

## Problem Set #2

Jules Tucher

2025-12-09

Please briefly discuss why you chose these data and what they contain. Also, please identify and discuss the “outcome” and “treatment” variables of interest to you.

I chose these data because it's the same dataset that I am working with for my replication final paper for this class. The dataset contains ~78,000 observations at the couple-level, where each observation contains data for one male-female heterosexual couple in Germany. The main research question surrounds the effect on overall life satisfaction as a result of a wife outearning her husband. While the full dataset includes covariates for both the husband and the wife, this exercise will focus on the covariates and outcome for the male spouse. Additionally, due to limited computational resources, I pull a random sample of 10% of the full sample for this exercise. Given power analysis in the pre-analysis plan, a sample size of around 7,800 should be sufficient to estimate treatment effect. The treatment variable is a binary indicator variable representing whether the wife out-earns the husband. The outcome variable for this assignment is overall life satisfaction measured on an ordinal scale.

```
# pull random sample to reduce complexity (10% should be sufficient, given power analysis)
set.seed(5678)
matching_sample <- dplyr::slice_sample(analysis_data, prop=0.1)

# summarize treatment variable
matching_sample %>%
  dplyr::summarize(
    across(
      all_of(c("treatment")),
      list(mean = mean, sd = sd),
      .names = "{.col}_{.fn}"
    ),
    n=n()
  )
```

```
## # A tibble: 1 x 3
##   treatment_mean treatment_sd      n
##   <dbl>          <dbl> <int>
## 1      0.155        0.362  7873
```

From the summary, we can see that around 15.5% of the sample ( $N = 7,873$ ) is in the treatment group (i.e. where the wife outearns the husband).

Then, explain which variables are included in your balance and/or X matrices and why. Describe the settings you chose for the various GenMatch arguments (e.g. what estimand have you selected and why?).

Variables included for balance are the covariates we are interested in for the RDD: wage, age, number of children, and whether the husband has a college degree. I will also include region (East or West Germany) as a matching control, since this is a variable of interest in the analysis. Finally, since we are mimicking a regression discontinuity design (RD), include a scaled version of the running variable as a matching control. This will preserve the RD constraint of only including observations close to the threshold. See the following distributions for those five variables.

```
# store names of covariate variables-- male and female
covs = c("age", "wage", "nchildren", "college", "west")

# summarize matching variables
matching_sample %>%

# generate mean and SD of each matching variable for full sample
dplyr::summarize(
  across(
    all_of(covs),
    list(min=min, mean=mean, max=max),
    .names = "{.col}_{.fn}"
  )
) %>%

# reshape for readability
pivot_longer(everything()) %>%
separate(col = "name", into = c("variable", "stat"), sep = "_m") %>%
pivot_wider(id_cols = "variable", names_from = "stat", names_prefix = "m")
```

```
## # A tibble: 5 x 4
##   variable    min      mean      max
##   <chr>      <dbl>    <dbl>   <dbl>
## 1 age         20     45.2     64
## 2 wage        50 35359. 1440000
## 3 nchildren    0     0.977     6
## 4 college     0     0.347     1
## 5 west        0     0.797     1
```

Set parameters for GenMatch method according to design characteristics. The treatment vector is pulled directly from the data. The covariates matrix (equal, in this case, to the balance matrix), includes the relevant covariates. Declare and include a vector of caliper measures, which will constrict matching distances on included variables. I set caliper distances only for the region covariate and the standardized running variable. Caliper for region (value is 0) is included to prevent East-West matches. Caliper for standardized running variable (value is 0.25 SD) to maintain RD design and match close to the threshold. Use M = 1 for 1:1 matching. Since we are interested in the impact of the wife out earning her husband among actual treated couples, set the estimand to ATT. Population size, max generations, and wait generations are set to optimize runtime while also having meaningful results. Use replacement and allow for ties, since we have a relatively small proportion of the sample in the treatment group.

```

# treatment vector
t = as.vector(matching_sample$treatment)

# covariate matrix
X = as.matrix(with(matching_sample,
                    cbind(age, wage, nchildren, college, west, scale(diff_wage_OR))))

# caliper vector -- only care about region and running variable
caliper_vec <- c(NA, NA, NA, NA, 0, 0.25)

# run matching algorithm for weights
genout = GenMatch(Tr = t,
                  X = X,
                  estimand = "ATT",
                  M = 1,
                  pop.size = 200,
                  max.generations = 15,
                  wait.generations = 3,
                  caliper = caliper_vec,
                  replace = TRUE,
                  ties = TRUE,
                  print.level = 1)

```

```

##
##
## Tue Dec 9 11:55:42 2025
## Domains:
## 0.000000e+00 <= X1 <= 1.000000e+03
## 0.000000e+00 <= X2 <= 1.000000e+03
## 0.000000e+00 <= X3 <= 1.000000e+03
## 0.000000e+00 <= X4 <= 1.000000e+03
## 0.000000e+00 <= X5 <= 1.000000e+03
## 0.000000e+00 <= X6 <= 1.000000e+03
##
## Data Type: Floating Point
## Operators (code number, name, population)
## (1) Cloning..... 22
## (2) Uniform Mutation..... 25
## (3) Boundary Mutation..... 25
## (4) Non-Uniform Mutation..... 25
## (5) Polytope Crossover..... 25
## (6) Simple Crossover..... 26
## (7) Whole Non-Uniform Mutation..... 25
## (8) Heuristic Crossover..... 26
## (9) Local-Minimum Crossover..... 0
##
## SOFT Maximum Number of Generations: 15
## Maximum Nonchanging Generations: 3
## Population size : 200
## Convergence Tolerance: 1.000000e-03
##
## Not Using the BFGS Derivative Based Optimizer on the Best Individual Each Generation.
## Not Checking Gradients before Stopping.

```

```

## Using Out of Bounds Individuals.
##
## Maximization Problem.
##
##
## Generation#      Solution Values (lexical)
##
##      0  0.000000e+00  3.429738e-286  1.434228e-02  4.525277e-02  8.418580e-01  8.470073e-01  9.9835
##
##      1  0.000000e+00  3.429738e-286  2.799155e-02  4.525277e-02  8.828268e-01  8.915798e-01  9.9952
##
##      2  0.000000e+00  3.429738e-286  3.283436e-02  4.525277e-02  8.053778e-01  8.234528e-01  9.9952
##
##      3  0.000000e+00  3.429738e-286  3.839809e-02  4.525277e-02  7.589051e-01  9.241115e-01  9.9554
##
##      4  0.000000e+00  3.429738e-286  3.839809e-02  8.302895e-02  7.421408e-01  7.495531e-01  9.9554
##
##      5  0.000000e+00  3.429738e-286  4.476825e-02  4.525277e-02  8.253910e-01  9.854248e-01  9.9952
##
##      6  0.000000e+00  3.429738e-286  4.476825e-02  1.571408e-01  8.257001e-01  9.557247e-01  9.9952
##
##      7  0.000000e+00  3.429738e-286  4.476825e-02  1.571408e-01  8.446679e-01  9.113508e-01  9.9952
##
## 'wait.generations' limit reached.
## No significant improvement in 3 generations.
##
## Solution Lexical Fitness Value:
## 0.000000e+00  3.429738e-286  4.476825e-02  1.571408e-01  8.446679e-01  9.113508e-01  9.995295e-01  1
##
## Parameters at the Solution:
##
## X[ 1] : 2.242484e+02
## X[ 2] : 1.984563e+02
## X[ 3] : 4.709080e+02
## X[ 4] : 4.186388e+01
## X[ 5] : 4.813252e+02
## X[ 6] : 5.937708e+02
##
## Solution Found Generation 7
## Number of Generations Run 11
##
## Tue Dec  9 12:09:02 2025
## Total run time : 0 hours 13 minutes and 20 seconds

```

Present your key balance metrics both numerically and graphically (e.g. Q-Q plots or t-test plots). Was there improvement?

Use the `MatchBalance` function from the `Matching` package to numerically present metrics.

```

# use match to calculate estimand
mout <- Match(
  Y = matching_sample$overalllifesat,
  Tr = t,

```

```

X = X,
Weight.matrix = genout,
estimand = "ATT",
M = 1
)

mb <- MatchBalance(
  treatment ~ age + wage + nchildren + college + west + diff_wage_OR,
  data = matching_sample,
  match.out = mout
)

```

```

##
## ***** (V1) age *****
##
##          Before Matching      After Matching
## mean treatment.....      45.692      45.692
## mean control.....      45.115      45.661
## std mean diff.....      5.8285      0.31451
##
## mean raw eQQ diff.....      0.73956      0.25995
## med  raw eQQ diff.....      1          0
## max  raw eQQ diff.....      2          3
##
## mean eCDF diff.....      0.016521      0.0060076
## med  eCDF diff.....      0.0089547      0.0048741
## max  eCDF diff.....      0.045117      0.030057
##
## var ratio (Tr/Co).....      1.1272      1.0137
## T-test p-value.....      0.059115      0.55384
## KS Bootstrap p-value..      0.024      0.476
## KS Naive p-value.....      0.029995      0.63443
## KS Statistic.....      0.045117      0.030057
##
##
## ***** (V2) wage *****
##
##          Before Matching      After Matching
## mean treatment.....      20176      20176
## mean control.....      38146      20909
## std mean diff.....      -138.42      -5.6496
##
## mean raw eQQ diff.....      18671      1604.5
## med  raw eQQ diff.....      13722      1235
## max  raw eQQ diff.....      1332000      21594
##
## mean eCDF diff.....      0.24805      0.040289
## med  eCDF diff.....      0.27832      0.027214
## max  eCDF diff.....      0.39186      0.10236
##
## var ratio (Tr/Co).....      0.17606      1.351
## T-test p-value..... < 0.000000000000000222      0.0000052753
## KS Bootstrap p-value.. < 0.000000000000000222      < 0.000000000000000222
## KS Naive p-value..... < 0.000000000000000222      0.000005012
## KS Statistic.....      0.39186      0.10236

```

```

##
##
## ***** (V3) nchildren *****
##           Before Matching           After Matching
## mean treatment..... 0.73464      0.73464
## mean control..... 1.0218      0.73464
## std mean diff..... -31.041      0
##
## mean raw eQQ diff..... 0.28829      0
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 1      0
##
## mean eCDF diff..... 0.041022      0
## med eCDF diff..... 0.0079471      0
## max eCDF diff..... 0.12532      0
##
## var ratio (Tr/Co)..... 0.80031      1
## T-test p-value..... < 0.000000000000000222      1
## KS Bootstrap p-value.. < 0.000000000000000222      1
## KS Naive p-value..... 0.000000000000016891      1
## KS Statistic..... 0.12532      0
##
##
## ***** (V4) college *****
##           Before Matching           After Matching
## mean treatment..... 0.3276      0.3276
## mean control..... 0.35087      0.3276
## std mean diff..... -4.9564      0
##
## mean raw eQQ diff..... 0.022932      0
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 1      0
##
## mean eCDF diff..... 0.011636      0
## med eCDF diff..... 0.011636      0
## max eCDF diff..... 0.023272      0
##
## var ratio (Tr/Co)..... 0.96779      1
## T-test p-value..... 0.11251      1
##
##
## ***** (V5) west *****
##           Before Matching           After Matching
## mean treatment..... 0.61343      0.61343
## mean control..... 0.83073      0.61343
## std mean diff..... -44.604      0
##
## mean raw eQQ diff..... 0.21704      0
## med raw eQQ diff..... 0      0
## max raw eQQ diff..... 1      0
##
## mean eCDF diff..... 0.10865      0
## med eCDF diff..... 0.10865      0
## max eCDF diff..... 0.2173      0

```

```
##
## var ratio (Tr/Co).....      1.6875              1
## T-test p-value..... < 0.0000000000000000222      1
##
##
## ***** (V6) diff_wage_OR *****
##                               Before Matching      After Matching
## mean treatment.....          10996              10996
## mean control.....          -23077             -727.22
## std mean diff.....          229.35              78.912
##
## mean raw eQQ diff.....        34808              11738
## med  raw eQQ diff.....        28176              7300
## max  raw eQQ diff.....       1426861             341202
##
## mean eCDF diff.....          0.44814             0.42406
## med  eCDF diff.....          0.4225              0.3879
## max  eCDF diff.....          1                  1
##
## var ratio (Tr/Co).....        0.25002             150.84
## T-test p-value..... < 0.0000000000000000222      < 0.0000000000000000222
## KS Bootstrap p-value.. < 0.0000000000000000222      < 0.0000000000000000222
## KS Naive p-value..... < 0.0000000000000000222      < 0.0000000000000000222
## KS Statistic.....          1                  1
##
##
## Before Matching Minimum p.value: < 0.0000000000000000222
## Variable Name(s): wage nchildren west diff_wage_OR  Number(s): 2 3 5 6
##
## After Matching Minimum p.value: < 0.0000000000000000222
## Variable Name(s): wage diff_wage_OR  Number(s): 2 6
```

From this numerical presentation of unmatched and matched samples, we can see that the standard mean difference between treatment and control groups is lower after matching for all balance covariates.

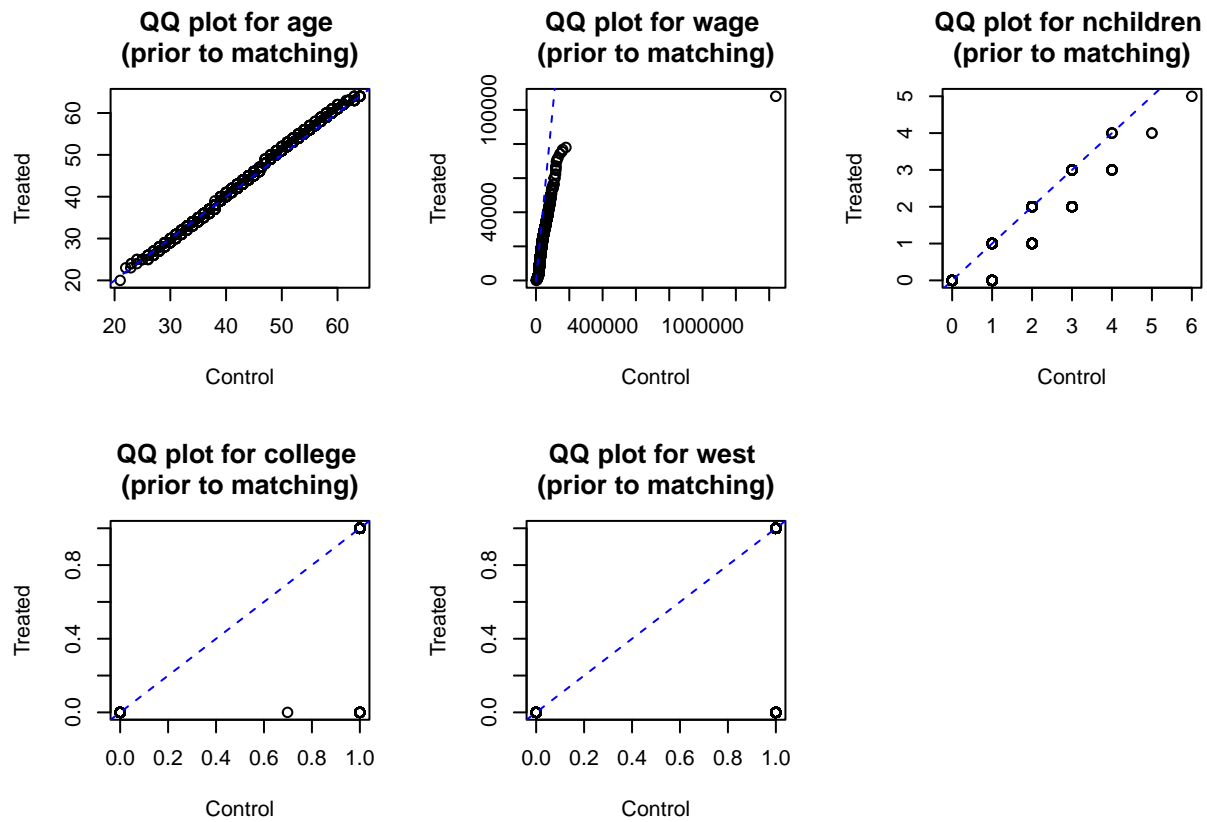
Also produce QQ-plots for each covariate before and after matching.

```
par(mfrow = c(2,3))

for (cov in covs) {
  # pull out covariate values from before
  x_treated <- matching_sample[matching_sample$treatment == 1,][[cov]]
  x_control <- matching_sample[matching_sample$treatment == 0,][[cov]]

  qqplot(x_control,
          x_treated,
          xlab = "Control",
          ylab = "Treated",
          main = paste("QQ plot for", cov, "\n(prior to matching)"))
  abline(0, 1, col = "blue", lty = 2)
}

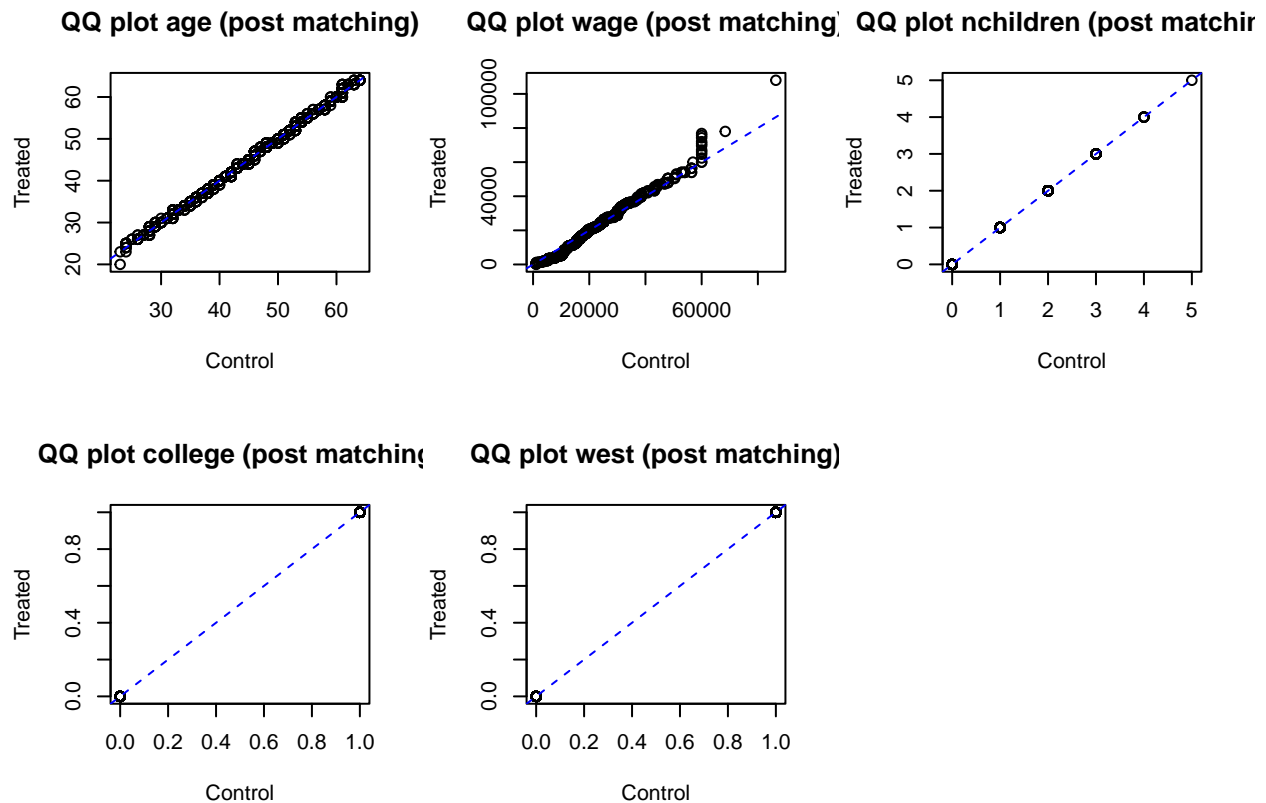
par(mfrow = c(1,1))
```



```
par(mfrow = c(2,3))
for (cov in covs) {
  x_treated <- matching_sample[[cov]][mout$index.treated]
  x_control <- matching_sample[[cov]][mout$index.control]

  qqplot(x_control,
          x_treated,
          xlab = "Control",
          ylab = "Treated",
          main = paste("QQ plot", cov, "(post matching)"))
  abline(0, 1, col = "blue", lty = 2)
}
par(mfrow = c(1,1))
```





Comparing QQ plots of covariates before and after matching tells a similar story. All covariates show visible improvement in balance.

**Did you find a treatment effect? Was this finding different from what you find without matching?**

Summarizing the matching output provides the estimate of the relevant estimand, ATT:

```
summary(mout)
```

```
##
## Estimate... -0.15192
## AI SE..... 0.12984
## T-stat..... -1.1701
## p.val..... 0.24197
##
## Original number of observations..... 7873
## Original number of treated obs..... 1221
## Matched number of observations..... 1221
## Matched number of observations (unweighted). 1231
```

While the estimate for ATT is negative and of expected magnitude, it is not statistically significant, so cannot conclude any treatment effect of wife outearning husband on husband's overall life satisfaction.

Without matching, we can compare outcomes of the treatment and control group using a two-sided t-test.

```
t.test(matching_sample[matching_sample$treatment==1,][["overalllifesat"]],
       matching_sample[matching_sample$treatment==0,][["overalllifesat"]])
```

```
##
## Welch Two Sample t-test
##
## data:  matching_sample[matching_sample$treatment == 1, ][["overalllifesat"]] and matching_sample[mat
## t = -7.5575, df = 1547.9, p-value = 0.00000000000006996
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.5307788 -0.3120328
## sample estimates:
## mean of x mean of y
## 6.955774 7.377180
```

Results from the t-test, using data prior to matching, demonstrate a statistically significant negative treatment effect on the treated.

For additional comparison, estimate the ATT (local average treatment effect) using OLS regression with an analogous linear regression model as specified in the pre-analysis plan.

```
model.1 <- lm(
  overalllifesat ~
    treatment +
    # running variable
    diff_wage_OR +
    # treatment-rv interaction
    treatment * diff_wage_OR +

    # covariate controls
    age +
    wage +
    nchildren +
    college +
    west
  ,
  data = matching_sample
)
summary(model.1)
```

```
##
## Call:
## lm(formula = overalllifesat ~ treatment + diff_wage_OR + treatment *
##     diff_wage_OR + age + wage + nchildren + college + west, data = matching_sample)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.6756 -0.6203  0.3162  0.9016  3.5759
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   7.116138349  0.107479194  66.209 < 0.0000000000000002
## treatment    -0.410828908  0.061294783  -6.703   0.00000000000219
```

```

## diff_wage_OR      0.000014157  0.000001616   8.761 < 0.00000000000000002
## age              -0.012245403  0.002022059  -6.056   0.0000000014604
## wage             0.000017982  0.000001571  11.444 < 0.00000000000000002
## nchildren        0.040443162  0.018634530   2.170   0.0300
## college          0.078384479  0.039580074   1.980   0.0477
## west             0.463964221  0.044762719  10.365 < 0.00000000000000002
## treatment:diff_wage_OR -0.000004410  0.000003361  -1.312   0.1895
##
## (Intercept)      ***
## treatment        ***
## diff_wage_OR     ***
## age              ***
## wage             ***
## nchildren        *
## college          *
## west             ***
## treatment:diff_wage_OR
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.549 on 7864 degrees of freedom
## Multiple R-squared:  0.05207,    Adjusted R-squared:  0.0511
## F-statistic: 53.99 on 8 and 7864 DF,  p-value: < 0.000000000000000022

```

From the OLS model summary, I do find a treatment effect, where wife outearning husband has a negative effect on overall life satisfaction.

Matching may have reduced statistical power because the total number of couples in the matched sample is significantly smaller ( $N = 2,442$ ) than the unmatched sample ( $N = 7,873$ ). With additional computational resources, using a larger proportion of the initial sample would include a higher number of treated couples, leading to a larger matched sample and more statistical power.