We have got three instruments to research for the conditions which might lead to recurrence of breast cancer for patients within five years of their first diagnosis.We analysed 50 patients chart record and concluded following tables based on overweight and breast cancer recurrence relation and Age Group vs breast cancer recurrence:

Statistical Analysis of 50 patients chart:

Obesity vs Breast Cancer Recurrence			
	Did cancer return after you completed treatment? (B)		
Are you more than 30 pounds overweight? ***(O)***	***YES***	***NO***	Total
YES	6	27	33
NO	2	15	17
Total	8	42	50

The above table can be tested for chi square test since the independent variable "O" is of type nominal and dependent variable "B" is also of nominal.
We can see that our dependent variable "B" has two categories "YES" and "NO".Moreover, the independent variable "O" is also nominal and has two levels "YES / NO".The Chi Square test is only possible in this case
Let us make our Hypotheses:

NULL Hypothesis H0 : Obesity is related with BC recurrence
Alternative Hypothesis Ha : Obesity is not related with BC recurrence

Let us consider confidence level of 95%, hence ALPHA = 0.05
Degree of Freedom = (Number of Rows - 1) * (Number of Columns - 1) = (2-1) * (2-1) = 1
Standard Chi Square value from the table = 3.841

Observation (O)	Expected (E)	O - E	(O - E)^2	(O - E)^2 / E
6	5.280000	0.720000	0.518400	0.098182
27	27.720000	-0.720000	0.518400	0.018701
2	2.720000	-0.720000	0.518400	0.190588
15	14.280000	0.720000	0.518400	0.036303
Total		0		0.343774

The calculated Chi Square = 0.343774

Since , Standard Chi Square value from the table is 3.841 which is greater than the calculated Chi Square 0.343774, so we fail to reject null hypothesis.
From this Chi Square test we conclude that there is much evidence from the provided 50 patients chart that obesity is related to the breast cancer recurrence.

Age vs Breast Cancer Recurrence		
	Did cancer return after you completed treatment? (B)	
Age Group (A)	***YES***	***NO***
40 >	0	0
40 - 50	2	6
50 - 60	3	13
60 - 70	2	15
70 - 80	1	8
80 <	0	0

The above table can be tested for chi square since the independent variable "A" is of type interval and dependent variable "B" is of nominal.

To observe relation between Breast Cancer recurrence for responses of (YES) vs Age group, we can find two statistics at very first:

Mean = (45 * 2 + 55 * 3 + 65 * 2 + 75 * 1) / 19 = 24.21
Standard Deviation = SQRT [((45 - 24.21)^2 * 2 + (55 - 24.21)^2 * 3 + (65 - 24.21)^2 * 2 + (75 - 24.21)^2 * 1)/(19-1)] = 23.11

Calculate corresponding Z-Values= (Boundary Value - Mean) / Standard Deviation	
Boundary Value	***Z- Value***
40	0.6832
50	1.1160
60	1.5500
70	1.9814
80	2.4141

Areas of intervals:

Intervals	***Area of Z Curve for this interval***	***Expected Frequency (E)***
40-50	0.4328	8.2232
50-60	0.434	8.246
60-70	0.4314	8.1966
70-80	0.4327	8.2213

Calculating Chi Square:

Intervals	***Observed Frequency for "YES" (O)***	***Expected Frequency (E)***	***(O-E)^2/E***
40-50	2	8.2232	4.71
50-60	3	8.246	3.34
60-70	2	8.1966	4.68
70-80	1	8.2213	6.34
Total			19.073

There are 4 age intervals hence 4 categories.Degree of freedom will be 3.
For ALPHA = 0.05, Standard Chi Square = 7.815
The calculated Chi Square = 19.073

The calculated Chi square is 19.073 which is greater than the standard chi square 7.815,hence rejecting null hypothesis and conclude that there is no connection between BC recurrence and any age group.

Statistical Analysis of 19 Questionnaires:

From the second instrument of 19 questionnaire , five tables have been prepared in order to know the factors which might lead to breast cancer recurrence.

Overweight vs Breast Cancer recurrence			
	Are you more than 30 pounds overweight? (O)		
BC Recurrence (B)	***YES***	***NO***	Total
YES	3	0	3

NO	9	7	16
Total	12	7	19

We can see that our dependent variable “B” has two categories “YES” and “NO”.Moreover, the independent variable “O” is also nominal and has two levels “YES / NO”.The Chi Square test is only possible in this case
Let us make our Hypotheses:

NULL Hypothesis H0 : Obesity is related with BC recurrence

Alternative Hypothesis Ha : Obesity is not related with BC recurrence

Let us consider confidence level of 95%, hence ALPHA = 0.05
Degree of Freedom = (Number of Rows - 1) * (Number of Columns - 1) = (2-1) * (2-1) = 1
Standard Chi Square value from the table = 3.841

Observation (O)	Expected (E)	O - E	(O - E)^2	(O - E)^2 / E
3	1.895	1.105	1.221	0.644
0	1.105	-1.105	1.221	1.105
9	10.105	-1.105	1.221	0.121
7	5.895	1.105	1.221	0.207
Total		0		2.077

The calculated Chi Square = 2.077

Since , Standard Chi Square value from the table is 3.841 which is greater than the calculated Chi Square 2.077, so we fail to reject null hypothesis.
From this Chi Square test we conclude that there is much evidence from the provided questionnaire that obesity is related to the breast cancer recurrence.

Physical Activity vs Breast Cancer recurrence					
	Current physical activity during an average work day (P)				
BC Recurrence (B)	***Sitting***	***Walking***	***Fairly Active***	***Very Active***	Total
YES	1	1	1	0	3
NO	0	1	9	6	16
Total	1	2	10	6	19

We can see that our dependent variable “B” has two categories “YES” and “NO”.Moreover, the independent variable “P” is also nominal and has four levels “sitting / walking / fairly active / very active”.The Chi Square test is only possible in this case
Let us make our Hypotheses:

NULL Hypothesis H0 : Physical activity is related with BC recurrence

Alternative Hypothesis Ha : Physical activity is not related with BC recurrence

Let us consider confidence level of 95%, hence ALPHA = 0.05
Degree of Freedom = (Number of Rows - 1) * (Number of Columns - 1) = (2-1) * (4-1) = 3
Standard Chi Square value from the table = 7.815

Observation (O)	Expected (E)	O - E	(O - E)^2	(O - E)^2 / E
1	0.158	0.842	0.709	4.487
1	0.316	0.684	0.468	1.481
1	1.579	-0.579	0.335	0.212
0	0.947	-0.947	0.897	0.947
0	0.842	-0.842	0.709	0.842
1	1.684	-0.684	0.468	0.278
9	8.421	0.579	0.335	0.40
6	5.053	0.947	0.897	0.177
Total		0		8.464

The calculated Chi Square = 8.464

Since , Standard Chi Square value from the table is 7.815 which is less than the calculated Chi Square 8.464, so we reject our null hypothesis.
From this Chi Square test we conclude that from the provided questionnaire that physical activity is not related to the breast cancer recurrence.

Daily Serving Counts of Fruit vs Breast Cancer recurrence				
	How many serving of fruits per day (P)			
BC Recurrence (B)	***None***	***1-3***	***More than 4***	Total
YES	3	0	0	3
NO	3	9	4	16
Total	6	9	4	19

We can see that our dependent variable “B” has two categories “YES” and “NO”.Moreover, the independent variable “P” is also nominal and has three levels “None / 1-3 / More than 4”.The Chi Square test is only possible in this case
Let us make our Hypotheses:

NULL Hypothesis H0 : Daily serving counts of fruits is related with BC recurrence
Alternative Hypothesis Ha : Daily serving counts of fruits is not related with BC recurrence

Let us consider confidence level of 95%, hence ALPHA = 0.05
Degree of Freedom = (Number of Rows - 1) * (Number of Columns - 1) = (2-1) * (4-1) = 3
Standard Chi Square value from the table = 7.815

Observation (O)	Expected (E)	O - E	(O - E)^2	(O - E)^2 / E
3	0.947368	2.052632	4.213296	4.447368
0	1.421053	-1.421053	2.019391	1.421053
0	0.631579	-0.631579	0.398892	0.631579
3	5.052632	-2.052632	4.213296	0.833882
9	7.578947	1.421053	2.019391	0.266447
4	3.368421	0.631579	0.398892	0.118421
Total		0		7.718750

The calculated Chi Square = 7.718750

Since , Standard Chi Square value from the table is 7.815 which is greater than the calculated Chi Square 7.718750, so we failed to reject our null hypothesis.
From this Chi Square test we conclude that from the provided questionnaire that daily serving of fresh fruits is related to the breast cancer recurrence.

Eating Habits vs Breast Cancer recurrence				
	Do you think what you eat is related to health problems such as breast cancer? (P)			
BC Recurrence (B)	***YES***	***NO***	***Not Sure***	Total
YES	0	1	2	3
NO	14	0	2	16
Total	14	1	4	19

We can see that our dependent variable “B” has two categories “YES” and “NO”.Moreover, the independent variable “P” is also nominal and has three levels “Yes / No / Not sure”.The Chi Square test is only possible in this case
Let us make our Hypotheses:

NULL Hypothesis H0 : Eating habits are related with BC recurrence

Alternative Hypothesis Ha : Eating habits are not related with BC recurrence

Let us consider confidence level of 95%, hence ALPHA = 0.05
Degree of Freedom = (Number of Rows - 1) * (Number of Columns - 1) = (2-1) * (4-1) = 3
Standard Chi Square value from the table = 7.815

Observation (O)	Expected (E)	O - E	(O - E)^2	(O - E)^2 / E
0	2.210526	-2.210526	4.886427	2.210526
1	0.157895	0.842105	0.709141	4.491228
2	0.631579	1.368421	1.872576	2.964912
14	11.789474	2.210526	4.886427	0.414474
0	0.842105	-0.842105	0.709141	0.842105
2	3.368421	-1.368421	1.872576	0.555921
Total		0		11.479167

The calculated Chi Square = 11.479167

Since , Standard Chi Square value from the table is 7.815 which is less than the calculated Chi Square 11.479167, so we reject our null hypothesis.
From this Chi Square test we conclude that from the provided questionnaire that eating habits are not related to the breast cancer recurrence.

Fat Intake vs Breast Cancer recurrence				
	Do you think some health conditions can be caused by eating too much fat? (P)			
BC Recurrence (B)	***YES***	***NO***	***Not Sure***	Total
YES	1	1	1	3
NO	13	1	2	16
Total	14	2	3	19

We can see that our dependent variable “B” has two categories “YES” and “NO”.Moreover, the independent variable “P” is also nominal and has three levels “Yes / No / Not sure”.The Chi Square test is only possible in this case

Let us make our Hypotheses:

NULL Hypothesis H0 : Fat intake is related with BC recurrence

Alternative Hypothesis Ha : Fat intake is not related with BC recurrence

Let us consider confidence level of 95%, hence ALPHA = 0.05
Degree of Freedom = (Number of Rows - 1) * (Number of Columns - 1) = (2-1) * (4-1) = 3
Standard Chi Square value from the table = 7.815

Observation (O)	Expected (E)	O - E	(O - E)^2	(O - E)^2 / E
1	2.210526	-1.210526	1.465374	0.662907
1	0.315789	0.684211	0.468144	1.482456
1	0.473684	0.526316	0.277008	0.584795
13	11.789474	1.210526	1.465374	0.124295
1	1.684211	-0.684211	0.468144	0.277961
2	2.526316	-0.526316	0.277008	0.109649
Total		0		3.242063

The calculated Chi Square = 3.242063

Since , Standard Chi Square value from the table is 7.815 which is greater than the calculated Chi Square 3.242063, so we failed to reject our null hypothesis.
From this Chi Square test we conclude that from the provided questionnaire that fat intake is related to the breast cancer recurrence.

T-test Statistical Analysis of 50 patients chart vs 6 case studies:

BMI vs Breast Cancer Recurrence for 50 patients		
Patient Sl.No.	**BMI (B)**	***Cancer returned after completing treatment. (R)***
1	*35.00*	yes
2	*24.20*	yes
3	*35.62*	yes
4	*25.56*	yes
5	*40.44*	yes
6	*38.90*	yes
7	*24.50*	yes

8	*36.56*	yes

T - Test

Table 1: First Sample Statistics

N	Mean	Std deviation	Std Mean Error
8	32.5975	6.73	2.38

BMI vs Breast Cancer Recurrence for 6 patients from case studies		
Patient Sl.No.	**BMI (B)**	***Cancer returned after completing treatment. (R)***
1	*30.26*	yes
2	*20.34*	yes
3	*22.49*	yes

4	*38.03*	yes
5	*40.44*	yes
6	*36.85*	yes

T - Test
Table 2: Second Sample Statistics

N	Mean	Std deviation	Std Mean Error
6	31.4017	8.47	3.458

Assumption 1: Are these independent samples? Yes, since the samples are taken from the two different charts or hospitals which are not related.

Assumption 2: Are these large samples or a normal population? We have

N1=8 < 30
N2=6 < 30

We do not have large enough samples and thus we need to check the normality assumption from both populations.

Assumption 3:
Do the populations have equal variance?
Yes, since var1 and var2 are not that different.

How do conclude this?

By using a rule of thumb where the ratio of the two sample standard deviations is from 0.5 to 2. (They are not that different as s1/s2 = 6.73/8.47 = 0.79 is quite close to 1)

We can thus proceed with the pooled t-test.

Let

x1 denote the mean for the 1st table and
x2 denote the mean for the 2nd table.

Step 1.

Null Hypothesis H0 : Mu1 - Mu2 = 0,
Alternate Hypothesis Ha : Mu1 - Mu2 != 0

Step 2. Significance level:

Alpha = 0.05

Step 3. Compute the t-statistic:

Sp = SQRT [((N1-1)*S1^2 + (N2-1) * S2^2)/(N1+N2-2)]
= SQRT[((8-1)*6.73^2 + (6-1)*8.47^2)/(8+6-2)]
= 7.5042

T = ((x1-x2) - (Mu1-Mu1)) / (Sp * SQRT(1/N1 + 1/N2))
= ((32.5975 - 31.4017) - (0)) / (7.5042 * SQRT(1/8 + 1/6))
=0.295

Step 4. Critical value:

Two-tailed test
Critical value = t(ALPHA/2) = t(0.025)

Degrees of freedom = 8 + 6 - 2 = 12
t(0.025) = 2.179
Rejection region = t < - 2.179 or t > 2.179

Step 5. Check to see if the value of the test statistic falls in the rejection region and decide whether to reject Ho.

T = 0.295 , which doesn't fall into rejection region
Hence, Failed to Reject H0 at ALPHA = 0.05

Step 6. State the conclusion in words of this t-test.

Table 3: Two sample t-test result

					95% of confidence interval	
	t(0.025)	df	Significance level (2-tailed)	Mean difference	Lower	Upper
Overweight	0.295	12	0.025	1.1958	-2.179	2.179

At 5% level of significance, the data provide sufficient evidence that the Mean BMI of both tables ,32.5975 and 31.4017 , are more or less equal and it's responsible to recur breast cancer.

Pearson Correlation between BMI and the gap interval (in years) when breast cancer recurred after first diagnosis:

A Pearson correlation was used to test the relationship between BMI and the gap interval (in years) when breast cancer recurred after the first diagnosis: this relationship was tested using Pearson Correlation Coefficient level of (- 0.4051092) for BMI's and the interval duration when breast cancer recurred for 14 patients (8 from 50 patient chart and 6 from the case studies). The test showed the approximately linear correlated relationship between a high BMI and incident of breast cancer recurrence.

BMI and the duration when breast cancer recurred for 14 patients (8 from 50 patient chart and 6 from the case studies)		
Patient Sl.No.	**BMI (X)**	***Duration in years between first diagnosis and recurrecurrence of BC. (Y)***
1	*35.00*	1
2	*24.20*	2
3	*35.62*	1
4	*25.56*	2
5	*40.44*	3
6	*38.90*	2
7	*24.50*	8

8	*36.56*	3
9	*32.60*	3
10	*20.34*	5
11	*22.49*	1
12	*38.03*	0.66
13	*40.44*	2
14	*36.85*	2

Scatter Plot of BMI vs interval duration of BC recurrence:

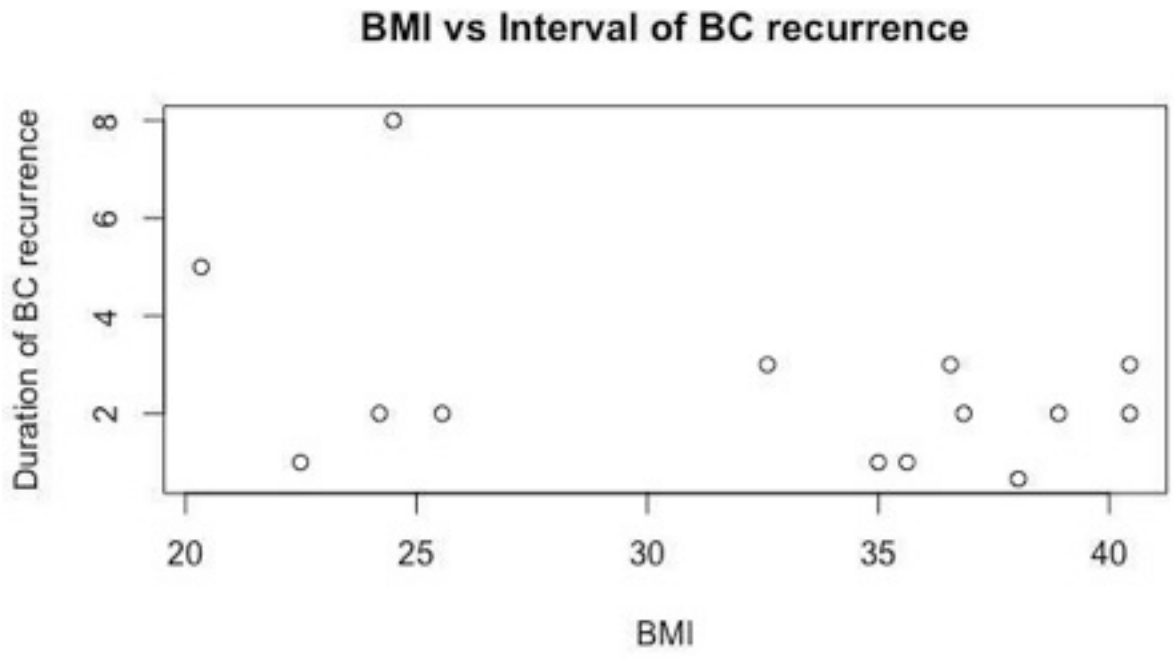

Figure: Scatter Plot shows a negative correlation between BMI and interval duration of breast cancer recurrence

The BMI compared with interval duration of breast cancer recurrence. The scatter plot shows correlation with a high BMI and breast cancer recurrence period. The graph depicts that as x-axis is increased, the time interval of breast cancer recurrence is decreased. This shows both parameters are negatively correlated. That means higher BMI will cause recurrence of breast cancer sooner.

The formula of Pearson Correlation Coefficient using above table :

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

Pearson Correlation Coefficient = -0.4051092

Reference : https://stats.idre.ucla.edu/other/mult-pkg/whatstat/