Shrinkage:
When Others Help You Become a Better You

Stephan Lewandowsky

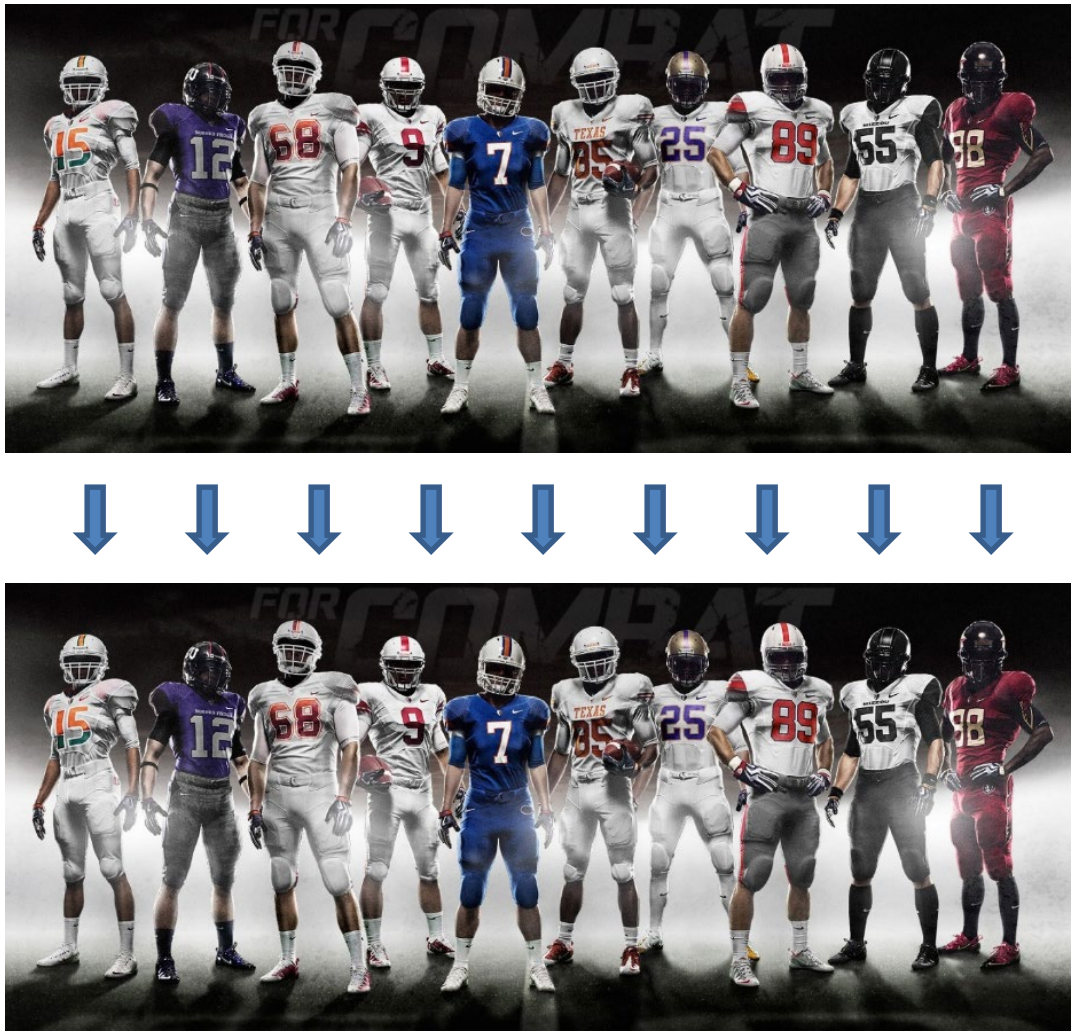
Your Task

- Predict second-half performance from first half

Three Options:
1. Aggregate Analysis

....

Three Options:
2. Individual Analysis



Three Options:
3. Huh….?

- Aggregation has obvious problems
- Predicting each individual's performance on the basis of that individual's performance seems obvious and optimal?
- We can do better than that ….

Three Options:
3. Shrinkage Estimators

- Prediction can be improved by adjusting each individual's prediction y_i towards the mean across individuals $\bar{y}$:

$$y_{i2} = \overline{y_1} + c(y_{i1} - \overline{y_1})$$

  where $c < 1$.

- "James-Stein estimator"
- Consider others in addition to yourself to predict your own performance

James-Stein Estimator in R

How Much Shrinkage?

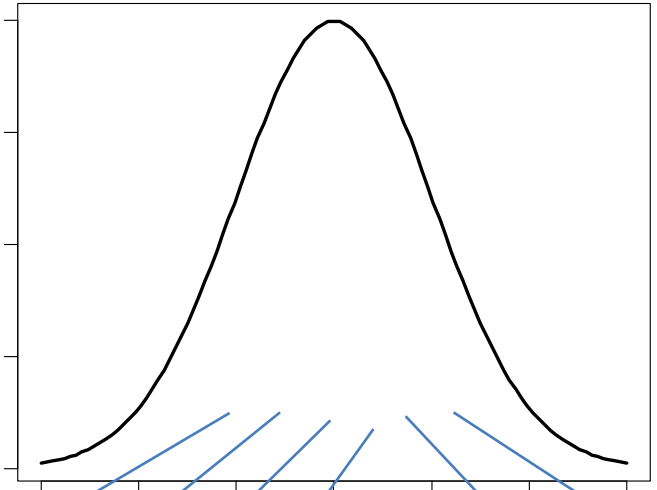
- Intuitions:
  - the more means there are, the more they should be considered → more shrinkage
  - the more precisely a mean is known, the less it should be affected by others → less shrinkage
  - the more discrepant the means are, the less likely they are to measure something common → less shrinkage

$$C = \left(1 - \frac{(m-3)\sigma^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2}\right)$$

Multilevel or Hierarchical Modeling

- Consider all subjects individually, but exploit dependence between them
- Key aspect of hierarchical modeling
  - model recognizes individual variation
  - but assumes there is an orderly distribution governing that variation
- Exploit shrinkage

Parent Distribution

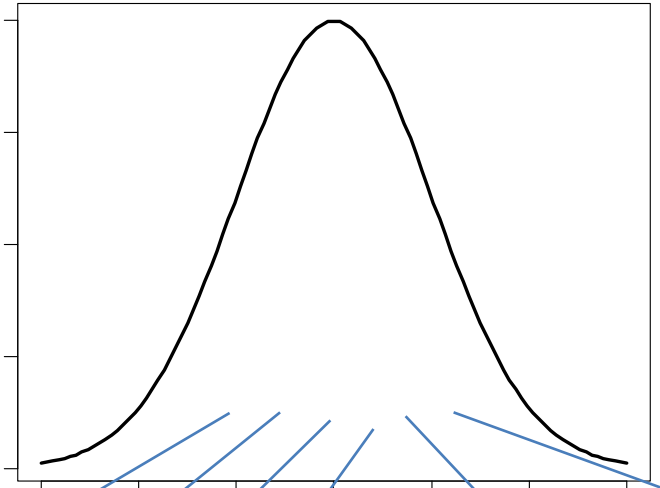


Hyper parameters

$(\mu_\theta, \sigma_\theta)$

θ_1 θ_2 θ_3 θ_4 θ_5 θ_6 θ_7 … θ_n

Parent Distribution



Hyper parameters

$(\mu_\theta, \sigma_\theta)$

θ_1 θ_2 θ_3 θ_4 θ_5 θ_6 θ_7 … θ_n

Hierarchical Bayesian Models

Give you even more than everything you could possibly want in life …