

HW week 12

w203: Statistics for Data Science

w203 teaching team

OLS Inference

The file `videos.txt` contains data scraped from Youtube.com. You wish to measure how much the quality of a video affects the number of views. To proxy video quality, you have the variable *rate* which represents average user rating. For the number of views, you have the variable *views*.

1. Using diagnostic plots and numerical tools, build a *single* model specification for this task.
2. Explain your result in terms of statistical and practical significance.
3. Identify 3 omitted variables and estimate whether they are biasing the effect you measure towards zero or away from zero.