BREAK
THROUGH
TECH

Unwrapping Customer Delight

Milestone #4 Meeting: The Analysis Phase

The Estée Lauder Companies

October 17, 2025

Why We Care About the Average Treatment Effect (ATE)

- In experiments (A/B tests, clinical trials, campaigns), we want to know:
  “Did the treatment cause a measurable change?”

- The **Average Treatment Effect (ATE)** captures the *causal impact* of a treatment across all participants

- Defined as: **ATE** = $E[Y(1)] - E[Y(0)]$
  where $Y(1)$ = outcome if treated, $Y(0)$ = outcome if not treated

- Because we can’t observe both outcomes for the same unit, we estimate this using **control vs. treatment groups**

- **Relevance to our project:**

  - Treatment = sending customers a “gift”
  - Outcome = their *post-intervention* revenue ($)
  - ATE = average change in revenue attributable to the gift

How We Estimate the ATE

- Thanks to **randomization**, treatment assignment is independent of other customer characteristics.

- This lets us estimate:

  ATE = $Y_{treatment} - Y_{control}$

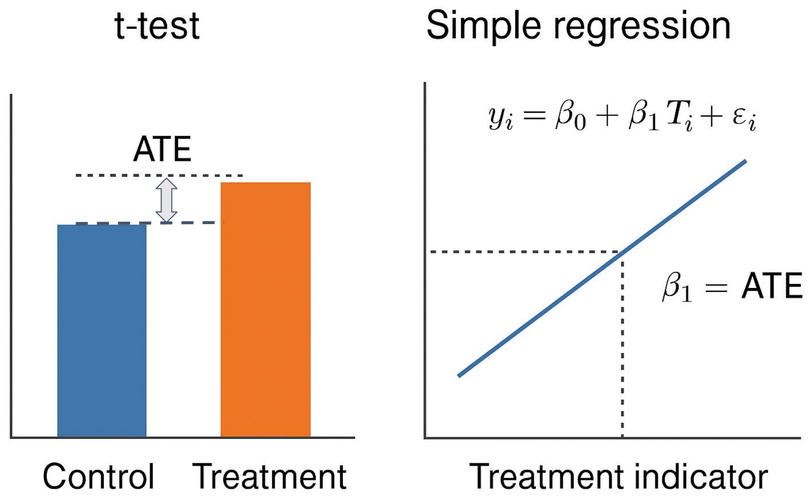
  the difference in average outcomes between treatment and control.

- Two *equivalent* ways to compute this:

  - **Two-sample t-test**: compares group means directly

  - **Simple regression**: $y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$

    - T_i: treatment indicator (1 = treated, 0 = control)
    - β_1: estimate of the ATE



- Both approaches produce the same estimate when data are randomized — they differ only in *implementation*, not in math.

Why the Regression Approach Is Useful

- Even though the regression and t-test yield the same ATE now, regression offers **important advantages**:
  - **Extensibility** – we can later add control variables (covariates) or interaction terms
  - **Robustness** – we can use heteroskedasticity-robust standard errors (like "HC0") when group variances differ
  - **Consistency of workflow** – it's easier to build on this structure when moving to more advanced estimators (like MLRATE).
- Conceptually, think of the regression as a *generalized t-test framework*:
  - One model type, many extensions
  - You'll use the same pattern for the variance-reduced estimator in Milestone #6

Computing the Unadjusted ATE and Confidence Interval

```python
import statsmodels.api as sm

# Suppose you have variables 'revenue' (outcome) and
# 'assignment' (with 0=control, 1=treatment) in memory

# 1. Define model: revenue = b0 + b1 * assignment + error
X = sm.add_constant(assignment)
y = revenue

# 2. Fit OLS regression
model = sm.OLS(y, X).fit(cov_type="HC0")
print(model.summary(xname=["const", "T"]))

# 3. Extract results
ate = model.params[1] # index 1 corresponds to 'assignment'
ci_lower, ci_upper = model.conf_int()[1]

print(f"ATE: {ate:.4f}")
print(f"95% CI: [{ci_lower:.4f}, {ci_upper:.4f}]")
```

- The **95% CI** gives a range of plausible values for the *true* ATE, based on the sample data
- **Frequentist interpretation:**
  - If we were to repeat this experiment many times, **95% of those CIs would contain the true ATE**
  - It does **not** mean there's a "95% chance" that *this* interval contains the true value
  - The truth is fixed; the CI is random across repeated samples
- **Practical intuition:**
  - A narrow CI → precise estimate (low variance)
  - A wide CI → noisy estimate (uncertain effect)
  - If the CI excludes 0 → effect is statistically significant

Computing the Unadjusted ATE

See the below OLS results from two analyses of the same dataset, one with a larger sample size. **The true effect is 20$**.

- *Note: this OLS summary is for illustrative purposes only, using a separate dataset not related to your project.*
- P-value: The probability that in reality there is no treatment effect (i.e. the null hypothesis is true) and that the observed difference in means is purely by random chance.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.232
Model:                            OLS   Adj. R-squared:                  0.230
Method:                 Least Squares   F-statistic:                     90.12
Date:                Tue, 10 Sep 2024   Prob (F-statistic):           5.89e-51
Time:                        14:30:25   Log-Likelihood:                -3330.3
No. Observations:                 898   AIC:                             6669.
Df Residuals:                     894   BIC:                             6688.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         11.1637     23.366      0.478      0.633     -34.695      57.022
x              1.0777      0.301      3.575      0.000       0.486       1.669
treated       10.4790     36.389      0.288      0.773     -60.939      81.897
x:treated     -0.0642      0.453     -0.142      0.887      -0.954       0.825
==============================================================================
Omnibus:                        0.884   Durbin-Watson:                   1.928
Prob(Omnibus):                  0.643   Jarque-Bera (JB):                0.965
Skew:                          -0.066   Prob(JB):                        0.617
Kurtosis:                       2.909   Cond. No.                     1.09e+04
==============================================================================
```

Treatment effect IS NOT statistically significant

- p = 0.773
- There is a **~77% chance** that the null hypothesis is actually true (i.e. no treatment effect) and that we observed a **$10.48 effect** in this sample **by pure chance**.
- Notice the smaller sample size and R-squared, and how the resulting decrease in power masks the true (existing) effect.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.864
Model:                            OLS   Adj. R-squared:                  0.864
Method:                 Least Squares   F-statistic:                 2.943e+04
Date:                Tue, 10 Sep 2024   Prob (F-statistic):               0.00
Time:                        14:50:25   Log-Likelihood:                -51492.
No. Observations:               13861   AIC:                         1.030e+05
Df Residuals:                   13857   BIC:                         1.030e+05
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         10.2356      0.340     30.062      0.000       9.568      10.903
x              1.0946      0.007    161.271      0.000       1.081       1.108
treated       21.2567      1.597     13.310      0.000      18.126      24.387
x:treated     -0.2034      0.017    -11.894      0.000      -0.237      -0.170
==============================================================================
Omnibus:                        0.104   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.949   Jarque-Bera (JB):                0.123
Skew:                          -0.000   Prob(JB):                        0.940
Kurtosis:                       2.985   Cond. No.                     1.27e+03
==============================================================================
```

Treatment effect IS statistically significant

- p = 0
- There is a **0% chance** that the null hypothesis is true (i.e. no treatment effect) and that we observed a **$21.26 effect** in this sample **by pure chance**.
- With a sufficient size, this sample has enough power to successfully detect the true (existing) effect.

Key Takeaways

- The **ATE** quantifies the *causal lift* from treatment in randomized experiments
- In a simple randomized design with only a treatment indicator and an intercept, the **regression coefficient on treatment** is *algebraically identical* to the **difference in group means**
- **Regression** is preferred in practice because it's flexible, robust, and easy to extend
- The **confidence interval** expresses the *precision* of our estimate — not the confidence in our model
- ➡️ Looking ahead to Milestone #6:
  - We'll introduce the MLRATE estimator, which enhances precision by incorporating information from pre-treatment covariates that help explain variation in outcomes
  - The idea as always: use machine learning to explain outcome variance and thereby *reduce noise* in the ATE estimate

Estimating Standard ATE

- **Load data:** use experiment_results_*.parquet
  - Use assignment as treatment indicator and revenue as post-experiment outcome
- Fit OLS model (as specified in slides 3 and 5)
  - Estimate the **unadjusted treatment effect** using statsmodels.OLS with robust SEs (cov_type=HC0)
  - Generate a results summary printout (similar to those on slide 6) and extract:
    - Coefficient on assignment → your ATE
    - Its 95% confidence interval, and compute the CI width
- Reflect on findings:
  - What does the estimated ATE suggest about the campaign's impact?
  - Is the CI too wide or narrow? Does it include 0? What are the implications of this for business decision-making?
  - What is the p-value for the treatment coefficient? Does it provide strong evidence to reject the null?
  - How might sample size or variance have influenced your results?
- **Meeting will also cover:**
  - MLRATE approach for ATE estimation

Project milestones and timeline

These are the milestones for your Challenge Project. They include the CRISP-DM process steps you learned about in your ML Foundations course. In addition, there is an educational component in the front-end.

MILESTONE 1

- Review material from "Helpful resources" slide
- Q&A with challenge advisors

MILESTONE 2

- EDA on pre-experiment data
- Present EDA findings and data preprocessing

MILESTONE 3

- Power analysis
  - Determine the variance-reduced sample size for the upcoming RCT
  - Compare to standard t-test sample size

MILESTONE 4

- You will have received the experiment data
- Perform EDA on experimental data
- Present EDA findings and data preprocessing

MILESTONE 5

- Estimate ATE using two approaches:
  - First by using the standard method to estimating the appropriate OLS regression coefficient

MILESTONE 6

- Second, estimate ATE with MLRATE
- Compare to standard ATE from OLS

MILESTONE 7

- Deliver final project presentation: RCT, MLRATE, key findings
- Present quantified gift impact and campaign recommendations
- Final Q&A and project recap