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

Arthur Capelier-Mourguy

Lancaster University

17th of July 2018

Outline

Defining the problem

What you might see

We defined a regression model `Score ~ Condition*PrePost`.

Defining the problem

What you might see

We defined a regression model
`Score ~ Condition + PrePost + Condition:PrePost`.

Defining the problem

What you might see

We defined a regression model
Score ∼ Condition + PrePost + Condition: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
Score $\sim$ Condition + PrePost + Condition: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
Score $\sim$ Condition + PrePost + Condition: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
Score $\sim$ Condition + PrePost + Condition: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
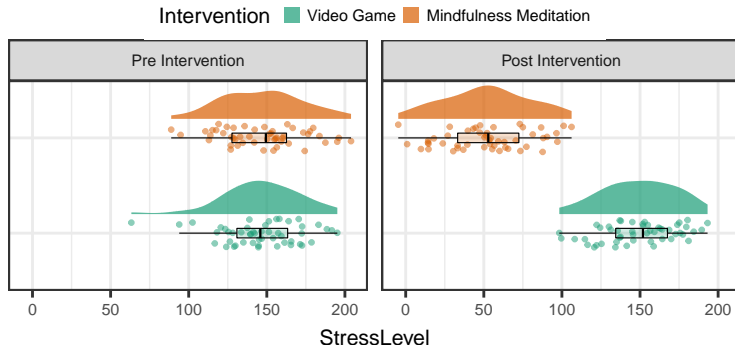
What this talk is not about

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

The simulated data

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

The simulated data

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

ANOVA and regression results

ANOVA and regression results

ANOVA results

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

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

ANOVA and regression results

ANOVA results

`aov(StressLevel ~ Intervention*PrePost)`

Parameter	Sum Square	F value	$Pr(> \mathrm{F})$
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)	150.703	3.725	40.453	$< 2e{-}16$
Intervention	2.484	5.269	0.472	0.638
PrePost	-106.182	5.269	-20.154	$< 2e{-}16$
Intervention:PrePost	90.960	7.451	12.172	$< 2e{-}16$

ANOVA and regression results

ANOVA results

`aov(StressLevel ~ Intervention*PrePost)`

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

Regression results

`lm(StressLevel ~ Intervention*PrePost)`

Parameter	Estimate	Std. Error	t value	$Pr(>|t|)$
(Intercept)	150.703	3.725	40.453	$< 2\mathrm{e}{-}16$
InterventionVideo Game	2.484	5.269	0.472	0.638
PrePostPost Intervention	-106.182	5.269	-20.154	$< 2\mathrm{e}{-}16$
Intervention:PrePost	90.960	7.451	12.172	$< 2\mathrm{e}{-}16$

Graphically understanding the regression results

Changes to the simulated data

Regression results

Graphically understanding the regression results

The experiment in a nutshell

Impact of the choice of reference levels

What's the take home message?