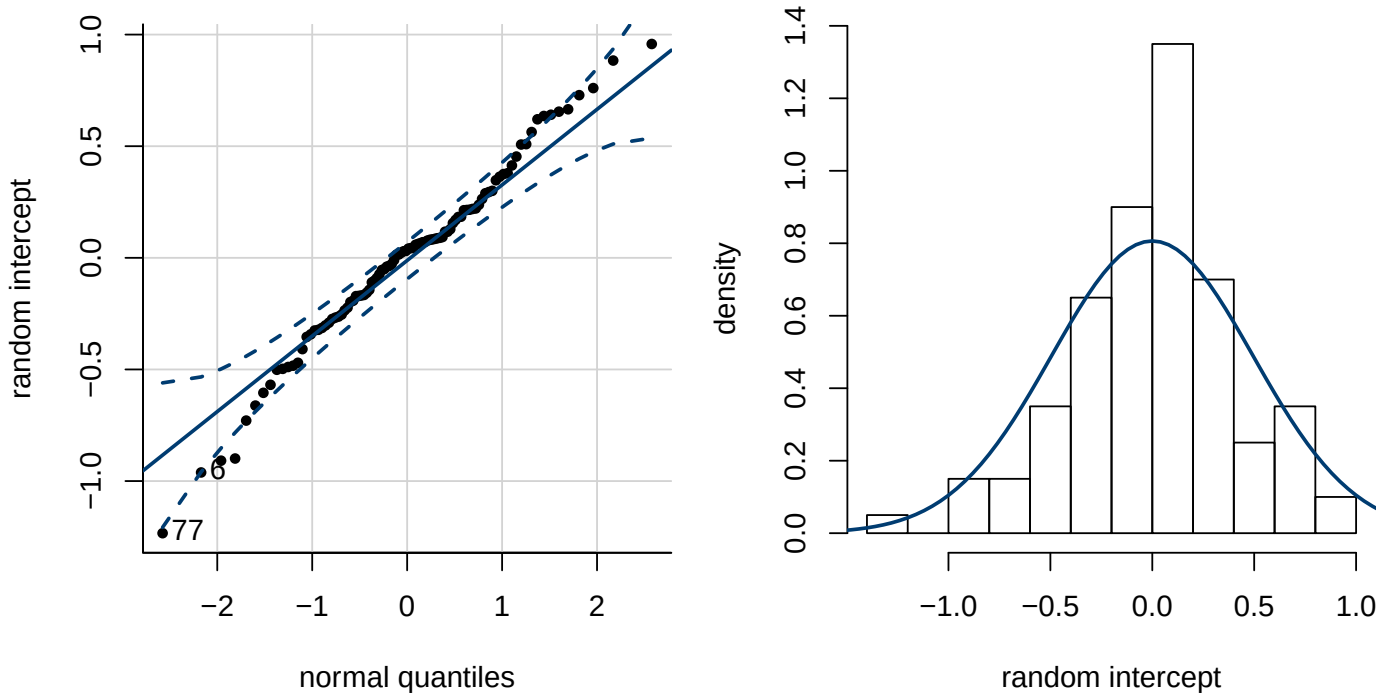
Predictions of the random effects

Solution for Random Effects

Effect	idunit	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	1	0.2143	0.2933	913	0.73	0.4651
Intercept	2	0.08777	0.3325	913	0.26	0.7919
Intercept	3	-0.4830	0.2731	913	-1.77	0.0774
Intercept	4	0.4537	0.2598	913	1.75	0.0811
Intercept	5	-0.3024	0.2667	913	-1.13	0.2572
⋮						
Intercept	96	-0.5014	0.2564	913	-1.96	0.0508
Intercept	97	-0.07346	0.2810	913	-0.26	0.7938
Intercept	98	-0.2631	0.3189	913	-0.82	0.4096
Intercept	99	0.5634	0.3567	913	1.58	0.1146
Intercept	100	0.7287	0.2801	913	2.60	0.0094

Histogram of random effects

We can plot histograms and quantile-quantile plots of the predicted random intercepts.



These can help us check the normality assumption of the random effects (think of these as further residual diagnostics). Note that (by construction), the average of these random effects is always zero.

Predictions for observations Y

- With `proc mixed`, we can save the values for all observations in the data file:
  - Predictions for the mean of the population (fixed effects),
  - Individual predictions (fixed and random effects).
- This is done using the options `outpm` and `outp`, respectively, in the `model` command.
- **Trick**: in SAS, if you want predictions for new individuals, you can just include these in the data file with a missing response (with "."). These individuals will not be used in the estimation of the model

Prediction for new employees

We assume that we want to get predictions for two new employees, one of whom is part of a unit already present in the dataset (`idunit=1`) and one that is part of a unit not in the original dataset (`idunit=101`).

SAS code to input two new observations

```
data newdata;
input nunit idunit idemployee yrserv sex
      motiv agemanager;
cards;
9 1 10 5 0 . 40
9 101 1 5 0 . 40;
run;

/* Merge observations with database */
data motivation;
set statmod.motivation newdata;
run;
```

Code to fit the model and get predictions

SAS code to output predictions from a mixed model

```
proc mixed data=motivation;
class idunit;
model motiv = sex yrserv agemanager nunit
      / solution outp=prediction outpm=mean;
random intercept / subject=idunit type=vc;
run;
```

- The data file used is `data=motivation`, which contains the 1018 observations, but only the 1016 observations from the original file are used in fitting the model.
- However, predictions will be made for all 1018 observations in the files `mean` and `prediction`.

Population mean of the two new subject

Output in file `mean`:

idunit	Pred	StdErrPred
1	12.2321	0.094962
101	12.2321	0.094962

- The fitted mean (12.23) is the same in both cases because only the fixed effects were used and the two employees have the same values for the explanatory variables.

Predictions for the two new subjects

Output in file `prediction`:

idunit	Pred	StdErrPred
1	12.4465	0.29287
101	12.2321	0.50376

- This time, the random effects are used if they're available. Since unit 1 was present in the model fitting, its random effect is used in making the prediction (12.45).
- However, the unit 101 was absent when fitting the model. Therefore, the prediction for the employee in unit 101 is only based on the fixed effects in the model, meaning that we get the same predicted value (12.23) as before.
- The standard errors for the individual predictions are larger, reflecting the added individual uncertainty arising from the errors and the random effects.