Mixed effects models

Readings for today

- Bates, Douglas M. "lme4: Mixed-effects modeling with R." (2010): 470-474.
- Yarkoni, T. (2019). The generalizability crisis. PsyArXiv

Topics

1. Fixed vs. random effects
2. Towards increased generalizability
3. Random effects vs. nuisance terms

Fixed vs. random effects

Example: random factors

Q: Does having distracting stimuli on classroom walls impact learning?

Variables:

- Y_{recall}: item recall after 5 minutes
- $X_{walls} : \begin{cases} 1, & \text{distracting walls} \\ 0, & \text{blank walls} \end{cases}$
- Z_{class}: categorical classroom ID

Bad model: $Y_{recall} = \hat{\beta}_0 + \underbrace{\hat{\beta}_1 X_{walls}}_{\text{main effect of wall}} + \underbrace{\hat{\beta}_2 Z_{class}}_{\text{main effect of classroom}}$

If classroom ID is an arbitrary label, what does a unit change in Z_{class} mean with regards to unit changes in Y_{recall}?

(Bates 2010)

Mixed effects models

$$Y = \underbrace{\sum_{j=1}^{p} \hat{\beta}_j X}_{\text{fixed effects}} + \underbrace{\sum_{k=1}^{q} \hat{\nu}_k Z}_{\text{random effects}}$$

Fixed Effects: Variables whose relationship with Y are stationary and covary in a meaningful way.

Random Effects: Qualitative variables that are not of primary interest but have a systematic influence on Y.

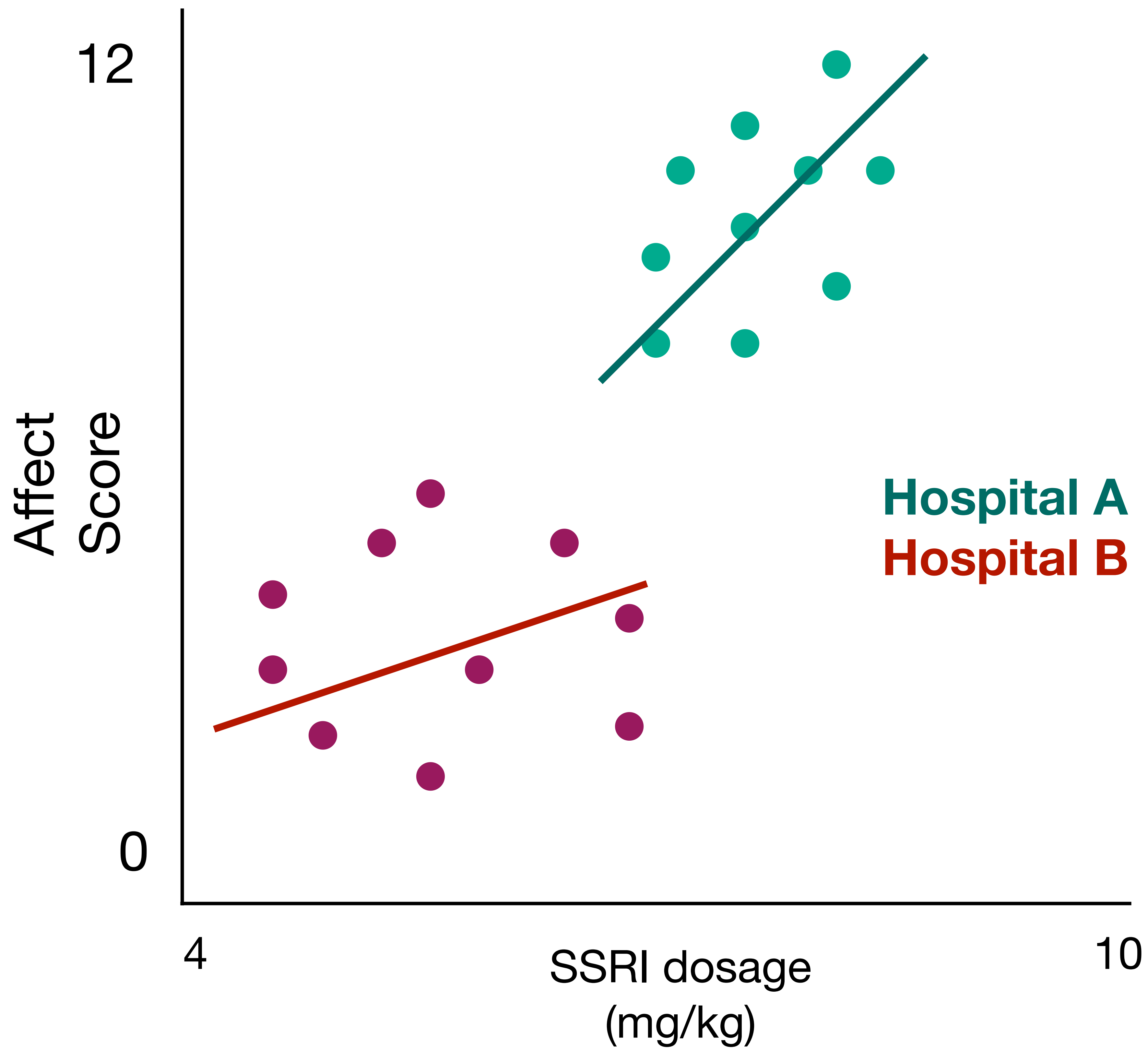
(Bates 2010)

Example: random effect

Q: Does SSRI dosage increase positive affect?

Variables:

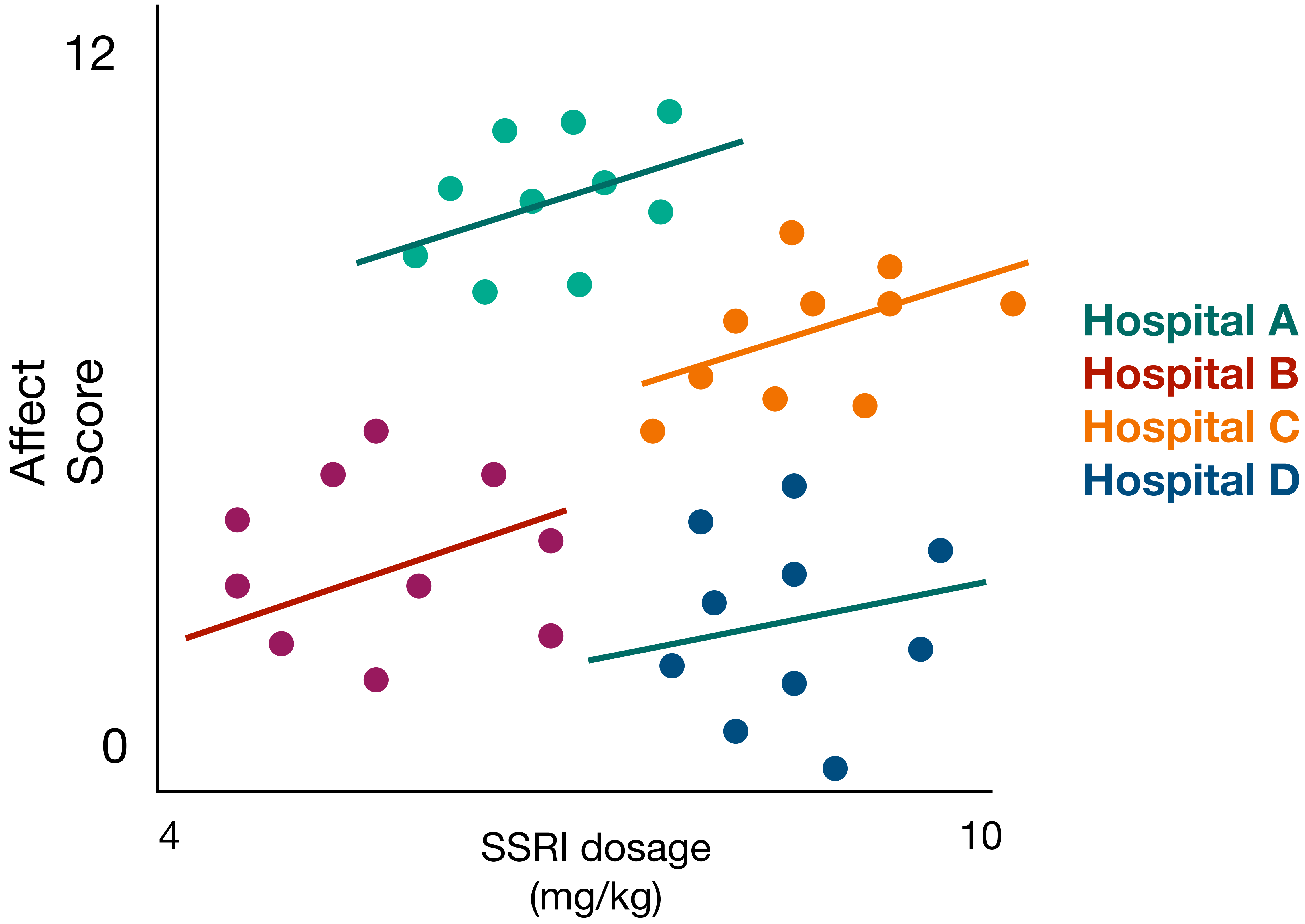
- Y_{affect}: positive affect score
- X_{SSRI}: dosage as mg/kg body weight
- $Z_{hospital}$: hospital ID



(Yarkoni 2019)

Types of random effects

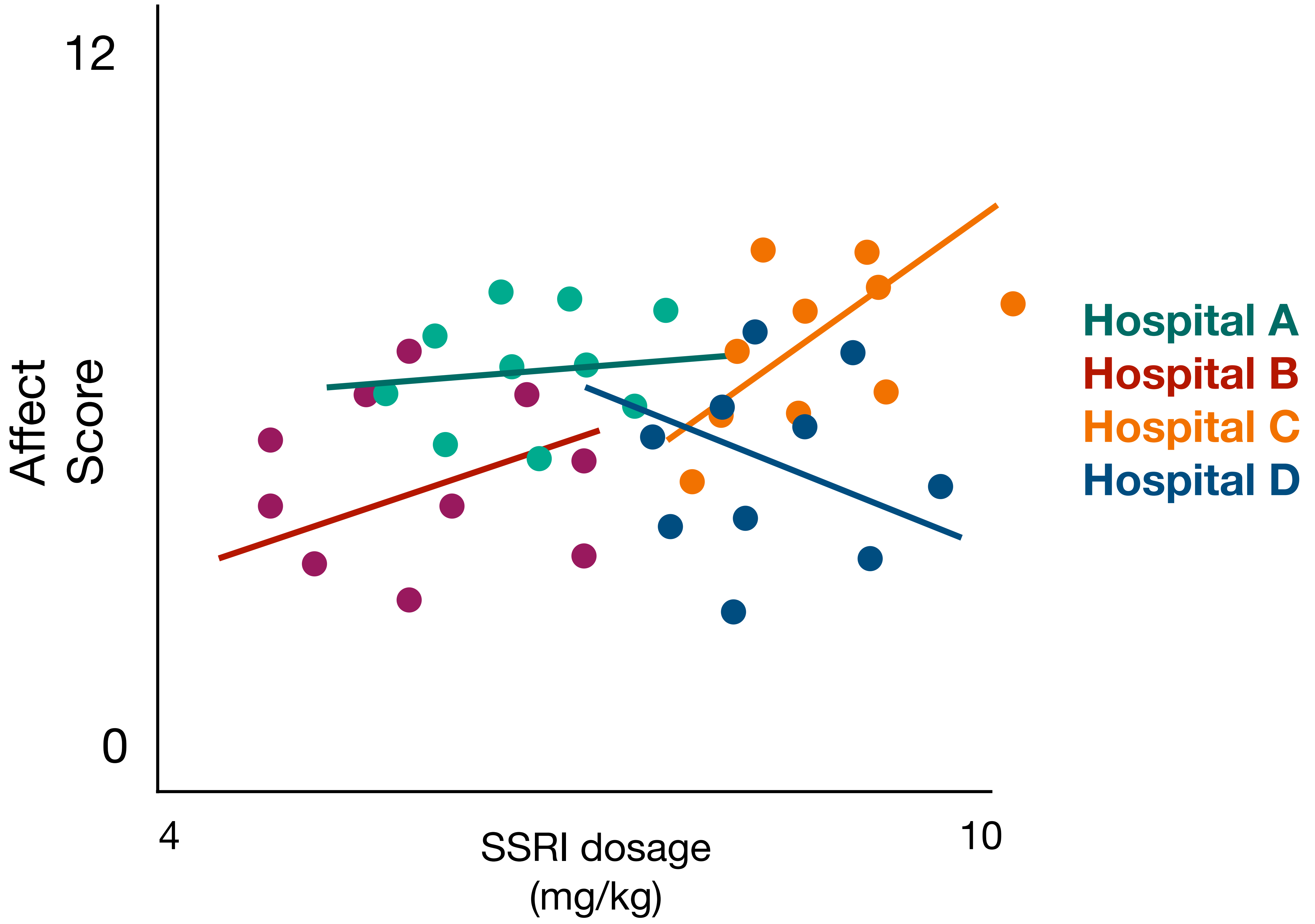
<u>Mean effects:</u> The random factor shifts $\bar{Y}$ (i.e., $\hat{\beta}_0$), but does not impact the nature of $\hat{\beta}_i X$.



(Bates 2010)

Types of random effects

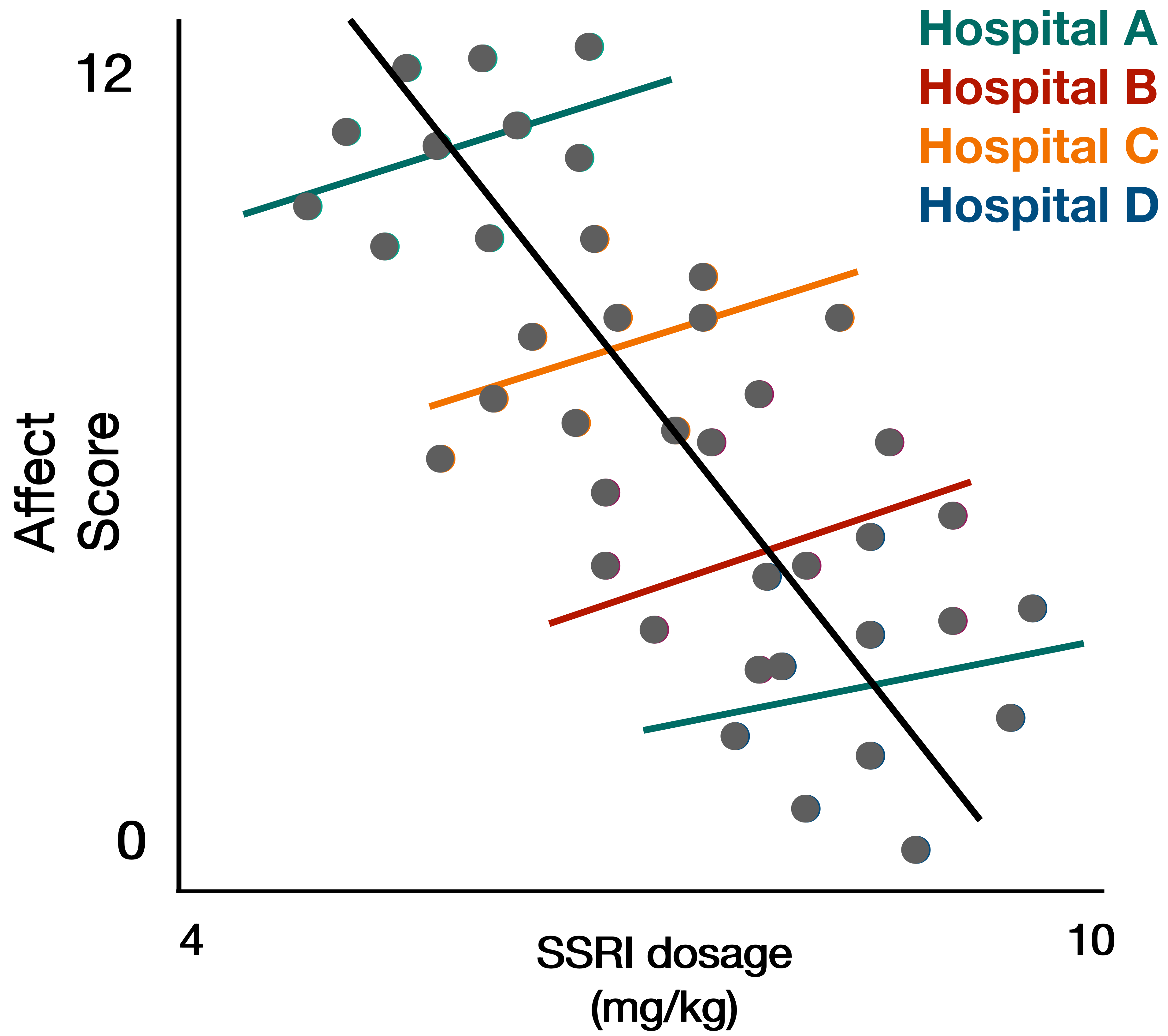
<u>Slope effects:</u> The random factor shifts $\hat{\beta}_i X$, but does not impact $\bar{Y}$.



(Bates 2010)

Simpson’s Paradox

Across batch (group) trends conflict with within batch (group) trends



Towards increased generalizability

Random variables

<u>Examples:</u>

- Subject ID
- Voxel
- Neuron
- Classroom
- Data collection site
- EEG system
- Stimulus presentation software

- Image stimuli
- Testing computer
- Operating system
- Testing room

There are many random factors that contribute to experimental effects that we often overlook.

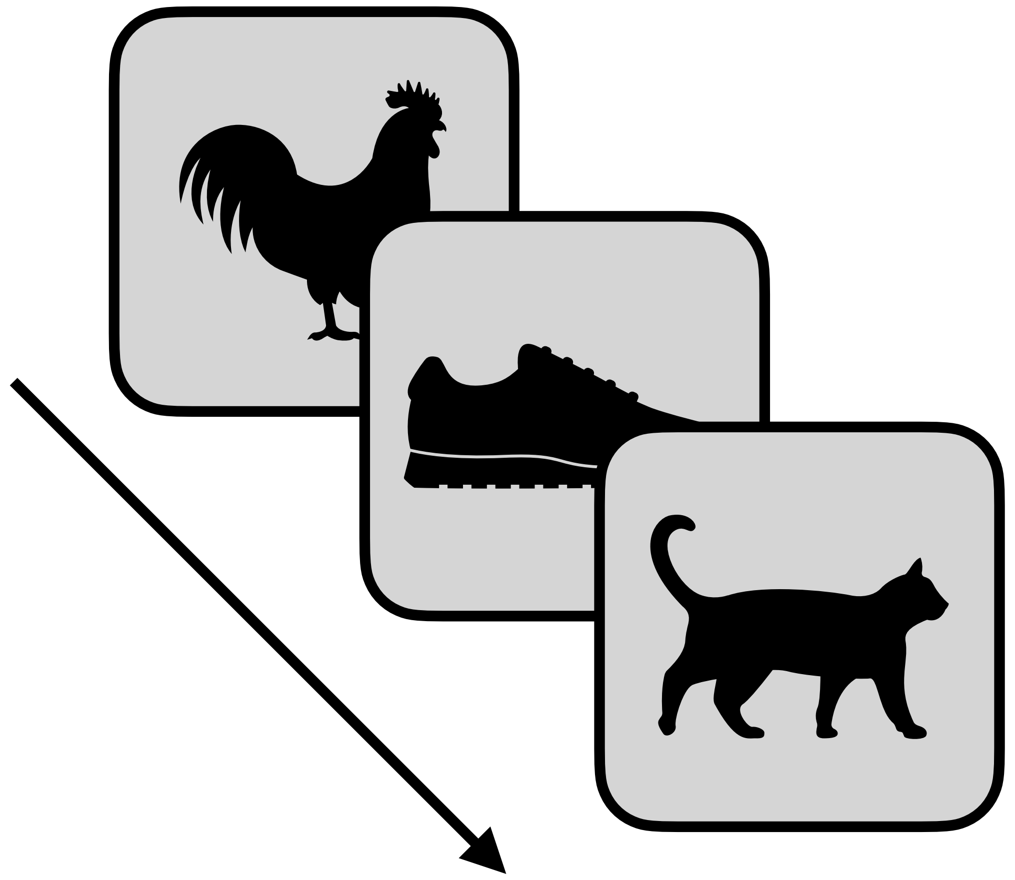
(Yarkoni 2019)

Example: generalizable model

Q: Do people respond faster to animate objects than inanimate objects?

Variables:

- Y_{RT}: reaction time
- $X_{animacy}: \begin{cases} 1, & \text{animate} \\ 0, & \text{inanimate} \end{cases}$
- Z_{image}: JPG image ID number



Traditional model:

$$Y_{RT} = \hat{\beta}_0 + \hat{\beta}_1 X_{animacy}$$

Full model:

$$Y_{RT} = \hat{\beta}_0 + \underbrace{\hat{\beta}_1 X_{animacy}}_{\text{hypotheses}} + \underbrace{\hat{\nu}_2 Z_{image}}_{\text{incidental}}$$

(Yarkoni 2019)

Considerations

Incidental factors:

- Image stimuli
- Testing computer
- Operating system
- Testing room
- Research assistant
- Monitor type
- Electrode
- Scanner

Limitations:

- Statistical power (more factors = greater model variance)
- Random effects are largely uninterpretable

(Yarkoni 2019)

Random effects vs nuisance terms

The problem of confounds

<u>Confounds:</u> Variables that influence the $Y = \hat{f}(X)$ relationship but are not needed to evaluate the central hypothesis.

Random Effects

Nuisance Factors

Nuisance factors

Two types of X:

Units are meaningful, not random

- $X_{target} \rightarrow$ factors of interest
- $X_{nuisance} \rightarrow$ factors that impact $\hat{\beta}_i X_{target}$

$$Y = \hat{\beta}_0 + \underbrace{\hat{\beta}_1 X_{target}}_{\text{hypothesis}} + \underbrace{\hat{\beta}_2 X_{nuisance}}_{\text{incidental}}$$

Collinearity

A correlation between two or more predictor variables in X.

<u>Strong case:</u> $Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$

<u>Set:</u> $X_1 = X_2 \rightarrow \rho(X_1, X_2) = 1$

<u>Then:</u>
$$
\begin{aligned}
Y &= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \\
&= \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_1 \\
&= \hat{\beta}_0 + X_1 \underbrace{(\hat{\beta}_1 + \hat{\beta}_2)}_{\tilde{\beta}_1} \\
&= \hat{\beta}_0 + \tilde{\beta}_1 X_1
\end{aligned}
$$

Nuisance factors <u>can</u> leverage the effect that collinearity has on $\hat{f}(X)$ in order to usefully temper confounded relationships between X_{target} and Y.

(James et al. 2013)

Nuisance vs. Random

OLS: $Y = \hat{\beta}_0 + \hat{\beta}_1 X_{target} + \hat{\beta}_2 X_{nuisance}$

Objective: $\min(\|Y - \hat{\beta}_0 - \hat{\beta}_1 X_{target} - \hat{\beta}_2 X_{nuisance}\|^2)$

Mixed Effects: $Y = \hat{\beta}_1 X_{target} + \hat{\nu}_2 Z_{random}$

Objective: $\min(\|Y - \hat{\beta}_0 - \hat{\beta}_1 X_{target} - \hat{\nu}_1 \Lambda_\theta Z_{random}\|^2 + \|\hat{\nu}\|^2)$

covariance matrix that explains structure in Z_{random}

$\hat{\nu} \sim N(0, \Sigma_\theta), \Sigma_\theta = \sigma \Lambda_\theta \Lambda_\theta'$

(Bates 2010)

Take home message

- Accounting for random effects in your model, as well as nuisance factors, can improve the generalizability of your results by accounting for incidental influences on your outcome measures.