

MATH 60604
Modélisation statistique
§ 6f - Prédications pour modèles mixtes

HEC Montréal
Département de sciences de la décision

- Les effets aléatoires b sont des **variables aléatoires** et non pas des paramètres (c'est-à dire des quantités fixes, mais inconnues).
- On peut toutefois obtenir des **prédictions** pour ces effets aléatoires.
- Une fois qu'on a prédit les effets aléatoires b et estimé les paramètres des effets fixes β , on peut prédire la moyenne conditionnelle des Y .

- S'il n'y a pas d'effets aléatoires dans le modèle (par exemple si on a seulement utilisé une structure de covariance pour les erreurs), alors on est dans la même situation qu'en régression linéaire ordinaire.
- C'est-à-dire, la prédiction pour Y_{ij} est

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

- Cette quantité est aussi l'estimation des moyennes (au niveau de la population) de la variable réponse.

- S'il y a des effets aléatoires dans le modèle, l'estimation de la **moyenne marginale** (au niveau de la **population**) de la variable réponse pour un individu ayant les caractéristiques de l'individu j du groupe i est

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

- On peut aussi prédire des **valeurs** de la variable réponse de l'individu j du groupe i : par exemple, dans un modèle avec effet aléatoire sur l'ordonnée à l'origine b_i du groupe i ,

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

- Par contre, si on désire obtenir une prédiction pour une observation d'un **nouvel** individu/groupe dont les caractéristiques ne sont pas présentes dans l'échantillon d'apprentissage, alors on n'a pas le choix que d'utiliser la prédiction de la moyenne, car l'estimation de l'effet aléatoire n'est pas disponible pour ce nouveau groupe.

Code SAS pour l'ordonnée à l'origine aléatoire

```
proc mixed data=modstat.mobilisation;  
class idunite;  
model mobilisation = sexe anciennete agegest nunite / solution;  
random intercept / subject=idunite type=vc solution;  
ods output Mixed.SolutionR=re;  
run;
```

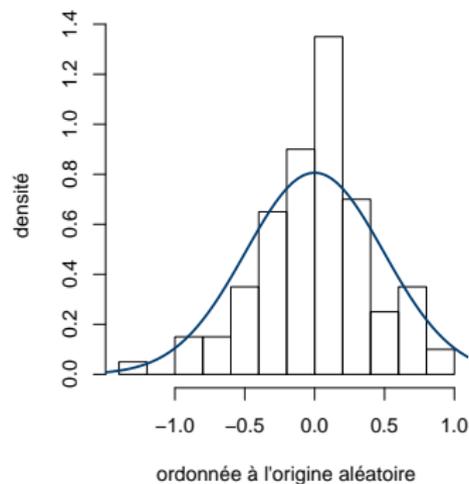
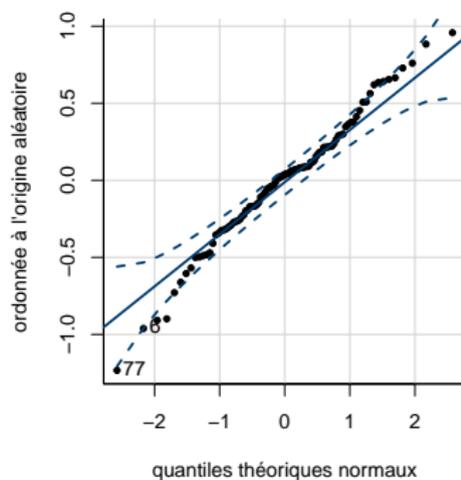
L'option `solution` à la commande `random` est utilisée pour avoir les prédictions des effets aléatoires dans la sortie. La commande `ods output` sauvegarde ces derniers afin qu'on puisse produire des diagnostics graphiques pour les effets aléatoires.

Prédiction des effets aléatoires

Solution pour effets aléatoires						
Effet	idunite	Estimation	Err type		Valeur du test t	Pr > t
			Préd	DDL		
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

Vérification des effets aléatoires

on peut tracer un histogramme et produire un diagramme quantile-quantile normal des ordonnées à l'origine prédites.



Ces diagnostics graphiques permettent de vérifier l'hypothèse de normalité des effets aléatoires (considérez-les comme des diagnostics des résidus supplémentaires). Il est important de noter que, par construction, la moyenne des effets aléatoires est toujours zéro.

- Avec `proc mixed`, on peut sauvegarder les valeurs **pour toutes les observations dans un jeu de données**:
 - prédiction de la moyenne de la population (effets fixes),
 - prédictions individuelles (effets fixes et aléatoires).
- Ceci se fait en ajoutant les options `outpm` et `outp`, respectivement, à la commande `model`.
- **Truc SAS**: Si vous désirez obtenir des prédictions pour des nouveaux individus, il suffit de les inclure dans la base de données en laissant des valeurs manquantes (".") pour SAS) pour la variable réponse. Ainsi, ces individus ne vont pas être utilisés pour ajuster le modèle, mais des valeurs prédites pour eux seront quand même produites.

On veut obtenir des prédictions pour deux nouveaux employés, l'un qui fait partie d'une unité connue (`idunite=1`) et un autre qui fait partie d'une unité absente des données originales (`idunite=101`).

Code SAS pour imputer deux nouvelles observations

```
data newdata;
input nunite idunite idemployee anciennete sexe
      mobilisation agegest;
cards;
9 1 10 5 0 . 40
9 101 1 5 0 . 40;
run;

/* Fusionner observations */
data mobilisation;
set modstat.mobilisation newdata;
run;
```

Code SAS pour obtenir des prédictions du modèle mixte

```
proc mixed data=mobilisation;  
class idunite;  
model mobilisation = sexe anciennete agegest nunite  
      / solution outp=prediction outpm=mean;  
random intercept / subject=idunite type=vc;  
run;
```

- Le jeu de données `mobilisation` contient les 1018 observations, mais seulement les 1016 observations originales seront utilisées pour ajuster le modèle.
- Par contre, les prédictions seront produites pour toutes les 1018 observations dans les fichiers `mean` et `prediction`.

Moyenne de la population pour les deux nouveaux individus

Sortie du fichier `mean`:

idunite	Pred	StdErrPred
1	12.2321	0.094962
101	12.2321	0.094962

- La moyenne de la population (12,23) est la même pour les deux individus parce que seuls les effets fixes sont pris en compte et les employés ont les même valeurs explicatives.

Prédictions pour deux nouveaux individus

Sortie du fichier `prediction`:

idunite	Pred	StdErrPred
1	12.4465	0.29287
101	12.2321	0.50376

- Cette fois ci, les effets aléatoires sont utilisés pour autant qu'ils soient disponibles. Puisque l'unité 1 était présente lors de l'ajustement, l'effet aléatoire du groupe est utilisé dans la prédiction (12,45).
- En revanche, l'unité 101 n'était pas disponible et donc la prédiction pour l'employé de l'unité 101 est basée seulement sur la moyenne marginale de la population (effets fixes) (12.23).
- Dans les deux cas, les erreurs-type des prédictions individuelles sont plus larges que celles de la moyenne, ce qui reflète l'incertitude additionnelle des erreurs et des effets aléatoires.