Niranjan Poudel
STAT 5100
Spring 2019

Homework 6

1) The coefficients estimate from the logistic model fit are as follow:

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.1772	2.9824	0.1558	0.6930
age	1	0.0728	0.0304	5.7401	0.0166
healthawareness	1	-0.0990	0.0335	8.7419	0.0031
male	1	0.4339	0.5218	0.6917	0.4056

From above SAS output the predictive equation for L is as follow:

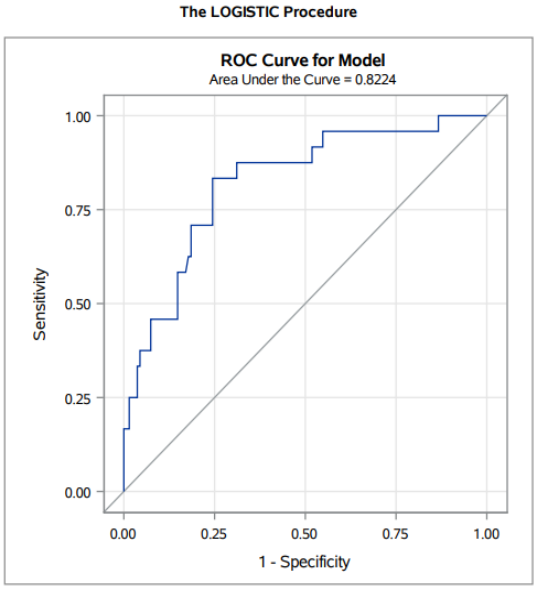
$$L = -1.1772 + 0.0728\,(Age) - 0.0990\,(healthawareness) + 0.4339\,(male)$$

2) Here we have from the SAS output the coefficient estimate for age,

$$b_{age} = 0.0728$$

This can be interpreted as one-unit increase in Age while holding all the other predictor constant will multiply the odds of whether they received a flu shot or not by $e^{0.0728} = 1.076$.

3) The ROC curve for the model is as follow:

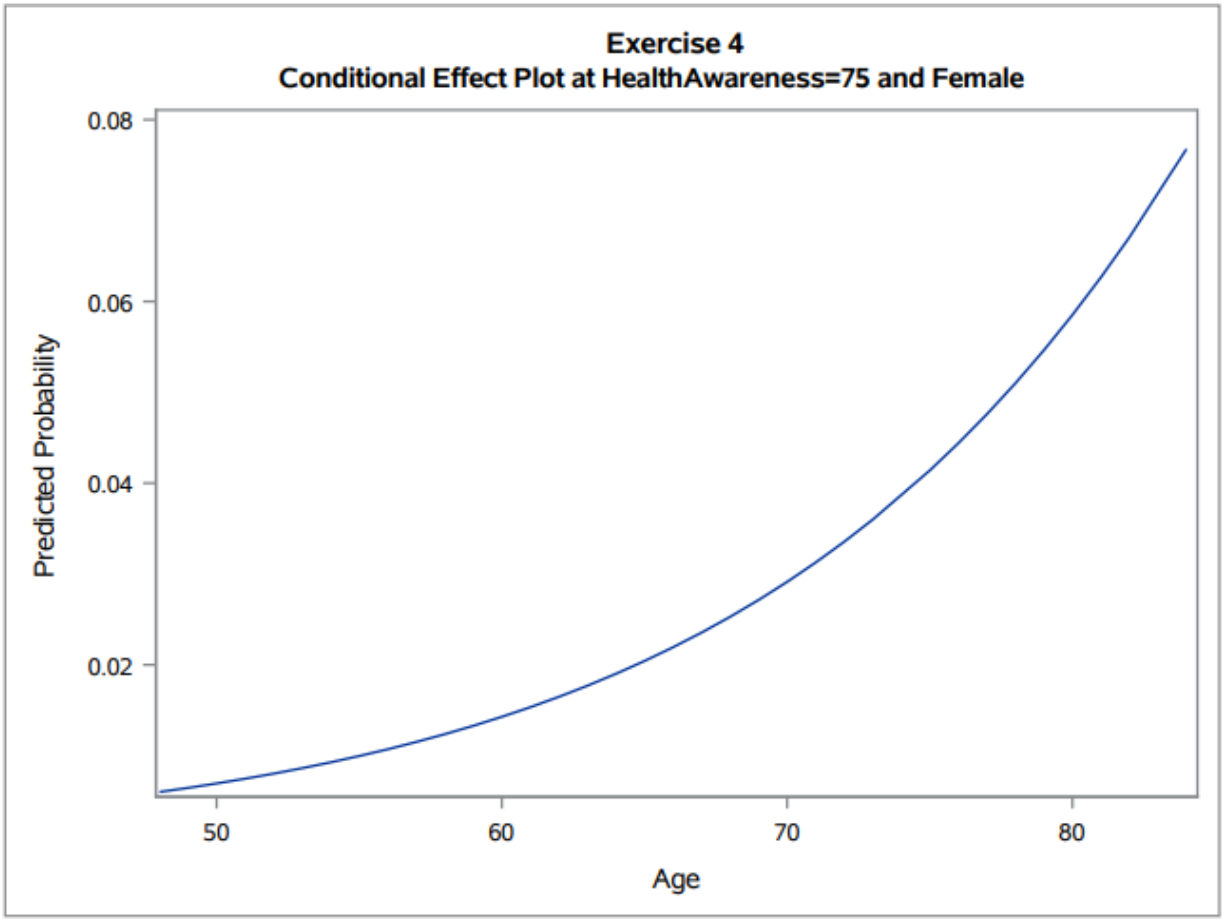


The area under the curve is 0.8224 which is much greater than 0.5 this implies good concordance between the predicted and observed event, the model has decent predictive ability.

Niranjan Poudel
STAT 5100
Spring 2019

Homework 6

4) The condition effect curve is as follow:



From above condition effect curve the predicted probability of getting the flu shot for a female with health awareness 75 and age 59 would get the flu shot increases with the age. The relationship is increasing relationship.

5) For a female with health awareness 75 we have L= -1.1772 + 0.0728 * 59 – 0.0990* 75 = -4.307

So now to transform it to the probability scale we have

$P = 1/ (1 + e^{-L}) = 1/ (1 + e^{4.307}) = 0.013$

Therefore, the predicted probability is 1.3 %.

6) Following is the parameter estimate of the regression of rating on price.

Niranjan Poudel
STAT 5100
Spring 2019

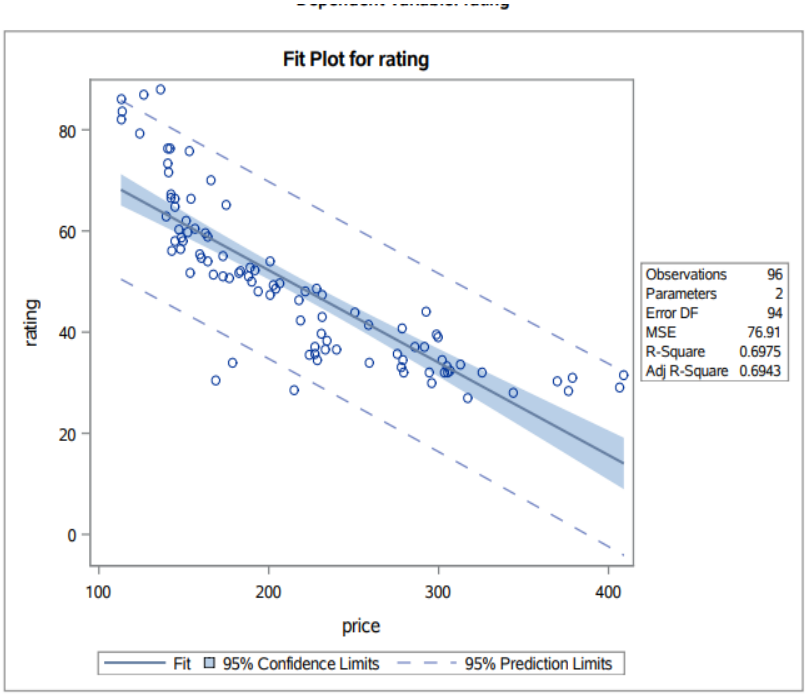
Homework 6

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	88.80015	2.82573	31.43	<.0001
price	1	-0.18281	0.01242	-14.72	<.0001

From the above table the estimated regression function is

Rating = 88.8 – 0.1828 * price

The p-value of the estimate for the coefficient of price is < 0.0001 which is less than 0.005, this implies that there is effect (is significant) of the price in the rating. In the given case the rating has decreasing effect with increase in price. The fit plot is as follow:

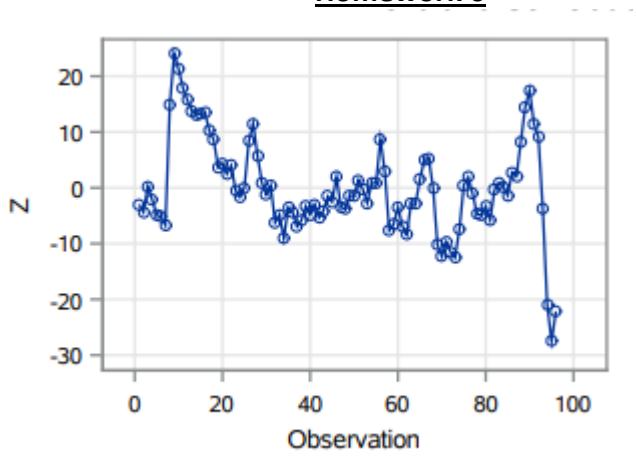


7)

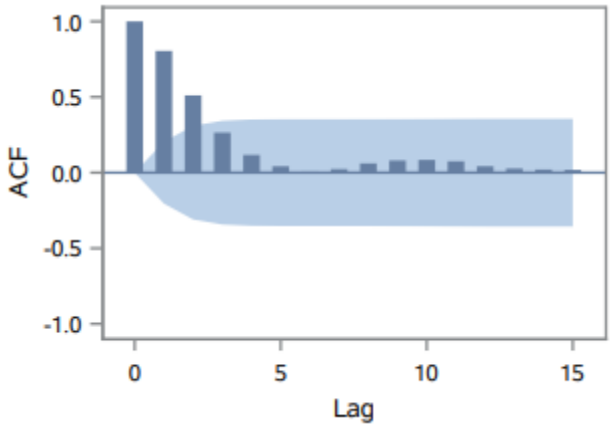
a. The plots of Z vs time, SAC (ACF) of Z and SPAC(PACF) of Z are reported below.

Niranjan Poudel
STAT 5100
Spring 2019

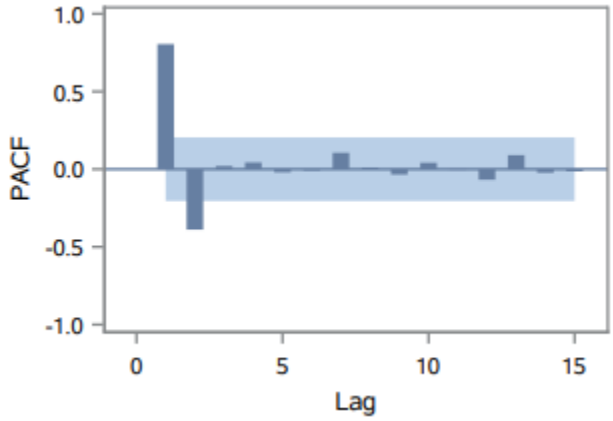
Homework 6



The plot of Z vs time



The plot SAC (ACF) of Z



The plot of PACF of Z

From above plots we can see there is no constant mean and variation of the Z over time these suggests that there is no stationarity in Z from Z vs time plot. The auto correlation plots die down pretty quickly after lag 2 and cuts of fairly quickly shows the data are stationary. The

Homework 6

partial autocorrelation plot does not suggest about the stationarity in Z but is used to detect the dependency.

b. Following is the lag-6 P-value for "Autocorrelation Check for White Noise" for Z

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	98.95	6	<.0001	0.805	0.511	0.264	0.117	0.042	0.007
12	101.68	12	<.0001	0.023	0.061	0.080	0.084	0.075	0.042

From the above auto correlation table, the p-value is < 0.0001 which is less than 0.05 which provides the evidence to suggest that there is dependency structure in the data.

c. The graph of PACF cuts of after lag 2 and ACF graph dies down in an exponential decay but not in a form of sine waves. So from PACF plot we can say that there is AR (2) dependence structure.

8)

a.

i) The p-values for two AR model parameters are as follow:

Unconditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	51.20183	11.60930	4.41	<.0001	0	rating	0
AR1,1	1.10365	0.10399	10.61	<.0001	1	rating	0
AR1,2	-0.13152	0.10500	-1.25	0.2135	2	rating	0
NUM1	-0.01472	0.02049	-0.72	0.4745	0	price	0

For AR (1,1) the p- value is less than 0.001 and for AR (1,2) the p-value is 0.2135. The p-value of 0.2135 is not significant so AR (1,2) can be dropped from the model. The p-value of AR (1,1) is significant.

ii) The diagnostics for the goodness of fits are as follow:

Constant Estimate	1.427211
Variance Estimate	13.44695
Std Error Estimate	3.667008
AIC	528.8884
SBC	539.1458
Number of Residuals	96

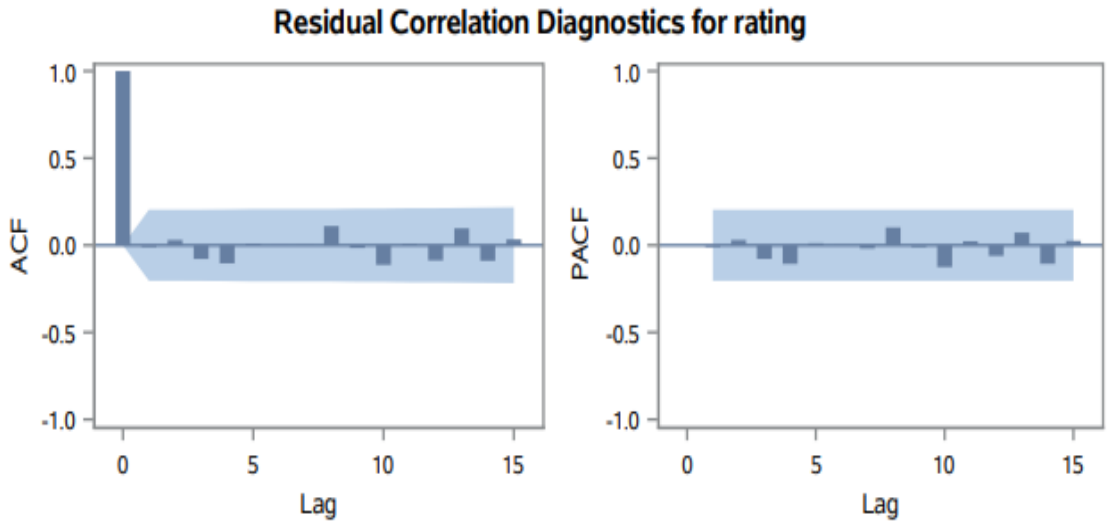
Niranjan Poudel
STAT 5100
Spring 2019

Homework 6

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.77	4	0.7778	-0.007	0.033	-0.076	-0.102	0.011	0.002
12	5.20	10	0.8772	0.001	0.114	-0.010	-0.107	0.013	-0.082
18	11.05	16	0.8061	0.101	-0.084	0.038	-0.072	0.136	0.086
24	15.46	22	0.8419	0.123	-0.114	0.007	-0.068	-0.051	0.008

The **S value** of the diagnostics is 3.667, the Q* value is 1.77 and the p-value associated with it is 0.7778 which suggests the adequacy of the model, suggest that there is no dependence remaining in the model.

iii) The RSAC and RSPAC plots are as follow:

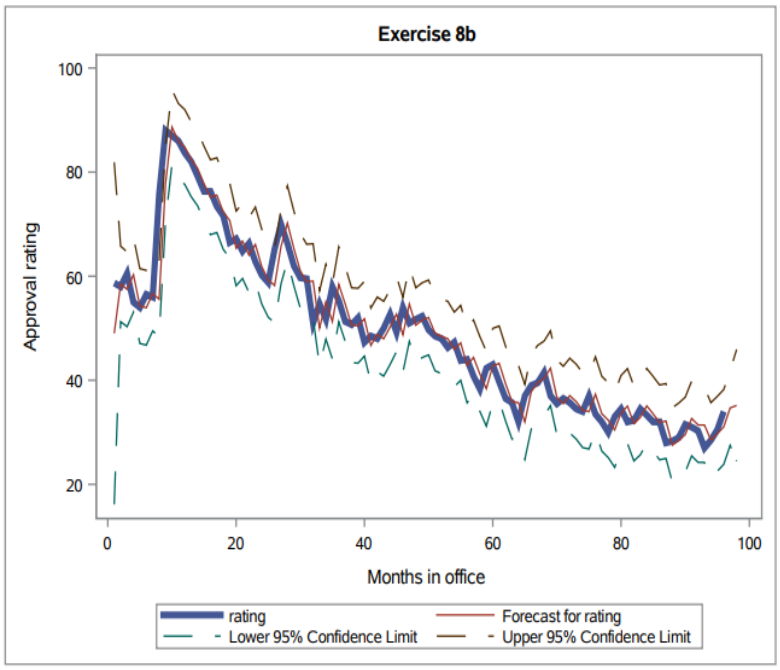


These plots have no spikes which suggests that there is no dependency in the remaining in the model and model is adequate.

b. The forecast plot of ratings vs months in office through March 2009 for model 1, with 95% confidence bounds is as follow:

Niranjan Poudel
STAT 5100
Spring 2019

Homework 6



The plots show that the forecast is somewhat (not totally) consistent with the actual rating which shows a good fit of the model.

c.

Unconditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	51.20183	11.60930	4.41	<.0001	0	rating	0
AR1,1	1.10365	0.10399	10.61	<.0001	1	rating	0
AR1,2	-0.13152	0.10500	-1.25	0.2135	2	rating	0
NUM1	-0.01472	0.02049	-0.72	0.4745	0	price	0

From the above table the p-value of the effect of the price on the rating is 0.4745 which is greater than the significance level of 0.05 which suggest effect of price on rating is not significant.

Niranjan Poudel
STAT 5100
Spring 2019

Homework 6

9) a. The p-value for MA model parameter MA (1,1) is as follow:

Unconditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MU	87.00816	3.26413	26.66	<.0001	0	rating	0
MA1,1	-0.78214	0.06576	-11.89	<.0001	1	rating	0
NUM1	-0.17486	0.01434	-12.19	<.0001	0	price	0

From the above table the p-value for the parameter is <0.0001 which suggest that it is significant and cannot be dropped from the model.

b. The numerical diagnostics and goodness of fit are as follow:

Constant Estimate	87.00816
Variance Estimate	33.98243
Std Error Estimate	5.829445
AIC	614.8154
SBC	622.5085
Number of Residuals	96

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	53.12	5	<.0001	0.447	0.529	0.113	0.185	-0.004	0.070
12	56.48	11	<.0001	0.001	0.099	0.070	0.082	0.096	0.020
18	59.54	17	<.0001	0.077	-0.014	0.092	-0.041	0.062	-0.077
24	62.83	23	<.0001	0.005	-0.133	0.020	-0.051	0.073	-0.015

From the above tables we have S= 5.829445

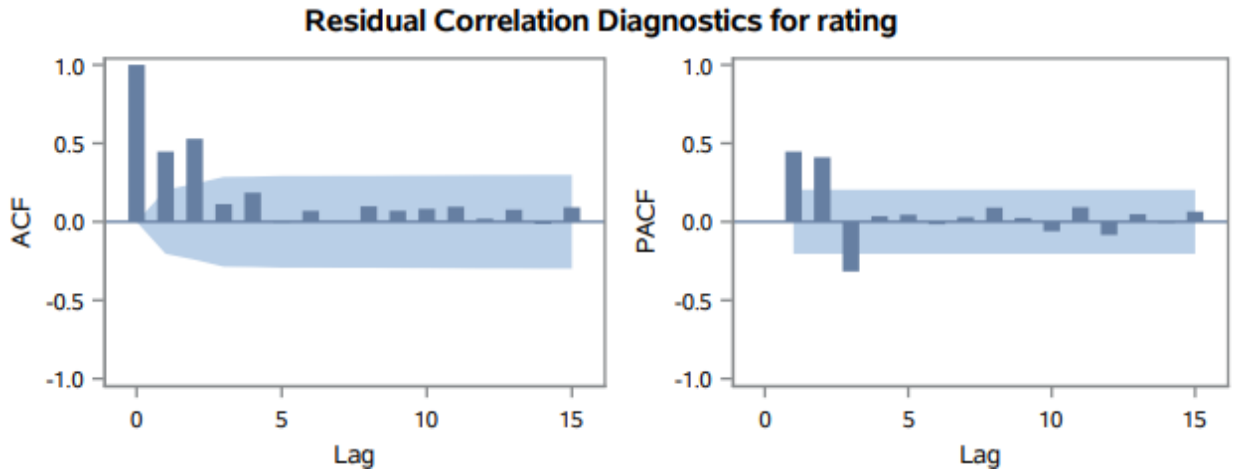
$Q^* = 53.12$

P-value < 0.0001 which is less than 0.05, suggests dependency remaining in the model (is inadequate).

Niranjan Poudel
STAT 5100
Spring 2019

Homework 6

c) The RSAC and RSPAC plot are reported as follow:



The plot of RSPAC and RSAC have spikes remaining which suggest that there is still dependency remaining in the model.

10) There may be significant relationship between the monthly average gas prices and approval ratings if the model is created adequately including all the parameters, but with the two model above the average monthly gas prices there is no significant relationship. In the response of Exercise 6 we say that price is significant is predicting the rating using the simple linear regression model where as in the exercise 8 (c) we look for the time series relation of the significance of the price on the rating, whether it is related on the basis of time with the flow of the moth or not. It might be significant in the simple prediction model but may not be significant in the time series model.