Springboard Capstone Project I:

Healthcare Provider Fraud Detection: Inferential Statistics

Zheng Zhang

April 8, 2020

In order to identify the factors that could distinguish the potentially fraudulent providers from the others, we looked at both provider level variable (such as the indicator variable that flags the providers as potentially fraudulent) as well as patient level variables such as gender and baseline chronic conditions. After the extensive exploratory data analyses, we have identified several factors that may impact the claim amount, specifically:

*I. Patient level factors*

1. Gender: Question: Is there a difference of mean claim amount between man and woman?

Approach: We used two-sample t-test to compare inpatient or outpatient mean claim amount between men and women.

Results: There is no statistical significant difference for mean inpatient claim amount ($3641 vs. $3675, P=0.51). However, there is a statistical significant difference for mean outpatient claim amount ($1278 vs. $1313, P=0.009), albeit the difference is small ($35).

1. Chronic disease: Is there difference of mean claim amount between patients with chronic disease and those without?

Approach: We used two-sample t-test to compare inpatient or outpatient mean claim amount between patients with the condition vs. those without.

Result: Across the board, the patients with the condition have significant higher claim amounts compared with the patients without on average, as shown in the following tables.

2a. Inpatient:

|  |  |  |  |
| --- | --- | --- | --- |
| Condition | Yes | No | P-value |
| Alzheimer | 5371 | 2809 | <0.0001 |
| Heart Failure | 5422 | 1943 | <0.0001 |
| Kidney Disease | 7501 | 1916 | <0.0001 |
| Cancer | 6068 | 3332 | <0.0001 |
| Obstructive Pulmonary | 7362 | 2510 | <0.0001 |
| Depression | 5022 | 2909 | <0.0001 |
| Diabetes | 4871 | 1831 | <0.0001 |
| Ischemic Heart | 4698 | 1497 | <0.0001 |
| Osteoporosis | 4608 | 3301 | <0.0001 |
| Rheumatoid Arthritis | 5102 | 3162 | <0.0001 |
| Stroke | 8111 | 3278 | <0.0001 |

2b. Outpatient:

|  |  |  |  |
| --- | --- | --- | --- |
| Condition | Yes | No | P-value |
| Alzheimer | 1623 | 1136 | <0.0001 |
| Heart Failure | 1671 | 935 | <0.0001 |
| Kidney Disease | 2062 | 951 | <0.0001 |
| Cancer | 1788 | 1231 | <0.0001 |
| Obstructive Pulmonary | 1811 | 1139 | <0.0001 |
| Depression | 1604 | 1130 | <0.0001 |
| Diabetes | 1607 | 831 | <0.0001 |
| Ischemic Heart | 1521 | 834 | <0.0001 |
| Osteoporosis | 1495 | 1224 | <0.0001 |
| Rheumatoid Arthritis | 1548 | 1212 | <0.0001 |
| Stroke | 1925 | 1244 | <0.0001 |

*II. Patient level factors: Potentially Fraud vs. Not*

Question: Is there difference of mean number of claims and the total claim amount between the providers who are flagged as “potentially fraud” vs. those that are not?

Approach: We used two-sample t-test to compare inpatient or outpatient mean claim number and the total amount between providers who are flagged as “potentially fraud” vs. those who are not.

Result: Across the board, the providers that are potentially fraudulent have filed significantly more claims and significant higher total claim amount compared to those who are not flagged on average, as shown in the following tables.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Fraud | Not-Fraud | P-value |
| Number of claims  - inpatient | 46.2 | 3.5 | <0.0001 |
| Number of claims  - outpatient | 374 | 67 | <0.0001 |
| Claim amount  - inpatient($) | 476,855 | 34,056 | <0.0001 |
| Claim amount  - outpatient($) | 107,495 | 19,138 | <0.0001 |