

## Transformer test loss vs. learning rate and number of samples

