Problem Set 2

Cem Kozanoglu

October 4, 2025

1 Question 1: Average Treatment Effect Estimation

This section analyzes the treatment effects of a job training program using data from the Riverside GAIN experiment. I estimate the Average Treatment Effect (ATE) using two different approaches: (1) the overall difference-in-means estimator applied to the full sample, and (2) a weighted average of subgroup-specific treatment effects. The analysis covers two outcome periods: earnings in year 1 and earnings in year 4 following program assignment.

1.1 Overall ATE Estimation

The overall Average Treatment Effect is estimated using the standard difference-in-means estimator:

$$\hat{\tau}_{overall} = \bar{Y}_1 - \bar{Y}_0 = \frac{1}{M} \sum_{i:W_i=1} Y_i - \frac{1}{N-M} \sum_{i:W_i=0} Y_i$$

where Y_i represents earnings, W_i is the treatment indicator, M is the number of treated units, and $N - M$ is the number of control units.

The variance estimator for the overall ATE is:

$$\widehat{Var}(\hat{\tau}_{overall}) = \frac{s_1^2}{M} + \frac{s_0^2}{N-M}$$

where s_1^2 and s_0^2 are the sample variances for treated and control groups, respectively.

Table **??** presents the estimated treatment effects for both outcome periods using our ATE estimators.

Table 1: ATE Estimates with Neyman Variance

Outcome	ATE Estimate	Neyman Variance	Neyman Std. Error
Earnings Year 1	1.1362	0.0180	0.1342
Earnings Year 4	1.2323	0.0619	0.2488

Table 2: ATE Estimates with Homoskedastic Variance

Outcome	ATE Estimate	Homoskedastic Variance	Homoskedastic Std. Error
Earnings Year 1	1.1362	0.0290	0.1703
Earnings Year 4	1.2323	0.0786	0.2804

Using a 90% confidence interval (z = 1.645) with Neyman Variance, both treatment effects are statistically significant:

- Year 1 Overall ATE: $1.1362 \pm 1.645 \times 0.1342 = [0.916, 1.357]$

- Year 4 Overall ATE: $1.2323 \pm 1.645 \times 0.2488 = [0.823, 1.642]$

1.2   Subgroup-Weighted ATE Estimation

I also estimate the ATE as a weighted average of subgroup-specific treatment effects, where subgroups are defined by education status (high school completion) and presence of small children:

$$\hat{\tau}_{aggregate} = \sum_{g=1}^{4} w_g \hat{\tau}_g$$

where $w_g = \frac{N_g}{N}$ is the proportion of observations in subgroup g, and $\hat{\tau}_g$ is the subgroup-specific ATE.

The variance estimator for the aggregate ATE is:

$$\widehat{Var}(\hat{\tau}_{aggregate}) = \sum_{g=1}^{4} w_g^2 \widehat{Var}(\hat{\tau}_g)$$

Table 3 shows the heterogeneous treatment effects across the four demographic subgroups for year 1 earnings.

Table 3: Subgroup Analysis - Earnings Year 1

Subgroup	ATE	Variance	Sample Size	Weight
Education + Child	1.4306	0.1784	519	0.0958
Education + No Child	1.5040	0.0547	2,316	0.4274
No Education + Child	0.4951	0.3488	368	0.0679
No Education + No Child	0.7640	0.0289	2,216	0.4089

The results reveal substantial treatment effect heterogeneity. Individuals with high school education experience much larger treatment effects (1.43-1.50) compared to those without high school education (0.50-0.76). The largest treatment effect is observed for educated individuals without small children (1.5040), while the smallest effect is for non-educated individuals with small children (0.4951).

Figure 1 illustrates this heterogeneity, clearly showing the education-based divide in treatment effectiveness.
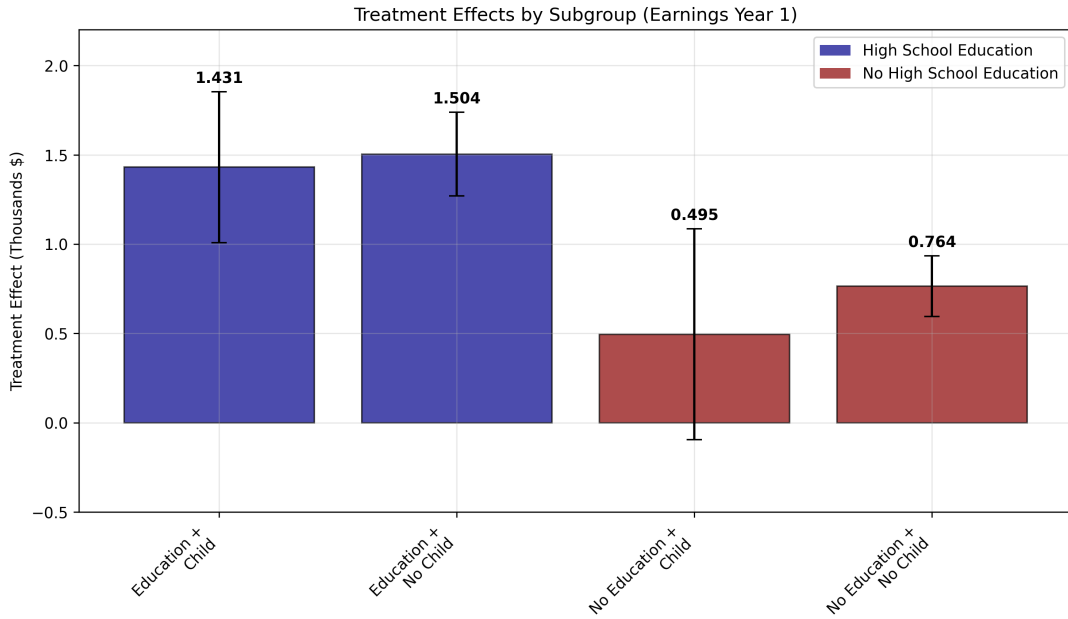

Figure 1: Treatment Effects by Subgroup (Earnings Year 1)

We observe fairly similar results for earnings in year 4, as shown in Table 4 and Figure 2. The pattern of heterogeneity remains consistent, with educated individuals benefiting more from the program.

Table 4: Subgroup Analysis - Earnings Year 4

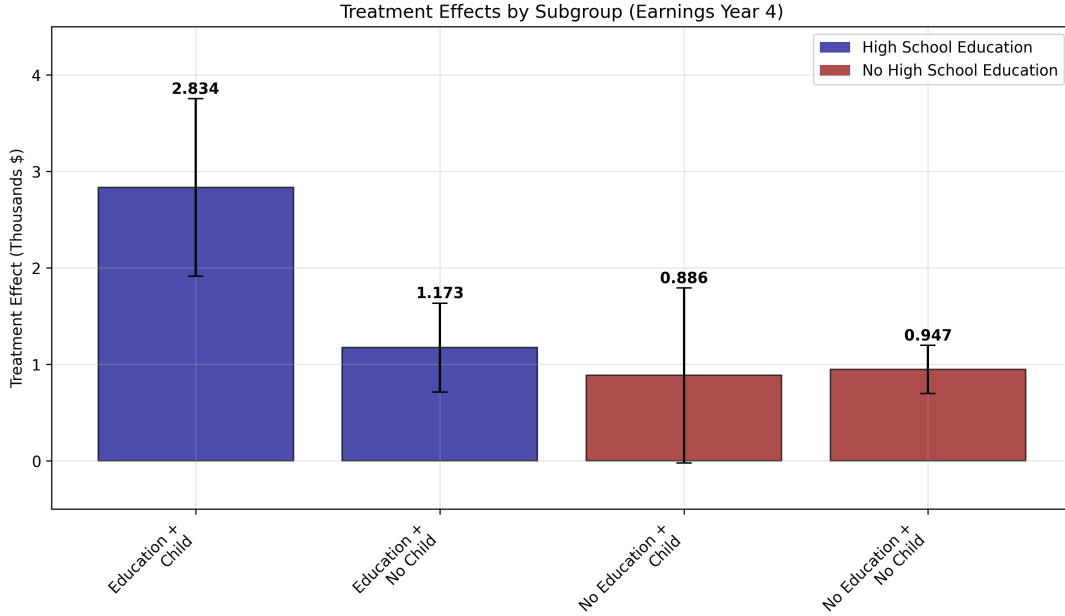| Subgroup	ATE	Variance	Sample Size	Weight
Education + Child	2.8341	0.846568	519	0.0958
Education + No Child	1.1731	0.213235	2,316	0.4274
No Education + Child	0.8861	0.823355	368	0.0679
No Education + No Child	0.9471	0.062023	2,216	0.4089

Figure 2: Treatment Effects by Subgroup (Earnings Year 4)

1.3 Comparison of ATE Estimation Methods

Another observation is the difference between the overall ATE and the aggregate ATE calculated from the subgroups. Table 5 summarizes the comparison of both ATE estimation methods for both outcome periods.

Table 5: Comparison of ATE Estimation Methods

Outcome	Method	ATE Estimate	Variance	Std. Error
Earnings Year 1	Overall ATE	1.1362	0.0180	0.1342
Earnings Year 1	Aggregate ATE	1.1258	0.0181	0.1344
Earnings Year 4	Overall ATE	1.2323	0.0619	0.2488
Earnings Year 4	Aggregate ATE	1.2203	0.0609	0.2468

The small but systematic differences between the overall and aggregate ATE estimates arise because treatment assignment is not perfectly balanced across subgroups. Specifically, the treatment proportions vary as follows:

- Education + Child: 79.0% treated

- Education + No Child: 81.3% treated

- No Education + Child: 83.2% treated

- No Education + No Child: 80.5% treated

- Overall: 80.9% treated

Subgroups with higher treatment proportions (e.g., No Education + Child) tend to have lower treatment effects, which pulls the overall ATE down slightly compared to the aggregate ATE that weights subgroups equally by their population share. This imbalance in treatment assignment across subgroups leads to the observed differences in ATE estimates.

We can interpret these differences by considering stratification. It appears that despite the miniscule differences, the experiment was stratified ex-ante with respect to education and presence of small children. This stratification likely aimed to ensure balanced representation across these key demographic factors, which are known to influence treatment effects. The slight variations in treatment proportions across subgroups suggest that the randomization process maintained overall balance while allowing for some natural variation within strata. This stratified design helps improve the precision of ATE estimates by controlling for confounding variables related to education and family status. If the stratification had been done ex-post, we would not have been able to be assured that this was a positive contribution to the design of the experiment.

2 Question 2

2.1 Variance of the Difference Between the ATE Estimator and Sample ATT

We define the average effect for the treated (ATT) as:

$$\tau_t = \frac{1}{M} \sum_{i:W_i=1} (Y_i(1) - Y_i(0))$$

Let $\hat{\tau}$ denote the difference-in-means estimator for the ATE, and let τ_t denote the true Sample ATT (τ_{SATT}). We want to find an unbiased estimator for $\text{Var}(\hat{\tau} - \tau_t)$.

1. Point Estimator for ATT

The Sample ATT is unbiased for the Sample ATE, $E[\tau_t|Y(0), Y(1)] = \tau$. Relatedly, the difference-in-means (DiM) estimator is unbiased for the ATE, meaning $E[\hat{\tau} - \tau|Y(0), Y(1)] = 0$. Given this relationship, we can use the DiM estimator $\hat{\tau}$ as our estimator for the ATT parameter τ_t.

2. Unbiasedness of the Estimator

As established in the course materials (section 2 slides), the DiM estimator is an unbiased estimator for the Sample ATT. Therefore, the expectation of the estimation error is zero.

$$E_W[\hat{\tau} - \tau_t|Y(0), Y(1)] = 0$$

3. Derivation of the Variance

We begin with the definition of variance, conditional on the potential outcomes.

$$\text{Var}(\hat{\tau} - \tau_t|Y(0), Y(1)) = E[(\hat{\tau} - \tau_t)^2|Y(0), Y(1)] - (E[\hat{\tau} - \tau_t|Y(0), Y(1)])^2$$
$$= E[(\hat{\tau} - \tau_t)^2|Y(0), Y(1)] - 0^2 \quad \text{by unbiasedness}$$

The algebraic simplification of the estimation error is:

$$\hat{\tau} - \tau_t = \frac{N}{M(N-M)} \sum_{i=1}^{N} W_i Y_i(0) - \frac{1}{N-M} \sum_{i=1}^{N} Y_i(0)$$

Since the second term is a constant with respect to the randomization W, the variance is:

$$\text{Var}(\hat{\tau} - \tau_t | Y(0), Y(1)) = \text{Var}\left(\frac{N}{M(N-M)} \sum_{i=1}^{N} W_i Y_i(0) \,\middle|\, Y(0), Y(1)\right)$$

$$= \frac{N^2}{M^2(N-M)^2} \text{Var}\left(\sum_{i=1}^{N} W_i Y_i(0) \,\middle|\, Y(0), Y(1)\right)$$

The variance of the summation term has been previously shown to be:

$$\text{Var}\left(\sum_{i=1}^{N} W_i Y_i(0)\right) = \frac{M(N-M)}{N} S_0^2$$

Substituting this result back in, we get:

$$\text{Var}(\hat{\tau} - \tau_t | Y(0), Y(1)) = \frac{N^2}{M^2(N-M)^2} \cdot \frac{M(N-M)}{N} S_0^2$$

$$= \frac{N}{M(N-M)} S_0^2$$

4. An Unbiased Estimator for the Variance

To form an unbiased estimator for the variance derived above, we replace the unknown population parameter S_0^2 with its unbiased sample estimator, s_c^2.

$$\widehat{\text{Var}}(\hat{\tau} - \tau_t) = \frac{N}{M(N-M)} s_c^2$$

This estimator is unbiased because the sample variance of the control group outcomes, s_c^2, is an unbiased estimator for the population variance of the control potential outcomes, S_0^2.

$$E[s_c^2] = S_0^2$$

2.2 Variance of the Difference Between the ATE Estimator and Sample ATC

The average effect for the controls (ATC) is defined as:

$$\tau_c = \frac{1}{N-M} \sum_{i:W_i=0} (Y_i(1) - Y_i(0))$$

We want to find an unbiased estimator for $\text{Var}(\hat{\tau} - \tau_c)$.

The derivation in this section will mirror exactly that of part (a).

1. Point Estimator for ATC

The Sample ATC is unbiased for the Sample ATE, $E[\tau_c|Y(0), Y(1)] = \tau$. Relatedly again, the difference-in-means (DiM) estimator is unbiased for the ATE, meaning $E[\hat{\tau} - \tau|Y(0), Y(1)] = 0$. Given this relationship, we can use the DiM estimator $\hat{\tau}$ as our estimator for the ATC parameter τ_c.

2. Unbiasedness of the Estimator

As established in the course materials (section 2 slides again), the DiM estimator is an unbiased estimator for the Sample ATC. Therefore, the expectation of the estimation error is zero.

$$E_W[\hat{\tau} - \tau_c|Y(0), Y(1)] = 0$$

3. Derivation of the Variance

We begin with the definition of variance, conditional on the potential outcomes.

$$\text{Var}(\hat{\tau} - \tau_c|Y(0), Y(1)) = E[(\hat{\tau} - \tau_c)^2|Y(0), Y(1)] - (E[\hat{\tau} - \tau_c|Y(0), Y(1)])^2$$

$$= E[(\hat{\tau} - \tau_c)^2|Y(0), Y(1)] \quad \text{by unbiasedness}$$

The algebraic simplification of the estimation error is:

$$\hat{\tau} - \tau_c = \frac{N}{M(N-M)} \sum_{i=1}^{N} W_i Y_i(1) - \frac{1}{N-M} \sum_{i=1}^{N} Y_i(1)$$

Since the second term is a constant with respect to the randomization W, the variance is:

$$\text{Var}(\hat{\tau} - \tau_c | Y(0), Y(1)) = \text{Var}\left(\frac{N}{M(N-M)} \sum_{i=1}^{N} W_i Y_i(1) \middle| Y(0), Y(1) \right)$$

$$= \frac{N^2}{M^2(N-M)^2} \text{Var}\left(\sum_{i=1}^{N} W_i Y_i(1) \middle| Y(0), Y(1) \right)$$

The rest of the calculations are exactly symmetric to that in section (a), but with the roles of treated and control groups reversed. Extrapolating from previous work, we get:

$$\text{Var}(\hat{\tau} - \tau_c | Y(0), Y(1)) = \frac{N^2}{M^2(N-M)^2} \cdot \frac{M(N-M)}{N} S_1^2$$

$$= \frac{N}{M(N-M)} S_1^2$$

4. An Unbiased Estimator for the Variance

To form an unbiased estimator for the variance, we follow the same approach as in section (a), replacing S_1^2 with its unbiased sample estimator s_t^2:

$$\widehat{\text{Var}}(\hat{\tau} - \tau_c) = \frac{N}{M(N-M)} s_t^2$$

2.3   Bonus Question

Yes, we can use the results from parts (a) and (b) to form a conservative estimator for the variance of the difference-in-means estimator $\hat{\tau}$. Intuitively, we can think of it as the following: our overall treatment effect is a weighted average of our treatment effect on the treated and our treatment effect on the controls. Thus one way to estimate the variance of the ATE would be to sum the variance of the ATT and the ATU; yet this will be a biased approximation because the ATT and the ATU are not independent. Since them

being correlated would reduce variance overall, our approximation based on summing up the two variances will be conservative and overestimate the actual variance. This exactly mirrors the logic of the Neyman Variance estimator, and they are in fact mathematically equivalent as well.

Mathematical Derivation

To show this relationship formally, we start with the sum of the variances from parts (a) and (b):

$$\widehat{\text{Var}}(\hat{\tau} - \tau_c) + \widehat{\text{Var}}(\hat{\tau} - \tau_{ATC}) = \frac{N}{M(N-M)} s_c^2 + \frac{N}{M(N-M)} s_t^2$$
$$= \frac{N}{M(N-M)} (s_c^2 + s_t^2)$$

The Neyman variance estimator is given by:

$$\widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau}) = \frac{s_t^2}{M} + \frac{s_c^2}{N-M}$$

To show these are equivalent, we can rewrite the Neyman variance with a common denominator:

$$\widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau}) = \frac{s_t^2}{M} + \frac{s_c^2}{N-M}$$
$$= \frac{s_t^2(N-M) + s_c^2 M}{M(N-M)}$$

Applying the correction factors $\alpha = \frac{M}{N}$ and $\beta = \frac{N-M}{N}$ to our sum:

$$\alpha \cdot \widehat{\text{Var}}(\hat{\tau} - \tau_c) + \beta \cdot \widehat{\text{Var}}(\hat{\tau} - \tau_{ATC})$$
$$= \frac{M}{N} \cdot \frac{N s_c^2}{M(N-M)} + \frac{N-M}{N} \cdot \frac{N s_t^2}{M(N-M)}$$
$$= \frac{s_c^2}{N-M} + \frac{s_t^2}{M}$$
$$= \widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau})$$

Therefore, the corrected sum equals the Neyman variance:

$$\frac{M}{N} \cdot \widehat{\text{Var}}(\hat{\tau} - \tau_c) + \frac{N-M}{N} \cdot \widehat{\text{Var}}(\hat{\tau} - \tau_{ATC}) = \widehat{\text{Var}}_{\text{Neyman}}(\hat{\tau})$$

The correction factors $\frac{M}{N}$ and $\frac{N-M}{N}$ represent the proportions of treated and control units in the sample, respectively, providing the proper weighting to make this relationship exact.