Exercise 1: Experiments
(Applied Microeconometrics)

Alexander Eriksson Byström Erik Alm Ramberg Jacob Lindberg
Nathalie Hånell

1 Descriptive Statistics

1.A Generating Variables

See attached STATA code.

1.B Descriptive Statistics of New Variables

Table 1 presents the means of student characteristics across the three treatment arms: small classes, regular classes, and regular classes with a teacher's aide. The final column reports the P-value from a joint F-test for the equality of means across all three categories.

1.C Testing for Randomization (Balance Test)

To see if the randomization actually worked, we need to check if the student groups looked the same at the starting line. We test the null hypothesis (H_0) that there are no significant differences in the means of baseline characteristics, specifically Free Lunch, Race, and Age, across the different treatment groups.

To do this, we run a separate OLS regression for each characteristic. Each row in Table 1 essentially summarizes one of these models. For example, the row for "Free Lunch" represents the following regression:

$$\text{Free Lunch}_i = \underbrace{\beta_0}_{\text{Small Mean}} + \beta_1(\text{Regular}_i) + \beta_2(\text{Regular+Aide}_i) + \varepsilon_i \tag{1.1}$$

For each model, we perform an F-test for the joint significance of the treatment indicators ($H_0 : \beta_1 = \beta_2 = 0$). As long as we fail to reject this null (i.e., we get a high p-value), we can conclude that the groups are "balanced". This means no specific class type was accidentally stacked with students of a certain background, ensuring our different class sizes have similar pre-treatment status.

1.C.1 Interpretation of Results

The results generally support the claim that the initial assignment in Kindergarten was random. In Panel A, the P-values for Free Lunch (0.246), Race (0.238), and Age (0.341) are all well above the standard 5% significance level. This indicates that we fail to reject the null hypothesis of balanced characteristics, suggesting that students were distributed across class types without systematic bias.

However, as the experiment progressed into later grades (Panels B, C, and D), several P-values for Race and Age became statistically significant ($P < 0.05$). While the lack of a clear, consistent pattern across all variables suggests that balance was maintained in many dimensions, these significant F-values indicate that the data is not perfectly random in later years. This potential "unbalancing" is likely due to non-random attrition or student mobility between class types after the initial Kindergarten assignment. Despite these exceptions, the overall results for Kindergarten provide strong evidence for the integrity of the initial experimental design.

Table 1: Student Characteristics and Joint P-values by Grade

Variable	Small	Regular	Regular+Aide	Joint P-value
Panel A: Kindergarten				
Free Lunch	0.468	0.480	0.495	0.246
White/Asian	0.686	0.676	0.660	0.238
Age in 1985	5.257	5.238	5.239	0.341
Average Score	465.803	458.943	459.616	0.000
Class size	15.396	22.375	23.257	0.000
Panel B: Grade 1				
Free Lunch	0.438	0.468	0.440	0.220
White/Asian	0.689	0.630	0.704	0.000
Age in 1985	5.333	5.423	5.424	0.000
Average Score	534.646	519.456	525.740	0.000
Class size	15.690	22.689	23.420	0.000
Panel C: Grade 2				
Free Lunch	0.411	0.428	0.406	0.587
White/Asian	0.684	0.645	0.652	0.033
Age in 1985	5.428	5.517	5.509	0.000
Average Score	588.757	578.729	581.359	0.000
Class size	15.310	23.464	23.469	0.000
Panel D: Grade 3				
Free Lunch	0.383	0.391	0.405	0.614
White/Asian	0.690	0.677	0.656	0.075
Age in 1985	5.487	5.545	5.554	0.003
Average Score	621.863	614.685	614.548	0.000
Class size	15.714	23.610	24.435	0.000

Note: Joint P-values refer to the F-test for equality of means across the three class types. Average score is the mean of reading and math scaled scores.

Comparing Table 1 in this paper with the original Table 1 from Krueger (1999), we find that our results are mostly the same but vary slightly due to differences in the dataset.

2 Random Assignment

2.A Treatment Randomly Assigned to Students and Teachers?

Balance holding across grades should imply randomness; if we fail to reject the null, there is no statistical difference between the beta coefficients. This suggests that the differences in outcome variables (e.g., school lunch, race, and age) are randommeaning they are independent of class-size assignment.

However, this statistical insignificance does not prove randomness; it only indicates that we cannot reject the possibility of it. True randomness is derived from the experimental design itself, not merely from the statistical results. While these results provide increased confidence in the study's integrity, they come with caveats: based on the F-test results, we cannot reject the null, meaning randomness is possible, but this conclusion remains a statistical inference rather than an absolute certainty.

References

Krueger, Alan B. (1999). “Experimental Estimates of Education Production Functions”. In: *Quarterly Journal of Economics*, pp. 497–532.