STA 303/1002-Methods of Data Analysis II

Sections L0101& L0201, Winter 2018

Shivon Sue-Chee

UNIVERSITY OF TORONTO

January 23-25, 2018

In the presence of "Interactions"

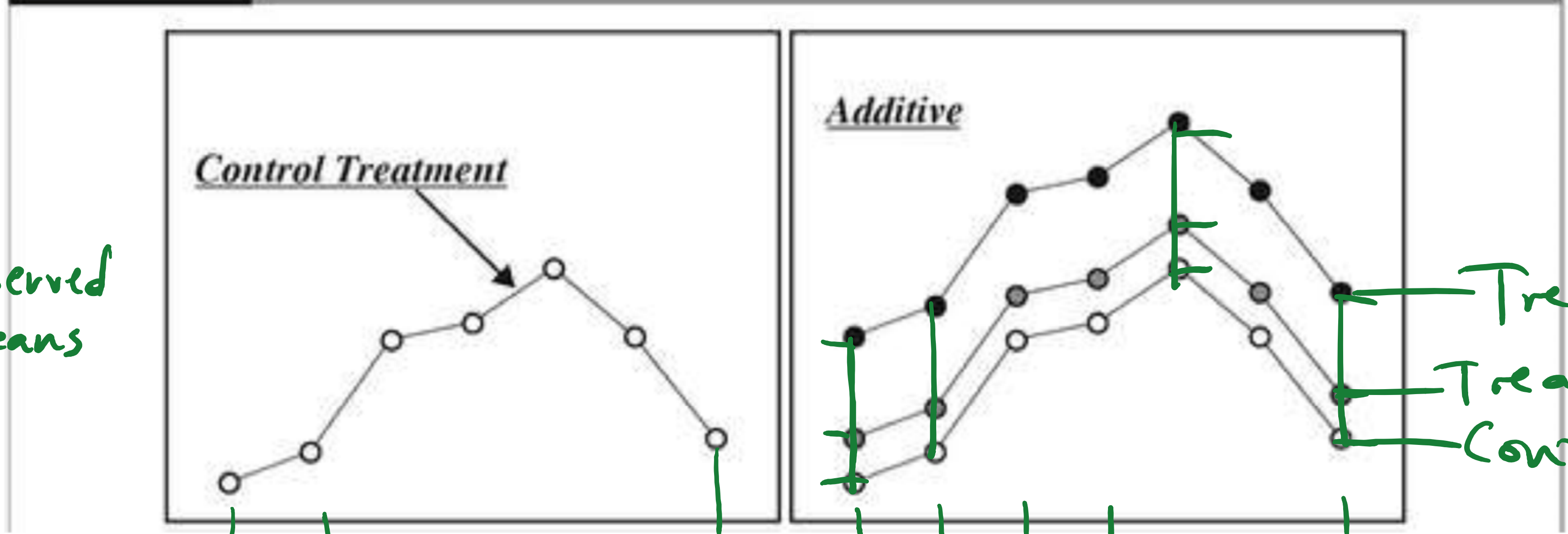
- Hard to answer questions about the main factor effects
- Communicate a table of estimated means, $\hat{y}_i$
- Have separate models of Y against one factor for the different levels of the other factor

References:

- The Statistical Sleuth, 3rd edition by Ramsey and Schafer
- `https://cran.r-project.org/web/packages/Sleuth3/vignettes/chapter13-HortonMosaic.pdf`

In the presence of "Interactions"

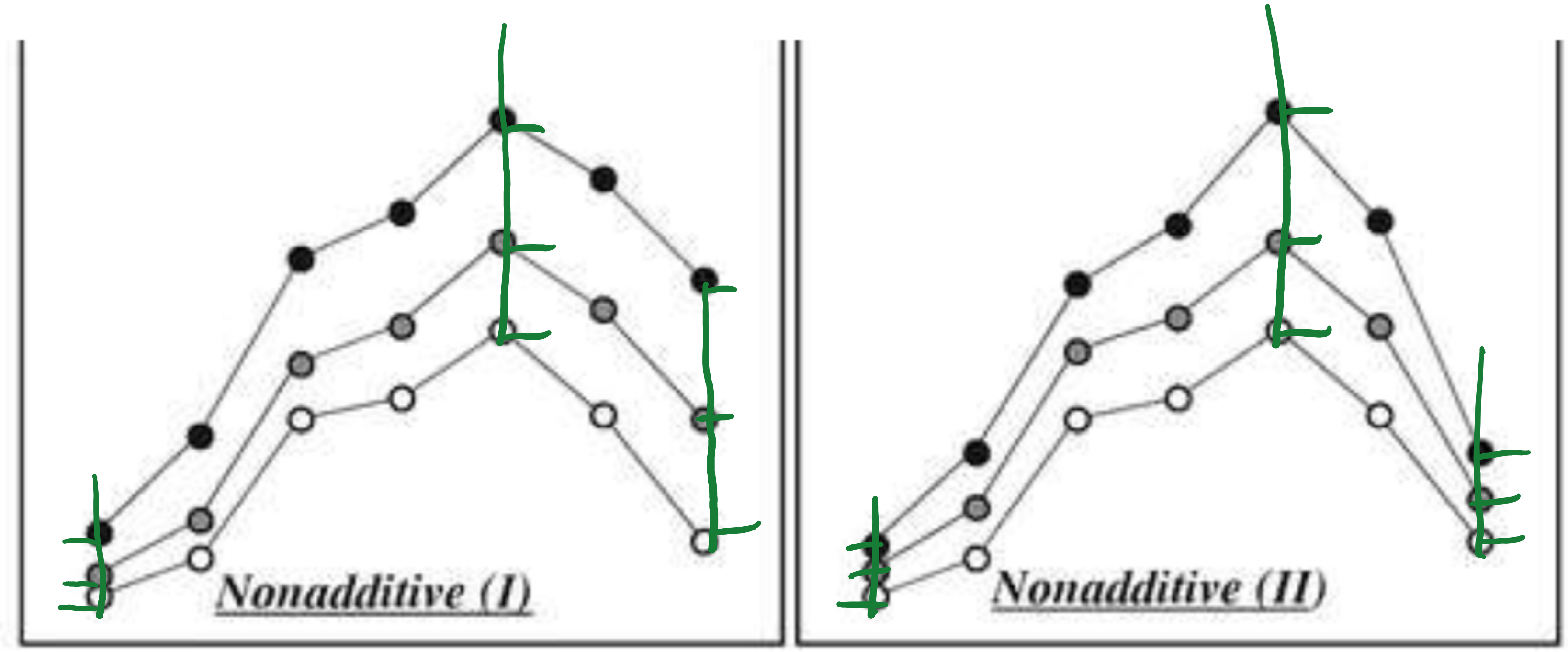
DISPLAY 13.21 Hypothetical treatment curves plotted against another factor, illustrating additive and some nonadditive conditions



Factors

1. Treatment – 3 levels (Control, Treat ⊘, Treat ●)
2. Factor 2 – 7 levels (1, 2, ..., 7)

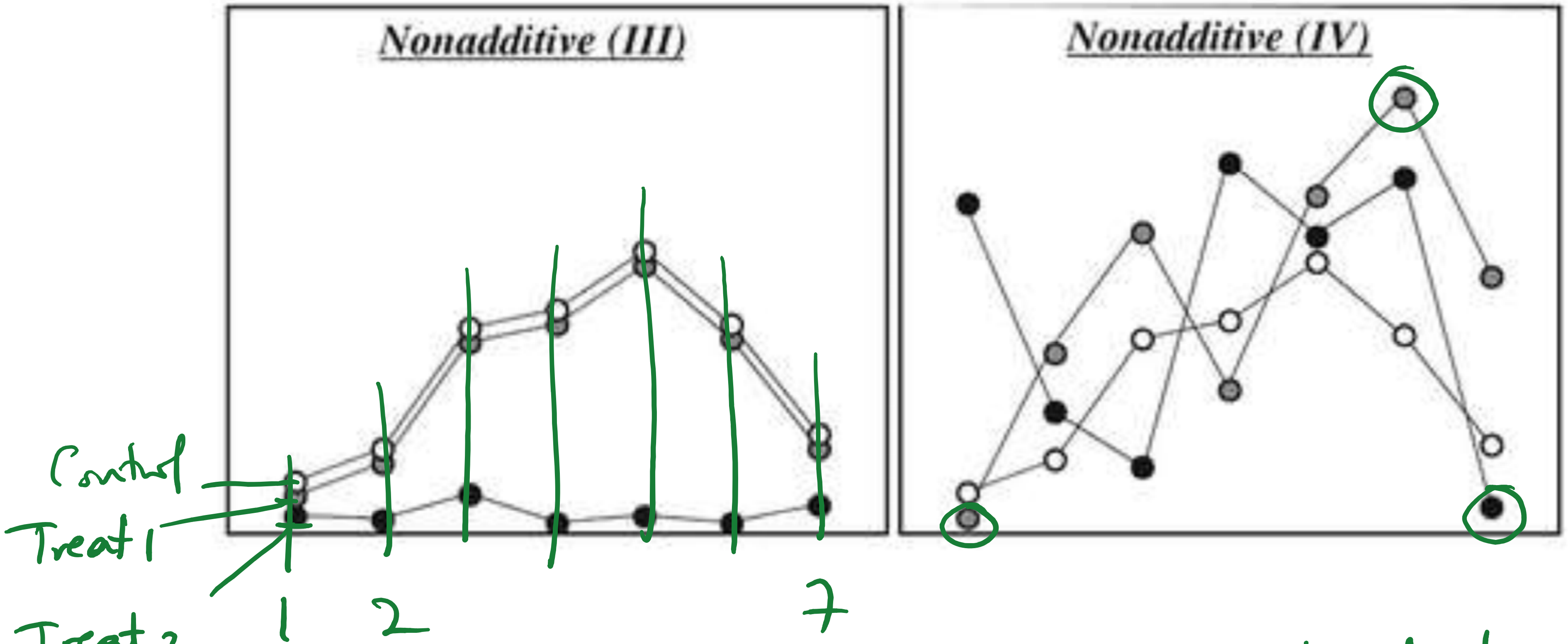
In the presence of "Interactions"



Systematic increase in treatment effect as we increase level of Factor 2

For small observed means, the treatment effect is smaller and vice versa

In the presence of "Interactions"



Treat 2's effect is different from that of Control or Treat 1 as we vary Factor 2's level.

Very complicated interaction
- Identify a few cases
- Add an additional factor

Should insignificant block effects be kept in the model?

– a factor
– not treatment factor

- General advice is to drop insignificant terms
- For data from a randomized block experiment, block effects should be maintained
- Ensure that the control exercised by blocking is maintained in the analysis.

Case Study II-The Pygmalion Effect

- *Pygmalion effect*- high expectations of a supervisor or teacher translate to improved performance by subordinates or students
- Data:

	Treatments		
Company	Pygmalion	Control	
1	80.0	63.2	69.2
2	83.9	63.1	81.5
3	68.2	76.2	
4	76.5	59.5	73.5
5	87.8	73.9	78.5
6	89.8	78.9	84.7
7	76.1	60.6	69.6
8	71.5	67.8	73.2
9	69.5	72.3	73.9
10	83.7	63.7	77.7

$\frac{63.2+69.2}{2} = 66.2$

$\rightarrow 76.2$

Case Study II: Additive model summary

$\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{10}$

```
Call:
lm(formula = Score ~ company + treat)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    68.39316    3.89308  17.568 8.92e-13 ***
companyC10      4.23333    5.36968   0.788   0.4407
companyC2       5.36667    5.36968   0.999   0.3308
companyC3       0.19658    6.01886   0.033   0.9743
companyC4      -0.96667    5.36968  -0.180   0.8591
companyC5       9.26667    5.36968   1.726   0.1015
companyC6      13.66667    5.36968   2.545   0.0203 *
companyC7      -2.03333    5.36968  -0.379   0.7094
companyC8       0.03333    5.36968   0.006   0.9951
companyC9       1.10000    5.36968   0.205   0.8400
treatPygmalion  7.22051    2.57951   2.799   0.0119 *

Residual standard error: 6.576 on 18 degrees of freedom
Multiple R-squared:  0.5647,Adjusted R-squared:  0.3228
F-statistic: 2.335 on 10 and 18 DF,  p-value: 0.0564
```

$\hat{y} \mid (\text{Treat}, \text{Company}) = \hat{\beta}_0 + \hat{\beta}_{treat} + \hat{\beta}_{comp.}$

Case Study II: Additive model summary

```
Call:
lm(formula = Score ~ treat + company)

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      68.39316    3.89308  17.568 8.92e-13 ***
treatPygmalion    7.22051    2.57951   2.799   0.0119 *
companyC10        4.23333    5.36968   0.788   0.4407
companyC2         5.36667    5.36968   0.999   0.3308
companyC3         0.19658    6.01886   0.033   0.9743
companyC4        -0.96667    5.36968  -0.180   0.8591
companyC5         9.26667    5.36968   1.726   0.1015
companyC6        13.66667    5.36968   2.545   0.0203 *
companyC7        -2.03333    5.36968  -0.379   0.7094
companyC8         0.03333    5.36968   0.006   0.9951
companyC9         1.10000    5.36968   0.205   0.8400

Residual standard error: 6.576 on 18 degrees of freedom
Multiple R-squared:  0.5647,Adjusted R-squared:  0.3228
F-statistic: 2.335 on 10 and 18 DF,  p-value: 0.0564
```

1 (treat), 2 (company)

$\hat\beta_0$, $\hat\beta_1$, $\hat\beta_2$, $\hat\beta_3$, $\hat\beta_{10}$

$\hat\beta_1 \pm t^* se(\hat\beta_1)$.

H_0: $\beta_1 = \beta_2 = \beta_3 = \cdots = \beta_{10} = 0 \Rightarrow E(y) = \beta_0$

H_a: at least 1 β not 0 $(p = 0.0564) < (\alpha = 0.10)$

Estimated Mean Response from Additive Model

$\hat{y}(\text{Treat}, \text{Comp}) = \hat{\beta}_0 + \hat{\beta}_T + \hat{\beta}_{comp}$.

Company	Pygmalion($\mathbb{1}_{PYG,i} = 1$)	Control($\mathbb{1}_{PYG,i} = 0$)
1	68.39+7.22 $= \hat{\beta}_0 + \hat{\beta}_1$	68.39 $= \hat{\beta}_0$
2	68.39+7.22+5.37 $= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3$	68.39+5.37
3	$\vdots$	$\vdots$
4		
5		
6		
7		
8		
9		
10	68.39+7.22+4.23	68.39+4.23 $= \hat{\beta}_0 + \hat{\beta}_2$

$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2$

Observed Group means vs Estimated means

$\hat{\beta}_0 + \hat{\beta}_1$ $\hat{\beta}$

	Observed Means			*Estimated* Means	
Company	Pyg	Control	$n_{control}$	Pyg	Control
1	80.0	66.2	2	75.61	68.39 — $\hat{\beta}_0$
2	83.9	72.3	2	80.98	73.76
3	68.2	76.2	$\boxed{1}$	75.81	68.59
4	76.5	66.5	2	74.65	67.43
5	87.8	76.2	2	84.88	77.66
6	89.8	81.8	2	89.28	82.06
7	76.1	65.1	2	73.58	66.36
8	71.5	70.5	2	75.65	68.43
9	69.5	73.1	2	76.71	69.49
10	83.7	70.7	2	79.85	72.63
means	78.70	71.63		78.70	71.48

from 10 19 from 10 From 10 estimated means

Parameter estimation and Unbalanced design

- Estimated means for treatments are averages over 10 companies
- Observed Means vs Estimated means: Not the same because there are unequal number of control observations per company. Company 3 has 1 control platoon; other companies have 2.
- The design is nearly balanced.
- Affects constant variance assumption and variance estimate
- Consider any evidence as exploratory
- Consider *weighted* least squares regression

Measuring treatment effect

$$(\bar{x}_1 - \bar{x}_2) \pm t^*_{27,\,0.025}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
> qt(1-0.05/2,df=27)
> sqrt((9*var(Score[treat=="Pygmalion"])+18*var(Score[treat=="Control"]))/27)
> t.test(Score[treat=="Pygmalion"], Score[treat=="Control"], var.equal=T)

[1] 2.051831
[1] 7.356078
```

10   19

$n_1 + n_2 - 2 = 10 + 19 - 2 = 27.$

```
Two Sample t-test

data:  Score[treat == "Pygmalion"] and Score[treat == "Control"]
t = 2.4595, df = 27, p-value = 0.0206
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  1.171707 12.965135
sample estimates:
mean of x mean of y
 78.70000  71.63158
```

← does not include 0.

Conclusions

- There is evidence of a difference in mean score between pygmalion and control platoons (p=0.0119). (Consider this as weak evidence since we have some concerns about variance estimates.)
- Confidence Intervals for the difference in mean score between pygmalion and control platoons:
  - Pooled 2-sample t:

$$(78.7 - 71.6) \pm 2.05(7.36)\sqrt{(1/10 + 1/19)} = (1.17, 12.96)$$

Similar since design is nearly balanced.

  - Least-squares approach (Additive model):

$18 = 10 + 10 - 2 = df$

$78.70 - 71.48 = 7.22$

$$\underbrace{7.22}_{\hat{\beta}_1} \pm 2.101\underbrace{(2.5795)}_{se(\hat{\beta}_1)} = (1.8, 12.6)$$

- On average, pygmalion platoons (mean=78.7) scored higher than control platoons (mean=71.6).