Mohnish Gandhi
Emily Mi

Stats 371H 9:30 – 11:00 AM Class Notes

The main concept of today's class was an Interaction Term

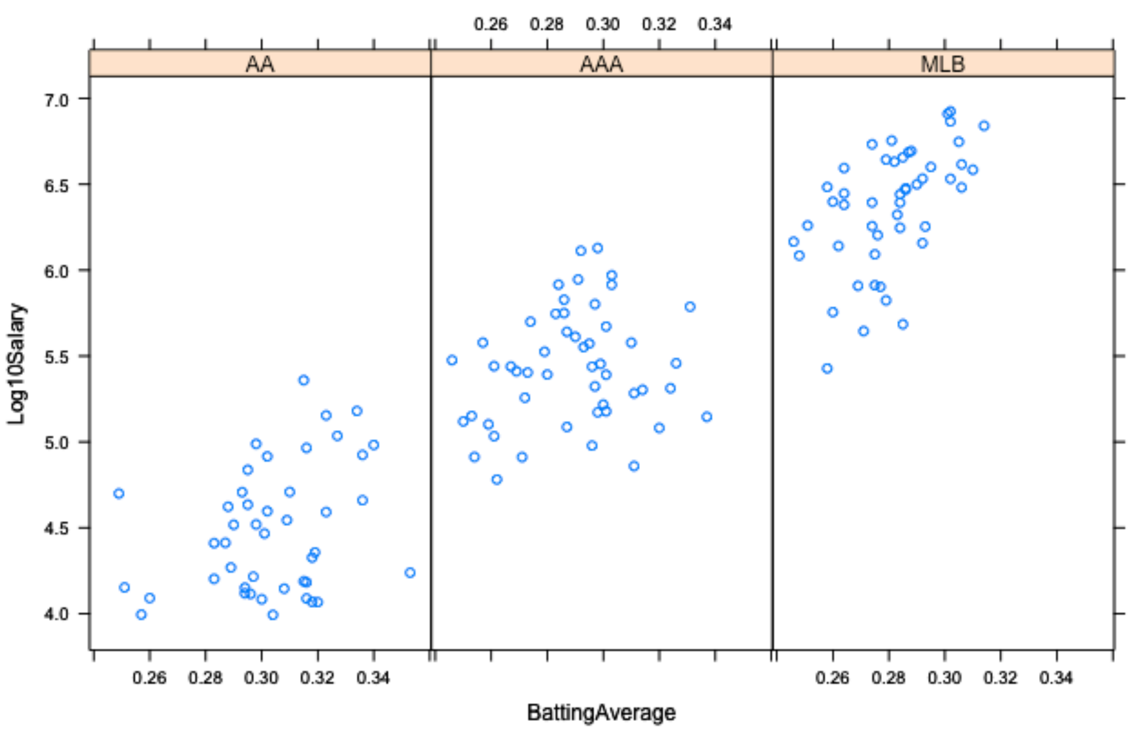
Interaction: A situation in which the sum of the parts is less than the whole

Example: When someone is riding a bike, both the gear and the hill will effect how easily the pedals turn. Eventually, it may become impossible to pedal. This means that one of the variables likely modulated the other variable.

Question: Is the slope of the batting average different depending on the leagues

Objective: Fit a linear model between class, batting average, and interaction between class and batting average

Syntax: "Class: Batting Average"



Mohnish Gandhi
Emily Mi

This is the output for the baseline off set form.

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                2.8488     0.6723   4.238 4.14e-05 ***
BattingAverage             5.3985     2.2083   2.445   0.0158 *
ClassAAA                   1.7936     0.9139   1.963   0.0517 .
ClassMLB                   0.3148     1.0464   0.301   0.7640
BattingAverage:ClassAAA   -2.6468     3.0739  -0.861   0.3907
BattingAverage:ClassMLB    6.0100     3.6054   1.667   0.0978 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3279 on 136 degrees of freedom
Multiple R-squared:  0.8514,	Adjusted R-squared:  0.8459
F-statistic: 155.8 on 5 and 136 DF,  p-value: < 2.2e-16
```

Rscript:

lm3 = lm(Log10Salary ~ BattingAverage + Class + Class:BattingAverage, data=baseballsalary)
summary(lm3)
plot(Log10Salary ~ BattingAverage + Class + Class:BattingAverage, data=baseballsalary)

Mohnish Gandhi
Emily Mi

3 separate regression lines

MLB : $y_i = B_0^1 + B_1^1 x_i + e_i$

AAA : $y_i = B_0^2 + B_1^2 x_i + e_i$

AA : $y_i = B_0^3 + B_1^3 x_i + e_i$

B_0^1
B_0^2
B_0^3
⋮
B_1

$\boxed{\text{Straight up}}$

idea: rewrite above 3 regression lines as 1 line but with dummy variables

Key: X_{i2} = batting average
X_i = class

$\mathbb{1}\{X_{i1} = MLB\} = \begin{cases} 1 & \text{if } X_i = MLB \\ 0 & \text{otherwise} \end{cases}$

dummy variable

$\boxed{\text{Baseline offset form:}}$

$$y_i = B_0 + B_2 X_{i2} + B_1^{AAA} \cdot \mathbb{1}\{X_{i1} = AAA\} + B_1^{MLB} \cdot \mathbb{1}\{X_{i1} = MLB\} + \underbrace{\gamma_1^{AAA} \cdot \mathbb{1}\{X_{i1} = AAA\} \cdot X_{i2} + \gamma_1^{MLB} \cdot \mathbb{1}\{X_{i1} = MLB\} \cdot X_{i2}}_{\text{interaction}}$$

algebra to prove the 1 regression eqn is reconstructed by all 3

when $X_{i1} = AA$:
$y_i = B_0 + B_2 X_{i2} + e_i$ (b/c all the interactions = 0 when you plug it in)

when $X_{i1} = AAA$:
$y_i = B_0 + B_2 X_{i2} + B_1^{AAA} \cdot 1 + \gamma_1^{AAA} \cdot 1 \cdot X_{i2} + e_i$ grouped
$y_i = [\underset{\text{baseline}}{B_0} + \underset{\text{offset}}{B_1^{AAA}}] + [\underset{\text{baseline}}{B_2} + \underset{\text{offset}}{\gamma_1^{AAA}}] X_{i2} + e_i$

when $X_{i1} = MLB$:
$y_i = B_0 + B_2 X_{i2} + B_1^{MLB} + \gamma^{MLB} \cdot X_{i2} + e_i$
$= [B_0 + B_1^{MLB}] + [B_2 + \gamma_1^{MLB}] \cdot X_{i2} + e_i$

· if offset = 0 → no offset → back to baseline