New York University Shanghai

Problem Set 4

BUSF-SHU 210: Business Analytics (Spring 2018)

Due at 10:00PM on Tuesday, April 10

Solutions

Instructions

Please submit a zip file of a report together with the accompanying R codes on NYU Classes. Please also bring a hard copy of your report to class. The total score of the problem set is 7 points.

1. Causal Inference with Instrumental Variable (1.5 points)

Facebook is considering establishing a new feature that could facilitate the process of searching and connecting to a new friend. The company runs an online randomized experiment to examine the **whether users' adoption of this new feature could improve their satisfaction**. To do so, Facebook randomly selects a sample of 10,000 users. Within this sample, a random group of users are selected into the treatment group. Facebook sends an encouragement message to each individual in the treatment group. The encouragement message advocates the new feature and encourages the user to adopt it. The rest of the users in the sample are in the control group, to which Facebook sends nothing. Facebook cannot control which user will adopt their new feature, but believes that the adoption of the new feature is positively correlated with receiving the encouragement message. Users who are initially more satisfied with Facebook before the experiments will be more likely to adopt the new feature. At the end of the experiment, Facebook conducts a survey that asks each user in the experiment to report his/her satisfaction level of Facebook then. We assume that each user truthfully reports his/her satisfaction level of this online social media.

The data associated with the experiment described above is stored in `NewFeature.csv`. Each row in this data set represents a user. The data has 3 variables:

- $Encouragement$: An indicator of whether the user has received the encouragement. $Encouragement = 1$ if the user received the encouragement; $Encouragement = 0$ if the user did not receive the encouragement.
- $Satisfaction$: The reported satisfaction level. A higher value means the user is more satisfied. 0 means not satisfied at all; 10 means fully satisfied.
- $Adoption$: An indicator of whether the user has adopted the new feature. $Adoption = 1$ if the user adopted the new feature; $Adoption = 1$ if the user did not adopt the new feature.

Please briefly answer the following questions:

(a) (0.5 points) Run a linear regression with $Satisfaction$ as the outcome and $Adoption$ as the covariate. Assume $\hat{\beta}_A$ is the fitted coefficient of $Adoption$. What is the value of $\hat{\beta}_A$? Is $\hat{\beta}_A$ an unbiased estimation of the causal effect of $Adoption$ on $Satisfaction$? Why or why not? If not, will this be an overestimation or an underestimation of the true causal effect?

All the codes for Question 1 can be found in the *R* script PS4_Q1.R.

The summary of the fitted basic linear regression model is given in Figure 1. Hence, the estimated causal effect of $Adoption$ on $Satisfaction$ is $\hat{\beta}_A = 3.47$. This is an overestimation of the true causal effect of adoption. Because the omitted variable *initial satisfaction* would be positively correlated with both adoption and the post-usage satiafaction. Hence, the estimated value 3.47 is the "sum" of the direct effect of the adoption of the new feature and the indirect from initial satisfaction.

GRADING. 0.5 points in total. 0.2 points for using regression to report the estimated causal effect of new feature adoption and user satisfaction. 0.3 points for correctly identifying the overestimation and explaining why.

```
Call:
lm(formula = Satisfaction ~ Adoption, data = NewFeature)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.1261 -1.1261 -0.1261  0.8739  4.3488 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.65121    0.01935   240.4   <2e-16 ***
Adoption     3.47492    0.02863   121.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.426 on 9998 degrees of freedom
Multiple R-squared:  0.5958,
  Adjusted R-squared:  0.5957 
F-statistic: 1.473e+04 on 1 and 9998 DF,  p-value: < 2.2e-16
```

Figure 1: Basic Linear Regression Model

(b) (0.5 points) Facebook uses $Encouragement$ as an instrumental variable to establish the causal effect of $Adoption$ on $Satisfaction$. Discuss why $Encouragement$ could be a valid instrumental variable in this setting.

First, the users who received the encouragement will be more likely to adopt the new feature. From the data, we calculate that the correlation between *Adoption* and *Encouragement* is 0.2248. Hence, the strong first-stage assumption is satisfied. On the other hand, the encouragement is completely randomly assigned to the sample users, so it is uncorrelated with the omitted variable, *initial satisfaction*. So the exclusion restriction assumption is also satisfied. Therefore, the variable *Encouragement* may be a valid instrumental variable to establish the causal effect of new feature adoption on user satisfaction.

GRADING. 0.5 points in total. 0.3 points for the strong first-stage assumption (0.2 points for the discussion and 0.1 points for calculating the correlation or running a regression between these two variables). 0.2 points for the exclusion restriction assumption.

(c) (0.5 points) Use the two-stage least-square method to estimate the unbiased causal effect of *Adoption* on $Satisfaction$. Please obtain the estimation results using two different approaches, one with two linear regression steps, and the other with a single integrated step. Take a screen shot of the regression models you build. We use $\hat{\gamma}_A$ to denote the true causal effect of *Adoption* on $Satisfaction$. What is the value of $\hat{\gamma}_A$? How do you interpret $\hat{\gamma}_A$?

First, we use the two-step method to estimate $\hat{\gamma}_A$. The first step is to run a linear regression using *Adoption* as the first step and *Encouragement* as the second step. See Figure 2 for a summary of this model. The second step is to run a linear regression using $Satisfaction$ as the outcome first step and the fitted value of adoption in the first stage $\widehat{Adoption}$ as the covariate. Then, the estimated coefficient of $\widehat{Adoption}$ is an unbiased estimation of the causal effect of *Adoption* on $Satisfaction$. See Figure 3 for a summary of this model. Hence, $\hat{\gamma}_A = 0.91$.

The second approach is to have an integrated estimation of the 2-stage-least-squares. The model summary is presented in Figure 4. As you can see, both approaches lead to the same estimation $\hat{\gamma}_A = 0.91$.

The interpretation of $\hat{\gamma}_A$ is that, on average, adopting the new feature can lead to an increase of 0.91 in the user satisfaction.

GRADING. 0.5 points in total. 0.2 points for each approach. 0.1 points for the interpretation of $\hat{\gamma}_A$.

```
Call:
lm(formula = Adoption ~ Encouragement, data = NewFeature)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5685 -0.3445 -0.3445  0.4315  0.6555

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.344489   0.006872   50.13   <2e-16 ***
Encouragement 0.223974   0.009708   23.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4854 on 9998 degrees of freedom
Multiple R-squared:  0.05054,
  Adjusted R-squared:  0.05045
F-statistic: 532.2 on 1 and 9998 DF,  p-value: < 2.2e-16
```

Figure 2: 2SLS: First Stage

```
Call:
lm(formula = Satisfaction ~ Adoption1, data = NewFeature)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3397 -1.3397 -0.1363  1.6603  3.8637

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.82335    0.09408   61.90  < 2e-16 ***
Adoption1    0.90836    0.20007    4.54 5.68e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.24 on 9998 degrees of freedom
Multiple R-squared:  0.002058,
  Adjusted R-squared:  0.001958
F-statistic: 20.61 on 1 and 9998 DF,  p-value: 5.684e-06
```

Figure 3: 2SLS: Second Stage

```
Call:
ivreg(formula = Satisfaction ~ Adoption | Encouragement, data = NewFeature)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8234 -0.8234  0.1766  1.2683  3.2683

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.82335    0.08042  72.412  < 2e-16 ***
Adoption     0.90836    0.17102   5.311 1.11e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.915 on 9998 degrees of freedom
Multiple R-Squared: 0.2708,
  Adjusted R-squared: 0.2707
Wald test: 28.21 on 1 and 9998 DF,  p-value: 1.112e-07
```

Figure 4: 2SLS: Integrated Estimation

2. Application of For Loop in R (0.5 points)

The loop structure is very important in R and other programing languages. In this problem, you are going to use `for` loop, which we learned in the lecture of Session 9, to solve two simple mathematical problems. You may use the commend `format(Num,scientific=FALSE,big.mark=",")` to change the format of your results if the number, `Num`, is too large.

(a) (0.2 points) A sequence $\{A_n : n \geq 1\}$ satisfies that $A_1 = 1$, $A_2 = 1$, and $A_{n+2} = A_{n+1} + A_n + n$ for $n \geq 1$. Use R to find the value of A_{100}.

> The code of Question 2 is given in PS4_Q2.R. $A_100 = 1,281,597,540,372,340,801,536.$
>
> <u>Grading</u>. 0.2 points in total. All or nothing.

(b) (0.3 points) We define $\binom{n}{m}$ as the coefficient of binomial expansion for $(1+x)^n$ ($m = 0, 1, 2, \cdots, n$). Hence, $\binom{n}{m}$ is the coefficient of x^m in $(1+x)^n$, where $m = 0, 1, 2, \cdots, n$. We know from the property of binomial coefficients that $\binom{n}{m} + \binom{n}{m+1} = \binom{n+1}{m+1}$ for all n and all $0 \leq m \leq n-1$. Please use R to calculate $\binom{100}{50}$. (*Hint:* You may need to use a matrix to store your calculations of $\binom{n}{m}$.)

> $\binom{100}{50} = 100,891,344,545,564,237,256,087,044,096.$
>
> <u>Grading</u>. 0.3 points in total. 0.2 points for correctly implementing the recursive structure in R. 0.1 points for getting the correct answer.

3. Steakhouse Staff Scheduling (1 Point)

Western Family Steakhouse offers a variety of low-cost meals and quick service. Other than management, the steakhouse operates with two full-time employees who work 8 hours per day. The rest of the employees are part-time employees who are scheduled for 4-hour shifts during peak meal times[1]. On Saturdays the steakhouse is open from 11:00am to 10:00pm. Management wants to develop a schedule for part-time employees that will minimize labor costs and still provide excellent customer service. The average wage rate for the part-time employee is \$7.60 per hour. The total number of full-time and par-time employees needed varies with the time of day as shown in the following table

Time	Total Number of Employees Needed
11am-noon	9
noon-1pm	9
1pm-2pm	9
2pm-3pm	3
3pm-4pm	3
4pm-5pm	3
5pm-6pm	6
6pm-7pm	12
7pm-8pm	12
8pm-9pm	7
9pm-10pm	7

One full-time employee comes on duty at 11am, works 4 hours, takes an hour off, and returns for another 4 hours. The other full-time employee comes to work at 1pm and works the same 4-hours-on, 1-hour-off, 4-hours-on pattern.

(a) (0.5 points) What is the minimum-cost schedule for part-time employees? Please formulate this problem as a linear program and solve it using *R*. Clearly state the decision variables, the objective function, and the constraints for the linear program.

[1] 4-hour shifts mean that the part-time employee works for a consecutive period of 4 hours.

The code of Question 3 is given in PS4_Q3.R. We define the number of part-time employees who start working at the beginning of each hour as follows:

- x_1 = Number of employees from 11am to 3pm
- x_2 = Number of employees from noon to 4pm
- x_3 = Number of employees from 1pm to 5pm
- x_4 = Number of employees from 2pm to 6pm
- x_5 = Number of employees from 3pm to 7pm
- x_6 = Number of employees from 4pm to 8pm
- x_7 = Number of employees from 5pm to 9pm
- x_8 = Number of employees from 6pm to 10pm

The objective function is to minimize the total cost of recruiting part-time employees:

$$\min\ 7.6 \times 4(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8) = 30.2(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8)$$

The constraints are that, in each hour, the total number of employees meet the requirements of the Steakhouse. Hence, we have the following constraints:

- $x_1 + 1 \geq 9$: 11am-noon
- $x_1 + x_2 + 1 \geq 9$: noon-1pm
- $x_1 + x_2 + x_3 + 2 \geq 9$: 1pm-2pm
- $x_1 + x_2 + x_3 + x_4 + 2 \geq 3$: 2pm-3pm
- $x_2 + x_3 + x_4 + x_5 + 1 \geq 3$: 3pm-4pm
- $x_3 + x_4 + x_5 + x_6 + 2 \geq 3$: 4pm-5pm
- $x_4 + x_5 + x_6 + x_7 + 1 \geq 6$: 5pm-6pm
- $x_5 + x_6 + x_7 + x_8 + 2 \geq 12$: 6pm-7pm
- $x_6 + x_7 + x_8 + 2 \geq 12$: 7pm-8pm
- $x_7 + x_8 + 1 \geq 7$: 8pm-9pm
- $x_8 + 1 \geq 7$: 9pm-10pm
- $x_i \geq 0$ for all i: Non-negativity constraint

Putting everything together, we have the linear program can be written as:

$$
\begin{aligned}
&\min 30.4(x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8) \\
&\text{Subject to} \\
&\quad x_1 \geq 8 \\
&\quad x_1 + x_2 \geq 8 \\
&\quad x_1 + x_2 + x_3 \geq 7 \\
&\quad x_1 + x_2 + x_3 + x_4 \geq 1 \\
&\quad x_2 + x_3 + x_4 + x_5 \geq 2 \\
&\quad x_3 + x_4 + x_5 + x_6 \geq 1 \\
&\quad x_4 + x_5 + x_6 + x_7 \geq 5 \\
&\quad x_5 + x_6 + x_7 + x_8 \geq 10 \\
&\quad x_6 + x_7 + x_8 \geq 10 \\
&\quad x_7 + x_8 \geq 6 \\
&\quad x_8 \geq 6 \\
&\quad x_i \geq 0 \text{ for all } i
\end{aligned}
$$

Then, we use R to solve this linear program. The optimal recruiting plan is $x_1^* = 8$, $x_2^* = 0$, $x_3^* = 0$, $x_4^* = 0$, $x_5^* = 2$, $x_6^* = 4$, $x_7^* = 0$, and $x_8^* = 6$, with the minimum cost equal to \$608.

Grading Scheme: 0.5 points in total. 0.3 points for correctly formulating the problem. 0.2 points for using R to solve the linear program.

(b) (0.5 points) Assume that part-time employees can be assigned either a 3-hour or a 4-hour shift. What is the cost savings compared to the previous schedule you give in part (a)?

We define the number of part-time employees (in different shifts) who start working at the beginning of each hour as follows:

- $y_1 =$ Number of employees from 11am to 3pm
- $y_2 =$ Number of employees from noon to 4pm
- $y_3 =$ Number of employees from 1pm to 5pm
- $y_4 =$ Number of employees from 2pm to 6pm
- $y_5 =$ Number of employees from 3pm to 7pm
- $y_6 =$ Number of employees from 4pm to 8pm
- $y_7 =$ Number of employees from 5pm to 9pm
- $y_8 =$ Number of employees from 6pm to 10pm
- $y_9 =$ Number of employees from 11am to 2pm
- $y_{10} =$ Number of employees from noon to 3pm
- $y_{11} =$ Number of employees from 1pm to 4pm
- $y_{12} =$ Number of employees from 2pm to 5pm
- $y_{13} =$ Number of employees from 3pm to 6pm
- $y_{14} =$ Number of employees from 4pm to 7pm
- $y_{15} =$ Number of employees from 5pm to 8pm
- $y_{16} =$ Number of employees from 6pm to 9pm
- $y_{17} =$ Number of employees from 7pm to 10pm

The objective function is to minimize the total cost of recruiting part-time employees:

$$\min\ 30.4(y_1+y_2+y_3+y_4+y_5+y_6+y_7+y_8)+22.8(y_9+y_{10}+y_{11}+y_{12}+y_{13}+y_{14}+y_{15}+y_{16}+y_{17})$$

The constraints are that, in each hour, the total number of employees meet the requirements of the Steakhouse. Hence, we have the following constraints:

- $y_1 + y_9 + 1 \geq 9$: 11am-noon
- $y_1 + y_2 + y_9 + y_{10} + 1 \geq 9$: noon-1pm
- $y_1 + y_2 + y_3 + y_9 + y_{10} + y_{11} + 2 \geq 9$: 1pm-2pm
- $y_1 + y_2 + y_3 + y_4 + y_{10} + y_{11} + y_{12} + 2 \geq 3$: 2pm-3pm
- $y_2 + y_3 + y_4 + y_5 + y_{11} + y_{12} + y_{13} + 1 \geq 3$: 3pm-4pm
- $y_3 + y_4 + y_5 + y_6 + y_{12} + y_{13} + y_{14} + 2 \geq 3$: 4pm-5pm
- $y_4 + y_5 + y_6 + y_7 + y_{13} + y_{14} + y_{15} + 1 \geq 6$: 5pm-6pm
- $y_5 + y_6 + y_7 + y_8 + y_{14} + y_{15} + y_{16} + 2 \geq 12$: 6pm-7pm
- $y_6 + y_7 + y_8 + y_{15} + y_{16} + y_{17} + 2 \geq 12$: 7pm-8pm

- $y_7 + y_8 + y_{16} + y_{17} + 1 \geq 7$: 8pm-9pm
- $y_8 + y_{17} + 1 \geq 7$: 9pm-10pm
- $y_i \geq 0$ for all i: Non-negativity constraint

Putting everything together, we have the linear program can be written as:

$$\begin{aligned}
&\min 30.4(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8) \\
&\qquad + 22.8(y_9 + y_{10} + y_{11} + y_{12} + y_{13} + y_{14} + y_{15} + y_{16} + y_{17}) \\
&\text{Subject to} \\
&\qquad y_1 + y_9 \geq 8 \\
&\qquad y_1 + y_2 + y_9 + y_{10} \geq 8 \\
&\qquad y_1 + y_2 + y_3 + y_9 + y_{10} + y_{11} \geq 7 \\
&\qquad y_1 + y_2 + y_3 + y_4 + y_{10} + y_{11} + y_{12} \geq 1 \\
&\qquad y_2 + y_3 + y_4 + y_5 + y_{11} + y_{12} + y_{13} \geq 2 \\
&\qquad y_3 + y_4 + y_5 + y_6 + y_{12} + y_{13} + y_{14} \geq 1 \\
&\qquad y_4 + y_5 + y_6 + y_7 + y_{13} + y_{14} + y_{15} \geq 5 \\
&\qquad y_5 + y_6 + y_7 + y_8 + y_{14} + y_{15} + y_{16} \geq 10 \\
&\qquad y_6 + y_7 + y_8 + y_{15} + y_{16} + y_{17} \geq 10 \\
&\qquad y_7 + y_8 + y_{16} + y_{17} \geq 6 \\
&\qquad y_8 + y_{17} \geq 6 \\
&\qquad y_i \geq 0 \text{ for all } i
\end{aligned}$$

Then, we use R to solve this linear program. The optimal recruiting plan is $y_1^* = 0$, $y_2^* = 0$, $y_3^* = 0$, $y_4^* = 0$, $y_5^* = 0$, $y_6^* = 0$, $y_7^* = 0$, $y_8^* = 6$, $y_9^* = 8$, $y_{10}^* = 0$, $y_{11}^* = 1$, $y_{12}^* = 0$, $y_{13}^* = 1$, $y_{14}^* = 0$, $y_{15}^* = 4$, $y_{16}^* = 0$, and $y_{17}^* = 0$ with the minimum cost equal to \$501.6. Hence, the cost reduction is 608-500.4=\$107.6.

Grading Scheme: 0.5 points in total. 0.3 points for correctly formulating the problem. 0.2 points for using R to solve the linear program.

4. Gasoline Blending (2 points)

Gasoline blending is part of oil refineries, where crude oil is processed and refined into more useful products, such as gasoline and diesel fuel. Consider three products: super gasoline, regular gasoline, and diesel fuel, all of which can be made by mixing three crude oils: Crude A, Crude B, and Crude C. Each product is distinguished by its octane rating, which measures the quality of the fuel, and its iron content, which is a contaminant in the product. The crude oils each have a respective octane rating and iron content as well. The following table shows the required octane ratings and iron contents for each of the final products, as well as the known octane ratings and iron contents for each of the crude oils:

Product/Oil	Octane Rating	Iron Content
Super Gasoline	≥ 10	≤ 1
Regular Gasoline	≥ 8	≤ 2
Diesel Gasoline	≥ 6	≤ 1
Crude A	12	0.5
Crude B	6	2.0
Crude C	8	3.0

The product must meet these standards for octane ratings and iron content. The octane rating and iron content of a product is the weighted average of the octane rating and iron content of the crude oils that produce it. For example, if we produce 35 barrels of regular gasoline using 20 barrels of Crude A, 5 barrels of Crude B, and 10 barrels of Crude C, the Octane Rating of this product would be:

$$\frac{20 \times 12 + 5 \times 5 + 10 \times 8}{35} = 10$$

Analogously, the iron average content of the regular gasoline would be:

$$\frac{20 \times 0.5 + 5 \times 2.0 + 10 \times 3.0}{35} = 1.43$$

The objective of the oil refinery company is to **maximize its profit**. The following table gives the sales price (revenue) for one barrel of each product, and the demand for each product:

Product	Sales Price	Demand
Super Gasoline	\$70	3,000 Barrels
Regular Gasoline	\$60	2,000 Barrels
Diesel Gasoline	\$50	1,000 Barrels

The cost of each crude oil is given in the following table:

Oil	Purchase Cost
Crude A	\$45
Crude B	\$35
Crude C	\$25

The refinery company can only purchase 5,000 barrels of each crude oil, and has an oil refining capacity of no more than 14,000 barrels in total. That is the total amount of final products the company produces cannot exceed 14,000 barrels. One barrel of crude oil makes one barrel of the final product (there is no loss in the conversion).

(a) (0.5 points) Formulate the planning problem of the refinery company as a linear program. Please clearly state the decision variables, the objective function, and the constraints.

The *R* code for Question 4 is given in PS4_Q4.R. We define the following decision variables:

- x_{1a} = Number of barrels of Crude Oil A to produce Super Gasoline
- x_{1b} = Number of barrels of Crude Oil B to produce Super Gasoline
- x_{1c} = Number of barrels of Crude Oil C to produce Super Gasoline
- x_{2a} = Number of barrels of Crude Oil A to produce Regular Gasoline
- x_{2b} = Number of barrels of Crude Oil B to produce Regular Gasoline
- x_{2c} = Number of barrels of Crude Oil C to produce Regular Gasoline
- x_{3a} = Number of barrels of Crude Oil A to produce Diesel Gasoline
- x_{3b} = Number of barrels of Crude Oil B to produce Diesel Gasoline
- x_{3c} = Number of barrels of Crude Oil C to produce Diesel Gasoline

The objective function is to maximize the total profit:

$$\begin{aligned}
\max\ & 70(x_{1a}+x_{1b}+x_{1c})+60(x_{2a}+x_{2b}+x_{2c})+50(x_{3a}+x_{3b}+x_{3c})-45(x_{1a}+x_{2a}+x_{3a}) \\
& -35(x_{1b}+x_{2b}+x_{3b})-25(x_{1c}+x_{2c}+x_{3c}) \\
=& 25x_{1a}+35x_{1b}+45x_{1c}+15x_{2a}+25x_{2b}+35x_{2c}+5x_{3a}+15x_{3b}+25x_{3c}
\end{aligned}$$

We establish the following constraints for the this problem:

- $12x_{1a}+6x_{1b}+8x_{1c} \geq 10(x_{1a}+x_{1b}+x_{1c})$: The octane rating requirement for Super Gasoline
- $0.5x_{1a}+2x_{1b}+3x_{1c} \leq x_{1a}+x_{1b}+x_{1c}$: The iron content requirement for Super Gasoline
- $12x_{2a}+6x_{2b}+8x_{2c} \geq 8(x_{2a}+x_{2b}+x_{2c})$: The octane rating requirement for Regular Gasoline
- $0.5x_{2a}+2x_{2b}+3x_{2c} \leq 2(x_{2a}+x_{2b}+x_{2c})$: The iron content requirement for Regular Gasoline
- $12x_{3a}+6x_{3b}+8x_{3c} \geq 6(x_{3a}+x_{3b}+x_{3c})$: The octane rating requirement for Diesel Gasoline
- $0.5x_{3a}+2x_{3b}+3x_{3c} \leq x_{3a}+x_{3b}+x_{3c}$: The iron content requirement for Diesel Gasoline
- $x_{1a}+x_{1b}+x_{1c} \leq 3000$: The demand constraint for Super Gasoline
- $x_{2a}+x_{2b}+x_{2c} \leq 2000$: The demand constraint for Regular Gasoline
- $x_{3a}+x_{3b}+x_{3c} \leq 1000$: The demand constraint for Diesel Gasoline
- $x_{1a}+x_{2a}+x_{3a} \leq 5000$: The quantity constraint for Crude Oil A
- $x_{1b}+x_{2b}+x_{3b} \leq 5000$: The quantity constraint for Crude Oil B
- $x_{1c}+x_{2c}+x_{3c} \leq 5000$: The quantity constraint for Crude Oil C
- $x_{1a}+x_{1b}+x_{1c}+x_{2a}+x_{2b}+x_{2c}+x_{3a}+x_{3b}+x_{3c} \leq 14000$: The production capacity constraint
- $x_{1a}, x_{1b}, x_{1c}, x_{2a}, x_{2b}, x_{2c}, x_{3a}, x_{3b}, x_{3c} \geq 0$: Non-negativity constraint

Putting everything together, we formulate the problem as the following linear program:

$$\max\ 25x_{1a} + 35x_{1b} + 45x_{1c} + 15x_{2a} + 25x_{2b} + 35x_{2c} + 5x_{3a} + 15x_{3b} + 25x_{3c}$$

Subject to

$$
\begin{aligned}
&2x_{1a} - 4x_{1b} - 2x_{1c} \geq 0 \\
&0.5x_{1a} - x_{1b} - 2x_{1c} \geq 0 \\
&4x_{2a} - 2x_{2b} \geq 0 \\
&1.5x_{2a} - x_{2c} \geq 0 \\
&6x_{3a} + 2x_{3c} \geq 0 \\
&0.5x_{3a} - x_{3b} - 2x_{3c} \geq 0 \\
&x_{1a} + x_{1b} + x_{1c} \leq 3000 \\
&x_{2a} + x_{2b} + x_{2c} \leq 2000 \\
&x_{3a} + x_{3b} + x_{3c} \leq 1000 \\
&x_{1a} + x_{2a} + x_{3a} \leq 5000 \\
&x_{1b} + x_{2b} + x_{3b} \leq 5000 \\
&x_{1c} + x_{2c} + x_{3c} \leq 5000 \\
&x_{1a} + x_{1b} + x_{1c} + x_{2a} + x_{2b} + x_{2c} + x_{3a} + x_{3b} + x_{3c} \leq 14000 \\
&x_{1a}, x_{1b}, x_{1c}, x_{2a}, x_{2b}, x_{2c}, x_{3a}, x_{3b}, x_{3c} \geq 0
\end{aligned}
$$

Grading Scheme: 0.5 points in total. 0.1 points for the decision variables.0.1 points for the objective function. 0.3 points for the constraints.

(b) (0.5 points) Use R to solve the linear program you build in part (a). What is the optimal production planning and the optimal profit for the refinery company? Please also report the dual variable for each of the constraints.

We use R to solve the linear program. The optimal decisions are: $x^*_{1a} = 2400$, $x^*_{1b} = 0$, $x^*_{1c} = 600$, $x^*_{2a} = 800$, $x^*_{2b} = 0$, $x^*_{2c} = 1200$, $x^*_{3a} = 800$, $x^*_{3b} = 0$, $x^*_{3c} = 200$. The associated optimal profit is \$150000. The shadow prices of each constraint is given by:

$$(0, -8, 0, -8, 0, -8, 29, 27, 9, 0, 0, 0, 0, 0, -2, 0, 0, -2, 0, 0, -2, 0)$$

Grading Scheme: 0.5 points in total. 0.3 points for the implementing the model in R. 0.2 points for getting the correct answers.

(c) (0.5 points) Assume that it costs \$2 per unit of increase in demand to advertise a product, regardless of which type of the product being advertised. Which type of gasoline do your recommend to advertise first? Base your recommendation on solid quantitative evidence.

The shadow price for the demand constraint of Super Gasoline is 29, that for the demand constraint of Regular Gasoline is 27, and that for the demand constraint of Diesel is 9. Hence, we recommend advertising on Super Gasoline.

Grading Scheme: 0.2 points for using the shadow prices. 0.3 points for getting the correct answer.

(d) (0.5 points) The manager of the refinery company is considering introducing a new product Premium Gasoline, which requires Octane Rating at least 9 and Iron Content not exceeding 1. The price of this new product is \$65. The demand for the new Premium Gasoline is estimated to be 3,000 barrels, and its introduction will cannibalize the demand for original products. Specifically the demand for Super Gasoline will decrease to 1,500, whereas the demand for regular gasoline will decrease to $1,000$. The demand for Diesel Gasoline won't change. Introducing this new product would require some initial investment of \$3,000 on product line adjustment and advertising. Do you recommend introducing this new product? Why or why not?

We define the following decision variables:

- y_{1a} = Number of barrels of Crude Oil A to produce Super Gasoline
- y_{1b} = Number of barrels of Crude Oil B to produce Super Gasoline
- y_{1c} = Number of barrels of Crude Oil C to produce Super Gasoline
- y_{2a} = Number of barrels of Crude Oil A to produce Regular Gasoline
- y_{2b} = Number of barrels of Crude Oil B to produce Regular Gasoline
- y_{2c} = Number of barrels of Crude Oil C to produce Regular Gasoline
- y_{3a} = Number of barrels of Crude Oil A to produce Diesel Gasoline
- y_{3b} = Number of barrels of Crude Oil B to produce Diesel Gasoline
- y_{3c} = Number of barrels of Crude Oil C to produce Diesel Gasoline
- y_{4a} = Number of barrels of Crude Oil A to produce Premium Gasoline
- y_{4b} = Number of barrels of Crude Oil B to produce Premium Gasoline
- y_{4c} = Number of barrels of Crude Oil C to produce Premium Gasoline

The objective function is to maximize the total profit:

$$\begin{aligned}\max\ & 70(y_{1a}+y_{1b}+y_{1c})+60(y_{2a}+y_{2b}+y_{2c})+50(y_{3a}+y_{3b}+y_{3c})+65(y_{4a}+y_{4b}+y_{4c})\\ &-45(y_{1a}+y_{2a}+y_{3a}+y_{4a})-35(y_{1b}+y_{2b}+y_{3b}+y_{4b})-25(y_{1c}+y_{2c}+y_{3c}+y_{4c})-3000\\ =&25y_{1a}+35y_{1b}+45y_{1c}+15y_{2a}+25y_{2b}+35y_{2c}+5y_{3a}+15y_{3b}+25y_{3c}+20y_{4a}+30y_{4b}+40y_{4c}\\ &-3000,\end{aligned}$$

where 3000 is the cost of advertising.

We establish the following constraints for the this problem:

- $12y_{1a}+6y_{1b}+8y_{1c}\geq 10(y_{1a}+y_{1b}+y_{1c})$: The octane rating requirement for Super Gasoline
- $0.5y_{1a}+2y_{1b}+3y_{1c}\leq y_{1a}+y_{1b}+y_{1c}$: The iron content requirement for Super Gasoline
- $12y_{2a}+6y_{2b}+8y_{2c}\geq 8(y_{2a}+y_{2b}+y_{2c})$: The octane rating requirement for Regular Gasoline
- $0.5y_{2a}+2y_{2b}+3y_{2c}\leq 2(y_{2a}+y_{2b}+y_{2c})$: The iron content requirement for Regular Gasoline
- $12y_{3a}+6y_{3b}+8y_{3c}\geq 6(y_{3a}+y_{3b}+y_{3c})$: The octane rating requirement for Diesel Gasoline
- $0.5y_{3a}+2y_{3b}+3y_{3c}\leq y_{3a}+y_{3b}+y_{3c}$: The iron content requirement for Diesel Gasoline
- $12y_{4a}+6y_{4b}+8y_{4c}\geq 9(y_{4a}+y_{4b}+y_{4c})$: The octane rating requirement for Premium Gasoline
- $0.5y_{4a}+2y_{4b}+3y_{4c}\leq y_{4a}+y_{4b}+y_{4c}$: The iron content requirement for Premium Gasoline
- $y_{1a}+y_{1b}+y_{1c}\leq 1500$: The demand constraint for Super Gasoline
- $y_{2a}+y_{2b}+y_{2c}\leq 1000$: The demand constraint for Regular Gasoline
- $y_{3a}+y_{3b}+y_{3c}\leq 1000$: The demand constraint for Diesel Gasoline
- $y_{4a}+y_{4b}+y_{4c}\leq 3000$: The demand constraint for Premium Gasoline
- $y_{1a}+y_{2a}+y_{3a}+y_{4a}\leq 5000$: The quantity constraint for Crude Oil A
- $y_{1b}+y_{2b}+x_{3b}+y_{4b}\leq 5000$: The quantity constraint for Crude Oil B
- $y_{1c}+y_{2c}+y_{3c}+y_{4c}\leq 5000$: The quantity constraint for Crude Oil C

- $y_{1a} + y_{1b} + y_{1c} + y_{2a} + y_{2b} + y_{2c} + y_{3a} + y_{3b} + y_{3c} + y_{4a} + y_{4b} + y_{4c} \leq 14000$: The production capacity constraint
- $y_{1a}, y_{1b}, y_{1c}, y_{2a}, y_{2b}, y_{2c}, y_{3a}, y_{3b}, y_{3c}, y_{4a}, y_{4b}, y_{4c} \geq 0$: Non-negativity constraint

Putting everything together, we formulate the problem as the following linear program:

$$\max\ 25y_{1a} + 35y_{1b} + 45y_{1c} + 15y_{2a} + 25y_{2b} + 35y_{2c} + 5y_{3a} + 15y_{3b} + 25y_{3c} + 20y_{4a} + 30y_{4b} + 40y_{4c} - 3000$$

Subject to

$$
\begin{aligned}
&2y_{1a} - 4y_{1b} - 2y_{1c} \geq 0\\
&0.5y_{1a} - y_{1b} - 2y_{1c} \geq 0\\
&4y_{2a} - 2y_{2b} \geq 0\\
&1.5y_{2a} - y_{2c} \geq 0\\
&6y_{3a} + 2y_{3c} \geq 0\\
&0.5y_{3a} - y_{3b} - 2y_{3c} \geq 0\\
&3y_{4a} - 3y_{4b} - y_{4c} \geq 0\\
&0.5y_{4a} - y_{4b} - 2y_{4c} \geq 0\\
&y_{1a} + y_{1b} + y_{1c} \leq 1500\\
&y_{2a} + y_{2b} + y_{2c} \leq 1000\\
&y_{3a} + y_{3b} + y_{3c} \leq 1000\\
&y_{4a} + y_{4b} + y_{4c} \leq 3000\\
&y_{1a} + y_{2a} + y_{3a} + y_{4a} \leq 5000\\
&y_{1b} + y_{2b} + y_{3b} + y_{4b} \leq 5000\\
&y_{1c} + y_{2c} + y_{3c} + y_{4c} \leq 5000\\
&y_{1a} + y_{1b} + y_{1c} + y_{2a} + y_{2b} + y_{2c} + y_{3a} + y_{3b} + y_{3c} + y_{4a} + y_{4b} + y_{4c} \leq 14000\\
&y_{1a}, y_{1b}, y_{1c}, y_{2a}, y_{2b}, y_{2c}, y_{3a}, y_{3b}, y_{3c}, y_{4a}, y_{4b}, y_{4c} \geq 0
\end{aligned}
$$

Then, we solve this linear program in R and obtain the optimal solution as follows: $y^*_{1a} = 1200$, $y^*_{1b} = 0$, $y^*_{1c} = 300$, $y^*_{2a} = 400$, $y^*_{2b} = 0$, $y^*_{2c} = 600$, $y^*_{3a} = 800$, $y^*_{3b} = 0$, $y^*_{3c} = 200$, $y^*_{4a} = 2400$, $y^*_{4b} = 0$, $y^*_{4c} = 600$. The associated optimal profit is $\$148,500 < \$150,000$. Hence, it is optimal not to introduce the new product, Premium Gasoline.

Grading Scheme: 0.5 points in total. 0.2 points for formulating the linear program model. 0.2 for solving the linear program model. 0.1 points in giving the correct recommendation.

5. Radiation Therapy (2 points)

Radiation therapy is a major treatment method for cancer patients. In radiation therapy, beams of high energy photons (γ-ray) are fired into the patient to kill cancerous cells. An important breakthrough in the history of radiation therapy is the so-called intensity modulated radiation therapy (IMRT), invented in early 1980s. To reach and effective destroy the tumor, radiation passes through healthy tissues, and damages healthy cells as well. This will seriously affect the post-treatment quality of life for the patient. The goal is thus to have the tumor cells receiving the radiation as much as possible, and the healthy tissues as little as possible. In IMRT, we achieve this goal by adjusting the intensity profile of each beam of $\gamma-$ray.

Specifically, based on CT scan of the patient, a radiation oncologist contours the tumor and the structures of the nearby healthy tissues. Then, each structure is discretized into voxels, typically $4mm \times 4mm \times 4mm$. Let's see a simple example as shown in Figure 5. This is a simplification for a tumor on the neck near the spinal cord. Tumors in this area are hard to perform surgery, so Radiation Therapy is the main treatment. The voxels in pink represent tumor tissues, Spinal Cord is represented by dark green voxels, and other healthy tissues are represented in light green tissues. The radiation oncologist plans to use 2 $\gamma-$ray beams, each with 3 beamlets, to perform the radiation therapy for this patient. We can control the intensity of each breamlet to minimize the total dose of healthy tissues (spinal cord and others), and to have the tumor tissues receiving the effective radiation doses. More specifically, the dose of radiation received by each spinal cord voxel is at most 7Gy (Gray). Otherwise, the spinal cord may receive some inevitable damages. Likewise, the dose of radiation received by each voxel of other healthy tissues should be at most 10Gy. On the other hand, the dose of radiation received by each tumor tissue voxel must be at least 7Gy, in order to effectively damage cancerous cells. When you design an IMRT treatment plan, you have to satisfy these requirements.

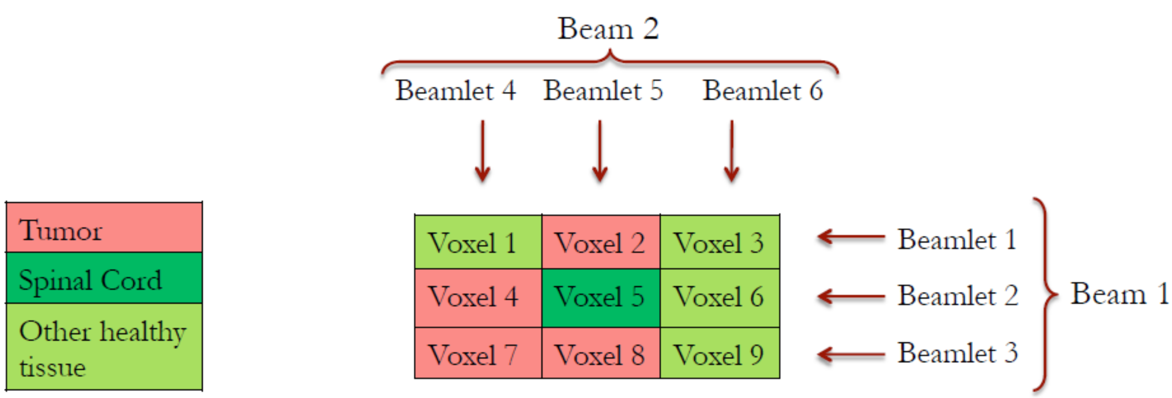

Figure 5: Radiation Therapy

Different voxels have different absorption rates, depending on the nature of the tissue and the location, as shown in Figures 6 and 7. For example, if the intensity Beamlet 4 is 2Gy, the dose of radiation received by Voxel 1 will be $2 \times 1 = 2$Gy, the dose of radiation received by Voxel 4 will be $2 \times 2 = 4$Gy, and the dose received by Voxel 7 will be $2 \times 1 = 1$Gy. The total dose received by a voxel will be the sum of doses received by this voxel from all the beamlets that pass this voxel.

For example, the dose received by Voxel 1 should be the sum of the dose from Beamlet 4 and that from Beamlet 1.

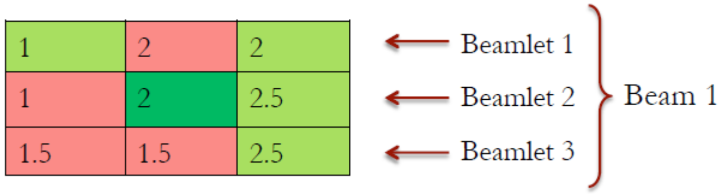

Figure 6: Absorption Rate: Beam 1

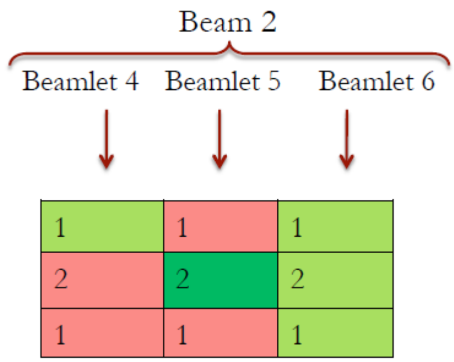

Figure 7: Absorption Rate: Beam 2

Suppose you are trying to assist the radiation oncologist to design an IMRT plan for this patient. Specifically, you need to develop optimization models to decide the intensity for each $\gamma-$ray beamlets shown in Figures 5, 6, and 7.

(a) (0.5 points) Suppose your goal is to minimize the total doses of healthy tissues. Formulate the IMRT design problem as a linear program. Clearly state the decision variables, the objective function, and constraints. What is the optimal intensity for each beamlet?

The R code for Question 5 is given in PS4_Q5.R. We define the following decision variables:

- x_1 = radiation intensity for beamlet 1
- x_2 = radiation intensity for beamlet 2
- x_3 = radiation intensity for beamlet 3
- x_4 = radiation intensity for beamlet 4
- x_5 = radiation intensity for beamlet 5
- x_6 = radiation intensity for beamlet 6

The objective function is to minimize the total doses received by healthy tissues:

$$\begin{aligned}\min\ & (1+2)x_1 + (2+2.5)x_2 + 2.5x_3 + x_4 + 2x_5 + (1+2+1)x_6 \\ =& 3x_1 + 4.5x_2 + 2.5x_3 + x_4 + 2x_5 + 4x_6\end{aligned}$$

We establish the following constraints for the this problem:

- $x_1 + x_4 \leq 10$: Radiation dose of Voxel 1 is at most 10
- $2x_1 + x_5 \geq 7$: Radiation dose of Voxel 2 is at least 7
- $2x_1 + x_6 \leq 10$: Radiation dose of Voxel 3 is at most 10
- $x_2 + 2x_4 \geq 7$: Radiation dose of Voxel 4 is at least 7
- $2x_2 + 2x_5 \leq 7$: Radiation dose of Voxel 5 is at most 7
- $2.5x_2 + 2x_6 \leq 10$: Radiation dose of Voxel 6 is at most 10
- $1.5x_3 + x_4 \geq 7$: Radiation dose of Voxel 7 is at least 7
- $1.5x_3 + x_5 \geq 7$: Radiation dose of Voxel 8 is at least 7
- $2.5x_3 + x_6 \leq 10$: Radiation dose of Voxel 9 is at most 10
- $x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$: Non-negativity constraint

Putting everything together, we formulate the problem as the following linear program:

$$\min\ 3x_1 + 4.5x_2 + 2.5x_3 + x_4 + 2x_5 + 4x_6$$

Subject to

$$x_1 + x_4 \leq 10$$
$$2x_1 + x_5 \geq 7$$
$$2x_1 + x_6 \leq 10$$
$$x_2 + 2x_4 \geq 7$$
$$2x_2 + 2x_5 \leq 7$$
$$2.5x_2 + 2x_6 \leq 10$$
$$1.5x_3 + x_4 \geq 7$$
$$1.5x_3 + x_5 \geq 7$$
$$2.5x_3 + x_6 \leq 10$$
$$x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$$

We use R to solve this linear program. We obtain the optimal solution as: $x_1^* = 1.75$, $x_2^* = 0$, $x_3^* = 2.33$, $x_4^* = 3.5$, $x_5^* = 3.5$, $x_6^* = 0$.

Grading Scheme: 0.5 points in total. 0.3 for formulating the LP. 0.2 points for solving the LP model.

(b) (0.5 points) Suppose your goal is to maximize the total doses of tumor tissues. Formulate the IMRT design problem as a linear program. Clearly state the decision variables, the objective function, and constraints. What is the optimal intensity for each beamlet?

We define the following decision variables:

- x_1 = radiation intensity for beamlet 1
- x_2 = radiation intensity for beamlet 2
- x_3 = radiation intensity for beamlet 3
- x_4 = radiation intensity for beamlet 4
- x_5 = radiation intensity for beamlet 5
- x_6 = radiation intensity for beamlet 6

The objective function is to maximize the total doses received by tumor tissues:

$$\max\ 2x_1 + x_2 + 3x_3 + 3x_4 + 2x_5$$

We establish the following constraints for the this problem:

- $x_1 + x_4 \leq 10$: Radiation dose of Voxel 1 is at most 10
- $2x_1 + x_5 \geq 7$: Radiation dose of Voxel 2 is at least 7
- $2x_1 + x_6 \leq 10$: Radiation dose of Voxel 3 is at most 10
- $x_2 + 2x_4 \geq 7$: Radiation dose of Voxel 4 is at least 7
- $2x_2 + 2x_5 \leq 7$: Radiation dose of Voxel 5 is at most 7
- $2.5x_2 + 2x_6 \leq 10$: Radiation dose of Voxel 6 is at most 10
- $1.5x_3 + x_4 \geq 7$: Radiation dose of Voxel 7 is at least 7
- $1.5x_3 + x_5 \geq 7$: Radiation dose of Voxel 8 is at least 7
- $2.5x_3 + x_6 \leq 10$: Radiation dose of Voxel 9 is at most 10
- $x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$: Non-negativity constraint

Putting everything together, we formulate the problem as the following linear program:

$$\begin{aligned}
\max\ & 2x_1 + x_2 + 3x_3 + 3x_4 + 2x_5 \\
\text{Subject to}\ & \\
& x_1 + x_4 \leq 10 \\
& 2x_1 + x_5 \geq 7 \\
& 2x_1 + x_6 \leq 10 \\
& x_2 + 2x_4 \geq 7 \\
& 2x_2 + 2x_5 \leq 7 \\
& 2.5x_2 + 2x_6 \leq 10 \\
& 1.5x_3 + x_4 \geq 7 \\
& 1.5x_3 + x_5 \geq 7 \\
& 2.5x_3 + x_6 \leq 10 \\
& x_1, x_2, x_3, x_4, x_5, x_6 \geq 0
\end{aligned}$$

We use R to solve this linear program. We obtain the optimal solution as: $x_1^* = 1.75$, $x_2^* = 0$, $x_3^* = 4$, $x_4^* = 8.25$, $x_5^* = 3.5$, $x_6^* = 0$.

Grading Scheme: 0.5 points in total. 0.3 for formulating the LP. 0.2 points for solving the LP model.

(c) (0.5 points) Now we take a more comprehensive perspective of designing the IMRT treatment plan. Specifically, we try to balance the benefit of destroying cancerous cells and the harm of damaging healthy tissues. For a given IMRT plan, we use D_T to denote the total dose of radiation received by the tumor, and D_H as the total dose of radiation received by healthy tissues. For a given weighting parameter α, the goal is to maximize

$$\alpha D_T - (1-\alpha) D_H$$

Formulate the IMRT design problem as a linear program if we set $\alpha = 0.7$. Clearly state the decision variables, the objective function, and constraints. What is the optimal intensity for each beamlet? What are the total doses of radiation received by tumor, and the total doses of radiation received by healthy tissues, respectively, associated with your prescribed IMRT plan?

We define the following decision variables:

- x_1 = radiation intensity for beamlet 1
- x_2 = radiation intensity for beamlet 2
- x_3 = radiation intensity for beamlet 3
- x_4 = radiation intensity for beamlet 4
- x_5 = radiation intensity for beamlet 5
- x_6 = radiation intensity for beamlet 6

The objective function is to maximize the total weighted doses received by tumor and healthy tissues:

$$\begin{aligned} \max\ & \alpha(2x_1 + x_2 + 3x_3 + 3x_4 + 2x_5) - (1-\alpha)(3x_1 + 4.5x_2 + 2.5x_3 + x_4 + 2x_5 + 4x_6) \\ =&(5\alpha - 3)x_1 + (5.5\alpha - 4.5)x_2 + (5.5\alpha - 2.5)x_3 + (4\alpha - 1)x_4 + (4\alpha - 2)x_5 + (4\alpha - 4)x_6 \end{aligned}$$

We establish the following constraints for the this problem:

- $x_1 + x_4 \leq 10$: Radiation dose of Voxel 1 is at most 10
- $2x_1 + x_5 \geq 7$: Radiation dose of Voxel 2 is at least 7
- $2x_1 + x_6 \leq 10$: Radiation dose of Voxel 3 is at most 10
- $x_2 + 2x_4 \geq 7$: Radiation dose of Voxel 4 is at least 7
- $2x_2 + 2x_5 \leq 7$: Radiation dose of Voxel 5 is at most 7
- $2.5x_2 + 2x_6 \leq 10$: Radiation dose of Voxel 6 is at most 10
- $1.5x_3 + x_4 \geq 7$: Radiation dose of Voxel 7 is at least 7
- $1.5x_3 + x_5 \geq 7$: Radiation dose of Voxel 8 is at least 7
- $2.5x_3 + x_6 \leq 10$: Radiation dose of Voxel 9 is at most 10
- $x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$: Non-negativity constraint

Putting everything together, we formulate the problem as the following linear program:

$$\max\ (5\alpha - 3)x_1 + (5.5\alpha - 4.5)x_2 + (5.5\alpha - 2.5)x_3 + (4\alpha - 1)x_4 + (4\alpha - 2)x_5 + (4\alpha - 4)x_6$$

Subject to

$$\begin{aligned} & x_1 + x_4 \leq 10 \\ & 2x_1 + x_5 \geq 7 \\ & 2x_1 + x_6 \leq 10 \\ & x_2 + 2x_4 \geq 7 \\ & 2x_2 + 2x_5 \leq 7 \\ & 2.5x_2 + 2x_6 \leq 10 \\ & 1.5x_3 + x_4 \geq 7 \\ & 1.5x_3 + x_5 \geq 7 \\ & 2.5x_3 + x_6 \leq 10 \\ & x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \end{aligned}$$

We use R to solve this linear program for $\alpha = 0.7$. We obtain the optimal solution as: $x_1^* = 1.75$, $x_2^* = 0$, $x_3^* = 4$, $x_4^* = 8.25$, $x_5^* = 3.5$, $x_6^* = 0$. The associated dose received by the tumor tissues is 47.25Gy, and the associated dose received by healthy tissues is 30.5Gy.

Grading Scheme: 0.5 points in total. 0.2 for formulating the LP. 0.2 points for solving the LP model. 0.1 points for getting the associated doses of radiation received by the healthy and tumor tissues.

(d) (0.5 points) As in part (c), we still consider maximizing

$$\mathrm{Opt}(\alpha) = \alpha D_T - (1-\alpha) D_H$$

Now take $\alpha \in \mathcal{S} = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. For each $\alpha \in \mathcal{S}$, denote $T(\alpha)$ as the total doses of radiation received by the tumor, and $H(\alpha)$ as the total doses of radiation received by healthy tissues, respectively, associated with your IMRT plan that maximizes $\mathrm{Opt}(\alpha)$. Please calculate $T(\alpha)$ and $H(\alpha)$ for each $\alpha \in \mathcal{S}$. Also plot the points $(T(\alpha), H(\alpha))$ for all the values of $\alpha \in \mathcal{S}$. This figure is usually called the Pareto frontier in the literature. What observations do you have from the scatter plot figure? Can you give an interpretation for the parameter α?

The model is the same as Part (c), with different α's. We summarize the total doses of radiation received by tumor and healthy tissues in the following table:

α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$T(\alpha)$	28	28	28	42.25	42.25	47.25	47.25	47.25	47.25	47.25	47.25
$H(\alpha)$	21.58	21.58	21.58	26.33	26.33	30.5	30.5	30.5	30.5	30.5	30.5

If α increases, the total dose of tumor tissues and that of healthy tissues are both higher. The parameter α captures the trade-off between destroying tumor tissues and the protecting healthy tissues. If α is larger, we put more weight on destroying tumor tissues. If α is smaller, we put more weight on protecting healthy tissues.

Grading Scheme: 0.5 points in total. 0.2 for formulating the LP. 0.2 points for getting the associated doses of radiation received by the healthy and tumor tissues for each α. 0.1 points for explaining the interpretation of α.