Sultan Qaboos University

College of Science

Department of Statistics

Diagnoses Related to Vitamin D Levels and Other Factors:

A study done among people in Oman

Submitted in partial fulfillment of the requirements for the

Degree of Bachelor's in Statistics

Done by:

Marwan Said Amur Al Omairi (ID: 120560)

Yousuf Mubarak Rashid Al Ismaili (ID: 119882)

Supervisor:

Dr. Maryam Al Alawi

April 2023

SULTAN QABOOS UNIVERSITY
COLLEGE OF SCIENCE/DEPARTMENT OF STATISTICS
STAT5557: PROJECT IN STATISTICS,
FINAL REPORT CLEARANCE FORM

Project Title Diagnoses Related to Vitamin D Levels and Other Factors: A study done among people in Oman

A. Student Information:

S.No	Student Id.	Student Name
1	119882	Yousuf Al Ismaili
2	120560	Marwan Al Amiri
3		

B. Supervisor Name: Maryam Al Alawi

C. Students acknowledgment:

I (we) certify that the attached final project report has been completed by me (us) and that anything which is copied from or based upon the work of others has its source clearly acknowledged in the text at the place where it appears in the report.

Student's Signature: (1) …………………………………. **Date**: 10 May 2023
(2) …………………………………. **Date:** 10 May 2023
(3) …………………………………. **Date:** ……………........

D. Late submission, Comments, Suggestions or Corrections:

If this is a **late submission** after **week 13** of the second semester of the project please enter
Week :
Comments, suggestions or corrections to the final report are clearly marked and written in the text at the place where it occurs on the report. Please enter a tick below:
[ ] Yes, there are corrections to be made to the final report.
[ ] No, there are no corrections to be made.

Signature (Project Committee): ……………………………………**Date:**………………

I (we) certify that all corrections as suggested by the *Project Examiner* have been made to the attached revised final report.

Student's Signature: (1) …………………………………. **Date:** ……………........
(2) …………………………………. **Date:** ……………........
(3) …………………………………. **Date:** ……………........

Supervisor's Signature:  **Date:** 11 May 2023

Abstract

This project investigates the relationship between vitamin D levels and various health diagnoses. Vitamin D is an essential nutrient that plays a crucial role in bone health and immune system function. Studies have suggested that low levels of vitamin D may increase the risk of a range of health problems, including osteoporosis, cancer, and cardiovascular disease.

In this study, we investigated the relation between common diagnosis and the levels of vitamin D in the body -that can be categorized as deficiency, insufficiency and normal-, along with other important factors such as patient's gender, age, and region. The diagnosis that are of interest; dilated cardiomyopathy, hypocalcemia, nonscaring hair loss, and rickets. In this project, we applied some statistical techniques such as Pearson chi-square test of independence and the binary logistic regression model to analyze and evaluate the risk factors that can increase the rates of having any of the above diagnosis. Based on our results, we found that levels of vitamin D can increase the rates of being diagnosed with cardiomyopathy. We also found that there is no clear relation between vitamin D and age, and vitamin D and gender.

Our findings have important implications for public health, particularly in areas where vitamin D deficiency is common. Further research is needed to better understand the optimal levels of vitamin D for various health outcomes and to identify the most effective strategies for improving vitamin D status. Overall, our study supports the importance of maintaining adequate vitamin D levels for optimal health.

Acknowledgments

First, we thank God for helping, guiding and providing us with the confidence and strength to complete this project on time. We would like to extend our greatest gratitude to Dr. Maryam Al Alawi, our academic supervisor, for her ongoing assistance and useful knowledge. We also want to express our thanks to all of the professors in the Department of Statistics, for their excellent instruction and for imparting better skills and knowledge to us. We are also grateful to SQUH for their trust to provide us with the required data, and a special thank to Mr. Bader Al Busaidi in SQUH, who made this process easy and in time. Finally, we really would like to express our gratitude to our families and close friends who have supported us from the beginning till now. Additionally, we pray to God that our research will be conducted in the proper manner and with the best results possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Non-communicable diseases (NCDs) are conditions or illnesses not brought on by infectious agents (non-infectious or non-transmissible). A confluence of genetic, physiological, environmental, and behavioral variables leads to NCDs, which frequently have a long lifespan. The main categories of NCDs are diabetes, cancer, chronic and vitamin deficiency, respiratory diseases (such as chronic obstructive pulmonary disease and asthma), and cardiovascular diseases (including heart attacks and stroke). Other NCDs exist, but they are far less common. Oman is a country located in the Middle East, and its residents have been shown to have a high prevalence of vitamin D deficiency. This is likely due to a combination of factors, including limited sun exposure, cultural practices that promote covering the skin, and a lack of dietary sources of vitamin D.

The country of Oman is located 21 degrees north of the equator. Its relatively close proximity to the equator makes it possible for people to produce vitamin D from sun exposure year-round. This is in contrast to countries that are further from the equator, where it's impossible to make vitamin D during the winter due to low UVB intensity. Despite this advantage, low vitamin D levels are often reported in the Middle East, likely due to heavy clothing use and other lifestyle factors. According to a study published in the Oman Medical Journal in 2011, the prevalence of vitamin D deficiency among Omani adults was around 71%. The study also found that women were more likely to be deficient than men and that older individuals were more likely to have a severe deficiency.

In response to this issue, the Omani government has launched initiatives to increase awareness about the importance of vitamin D and promote its consumption. This includes encouraging more outdoor activities and increasing the availability of vitamin D-fortified foods such as milk and cereals (Al-Kindi, 2011).

Vitamin D is an essential nutrient that plays a crucial role in maintaining strong bones and a healthy immune system. Vitamin D deficiency occurs when the body does not get enough of this nutrient.

A hormone called vitamin D regulates the amount of calcium in the bones and blood. Strong bones, muscles, and general health depend on it. Vitamin D is a lipid-soluble vitamin that acts as a hormone. The primary function of vitamin D is to maintain the equilibrium of calcium levels. It is now understood that vitamin D is crucial for both children and adults to maintain healthy bones (Holick, 1996).

Some of the common causes of vitamin D deficiency and insufficiency include: Lack of sun exposure The primary source of vitamin D is sunlight. When your skin is exposed to sunlight, it produces vitamin D. However, people who live in areas with little sunlight or who spend most of their time indoors may not get enough vitamin D. and While vitamin D can be obtained from certain foods, such as fatty fish, egg yolks, and fortified foods like milk and cereals, it can be difficult to get enough vitamin D through diet alone.

also, there are certain medical conditions, such as Crohn's disease, celiac disease, and liver disease, can interfere with the body's ability to absorb vitamin D. overall Vitamin D is a fat-soluble vitamin, which means that it is stored in fat tissue. People who are obese may have lower levels of vitamin D because the vitamin is stored in their fat tissue instead of circulating in their bloodstream. there are Certain medications, such as anticonvulsants and glucocorticoids, can interfere with vitamin D metabolism and absorption.

Symptoms of vitamin D deficiency can include bone pain, muscle weakness, fatigue, and an increased risk of fractures. In severe cases, vitamin D deficiency can lead to rickets in children and osteomalacia in adults.

Treatment for vitamin D deficiency typically involves increasing sun exposure, eating foods high in vitamin D, and taking vitamin D supplements. However, it's important to talk to your healthcare provider before starting any supplements, as vitamin D can interact with certain medications and can be harmful in excessive doses (Pearce and Cheetham, 2010).

1.1 background

1.1.1 Scientific Definition of Vitamin D

Vitamin D is a fat-soluble vitamin that plays an important role in regulating calcium and phosphorus metabolism in the body. Chemically, it is a steroid hormone precursor that is synthesized in the skin through exposure to sunlight or obtained from dietary sources such as fortified foods, fatty fish, and supplements. Methods for measuring 25(OH)D and 1,25(OH)2D in the blood were created when these metabolites were discovered. The barometer for vitamin D status is serum 25(OH)D. Whereas Serum 1,25(OH)2D does not indicate the presence of vitamin D and is frequently normal.(Pearce and Cheetham, 2010)

In the body, vitamin D undergoes two hydroxylation reactions to become an active hormone, known as calcitriol. Calcitriol binds to specific receptors in target tissues, such as the intestines, kidneys, and bones, to regulate calcium and phosphorus absorption, reabsorption, and utilization. Vitamin D also plays a role in modulating immune function, cell growth, and differentiation. Deficiency in vitamin D can lead to a variety of health problems, including bone disorders, immune dysfunction, and increased risk of chronic diseases such as diabetes, cardiovascular disease, and cancer.

1.1.2 Source of Vitamin D

The main sources of Vitamin D can be obtained in the following ways: through the skin, food, and supplements.

One does not need to burn or tan to receive vitamin D from the sun. In roughly half the time it takes for the skin to burn, the body can produce all the vitamin D it needs for a day. Most individuals in the world get at least some of the vitamin D they need through sunlight. The production of vitamin D after exposure to UV radiation is affected by several variables such as the season, day length, time of the day, cloud cover, smog, skin melanin content, and the use of sunscreen. Less vitamin D can be made from sunshine by older persons and people with darker skin. Because UV radiation cannot pass through the glass, sunlight exposure indoors through a window does not affect the production of vitamin D (Wacker and Holick, 2013). It is estimated that upwards of 80-90 % of the body's requirement for vitamin D comes from this source.

Unfortified foods hardly ever contain vitamin D. fish and tuna liver are rich sources of vitamin D. In addition, milk, yogurt, and cheese also have a good amount of this vitamin. At the same time, cereal and some bread products are among the many foods with vitamin D added to them. To help individuals get adequate Vitamin D in their diets, some countries fortify foods like milk, yogurt, and bread with vitamins, including vitamin D. In addition, many countries encourage their citizens to take a daily Vitamin D supplement, especially throughout the winter, because their exposure to the sun is minimal.(Wacker and Holick, 2013)

1.1.3 Level of Vitamin D

Measuring a person's circulating level of 25(OH)D is the only way to tell if they have enough vitamin D. Vitamin D levels might be expected, insufficient, or deficient. These levels are not fixed and vary from country to country based on different factors. For instance, in Oman, the group of 25(OH)D is defined as shown in Table 1.1.

Level of Vitamin D	**Adults**	**Children**
Deficiency	less than 50	less than 35
Insufficiency	50 - 75	35 - 50
Normal	above 75	above 50

Table 1.1: Distribution of vitamin D in adults and children under the three categories

To quantify 25(OH)D, a number of assays are employed and several techniques are applied. Vitamin D sufficiency and inadequacy can be determined using competitive protein binding tests for 25(OH)D and radioimmunoassays.

Vitamin D Deficiency

Today, vitamin D deficiency is acknowledged as a global epidemic, and in general vitamin D level that are less than 50 ng/ml [25(OH)D] for both adults and children is considered low. There is a need increase awareness among people to understand that moderate sun exposure is the primary source of vitamin D. For most people not exposing to the sun properly during the day is the main contributor to vitamin D deficiency. Very few foods naturally contain vitamin D, and those that are supplemented with it frequently fall short of meeting either adult's or child's needs. Rickets in children and osteopenia, osteoporosis, and fractures in adults are brought on by vitamin D deficiency. Lack of vitamin D has been linked to an increased risk of viral disorders, hypertension, autoimmune diseases, and common malignancies. To maximize vitamin-positive effects on health, blood levels of the vitamin must be above 75 nmol/L or 30 ng/mL. For both children and adults respectively, this may require a minimum of 800-1000 IU of vitamin D3/d in the absence of sufficient sun exposure. When given in physiologic amounts, vitamin D2 may be just as effective at sustaining circulating levels of 25-hydroxyvitamin-D (Pearce and Cheetham, 2010)

Vitamin D Insufficiency

Although insufficiency is better than deficiency of vitamin D levels, it is still assumed to be low levels of vitamin D and must be treated when found out. As we know, insufficiency is defined as a serum [25(OH)D] level of 50 to 75 ng per mL. In fact, these two low levels of vitamin D both are considered as a main contributor to some common diseases. For instance, Vitamin D insufficiency can also cause osteomalacia and osteoporosis, in addition to widespread muscle weakness, muscle aches, and bone pains. Although the kidney plays a critical role in its creation, it is now known that a range of different tissues has the enzymatic machinery to make 1,25-dihydroxy vitamin D, which is essential for regulating calcium, phosphorus, and bone metabolism.

It is commonly known that vitamin D affects how bones and calcium are metabolized. Vitamin D insufficiency is the root cause of rickets in infants and young children, as well as osteomalacia and osteoporosis in adults. However, it's becoming increasingly clear that vitamin D has a much more comprehensive range of physiological impacts than was previously believed. The vitamin D receptor is present in nearly all cells, and almost all of them respond when exposed to vitamin D. Approximately

3 % of the human genome is under the direct or indirect control of the vitamin D endocrine system. Vitamin D consequently impacts various physiological processes, including immune response, cellular integrity, nervous system functioning, and muscle performance. It is not difficult to imagine how significant disturbances in these basic systems may harm health (Rao et al., 1987).

1.2 Related Diseases to Vitamin D

1.2.1 Hypocalcemia

Vitamin D plays an important role in regulating calcium and phosphorus levels in the body, so a deficiency of this vitamin can lead to a variety of conditions related to bone health, including hypocalcemia. Hypocalcemia is a condition characterized by low levels of calcium in the blood. This can be caused by a variety of factors, including vitamin D deficiency, which impairs the body's ability to absorb calcium from the diet. Other causes of hypocalcemia include certain medications, kidney disease, and hormonal imbalances.

Symptoms of hypocalcemia can include muscle cramps, twitching, and spasms, as well as numbness or tingling in the fingers and toes. In severe cases, hypocalcemia can lead to seizures, cardiac arrhythmias, and other life-threatening complications. Treatment for hypocalcemia depends on the underlying cause of the condition. If vitamin D deficiency is the cause, increasing the intake of vitamin D through diet or supplements may be recommended. In some cases, calcium supplements may also be prescribed to help restore normal levels of calcium in it (Balasubramanian et al., 2006).

1.2.2 Cardiomyopathy

There is some evidence to suggest that vitamin D deficiency may be associated with an increased risk of cardiomyopathy, which is a disease relate to heart muscle. Cardiomyopathy can cause the heart to become enlarged, thickened, or rigid, which can interfere with its ability to pump blood effectively.

Vitamin D plays an important role in maintaining healthy bones and teeth, but it also has other functions in the body, including regulating immune function and reducing inflammation. Some research has suggested that vitamin D may also be involved in maintaining heart health.

A number of studies have found that low levels of vitamin D are associated with an increased risk of heart disease, including cardiomyopathy. However, more research is needed to fully understand the relationship between vitamin D and cardiomyopathy and to determine whether vitamin D supplementation can help prevent or treat this condition (Polat et al., 2015).

1.2.3 Rickets

Rickets is known as a bone disorder that occurs due to a deficiency of vitamin D, calcium, or phosphate in the body, which leads to weakened and soft bones. Vitamin D is essential for the body to absorb calcium and phosphate from the diet, which is necessary for bone growth and development. In children, rickets can cause growth retardation, skeletal deformities, and delayed tooth formation. In adults, vitamin D deficiency can lead to osteomalacia, a condition characterized by weak bones, muscle weakness, and bone pain (Ives and Brickley, 2014).

1.2.4 Hair loss

One such condition is alopecia areata, an autoimmune disorder that causes hair loss on the scalp and other areas of the body. Vitamin D has been found to play a role in regulating the immune system, and deficiencies in vitamin D have been linked to autoimmune diseases.

Another condition that can be related to vitamin D deficiency and hair loss is telogen effluvium, which is characterized by sudden hair loss due to a disruption in the hair growth cycle. Vitamin D is important for the normal growth and maintenance of hair follicles, and a deficiency can lead to hair thinning and loss. It is important to note that while vitamin D deficiency can contribute to hair loss, it is often not the sole cause. Other factors such as genetics, stress, and certain medications can also play a role (Vegesna et al., 2002).

1.2.5 Adult osteomalacia

Osteomalacia translates as (soft bones). Bones can become more brittle and more prone to breaking because of osteomalacia. Bone breaks down more quickly than it can be rebuilt due to a disease of reduced mineralization. Adults can suffer from this illness. When vitamin D levels are too low, it can lead to rickets in young people (Rao et al., 1987).

1.2.6 Fracture in body

A fracture is a crack or break in a bone. Most fractures are caused by applying pressure to a bone. Overuse or injury are the usual causes of fractures. The wounded area is frequently swollen, painful (particularly when used), bruised, and seems deformed, crooked, or out of place (De Laet et al., 2005).

1.3 Objectives of the Project

The main purpose of our study is to investigate the relationship between four non-communicable diseases with vitamin D and other factors such as gender, age, and region. The diseases that we will focus on are Hypocalcemia, Cardiomyopathy, Rickets, and Non-scaring Hair Loss. Each disease will be evaluated separately using the Chi-square test of independence and the binary logistic regression model.

Further, the project will investigate the relationship between different variables as follow:

- Association between age and vitamin D.
- Association between gender and vitamin D.
- Association between region and vitamin D.

1.4 Organization of the Report

This report comprises four chapters, each with a distinct focus. Chapter 1 provides an introduction to the study, presenting the background, research problem, and objectives. In Chapter 2, a brief description of the data and the descriptive statistical methods employed are presented, along with statistical summaries. Chapter 3 details the application of various statistical methods to analyze the data. Finally, Chapter 4 presents the project's concluding remarks, highlighting major findings and major points.

Chapter 2

Descriptive Statistics

2.1 The Project Data

2.1.1 Data Collection

The data of this project were obtained from the SQU hospital. The data mainly consists of diagnosis with patients, measurements of the Vitamin D test D[25(OH)D], and general information about the patients such as age, gender, etc. The procedure in the hospital is that if one patient is diagnosed with a specific disease, his/her doctor will require him to do a vitamin D test. Because, in their records these diagnoses might be related directly to vitamin D levels. However, after testing not all the patients appear to have low vitamin D.

The hospital then keeps records of all patients who do the Vitamin D test D[25(OH)D] over time. The patients who's results show low levels of vitamin D are supplied with medications and supplements to boost their vitamin D levels, then they are required to repeat the test later. Although, our records consists of data from 2010 to 2022, there are few cases that relate to the same patients.

We received the data from the hospital in an Excel program. The data includes different variables such as collected date of Vitamin D test, diagnosis name, diagnosis medical name, lab episode (unique number for each patient)[1], the type of test (all is D[25(OH)D]), the result of vitamin D, region, age,

[1]The unique number of patients is confidential and do not reveal any information about the patients, and it is different

gender, and year of hospital visit. We also searched for any statistical research, read literature reviews related to this subject, and tried to gather enough information about vitamin D from different websites to see what are the main variables that must be considered when studying about vitamin D and the related diagnosis.

2.1.2 Data Cleaning

Originally the data consists of around 1300 patients, after cleaning the data we ended with 1193 patients. There are different treatments we had to do to put the data in its best format, these can be summarized below:

- We omitted all (NA), which means an empty set or missing value.
- There are some branching diseases, so we have collected them into specific conditions to make it easy.
- We deleted all duplicates, patients who have same lab episode (less than 20 cases).
- We added a new variable which is the level of vitamin D (Normal deficiency insufficiency), based on the definition of vitamin D categories in the SQUH and by using the numerical vitamin D measurements
- We created dummy variables for each of the diseases as (0 for no disease, and 1 for the disease).
- We rephrased the names of the regions and combined the Wilayat into one council city (Muhafadha).
- We discarded all unneeded variables, such as collected date, diagnosis medical name, type of test.

2.1.3 Variable Description

The variables of the data can be summarized in the table below.

from the hospital patient ID.

Variable name	Variable distribution	Data type
Vitamin D measures	D[25(OH)D] results	Numerical
Age	Age of patient	Numerical
Diagnosis	Types of related diseases	Categorical
Region	Patient hometown	Categorical
Sex	Patient gender (Male, Female)	Categorical
Year	Year of the test (2010-2022)	Categorical
Age group	Patient age group (Adults, children)	Categorical
Levels of Vitamin D	Normal, insufficiency, deficiency	Categorical

Table 2.1: The table summarizes all the variables in the data

Explanatory Variables

Selected demographic characteristics of the respondents were considered as the explanatory variables in this study. These include:

Gender: Male or Female.

Age: Adults above 14 years old, while children from 0 to 14 years old, this definition of age groups was taken from the source of the data (SQUH).

Region of residence: There are patients from all the 11 governorates of Oman; Muscat, Dhofar, Al-Dakhlia, North Sharqiah, South Sharqiah, North Batina, South Batina, Al-Dhahirah, Musandam, Al-Wasta and Al Buraymi.

In addition to the above, the main explanatory variable of interest in this study is the level of vitamin D.

Response Variables

As per our main purpose of this research, we want to investigate the relationship between the different diagnoses and the risk factors. Therefore, the diagnoses will be the response variable. However, the data set consists of a range of diseases, for that reason we will apply some descriptive statistics to choose the most common diagnoses, and focus on them.

2.2 Descriptive Statistics

Descriptive statistics is always the first phase for any researcher to investigate the data in hand and to create initial ideas about the data. In this section, we will use some methods to describe the data graphically and numerically. For the numerical variables in the data, we will compute the mean, median and standard deviation, while the graphical analysis will contain a scatter plot. On the other hand for the categorical variables, we will explore them graphically with bar graphs and pie charts for each variable depending on the variable type. By these methods, we just describe and summarize data but do not draw final conclusions.

2.2.1 Descriptive statistics

Numerical Data

The numerical and graphical methodologies used to arrange, present, and analyze data are known as descriptive statistics that usually summarize the central tendency, variation and confidence intervals of the data. The mean or average is found by summing all the numbers in the data set and then dividing the total by a number of values. The standard deviation is the square root of the variance. The SD of a dataset is a measure of its dispersion in relation to its mean and how to spread out numbers. The median is the middle value in a data list that has been ordered descending or ascending and it separates the lower half from the higher half of a data set. In ordered data, the maximum is the highest observation, while the minimum is the lowest observation. The confidence interval is the range of values that you expect your estimate to fall between a certain percentage of the time if you run your experiment again or re-sample the population in the same way.

Categorical Data

Categorical variables represent the data that can be sorted into groups or categories. It can be nominal or ordinal data. Analysis of categorical data generally involves the tables. A two-way table presents categorical data by representing two categorical variables, denoted by X and Y. Graphically the categorical data can be represented by Bar graphs, and Pie charts. In a bar chart, each category is represented by a separate vertical or horizontal bar, the length of which is proportional to the frequency of that category. To indicate that the data set isn't continuous, the bars are separated or split by small gaps. The "pie" in a pie chart is divided into different sections, one per for each category, with the area for every section proportional to the frequency in that category.

2.2.2 Numerical Data in the Study

There are only 2 numerical variables, age, and vitamin D, thus we will first explore them numerically by basic statistics measures, while some graphs will be presented later. Also, they will come again in the report as categorical variables. We notice that the median age of the patient is 34 years which means half of the patient is above 34 years old and half is below, and the mean is 37.551. The maximum age of the patient is 92 years old. Also, the minimum age of the patient is one year old, and the standard deviation is 21.176. In the level of Vitamin D, the deficiency for adults is less than 50 (ng/ml), and for children is less than 35 (ng/ml). The insufficiency is between 50-70 (ng/ml) for adults and 35-50 (ng/ml) for children, and the normal is above 70 (ng/ml) for adults. Above 50 for children, we can see in the table the median is 64 (ng/ml), the mean is 71.311 (ng/ml), the maximum is 414 (ng/ml), and the minimum is 7 (ng/ml), and the standard deviation is 39.888.

	Minimum	**Maximum**	**Mean**	**Median**	**St. Deviation**
Age	1	92	34	37.551	21.176
Vitamin D measures	7	414	64	71.311	39.888

Table 2.2: The table shows the minimum, maximum, mean, median, and stander deviation of our data.

2.2.3 Variable: Diagnosis

Since our response variable is diagnoses, we will explore it first using a pie chart to display all the recorded diagnoses. Figure 2.1 shows that nearly one third of the patients had Hypocalcemia, which represents (31%) and the second most common diagnosis is Nonscarring hair loss with (30%). Dilated cardiomyopathy makes up (25%) of total disease, another notable diagnosis is Rickets with 10% of the total cases. The remaining diagnoses consist of a small percentage such as adult osteomalacia, fracture of unspecified body regions, and other specified disorders of bone and Fracture, all together making less than (4%). Based on this result, we will only focus on the 4 most common diagnoses in our data which are: Hypocalcemia, Dilated cardiomyopathy, Nonscaring hair loss, and Rickets. All the forthcoming analyses will focus on these 4 diagnoses.

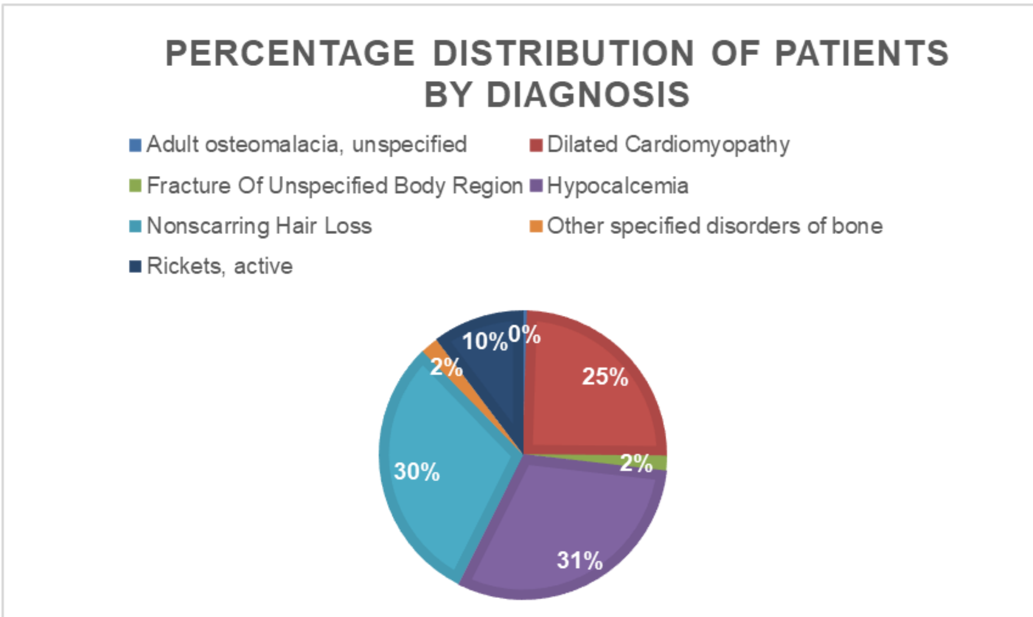


Figure 2.1: percentage distribution of patients by diagnosis

2.2.4 Variable: Gender

Table 2.3 provides the Omani patient by gender. It shows the percentage of male patients (33%) that are 389 patients which is less than the percentage of female patients (67%) that are 804 patients. The study had more female participants than males. The variable is also displayed as a pie chart Figure 2.2. We also looked at these numbers in terms of its distribution among gender, for example Figure 2.3 shows bar graphs of the variables gender and diagnosis.

Gender	Total number of respondents	Percent
Male	389	33%
Female	804	67%
Total	1193	100%

Table 2.3: Levels of vitamin D in adults and children under the three categories

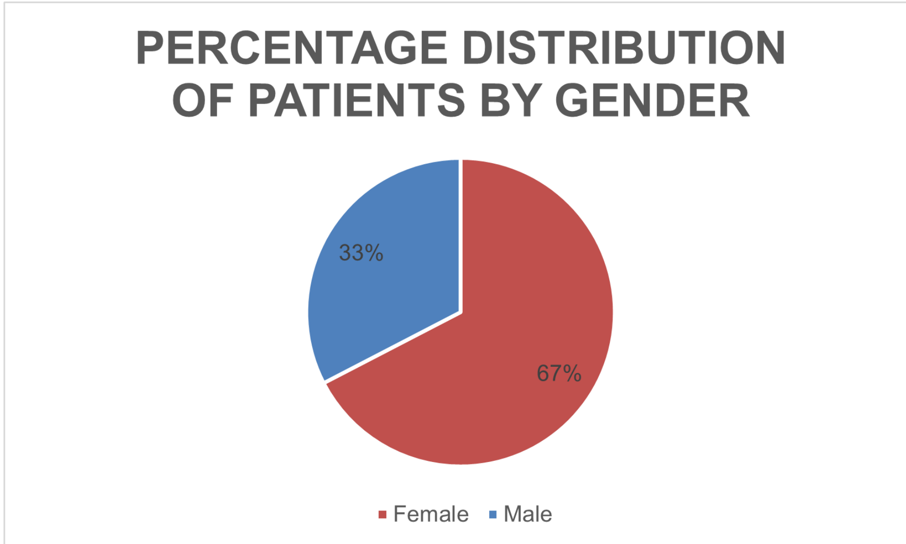


Figure 2.2: Gender distribution among patients

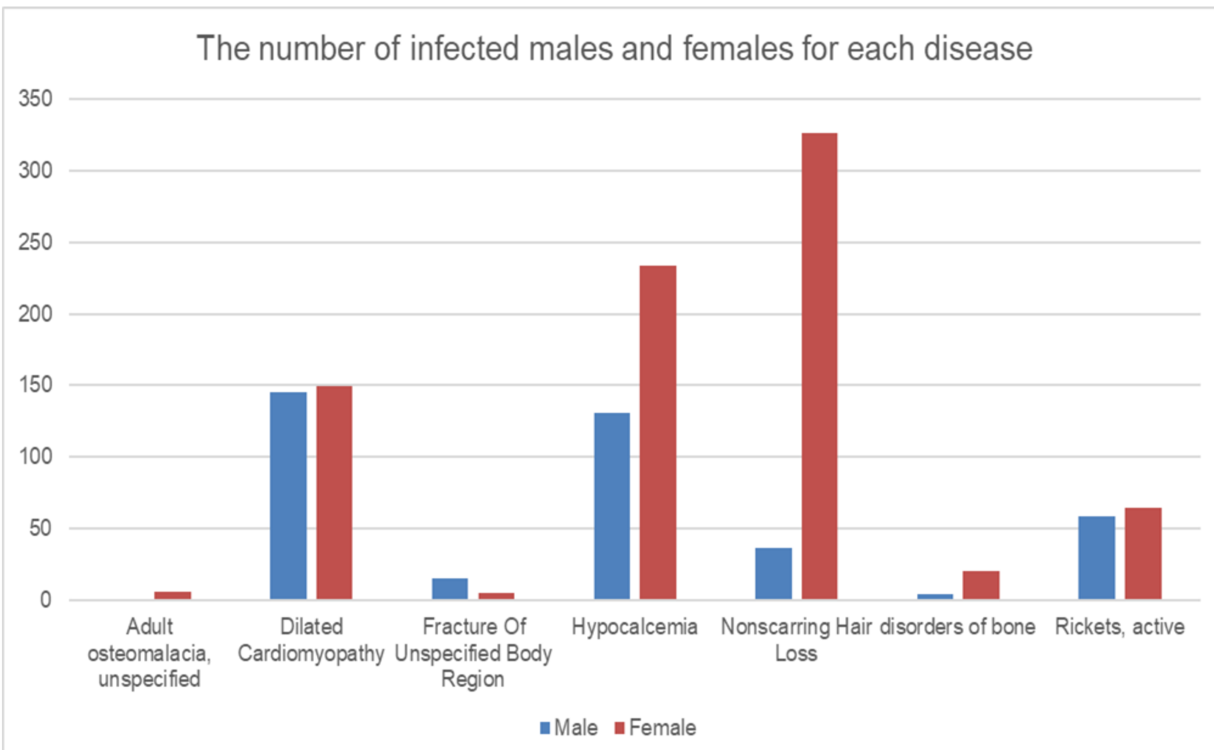


Figure 2.3: The number of infected males and females for each disease

2.2.5 Variable: Region

The figure 2.4 shows the percentage distribution of patients by region. The patients from the Muscat governorate were the most by (41%) and the fewer patients come from Dhofar by (2%) and Musandam by (0%), only 1 patient. This percentage is normal because most SQUH patients are from Muscat, while residents of Dhofar and Musandam are not able to come frequently due to the distance.

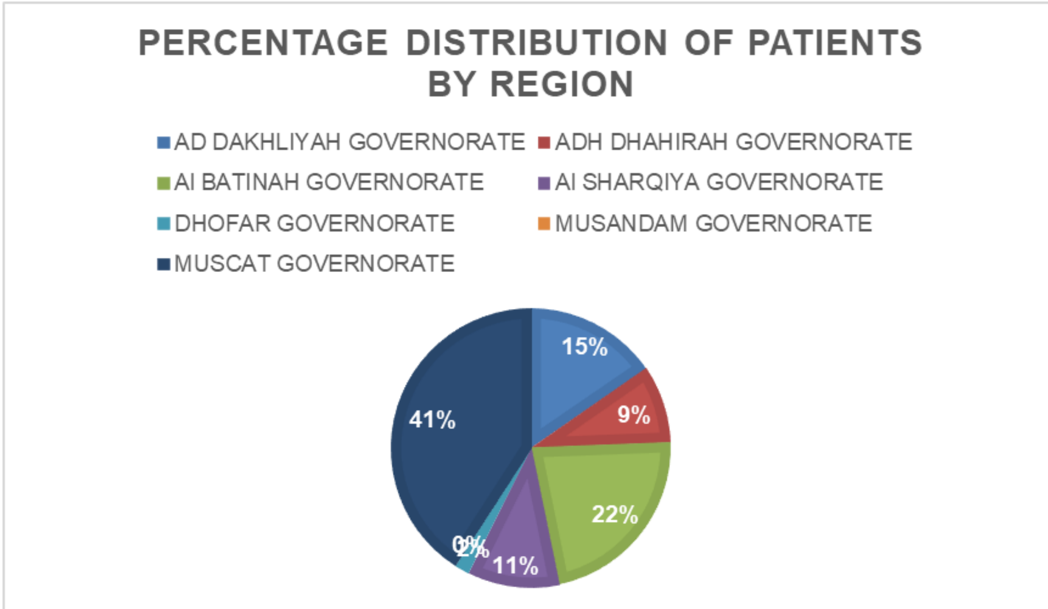


Figure 2.4: percentage distribution of patients by region

The number of infections of the four diseases among regions

The table below presents the number of patients of the four diagnoses for all the regions. It is clear that most patients are from Muscat, followed by Al Batinah, then Ad Dakhliyah. While there is only one patient from Musandam, and only 20 from Dhofar.

Region	Gender	Diagnosis				Total
		Cardiomyopa	Hypocalce	Nonscarrir	Rickets,	
AD DAKHLIYAH	Male	35	17	4	1	182
	Female	31	34	49	11	
ADH DHAHIRAH	Male	6	11	2	3	96
	Female	17	19	22	16	
AL BATINAH	Male	27	38	9	22	264
	Female	34	48	71	15	
AL SHARQIYA	Male	12	7	16	3	115
	Female	8	24	44	1	
DHOFAR	Male	3	2	0	0	20
	Female	10	5	0	0	
MUSANDAM	Male	0	0	0	0	1
	Female	0	0	1	0	
MUSCAT	Male	62	56	5	29	465
	Female	49	104	139	21	
Total		294	365	362	122	1143

Figure 2.5: The number of infected for gender and region in each disease

2.2.6 Variable: Age

The variable age is always of an interest when dealing with health data. In table (2.2) we saw that the youngest individual in the data is only 1 year old, while the oldest is 92 years old, and the mean is 34 years old. Although we have a good insight of age as a numerical variable in our data, we created a categorical variable of the age groups to be either adult or children, as defined by the SQUH. Figure (2.6) shows a pie chart of the patients groups. The difference between adults and children is clear, where (84%) are adult and (16%) are children individuals. In addition, table (2.4) displays the numbers of adults and children in the data.

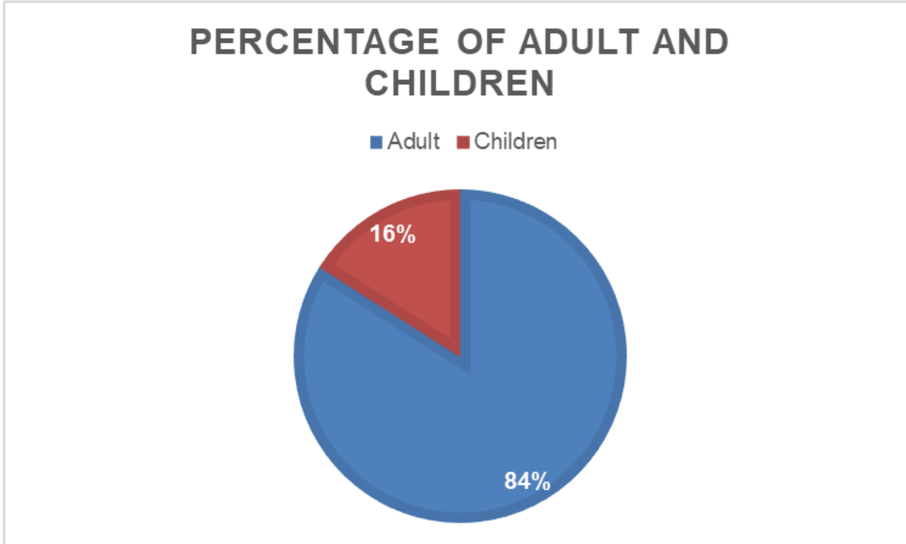


Figure 2.6: percentage of adult and children

Age group	Total number
Adult	1002
Children	191
Total	1193

Table 2.4: Numbers of adult and children

2.2.7 Variable: Vitamin D

The main explanatory variable in this data is Vitamin D, the D[25(OH)D] results have been recorded as numerical values range from 7 (minimum) to 414 (maximum) with mean 64 ng/ml, as shown in table (2.2). However, the common definition of vitamin D is Normal, insufficiency, and deficiency. Thus, we categorized the numerical values into the 3 groups of vitamin D, based on the distribution of vitamin D as defined in Oman (Table 1.1). Figure (2.7) shows that almost half of the people in the data are normal, and it represents 529 patients (44%) of the total number, while Deficiency and Insufficiency are the almost same number of patients and they represent both (28%).

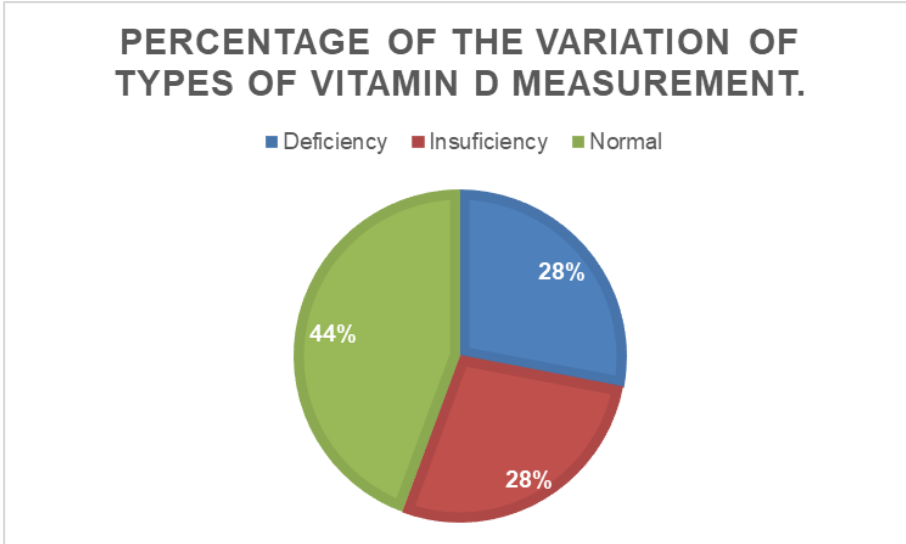


Figure 2.7: types of Vitamin D measurement

Vitamin D level	**Total number**
Deficiency	334
Insufficiency	330
Normal	529
Total	1193

Table 2.5: Numbers of patients in each level of vitamin D

Vitamin D for both adult and children

Further, we wanted to have an insight into the distribution of the numerical values of vitamin D. Thus, we looked at the data graphically by plotting the vitamin D values on the x-axis, and its frequencies on the y-axis as shown below.

Figure (2.8) provides information related to the distribution of vitamin D levels in both adults and children. It indicates that most individuals fall within the range of vitamin D levels that lie between insufficiency and normality, with the concentration centered around level 50 ng/ml. The figure also show a very clear outliers, which are the values above 400 ng/ml. However, since there is no identified limit for vitamin D levels, these records might be considered as normal vitamin D levels. We thought it is also interesting to look at the same graph but for adults and children separately.

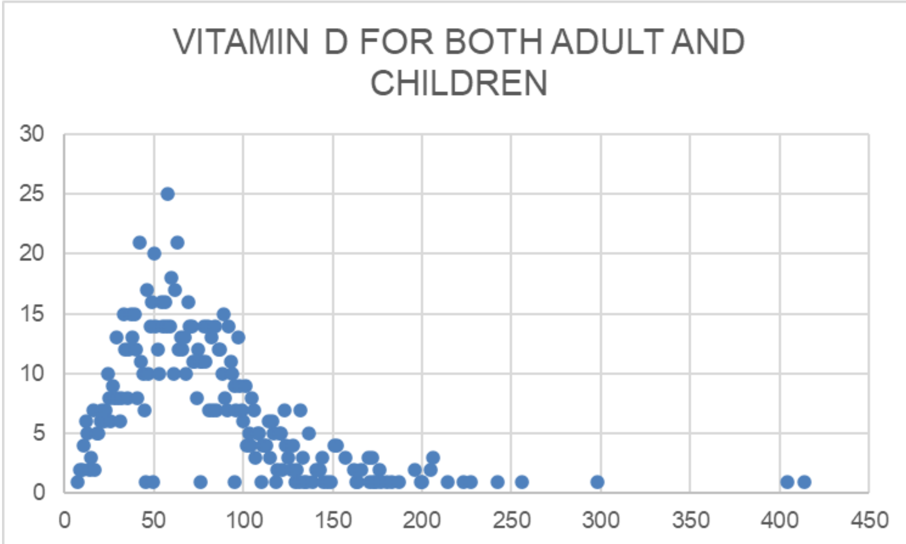


Figure 2.8: Vitamin D for both adult and children

Vitamin D for adult

Figure (2.9) provides information about the distribution of vitamin D levels in adults, more than half of adults have an average of over 50 of vitamin D, which falls between deficiency and normal levels. It is worth noting that this figure is similar to the full figure above.

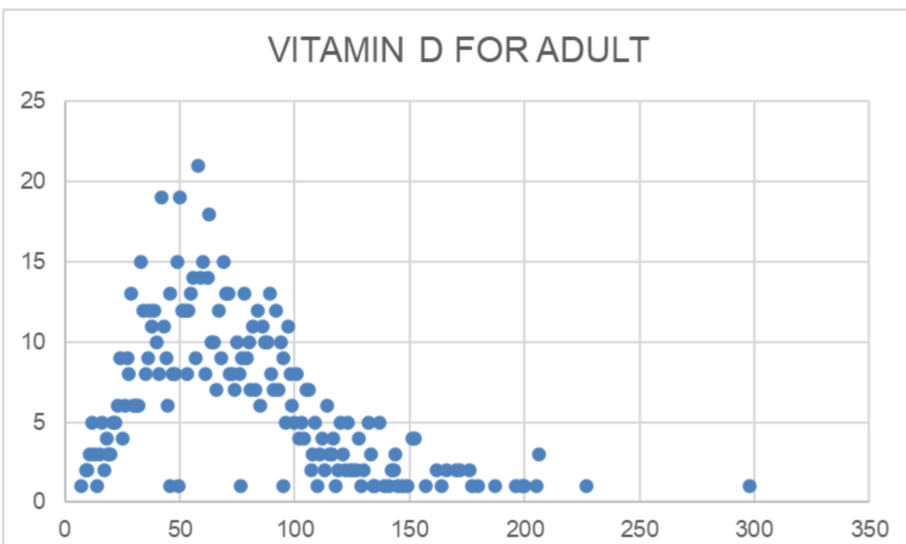


Figure 2.9: Vitamin D for adult

Vitamin D for children

Figure (2.10) provides information about the distribution of vitamin D levels in children. Since the numbers of children is less in our data, the figure consists of less frequency values, however, it should be mentioned that the outliers (values above 400) are actually coming form the children category. This notice lead us to suggest that the outlier could be an error in recording the numbers!

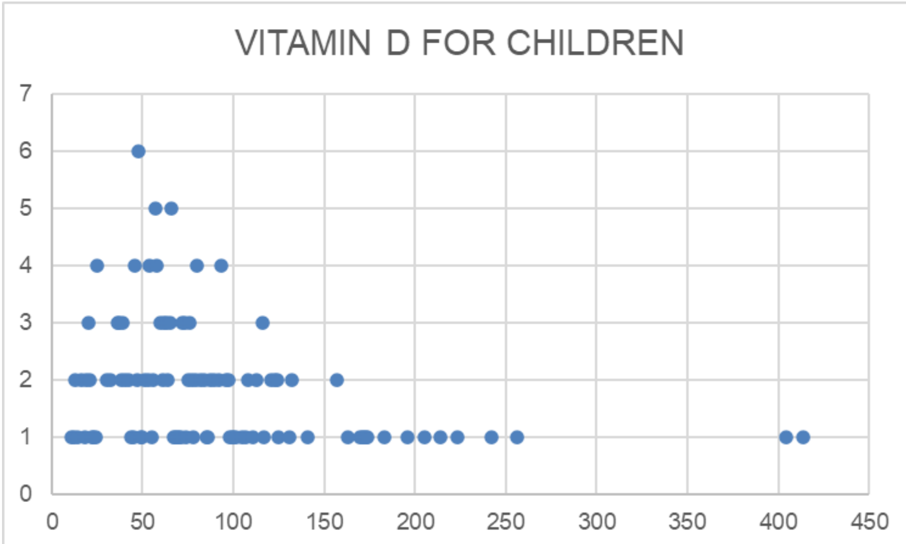


Figure 2.10: Vitamin D for children

2.3 Research Focus

2.3.1 Response Variable

Based on the descriptive statistics, we found that we should limit our investigation to 4 main diseases that are: Hypocalcemia, Nonscarring hair loss, Cardiomyopathy, and rickets. The diseases will be analyzed with some categorical variables. A brief description of these diagnoses is given below.

Hypocalcemia

Hypocalcemia is a treatable condition that happens when the levels of calcium in the blood are too low. Many different health conditions can cause hypocalcemia, and it is often caused by abnormal levels of parathyroid hormone (PTH) or vitamin D in the body.

Nonscarring hair loss

Vitamin D is metabolized in the skin by keratinocytes. These are skin cells that process keratin, a protein in hair, nails, and skin. When the body does not have enough vitamin D, keratinocytes in hair follicles have trouble regulating hair growth and shedding.

Cardiomyopathy

Cardiomyopathy is a disease that affects the heart muscle, making it difficult for the heart to pump blood effectively. Recent studies have suggested that vitamin D deficiency may contribute to the development of cardiomyopathy. Vitamin D is important for maintaining healthy bones, but it also plays a role in regulating the immune system and reducing inflammation, which can be beneficial for heart health. Low levels of vitamin D have been linked to an increased risk of developing cardiomyopathy and other cardiovascular diseases. Therefore, ensuring adequate vitamin D intake may be an important strategy for preventing or managing cardiomyopathy.

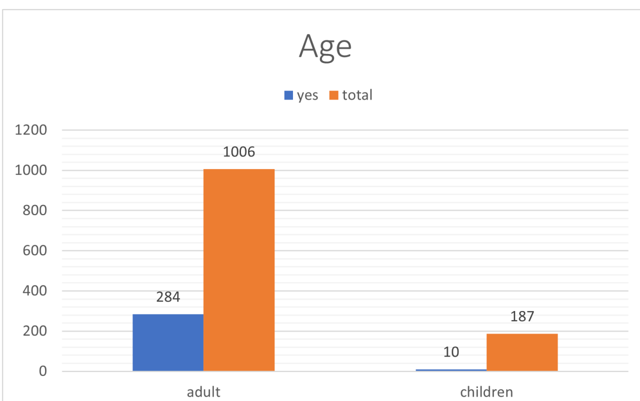
Rickets

Rickets is the softening and weakening of bones in children, usually because of an extreme and prolonged vitamin D deficiency. Rare inherited problems also can cause rickets. Vitamin D helps the child's body absorb calcium and phosphorus from food.

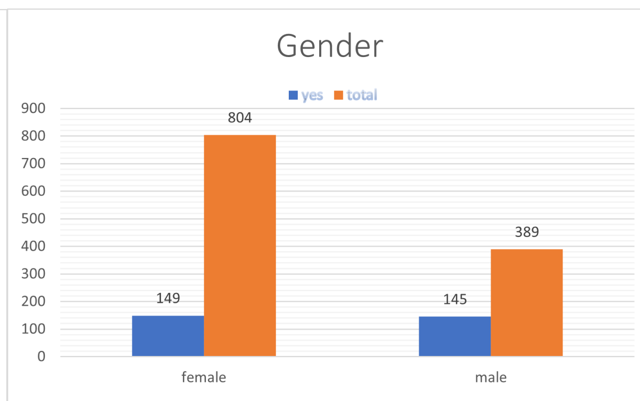
2.3.2 Response Variable & Categorical Variables

Cardiomyopathy

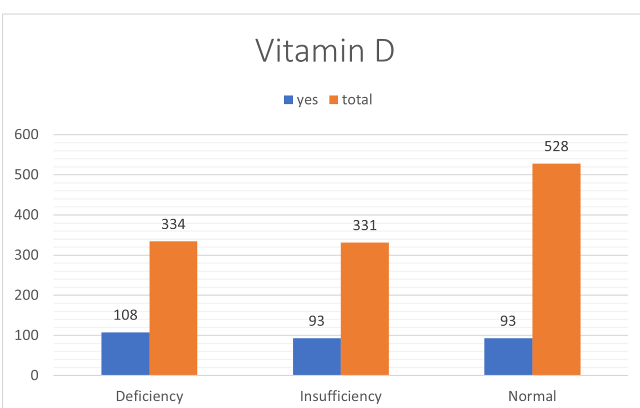
Figure 2.11 shows bar charts of Cardiomyopathy with four risk factors; gender, age, vitamin D, and region. It is clear that the number of having Cardiomyopathy among males and females is approximately the same, however, they represent two different ratios among each gender, thus we can say that the incidence of having the diagnosis with a male (%37.27) is more than female (%18.53). Similarly, the percentage of diagnosed in adults (%28.23) is more than in children (%5.3). In addition, there are many patients with Cardiomyopathy from Muscat compared to the other regions, however, in terms of the rations of having the diagnosis it is actually less in Muscat compared to the other regions. Overall, there are fluctuating ratios of having or not having the diagnosis in the different regions. The incidence of having Cardiomyopathy is approximately the same in the group of deficiency (%32.33) and insufficiency (%28.1) in vitamin D levels, while it is less in the Normal level (%17.6).



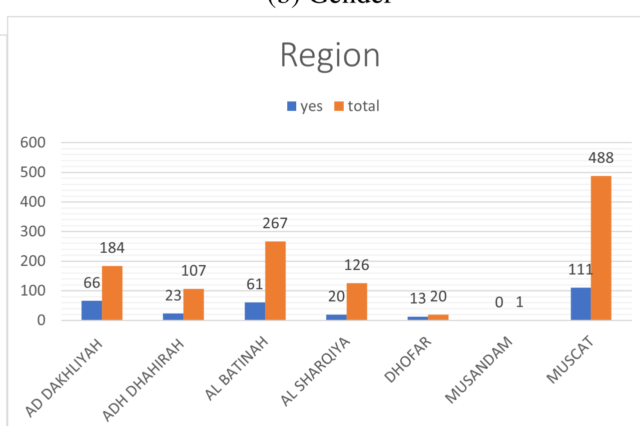
(a) Age



(b) Gender



(c) Vitamin D

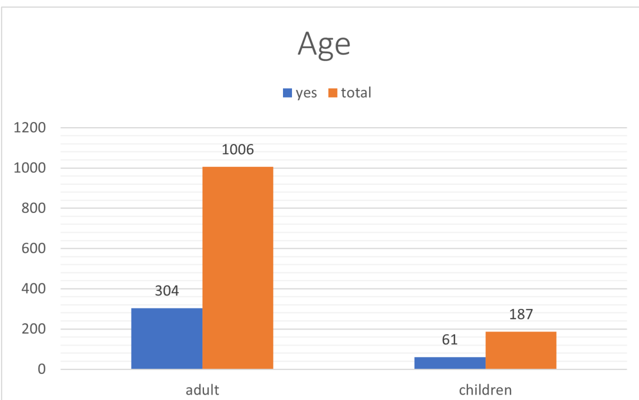


(d) Regions

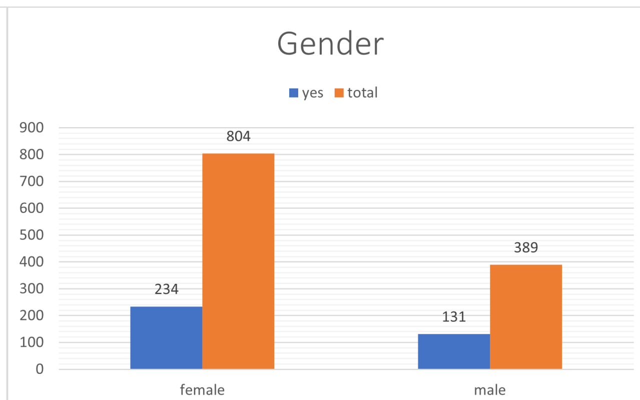
Figure 2.11: This figure shows the bar graph for Cardiomyopathy with the different variables

Hypoglycemia

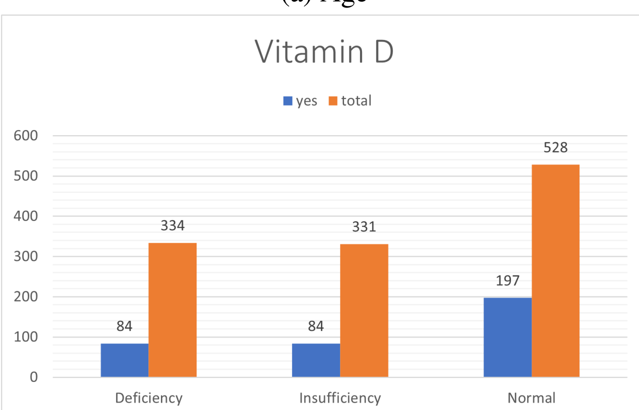
Figure 2.12 shows bar charts of Hypoglycemia with four risk factors; gender, age, vitamin D, and region. It is clear that the number of having Hypoglycemia among males and females is different. In terms of ratios, we can say that the incidence of having the diagnosis with a male is (%33.6) while it is (%29.1) for female. Similarly, the percentage of diagnosed in adults (%30.2) is slightly less than in children (%32.6). In addition, there are many patients with Hypoglycemia from Muscat compared to the other regions, however, in terms of the rations of having the diagnosis it is actually less in Muscat compared to the other regions. Overall, there are fluctuating ratios of having or not having the diagnosis in the different regions. The incidence of having Hypoglycemia is approximately the same in the group of deficiency(%25.14) and insufficiency(%25.37) in vitamin D levels, while it is more in the Normal level(%37.31) of vitamin D.



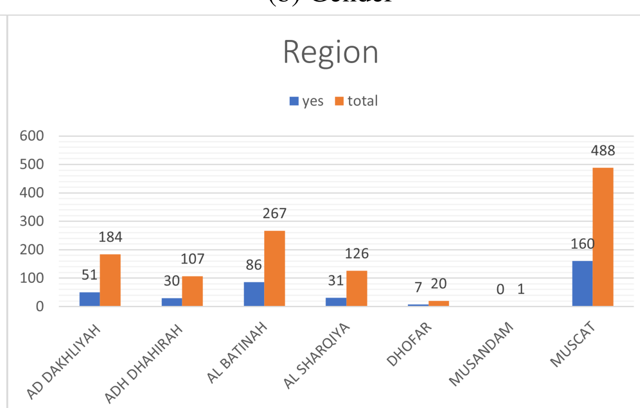
(a) Age



(b) Gender



(c) Vitamin D

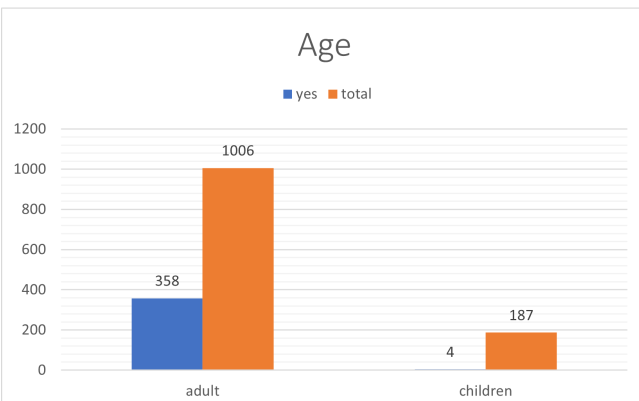


(d) Regions

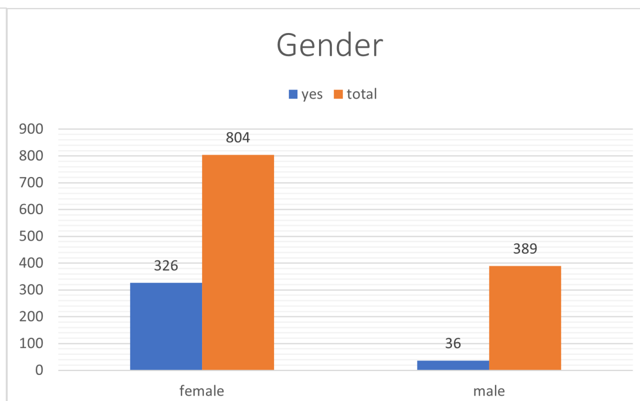
Figure 2.12: This figure shows the bar graph for Hypoglycemia with the different variables

Non-scaring Hair loss

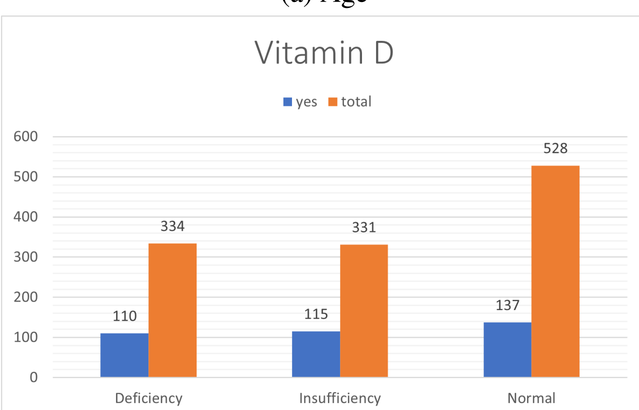
Figure 2.13 shows bar charts of Hair loss with four risk factors; gender, age, vitamin D, and region. It is very clear that the number of having Hair loss among females is more than in males. In fact, based on the rations, we can say that the incidence of having the diagnosis with female is (%40.5) while for male, it is only (%9.2). Similarly, the percentage of diagnosed in adults(%35.5) is more than in children, which is only (%2.1). In addition, there are many patients with Hair loss from Muscat compared to the other regions, however, in terms of the rations of having the diagnosis it is actually less in Muscat compared to the other regions. Overall, there are fluctuating ratios of having or not having the diagnosis in the different regions. The incidence of having Hair loss is approximately the same in the group of deficiency(%32.9) and insufficiency(%34.7) in vitamin D levels, while it is less (%25.9) in the Normal level of vitamin D.



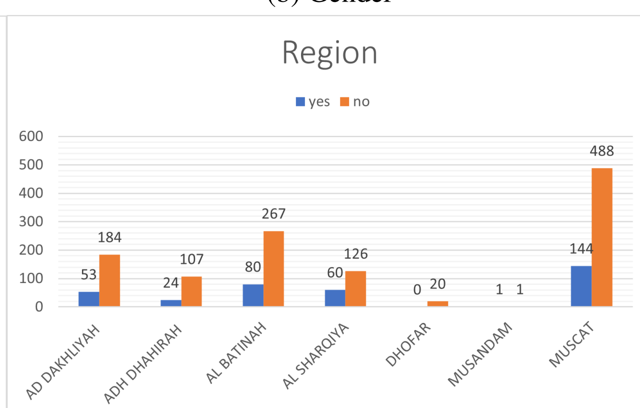
(a) Age



(b) Gender



(c) Vitamin D

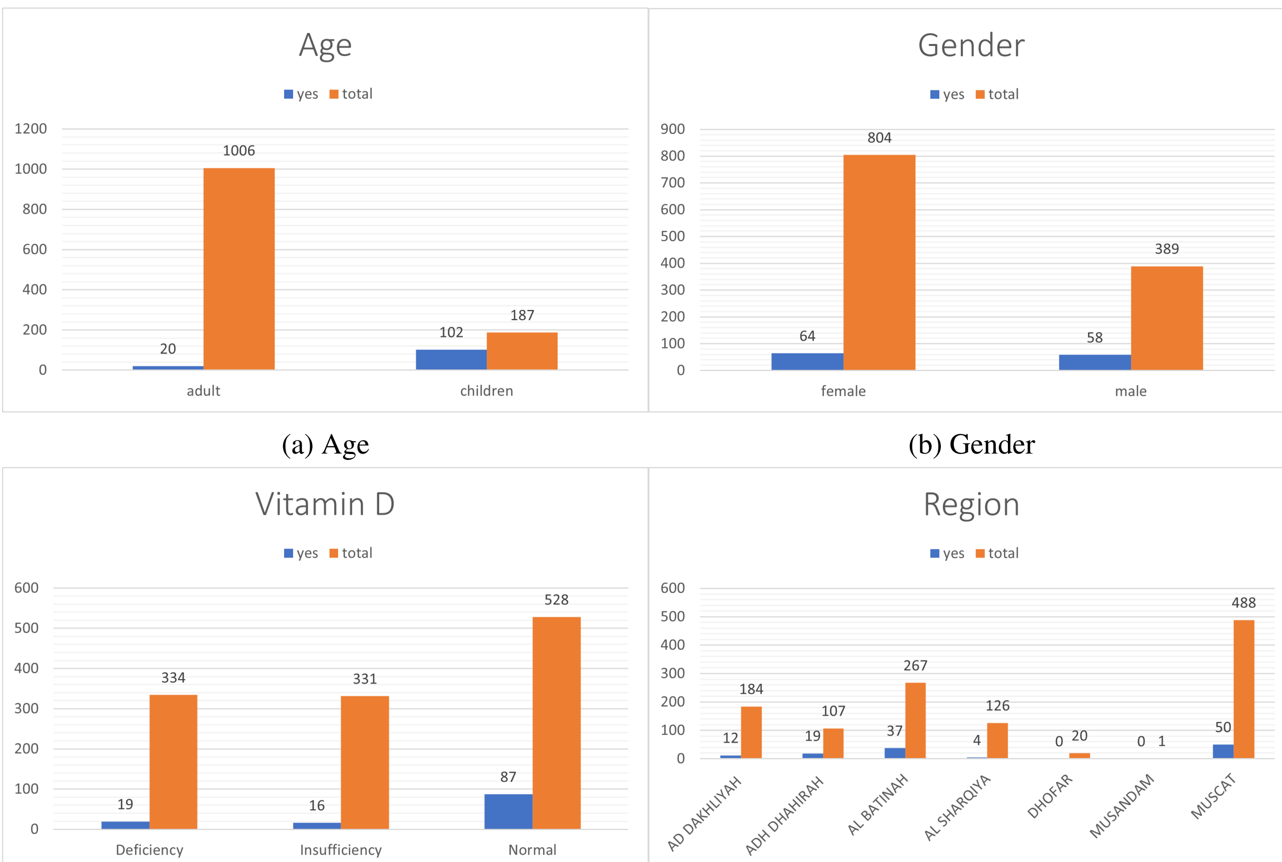


(d) Regions

Figure 2.13: This figure shows the bar graph for Hair loss with the different variables

Rickets

Figure 2.14 shows bar charts of Rickets with four risk factors; gender, age, vitamin D, and region. It is clear that the number of having Rickets among males and females is approximately the same, however, they represent two different ratios among each gender since there are much more females with no Rickets, thus we can say that the incidence of having the diagnosis with a male is more than female which is in males (14.9%) and in females (7.9%). Similarly, the percentage of diagnosed in children is very high, which is (54.54%) while it is only (1.9%) in adults. In addition, there are many patients with Rickets from Muscat compared to the other regions, however, in terms of the rations of having the diagnosis it is actually less in Muscat compared to the other regions. Overall, there are fluctuating ratios of having or not having the diagnosis in the different regions. The incidence of having Rickets is approximately the same in the group of deficiency (%5.6) and insufficiency (%4.8) in vitamin D levels, while it is more (%16.4) in the Normal level of vitamin D!



(a) Age (b) Gender

(c) Vitamin D (d) Regions

Figure 2.14: This figure shows the bar graph for Rickets with the different variables

2.4 Chapter Summary

Finally, we know that descriptive statistics just describes and summarizes data but do not allow us to draw conclusions about the whole population from which we took the sample. Descriptive statistics are simply describing what is or what the data shows. We will analyze the data in the next chapter. In short and based on the work we did in this chapter we can summarize the chapter as follow:

- More than three-quarters of patients are adults were adults (14 years old or above).
- One-third of individuals are male, and two-thirds are female.
- Slightly less than 50% of the data are for an average level of vitamin D, and then precisely the same number for insufficiency and deficiency.
- Levels of vitamin d ranges from 9 ng/ml to above 400 ng/ml, with most patients around 50 - 80 ng/ml.
- Patients from Muscat more than 40% of the total cases.
- Hypocalcemia is the most common disease, the other diseases come next in order, non-scaring hair loss, cardiomyopathy and rickets.
- Non-scarring hair loss affects more women than men, while dilated cardiomyopathy and rickets are effecting men and women similarly.
- Deficiency means a very low vitamin level, and people should get treatment to enhance their vitamin D levels. Inefficiency is also considered that the person is at risk of going down to deficiency. Therefore, it should be considered and must be improved.

Chapter 3

Methodologies and Inferential Statistics

3.1 Methodologies

In this chapter, we present the methodology that we used to further analyze our data. In the previous chapter, we used descriptive statistics to give us insight into our data. Whereas to understand the data more and to answer our project objectives we will apply different statistical techniques in this chapter. Starting with the chi-square test, moving to the logistic linear regression, and finalizing with simple linear regression.

3.1.1 Chi-square

The Chi-square test of independence is one of the most useful statistics for testing hypotheses when the variables are nominal. It is used to test if there is a relationship between two categorical variables. This test utilizes a contingency table to analyze the data. A contingency table is an arrangement in which data is classified according to two categorical variables. The categories for one variable appear in the rows, and the categories for other variables appear in columns. Each cell reflects the total count of cases for a specific pair of categories. (SPSS Tutorials: Chi-Square Test of Independence, n.d.).

Assumption:

1- Two categorical variables.

2- Two or more categories (groups) for each variable.

3- Independence of observations.

- There is no relationship between the subjects in each group.
- The categorical variables are not "paired" in any.

4- Relatively large sample size.

- Expected frequencies for each cell are at least 1.
- Expected frequencies should be at least 5- for the majority (80%) of the cells.

Hypothesis:

H_o: The response variable and each of the explanatory variables are not associated

H_1: There is an association between the two variables.

Test Statistics:

The test statistic for the Chi-Square Test of Independence is denoted, X , and is computed as:

$$\chi^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{(O_{ij}-e_{ij})^2}{e_{ij}} \tag{3.1}$$

where:

- Oij is the observed cell count in the ith row and jth column of the table.

- eij is the expected cell count in the ith row and jth column of the table computed as

$$e_{ij} = \frac{\text{row i total} \times \text{col j total}}{\text{grand total}} \tag{3.2}$$

The quantity $(O_{ij} - e_{ij})$ is sometimes referred to as the residual of the cell (i,j), denoted r_{ij}. The calculated χ^2 value is then compared to the critical value from the χ^2 distribution table with degrees of freedom $df = (R-1)\times(C-1)$ and chosen confidence level. If the calculated χ^2 values > critical χ^2 value, then we reject the null hypothesis. (SPSS Tutorials: Chi-Square Test of Independence, n.d.)

3.1.2 Binary Logistic Regression

Whenever you want to predict a categorical variable like binary data because it is the most common type of categorical data (Pass/Fail, Yes/No) based on the set of independent variables, we can use a classification algorithm called logistic regression. Thus, we denote Y as a binary response variable (Yes/No), P(Y=1) will be the model of yes or success probability, and P(Y=0) will be the model of no or failure probability. Furthermore, P(Y=1) is dependent on X which is an explanatory variable(s). As previously stated, logistic regression models are defined as =E(Y) by a linear predictor.

$$Logit(\mu) = \log(\frac{\mu}{1-\mu}) \tag{3.3}$$

$$Logit(\pi(x)) = \log(\frac{\pi(x)}{1-\pi(x)}) = \alpha + X^T\beta \tag{3.4}$$

Where:

- $\pi(x) =$ the odds of one category or group.
- $X^T =$ a vector of predictors
- $\alpha =$ the intercept or (constant)
- $\beta =$ the slope or (coefficients) of the explanatory variables

Remarks:

- Odds are $\pi(x) = \frac{p}{1-p} = e^{\alpha+\beta_i}$
- Odds ratio (OR) is $\frac{\pi(x)}{1-\pi(x)} = \frac{\frac{p_i}{1-p_i}}{\frac{p_o}{1-p_o}} = e^{\beta_i}$
- When the odds at $x+1$ equals the odds at xe^{β}. In other words, the odds of success changes by a factor of e^{β} for every unit increase in x.
- The change is a decrease if $e^{\beta} < 1(\beta < 0)$ and an increase if $e^{\beta} > 1(\beta > 0)$. The odds are independent of x if $\beta = 0$.

Hypothesis:

H_o: All β coefficients are = 0 means that there is no association between the explanatory and the response variables.

H_1: Assumes that β coefficients are not 0 means that there is an association between the explanatory and the response variables.

Interpret

When the P-value is less than or equal to the α level of significance, we reject H_o, and we conclude that there is a statistically significant association between the explanatory and the response variables.

3.1.3 Simple Linear Regression

A simple linear regression model is a model that consists of only one independent variable x that has a linear relationship with a response y, that is both variables x and y consists of continuous values. The model can be written as:

$$y = \beta_o + \beta_1 x + \varepsilon \tag{3.5}$$

where the intercept β_o and the slope β_1 are unknown constants, while ε is a random error term. This model allows us to investigate if there is a positive linear relationship or a negative linear relationship, or no relationship between the two variables.

3.2 Analysis

3.2.1 Chi-square

The chi-square test is applied when we have two categorical variables to test the association between them. Here we will test the association between the dependent variable (disease) and the independent variables (the risk factor).

Analysing Cardiomyopathy disease data

Cardiomyopathy is a response variable with a (yes /no) outcome. The table shows the association between having Cardiomyopathy and some independent risk factor variables.

Variables		Cardiomyopathy		
		Value of Pearson Chi-Square	P-Value	Decision
Gender	Female	49.591	<0.01	Reject Ho
	Male			
Age categorical	Adult	44.463	<0.01	Reject Ho
	Children			
Region	AD DAKHLIYAH	37.554	<0.01	Reject Ho
	ADH DHAHIRAH			
	AL BATINAH			
	AL SHARQIYA			
	DHOFAR			
	MUSANDAM			
	MUSCAT			
Vitamin D categorical	Deficiency	26.817	<0.01	Reject Ho
	Insufficiency			
	Normal			

Figure 3.1: Person Chi-square Test for Testing the Association between Cardiomyopathy Disease and Some Risk Factors.

The table(3.1) shows the association between having Cardiomyopathy and all independent risk factor variables. Depending on the information given it is very clear that there is an association detected between having Cardiomyopathy and all the risk factors. Therefore, we reject the null hypothesis and conclude that there is a statistically significant association between gender and Cardiomyopathy (P-value = <0.01), age group and Cardiomyopathy (P-value = <0.01), region and Cardiomyopathy (P-value = <0.01), and vitamin D category (P-value = <0.01).

Dilated Cardiomyopathy			
		Yes	No
Gender	Male	5.018	-2.87
	Female	-3.491	1.996
Age	Adult	2.292	-1.311
	Children	-5.315	3.04
Region	AD DAKHLIYAH	3.0674	-1.7542
	ADH DHAHIRAH	-0.656	0.3752
	AL BATINAH	-0.5916	0.3383
	AL SHARQIYA	-1.9832	1.1341
	DHOFAR	3.6356	-2.0791
	MUSANDAM	-0.4964	0.2839
	MUSCAT	-0.8445	0.483
Vitamin D	Deficiency	2.832	-1.619
	Insufficiency	1.265	-0.724
	Normal	-3.254	1.861

Figure 3.2: Standardized Residuals of Cardiomyopathy Disease with Gender, Region, Age and Vitamin D.

Based on the information presented in figure(3.2) above, the distribution of the residuals is mostly within 2. However, there are several cases where extreme residual values are observed. For instance, the table shows that the incidence of Cardiomyopathy among males is higher (5.018) than the expected value, while among females, it is lower (-3.491). Additionally, the number of children with Cardiomyopathy is lower (-5.315) than the observed value. also, AD DAKHLIYAH and DHOFAR patients have a higher incidence of Cardiomyopathy (3.0674 and 3.6356, respectively) than expected. Conversely, fewer patients with normal vitamin D levels have Cardiomyopathy (-3.254) compared to the expected value. Furthermore, it is worth noting that the observed number of children without Cardiomyopathy is more (3.04) than expected.

Analysing Hypocalcemia disease data

Hypocalcemia is a response variable with a (yes /no) outcome. The table shows the association between hypocalcemia and some independent risk factor variables.

Variables		Hypocalcemia		
		Value of Pearson Chi-Square	P-Value	Decision
Gender	F	2.58	0.108	Fail to reject Ho
	M			
Age categorical	Adult	0.428	0.513	Fail to reject Ho
	Children			
Region	AD DAKHLIYAH	5.233	0.514	Fail to reject Ho
	ADH DHAHIRAH			
	AL BATINAH			
	AL SHARQIYA			
	DHOFAR G			
	MUSANDAM			
	MUSCAT			
Vitamin D categorical	Deficiency	20.121	<0.01	Reject Ho
	Insufficiency			
	Normal			

Figure 3.3: Person Chi-square Test for Testing the Association between Hypocalcemia Disease and Some Risk Factors.

The above figure (3.3) shows the association between hypocalcemia and some independent risk factor variables. Depending on this information, it is very clear that there is no association detected between having Hypocalcemia and some risk factor. Therefore, we fail to reject the null hypothesis and conclude that there is no statistically significant association between gender and Hypocalcemia, age categorical and Hypocalcemia, and region and Hypocalcemia. The only variable that shows a significant association is vitamin D categories where (P-Value =<0.01), thus we can say there is a significant association between Hypocalcemia disease and vitamin D.

Hypocalcemia			
		Yes	No
Gender	Male	-7.551	4.984
	Female	5.252	-3.467
Age	Adult	3.019	-1.992
	Children	-7.002	4.621
Region	AD DAKHLIYAH	-0.3791	0.2502
	ADH DHAHIRAH	-1.4861	0.9808
	AL BATINAH	-0.1131	0.0746
	AL SHARQIYA	3.5203	-2.3234
	DHOFAR	-2.4635	1.6259
	MUSANDAM	1.2645	-0.8346
	MUSCAT	0.335	0.2211
Vitamin D	Deficiency	0.8594	-0.5672
	Insufficiency	1.4531	-0.959
	Normal	-1.834	1.2105

Figure 3.4: Standardized Residuals of Hypocalcemia Disease with Gender, Region, Age and Vitamin D.

Based on the information presented in figure(3.4) above, the distribution of the residuals is mostly within 2. However, there are several cases where extreme residual values are observed. For instance, the table shows that the incidence of Hypocalcemia among males is lower (-7.551) than the expected value. while among females, it is higher (5.252). Additionally, the number of children with Hypocalcemia is lower (-7.002) than the expected value. While among adults, it is higher (3.019). Also, AL SHARQIYA patients have a higher incidence of Hypocalcemia (3.5203) than the expected value. Furthermore, it is worth noting that the observed number of males without Hypocalcemia is more(4.984) than the expected value. And females without Hypocalcemia less (-3.467) Additionally, the number of children without Hypocalcemia is (4.621) more than expected.

Analysing Nonscarring Hair Loss disease data

Nonscarring Hair Loss is a response variable with a (yes /no) outcome. The table shows the association between having Nonscarring Hair Loss and some independent risk factor variables.

Variables		Nonscarring Hair Loss		
		Value of Pearson Chi-Square	P-Value	Decision
Gender	F	121.458	<0.01	Reject Ho
	M			
Age categorical	Adult	83.463	<0.01	Reject Ho
	Children			
Region	AD DAKHLIYAH	32.355	<0.01	Reject Ho
	ADH DHAHIRAH			
	AL BATINAH			
	AL SHARQIYA			
	DHOFAR			
	MUSANDAM			
	MUSCAT			
Vitamin D categorical	Deficiency	8.921	0.12	Fail to reject Ho
	Insufficiency			
	Normal			

Figure 3.5: Person Chi-square Test for Testing the Association between Nonscarring Hair Loss Disease and Some Risk Factors.

Figure (3.5) above shows the association between having Nonscarring Hair Loss and some independent risk factor variables. Depending on the information, it is very clear that there is an association detected between having Nonscarring Hair Loss and some risk factors. Therefore, we reject the null hypothesis and conclude that there is a statistically significant association between Gender (P-value = <0.01), age categorical (P-value = <0.01), and region (P-value = <0.01), and Nonscarring Hair Loss. However, we can not say there an association between Nonscaring hair loss and vitamin D categories because we fail to reject Ho. And conclude that there is no statistically significant association between vitamin D categorical and Nonscarring Hair Loss.

Nonscarring Hair Loss			
		Yes	No
Gender	Male	1.0986	-0.7294
	Female	-0.7642	0.5074
Age	Adult	-0.2159	0.1433
	Children	0.5007	-0.3324
Region	AD DAKHLIYAH	-0.7057	0.4686
	ADH DHAHIRAH	-0.4783	0.3176
	AL BATINAH	0.477	-0.3167
	AL SHARQIYA	-1.216	0.8073
	DHOFAR	0.3561	0.2365
	MUSANDAM	-0.5531	0.3672
	MUSCAT	0.8753	-0.5812
Vitamin D	Deficiency	-1.799	1.195
	Insufficiency	-1.716	1.139
	Normal	2.79	-1.852

Figure 3.6: Standardized Residuals of Nonscarring Hair Loss Disease with Gender, Region, Age and Vitamin D.

Based on the information presented in Figure (3.6) above, the distribution of the residuals is mostly within 2. However, there is only one case where a residual value above 2 has been observed. The table shows that the incidence of Nonscarring Hair Loss among the normal levels of vitamin D higher (2.79) than expected, so that means there are more patients suffering from Nonscarring Hair Loss with normal vitamin D levels than the expected value.

Analysing Rickets disease data

Rickets is a response variable with a (yes /no) outcome. The table shows the association between having Rickets and some independent risk factor variables.

variables		Rickets		
		Value of Pearson Chi-Square	P-Value	Decision
Gender	F	13.793	<0.01	Reject Ho
	M			
Age categorical	Adult	474.458	<0.01	Reject Ho
	Children			
Region	AD DAKHLIYAH	22.413	<0.01	Reject Ho
	ADH DHAHIRAH			
	AL BATINAH			
	AL SHARQIYA			
	DHOFAR			
	MUSANDAM			
	MUSCAT			
Vitamin D categorical	Deficiency	40.448	<0.01	Reject Ho
	Insufficiency			
	Normal			

Figure 3.7: Person Chi-square Test for Testing the Association between Rickets Disease and Some Risk Factors.

Figure (3.7) above shows the association between having Rickets and some independent risk factor variables. Depending on the information, it is very clear that there is an association detected between having Rickets and all the risk factors. Therefore, we reject the null hypothesis and conclude that there is a statistically significant association between gender (P-value = <0.01), age groups (P-value = <0.01), and vitamin D categories (P-value = <0.01), with Rickets. Also, we can say there is a region effect as well because we can reject H_o and conclude that there is a statistically significant association between region and Rickets.

Rickets			
		Yes	No
Gender	Male	-2.0093	0.6782
	Female	2.8887	-0.975
Age	Adult	-8.171	2.758
	Children	18.952	-6.396
Region	AD DAKHLIYAH	-1.5714	0.5304
	ADH DHAHIRAH	2.4359	-0.8222
	AL BATINAH	1.8555	-0.6263
	AL SHARQIYA	-2.4753	0.8354
	DHOFAR	-1.4301	0.4827
	MUSANDAM	-0.3198	0.1079
	MUSCAT	0.0135	-0.0046
Vitamin D	Deficiency	-2.593	0.875
	Insufficiency	-3.068	1.035
	Normal	4.492	-1.516

Figure 3.8: Standardized Residuals of Rickets Disease with Gender, Region, Age and Vitamin D.

Based on the information presented in the above figure(3.8), the distribution of the residuals is mostly within 2. However, there are some cases where extreme residual values are observed. For instance, the table shows that the observed number of Rickets among children is higher(18.952) than expected. This indicates there are more children with rickets observed compared to expected. additionally, the number of adults with rickets is lower(-8.171) than expected. Also, more patients with normal vitamin D levels have rickets (4.482) compared to the expected value.

Vitamin D and categorical risk factors

Although our main concern is to study the relation between the 4 diagnosis with the risk factors; vitamin D levels, age, gender, and region. We are also interested in investigating the relation between vitamin D levels and the other factors. Therefore, we applied chi-square test on vitamin D levels and gender, and vitamin D levels and region, since all these variables are categorical. Further, we will explore the relation between vitamin D levels as numeric and age as numeric in section 3.2.3.

variables		Vitamin D		
		Value of Pearson Chi-Square	P-Value	Decision
Gender	Female	2.985	0.225	Fail to reject Ho
	Male			
Region	AD DAKHLIYAH	17.316	0.138	Fail to reject Ho
	ADH DHAHIRAH			
	AL BATINAH			
	AL SHARQIYA			
	DHOFAR			
	MUSANDAM			
	MUSCAT			

Figure 3.9: Person Chi-square Test for Testing the Association between Rickets Disease and Some Risk Factors.

Figure (3.9) shows the results of applying chi-square for vitamin D and two independent risk factors which are gender and region. Based on the information, it is very clear that there is no association detected between having vitamin D and these risk factors. Therefore, we fail to reject the null hypothesis and conclude that there is no statistically significant association between gender, region, and Vitamin D.

3.2.2 Binary Logistic Regression

In this section, we will study the association between the dependent variable and the group of independent variables. Where we can assume the following model for all the four diagnosis in this study, which will represent the response variable.

The full logistic model: $Logit[P(Y=1)] = \alpha + \beta_1 G + \beta_2 R + \beta_3 A + \beta_4 D + \beta_5 RD$

Where,

α = The intercept

β = Risk factors

G= Gender

R= Region

A= Age

RD = result of vitamin D

Analysing Cardiomyopathy disease data

In this part We study the relationship between cardiomyopathy and some risk factors. We fitted a logistic regression model because the response variable is binary. First, when fitting the full model we looked at the omnibus tests of model coefficients to test if there is at least one difference from zero. This test is sometimes referred to as the regression significance test. The hypothesis form of this test is:

Ho: all equal to zero.

Ha: There is at least one difference from zero.

As we can see from the Table below, there is a significant effect on the coefficients of the cardiomyopathy model. So, we can reject Ho which means there is at least one difference from zero. In other words, we can say that at least one of the predictor variables is affecting the response variable.

	Chi-square	df	Sig.
Step	165.500	10	0.000
Block	165.500	10	0.000
Model	165.500	10	0.000

Table 3.1: Omnibus Table of Cardiomyopathy

The full logistic model is:

Logit [P(Y=1)] $= -3.732 + (1.218)G + (0.524)R_1 + (0.076)R_2 + (0.019)R_3 - (0.623)R_4 + (1.792)R_5 + (-20.084)R_6 + (1.905)A + (.708)RD + (0.449)RD_2$

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I for EXP(B)	
							Lower	Upper
Gender								
Male	1.218	0.150	65.673	1	0.000	3.382	2.519	4.541
Female						1		
Region								
AD DAKHLIYAH	0.524	0.199	6.952	1	0.008	1.689	1.144	2.493
ADH DHAHIRAH	0.076	0.276	0.076	1	0.783	1.079	0.629	1.852
AL BATINAH	0.019	0.195	0.009	1	0.923	1.019	0.696	1.492
AL SHARQIYA	-0.623	0.281	4.928	1	0.026	0.536	0.310	0.930
DHOFAR	1.792	0.498	12.927	1	0.000	6.001	2.259	15.940
MUSANDAM	-20.084	40192.970	0.000	1	1.000	0.000	0.000	
MUSCAT						1		
Age group								
Adult	1.905	0.343	30.794	1	0.000	6.720	3.429	13.171
Children						1.000		
Vitamin D								
Deficiency	0.708	0.175	16.303	1	0.000	2.030	1.440	2.863
Insuficiency	0.449	0.180	6.204	1	0.013	1.567	1.101	2.232
Normal						1		
Constant								
Constant	-3.732	0.358	108.958	1	0.000	0.024		

Figure 3.10: Full Model of Cardiomyopathy Disease vs. Some Factors.

The table above presents the multiple logistic regression analysis of cardiomyopathy, showing the effects of factors on the occurrence of cardiomyopathy. The results indicate gender, region, age group, and Vitamin D as significant predictors of having cardiomyopathy rather than normal. For instance, gender is a significant effect to have cardiomyopathy rather than normal, the odds of having cardiomyopathy rather than normal for males were found to have 3.382 times more likely to have cardiomyopathy rather than normal for females. (Odds ratio = 3.382, 95% CI: (2.519, 4.541), P-value = 0.000). Regions appeared as a significant predictor of having cardiomyopathy rather than normal, for respondents from Ad dakhliyah, Al sharqiya, and Dhofar was significant to have cardiomyopathy rather than normal compared to Muscat. For example, the odds of having cardiomyopathy rather than normal for those who live in Ad dakhliyah is 1.689 times more likely to have cardiomyopathy rather than normal for those who live in Muscat. The age of the respondent shows a significant positive association with having cardiomyopathy rather than normal, the odds of having cardiomyopathy rather than normal for adult people is 6.720 times more likely to have cardiomyopathy rather than normal for children. (OR=6.720, 95% C: (3.429, 13.171), P-value=0.000).

Vitamin D shows a significant association with cardiomyopathy rather than normal. For example, the odds of having cardiomyopathy rather than normal for patients with a deficiency in vitamin D is 2.030 times more likely to have cardiomyopathy rather than normal for patients with normal vitamin D levels. (Odds ratio= 2.030, 95% CI:(1.440,2.863), P-value = 0.000).

Analyzing Hypocalcemia Disease Data

Again multiple logistic regression is used to study the relationship between having hypocalcemia and some risk factors. We started by looking at omnibus tests that is provided by the SPSS. The test is previously described. For the sake of the presentation, we will omit some of the discussions provided in the previous section.

	Chi-square	df	Sig.
Step	27.269	10	0.000
Block	27.269	10	0.000
Model	27.269	10	0.000

Table 3.2: Omnibus Table of hypocalcemia

From the Table, we can reject the hypothesis of omnibus tests. Then, there is at least one effect on hypocalcemia disease. Thus, at least one of the risk factors is important for predicting the response variable.

The full logistic model is:

Logit [P(Y=1)] $= -0.579 + (0.198)G + (-0.239)R_1 + (-0.176)R_2 + (-0.050)R_3 + (-0.392)R_4 + (0.214)R_5 + (-20.159)R_6 + (0.112)A + (-0.577)RD + (-0.559)RD_2$

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I for EXP(B)	
							Lower	Upper
Gender								
Male	0.198	0.135	2.132	1	0.144	1.219	0.935	1.589
Female						1		
Region								
AD DAKHLIYAH	-0.239	0.194	1.528	1	0.216	0.787	0.538	1.151
ADH DHAHIRAH	-0.176	0.239	0.546	1	0.460	0.838	0.525	1.339
AL BATINAH	-0.050	0.165	0.090	1	0.764	0.952	0.688	1.316
AL SHARQIYA	-0.392	0.230	2.886	1	0.089	0.676	0.430	1.062
DHOFAR	0.214	0.484	0.195	1	0.659	1.238	0.479	3.200
MUSANDAM	-20.159	40192.970	0.000	1	1.000	0.000	0.000	
MUSCAT						1		
Age group								
Adult	0.112	0.181	0.383	1	0.536	1.119	0.784	1.596
Children						1		
Vitamin D								
Deficiency	-0.577	0.158	13.268	1	0.000	0.562	0.412	0.766
Insuficiency	-0.559	0.159	12.402	1	0.000	0.572	0.419	0.780
Normal						1		
Constant								
Constant	-0.579	0.191	9.199	1	0.002	0.560		

Figure 3.11: Full Model of Hypocalcemia Disease vs. Some Factors.

The above table presents the multiple logistic regression analysis of hypocalcemia, showing the effects of factors on the occurrence of hypocalcemia. The results indicate Vitamin D as a significant predictor of having hypocalcemia rather than normal. Gender has no significant effect to have hypocalcemia rather than normal, (Odds ratio = 1.219, 95% CI: (0.935, 1.589), P-value = 0.144). Regions appeared as a not significant predictor of having hypocalcemia rather than normal compared to Muscat. The age of the respondent shows a not significant positive association with having hypocalcemia rather than normal. Vitamin D levels shows a significant association with hypocalcemia rather than normal. For example, the odds of having hypocalcemia rather than normal for people with deficient vitamin D levels is 0.438 times less likely to have hypocalcemia rather than normal for people with normal vitamin D, (Odds ratio = 0.562, 95% CI: (0.412, 0.766), P-value = 0.000). This means that for people with vitamin D deficiency, they are less likely to have Hypocalcemia than normal people, the same results hold for vitamin D insufficiency.

Analyzing Nonscarring Hair Loss Disease Data

Again multiple logistic regression is used to study the relationship between having Nonscarring Hair Loss and some risk factors. We started by looking at omnibus tests that are provided by the SPSS. The test is previously described. For the sake of the presentation, we will omit some of the discussions provided in the previous section.

	Chi-square	df	Sig.
Step	301.630	10	0.000
Block	301.630	10	0.000
Model	301.630	10	0.000

Table 3.3: Omnibus Table of Nonscarring Hair Loss

From the Table, we can reject the hypothesis of omnibus tests. Then, there is at least one effect on Nonscarring Hair Loss disease. Thus, at least one of the risk factors is important for predicting the response variable.

The full logistic model is:

Logit [P(Y=1)] $= -3.677 + (-2.044)G + (-0.128)R_1 + (-0.450)R_2 + (0.414)R_3 + (1.178)R_4 + (-20.670)R_5 + (21.463)R_6 + (3.325)A + (0.092)RD + (0.254)RD_2$

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Gender								
Male	-2.044	0.200	103.966	1	0.000	0.130	0.087	0.192
Female						1		
Region								
AD DAKHLIYAH	-0.128	0.205	0.391	1	0.532	0.879	0.588	1.315
ADH DHAHIRAH	-0.450	0.271	2.743	1	0.098	0.638	0.375	1.086
AL BATINAH	0.414	0.191	4.703	1	0.030	1.513	1.041	2.201
AL SHARQIYA	1.178	0.244	23.330	1	0.000	3.249	2.014	5.240
DHOFAR	-20.670	8649.824	0.000	1	0.998	0.000	0.000	
MUSANDAM	21.463	40192.969	0.000	1	1.000	209.7	0.000	
MUSCAT						1		
Age group								
Adult	3.325	0.519	40.982	1	0.000	27.799	10.044	76.939
Children						1		
Vitamin D								
Deficiency	0.092	0.172	0.284	1	0.594	1.096	0.783	1.535
Insuficiency	0.254	0.173	2.162	1	0.141	1.290	0.919	1.811
Normal						1		
Constant								
Constant	-3.677	0.528	48.548	1	0.000	0.025		

Figure 3.12: Full Model of Nonscarring Hair Loss Disease vs. Some Factors.

The table above presents the multiple logistic regression analysis of Nonscarring Hair Loss, showing the effects of factors on the occurrence of Nonscarring Hair Loss. The results indicate gender, region, and age group as significant predictors of having Nonscarring Hair Loss rather than normal. Gender is a significant effect to have Nonscarring Hair Loss rather than normal, the odds of having Nonscarring Hair Loss rather than normal for males were found to have 0.130 compared to the odds of having Nonscarring Hair Loss rather than normal for females. (Odds ratio = 0.130, 95% CI: (0.087, 0.192), P-value = 0). Regions appeared as a significant predictor of having Nonscarring Hair Loss rather than normal compared to Muscat. for respondents from Al batinah, Al sharqiya, was significant to have Nonscarring Hair Loss rather than normal. For example, the odds of having Nonscarring Hair Loss rather than normal for those who live in Al batinah is 1.513 times more likely to have Nonscarring Hair Loss rather than normal for those who live in Muscat. The age of the respondent shows a significant positive association with having Nonscarring Hair Loss rather than normal, the odds of having Nonscarring Hair Loss rather than normal for adult people is 27.799 times more likely to have Nonscarring Hair Loss rather than normal for children, (OR=27.799, 95% C: (10.044, 76.939)).

Vitamin D is found as a not significant effect for having Nonscarring Hair Loss rather than normal.

Analyzing Rickets Disease Data

Again multiple logistic regression is used to study the relationship between having Rickets and some risk factors. We started by looking at omnibus tests that are provided by the SPSS. The test is previously described. For the sake of the presentation, we will omit some of the discussions provided in the previous section.

	Chi-square	df	Sig.
Step	359.874	10	0.000
Block	359.874	10	0.000
Model	359.874	10	0.000

Table 3.4: Omnibus Table of Rickets

From the Table, we can reject the hypothesis of omnibus tests. Then, there is at least one effect on Rickets disease. Thus, at least one of the risk factors is important for predicting the response variable. The full logistic model is:

Logit [P(Y=1)] $= 0.139 + (0.533)G + (0.750)R_1 + (0.861)R_2 + (-0.202)R_3 + (-1.526)R_4 + (-17.031)R_5 + (-16.910)R_6 + (-4.170)A + (-0.261)RD + (-0.473)RD_2$

	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Gender								
Male	0.533	0.265	4.052	1	0.044	1.705	1.014	2.865
Female						1		
Region								
AD DAKHLIYAH	0.750	0.441	2.894	1	0.089	2.117	0.892	5.023
ADH DHAHIRAH	0.861	0.431	4.000	1	0.046	2.366	1.017	5.502
AL BATINAH	-0.202	0.312	0.420	1	0.517	0.817	0.443	1.506
AL SHARQIYA	-1.526	0.605	6.355	1	0.012	0.217	0.066	0.712
DHOFAR	-17.031	8947.437	0.000	1	0.998	0.000	0.000	
MUSANDAM	-16.910	40192.970	0.000	1	1.000	0.000	0.000	
MUSCAT						1		
Age group								
Adult	-4.170	0.309	181.579	1	0.000	0.015	0.008	0.028
Children						1		
Vitamin D								
Deficiency	-0.261	0.346	0.570	1	0.450	0.770	0.391	1.517
Insuficiency	-0.473	0.358	1.748	1	0.186	0.623	0.309	1.256
Normal						1		
Constant								
Constant	0.139	0.276	0.251	1	0.616	1.149		

Figure 3.13: Full Model of Rickets Disease vs. Some Factors.

The above table presents the multiple logistic regression analysis of Rickets, showing the effects of factors on the occurrence of Rickets. The results indicate gender, region, and age group as significant predictors of having Rickets rather than normal. For gender factors, gender was a significant effect to have Rickets rather than normal, the odds of having Rickets rather than normal for males were found to have 1.705 times more likely to have Rickets rather than normal for females, (Odds ratio = 1.705, 95% CI: (1.014, 2.865), P-value = 0.044). For region factors, regions appeared as a significant predictor of having Rickets rather than normal, for respondents from Adh dhahirah, Al sharqiya was significant to have Rickets rather than normal. For example, the odds of having Rickets rather than normal for those who live in Adh dhahirah is 2.366 times more likely to have Rickets rather than normal for those who live in Muscat. The age of the respondent shows a significant positive association with having Rickets rather than normal, the odds of having Rickets rather than normal for adult people is 0.015 times less likely to have Rickets rather than normal for children, (OR=0.015, 95% C: (0.008, 0.028), P-value=0.000). whereas vitamin D shows a not significant association with Rickets rather than normal.

3.2.3 Simple Linear Regression

As discussed previously that one of our objectives is to see if there any association between age and vitamin D. Using the available numeric values of these two variables, we built a simple linear regression model where age is the expalnatory variable, while vitamin D is the response. The results of our model are given as follow:

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error			
(Constant)	73.698	2.351		31.342	0.000
Age	-0.064	0.055	-0.034	-1.165	0.244

Figure 3.14: Linear Regression between age and vitamin D

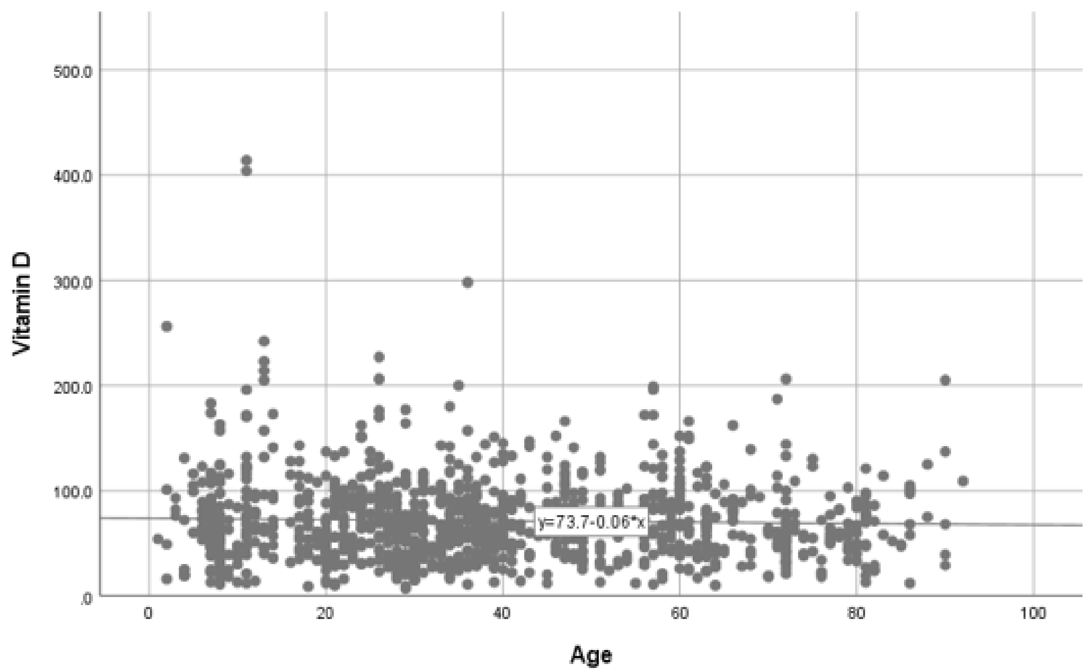


Figure 3.15: Linear Regression between age and vitamin D

The unstandardized coefficient for age is -0.064, which means that for every one-unit increase in age, there is a decrease of 0.064 units in vitamin D, holding all other variables constant. However, the standardized coefficient for age is -0.034, which measures vitamin D's effect on age after scaling both variables by their standard deviations. This standardized coefficient indicates that a one standard deviation increase in vitamin D is associated with a 0.034 standard deviation decrease in age.

The t-value for age is -1.165, which is the ratio of the coefficient to its standard error. This value indicates that the coefficient is not significantly different from zero, as the significance level (p-value) is 0.244, which is greater than the commonly used threshold of 0.05. Therefore, we cannot conclude that there is a statistically significant linear relationship between vitamin D and age in this regression model. Moreover, looking at Figure (3.15) we can see that the vitamin D values are flat along the age axis.

3.3 Chapter Summary

From this chapter, we can see that each type of data requires a different analysis method. In our case, we used the chi-square, binary logistic regression, and simple linear regression methods. Overall, all diseases were modeled with common risk factors, including gender, age, vitamin D, and region. However, the analysis revealed that the effect of these risk factors on each disease varies.

For instance, gender was found to be significant in all diseases except hypocalcemia, where it did not have an effect. Vitamin D was the only significant risk factor in hypocalcemia, while age was significant in Cardiomyopathy, Nonscarring Hair Loss, and Rickets. Therefore, when assessing a patient's risk of developing these conditions or developing treatment plans, their age should be taken into account.

On the other side, vitamin D was not significant in hair loss, which was unexpected. We also investigated the impact of region on disease prevalence but found that it was not a significant factor in any of the diseases we studied. Therefore, it is clear from our data that regions do not have a significant effect on the odds of developing these diseases.

Further, we studies the effect of gender, region and age on vitamin D levels, but found that all the three factors has no significant association with vitamin D.

Chapter 4

Conclusion and Recommendation

Since we started our final year project, we were interested in data that consists of vitamin levels in the body and their effect on the person's health. Therefore, we looked for data from the SQU hospital in Oman. After discussing the topic with the source of data, we found that it is more convenient to handle data about vitamin D, as they have recommended us. Thus, we became interested in exploring "the most common diagnosis that relates to vitamin D levels in Oman". The reason behind choosing this topic is that low level of vitamin D is considered a public health issue that causes serious health problems. Therefore, we also sought to know other factors that could affect the incidence of getting these diagnoses.

4.1 Conclusion

In this chapter, we summarize the work that we have done in the previous chapters and display the recommendation and possible future improvements. In chapter one, we discussed generally vitamin D related diseases and their risk factors and we presented the main objective of this project which is to study the relationship between four diseases related to vitamin D with some of the risk factors that we pointed out in the chapter. Also, we summarized a range of results from some studies that focus on vitamin D factors, and related diseases.

Before starting chapter two, we explored the data of our project and we did missing values cleaning, deleted all duplicates, added new variables, and rephrased the names of the regions. In chapter two, we discussed the data background then we summarized the variables of the data numerically and graphically to get some insights about our data.

Chapter three is the core of our study because it includes methodologies and the application of these methodologies that were used to analyze the data. In this chapter, we used mainly two statistical methods to analyze our data. These methods are the chi-square test for an independent test of a relationship between two categorical variables and the binary logistic regression model to explain the relationship between one dependent binary variable (diagnosis) and vitamin D levels along with other factors. Based on our analysis, we can conclude the following about each disease of interest in our study.

Cardiomyopathy

it is clear that there is a significant association between having cardiomyopathy and several risk factors, including gender, age, region, and vitamin D levels. Specifically, our research revealed that more of the participants with cardiomyopathy were female, indicating that gender is a potential risk factor for developing this condition. Additionally, our findings suggest that age and region may also be important risk factors for cardiomyopathy.

Hypocalcemia

Although higher rates of having Hypocalcemia have been always linked to vitamin D deficiency, in our data we could not show that. In fact, it turned out to be such that higher rates of Hypocalcemia are associated with people who maintain normal vitamin D levels. There could be different reasons behind this conclusion, however, it would need more investigations. It is worth mentioning that, Hypocalcemia has a significant association with calcium levels in the body, while vitamin D plays a critical role in calcium homeostasis.

In addition, we found that there is no significant association between hypocalcemia and gender, age, or region. This indicates that vitamin D deficiency may be the primary risk factor for hypocalcemia and that other factors such as age, gender, and region may not be as significant in contributing to this condition.

Nonscarring Hair Loss

Based on our research, we did not find any significant association between Vitamin D and Nonscarring Hair Loss disease as a risk factor. However, we did observe significant associations between Nonscarring Hair Loss disease and all other factors we examined. Notably, gender had a high impact on having Nonscarring Hair Loss disease, with more affected being females. our study suggests that Vitamin D may not play a significant role in the development of Nonscarring Hair Loss disease. Nevertheless, our findings highlight the importance of considering other factors, especially gender, in identifying individuals at risk for this condition. Future studies may explore additional risk factors and further investigate the role of gender, age, and interaction of gender and age in Nonscarring Hair Loss disease.

Rickets

Based on our research, we found significant associations between having Rickets and all risk factors examined, including gender, age, region, and vitamin D status. Our findings suggest that individuals who are male, or children are at higher risk for developing Rickets. Notably, our study revealed that more of the individuals affected by Rickets were children. This observation underscores the importance of monitoring vitamin D status and other risk factors among adults, as well as children, to prevent and manage Rickets. our study highlights the significant role of various risk factors in the development of Rickets. Our findings provide important insights into the population groups that may be most vulnerable to this condition and underscore the need for public health interventions to improve vitamin D status and address other risk factors.

4.2 Recommendation

Our project shows that the deficiency of vitamin D has a significant effect on the incidence of being diagnosed with Cardiomyopathy, which matches the results from a range of studies in this area.

Although our study did not find a significant association between vitamin D and the other diagnosis, it might be still indirectly linked to the disease. For instance, vitamin D has a positive relationship with the calcium level in the body, less calcium means more likely to have Hypocalcemia.

Therefore, it is very important to maintain appropriate levels of vitamin D, the government should increase its focus on specialized health education programs that can influence lifestyle choices, particularly with regard to sufficient exposure to sunlight and consuming foods rich in vitamins.

One issue we encountered in our study was the high number of inaccurate data points, such as blank cells and repetitions in some patients. To enhance the understanding of vitamin D levels in Oman, future research should consider more detailed factors, such as the duration of time patients spend in sunlight, which is the primary source of vitamin D, as well as the weight of the patients, which can affect the absorption and metabolism of vitamin D in the body.

To address the issue of inaccurate data, we recommend that the Ministry of Health in Oman implement more rigorous quality control measures when entering data into the system. This will help ensure that the data is reliable and can be used effectively for research purposes. By addressing these issues, future research on vitamin D levels in Oman can provide more accurate and detailed insights into the status of vitamin D deficiency in the population. This, in turn, will enable policymakers and healthcare professionals to implement more effective interventions and improve the health outcomes of individuals in Oman.

Bibliography

Al-Kindi, M. K. (2011). Vitamin d status in healthy omani women of childbearing age: study of female staff at the royal hospital, muscat, oman. *Sultan Qaboos University Medical Journal 11*(1), 56.

Balasubramanian, S., S. Shivbalan, and P. S. Kumar (2006). Hypocalcemia due to vitamin d deficiency in exclusively breastfed infants. *Indian pediatrics 43*(3), 247.

De Laet, C., J. Kanis, A. Od&39;en, H. Johanson, O. Johnell, P. Delmas, J. Eisman, H. Kroger, S. Fujiwara, P. Garnero, et al. (2005). Body mass index as a predictor of fracture risk: a meta-analysis. *Osteoporosis international 16*(11), 1330–1338.

Holick, M. F. (1996). Vitamin d and bone health. *The Journal of nutrition 126*(4), 1159S–1164S.

Ives, R. and M. Brickley (2014). New findings in the identification of adult vitamin d deficiency osteomalacia: results from a large-scale study. *International Journal of Paleopathology 7*, 45–56.

Pearce, S. H. and T. D. Cheetham (2010). Diagnosis and management of vitamin d deficiency. *Bmj 340*.

Polat, V., E. Bozcali, T. Uygun, S. Opan, and O. Karakaya (2015). Low vitamin d status associated with dilated cardiomyopathy. *International journal of clinical and experimental medicine 8*(1), 1356.

Rao, D. S., A. Parfitt, A. Villanueva, P. Dorman, and M. Kleerekoper (1987). Hypophosphatemic osteomalacia and adult fanconi syndrome due to light-chain nephropathy: another form of oncogenous osteomalacia. *The American journal of medicine 82*(2), 333–338.

Vegesna, V., J. O'Kelly, M. Uskokovic, J. Said, N. Lemp, T. Saitoh, T. Ikezoe, L. Binderup, and H. P. Koeffler (2002). Vitamin d3 analogs stimulate hair growth in nude mice. *Endocrinology 143*(11), 4389–4396.

Wacker, M. and M. F. Holick (2013). Sunlight and vitamin d: A global perspective for health. *Dermato-endocrinology 5*(1), 51–108.