

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

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

cpu_hours	-0.166348 +/- 0.00494929 (stderr) (p=5.10213e-10, Welch's t-test)
gpu_hours	-0.166124 +/- 0.00318604 (stderr) (p=1.53706e-09, Welch's t-test)
estimated_carbon_impact_kg	-0.0124861 +/- 0.000399377 (stderr) (p=2.53384e-10, Welch's t-test)
total_power	-0.0438578 +/- 0.0014175 (stderr) (p=3.19039e-09, Welch's t-test)
kw_hr_gpu	-0.0210945 +/- 0.000468421 (stderr) (p=8.23936e-12, Welch's t-test)
kw_hr_cpu	-0.00666361 +/- 0.000482332 (stderr) (p=2.59334e-06, Welch's t-test)
exp_len_hours	-0.177935 +/- 0.00559932 (stderr) (p=2.43238e-11, Welch's t-test)
average_realtime_carbon_intensity	57.5477 +/- 7.76141 (stderr) (p=0.000116966, Welch's t-test)

## 1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-0.831738888888889
gpu_hours	-0.8306200807991735
estimated_carbon_impact_kg	-0.062430654725577024
total_power	-0.21928878185872638
kw_hr_gpu	-0.10547232299913036
kw_hr_cpu	-0.033318045265886345
exp_len_hours	-0.8896733689970441
average_realtime_carbon_intensity	287.7383341066295