Q1

Model A:

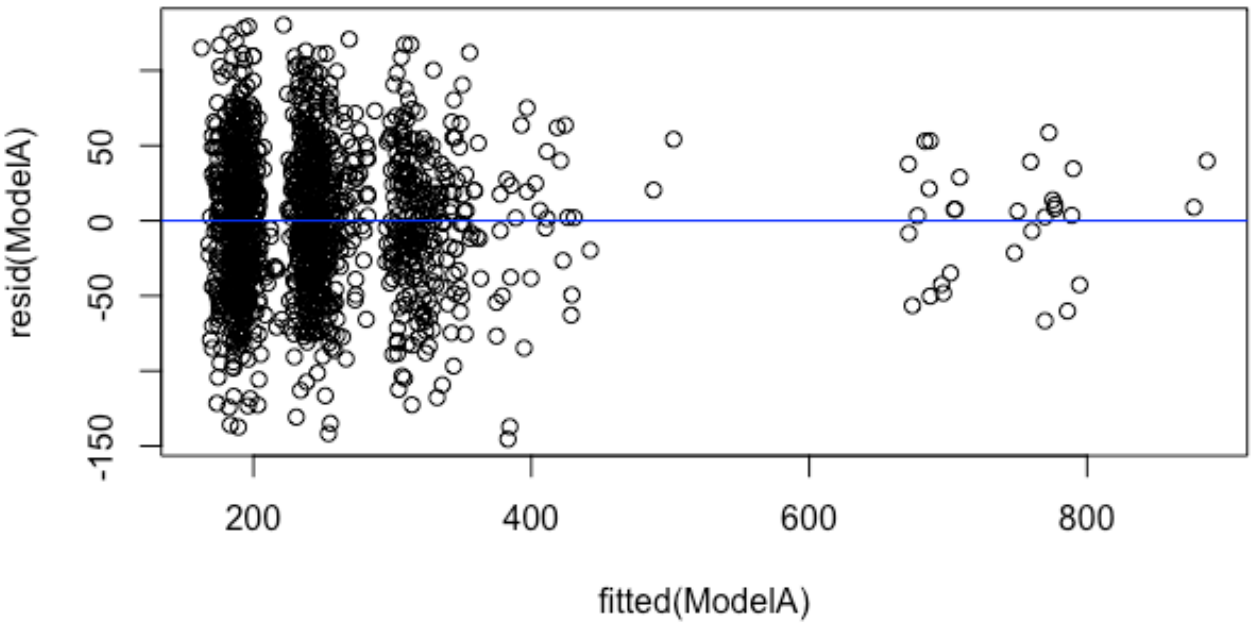
a) (10 points) Copy and paste the R code, the regression output, and the plots.

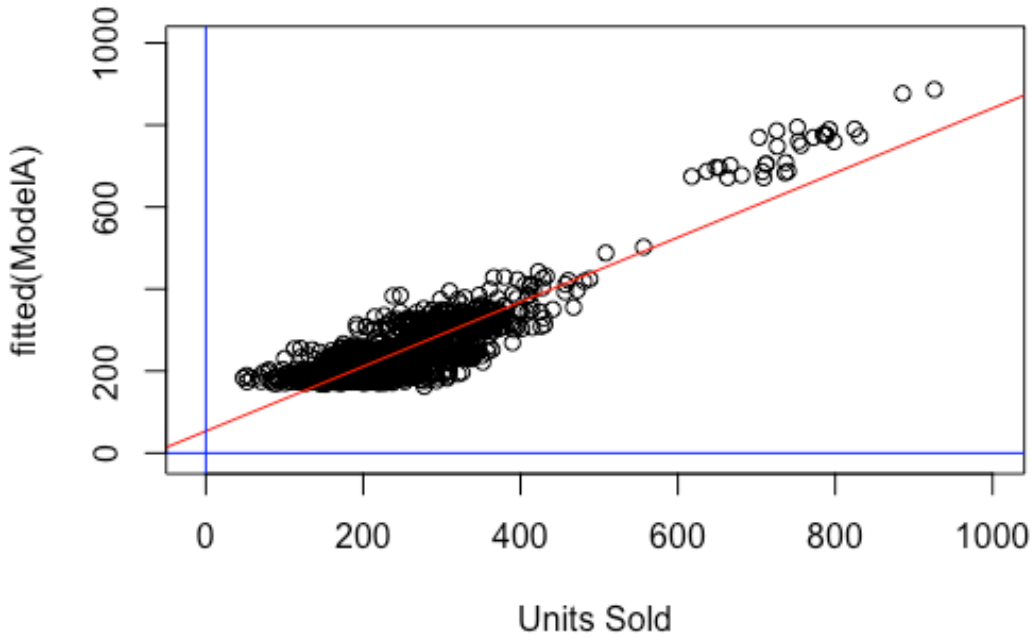
Solutions:

```r
ModelA <- lm(`Units Sold` ~ `Average Retail Price`+ `Sales Rep`+Endcap+
Demo+`Demo1-3`+`Demo4-5`+Natural+Fitness+ Region)
summary(ModelA)

plot(resid(ModelA)~fitted(ModelA))
abline(h=0,col='blue')

plot(`Units Sold`,fitted(ModelA),xlim = c(-10,1000),ylim = c(-10,1000))
abline(lm(fitted(ModelA)~`Units Sold`),col = 'red')
abline(h=0,col='blue')
abline(v=0,col='blue')

vif(ModelA)
```

```
Call:
lm(formula = `Units Sold` ~ `Average Retail Price` + `Sales Rep` +
    Endcap + Demo + `Demo1-3` + `Demo4-5` + Natural + Fitness +
    Region)

Residuals:
     Min       1Q   Median       3Q      Max
-145.452  -34.040   -0.356   33.174  130.333

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            273.1565    12.3820  22.061  < 2e-16 ***
`Average Retail Price` -21.0172     3.1245  -6.727 2.56e-11 ***
`Sales Rep`             59.7737     3.0564  19.557  < 2e-16 ***
Endcap                 441.8848     9.1190  48.458  < 2e-16 ***
Demo                   107.5483     5.7434  18.725  < 2e-16 ***
`Demo1-3`               73.9177     3.7541  19.690  < 2e-16 ***
`Demo4-5`               71.7364     5.0602  14.177  < 2e-16 ***
Natural                  0.5406     1.3853   0.390    0.696
Fitness                  0.3336     0.8321   0.401    0.689
Region                  -0.4768     0.4759  -1.002    0.317
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 48.26 on 1349 degrees of freedom
Multiple R-squared:  0.7871,  Adjusted R-squared:  0.7857
F-statistic: 554.2 on 9 and 1349 DF,  p-value: < 2.2e-16
```





b) (4 points) Discuss the performance and validity of the model, and how to improve and refine the model.

Solutions:

From the summary report, we can see that the Multiple R-squared: 0.7871 and Adjusted R-squared: 0.7857 are both high and there is no big difference between them, which means the performance of this model is good. However, we can find that the p-values of Region, Natural, Fitness are high, which suggests that changes in the these features are not associated much with changes in the Unit Sold.

We used **variance inflation factor** (**VIF**) to measures how much the variance of a regression coefficient is inflated due to multicollinearity in the model and we can see there is not presence of multicollinearity(close to 1).

```
> vif(ModelA)
`Average Retail Price`          `Sales Rep`               Endcap
              1.226387             1.347361             1.047635
                 Demo           `Demo1-3`            `Demo4-5`
              1.028945             1.082909             1.028034
              Natural              Fitness               Region
              1.065560             1.033260             1.201939
```

From the resid vs fitted plot, we can see that these points are almost randomly distributed around the straight line, and there is no obvious nonlinear trend. However, we can find the sign of non constant variance, that is, the residual decreases with the increase of x-axis.

The p-value is small enough, which means it is a valid model.

Model B:

To update the ModelA, we did some changes to the regression function.

- Since p-value refers to the hypothesis of the significance level and generally we take 5% significance level by default, we dropped features 'Region', 'Natural' and 'fitness'.
- To Fix Heteroscedasticity, we used exponential transformation on 'Units Sold'.
- We also introduced interaction terms 'Demo*Demo1-3*Demo4-5'.

```
ModelB <- lm(I(`Units Sold`^(1.5)) ~ `Average Retail Price` + `Sales Rep`+Endcap+ Demo+`Demo1-3`+`Demo4-5`+ Demo*`Demo1-3`*`Demo4-5`)
```

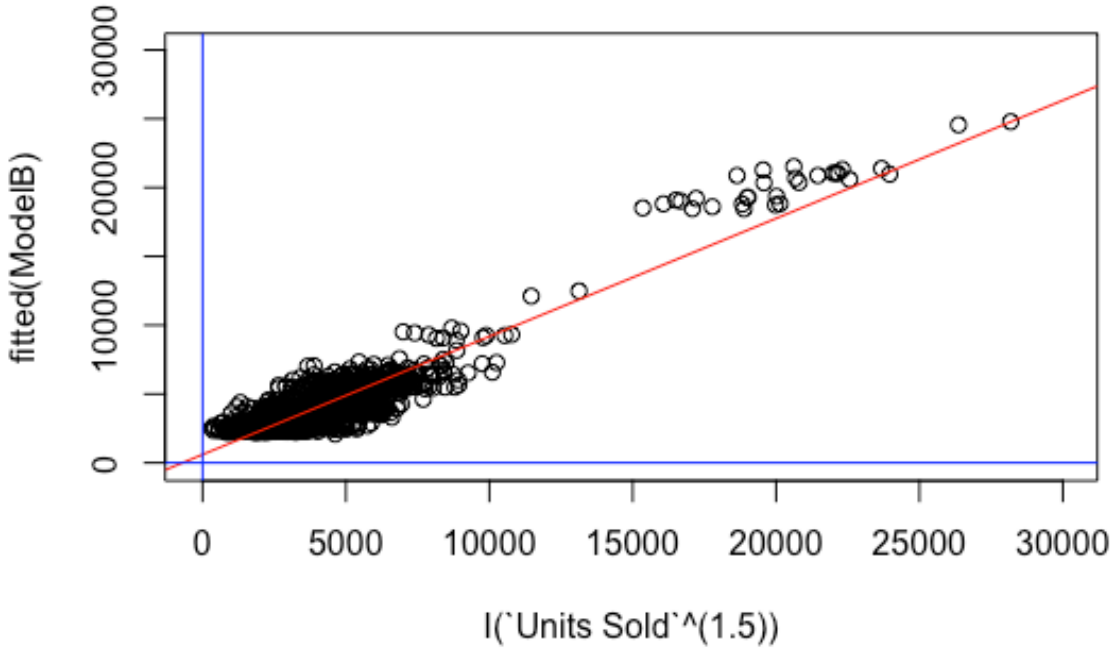
a) (6 points) Copy and paste the regression output, and the plots.

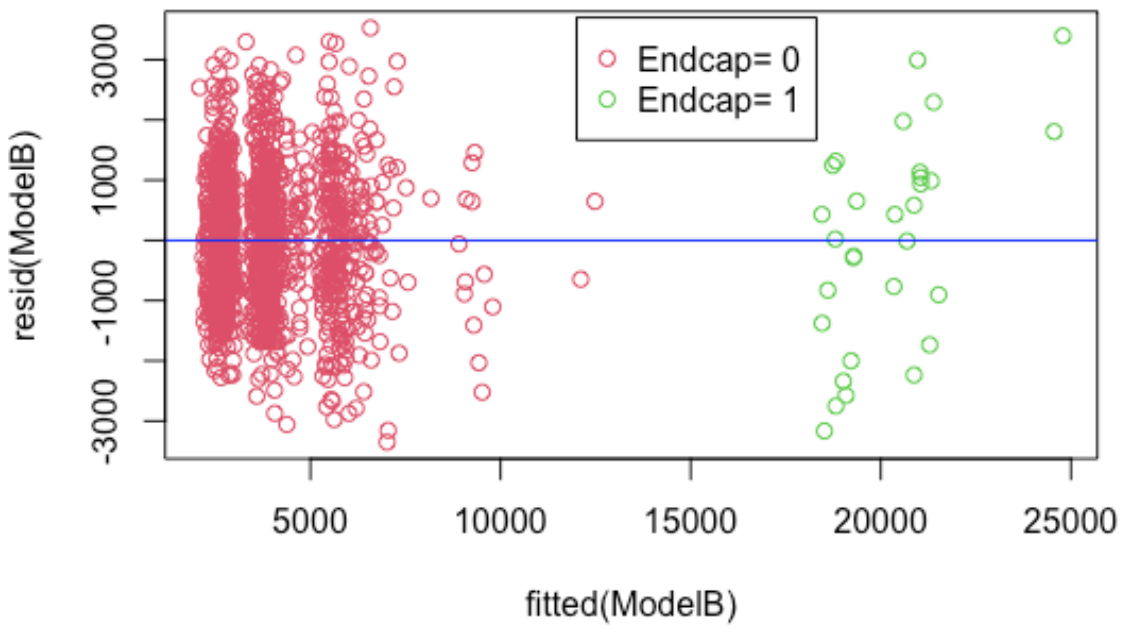
Solutions:

```
summary(ModelB)

plot(resid(ModelB)~fitted(ModelB), col=(2:3)[factor(Endcap)])
legend(12000,3700 ,legend = paste("Endcap=", 0:1), col=2:3, pch=1)
abline(h=0,col='blue')

plot(I(`Units Sold`^(1.5)),fitted(ModelB),xlim = c(-100,30000),ylim = c(-100,30000))
abline(lm(fitted(ModelB)~I(`Units Sold`^(1.5))),col = 'red')
abline(h=0,col='blue')
abline(v=0,col='blue')

```

```
Call:
lm(formula = I(`Units Sold`^(1.5)) ~ `Average Retail Price` + 
    `Sales Rep` + Endcap + Demo + `Demo1-3` + `Demo4-5` + Demo * 
    `Demo1-3` * `Demo4-5`)

Residuals:
    Min      1Q  Median      3Q     Max 
-3351.8  -780.2   -63.3   743.9  3526.8 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                4631.65     284.76  16.265  < 2e-16 ***
`Average Retail Price`     -498.94      71.24  -7.004 3.92e-12 ***
`Sales Rep`                1314.47      69.98  18.785  < 2e-16 ***
Endcap                    14930.34     214.00  69.767  < 2e-16 ***
Demo                       2678.45     153.86  17.408  < 2e-16 ***
`Demo1-3`                  1851.28      95.16  19.455  < 2e-16 ***
`Demo4-5`                  1826.27     131.87  13.849  < 2e-16 ***
Demo:`Demo1-3`             1011.29     350.17   2.888  0.00394 ** 
Demo:`Demo4-5`             -274.30     826.65  -0.332  0.74007    
`Demo1-3`:`Demo4-5`        -369.50     331.93  -1.113  0.26582    
Demo:`Demo1-3`:`Demo4-5`   1734.85    1230.14   1.410  0.15869    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
25
26 Residual standard error: 1135 on 1348 degrees of freedom
27 Multiple R-squared:  0.858, Adjusted R-squared:  0.857
28 F-statistic: 814.8 on 10 and 1348 DF,  p-value: < 2.2e-16
```





b) (2 points) Discuss the validity of the model.

Solutions:

From the summary report, we can see that the Multiple R-squared: 0.858 and Adjusted R-squared: 0.857 are both higher than those of Model A, and there is no big difference between them, which means the performance of this model is good.

The p-value is small enough, which means it is a valid model.

c) (2 points) Does the in-store demo program boost the sales? If so, for how long does the sales lift last? Explain your answer.

Solutions:

Yes, the in-store demo program boost the sales. We can see from the summary that p values of all demo features are low, which means they have huge impact on 'Unit Sold'.

The sales lift last 1-3 weeks. From the summary report, in the demo combination at different times, only 'Demo: Demo1-3' 's p-value is low, which means its change is associated with the change of Unit Sold.

d) (2 points) Does the placement of the product within the store (endcap promotion) affect the sales? Explain your answer.

Solutions:

Yes, it does.

From the summary report, the t-value of endcap is the largest, which means we can more likely to declare a relationship between speed and distance exists.

In the second plot, the green points are with endcap promotion and red ones are without endcap promotion. We can clearly see that the Unit Sold of green points are clearly higher for that of red ones.

e) (2 points) What other factors affect the sales of GoodMorning product? Explain your answer.

Solutions:

The p-values of Average Retail Price and Sales Rep are both low, which means these two factors also affect the sales of GoodMorning product.

f) (2 points) Based on the regression output, what are your recommendations to GoodMorning management?

Solutions:

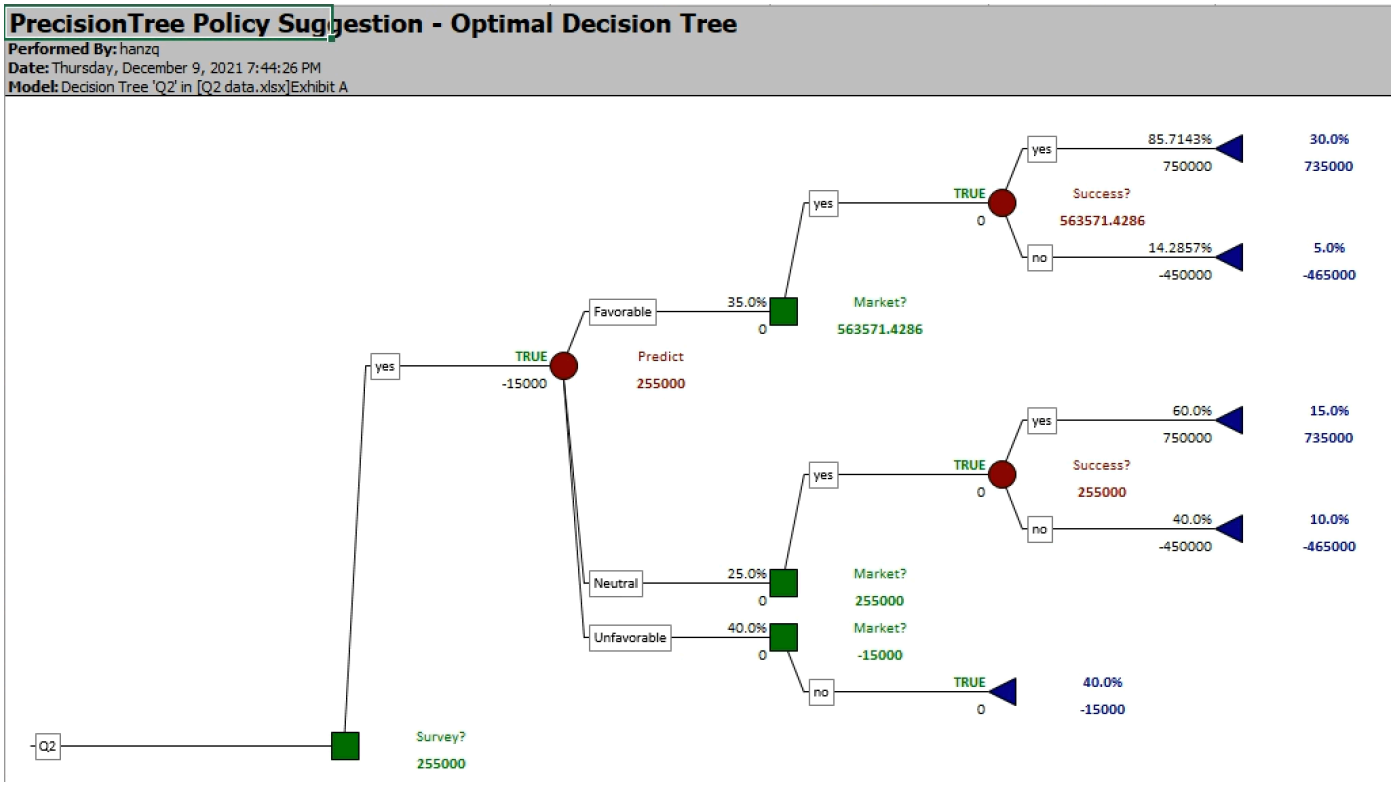
I think they should pay attention to the average retail price, sales representative, endcap, demo within 1-3 weeks. Among them, endcap has the most important impact.

Q2

a) (10 pts) Construct a decision tree for this problem (Exhibit A). Generate the optimal decision strategy tree and paste the copy on your word document.

- ***Does the optimal strategy involve conducting the survey?***
- ***What is the EMV under the optimal strategy?***

Solutions:



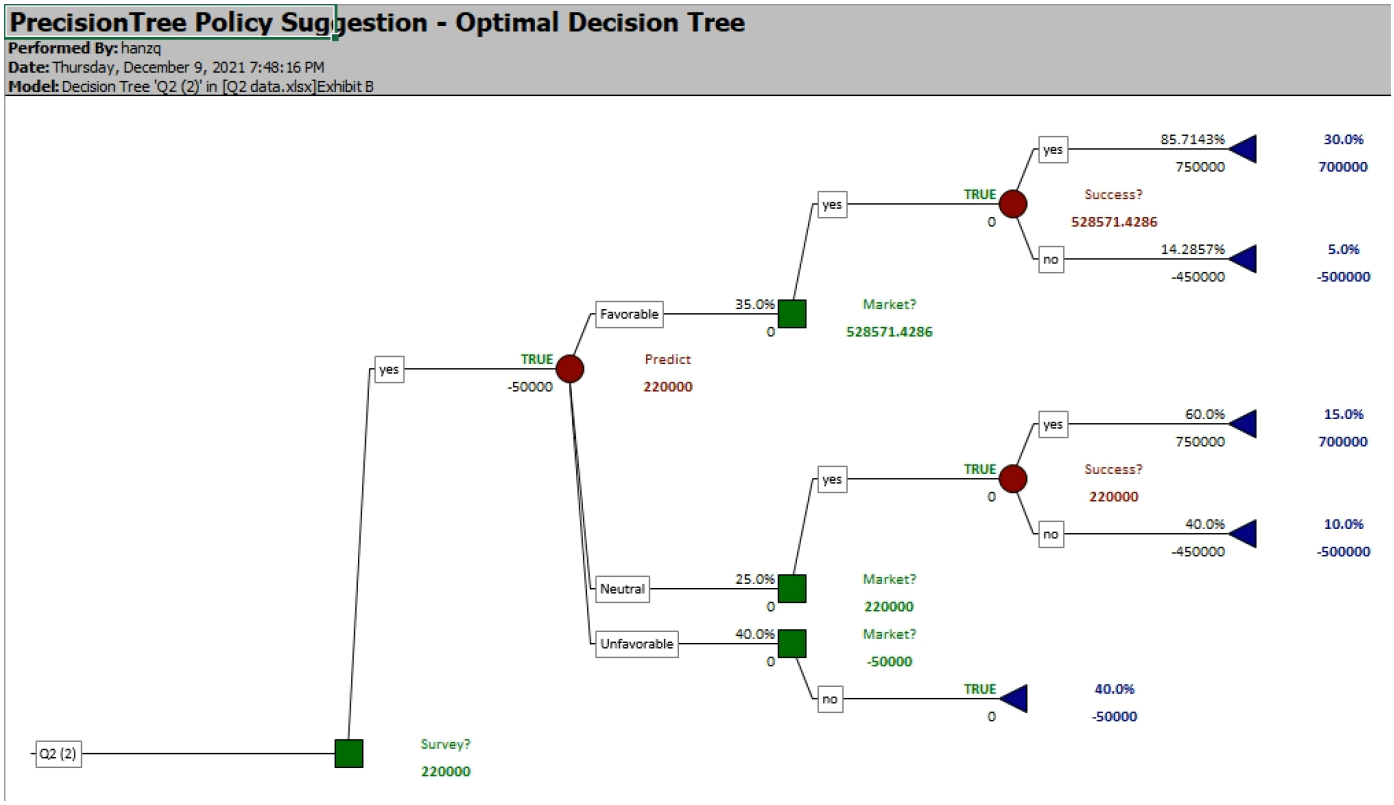
The EMV is 255000.

b) (5 pts) Suppose that the total cost of administering this survey is $50,000.

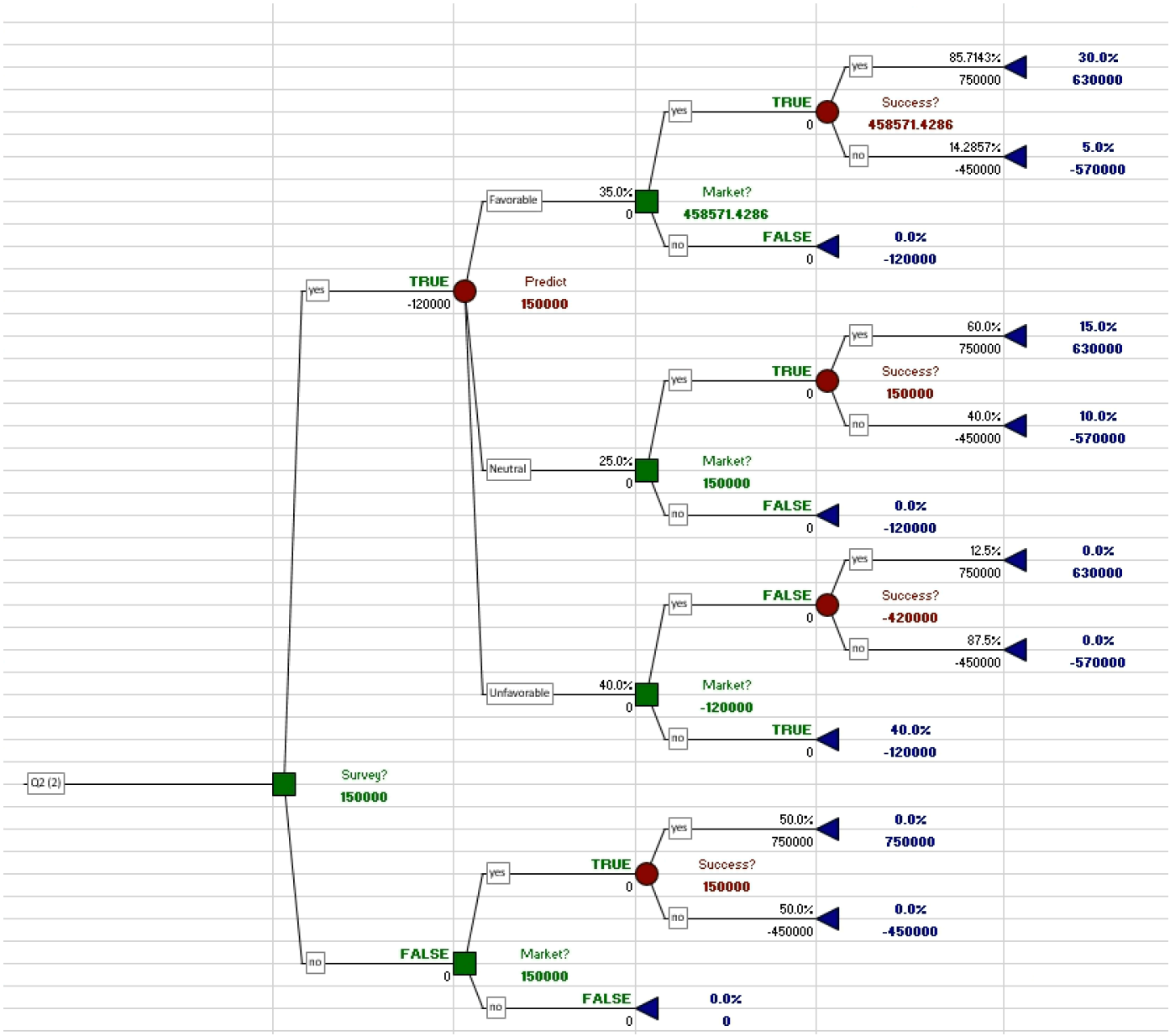
- ***Does this change the company's decision?***
- ***What is the maximum amount that the company is willing to pay for the survey?***

Solutions:

It does not change the company's decision. The new decision is as follows.



The maximum amount that the company is willing to pay for the survey is 120000. In that case the EMVs of doing the survey or not are the same: 150000. The new decision tree is as follows.



c) (5 pts) Conduct a sensitivity analysis on between 0.3 and 0.9 with 10 steps. Attached the strategy graph (Exhibit B) and paste the copy on your word document.

- ***What is the approximate value of p that changes the optimal strategy?***
- ***Explain the results in detail.***

Solutions:

The approximate value of p that changes the optimal strategy is 0.77.

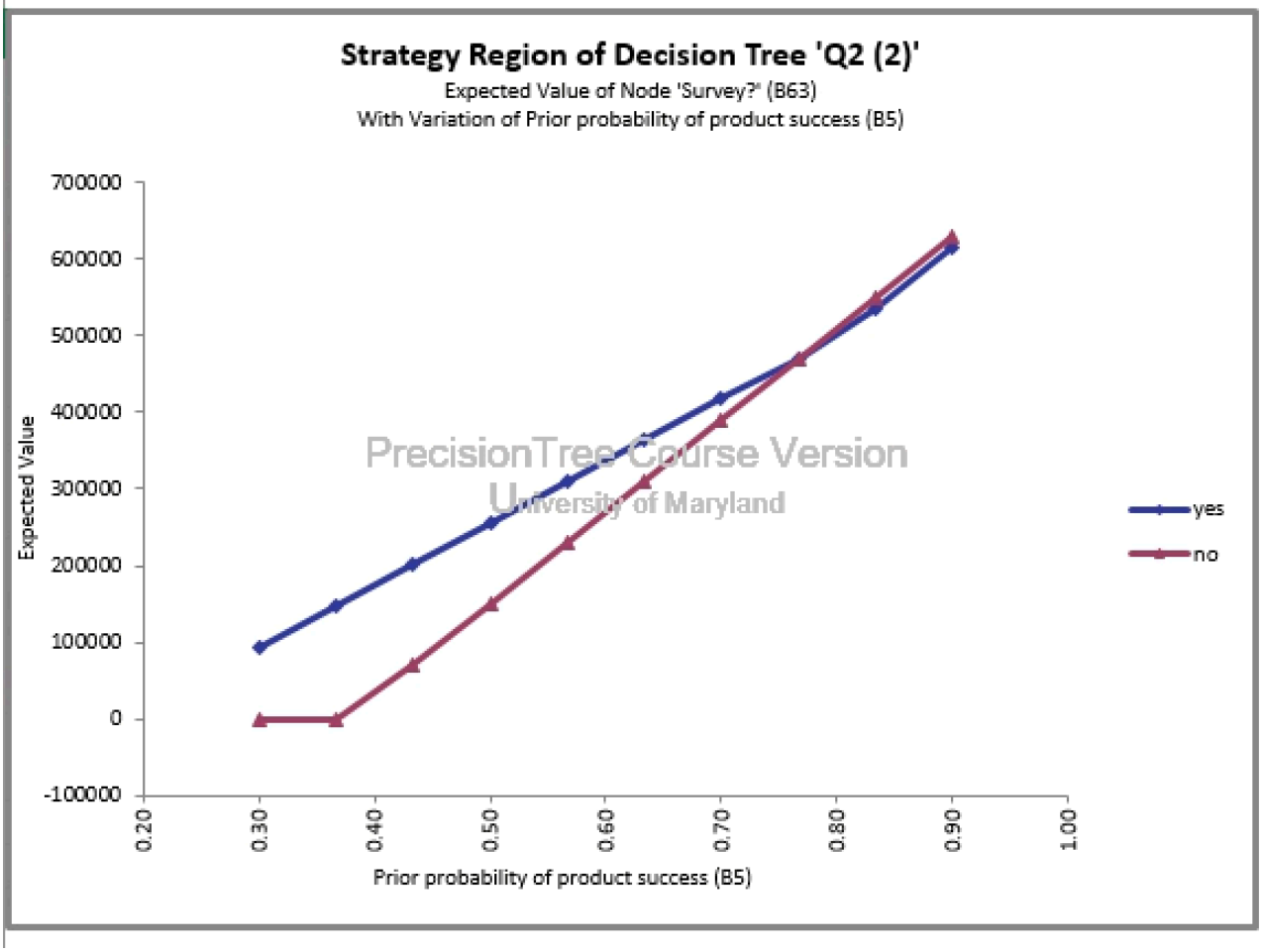
For the red line(without survey), at first the company choose not to market the product so the expected value is 0. And then start to choose to market with the increase of p.

For the blue line(with survey), the company keep the strategy all the time: market the product if the prediction is favorable or neutral, not market the product if the prediction result is unfavorable.

When the probability of product success is high,

PrecisionTree Sensitivity Analysis - Strategy Region

Performed By: hanzq
Date: Thursday, December 9, 2021 7:53:50 PM
Output: Decision Tree 'Q2 (2)' (Expected Value of Entire Model)
Input: Prior probability of product success (B5)

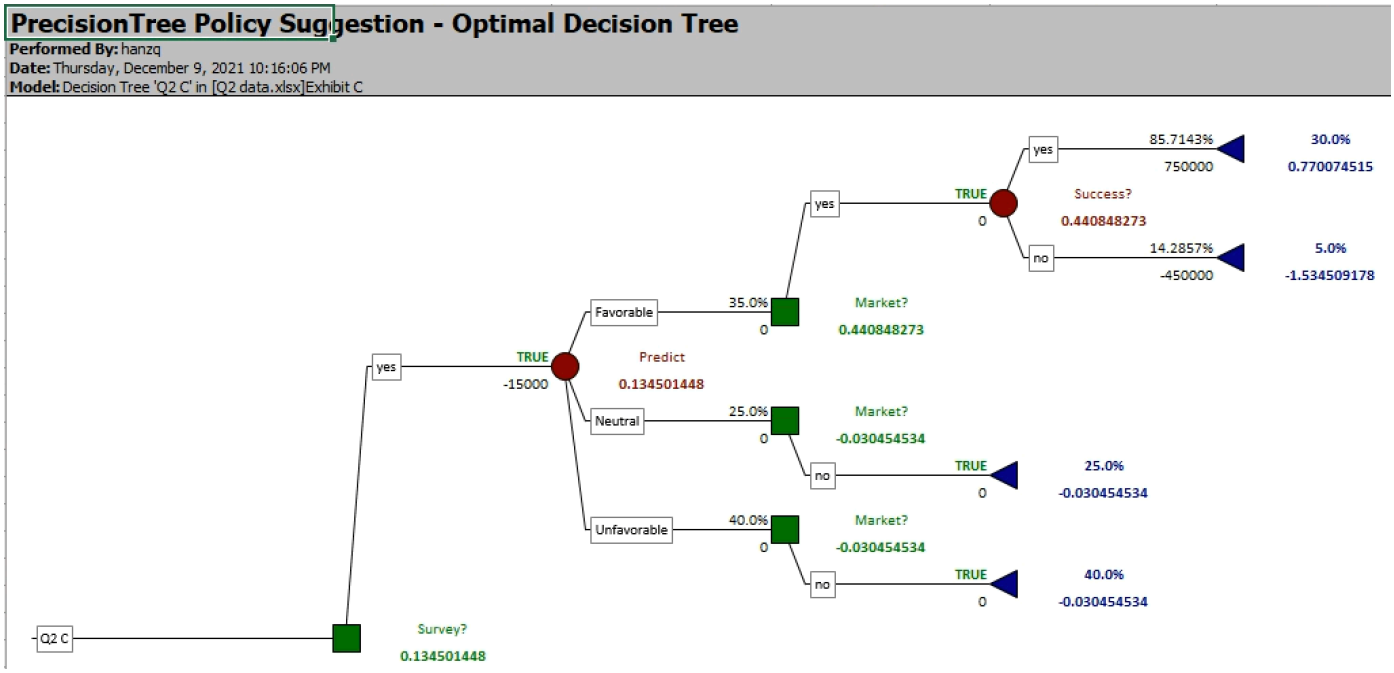


Now, we would like to find the strategy that maximizes the company's expected utility with the risk tolerance R= 500,000.

d) (5 pts) Generate the optimal decision strategy tree and paste the copy on your word document. Does this change the company's decision?

Solutions:

The company still to choose do the survey. However, they change the market decisions from different prediction result. Now the company will market the product only when the prediction result is favorable.



<u>e) (5 pts) Conduct a sensitivity analysis on p: between 0.3 and 0.9 with 10 steps. Attached the strategy graph (Exhibit C) and paste the copy on your word document. Explain the results. Particularly, explain how the second stage decision changes as p increases.</u>

For the red line(without survey), when the prob is low, the company will not market the price. However, when the probability is higher, they changed to market the product.

For the blue line(with survey), when the probability is low, the company will market the product only when the prediction result is favorable. Then when the probability is higher(about 0.57), the company's strategy changes to market the product when the prediction result is favorable or neutral.

Solutions:

PrecisionTree Sensitivity Analysis - Strategy Region

Performed By: hanzq
Date: Thursday, December 9, 2021 10:16:42 PM
Output: Decision Tree 'Q2 C' (Expected Utility of Entire Model)
Input: Prior probability of product success (B5)