Assessing key cost drivers associated with inpatient admission of patients with condition X

Contents

Our Data

A summary of our dataset .

Our Data

- Clinical and financial data of patients hospitalised for a certain (X) condition
- 3,400 inpatient admission
- 3,000 unique patients
- 5 year data (2011 - 2015)
- Per row per inpatient admission

Clinical (Admission)

- Date of admission/discharge
- Pre-Op Medication
- Symptoms
- Lab Results
- Weight / Height

Patient

- Demographics
- Medical History

Financial

- Total Bill per admission

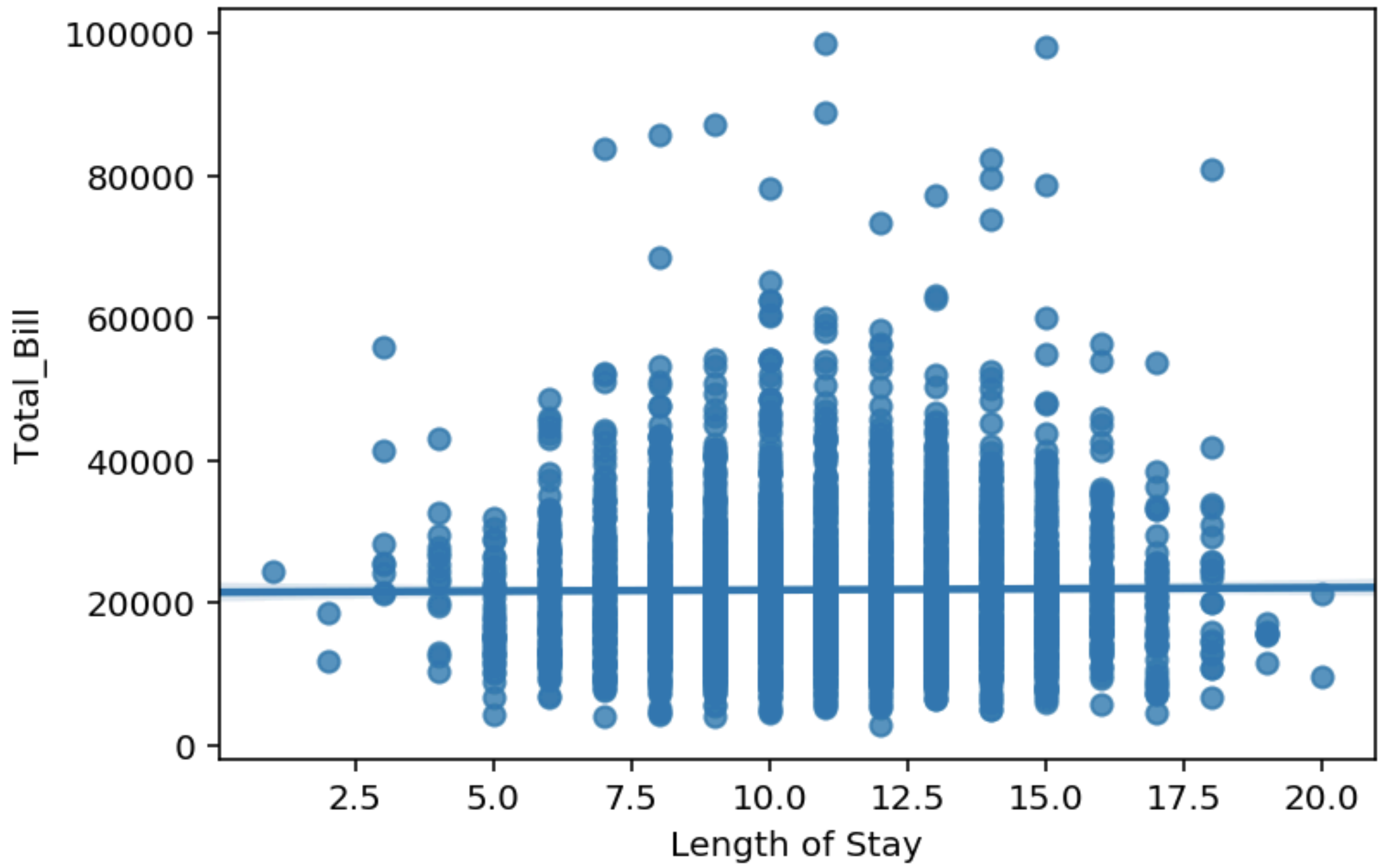
New Variables

- Length of Stay
- Age
- BMI
- BMI Risk Level
- Readmission per year
- Year of admission
- Number of Medical History
- Number of Symptoms
- Number of Pre-Op Medications

Clean-Up

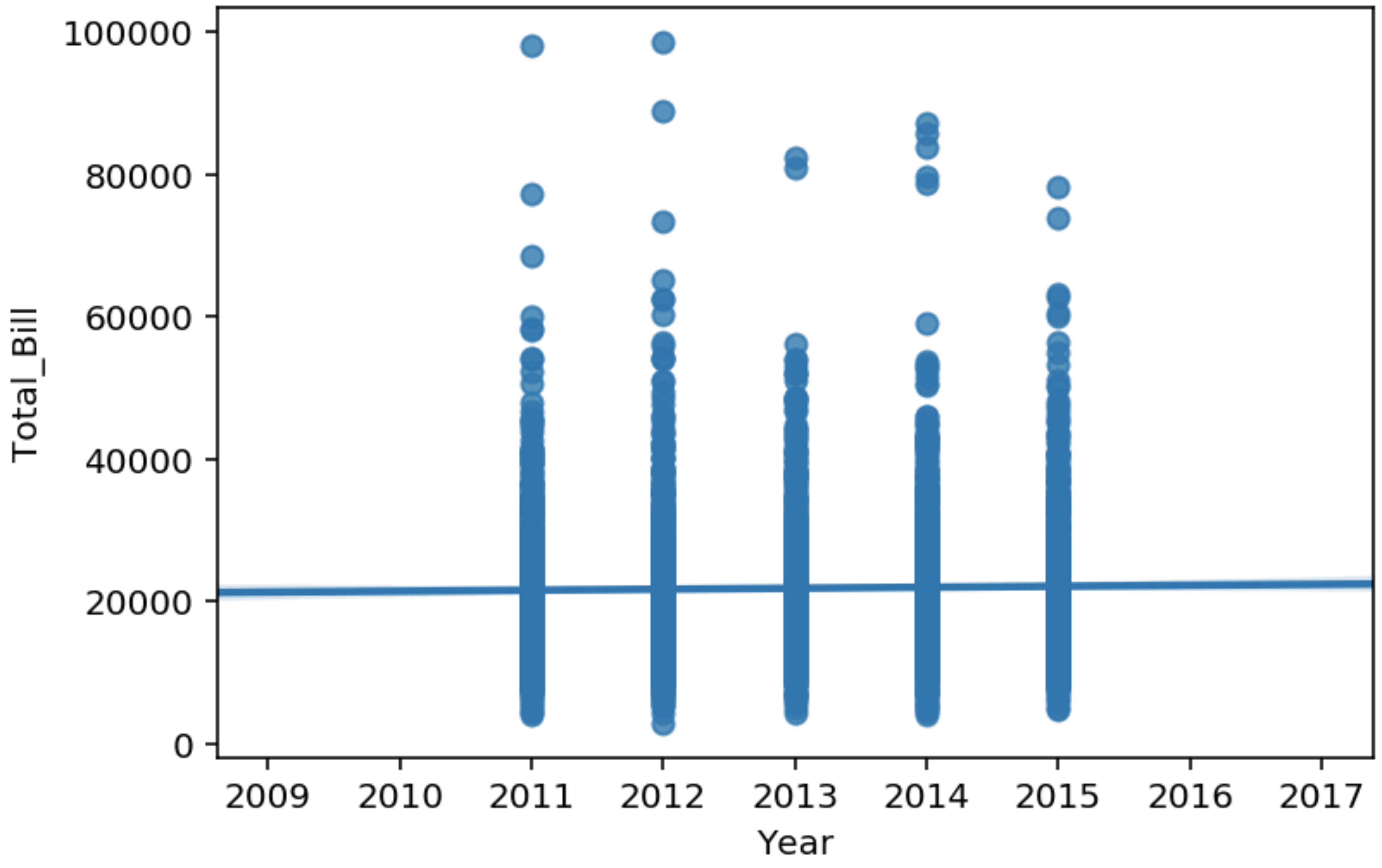
- Standardising strings
- Converting into numeric
- Imputing null values

Length of Stay



Pearson Correlation Coefficient
R = 0.0090
P-value = 0.5996

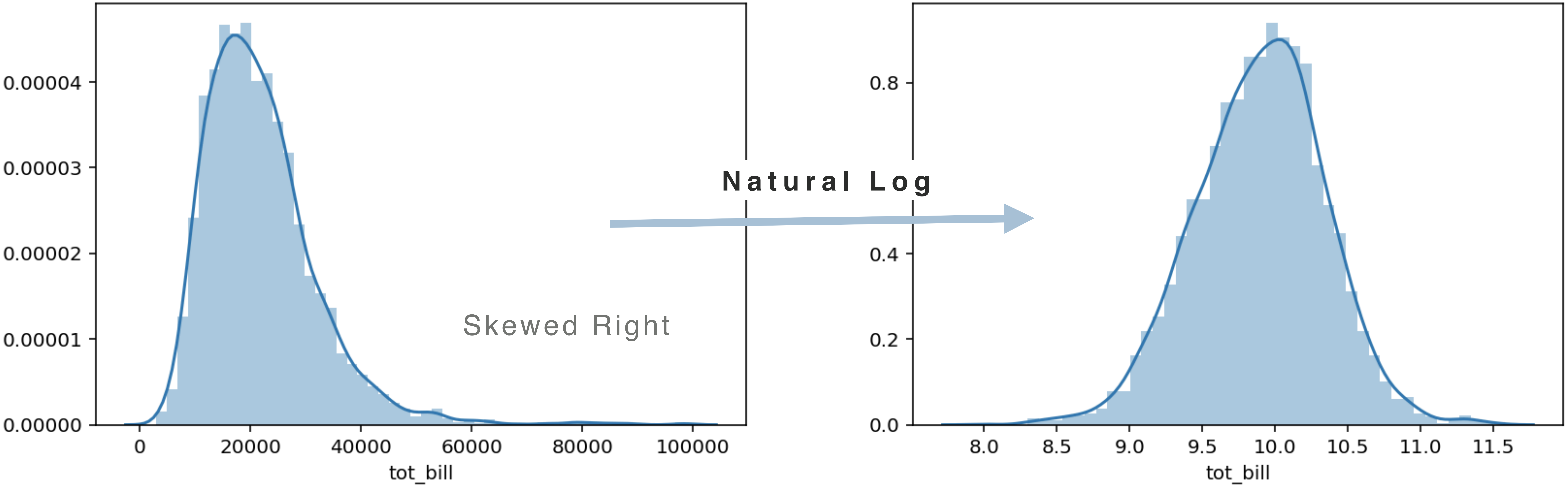
Year of Admission



Pearson Correlation Coefficient
R = 0.019
P-value = 0.2632

Target Variable

Distribution



Total Bill **Total Bill (Log)**

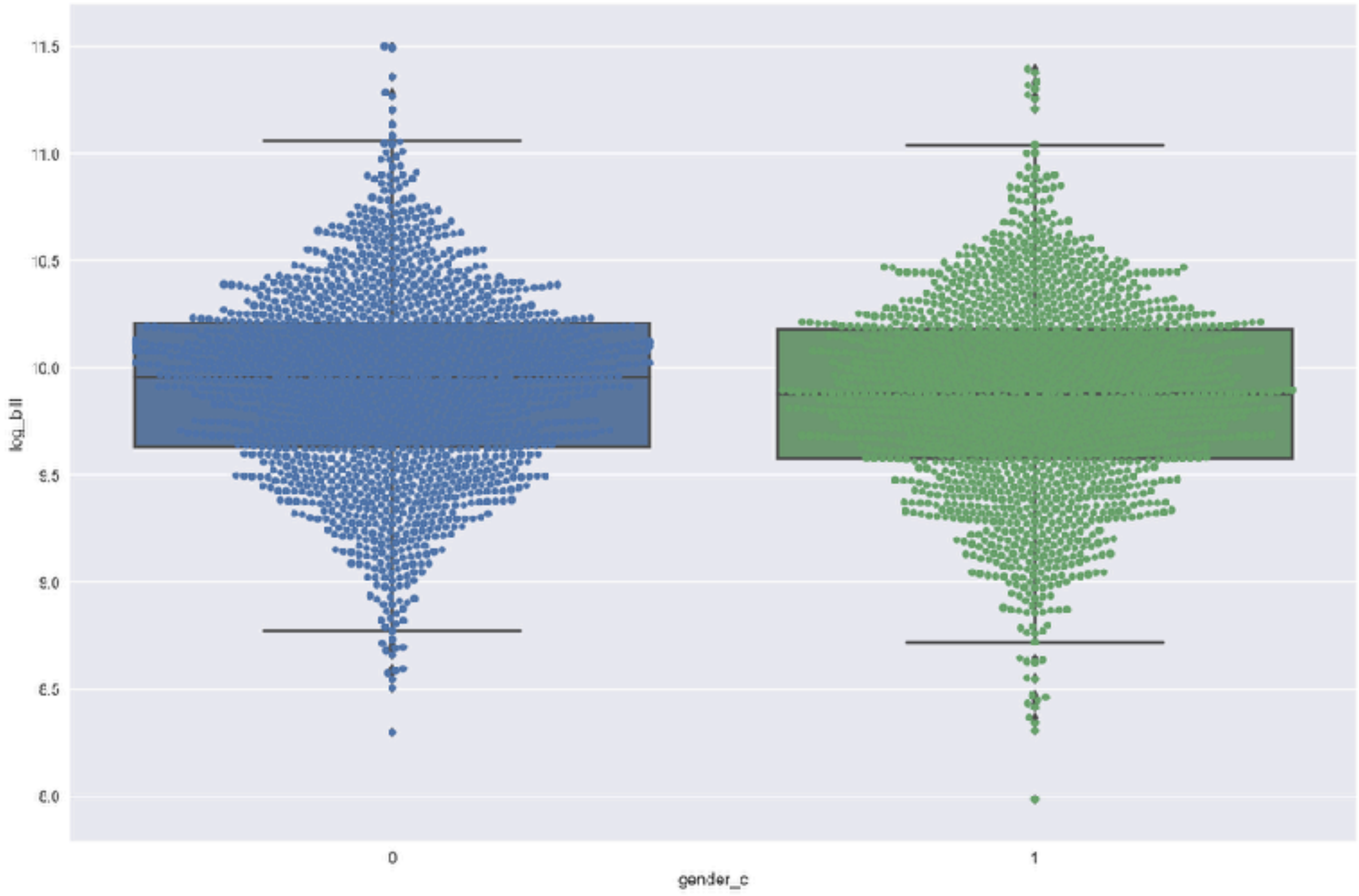
To ensure approximate normality

Demographics

		Count (per patient)	Total Bill (per adm)	
			Mean	P-Value
Total		**3,000**	**21,859**	
Gender	Female (1)	1,497	21,273	**<0.01***
	Male (0)	1,503	22,446	
Resident Status	Singaporean (0)	2,392	20,211	**<0.01***
	PR (1)	465	24,370	
	Foreigner (2)	143	41,704	
Race	Indian (1)	295	23,682	**<0.01***
	Chinese (2)	1,915	19,118	
	Malay (3)	629	29,506	
	Others (4)	161	21,320	
Age	< 55	1,755	19,334	**<0.01***
	> 55	1,252	25,398	

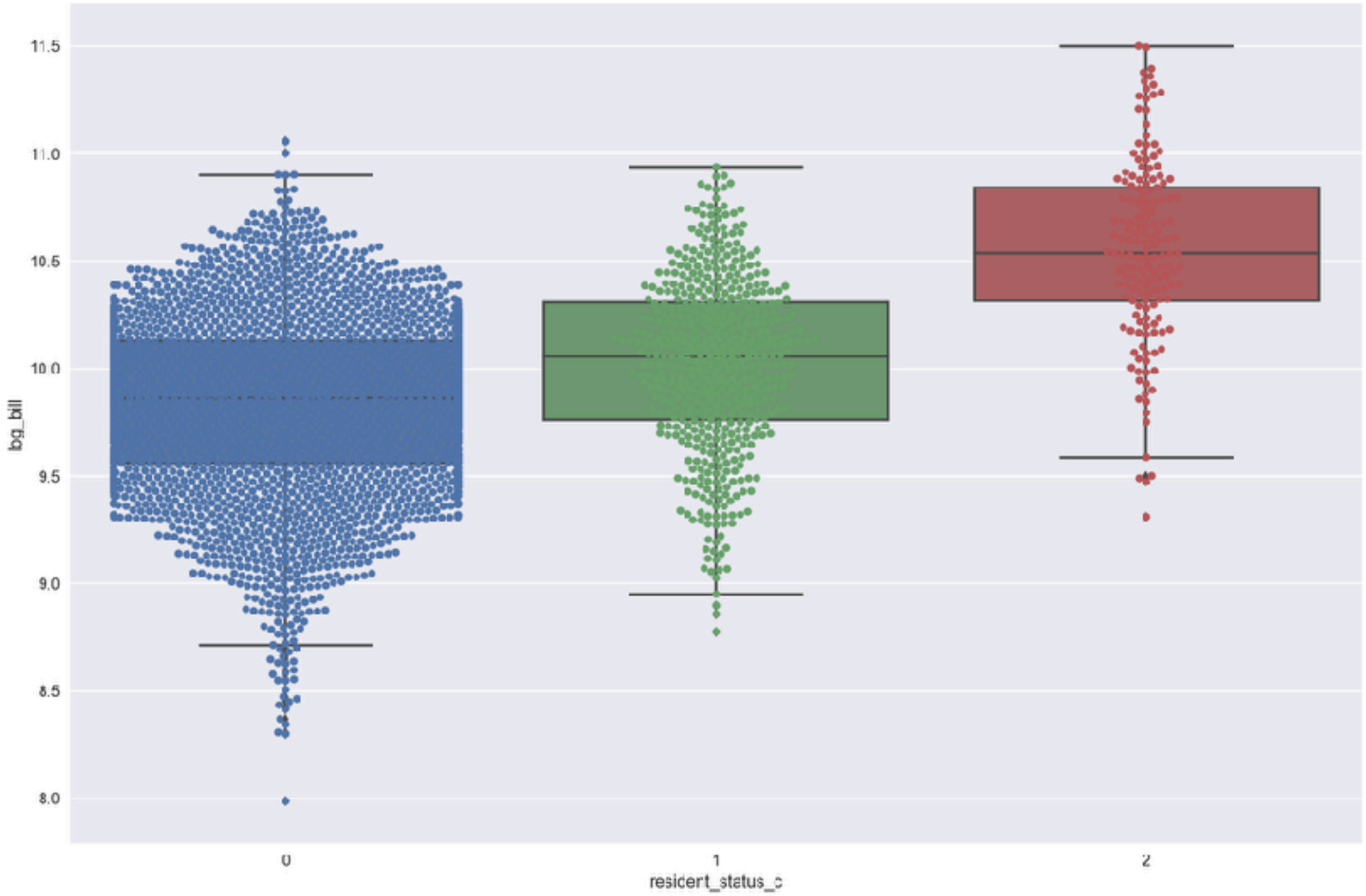
* Not tested for homogeneity and normality assumptions due to time constraint. One-way ANOVA is used

Gender



Male Female

Resident Status

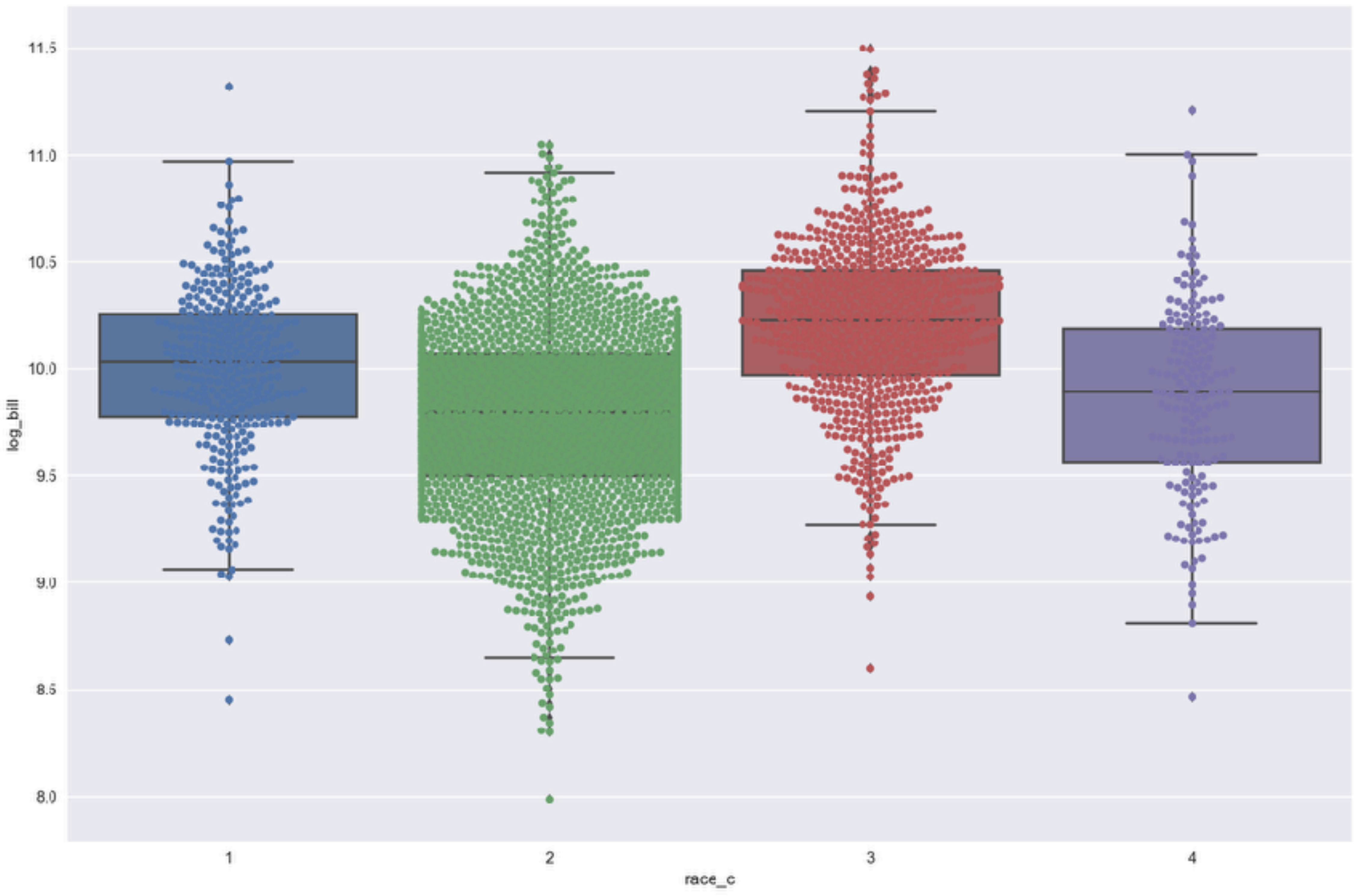


Singaporean Permanent Resident Foreigners

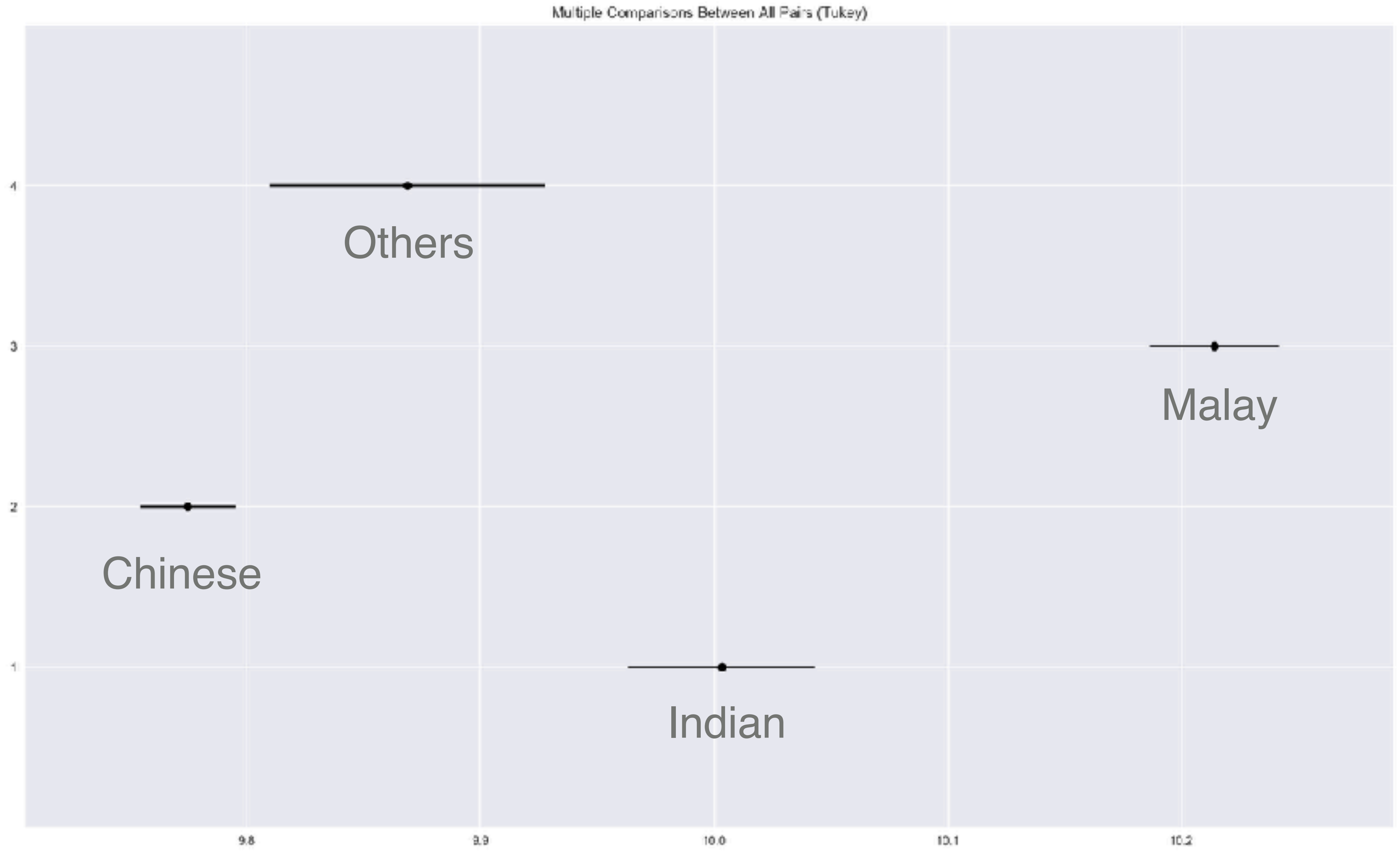
- Almost equal representation between male and female
- Male having 5.5% higher mean than female

- Singaporean have higher subsidies than PR and foreigners

Race



Indian Chinese Malay Others



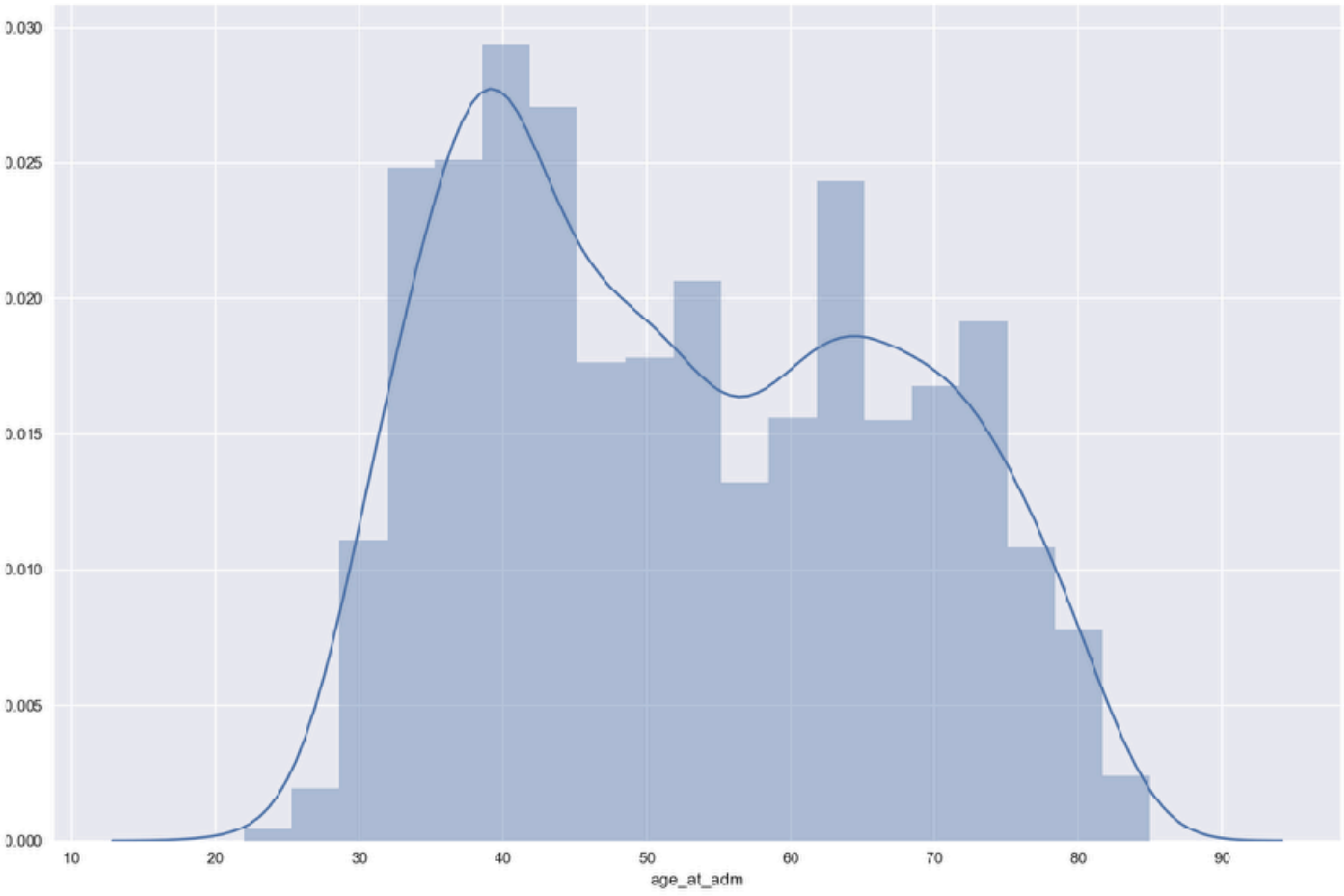
Multiple Comparison of Means - Tukey HSD,FWER=0.05

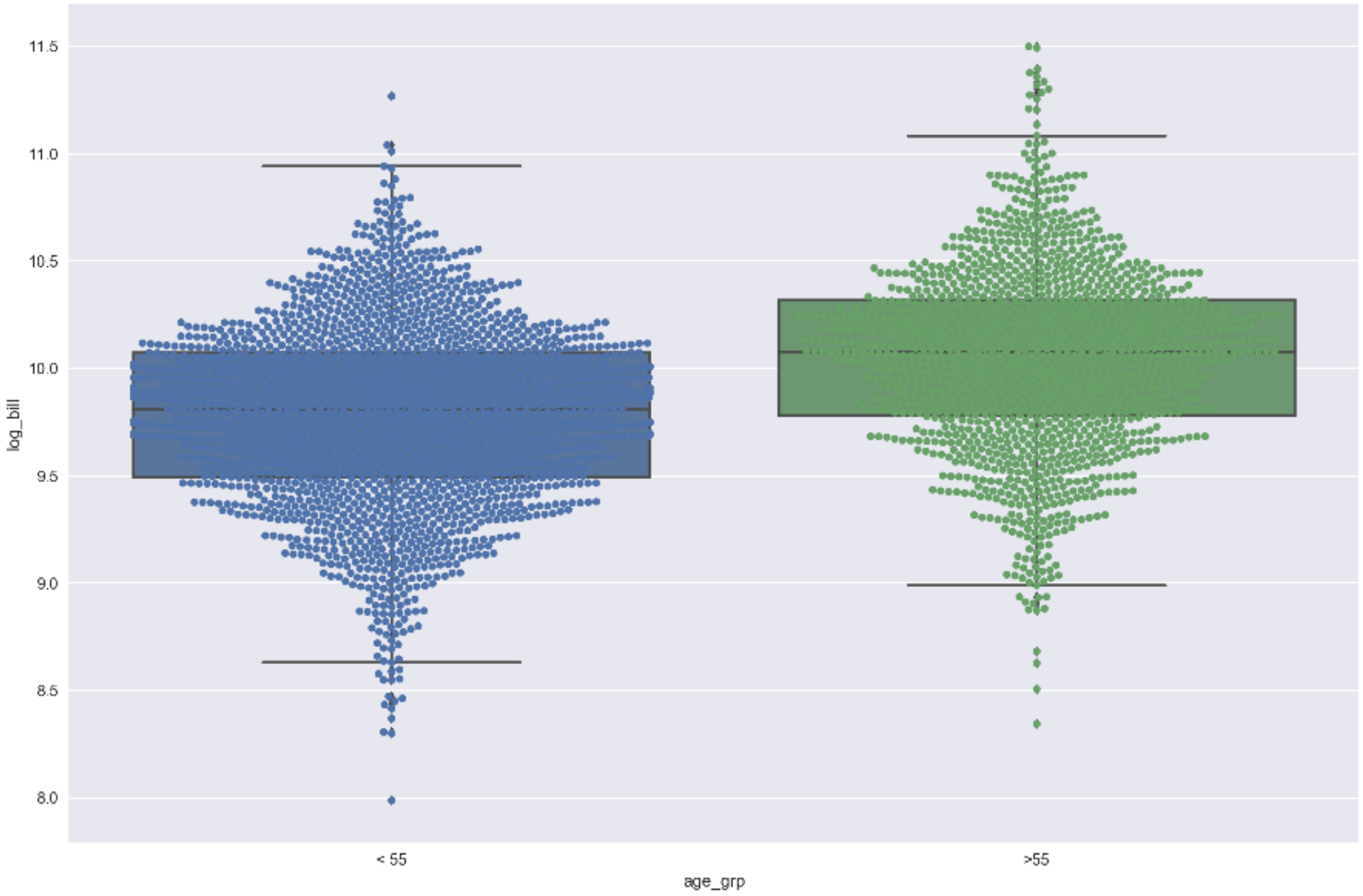
group1	group2	meandiff	lower	upper	reject
1	2	-0.2285	-0.2896	-0.1673	True
1	3	0.2108	0.1415	0.2801	True
1	4	-0.1343	-0.2308	-0.0379	True
2	3	0.4393	0.3936	0.4849	True
2	4	0.0941	0.013	0.1753	True
3	4	-0.3451	-0.4325	-0.2577	True

- There are more Chinese represented in our data
- Malays have the highest mean out of all races

• Not tested for homogeneity assumptions due to time constraint. Tukey's Honestly Significant Difference

Age





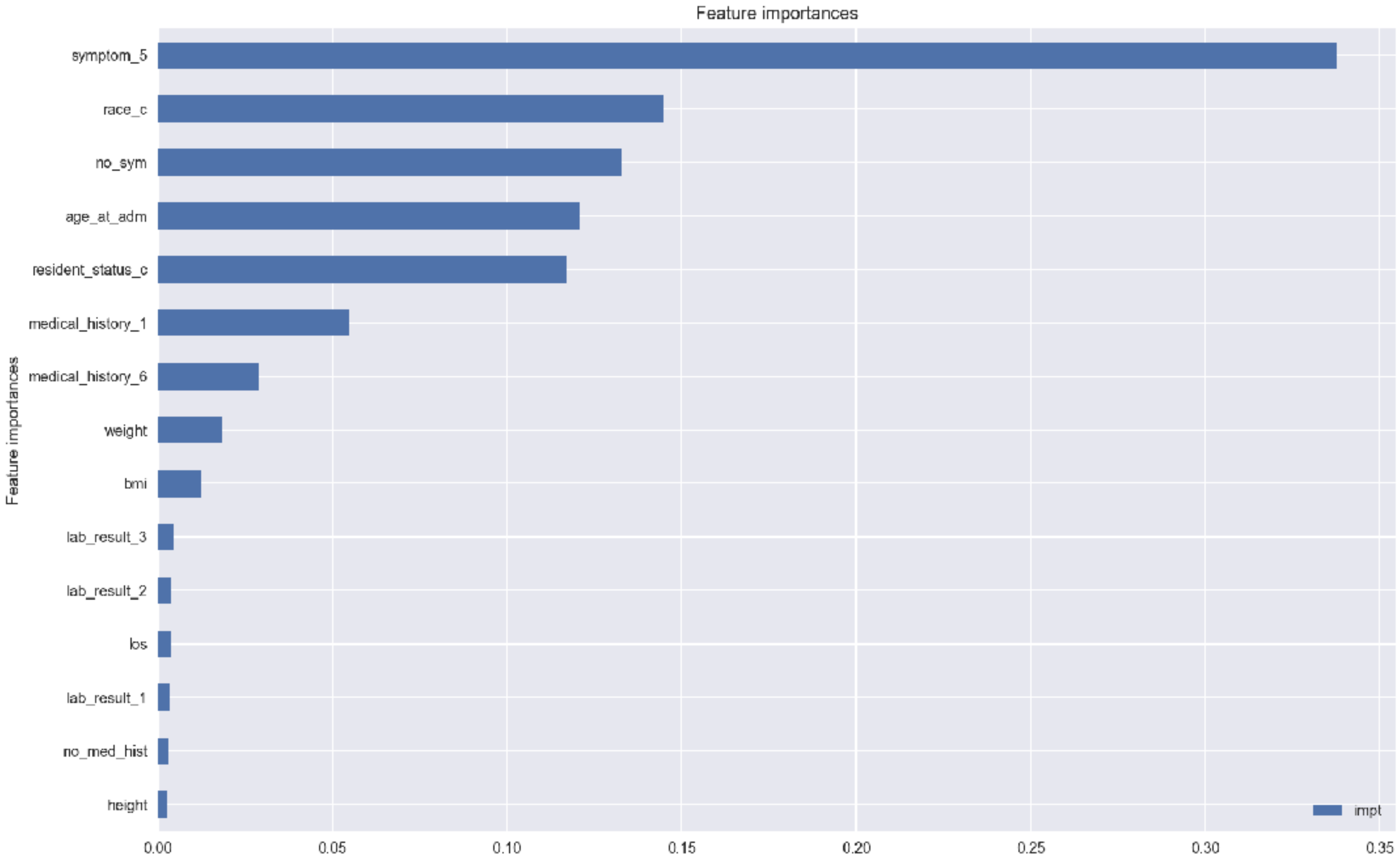
- Bimodal Distribution - We have 2 groups - above 55 and below 55
- Patients > 55 years old have ~31% higher mean total bill than those < 55

Which variable matters?

A quick way to help us focus on what matters .

Decision Tree

A quick dirty way to help zoom into important variables

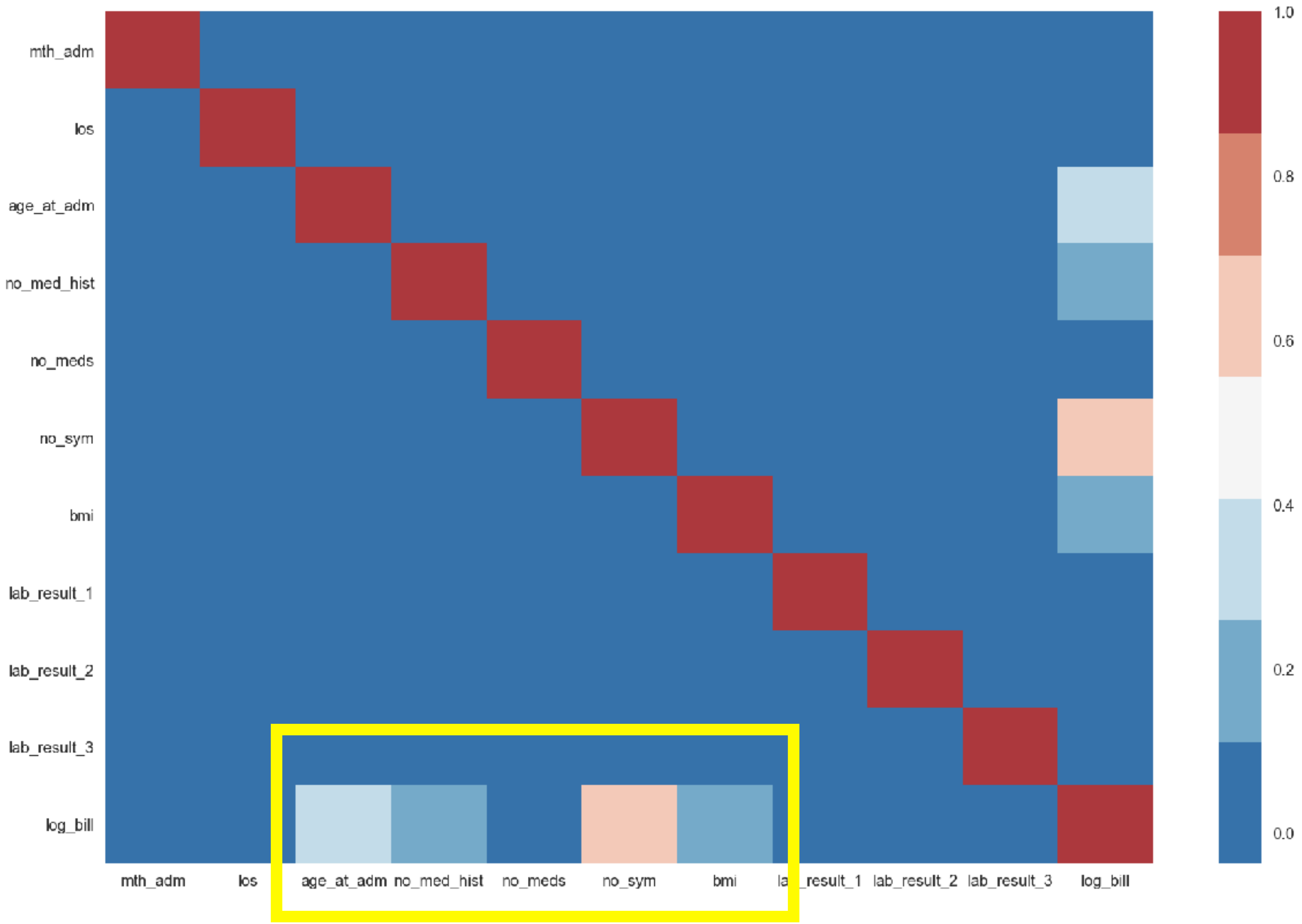


Top 10

1. Symptom 5
2. Race
3. Number of Symptoms
4. Age
5. Resident Status
6. Medical History 1
7. Medical History 6
8. BMI (Height & Weight)
9. Lab Result 3
10. Lab Result 2

Correlation Coefficient Matrix

On continuous variables



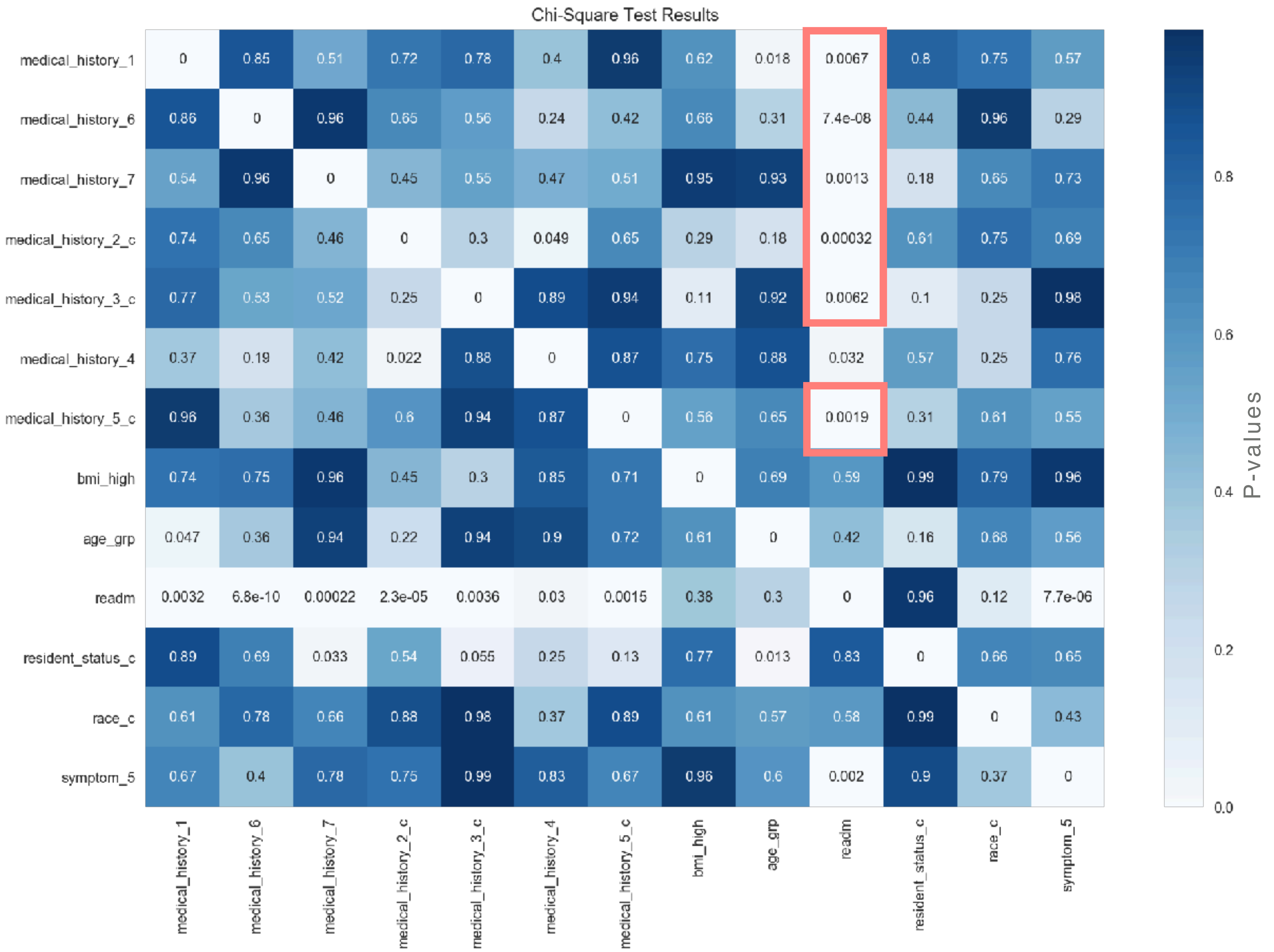
Some positive correlation

- Age
- Number of medical history
- Number of symptoms
- BMI

Deep Dive

Into each variables .

Number of encounters / year: CHI-Square Test



- The presence of medical history have high correlation with > 1 encounter per year
- P-values < 0.01

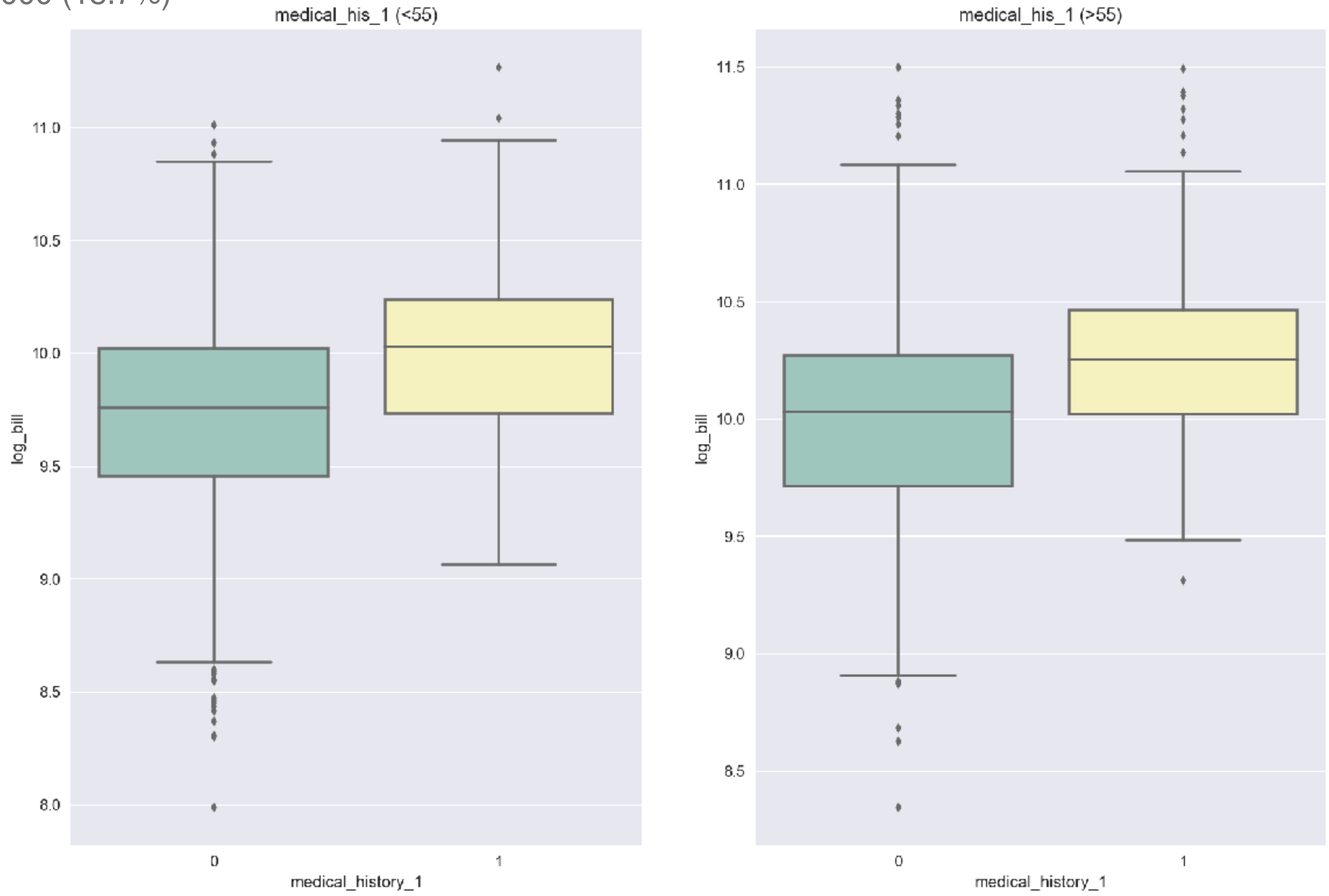
Medical History

		Count	Total Bill	
			Mean	P-Value
Total		*3,314*	*21,859*	
Number of Med History	0	795	19,497	**<0.01***
	1	1,258	21,463	
	2	800	24,696	
	3	310	26,532	
	4	47	31,191	
	5	7	36,853	
Med Hist 1		**562**	26,850	**<0.01***
Med Hist 2		953	22,357	0.086
Med Hist 3		459	22,205	0.023
Med Hist 4		175	21,480	0.943
Med Hist 5		196	23,129	0.018
Med Hist 6		**839**	24,175	**<0.01***
Med Hist 7		842	22,484	0.040

- Not tested for homogeneity and normality assumptions due to time constraint. One-way ANOVA is used

Medical History 1

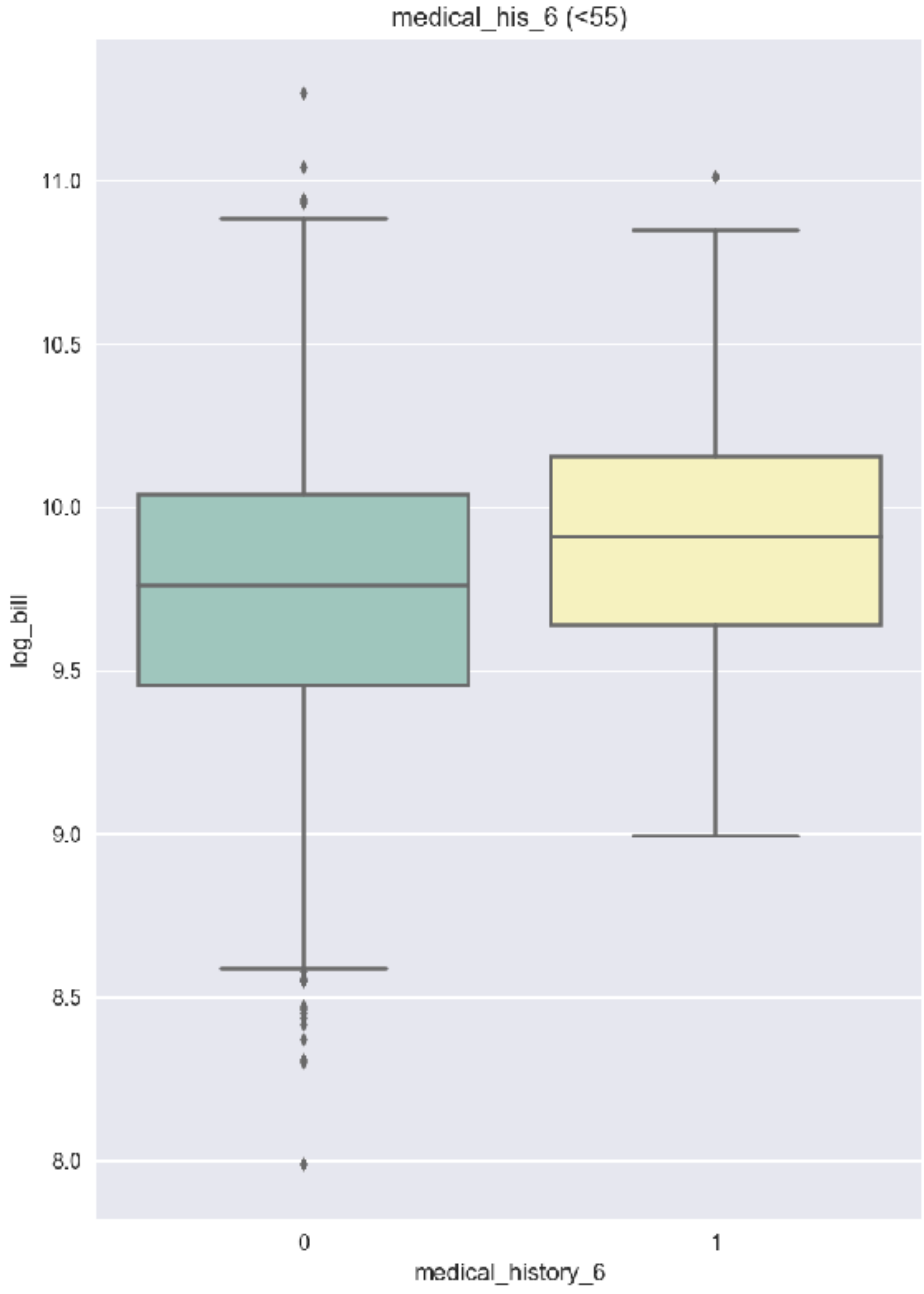
Total Count: 562 / 3,000 (18.7%)

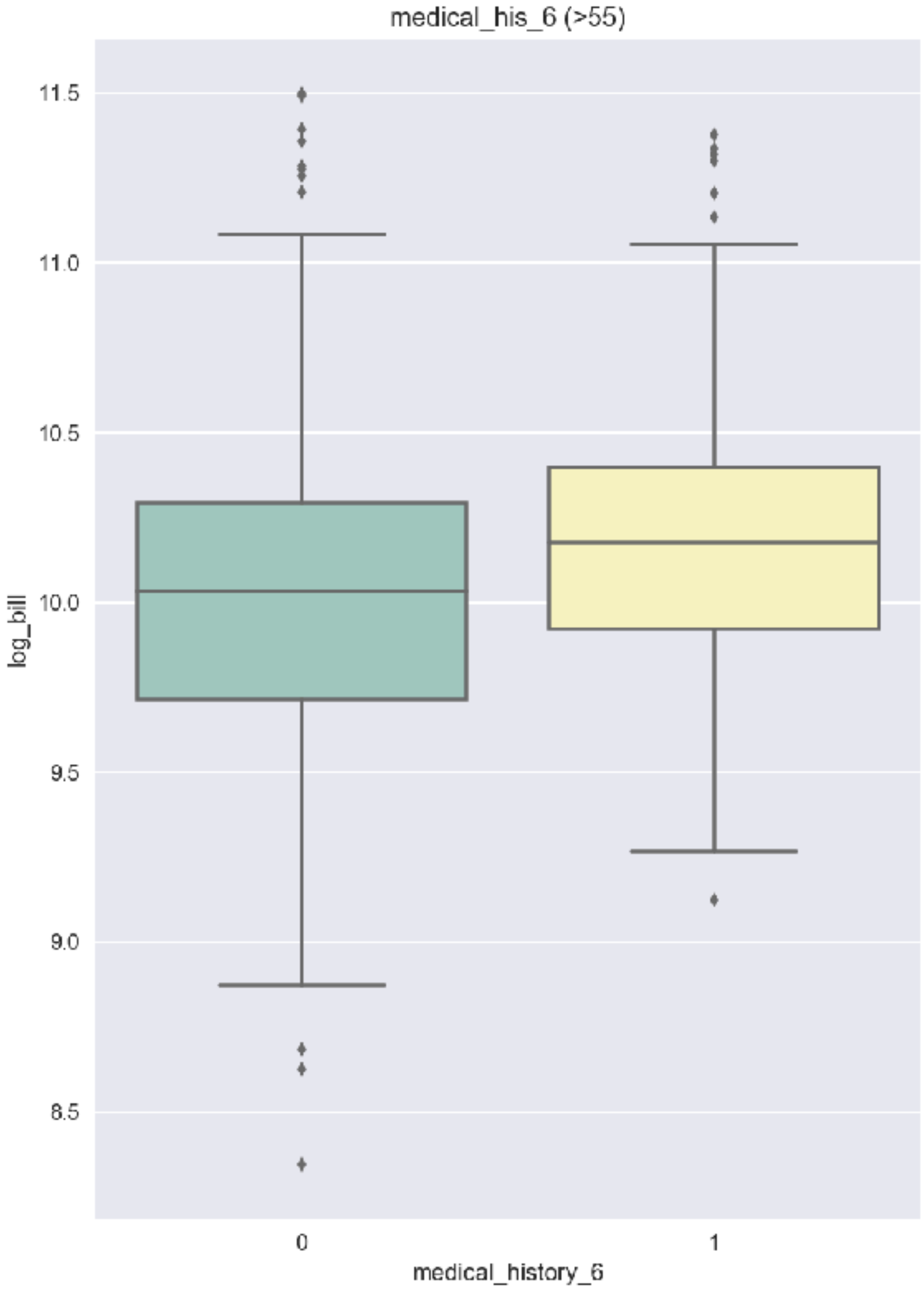


- 18.7% of patients with medical history 1 contributes 21.7% of the total bills.
- P-Value < 0.01 for both age groups

Medical History 6

Count: 839/ 3,000 (28.0%)

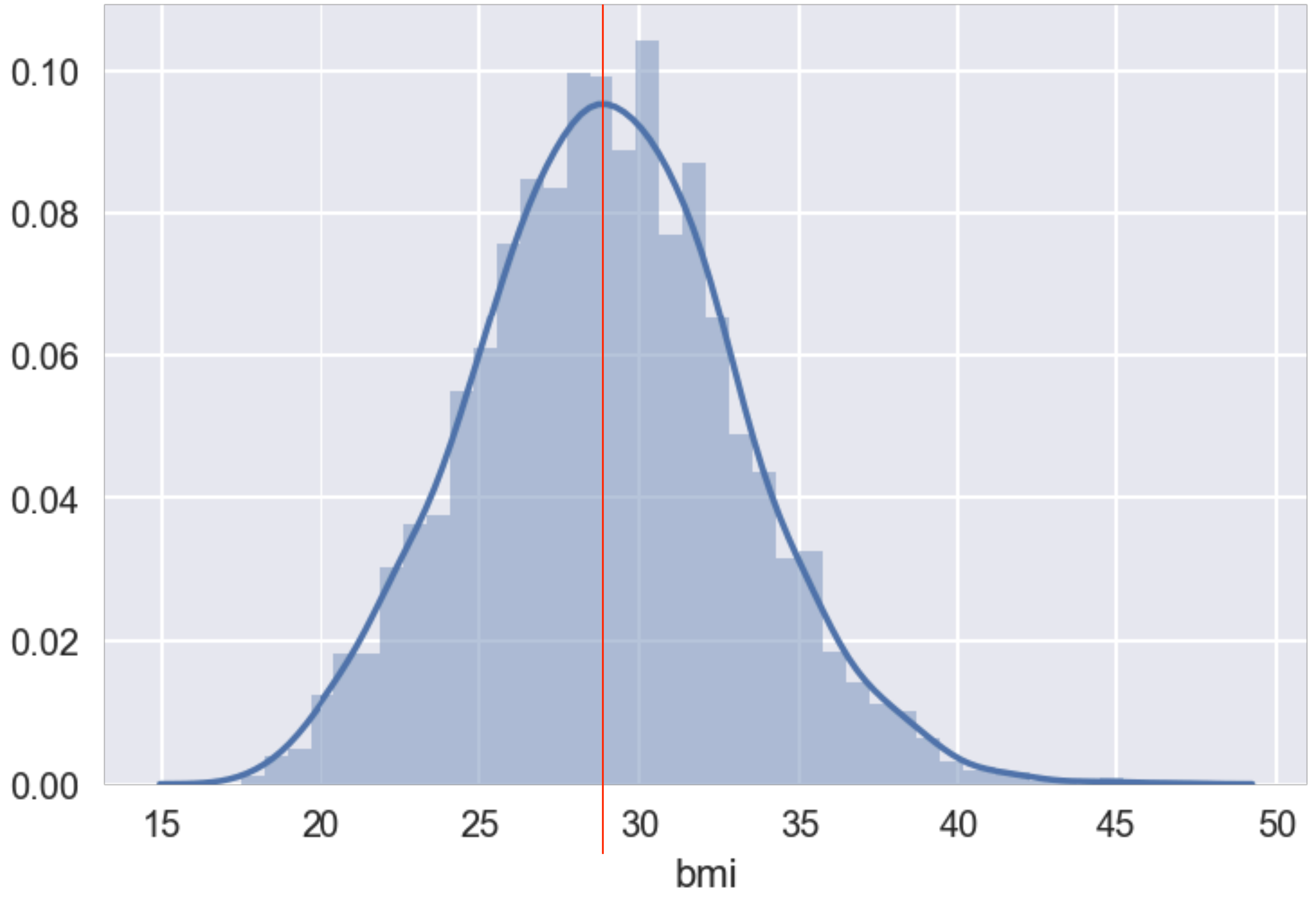


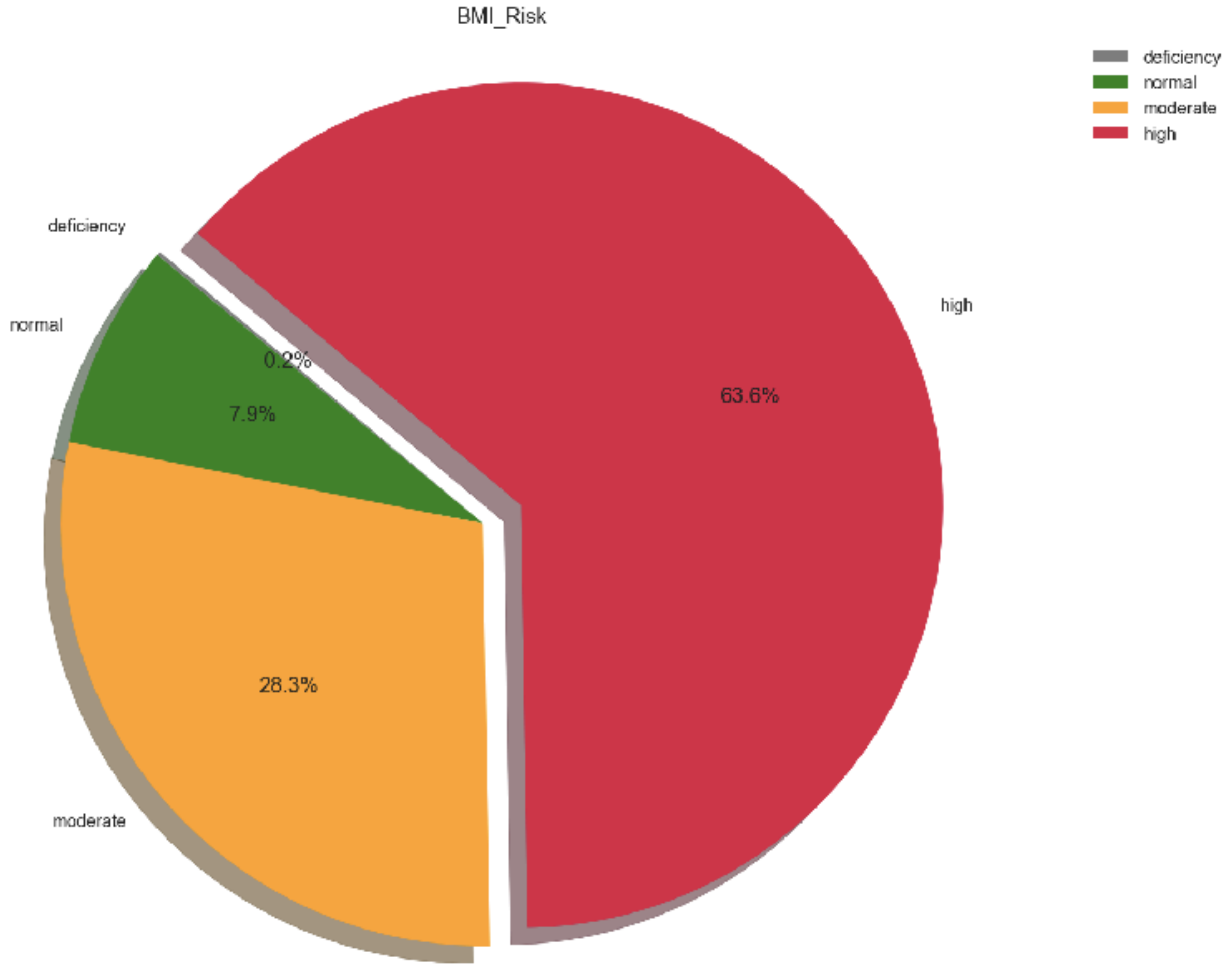


- P-Value < 0.01 for both age groups

Body Mass Index (BMI)

$$BMI = \frac{Weight\,(kg)}{[Height(m)]^2}$$



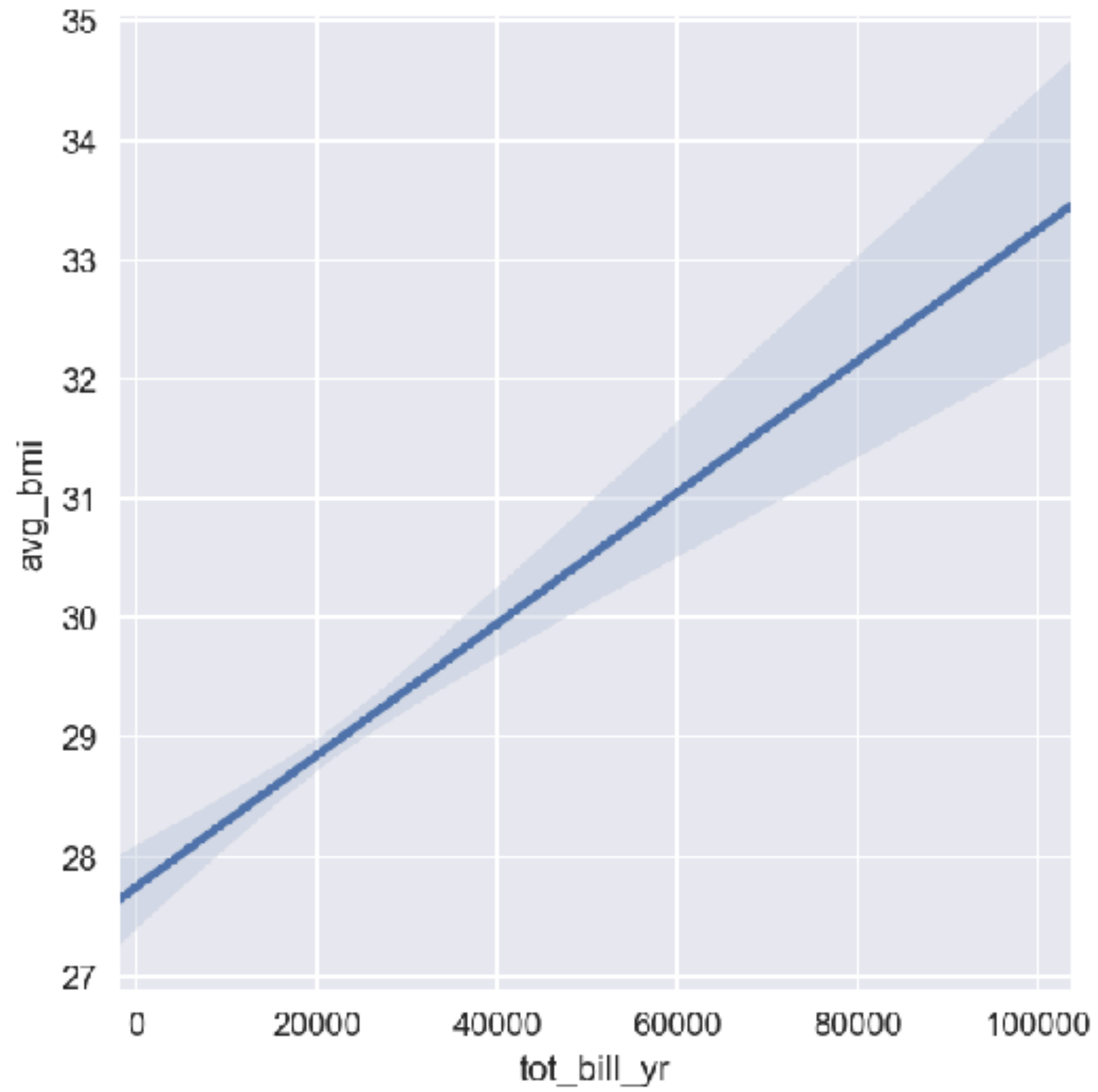


- Mean BMI = 28.95 (high risk)
- **63.6**% patient population have high BMI risk

Body Mass Index (BMI)

$$BMI = \frac{Weight\ (kg)}{[Height(m)]^2}$$

		Total Bill	
		Mean	**P-Value**
BMI Risk Level	Deficiency	11,200	**<0.01***
	Normal	19,389	
	Moderate	20,774	
	High	22,651	



R: 0.140, P-value: <0.01*

- There is significance between BMI Risk and total bill per year at 0.01 significance level

- Positive correlation between BMI and total bill

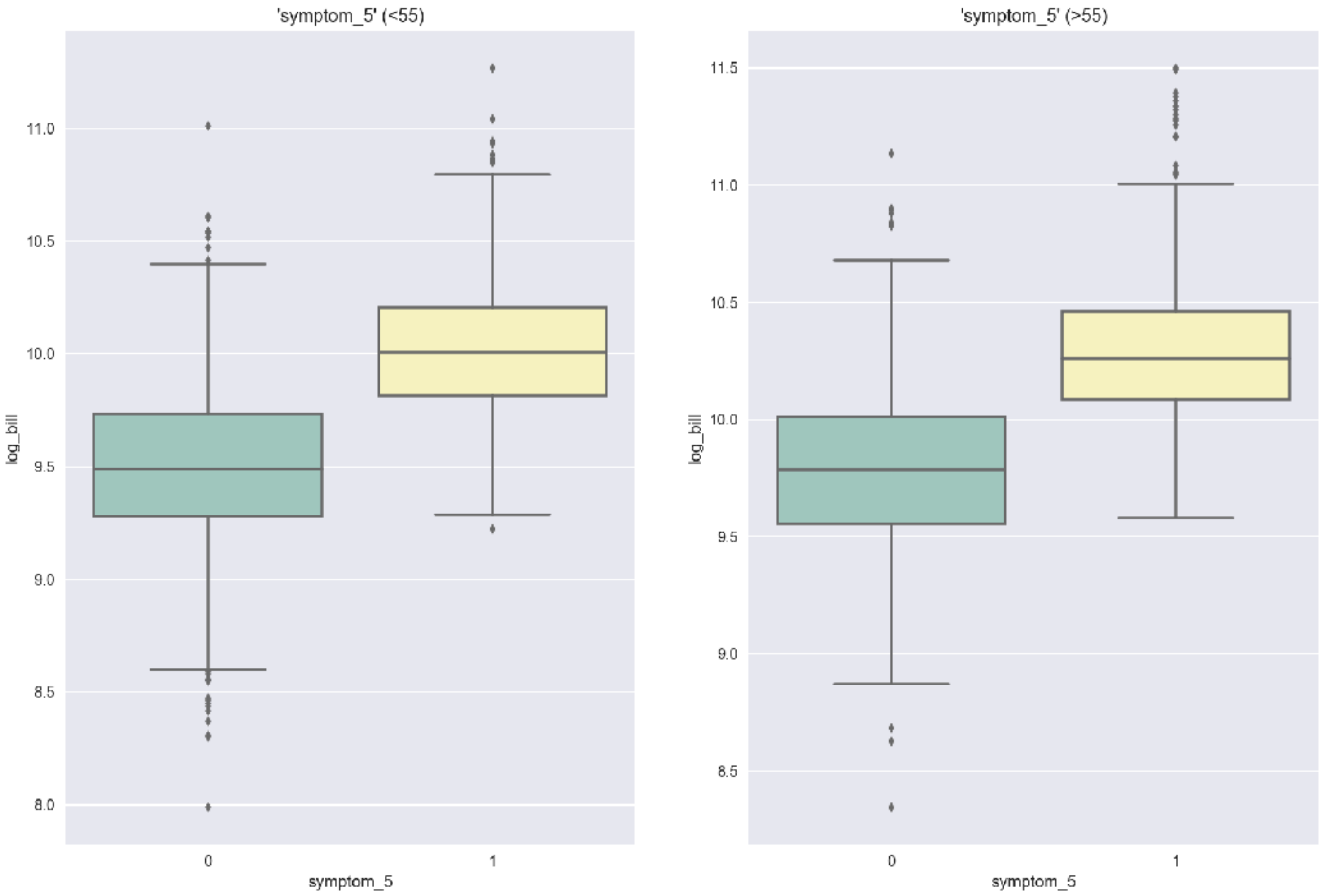
- Not tested for homogeneity and normality assumptions due to time constraint. One-way ANOVA is used

Symptoms

		Count	Total Bill	
			Mean	P-Value
Total		*3,314*	*21,859*	
Number of Symptoms	0	25	7477	**<0.01***
	1	209	12,063	
	2	748	16,637	
	3	1,196	21,296	
	4	931	26,421	
	5	291	31,267	
Symptom 1		2,107	22,879	**<0.01***
Symptom 2		2,252	23,003	**<0.01***
Symptom 3		1,852	23,567	**<0.01***
Symptom 4		2,470	22,670	**<0.01***
Symptom 5		1,791	26,832	**<0.01***

- Not tested for homogeneity and normality assumptions due to time constraint. One-way ANOVA is used

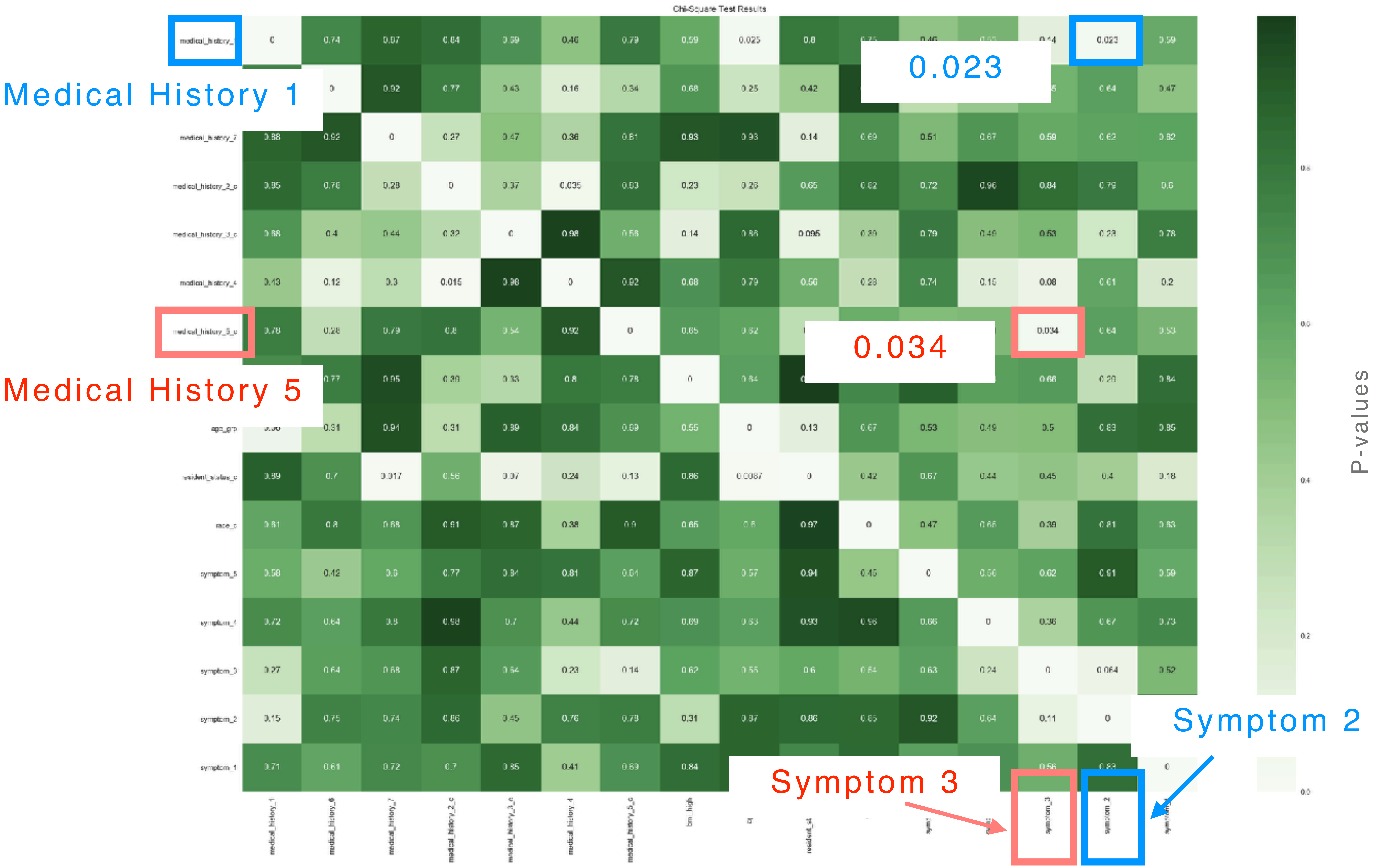
Symptom 5



- P-Value < 0.01 for both age groups

- Not tested for homogeneity and normality assumptions due to time constraint. One-way ANOVA is used

Symptoms



CHI Squared Test: Not significantly correlated to other variables

Model & Interpretation

Explainability matters .

Variance Inflation Factor

To check for multicollinearity between variables

	VIF Factor	features
0	1.0	medical_history_1
1	1.0	medical_history_6
2	2.0	bmi_high
3	2.0	age_grp
4	1.0	foreigner
5	1.0	pr
6	1.0	malay
7	1.0	indian
8	1.0	other
9	2.0	symptom_5
10	3.0	symptom_4
11	2.0	symptom_3
12	3.0	symptom_2
13	2.0	symptom_1

No variables have a VIF of more than 5

Linear Regression Model

Y = β0 + β1 medical history 1 + β2 medical history 6 + β3 high bmi + β4 age > 55 +
β5 foreigner + β6 pr + β7 malay + β8 indian, β9 other +
β10 symptom_5 + β11 symptom_4 + β12 symptom_3 + β13 symptom_2 + β14 symptom_1

R-squared = *0.931*

Parameter	Estimate	Standard Error	P-Value
Intercept	8.37	0.007	<0.01*
Medical History 1	0.28	0.005	<0.01*
Medical History 6	0.17	0.005	<0.01*
High BMI risk	0.11	0.004	<0.01*
Age > 55	0.25	0.004	<0.01*

Total bill for patients with condition X and have / is :

- medical history 1 is **~ 32 %** higher
- medical history 6 is **~ 18 %** higher
- high bmi risk is ~ **11.12 %** higher
- > 55 years old is ~ **28.24** % higher

compared to those without / isn't

Linear Regression Model

Y = β0 + β1 medical history 1 + β2 medical history 6 + β3 high bmi + β4 age > 55 +
β5 foreigner + β6 pr + β7 malay + β8 indian, β9 other +
β10 symptom_5 + β11 symptom_4 + β12 symptom_3 + β13 symptom_2 + β14 symptom_1

R-squared = *0.931*

Parameter	Estimate	Standard Error	P-Value
Foreigner	0.70	0.01	<0.01*
Permanent Resident (PR)	0.18	0.006	<0.01*
Malay	0.43	0.005	<0.01*
Indian	0.19	0.007	<0.01*
Other	0.09	0.009	<0.01*

Linear Regression Model

Y = β0 + β1 medical history 1 + β2 medical history 6 + β3 high bmi + β4 age > 55 +
β5 foreigner + β6 pr + β7 malay + β8 indian, β9 other +
β10 symptom_5 + β11 symptom_4 + β12 symptom_3 + β13 symptom_2 + β14 symptom_1

R-squared = *0.931*

Parameter	Estimate	Standard Error	P-Value
Symptom 5	0.51	0.004	<0.01*
Symptom 4	0.18	0.005	<0.01*
Symptom 3	0.20	0.004	<0.01*
Symptom 2	0.19	0.004	<0.01*
Symptom 1	0.14	0.004	<0.01*

Total bill for patients with condition X and presented with :

- Symptom 5 is **~ 67 %** higher
- Symptom 4 is **~ 20 %** higher
- Symptom 3 is ~ **22 %** higher
- Symptom 2 is ~ **21** % higher
- Symptom 1 is ~ **15**% higher

compared to those who didn’t

Use Cases

Insights are only valuable if they are used .

Use Case Discussion

Findings suggest that admission cost per year are significantly affected by

1. Number of admitted encounters,
2. Medical history,
3. Body Mass Index
4. Age
5. Demographics
6. Treatment/Care of Symptom 5

Use Case Discussion

Findings suggest that admission cost per year are significantly affected by

1. Number of admitted encounters
2. Medical history
3. Body Mass Index
4. Age
5. Demographics
6. Treatment/Care of Symptom 5

- Cost per year increases with frequent admissions
- Patients with existing co-morbidities have tend to have readmissions
- Consider continuity of care for these patients after discharge to prevent readmission

Use Case Discussion

Findings suggest that admission cost per year are significantly affected by

1. Number of admitted encounters
2. Medical history
3. Body Mass Index
4. Age
5. Demographics
6. Treatment/Care of Symptom 5

- Increases complexity, more complications, more need for healthcare attention
- Preventive care for patients without existing conditions through regular screenings
- Medical History 1 and 6 have bigger impact than others
- Not enough information

Use Case Discussion

Findings suggest that admission cost per year are significantly affected by

1. Number of admitted encounters
2. Medical history
3. Body Mass Index
4. Age
5. Demographics
6. Treatment/Care of Symptom 5

- Not surprising that patients with high health risk to have more complications and require more interventions
- Better manage patient's BMI by encouraging them to participate in health programmes
- Frequent health screenings

Use Case Discussion

Findings suggest that admission cost per year are significantly affected by

1. Number of admitted encounters
2. Medical history
3. Body Mass Index
4. Age
5. Demographics
6. Treatment/Care of Symptom 5

- Expected that elderly patients would need more medical attention than younger population
- Non-Singaporeans have less subsidies
- Non-modifiable variables, but allows healthcare providers/ministry of health to target population for policy planning and interventions

Use Case Discussion

Findings suggest that admission cost per year are significantly affected by

1. Number of admitted encounters
2. Medical history
3. Body Mass Index
4. Age
5. Demographics
6. Treatment/Care of Symptom 5

- Relook into the cost of treatment/care
- Opportunity in cheaper alternative treatments
- Epidemiology study into symptom 5

Insights output format

- **Report with insights**
  - How each variable impacts
  - To what extent does each variable impact population
  - Which population will have higher risk
  - Comparison between hospitals

- **Prediction Model**
  - Prevalence of symptoms will not be relevant as a variable
  - Use specific age and bmi may improve accuracy of the model

Extensions…

- Deep dive into why race is a key factor in cost (due to diet?)
- Find out if subsidies is a factor to cost differences in resident status
- More financial data such as ward class, bill type will be useful
- Given the context of the de-identified variables may have more clinically relevant insights
- Look into clinical data for more use cases

- Test for homogeneity and normality assumptions
- Improve on visualisation and story-telling