STATISTICAL ANALYSIS ON DATA OF THYROID, VITAMIN D & B12, BONE MINERAL DENSITY TESTS

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

FACULTY OF SCIENCE

DEPARTMENT OF STATISTICS

Guide: Prof. Rupal Shah

Submitted By: Avanika Bhati
Bhumika Karia
Vishruti Karia
Vishwajeet Jadeja

CERTIFICATE

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

FACULTY OF SCIENCE

DEPARTMENT OF STATISTICS

This is to certify that "Avanika Bhati, Vishruti Karia, Bhumika Karia, Vishwajeet Jadeja" have completed the project Entitled "Statistical Analysis on data of Thyroid, Vitamin D, Vitamin B12 and Bone Mineral Density Tests" As a team in academic year 2019-2020 and submitted the work to the department of statistics in partial fulfilment for the degree of masters in science with statistics. Throughout the semester the work has been done with sincerity and Enthusiasm.

Prof. Vipul Kalamkar
(Head of Department)

Prof. Rupal Shah
(Guide)

ACKNOWLEDGMENT

We would like to express our deepest appreciation to the Faculty of Statistics, Maharaja Sayajirao University, for providing us the facility to complete this project on 'Statistical Analysis on data of Thyroid, Vitamin B12 & D and bone mineral density tests'. It not only taught us to apply our theoretical knowledge to use but the research that was undertaken during this period also helped us recognize new things which we're really thankful for.

A special gratitude I give to our respected Professors for approving our topic of research, without their kind and encouraging words this Project wouldn't have been completed.

We are extremely privileged to mention, that the success and the Final outcome of this project required a lot of guidance and assistance of our guide Prof. Rupal Shah. All that we have done and learned under her supervision cannot go unnoticed.

A special vote of thanks goes to our batchmates of MSc Final class (2020), who never hesitated in giving suggestions and absolutely needed criticism, when asked for.

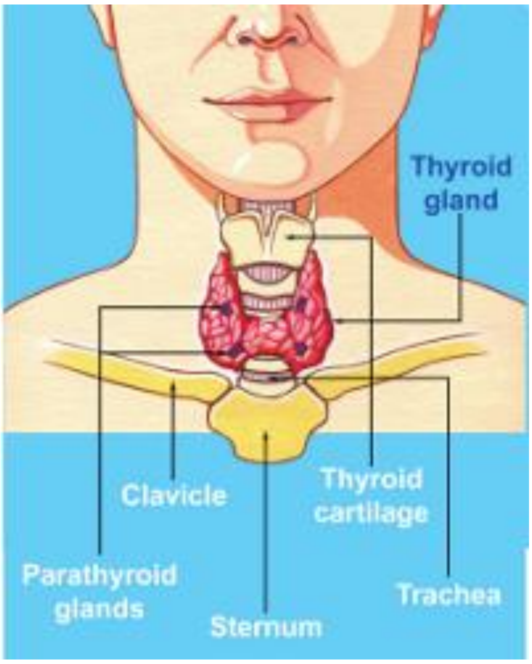
At last thank you everyone for investing their precious time in helping us with this project.

CONTENTS

INTRODUCTION

What is Thyroid?

- A large ductless gland in the neck which secretes hormones regulating growth and development through the rate of metabolism.
- The thyroid is a butterfly-shaped gland that sits low on the front of the neck. Your thyroid lies below your Adam's apple, along the front of the windpipe. The thyroid has two side lobes, connected by a bridge (isthmus) in the middle. When the thyroid is its normal size, you can't feel it.
- Brownish-red in colour, the thyroid is rich with blood vessels. Nerves important for voice quality also pass through the thyroid.
- The thyroid secretes several hormones, collectively called thyroid hormones. The main hormone is thyroxin, also called T4.
- The thyroid secretes several hormones, collectively called thyroid hormones. The main hormone is thyroxin, also called T4.



Types of Thyroid:

1. Hypothyroidism

Hypothyroidism results from the thyroid gland producing an insufficient amount of thyroid hormone. It can develop from problems within the thyroid gland, pituitary gland, or hypothalamus.

2. Hyperthyroidism

Hyperthyroidism is a condition of the thyroid. The thyroid is a small, butterfly-shaped gland located at the front of your neck. It produces tetraiodothyronine (T4) and triiodothyronine (T3), which are two primary hormones that control how your cells use energy. Your thyroid gland regulates your metabolism through the release of these hormones.

WHAT IS VITAMIN B12?

- Vitamin B12, also known as cobalamin, is an important water-soluble vitamin.
- It plays an essential role in the production of your red blood cells and DNA, as well as the proper functioning of your nervous system.
- Vitamin B12 is naturally found in animal foods, including meats, fish, poultry, eggs and dairy.

WHAT IS VITAMIN D?

- Vitamin D is a fat-soluble vitamin that is naturally present in very few foods, added to others, and available as a dietary supplement.
- It is also produced endogenously when ultraviolet rays from sunlight strike the skin and trigger vitamin D synthesis.
- Vitamin D promotes calcium absorption in the gut and maintains adequate serum calcium and phosphate concentrations to enable normal mineralization of bone and to prevent hypocalcaemia tetany.
- It is also needed for bone growth and bone remodelling by osteoblasts and osteoclasts without sufficient vitamin D, bones can become thin, brittle, or misshapen.
- Together with calcium, vitamin D also helps protect older adults from osteoporosis.

WHAT IS BONE MINERAL DENSITY?

- A measure of the amount of minerals (mostly calcium and phosphorous) contained in a certain volume of bone.
- Bone mineral density measurements are used to diagnose osteoporosis (a condition marked by decreased bone mass), to see how well osteoporosis treatments are working, and to predict how likely the bones are to break.
- Low bone mineral density can occur in patients treated for cancer. Also called BMD, bone density, and bone mass.

BMD TEST

- A bone mineral density test uses X-rays to measure the amount of minerals namely calcium in your bones. This test is important for people who are at risk for osteoporosis, especially women and older adults.

- The test is also referred to as dual energy X-ray absorptiometry (DXA). It’s an important test for osteoporosis, which is the most common type of bone disease. Osteoporosis causes your bone tissue to become thin and frail over time and leads to disabling fractures.

OBJECTIVES

- To identify that which factor is most responsible for Thyroid.

- Relationships between Eating Habits with Vitamin -B12.

- Association of Thyroid with Age, Gender, Vitamin -B12.

- Association of BMD with Age and Gender.

DATA

Type of data: Primary data

Location: MSU health centre

Method: Questionnaire

- There was a camp organised by MSU health centre.
- These camps are organised thrice in a year.
- The patients in these were mostly the campus student, teachers, staff members of university and also some retired professors.
- We have collected the data of the Dec 2019 camp.
- The camp was held on 27 Dec 2019.
- Here is the poster of the camp describing the test which are conducted and their prices.

University Health Centre
The Maharaja Sayajirao University of Baroda
Opp Rosary School, Pratapgunj Vadodara – 390 002.Gujarat. India.

Tel: (+91 0265)2791616

CAMP

↓

(A) BLOOD TEST

(i) Thyroid test

- TSH.................. Rs. 70/-
- T4.................... Rs. 70/-
- T3.................... Rs. 70/-

(ii) Serum B12 = Rs 280/-

(iii) Serum Vitamin-D = Rs 500/-

(B) BMD (BONE MINERAL DENSITY)

→ Rs 10/-

Date:-
27/12/2019
(Friday)

Time: - *Blood Test* (9:00 to 12:00 PM)
BMD (9:00 to 2:00 PM)

Camp Venue: - University Health Centre
Opp. Rosary School,
Pratapgunj, Vadodara

Register in Room No.

થાઈરોઈડ નિદાન કેમ્પ

- જે દર્દીને આ કેમ્પમાં નિદાન કરાવવાની ઇચ્છા હોય, તેમને પોતાના નામ વહેલી તકે નોંધાવી લેવા.
- નીચે જણાવેલ તકલીફો કે ચિન્હોવાળા દર્દીને (સ્ત્રી/પુરુષ) થાઈરોઈડ ની બિમારી હોવાની શકયતા છે.

☛ હાયપોથાઈરોડીઝમ :

- વજનમાં વધારો
- અનિયમિત અથવા ભારે પ્રમાણમાં માસિક, બાળકો ન થવા.
- ખરબચડા, સૂકાવાળ, વાળ ઉતરવા
- અવાજ ઘોંઘરો થવો.
- શરીરમાં સોજો થવો
- અશકિત લાગવી.

☛ હાયપરથાઈરોડીઝમ :

- આંખોના ડોળા બહાર આવવા
- ધ્રુજારી કે કંપનની તકલીફ
- ગભરામણ થવી
- વજનમાં ઘટાડો થવો
- વધુ પડતી ગરમી લાગવી.
- ગોઈટર (ગરદન પર સોજો)

☛ કેમ્પનું સ્થળ :-

એમ. એસ. યુનિવર્સીટી ડીસ્પેન્સરી
રોઝરી સ્કુલની સામે,
પ્રતાપગંજ,
વડોદરા.

તારીખ :-
૨૭-૧૨-૨૦૧૯,
શુક્રવાર

સમય :-
સવારે
૯-૦૦ થી ૧૨-૦૦

☛ થાઈરોઈડ ટેસ્ટ TSH રાહત દરે Rs. 70/- (બજાર ભાવ રૂા. ૨૫૦/-) માં અને વિટામીન B12 ફકત રૂા. ૨૮૦/- માં કરી આપવામાં આવશે. વિટામીન D ફકત રૂા. ૫૦૦/- માં કરી આપવામાં આવશે.

QUESTIONNAIRE

Medical Questionnaire for Thyroid, Vitamin B12, Vitamin D and Bone Mineral Density

Name:	**Height:**	**Occupation:**
Age:	**Weight:**	**Mobile No.:**
Gender:	**Blood Group:**	**Address:**

Vegetarian	**Non - Vegetarian**	**Eggetarian**	**Vegan**

Q1.	Any significant change in Weight?	**Yes**	**No**
	If Yes,	**Gain**	**Loss**
Q2.	Any Swelling in the Neck?	**Yes**	**No**
Q3.	Felt any significant changes in Heart Rate?	**Yes**	**No**
	If Yes,	**Slower**	**Faster**
Q4.	Do you have nerve cramps?	**Yes**	**No**
Q5.	Do you suffer from Hypertension?	**Yes**	**No**
Q6.	Feeling changes in Body Temperature around you?	**Yes**	**No**
	If Yes,	**Cold**	**Hot**
Q7.	Your Skin getting Dry?	**Yes**	**No**
Q8.	Your Nails getting Brittle?	**Yes**	**No**
Q9.	Feeling any Numbness or Tingling in Hands?	**Yes**	**No**
Q10.	Having Constipation problems?	**Yes**	**No**
Q11.	Having problems in Normal Vision?	**Yes**	**No**
Q12.	Having Diarrhoea?	**Yes**	**No**
Q13.	Observed Hair Loss?	**Yes**	**No**
Q14.	Feeling Weakness in Muscles?	**Yes**	**No**
Q15.	Do you feel any changes in your Appetite?	**Yes**	**No**
Q16.	Do you get Sick or Infected Often?	**Yes**	**No**
Q17.	Feeling Fatigue and Tiredness?	**Yes**	**No**
Q18.	Getting Depressed easily?	**Yes**	**No**
Q19.	Do you have Impaired Wound Healing?	**Yes**	**No**
Q20.	Having Back Pain?	**Yes**	**No**
Q21.	Having problems like depression, memory loss, or behavioural changes?	**Yes**	**No**

Q22.	Feeling tiredness, or light-headedness?	**Yes**	**No**
Q23.	Having shortness of breath?	**Yes**	**No**
Q24.	Skin Paling?	**Yes**	**No**
Q25.	Having smooth tongue?	**Yes**	**No**
Q26.	Felt any significant Loss of height over time?	**Yes**	**No**
Q27.	Your body posture is Stooped?	**Yes**	**No**
Q28.	Your bones break much easily than expected?	**Yes**	**No**
Q29.	You generally consume Junk Food?	**Yes**	**No**
Q30.	Do you take any Supplements?	**Yes**	**No**
Q31.	Do you suffer from Pain?	**Yes**	**No**
	If Yes,	**Chest pain**	**Muscle pain**
		Joint pain	**Ribs pain**
Q32.	Do you consume alcohol?	**Yes**	**No**
Q33.	Types of work	**Office**	**Household**
		Outdoor	
Q34.	Do you take RO water?	**Yes**	**No**
Q35.	Taking any kind of drugs?	**Yes**	**No**
	If Yes, Name		
Q36.	For Females, excessive blood loss during menstruation.	**Yes**	**No**
Q37.	Any history of surgery?	**Yes**	**No**
	If Yes, Name		
Q38.	Any Family History of Thyroid?	**Yes**	**No**
	If yes, then taking any kind of medicine? Specify Name.		

DATA DESCRIPTION

- Some questions are based on symptoms and rest are based on the individual's habits and their daily schedules.
- There are total 38 questions.

Questions	Description
1 to 16	Related to thyroid symptoms.
17 to 20	Related to vitamin D.
20 to 25	Related to vitamin B12.
26 to 28	Related to Bone Mineral Density.
29 to 38	Related to day to day routine of an individual.

BELOW PICTURES ARE THE SAMPLES OF THE COLLECTED DATA

- 1ST PICTURE SHOWS THE PERSONAL INFORMATION OF AN INDIVIDUAL

A	B	C	D	E	F	G	H	I	J	K
Name	**Age**	**Age Group**	**Gender**	**Height**	**Weight**	**Blood Group**	**Occupation**	**Mobile No.**	**Address**	**Eating Habit**
Ruchi Jain	38	31 to 50	Female	161	105	B+	Household	8000985673		Vegetarian
Manish Jain	38	31 to 50	Male	182	108	O-	Service	9328908932		Vegetarian
Nisha Jagani	19	11 to 30	Female	158	48	B+		7878179879		Eggetarian
Pooja	20	11 to 30	Female	161	54	A+		9723019172	Fatehgunj	Vegetarian
Neha	25	11 to 30	Female	125	51	A+	Teaching Asst.	9638267266	MSU Quarter	Non - Vegetarian
Lizy	51	51 to 70	Female	160	69.4	B+	Professor	7405841294	Fatehgunj	Non - Vegetarian

BELOW THREE PICTURES SHOWS CODED DATA

L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
Q1		**Q2**	**Q3**		**Q4**	**Q5**	**Q6**		**Q7**	**Q8**	**Q9**	**Q10**	**Q11**	**Q12**	**Q13**	**Q14**
1	Gain	1	1	Faster	1	1	0		0	0	1	1	0	0	1	1
0		0	0		1	1	1	Hot	0	0	0	0	0	0	1	1
0		0	0		1	0	0		1	1	1	1	1	0	1	0
0		0	0		0	0	0		0	0	0		1	0	0	1
1	Loss	0	0		0	1	0		0	1	0	0	0	0	0	1

AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
Q15	**Q16**	**Q17**	**Q18**	**Q19**	**Q20**	**Q21**	**Q22**	**Q23**	**Q24**	**Q25**	**Q26**	**Q27**	**Q28**	**Q29**	**Q30**
1	1	1	1	0	1	1	1	1	0	0	0	1	1	1	0
	0	1	1		1	0	1	0	0	0	0	0	0	1	0
0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	0
0	0	1	1	0	1	0	0	0	0	0	0	0	1	1	
0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0

AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD
Q31		**Q32**	**Q33**	**Q34**	**Q35**		**Q36**	**Q37**		**Q38**	
1	J	0	Household	0	1	n 35mg/Macp	0	0		0	
0		0	Office	0	0		0	1		0	
1	M	0		0	0		1	0		0	
0		0	Household	1	0		0			0	
1	M	0	Outdoor	1	0		0	0		0	

BELOW PICTURE SHOWS THE RESULTS OF TSH, VIT B12, VIT D & BMD TEST OF AN INDIVIDUAL

BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR
TSH	**T3**	**T4**	**Vitamin B12**	**Vitamin D**	**BMD**			**TSH**	**T3**	**T4**	**B12**	**D**	**BMD**
1.482			291		0.33	0.72		0			0		Normal
2.47			177	9.97	-1.46	-1.23		0			**1**	**Deficient**	**Osteopenia**
			292								0		
1.688								0					
2.447	1.06	7.9						0	0	0			
2.88			534	6.59				0			0	**Deficient**	

DATABASE

Here we have created the database of health center data

1st picture shows the tables:

1. Profile
2. Questions
3. Reports

```
Enter password: *****
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 17
Server version: 8.0.17 MySQL Community Server - GPL

Copyright (c) 2000, 2019, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use health_camp
Database changed
mysql> show tables;
+-----------------------+
| Tables_in_health_camp |
+-----------------------+
| profile               |
| questions             |
| reports               |
+-----------------------+
3 rows in set (0.02 sec)

mysql>
```

```
mysql> select * from profile;
```

P_Id	Name	Age	Gender	Mobile_No	Eating_Habits
P001	Ruchi Jain	38	Female	8000985673	Vegetarian
P002	Manish Jain	38	Male	9328908932	Vegetarian
P003	Nisha Jagani	19	Female	7878179879	Eggetarian
P004	Pooja	20	Female	9723019172	Vegetarian
P005	Neha	25	Female	9638267266	Non - Vegetarian
P006	Lizy	51	Female	7405841294	Non - Vegetarian
P007	Dia J.S. Sheth	15	Female	9426324794	Vegetarian
P008	Masha J.S. Sheth	16	Female	9426324794	Eggetarian
P009	Anit Thomas	46	Female	8469888985	Vegetarian
P010	D P Chattopadhyay	60	Male	NULL	Vegetarian
P011	Madhuri Pandey	55	Female	8128058426	Vegetarian
P012	Savitri Chellani	54	Female	9898376029	Non - Vegetarian
P013	Patil Nilesh	40	Male	9904076289	Vegetarian
P014	Bhavika	37	Female	9429948499	Vegetarian
P015	S.B. Saxena	82	Male	9427857201	Non - Vegetarian
P016	Jyotsna Pathak	65	Female	9913599381	Vegetarian
P017	Patel Shivani	25	Female	9978458669	Vegetarian
P018	Suman Saxena	73	Female	9427857201	Non - Vegetarian
P019	G A Parmar	49	Male	9898085152	Non - Vegetarian
P020	Ajay Roy	41	Male	9427282504	Non - Vegetarian
P021	J R Shah	83	Male	9979345551	Vegetarian
P022	Dhanuska Roy	11	Female	9427282504	Non - Vegetarian
P023	Vinay Singh	46	Male	9427453261	Non - Vegetarian
P024	Lata Shanmungan	47	Female	7778090333	Vegan
P025	Kishan Chandran	78	Male	78340514028	Non - Vegetarian
P026	Padmaben Kishan	72	Female	78340514028	Non - Vegetarian
P027	Yogendra Pathak	66	Male	9913599387	Vegetarian
P028	G Venkat	57	Male	NULL	Vegetarian
P029	P K Jha	53	Male	9825032877	Non - Vegetarian
P030	Prof D K Basa	73	Male	9099944274	Non - Vegetarian
P031	Jayant S Pathak	78	Male	9428820177	Vegetarian
P032	Dr Sheetal Sheth	49	Female	9426324794	Vegetarian
P033	Jagdish solanki	58	Male	9824174901	Vegetarian
P034	Sunil Ratndir	50	Male	9825711781	Vegetarian
P035	tara mehta	79	Female	NULL	Vegetarian
P036	Sapna	44	Female	9824985394	Non - Vegetarian
P037	Dr kishan	34	Male	9016649497	Vegetarian
P038	Tejal	39	Female	8180836177	Vegetarian
P039	Sapun	34	Male	9898263371	Vegetarian
P040	Janak Thakkur	29	Male	9510408298	Vegetarian
P041	Bhumika karia	23	Female	8490055364	Vegetarian
P042	Vishwajeet	22	Male	7066889971	Vegetarian
P043	Vishruti karia	22	Female	7574971273	Vegetarian
P044	Avanika bhati	23	Female	7383956193	Non - Vegetarian
P045	Gaurav dirvedy	20	Male	7016245312	Vegetarian
P046	Kruti shah	31	Female	9727204689	Vegetarian

```
mysql> select * from questions;
```

P_Id	R_Id	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38
P001	R001	1	1	1	1	1	0	0	0	1	1	0	0	1	1	1	1	1	0	1	0	Household	0	1	0	0	0
P002	R002	0	0	0	1	1	1	0	0	0	0	0	0	1	1	NULL	0	1	0	0	0	Office	0	0	0	1	0
P003	R003	0	0	0	1	0	0	1	1	1	1	1	0	1	0	0	1	1	0	1	0	NULL	0	0	1	0	0
P004	R004	0	0	0	0	0	0	0	0	0	NULL	1	0	0	1	0	0	1	NULL	0	0	Household	1	0	0	NULL	0
P005	R005	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	Outdoor	1	0	0	0	0
P006	R006	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	Office	1	0	0	0	0
P007	R007	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	NULL	0	0	1	0	0
P008	R008	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	NULL	0	0	1	0	0
P009	R009	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Household	0	0	0	0	0
P010	R010	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Office	1	1	NULL	0	0
P011	R011	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	Household	1	0	1	0	0
P012	R012	1	0	1	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0	1	0	Office	0	1	0	0	0
P013	R013	0	0	1	1	0	0	1	1	0	1	1	0	1	1	0	0	0	0	0	0	Office	1	0	0	1	1
P014	R014	0	0	0	1	0	0	1	1	1	0	1	0	1	1	1	0	0	0	1	0	Office	1	0	0	0	0
P015	R015	1	0	0	1	1	1	0	1	0	1	1	0	0	1	0	0	0	1	0	0	NULL	0	0	0	0	0
P016	R016	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	Household	1	0	0	0	0
P017	R017	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Office	1	0	0	0	0
P018	R018	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0	1	0	0	NULL	0	NULL	0	0	0
P019	R019	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Office	1	0	NULL	0	0
P020	R020	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	Office	1	0	NULL	0	0
P021	R021	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	1	1	0	Office	0	0	NULL	1	0
P022	R022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	Outdoor	1	0	NULL	0	0
P023	R023	1	0	1	0	0	0	1	1	1	1	1	0	1	1	0	1	0	1	1	0	Office	1	0	NULL	1	0
P024	R024	1	1	1	1	0	NULL	1	1	1	1	NULL	0	NULL	1	1	1	0	NULL	1	0	Household	0	NULL	0	1	0
P025	R025	0	0	1	1	0	0	1	0	0	1	1	0	1	1	1	0	0	1	1	0	Household	0	0	NULL	0	0
P026	R026	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	1	0	Household	1	0	0	0	0
P027	R027	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	Household	1	0	NULL	0	0
P028	R028	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	Office	1	0	NULL	0	0
P029	R029	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Office	1	1	0	0	0
P030	R030	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	1	0	Household	1	1	NULL	1	0
P031	R031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Household	NULL	NULL	NULL	NULL	NULL
P032	R032	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	Outdoor	0	0	0	1	0
P033	R033	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	Office	1	1	0	0	0
P034	R034	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	Office	1	0	0	1	0
P035	R035	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	Office	0	1	0	0	0
P036	R036	0	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	Office	0	0	0	1	1
P037	R037	0	0	0	1	0	0	1	0	1	0	0	0	1	1	0	0	0	0	1	0	Office	1	0	NULL	1	1
P038	R038	1	0	1	0	1	0	1	1	0	0	NULL	1	1	0	0	NULL	0	1	0	0	Office	1	1	1	1	0
P039	R039	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Office	1	0	0	0	0
P040	R040	0	0	1	0	1	1	1	0	1	0	1	0	0	0	1	1	NULL	0	0	0	Office	1	0	NULL	0	0
P041	R041	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	Outdoor	1	0	0	0	1
P042	R042	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	Outdoor	1	0	NULL	0	0
P043	R043	1	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	0	Outdoor	1	1	1	0	0
P044	R044	1	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	Outdoor	1	0	0	1	0
P045	R045	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	NULL	0	0	NULL	0	0
P046	R046	1	0	0	0	0	0	0	1	1	1	1	0	1	1	0	0	0	0	1	0	Office	0	0	0	0	1

```
mysql> select * from reports;
```

R_Id	TSH	B12	D	BMD	R_TSH	R_B12	R_D	R_BMD
R001	1.482	291	NULL	0.330	0	0	NULL	Normal
R002	2.470	177	9.970	-1.460	0	1	Deficient	Osteopenia
R003	NULL	292	NULL	NULL	NULL	0	NULL	NULL
R004	1.688	NULL	NULL	NULL	0	NULL	NULL	NULL
R005	2.447	NULL	NULL	NULL	0	NULL	NULL	NULL
R006	2.880	534	6.590	NULL	0	0	Deficient	NULL
R007	NULL	NULL	19.260	-0.090	NULL	NULL	Insufficient	Normal
R008	NULL	NULL	30.460	-0.520	NULL	NULL	Sufficient	Normal
R009	2.613	NULL	NULL	NULL	0	NULL	NULL	NULL
R010	6.854	NULL	NULL	NULL	1	NULL	NULL	NULL
R011	NULL	516	15.520	NULL	NULL	0	Insufficient	NULL
R012	NULL	NULL	NULL	-1.760	NULL	NULL	NULL	Osteopenia
R013	NULL	NULL	NULL	-1.940	NULL	NULL	NULL	Osteopenia
R014	NULL	NULL	NULL	0.330	NULL	NULL	NULL	Normal
R015	NULL	727	17.540	NULL	NULL	0	Insufficient	NULL
R016	9.372	1757	22.920	NULL	1	1	Insufficient	NULL
R017	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
R018	1.474	763	28.840	-1.820	0	0	Insufficient	Osteopenia
R019	5.324	162	11.520	-2.030	1	1	Insufficient	Osteopenia
R020	NULL	310	7.460	NULL	NULL	0	Deficient	NULL
R021	NULL	NULL	NULL	-0.480	NULL	NULL	NULL	Normal
R022	1.465	268	11.890	NULL	0	0	Insufficient	NULL
R023	1.898	706	12.820	NULL	0	0	Insufficient	NULL
R024	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
R025	NULL	NULL	13.870	NULL	NULL	NULL	Insufficient	NULL
R026	NULL	NULL	14.720	NULL	NULL	NULL	Insufficient	NULL
R027	4.256	820	13.560	-1.260	0	NULL	NULL	NULL
R028	1.758	651	23.060	-0.970	0	0	Insufficient	Normal
R029	NULL	634	24.240	NULL	NULL	0	Insufficient	NULL
R030	2.320	NULL	NULL	-1.230	0	NULL	NULL	Osteopenia
R031	NULL	614	NULL	-0.690	NULL	0	NULL	Normal
R032	2.326	NULL	NULL	-0.550	0	NULL	NULL	Normal
R033	NULL	NULL	NULL	-1.600	NULL	NULL	NULL	Osteopenia
R034	NULL	NULL	NULL	-1.600	NULL	NULL	NULL	Osteopenia
R035	NULL	NULL	29.920	NULL	NULL	NULL	Insufficient	NULL
R036	2.216	678	NULL	NULL	0	0	NULL	NULL
R037	NULL	NULL	NULL	-0.450	NULL	NULL	NULL	Normal
R038	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL
R039	NULL	NULL	NULL	-0.660	NULL	NULL	NULL	Normal
R040	NULL	NULL	NULL	-0.690	NULL	NULL	NULL	Normal
R041	NULL	NULL	NULL	-0.390	NULL	NULL	NULL	Normal
R042	NULL	NULL	NULL	0.060	NULL	NULL	NULL	Normal
R043	NULL	471	NULL	2.090	NULL	0	NULL	Normal
R044	NULL	NULL	NULL	0.090	NULL	NULL	NULL	Normal
R045	NULL	NULL	NULL	2.710	NULL	NULL	NULL	Normal
R046	2.181	491	27.740	-0.300	0	0	Insufficient	Normal

```
mysql> desc profile;
```

Field	Type	Null	Key	Default	Extra
P_Id	char(4)	NO	PRI	NULL	
Name	varchar(50)	YES		NULL	
Age	int(2)	YES		NULL	
Gender	enum('Male','Female')	YES		NULL	
Mobile_No	varchar(20)	YES		NULL	
Eating_Habits	varchar(20)	YES		NULL	

6 rows in set (0.02 sec)

```
mysql> desc questions;
```

Field	Type	Null	Key	Default	Extra
P_Id	char(4)	NO	MUL	NULL	
R_Id	char(4)	NO	PRI	NULL	
Q1	int(1)	YES		NULL	
Q2	int(1)	YES		NULL	
Q3	int(1)	YES		NULL	
Q4	int(1)	YES		NULL	
Q5	int(1)	YES		NULL	
Q6	int(1)	YES		NULL	
Q7	int(1)	YES		NULL	
Q8	int(1)	YES		NULL	
Q9	int(1)	YES		NULL	
Q10	int(1)	YES		NULL	
Q11	int(1)	YES		NULL	
Q12	int(1)	YES		NULL	
Q13	int(1)	YES		NULL	
Q14	int(1)	YES		NULL	
Q15	int(1)	YES		NULL	
Q16	int(1)	YES		NULL	
Q29	int(1)	YES		NULL	
Q30	int(1)	YES		NULL	
Q31	int(1)	YES		NULL	
Q32	int(1)	YES		NULL	
Q33	varchar(15)	YES		NULL	
Q34	int(1)	YES		NULL	
Q35	int(1)	YES		NULL	
Q36	int(1)	YES		NULL	
Q37	int(1)	YES		NULL	
Q38	int(1)	YES		NULL	

28 rows in set (0.00 sec)

MySQL 8.0 Command Line Client

```
mysql> desc reports;
+-------+-------------+------+-----+---------+-------+
| Field | Type        | Null | Key | Default | Extra |
+-------+-------------+------+-----+---------+-------+
| R_Id  | char(4)     | NO   | PRI | NULL    |       |
| TSH   | float(5,3)  | YES  |     | NULL    |       |
| B12   | int(4)      | YES  |     | NULL    |       |
| D     | float(5,3)  | YES  |     | NULL    |       |
| BMD   | float(5,3)  | YES  |     | NULL    |       |
| R_TSH | varchar(1)  | YES  |     | NULL    |       |
| R_B12 | varchar(1)  | YES  |     | NULL    |       |
| R_D   | varchar(20) | YES  |     | NULL    |       |
| R_BMD | varchar(20) | YES  |     | NULL    |       |
+-------+-------------+------+-----+---------+-------+
9 rows in set (0.00 sec)

mysql>
```

The purpose of creating this database is;

It is possible for users to create, edit and update data in database files. Once created, it possible to store and retrieve data from those database files.

Joining Python with Database
(TSH, Vitamin B12 & D, BMD)

Thyroid TSH Analysis

In [377]:
```
import MySQLdb
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import seaborn as sns
```

In [378]:
```
conn = MySQLdb.connect(host = 'localhost', user = 'root', passwd = 'mysql')
```

In [379]:
```
cursor = conn.cursor()
```

In [380]:
```
cursor.execute('use health_camp')
```

Out[380]: 0

In [381]:
```
cursor.execute('select age, eating_habits, gender, Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q29, Q30, Q
TSH = pd.DataFrame(list(cursor.fetchall()), columns = ['Age', 'Eating Habits', 'Gender', 'Q1', 'Q2', 'Q3', 'Q4', 'Q5', 'Q6', 'Q7', 'Q
TSH.head()
```

Out[381]:

	Age	Eating Habits	Gender	Q1	Q2	Q3	Q4	Q5	Q6	Q7	...	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	R_TSH	R_B12
0	38	Vegetarian	Female	1	1	1	1	1	0	0	...	1	0	Household	0	1.0	0.0	0.0	0	0	0
1	38	Vegetarian	Male	0	0	0	1	1	1	0	...	0	0	Office	0	0.0	0.0	1.0	0	0	1
2	20	Vegetarian	Female	0	0	0	0	0	0	0	...	0	0	Household	1	0.0	0.0	NaN	0	0	None
3	25	Non - Vegetarian	Female	1	0	0	0	1	0	0	...	1	0	Outdoor	1	0.0	0.0	0.0	0	0	None
4	51	Non - Vegetarian	Female	0	0	0	0	1	0	0	...	1	0	Office	1	0.0	0.0	0.0	0	0	0

5 rows × 31 columns

Vitamin B12 Analysis

In [1]:
```
import MySQLdb
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

In [2]:
```
conn = MySQLdb.connect(host = 'localhost', user = 'root', passwd = 'mysql')
```

In [3]:
```
cursor = conn.cursor()
```

In [4]:
```
cursor.execute('use health_camp')
```

Out[4]: 0

In [5]:
```
cursor.execute('select age, eating_habits, gender, Q29, Q30, Q31, Q32, Q33, Q34, Q35, Q36, Q37, Q38, R_TSH, R_B12 from profile, quest
B12 = pd.DataFrame(list(cursor.fetchall()), columns = ['Age', 'Eating Habits', 'Gender', 'Q29', 'Q30', 'Q31', 'Q32', 'Q33', 'Q34', 'Q
B12.head()
```

Out[5]:

	Age	Eating Habits	Gender	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	R_TSH	R_B12
0	38	Vegetarian	Female	1	0.0	1	0	Household	0	1.0	0.0	0.0	0	0	0
1	38	Vegetarian	Male	1	0.0	0	0	Office	0	0.0	0.0	1.0	0	0	1
2	20	Vegetarian	Female	1	NaN	0	0	Household	1	0.0	0.0	NaN	0	0	None
3	25	Non - Vegetarian	Female	1	0.0	1	0	Outdoor	1	0.0	0.0	0.0	0	0	None
4	51	Non - Vegetarian	Female	0	0.0	1	0	Office	1	0.0	0.0	0.0	0	0	0

Vitamin D Analysis

In [1]:
```
import MySQLdb
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

In [2]:
```
conn = MySQLdb.connect(host = 'localhost', user = 'root', passwd = 'mysql')
```

In [3]:
```
cursor = conn.cursor()
```

In [4]:
```
cursor.execute('use health_camp')
```

Out[4]: 0

In [5]:
```
cursor.execute('select age, eating_habits, gender, Q29, Q30, Q31, Q32, Q33, Q34, Q35, Q36, Q37, Q38, R_D, R_BMD from profile, questio
D = pd.DataFrame(list(cursor.fetchall()), columns = ['Age', 'Eating Habits', 'Gender', 'Q29', 'Q30', 'Q31', 'Q32', 'Q33', 'Q34', 'Q35
D.head()
```

Out[5]:

	Age	Eating Habits	Gender	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	R_D	R_BMD
0	38	Vegetarian	Male	1	0	0	0	Office	0	0.0	0.0	1	0.0	Deficient	Osteopenia
1	51	Non - Vegetarian	Female	0	0	1	0	Office	1	0.0	0.0	0	0.0	Deficient	None
2	15	Vegetarian	Female	1	0	1	0	None	0	0.0	1.0	0	0.0	Insufficient	Normal
3	16	Eggetarian	Female	1	0	1	0	None	0	0.0	1.0	0	0.0	Sufficient	Normal
4	55	Vegetarian	Female	0	1	1	0	Household	1	0.0	1.0	0	0.0	Insufficient	None

Bone Mineral Density Analysis

In [90]:
```
import MySQLdb
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

In [91]:
```
conn = MySQLdb.connect(host = 'localhost', user = 'root', passwd = 'mysql')
```

In [92]:
```
cursor = conn.cursor()
```

In [93]:
```
cursor.execute('use health_camp')
```

Out[93]: 0

In [94]:
```
cursor.execute('select age, eating_habits, gender, Q29, Q30, Q31, Q32, Q33, Q34, Q35, Q36, Q37, Q38, R_BMD from profile, questions, r
BMD = pd.DataFrame(list(cursor.fetchall()), columns = ['Age', 'Eating Habits', 'Gender', 'Q29', 'Q30', 'Q31', 'Q32', 'Q33', 'Q34', 'Q
BMD.head()
```

Out[94]:

	Age	Eating Habits	Gender	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	R_BMD
0	38	Vegetarian	Male	1	0	0	0	Office	0	0.0	0.0	1	0.0	Osteopenia
1	51	Non - Vegetarian	Female	0	0	1	0	Office	1	0.0	0.0	0	0.0	None
2	15	Vegetarian	Female	1	0	1	0	None	0	0.0	1.0	0	0.0	Normal
3	16	Eggetarian	Female	1	0	1	0	None	0	0.0	1.0	0	0.0	Normal
4	55	Vegetarian	Female	0	1	1	0	Household	1	0.0	1.0	0	0.0	None

DATA VISUALISATION

Vitamin B12 vs Gender

- **Here, 0 indicates Vitamin B12 Sufficient and 1 indicates Vitamin B12 Deficient**

In [11]:
```
pivot1 = pd.pivot_table(B12, values='Count', index=['R_B12'], columns=['Gender'], aggfunc=np.sum)
pivot1
```

Out[11]:

Gender	Female	Male
R_B12		
0	19	9
1	4	5

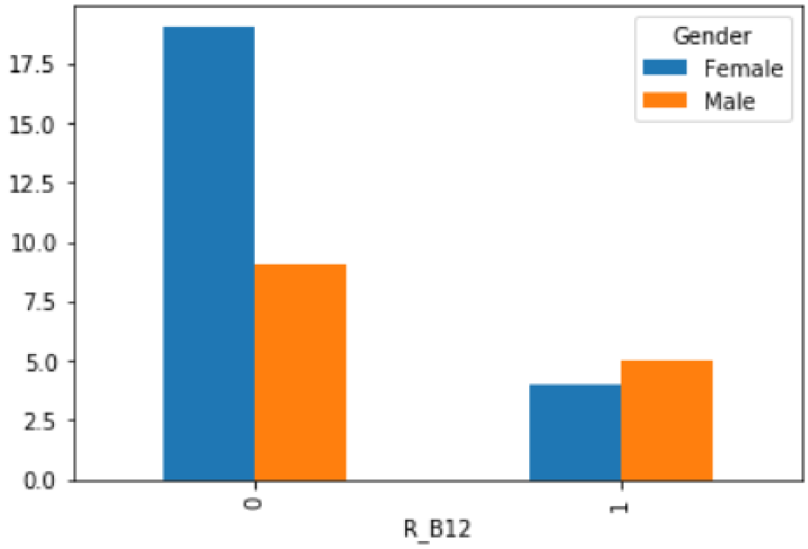
In [14]:
```
Total_B12 = [pivot1.iloc[0, :] + pivot1.iloc[1, :]]
Total_B12 = pd.DataFrame(Total_B12, index = ['Total B12'])
piv = pivot1.append(Total_B12)
Total_Gender = pivot1['Female'] + pivot1['Male']
piv['Total Gender'] = Total_Gender
Percentage_0 = [(piv.iloc[0, :]/piv.iloc[2, :])*100]
Percentage_1 = [(piv.iloc[1, :]/piv.iloc[2, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
Percentage_R0 = (piv['Female']/piv['Total Gender'])*100
Percentage_R1 = (piv['Male']/piv['Total Gender'])*100
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
piv['Percentage Female'] = Percentage_R0
piv['Percentage Male'] = Percentage_R1
piv
```

Out[14]:

Gender	Female	Male	Total Gender	Percentage Female	Percentage Male
0	19.000000	9.000000	28.0	67.857143	32.142857
1	4.000000	5.000000	9.0	44.444444	55.555556
Total B12	23.000000	14.000000	NaN	NaN	NaN
Percentage 0	82.608696	64.285714	NaN	NaN	NaN
Percentage 1	17.391304	35.714286	NaN	NaN	NaN

In [13]:
```
pivot1.plot(kind='bar')
```

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x181c7a03d30>



- **The above Bar Diagram represents the count of patients who have done Vitamin B12 test.**
- **As per data we can conclude that 44.44% Females are Vitamin B12 Deficient and 55.56% Males are Vitamin B12 Deficient.**
- **Also, Vitamin B12 Deficiency Percentage in Females is 17.4% and in Males is 35.7%.**
- **Hence, Males have more Vitamin B12 Deficiency than Females.**

BMD vs Age Group

Here we divided age groups in 4 groups as shown in table

In [120]:
```
Age_Group = []
for i in range(len(BMD['Age'])):
    if BMD.iloc[i, 0] <= 30:
        Age_Group.append('11 to 30')
    elif BMD.iloc[i, 0] <= 50:
        Age_Group.append('31 to 50')
    elif BMD.iloc[i, 0] <= 70:
        Age_Group.append('51 to 70')
    else:
        Age_Group.append('More than 70')
```

In [121]:
```
Age_Group = pd.DataFrame([Age_Group]).T
Age_Group.head()
```

Out[121]:

	0
0	31 to 50
1	51 to 70
2	11 to 30
3	11 to 30
4	51 to 70

In [124]:
```
pivot6 = pd.pivot_table(BMD, values='Count', index=['R_BMD'], columns=['Age_Group'], aggfunc=np.sum)
pivot6.fillna(0, inplace = True)
pivot6
```

Out[124]:

Age_Group R_BMD	11 to 30	31 to 50	51 to 70	More than 70
Normal	2.0	5.0	6.0	1.0
Osteopenia	0.0	2.0	5.0	1.0
Osteoporotic	0.0	0.0	1.0	0.0

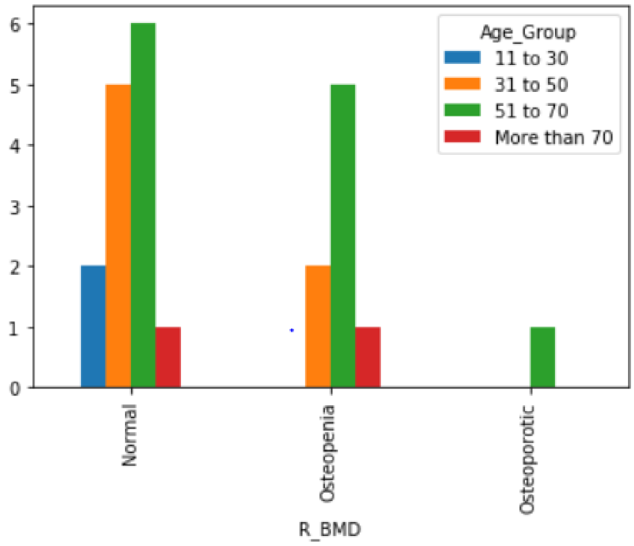
In [127]:
```
Total_BMD = [pivot6.iloc[0, :] + pivot6.iloc[1, :] + pivot6.iloc[2, :]]
Total_BMD = pd.DataFrame(Total_BMD, index = ['Total BMD'])
piv = pivot6.append(Total_BMD)
Total_AG = pivot6['11 to 30'] + pivot6['31 to 50'] + pivot6['51 to 70'] + pivot6['More than 70']
piv['Total Age Groups'] = Total_AG
Percentage_Norm = [(piv.iloc[0, :]/piv.iloc[3, :])*100]
Percentage_Ostpn = [(piv.iloc[1, :]/piv.iloc[3, :])*100]
Percentage_Ostpr = [(piv.iloc[2, :]/piv.iloc[3, :])*100]
Percentage_Norm = pd.DataFrame(Percentage_Norm, index = ['Percentage Normal'])
Percentage_Ostpn = pd.DataFrame(Percentage_Ostpn, index = ['Percentage Osteopenia'])
Percentage_Ostpr = pd.DataFrame(Percentage_Ostpr, index = ['Percentage Osteoporotic'])
Percentage_R1 = (piv['11 to 30']/piv['Total Age Groups'])*100
Percentage_R2 = (piv['31 to 50']/piv['Total Age Groups'])*100
Percentage_R3 = (piv['51 to 70']/piv['Total Age Groups'])*100
Percentage_R4 = (piv['More than 70']/piv['Total Age Groups'])*100
piv = piv.append(Percentage_Norm)
piv = piv.append(Percentage_Ostpn)
piv = piv.append(Percentage_Ostpr)
piv['Percentage 11 to 30'] = Percentage_R1
piv['Percentage 31 to 50'] = Percentage_R2
piv['Percentage 51 to 70'] = Percentage_R3
piv['Percentage More than 70'] = Percentage_R4
piv
```

Out[127]:

Age_Group	11 to 30	31 to 50	51 to 70	More than 70	Total Age Groups	Percentage 11 to 30	Percentage 31 to 50	Percentage 51 to 70	Percentage More than 70
Normal	2.0	5.000000	6.000000	1.0	14.0	14.285714	35.714286	42.857143	7.142857
Osteopenia	0.0	2.000000	5.000000	1.0	8.0	0.000000	25.000000	62.500000	12.500000
Osteoporotic	0.0	0.000000	1.000000	0.0	1.0	0.000000	0.000000	100.000000	0.000000
Total BMD	2.0	7.000000	12.000000	2.0	NaN	NaN	NaN	NaN	NaN
Percentage Normal	100.0	71.428571	50.000000	50.0	NaN	NaN	NaN	NaN	NaN
Percentage Osteopenia	0.0	28.571429	41.666667	50.0	NaN	NaN	NaN	NaN	NaN
Percentage Osteoporotic	0.0	0.000000	8.333333	0.0	NaN	NaN	NaN	NaN	NaN

```
In [126]: pivot6.plot(kind='bar')
```

```
Out[126]: <matplotlib.axes._subplots.AxesSubplot at 0x1dca9e5a278>
```



- **The above Bar Diagram represents the counts of the patients of different age groups who had done BMD test.**
- **As per data we can conclude that as the age increases, the risk of osteopenia increases.**

BMD vs Gender

In [100]:
```
pivot1 = pd.pivot_table(BMD, values='Count', index=['R_BMD'], columns=['Gender'], aggfunc=np.sum)
pivot1.fillna(0, inplace = True)
pivot1
```

Out[100]:

Gender	Female	Male
R_BMD		
Normal	10.0	4.0
Osteopenia	3.0	5.0
Osteoporotic	1.0	0.0

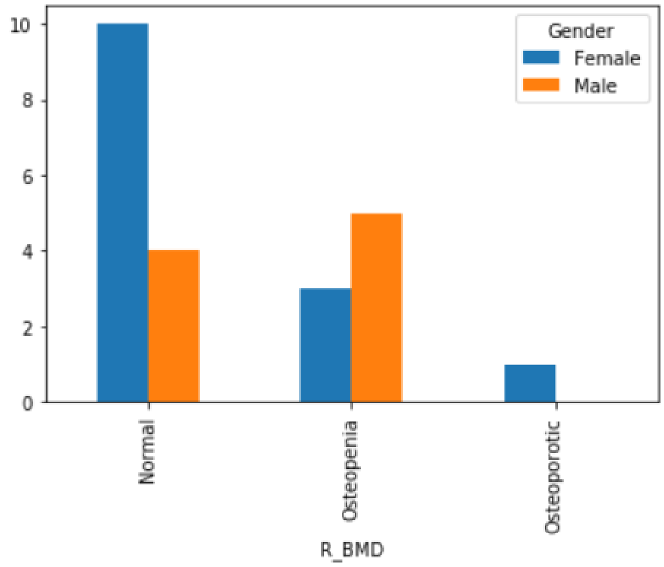
In [103]:
```
Total_BMD = [pivot1.iloc[0, :] + pivot1.iloc[1, :] + pivot1.iloc[2, :]]
Total_BMD = pd.DataFrame(Total_BMD, index = ['Total BMD'])
piv = pivot1.append(Total_BMD)
Total_Gender = pivot1['Female'] + pivot1['Male']
piv['Total Gender'] = Total_Gender
Percentage_Norm = [(piv.iloc[0, :]/piv.iloc[3, :])*100]
Percentage_Ostpn = [(piv.iloc[1, :]/piv.iloc[3, :])*100]
Percentage_Ostpr = [(piv.iloc[2, :]/piv.iloc[3, :])*100]
Percentage_Norm = pd.DataFrame(Percentage_Norm, index = ['Percentage Normal'])
Percentage_Ostpn = pd.DataFrame(Percentage_Ostpn, index = ['Percentage Osteopenia'])
Percentage_Ostpr = pd.DataFrame(Percentage_Ostpr, index = ['Percentage Osteoporotic'])
Percentage_R0 = (piv['Female']/piv['Total Gender'])*100
Percentage_R1 = (piv['Male']/piv['Total Gender'])*100
piv = piv.append(Percentage_Norm)
piv = piv.append(Percentage_Ostpn)
piv = piv.append(Percentage_Ostpr)
piv['Percentage Female'] = Percentage_R0
piv['Percentage Male'] = Percentage_R1
piv
```

Out[103]:

Gender	Female	Male	Total Gender	Percentage Female	Percentage Male
Normal	10.000000	4.000000	14.0	71.428571	28.571429
Osteopenia	3.000000	5.000000	8.0	37.500000	62.500000
Osteoporotic	1.000000	0.000000	1.0	100.000000	0.000000
Total BMD	14.000000	9.000000	NaN	NaN	NaN
Percentage Normal	71.428571	44.444444	NaN	NaN	NaN
Percentage Osteopenia	21.428571	55.555556	NaN	NaN	NaN
Percentage Osteoporotic	7.142857	0.000000	NaN	NaN	NaN

In [102]:
```
pivot1.plot(kind='bar')
```

Out[102]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1dca8c4ac18>
```



- **The above Bar Diagram represents the counts of the patients who have done BMD test.**
- **As per data we can conclude that 62.5% males are suffering from osteopenia and 37.5% females are suffering from osteopenia.**
- **Also, Osteopenia Percentage in Females is 21.42% and in Males is 55.56%.**
- **As per data we can conclude that as the age increases, the risk of osteopenia increases.**

TSH vs Gender

In [387]:
```
pivot1 = pd.pivot_table(TSH, values='Count', index=['R_TSH'], columns=['Gender'], aggfunc=np.sum)
pivot1
```

Out[387]:

Gender	Female	Male
R_TSH		
0	35	13
1	8	7

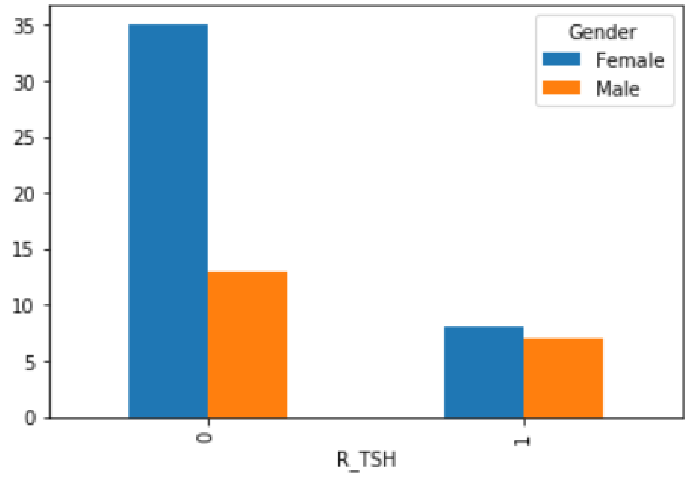
In [390]:
```
Total_TSH = [pivot1.iloc[0, :] + pivot1.iloc[1, :]]
Total_TSH = pd.DataFrame(Total_TSH, index = ['Total TSH'])
piv = pivot1.append(Total_TSH)
Total_Gender = pivot1['Female'] + pivot1['Male']
piv['Total Gender'] = Total_Gender
Percentage_0 = [(piv.iloc[0, :]/piv.iloc[2, :])*100]
Percentage_1 = [(piv.iloc[1, :]/piv.iloc[2, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
Percentage_R0 = (piv['Female']/piv['Total Gender'])*100
Percentage_R1 = (piv['Male']/piv['Total Gender'])*100
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
piv['Percentage Female'] = Percentage_R0
piv['Percentage Male'] = Percentage_R1
piv
```

Out[390]:

Gender	Female	Male	Total Gender	Percentage Female	Percentage Male
0	35.000000	13.0	48.0	72.916667	27.083333
1	8.000000	7.0	15.0	53.333333	46.666667
Total TSH	43.000000	20.0	NaN	NaN	NaN
Percentage 0	81.395349	65.0	NaN	NaN	NaN
Percentage 1	18.604651	35.0	NaN	NaN	NaN

```
In [389]: pivot1.plot(kind='bar')
```

```
Out[389]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbee8ebe0>
```



- 53.33% Females are Thyroid Positive where as 46.67% Males are Thyroid Positive.
- The above graph indicates that the risk of having thyroid is higher in Females as compared to Males.

TSH vs Vitamin B12

- **Here, in column index, 0 = not having B12 deficiency and 1=having B12 deficiency.**
- **In row index, 0 = not having thyroid and 1=having thyroid.**

In [402]:
```
pivot4 = pd.pivot_table(TSH, values = 'Count', index = ['R_TSH'], columns = ['R_B12'], aggfunc = np.sum)
pivot4
```

Out[402]:

R_B12	0	1
R_TSH		
0	23	6
1	5	3

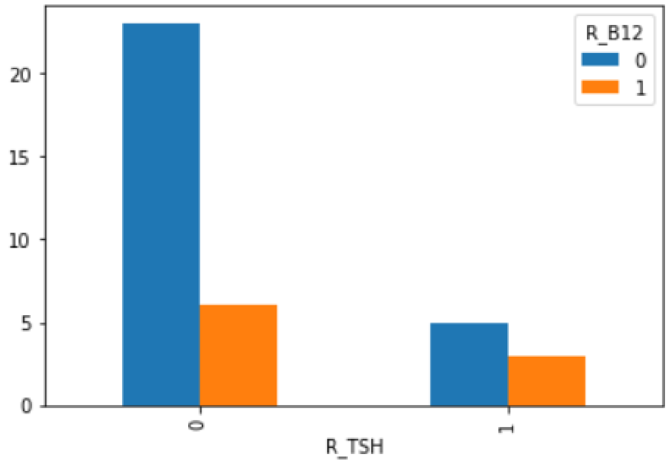
In [405]:
```
Total_TSH = [pivot4.iloc[0, :] + pivot4.iloc[1, :]]
Total_TSH = pd.DataFrame(Total_TSH, index = ['Total TSH'])
piv = pivot4.append(Total_TSH)
Total_B12 = pivot4.iloc[:, 0] + pivot4.iloc[:, 1]
piv['Total B12'] = Total_B12
Percentage_0 = [(piv.iloc[0, :]/piv.iloc[2, :])*100]
Percentage_1 = [(piv.iloc[1, :]/piv.iloc[2, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
Percentage_R0 = (piv.iloc[:, 0]/piv['Total B12'])*100
Percentage_R1 = (piv.iloc[:, 1]/piv['Total B12'])*100
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
piv['Percentage 0'] = Percentage_R0
piv['Percentage 1'] = Percentage_R1
piv
```

Out[405]:

R_B12	0	1	Total B12	Percentage 0	Percentage 1
0	23.000000	6.000000	29.0	79.310345	20.689655
1	5.000000	3.000000	8.0	62.500000	37.500000
Total TSH	28.000000	9.000000	NaN	NaN	NaN
Percentage 0	82.142857	66.666667	NaN	NaN	NaN
Percentage 1	17.857143	33.333333	NaN	NaN	NaN

```
In [404]: pivot4.plot(kind = 'bar')
```

```
Out[404]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbefedac8>
```



- **From the above Bar Diagram, we can say that there are only three patients who are suffering from Thyroid as well as vitamin B12 deficiency.**

TSH vs Age Group

- **Here we divided age groups in 4 groups as shown in table**

```
In [431]: pivot9 = pd.pivot_table(TSH, values='Count', index=['R_TSH'], columns=['Age_Group'], aggfunc=np.sum)
          pivot9.fillna(0, inplace = True)
          pivot9
```

Out[431]:

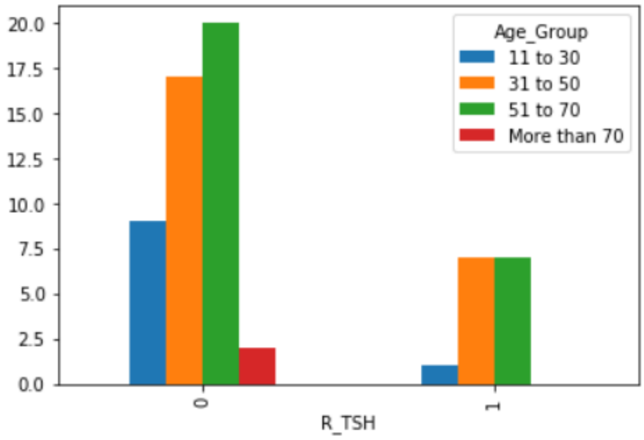
Age_Group	11 to 30	31 to 50	51 to 70	More than 70
R_TSH				
0	9.0	17.0	20.0	2.0
1	1.0	7.0	7.0	0.0

```
In [434]: Total_TSH = [pivot9.iloc[0, :] + pivot9.iloc[1, :]]
          Total_TSH = pd.DataFrame(Total_TSH, index = ['Total TSH'])
          piv = pivot9.append(Total_TSH)
          Total_AG = pivot9['11 to 30'] + pivot9['31 to 50'] + pivot9['51 to 70'] + pivot9['More than 70']
          piv['Total Age Groups'] = Total_AG
          Percentage_0 = [(piv.iloc[0, :]/piv.iloc[2, :])*100]
          Percentage_1 = [(piv.iloc[1, :]/piv.iloc[2, :])*100]
          Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
          Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
          Percentage_R1 = (piv['11 to 30']/piv['Total Age Groups'])*100
          Percentage_R2 = (piv['31 to 50']/piv['Total Age Groups'])*100
          Percentage_R3 = (piv['51 to 70']/piv['Total Age Groups'])*100
          Percentage_R4 = (piv['More than 70']/piv['Total Age Groups'])*100
          piv = piv.append(Percentage_0)
          piv = piv.append(Percentage_1)
          piv['Percentage 11 to 30'] = Percentage_R1
          piv['Percentage 31 to 50'] = Percentage_R2
          piv['Percentage 51 to 70'] = Percentage_R3
          piv['Percentage More than 70'] = Percentage_R4
          piv
```

Out[434]:

Age_Group	11 to 30	31 to 50	51 to 70	More than 70	Total Age Groups	Percentage 11 to 30	Percentage 31 to 50	Percentage 51 to 70	Percentage More than 70
0	9.0	17.000000	20.000000	2.0	48.0	18.750000	35.416667	41.666667	4.166667
1	1.0	7.000000	7.000000	0.0	15.0	6.666667	46.666667	46.666667	0.000000
Total TSH	10.0	24.000000	27.000000	2.0	NaN	NaN	NaN	NaN	NaN
Percentage 0	90.0	70.833333	74.074074	100.0	NaN	NaN	NaN	NaN	NaN
Percentage 1	10.0	29.166667	25.925926	0.0	NaN	NaN	NaN	NaN	NaN

In [433]:
```
pivot9.plot(kind='bar')
```
Out[433]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbf1f8400>



- **As per data we can conclude that as the age increases, the risk of thyroid is also increasing.**

Vitamin B12 vs Age Group

Here we divided age groups in 4 groups as shown in table

In [435]:
```
pivot10 = pd.pivot_table(TSH, values='Count', index=['R_B12'], columns=['Age_Group'], aggfunc=np.sum)
pivot10.fillna(0, inplace = True)
pivot10
```

Out[435]:

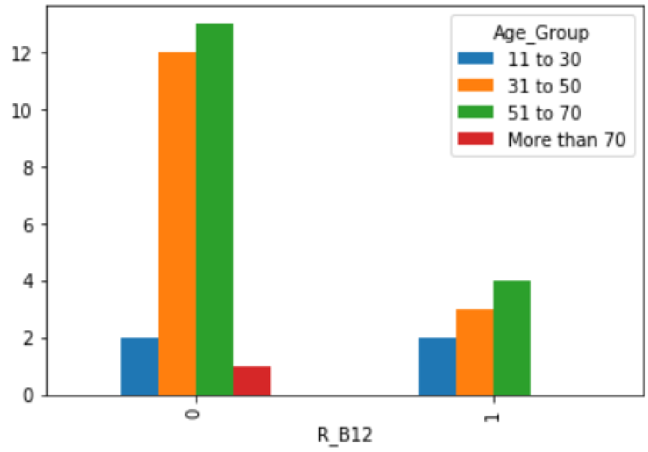
Age_Group	11 to 30	31 to 50	51 to 70	More than 70
R_B12				
0	2.0	12.0	13.0	1.0
1	2.0	3.0	4.0	0.0

In [438]:
```
Total_B12 = [pivot10.iloc[0, :] + pivot10.iloc[1, :]]
Total_B12 = pd.DataFrame(Total_B12, index = ['Total B12'])
piv = pivot10.append(Total_B12)
Total_AG = pivot10['11 to 30'] + pivot10['31 to 50'] + pivot10['51 to 70'] + pivot10['More than 70']
piv['Total Age Groups'] = Total_AG
Percentage_0 = [(piv.iloc[0, :]/piv.iloc[2, :])*100]
Percentage_1 = [(piv.iloc[1, :]/piv.iloc[2, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
Percentage_R1 = (piv['11 to 30']/piv['Total Age Groups'])*100
Percentage_R2 = (piv['31 to 50']/piv['Total Age Groups'])*100
Percentage_R3 = (piv['51 to 70']/piv['Total Age Groups'])*100
Percentage_R4 = (piv['More than 70']/piv['Total Age Groups'])*100
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
piv['Percentage 11 to 30'] = Percentage_R1
piv['Percentage 31 to 50'] = Percentage_R2
piv['Percentage 51 to 70'] = Percentage_R3
piv['Percentage More than 70'] = Percentage_R4
piv
```

Out[438]:

Age_Group	11 to 30	31 to 50	51 to 70	More than 70	Total Age Groups	Percentage 11 to 30	Percentage 31 to 50	Percentage 51 to 70	Percentage More than 70
0	2.0	12.0	13.000000	1.0	28.0	7.142857	42.857143	46.428571	3.571429
1	2.0	3.0	4.000000	0.0	9.0	22.222222	33.333333	44.444444	0.000000
Total B12	4.0	15.0	17.000000	1.0	NaN	NaN	NaN	NaN	NaN
Percentage 0	50.0	80.0	76.470588	100.0	NaN	NaN	NaN	NaN	NaN
Percentage 1	50.0	20.0	23.529412	0.0	NaN	NaN	NaN	NaN	NaN

In [437]:
```
pivot10.plot(kind='bar')
```
Out[437]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1fcbf249780>
```



- **As per data we can conclude that as the age increases, the risk of Vitamin B12 deficiency is also increases.**

TSH Constant and Gender vs Two Age Groups

In [439]:
```
A_G = []
for i in range(len(TSH['Age'])):
    if TSH.iloc[i, 0] <= 50:
        A_G.append('Less than or equal to 50')
    else:
        A_G.append('More than 50')
```

In [440]:
```
A_G = pd.DataFrame([A_G]).T
A_G.head()
```

Out[440]:

	0
0	Less than or equal to 50
1	Less than or equal to 50
2	Less than or equal to 50
3	Less than or equal to 50
4	More than 50

In [452]:
```
pivot13 = pd.pivot_table(TSH, values='Count', index=['R_TSH', 'Gender'], columns=['A_G'], aggfunc=np.sum)
pivot13
```

Out[452]:

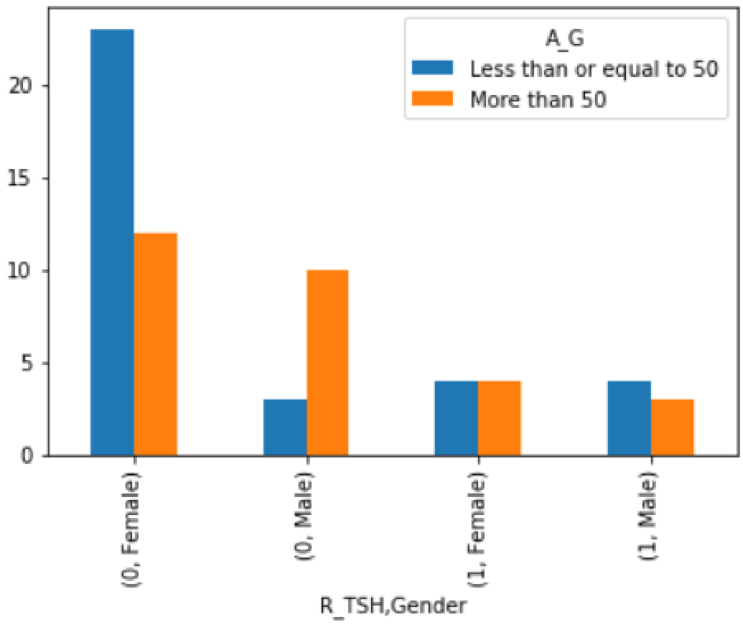
	A_G	**Less than or equal to 50**	**More than 50**
R_TSH	**Gender**		
0	**Female**	23	12
	Male	3	10
1	**Female**	4	4
	Male	4	3

In [455]:
```
Total_0 = [pivot13.loc['0'].iloc[0, :] + pivot13.loc['0'].iloc[1, :]]
Total_0 = pd.DataFrame(Total_0, index = ['Total 0'])
piv = pivot13.append(Total_0)
Total_1 = [pivot13.loc['1'].iloc[0, :] + pivot13.loc['1'].iloc[1, :]]
Total_1 = pd.DataFrame(Total_1, index = ['Total 1'])
piv = piv.append(Total_1)
Total_Female = [pivot13.iloc[0, :] + pivot13.iloc[2, :]]
Total_Female = pd.DataFrame(Total_Female, index = ['Total Female'])
piv = piv.append(Total_Female)
Total_Male = [pivot13.iloc[1, :] + pivot13.iloc[3, :]]
Total_Male = pd.DataFrame(Total_Male, index = ['Total Male'])
piv = piv.append(Total_Male)
Total_TSH_Gender = [pivot13.iloc[0, :] + pivot13.iloc[1, :] + pivot13.iloc[2, :] + pivot13.iloc[3, :]]
Total_TSH_Gender = pd.DataFrame(Total_TSH_Gender, index = ['Total TSH and Gender'])
piv = piv.append(Total_TSH_Gender)
Total_AG = pivot13['Less than or equal to 50'] + pivot13['More than 50']
pivot13['Total Age Groups'] = Total_AG
Percentage_0 = [((piv.iloc[0, :] + piv.iloc[1, :])/piv.iloc[8, :])*100]
Percentage_1 = [((piv.iloc[2, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
Percentage_Female = [((piv.iloc[0, :] + piv.iloc[2, :])/piv.iloc[8, :])*100]
Percentage_Male = [((piv.iloc[1, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_Female = pd.DataFrame(Percentage_Female, index = ['Percentage Female'])
Percentage_Male = pd.DataFrame(Percentage_Male, index = ['Percentage Male'])
piv = piv.append(Percentage_Female)
piv = piv.append(Percentage_Male)
Percentage_R1 = (piv['Less than or equal to 50']/piv['Total Age Groups'])*100
Percentage_R2 = (piv['More than 50']/piv['Total Age Groups'])*100
piv['Percentage Less than or equal to 50'] = Percentage_R1
piv['Percentage More than 50'] = Percentage_R2
piv
```

A_G	Less than or equal to 50	More than 50	Total Age Groups	Percentage Less than or equal to 50	Percentage More than 50
(Female, 0)	23.000000	12.000000	35.000000	65.714286	34.285714
(Female, 1)	4.000000	4.000000	8.000000	50.000000	50.000000
(Male, 0)	3.000000	10.000000	13.000000	23.076923	76.923077
(Male, 1)	4.000000	3.000000	7.000000	57.142857	42.857143
Total Female	27.000000	16.000000	43.000000	62.790698	37.209302
Total Male	7.000000	13.000000	20.000000	35.000000	65.000000
Total 0	26.000000	22.000000	48.000000	54.166667	45.833333
Total 1	8.000000	7.000000	15.000000	53.333333	46.666667
Total TSH and Gender	34.000000	29.000000	63.000000	53.968254	46.031746
Percentage Female	79.411765	55.172414	68.253968	116.347469	80.834002
Percentage Male	20.588235	44.827586	31.746032	64.852941	141.206897
Percentage 0	76.470588	75.862069	76.190476	100.367647	99.568966
Percentage 1	23.529412	24.137931	23.809524	98.823529	101.379310

In [453]:
```
pivot13.plot(kind='bar')
```

Out[453]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbf392e48>



- **57.14% Males are diagnosed with Thyroid Positive in Age Group of Less than or Equal to 50 Years, whereas 42.85% Males are diagnosed with Thyroid Positive in Age Group of More than 50 Years.**

TSH and Gender Constant vs Age Groups

In [457]:
```
pivot14 = pd.pivot_table(TSH, values='Count', index=['Gender', 'R_TSH'], columns=['A_G'], aggfunc=np.sum)
pivot14
```

Out[457]:

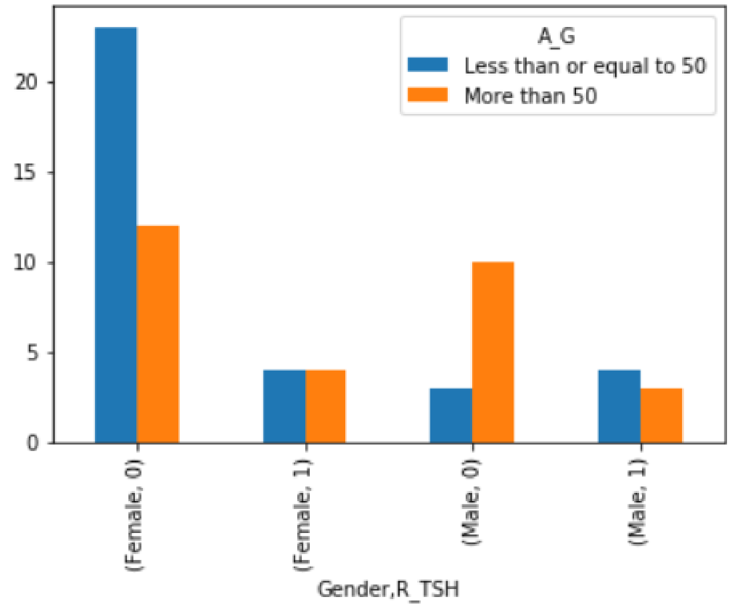
	A_G	Less than or equal to 50	More than 50
Gender	**R_TSH**		
Female	**0**	23	12
	1	4	4
Male	**0**	3	10
	1	4	3

In [460]:
```
Total_F = [pivot14.loc['Female'].iloc[0, :] + pivot14.loc['Female'].iloc[1, :]]
Total_F = pd.DataFrame(Total_F, index = ['Total Female'])
piv = pivot14.append(Total_F)
Total_M = [pivot14.loc['Male'].iloc[0, :] + pivot14.loc['Male'].iloc[1, :]]
Total_M = pd.DataFrame(Total_M, index = ['Total Male'])
piv = piv.append(Total_M)
Total_0 = [pivot14.iloc[0, :] + pivot14.iloc[2, :]]
Total_0 = pd.DataFrame(Total_0, index = ['Total 0'])
piv = piv.append(Total_0)
Total_1 = [pivot14.iloc[1, :] + pivot14.iloc[3, :]]
Total_1 = pd.DataFrame(Total_1, index = ['Total 1'])
piv = piv.append(Total_1)
Total_TSH_Gender = [pivot14.iloc[0, :] + pivot14.iloc[1, :] + pivot14.iloc[2, :] + pivot14.iloc[3, :]]
Total_TSH_Gender = pd.DataFrame(Total_TSH_Gender, index = ['Total TSH and Gender'])
piv = piv.append(Total_TSH_Gender)
Total_AG = pivot14['Less than or equal to 50'] + pivot14['More than 50']
pivot14['Total Age Groups'] = Total_AG
Percentage_F = [((piv.iloc[0, :] + piv.iloc[1, :])/piv.iloc[8, :])*100]
Percentage_M = [((piv.iloc[2, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_F = pd.DataFrame(Percentage_F, index = ['Percentage Female'])
Percentage_M = pd.DataFrame(Percentage_M, index = ['Percentage Male'])
piv = piv.append(Percentage_F)
piv = piv.append(Percentage_M)
Percentage_0 = [((piv.iloc[0, :] + piv.iloc[2, :])/piv.iloc[8, :])*100]
Percentage_1 = [((piv.iloc[1, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
Percentage_R1 = (piv['Less than or equal to 50']/piv['Total Age Groups'])*100
Percentage_R2 = (piv['More than 50']/piv['Total Age Groups'])*100
piv['Percentage Less than or equal to 50'] = Percentage_R1
piv['Percentage More than 50'] = Percentage_R2
piv
```

A_G	Less than or equal to 50	More than 50	Total Age Groups	Percentage Less than or equal to 50	Percentage More than 50
(Female, 0)	23.000000	12.000000	35.000000	65.714286	34.285714
(Female, 1)	4.000000	4.000000	8.000000	50.000000	50.000000
(Male, 0)	3.000000	10.000000	13.000000	23.076923	76.923077
(Male, 1)	4.000000	3.000000	7.000000	57.142857	42.857143
Total Female	27.000000	16.000000	43.000000	62.790698	37.209302
Total Male	7.000000	13.000000	20.000000	35.000000	65.000000
Total 0	26.000000	22.000000	48.000000	54.166667	45.833333
Total 1	8.000000	7.000000	15.000000	53.333333	46.666667
Total TSH and Gender	34.000000	29.000000	63.000000	53.968254	46.031746
Percentage Female	79.411765	55.172414	68.253968	116.347469	80.834002
Percentage Male	20.588235	44.827586	31.746032	64.852941	141.206897
Percentage 0	76.470588	75.862069	76.190476	100.367647	99.568966
Percentage 1	23.529412	24.137931	23.809524	98.823529	101.379310

```
In [458]: pivot14.plot(kind='bar')

Out[458]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbf4202b0>
```



- **57.14% Males are diagnosed with Thyroid Positive in Age Group of Less than or Equal to 50 Years, whereas 42.85% Males are diagnosed with Thyroid Positive in Age Group of More than 50 Years.**

B12 Constant and Gender vs Age Groups

In [462]:
```
pivot15 = pd.pivot_table(TSH, values='Count', index=['R_B12', 'Gender'], columns=['A_G'], aggfunc=np.sum)
pivot15
```

Out[462]:

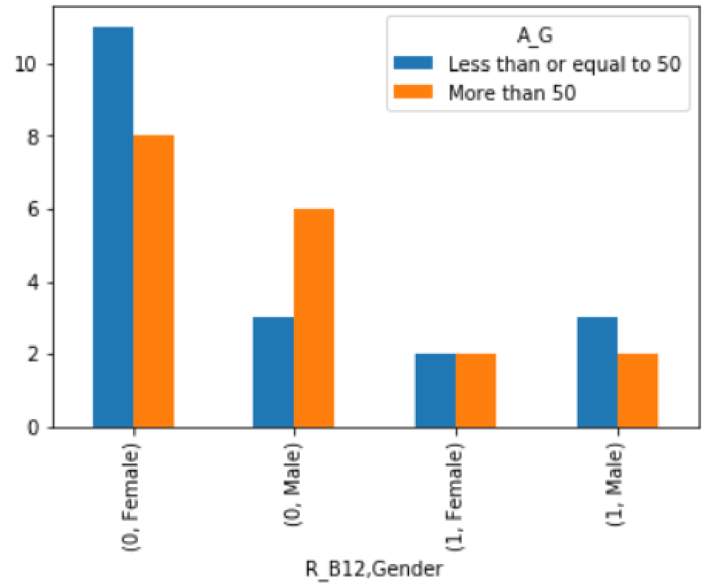
R_B12	Gender	Less than or equal to 50	More than 50
0	Female	11	8
	Male	3	6
1	Female	2	2
	Male	3	2

In [465]:
```
Total_0 = [pivot15.loc['0'].iloc[0, :] + pivot15.loc['0'].iloc[1, :]]
Total_0 = pd.DataFrame(Total_0, index = ['Total 0'])
piv = pivot15.append(Total_0)
Total_1 = [pivot15.loc['1'].iloc[0, :] + pivot15.loc['1'].iloc[1, :]]
Total_1 = pd.DataFrame(Total_1, index = ['Total 1'])
piv = piv.append(Total_1)
Total_Female = [pivot15.iloc[0, :] + pivot15.iloc[2, :]]
Total_Female = pd.DataFrame(Total_Female, index = ['Total Female'])
piv = piv.append(Total_Female)
Total_Male = [pivot15.iloc[1, :] + pivot15.iloc[3, :]]
Total_Male = pd.DataFrame(Total_Male, index = ['Total Male'])
piv = piv.append(Total_Male)
Total_B12_Gender = [pivot15.iloc[0, :] + pivot15.iloc[1, :] + pivot15.iloc[2, :] + pivot15.iloc[3, :]]
Total_B12_Gender = pd.DataFrame(Total_B12_Gender, index = ['Total B12 and Gender'])
piv = piv.append(Total_B12_Gender)
Total_AG = pivot15['Less than or equal to 50'] + pivot15['More than 50']
pivot15['Total Age Groups'] = Total_AG
Percentage_0 = [((piv.iloc[0, :] + piv.iloc[1, :])/piv.iloc[8, :])*100]
Percentage_1 = [((piv.iloc[2, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
Percentage_Female = [((piv.iloc[0, :] + piv.iloc[2, :])/piv.iloc[8, :])*100]
Percentage_Male = [((piv.iloc[1, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_Female = pd.DataFrame(Percentage_Female, index = ['Percentage Female'])
Percentage_Male = pd.DataFrame(Percentage_Male, index = ['Percentage Male'])
piv = piv.append(Percentage_Female)
piv = piv.append(Percentage_Male)
Percentage_R1 = (piv['Less than or equal to 50']/piv['Total Age Groups'])*100
Percentage_R2 = (piv['More than 50']/piv['Total Age Groups'])*100
piv['Percentage Less than or equal to 50'] = Percentage_R1
piv['Percentage More than 50'] = Percentage_R2
piv
```

A_G	Less than or equal to 50	More than 50	Total Age Groups	Percentage Less than or equal to 50	Percentage More than 50
(0, Female)	11.000000	8.000000	19.000000	57.894737	42.105263
(0, Male)	3.000000	6.000000	9.000000	33.333333	66.666667
(1, Female)	2.000000	2.000000	4.000000	50.000000	50.000000
(1, Male)	3.000000	2.000000	5.000000	60.000000	40.000000
Total 0	14.000000	14.000000	28.000000	50.000000	50.000000
Total 1	5.000000	4.000000	9.000000	55.555556	44.444444
Total Female	13.000000	10.000000	23.000000	56.521739	43.478261
Total Male	6.000000	8.000000	14.000000	42.857143	57.142857
Total B12 and Gender	19.000000	18.000000	37.000000	51.351351	48.648649
Percentage 0	73.684211	77.777778	75.675676	97.368421	102.777778
Percentage 1	26.315789	22.222222	24.324324	108.187135	91.358025
Percentage Female	68.421053	55.555556	62.162162	110.068650	89.371981
Percentage Male	31.578947	44.444444	37.837838	83.458647	117.460317

```
In [463]: pivot15.plot(kind='bar')

Out[463]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbf4c25c0>
```



- **60% Males are diagnosed with Vitamin B12 Deficiency in Age Group of Less than or Equal to 50 Years, whereas 40% Males are diagnosed with Vitamin B12 Deficiency in Age Group of More than 50 Years.**

B12 and Gender Constant vs Age Groups

In [472]:
```
pivot16 = pd.pivot_table(TSH, values='Count', index=['Gender', 'R_B12'], columns=['A_G'], aggfunc=np.sum)
pivot16
```

Out[472]:

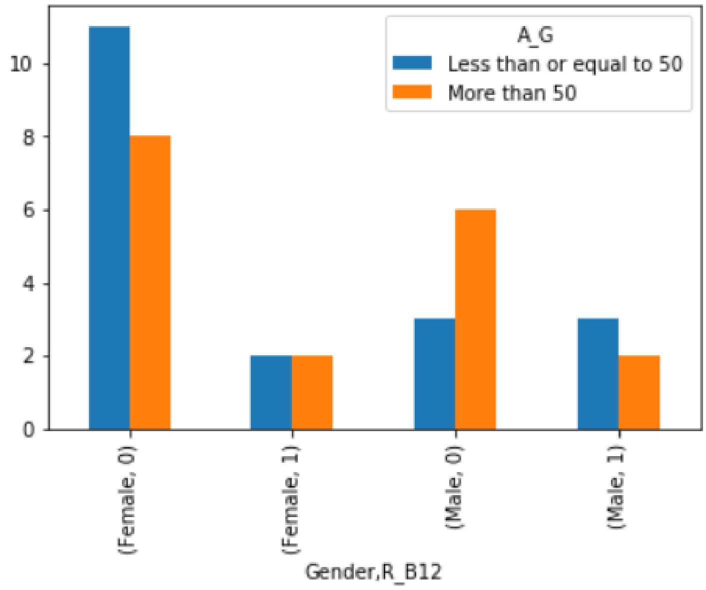
	A_G	Less than or equal to 50	More than 50
Gender	R_B12		
Female	0	11	8
	1	2	2
Male	0	3	6
	1	3	2

In [475]:
```
Total_F = [pivot16.loc['Female'].iloc[0, :] + pivot16.loc['Female'].iloc[1, :]]
Total_F = pd.DataFrame(Total_F, index = ['Total Female'])
piv = pivot16.append(Total_F)
Total_M = [pivot16.loc['Male'].iloc[0, :] + pivot16.loc['Male'].iloc[1, :]]
Total_M = pd.DataFrame(Total_M, index = ['Total Male'])
piv = piv.append(Total_M)
Total_0 = [pivot16.iloc[0, :] + pivot16.iloc[2, :]]
Total_0 = pd.DataFrame(Total_0, index = ['Total 0'])
piv = piv.append(Total_0)
Total_1 = [pivot16.iloc[1, :] + pivot16.iloc[3, :]]
Total_1 = pd.DataFrame(Total_1, index = ['Total 1'])
piv = piv.append(Total_1)
Total_B12_Gender = [pivot16.iloc[0, :] + pivot16.iloc[1, :] + pivot16.iloc[2, :] + pivot16.iloc[3, :]]
Total_B12_Gender = pd.DataFrame(Total_B12_Gender, index = ['Total B12 and Gender'])
piv = piv.append(Total_B12_Gender)
Total_AG = pivot16['Less than or equal to 50'] + pivot16['More than 50']
pivot16['Total Age Groups'] = Total_AG
Percentage_F = [((piv.iloc[0, :] + piv.iloc[1, :])/piv.iloc[8, :])*100]
Percentage_M = [((piv.iloc[2, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_F = pd.DataFrame(Percentage_F, index = ['Percentage Female'])
Percentage_M = pd.DataFrame(Percentage_M, index = ['Percentage Male'])
piv = piv.append(Percentage_F)
piv = piv.append(Percentage_M)
Percentage_0 = [((piv.iloc[0, :] + piv.iloc[2, :])/piv.iloc[8, :])*100]
Percentage_1 = [((piv.iloc[1, :] + piv.iloc[3, :])/piv.iloc[8, :])*100]
Percentage_0 = pd.DataFrame(Percentage_0, index = ['Percentage 0'])
Percentage_1 = pd.DataFrame(Percentage_1, index = ['Percentage 1'])
piv = piv.append(Percentage_0)
piv = piv.append(Percentage_1)
Percentage_R1 = (piv['Less than or equal to 50']/piv['Total Age Groups'])*100
Percentage_R2 = (piv['More than 50']/piv['Total Age Groups'])*100
piv['Percentage Less than or equal to 50'] = Percentage_R1
piv['Percentage More than 50'] = Percentage_R2
piv
```

A_G	Less than or equal to 50	More than 50	Total Age Groups	Percentage Less than or equal to 50	Percentage More than 50
(Female, 0)	11.000000	8.000000	19.000000	57.894737	42.105263
(Female, 1)	2.000000	2.000000	4.000000	50.000000	50.000000
(Male, 0)	3.000000	6.000000	9.000000	33.333333	66.666667
(Male, 1)	3.000000	2.000000	5.000000	60.000000	40.000000
Total Female	13.000000	10.000000	23.000000	56.521739	43.478261
Total Male	6.000000	8.000000	14.000000	42.857143	57.142857
Total 0	14.000000	14.000000	28.000000	50.000000	50.000000
Total 1	5.000000	4.000000	9.000000	55.555556	44.444444
Total B12 and Gender	19.000000	18.000000	37.000000	51.351351	48.648649
Percentage Female	68.421053	55.555556	62.162162	110.068650	89.371981
Percentage Male	31.578947	44.444444	37.837838	83.458647	117.460317
Percentage 0	73.684211	77.777778	75.675676	97.368421	102.777778
Percentage 1	26.315789	22.222222	24.324324	108.187135	91.358025

```
In [473]: pivot16.plot(kind='bar')
```

```
Out[473]: <matplotlib.axes._subplots.AxesSubplot at 0x1fcbf5ca860>
```



- **60% Males are diagnosed with Vitamin B12 Deficiency in Age Group of Less than or Equal to 50 Years, whereas 40% Males are diagnosed with Vitamin B12 Deficiency in Age Group of More than 50 Years.**

Vitamin D v/s Gender

In [11]:
```
pivot1 = pd.pivot_table(D, values='Count', index=['R_D'], columns=['Gender'], aggfunc=np.sum)
pivot1
```

Out[11]:

Gender	Female	Male
R_D		
Deficient	4	3
Insufficient	23	16
Sufficient	4	3

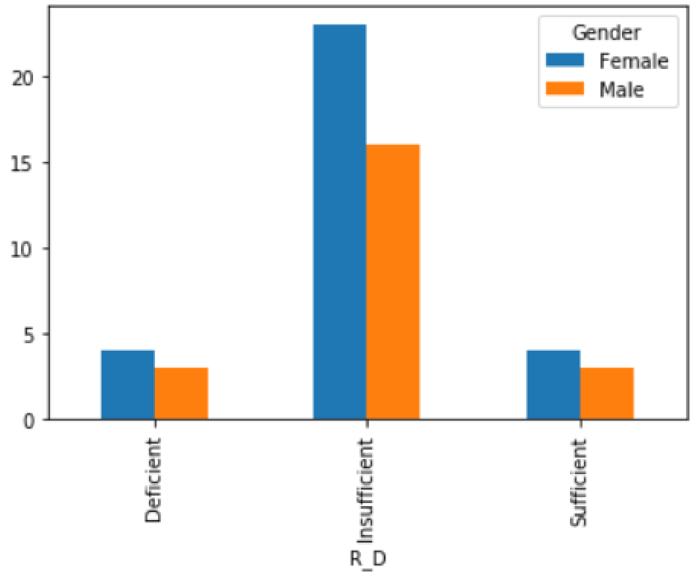
In [14]:
```
Total_D = [pivot1.iloc[0, :] + pivot1.iloc[1, :] + pivot1.iloc[2, :]]
Total_D = pd.DataFrame(Total_D, index = ['Total D'])
piv = pivot1.append(Total_D)
Total_Gender = pivot1['Female'] + pivot1['Male']
piv['Total Gender'] = Total_Gender
Percentage_Def = [(piv.iloc[0, :]/piv.iloc[3, :])*100]
Percentage_Insuff = [(piv.iloc[1, :]/piv.iloc[3, :])*100]
Percentage_Suff = [(piv.iloc[2, :]/piv.iloc[3, :])*100]
Percentage_Def = pd.DataFrame(Percentage_Def, index = ['Percentage Def'])
Percentage_Insuff = pd.DataFrame(Percentage_Insuff, index = ['Percentage Insuff'])
Percentage_Suff = pd.DataFrame(Percentage_Suff, index = ['Percentage Suff'])
Percentage_R0 = (piv['Female']/piv['Total Gender'])*100
Percentage_R1 = (piv['Male']/piv['Total Gender'])*100
piv = piv.append(Percentage_Def)
piv = piv.append(Percentage_Insuff)
piv = piv.append(Percentage_Suff)
piv['Percentage Female'] = Percentage_R0
piv['Percentage Male'] = Percentage_R1
piv
```

Out[14]:

Gender	Female	Male	Total Gender	Percentage Female	Percentage Male
Deficient	4.000000	3.000000	7.0	57.142857	42.857143
Insufficient	23.000000	16.000000	39.0	58.974359	41.025641
Sufficient	4.000000	3.000000	7.0	57.142857	42.857143
Total D	31.000000	22.000000	NaN	NaN	NaN
Percentage Def	12.903226	13.636364	NaN	NaN	NaN
Percentage Insuff	74.193548	72.727273	NaN	NaN	NaN
Percentage Suff	12.903226	13.636364	NaN	NaN	NaN

In [13]:
```
pivot1.plot(kind='bar')
```

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1cd5eb9f668>



- **57.14% Females are Vitamin D Deficient whereas 42.85% Males are Vitamin D Deficient.**
- **58.97% Females have Vitamin D Insufficiency whereas 41.02% Males have Vitamin D Insufficiency.**

ASSOCIATIONS

- Correlation analysis explores the association between two or more variables and makes inferences about the strength of the relationship.

- First, the researcher should know that measures of association are not the same as measures of statistical significance. It is possible for a weak association to be statistically significant; it is also possible for a strong association to not be statistically significant.

- For measures of association, a value of zero signifies that no relationship exists. In a correlation analysis, if the coefficient (r) has a value of one, it signifies a perfect relationship on the variables of interest.

- In regression analyses, if the standardized beta weight (β) has a value of one, it also signifies a perfect relationship on the variables of interest. The researcher should note that bivariate measures of association (e.g., Pearson correlations) are inappropriate for curvilinear relationships or discontinuous relationships.

ALL THE BELOW ASSOCIATONS ARE DONE IN SPSS

○ TSH V/S B12 (1)

Ho: There is no Association between TSH and B12.
vs
H1: There is Association between TSH and B12.

1. Most measures of association are scaled Measures of so that they reach a maximum numerical value of 1 when the two variables have a perfect relationship with each other.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
TSH * B12	34	28.8%	84	71.2%	118	100.0%

TSH * B12 Crosstabulation

			B12		Total
			0	1	
TSH	0	Count	23_a	6_a	29
		Expected Count	21.3	7.7	29.0
	1	Count	2_a	3_a	5
		Expected Count	3.7	1.3	5.0
Total		Count	25	9	34
		Expected Count	25.0	9.0	34.0

Each subscript letter denotes a subset of B12 categories whose column proportions do not differ significantly from each other at the .05 level.

○ TSH V/S B12 (2)

1. The **chi square test for association** is used to find a relationship between two categorical variables. As well as association, the test can be used to demonstrate non-association as well.

2. **Fisher's Exact Test** is used as an alternative to chi-square when you have a small sample size.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.386[a]	1	.066		
Continuity Correction[b]	1.667	1	.197		
Likelihood Ratio	2.999	1	.083		
Fisher's Exact Test				.102	.102
Linear-by-Linear Association	3.286	1	.070		
N of Valid Cases	34				

Since, p-value = 0.066 > 0.05

We Reject Ho, that is, there is no Association between TSH and B12.

TSH V/S GENDER (1)

Ho: There is no Association between TSH and Gender.
vs
H1: There is Association between TSH and Gender.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
TSH * Gender	57	48.3%	61	51.7%	118	100.0%

TSH V/S GENDER (2)

TSH * Gender Crosstabulation

			Gender		Total
			MALE	FEMALE	
TSH	0	Count	13a	35a	48
		Expected Count	15.2	32.8	48.0
	1	Count	5a	4a	9
		Expected Count	2.8	6.2	9.0
Total		Count	18	39	57
		Expected Count	18.0	39.0	57.0

Each subscript letter denotes a subset of Gender categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.844[a]	1	.092		
Continuity Correction[b]	1.678	1	.195		
Likelihood Ratio	2.659	1	.103		
Fisher's Exact Test				.124	.100
Linear-by-Linear Association	2.794	1	.095		
N of Valid Cases	57				

Since, p-value = 0.092 > 0.05

We Reject Ho, that is, there is no Association between TSH and Gender.

VITAMIN B12 V/S EATING HABITS

Ho: There is no Association between Vitamin B12 and Eating Habits.
vs
H1: There is Association between Vitamin B12 and Eating Habits.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
B12 * Eating Habit	62	52.5%	56	47.5%	118	100.0%

B12 * Eating Habit Crosstabulation

			Eating Habit			Total
			Vegetarian	Non - Vegetarian	Eggetarian	
B12	0	Count	29a	12a	2a	43
		Expected Count	30.5	10.4	2.1	43.0
	1	Count	15a	3a	1a	19
		Expected Count	13.5	4.6	.9	19.0
Total		Count	44	15	3	62
		Expected Count	44.0	15.0	3.0	62.0

Each subscript letter denotes a subset of Eating Habit categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.056[a]	2	.590
Likelihood Ratio	1.118	2	.572
Linear-by-Linear Association	.480	1	.489
N of Valid Cases	62		

Since, p-value = 0.592 > 0.05

We Reject Ho, that is, there is no Association between Vitamin B12 and Eating Habits.

BMD V/S AGE GROUPS

Ho: There is no Association between BMD and Age Groups.

vs

H1: There is Association between BMD and Age Groups.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
BMD * Age Group	48	40.7%	70	59.3%	118	100.0%

BMD * Age Group Crosstabulation

			Age Group				Total
			11 to 30	31 to 50	51 to 70	71 to 90	
BMD	Normal	Count	8a	13a, b	8b	4a, b	33
		Expected Count	5.5	12.4	11.0	4.1	33.0
	Osteopenia	Count	0a	5a, b	7b	2a, b	14
		Expected Count	2.3	5.3	4.7	1.8	14.0
	Osteoporotic	Count	0a	0a	1a	0a	1
		Expected Count	.2	.4	.3	.1	1.0
Total		Count	8	18	16	6	48
		Expected Count	8.0	18.0	16.0	6.0	48.0

Each subscript letter denotes a subset of Age Group categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.538[a]	6	.274
Likelihood Ratio	9.855	6	.131
N of Valid Cases	48		

Since, p-value = 0.274 > 0.05

We Reject Ho, that is, there is no Association between BMD and Age Groups.

BMD V/S GENDER

Ho: There is no Association between BMD and Gender.

vs

H1: There is Association between BMD and Gender.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
BMD * Gender	48	40.7%	70	59.3%	118	100.0%

BMD * Gender Crosstabulation

			Gender		Total
			MALE	FEMALE	
BMD	Normal	Count	15a	18a	33
		Expected Count	16.5	16.5	33.0
	Osteopenia	Count	9a	5a	14
		Expected Count	7.0	7.0	14.0
	Osteoporotic	Count	0a	1a	1
		Expected Count	.5	.5	1.0
Total		Count	24	24	48
		Expected Count	24.0	24.0	48.0

Each subscript letter denotes a subset of Gender categories whose column proportions do not differ significantly from each other at the .05 level.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.416[a]	2	.299
Likelihood Ratio	2.818	2	.244
Linear-by-Linear Association	.309	1	.578
N of Valid Cases	48		

Since, p-value = 0.299 > 0.05

We Reject Ho, that is, there is no Association between BMD and Gender.

ODDS RATIO

- An **Odds Ratio** (**OR**) is a statistic that quantifies the strength of the association between two events, A and B.

- The odds ratio is defined as the ratio of the odds of A in the presence of B and the odds of A in the absence of B, or equivalently (due to symmetry), the ratio of the odds of B in the presence of A and the odds of B in the absence of A.

- Two events are independent if and only if the OR equals 1

- Odds of an event happening is defined as the likelihood that an event will occur, expressed as a proportion of the likelihood that the event will not occur.

- Calculated in case-control studies as incidence of outcome is not known
- OR >1 indicates increased occurrence of event

- OR <1 indicates decreased occurrence of event (protective exposure)
- Look at CI and P value for statistical significance of value

- In rare outcomes OR = RR (RR = Relative Risk). This applies when the incidence of disease is < 10%

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

COEFFICIENT OF ASSOCIATION:

Yule's Coefficient of Association measures the strength and direction of association. "Association" means that the attributes have some degree of agreement.

2×2 Contingency Table

2×2 Contingency Table

Attribute A ⇩	Attribute B		Total
	Yes B	No β	
Yes A	(AB)	$(A\beta)$	(A)
No α	(αB)	$(\alpha\beta)$	(α)
Total	(B)	(β)	N

Yule's coefficient: $Q = \dfrac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$

Note 1: The usage of the symbol α is not to be confused with level of significance.

Note 2: (AB): Number with attributes AB etc.

This coefficient ranges from –1 to +1. The values between –1 and 0 indicate inverse relationship (association) between the attributes. The values between 0 and +1 indicate direct relationship (association) between the attributes.

TSH v/s Gender

```
In [391]: Q = ((piv.iloc[1, 0] * piv.iloc[0, 1]) - (piv.iloc[1, 1] * piv.iloc[0, 0]))/((piv.iloc[1, 0] * piv.iloc[0, 1]) + (piv.iloc[1, 1] * pi
          print("Yule's Coefficient of Association: ", Q)
          O = (piv.iloc[1, 0]*piv.iloc[0, 1])/(piv.iloc[0, 0]*piv.iloc[1, 1])
          print("Odd's Ratio: ", O)
```

```
Yule's Coefficient of Association:  -0.4040114613180516
Odd's Ratio:  0.42448979591836733
```

There is Negative Association between Thyroid Positive and Females.

The estimated Odds of Females for Thyroid Positive were 0.42 times the estimated Odds for Thyroid Negative.

1. TSH v/s B12

```
In [406]: Q = ((piv.iloc[1, 1] * piv.iloc[0, 0]) - (piv.iloc[1, 0] * piv.iloc[0, 1]))/((piv.iloc[1, 1] * piv.iloc[0, 0]) + (piv.iloc[1, 0] * pi
          print("Yule's Coefficient of Association: ", Q)
          O = (piv.iloc[1, 0]*piv.iloc[0, 1])/(piv.iloc[0, 0]*piv.iloc[1, 1])
          print("Odd's Ratio: ", O)
```

```
Yule's Coefficient of Association:  0.3939393939393939
Odd's Ratio:  0.43478260869565216
```

There is Positive Association between Thyroid Positive and Vitamin B12 Deficient.

The estimated Odds of Vitamin B12 Deficient for Thyroid Positive were 0.43 times the estimated Odds for Thyroid Negative.

3.TSH CONSTANT & Gender V/S Age Group

In [456]:
```
Q_0 = ((piv.iloc[0, 0] * piv.iloc[1, 1]) - (piv.iloc[0, 1] * piv.iloc[1, 0]))/((piv.iloc[0, 0] * piv.iloc[1, 1]) + (piv.iloc[0, 1] *
print("Yule's Coefficient of Association of TSH Negative and Age Groups: ", Q_0)
Q_1 = ((piv.iloc[2, 0] * piv.iloc[3, 1]) - (piv.iloc[3, 0] * piv.iloc[2, 1]))/((piv.iloc[2, 0] * piv.iloc[3, 1]) + (piv.iloc[3, 0] *
print("Yule's Coefficient of Association of TSH Positive and Age Groups: ", Q_1)
Q_TSHvsAG = ((piv.iloc[5, 0] * piv.iloc[4, 1]) - (piv.iloc[4, 0] * piv.iloc[5, 1]))/((piv.iloc[5, 0] * piv.iloc[4, 1]) + (piv.iloc[4,
print("Yule's Coefficient of Association of TSH and Age Group: ", Q_TSHvsAG)
O_0 = (piv.iloc[0, 0]*piv.iloc[1, 1])/(piv.iloc[1, 0]*piv.iloc[0, 1])
print("Odd's Ratio of TSH Negative and Age Groups: ", O_0)
O_1 = (piv.iloc[2, 0]*piv.iloc[3, 1])/(piv.iloc[3, 0]*piv.iloc[2, 1])
print("Odd's Ratio of TSH Positive and Age Groups: ", O_1)
O_TSHvsAG = (piv.iloc[5, 0]*piv.iloc[4, 1])/(piv.iloc[4, 0]*piv.iloc[5, 1])
print("Odd's Ratio of TSH and Age Groups: ", O_TSHvsAG)
```

```
Yule's Coefficient of Association of TSH Negative and Age Groups:  0.7293233082706767
Yule's Coefficient of Association of TSH Positive and Age Groups:  -0.14285714285714285
Yule's Coefficient of Association of TSH and Age Group:  -0.01675977653631285
Odd's Ratio of TSH Negative and Age Groups:  6.388888888888889
Odd's Ratio of TSH Positive and Age Groups:  0.75
Odd's Ratio of TSH and Age Groups:  0.967032967032967
```

1. There is Positive Association between Females and Age Group of Less than and equal to 50 Years having Thyroid Negative.
2. There is Negative Association between Females and Age Group of Less than and equal to 50 Years having Thyroid Positive.
3. There is Negative Association between Thyroid Negative and Age Group of Less than and equal to 50 Years.
4. The estimated Odds of Age Group of Less than and equal to 50 Years for Females were 6.388 times the estimated Odds for Males having Thyroid Negative.
5. The estimated Odds of Age Group of Less than and equal to 50 Years for Females were 0.75 times the estimated Odds for Males having Thyroid Positive.
6. The estimated Odds of Age Group of Less than and equal to 50 Years for Thyroid Positive were 0.96 times the estimated Odds for Thyroid Negative.

4. TSH & Gender Constant V/S Age Group

In [461]:
```
Q_F = ((piv.iloc[1, 0] * piv.iloc[0, 1]) - (piv.iloc[0, 0] * piv.iloc[1, 1]))/((piv.iloc[1, 0] * piv.iloc[0, 1]) + (piv.iloc[0, 0] *
print("Yule's Coefficient of Association of Female and Age Groups: ", Q_F)
Q_M = ((piv.iloc[3, 0] * piv.iloc[2, 1]) - (piv.iloc[2, 0] * piv.iloc[3, 1]))/((piv.iloc[3, 0] * piv.iloc[2, 1]) + (piv.iloc[2, 0] *
print("Yule's Coefficient of Association of Male and Age Groups: ", Q_M)
Q_GvsAG = ((piv.iloc[4, 0] * piv.iloc[5, 1]) - (piv.iloc[5, 0] * piv.iloc[4, 1]))/((piv.iloc[4, 0] * piv.iloc[5, 1]) + (piv.iloc[5, 0
print("Yule's Coefficient of Association of Gender and Age Group: ", Q_GvsAG)
O_F = (piv.iloc[1, 0] * piv.iloc[0, 1])/(piv.iloc[0, 0] * piv.iloc[1, 1])
print("Odd's Ratio of Female and Age Groups: ", O_F)
O_M = (piv.iloc[3, 0] * piv.iloc[2, 1])/(piv.iloc[2, 0] * piv.iloc[3, 1])
print("Odd's Ratio of Male and Age Groups: ", O_M)
O_GvsAG = (piv.iloc[4, 0] * piv.iloc[5, 1])/(piv.iloc[5, 0] * piv.iloc[4, 1])
print("Odd's Ratio of TSH and Age Groups: ", O_GvsAG)
```

```
Yule's Coefficient of Association of Female and Age Groups:  -0.3142857142857143
Yule's Coefficient of Association of Male and Age Groups:  0.6326530612244898
Yule's Coefficient of Association of Gender and Age Group:  0.5161987041036717
Odd's Ratio of Female and Age Groups:  0.5217391304347826
Odd's Ratio of Male and Age Groups:  4.444444444444445
Odd's Ratio of TSH and Age Groups:  3.1339285714285716
```

1. There is Negative Association between Thyroid Positive and Age Group of Less than and equal to 50 Years having Gender as Females.
2. There is Positive Association between Thyroid Positive and Age Group of Less than and equal to 50 Years having Gender as Males.
3. There is Positive Association between Females and Age Group of Less than and equal to 50 Years.
4. The estimated Odds of Age Group of Less than and equal to 50 Years for Thyroid Positive were 0.521 times the estimated Odds for Thyroid Negative for Females.
5. The estimated Odds of Age Group of Less than and equal to 50 Years for Thyroid Positive were 4.44 times the estimated Odds for Thyroid Negative for Males.
6. The estimated Odds of Age Group of Less than and equal to 50 Years for Females were 3.122 times the estimated Odds for Males.

1. B12 Constant & Gender V/S Age Group

```
In [466]: Q_0 = ((piv.iloc[0, 0] * piv.iloc[1, 1]) - (piv.iloc[0, 1] * piv.iloc[1, 0]))/((piv.iloc[0, 0] * piv.iloc[1, 1]) + (piv.iloc[0, 1] *
print("Yule's Coefficient of Association of B12 Negative and Age Groups: ", Q_0)
Q_1 = ((piv.iloc[2, 0] * piv.iloc[3, 1]) - (piv.iloc[3, 0] * piv.iloc[2, 1]))/((piv.iloc[2, 0] * piv.iloc[3, 1]) + (piv.iloc[3, 0] *
print("Yule's Coefficient of Association of B12 Positive and Age Groups: ", Q_1)
Q_B12vsAG = ((piv.iloc[5, 0] * piv.iloc[4, 1]) - (piv.iloc[4, 0] * piv.iloc[5, 1]))/((piv.iloc[5, 0] * piv.iloc[4, 1]) + (piv.iloc[4,
print("Yule's Coefficient of Association of B12 and Age Group: ", Q_B12vsAG)
O_0 = (piv.iloc[0, 0]*piv.iloc[1, 1])/(piv.iloc[1, 0]*piv.iloc[0, 1])
print("Odd's Ratio of B12 Negative and Age Groups: ", O_0)
O_1 = (piv.iloc[2, 0]*piv.iloc[3, 1])/(piv.iloc[3, 0]*piv.iloc[2, 1])
print("Odd's Ratio of B12 Positive and Age Groups: ", O_1)
O_TSHvsAG = (piv.iloc[5, 0]*piv.iloc[4, 1])/(piv.iloc[4, 0]*piv.iloc[5, 1])
print("Odd's Ratio of B12 and Age Groups: ", O_TSHvsAG)
```

```
Yule's Coefficient of Association of B12 Negative and Age Groups:  0.4666666666666667
Yule's Coefficient of Association of B12 Positive and Age Groups:  -0.2
Yule's Coefficient of Association of B12 and Age Group:  0.1111111111111111
Odd's Ratio of B12 Negative and Age Groups:  2.75
Odd's Ratio of B12 Positive and Age Groups:  0.6666666666666666
Odd's Ratio of B12 and Age Groups:  1.25
```

1. There is Positive Association between Females and Age Group of Less than and equal to 50 Years not having Vitamin B12 Deficiency.
2. There is Negative Association between Females and Age Group of Less than and equal to 50 Years having Vitamin B12 Deficiency.
3. There is Positive Association between Vitamin B12 Deficient and Age Group of Less than and equal to 50 Years.
4. The estimated Odds of Age Group of Less than and equal to 50 Years for Females were 2.75 times the estimated Odds for Males not having Vitamin B12 Deficiency.
5. The estimated Odds of Age Group of Less than and equal to 50 Years for Females were 0.66 times the estimated Odds for Males having Vitamin B12 Deficiency.
6. The estimated Odds of Age Group of Less than and equal to 50 Years for Thyroid Positive were 1.25 times the estimated Odds for Thyroid Negative

6. B12 AND GENDER CONSTANT V/S AGE GROUP

In [476]:
```
Q_F = ((piv.iloc[1, 0] * piv.iloc[0, 1]) - (piv.iloc[0, 0] * piv.iloc[1, 1]))/((piv.iloc[1, 0] * piv.iloc[0, 1]) + (piv.iloc[0, 0] *
print("Yule's Coefficient of Association of Female and Age Groups: ", Q_F)
Q_M = ((piv.iloc[3, 0] * piv.iloc[2, 1]) - (piv.iloc[2, 0] * piv.iloc[3, 1]))/((piv.iloc[3, 0] * piv.iloc[2, 1]) + (piv.iloc[2, 0] *
print("Yule's Coefficient of Association of Male and Age Groups: ", Q_M)
Q_GvsAG = ((piv.iloc[4, 0] * piv.iloc[5, 1]) - (piv.iloc[5, 0] * piv.iloc[4, 1]))/((piv.iloc[4, 0] * piv.iloc[5, 1]) + (piv.iloc[5, 0
print("Yule's Coefficient of Association of Gender and Age Group: ", Q_GvsAG)
O_F = (piv.iloc[1, 0] * piv.iloc[0, 1])/(piv.iloc[0, 0] * piv.iloc[1, 1])
print("Odd's Ratio of Female and Age Groups: ", O_F)
O_M = (piv.iloc[3, 0] * piv.iloc[2, 1])/(piv.iloc[2, 0] * piv.iloc[3, 1])
print("Odd's Ratio of Male and Age Groups: ", O_M)
O_GvsAG = (piv.iloc[4, 0] * piv.iloc[5, 1])/(piv.iloc[5, 0] * piv.iloc[4, 1])
print("Odd's Ratio of TSH and Age Groups: ", O_GvsAG)
```

```
Yule's Coefficient of Association of Female and Age Groups:  -0.15789473684210525
Yule's Coefficient of Association of Male and Age Groups:  0.5
Yule's Coefficient of Association of Gender and Age Group:  0.2682926829268293
Odd's Ratio of Female and Age Groups:  0.7272727272727273
Odd's Ratio of Male and Age Groups:  3.0
Odd's Ratio of TSH and Age Groups:  1.7333333333333334
```

1. There is Negative Association between Vitamin B12 Deficient and Age Group of Less than and equal to 50 Years having Gender as Females.
2. There is Positive Association between Vitamin B12 Deficient and Age Group of Less than and equal to 50 Years having Gender as Males.
3. There is Positive Association between Females and Age Group of Less than and equal to 50 Years.
4. The estimated Odds of Age Group of Less than and equal to 50 Years for Vitamin B12 Deficient were 0.72 times the estimated Odds for Vitamin B12 non - Deficient for Females.
5. The estimated Odds of Age Group of Less than and equal to 50 Years for Vitamin B12 Deficient were 3.0 times the estimated Odds for Vitamin B12 non - Deficient for Males.
6. The estimated Odds of Age Group of Less than and equal to 50 Years for Females were 1.733 times the estimated Odds for Males.

7. B12 vs Gender

```
In [15]: Q = ((piv.iloc[1, 0] * piv.iloc[0, 1]) - (piv.iloc[1, 1] * piv.iloc[0, 0]))/((piv.iloc[1, 0] * piv.iloc[0, 1]) + (piv.iloc[1, 1] * pi
         print("Yule's Coefficient of Association: ", Q)
         O = (piv.iloc[1, 0]*piv.iloc[0, 1])/(piv.iloc[0, 0]*piv.iloc[1, 1])
         print("Odd's Ratio: ", O)
```

```
Yule's Coefficient of Association:  -0.45038167938931295
Odd's Ratio:  0.37894736842105264
```

There is Negative Association between Vitamin B12 Deficient and Females.

The estimated Odds of Females for Vitamin B12 Deficient were 0.37 times the estimated Odds for Vitamin B12 non - Deficient.

LOGISTIC REGRESSION

Logistic regression predicts the probability of an outcome that can only have two values (i.e. a dichotomy). The prediction is based on the use of one or several predictors (numerical and categorical). A linear regression is not appropriate for predicting the value of a binary variable for two reasons:

- A linear regression will predict values outside the acceptable range (e.g. predicting probabilities outside the range 0 to 1)
- Since the dichotomous experiments can only have one of two possible values for each experiment, the residuals will not be normally distributed about the predicted line.

Logistic regression is similar to a linear regression, but the curve is constructed using the natural logarithm of the "odds" of the target variable, rather than the probability. Moreover, the predictors do not have to be normally distributed or have equal variance in each group.

logistic regression produces a logistic curve, which is limited to values between 0 and 1.

Logistic regression uses "odds" instead of probability.

If p is the probability that loan is in default, then odds of loan being defaulted is p(1-p).

The Logistic regression is given by,

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

Forward Logistic Regression

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	31	49.2
	Missing Cases	32	50.8
	Total	63	100.0
Unselected Cases		0	.0
Total		63	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Categorical Variables Codings

		Frequency	Parameter coding (1)	(2)	(3)
Age Group	11 to 30	3	1.000	.000	.000
	31 to 50	11	.000	1.000	.000
	51 to 70	16	.000	.000	1.000
	71 to 90	1	.000	.000	.000
Eating Habit	Eggetari	1	1.000	.000	
	Non - Ve	7	.000	1.000	
	Vegetari	23	.000	.000	
Q9	0	16	1.000		
	1	15	.000		
Q8	0	19	1.000		
	1	12	.000		
Q7	0	18	1.000		
	1	13	.000		
Q6	0	24	1.000		
	1	7	.000		
Q5	0	18	1.000		
	1	13	.000		
Q4	0	14	1.000		
	1	17	.000		
Q3	0	22	1.000		
	1	9	.000		
Q2	0	24	1.000		
	1	7	.000		
Q1	0	15	1.000		
	1	16	.000		
Gender	Female	22	1.000		
	Male	9	.000		
Q10	0	17	1.000		
	1	14	.000		
B12	0	24	1.000		
	1	7	.000		
Q12	0	24	1.000		
	1	7	.000		
Q38	0	23	1.000		
	1	8	.000		
Q36	0	27	1.000		
	1	4	.000		
Q34	0	10	1.000		
	1	21	.000		
Q32	0	28	1.000		
	1	3	.000		

Categorical Variables Codings

		Frequency	Parameter coding (1)	(2)	(3)
Q31	0	15	1.000		
	1	16	.000		
Q30	0	20	1.000		
	1	11	.000		
Q29	0	24	1.000		
	1	7	.000		
Q13	0	9	1.000		
	1	22	.000		
Q14	0	7	1.000		
	1	24	.000		
Q15	0	23	1.000		
	1	8	.000		
Q16	0	21	1.000		
	1	10	.000		
Q11	0	14	1.000		
	1	17	.000		

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted TSH 0	TSH 1	Percentage Correct
Step 0	TSH	0	25	0	100.0
		1	6	0	.0
	Overall Percentage				80.6

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-1.427	.455	9.855	1	.002	.240

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	AgeGroup	1.406	3	.704
		AgeGroup(1)	.797	1	.372
		AgeGroup(2)	.685	1	.408
		AgeGroup(3)	.008	1	.930
		Gender(1)	.552	1	.457
		EatingHabit	.439	2	.803
		EatingHabit(1)	.248	1	.618
		EatingHabit(2)	.149	1	.700
		Q1(1)	2.998	1	.083
		Q2(1)	3.200	1	.074
		Q3(1)	5.114	1	.024
		Q4(1)	.420	1	.517
		Q5(1)	.199	1	.656
		Q6(1)	3.200	1	.074
		Q7(1)	.199	1	.656
		Q8(1)	.400	1	.527
		Q9(1)	.008	1	.930
		Q10(1)	.070	1	.791
		Q11(1)	.070	1	.791
		Q12(1)	.149	1	.700
		Q13(1)	.552	1	.457
		Q14(1)	.149	1	.700
		Q15(1)	.325	1	.569
		Q16(1)	1.072	1	.301
		Q29(1)	.492	1	.483
		Q30(1)	3.160	1	.075
		Q31(1)	.008	1	.930
		Q32(1)	.797	1	.372
		Q34(1)	.828	1	.363
		Q36(1)	1.102	1	.294
		Q38(1)	.325	1	.569
		B12(1)	.492	1	.483
	Overall Statistics		31.000	30	.415

Block 1: Method = Forward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	4.693	1	.030
	Block	4.693	1	.030
	Model	4.693	1	.030

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	25.769[a]	.140	.225

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted TSH 0	Predicted TSH 1	Percentage Correct
Step 1	TSH	0	25	0	100.0
		1	6	0	.0
	Overall Percentage				80.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Q3(1)	-2.079	1.000	4.324	1	.038	.125
	Constant	-.223	.671	.111	1	.739	.800

a. Variable(s) entered on step 1: Q3.

Model if Term Removed

Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Step 1	Q3	-15.231	4.693	1	.030

Variables not in the Equation

			Score	df	Sig.
Step 1	Variables	AgeGroup	1.212	3	.750
		AgeGroup(1)	1.061	1	.303
		AgeGroup(2)	.059	1	.808
		AgeGroup(3)	.252	1	.616
		Gender(1)	.354	1	.552
		EatingHabit	.295	2	.863
		EatingHabit(1)	.105	1	.746
		EatingHabit(2)	.168	1	.682
		Q1(1)	.335	1	.563
		Q2(1)	.508	1	.476
		Q4(1)	.117	1	.732
		Q5(1)	.165	1	.684
		Q6(1)	2.221	1	.136
		Q7(1)	.003	1	.959
		Q8(1)	.253	1	.615
		Q9(1)	.252	1	.616
		Q10(1)	.469	1	.494
		Q11(1)	.098	1	.754
		Q12(1)	.000	1	.989
		Q13(1)	.046	1	.830
		Q14(1)	.335	1	.563
		Q15(1)	.763	1	.382
		Q16(1)	.120	1	.729
		Q29(1)	.234	1	.629
		Q30(1)	2.667	1	.102
		Q31(1)	.351	1	.554
		Q32(1)	1.061	1	.303
		Q34(1)	1.926	1	.165
		Q36(1)	2.145	1	.143
		Q38(1)	.253	1	.615
		B12(1)	1.759	1	.185
	Overall Statistics		31.000	29	.365

From the above applied Forward Logistic Regression, we have an attribute called “Q3” significant amongst all with p-Value 0.038.

CONCLUSION

1. The most significant factor for Thyroid disease is change in Heart Rate.
2. There is no Relationship between Vitamin B12 and Eating Habits.
3. There is no Association between TSH with Age, Gender, Vitamin B12 respectively.
4. There is no Association between BMD with Age, Gender respectively.

REFERENCES

- **Categorical Data Analysis Second Edition by ALAN AGRESTI (Chapter 2,3)**
- **https://www.medicinenet.com/script/main/art.asp?articlekey=5778**
- **https://www.webmd.com/diet/vitamin-b12-deficiency-symptoms-causes#1**
- **https://www.healthline.com/health/bone-mineral-density-test#purpose**
- **https://medlineplus.gov/vitaminddeficiency.html#:~:text=Vitamin%20D%20deficiency%20can%20lead,children%2C%20it%20can%20cause%20rickets.**
- **https://medichecks.com/blogs/news/vitamin-b12-could-it-be-affecting-your-health**
- **https://statistics.laerd.com/spss-tutorials/chi-square-test-for-association-using-spss-statistics.php**
- **https://tutorials.methodsconsultants.com/posts/logistic-regression-in-spss/**
- **http://www.brainkart.com/article/Yule---s-Coefficient-of-Association_39250/**