Assignment 3
44468423
Gautham Meenakshisundaram

Prevalence of adverse dichotomous outcomes in the two groups and RR (95% CI).

Outcome	Exp (n=99)	Con (n=104)	RR (95% CI)
Fever, n (%)	22 (22)	41 (39)	0.56 (0.36 to 0.87)

a) Using Analyze .. Descriptive statistics .. Crosstabs .. , obtain a table from which the above result can be calculated, and show the calculation. (Don't attempt to calculate the 95% CI for RR.) Perform an appropriate hypothesis test.

Group * Fever Crosstabulation

Count

		Fever		
		no	yes	Total
Group	Con	63	41	104
	Exp	77	22	99
Total		140	63	203

Probability of Fever in Con group = No. of patients in Con group with Fever / No. of patients in Con group
= 41 / 104
= 0.39
Probability of Fever in Exp group = No. of patients in Exp group with Fever / No. of patients in Exp group
= 22 / 99
= 0.22
The relative risk of Fever in Exp group compared to Con group is
RR = Probability of Fever in Exp group / Probability of Fever in Con group
= 0.22 / 0.39
= 0.56

Since we are dealing with two categorical variables, the Chi-Square Test of Association is the appropriate hypothesis test to test the association between them.

H_0 : Fever is not associated with Group
H_1 : Fever is associated with Group

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7.011[a]	1	.008		
Continuity Correction[b]	6.231	1	.013		
Likelihood Ratio	7.100	1	.008		
Fisher's Exact Test				.010	.006
N of Valid Cases	203				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.72.
b. Computed only for a 2x2 table

The expected cell count for each cell is greater than 5 which satisfies one of the assumptions.

Since it is a 2x2 table, the degrees of freedom (df) for the test statistic is
df = (rows - 1) * (columns - 1)
= (2 - 1) * (2 - 1)
df = 1

Pearson Chi-Square provides a test statistic $X^2(1)$ of 7.011

Since the p-value is less than 0.05 at 0.008, we reject the null hypothesis and conclude that there is an association Fever and Group.

Assignment 3
44468423
Gautham Meenakshisundaram

b) Perform a logistic regression corresponding to the analysis in (a). What is the effect of the treatment on the incidence of Fever? Compare this with the effect given by Zhang et al .. is it the same? Comment.

Before performing logistic regression, let's look at the frequency tables of the categorical variables.

Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Con	104	51.2	51.2	51.2
	Exp	99	48.8	48.8	100.0
	Total	203	100.0	100.0	

Fever

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	140	69.0	69.0	69.0
	yes	63	31.0	31.0	100.0
	Total	203	100.0	100.0	

It seems like Exp might be the reference category in Group which might not be ideal as Con is the baseline in Group for any treatments. Recode the former as 0 and the latter as 1 to change the reference category.

In Fever, recode its non-incidence denoted by no as 0 and its incidence denoted by yes as 1.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Group2(1)	-.823	.314	6.866	1	.009	.439
	Constant	-.430	.201	4.583	1	.032	.651

a. Variable(s) entered on step 1: Group2.

In the logistic regression model, Group is significant with its p-value less than 0.05 at 0.009

Since Con group is the reference category, the beta coefficient of Group represents the Exp group.

$y_i \sim$ Bernoulli (π_i)   where i = 1,….., n

$\ln(\hat{\pi}_i / 1 - \hat{\pi}_i) = - 0.430 - 0.823\ x_{i1}$

where,

x_{i1} = {1 if patient belonging to Exp group, 0 otherwise
$\hat{\pi}_i$ = probability of a patient having Fever(1)

The odds of a patient belonging to Exp group having Fever decreases by 56% ($e^B = e^{-0.823} = 0.439$), when compared to patients in Con group.

Zhang et al observes with a relative risk of 0.56 that only 22% of Exp group had Fever when compared to 42% of Con group. The parameter interpretation of the logistic regression is in accordance with the observation of Zhang et al.

c) Model carefully the occurrence of Fever as a function of treatment and appropriate baseline and demographic covariates.

c(i) perform an initial exploratory analysis;

Categorical outcome variable vs Categorical predictor variables

Group2 * Fever2 Crosstabulation

			Fever2 .00	Fever2 1.00	Total
Group2	.00	Count	77	22	99
		% within Group2	77.8%	22.2%	100.0%
	1.00	Count	63	41	104
		% within Group2	60.6%	39.4%	100.0%
Total		Count	140	63	203
		% within Group2	69.0%	31.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	7.011[a]	1	.008		
Continuity Correction[b]	6.231	1	.013		
Likelihood Ratio	7.100	1	.008		
Fisher's Exact Test				.010	.006
Linear-by-Linear Association	6.977	1	.008		
N of Valid Cases	203				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.72.
b. Computed only for a 2x2 table

22% of patients in Exp group(0) had Fever when compared to 40% of patients in Con group(1). As noted previously, Group is significant with a p-value less than 0.05 at 0.008.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	F	79	38.9	38.9	38.9
	M	124	61.1	61.1	100.0
	Total	203	100.0	100.0	

Site of surgery

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	abdo	99	48.8	48.8	48.8
	thor	104	51.2	51.2	100.0
	Total	203	100.0	100.0	

There is no need to change the reference category for Gender and Site of surgery but their categories need to be recoded - F(0), M(1), abdo(0), thor(1) - for logistic regression.

Gender2 * Fever2 Crosstabulation

			Fever2 .00	Fever2 1.00	Total
Gender2	.00	Count	56	23	79
		% within Gender2	70.9%	29.1%	100.0%
	1.00	Count	84	40	124
		% within Gender2	67.7%	32.3%	100.0%
Total		Count	140	63	203
		% within Gender2	69.0%	31.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.223[a]	1	.637		
Continuity Correction[b]	.100	1	.752		
Likelihood Ratio	.224	1	.636		
Fisher's Exact Test				.756	.377
Linear-by-Linear Association	.222	1	.638		
N of Valid Cases	203				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 24.52.
b. Computed only for a 2x2 table

29% of female patients(0) had Fever when compared to 32% of male patients(1). Gender is not significant with a p-value greater than 0.05 at 0.637.

Siteofsurgery2 * Fever2 Crosstabulation

			Fever2 .00	Fever2 1.00	Total
Siteofsurgery2	.00	Count	70	29	99
		% within Siteofsurgery2	70.7%	29.3%	100.0%
	1.00	Count	70	34	104
		% within Siteofsurgery2	67.3%	32.7%	100.0%
Total		Count	140	63	203
		% within Siteofsurgery2	69.0%	31.0%	100.0%

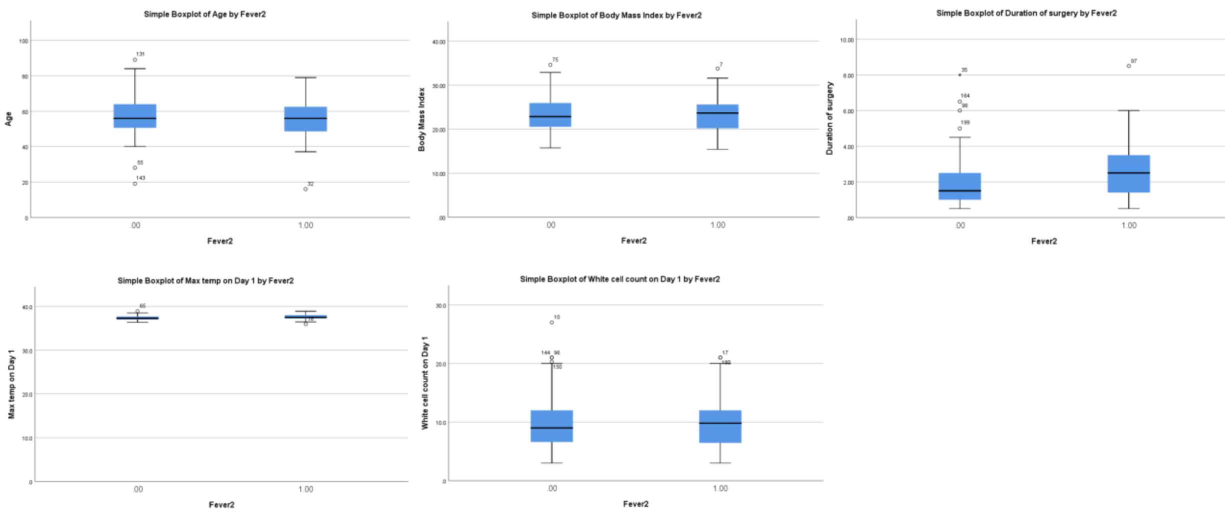
Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.274[a]	1	.601		
Continuity Correction[b]	.138	1	.710		
Likelihood Ratio	.274	1	.601		
Fisher's Exact Test				.650	.355
Linear-by-Linear Association	.272	1	.602		
N of Valid Cases	203				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 30.72.
b. Computed only for a 2x2 table

29% of patients operated in abdomen(0) had Fever when compared to 33% of patients operated in thorax(1). Site of surgery is not significant with a p-value greater than 0.05 at 0.601.

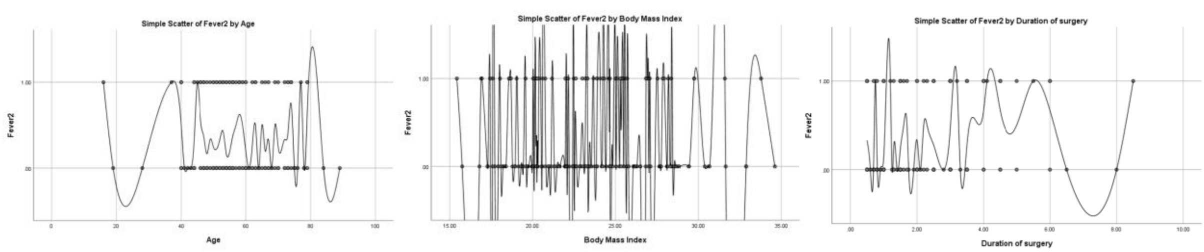
Boxplots of Categorical outcome variable vs Continuous predictor variables



There seems to be a increasing trend in the boxplot between Duration of Surgery and Fever, indicating an association between them. It can be interpreted as patients being susceptible to Fever when their Duration of Surgery is longer, on an average.

The boxplots between Fever and other variables does not show any increasing or decreasing trend.

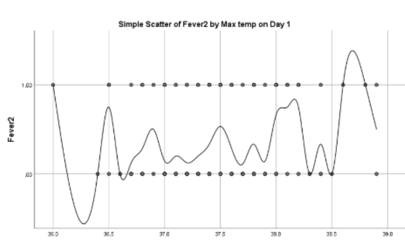
Scatterplots of Categorical outcome variable vs Continuous predictor variables



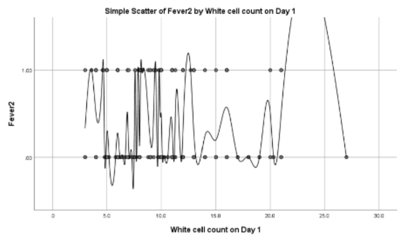
Fever ~ Age: Most of the patients are between the ages 40 and 80 and there seems to be less incidence of fever, on an average.

Fever ~ BodyMassIndex: No obvious pattern observed.

Fever ~ Durationofsurgery: On an average, less incidence of fever when Durationofsurgery is > 3 hours; and high incidence of fever when Durationofsurgery is 3-6 hours;



Fever ~ Maxtemponday1: The incidence of fever increase as the Maxtemponday1 increases, on an average.



Fever ~ Whitecellcountonday1: The incidence of fever is less when the Whitecellcountonday1 is between 13 and 20 which might be the ideal range.

c(ii) using a model selection criterion, develop an appropriate model. Give a table showing your workings, and give the table of parameter estimates for your final model.

Model selection criterion: AIC

Model	**p**	**-2L(p)**	**AIC -2L(p) + 2p**	**p-values of variables**
Fever ~ Group	2	244.367	248.367	significant
Fever ~ Gender	2	251.243	255.243	not significant
Fever ~ Siteofsurgery	2	251.193	255.193	not significant
Fever ~ Age	2	250.795	254.795	not significant
Fever ~ Bodymassindex	2	251.442	255.442	not significant
Fever ~ Durationofsurgery	2	239.952	**243.952**	**significant**
Fever ~ Maxtemponday1	2	245.110	249.110	significant
Fever ~ Whitecellcountonday1	2	251.441	255.441	not significant

Model	**p**	**-2L(p)**	**AIC -2L(p) + 2p**	**p-values of variables**
Fever ~ Durationofsurgery + Group	3	233.192	239.192	significant
Fever ~ Durationofsurgery + Gender	3	239.898	245.898	not significant
Fever ~ Durationofsurgery + Siteofsurgery	3	239.863	245.863	not significant
Fever ~ Durationofsurgery + Age	3	238.186	244.186	not significant
Fever ~ Durationofsurgery + Bodymassindex	3	239.891	245.891	not significant
Fever ~ Durationofsurgery + Maxtemponday1	3	231.305	**237.305**	**significant**
Fever ~ Durationofsurgery + Whitecellcountonday1	3	239.774	245.774	not significant

Model	**p**	**-2L(p)**	**AIC -2L(p) + 2p**	**p-values of variables**
Fever ~ Durationofsurgery + Maxtemponday1 + Group	4	225.780	**233.780**	**significant**
Fever ~ Durationofsurgery + Maxtemponday1 + Gender	4	231.305	239.305	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Siteofsurgery	4	231.216	239.216	not significant

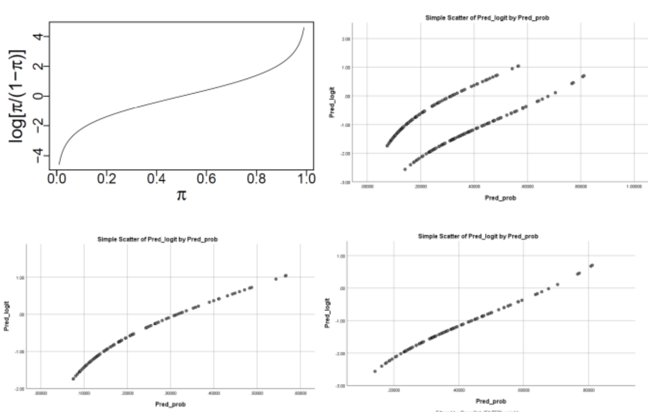
Fever ~ Durationofsurgery + Maxtemponday1 + Age	4	229.960	237.960	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Bodymassindex	4	231.222	238.222	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Whitecellcountonday1	4	229.885	237.885	not significant

Model	**p**	**-2L(p)**	**AIC -2L(p) + 2p**	**p-values of variables**
Fever ~ Durationofsurgery + Maxtemponday1 + Group + Gender	5	225.780	235.780	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Group + Siteofsurgery	5	225.697	235.697	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Group + Age	5	224.260	234.260	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Group + Bodymassindex	5	225.682	235.682	not significant
Fever ~ Durationofsurgery + Maxtemponday1 + Group + Whitecellcountonday1	5	223.614	**233.614**	**not significant**

Fever against Durationofsurgery, Maxtemponday1, Group and Whitecellcountonday1 has produced the least AIC score but Whitecellcountonday1 is not significant with its p-value grater than 0.05 at 0.151.

Therefore, we will retain the previously selected model Fever against Durationofsurgery, Maxtemponday1 and Group which has all its covariates to be significant and AIC at 233.780.

Model Fit

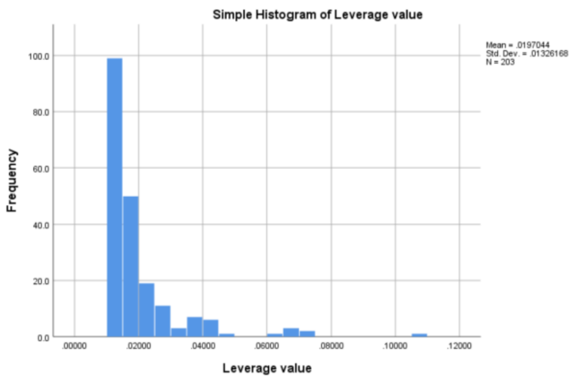


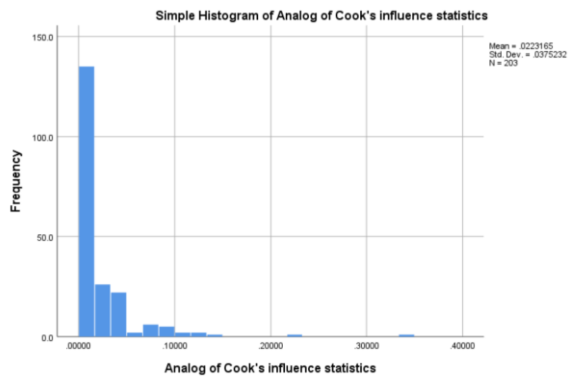
An ideal logistic regression model would produce a non-linear and sigmoidal curve as shown in the first figure. The selected model does not show a perfect sigmoidal curve, especially at the top and bottom, but does show some hints with its curvatures.

On an average, the Exp group (0) is having lower probabilities of fever when compared to Con group (1).

Assignment 3
44468423
Gautham Meenakshisundaram

Outlier Diagnostics





There are no leverage h_{ii} values more than 0.105 which does not satisfy the required moderate leverage values of 0.2 - 0.5 and high leverage values of > 0.5.

There are no Cook's distance C_i values greater than 0.340 which satisfies the requirement of $C_i < 1$.

$DFBETA_{ij} > 2/sqrt(n)$
$DFBETA_{ij} > 2/sqrt(203) = 0.140$
There is only one $DFBETA_{ij}$ value greater than the absolute value of 0.140 at -0.146 but it does not seem too extreme, considering the range of values.

To conclude, there doesn't seem to be any influential observations.

Final model

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Duration of surgery	.403	.115	12.343	1	.000	1.496
	Max temp on Day 1	.889	.333	7.129	1	.008	2.432
	Group2(1)	-.767	.331	5.374	1	.020	.465
	Constant	-34.648	12.541	7.633	1	.006	.000

a. Variable(s) entered on step 1: Duration of surgery, Max temp on Day 1, Group2.

c(iii) provide interpretations of the parameter estimates in your model. Do these agree with your investigations in (i)?

Since Con group is the reference category, the beta coefficient of Group represents the Exp group.

$y_i \sim$ Bernoulli (π_i) where i = 1,….., n

$\ln(\hat{\pi}_i / 1 - \hat{\pi}_i) = - 34.648 - 0.767\, x_{i1} + 0.889\, x_{i2} + 0.403\, x_{i3}$

where,

x_{i1} = {1 if patient belonging to Exp group, 0 otherwise
x_{i2} = Maximum temperature on day 1
x_{i3} = Duration of surgery
$\hat{\pi}_i$ = probability of a patient having Fever(1)

The odds of a patient in Exp group having Fever decreases by 54% ($e^B = e^{-0.767} = 0.465$), when compared to patients in Con group.

> *This effect of Group on Fever was suspected in c(i) where the Chi-Square Test of Association between Fever and Group showed significance at 0.008.*

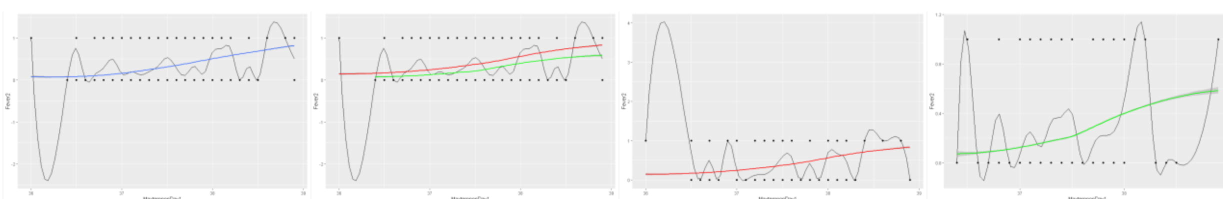
For every increase in degree Celsius in the Maximum temperature on day 1, the odds of a patient having Fever increases by 143% ($e^B = e^{0.889} = 2.432$).

> *This effect of Maximum temperature on day 1 on Fever was suspected in c(i) where the scatterplot showed that the incidence of fever increase as the Maximum temperature on day 1 increases, on an average.*

For every increase in hour in the Duration of surgery, the odds of a patient having Fever increases by 50% ($e^B = e^{0.403} = 1.496$).

> *This effect of Duration of surgery on Fever was suspected in c(i) where the boxplot and scatterplot showed high incidence of fever when Duration of surgery is longer, 3-6 hours to be specific.*

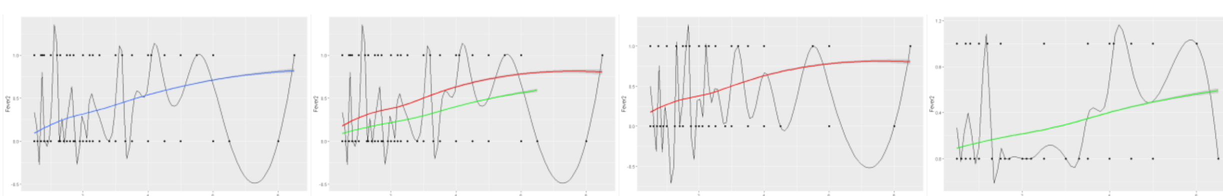
Fever ~ Maxtemponday1 by Group:



The predicted probabilities (blue) of the final model are able to be fitted across the spline which provides a summary of the relationship between Fever and Maxtemponday1.

On an average, the predicted probabilities of Exp group (green) are lower than the predicted probabilities of Con group (red). It indicates that patients in Exp group have less incidence of fever, when compared to Con group.

Fever ~ Durationofsurgery by Group:



The predicted probabilities (blue) of the final model are able to be fitted across the spline which provides a summary of the relationship between Fever and Durationofsurgery.

On an average, the predicted probabilities of Exp group (green) are lower than the predicted probabilities of Con group (red). Again, it indicates that patients in Exp group have less incidence of fever, when compared to Con group.

c(iv) Using your final model, compute the fitted probability of Fever of a patient with the following profile.
Group : control
Gender : male
Age : 45 years
BodyMassIndex : 21.5
Siteofsurgery : abdominal
Durationofsurgery : 2 hours
MaxtempDay1 : 37.2C
WhitecellcountDay1 : 10

$\ln(\hat{\pi}_i / 1 - \hat{\pi}_i) = -34.648 - 0.767\, x_{i1} + 0.889\, x_{i2} + 0.403\, x_{i3}$

From the above model equation, $\hat{n}_i = -34.648 - 0.767\, x_{i1} + 0.889\, x_{i2} + 0.403\, x_{i3}$

$= -34.648 - (0.767 * 0) + (0.889 * 37.2) + (0.403 * 2)$

$\hat{n}_i = -0.771$

Fitted probability of Fever $\hat{\pi}_i = e^n / (1 + e^n)$

$= e^{-0.771} / (1 + e^{-0.771})$

$= 0.463 / (1 + 0.463)$

$\hat{\pi}_i = 0.316$

c(v) Give the classification table, using a cut-off of the prior probability of Fever. What is the sensitivity and specificity?

Fever2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	140	69.0	69.0	69.0
	1.00	63	31.0	31.0	100.0
	Total	203	100.0	100.0	

Prior Probability of Fever = Total number of patients having Fever(1) / Total number of patients

= 63 / 203

= 0.310

Classification Table[a]

			Predicted		
			Fever2		Percentage Correct
	Observed		.00	1.00	
Step 1	Fever2	.00	92	48	65.7
		1.00	17	46	73.0
	Overall Percentage				68.0

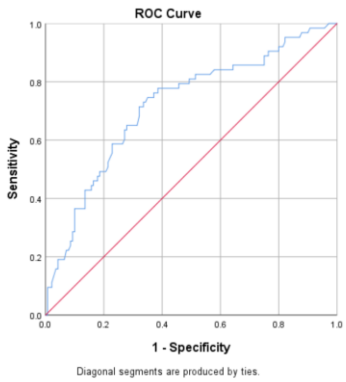
a. The cut value is .310

The number of patients that were correctly predicted as not having Fever(0) over the total number of patients not having Fever(0).
Specificity = 92 / (92 + 48) = 0.66

The number of patients that were correctly predicted as having Fever(1) over the total number of patients having Fever(1).
Sensitivity = 46 / (17 + 46) = 0.73

The model is not perfectly balanced but is just good enough to predict patients not having Fever(0) and having Fever(1), with the cut off probability of 0.310 which is the prior probability of Fever.

c(vi) Construct the ROC curve and give the area under the curve. What does the area under the curve indicate in this case?



Area Under the Curve

Test Result Variable(s): Predicted probability

Area
.719

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Area under curve > 0.5 at 0.719
The model seems to have useful predictive ability, indicating that there is a 72% chance that the model will be able to predict patients without Fever and with Fever correctly.

d) Is there a large difference between the unadjusted and adjusted estimates of treatment effect on occurrence of Fever? Explain your answer.

Unadjusted model

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Group2(1)	-.823	.314	6.866	1	.009	.439
	Constant	-.430	.201	4.583	1	.032	.651

a. Variable(s) entered on step 1: Group2.

Adjusted model

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Group2(1)	-.767	.331	5.374	1	.020	.465
	Duration of surgery	.403	.115	12.343	1	.000	1.496
	Max temp on Day 1	.889	.333	7.129	1	.008	2.432
	Constant	-34.648	12.541	7.633	1	.006	.000

a. Variable(s) entered on step 1: Group2, Duration of surgery, Max temp on Day 1.

All the variables in both the models are significant. Since Con group is the reference category, the beta coefficient of Group represents the Exp group.

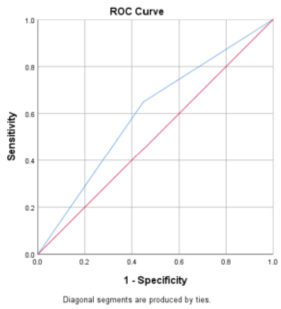
In the unadjusted model, the odds of a patient belonging to Exp group having Fever decreases by 56% ($e^B = e^{-0.823} = 0.439$), when compared to patients in Con group.
In the adjusted model, the odds of a patient belonging to Exp group having Fever decreases by 54% ($e^B = e^{-0.767} = 0.465$), when compared to patients in Con group.

There seems to be a small difference of 2% between the unadjusted and adjusted estimates on patients in Exp group having Fever.

This difference in estimates could be attributed to the additional two covariates in the adjusted model which are predictive of the response variable, resulting in the change of the coefficient of existing covariate in the model. The estimated beta coefficient of an existing covariate in regression model might change if the newly added covariate(s) is correlated / associated with the existing covariate <u>and</u> the response variable.

Assignment 3
44468423
Gautham Meenakshisundaram

Unadjusted model



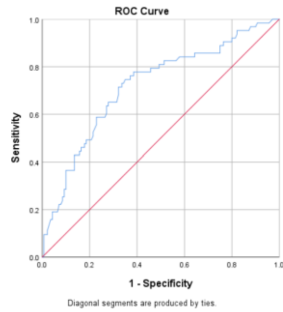
Area Under the Curve

Test Result Variable(s): Predicted probability

Area
.600

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Adjusted model



Area Under the Curve

Test Result Variable(s): Predicted probability

Area
.719

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Moreover, the adjusted model could be preferred as its predictive ability is much better than the unadjusted model.