Inferential Statistics: Final Project(Udacity) - Introduction

Introduction and Description:	**Importance of analyzing and understanding the results:**	**Things to know:**
The dataset contains cases from a study that was conducted between 1958 and 1970 at the University of Chicago's Billings Hospital on the survival of patients who had undergone surgery for breast cancer.	Analyzing the given dataset can help us answer some important questions like:	Positive axillary nodes: A positive axillary lymph node is a lymph node in the area of the armpit (axilla) to which cancer has spread.
Number of Instances: 306	If there is a particular age which is more prone to the disease.	
Number of Attributes: 4 (including the class attribute)	Or are there high chances of having a positive treatment if operated at an earlier age.	Source: https://en.wikipedia.org/wiki/Positive_axillary_lymph_node
Attribute Information: 1. Age of patient at time of operation (numerical) 2. Patient's year of operation (year - 1900, numerical) 3. Number of positive axillary nodes detected (numerical) 4. Survival status (class attribute) 1 = the patient survived 5 years or longer 2 = the patient died within 5 year	Are we becoming better at treating breast cancer with time.	
Missing Attribute Values: None	Is there a particular age having chances of more severe level of the disease.	
Source: http://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival	Many questions like above can be answered by analyzing the given dataset, which can be very helpful to understand the behaviour and pattern between the disease and the concerned patients. Having these answers the doctors/government can proceed in a way to tackle the situation more effectively.	

Inferential Statistics: Final Project(Udacity) - Dataset

Age of patient at time of operation	Patient's year of operation	Number of positive axillary nodes detected	Survival status
30	64	1	1
30	62	3	1
30	65	0	1
31	59	2	1
31	65	4	1
33	58	10	1
33	60	0	1
34	59	0	2
34	66	9	2
34	58	30	1
34	60	1	1
34	61	10	1
34	67	7	1
34	60	0	1
35	64	13	1
35	63	0	1
36	60	1	1
36	69	0	1
37	60	0	1
37	63	0	1
37	58	0	1
37	59	6	1
37	60	15	1
37	63	0	1
38	69	21	2
38	59	2	1
38	60	0	1
38	60	0	1

38	62	3	1
38	64	1	1
38	66	0	1
38	66	11	1
38	60	1	1
38	67	5	1
39	66	0	2
39	63	0	1
39	67	0	1
39	58	0	1
39	59	2	1
39	63	4	1
40	58	2	1
40	58	0	1
40	65	0	1
41	60	23	2
41	64	0	2
41	67	0	2
41	58	0	1
41	59	8	1
41	59	0	1
41	64	0	1
41	69	8	1
41	65	0	1
41	65	0	1
42	69	1	2
42	59	0	2
42	58	0	1
42	60	1	1
42	59	2	1
42	61	4	1

Inferential Statistics: Final Project(Udacity) - Dataset

42	62	20	1
42	65	0	1
42	63	1	1
43	58	52	2
43	59	2	2
43	64	0	2
43	64	0	2
43	63	14	1
43	64	2	1
43	64	3	1
43	60	0	1
43	63	2	1
43	65	0	1
43	66	4	1
44	64	6	2
44	58	9	2
44	63	19	2
44	61	0	1
44	63	1	1
44	61	0	1
44	67	16	1
45	65	6	2
45	66	0	2
45	67	1	2
45	60	0	1
45	67	0	1
45	59	14	1
45	64	0	1
45	68	0	1
45	67	1	1
46	58	2	2

Inferential Statistics: Final Project(Udacity) - Dataset

46	69	3	2
46	62	5	2
46	65	20	2
46	62	0	1
46	58	3	1
46	63	0	1
47	63	23	2
47	62	0	2
47	65	0	2
47	61	0	1
47	63	6	1
47	66	0	1
47	67	0	1
47	58	3	1
47	60	4	1
47	68	4	1
47	66	12	1
48	58	11	2
48	58	11	2
48	67	7	2
48	61	8	1
48	62	2	1
48	64	0	1
48	66	0	1
49	63	0	2
49	64	10	2
49	61	1	1
49	62	0	1
49	66	0	1
49	60	1	1
49	62	1	1

Inferential Statistics: Final Project(Udacity) - Dataset

49	63	3	1
49	61	0	1
49	67	1	1
50	63	13	2
50	64	0	2
50	59	0	1
50	61	6	1
50	61	0	1
50	63	1	1
50	58	1	1
50	59	2	1
50	61	0	1
50	64	0	1
50	65	4	1
50	66	1	1
51	59	13	2
51	59	3	2
51	64	7	1
51	59	1	1
51	65	0	1
51	66	1	1
52	69	3	2
52	59	2	2
52	62	3	2
52	66	4	2
52	61	0	1
52	63	4	1
52	69	0	1
52	60	4	1
52	60	5	1
52	62	0	1

Inferential Statistics: Final Project(Udacity) - Dataset

52	62	1	1
52	64	0	1
52	65	0	1
52	68	0	1
53	58	4	2
53	65	1	2
53	59	3	2
53	60	9	2
53	63	24	2
53	65	12	2
53	58	1	1
53	60	1	1
53	60	2	1
53	61	1	1
53	63	0	1
54	60	11	2
54	65	23	2
54	65	5	2
54	68	7	2
54	59	7	1
54	60	3	1
54	66	0	1
54	67	46	1
54	62	0	1
54	69	7	1
54	63	19	1
54	58	1	1
54	62	0	1
55	63	6	2
55	68	15	2
55	58	1	1

Inferential Statistics: Final Project(Udacity) - Dataset

55	58	0	1
55	58	1	1
55	66	18	1
55	66	0	1
55	69	3	1
55	69	22	1
55	67	1	1
56	65	9	2
56	66	3	2
56	60	0	1
56	66	2	1
56	66	1	1
56	67	0	1
56	60	0	1
57	61	5	2
57	62	14	2
57	64	1	2
57	64	9	1
57	69	0	1
57	61	0	1
57	62	0	1
57	63	0	1
57	64	0	1
57	64	0	1
57	67	0	1
58	59	0	1
58	60	3	1
58	61	1	1
58	67	0	1
58	58	0	1
58	58	3	1

58	61	2	1
59	62	35	2
59	60	0	1
59	63	0	1
59	64	1	1
59	64	4	1
59	64	0	1
59	64	7	1
59	67	3	1
60	59	17	2
60	65	0	2
60	61	1	1
60	67	2	1
60	61	25	1
60	64	0	1
61	62	5	2
61	65	0	2
61	68	1	2
61	59	0	1
61	59	0	1
61	64	0	1
61	65	8	1
61	68	0	1
61	59	0	1
62	59	13	2
62	58	0	2
62	65	19	2
62	62	6	1
62	66	0	1
62	66	0	1
62	58	0	1

63	60	1	2
63	61	0	1
63	62	0	1
63	63	0	1
63	63	0	1
63	66	0	1
63	61	9	1
63	61	28	1
64	58	0	1
64	65	22	1
64	66	0	1
64	61	0	1
64	68	0	1
65	58	0	2
65	61	2	2
65	62	22	2
65	66	15	2
65	58	0	1
65	64	0	1
65	67	0	1
65	59	2	1
65	64	0	1
65	67	1	1
66	58	0	2
66	61	13	2
66	58	0	1
66	58	1	1
66	68	0	1
67	64	8	2
67	63	1	2
67	66	0	1

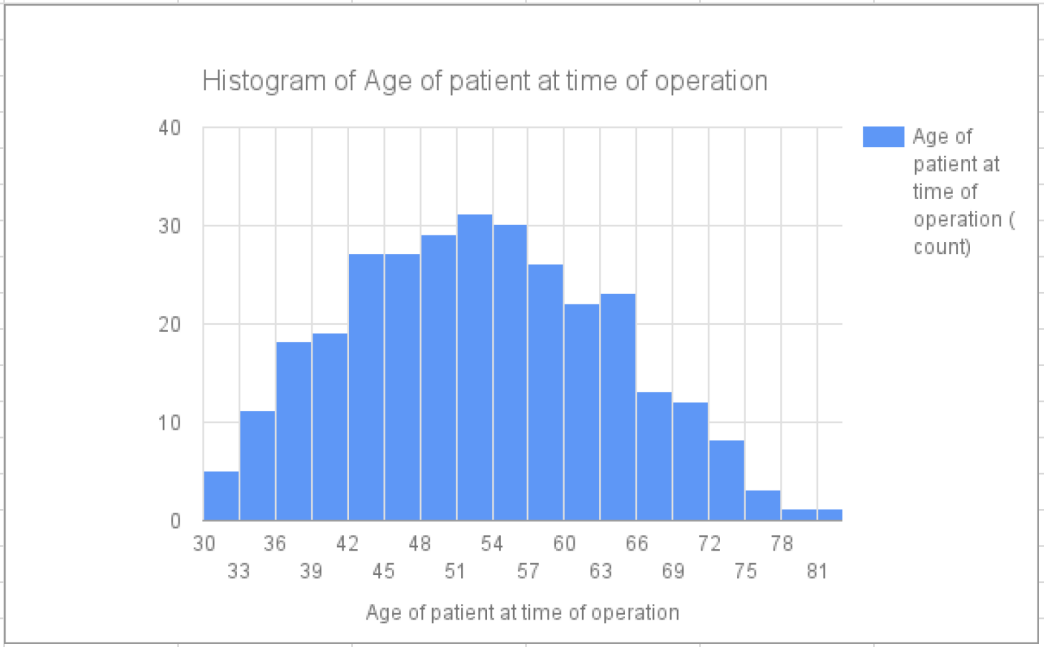
Inferential Statistics: Final Project(Udacity) - Dataset

67	66	0	1
67	61	0	1
67	65	0	1
68	67	0	1
68	68	0	1
69	67	8	2
69	60	0	1
69	65	0	1
69	66	0	1
70	58	0	2
70	58	4	2
70	66	14	1
70	67	0	1
70	68	0	1
70	59	8	1
70	63	0	1
71	68	2	1
72	63	0	2
72	58	0	1
72	64	0	1
72	67	3	1
73	62	0	1
73	68	0	1
74	65	3	2
74	63	0	1
75	62	1	1
76	67	0	1
77	65	3	1
78	65	1	2
83	58	2	2

Inferential Statistics: Final Project(Udacity) - Research Question & Hypothesis

Description	Specific Question	Possible Lurking Variables	Hypothesis	Prediction
Distribution of age of pateint at the time of operation, to see the trend on which age the disease often affects	What are the descriptive statistics for the age of pateints at the time of operation?	There might be lesser number of samples for higher age.	NA	-
Distribution of year of treatment, to see the trend	What are the descriptive statistics for the patients year of operation?	-	NA	-
Distribution of number of positive axillary nodes detected at the time of operation, to see the trend	What are the descriptive statistics for the number of positive axillary nodes detected?	-	NA	-
Releationship between age of patient at the time of operation and number of positive axillary nodes detected	How does the age of patient affect the number of positive axillary nodes detected? Are they linearly related?	There might be lesser number of samples for higher age group as few people only live that long.	H0: B1 = B2 = 0 HA: Atleast one B is not zero. where **B1** is the population slope for the predictor **age of patient** and **B2** is the population slope for the predictor **year of operation**. **Number of axillary nodes detected** is the **response variable** here. The null hypothesis claims that there is no significant correlation at all. That is, all of the coefficients are zero and none of the variables belong in the model. The alternative hypothesis is that at least one of the variables belongs in the model.	These two questions will be answered together using a multiple regression model. I support alternative hypothesis here. I think age of patient will be having some significant linear relationship with number of positive axillary nodes detected.
Change in number of positive axillary nodes detected with the year in which operation is performed, to know if the disease is becoming more or less severe with time. Or are we becoming more/less immune to disease with time. Or if we are becoming more able to detect the positive axillary nodes with time. The last one is quite an unlikely question though.	Is the disease becoming more/less severe with time? Is there a pattern?	-		
Relationship between age of patient at the time of operation and survival status post operation	Can we say that the result of treatment changes with change in the age of patients?	There might be lesser number of samples for higher age group as few people only live that long. Also, if the patients are suffereing from some other disease it might affect the result of treatment.	The main null hypothesis of a multiple logistic regression is that there is no relationship between the X variables and the Y variable; in other words, the Y values we predict from our multiple logistic regression equation are no closer to the actual Y values than we would expect by chance.	These three questions will be answered by a single model using mulitple regression. I will go for alternative hypothesis here.
Change in survival status of patients with the year in which operation is performed	Can we predict the result of treatment, based on the year in which it is being performed?	If the patients are suffereing from some other disease it might affect the result of treatment.		
Relationship between number of positive axillary nodes detected and survival status	Is there any relationship between number of positive axillary nodes detected and survival rate?	If the patients are suffereing from some other disease it might affect the result of treatment.		

Inferential Statistics: Final Project(Udacity) - Results & Conclusions 2

Age of patient at time of operation		What are the descriptive statistics for the age of pateints at the time of operation?			
30					
30				*Age of patient at time of operation*	
30					
31				Mean	**52.45751634**
31				Standard Error	0.6175922642
33				Median	**52**
33				Mode	**52**
34				Standard Deviati	**10.80345235**
34				Sample Variance	116.7145827
34				Kurtosis	-0.5893930384
34				Skewness	0.1465050565
34				Range	53
34				Minimum	30
34				Maximum	83
35				Sum	16052
35				Count	306
36				Largest(1)	**83**
36				Smallest(1)	**30**
37				Confidence Leve	1.215280944
37		Ranges from 30 to 83, but most values are around 52.5, plus or minus 10.5.			
37					
37					
37		The above data gives descriptive statistics about the age of patients being diagonsed with Breast Cancer.			
37					
38		The distribution is very normal and have central tendency as 52. So we can say the average age of a person undergoing a breast cancer operation is 52 years. Thus, it can be indirectly inferred that the age group 41-63 is most prone to breast cancer.			
38					
38					
38					
38					
38					
38					
38					

38										
38										
39										
39										
39										
39										
39										
39										
40										
40										
40										
41										
41										
41										
41										
41										
41										
41										
41										
41										
41										
41										
42										
42										
42										
42										
42										
42										
42										
42										
42										
42										
43										
43										
43										
43										
43										
43										

43										
43										
43										
43										
43										
44										
44										
44										
44										
44										
44										
44										
44										
45										
45										
45										
45										
45										
45										
45										
45										
45										
45										
46										
46										
46										
46										
46										
46										
46										
47										
47										
47										
47										
47										
47										
47										
47										

47										
47										
47										
48										
48										
48										
48										
48										
48										
48										
49										
49										
49										
49										
49										
49										
49										
49										
49										
49										
49										
50										
50										
50										
50										
50										
50										
50										
50										
50										
50										
50										
50										
51										
51										
51										
51										

51										
51										
52										
52										
52										
52										
52										
52										
52										
52										
52										
52										
52										
52										
52										
52										
52										
53										
53										
53										
53										
53										
53										
53										
53										
53										
53										
53										
53										
54										
54										
54										
54										
54										
54										
54										
54										
54										

54										
54										
54										
54										
55										
55										
55										
55										
55										
55										
55										
55										
55										
55										
55										
56										
56										
56										
56										
56										
56										
56										
56										
57										
57										
57										
57										
57										
57										
57										
57										
57										
57										
57										
58										
58										
58										
58										

58										
58										
58										
59										
59										
59										
59										
59										
59										
59										
59										
59										
60										
60										
60										
60										
60										
60										
60										
61										
61										
61										
61										
61										
61										
61										
61										
61										
61										
62										
62										
62										
62										
62										
62										
62										
63										
63										
63										

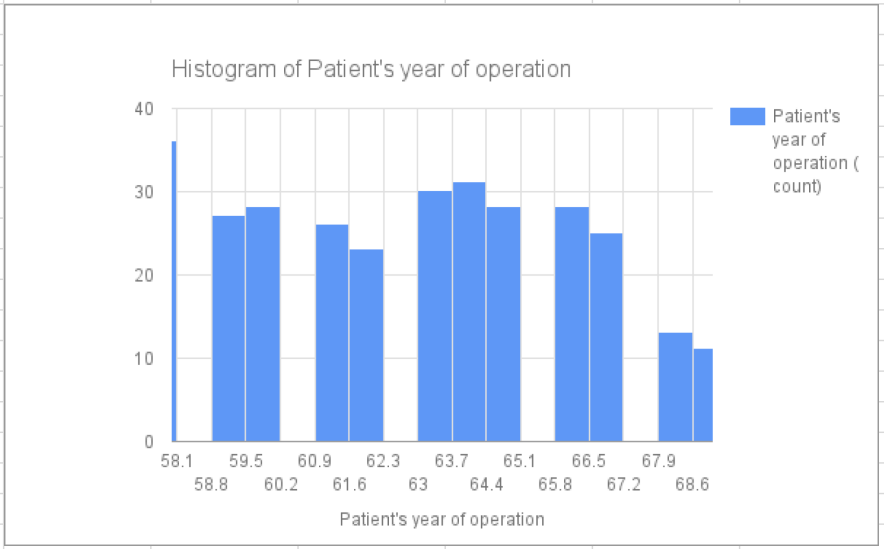
63										
63										
63										
63										
63										
64										
64										
64										
64										
64										
65										
65										
65										
65										
65										
65										
65										
65										
65										
65										
65										
66										
66										
66										
66										
66										
67										
67										
67										
67										
67										
67										
68										
68										
69										
69										
69										

69										
70										
70										
70										
70										
70										
70										
70										
71										
72										
72										
72										
72										
73										
73										
74										
74										
75										
76										
77										
78										
83										

Inferential Statistics: Final Project(Udacity) - Results & Conclusions 3

Patient's year of operation	
64	
62	
65	
59	
65	
58	
60	
59	
66	
58	
60	
61	
67	
60	
64	
63	
60	
69	
60	
63	
58	
59	
60	
63	
69	
59	
60	
60	
62	
64	
66	
66	
60	
67	
66	
63	
67	
58	
59	
63	

What are the descriptive statistics for the patients year of operation?



Patient's year of operation	
Mean	**62.85294118**
Standard Error	0.1857561008
Median	**63**
Mode	**58**
Standard Deviati	**3.249404663**
Sample Variance	10.55863067
Kurtosis	-1.118825679
Skewness	0.07875486014
Range	11
Minimum	58
Maximum	69
Sum	19233
Count	306
Largest(1)	**69**
Smallest(1)	**58**
Confidence Leve	0.3655257077

Ranges from 58 to 69.

The above data gives descriptive statistics about the year of patients being diagonsed with Breast Cancer.

There is around 3 times decrease in the count in the year 1968 as compared to that of 1958.

Also the distribution is quite asymmetrical and its very difficult to deduce any specific information from it.

58												
58												
65												
60												
64												
67												
58												
59												
59												
64												
69												
65												
65												
69												
59												
58												
60												
59												
61												
62												
65												
63												
58												
59												
64												
64												
63												
64												
64												
60												
63												
65												
66												
64												
58												
63												
61												
63												
61												
67												
65												
66												
67												

60												
67												
59												
64												
68												
67												
58												
69												
62												
65												
62												
58												
63												
63												
62												
65												
61												
63												
66												
67												
58												
60												
68												
66												
58												
58												
67												
61												
62												
64												
66												
63												
64												
61												
62												
66												
60												
62												
63												
61												
67												
63												
64												

59												
61												
61												
63												
58												
59												
61												
64												
65												
66												
59												
59												
64												
59												
65												
66												
69												
59												
62												
66												
61												
63												
69												
60												
60												
62												
62												
64												
65												
68												
58												
65												
59												
60												
63												
65												
58												
60												
60												
61												
63												
60												
65												

65												
68												
59												
60												
66												
67												
62												
69												
63												
58												
62												
63												
68												
58												
58												
58												
66												
66												
69												
69												
67												
65												
66												
60												
66												
66												
67												
60												
61												
62												
64												
64												
69												
61												
62												
63												
64												
64												
67												
59												
60												
61												
67												

58												
58												
61												
62												
60												
63												
64												
64												
64												
64												
67												
59												
65												
61												
67												
61												
64												
62												
65												
68												
59												
59												
64												
65												
68												
59												
59												
58												
65												
62												
66												
66												
58												
60												
61												
62												
63												
63												
66												
61												
61												
58												
65												

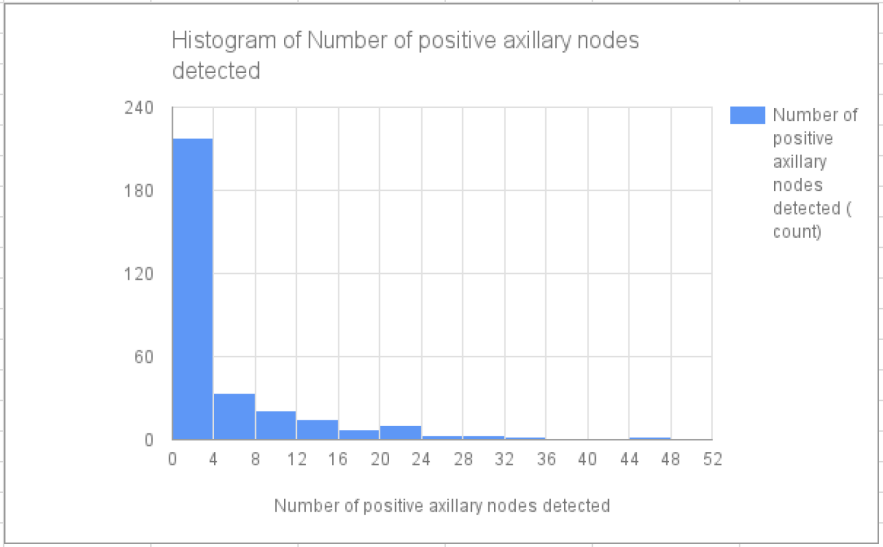
66												
61												
68												
58												
61												
62												
66												
58												
64												
67												
59												
64												
67												
58												
61												
58												
58												
68												
64												
63												
66												
66												
61												
65												
67												
68												
67												
60												
65												
66												
58												
58												
66												
67												
68												
59												
63												
68												
63												
58												
64												
67												
62												

68												
65												
63												
62												
67												
65												
65												
58												

Inferential Statistics: Final Project(Udacity) - Results & Conclusions 4

Number of positive axillary nodes detected
1
3
0
2
4
10
0
0
9
30
1
10
7
0
13
0
1
0
0
0
6
15
0
21
2
0
0
3
1
0
11
1
5
0
0
0
0
2

What are the descriptive statistics for the number of positive axillary nodes detected?



Number of positive axillary nodes detected

Mean	**4.026143791**
Standard Error	0.4110051347
Median	**1**
Mode	**0**
Standard Deviati	**7.189653506**
Sample Variance	51.69111754
Kurtosis	11.73087691
Skewness	2.983822904
Range	52
Minimum	0
Maximum	52
Sum	1232
Count	306
Largest(1)	**52**
Smallest(1)	**0**
Confidence Leve	0.808764515

Ranges from 0 to 52, with most values at the low end. The median is 1.

The above data gives descriptive statistics about the number of positive axillary nodes detetcted for pateints with Breast Cancer being diagonsed.

The distribution is highly positively skewed having mean, median and mode all fell at extreme left end.
This means, in general pateints having Breast Cancer being diagoned have around 0 to 4 number of positive axillary nodes.

4												
2												
0												
0												
23												
0												
0												
0												
8												
0												
0												
8												
0												
0												
1												
0												
0												
1												
2												
4												
20												
0												
1												
52												
2												
0												
0												
14												
2												
3												
0												
2												
0												
4												
6												
9												
19												
0												
1												
0												
16												
6												
0												

1												
0												
0												
14												
0												
0												
1												
2												
3												
5												
20												
0												
3												
0												
23												
0												
0												
0												
6												
0												
0												
3												
4												
4												
12												
11												
11												
7												
8												
2												
0												
0												
0												
10												
1												
0												
0												
1												
1												
3												
0												
1												
13												

0												
0												
6												
0												
1												
1												
2												
0												
0												
4												
1												
13												
3												
7												
1												
0												
1												
3												
2												
3												
4												
0												
4												
0												
4												
5												
0												
1												
0												
0												
0												
4												
1												
3												
9												
24												
12												
1												
1												
2												
1												
0												
11												

23												
5												
7												
7												
3												
0												
46												
0												
7												
19												
1												
0												
6												
15												
1												
0												
1												
18												
0												
3												
22												
1												
9												
3												
0												
2												
1												
0												
0												
5												
14												
1												
9												
0												
0												
0												
0												
0												
0												
0												
0												
3												
1												

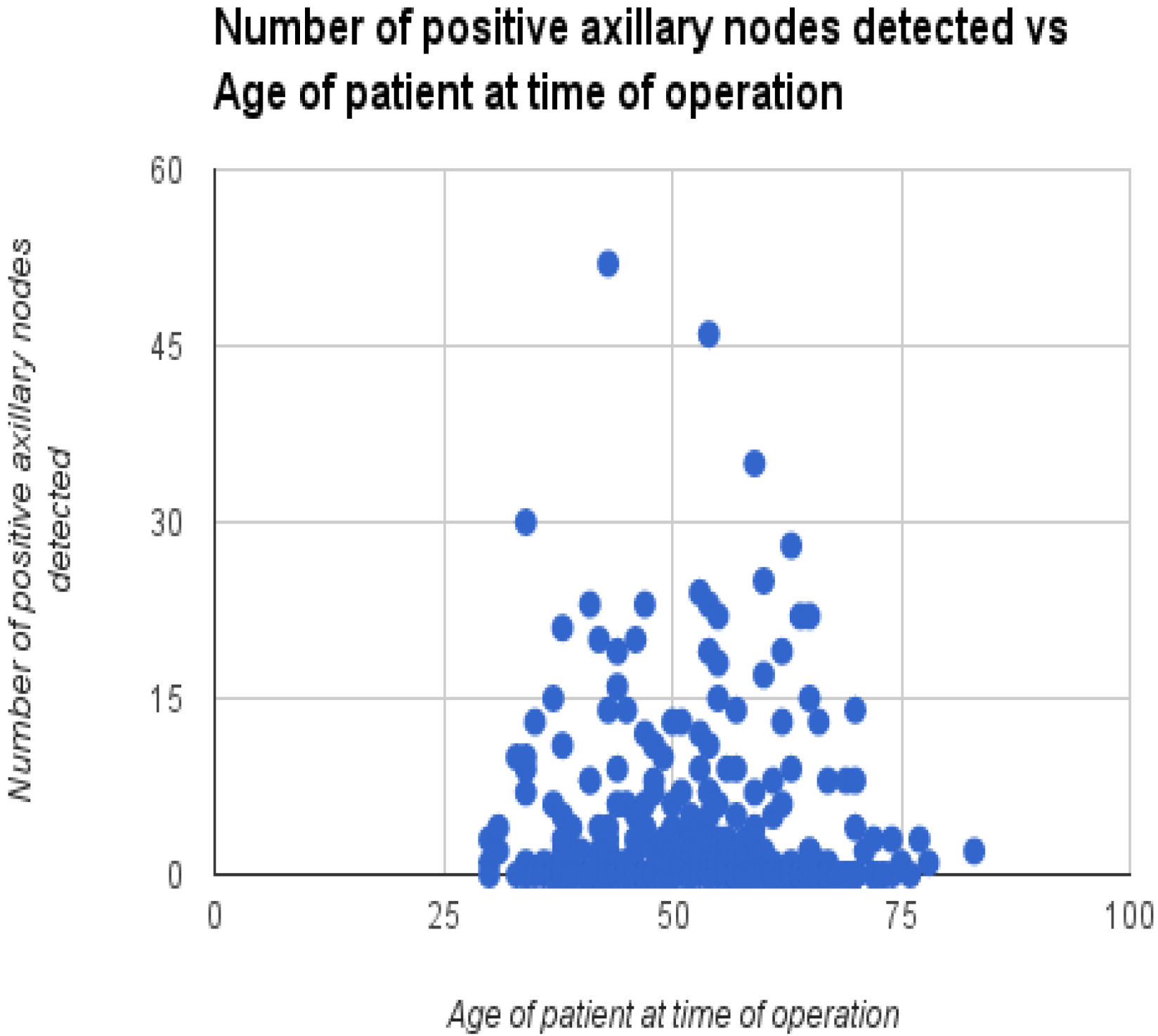
0												
0												
3												
2												
35												
0												
0												
1												
4												
0												
7												
3												
17												
0												
1												
2												
25												
0												
5												
0												
1												
0												
0												
0												
8												
0												
0												
13												
0												
19												
6												
0												
0												
0												
1												
0												
0												
0												
0												
0												
9												
28												
0												

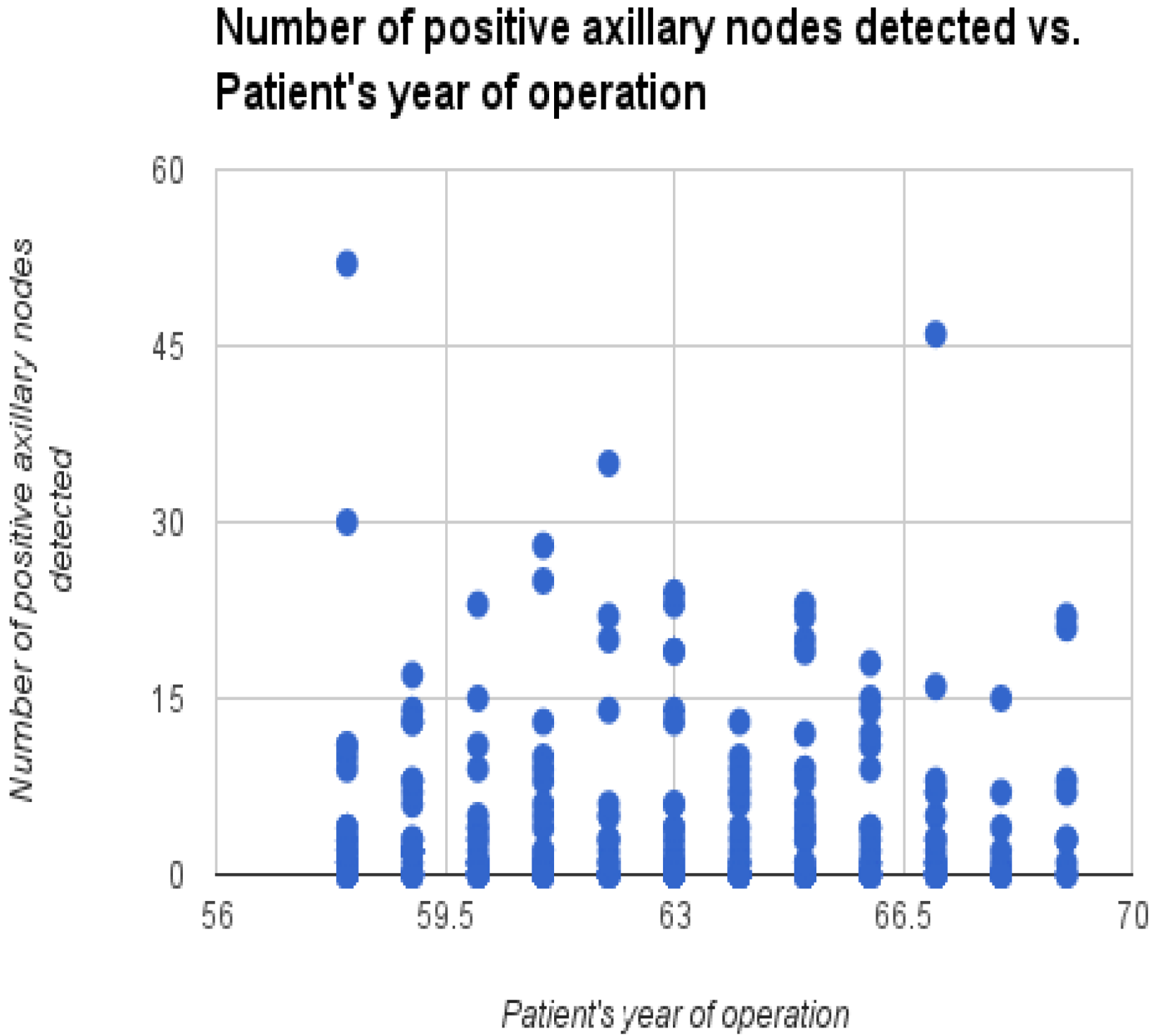
22												
0												
0												
0												
0												
2												
22												
15												
0												
0												
0												
2												
0												
1												
0												
13												
0												
1												
0												
8												
1												
0												
0												
0												
0												
0												
0												
8												
0												
0												
0												
0												
4												
14												
0												
0												
8												
0												
2												
0												
0												
0												
3												

0												
0												
3												
0												
1												
0												
3												
1												
2												

Inferential Statistics: Final Project(Udacity) - Experimental Design 6, 7

How does the age of patient affect the number of positive axillary nodes detected? **Are they linearly related?**
Is the disease becoming more/less severe with time? **Is there a pattern?**
We will perform multiple regression here. For that, we have two predictor variables namely, age of patient and year of operation. And the response will be the number of positive axillary nodes detected.
And we will see whether there is any linear relationship between the predictor(s) and the response.
If yes, we will try to predict answers to some questions.


Number of positive axillary nodes detected vs Age of patient at time of operation
Number of positive axillary nodes detected
60
45
30
15
0
0
25
50
75
100
Age of patient at time of operation


Number of positive axillary nodes detected vs. Patient's year of operation
Number of positive axillary nodes detected
60
45
30
15
0
56
59.5
63
66.5
70
Patient's year of operation

Inferential Statistics: Final Project(Udacity) - Results & Conclusions 6, 7 c

Age of patient at time of operation	Patient's year of operation	Number of positive axillary nodes detected
30	64	1
30	62	3
30	65	0
31	59	2
31	65	4
33	58	10
33	60	0
34	59	0
34	66	9
34	58	30
34	60	1
34	61	10
34	67	7
34	60	0
35	64	13
35	63	0
36	60	1
36	69	0
37	60	0
37	63	0
37	58	0
37	59	6
37	60	15
37	63	0
38	69	21
38	59	2
38	60	0
38	60	0
38	62	3
38	64	1
38	66	0
38	66	11
38	60	1
38	67	5
39	66	0

In the first plot, there seems to be **very very low negative slope**, but its hard to commit to by just seeing the scatterplot, whereas in the second, the plot is absolutely random. Let us perform a linear regression here using XLMiner Analysis Toolpak and see the actual results:

SUMMARY OUTPUT

Regression Statistics

Multiple R	**0.06320464505**
R Square	**0.003994827156**
Adjusted R Squa	-0.00257946441
Standard Error	7.198920262
Observations	306

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	62.98160943	31.49080471	0.6076437459	0.5452949654
Residual	303	15702.80924	51.82445294		
Total	305	15765.79085			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95%*	*Upper 95%*
Intercept	5.972403447	8.087937964	0.7384333898	0.4608226723	-9.943235257	21.88804215	-9.943235257	21.88804215
Age of patient at	-0.0421570565	0.03830914368	-1.100443718	**0.2720123216**	-0.1175427101	0.03322859713	-0.1175427101	0.03322859713
Patient's year of	0.004219290598	0.127368257	0.03312670439	**0.9735953316**	-0.2464190266	0.2548576078	-0.2464190266	0.2548576078

Regression Line Equation: **yhat = 5.97 - 0.04*Age + 0.00*Year = 5.97 - 0.04*Age**

So, we can note that P-values for both the predictors is pretty high, which means none of the independent variables are actually related to dependent variable in a significant linear way.

Also the correlation coefficiemt(r) is just 0.06, which is again a sign of **absolutely no relationship** between the inputs and the outcome.
And thus the coefficient of determination(r^2) is also just 0.41%.

This shows that the two predictors cannot be there in the model, as can be seen even the adjusted r is -ve.
Moreover, if we see the equation the coefficient for patient's year of operation is 0, so it means that at least patient's year of operation surely doesn't belong in this model.

Let us try removing the predictor patient's year of operation and see if we can see any significant increase in adjusted r.

SUMMARY OUTPUT

Regression Statistics

Multiple R	**0.06317610239**
R Square	**0.003991219913**
Adjusted R Squa	**0.000714875241**

Inferential Statistics: Final Project(Udacity) - Results & Conclusions 6, 7 c

39	63	0		Standard Error	7.187083194						
39	67	0		Observations	306						
39	58	0									
39	59	2		ANOVA							
39	63	4			*df*	*SS*	*MS*	*F*	*Significance F*		
40	58	2		Regression	1	62.92473839	62.92473839	1.218192929	0.2705884738		
40	58	0		Residual	304	15702.86611	51.65416484				
40	65	0		Total	305	15765.79085					
41	60	23									
41	64	0			*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95%*	*Upper 95%*
41	67	0		Intercept	6.231638161	2.040042275	3.054661286	0.002452693052	2.217246896	10.24602943	2.217246896	10.24602943
41	58	0		Age of patient at	-0.04204343865	0.03809256275	-1.103717776	**0.2705884776**	-0.1170019103	0.03291503305	-0.1170019103	0.03291503305
41	59	8									
41	59	0		Regression Line Equation:		**yhat = 6.23 - 0.04*Age**					
41	64	0									
41	69	8		Even after we removed the predictor with higher P-value, we still are having adjusted r value close to 0, so we cannot say that number of positive nodes detected is significantly related to either of the two predictors in a linear way, and hence we **fail to reject the null** and retain it.							
41	65	0									
41	65	0		It means regardless of the age of patient, the patient can have any number of postitive axillary nodes present in their body. Based on age of patient, its hard to predict the severity of the disease in a patient and we should not use the above model.							
42	69	1									
42	59	0		And there is no linear relationship that shows whether the disease is becoming more/less severe with time.							
42	58	0									
42	60	1		Further experiments can be conducted to know the various factors that actually affects the number of positive axillary nodes present in a patient, so that measures can be taken in advance to reduce the chances of getting a severe level of disease or getting a disease altogether.							
42	59	2									
42	61	4									
42	62	20									
42	65	0									
42	63	1									
43	58	52									
43	59	2									
43	64	0									
43	64	0									
43	63	14									
43	64	2									
43	64	3									
43	60	0									
43	63	2									
43	65	0									
43	66	4									
44	64	6									
44	58	9									
44	63	19									

44	61	0										
44	63	1										
44	61	0										
44	67	16										
45	65	6										
45	66	0										
45	67	1										
45	60	0										
45	67	0										
45	59	14										
45	64	0										
45	68	0										
45	67	1										
46	58	2										
46	69	3										
46	62	5										
46	65	20										
46	62	0										
46	58	3										
46	63	0										
47	63	23										
47	62	0										
47	65	0										
47	61	0										
47	63	6										
47	66	0										
47	67	0										
47	58	3										
47	60	4										
47	68	4										
47	66	12										
48	58	11										
48	58	11										
48	67	7										
48	61	8										
48	62	2										
48	64	0										
48	66	0										
49	63	0										
49	64	10										
49	61	1										
49	62	0										
49	66	0										

49	60	1										
49	62	1										
49	63	3										
49	61	0										
49	67	1										
50	63	13										
50	64	0										
50	59	0										
50	61	6										
50	61	0										
50	63	1										
50	58	1										
50	59	2										
50	61	0										
50	64	0										
50	65	4										
50	66	1										
51	59	13										
51	59	3										
51	64	7										
51	59	1										
51	65	0										
51	66	1										
52	69	3										
52	59	2										
52	62	3										
52	66	4										
52	61	0										
52	63	4										
52	69	0										
52	60	4										
52	60	5										
52	62	0										
52	62	1										
52	64	0										
52	65	0										
52	68	0										
53	58	4										
53	65	1										
53	59	3										
53	60	9										
53	63	24										
53	65	12										

53	58	1										
53	60	1										
53	60	2										
53	61	1										
53	63	0										
54	60	11										
54	65	23										
54	65	5										
54	68	7										
54	59	7										
54	60	3										
54	66	0										
54	67	46										
54	62	0										
54	69	7										
54	63	19										
54	58	1										
54	62	0										
55	63	6										
55	68	15										
55	58	1										
55	58	0										
55	58	1										
55	66	18										
55	66	0										
55	69	3										
55	69	22										
55	67	1										
56	65	9										
56	66	3										
56	60	0										
56	66	2										
56	66	1										
56	67	0										
56	60	0										
57	61	5										
57	62	14										
57	64	1										
57	64	9										
57	69	0										
57	61	0										
57	62	0										
57	63	0										

57	64	0										
57	64	0										
57	67	0										
58	59	0										
58	60	3										
58	61	1										
58	67	0										
58	58	0										
58	58	3										
58	61	2										
59	62	35										
59	60	0										
59	63	0										
59	64	1										
59	64	4										
59	64	0										
59	64	7										
59	67	3										
60	59	17										
60	65	0										
60	61	1										
60	67	2										
60	61	25										
60	64	0										
61	62	5										
61	65	0										
61	68	1										
61	59	0										
61	59	0										
61	64	0										
61	65	8										
61	68	0										
61	59	0										
62	59	13										
62	58	0										
62	65	19										
62	62	6										
62	66	0										
62	66	0										
62	58	0										
63	60	1										
63	61	0										
63	62	0										

63	63	0										
63	63	0										
63	66	0										
63	61	9										
63	61	28										
64	58	0										
64	65	22										
64	66	0										
64	61	0										
64	68	0										
65	58	0										
65	61	2										
65	62	22										
65	66	15										
65	58	0										
65	64	0										
65	67	0										
65	59	2										
65	64	0										
65	67	1										
66	58	0										
66	61	13										
66	58	0										
66	58	1										
66	68	0										
67	64	8										
67	63	1										
67	66	0										
67	66	0										
67	61	0										
67	65	0										
68	67	0										
68	68	0										
69	67	8										
69	60	0										
69	65	0										
69	66	0										
70	58	0										
70	58	4										
70	66	14										
70	67	0										
70	68	0										
70	59	8										

70	63	0										
71	68	2										
72	63	0										
72	58	0										
72	64	0										
72	67	3										
73	62	0										
73	68	0										
74	65	3										
74	63	0										
75	62	1										
76	67	0										
77	65	3										
78	65	1										
83	58	2										

Inferential Statistics: Final Project(Udacity) - Experimental Design 9, 10, 11

Can we say that the result of treatment changes with change in the age of patients?

Can we predict the result of treatment, based on the year in which it is being performed?

Is there any relationship between number of positive axillary nodes detected and survival rate?

For answering the above questions, we will use logistic regression, as the response variable here is survival status, which is a nonimal data type, so we will code the values of survival status as binary success and failure and then will perform logistic regression.

We will code 2 as 0, which will represent a failure, and 1 as 1 which will represent a success.

Success(1) here means that the patient survived 5 years or longer.
Failure(0) means the patient died within 5 year

Inferential Statistics: Final Project(Udacity) - Results & Conclusions 9, 10, 11

Age of patient at time of operation	Patient's year of operation	Number of positive axillary nodes detected	Survival status	Binary Survival Status
30	64	1	1	1
30	62	3	1	1
30	65	0	1	1
31	59	2	1	1
31	65	4	1	1
33	58	10	1	1
33	60	0	1	1
34	59	0	2	0
34	66	9	2	0
34	58	30	1	1
34	60	1	1	1
34	61	10	1	1
34	67	7	1	1
34	60	0	1	1
35	64	13	1	1
35	63	0	1	1
36	60	1	1	1
36	69	0	1	1
37	60	0	1	1
37	63	0	1	1
37	58	0	1	1
37	59	6	1	1
37	60	15	1	1
37	63	0	1	1
38	69	21	2	0
38	59	2	1	1
38	60	0	1	1
38	60	0	1	1
38	62	3	1	1
38	64	1	1	1
38	66	0	1	1
38	66	11	1	1
38	60	1	1	1
38	67	5	1	1
39	66	0	2	0
39	63	0	1	1
39	67	0	1	1
39	58	0	1	1
39	59	2	1	1
39	63	4	1	1
40	58	2	1	1
40	58	0	1	1
40	65	0	1	1
41	60	23	2	0
41	64	0	2	0
41	67	0	2	0

SUMMARY OUTPUT

Regression Statistics

Chi Square	25.43171012
Residual Dev.	328.2564282
# of iterations	5
Observations	306

	Coefficients	Standard Error	P-value	Odd Ratio	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.861625254	2.67519704	0.4865018441	6.434185462	0.03399082662	1217.938682	0.03399082662	1217.938682
Age of patient at	-0.01989934744	0.01273519344	**0.1181590139**	0.9802973378	0.9561314571	1.005074002	0.9561314571	1.005074002
Patient's year of	0.009783860489	0.04201345175	**0.8158588801**	1.009831879	0.930009022	1.096505947	0.930009022	1.096505947
Number of positi	-0.08844243662	0.01984925163	**0.000008361742**	0.9153558002	0.880428767	0.9516684056	0.880428767	0.9516684056

Can continue furter and make any comments, only once have clear understanding of Logistic Regression.
The above output is simply computed by XLMiner Analysis Toolpak.

41	58	0	1	1										
41	59	8	1	1										
41	59	0	1	1										
41	64	0	1	1										
41	69	8	1	1										
41	65	0	1	1										
41	65	0	1	1										
42	69	1	2	0										
42	59	0	2	0										
42	58	0	1	1										
42	60	1	1	1										
42	59	2	1	1										
42	61	4	1	1										
42	62	20	1	1										
42	65	0	1	1										
42	63	1	1	1										
43	58	52	2	0										
43	59	2	2	0										
43	64	0	2	0										
43	64	0	2	0										
43	63	14	1	1										
43	64	2	1	1										
43	64	3	1	1										
43	60	0	1	1										
43	63	2	1	1										
43	65	0	1	1										
43	66	4	1	1										
44	64	6	2	0										
44	58	9	2	0										
44	63	19	2	0										
44	61	0	1	1										
44	63	1	1	1										
44	61	0	1	1										
44	67	16	1	1										
45	65	6	2	0										
45	66	0	2	0										
45	67	1	2	0										
45	60	0	1	1										
45	67	0	1	1										
45	59	14	1	1										
45	64	0	1	1										
45	68	0	1	1										
45	67	1	1	1										
46	58	2	2	0										
46	69	3	2	0										
46	62	5	2	0										
46	65	20	2	0										
46	62	0	1	1										
46	58	3	1	1										
46	63	0	1	1										

47	63	23	2	0										
47	62	0	2	0										
47	65	0	2	0										
47	61	0	1	1										
47	63	6	1	1										
47	66	0	1	1										
47	67	0	1	1										
47	58	3	1	1										
47	60	4	1	1										
47	68	4	1	1										
47	66	12	1	1										
48	58	11	2	0										
48	58	11	2	0										
48	67	7	2	0										
48	61	8	1	1										
48	62	2	1	1										
48	64	0	1	1										
48	66	0	1	1										
49	63	0	2	0										
49	64	10	2	0										
49	61	1	1	1										
49	62	0	1	1										
49	66	0	1	1										
49	60	1	1	1										
49	62	1	1	1										
49	63	3	1	1										
49	61	0	1	1										
49	67	1	1	1										
50	63	13	2	0										
50	64	0	2	0										
50	59	0	1	1										
50	61	6	1	1										
50	61	0	1	1										
50	63	1	1	1										
50	58	1	1	1										
50	59	2	1	1										
50	61	0	1	1										
50	64	0	1	1										
50	65	4	1	1										
50	66	1	1	1										
51	59	13	2	0										
51	59	3	2	0										
51	64	7	1	1										
51	59	1	1	1										
51	65	0	1	1										
51	66	1	1	1										
52	69	3	2	0										
52	59	2	2	0										
52	62	3	2	0										
52	66	4	2	0										

52	61	0	1	1										
52	63	4	1	1										
52	69	0	1	1										
52	60	4	1	1										
52	60	5	1	1										
52	62	0	1	1										
52	62	1	1	1										
52	64	0	1	1										
52	65	0	1	1										
52	68	0	1	1										
53	58	4	2	0										
53	65	1	2	0										
53	59	3	2	0										
53	60	9	2	0										
53	63	24	2	0										
53	65	12	2	0										
53	58	1	1	1										
53	60	1	1	1										
53	60	2	1	1										
53	61	1	1	1										
53	63	0	1	1										
54	60	11	2	0										
54	65	23	2	0										
54	65	5	2	0										
54	68	7	2	0										
54	59	7	1	1										
54	60	3	1	1										
54	66	0	1	1										
54	67	46	1	1										
54	62	0	1	1										
54	69	7	1	1										
54	63	19	1	1										
54	58	1	1	1										
54	62	0	1	1										
55	63	6	2	0										
55	68	15	2	0										
55	58	1	1	1										
55	58	0	1	1										
55	58	1	1	1										
55	66	18	1	1										
55	66	0	1	1										
55	69	3	1	1										
55	69	22	1	1										
55	67	1	1	1										
56	65	9	2	0										
56	66	3	2	0										
56	60	0	1	1										
56	66	2	1	1										
56	66	1	1	1										
56	67	0	1	1										

56	60	0	1	1										
57	61	5	2	0										
57	62	14	2	0										
57	64	1	2	0										
57	64	9	1	1										
57	69	0	1	1										
57	61	0	1	1										
57	62	0	1	1										
57	63	0	1	1										
57	64	0	1	1										
57	64	0	1	1										
57	67	0	1	1										
58	59	0	1	1										
58	60	3	1	1										
58	61	1	1	1										
58	67	0	1	1										
58	58	0	1	1										
58	58	3	1	1										
58	61	2	1	1										
59	62	35	2	0										
59	60	0	1	1										
59	63	0	1	1										
59	64	1	1	1										
59	64	4	1	1										
59	64	0	1	1										
59	64	7	1	1										
59	67	3	1	1										
60	59	17	2	0										
60	65	0	2	0										
60	61	1	1	1										
60	67	2	1	1										
60	61	25	1	1										
60	64	0	1	1										
61	62	5	2	0										
61	65	0	2	0										
61	68	1	2	0										
61	59	0	1	1										
61	59	0	1	1										
61	64	0	1	1										
61	65	8	1	1										
61	68	0	1	1										
61	59	0	1	1										
62	59	13	2	0										
62	58	0	2	0										
62	65	19	2	0										
62	62	6	1	1										
62	66	0	1	1										
62	66	0	1	1										
62	58	0	1	1										
63	60	1	2	0										

63	61	0	1	1										
63	62	0	1	1										
63	63	0	1	1										
63	63	0	1	1										
63	66	0	1	1										
63	61	9	1	1										
63	61	28	1	1										
64	58	0	1	1										
64	65	22	1	1										
64	66	0	1	1										
64	61	0	1	1										
64	68	0	1	1										
65	58	0	2	0										
65	61	2	2	0										
65	62	22	2	0										
65	66	15	2	0										
65	58	0	1	1										
65	64	0	1	1										
65	67	0	1	1										
65	59	2	1	1										
65	64	0	1	1										
65	67	1	1	1										
66	58	0	2	0										
66	61	13	2	0										
66	58	0	1	1										
66	58	1	1	1										
66	68	0	1	1										
67	64	8	2	0										
67	63	1	2	0										
67	66	0	1	1										
67	66	0	1	1										
67	61	0	1	1										
67	65	0	1	1										
68	67	0	1	1										
68	68	0	1	1										
69	67	8	2	0										
69	60	0	1	1										
69	65	0	1	1										
69	66	0	1	1										
70	58	0	2	0										
70	58	4	2	0										
70	66	14	1	1										
70	67	0	1	1										
70	68	0	1	1										
70	59	8	1	1										
70	63	0	1	1										
71	68	2	1	1										
72	63	0	2	0										
72	58	0	1	1										
72	64	0	1	1										

72	67	3	1	1										
73	62	0	1	1										
73	68	0	1	1										
74	65	3	2	0										
74	63	0	1	1										
75	62	1	1	1										
76	67	0	1	1										
77	65	3	1	1										
78	65	1	2	0										
83	58	2	2	0										