Histogram of p_y before data cleaning