

Regression Vs. ANOVA: Is a main effect really a main effect?

Arthur Capelier-Mourguy

Lancaster University

17th of July 2018

Outline

- 1 Introduction
 - Defining the problem
 - Content of this talk
- 2 Toy Example
 - Using categorical variables only
 - Using continuous variables
- 3 Real Data Example
 - Methods
 - Results
- 4 Conclusion

Defining the problem

What you might see

We defined a regression model $\text{Score} \sim \text{Condition} * \text{PrePost}$.

Defining the problem

What you might see

We defined a regression model

$\text{Score} \sim \text{Condition} + \text{PrePost} + \text{Condition}:\text{PrePost}.$

Defining the problem

What you might see

We defined a regression model

$\text{Score} \sim \text{Condition} + \text{PrePost} + \text{Condition}:\text{PrePost}.$

We found a significant main effect of Condition, with higher scores in the group A than in the group B.

Defining the problem

What you might see

We defined a regression model

$\text{Score} \sim \text{Condition} + \text{PrePost} + \text{Condition}:\text{PrePost}.$

We found a significant main effect of Condition, with higher scores in the group A than in the group B.

[Table with parameter estimates and statistics]

Defining the problem

What you might see

We defined a regression model

$\text{Score} \sim \text{Condition} + \text{PrePost} + \text{Condition}:\text{PrePost}.$

We found a significant main effect of Condition, with higher scores in the group A than in the group B.

[Table with parameter estimates and statistics]

- What does the regression model actually do?
- What do the parameter values in the table mean?
- What does “main effect” mean in the context of a regression?

Defining the problem

What you might see

We defined a regression model

$\text{Score} \sim \text{Condition} + \text{PrePost} + \text{Condition}:\text{PrePost}.$

We found a significant main effect of Condition, with higher scores in the group A than in the group B.

[Table with parameter estimates and statistics]

- What does the regression model actually do?
- What do the parameter values in the table mean?
- What does “main effect” mean in the context of a regression?

All stats in R have the same syntax

What to expect from this talk?

What this talk is about

- Demonstrate how ANOVA and regression results differ
- Detail what parameters in a regression model mean and do

What to expect from this talk?

What this talk is about

- Demonstrate how ANOVA and regression results differ
- Detail what parameters in a regression model mean and do

What this talk is not about

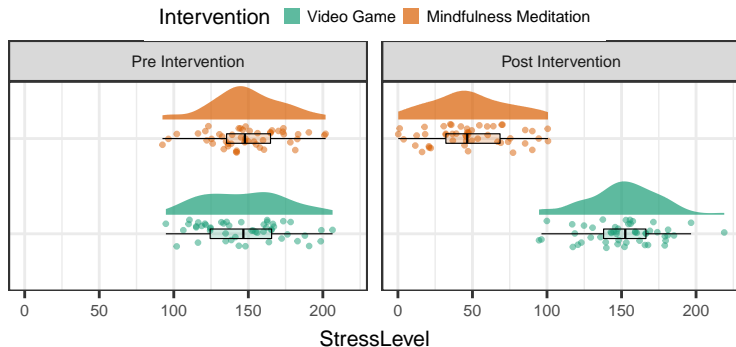
- How to use R
- How to build a good mixed-effects model
- The p -value debate

The simulated data with two categorical variables

Assessing stress levels after and before a 30 minutes intervention, “mindfulness meditation” or “video games”.

The simulated data with two categorical variables

Assessing stress levels after and before a 30 minutes intervention, “mindfulness meditation” or “video games”.



“Raincloud” plot: <https://micahallen.org/2018/03/15/introducing-raincloud-plots/>

ANOVA and regression results

ANOVA and regression results

ANOVA results

```
aov(StressLevel ~ Intervention*PrePost)
```

Parameter	Sum Square	F value	$Pr(> F)$
Intervention	114381	164.8	$< 2e-16$
PrePost	185059	266.7	$< 2e-16$
Intervention:PrePost	102808	148.2	$< 2e-16$

ANOVA and regression results

ANOVA results

```
aov(StressLevel ~ Intervention*PrePost)
```

Parameter	Sum Square	F value	$Pr(> F)$
Intervention	114381	164.8	$< 2e-16$
PrePost	185059	266.7	$< 2e-16$
Intervention:PrePost	102808	148.2	$< 2e-16$

Regression results

```
lm(StressLevel ~ Intervention*PrePost)
```

Parameter	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	147.305	3.674	40.099	$< 2e-16$
Intervention	1.801	5.195	0.347	0.729
PrePost	3.553	5.195	0.684	0.495
Intervention:PrePost	-104.193	7.347	-14.182	$< 2e-16$

ANOVA and regression results

ANOVA results

```
aov(StressLevel ~ Intervention*PrePost)
```

Parameter	Sum Square	F value	$Pr(> F)$
Intervention	114381	164.8	$< 2e-16$
PrePost	185059	266.7	$< 2e-16$
Intervention:PrePost	102808	148.2	$< 2e-16$

Regression results

```
lm(StressLevel ~ Intervention*PrePost)
```

Parameter	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	147.305	3.674	40.099	$< 2e-16$
InterventionMeditation	1.801	5.195	0.347	0.729
PrePostPost Intervention	3.553	5.195	0.684	0.495
Intervention:PrePost	-104.193	7.347	-14.182	$< 2e-16$

Graphically understanding the regression results

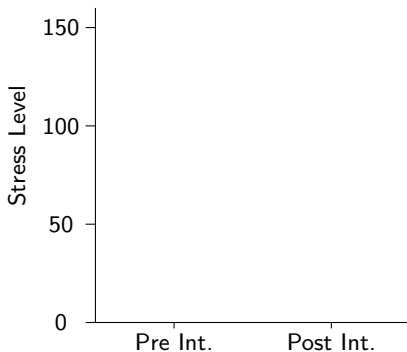
$$\begin{aligned} \text{Stress} &= 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo} \\ \text{Int} &= \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases} \end{aligned}$$

Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Video Game Pre Int., $\text{Stress} = 147.5 + 2 \times 0 + 3.5 \times 0 - 104 \times 0 \times 0$

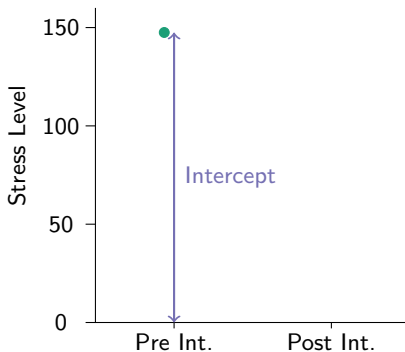


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Video Game Pre Int., $\text{Stress} = 147.5$

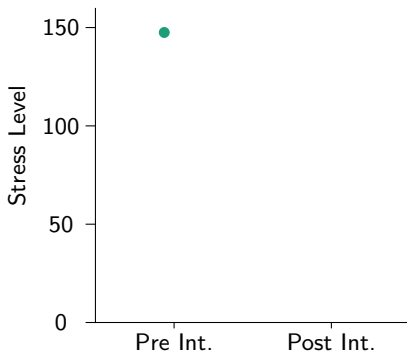


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Meditation Pre Int., $\text{Stress} = 147.5 + 2 \times 1 + 3.5 \times 0 - 104 \times 1 \times 0$

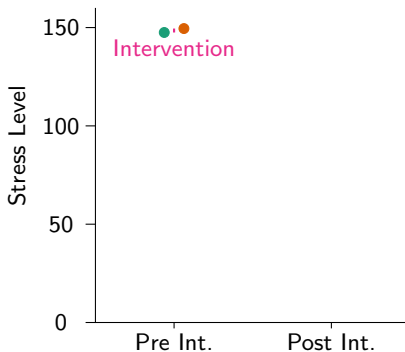


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Meditation Pre Int., $\text{Stress} = 147.5 + 2$

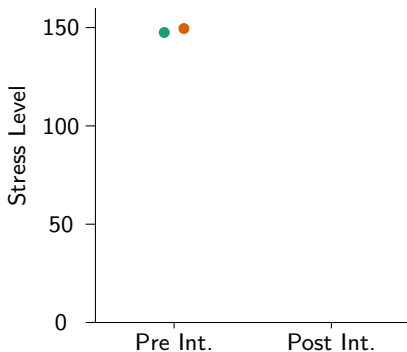


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Video Game Post Int., $\text{Stress} = 147.5 + 2 \times 0 + 3.5 \times 1 - 104 \times 0 \times 1$

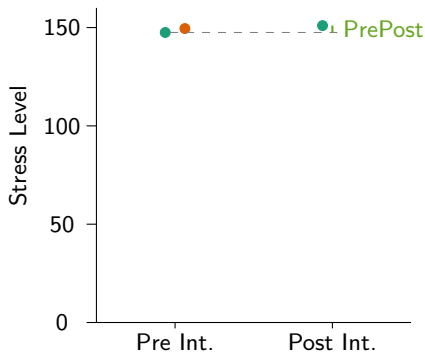


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Video Game Post Int., $\text{Stress} = 147.5 + 3.5$

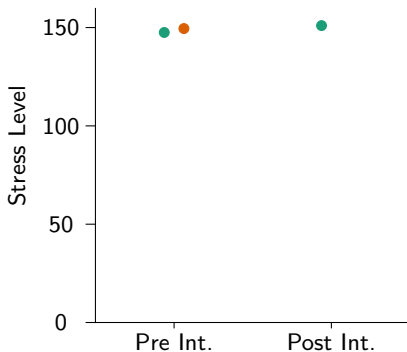


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Meditation Post Int., $\text{Stress} = 147.5 + 2 \times 1 + 3.5 \times 1 - 104 \times 1 \times 1$

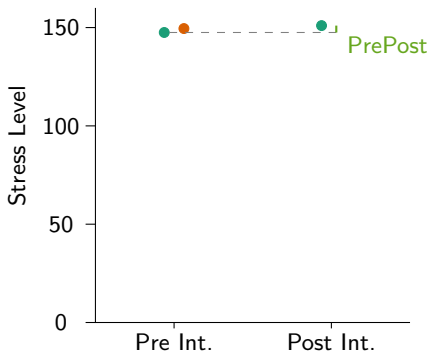


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Meditation Post Int., $\text{Stress} = 147.5 + 2 + 3.5 - 104$

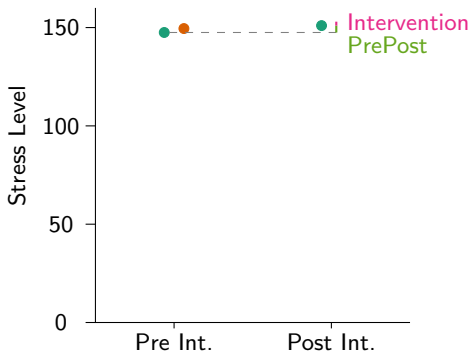


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Meditation Post Int., $\text{Stress} = 147.5 + 2 + 3.5 - 104$

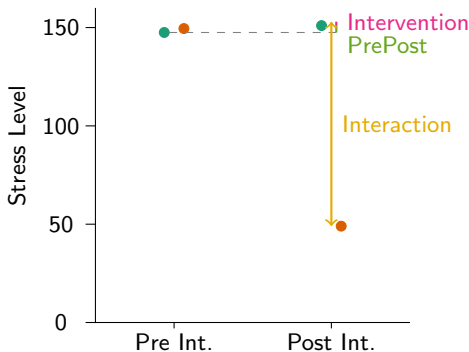


Graphically understanding the regression results

$$\text{Stress} = 147.5 + 2 \times \text{Int} + 3.5 \times \text{PrPo} - 104 \times \text{Int} \times \text{PrPo}$$

$$\text{Int} = \begin{cases} 0 & \text{if Video Game} \\ 1 & \text{if Meditation} \end{cases}, \quad \text{PrPo} = \begin{cases} 0 & \text{if Pre Intervention} \\ 1 & \text{if Post Intervention} \end{cases}$$

For Meditation Post Int., $\text{Stress} = 147.5 + 2 + 3.5 - 104$



Impact of the choice of reference levels

Impact of the choice of reference levels

Let's redefine Intervention: $0 = \text{Meditation}$, $1 = \text{Video Game}$.

Impact of the choice of reference levels

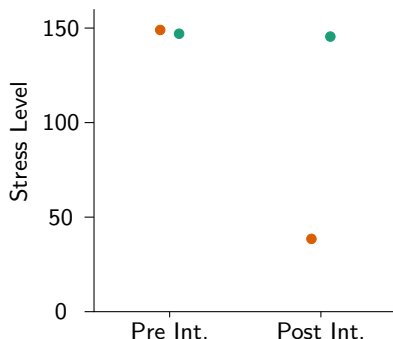
Let's redefine Intervention: 0 = Meditation, 1 = Video Game.

Parameter	Estimate	Std. Error	<i>t</i> value	$Pr(> t)$
(Intercept)	149.106	3.674	40.589	$< 2e-16$
InterventionVideo Game	-1.801	5.195	-0.347	0.729
PrePostPost Intervention	-100.640	5.195	-19.372	$< 2e-16$
Intervention:PrePost	104.193	7.347	14.182	$< 2e-16$

Impact of the choice of reference levels

Let's redefine Intervention: 0 = Meditation, 1 = Video Game.

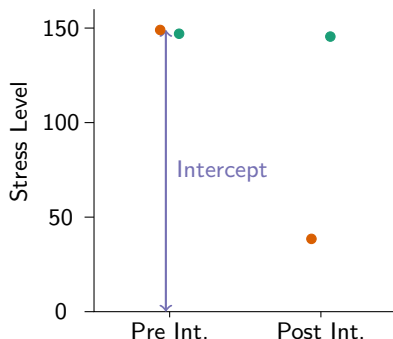
Parameter	Estimate	Std. Error	<i>t</i> value	$Pr(> t)$
(Intercept)	149.106	3.674	40.589	$< 2e-16$
InterventionVideo Game	-1.801	5.195	-0.347	0.729
PrePostPost Intervention	-100.640	5.195	-19.372	$< 2e-16$
Intervention:PrePost	104.193	7.347	14.182	$< 2e-16$



Impact of the choice of reference levels

Let's redefine Intervention: 0 = Meditation, 1 = Video Game.

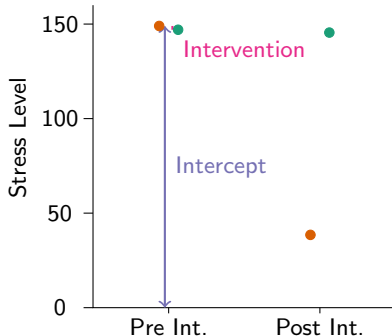
Parameter	Estimate	Std. Error	<i>t</i> value	$Pr(> t)$
(Intercept)	149.106	3.674	40.589	$< 2e-16$
InterventionVideo Game	-1.801	5.195	-0.347	0.729
PrePostPost Intervention	-100.640	5.195	-19.372	$< 2e-16$
Intervention:PrePost	104.193	7.347	14.182	$< 2e-16$



Impact of the choice of reference levels

Let's redefine Intervention: 0 = Meditation, 1 = Video Game.

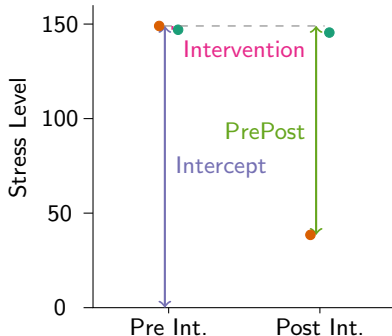
Parameter	Estimate	Std. Error	<i>t</i> value	$Pr(> t)$
(Intercept)	149.106	3.674	40.589	$< 2e-16$
InterventionVideo Game	-1.801	5.195	-0.347	0.729
PrePostPost Intervention	-100.640	5.195	-19.372	$< 2e-16$
Intervention:PrePost	104.193	7.347	14.182	$< 2e-16$



Impact of the choice of reference levels

Let's redefine Intervention: 0 = Meditation, 1 = Video Game.

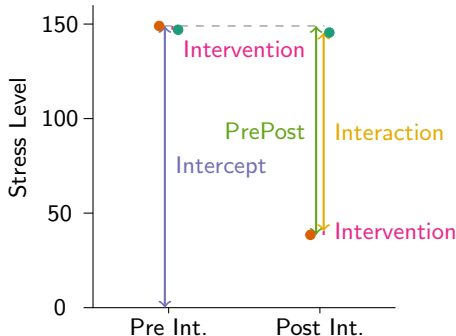
Parameter	Estimate	Std. Error	<i>t</i> value	$Pr(> t)$
(Intercept)	149.106	3.674	40.589	$< 2e-16$
InterventionVideo Game	-1.801	5.195	-0.347	0.729
PrePostPost Intervention	-100.640	5.195	-19.372	$< 2e-16$
Intervention:PrePost	104.193	7.347	14.182	$< 2e-16$



Impact of the choice of reference levels

Let's redefine Intervention: 0 = Meditation, 1 = Video Game.

Parameter	Estimate	Std. Error	t value	$Pr(> t)$
(Intercept)	149.106	3.674	40.589	$< 2e-16$
InterventionVideo Game	-1.801	5.195	-0.347	0.729
PrePostPost Intervention	-100.640	5.195	-19.372	$< 2e-16$
Intervention:PrePost	104.193	7.347	14.182	$< 2e-16$

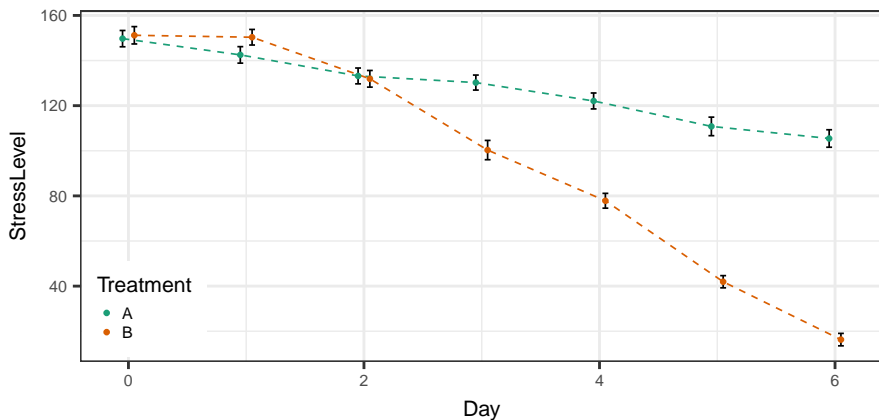


A slightly less meaningful example with a continuous variable

Assessing the difference between treatment A and treatment B in reducing stress levels each day over a week.

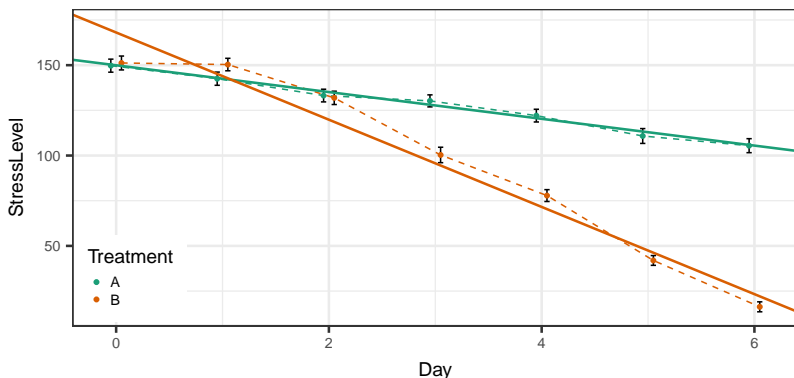
A slightly less meaningful example with a continuous variable

Assessing the difference between treatment A and treatment B in reducing stress levels each day over a week.



Regression results

Parameter	Estimate	Std. Error	<i>t</i> value	$Pr(> t)$
(Intercept)	149.9317	2.5003	59.967	$< 2e-16$
TreatmentB	18.1489	3.5359	5.133	$3.71e-7$
Day	-7.4051	0.6934	-10.679	$< 2e-16$
TreatmentB:Day	-16.7244	0.9807	-17.054	$< 2e-16$



The experiment in a nutshell

Categorisation and labelling in 15-month-old infants.

The experiment in a nutshell

Categorisation and labelling in 15-month-old infants.

Conditions Label and no-label.

The experiment in a nutshell

Categorisation and labelling in 15-month-old infants.

Conditions Label and no-label.

Familiarisation Snake-like animal with a head and a tail.

Label condition: categories defined by the tail.

The experiment in a nutshell

Categorisation and labelling in 15-month-old infants.

Conditions Label and no-label.

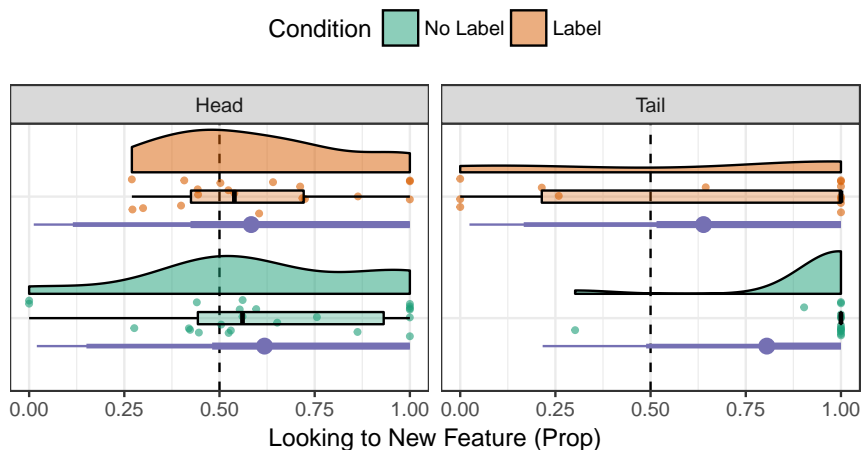
Familiarisation Snake-like animal with a head and a tail.

Label condition: categories defined by the tail.

Novelty Preference Did they encode the tail? The head?

One old animal against one animal with a new head/tail.

Choosing reference levels



What's the take home message?

What's the take home message?

- Don't assume you know what your model's parameters mean

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions
- Ask online

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions
- Ask online: StackOverflow

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions
- Ask online: StackOverflow, Cross Validated

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions
- Ask online: StackOverflow, Cross Validated, Github

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions
- Ask online: StackOverflow, Cross Validated, Github, Twitter

What's the take home message?

- Don't assume you know what your model's parameters mean
- Break down the model formula case-by-case
- Plot model predictions
- Ask online: StackOverflow, Cross Validated, Github, Twitter

Thanks for listening!