QUASI-EXPERIMENT

PROF. XINXIN LI

QUASI-EXPERIMENT

- A quasi-experiment, again, examines the interventional effect of a treatment, but is done without random assignment.
  - is useful when random assignment is difficult or infeasible to implement
  - can be used to examine the effect in retrospect

- With random assignment, participants have the same chance of being assigned to the treatment and control groups, so their differences before treatment are due to chance, and their differences after treatment can be attributed to the treatment effect.

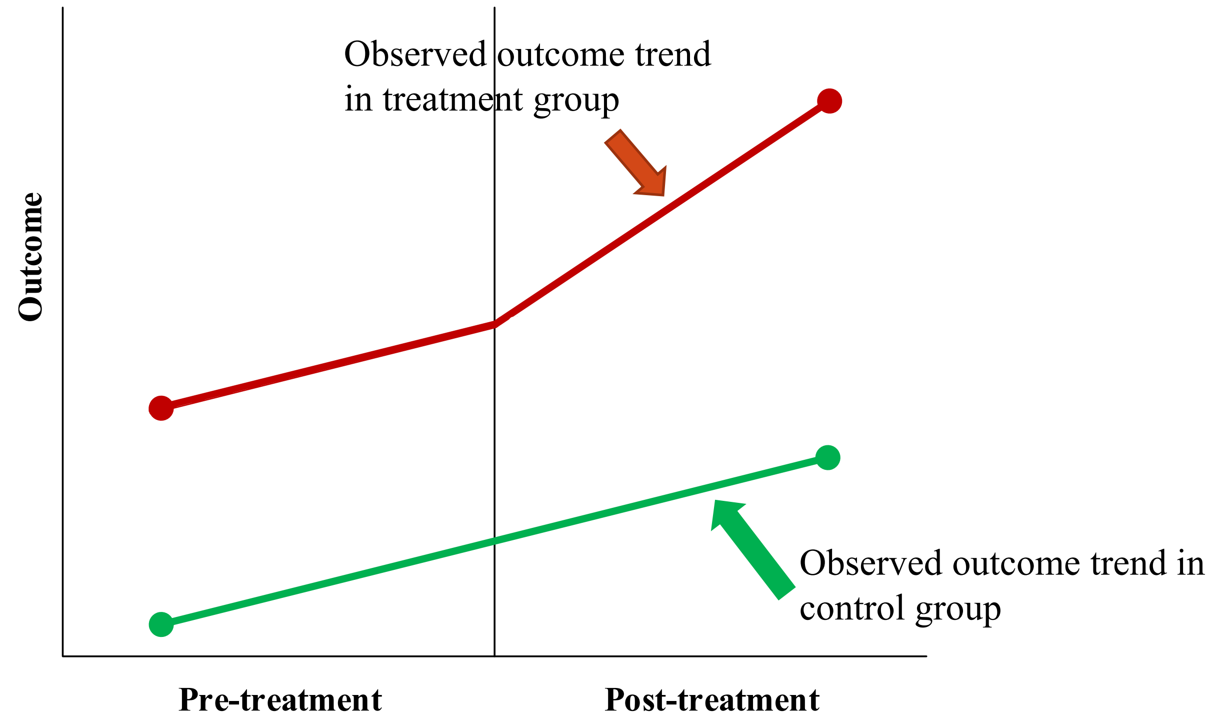
- Without random assignment, it is important to validate whether the participants in the treatment group and those in the control group are comparable, and a participant's chance of being in the treatment group is not correlated with the treatment.

- Method for validity check: Difference in Differences (DID)

DIFFERENCE IN DIFFERENCES (DID)

- Makes use of longitudinal data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect.

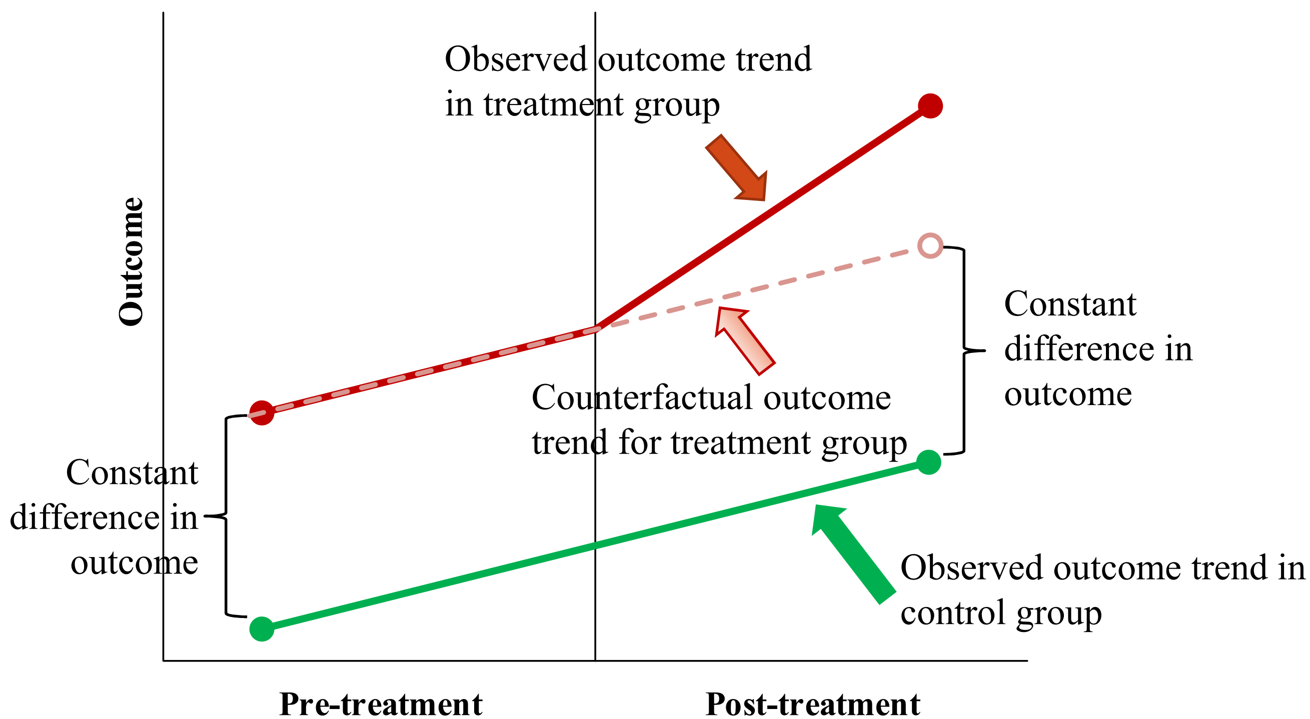
$$Y = \beta_0 + \beta_1 Time + \beta_2 Treatment + \beta_3 Time * Treatment + \varepsilon$$



DIFFERENCE IN DIFFERENCES (DID)

- Makes use of longitudinal data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect.

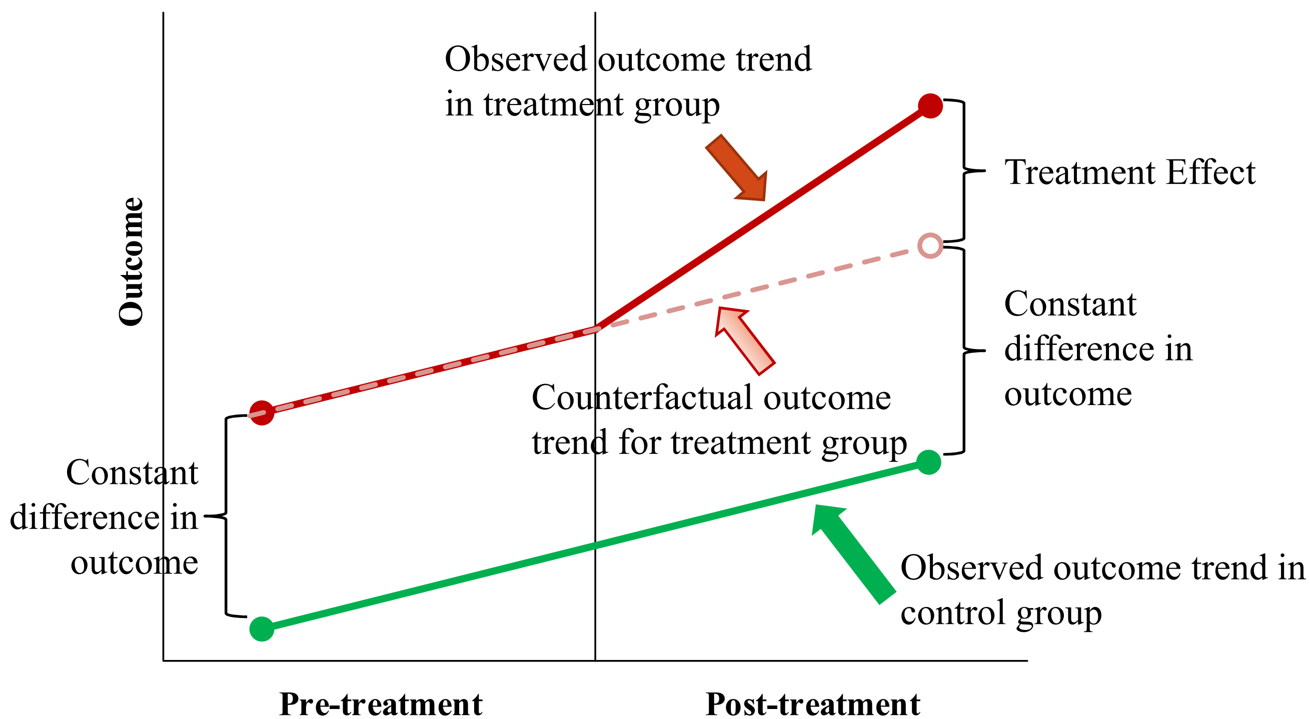
$$Y = \beta_0 + \beta_1 Time + \beta_2 Treatment + \beta_3 Time * Treatment + \varepsilon$$



DIFFERENCE IN DIFFERENCES (DID)

- Makes use of longitudinal data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect.

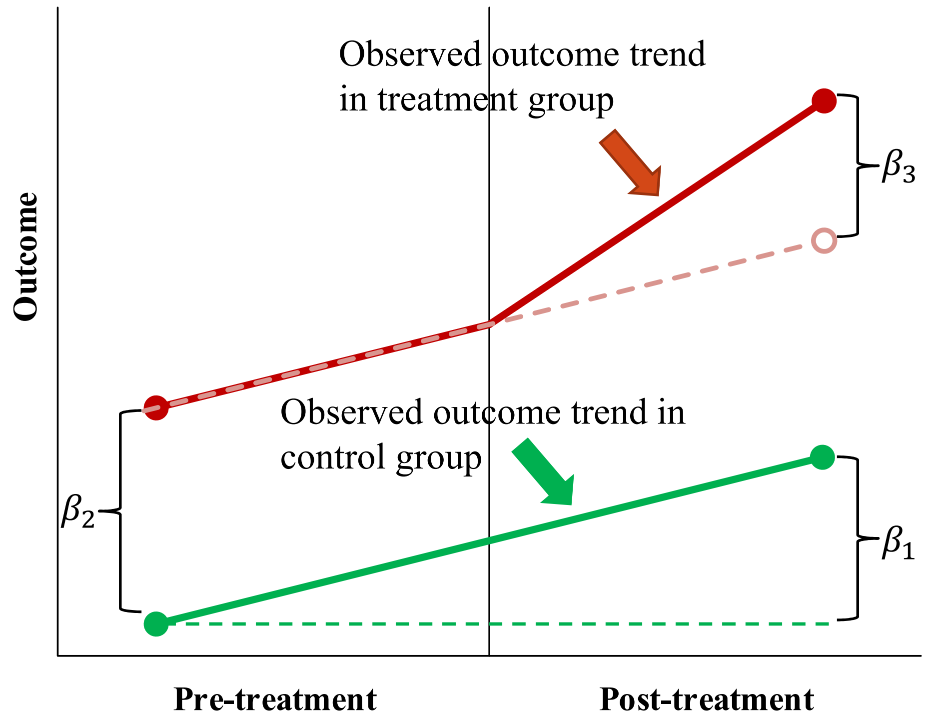
$$Y = \beta_0 + \beta_1 Time + \beta_2 Treatment + \beta_3 Time * Treatment + \varepsilon$$



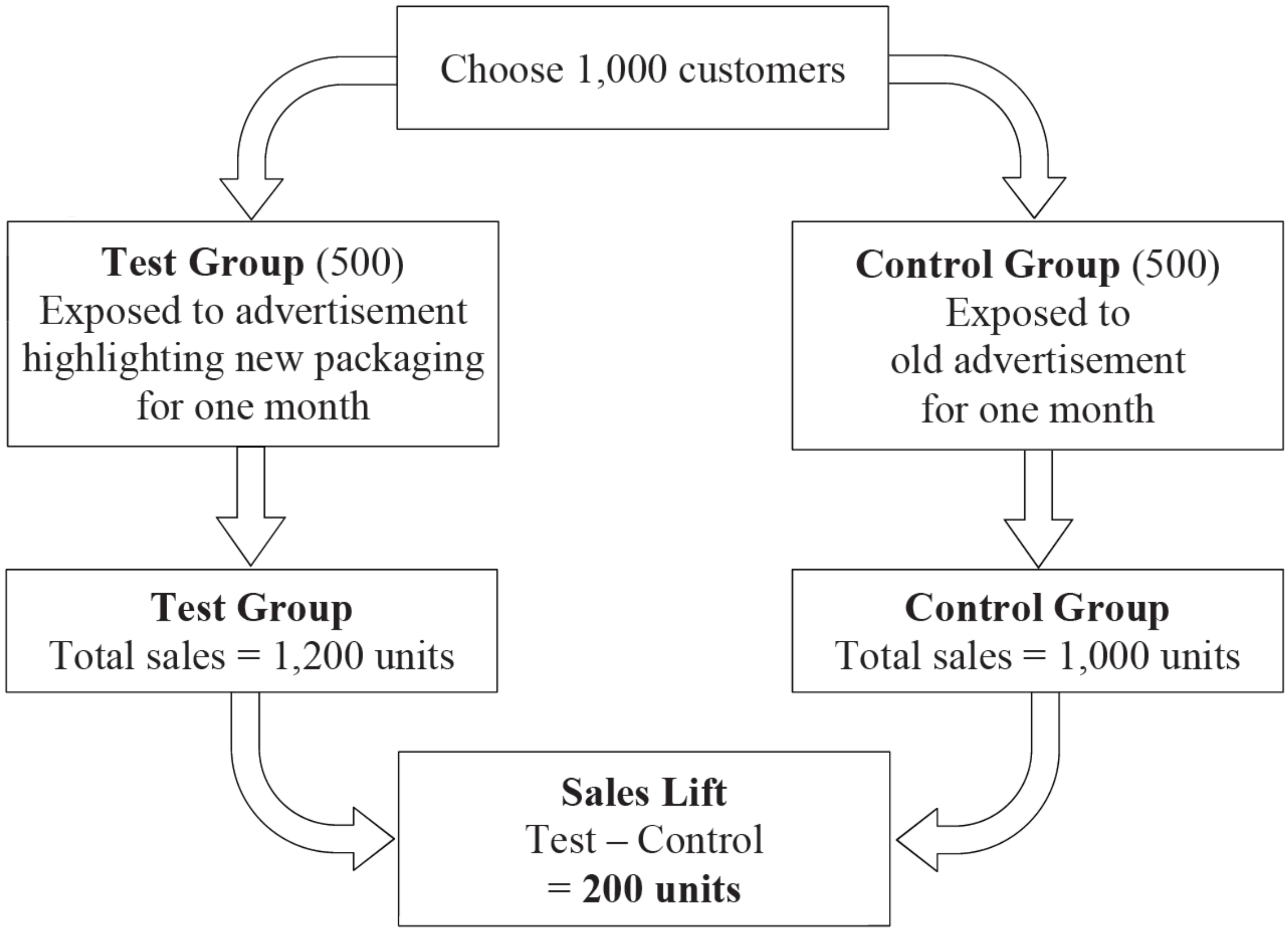
DIFFERENCE IN DIFFERENCES (DID)

- Makes use of longitudinal data from treatment and control groups to obtain an appropriate counterfactual to estimate a causal effect.

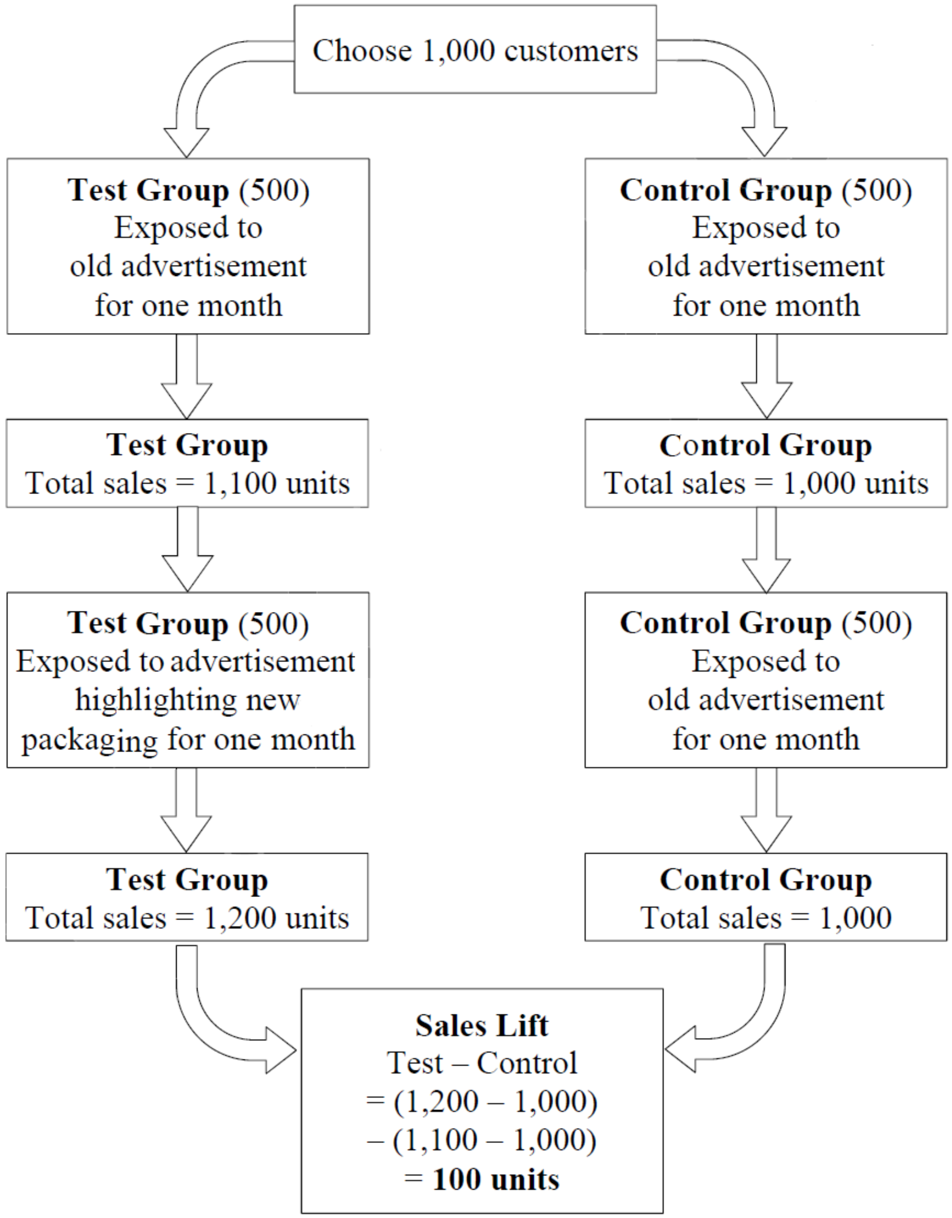
$$Y = \beta_0 + \beta_1 Time + \beta_2 Treatment + \beta_3 Time * Treatment + \varepsilon$$



AFTER-ONLY EXPERIMENT

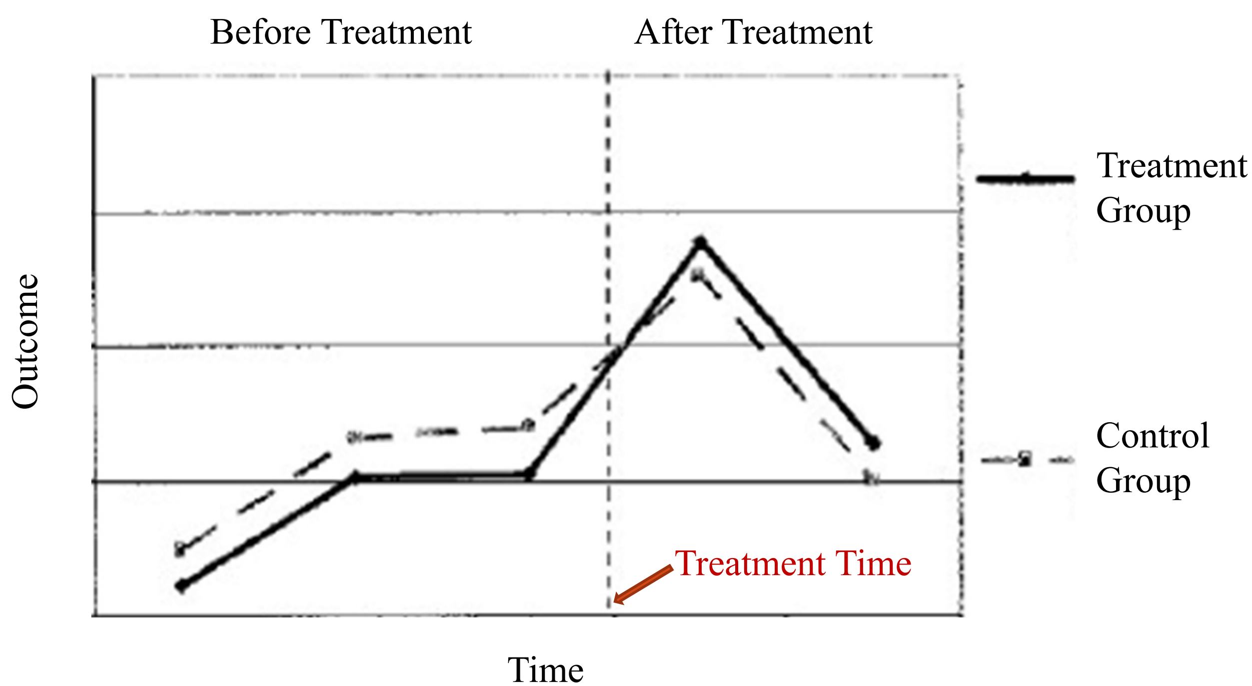


BEFORE-AFTER EXPERIMENT



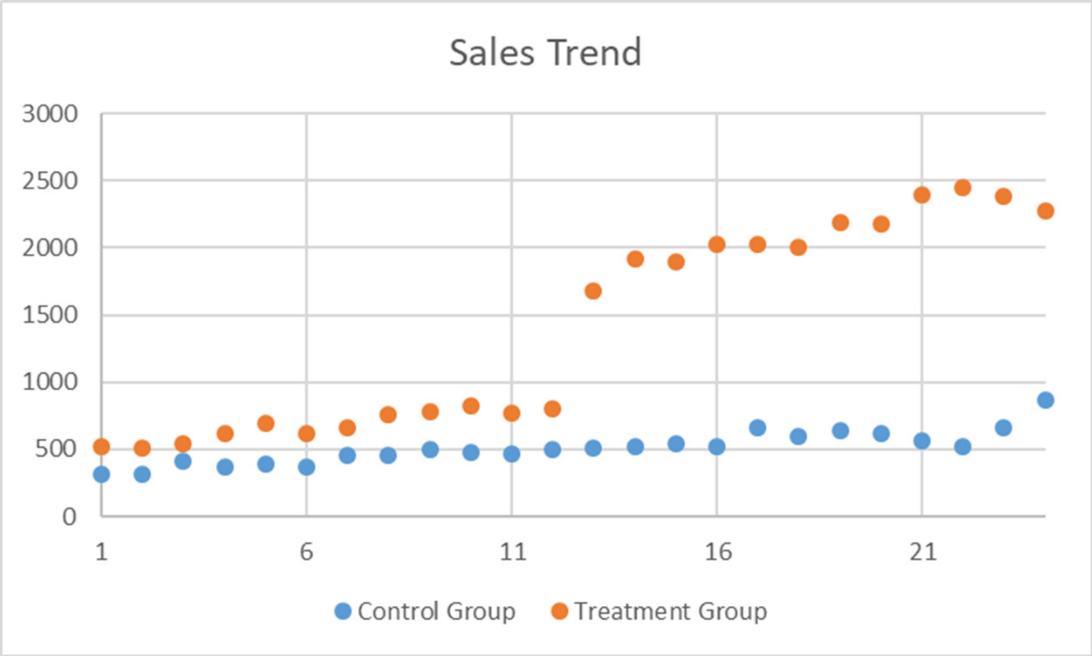
PARALLEL TREND ASSUMPTION

- A critical assumption to ensure validity of DID models (treatment unrelated to outcome at baseline)



IS AD EFFECTIVE? A DID EXERCISE

- Two groups (DID.xlsx)
  - Control group: 50 locations
  - Treatment group: 50 locations
  - Started to offer ads to Treatment group on day #13
  - Variables
    - locationID: ID of the locations
    - treated: $= \begin{cases} 1 & \text{treatment group} \\ 0 & \text{control group} \end{cases}$
    - period: day # (from 1 to 24)
    - after: $= \begin{cases} 1 & \text{since day \#13} \\ 0 & \text{before day \#13} \end{cases}$
    - sales: daily sales
- How to set up the model?



DID ANALYSIS

$$Sales = \beta_0 + \beta_1 After + \beta_2 Treated + \beta_3 After * Treated + \varepsilon$$

```
Call:
lm(formula = sales ~ as.factor(after) + as.factor(treated) +
    as.factor(treated):as.factor(after), data = DID)

Residuals:
    Min      1Q  Median      3Q     Max
-28.373  -5.477  -0.477   4.970  32.627

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                             8.3383     0.3342  24.953  < 2e-16 ***
as.factor(after)1                       3.6917     0.4726   7.812 8.36e-15 ***
as.factor(treated)1                     5.1383     0.4726  10.873  < 2e-16 ***
as.factor(after)1:as.factor(treated)1  25.2050     0.6683  37.714  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.185 on 2396 degrees of freedom
Multiple R-squared:  0.7342,   Adjusted R-squared:  0.7339
F-statistic:  2206 on 3 and 2396 DF,  p-value: < 2.2e-16
```

VALIDATING PARALLEL TREND ASSUMPTION

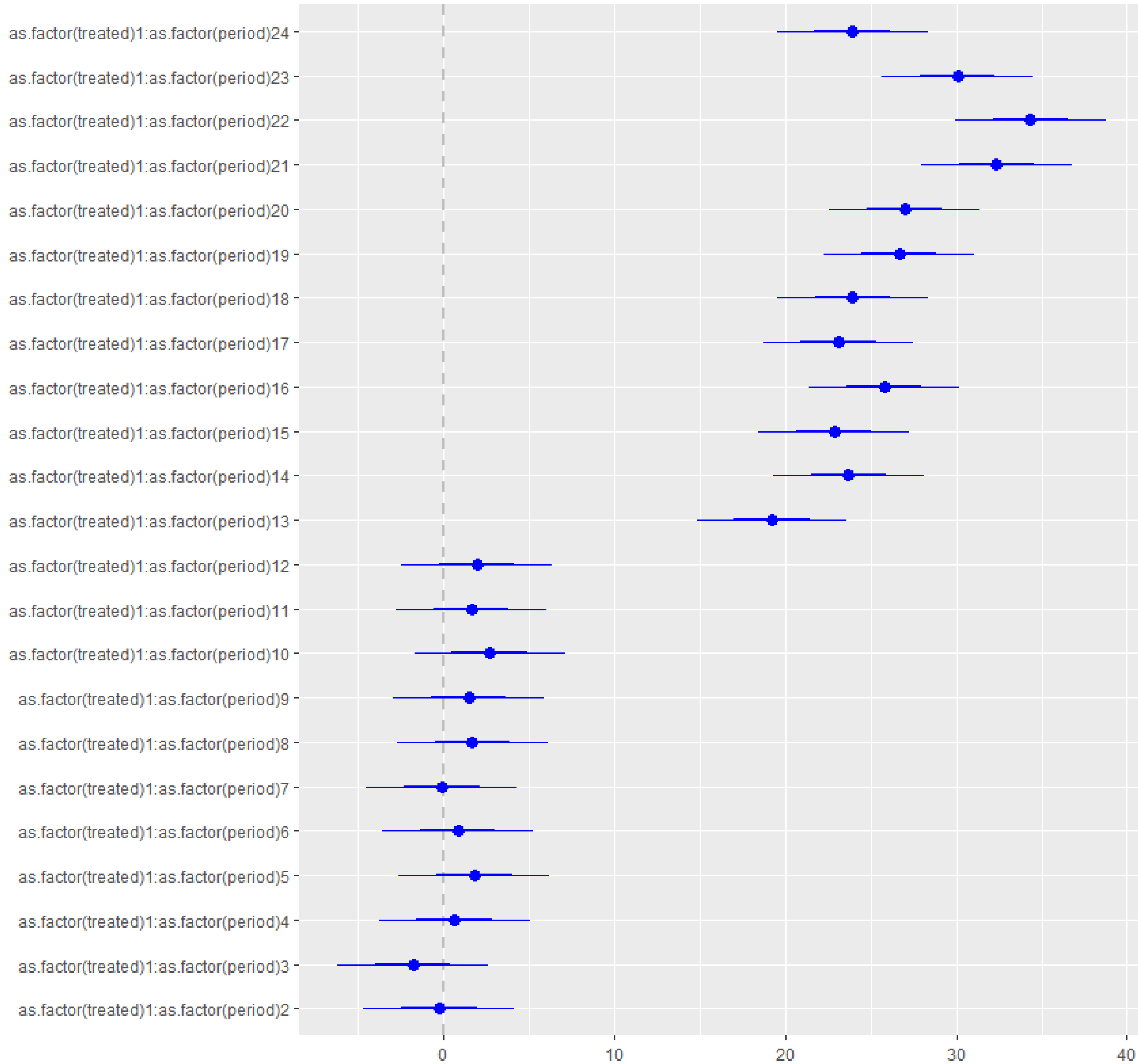
$$Sales = \alpha_0 + \alpha_1 Treated + \alpha_2 Period + \alpha_3 Treated * Period + \varepsilon$$

```
Call:
lm(formula = sales ~ as.factor(treated) + as.factor(period) +
    as.factor(treated):as.factor(period))

Residuals:
   Min      1Q  Median      3Q     Max
-26.56  -5.38  -0.25    5.04   29.50

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                                6.240      1.101   5.665 1.65e-08 ***
as.factor(treated)1                        4.260      1.558   2.735 0.006288 **
as.factor(period)2                        -0.020      1.558  -0.013 0.989757
 ......
as.factor(treated)1:as.factor(period)2    -0.260      2.203  -0.118 0.906056
as.factor(treated)1:as.factor(period)3    -1.780      2.203  -0.808 0.419154
as.factor(treated)1:as.factor(period)4     0.640      2.203   0.291 0.771438
as.factor(treated)1:as.factor(period)5     1.800      2.203   0.817 0.413948
as.factor(treated)1:as.factor(period)6     0.840      2.203   0.381 0.703001
as.factor(treated)1:as.factor(period)7    -0.100      2.203  -0.045 0.963796
as.factor(treated)1:as.factor(period)8     1.680      2.203   0.763 0.445756
as.factor(treated)1:as.factor(period)9     1.460      2.203   0.663 0.507544
as.factor(treated)1:as.factor(period)10    2.700      2.203   1.226 0.220447
as.factor(treated)1:as.factor(period)11    1.620      2.203   0.735 0.462169
as.factor(treated)1:as.factor(period)12    1.940      2.203   0.881 0.378590
as.factor(treated)1:as.factor(period)13   19.220      2.203   8.725  < 2e-16 ***
as.factor(treated)1:as.factor(period)14   23.700      2.203  10.759  < 2e-16 ***
as.factor(treated)1:as.factor(period)15   22.860      2.203  10.377  < 2e-16 ***
as.factor(treated)1:as.factor(period)16   25.800      2.203  11.712  < 2e-16 ***
as.factor(treated)1:as.factor(period)17   23.120      2.203  10.495  < 2e-16 ***
as.factor(treated)1:as.factor(period)18   23.940      2.203  10.868  < 2e-16 ***
as.factor(treated)1:as.factor(period)19   26.660      2.203  12.102  < 2e-16 ***
as.factor(treated)1:as.factor(period)20   26.980      2.203  12.248  < 2e-16 ***
as.factor(treated)1:as.factor(period)21   32.360      2.203  14.690  < 2e-16 ***
as.factor(treated)1:as.factor(period)22   34.360      2.203  15.598  < 2e-16 ***
as.factor(treated)1:as.factor(period)23   30.080      2.203  13.655  < 2e-16 ***
as.factor(treated)1:as.factor(period)24   23.920      2.203  10.858  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

VALIDATING PARALLEL TREND ASSUMPTION



SENSITIVITY ANALYSIS

- Perform a "placebo" analysis, e.g., by
  - Using data for control and treatment groups from previous years
  - Using a population that is not supposed to be affected by the treatment as a "fake" treatment group
  - If the new analysis yields a significant result, the original DID analysis can be biased

- Use a different control group
  - Result should be similar, otherwise the original DID is questionable

- Perform a falsification test by using an outcome variable that is not supposed to be affected by the treatment
  - If the result is significant, the original DID analysis is questionable