3) We will be working a lot with matching based techniques. One of the best thinkers/writers on the topic of matching is Elizabeth Stuart from Johns Hopkins. For this problem, take a look at her paper: "Matching Methods for Causal Inference: A Review and a Look Forward." In lecture 01 you were introduced to "balance tables" (a.k.a. "Table 1") which summarizes the covariate distribution of the observations. A handful of questions: (a) as concisely as possible, state why we focus on balance assessments as part of our argumentation when attempting to perform causal inference, (b) in addition to a balance table, name other tools used to report balance, (c) why do we use standardized mean differences instead of p-values to assess balance when assessing the quality of a match design?, and (d) why is it kinda weird to use a p-value of the covariates in a randomized trial to assess balance?

(a) If the contrast groups start of imbalanced pre-intervention then it is hard to attributed a difference in the outcome post-intervention to the intervention. Note: this kind of argument means balance is a necessary (but not sufficient) condition for causal inference.

(b) Histogram of covariates and/or propensity score, Love plot of pre- and post-matched sets of covariates.

(c) See section 4.1 of the paper (page 11), the paragraph starting "Although common, hypothesis tests and p-values…" for the two main arguments against p-values. The argument for SMDs is that they standardized the observed difference between the groups relative to the observed standard deviation, making the difference contextually meaningful ("small" or "large" relative to a standard). This standardization is a common technique in statistics, most famously deployed in Cohen's d.

(d) This argument doesn't explicitly appear in Stuart's paper. But it is in there. In a randomized trial, the researcher randomized the observations. The p-value assesses the covariate balance under the hypothesis that the data were randomized, which we know is true by construction. So what happens if we get a small p-value (say p-value << 0.05)? Would we reject the null hypothesis? But we know for a fact the null is true because we made it true by randomizing, and the baseline covariates can't be changed by the randomization. The usual logic deployed in a hypothesis test aren't appropriate in assessing balance of a randomization. This is a bit more hand-wavy: it is not clear that the p-value is answering a meaningful question – we might want something like: "will the two groups produce the same average outcome?", but the p-value is answering a question more like: "are the two groups different enough that we can reliably tell them apart?"