	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade	p
0	1	8	c	3	Z	3	radiotherapy	R	F	2.0	
1	2	2	c	5	Z	2	radiotherapy	S	F	2.0	
2	3	10	e	1	X	2	anesthesia	S	E	2.0	
3	4	26	b	2	Y	2	radiotherapy	R	D	2.0	
4	5	26	b	2	Y	2	radiotherapy	S	D	2.0	

Preliminary Exploration

*The first set of test variables are consistent across regions*further analysis could be done for performance across different regions on similar patient types

Test variables:

- Age
- Type of Admission
- severity of Illness
- Admission_Deposit
- Bed Grade
- Hospital type code

case_id	bed_grade	patientid	admission_type	severity	visitors	age	deposit	stay
1	2.0	31397	Emergency	Extreme	2	51-60	4911.0	0-10
2	2.0	31397	Trauma	Extreme	2	51-60	5954.0	41-50

case_id	bed_grade	patientid	admission_type	severity	visitors	age	deposit	stay
3	2.0	31397	Trauma	Extreme	2	51-60	4745.0	31-40

```
table info...

<class 'pandas.core.frame.DataFrame'>
Int64Index: 318438 entries, 1 to 318438
Data columns (total 8 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   bed_grade       318325 non-null  float64
 1   patientid       318438 non-null  int64
 2   admission_type  318438 non-null  object
 3   severity        318438 non-null  object
 4   visitors        318438 non-null  int64
 5   age             318438 non-null  object
 6   deposit         318438 non-null  float64
 7   stay            318438 non-null  object
dtypes: float64(2), int64(2), object(4)
memory usage: 21.9+ MB
None

# of unique patientids...
92017

41-50     63749
31-40     63639
51-60     48514
21-30     40843
71-80     35792
61-70     33687
11-20     16768
81-90      7890
0-10       6254
91-100     1302
Name: age, dtype: int64
```

- **The object data type for age and stay need to be converted to a numeric to represent each group.**

case_id	bed_grade	patientid	admission_type	severity	visitors	age	deposit	stay

case_id	bed_grade	patientid	admission_type	severity	visitors	age	deposit	stay
1	2.0	31397	Emergency	Extreme	2	55	4911.0	5.0
2	2.0	31397	Trauma	Extreme	2	55	5954.0	45.0
3	2.0	31397	Trauma	Extreme	2	55	4745.0	35.0

```
Null value counts...
False    311755
True       6683
Name: stay, dtype: int64
```

case_id	bed_grade	patientid	admission_type	severity	visitors	age	deposit	stay
1	2.0	31397	Emergency	Extreme	2	55	4911.0	5.0
2	2.0	31397	Trauma	Extreme	2	55	5954.0	45.0
3	2.0	31397	Trauma	Extreme	2	55	4745.0	35.0
4	2.0	31397	Trauma	Extreme	2	55	7272.0	45.0
5	2.0	31397	Trauma	Extreme	2	55	5558.0	45.0

```
Null value counts after fill...
False    311644
Name: stay, dtype: int64

Unique severity values...
['Extreme' 'Moderate' 'Minor']

Unique admission types...
['Emergency' 'Trauma' 'Urgent']

Age groups...
[55 75 35 45 85 65 25 15  5 95]
```

	stay	
	median	mean	count

age	severity	admission_type			
		Emergency	25.000000	29.848485	198
	Extreme	Trauma	25.000000	34.774920	311
		Urgent	25.000000	32.500000	104
		Emergency	15.000000	22.459716	1055
5	Minor	Trauma	25.000000	26.327543	806
		Urgent	15.000000	23.767123	657
		Emergency	25.000000	27.593607	1095
	Moderate	Trauma	25.000000	31.970213	1175
		Urgent	25.000000	28.014805	743
		Emergency	25.000000	28.721973	446
	Extreme	Trauma	25.000000	33.804781	502
		Urgent	25.000000	30.817175	361
		Emergency	15.000000	23.529231	3250
15	Minor	Trauma	25.000000	27.502670	2809
		Urgent	25.000000	25.114643	1919
		Emergency	25.000000	26.356275	2470
	Moderate	Trauma	25.000000	30.243986	2910
		Urgent	25.000000	27.966102	1888
		Emergency	25.000000	30.803443	1394
	Extreme	Trauma	25.000000	34.199567	1849
		Urgent	25.000000	30.250278	899
25		Emergency	25.000000	25.021952	5922
	Minor	Trauma	25.000000	28.688867	6216
		Urgent	25.000000	25.499195	3105

		Emergency	25.000000	26.842796	7239
	Moderate	Trauma	25.000000	31.142997	9217
		Urgent	25.000000	28.355825	4446
		Emergency	25.000000	30.264680	3457
	Extreme	Trauma	25.000000	34.721519	4740
		Urgent	25.000000	31.717774	1761
		Emergency	25.000000	25.768664	6791
35	Minor	Trauma	25.000000	28.982186	6961
		Urgent	25.000000	25.814961	2540
		Emergency	25.000000	27.779865	12774
	Moderate	Trauma	25.000000	31.860559	17312
		Urgent	25.000000	28.652961	6299
		Emergency	25.000000	31.224745	4405
	Extreme	Trauma	25.000000	35.596949	6751
		Urgent	25.000000	32.961853	1835
		Emergency	25.000000	25.740140	5553
45	Minor	Trauma	25.000000	29.146829	5503
		Urgent	25.000000	25.932401	1716
		Emergency	25.000000	28.495138	13164
	Moderate	Trauma	25.000000	32.831416	18519
		Urgent	25.000000	29.308157	4965
		Emergency	25.000000	32.477350	3532
	Extreme	Trauma	25.000000	35.454630	5345
55		Urgent	25.000000	32.259696	1186
	Minor	Emergency	25.000000	26.654292	4310

		Trauma	25.000000	30.115526	4631
		Urgent	25.000000	25.598761	1453
		Emergency	25.000000	29.421776	9460
	Moderate	Trauma	25.000000	33.171514	14028
		Urgent	25.000000	29.959738	3353
		Emergency	25.000000	31.967796	2391
	Extreme	Trauma	35.000000	36.265350	3746
		Urgent	25.000000	33.800989	809
		Emergency	25.000000	26.224554	3193
65	Minor	Trauma	25.000000	29.395265	3886
		Urgent	25.000000	26.156863	1020
		Emergency	25.000000	30.414201	6084
	Moderate	Trauma	25.000000	33.598993	9536
		Urgent	25.000000	31.478608	2127
		Emergency	25.000000	33.569045	2397
	Extreme	Trauma	35.000000	36.542568	3559
		Urgent	25.000000	35.081522	736
		Emergency	25.000000	28.097740	3806
75	Minor	Trauma	25.000000	30.483733	4672
		Urgent	25.000000	27.452652	1056
		Emergency	25.000000	31.460967	6725
	Moderate	Trauma	25.000000	33.966304	9645
		Urgent	25.000000	31.086548	2126
		Emergency	35.000000	37.950392	766
85	Extreme	Trauma	35.000000	38.897059	816

		Urgent	35.000000	36.866667	150
		Emergency	25.000000	31.503885	901
	Minor	**Trauma**	25.000000	32.692308	780
		Urgent	25.000000	31.779661	118
		Emergency	35.000000	35.529070	1720
	Moderate	**Trauma**	35.000000	36.455664	1951
		Urgent	25.000000	34.275766	359
		Emergency	35.000000	37.913386	127
	Extreme	**Trauma**	35.000000	38.623188	138
		Urgent	35.000000	41.956522	23
		Emergency	25.000000	29.751773	141
95	**Minor**	**Trauma**	25.000000	31.534653	101
		Urgent	35.000000	33.571429	21
		Emergency	35.000000	37.211221	303

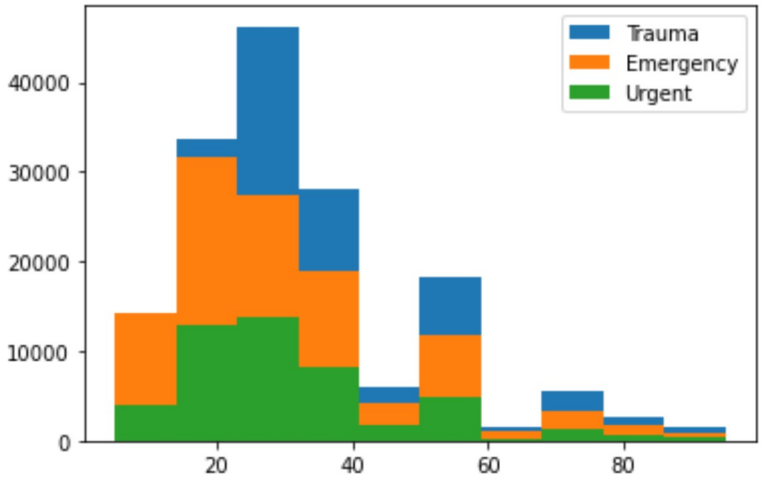
Preliminary Analysis:

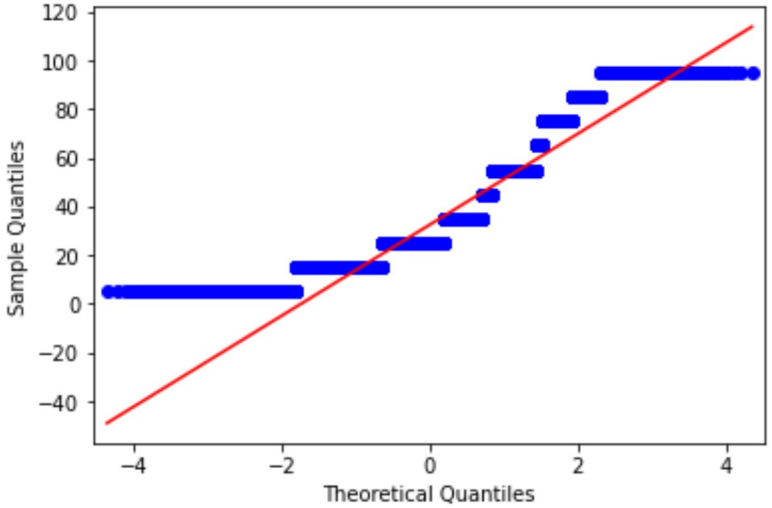
- trauma in each group tends to have the highest average stay regardless of admission type or age
- youngest age group with extreme severity of any admission type has a higher average stay
- starting at ages 55-65+ anything moderate or extreme results on average in a higher length of stay
- within each age group, severity looks to play a significantly greater role in determining a higher avg stay length
- admission type looks to have a small relationship to higher stay lengths
- ages 35-55 seem to show the highest incidence of visit counts, mainly of moderate/emergency trauma & the same pattern looks to trail off at age groups to either side

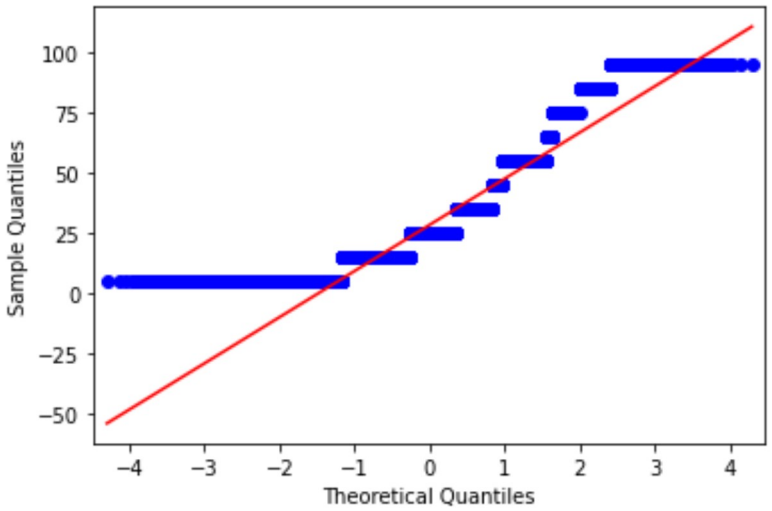
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 311644 entries, 1 to 318438
Data columns (total 8 columns):
 #   Column          Non-Null Count   Dtype  
---  ------          --------------   -----  
 0   bed_grade       311644 non-null  float64
 1   patientid       311644 non-null  int64  
```

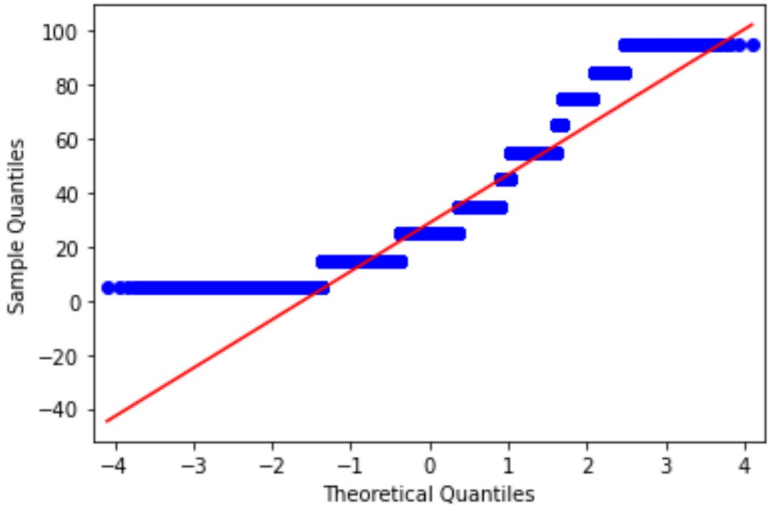
```
 2   admission_type  311644 non-null  object
 3   severity        311644 non-null  object
 4   visitors        311644 non-null  int64
 5   age             311644 non-null  int64
 6   deposit         311644 non-null  float64
 7   stay            311644 non-null  float64
dtypes: float64(3), int64(3), object(2)
```

Admission Type









Jarque-Bera Test

- probability that kurtosis and skewness matches that of a noraml distribution

H_o = both match that of a normal distribution

```
trauma
0.0
emergency
0.0
urgent
```

```
0.0
```

Analysis: null hypothesis is rejected; data is not normally distributed

Kruskal-Wallis Test

H_o = none of the group medians are significantly different

```
True
```

Analysis: null hypothesis that the medians are the same is rejected

Kruskal pair testing

```
[ True  True  True]
[0.00000000e+00 0.00000000e+00 3.78133696e-22]
0.016666666666666666
```

Analysis: After the correction, the null hypotheses that the medians are all the same can be rejected

Post hoc effect size analysis

Confidence interval for 2 group median test

Trauma vs. Urgent confidence interval

```
100%                                               5000/5000 [01:48<00:00, 46.20it/s]
Trauma vs. Urgent
Bootstrapped confidence interval: 0.0, 0.0
```

Analysis: The confidence interval between median difference of trauma and urgent is 0. There is essentially 0 effect size.

Trauma vs. Emergency confidence interval

```
100%                                               5000/5000 [01:10<00:00, 70.56it/s]
Trauma vs. Emergency
Bootstrapped confidence interval: 0.0, 0.0
```

Analysis: The confidence interval between median difference of trauma and urgent is 0. There is essentially 0 effect size.

Common language effect size test

CLES test: Trauma vs. Urgent

```
0.45872462896879657
```

Analysis: When admitting a patient as **trauma**, 46% of the time we would expect a longer stay than if they were admitted as **urgent**.

CLES Test: Trauma vs. Emergency

```
0.4805367036736003
```

Analysis: When admitting a patient as **trauma**, 48% of the time we would expect a longer stay than if they were admitted as **emergency**.

CLES Test: Emergency vs. Urgent

```
0.4188720317878009
```

Analysis: When admitting a patient as **emergency**, 42% of the time we would expect a longer stay than if they were admitted as **urgent**

Conclusion: Even though we were able to reject the null hypothesis that the group medians are the same, but the confidence interval for the size of the effect is 0.

Age vs. LOS

Minor severity

	stay	
	mean	count
age		
5	24.038920	2518
15	25.309601	7978
25	26.614512	15243
35	27.148907	16292
45	27.233793	12772
55	28.048874	10394
65	27.737375	8099

	stay	
	mean	count

Moderate severity

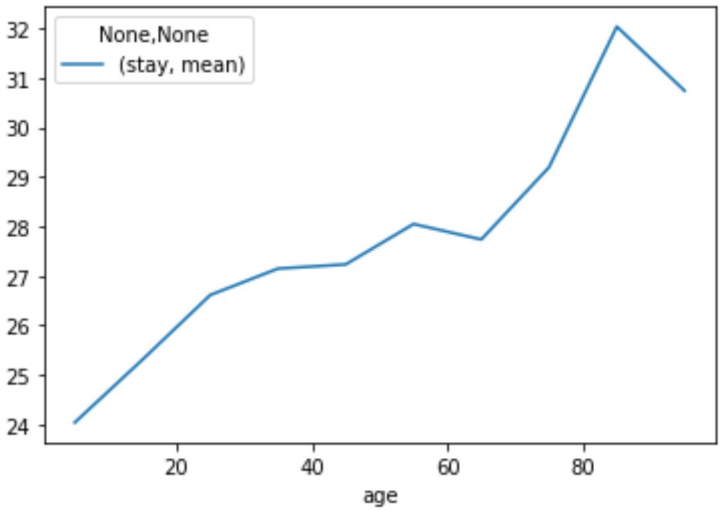
	stay	
	mean	count
age		
5	29.404248	3013
15	28.331040	7268
25	29.060855	20902
35	29.872612	36385
45	30.796496	36648
55	31.448717	26841
65	32.253057	17747
75	32.724373	18496
85	35.866005	4030
95	34.694767	688

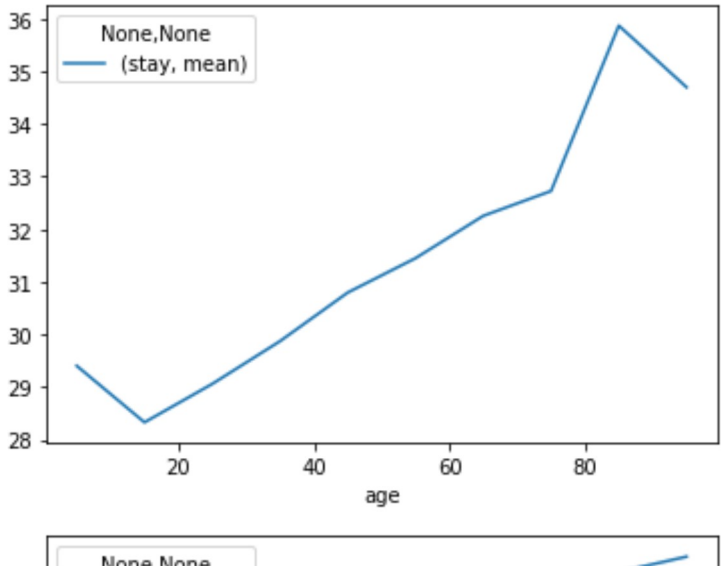
Extreme severity

	stay	
	mean	count
age		
5	32.797716	613
15	31.249045	1309
25	32.199421	4142

	stay	
	mean	count
age		
35	32.643101	9958
45	33.742206	12991
55	34.033092	10063

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c502ffb1c0>
```





Analysis:

- Between the aggregate mean table and the line plot, there seems to be a definite correlation between age and length of stay.
- there does seem to be more of a correlation in the minor and moderate groups than in the extreme group

Correlation coefficients

```
Minor
(0.07056266314133419, 3.766795192972117e-94)
Moderate
(0.075292102384955, 1.1354832776736747e-214)
Extreme
(0.05987691272507544, 1.165543385469099e-44)
```

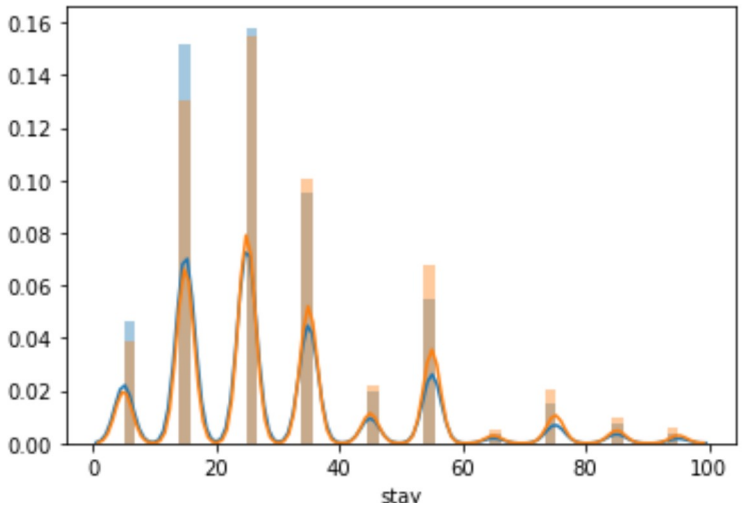
Bottom 20th percentile for age

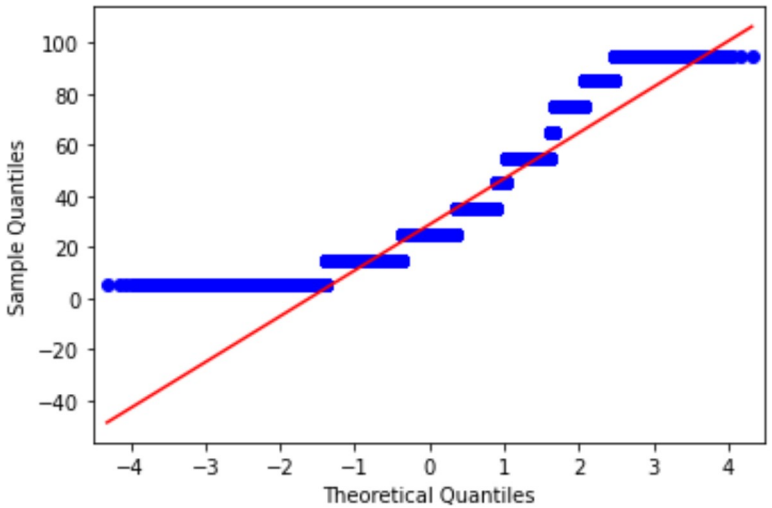
Moderate

```
35.0
```

Normality checks

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c503c0eee0>
```


0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
stay


100
80
60
40
20
0
−20
−40
Sample Quantiles
−4
−3
−2
−1
0
1
2
3
4
Theoretical Quantiles

Jarque-Bera Test

H_null = the kurtosis and skewness match that of a normal distribution

```
reject null: True
reject null: True
```

- both groups do not match the kurtosis and skewness of a normal distribution

Kruskal-Wallis Test

H_null = the medians of both groups are the same

```
reject null: True
```

- the medians of both groups are significantly different

100% 5000/5000 [00:56<00:00, 88.53it/s]

```
Bottom 20th percentile confidence interval: 0.0, 0.0
```

Analysis:

- moderate severity group
- The median confidence interval for the bottom 20th age percentile is 0.0

Common language effect size

```
0.443294103974258
```

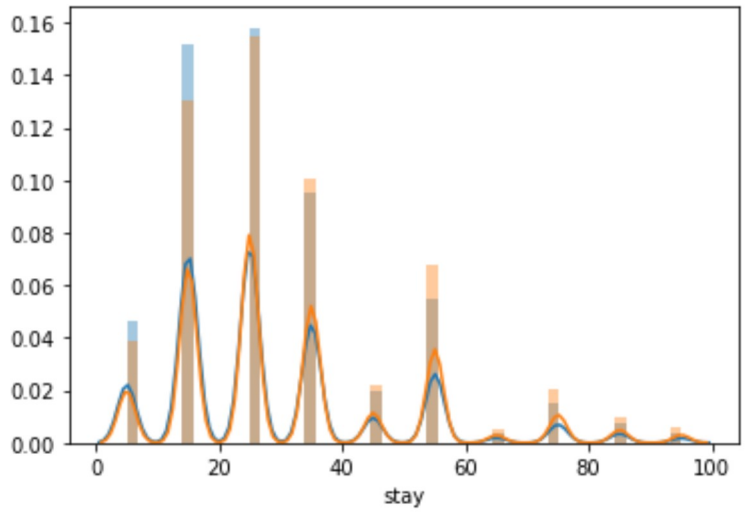
Analysis: 44% of the time when a patient in the bottom 20th age percentile is admitted with a moderate condition they will have a shorter length of stay

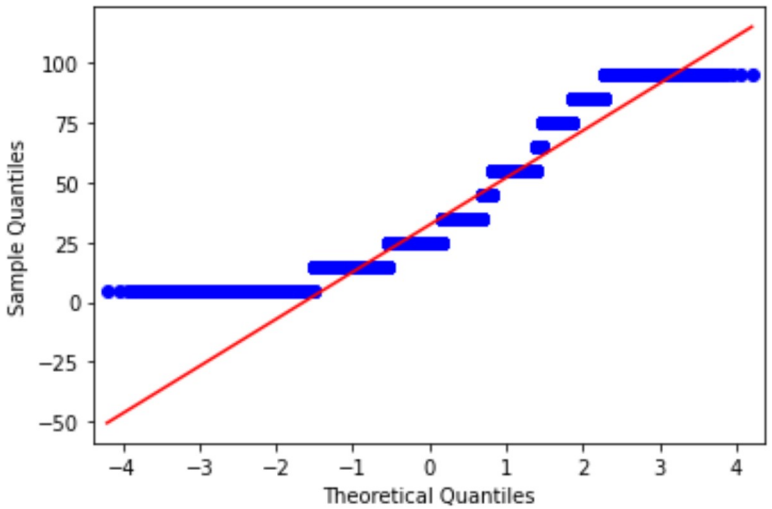
Top 20th percentile for age

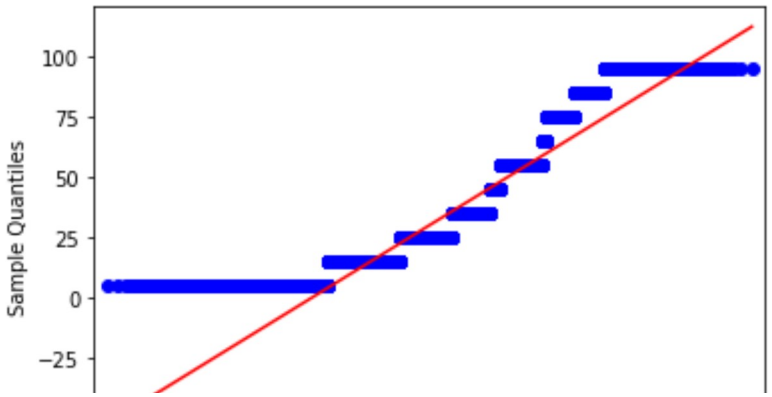
```
65.0
```

Normality Checks

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c5074ee070>
```







Jarque-Bera Test

H_null = the distributions match that of a normal distribution

```
reject null: True
reject null: True
```

Kruskal-Wallis Test

H_null = the medians of both groups are the same

```
reject null: True
```

- the medians for both groups are statistically different

100% 5000/5000 [00:56<00:00, 88.88it/s]

```
Top 20th percentile confidence interval: 0.0, 0.0
```

Analysis:

- moderate severity group
- The median confidence interval for the top 20th age percentile is 0.0

Common language effect size

```
0.43957878069541634
```

Analysis: 44% of the time when a patient in the top 20th percentile for age is admitted with a moderate condition, they will have a longer length of stay.

Bed grade vs. LOS

	stay	
	mean	count
bed_grade		
1.0	23.095488	1906
2.0	24.229009	18353
3.0	28.072340	43475
4.0	28.801872	21158

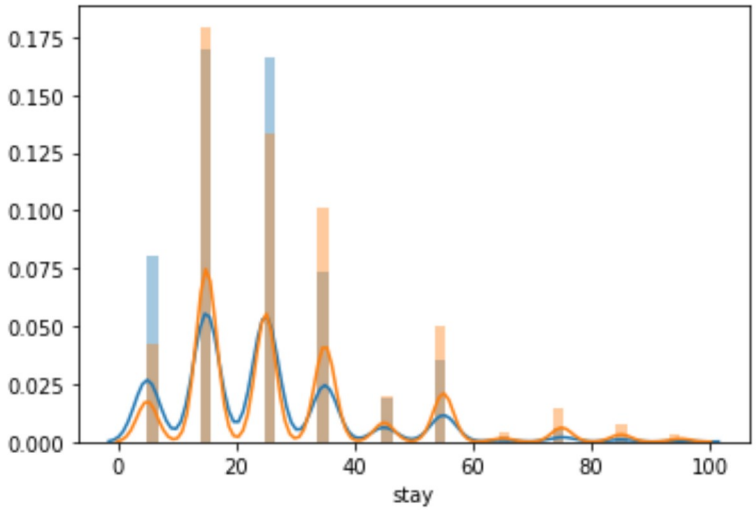
- it looks like in minor cases there may be some relationship between stay length and bed grade

	stay	
	mean	count
bed_grade		
1.0	27.046941	10396
2.0	30.684922	82659
3.0	31.448417	51577
4.0	31.698313	27386

	stay	
	mean	count
bed_grade		
1.0	35.311768	13664
2.0	33.470483	21242
3.0	33.264912	12691

stay

Minor



- just like the other groups these don't appear to be normally distributed

Kruskal-Wallis Test

h_null = the medians for both groups are the same

```
reject null: True
```

Median confidence interval

100% 10000/10000 [00:17<00:00, 556.36it/s]

```
Top 20th percentile confidence interval: 0.0, 0.0
```

Analysis:

- groups for bed grades 1-2 and 3-4
- The median confidence interval for the difference in stay length is 0.0

Common language effect size

```
0.45402034982669814
```

Analysis: 46% of the time when a person is admitted to room with a bed grade of 1-2 for a minor condition, they will experience a shorter length of stay.