Multivariable Linear Regression

Q create power scores for college basketball teams which take into account strength of schedule, score diff., home court.

Variables i index of i^{th} game in the dataset

Teams are $\{1, \dots, N\}$

Y_i = score differential of game i (observed)

= Home team score − Away Team Score

$\beta_{H(i)}$ = (unknown) power score of the Home Team in game i

$\beta_{A(i)}$ = power score Away team in game i

data generating process

how is the data generated?

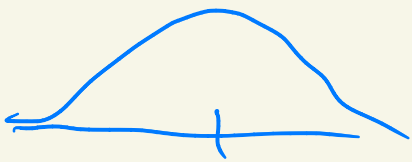
- schedule
- score differential

Model

$$\boxed{Y_i = \beta_0 + \beta_{H(i)} - \beta_{A(i)} + \varepsilon_i}$$

ε_i is mean-zero noise, $\mathbb{E}\varepsilon_i = 0$

$$\mathbb{E}Y_i = \underbrace{\beta_0}_{\text{home court advantage}} + (\underset{90}{\beta_{H(i)}} - \underset{80}{\beta_{A(i)}})$$



Teams $1, \dots, N$

game i, $H(i)$ = home team index

One β (power score) for each team

$\beta_1, \dots, \beta_N$

<u>ex</u> if team 2 @ team 7 in game i

then $\beta_{H(i)} = \beta_7$, $\beta_{A(i)} = \beta_2$

$$\left(\beta_{H(i)} - \cancel{\text{Avg } \beta_{H \, \text{score}}}^{\,C_1}\right) - \left(\beta_{A(i)} - \cancel{\text{Avg } \beta_{A \, \text{score}}}^{\,C_2}\right)$$

observed difference in pts

(unobserved) difference in strength

$$\beta_{H(i)} / \underbrace{\left(\text{Avg } \beta_{H \, \text{score}}\right)}_{C_3} - \beta_{A(i)} / \left(\text{Avg } \beta_{A \, \text{score}}\right)$$

$$1: \; Y_i = \beta_0 + \left(\beta_{H(i)} - C_1\right) + \left(\beta_{A(i)} - C_2\right)$$

$$= \left(\beta_0 + C_1 + C_2\right) + \beta_{H(i)} - \beta_{A(i)}$$

$$2: \; Y_i = \beta_0 + \beta_{H(i)} - \beta_{A(i)}$$

model 2 off. and def. power scores

game i

points scored by Home Team S_i

points allowed by Home Team Y_i

Home Tm off power score $\beta_{H(i)}$

def power score $\alpha_{H(i)}$

Away Tm off power score $\beta_{A(i)}$
def power score $\alpha_{A(i)}$

$$
\begin{cases}
S_i = \beta_0 + \beta_{H(i)} - \alpha_{A(i)} + \varepsilon_i \\
Y_i = \beta_1 + \beta_{A(i)} - \alpha_{H(i)} + \varepsilon_i
\end{cases}
$$

$\mathbb{E}\varepsilon_i = 0$

Lakers = Home
Warriors = Away

Lakers offense vs. Warriors defense $\longrightarrow$ points scored by Lakers

Lakers defense vs. Warriors offense $\longrightarrow$ points allowed by Lakers = Points scored Warriors

$$\boxed{Y_i = \beta_0 + \beta_{H(i)} - \beta_{A(i)} + \varepsilon_i}$$

β's are unknown $\rightarrow$ need to be estimated

Y_i known

schedule $H(i)$, $A(i)$ known

$\downarrow$

data

Season	WLoc	WTeamName	LTeamName	ScoreDiff	WScore	LScore
<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
2023	H	DePaul	Loyola MD	6	72	66
2023	H	Duke	Jacksonville	27	71	44
2023	A	Evansville	Miami OH	-4	78	74
2023	A	FL Gulf Coast	USC	-13	74	61
2023	H	Florida	Stony Brook	36	81	45
2023	H	Florida Intl	Houston Chr	11	77	66

$$\left[\begin{aligned}
Y_1 &= \beta_0 + \beta^{1}_{DePaul} - \beta^{2}_{Loyola} + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta^{3}_{Duke} - \beta^{4}_{Jack} + \varepsilon_2 \\
Y_3 &= \beta_0 + \beta^{5}_{Miami} - \beta^{6}_{Evansville} + \varepsilon_3
\end{aligned}\right.$$

$$\begin{bmatrix} \quad & \vdots & \quad \end{bmatrix}$$

Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

Before: Teams $1, \dots, N$, $\beta_1, \dots, \beta_N$

In Matrix-Vector Form

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \end{bmatrix}}_{\substack{y \\ \text{(data)}}} = \underbrace{\begin{bmatrix} x_{10}, x_{11}, x_{12}, x_{13}, \dots \\ \\ \\ \\ \end{bmatrix}}_{\substack{X \\ \text{(data)}}} \cdot \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix}}_{\substack{\beta \\ \text{Unknown} \\ \text{parameters} \\ \text{to be} \\ \text{estimated}}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \end{bmatrix}}_{\substack{\varepsilon \\ \text{noise}}}$$

$$Y_1 = \left(\beta_0 + \overset{1}{\beta_{DePaul}} - \overset{2}{\beta_{Loyola}}\right) + \varepsilon_1$$

$$x_{10}\cdot\beta_0 + x_{11}\cdot\beta_1 + x_{12}\cdot\beta_2 + x_{13}\cdot\beta_3 + \dots + x_{1N}\beta_N$$

$$= \beta_0 + \beta_1 - \beta_2$$

$$x_{10}=1,\quad x_{11}=1,\quad x_{12}=-1,\quad x_{1,\text{other}}=0$$

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ \end{bmatrix}}_{y} = \underbrace{\overset{(\text{num games})\times(N+1)}{\begin{bmatrix} 1 & 1 & -1 & 0 & \cdots & 0 & \\ 1 & 0 & 0 & 1 & -1 & 0 & \cdots 0 \\ 1 & & & & & & \\ 1 & & & & & & \\ \vdots & & & & & & \end{bmatrix}}}_{X} \cdot \underbrace{\overset{(N+1)\times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix}}}_{\beta} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \end{bmatrix}}_{\varepsilon}$$

y (data)

score diff

X (data)

Schedule matrix

β Unknown parameters to be estimated

power scores

ε noise

noise/randomness

$$\boxed{y = X\beta + \varepsilon}$$

$$\underbrace{y}_{(\text{num games})\times 1} \quad \underbrace{X}_{(\text{num games})\times(N+1)} \quad \underbrace{\beta}_{N+1} \quad \underbrace{\varepsilon}_{(\text{num games})\times 1}$$

observed data: y, X

estimate: β

Simple linear regression $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

X_{ij} = value of the X matrix at Row i (game i) and Column j $\left(\begin{array}{l}\text{if } j=0 \rightarrow \text{intercept cl}m \\ \text{if } j\neq 0 \rightarrow \text{team } j\end{array}\right)$

$$= \begin{cases} \text{if } j=0 & X_{ij}=1 \\ \text{if } j=H(i) & X_{ij}=1 \\ \text{if } j=A(i) & X_{ij}=-1 \\ \text{else} & X_{ij}=0 \end{cases}$$

$$y = X\beta + \varepsilon$$

$$\mathbb{E}y = X\beta$$

```
> df_ncaamb2[1:5,]
# A tibble: 5 x 7
  Season WTeamName   LTeamName   WScore LScore WLoc  ScoreDiff
   <dbl> <chr>       <chr>        <dbl>  <dbl> <chr>     <dbl>
1   2023 Abilene Chr Jackson St      65     56 H             9
2   2023 Akron       S Dakota St     81     80 H             1
3   2023 Alabama     Longwood        75     54 H            21
4   2023 Arizona     Nicholls St    117     75 H            42
5   2023 Arizona St  Tarleton St     62     59 H             3
> X[1:5,c(1:5,131)]
     (Intercept) Abilene Chr Air Force Akron Alabama Jackson St
[1,]           1           1         0     0       0         -1
[2,]           1           0         0     1       0          0
[3,]           1           0         0     0       1          0
[4,]           1           0         0     0       0          0
[5,]           1           0         0     0       0          0
```

$$\boxed{y = X\beta + \varepsilon}$$

Solve for β.
estimate β.

Least Squares: Minimize Residual Sum of Squares
(~~Mean~~ Sum Squared Error)

SLR: $RSS(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - \underbrace{(\beta_0 + \beta_1 x_i)}_{\hat{y}_i}\right)^2$

Now:

$$RSS(\beta) = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$$

$$= \sum_{i=1}^{n} \left(y_i - \begin{pmatrix} i\text{th Row} \\ \text{of } X \end{pmatrix} \cdot \beta\right)^2$$

$$\begin{cases} x_i = i^{th} \text{ column of } X \\ x_i^T = i^{th} \text{ Row of } X \\ x_i^T\beta = \beta_0 x_{10} + \beta_1 x_{11} + \cdots + \beta_N x_{1N} \end{cases}$$

$$= \sum_{i=1}^{n} \underbrace{\left(y_i - x_i^T\beta\right)^2}_{a_i^2}$$

$$= \underbrace{(y - X\beta)^T \cdot (y - X\beta)}_{a^T a}$$

<u>Fact</u> $\sum_{i=1}^{n} a_i^2 = a^T a$

<u>Pf</u>

$$a^T a = \underset{1\times n}{[a_1\ a_2 \cdots a_n]} \underset{n\times 1}{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}} = a_1 \cdot a_1 + a_2 \cdot a_2 + \cdots + a_n \cdot a_n$$

$$= a_1^2 + a_2^2 + \cdots + a_n^2$$

$$= \sum_{i=1}^{n} a_i^2 .$$

$$= (y - X\beta)^T (y - X\beta)$$

$$= \left(y^T - (X\beta)^T\right)(y - X\beta)$$

$$= y^T y - y^T (X\beta) - (X\beta)^T y - (X\beta)^T (X\beta)$$

$$= y^T y - 2\underbrace{(X\beta)}_{a}{}^T \underbrace{y}_{b} + (X\beta)^T (X\beta)$$

because $a^T b = b^T a = a_1 b_1 + \dots + a_n b_n$

$$RSS(\beta) = y^T y - 2\beta^T X^T y - \beta^T X^T X \beta$$

because $(X\beta)^T = \beta^T \cancel{X}^T$

do this yourself later

SLR: $RSS'(\beta) = 0$ and solve for β

Now: $\overset{\text{Multivariable}}{\text{Calculus}}$ $\longrightarrow$ Gradient

$$\nabla_\beta RSS(\beta) = 0 \quad \text{and solve}$$

$$\left[\frac{d}{d\beta_1} RSS(\beta_1)\right]$$

$$\nabla_\beta RSS(\beta) =$$

$$\nabla_\beta \left(\cancelto{0}{y^T y} - 2\beta^T X^T y - \beta^T X^T X \beta \right)$$

* first term: $-2\beta^T \underbrace{X^T y}_{a} \longrightarrow \beta^T a$

$$\nabla_\beta(\beta^T a) = \left(\frac{\partial}{\partial \beta_1} \beta^T a, \frac{\partial}{\partial \beta_2} \beta^T a, \dots, \frac{\partial}{\partial \beta_n} \beta^T a \right)$$

$$\beta^T a = \beta_1 a_1 + \dots + \beta_n a_n$$

$$= (a_1, a_2, \dots, a_n) = a$$

$$\nabla_\beta\left(-2\beta^T X^T y\right) = -2\left(X^T y\right)$$

✱ Second term: $-\beta^T X^T X\beta \rightarrow \underset{1\times n}{\beta^T}\ \underset{n\times n}{A}\ \underset{n\times 1}{\beta}$

$$\nabla_\beta\left(\beta^T A\beta\right) = \left(\frac{\partial}{\partial\beta_1}\beta^T A\beta, \ldots, \frac{\partial}{\partial\beta_n}\beta^T A\beta\right)$$

$$\beta^T \underbrace{\overset{n\times n}{A}\ \overset{n\times 1}{\beta}}_{n\times 1} =$$

$A = X^T X$ symmetric $A_{ij} = A_{ji}$

$$\underbrace{\begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_n\end{bmatrix}}_{1\times n}\underbrace{\begin{bmatrix}a_{11}\beta_1 + a_{12}\beta_2 + \cdots + a_{1n}\beta_n \\ a_{21}\beta_1 + a_{22}\beta_2 + \cdots + a_{2n}\beta_n \\ \vdots \\ a_{n1}\beta_1 + a_{n2}\beta_2 + \cdots + a_{nn}\beta_n\end{bmatrix}}_{n\times 1}$$

$$= \beta_1\left(a_{11}\beta_1 + a_{12}\beta_2 + \cdots + a_{1n}\beta_n\right)$$
$$+ \beta_2\left(a_{21}\beta_1 + a_{22}\beta_2 + \cdots + a_{2n}\beta_n\right)$$

$+ \cdots$

$+ \beta_n (a_{n1} \beta_1 + a_{n2} \beta_2 + \cdots + a_{nn} \beta_n)$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j a_{ij} = \beta^T A \beta$$

$$\nabla_\beta (\beta^T A \beta) = \left(\frac{\partial}{\partial \beta_1} (\beta^T A \beta), \ldots, \frac{\partial}{\partial \beta_n} (\beta^T A \beta) \right)$$

$$= \left(\frac{\partial}{\partial \beta_1} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j a_{ij} \right), \cdots \right)$$

$$= \left(\frac{\partial}{\partial \beta_1} \left(\begin{array}{l} \beta_1^2 a_{11} \\ + 2\beta_1 \beta_2 a_{12} \\ + 2\beta_1 \beta_3 a_{13} \\ + \cdots \\ + 2\beta_1 \beta_n a_{1n} \end{array} \right), \cdots \right)$$

$$= \begin{pmatrix} 2\beta_1 a_{11} + 2\beta_2 a_{12} + \cdots + 2\beta_n a_{1n} \\ \vdots \end{pmatrix}$$

$$= \Big(\underbrace{2\sum_{j=1}^{n} \beta_j a_{1j}}_{\beta_1 a_{11} + \beta_2 a_{12} + \cdots + \beta_n a_{1n}},\ 2\sum_{j=1}^{n} \beta_j a_{2j},\ \cdots,\ 2\sum_{j=1}^{n} \beta_j a_{nj} \Big)$$

$\partial / \partial \beta_1$ $\qquad$ $\partial / \partial \beta_2$ $\qquad$ $\partial / \partial \beta_n$

$$= \left(2\beta^T A_1,\ 2\beta^T A_2,\ \ldots,\ 2\beta^T A_n \right)$$

$$= 2\beta^T \left(A_1, A_2, \ldots, A_n \right)$$

$$= 2\beta^T A$$

<u>Fact</u> $\quad \nabla_\beta (\beta^T A \beta) = 2\beta^T A$ if A symmetric

therefore $D_\beta(+\beta^T X^T X \beta) = +2\beta^T X^T X$

therefore

$$D_\beta RSS(\beta) =$$

$$= D_\beta\left(\cancel{y^T y}^{\,0} - 2\beta^T X^T y + \beta^T X^T X \beta\right)$$

$$= -2X^T y + 2\beta^T X^T X = 0$$

$$\Longrightarrow \quad \beta^T X^T X = X^T y$$

$$\Longrightarrow \quad \underbrace{(X^T X)}_{\substack{n\times n \\ \text{symmetric} \\ \text{matrix}}} \cdot \beta = X^T y$$

$$\Longrightarrow \quad \boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

$$\boxed{y = X\beta + \varepsilon}$$

$$y = X\beta \quad , \quad X \text{ not square}$$

$$X^T y = X^T X \beta$$

$$(X^T X)^{-1} X^T y = \hat{\beta}$$

X : known matrix (data)

y : known vector (data)

$\hat{\beta}$ = estimated power scores $= (X^T X)^{-1} X^T y$

Takeaways

- English → Math
- Q: wanting power scores
- variables
- model
- estimation: multivariable linear regression
- actually find the formula to estimate the power scores
- get data, code it up, view results

SLR: $\hat{\beta}_1 = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \approx \dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}$

Now: $\hat{\beta} = (X^TX)^{-1}X^Ty \approx \dfrac{X^Ty}{(X^TX)} \approx \dfrac{\mathrm{cov}(X,y)}{\mathrm{var}(X)}$