1 Experiment Set Difference: Conv (25-50 words) v. Conv (3-10 words)

1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	0.236844 +/- 0.00941818 (stderr) (p=3.24998e-07, Welch's t-test)
gpu_hours	0.190105 +/- 0.00122813 (stderr) (p=1.85895e-11, Welch's t-test)
estimated_carbon_impact_kg	0.0160897 +/- 0.000519002 (stderr) (p=7.09323e-09, Welch's t-test)
total_power	0.046412 +/- 0.00137469 (stderr) (p=1.31286e-09, Welch's t-test)
kw_hr_gpu	0.0238534 +/- 0.000585434 (stderr) (p=5.04525e-10, Welch's t-test)
kw_hr_cpu	0.00552135 +/- 0.000302024 (stderr) (p=1.78828e-08, Welch's t-test)
exp_len_hours	0.234226 +/- 0.00970021 (stderr) (p=7.99905e-08, Welch's t-test)
average_realtime_carbon_intensity	-11.1796 +/- 3.8501 (stderr) (p=0.0026362, Welch's t-test)

1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	1.1842222222222223
gpu_hours	0.9505255814841831
estimated_carbon_impact_kg	0.0804483640218587
total_power	0.23206024355232593
kw_hr_gpu	0.11926683125994801
kw_hr_cpu	0.027606740608612686
exp_len_hours	1.1711313005950714
average_realtime_carbon_intensity	-55.89817765172688