

this test doesn't make sense –
why would we do bootstrap rather than
just to measure how likely is that we get the result
as extreme if we just sample two groups from general population
which is exactly what permutation test does?

