1 Experiment Set Difference: Conv (3-10 words) v. Conv (3-50 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	-0.155504 +/- 0.00967989 (stderr) (p=2.72289e-06, Welch's t-test)
gpu_hours	-0.123129 +/- 0.00263609 (stderr) (p=3.84025e-10, Welch's t-test)
estimated_carbon_impact_kg	-0.0112125 +/- 0.000557686 (stderr) (p=1.3982e-07, Welch's t-test)
total_power	-0.0303978 +/- 0.00151308 (stderr) (p=7.37246e-08, Welch's t-test)
kw_hr_gpu	-0.0155905 +/- 0.000714091 (stderr) (p=1.1381e-07, Welch's t-test)
kw_hr_cpu	-0.00364857 +/- 0.000271309 (stderr) (p=4.52414e-08, Welch's t-test)
exp_len_hours	-0.153872 +/- 0.00996196 (stderr) (p=9.41656e-07, Welch's t-test)
average_realtime_carbon_intensity	-5.07801 +/- 1.69638 (stderr) (p=0.00421635, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-0.7775194444444444
gpu_hours	-0.6156468614782007
estimated_carbon_impact_kg	-0.05606237365737079
total_power	-0.1519890007138038
kw_hr_gpu	-0.07795272568419709
kw_hr_cpu	-0.018242844387830637
exp_len_hours	-0.7693579218122694
average_realtime_carbon_intensity	-25.390055544718507