

PHP 2510 Final Project

Kuan-Min Lee

Methods

Before diving directly into analyze length of stay for each insurance company, several tests are taken place to make sure the data distribution of the sample from each insurance company are the same or statistically similar to each to ensure the fairness between the samples from two insurance company.

For the patient variable analysis, the influence of age, sex, and race are considered. For age analysis, the mean, standard deviation, maximum, and minimum value of age for each company are first calculated. After that, a t test of the age distributions from two companies is conducted. To test if the gender can be a factor that can influence the length of stay of the patient, a t test between the length of stay for male and female is taken to observe whether the lengths of stay for male and female are different. Lastly, for race analysis, first, the bar graphs for each insurance company are shown.

For the hospital variable analysis, the influence of hospital type is considered as a possible factor that influences the length of stay of a patient. In order to see if hospital type does have influence on the length of stay, the mean, standard deviation, maximum, and minimum of the length of stay for each hospital type, public, and private. And then, the t test is conducted to see if the mean of length of stay for each hospital type are the same.

After making sure the dataset from both insurance companies are statistically similar to each other, the mean, standard deviation, maximum, and minimum of the length of stay for both insurance companies are calculated. After that, a t test is conducted to see if the length of stay from each company are the same.

Results

Patient Variables Analysis:

Table 1: Statistical Information for Ages from Two Insurance Company		
	Insurer A	Insurer B
Mean of Age	6.85	6.57
Standard Deviation of Age	4.45	4.24
Maximum of Age	17	17
Minimum of Age	2	2

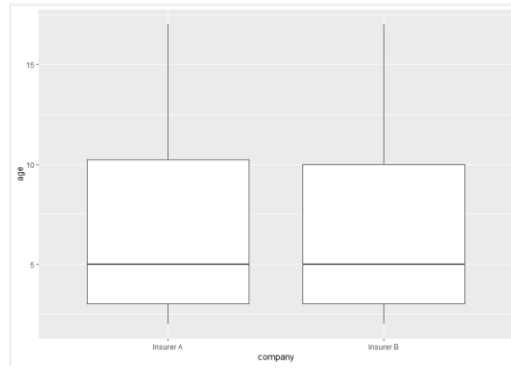


Figure 1: Boxplot for Two Insurance Companies (Left: Insurer A, Right: Insurer B)

Table 2: Two Sample T Test Information for Age in Insurer A and B	
p-value	0.362
95% Confidence Interval	[-0.3254 0.8903]
Hypothesis Accepted	Null Hypothesis

From the above table, we can make the conclusion that Age distribution from two insurance companies are not statistically different.

Table 3: Statistical Information for Length of Stay of Male and Female		
	Male	Female
Mean	2.59	2.60
Standard Deviation of Age	1.37	1.43
Maximum of Age	9	8
Minimum of Age	2	2

Table 4: Two Sample T Test Information for Length of Stay for Male and Female	
p-value	0.9263
95% Confidence Interval	[-0.2149 0.1955]
Hypothesis Accepted	Null Hypothesis

From the above table, we can make the conclusion that length of stay for two genders companies are not statistically different, which leads to the conclusion that gender has little effect for length of stay.

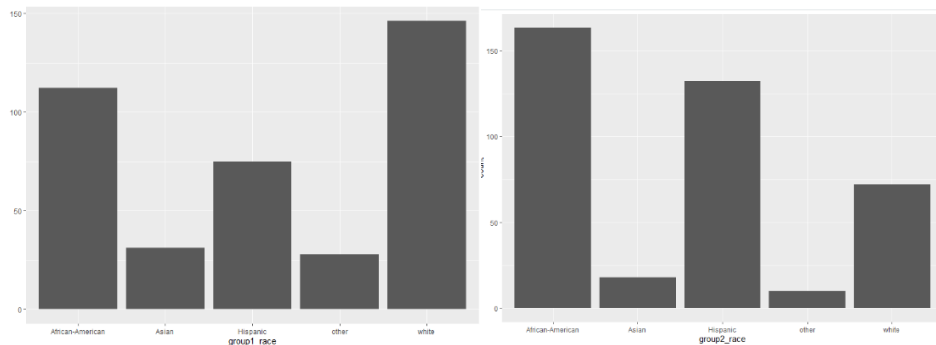


Figure 2: Histogram for Races (Left: Insurer A, Right: Insurer B)

Hospital Variables Analysis:

Table 5: Statistical Information for Length of Stay of Public and Private Hospital		
	Private	Public
Mean	2.59	2.63
Standard Deviation of Age	1.35	1.56
Maximum of Age	8	9
Minimum of Age	1	1

Table 6: Two Sample T Test Information for Length of Stay for Public and Private Hospital	
p-value	0.7873
95% Confidence Interval	[-0.3279 0.2489]
Hypothesis Accepted	Null Hypothesis

Overall Analysis:

Table 7: Statistical Information for Length of Stay of Insurer A and B		
	Insurer A	Insurer B
Mean	2.30	2.89
Standard Deviation of Age	1.16	1.52
Maximum of Age	8	9
Minimum of Age	1	1

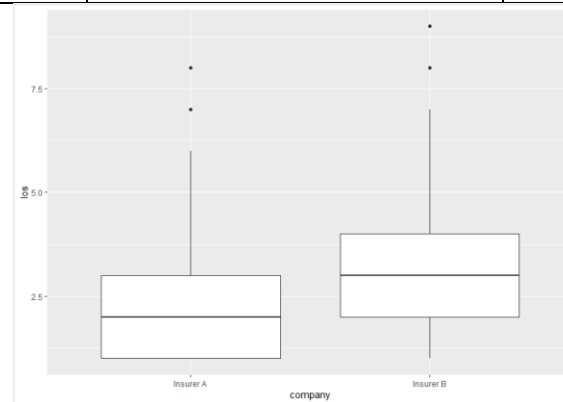


Figure 4: Boxplot for Length of Stay (Left: Insurer A, Right: Insurer B)

Table 8: Two Sample T Test Information for Length of Stay for Insurer A and B	
p-value	2.032e-09
95% Confidence Interval	[-0.7776 -0.3976]
Hypothesis Accepted	Alternative Hypothesis

From the above table, we can make the conclusion that length of stay from two insurance companies is statistically different from each other, and based on the statistical information, the Insurer A seems to have the lower value for length of stay.

Discussion:

In this project, based on the final t test information, one of the insurer companies does manage to have the lower time for length of stay. However, all the variables from the dataset that has been considered as possible influential factors on the length of stay are proved to have no influence on the length of stay, which is something that isn't expected. For example, the gender might have some influences on the length of stay since the immune ability might be different for different genders. But from this dataset, it shows that there's no difference between genders. Therefore, this might be questionable and might be one of the weakness of this dataset. As for the future enhancement, it might be better to have data point that can prove this relationship.