Baseball III

Produced by Dr. Mario | UNC STOR 538

Linear Weights

- **Multiple Linear Regression**

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon$$

Linear Weights

S = Single
D = Double
T = Triple
HR = Home Run
BB = Walk
HBP = Hit-by-Pitch
SB = Stolen Base
CS = Caught Stealing

- **Baseball Application**
  - $Y = Runs\ for\ the\ Season$
  - $\vec{X} = [BB + HBP, S, D, T, HR, SB, CS]'$
  - $Y = \vec{X}'\vec{\beta} + \vec{\epsilon}$
  - $\hat{Y} = Predicted\ Runs$
  - $\hat{Y} = \vec{X}'\hat{\vec{\beta}}$

Linear Weights

- **Estimated Linear Weights Using Least Squares**

Predictor	Estimate
Constant	-411.81
Single	0.46
Double	0.81
Triple	1.07
HR	1.43
BB+HBP	0.33
SB	0.25
CS	-0.25

$$n = 210$$
$$R^2 = 0.90$$
$$Adj.\,R^2 = 0.90$$

Used to Be Insignificant

Doesn't Add Marginal Value

Linear Weights

- Important Information From Linear Regression

	Coefficients	Standard Error	t Stat	P–value
Intercept	–411.8133561	33.00675506	–12.47663866	7.3423E–27
BB+HBP	0.326171191	0.026991877	12.08405016	1.1813E–25
1B	0.459107774	0.028209869	16.2747222	1.325E–38
2B	0.805141015	0.070539419	11.41405797	1.31E–23
3B	1.072129559	0.185083303	5.792686554	2.6244E–08
HR	1.428105264	0.052270693	27.32133795	9.1608E–70
SB	0.250044999	0.063490957	3.938277396	0.00011296
CS	–0.254380304	0.190576335	–1.334794818	0.18344599

Linear Weights

- **Important Information From Linear Regression**
  - **Removal of Insignificant Variables**

	Coefficients	Standard Error	t Stat	P–value
Intercept	–422.3214856	32.11582993	–13.14994775	5.654E–29
BB+HBP	0.328427033	0.026990732	12.16814092	6.1158E–26
1B	0.462425312	0.028154216	16.4247273	3.9961E–39
2B	0.809004928	0.070615562	11.45646795	9.2244E–24
3B	1.056646807	0.185074775	5.709296723	3.9868E–08
HR	1.432093994	0.052285581	27.38984579	4.1936E–70
SB	0.204454976	0.05362427	3.812732098	0.00018226

- $MAD = 17.15$ **(Now) vs.** $MAD = 26$ **(Bill James)**

Linear Weights

- **Historical Progression**

Event	1916 Lane	1950-1960 Lindsay	1978 Palmer	1989 Boswell	Now
BB + HBP	.164	—	.33	.33	0.33
Single	.457	.41	.46	.47	0.46
2B	.786	.82	.8	.78	0.81
Triple	1.15	1.06	1.02	1.09	1.06
Home Runs	1.55	1.42	1.4	1.4	1.43
Outs	—	—	– .25	—	-
SB	—	—	.3	.3	0.20
CS	—	—	– .6	—	-

Linear Weights

- Evaluation of Hitters
  - Imagine if Team Had Only Mike Trout (2016)
  - Approximately,

$$26.72 \times 162 = 4329\ Outs\ Per\ Season$$

  - Trout Hit 29 HR and Had 366.118 Outs
  - Therefore, Trout Hit

$$\frac{29}{366.118} = 0.079\ Home\ Runs\ Per\ Out$$

  - Scaling Up, We Expect a Team of Trouts to Hit on Average

$$4329 \times \frac{29}{366.118} = 342.9\ Home\ Runs\ Per\ Season$$

  - Using Linear Weights, We Expect 1,588.07 Runs Per Season which Can Be Thought of 9.80 Runs Per Game

Linear Weights

- OBP, SLG, OPS, and Runs Created
  - *Moneyball* Highlights the Importance of OBP
  - From 2010-2016, Average OBP was 32%
  - Purpose of OPS = Value Power Hitters
  - Recall:

$$OPS = OBP + SLG$$
$$= 1 \times OBP + 1 \times SLG$$

Equal Weights

- Which Covariate (OBP or SLG) is Better for Predicting Runs?

Linear Weights

- **OBP, SLG, OPS, and Runs Created**
  - **Multiple Regression (2010-2016 Team Data)**

$$Runs = \beta_0 + \beta_1(SLG) + \beta_2(OBP) + \epsilon$$

	Coefficients	Standard Error	t Stat	P–value
Intercept	–738.7520251	43.82154709	–16.85819133	1.04367E–40
OBP	2338.121668	191.8515917	12.18713719	4.14782E–26
SLG	1707.332494	92.94672979	18.3689356	2.39874E–45

$$n = 210 \;\&\; R^2 = 0.89 \;\&\; Adj.R^2 = 0.88$$

- **Summary: OBP is More Important Than SLG (1.4 Times More)**

Linear Weights

- Runs Created Above Average
  - How Many More Runs if Average Team Added a Player?
  - Average Team (2010-2016) Versus Bryant (2016)

Hit Type	Average Team	Bryant 2016
Single	939.83	99
Double	276.2	35
Triple	29.16	3
HR	159.36	39
BB+HBP	544.59	93
SB	95.08	8
Outs	4328.64	416.15

Linear Weights

- Runs Created Above Average

Hit Type	Average Team	Bryant	Bryant + Team
Single	939.83	99	948.48
Double	276.2	35	284.64
Triple	29.16	3	29.36
HR	159.36	39	183.04
BB+HBP	544.59	93	597.33
SB	95.08	8	93.94

Linear Weights

- **Runs Created Above Average**
  - **If Added, Rest of Players Will Cost an Approximate**

    $$4328.64 - 416.15 = 3912.49\ Outs$$

  - **For the Rest of The Team, This is Equivalent to**

    $$\frac{3912.49}{4328.64} = 90.4\%\ of\ Total\ Outs$$

  - **Singles With Bryant Added to Roster**

    $$Singles = 0.904(Singles\ of\ Team) + (Singles\ of\ Bryant)$$
    $$= 0.904(939.83) + (99) = 948.61$$

Linear Weights

- Runs Created Above Average
  - Predicted Runs of Average Team = 693.02
  - Predicted Runs of Bryant+Average Team = 751.08
  - Added Value of Bryant = 751.08-693.02 = 58 Runs Above Average

Final Inspiration

If you don’t like sports,
you may like baseball.

- Mahatma Mario