

7. Centering and interpretation

- Learn why centering continuous predictors can aid in interpreting model results
- Learn about how to center categorical variables (& why it can be difficult)

How centering helps interpretation

This is the average
AFD of clams when
LENGTH = 0 and
MONTH = 2

Can a clam ever
have a length of 0?

```
> mod2 <- lm(AFD ~ LENGTH + fMONTH, data = clams)
> summary(mod2)

Call:
lm(formula = AFD ~ LENGTH + fMONTH, data = clams)

Residuals:
    Min       1Q   Median       3Q      Max
-0.084185 -0.014142 -0.003910  0.009508  0.274515

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.142832  0.008056  -17.729 < 2e-16 ***
LENGTH       0.012674  0.000348   36.420 < 2e-16 ***
fMONTH3      0.024168  0.008703    2.777  0.00575 **
fMONTH4      0.001953  0.004339    0.450  0.65291
fMONTH9      0.006264  0.008548    0.733  0.46408
fMONTH11     0.002317  0.006810    0.340  0.73383
fMONTH12    -0.020546  0.004564   -4.501  8.92e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02901 on 391 degrees of freedom
Multiple R-squared:  0.8556,    Adjusted R-squared:  0.8534
F-statistic:  386 on 6 and 391 DF,  p-value: < 2.2e-16
```

Interpreting continuous variables

- Remember that your intercepts in your model are the values of your response **WHEN EVERYTHING IS SET TO ZERO**
- This might be biologically meaningless
 - It is impossible to have a body size of zero! (e.g. Clams)
 - How can you have a zero observation?
- Centering your variables can be useful – positive values mean its above average, negative below average. Intercept is now the value at the average

Interpreting categorical variables

- Remember, for categorical predictors R will just pick one level of your predictor to be the “intercept”. Effect estimates are therefore deviations from that “intercept”
 - To get the true effect of a certain level of the effect, you need to add together the intercept and that levels effect

Centering clams

- Please center LENGTH on the Clam data set, and re-run the model with LENGTH and fMONTH
- How do you now interpret these estimates?
- Remove the intercept for MONTH and re-interpret

```
Clams$LENGTH.center <- scale(Clams$LENGTH, center = T, scale = F)
```

```
mod1 <- lm(AFD ~ LENGTH.center + fMONTH, data = Clams)  
summary(mod1)
```

```
mod2 <- lm(AFD ~ LENGTH.center + fMONTH-1, data = Clams)  
summary(mod2)
```

Centered LENGTH

call:

```
lm(formula = AFD ~ LENGTH.center + fMONTH, data = Clams)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.084185	-0.014142	-0.003910	0.009508	0.274515

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.085116	0.002876	29.595	< 2e-16	***
LENGTH.center	0.012674	0.000348	36.420	< 2e-16	***
fMONTH3	0.024168	0.008703	2.777	0.00575	**
fMONTH4	0.001953	0.004339	0.450	0.65291	
fMONTH9	0.006264	0.008548	0.733	0.46408	
fMONTH11	0.002317	0.006810	0.340	0.73383	
fMONTH12	-0.020546	0.004564	-4.501	8.92e-06	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02901 on 391 degrees of freedom

Multiple R-squared: 0.8556, Adjusted R-squared: 0.8534

F-statistic: 386 on 6 and 391 DF, p-value: < 2.2e-16

Centered LENGTH + suppressed intercept

```
Call:
lm(formula = AFD ~ LENGTH.center + fMONTH - 1, data = Clams)

Residuals:
      Min       1Q   Median       3Q      Max
-0.084185 -0.014142 -0.003910  0.009508  0.274515

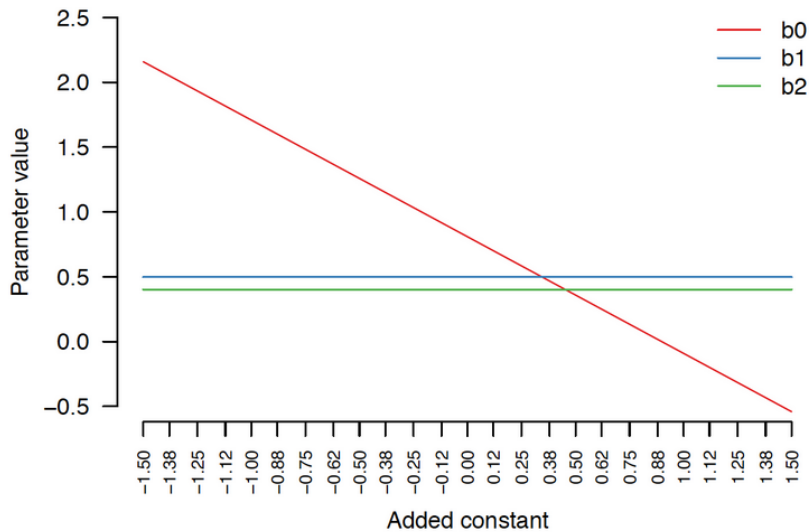
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
LENGTH.center  0.012674   0.000348   36.42  <2e-16 ***
fMONTH2        0.085116   0.002876   29.59  <2e-16 ***
fMONTH3        0.109284   0.007740   14.12  <2e-16 ***
fMONTH4        0.087069   0.002654   32.81  <2e-16 ***
fMONTH9        0.091381   0.007719   11.84  <2e-16 ***
fMONTH11       0.087434   0.006184   14.14  <2e-16 ***
fMONTH12       0.064570   0.004041   15.98  <2e-16 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02901 on 391 degrees of freedom
Multiple R-squared:  0.9355,    Adjusted R-squared:  0.9343
F-statistic: 810.1 on 7 and 391 DF,  p-value: < 2.2e-16

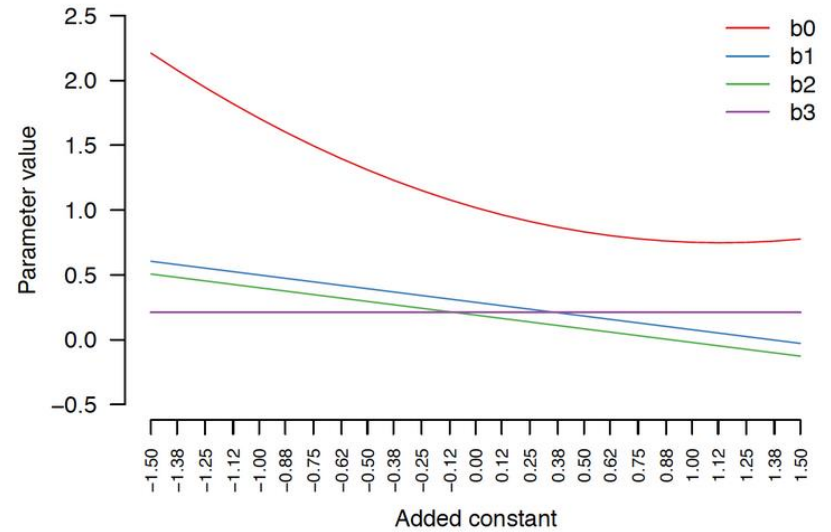
> |
```

But see

- <http://jmbh.github.io//CenteringPredictors/>
- If you only have main effects in your model then centering does not influence results, **BUT** if the predictors are involved in an interaction, centering **DOES** influence results!



Without an interaction, centering only influences the intercept (b_0); the effect of two main effects is the same



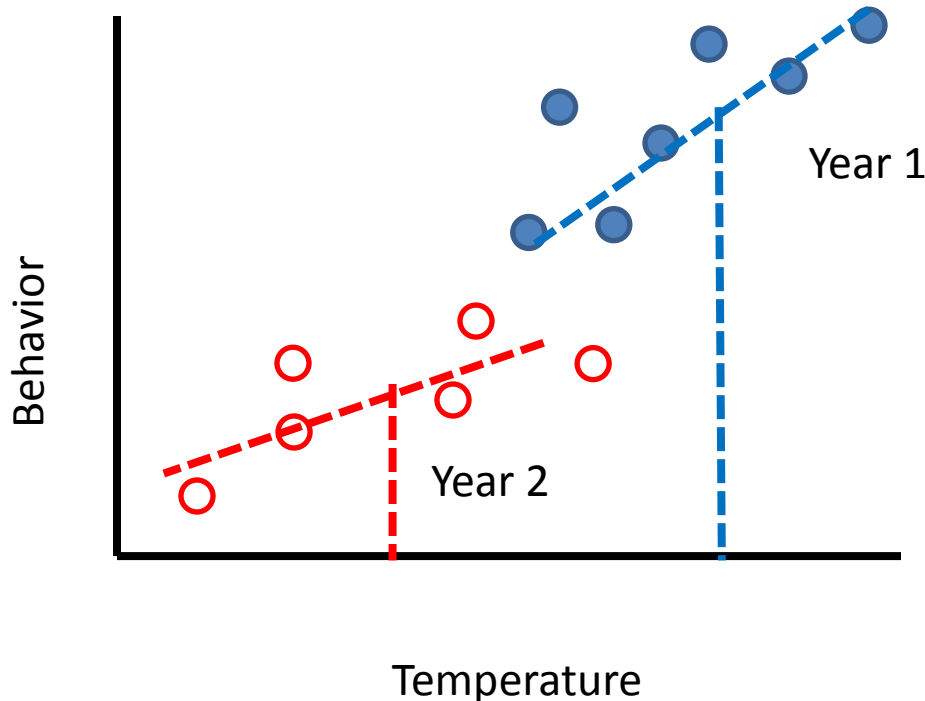
With an interaction, centering will maintain a constant interaction, but then alter the main effects and estimates of the intercept

Within-level centering

- If you have both categorical and continuous fixed effects and the continuous effect does not cover the same range for all levels of the categorical variable, then interpreting the effect of the continuous effect is very difficult!

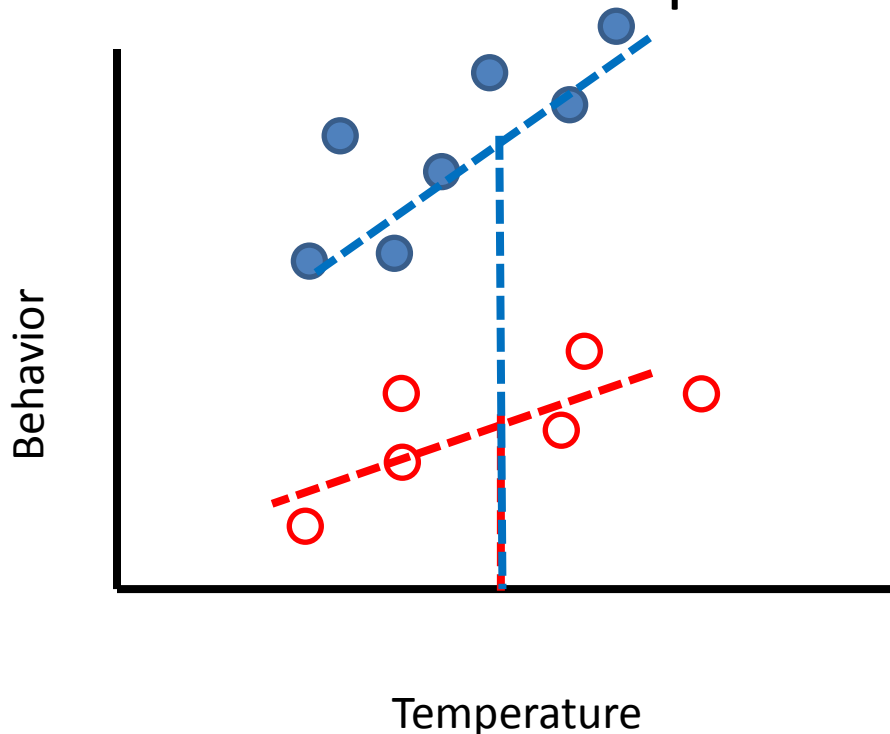
Example: between-year variance in temperature

- You are interested in how fish change their behavior in relation to water temperature. So you measure the behavior of fish multiple times over multiple years. **PROBLEM:** water temperature varies over the years



Example: between-year variance in temperature

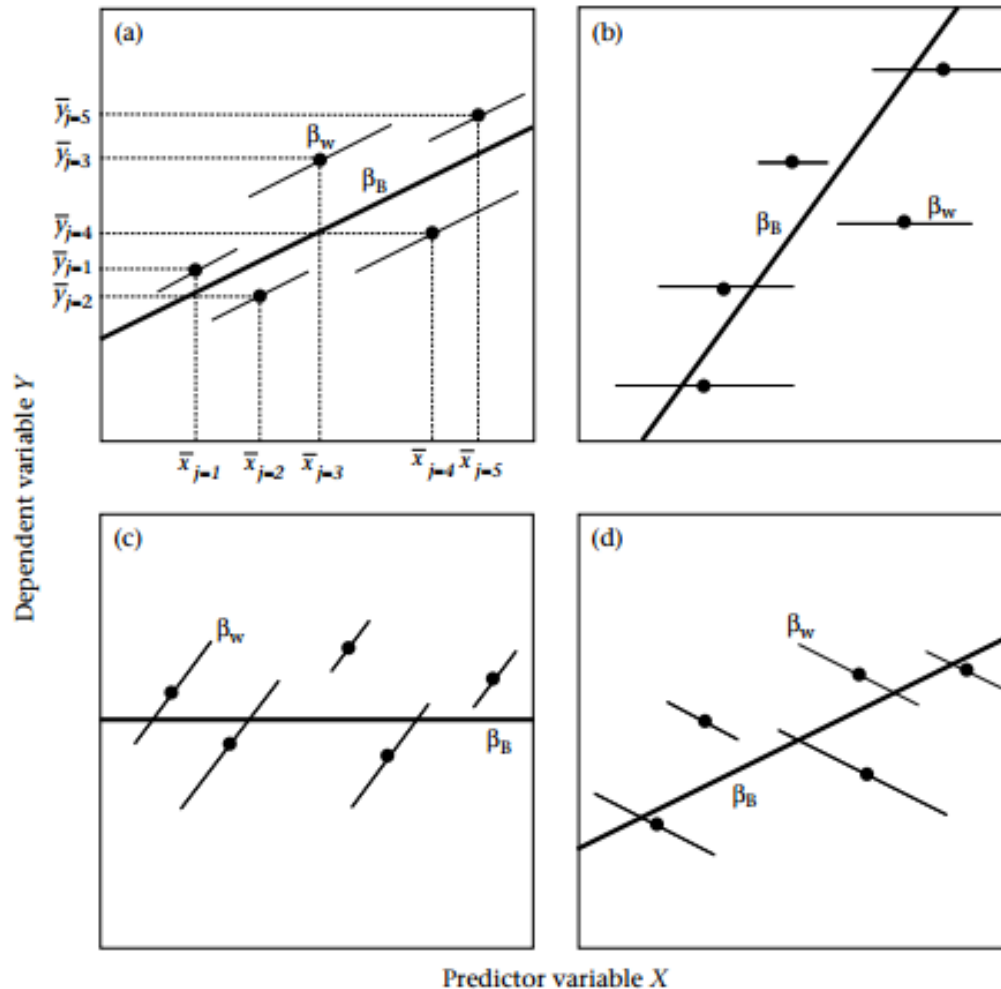
- You are interested in how fish change their behavior in relation to water temperature. So you measure the behavior of fish multiple times over multiple years. **PROBLEM:** water temperature varies over the years



- Two predictor variables:
- Absolute temperature
 - Year-centered temperature

Between- and within-individual centering

- Patterns of co-variation may actually vary within and between your levels
- Without within- and between-level centering, you mask these effects



Within a lake, fish spend the same time foraging regardless of resource abundance

Within a lake, fish spend LESS time foraging as resource abundance goes up

Van de Pol & Wright. 2009. A simple method for distinguishing within- and between-subject effects using mixed models. *Anim Behav* 77

Within subject centering

- Centering can be done in almost any way:
 - Within and across individual animals
 - Within and across lakes
 - Within and across years

Centering and interpretation

FURTHER READING on interpretation of estimates:

- Schielzeth 2010. Simple means to improve the interpretability of regression coefficients. *Methods Ecol Evol* 1

FURTHER READING on within/between individual centering:

- Van de Pol & Wright. 2009. A simple method for distinguishing within- and between-subject effects using mixed models. *Anim Behav* 77