

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

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

cpu_hours	-0.264345 +/- 0.00316559 (stderr) (p=2.21036e-14, Welch's t-test)
gpu_hours	-0.252667 +/- 0.00220196 (stderr) (p=6.66575e-14, Welch's t-test)
estimated_carbon_impact_kg	-0.0253424 +/- 0.0020437 (stderr) (p=2.2067e-05, Welch's t-test)
total_power	-0.0707536 +/- 0.00651002 (stderr) (p=3.8621e-05, Welch's t-test)
kw_hr_gpu	-0.0324651 +/- 0.000389227 (stderr) (p=5.63907e-15, Welch's t-test)
kw_hr_cpu	-0.0123157 +/- 0.00426054 (stderr) (p=0.00709199, Welch's t-test)
exp_len_hours	-0.275319 +/- 0.00406262 (stderr) (p=1.2072e-11, Welch's t-test)
average_realtime_carbon_intensity	2.25215 +/- 8.1213 (stderr) (p=0.65041, Welch's t-test)

## 1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-1.321725
gpu_hours	-1.2633338942855816
estimated_carbon_impact_kg	-0.12671224787761112
total_power	-0.35376812787680545
kw_hr_gpu	-0.16232561915635468
kw_hr_cpu	-0.06157825924668678
exp_len_hours	-1.3765927173031702
average_realtime_carbon_intensity	11.260726880213497