MATH 343 / 643 Homework #3

Natasha Watson

Monday 19th May, 2025

Problem 1

Consider the Poisson linear regression model with one feature, time:

$$Y_1, Y_2, \ldots, Y_n \mid t_1, t_2, \ldots, t_n \overset{ind}{\sim} \text{Poisson}\left(e^{\beta_0+\beta_1 t_i}\right)$$

and consider a Bayesian approach to inference.

(a) [easy] What is the parameter space for the two parameters of interest?

$$\beta_0, \beta_1 \in \mathbb{R}$$

(b) [easy] Assume a flat prior $f(\beta_0, \beta_1) \propto 1$. Find the kernel of the posterior distribution $f(\beta_0, \beta_1 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n)$.

$$f(\beta_0, \beta_1 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n) \propto f(\, y_1, \ldots, y_n, t_1, \ldots, t_n | \beta_0, \beta_1 \,) f(\beta_0, \beta_1)$$

Assume a flat prior $f(\beta_0, \beta_1) \propto 1$

$$f(\beta_0, \beta_1 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n) \propto f(\, y_1, \ldots, y_n, t_1, \ldots, t_n | \beta_0, \beta_1 \,)$$

$$\propto \Pi_{i=1}^{n} \frac{e^{-}\theta_i \theta_i^{y_i}}{y_i!}$$

Plug in $\theta_i = e^{\beta_0+\beta_1 t_i}$

$$\propto e^{n\bar{y}\beta_0+\beta_1 \sum_{i=1}^{n} t_i y_i - e^{\beta_0+\beta_1 t_i}}$$

(c) [easy] Find the log of the kernel of the posterior distribution.

$$\ln\left(f(\beta_0, \beta_1 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n)\right) \propto n\bar{y}\beta_0 + \beta_1 \sum_{i=1}^{n} t_i y_i - e^{\beta_0+\beta_1 t_i}$$

(d) [easy] Find the kernel of the conditional distribution $f(\beta_0 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n, \beta_1)$. Is it a brand name rv?

$$f(\beta_0 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n, \beta_1) \propto e^{n\bar{y}\beta_0 + \sum_{i=1}^n -e^{\beta_0+\beta_1 t_i}} \not\propto \text{known distribution}$$

(e) [easy] Find the kernel of the conditional distribution $f(\beta_1 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n, \beta_0)$. Is it a brand name rv?

$$f(\beta_1 \,|\, y_1, \ldots, y_n, t_1, \ldots, t_n, \beta_0) \propto e^{\beta_1 \sum_{i=1}^n t_i y_i - e^{\beta_0+\beta_1 t_i}} \not\propto \text{known distribution}$$

(f) [harder] [MA, not covered on the final] Given your answer in (a), the Supp $[\beta_0]$, provide a proposal distribution for the conditional distribution of β_0:

$q(\beta_0^* \,|\, \beta_{0_{t-1}}, y_1, \ldots, y_n, t_1, \ldots, t_n, \beta_1, \phi) =$

(g) [harder] [MA, not covered on the final] Given your answer in (a), the Supp $[\beta_1]$, provide a proposal distribution for the conditional distribution of β_1:

$q(\beta_1^* \,|\, \beta_{1_{t-1}} y_1, \ldots, y_n, t_1, \ldots, t_n, \beta_0, \phi) =$

Problem 2

This question is about basic causality, structural equation models and their visual representation as directed acyclic graphs (DAGs).

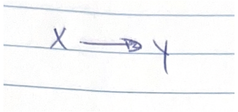
(a) [easy] We run a OLS to fit $\hat{y} = b_0 + b_1 x$ and find there is a statistically significant rejection of $H_0 : \beta_1 = 0$. If this test was decided correctly, what do we call the relationship between x and y? (The answer is one word).

correlation

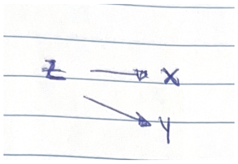
(b) [easy] If this test was decided incorrectly, what do we call the relationship between x and y? (The answer is two words).

spurious correlation

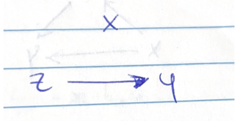
(c) [easy] Draw an example DAG where x causes y.



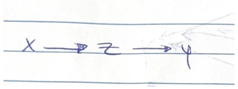
(d) [easy] Draw an example DAG where x is correlated to y but is not causal.



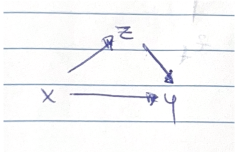
(e) [easy] Draw an example DAG that can result in a spurious correlation of x and y.



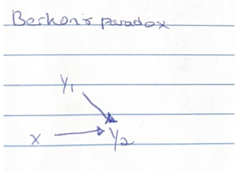
(f) [easy] Draw an example DAG where x causes y but its effect is fully blocked by z.



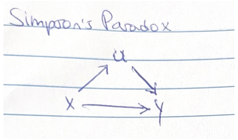
(g) [easy] Draw an example DAG where x causes y but its effect is partially blocked by z.



(h) [easy] Draw an example DAG that results in a Berkson's paradox between x and y_1. Denote the collider variable as y_2.



(i) [easy] Draw an example DAG that results in a Simpson's paradox between x and y. Denote the confounding variable as u.



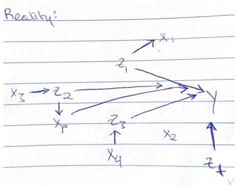
(j) [easy] In the previous Simpson's paradox DAG, provide an example structural equation for y and provide an example structural equation for x.

$$\hat{\boldsymbol{y}} = b_0 + b_1 x$$

If we control for z:

$$\hat{\boldsymbol{y}} = b_0 + b_1 x + b_2 z$$

(k) [easy] Consider observed covariates x_1, x_2, x_3 and phenomenon y. Draw a realistic DAG for this setting.



Problem 3

This question is about causal and correlational interpretations for generalized linear models.

(a) [easy] We run the following model on the diamonds dataset where y is the price of the diamond

```
> summary(lm(price ~ ., diamonds))

              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2184.477    408.197   5.352 8.76e-08 ***
carat        11256.978     48.628 231.494  < 2e-16 ***
cutGood        579.751     33.592  17.259  < 2e-16 ***
cutVery Good   726.783     32.241  22.542  < 2e-16 ***
cutPremium     762.144     32.228  23.649  < 2e-16 ***
cutIdeal       832.912     33.407  24.932  < 2e-16 ***
colorE        -209.118     17.893 -11.687  < 2e-16 ***
colorF        -272.854     18.093 -15.081  < 2e-16 ***
colorG        -482.039     17.716 -27.209  < 2e-16 ***
colorH        -980.267     18.836 -52.043  < 2e-16 ***
colorI       -1466.244     21.162 -69.286  < 2e-16 ***
colorJ       -2369.398     26.131 -90.674  < 2e-16 ***
claritySI2    2702.586     43.818  61.677  < 2e-16 ***
```

```
claritySI1    3665.472     43.634  84.005  < 2e-16 ***
clarityVS2    4267.224     43.853  97.306  < 2e-16 ***
clarityVS1    4578.398     44.546 102.779  < 2e-16 ***
clarityVVS2   4950.814     45.855 107.967  < 2e-16 ***
clarityVVS1   5007.759     47.160 106.187  < 2e-16 ***
clarityIF     5345.102     51.024 104.757  < 2e-16 ***
depth          -63.806      4.535 -14.071  < 2e-16 ***
table          -26.474      2.912  -9.092  < 2e-16 ***
x            -1008.261     32.898 -30.648  < 2e-16 ***
y                9.609     19.333   0.497    0.619
z              -50.119     33.486  -1.497    0.134
```

What is the interpretation of the b for `carat` (the unit of this feature is "carats")?

Holding all other variables in the model constant, an increase of 1 carat in diamond weight is associated with an \$ 11,256.98 increase in price.

(b) [difficult] What is the interpretation of the b for `cutIdeal` (note: the reference category for `cut` is `Fair`)?

Holding all other variables constant, a diamond with cut Ideal will be, on average, \$ 832.912 more than a diamond with cut Fair.

(c) [easy] We run the following model on the `Pima.tr2` dataset where y is 1 if the subject had diabetes or 0 if not.

```
> summary(glm(type ~ ., MASS::Pima.tr2, family = "binomial"))

             Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.773062   1.770386  -5.520 3.38e-08 ***
npreg        0.103183   0.064694   1.595  0.11073
glu          0.032117   0.006787   4.732 2.22e-06 ***
bp          -0.004768   0.018541  -0.257  0.79707
skin        -0.001917   0.022500  -0.085  0.93211
bmi          0.083624   0.042827   1.953  0.05087 .
ped          1.820410   0.665514   2.735  0.00623 **
age          0.041184   0.022091   1.864  0.06228 .
```

What is the interpretation of the b for `age` (the unit of this feature is age)?

For each additional year in age , the log-odds of having diabetes increases by 0.0412 when we hold all other variables constant.

We can express the effect of b_{age} more intuitively by exponentiating the coefficient to get the odds ratio i.e. $e^{0.0412} \approx 1.042$. Then perform another arithmetic to get the odds as a percentage i.e. $(1.0412 - 1) \times 100\% \approx 4.2\%$.

As such, we can also say: For each additional year in age, the odds of having diabetes increases by 4.2% when we hold all other predictors constant.

(d) [easy] What is the interpretation of the b for `glu` (the unit of this feature is mg/dL) if `glu` is known to be causal?

Causal interpretation means we interpret the effect of changing the glucose level rather than just observing a difference, assuming the model adjusts appropriately for confounders.

Therefore we say:

If blood sugar is increased by one mg/dL and all other measurements remain constant, then the log-odds of getting diabetes will resultingly increase by 0.032117 ± 0.006787 ($b_j \pm s_j$), assuming log-odds od getting diabetes is linear in the p covariates and the relationship remains stationary.

(e) [easy] We run the following model on the `phillippines` household dataset where y is the number of people living in a household.

```
> summary(MASS::glm.nb(total ~ ., read.csv("philippines_housing.csv")))

                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                        1.447108   0.088204  16.406  < 2e-16 ***
locationDavaoRegion               -0.011108   0.064367  -0.173  0.86298
locationIlocosRegion               0.053589   0.063284   0.847  0.39711
locationMetroManila                0.074016   0.056731   1.305  0.19201
locationVisayas                    0.131151   0.050440   2.600  0.00932 **
age                               -0.004896   0.001136  -4.309 1.64e-05 ***
roofPredominantly Strong Material  0.043376   0.052705   0.823  0.41051
```

What is the interpretation of the b for `age` (the unit of this feature is years)?

Note that we are now using a negative binomial regression and no longer using a logistic regression. However, we have a logit link.

We can say that a one-year increase in age leads to a 0.004896 decrease in the log of the expected number of people in the household, or, if we exponentiate b_{age}, subtract by 1 and then multiply by 100, we get the percent change i.e. $(e^{b_j} - 1) \times 100\%$.

In this case, $(e^{0.004896} - 1) \times 100\% \approx -0.49\%$. Therefore, we can say a one-year increase increase age leads to a 0.49% decrease in the number of people in the household.

(f) [easy] We run the following Weibull regression model on the `lung` dataset where y is survival of the patient.

```
> lung = na.omit(survival::lung)
> lung$status = lung$status - 1 #needs to be 0=alive, 1=dead
> summary(survreg(Surv(lung$time, lung$status) ~
      inst + sex + ph.ecog + ph.karno + wt.loss, lung))


               Value Std. Error     z       p
```

```
(Intercept)  7.13673    0.74732  9.55 < 2e-16
inst         0.02042    0.00877  2.33  0.0199
sex          0.39717    0.13852  2.87  0.0041
ph.ecog     -0.69588    0.15463 -4.50 6.8e-06
ph.karno    -0.01558    0.00749 -2.08  0.0376
wt.loss      0.00977    0.00525  1.86  0.0626
Log(scale)  -0.36704    0.07272 -5.05 4.5e-07
```

What is the interpretation of the b for `wt.loss` (the unit of this feature is lbs) if `wt.loss` is known to be causal?

If wt.loss increases by one pound and all other measurements remain constant, the log survival of the patient (years) will resultingly increase by 0.00977 ± 0.00525, assuming survival is Weibull distributed with log mean linear in the p covariates and the relationship is stationary.

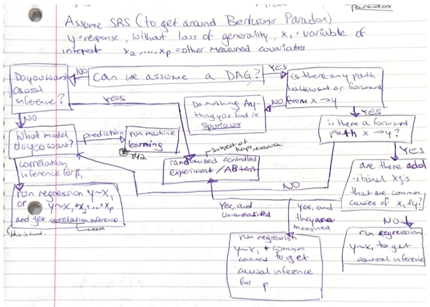
(g) [easy] What is the interpretation of the b for `ph.ecog` (the unit of this feature is mg/dL) if `ph.ecog` is known to be causal?

If ph.ecog increases by one mg/dL and all other measurements remain constant, the log survival of the patient (years) will resultingly decrease by 0.69588 ± 0.15463, assuming survival is Weibull distributed with log mean linear in the p covariates and the relationship is stationary.

Problem 4

This problem is about controlling values of variables to allow for causal inference.

(a) [easy] Redraw the "master decision tree" of what to do in every situation beginning with the root node of "Can we assume a DAG?"


Paradox
Assume SRS (to get around Berkson's Paradox)
y = response, without loss of generality, x_1 = variable of interest $x_2, \ldots, x_p$ = other measured covariates
Can we assume a DAG?
NO
YES
Do you want causal inference?
YES
NO
What model do you want?
prediction
run machine learning
☺ 342
correlation inference for β_1
run regression $y \sim x_1$ or $y \sim x_1 + x_2 + \ldots + x_p$ and get correlation inference
weak
this is hard.
is there any path backward or forward from $x \to y$
NO
Do nothing. Anything you find is Spurious
YES
is there a forward path $x \to y$?
NO
randomized controlled experiment / AB test
subject of Kap's research
YES
are there additional x_j's that are common causes of x_1 & y?
Yes, and Un-measured
Yes, and they are measured
run regression $y \sim x_1$ + common causes to get causal inference for p
NO
run regression $y \sim x_1$ to get causal inference

(b) [easy] Explain why controlling / manipulating the values of x allows for causal inference of x on y.

If we have controlled for all confounder (common causes among the covariates or lurking variables) then observed effect of x on y can be interpreted as causal.

Confounders are variables that affect both x and y and create spurious association between x and y. They cause issues in casual inference because they bias our estimate of the causal effect of x on y by mixing up the effect of x on y with the effect of z.

(c) [harder] Explain why a typical observational study (i.e. just collecting data and assembling it into $\mathbb{D}$) cannot allow for causal inference of x on y.

In observational studies, the researched does not intervene or control variables, and so there might be backdoor paths between x, y and z that cause spurious correlations. Because there are no controls, you cannot tell if x is actually causing y or is z is driving both x and y, and therefore you cannot use observational studies for causal inference.

(d) [easy] Give an example case (different from the one we spoke about in class) where controlling / manipulating the values of x is impossible.

If we were trying to understand the effects of early life environment on adult income, it would be impossible to control the circumstances of an adult's upbringing to see the effect on y, or to even control it if it is a covariate.

(e) [easy] Give an example case (different from the one we spoke about in class) where controlling / manipulating the values of x is unethical.

If we want to measure the effects of childhood trauma on social development, we can't just take groups of children and forcefully traumatize them to yield some usdeful data.

(f) [easy] Give an example case (different from the one we spoke about in class) where controlling / manipulating the values of x is impractical / unaffordable.

An example of where controlling covariates is unaffordable and/or impractical is if we are attempting to perform weather simulation and control. It is theoretically possible, but extremely expensive, technically complex and often impractical to implement at scale.

(g) [difficult] Assume in the `diamonds` dataset that the variable `cut` was manipulated by the experimenter prior to assessing the price y. This isn't absurd since raw diamonds can be cut differently but their color and clarity cannot be altered. Using the linear regression output from the previous problem, what is the interpretation of the b for `cutIdeal`. The reference category for this variable is `Fair`.

If cut of diamond is cutIdeal rather than cutFair and all other measurements are held constant, the price of the do=diamond will resultingly increase by an estimated \$832.912±\$33.407 dollars assuming cut is linear in the p covariates and the relationship remains stationary.

Problem 5

This problem is about randomized controlled trials (RCTs). Let n denote the number of subjects, let $\boldsymbol{w}$ denote the variable of interest which you seek causal inference for its effect.

Here we assume $\boldsymbol{w}$ is a binary allocation / assignment vector of the specific manipulation w_i for each subject (thus the experiment has "two arms" which is sometimes called a "treatment-control experiment" or "pill-placebo trial" or an "AB test". Let $\boldsymbol{y}$ denote the measurements of the phenomenon of interest for each subject and let $\boldsymbol{x}_{\cdot 1}, \ldots, \boldsymbol{x}_{\cdot p}$ denote the p baseline covariate measurements for each subject.

(a) [easy] How many possible allocations are there in this experiment?

$$2^n$$

(b) [easy] What are the three advantages of randomizing $\boldsymbol{w}$? We spoke about two main advantages and one minor advantage.

When we randomize $\boldsymbol{w}$ we know that it is independent of $\boldsymbol{U}$, the random variable that generates $\boldsymbol{u}_1 \ldots \boldsymbol{u}_n$. Consider:

$$\mathbb{E}_{\boldsymbol{W}}\left[\mathbb{E}_{\boldsymbol{\mathcal{E}}}\left[B_T\right]\right] = \mathbb{E}_{\boldsymbol{W}}\left[\beta_T + 2\frac{\beta_u}{n}(2\boldsymbol{u}^\top \boldsymbol{W} - \boldsymbol{u}^\top \boldsymbol{1})\right]$$

Notice, $2\frac{\beta_u}{n}(2\boldsymbol{u}^\top \boldsymbol{W} - \boldsymbol{u}^\top \boldsymbol{1} = 0$, so:

$$\mathbb{E}_{\boldsymbol{W}}\left[\mathbb{E}_{\boldsymbol{\mathcal{E}}}\left[B_T\right]\right] = \beta_T$$

Meaning, over all experiments, B_T is unbiased. This is the first advantage of randomizing $\boldsymbol{w}$.

The second advantage is that $\bar{u}_T - \bar{u}_C$ is "small". The third advantage is because, as per Fischer, it is a "reasoned basis for inference" i.e. the $\boldsymbol{W}$ variable has no effect on any response under the null hypothesis.

(c) [easy] In Fisher's Randomization test, what is the null hypothesis? Explain what this really means.

The strong or "sharp" null:

$$H_0 : y_i(w_i = 1) = y_i(w_i = 0)$$

For every subject, the outcome is the same regardless of treatment. This is stronger than just saying "treatment effect $=$ 0" as it is claiming no effect on anyone.

(d) [easy] Explain step-by-step how to run Fisher's Randomization test.

Under the null hypothesis that treatment has no effect, the observed outcomes are fixed and any difference in the means must be due to the random assignment of the treatment. To test this, begin by generating randomized reassignments of treatment labels keeping the number of treated units the same as in the original experiment.

Then randomly permute the treatment vector i.e. shuffle who is labeled as treatment vs. control. Repeat this a large number of times (i.e. 1000).

For each permuted treatment vector, compute the test statistic $b_T = \bar{y}_T - \bar{y}_C$ to build the randomization distribution and then compute p-values to draw a conclusion on whether to reject or not.

Assume now that Let $\boldsymbol{Y} = \beta_0 \boldsymbol{1}_n + \beta_T \boldsymbol{w} + \boldsymbol{\mathcal{E}}$ where $\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim}$ mean zero and has homoskedastic variance σ^2.

(e) [easy] What this the parameter of interest in causal inference? What is its name? The population average treatment effect (PATE):

$$\beta_T$$

(f) [easy] Assume we employ OLS to estimate β_T. We proved previously that OLS estimators are unbiased for any error distribution with mean zero. Find the $\mathbb{MSE}\left[B_T\right]$.

$$\mathbb{MSE}\left[B_T\right] = \mathbb{V}\text{ar}\left[B_T\right] = \sigma^2 \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}_{2,2}$$

(g) [easy] Prove that the optimal $\boldsymbol{w}$ has equal allocation to each arm.

$$\boldsymbol{w}_* = \arg\min \left\{\sigma^2 \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}_{2,2}\right\} = \arg\min \left\{\frac{1}{1-p_T} \frac{1}{p_T}\right\} \Rightarrow \arg\max \left\{p_T(1-p_T)\right\} \Rightarrow \boldsymbol{w} : p_T = \frac{1}{2}$$

(h) [easy] Explain how to run an experiment using the *completely randomized design.*

Once you define the population i.e. n = the total number of experimental units, you randomly assign each unit to an arm n_T units to treatment and $n_C = n - n_T$ units to control. This way, each subject has the same and independent chance of being assigned to any group.

Assume now that Let $\boldsymbol{Y} = \beta_0 \boldsymbol{1}_n + \beta_T \boldsymbol{w} + \beta_1 \boldsymbol{x}_{\cdot 1} + \ldots + \beta_p \boldsymbol{x}_{\cdot p} + \boldsymbol{\mathcal{E}}$ where $\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim}$ mean zero and have homoskedastic variance σ^2.

(i) [difficult] Prove that B_T is unbiased over the distribution of $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{W}$.

$\boldsymbol{Y} = f(\beta_T, \boldsymbol{w}, \boldsymbol{X}) +$ is called the population model assumption.

$$\mathbb{E}_{\boldsymbol{\mathcal{E}}}\left[\boldsymbol{Y}\right] = \mathbb{E}_{\boldsymbol{\mathcal{E}}}\left[\beta_0 \boldsymbol{1}_n + \beta_T \boldsymbol{w} + \boldsymbol{\mathcal{E}}\right] = \beta_0 \boldsymbol{1}_n + \beta_T \boldsymbol{w}$$

$$\Rightarrow \mathbb{E}\left[Y_i \mid w_i = 0\right] = \beta_0$$

$$\Rightarrow \mathbb{E}\left[Y_i \mid w_i = 1\right] = \beta_0 + \beta_T$$

The difference between the two isolates what we are looking for:

$$\mathbb{E}\left[Y_i \mid w_i = 1\right] - \mathbb{E}\left[Y_i \mid w_i = 0\right] = \beta_T$$

(j) [easy] What is the purpose using a *restricted design*? That is, using a set of allocations that is a subset of the full set of the completely randomized design.

When we restrict the design to not allow all $\binom{n}{\frac{n}{2}}$ allocations because some $\boldsymbol{w}$'s result in large $\bar{x}_{j,T} - \bar{x}_{j,C}$'s

(k) [harder] Explain how to run an experiment using Fisher's *blocking design* where you block on $\boldsymbol{x}_{\cdot 1}$, a factor with three levels and $\boldsymbol{x}_{\cdot 2}$, a factor with two levels.

If p =2 categories, we assume $L = 3$ blocks for the first co-variate ($\boldsymbol{x}_{\cdot 1}$) and $L = 2$ blocks for the second co-variate ($\boldsymbol{x}_{\cdot 2}$). Then we have B = 6 blocks total, where we parition our units into these blocks and then randomize the treatment within each block.

(l) [easy] What are the two main disadvantages to using Fisher's *blocking design*?

The two main disadvantages to running Fischer's blocking design is (1) the number of blocks increases exponentially in p and very soon the number of blocks outnumbers the number of data points; so you can only block a few features. (2) you get a loss of generalizability.

(m) [easy] Explain how to run an experiment using Student's *rerandomization design* where you let the imbalance metric be

$$\sum_{j=1}^{p} \frac{\left|\bar{x}_{j_T} - \bar{x}_{j_C}\right|}{s^2_{x_{j_T}}/(n/2) + s^2_{x_{j_C}}/(n/2)}$$

Student's rerandomization design involves rejecting randomizations that produce poor covariate balance. We define a distance function on all covariates we think matter i.e. $d(\boldsymbol{X}, \boldsymbol{w}) = \sum_{j=1}^{p} \left(\frac{\bar{x}_{s_T} - \bar{x}_{j_C}}{s_j}\right)$, starting with $\boldsymbol{w}$ from BCRD. If $d(\boldsymbol{X}, \boldsymbol{w}) > d_{\text{th}}$ then draw $\boldsymbol{w}$ again. Generate R $\boldsymbol{w}$'s and take the best subset of that.

(n) [easy] Explain how to run an experiment using the *pairwise matching design.*

We start with normalizing $\boldsymbol{X}$ such that all the covariates have average = 0 and standard deviation = 1.

We then define a distance function $d(\boldsymbol{x}_k, \boldsymbol{x}_l) = \sum_{j=1}^{p} big(x_{k,j} - x_{l,j}\big)^2$. Calculate $\mathcal{D}$, the n by n upper triangle matrix of all $d(\boldsymbol{x}_k, \boldsymbol{x}_l)$ i.e. $\frac{n^2}{2} - n$ computations.

Compute pairs $i_{(1)_1}, i_{(1)_2}, \ldots\ , i_{(\frac{n}{2})_1}, i_{(\frac{n}{2})_2}$ such that $d(x_{i_{(1)_1}}, x_{i_{(1)_2}}) + \cdots + d(x_{i_{(\frac{n}{2})_1}}, x_{i_{(\frac{n}{2})_2}})$ is minimal

Then, for each pair randomly make assignments with probability 50%.

(o) [easy] Does the pairwise matching design provide better imbalance on the observed covariates than the rerandomization design? Y/N

Yes, because it forces perfect balance within each pair by construction, while rerandomization merely selects from random assignments.