

# Peer-Feedback Survey

1 survey every Tuesday (constant link)



[https://ubc.ca1.qualtrics.com/jfe/form/SV\\_bvLa6xMXdRk1j6u](https://ubc.ca1.qualtrics.com/jfe/form/SV_bvLa6xMXdRk1j6u)

1. Go to jamboard link on Canvas announcement or below
2. For each of the descriptions below, move it under either fixed effects or random effects.
3. Yes you can, re-locate a sticky note that someone else placed if you think it's incorrect

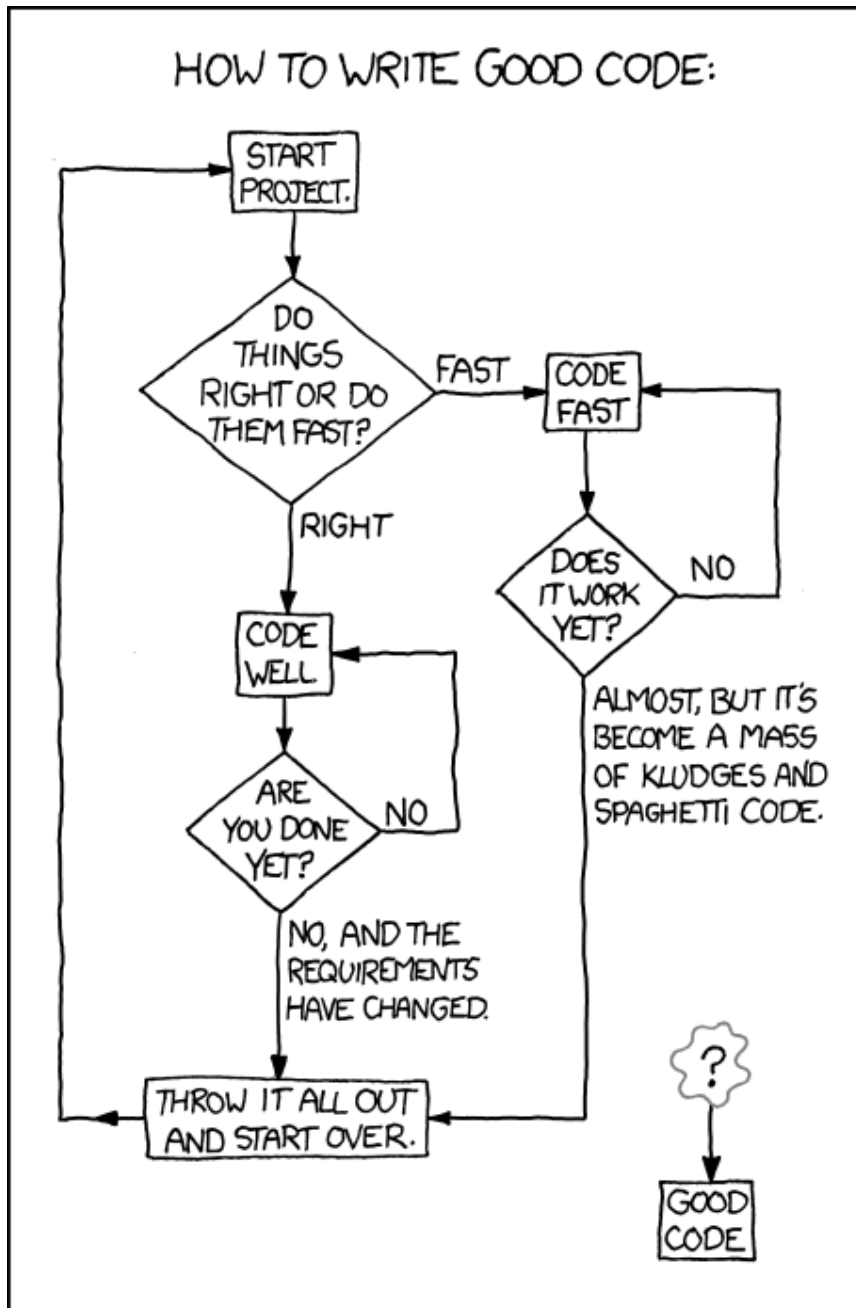
<https://jamboard.google.com/d/11BUhXK8GfFN1swCycxWd6SlvusWZCQnkYmt-C3NdBpQ/edit?usp=sharing>

# Jamboard

- **Slide 1:** Sort into fixed or random effects
- **Slide 2:** Do you expect to use LME models in your thesis?

<https://jamboard.google.com/d/11BUhXK8GfFN1swCycxWd6SlvusWZCQnkYmt-C3NdBpQ/edit?usp=sharing>

# BIOL 501: Mixed-Effects Models



LME

Linear mixed-effects models

Please get me out! It helps us all learn names

**FIRST NAME**  
Preferred pronouns

# Outline for today

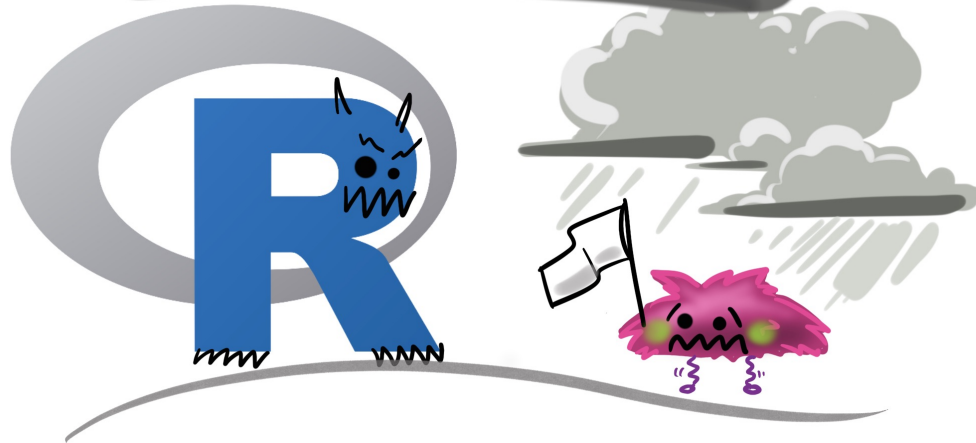
## Linear mixed-effects models (LME)

- Random vs. fixed effects
  - Why random effects are different and helpful
  - Unbalanced design
- How to tell if your study has random effects?
- LME model fitting
- Comparing a reduced (null) vs full model
- Assumptions of LME
- Confidence intervals

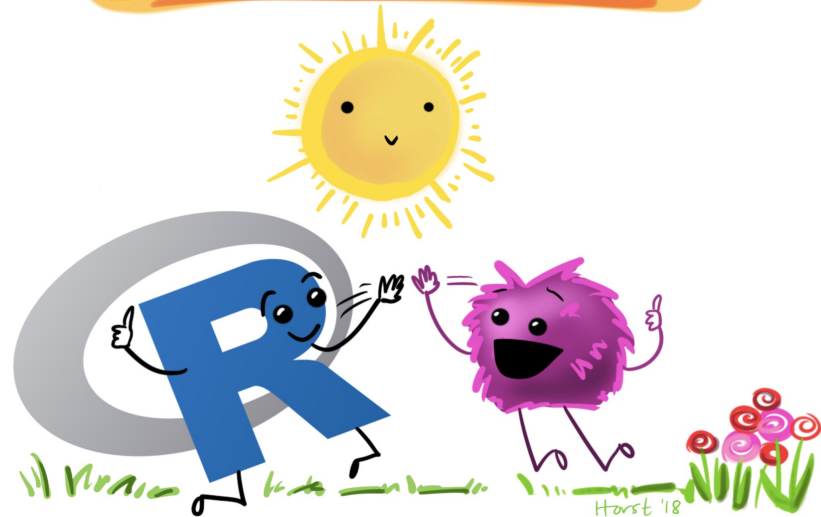
# If you only remember 4 things...

1. Mixed-effects models are very versatile!
2. How to identify random vs fixed effects
3. How to write simple lme or lmer fitted model
4. How to compare a full vs reduced (null) model

at first I was like...




...but now it's like...



# Linear Mixed-effects (LME) Models

- Linear models that include both fixed and random effects
- Model different error variance for each level of random variation.
- Can handle unequal sample size.
- Can handle repeated measures among and within groups
- *P*-values for tests of fixed effects are conservative when designs are unbalanced.
- Implemented with either
  - Lmer → lme4 and lmerTest packages
  - Lme → nlme package



It's called mixed because it mixes both fixed and random effects

# Random vs. Fixed Effects

# What are fixed effects

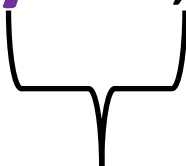
- **Predetermined categories of a variable of direct interest that are repeatable**
- Associated with entire population
- Associated with treatments controlled by the experimenter
- Conclusions reached in the study about differences among groups can be applied **only to the groups included in the study**



# Fixed Effects

- No random effects
- No repeated measures on same subject/animal/plant
- Residuals are the **only** source of random variation
- `lm()`, `glm()`, `gam()` models are only fixed effects

```
model1 <- lm(y ~ x, data=mydata)
```



**Fixed** part of formula  $y \sim x$

# Random Effects

- Randomly sampled categories of a variable, representing groups or clusters of measurements or units.
- Random effects occur when **repeated measurements** made on **randomly** sampled groups (e.g. on the same animal or plant)

```
model<-lme(y~x,random=~x|animal, data=mydata)
```

**Fixed**

**Random** part of the formula:  
describes the random effects and  
grouping structure

# Simple example of random effects

- The simplest example of a model with random effects is an observational study in which multiple measurements are taken on a random sample of individuals (or units, plots, etc) from a population. In this case “individual” is the random factor or group.
- Random group=animal/plant/ “individual”

# Comparison Code Syntax

## Fixed effects only vs. mixed-effects

Both are only testing linear relationships (not other equations)

```
model1 <- lm(y ~ x, data=mydata)
```

Use lm() for fixed effects only

**Fixed** part of formula  $y \sim x$

```
model <- lme(y ~ x, random=~x | animal, data=mydata)
```

Lme() or lmer() if  
mixed-effects

**Random** part of the formula:  
describes the random effects  
and grouping structure

# Random Effects

- Measurements from the same group/cluster/animal/plant are not independent random samples, but the animals are assumed to be randomly sampled from the “population” of groups
- Model different error variance for each level of random variation.

# Why add random effects?

- **Gives you greater inference to larger population.** Groups (e.g. 3 sea lions) assumed to be randomly sampled from population of groups (i.e. from entire population of sea lions). Therefore conclusions reached about groups **can be generalized** to this larger population when animal is a random effect.
- **Avoids pseudoreplication.** Allows for repeated measures among and within group (e.g. animal)

**If repeated measurements on animals or use animals multiple times → then yes you have repeated measurements**

# Another way to think about the difference between fixed vs random

- Fixed effects:
  - Yes, an experiment with the same treatment levels **could be repeated**
- Random effects:
  - Random effects could **not** be repeated exactly the same again

# Random Effects

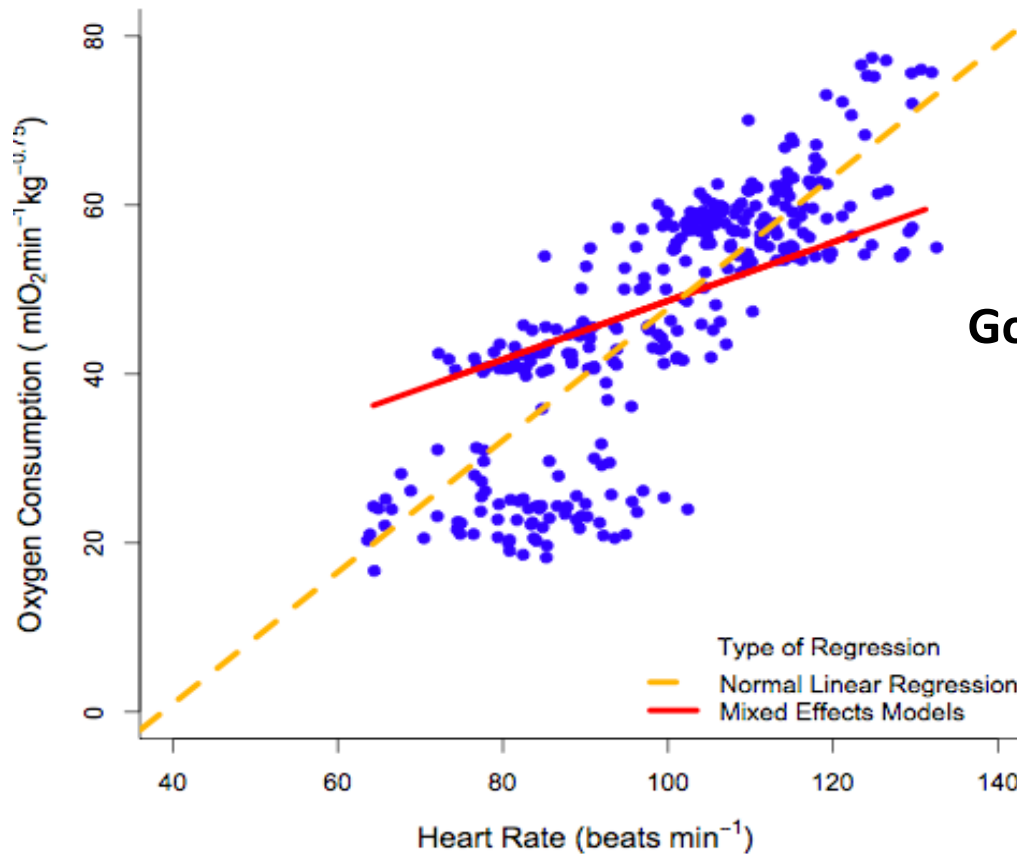
- **Use 1:** Sometimes random effects are noise or a nuisance—of no interest themselves
  - Field plots each assigned multiple treatments
- **Use 2:** Sometimes measuring the variance associated with each each level of random groupings is a purpose of the study
  - Subjects measured repeatedly (to specifically estimate repeatability)
  - Families of siblings (to estimate heritability etc)



# Analysis is different with random effects

- Unlike fixed groups, the means of the specific random groups are not of direct interest.
  - **Don't care (and don't plot) what each individual animal is doing**
  - Don't colour code your LME model by animal or grouping variable of individuals because that's not what the LME model is doing
- **Interest is focused on the variance among the random groups.**

LME models look different than `lm()` because our eyes can't easily see the random effects



**Goal is to infer about population NOT individuals**

- Treat animal as random sample of population
- Different than pooling of normal linear regressions
- Model animal differences in variance, errors etc.
- > power of with small n

## Sometimes reviewers don't understand LMEs

- Reviewer 3: “Did you find relationships for individual animals?”
- Response:
- Although this is an interesting question, we did not present the  $VO_2:ODBA$  relationships for individual animals. Animal was a random factor in the LME models to account for repeated measures and allow inferences from the sample population to the wild population. Plotting  $ODBA:VO_2$  for each animal individually violates one of the primary strengths of LME models.

Sometimes reviewers don't understand that the goal of LMEs is not to look at individual animals

- Reviewer 3: “Did you find relationships for individual animals?”
- Response:
- Although this is an interesting question, we did not present the VO<sub>2</sub>:ODBA relationships for individual animals. **Animal was a random factor** in the LME models to **account for repeated measures and allow inferences from the sample population to the wild population.** **Plotting ODBA: VO<sub>2</sub> for each animal individually violates one of the primary strengths of LME models.**
- Graphs depicting the statistical relationships of LME models are more difficult to visually interpret due to the random effects and repeated measures, particularly when readers are used to interpreting non-repeated measures linear regressions with only fixed effects. We have described LME models and provided several supporting references in the methods so that readers are able to read more detail on LME models if preferred

# Unbalanced designs with random effects

# Unbalanced designs

- Yes, LMEs can **handle unequal and unbalanced sample sizes**.
- When a design including random effects is unbalanced, the standard formulas for  $F$  statistics are not  $F$ -distributed. Standard ANOVA table calculations don't work with unequal sample sizes.
- With unbalanced designs, mixed models calculate approximate  $F$ -statistics and degrees of freedom for tests of fixed effects.
- $P$ -values for tests of fixed effects are conservative when designs are unbalanced. Results are approximate.

How to tell if I have random effects?

# How to know when you have random effects in your study

- Sampling design is nested: quadrats within transects; transects within woodlots; woodlots within districts.
- Measurements are clustered spatially or temporally within randomly sampled units i.e., in *blocks*, which are typically analyzed as random effects.
- Divide up plots, or families, clones, ponds, etc, and apply separate fixed treatments to subplots, or siblings, pond-halves, etc.
- Take measurements on related individuals.
- Measure subjects or other sampling units **repeatedly**.

In R, `lm()` is for fixed effects only. Do not use `lm()` if you have random effects. Watch out for pseudoreplication



Fitting a lme model

# lme (nlme package) vs lmer (lme4 package)

- Most of these examples today are lme and lmer examples are on the R tips page “Fit models to data”
- You can use either for learning purposes and BIOL 501
- Example below is null (reduced) model with numerical X & Y variables

```
null.model<-lme(y~x,random=~x|animal, data=mydata)
```

```
null.model<-lmer(y~x+(x|animal),data=mydata)
```

# More details on LME vs LMER

- Both packages model the variance structure of random effects explicitly.
- I prefer lme, but others prefer lmer. Either is ok.
- **Lmer:** Tests of treatment effects using the Kenward-Roger or Satterthwaite approximations for degrees of freedom are most accurate (available in lmerTest package).
- Differ in syntax and behind the scenes calculations but **majority of basic LME theory is the same**
- See R Tips Page for more on lmer

# Unpacking the **lme** model code into fixed vs random **library(nlme)**

**Fixed** part of formula  $y \sim x$

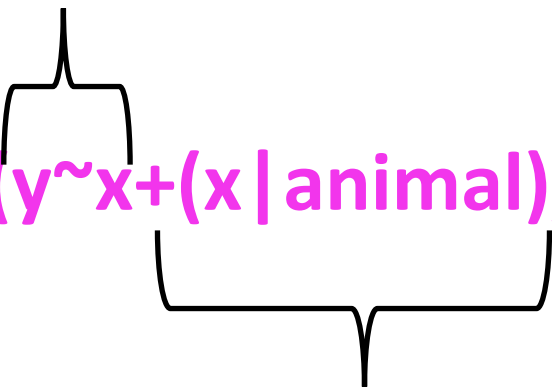
```
null.model<-lme(y~x,random=~x|animal, data=mydata)
```

**Random** part of the formula:  
describes the random effects  
and grouping structure

Unpacking the **lmer** model code into fixed vs random  
**library(lmerTest) within the lme4 package**

**Fixed** part of formula  $y \sim x$

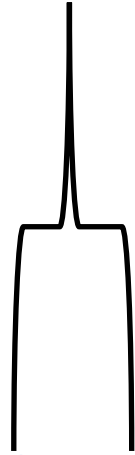
```
null.model<-lmer(y~x+(x|animal),data=mydata)
```



**Random** part of the formula:  
describes the random effects  
and grouping structure

## library(nlme)

**Fixed part** instructs R to fit a constant (intercept) based on the fitted values of the random groups (i.e. animal)



```
null.model<-lme(y~x,random=~x | animal, data=mydata)
```

# library(nlme)

Random part instructs R to fit a constant (intercept) to the multiple measurements within each animal. This yields a fitted value for each individual animal



```
null.model<-lme(y~x,random=~x | animal, data=mydata)
```

# Setting the likelihood: ML vs REML

- REML: restricted maximum likelihood is the *default*
  - `lme(y~x, method="REML")`
  - `Lmer(REML=TRUE)`
- ML: maximum likelihood
  - `lme(y~x, method="REML")`
  - `Lmer(REML=TRUE)`

**But**

- LME models with different fixed effects (full vs reduced) do not have comparable likelihoods when fitting using REML, **so must compare models with ML only for anova(null, model1) to be valid.**
- Once "best" model is found, re-run it with REML to obtain estimates of random effects

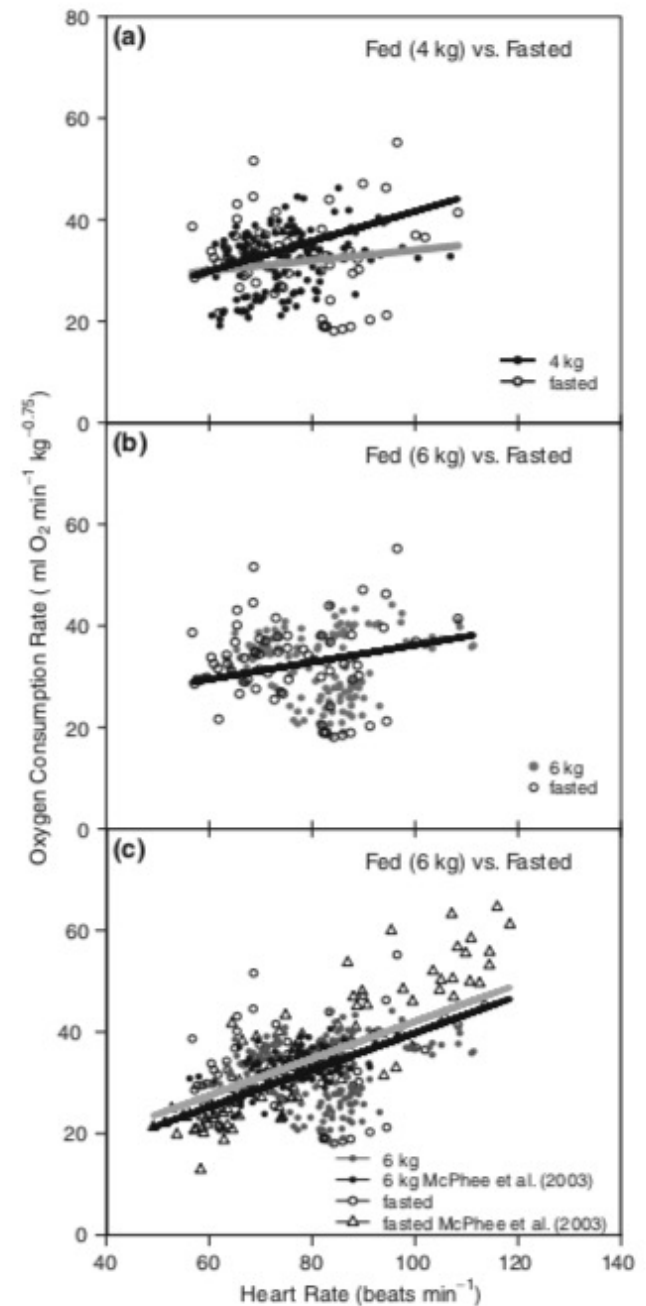


LME with numerical vs  
categorical X variable

# Can X predict Y ?

## Y=number, X=number

- Question: Can heart rate **predict** metabolic rate with a **linear relationship**? If so, is there a difference in this linear relationship among treatments?
- Random factor: animal
- N=7 Steller sea lions
- Repeated measurements among & within animals
- N=7 Steller sea lions
- Treatment: Fixed factor with 3 levels



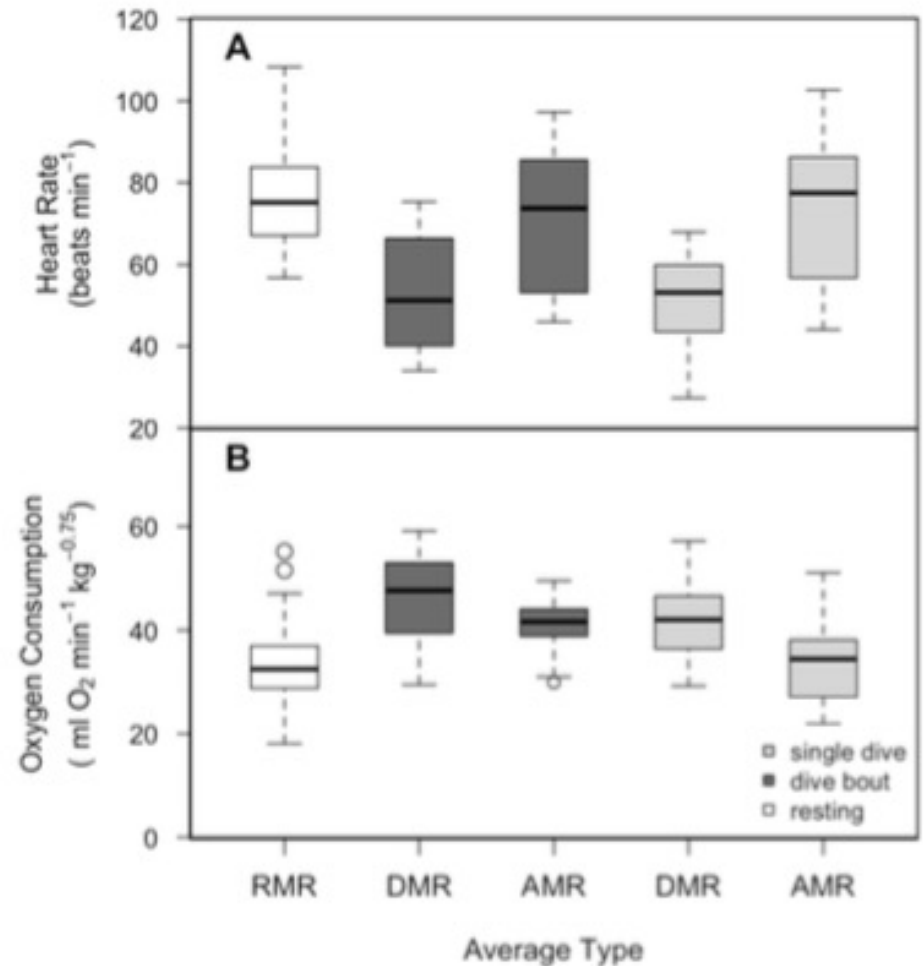
Yes predictive,  $Y=mx+b$

# Is there a difference among groups?

## Y=number, X=**category**

- Question: **Is there a difference** in metabolic rate among **treatments**?

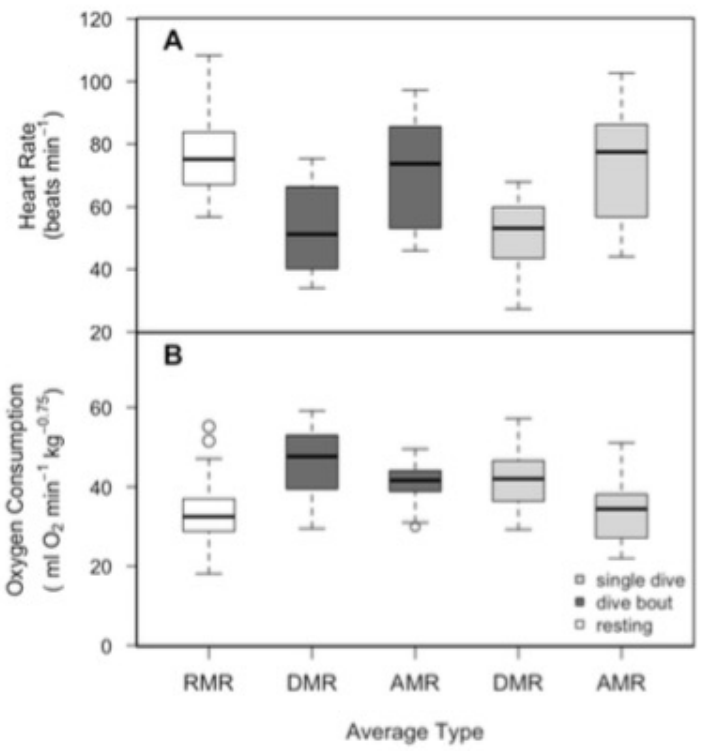
- Random factor: animal
- N=7 Steller sea lions
- Repeated measurements among & within animals
- Fixed factor of "Average Type" with 5 levels
- Fixed factor of "Dive type" with 3 levels



**Not predictive, no equation generated**

Is there a difference among groups?

Y=number, X=categorical

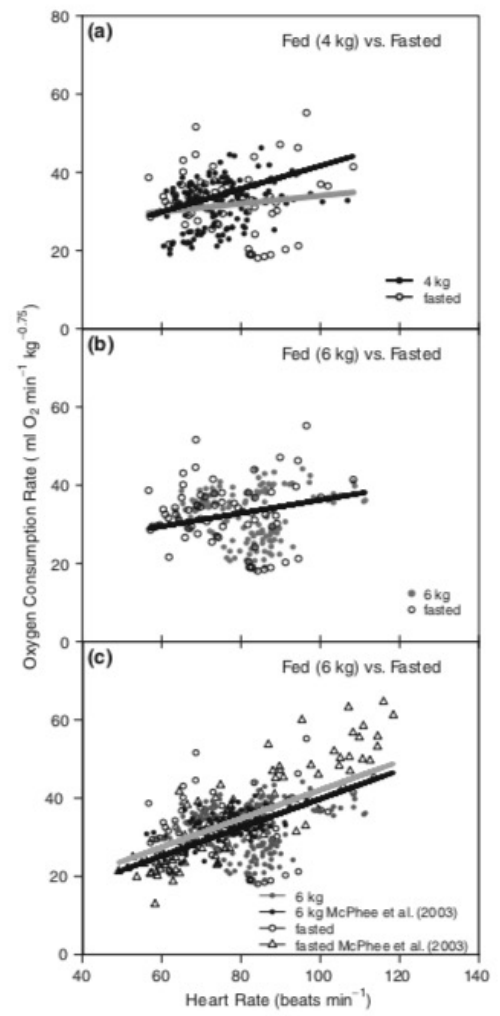


`lme(y~1,random=~1|animal, data=mydata)`

Young MSc Thesis Supplemental to Ch. 3, not published Fig. A3.2.

Can X predict Y ?

Y=number, X=number



`null.model<-lme(y~x,random=~x|animal, data=mydata)`

# lme (nlme package) vs lmer (lme4 package)

- Example below is null model with **numerical Y, categorical X** (only an intercept for a fixed effect)

```
null.model<-lme(y~1,random=~1|animal, data=mydata)
```

```
null.model<-lmer(y~1+(1|animal),data=mydata)
```

Comparing full vs  
reduced model

# Model comparison between a full and reduced lme model

- Anova() on 2 LME models compares the 2 LME model fits with Log likelihood ratio test (LRT)
- **\*\*must be comparing a reduced (null) vs. full model (aka hierarchically nested )otherwise test is invalid\***
- The full model contains the term of interest and the reduced model leaves it out.
  - Reduced and full model **only differ by this 1 thing**

# Anova(null,model1)

- **anova(model1)** An ANOVA executed on a single model generated a conditional F- test to determine the significance of model slope, intercept, and fixed effects.
- **anova(null,model1)** An ANOVA performed on two nested models (the fixed effect model hierarchically nested within the model without fixed effects) produced likelihood ratio tests (LRT) that compared the two models



# Lme model (X,Y, numerical)

1. Fit null (reduced model)
  - `null.model<-lme(y~x,random=~x|animal, data=mydata)`
  - Assess model fit- **“Is it linear”?**
    - `anova(null)`
2. Fit full (+factor1 model) ***that only varies by adding this 1 fixed factor***
  - `model1<-lme(y~x+fixed factor1,random=~x|animal, data=mydata)`
  - Assess model fit- **“Is it linear”?**
    - `anova(model1)`
3. Compare hierarchically nested models with LRT test and choose lowest AIC if p is sig
  - `Anova(null,model1)` – **“Which model is better?”** or **“Does adding this fixed factor improve the model better than the null model?”**

# Lme model (X,Y, numerical)

1. Fit full (+factor1+factor2 model) **that only varies by adding this 1 fixed factor**
  - `model2<-lme(y~x+fixed factor1+fixed factor2,random=~x|animal, data=mydata)`
  - Assess model fit- **“Is it linear”?**
    - `anova(model2)`
2. Compare hierarchically nested models with LRT test and choose lowest AIC if p is sig
  - `Anova(model1,model2)` – **“Which model is better?”** or **“Does adding this fixed factor improve the model better than the null model?”**

Could you compare the null model to model 2?  
Why or why not?

## #1.scatter plot (examine data)

```
plot(y ~ x, data = mydata)
```

## #2. Fit LME null model

```
# Extract coefficients and information from the model
```

```
#Add model line to scatter plot above
```

```
#Test model fit with anova (test hypothesis)→anova(null)
```

## #3. Fit Full (+factor model) model

```
# Extract coefficients and information from the model
```

```
#Add model line to scatter plot above
```

```
#Test model fit with anova (test hypothesis)→anova(model1)
```

```
#4. Model comparison between full (+factor) vs reduced (null) model→anova(null,model1)
```

Repeat during model building

# Assumptions

# Assumptions of LMEs

- As with all linear models: Residuals follow a normal distribution with equal variance.
- Groups (e.g. animals) are randomly sampled from a “population” of groups (i.e., are independent and sampled without bias).
- Group means have a normal distribution.

# Assumptions of LMEs

- Replicates within groups are also randomly sampled (i.e. independent and sampled without bias).
- No carry-over between repeated measurements on the same subject.
- Sphericity: the variances of the differences between all pairs of factor levels are equal. (Problems can arise when one of the factors is time.)

Plot of residual values vs fitted values with `plot(model1)` give you different output than when used it with `lm()`

Example R code below on how I check LME assumptions in R

```
#Examine the assumptions of the model
#this is similar to the plot(model) that we did with lm()
#ASSUMPTION 1:ERRORS ARE NORMALLY DISTRIBUTED AROUND ZERO--yes
plot(null,animal~resid(.),abline=0,
     main=" Are Errors Norm. dist around Zero?")
#2CHECK HOMOGENEITY OF THE VARIANCE
plot(null,resid(.,type="p")~fitted(.)|animal,animal=0.5, adj=-0.3,abline=0,
     main=" Homogeneity of variance --")
#normality for all ---ok
qqnorm(null,~resid(.),main=" check normality")
#normality sep by each animal
qqnorm(null,~resid(.)|animal,main=" check normality")
#4TO CHECK FOR AUTOCORRELATION
acf(null$residuals)
```

# Confidence and Error

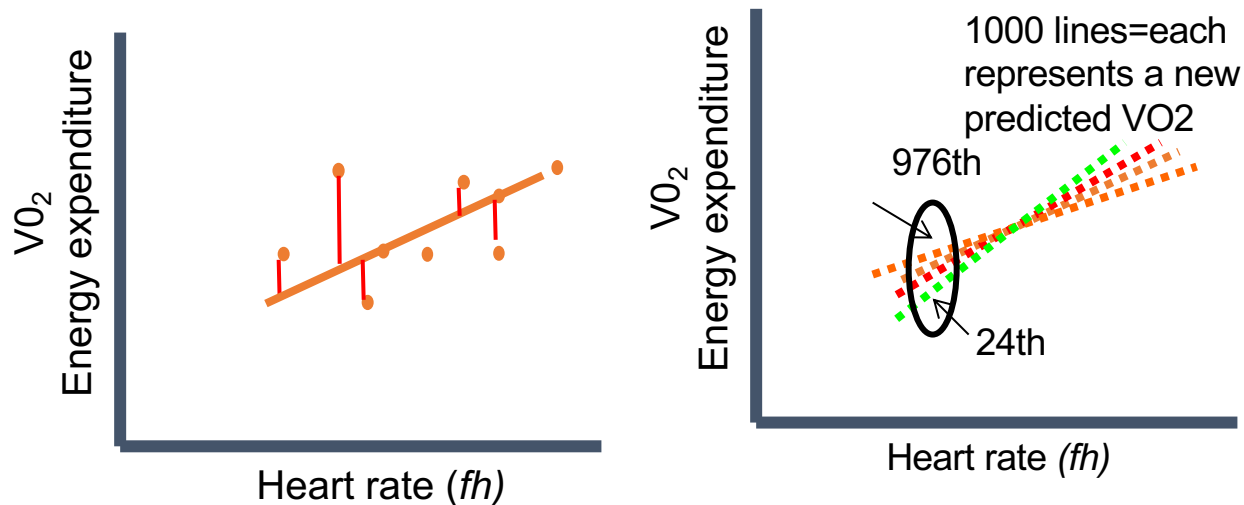


# 95% Confidence Intervals for **fixed effects**

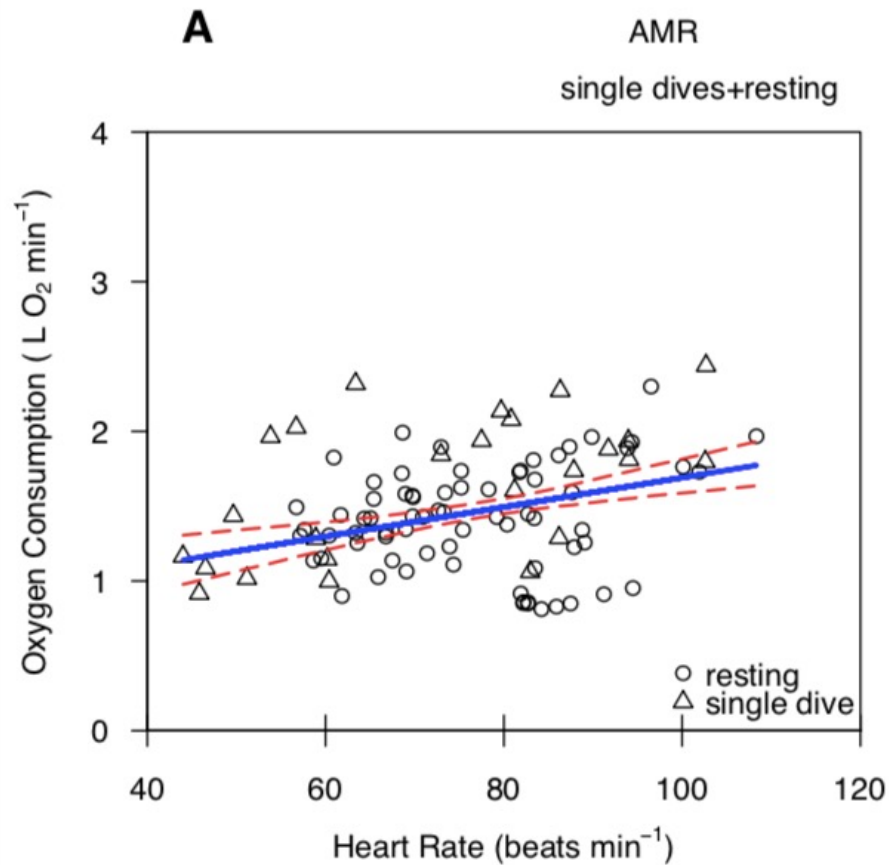
- **lmer use:** confint(z)
- **lme use:** intervals(z)

# Bootstrapping for 95% CI

- Fit model with lme
- Residuals represent **random error** in data that could not be modeled (difference between predicted value and true value)
- Took predicted values from model line and bootstrapped a random sample of the residuals
- Reassigned these residuals randomly to the predicted values and refit the LME model (preserved structure and repeated measures)
- In this way, the structure of the original model was conserved and the prediction errors were only associated with the dependent variable.
- Repeat 1000 times per model
- The ordered 24th and 976th bootstrap replicates represent lower and upper CI
  - Lower CI: 2.5% of 1000=25<sup>th</sup> replicate-1=24<sup>th</sup> replicate
  - Upper CI: 97.5% of 1000=975<sup>th</sup> replicate+1=976<sup>th</sup> replicate



# Example of plotted CI with bootstrapping



# What about $R^2$ for LME models?

- Unfortunately, there is **no direct equivalent of a traditional  $R^2$  for LME models**
- LME models have variance associated with both random factor (variation between-animals) and residual variance of fixed factors (within-animal variance)
- Difficult to compare  $R^2$  from LME models among studies

# Conditional and marginal $R^2$ for LME models

- Addresses some of the issues with calculating a goodness of fit on LME models by partitioning variance into fixed effects (marginal  $R^2$ ) and random effects (conditional  $R^2$ ).
- Improvement from a single “pseudo- $R^2$ ”, but caution should be used because it is still not a direct equivalent of traditional  $R^2$  from a non-mixed model and this function is still in developmental stages.
- We used the r.squaredGLMM package to calculate conditional and marginal  $R^2$  based on Nakagawa and Schielzeth (2013).

# Temporal Autocorrelation in lme

- Come see me during office hours if you have autocorrelation
- This is different than when time is your X variable,
- Mostly used for temporal correlation structures are *corAR1*, *corCAR1* (autoregressive and continuous autoregressive correlation structures of order 1), and *corCompSymm* (compound symmetry).

# Assignment #2

# Assignment #2: Due 18 March @9pm

- 5.5 weeks, but don't wait until the last minute
- Due after model selection intentionally
- Self-assessment with rubric *prior* to turning it in

- 1. Get your dataset soon**
- 2. Scatter plot to examine**
- 3. Decide if it is fixed only, mixed, or generalized**



# Assignment #2: Data set requirements

- **Linear, mixed, or generalized linear model in R.** You can choose which type of linear model to use, as long as it is appropriate for your dataset.
  - Fixed → Lm, glm, gam
  - Mixed → LME, GLMM, GAMM
- Only 1 response variable
- At least 1 categorical factor
- Include at least 1, and no more than 2, additional explanatory variables

# How to find data for Assignment #2

- Ask labmates, supervisor, other grad students for data
- Extract data that was used for something else other than a linear model
  - “How to find data for practicing R.docx” / Assignment #2 on Canvas
  - **Not** ok to redo a figure that’s already published and analyzed as a linear model, but you can extract data from published paper that was used for something else
- Yes it’s ok, if it’s your own thesis data, undergrad data

# Read all of the details and rubric on Canvas carefully

## Assignment #2: Analyze a linear model

Start Assignment

**Due** Mar 18 by 9p.m.    **Points** 25    **Submitting** a file upload  
**Available** Jan 9 at 9a.m. - Mar 18 at 10p.m.

**\*Important\*** Example Rubric available here to do self-assessment before submission ([BIOL 501 2022W2 Assignment #2 Rubric CANVAS 7Jan2023.pdf](#) ↓ )

This assignment is due **March 18, 9pm (Submit on Canvas)**

Obtain a data set and analyze it by fitting a linear, mixed, or generalized linear model in R. You can choose which type of linear model to use, as long as it is appropriate for your dataset.

- Obtain a data set from your supervisor or online data depository (e.g. datadryad.org).
- Include just one response variable.
- For the explanatory variables, include at least one categorical fixed factor, such as an experimental or observational treatment.
- **Include at least 1, and no more than 2**, additional explanatory variables (random or fixed factors, blocks, covariates, etc).

Prepare a thorough report on the analysis and interpretation of the data. Below I list some of the things to include in your report, but note that the list might not be complete and is variable depending on your model type and data set(i.e. you may need to include additional points to make a complete analysis based on your dataset). *HINT* It is helpful to think of this report as a mini version of a published manuscript with sections similar to the Introduction, Methods, Results, and Discussion.

# References on lme

Pinheiro and Bates is my #1 favourite reference. They wrote the lme package and Bates wrote lmer

- **lme()** [e-copy available]

- Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer

Linear mixed model implementation in lme4

Douglas Bates  
Department of Statistics  
University of Wisconsin – Madison

December 13, 2008

**Abstract**

We describe the form of the linear mixed-effects and generalized linear mixed-effects models fit by `lmer` and give details of the representation and the computational techniques used to fit such models. These techniques are illustrated on several examples.

- **lmer()**

Bates, D. (2010). *lme4: Mixed-effects modelling with R*. Springer

**lmer**—short handout by Bates (2008)  
e-copy available



This is an  
excellent  
resource too!

# Zuur (2007) and (2009)

- Zuur, Alain F., Elena N. Ieno, and Graham M. Smith. Analysing ecological data. Vol. 680. New York: Springer, 2007. **(Book 1)**
- Zuur, A.F. et al. (2009). Mixed effects models and extensions in ecology with R. Springer **(Book 2)**
- Have > 30 data sets attached for examples
- Go to: <https://www.highstat.com/index.php/mixed-effects-models-and-extensions-in-ecology-with-r>
- Download zipped data file called "ZuurDataMixedModelling"

# Additional Resources

- Quinn & Keough 2002. Experimental design and data analysis for biologists
  - Can be very useful in understanding design and assumptions of mixed-effects models (though it won't help you with modeling using maximum likelihood in R).
- Online books on mixed-effects models in R (see Books tab at course web site):
- Pinheiro and Bates (2000). Mixed-effects models in S and S-PLUS.
- Zuur et al (2009). Mixed effects models and extensions in ecology with R.
- Article:
- Bates, Douglas, et al. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67: 1-48.

Zuur, A.F. et al. (2009). *Mixed effects models and extensions in ecology with R*. Springer

- GLM, GAM, GLMM, GMM [e-copy available]
- Has 30 data sets attached for examples
  - Go to: <https://www.highstat.com/index.php/mixed-effects-models-and-extensions-in-ecology-with-r>
  - Download data

Workshop



<b>Measurement</b>	<b>What is the random group?</b>
Measure animals repeatedly	animal
Divide plots and apply a separate treatment to each plot	plot
Quadrats within transects (nested sample design)	Transects
Transects nested within woodlots	Wood lots

# Which mixed-effects package to use?

- lme function → nlme package
  - This is the one I primarily use
- lmer function → lme4 and lmerTest packages
  - R Tips page has more emphasis on lmer

# Workshop Thurs: Mixed models

## Imer examples (lme4 and lmerTest packages)

- Goldie's Vision (Y=numeric, **X=categorical**)
  - Y=sensitivity (numeric)
  - X=wavelength (categorical fixed factor)
  - No additional fixed factors
  - Random effect=fish with repeated measures on 5 fish
- Yukon yarrow (lmer) (Y=numeric, **X=categorical**)
  - Y=phenolics treatment (numeric). This is a plant defense compound.
  - X=treatment (categorical fixed factor, 4 levels)
  - Other fixed effects=Duration (categorical, 2 levels)
  - Random effect=plot

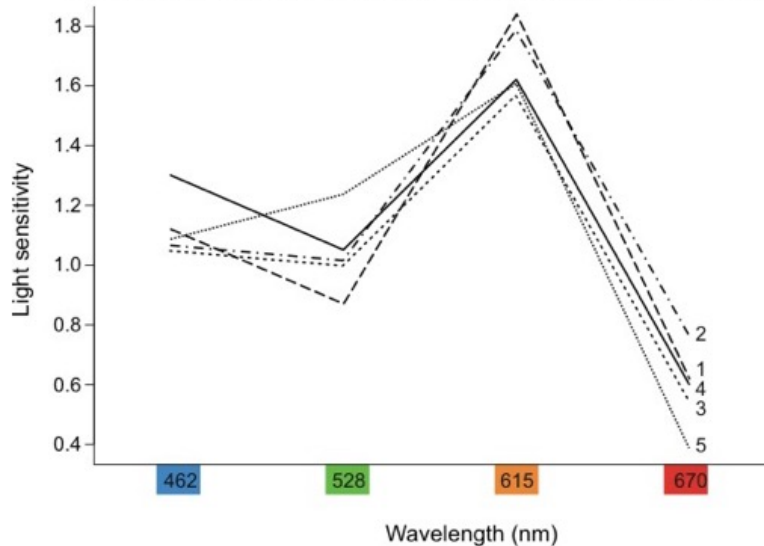
“Is there a difference?”,  
but can't predict Y with X  
using an equation if  
x=categorical

# Workshop Thurs: Mixed models

## Optional background information on Goldie's Vision in Workshop

### Example 2: "Subjects by treatment" repeated measures design

Cronly-Dillon and Muntz (1965) used the optomotor response to measure color vision in the goldfish. Each fish was tested at different wavelength in random order.



A large value indicates that the fish has high sensitivity --- it can detect a low light intensity.

#### Factors:

Wavelength (fixed, repeated measure)

Fish (random)



Light sensitivity of 5 goldfish to specific wavelengths of light.

# Workshop Thurs: Mixed models

## lme examples (nlme package)

Can x predict Y with a linear equation? Does this fixed factor change the relationship?

*Coming soon...*

- Can heart rate predict metabolic rate in sea lions? (numerical X and Y)
  - Y=metabolic rate (numerical)
  - X=heart rate (numerical)
  - Other fixed factors: dive type (3 levels) +dive depth (2 levels) +cumulative food fed (2 levels)
  - Random effect=animal (n=3 animals, repeated measures)

# Workshop Thurs: Mixed models

## lme examples (nlme package)

*Coming soon...*

- Lme example with Y=numerical and **X=categorical**

