

Mixed Effects Models

Carolyn Taylor
Applied Statistics and Data Science Group (ASDa)
Department of Statistics, UBC

March 21, 2017

Resources for statistical assistance

Department of Statistics at UBC:

www.stat.ubc.ca/how-can-you-get-help-your-data

SOS Program - An hour of free consulting to UBC graduate students. Funded by the Provost and VP Research Office.

STAT 551 - Stat grad students taking this course offer free statistical advice. Fall semester every academic year.

Short Term Consulting Service - Advice from Stat grad students. Fee-for-service on small projects (less than 15 hours).

Hourly Projects - ASDa professional staff. Fee-for-service consulting.

Outline

Need for Mixed Effects Models

Recall for Fixed Effects Linear Models

Random Effects Models

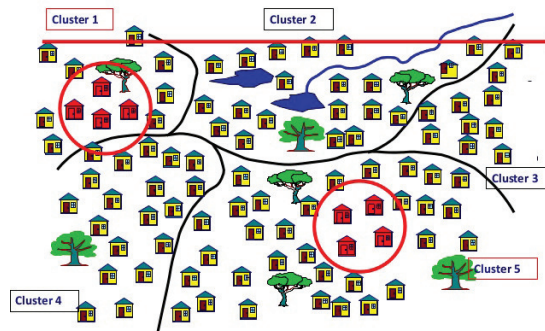
Mixed Effects Models

Why Mixed Models

Populations of measurements often fall naturally into groups

- ▶ patients from different hospitals participating in a medical study
- ▶ products from different production runs or batches
- ▶ repeated measures made on the same subject for a sample of different subjects
- ▶ plants or animals observed at different times
- ▶ soil nutrients and contamination at different depths from different sites

Cluster sampling



Sampling in groups may be an unavoidable part of the study but it also may ensure:

- ▶ a large enough sample can be collected within a reasonable time frame or budget
- ▶ results apply to a larger population of groups rather than to just those in one particular group

Why Mixed Models continued

Linear and generalized linear models make the critical assumption that the data to which they are fit consist of independent observations

It is typical that measurements from the same group or unit are more similar to one another than to observations made on different groups or units.

- ▶ different levels of a grouping variable may have naturally higher or lower response potential
- ▶ all measurements from the same cluster are exposed to that cluster's unique response potential
- ▶ measurements from the same cluster will tend to be more similar to one another than they are to measurements from different clusters that have different response potential

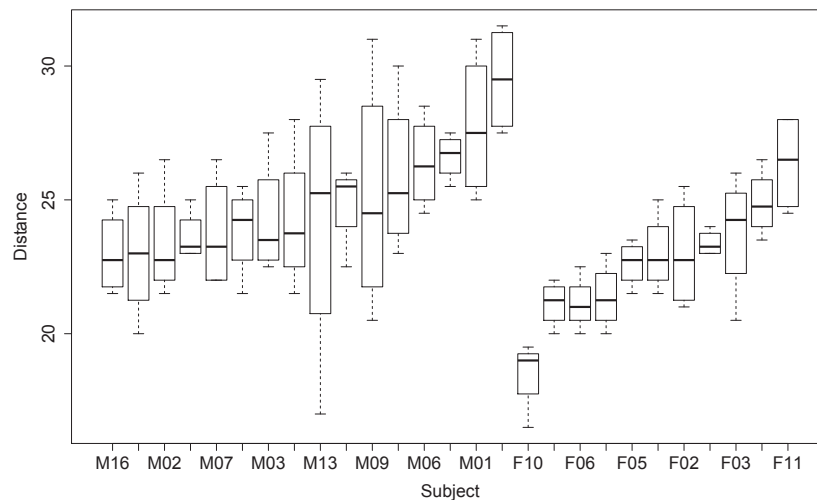
So data sampled in groups are typically not independent

Example: Orthodontic measurements over time

```
## Distance Age subject Sex
##      26.0  8    M01 Male
##      25.0 10    M01 Male
##      29.0 12    M01 Male
##      31.0 14    M01 Male
##      21.5  8    M02 Male
##      22.5 10    M02 Male
##      23.0 12    M02 Male
##      26.5 14    M02 Male
##      23.0  8    M03 Male
##      22.5 10    M03 Male
##      24.0 12    M03 Male
##      27.5 14    M03 Male
##      25.5  8    M04 Male
##      27.5 10    M04 Male
##      26.5 12    M04 Male
##      27.0 14    M04 Male
```

Inter-dependence of the within subject responses

Every subject has a slightly different distance, this affects all the responses from the same subject, rendering the within subject responses inter-dependent



Random effects

Consider all possible levels of a grouping variable (hospitals in a medical study, trees in a growth study)

- ▶ imagine that each level has a constant value associated with it that gets added to the overall mean, raising or lowering it by a unique amount
- ▶ this constant value is called the random effect of the level

Levels of a random effect used in a particular study are assumed to have been sampled from a population of possible levels whereas levels of a fixed effect are not

Assume they follow a normal distribution with mean zero

There can be more than one random effect in a given study. They can be nested or crossed.

Recall for Linear models

We assume $Y \sim N(\mu, \sigma^2)$ and the data are independent

μ and σ^2 are parameters that describe centre and spread of the distribution respectively

We can shift the normal distribution without changing its shape

This allows us to split the data into 2 parts, one fixed, the other random

- ▶ $Y_i = \mu + \varepsilon_i$
- ▶ μ is a fixed value and ε is the random error in our data

We assume $\varepsilon \sim N(0, \sigma^2)$

We can model both the fixed and random portions of our data

Fixed Effects Models

Regression, ANOVA, ANCOVA are all Fixed Effects Models

We model μ as a function of some predictors

Example: $\mu_{ij} = a + bx_{ij} + T_i$

x_{ij} is an observed numeric variable. a, b and T_i are fixed effects, T_i is associated with the levels of an unspecified categorical variable.

ε_{ij} is random described by a single parameter σ^2

Approximately 95% of the errors will be between -2σ and 2σ

We assume the errors are independent

Example: Does Age affect Distance?

One possible model for looking at this: $Distance_{ij} = \mu + Age_i + \varepsilon_{ij}$

Age - fixed effect, "systematic" part of the model

ε - "error term", "random" part of the model, represents the deviations from the predictions due to "random" factors that cannot be controlled, does not have any interesting structure

In Mixed Models:

- ▶ the "systematic" part of the model works just like with linear models
- ▶ complexity/structure is added to the "random" part of the model

Example continued

Another possible model: $Distance_{ijk} = \mu + Sex_i + Age_{ij} + \varepsilon_{ijk}$

- ▶ Age is treated as a categorical factor with four levels
- ▶ Sex is an additional fixed effect

For this data:

- ▶ distance from the pituitary to the pterygomaxillary fissure
- ▶ Age of the subject (8, 10, 12 and 14) (within subject)
- ▶ Subject id: 27 in total
- ▶ Sex, Male (n=16) or Female (n=11) (between subject)

BUT this Fixed Effects Model (and the previous one) assumes the “errors” are independent. This should not be assumed in this case and so both these models should not be considered initially.

Ignoring lack of independence

Consider the extreme situation with g groups each containing m observations where the m observations within each group are identical but different between the groups:

- ▶ a single measure from each group provides complete knowledge of that group
- ▶ multiple measurements in each group tell us nothing new and so are redundant
- ▶ effectively have a sample size of g meaningful observations
- ▶ sample is effectively much smaller than it appears ($n = g$ not $g \times m$)

If an analysis is done on all n observations assuming independence:

- ▶ variances for statistics will be too small
- ▶ resulting confidence intervals will be too narrow
- ▶ p-values will be smaller than they should be, inflating type I error rate

Random Effects Models

Random Effects Models allow us to model the variance of our data in a hierarchical way

Example: Suppose we want to measure aptitude of students using a standard test. We need a sample of students to take the test:

- ▶ Randomly select some schools (i)
- ▶ Randomly select a set of classes within each school (j)
- ▶ Randomly select some students from each class (k)

The overall average score is μ

The school score is $Y_i = \mu + \gamma_i$

The class score is $Y_{ij} = \mu + \gamma_i + \delta_{ij}$

The student score is $Y_{ijk} = \mu + \gamma_i + \delta_{ij} + \varepsilon_{ijk}$

Random Effects Models continued

Because we selected schools, classes and students randomly from a larger group we treat these effects as random. The assumptions for the random effects are

$\gamma \sim N(0, \sigma_s^2)$ (Between school variation).

$\delta \sim N(0, \sigma_c^2)$ (Between class but within school variation).

$\varepsilon \sim N(0, \sigma^2)$ (Between student but within class variation).

All random effects are independent

This model allows us to estimate how much variation in aptitude score between students is due to the school they attend and the class they are in

Mixed Effects Models

Mixed Effects Models combine both random and fixed effects into one model

Usually the fixed effects are the parameters of greatest interest in the model

The random effects serve to control for repeated measurements on the same sampled unit or within the same cluster

The fixed effects can exist at any level of the hierarchy and are tested at the appropriate level within the model

Mixed Effects Models do not require a balanced design which means the model can easily handle missing data. This is a huge advantage over traditional repeated measures ANOVA models.

Add fixed effects to our school example

Assume $\mu_{ijk} = \mu + A_i + B_{ij} + C_{ijk}$, where

- ▶ A is the province where the school is located
- ▶ B is the time of the class (Morning or Afternoon)
- ▶ C is the gender of the student

Province varies between schools, Time varies within school but between class, and Gender varies within class but between student

Our mixed effects model is

$$Y_{ijk} = \mu + A_i + B_{ij} + C_{ijk} + \gamma_i + \delta_{ij} + \varepsilon_{ijk}$$

Gender is tested at the student level, Time at the class level and Province at the school level

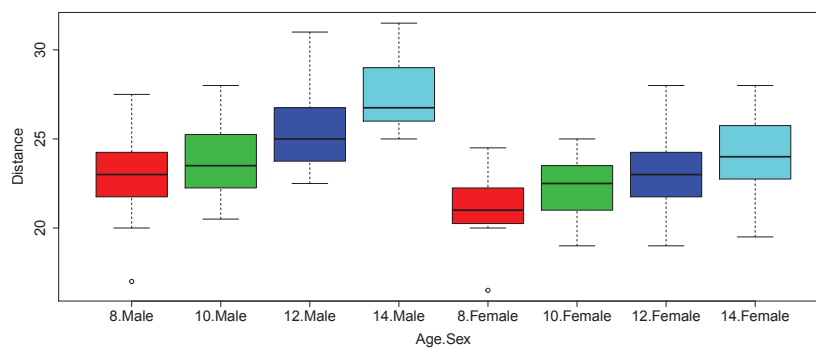
Example: Does Age affect Distance?

$$Distance_{ij} = \mu + Age_{ij} + subject_i + \varepsilon_{ij}$$

- ▶ gives some structure to the random "error" part of the model
- ▶ characterizes variation due to differences between subjects
- ▶ general error term ε still needed for the random noise that isn't due to differences between subjects

Mixed model - a model with a mix of fixed and random effects

Does Age affect Distance?



Lets fit distance as a function of Age and subject (ignoring gender for now)

- ▶ The first model uses fixed effects only, treating Age and Subject as fixed effects
- ▶ The second model treats subject as a random effect

Fixed Effects Model for within Subject factor

$$Distance_{ij} = \mu + Age_{ij} + Subject_i + \varepsilon_{ij}$$

```
## Anova Table, Response = distance
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Age       3  237.19   79.064  38.0396 2.986e-15
## Subject   26  518.38   19.938   9.5925 3.375e-15
## Residuals 78  162.12    2.078
```

```
##           Estimate Std. Error t value
## (Intercept)   22.185     0.277   79.96
## Age10         0.981     0.392    2.50
## Age12         2.463     0.392    6.28
## Age14         3.907     0.392    9.96
```

Mixed Effects Model for within subject factor

$$Distance_{ij} = \mu + Age_{ij} + subject_i + \varepsilon_{ij}$$

```
## Anova Table, Response = distance
##   NumDF DenDF F.value    Pr(>F)
## Age     3     78   38.04 2.998e-15
```

Random effects:

```
##           Var    SD
## subject   4.465  2.113
## Residual  2.078  1.442
```

- ▶ shows how much variability in Distance there is due to subject
- ▶ “Residual” is the variability not due to differences between subjects

Fixed effects:

##	Estimate	Std. Error	df	t value
## (Intercept)	22.185	0.492	43.4	45.07
## Age10	0.981	0.392	78.0	2.50
## Age12	2.463	0.392	78.0	6.28
## Age14	3.907	0.392	78.0	9.96

Age 8 is the reference level so Distance at Age 8 = 22.185

Distance at Age 10 = 22.185 + 0.981

The higher the age the longer the Distance

t value = Estimate / Std. Error

Comparing the results

The results from the 2 models are identical except for the error associated with the Intercept

Age is a within subject factor so the ANOVA test for Age and the estimated effects are the same as for the Mixed Model results. If the design was unbalanced the results would be similar but not the same.

In the Fixed Effects Model, we can compute the variance components using the expected means squares of the ANOVA table. This is non-trivial if the design is unbalanced.

In the Mixed Effects Model, we get the variance components directly

Models with only between subject factors

In order to control for repeated measures, we have to average the response over subject before we can use a Fixed Effects Model

If we do not average, we will overstate the significance of our between subject factor

$$Distance_{ij} = \mu + Sex_i + \varepsilon_{ij}$$

```
## Anova Table, Response = Distance
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Sex           1 140.46 140.465  19.157 2.831e-05
## Residuals 106  777.23   7.332

##           Estimate Std. Error t value
## (Intercept)    24.97     0.338   73.77
## SexFemale      -2.32     0.530   -4.38
```

Models with only between subject factors continued

```
## Anova Table, Response = average Distance
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Sex           1  35.116  35.116  9.2921 0.005375
## Residuals  25  94.479   3.779

##           Estimate Std. Error t value
## (Intercept)    24.97     0.486   51.38
## SexFemale      -2.32     0.761   -3.05
```

Mixed Effects Model approach

$$Distance_{ij} = \mu + Sex_i + subject_i + \varepsilon_{ij}$$

Mixed effects Model works without having to average and is accurate even if there is imbalance over the subjects

```
## Anova Table, Response = Distance
##      NumDF DenDF F.value  Pr(>F)
## Sex      1     25  9.2921 0.005375

##              Var    SD
## subject    2.547  1.596
## Residual   4.930  2.220

##              Estimate Std. Error df t value
## (Intercept)      24.97      0.486 25   51.38
## SexFemale         -2.32      0.761 25   -3.05
```

Both within and between subject factors in the model

For fitting a Fixed Effects only model, we cannot average over subject if there is a within subject factor in the model

If we do not average, the Fixed Effects Model won't realize that Sex is a between subject factor and so it won't be tested using the between subject variability (Subject). It will mistakenly use the within subject variability (Residuals) instead. This will overstate the significance of Sex.

$$Distance_{ij} = \mu + Sex_i + Age_{ij} + Subject_i + \varepsilon_{ij}$$

```
## Anova Table, Response = Distance
##      Df Sum Sq Mean Sq F value    Pr(>F)
## Age      3  237.19   79.064   38.040 2.986e-15
## Sex      1  140.46  140.465   67.581 3.513e-12
## Subject  25  377.91   15.117    7.273 6.062e-12
## Residuals 78  162.12    2.078
```

Using repeated measures ANOVA instead

This model works and is correct because we have complete data. If some subjects were not observed at all ages, this model could not be used.

```
##
## Error: Subject
##           Df Sum Sq Mean Sq F value Pr(>F)
## Sex         1  140.5   140.46   9.292 0.00538
## Residuals 25   377.9    15.12
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Age         3   237.2    79.06  38.04 2.99e-15
## Residuals 78   162.1     2.08
```

Mixed Effects Model approach

$$Distance_{ij} = \mu + Sex_i + Age_{ij} + subject_i + \varepsilon_{ij}$$

```
## Anova Table, Response = Distance
##      NumDF DenDF F.value Pr(>F)
## Age      3     78  38.040 2.998e-15
## Sex      1     25   9.292 0.005375

##           Var    SD
## subject  3.260 1.805
## Residual 2.078 1.442

##           Estimate Std. Error df t value
## (Intercept)  23.131    0.542 38  42.66
## Age10         0.981    0.392 78   2.50
## Age12         2.463    0.392 78   6.28
## Age14         3.907    0.392 78   9.96
## SexFemale    -2.321    0.761 25  -3.05
```

Within subject factor model if data is not balanced

Fixed Effects Model

```
## Anova Table, Response = Distance
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Age           3 195.04   65.014  27.9560 1.949e-11
## Subject       26 446.74   17.182   7.3884 9.542e-11
## Residuals     60 139.54    2.326
```

Mixed Effects Model

```
## Anova Table, Response = Distance
##      NumDF DenDF F.value    Pr(>F)
## Age       3 61.066  29.532 6.369e-12
```

Between subject factor model if data is not balanced

Fixed Effects Model on average observation

```
## Anova Table, Response = average Distance
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Sex           1 40.479   40.479  10.388 0.003512
## Residuals    25 97.419    3.897
```

Mixed Effects Model on raw data

```
## Anova Table, Response = Distance
##      NumDF DenDF F.value    Pr(>F)
## Sex       1 24.818   9.3844 0.005209
```

Mixed effects takes variation in sample size into account. Result is more accurate.

Interpretation of results is different

Together in a Mixed Effects Model

```
## Anova Table, Response = Distance
##      NumDF  DenDF F.value   Pr(>F)
## Age      3  61.892  29.740 4.972e-12
## Sex      1  24.928  10.285 0.003661

##              Var    SD
## subject    3.240  1.800
## Residual   2.322  1.524

##              Estimate Std. Error  df t value
## (Intercept)    23.16      0.568 40.6  40.75
## Age10           1.16      0.478 62.6   2.42
## Age12           2.53      0.471 61.3   5.37
## Age14           3.92      0.437 61.1   8.98
## SexFemale      -2.50      0.779 24.9  -3.21
```

Significance Testing

Unlike for Fixed Effects Models, Mixed Model parameters do not have nice asymptotic distributions to test against (no longer chi squared or F)

- ▶ inference on the parameters of a Mixed Effects Model rely on approximate distributions
- ▶ treat p-values with caution

Profiled confidence interval

- ▶ an interval which does not contain zero indicates the parameter is significant

Most reliable inferences for Mixed Models are done with Markov Chain Monte Carlo (MCMC) and parametric bootstrap tests

- ▶ both are computationally expensive and require longer run times
- ▶ parametric bootstrap is more intuitive and easier to generally apply

More on Mixed Effects Models

Suppose we have $Y_{ij} = \mu + \gamma_i + \delta_{ij}$ being the j th observation on the i th subject

$$\gamma \sim N(0, \sigma_B^2) \text{ and } \delta \sim N(0, \sigma_W^2)$$

We can compute the correlation between observations on the same and different subjects

$$\text{Cor}(Y_{1,1}, Y_{2,1}) = \text{Cor}((\gamma_1 + \delta_{11}), (\gamma_2 + \delta_{21})) = 0$$

$$\text{Cor}(Y_{1,1}, Y_{1,2}) = \text{Cor}((\gamma_1 + \delta_{11}), (\gamma_1 + \delta_{12})) = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

Observations on the same subject are correlated. This correlation is taken into account in Mixed Effects Models.

More than 1 Random Effect (Hierarchical or Crossed models)

We've demonstrated that we can fit 2 levels of data in the same model using a Mixed Effects Model where each level is defined by a random effect

A hierarchical model requires the random effects be nested to create a hierarchy

$$Y_{ijk} = \mu + \gamma_i + \delta_{ij} + \varepsilon_{ijk}$$

However, random effects can also be crossed

$$Y_{ijk} = \mu + \gamma_i + \delta_j + \varepsilon_{ijk}$$

Example: Student evaluation of Instructors

Students rate their lectures between 0 and 100. Each lecture is rated by multiple students and each student can rate multiple lectures.

The dataset has 21446 ratings from 2487 students. The lectures were taught by 318 instructors from 3 different departments. The goal is to test for a difference between departments and course type (service or not).

```
## Student Instructor dept service Score
##      1      1050     6    TRUE     35
##      3       494     9    TRUE     73
##      3       696     9    TRUE     29
##      7       588     4    TRUE     82
##      7      1843     4    TRUE     96
##      8        53     9    TRUE     62
##      9       281     6    TRUE     81
##      9      1770     6    TRUE      7
##     12       873     9    TRUE     47
```

Example continued

We need to account for repeated observations on the instructor and repeated observations by the student

```
## ANOVA Table, Response = Score
##              NumDF  DenDF F.value  Pr(>F)
## service              1 4293.1  39.305 3.983e-10
## dept                  2  350.1   4.839 0.008452
## service:dept          2 3976.8   6.897 0.001023

##              Est Std.Err   df t.val Pr(>|t|)
## (Intercept)  54.7     1.2 501.3 46.2  0.000
## serviceTRUE  -4.5     0.9 2331.9 -5.0  0.000
## dept4         2.4     1.6 521.4  1.5  0.137
## dept9         1.2     1.9 421.0  0.6  0.526
## serviceTRUE:dept4 3.7     1.3 3070.6  2.9  0.004
## serviceTRUE:dept9 -1.4    1.5 4471.9 -0.9  0.344
```

The fixed effects

The estimated group means are below:

```
## service dept Estimate SE DF
## FALSE 6 54.7 1.19 501
## TRUE 6 50.3 1.17 453
## FALSE 4 57.1 1.07 560
## TRUE 4 56.3 1.14 700
## FALSE 9 55.9 1.45 376
## TRUE 9 50.0 1.55 465
```

The fixed effects show that the lecture ratings are similar for Dept 6 and 9 but are higher for Dept 4. In general service courses receive a lower rating within the 3 departments.

The random effects and correlations

The variance components of the model are below:

```
## Var SD
## Student 49.72 7.051
## Instructor 98.35 9.917
## Residual 584.85 24.184
```

Observations from difference students on different instructors are uncorrelated

Same student on different instructors: $\rho = \frac{49.72}{584.85+49.72} = 0.08$

Different students on same instructor: $\rho = \frac{98.35}{584.85+98.35} = 0.14$

Each instructor/student combination appeared only once in the dataset so there is no same student same instructor correlation estimate

Testing Random Effects

Is the variance explained by the random effect significant?

Test the random effects using a likelihood ratio test

Refit the the model without a random effect and evaluate the change in the likelihood

```
## Likelihood ratio test for Student random effect
##   Chisq Chi Df Pr(>Chisq)
##  531.75     1 < 2.2e-16
```

```
## Likelihood ratio test for Instructor random effect
##   Chisq Chi Df Pr(>Chisq)
##  2508.5     1 < 2.2e-16
```

Test is approximate and overestimates the p-value (conservative)

Conclusions won't change for a more accurate test if the p-value is small enough to be significant

Mixed Effects model for non-normal responses

For distributions other than the normal model, we cannot separate the mean from the variance. The two are related.

Model parameters are related to some function of the mean (link function)

Dist	Link	Variance	Dist	Link	Variance
Poisson	$\log(\mu)$	μ	NB	$\log(\mu)$	$\mu + \mu^2/r$
Binomial	$\text{logit}(\mu)$	$\mu(1 - \mu)$	NB	$\log(\mu)$	$\tau^2\mu$

In this case we put the random effects within the link function

$$g(\mu_{ij}) = \alpha + \beta x_{ij} + \gamma_i \text{ where } \gamma_i \sim N(0, \sigma^2)$$

Example: Contagious bovine pleuropneumonia (CBPP)

Blood samples collected quarterly to look at changes in CBPP over time

We have repeated measures on herd of zebu cattle in Ethiopia. There are 15 herds with 3 or 4 repeated measures. The response is the number of cattle with CBPP.

```
## herd incidence size period
## 1 2 14 1
## 1 3 12 2
## 1 4 9 3
## 1 0 5 4
## 2 3 22 1
## 2 1 18 2
## 2 1 21 3
## 3 8 22 1
## 3 2 16 2
## 3 0 16 3
## 3 2 20 4
```

Example continued

$\text{logit}(\mu_{ij}) = \alpha + \text{Period}_{ij} + \text{herd}_i$ where $\text{herd}_i \sim N(0, \sigma^2)$

Test the hypothesis that the odds for CBPP are the same in all four periods

```
## Likelihood Ratio Test
##      Df  LRT  Pr(Chi)
## period 3 25.61 1.151e-05
```

```
##      Var  SD
## herd 0.4123 0.6421
```

Example continued

The odds of incidence for periods 2 to 4 are smaller than for period 1

The odds of CBPP status:

$$\text{period 1} \quad e^{-1.398} = 0.25, \frac{1}{0.25} = 4$$

$$\text{period 2} \quad e^{-1.398 + -0.992} = 0.09, \frac{1}{0.09} = 11$$

The odds ratio of CBPP status:

$$\text{period 2 to 1} \quad e^{-0.992} = 0.37, \frac{1}{0.37} = 2.7$$

$$\text{period 3 to 1} \quad e^{-1.128} = 0.32, \frac{1}{0.32} = 3.1$$

##	Estimate	Std. Error	z value	Pr(> z)
## (Intercept)	-1.398	0.231	-6.05	1.47e-09
## period2	-0.992	0.303	-3.27	1.07e-03
## period3	-1.128	0.323	-3.49	4.74e-04
## period4	-1.580	0.422	-3.74	1.82e-04

Resources for statistical assistance

Department of Statistics at UBC:

www.stat.ubc.ca/how-can-you-get-help-your-data

SOS Program - An hour of free consulting to UBC graduate students. Funded by the Provost and VP Research Office.

STAT 551 - Stat grad students taking this course offer free statistical advice. Fall semester every academic year.

Short Term Consulting Service - Advice from Stat grad students. Fee-for-service on small projects (less than 15 hours).

Hourly Projects - ASDa professional staff. Fee-for-service consulting.