# Multicompetitor and ranking models

# Goal for this set of lecture notes

- Normal models for measured outcomes in multicompetitor events
- Plackett-Luce model for rank orderings
- Applications

# Goal for this set of lecture notes

- The normal models are a special case of blocked one-way ANOVA. One reference (not specific to sports!) is
  - Section 9.2 of Albert, J., Rizzo, M. (2012). Analysis of Variance II. In: R by Example. Use R!. Springer, New York, NY. (this chapter is uploaded to the course Canvas site)
- Here are two references for the Plackett-Luce model:
  - Turner, H.L., van Etten, J., Firth, D. *et al.* (2020) Modelling rankings in R: the **PlackettLuce** package. *Comput Stat* **35**, 1027–1057.
  - The start of chapter 3 of Zhang, S. (2021). *Building upon Bradley-Terry and Plackett-Luce: some methods for modeling paired comparison and rank order data* (Doctoral dissertation, Harvard University).

# Introduction

Interest in measuring ability or merit in multicompetitor games/sports in which competitors receive numerical scores, or are rank-ordered, such as

- races (human, car, horses)
- poker
- gymnastics
- diving
- golf

# Introduction

We will focus on two types of outcomes:

- Multicompetitor games/sports in which outcomes are numerical (e.g., race times, point scores, etc.)
- Multicompetitor games/sports which result only in rank orderings

Even when outcomes are numerical, we can choose to use rank orderings and ignore the numerical values.

This might make sense is scenarios when we want a more robust analysis, particular to the distributional assumption for the numerical scores.

# Quantitative outcome models

The situation:

- Suppose we have $J$ competitors, a subset of whom compete in $n$ multiplayer games.
- For example, suppose we observe $J$ athletes who compete in $n$ races, and that we measure the finish time for each race.
- For game $i = 1, \dots, n$, let $y_{ij}$ be the score of player $j$ in game $i$ assuming the player competed.

# Quantitative outcome models

Assuming that the $y_{ij}$ are roughly normally distributed within game $i$, then the model we can assume is

$$y_{ij} = \underbrace{\beta_i}_{\text{Contribution of game } i} + \underbrace{\theta_j}_{\text{Contribution of player } j} + \underbrace{\varepsilon_{ij}}_{\text{Sampling error}}$$

where, independently across players and games,

$$\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$

Thus, the model assumes $\mathrm{E}(y_{ij}) = \beta_i + \theta_j$

# Quantitative outcome models

- This model assumes that the strength of competitor $j$ is represent by $\theta_j$.
- Also, the model assumes that game/race conditions may be distinct from each other, e.g., due to temperature, humidity, wind, or other environmental conditions.
- Thus the model controls for the effect of game $i$ by including $\beta_i$ as an additive term.

Problem: As specified, the model is unidentifiable! Adding the same value to all the $\theta_j$ and subtracting that value to all the $\beta_i$ results in the same mean for $y_{ij}$.

This seems to be a common theme in these linear models!

# Quantitative outcome models

Two solutions:

1. Impose a linear constraint on the $\theta_j$. This can be done by requiring that $\sum_{j=1}^{J} \theta_j = 0$.
2. Perform ridge regularization on the $\theta_j$ (but not on the $\beta_i$).

We will examine each solution briefly, since we have seen them in other contexts already.

# Quantitative outcome models

Suppose four players, $A, B, C$ and $D$, compete in two races.

The results were:

<u>Race 1</u>: $y_{1B} = 4.1, y_{1C} = 3.9, y_{1A} = 2.8, y_{1D} = 2.6$
<u>Race 2</u>: $y_{1B} = 3.8, y_{1A} = 3.5, y_{2C} = 3.3, y_{2D} = 2.7$

We now create two $X$ matrices encoding the players and the races, each with 8 rows $(= 4 \text{ players} \times 2 \text{ races})$.

# Quantitative outcome models

<u>Race 1</u>: $y_{1B} = 4.1, y_{1C} = 3.9, y_{1A} = 2.8, y_{1D} = 2.6$

<u>Race 2</u>: $y_{1B} = 3.8, y_{1A} = 3.5, y_{2C} = 3.3, y_{2D} = 2.7$

$$\boldsymbol{X}_{\boldsymbol{\theta}} = \begin{array}{cl} \begin{matrix} A & B & C & D \end{matrix} & \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \begin{matrix} \text{Race 1, place 1} \\ \text{Race 1, place 2} \\ \text{Race 1, place 3} \\ \text{Race 1, place 4} \\ \text{Race 2, place 1} \\ \text{Race 2, place 2} \\ \text{Race 2, place 3} \\ \text{Race 2, place 4} \end{matrix} \end{array} \qquad \boldsymbol{X}_{\boldsymbol{\beta}} = \begin{array}{cl} \begin{matrix} R1 & R2 \end{matrix} & \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} & \begin{matrix} \text{Race 1, place 1} \\ \text{Race 1, place 2} \\ \text{Race 1, place 3} \\ \text{Race 1, place 4} \\ \text{Race 2, place 1} \\ \text{Race 2, place 2} \\ \text{Race 2, place 3} \\ \text{Race 2, place 4} \end{matrix} \end{array} \qquad \boldsymbol{y} = \begin{bmatrix} 4.1 \\ 3.9 \\ 2.8 \\ 2.6 \\ 3.8 \\ 3.5 \\ 3.3 \\ 2.7 \end{bmatrix}$$

# Quantitative outcome models

Impose a linear constraint on the $\theta_j$ through $\sum_{j=1}^{J} \theta_j = 0$.

Let

$$\boldsymbol{W} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$\boldsymbol{X}_{\boldsymbol{\theta}}^{*} = \boldsymbol{X}_{\boldsymbol{\theta}}\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}$$

# Quantitative outcome models

Impose a linear constraint on the $\theta_j$ through $\sum_{j=1}^{J} \theta_j = 0$.

Now we can fit the model using $\boldsymbol{X}_{\boldsymbol{\theta}}^{*}$ and $\boldsymbol{X}_{\boldsymbol{\beta}}$ as the predictors.

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}_{\boldsymbol{\theta}}^{*}\boldsymbol{\theta}_{-\mathbf{1}} + \boldsymbol{X}_{\boldsymbol{\beta}}\boldsymbol{\beta} =$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ -1 & -1 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \theta_A \\ \theta_B \\ \theta_C \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ -1 & -1 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_A \\ \theta_B \\ \theta_C \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

# Quantitative outcome models

Impose a linear constraint on the $\theta_j$ through $\sum_{j=1}^{J} \theta_j = 0$.

Can now perform least-squares regression to obtain $\widehat{\boldsymbol{\theta}}_{-\mathbf{1}}$ and $\widehat{\boldsymbol{\beta}}$.

Once $\widehat{\boldsymbol{\theta}}_{-\mathbf{1}}$ is determined, can then compute $\widehat{\boldsymbol{\theta}} = \boldsymbol{W}\widehat{\boldsymbol{\theta}}_{-\mathbf{1}}$. See the normal model lecture notes.

```
  Normal.ability.est Normal.ability.stderr
1            -0.1875             0.6123724
2             0.6125             0.6123724
3             0.2625             0.6123724
4            -0.6875             0.6123724
```

# Quantitative outcome models

Perform ridge regularization on the $\theta_j$ (but not on the $\beta_i$).

This basically involves finding the maximum of the penalized log-likelihood, that is

$$\log \tilde{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{y}, \lambda) = \log \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{y}) - \lambda \sum_{j=1}^{J} \theta_j^2$$

where, up to an additive constant,

$$\log \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{y}) \\ = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left(\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{\theta}}\boldsymbol{\theta} - \boldsymbol{X}_{\boldsymbol{\beta}}\boldsymbol{\beta}\right)^T \left(\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{\theta}}\boldsymbol{\theta} - \boldsymbol{X}_{\boldsymbol{\beta}}\boldsymbol{\beta}\right)$$

# Example

```
7 x 1 sparse Matrix of class "dgCMatrix"
                    s0
(Intercept)      .
V1               3.350
V2               3.325
V3              -0.178
V4               0.582
V5               0.250
V6              -0.654
```

Run ridge regression using `glmnet`.

- Use cross-validation to estimate the penalty/tuning parameter, $\lambda$.
- Set `intercept=F`.
- Make sure to prevent the coefficients to the game indicators from being shrunk (use the `penalty.factor` argument)

GLMNET shrinks the coefficients a little bit when applying a ridge penalty.

```
   LM.est GLMNET.est
1 -0.1875    -0.1783
2  0.6125     0.5825
3  0.2625     0.2496
4 -0.6875    -0.6538
```

# Quantitative outcome models: Comments

- Attractive features of this model:
  - Doesn’t require every competitor to participate in every game.
  - No obvious bias if certain competitors choose which games in which to participate.
- Main concern: Normality assumption
  - Should perform residual diagnostics to assess the normality assumption, e.g., residual plots, QQ-plot of residuals to check for normality, etc.
  - Can always try transforming the scores (e.g., log transformation) to improve normality of the residuals.
  - Modern approaches can find a transformation automatically. See Che and Glickman (2024) for such an approach (applied to time-varying abilities).
  - If scores are suspected to follow some other distribution, the approach above can be adapted in a straightforward manner using generalized linear modeling.

# Example: PGA Tour golf game scores

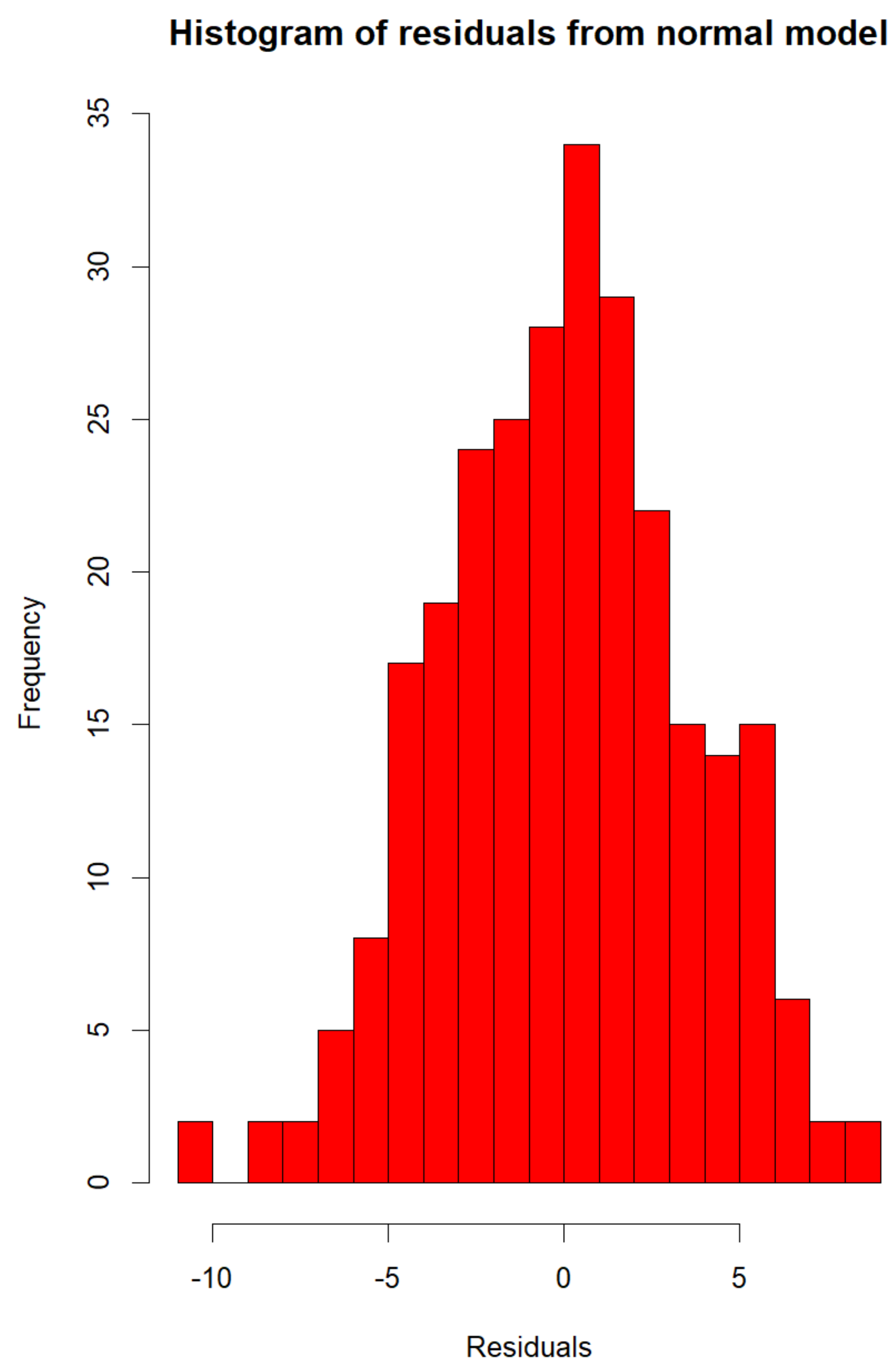
# Example: PGA Tour golf game scores

- Data set includes 15 active golfers who competed in 2019 PGA Tour golf tournaments.
- Recorded performances across 34 golf tournaments in 2019.
- Each golfer participated in 17 to 21 of the events, so fairly balanced participation.
- The score for each event is the total strokes over par, i.e., the number of strokes above what is considered expected for that golf course. Multiplied this by $-1$ in our analyses so that positive values indicate better results.

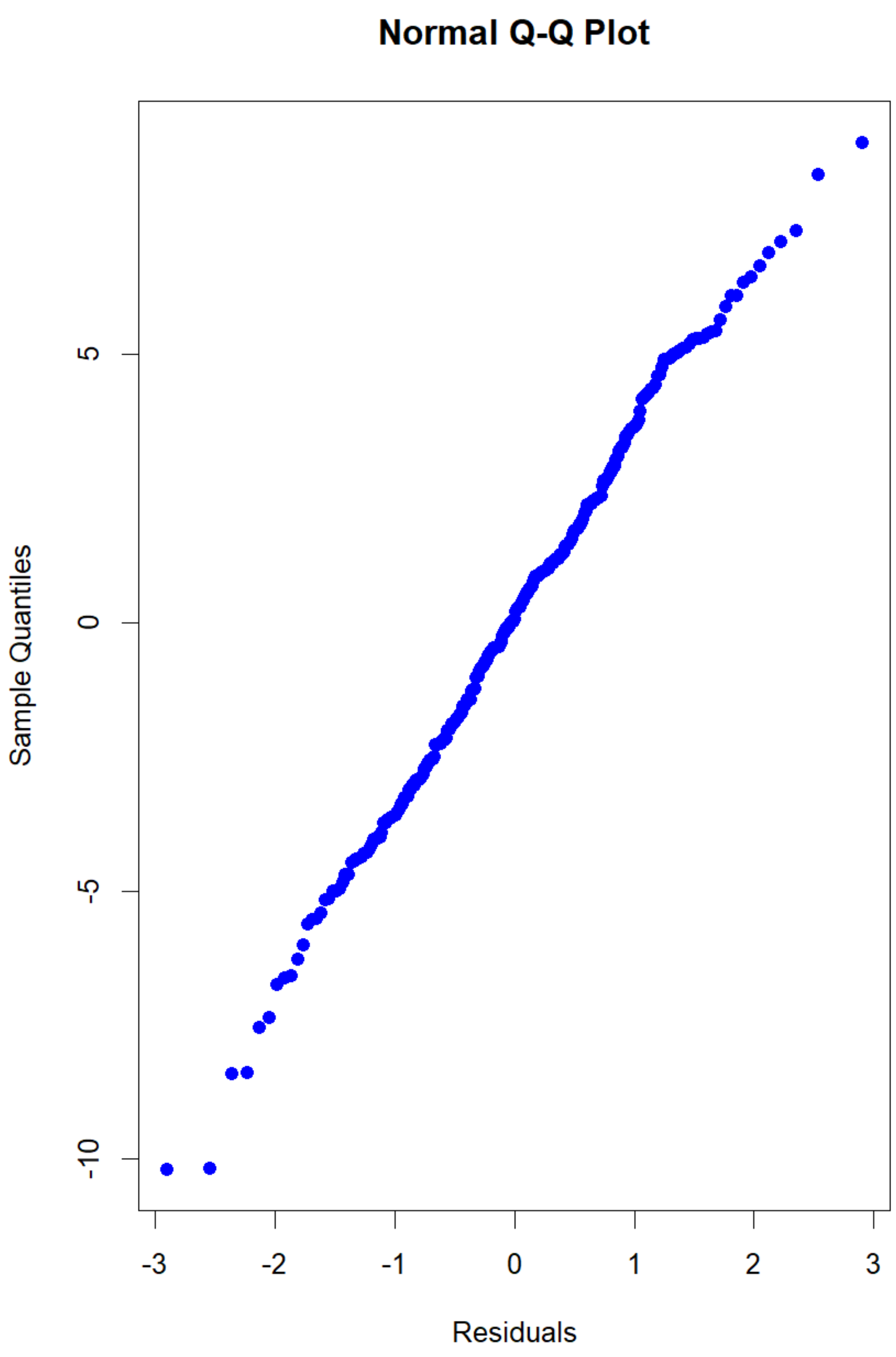
# Example: PGA Tour golf game scores

```
Rory Sabbatini
Lucas Glover
Vaughn Taylor
Brandt Snedeker
Webb Simpson
Billy Horschel
Kevin Kisner
Jason Kokrak
Nick Taylor
Patrick Reed
Hideki Matsuyama
Adam Schenk
J.T. Poston
Joaquin Niemann
Sungjae Im
```

Golfers in the data set



Histogram of residuals from normal model
Frequency
0
5
10
15
20
25
30
35
-10
-5
0
5
Residuals

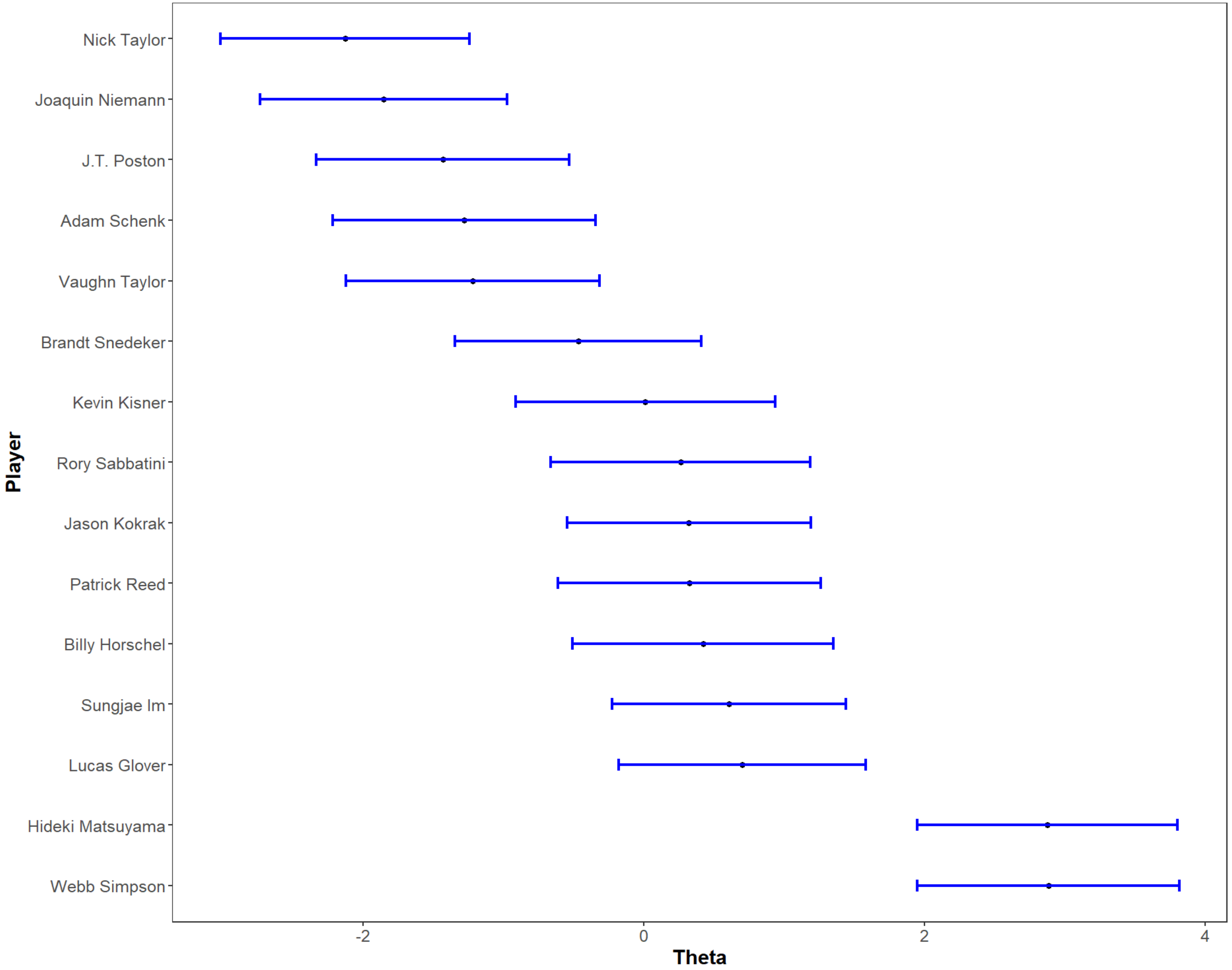



Normal Q-Q Plot
Sample Quantiles
-10
-5
0
5
-3
-2
-1
0
1
2
3
Residuals

# Example: PGA Tour golf game scores

| | Normal.ability.est | Normal.ability.stderr |
|---|---|---|
| Webb Simpson | 2.882 | 0.933 |
| Hideki Matsuyama | 2.874 | 0.929 |
| Lucas Glover | 0.700 | 0.879 |
| Sungjae Im | 0.605 | 0.833 |
| Billy Horschel | 0.420 | 0.929 |
| Patrick Reed | 0.324 | 0.936 |
| Jason Kokrak | 0.320 | 0.869 |
| Rory Sabbatini | 0.260 | 0.924 |
| Kevin Kisner | 0.010 | 0.927 |
| Brandt Snedeker | -0.469 | 0.879 |
| Vaughn Taylor | -1.221 | 0.904 |
| Adam Schenk | -1.282 | 0.937 |
| J.T. Poston | -1.434 | 0.901 |
| Joaquin Niemann | -1.857 | 0.879 |
| Nick Taylor | -2.131 | 0.888 |

# Example: PGA Tour golf game scores




Estimates with 1-SE intervals

# Ranking models

The situation: Same as before, but with one tweak

- Suppose we have $J$ competitors, a subset of whom compete in $n$ multiplayer games.
- For example, suppose we observe $J$ athletes who compete in $n$ races, and that we measure the finish time for each race.
- For game $i = 1, \dots, n$, we <u>do not observe</u> any score for player $j$ in game $i$, or at least we ignore any score measured. Instead, we <u>only observe the rank order</u> for players in game $i$.

# Ranking models: Why ignore scores if they exist?

- Clearly recording scores in a multicompetitor game (e.g., race times) is providing more information than just recording rank orderings.
- Stands to reason that modeling the scores is going to result in more accurate and more precise inferences about player strength, and more accurate predictions.
- However, the key issue in making accurate inferences about strength is that the score model is reliable!
  - If our choice of a model, e.g., normal distribution for scores, is inappropriate, inferences about player strengths may not be reliable.
  - Using rank orderings instead is a <u>robust approach</u> (robust to the choice of a score model). However, there may be some degradation in inferences relative to models for score outcomes.

# Ranking models: Motivating example

Suppose four players, $A, B, C$ and $D$, compete in two races.

The results were:

Race 1: $B \succ C \succ A \succ \mathrm{D}$

Race 2: $B \succ A \succ C \succ D$

These are just the rank orderings from the earlier stylized example, but ignoring the scores.

# Ranking models: Motivating example

Race 1: $B \succ C \succ A \succ \mathrm{D}$

Race 2: $B \succ A \succ C \succ D$

Would like to assume player abilities $\theta_A, \theta_B, \theta_C, \theta_D$ that can be estimated from the race outcomes.

But how?

# Ranking models: Motivating example

Idea:

Decompose a single multicompetitor event into all pairwise comparisons, and then use methods discussed previously (linear paired comparison models, such as Bradley-Terry) to estimate unknown strength parameters.

Race 1: $B \succ C \succ A \succ \mathrm{D}$

Race 2: $B \succ A \succ C \succ D$

would become

Race 1: $(B \succ C), (B \succ A), (B \succ D), (C \succ A), (C \succ D), (A \succ D)$

Race 2: $(B \succ A), (B \succ C), (B \succ D), (A \succ C), (A \succ D), (C \succ D)$

# Ranking models: Motivating example

Problem with this approach:

- Acts as though one multicompetitor game was equivalent to many head-to-head games.

- Analyzing the data this way may give accurate estimates of ability, but the uncertainty of the estimates will be grossly underestimated (that is, the estimates would appear overly precise).

- Furthermore, it is not obvious how to make probability predictions for future rankings.

- Finally, using a paired comparison approach is not necessary, as we shall soon see.

# Ranking models: A generative approach

Suppose players $A, B, C$ and $D$ are about to compete in game $i$.

Assume (unknown) player abilities $\theta_A, \theta_B, \theta_C, \theta_D$.

**Key idea**: For game $i$, let $y_{iA}, y_{iB}, y_{iC},$ and $y_{iD}$ be the latent numerical performances by the four players, where, independently,

$$\begin{gathered} y_{iA} \sim F(\theta_A) \\ y_{iB} \sim F(\theta_B) \\ y_{iC} \sim F(\theta_C) \\ y_{iD} \sim F(\theta_D) \end{gathered}$$

for some distribution $F$ that depends on the given parameter.

# Ranking models: A generative approach

Suppose that $y_{iB} > y_{iC} > y_{iA} > y_{iD}$. Then the <u>observed outcome</u> would be the rank ordering

$$B \succ C \succ A \succ D$$

Thus, if we knew the strength parameters, then to generate a particular rank ordering we need to

1. simulate latent numerical scores, and then
2. rank order the scores.

Of course, we never see the latent numerical scores.

This approach to modeling rank orderings is sometimes called the Thurstonian model for rankings.

# Ranking models: A generative approach

The probability of the rank ordering $B \succ C \succ A \succ D$, given the strength parameters, is given by

$$\Pr(B \succ C \succ A \succ D) = \Pr(y_{iB} > y_{iC} > y_{iA} > y_{iD} | \theta_A, \theta_B, \theta_C, \theta_D) = \int\int\int\int_{y_{iB} > y_{iC} > y_{iA} > y_{iD}} f(y_A|\theta_A)\, f(y_B|\theta_B) f(y_C|\theta_C) f(y_D|\theta_D)\, dy_A dy_B dy_C dy_D$$

where the $f$'s are the densities for the latent performance distributions.

Is this multiple integral feasible to calculate?

# Plackett-Luce model

Let the density for the latent performance distribution be

$$f(y|\theta) = \exp\left(-\left(y - \theta + e^{-(y-\theta)}\right)\right)$$

that is, the Gumbel distribution with location parameter $\theta$.

In other words, for our example we assume, independently,

$$f(y_A|\theta_A) = \exp\left(-\left(y_A - \theta_A + e^{-(y_A-\theta_A)}\right)\right)$$
$$\vdots$$
$$f(y_D|\theta_D) = \exp\left(-\left(y_D - \theta_D + e^{-(y_D-\theta_D)}\right)\right)$$

# Plackett-Luce model

This distributional assumption leads to

$\Pr(B \succ C \succ A \succ D) = \Pr(y_{iB} > y_{iC} > y_{iA} > y_{iD} | \theta_A, \theta_B, \theta_C, \theta_D) =$

$$\left(\frac{e^{\theta_B}}{e^{\theta_B}+e^{\theta_C}+e^{\theta_A}+e^{\theta_D}}\right)\left(\frac{e^{\theta_C}}{e^{\theta_C}+e^{\theta_A}+e^{\theta_D}}\right)\left(\frac{e^{\theta_A}}{e^{\theta_A}+e^{\theta_D}}\right) =$$

$$\Pr(B \text{ outperforms } C, A, D) \Pr(C \text{ outperforms } A, D) \Pr(A \text{ outperforms } D)$$

Important to note that this result relies on the non-trivial result that

$$\Pr(C \text{ outperforms } A, D) = \Pr(C \text{ outperforms } A, D | B \text{ performs best})$$

$$\Pr(A \text{ outperforms } D) = \Pr(A \text{ outperforms } D | B, C \text{ outperform } A \text{ and } D)$$

# Plackett-Luce model

In general, consider $J$ players with strength parameters $\theta_1, \theta_2, \ldots, \theta_J$.

Let $\{j_1, j_2, \ldots, j_M\} \subset \{1,2,\ldots,J\}$ be a permutation of the set of $M$ players where player $j_1$ comes in first place, player $j_2$ comes in second place, etc.

Then the <u>Plackett-Luce model</u> (also called the rank-order logit model) asserts

$$\Pr(j_1 \succ j_2 \succ \cdots \succ j_M) = \prod_{m=1}^{M} \frac{e^{\theta_{j_m}}}{\sum_{\ell=m}^{M} e^{\theta_{j_\ell}}}$$

When $M = 2$, the model becomes

$$\Pr(j_1 \succ j_2) = \prod_{m=1}^{2} \frac{e^{\theta_{j_m}}}{\sum_{\ell=m}^{2} e^{\theta_{j_\ell}}} = \frac{e^{\theta_{j_1}}}{e^{\theta_{j_1}} + e^{\theta_{j_2}}}$$

which is the Bradley-Terry model! So the BT model is a special case of PL.

# Plackett-Luce model: Comments

- The Plackett-Luce model satisfies Luce's "choice" axiom. This says that the relative rankings of a subset of players is independent of other players not in the rankings. That is, for example,
$$\Pr(B \succ C) \ = \Pr(B \succ C \,|A, D \text{ outperform } B, C)$$
- In fact, Yellott (1977) showed that the Gumbel distribution is the only distribution in a Thurstonian model for rankings that satisfies Luce's choice axiom.
- Gumbel distribution assumption also has the advantage of corresponding to probabilities of ranking orderings that can be written down in closed form.
- Assuming other distributions for performance distributions, e.g., normal, is much more complicated and does not correspond to closed-form expressions for the probability of a rank ordering.

# Plackett-Luce model: Covariates

The Plackett-Luce model can be easily adapted to include covariates.

Let $\boldsymbol{x}_{j_m}$ be a vector of (endogenous) covariates for competitor $j_m$ and let $\boldsymbol{u}$ be a vector of (exogenous) game-specific covariates. Then letting $\boldsymbol{\beta}$ be the associated vector of endogenous covariate effects, and letting $\boldsymbol{\gamma}_{j_m}$ be a vector of exogenous covariate effects, the modified Plackett-Luce model is given by

$$\Pr(j_1 \succ j_2 \succ \cdots \succ j_M) = \prod_{m=1}^{M} \frac{e^{\theta_{j_m} + \boldsymbol{x}_{j_m}{}^{T}\boldsymbol{\beta} + \boldsymbol{u}^{T}\boldsymbol{\gamma}_{j_m}}}{\sum_{\ell=m}^{M} e^{\theta_{j_\ell} + \boldsymbol{x}_{j_\ell}\boldsymbol{\beta}\boldsymbol{u}^{T}\boldsymbol{\gamma}_{j_\ell}}}$$

As with paired comparison models, it is important to distinguish between endogenous and exogenous covariates when estimating strength.

# Plackett-Luce model: Inference

For game $i$, the contribution to the likelihood is

$$\Pr\left(j_{i1} \succ j_{i2} \succ \cdots \succ j_{iM_i}\right) = \prod_{m=1}^{M_i} \frac{e^{\theta_{ij_m}}}{\sum_{\ell=m}^{M} e^{\theta_{ij_\ell}}}$$

so that the log-likelihood is

$$\log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \log \Pr\left(j_{i1} \succ j_{i2} \succ \cdots \succ j_{iM_i}\right) =$$

$$\sum_{i=1}^{n} \sum_{m=1}^{M_i} \left( \theta_{ij_m} - \log \sum_{\ell=m}^{M} e^{\theta_{ij_\ell}} \right)$$

# Plackett-Luce model: Inference

- Can perform maximum likelihood, but need to impose the usual linear constraint on the $\theta_j$ because the $\theta_j$ are unidentifiable otherwise.
- Can also perform ridge-penalized maximum likelihood. This involves maximizing

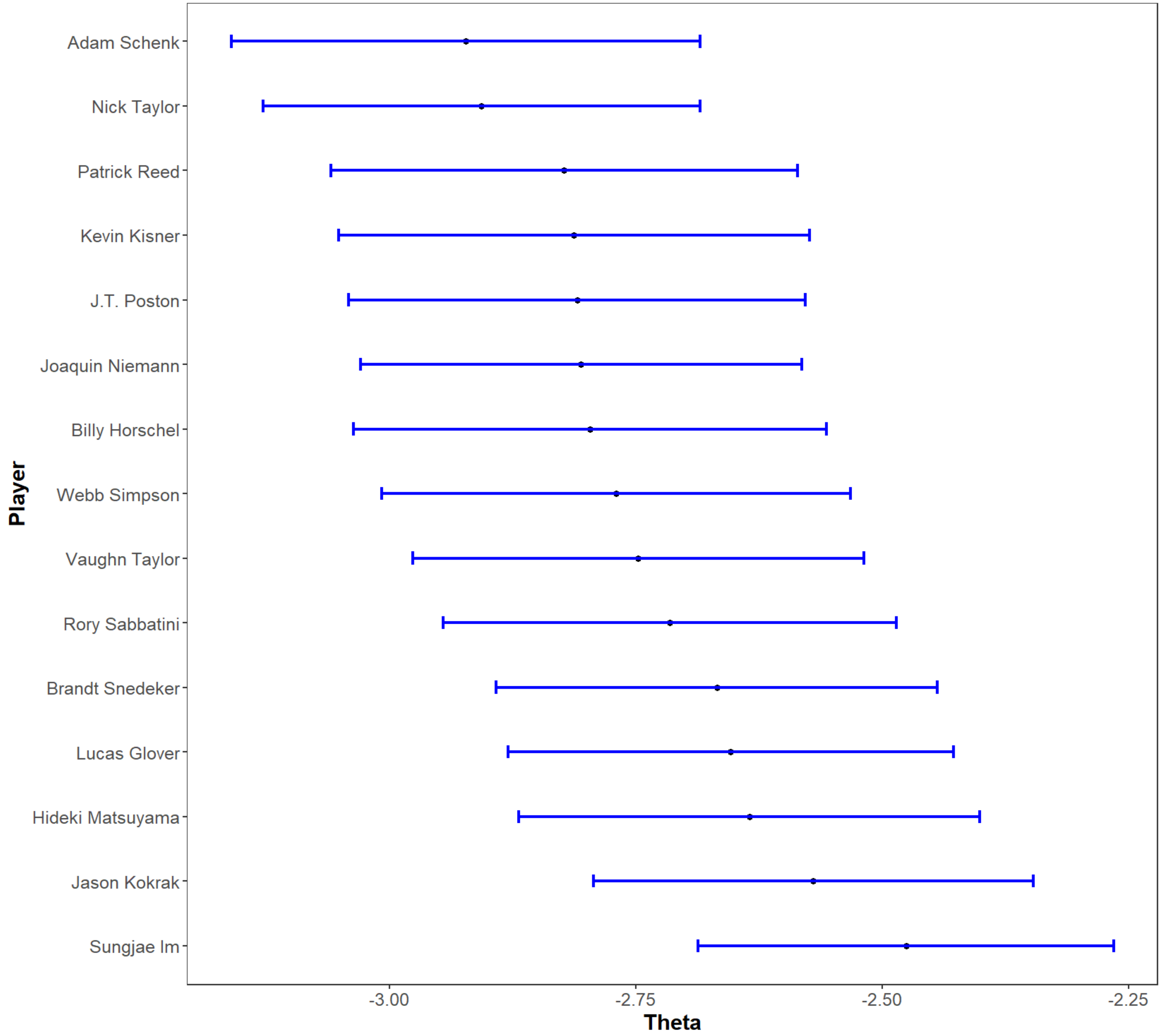
$$\log \tilde{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{y},\lambda) = \sum_{i=1}^{n} \sum_{m=1}^{M_i} \left( \theta_{i j_m} - \log \sum_{\ell=m}^{M} e^{\theta_{i j_\ell}} \right) - \lambda \sum_{j=1}^{J} \theta_j^2$$

- Can perform Bayesian inference with independent normal prior densities on the $\theta_j$, that is $\theta_j \sim N(0, \sigma^2)$.

# Example: PGA Tour golf game scores

| | PlackettLuce.est | PlackettLuce.stderr |
|---|---|---|
| Sungjae Im | -2.476 | 0.211 |
| Jason Kokrak | -2.570 | 0.223 |
| Hideki Matsuyama | -2.635 | 0.234 |
| Lucas Glover | -2.654 | 0.226 |
| Brandt Snedeker | -2.668 | 0.224 |
| Rory Sabbatini | -2.716 | 0.230 |
| Vaughn Taylor | -2.748 | 0.229 |
| Webb Simpson | -2.770 | 0.238 |
| Billy Horschel | -2.797 | 0.240 |
| Joaquin Niemann | -2.806 | 0.224 |
| J.T. Poston | -2.810 | 0.232 |
| Kevin Kisner | -2.813 | 0.239 |
| Patrick Reed | -2.823 | 0.237 |
| Nick Taylor | -2.907 | 0.222 |
| Adam Schenk | -2.923 | 0.238 |

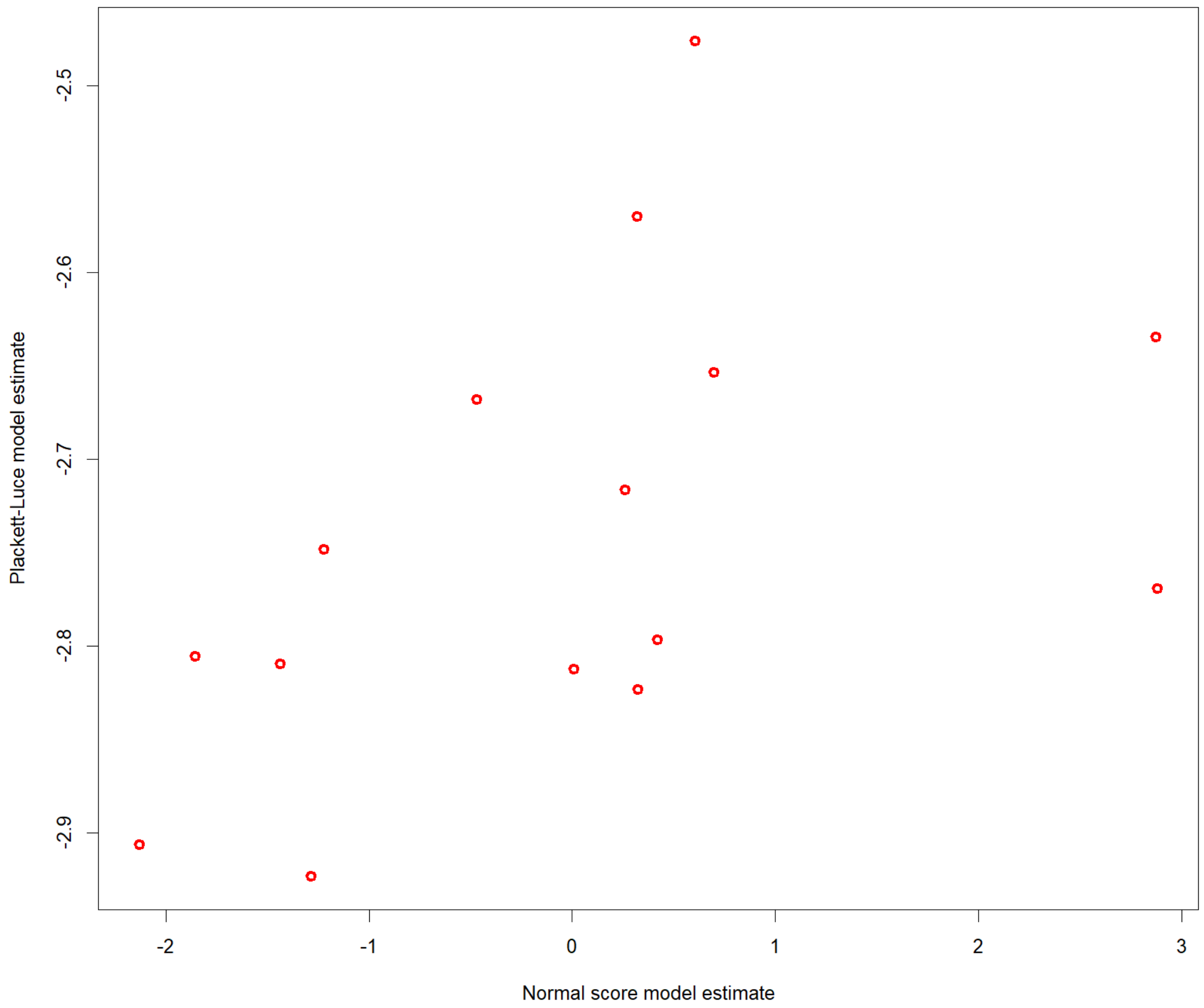
# Example: PGA Tour golf game scores




Estimates with 1-SE intervals

Overlap of intervals for Plackett-Luce model compared to normal model.

# Example: PGA Tour golf game scores




Relationship between normal estimates of ability and Plackett-Luce estimates.

Correlation of estimates is 0.482. Not surprising given the uncertainty in the Plackett-Luce estimates.

# References:


Glickman, M. E., & Hennessy, J. (2015). A stochastic rank ordered logit model for rating multi-competitor games and sports. Journal of Quantitative Analysis in Sports, 11(3), 131-144.

Luce, R. D. (1959). Individual Choice Behavior: A Theoretical Analysis. New York: Wiley.

Plackett, R. L. (1975). “The Analysis of Permutations.” Appl. Statist 24 (2): 193–202.

Yellott Jr, J. I. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. Journal of Mathematical Psychology, 15(2), 109-144.