

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

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

cpu_hours	-0.108841 +/- 0.00955952 (stderr) (p=1.91058e-05, Welch's t-test)
gpu_hours	-0.129537 +/- 0.00301578 (stderr) (p=3.428e-12, Welch's t-test)
estimated_carbon_impact_kg	-0.01413 +/- 0.002096 (stderr) (p=0.000145618, Welch's t-test)
total_power	-0.0403558 +/- 0.00662179 (stderr) (p=0.00027405, Welch's t-test)
kw_hr_gpu	-0.0168746 +/- 0.00071866 (stderr) (p=5.91824e-08, Welch's t-test)
kw_hr_cpu	-0.00866708 +/- 0.00426345 (stderr) (p=0.0232942, Welch's t-test)
exp_len_hours	-0.121447 +/- 0.00950496 (stderr) (p=1.1625e-05, Welch's t-test)
average_realtime_carbon_intensity	7.33016 +/- 7.96802 (stderr) (p=0.181638, Welch's t-test)

## 1.2 Cumulative Difference (Per Experiment Set)

cpu_hours	-0.5442055555555556
gpu_hours	-0.6476870328073809
estimated_carbon_impact_kg	-0.07064987422024034
total_power	-0.20177912716300167
kw_hr_gpu	-0.08437289347215758
kw_hr_cpu	-0.04333541485885614
exp_len_hours	-0.6072347954909008
average_realtime_carbon_intensity	36.650782424932004