

Final Report: Sales of Orthopedic Equipment

Yu-Hon Lin
May 11, 2020

Summary: Through factor analysis and principle component analysis, I predict sales of orthopedic products of hospitals where our sales are zero and identify five hospitals where our efforts will be rewarded. The forecasted sales of top 5 hospitals are \$301.60 for HID 24039 (Fort Myers, FL); \$290.46 for HID 83036 (Kinston, NC); \$241.79 for HID 99093 (Downey, CA); \$214.03 for HID 91054 (Whitfield, MS); and \$188.65 for HID 43065 (Sioux Falls, SD).

1. Introduction

There are 1996 of 4703 hospitals in the United States where our sales are zero. Among the 2707 hospitals who are doing business with us, the average sale is \$116.3 per hospital, as shown in Figure 1. Our goal is to identify hospitals with zero sales where our effort will be rewarded.

2. Dataset

Our dataset consists of 18 attributes where four of them are basic information (ZIP code, hospital ID, city, and state); eight are demographic information (number of beds, number of rehab beds, number of outpatient visits, administrative cost, revenue from inpatient, whether it is a teaching hospital, where there is a trauma unit, and whether there is a rehab unit); and five operation numbers (number of hip operations in Years 1 and 2; number of knee operations in Years 1 and 2; and number of femur operations in Year 2).

3. Data Transformation

Some of the variables need to be transformed because of severe skewness. As a result, the sales number (our response variable) has been transformed to $\ln(1 + \text{sales number})$; whereas the 10 quantitative variables have been transformed to $\ln(1 + c \times \text{variable})$, as shown in Table 1. Figure 2 illustrates the 11 histograms of transformed variables.

4. Dimension Reduction

We used principal components analysis (PCA) to reduce dimensionality of the transformed dataset. Figure 3 demonstrates that two components are sufficient for our prediction in this report.

5. Market Segmentation

We used second derivative method to obtain the number of clusters (25), as shown in Figure 4.

6. Estimating Potential Gain in Sales

We created a subset of hospitals with no current sales and transformed all variables the same way we transformed our training data. Based on the aforementioned PCA, we selected six variables consisting of BEDS, RBEDS, SIR, KNEE, REHAB, and FEMUR, as shown in Figure 5. We then used a linear model to predict the sales. We then transform the sales data back to dollar value.

7. Results

Figure 6 illustrates the summary of sales distribution (transformed back to dollar value). Despite a lower average than that of our existing customers (\$25.42 versus \$116.33), we identified five hospitals that have great potential sales. The forecasted sales of top 5 hospitals are \$301.60 for HID 24039 (Fort Myers, FL); \$290.46 for HID 83036 (Kinston, NC); \$241.79 for HID 99093 (Downey, CA); \$214.03 for HID 91054 (Whitfield, MS); and \$188.65 for HID 43065 (Sioux Falls, SD), as shown in Table 2.

Appendix

```
> summary(hosp.treat$SALES)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
 1.0    9.0    40.0 116.3  130.5 3918.0
```

Figure 1. average sales of current customers

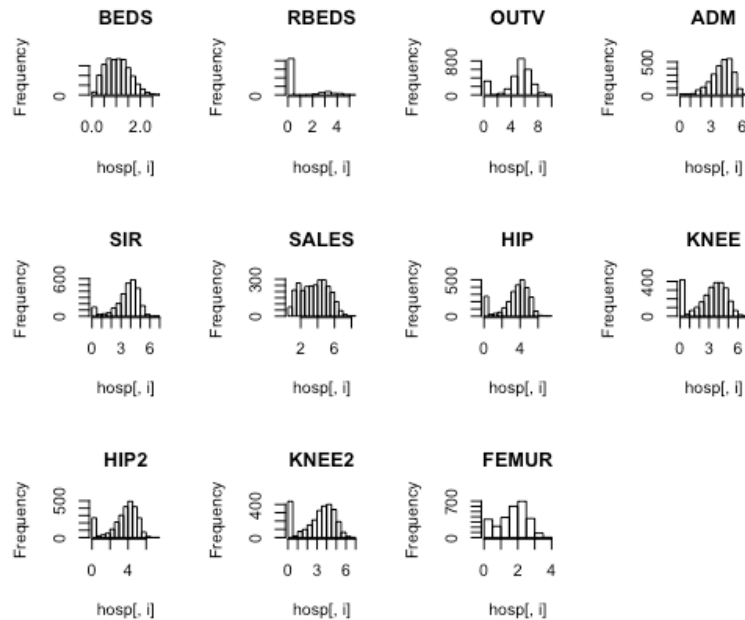


Figure 2. Histogram after data transformation

Table 1. Transformation Methods for Each Variable

Variable	Transformation
BEDS	$\ln(1 + 0.01 \times \text{BEDS})$
RBEDS	$\ln(1 + \text{RBEDS})$
OUTV	$\ln(1 + 0.01 \times \text{OUTV})$
ADM	$\ln(1 + 0.01 \times \text{ADM})$
SIR	$\ln(1 + 0.01 \times \text{SIR})$
SALES	$\ln(1 + \text{SALES})$
HIP	$\ln(1 + \text{HIP})$
KNEE	$\ln(1 + \text{KNEE})$
HIP2	$\ln(1 + \text{HIP2})$
KNEE2	$\ln(1 + \text{KNEE2})$
FEMUR	$\ln(1 + 0.1 \times \text{FEMUR})$

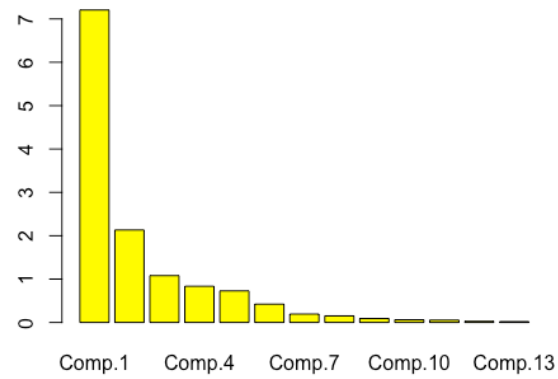


Figure 3. PCA result

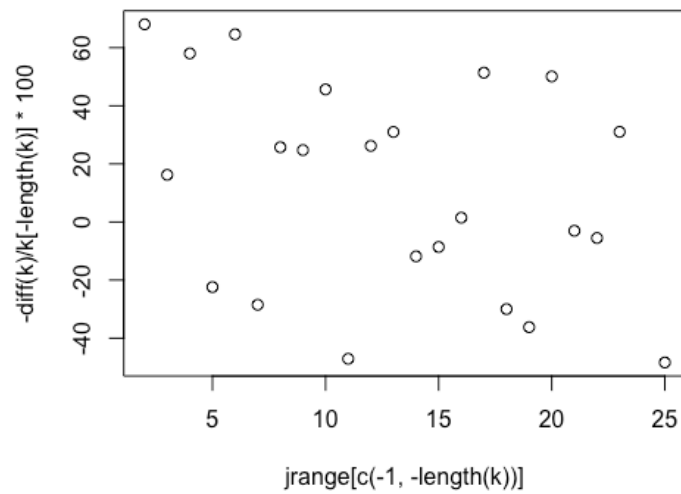


Figure 4. Cluster Analysis result using second derivative method

```
lm(formula = SALES ~ BEDS + RBEDS + SIR + KNEE + REHAB + FEMUR,
   data = hosp)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2574 -0.9883  0.1278  1.0884  4.1259

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.15323    0.09348   23.034 < 2e-16 ***
BEDS         0.45777    0.09211    4.970 7.12e-07 ***
RBEDS        0.60467    0.09596    6.301 3.44e-10 ***
SIR          -0.17008    0.04915   -3.460 0.000548 ***
KNEE         0.30370    0.02926   10.378 < 2e-16 ***
REHAB        -1.58830    0.31999   -4.964 7.35e-07 ***
FEMUR        0.33805    0.06875    4.917 9.33e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.396 on 2700 degrees of freedom
Multiple R-squared:  0.2463,    Adjusted R-squared:  0.2446
F-statistic:  147 on 6 and 2700 DF,  p-value: < 2.2e-16
```

Figure 5 Variables selected for linear model

```
> summary(hosp.test$SALES)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 3.932  9.565  18.517  25.416  32.395 301.603
```

Figure 6 Average predicted sales value of potential customers

Table 2. Top five prospective customers

	ZIP	HID		CITY	STATE	BEDS	RBEDS	OUTV	ADM	SIR	SALES	HIP
675	33901	24039	Fort Myers		FL	621	40	64495	25259	16973	301.60	333
2759	28501	83036	Kinston		NC	850	850	327541	87	0	290.46	0
2067	90242	99093	Downey		CA	610	151	164250	4594	2752	241.79	113
3091	39193	91054	Whitfield		MS	859	510	0	1985	0	214.03	0
1110	57105	43065	Sioux Falls		SD	476	28	71414	17090	13544	188.65	314
	KNEE	TH	TRAUMA	REHAB	HIP2	KNEE2	FEMUR					
675	413	0		1	1	349	372	208				
2759	0	0		0	1	0	0	0				
2067	74	1		0	1	82	76	11				
3091	0	1		0	1	0	0	0				
1110	406	1		0	1	318	354	115				