Raquel Ocasio

D207 – Exploratory Data Analysis

September 24, 2023

Western Governors University

Part A1

Is there a significant association between the martial status of a patient, and the means by which they are initially admitted to the hospital?

Part A2

An analysis of the data can benefit the stakeholders by providing insight into whether a patient's marital status is affecting how the patient is admitted to the hospital.

Part A3

The "Marital" and "Initial_admin" variables are relevant for answering the question.

Part B1

See attached code.

Part B2

```
Contingency Table
Initial_admin  Elective Admission  Emergency Admission  Observation Admission
Marital
Divorced                     492                  986                    483
Married                      499                 1051                    473
Never Married                528                  973                    483
Separated                    499                  998                    490
Widowed                      486                 1052                    507

The p-value is 0.5852742698130522
There is no significant association between the variables.

Expected frequencies:
 [[ 491.0344  992.266   477.6996]
 [ 506.5592 1023.638   492.8028]
 [ 496.7936 1003.904   483.3024]
 [ 497.5448 1005.422   484.0332]
 [ 512.068  1034.77    498.162 ]]
```

(Zach, 2021)

Part B3

The Chi-Squared test was chosen for this analysis because it can help determine if there is a significant relationship between categorical variables. The variables that need to be analyzed for the question are categorical.

Part C

Continuous Variables (Zach, 2021b)

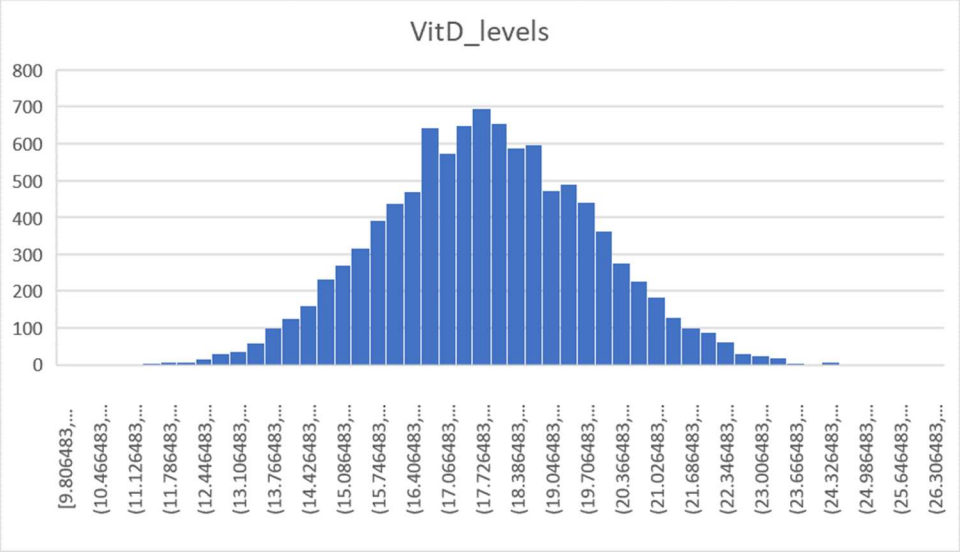
	VitD_levels	**Doc_visits**
Mean	17.96426165	5.0122
Median	17.95112226	5
Mode	18.13543091	5
Standard Deviation	2.017231042	1.045734439
Interquartile Range	2.721523888	2
Range	16.58796571	8

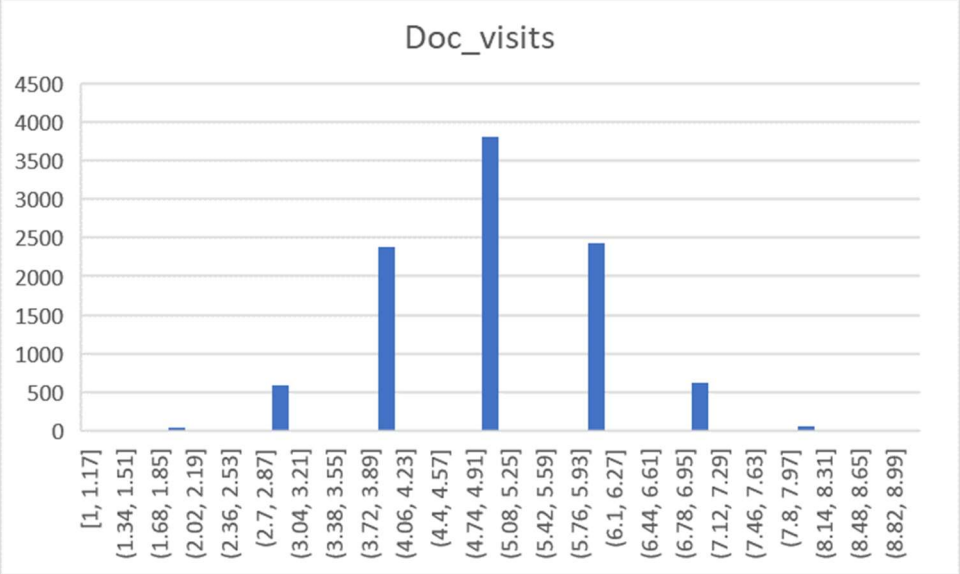
Categorical Variables (Paula's Web3 & Crypto, 2019)

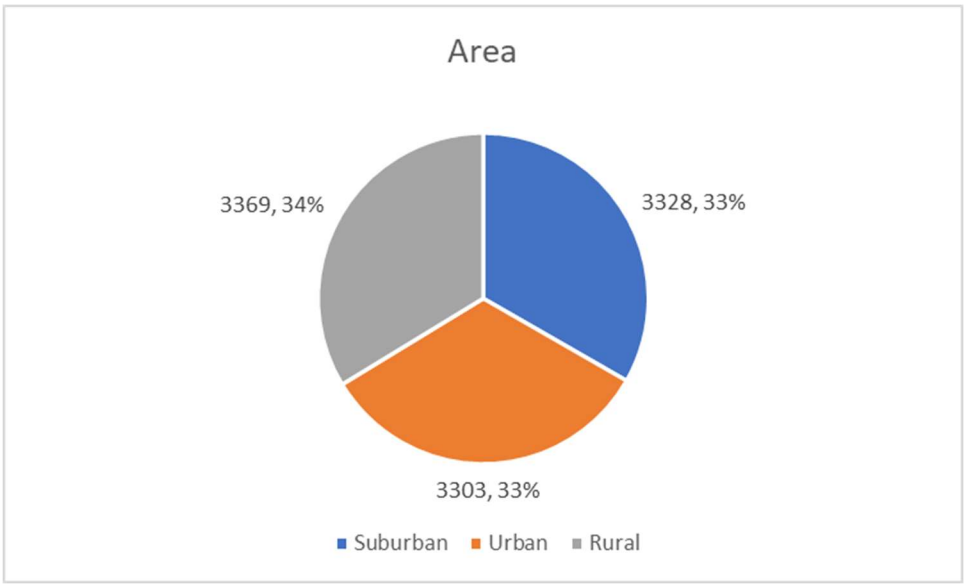
Area	**Count**	**Percentage**
Suburban	3328	33.3%
Urban	3303	33.0%
Rural	3369	33.7%

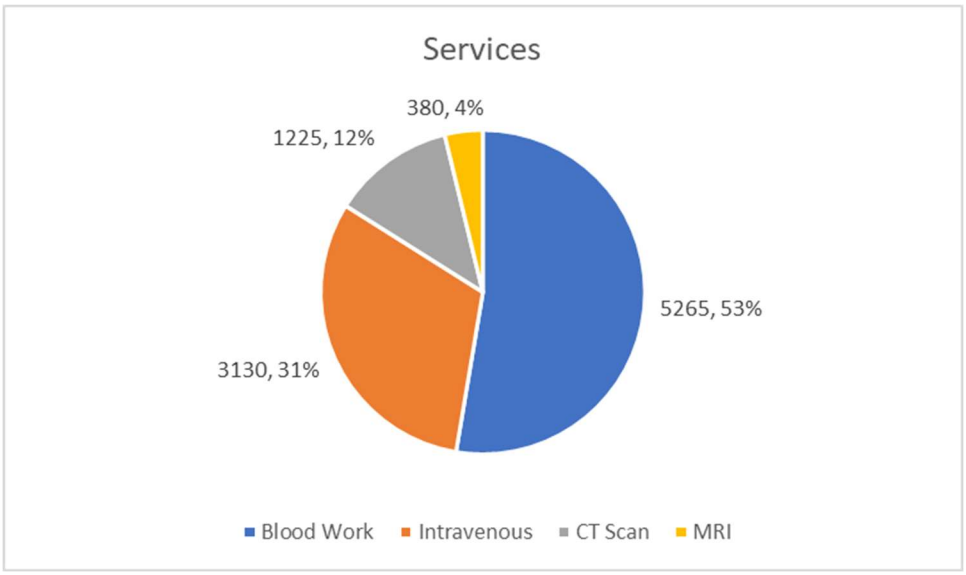
Services	**Count**	**Percentage**
Blood Work	5265	52.7%
Intravenous	3130	31.3%
CT Scan	1225	12.3%
MRI	380	3.8%

Part C1


VitD_levels
800
700
600
500
400
300
200
100
0
[9.806483,...
(10.466483,...
(11.126483,...
(11.786483,...
(12.446483,...
(13.106483,...
(13.766483,...
(14.426483,...
(15.086483,...
(15.746483,...
(16.406483,...
(17.066483,...
(17.726483,...
(18.386483,...
(19.046483,...
(19.706483,...
(20.366483,...
(21.026483,...
(21.686483,...
(22.346483,...
(23.006483,...
(23.666483,...
(24.326483,...
(24.986483,...
(25.646483,...
(26.306483,...


Doc_visits
4500
4000
3500
3000
2500
2000
1500
1000
500
0
[1, 1.17]
(1.34, 1.51]
(1.68, 1.85]
(2.02, 2.19]
(2.36, 2.53]
(2.7, 2.87]
(3.04, 3.21]
(3.38, 3.55]
(3.72, 3.89]
(4.06, 4.23]
(4.4, 4.57]
(4.74, 4.91]
(5.08, 5.25]
(5.42, 5.59]
(5.76, 5.93]
(6.1, 6.27]
(6.44, 6.61]
(6.78, 6.95]
(7.12, 7.29]
(7.46, 7.63]
(7.8, 7.97]
(8.14, 8.31]
(8.48, 8.65]
(8.82, 8.99]


Area
3369, 34%
3328, 33%
3303, 33%
Suburban
Urban
Rural


Services
380, 4%
1225, 12%
5265, 53%
3130, 31%
Blood Work
Intravenous
CT Scan
MRI

(Zach, 2021b)

Part D

Bivariate statistics for continuous variables (Zach, 2021c)

VitD_levels x Doc_visits

Correlation Coefficient	0.010210475							
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.010155965							
R Square	0.000103144							
Adjusted R Square	3.12394E-06							
Standard Error	1.045738435							
Observations	9999							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1.127724588	1.127725	1.031233	0.309893764			
Residual	9997	10932.40803	1.093569					
Total	9998	10933.53575						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.917525084	0.093718208	52.47139	0	4.73381853	5.101231638	4.73381853	5.101231638
19.1414657	0.005264717	0.005184377	1.015497	0.309894	-0.004897705	0.015427138	-0.004897705	0.015427138

Population x Age

Correlation Coefficient	-0.018986639							
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.018988025							
R Square	0.000360545							
Adjusted R Square	0.000260551							
Standard Error	20.63688064							
Observations	9999							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1535.585535	1535.586	3.605669	0.057611645			
Residual	9997	4257530.784	425.8808					
Total	9998	4259066.369						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	53.77519916	0.248678814	216.2436	0	53.28773863	54.2626597	53.28773863	54.2626597
2951	-2.64348E-05	1.39214E-05	-1.89886	0.057612	-5.37235E-05	8.53957E-07	-5.37235E-05	8.53957E-07

Bivariate statistics for categorical variables (Paula's Web3 & Crypto, 2019)

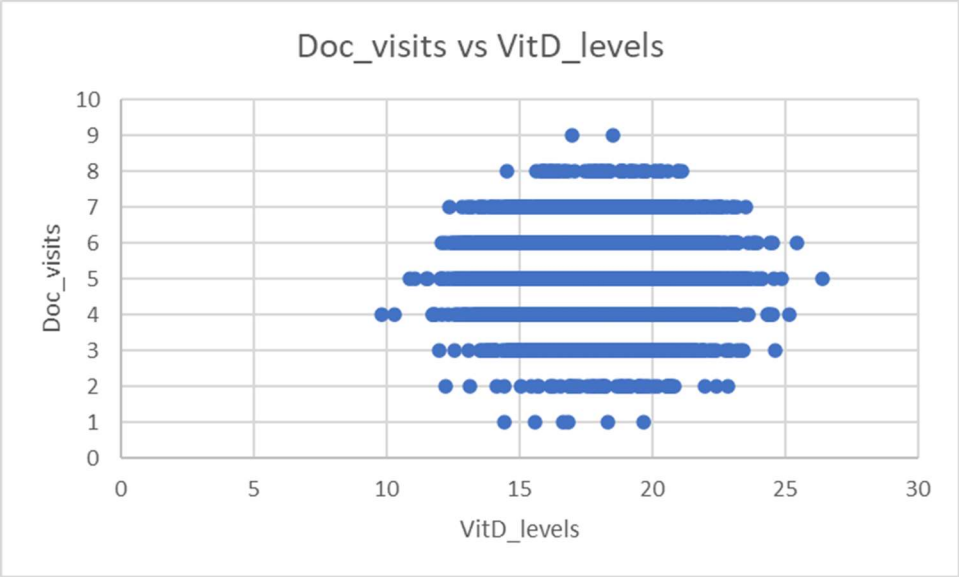
Services x Area

Services x Area	Column Lab			
Row Labels	Rural	Suburban	Urban	Grand Total
Blood Work	33.85%	32.86%	33.30%	100.00%
CT Scan	34.45%	33.06%	32.49%	100.00%
Intravenous	32.88%	34.12%	33.00%	100.00%
MRI	35.79%	32.89%	31.32%	100.00%
Grand Total	**33.69%**	**33.28%**	**33.03%**	**100.00%**

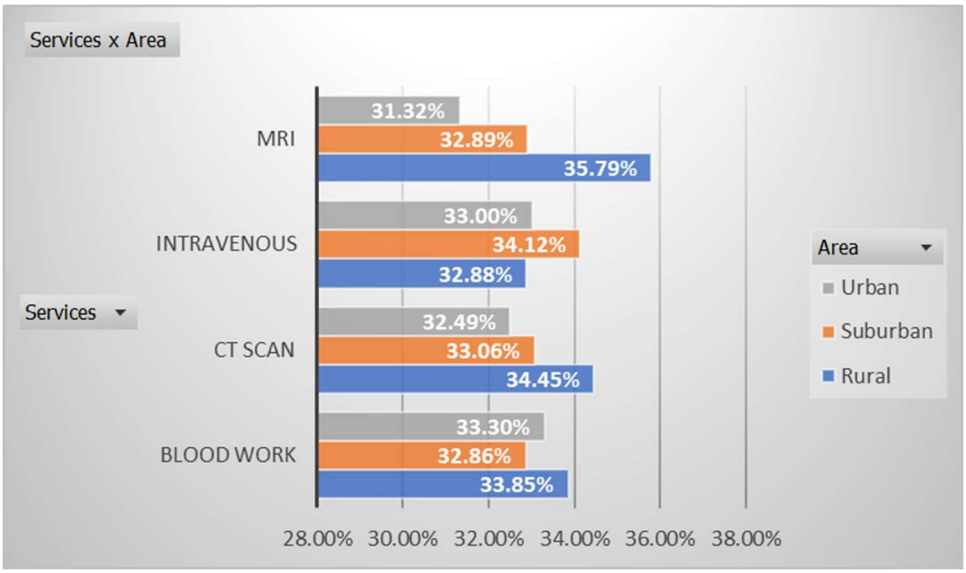
Marital x Gender

Marital x Gender	Column Labels					
Row Labels	Divorced	Married	Never Married	Separated	Widowed	Grand Total
Female	9.84%	9.93%	9.94%	10.17%	10.30%	50.18%
Male	9.22%	9.97%	9.57%	9.27%	9.65%	47.68%
Nonbinary	0.55%	0.33%	0.33%	0.43%	0.50%	2.14%
Grand Total	**19.61%**	**20.23%**	**19.84%**	**19.87%**	**20.45%**	**100.00%**

Part D1



(Zach, 2021c)



(Paula's Web3 & Crypto, 2019)

Part E1

The Chi-Squared test showed that the p-value was greater than 0.05, therefore there was no significant association between the Marital and Initial_admin variables.

Part E2

A limitation of the Chi-Squared test analysis is that it cannot determine if the variables have a causal relationship.

Part E3

Since there is no significant association between the variables, it is recommended that the organization consider other possible factors that determine how a patient is admitted to the hospital.

Part F

The Panopto video can be viewed at https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=555d23bb-e0fa-4e42-968b-b085016cd613

Part G

Zach. (2021b). How to perform univariate Analysis in Excel (With Examples). Statology. https://www.statology.org/univariate-analysis-excel/

Zach. (2021c). How to perform Bivariate Analysis in Excel (With Examples). Statology. https://www.statology.org/bivariate-analysis-in-excel/

[Paula's Web3 & Crypto]. (2019, April 22). Using Pivot tables to analyze Categorical Data [Video]. Youtube.com. Using Pivot tables to analyze Categorical Data

Zach. (2021). How to create a contingency table in Python. Statology. https://www.statology.org/contingency-table-python/

Part H

Not applicable.