

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

## 1.1 Average Treatment Effect (Per Experiment in Experiment Set)

cpu_hours	0.242976 +/- 0.0090914 (stderr) (p=1.20212e-06, Welch's t-test)
gpu_hours	0.177957 +/- 0.000575422 (stderr) (p=1.66408e-19, Welch's t-test)
estimated_carbon_impact_kg	0.0153946 +/- 0.000519173 (stderr) (p=8.97353e-09, Welch's t-test)
total_power	0.0428601 +/- 0.00128075 (stderr) (p=4.0351e-08, Welch's t-test)
kw_hr_gpu	0.0222813 +/- 0.000523414 (stderr) (p=1.35719e-07, Welch's t-test)
kw_hr_cpu	0.00484536 +/- 0.000330046 (stderr) (p=1.1213e-08, Welch's t-test)
exp_len_hours	0.227741 +/- 0.00903057 (stderr) (p=1.50087e-06, Welch's t-test)
average_realtime_carbon_intensity	21.8621 +/- 5.96295 (stderr) (p=0.000305154, Welch's t-test)

## 1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	1.214880555555557
gpu_hours	0.8897851658911947
estimated_carbon_impact_kg	0.07697291000911916
total_power	0.21430032018615155
kw_hr_gpu	0.1114062941878385
kw_hr_cpu	0.024226819854029552
exp_len_hours	1.138704908357726
average_realtime_carbon_intensity	109.31035582397271