Time Series Project
SOA Exam Statistics

I used concepts from Time Series to examine SOA preliminary exam statistics. This analysis includes development of correlograms of initial time series for a number of statistics as well as examination of correlograms of first and second differences. Further correlogram analysis was performed examining the effect of interventions on the time series. This projects objective is the same as time series analysis in general: to develop a compact description of data and test hypotheses.

Data: I used data provided by the SOA at http://www.soa.org/files/pdf/edu-archive-exam-per-results.pdf. I examined the data for the different exams and decided to focus on the preliminary exams as they are offered most frequently. The preliminary exams have changed some during the period this data covers, and I decided to only consider data from May 2000 to November 2008 (the most recent data provided). The preliminary exams over this period were 1/P, 2/FM, 3/M/MLC&MFE, and 4/C. I decided to eliminate exam 1/P from consideration in my data because while the other exams have only been offered in May/November, this exam is offered at other dates. I intended to examine pass rates and number of people taking the exam, and I hypothesized that the number of people taking an exam at a given level would be influenced by the number of hours the exam takes. I therefore examined the number of hours each of these tests took over the 18 testing periods in my data period and found the following:

	Exam Hours		
	fm/2	m/3	c/4
Nov-08	3.0	5.0	4.0
May-08	2.5	5.0	4.0
Nov-07	2.5	5.0	4.0
May-07	2.5	5.0	4.0
Nov-06	2.0	4.0	4.0
May-06	2.0	4.0	4.0
Nov-05	2.0	4.0	4.0
May-05	2.0	4.0	4.0
Nov-04	4.0	4.0	4.0
May-04	4.0	4.0	4.0
Nov-03	4.0	4.0	4.0
May-03	4.0	4.0	4.0
Nov-02	4.0	4.0	4.0
May-02	4.0	4.0	4.0
Nov-01	4.0	4.0	4.0
May-01	4.0	4.0	4.0
Nov-00	4.0	4.0	4.0
May-00	4.0	4.0	4.0

Based on my knowledge of the exams' hours and administration histories, I decided to focus on two of the exams: exam FM/2 and exam C/4. Exam C/4 remained 4 hours over this entire period while FM/2's hours changed with the transition from course 2 to exam FM and then twice more during the period of exam FM.

I developed four initial statistics for each of these two exams: the passing percentage (the number of people who passed the exam divided by the number who took the exam), the effective passing percentage (the number of people who passed the exam divided by the number of people who took the exam and got at least half of the score needed to pass), the percent effective (the percentage of those who took the exam who got at least half of the score needed to pass) and the number of people taking the exam. This data is given in the following tables:

Passing Percentage		
	fm/2	c/4
Nov-08	50.0%	43.6%
May-08	48.6%	47.0%
Nov-07	43.9%	49.9%
May-07	47.7%	42.7%
Nov-06	43.9%	56.4%
May-06	45.3%	53.2%
Nov-05	40.9%	50.6%
May-05	73.2%	52.8%
Nov-04	41.5%	51.4%
May-04	25.8%	50.0%
Nov-03	39.2%	51.1%
May-03	35.5%	50.5%
Nov-02	49.3%	57.6%
May-02	37.2%	44.3%
Nov-01	40.7%	42.7%
May-01	32.0%	40.6%
Nov-00	32.2%	37.0%
May-00	26.7%	33.8%
avg	41.9%	47.5%

Effective Passing Percentage

	fm/2	c/4
Nov-08	54.6%	45.3%
May-08	52.6%	49.4%
Nov-07	48.7%	51.8%
May-07	52.6%	44.9%
Nov-06	47.5%	59.3%
May-06	49.1%	56.0%
Nov-05	46.9%	52.7%
May-05	75.9%	55.0%
Nov-04	44.5%	53.9%
May-04	27.9%	52.1%
Nov-03	40.7%	53.5%
May-03	37.6%	51.8%
Nov-02	51.8%	60.3%
May-02	40.1%	47.7%
Nov-01	43.0%	46.5%
May-01	35.2%	43.3%
Nov-00	33.6%	41.0%
May-00	29.4%	37.3%
avg	45.1%	50.1%

Percent Effective

	fm/2	c/4
Nov-08	91.6%	96.3%
May-08	92.4%	95.1%
Nov-07	90.2%	96.2%
May-07	90.7%	95.0%
Nov-06	92.5%	95.1%
May-06	92.3%	95.1%
Nov-05	87.1%	96.1%
May-05	96.5%	96.1%
Nov-04	93.4%	95.5%
May-04	92.7%	95.9%
Nov-03	96.1%	95.5%
May-03	94.3%	97.4%
Nov-02	95.3%	95.6%
May-02	92.9%	92.9%
Nov-01	94.6%	92.0%
May-01	90.9%	93.8%
Nov-00	95.8%	90.1%
May-00	91.0%	90.7%
avg	92.8%	94.7%

	Number Taking	
	fm/2	c/4
Nov-08	3,968	1,698
May-08	4,847	1,848
Nov-07	3,792	1,857
May-07	4,043	2,079
Nov-06	4,444	2,050
May-06	4,824	2,119
Nov-05	4,436	1,785
May-05	5,261	1,573
Nov-04	3,525	2,006
May-04	3,656	1,728
Nov-03	3,356	1,610
May-03	2,710	1,215
Nov-02	2,758	1,283
May-02	2,549	1,272
Nov-01	2,115	1,149
May-01	2,115	1,008
Nov-00	1,952	963
May-00	1,903	913
avg	3,459	1,564

For each exam, there are thus 18 bi-annual values for each of the statistics being considered for examination for 9 years of data with 9 values each for November exams and for May exams. The first 10 dates for each exam correspond to Course 2/4 while the last 8 correspond to Exam FM/M.

Some data adjustments are discussed in later sections, but it's important to note that one of the purposes of the project is not to develop the optimal data. Therefore, I have not focused on possible areas where data may require adjustment, such as the anomalous May 2005 percentage passing and effective percentage passing for exam FM.

SEASONALITY:

I examined whether there was seasonality in the data for the exams in several ways. First, I developed the average for all 18 dates for each of the four absolute statistics, developed the average for the November dates for each of the four absolute statistics and developed the average for the May dates for each of the four absolute statistics. I then divided each half-year's pass rate by the average pass rate to get half-year relativities.

ABSOLUTES

Passing Percentage Average	fm/2	c/4	Effective Passing %Average	fm/2	c/4
November	42.4%	48.9%	November	45.7%	51.6%
May	41.3%	46.1%	May	44.5%	48.6%
Total	41.9%	47.5%	Total	45.1%	50.1%
Divided by Total Average					
	1.0127389	1.0297007		1.0134286	1.029718
	0.9872611	0.9702993		0.9865714	0.970282

Percent Effective Average	fm/2	c/4	Number Taking Average	fm/2	c/4
November	92.9%	94.7%	November	3,372	1,600
May	92.6%	94.7%	May	3,545	1,528
Total	92.8%	94.7%	Total	3,459	1,564
Divided by Total Average					
	1.0015906	1.0001594		0.9749092	1.022944
	0.9984094	0.9998406		1.0250908	0.977056

The bi-annual relativities for the absolute statistics were all within 3% of the total average, so from the absolute statistics, it does not appear that there is seasonality within any of this data. Because the data is highly stochastic, the use of all the years to adjust the relativities rather than just using single years is justified.

I examined the same information for the first differences of these statistics:

FIRST DIFFERENCES

Passing Percentage Average

	fm/2	c/4
November	1.1%	2.8%
May	1.7%	-2.0%
Total	1.4%	0.6%
Divided by Total Average		
	0.7782546	4.8956916
	1.2494635	-3.382653

Effective Passing %Average

	fm/2	c/4
November	1.2%	3.0%
May	1.8%	-2.4%
Total	1.5%	0.5%
	0.8170194	6.327778
	1.2058532	-4.99375

Percent Effective Average

	fm/2	c/4
November	0.3%	0.0%
May	-0.3%	0.7%
Total	0.0%	0.3%
Divided by Total Average		
	9.2371753	0.0912456
	-8.266822	2.0223487

Number Taking Average

	fm/2	c/4
	(174)	72
	453	17
Total	121	46
	-1.428787	1.554423
	3.732385	0.376274

As there are only two values per year, care must be taken not to draw too much information from this, but it does appear that for exam C, performance is more likely to improve in November than May for both the passing percentage and the effective passing percentage while there is a slight tendency in the opposite direction for exam FM. The percent effective did not exhibit the same behavior for the two exams. For exam FM, an increase in the percent effective is more likely to occur in November while for exam C it is more likely in May. The difference is more pronounced for exam FM. For both exams, growth in the number of people taking it seems slightly more pronounced for the May exams than for the November exams.

The seasonality can also be examined using the sample autocorrelations. In particular, the cyclical nature of the differences of some of the statistics will be explored in the correlograms. As an example of this, for the first differences of the pass rate for exam FM, the autocorrelation for each of the May exams are negative while with the exception of the first November autocorrelation, all the November ones are positive.

Support is thus given from the concept of seasonality towards considering a seasonal adjustment of AR(2) for these data sets. An alternative would be to attempt to detrend the data rather than to make a seasonal adjustment. This could be done by subtracting 0.33% from each May FM score and subtracting 2.39% from each November exam C score for both the passing percentage and subtracting 0.29% from each May effective passing percentage for FM and subtracting 2.66% from each November effective percentage for C. These are examples of possible adjustments, and are not supposed to indicate the ideal method. Using these examples, the following detrended data is produced:

DETRENDED DATA

	Passing Percentage fm/2	Passing Percentage c/4		Effective Passing Percentage fm/2	Effective Passing Percentage c/4
Nov-08	50.0%	41.2%	Nov-08	54.6%	42.6%
May-08	48.3%	47.0%	May-08	52.3%	49.4%
Nov-07	43.9%	47.5%	Nov-07	48.7%	49.1%
May-07	47.4%	42.7%	May-07	52.3%	44.9%
Nov-06	43.9%	54.0%	Nov-06	47.5%	56.6%
May-06	45.0%	53.2%	May-06	48.8%	56.0%
Nov-05	40.9%	48.2%	Nov-05	46.9%	50.0%
May-05	72.9%	52.8%	May-05	75.6%	55.0%
Nov-04	41.5%	49.0%	Nov-04	44.5%	51.2%
May-04	25.5%	50.0%	May-04	27.6%	52.1%
Nov-03	39.2%	48.7%	Nov-03	40.7%	50.8%
May-03	35.2%	50.5%	May-03	37.3%	51.8%
Nov-02	49.3%	55.2%	Nov-02	51.8%	57.6%
May-02	36.9%	44.3%	May-02	39.8%	47.7%
Nov-01	40.7%	40.3%	Nov-01	43.0%	43.8%
May-01	31.7%	40.6%	May-01	34.9%	43.3%
Nov-00	32.2%	34.6%	Nov-00	33.6%	38.3%
May-00	26.4%	33.8%	May-00	29.1%	37.3%
avg	41.7%	46.3%	avg	44.9%	48.8%

With this detrended data, the averages for both the absolute statistics and the first differences can be recomputed:

ABSOLUTES

	Passing Percentage Average fm/2	Passing Percentage Average c/4		Effective Passing %Average fm/2	Effective Passing %Average c/4
November	42.4%	46.5%	November	45.7%	48.9%
May	41.0%	46.1%	May	44.2%	48.6%
Total	41.7%	46.3%	Total	44.9%	48.8%
Divided by Total Average					
	1.016746	1.004666		1.016698	1.003258
	0.983254	0.995334		0.983302	0.996742

FIRST DIFFERENCES

	Passing Percentage Average fm/2	Passing Percentage Average c/4		Effective Passing %Average fm/2	Effective Passing %Average c/4
November	1.4%	0.4%	November	1.5%	0.3%
May	1.4%	0.4%	May	1.5%	0.3%
Total	1.4%	0.4%	Total	1.5%	0.3%
Divided by Total Average					
	1.004796	0.991603		1.001133	1.011652
	0.994604	1.009447		0.998725	0.986891

An area for possible further study relating to seasonality would be examining if the seasonality differs between the effective passing rate and the passing rate. This project's preliminary work indicates similar patterns, but this could be an area of interesting further study.

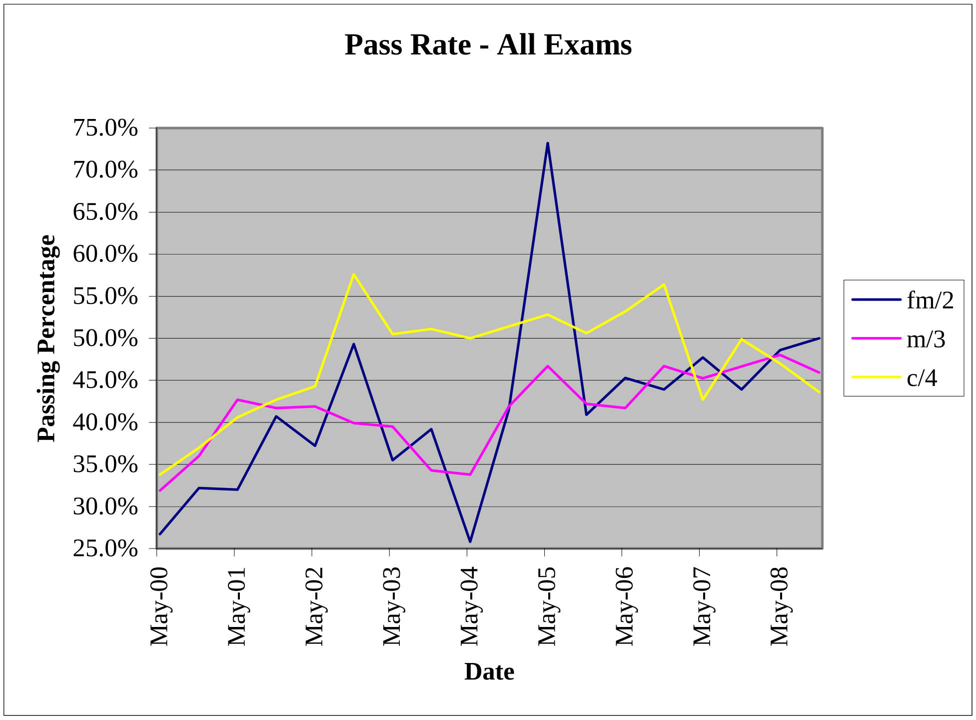
GRAPHS of ABSOLUTE DATA:

For the graphs, I have also included exam 3/M/MLC&MFE. The second intervention in the structure of this exam is the reason I opted not to focus on this exam for this project, but I did explore the best ways to deal with the data. For the percentage passing and effective, weighted averages seemed the obvious solution, but deciding on the best data to examine the percentage taking the exam in 2007 and 2008 when it was split into two pieces presented more of a challenge. I examined options including simply adding the people taking the two exams, taking the greater of the taking the exams and weighting the number of people taking the exam by the total number of hours. In the end, I decided that using the greater number yields the best results. In addition to qualitative assessment, I based this decision on the relative growth rates of the number of people taking 3/M/MLC&MFE under each of the options to the relative growth rates of exams FM and C:

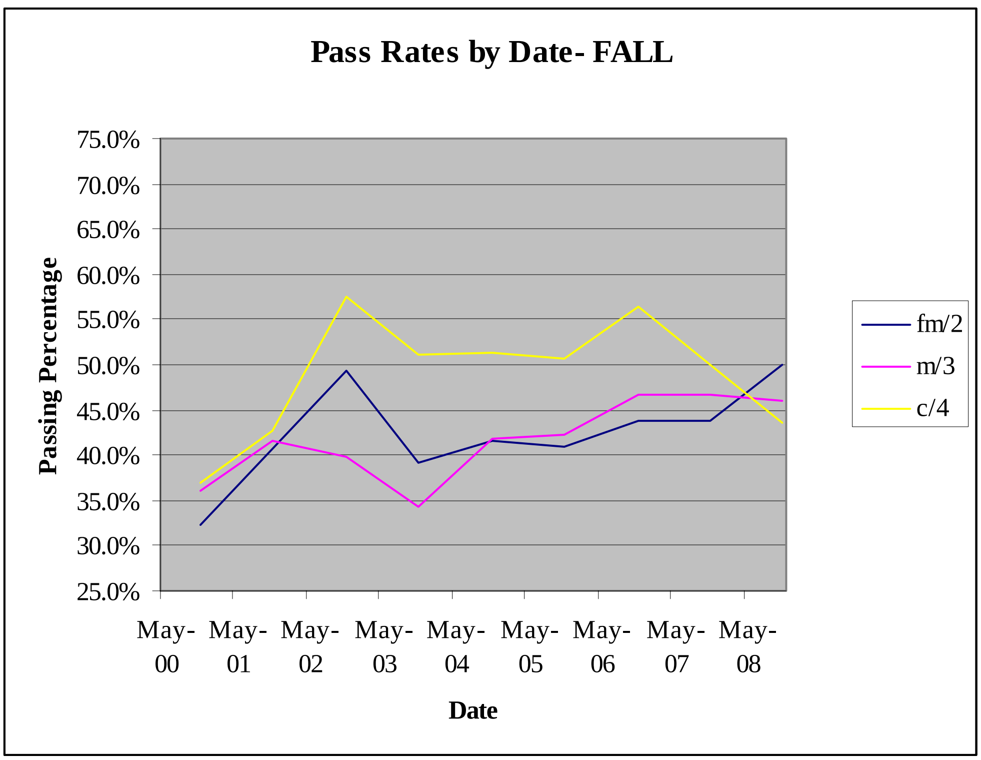
	FM	C	C/FM Average	m add	m greater of	m .8 ratio	
0							
1	1.10	1.15	1.12	1.05	1.05	1.05	
2	1.38	1.36	1.37	1.25	1.25	1.25	
3	1.57	1.51	1.54	1.46	1.46	1.46	M runs behind the
4	1.86	1.99	1.93	1.42	1.42	1.42	other two exams for
5	2.52	1.79	2.15	1.60	1.60	1.60	all the data period
6	2.40	2.22	2.31	2.05	2.05	2.05	before the split
7	2.03	2.10	2.07	2.75	1.60	2.20	
8	2.29	1.92	2.11	3.54	1.90	2.84	
from yr 6							
1	0.85	0.94	0.89	1.34	0.78	1.07	
2	0.95	0.87	0.91	1.73	0.93	1.38	Since M has run
							behind before and the
from yr 7							other two exams have
1	1.13	0.92	1.02	1.29	1.19	1.29	low or negative
							growth rates, the
							greater of option
							makes the most sense.

The above data is based on annual number of people taking the exam, combining one May and one November exam for each year. I also examined the data for just the November exams and just the May exams and found that the greater of adjustment appeared the strongest despite intuitively wanting the .8 times the total taking the two exams to work reflecting the fact that the two exams together are 5 hours in contrast to the 4 hours of exams they replaced.

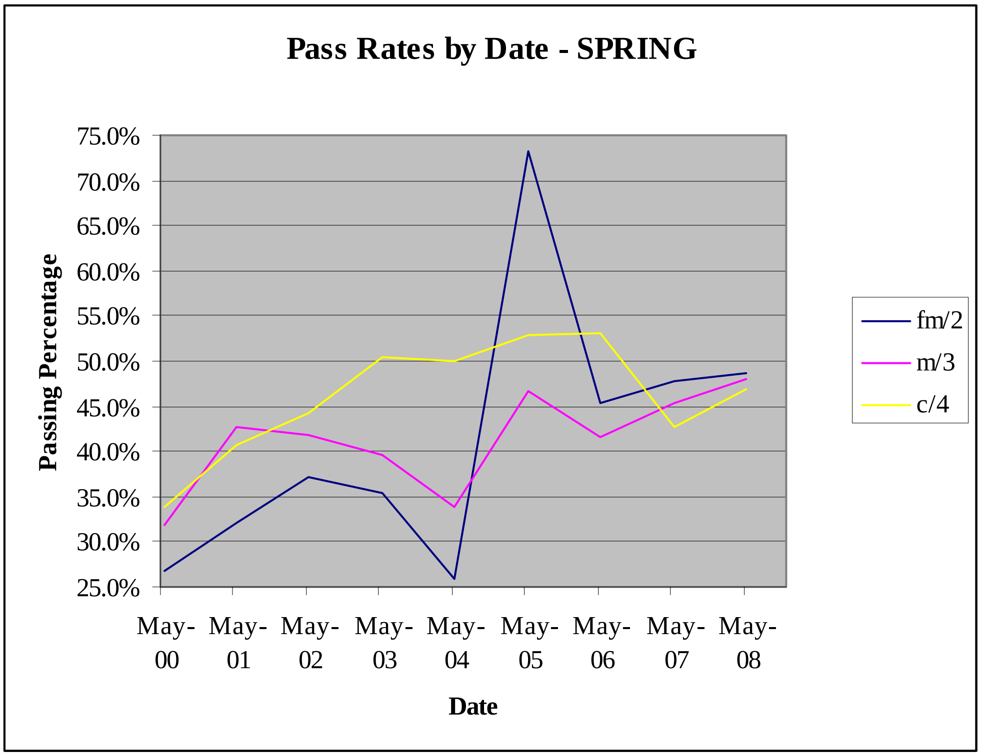
The data for exam 3/M/MLC&MFE is thus included in these graphs as described above.



Pass Rate - All Exams
75.0%
70.0%
65.0%
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
Passing Percentage
May-00
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4

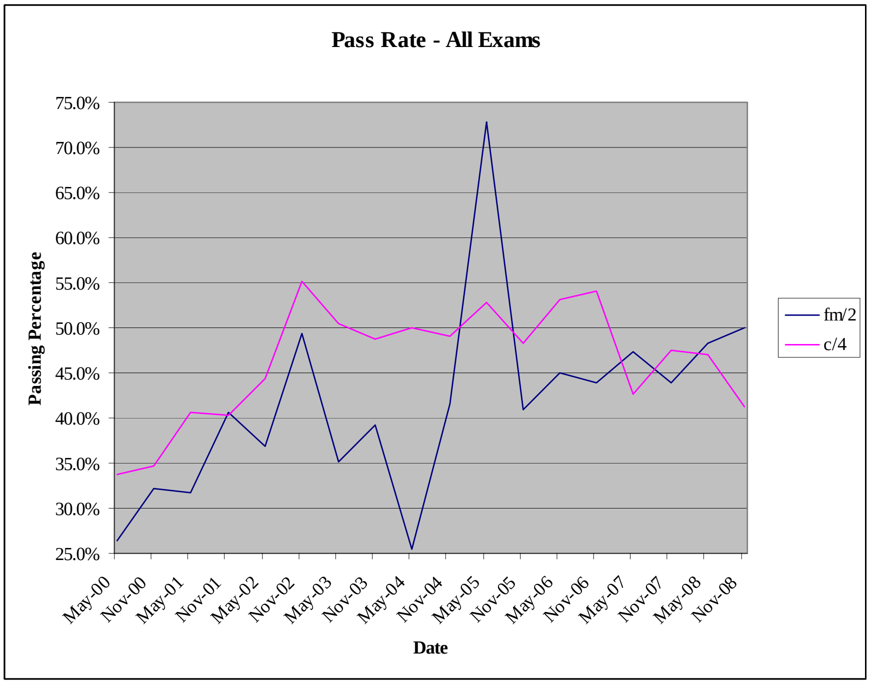


Pass Rates by Date- FALL
75.0%
70.0%
65.0%
60.0%
55.0%
50.0%
45.0%
40.0%
35.0%
30.0%
25.0%
Passing Percentage
May-00
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4

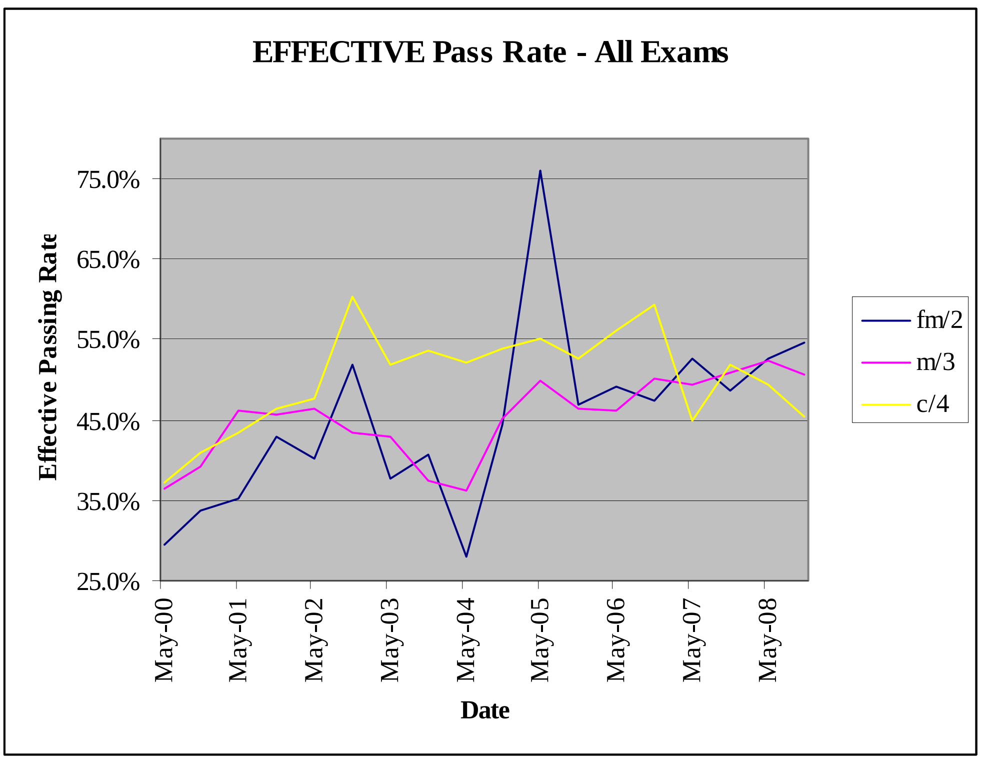


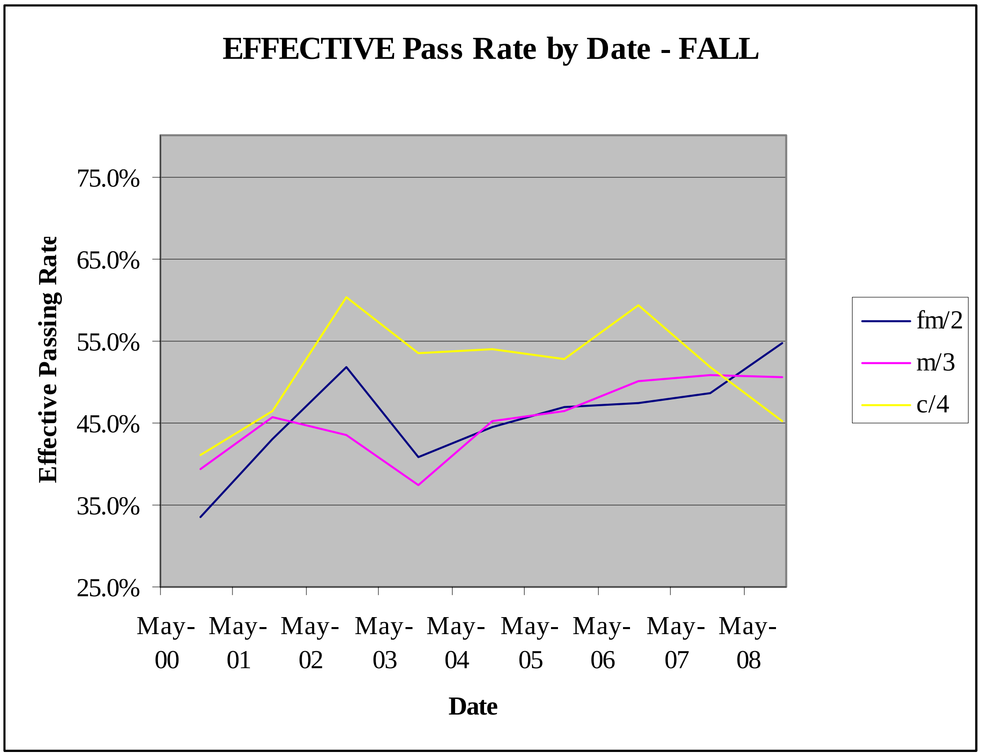
These first three graphs show the passing percentage for each exam: the first graph showing all dates, the second just the November dates and the last just the May dates. At first glance, the spring dates seem more volatile, but removing the previously discussed anomalous May 2005 value results in reasonably similar degrees of volatility between the fall and spring exams. In all three graphs, M/3 has the least variability. Exam FM/2 has the most variability and has both the highest and the lowest pass rates.

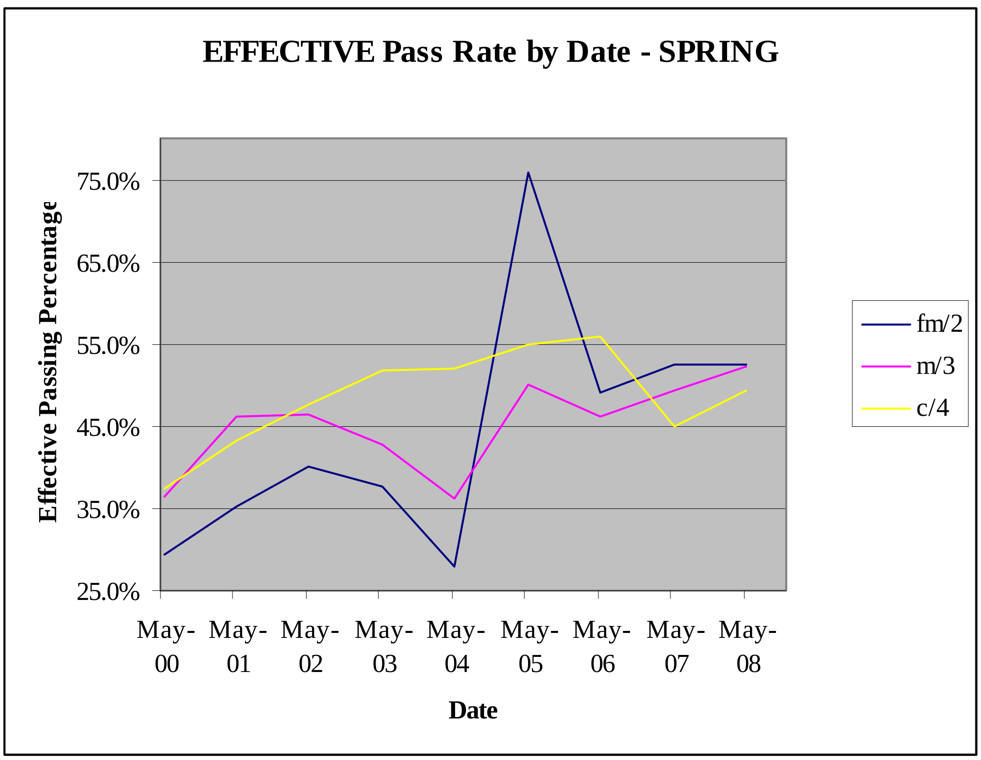
The suggested possible detrended data for exams FM and C were also graphed and can be compared to the unadjusted data:



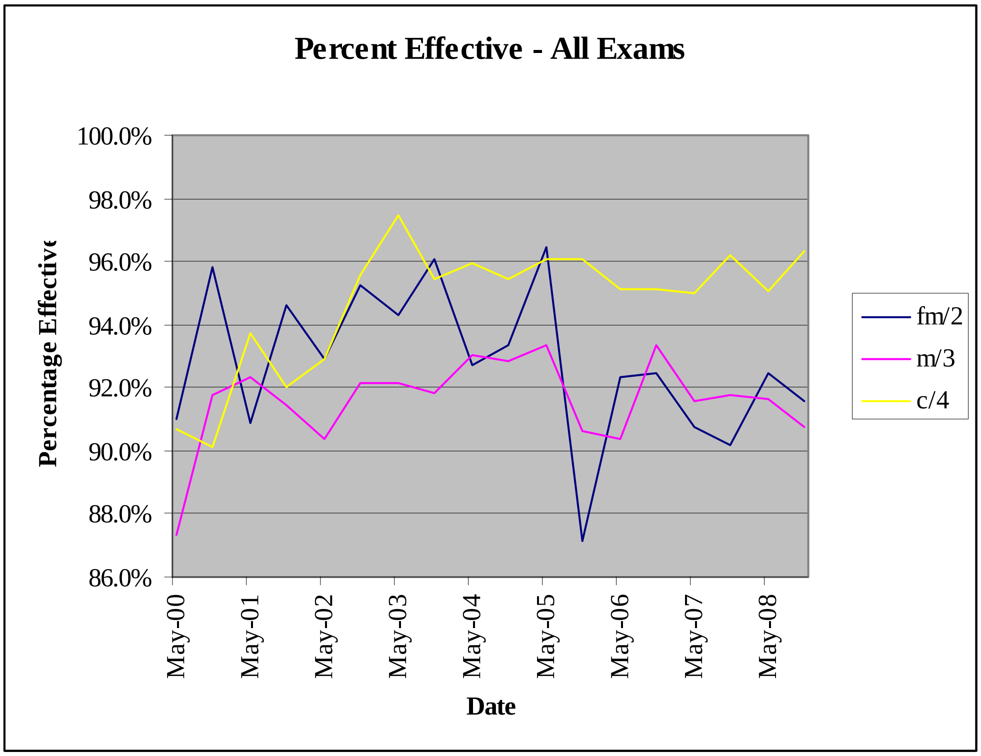
The patterns of the effective passing percentage are very similar to those of the total pass rates. This can be seen in the striking similarity of the shape of the following graphs of the effective passing percentages to the above passing percentages. The basic shapes of the graphs only varies around November 2005.

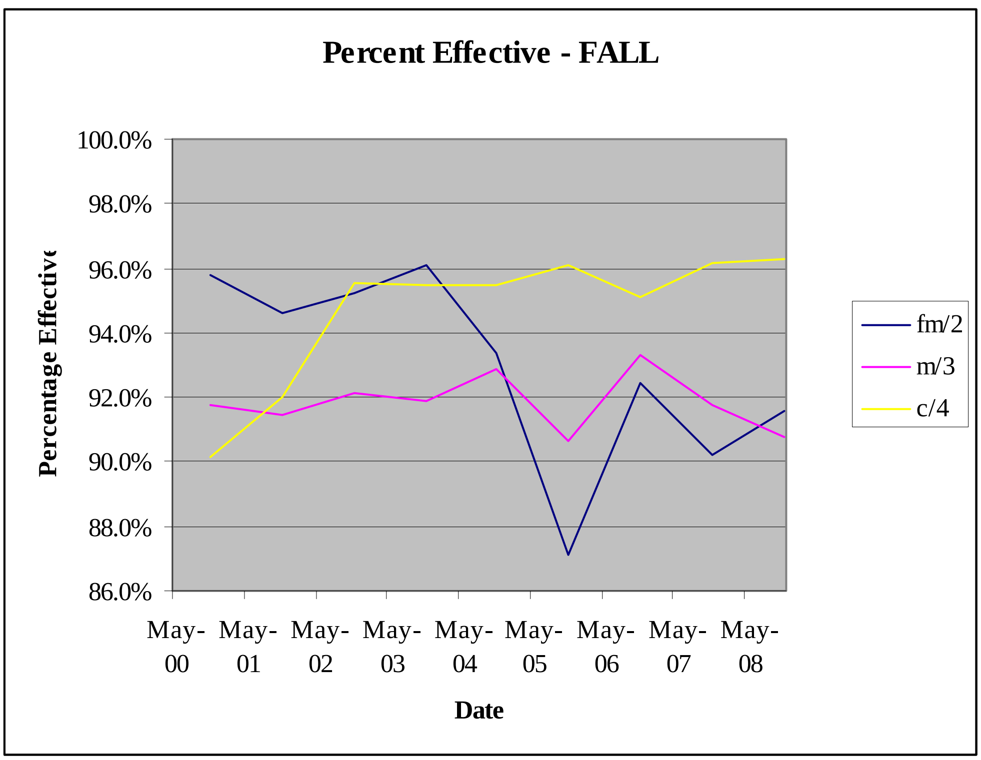


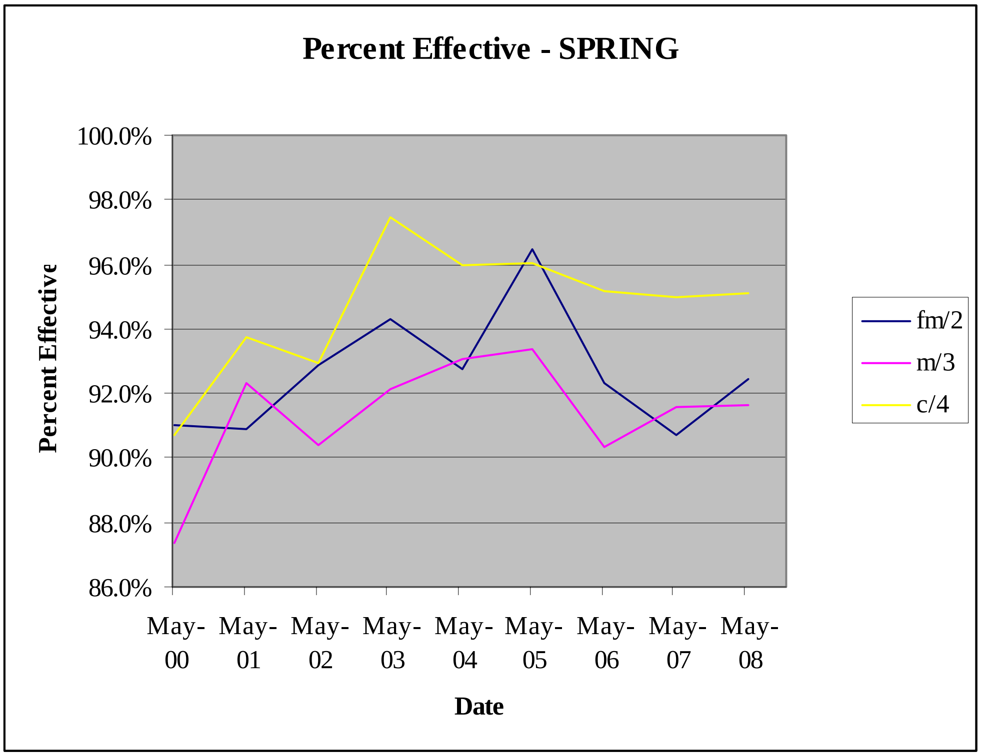




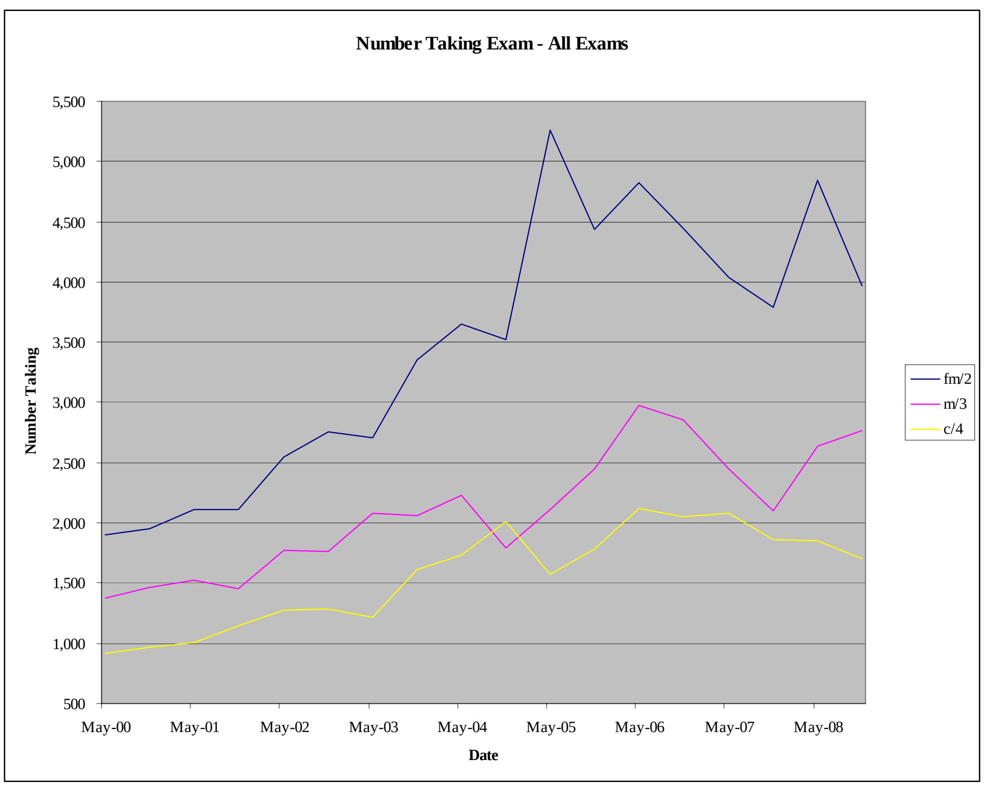
The percent effective shows differences between all the dates, the fall dates only and the spring dates only. These relate to the seasonality ideas previously addressed.

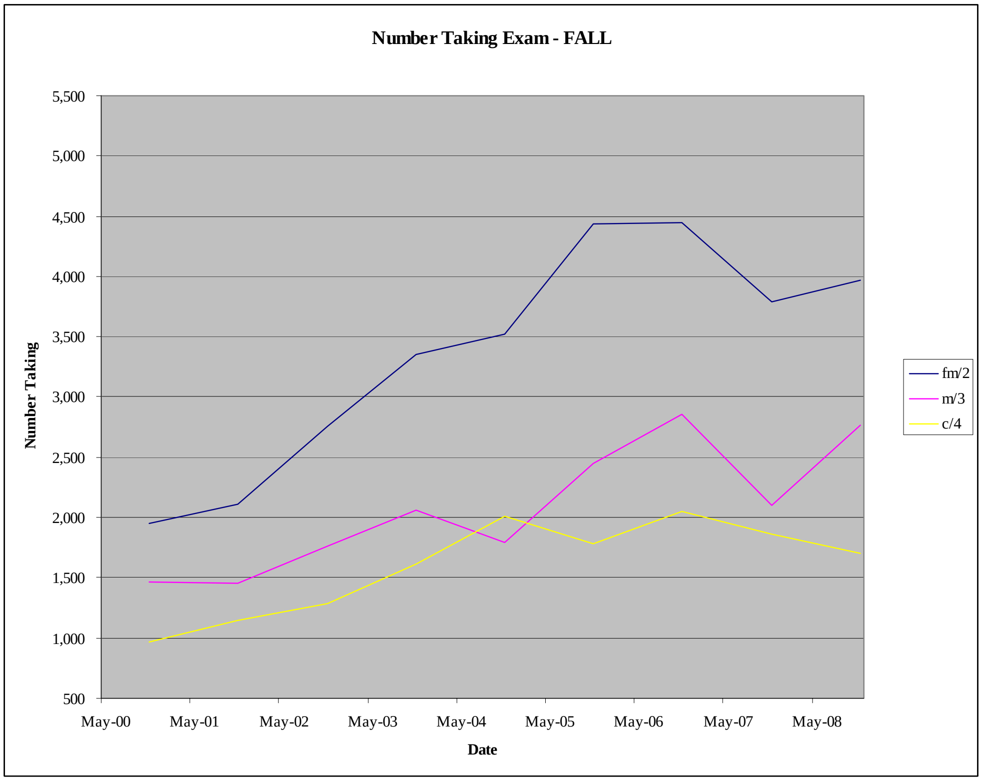

Percent Effective - All Exams
100.0%
98.0%
96.0%
94.0%
92.0%
90.0%
88.0%
86.0%
Percentage Effective
May-00
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4

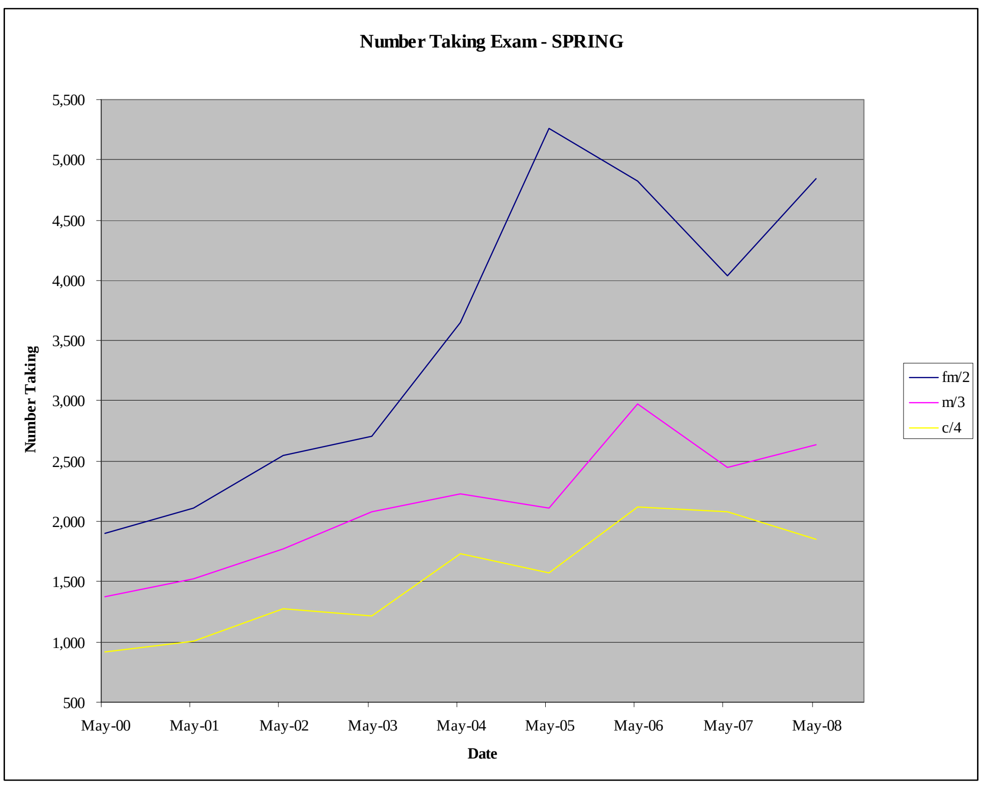

Percent Effective - FALL
100.0%
98.0%
96.0%
94.0%
92.0%
90.0%
88.0%
86.0%
Percentage Effective
May-00
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4



The data relating to the number of people taking each exam was graphed two ways, based on the actual numbers first and based on the logs of the number second.



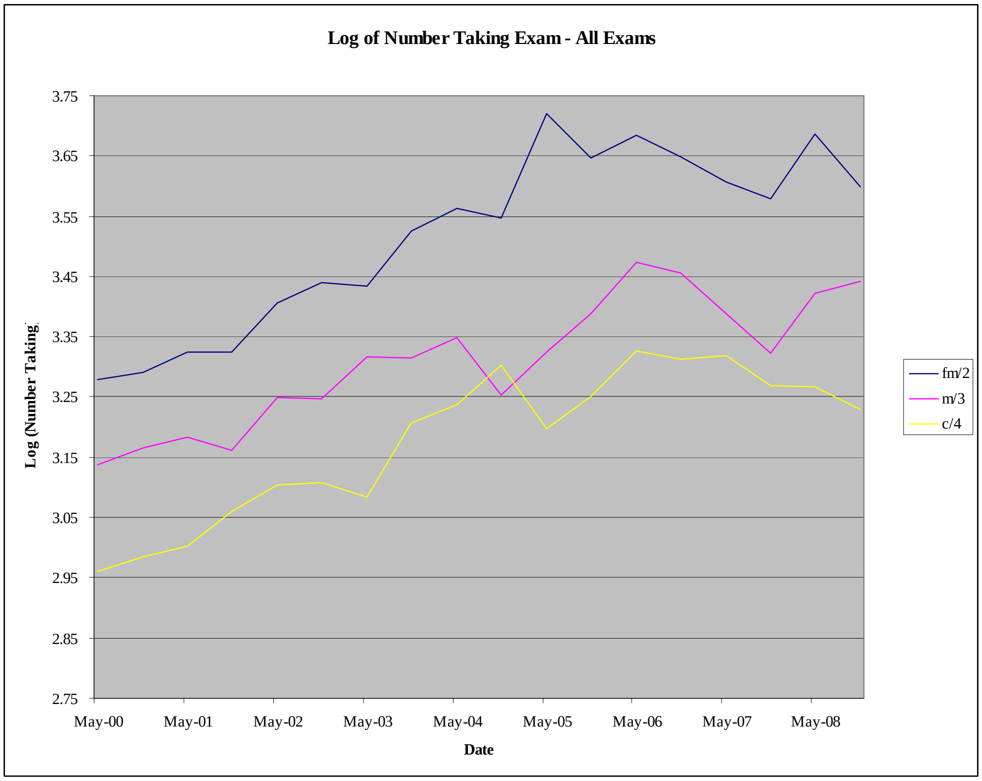

Number Taking Exam - FALL
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
Number Taking
May-00
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4

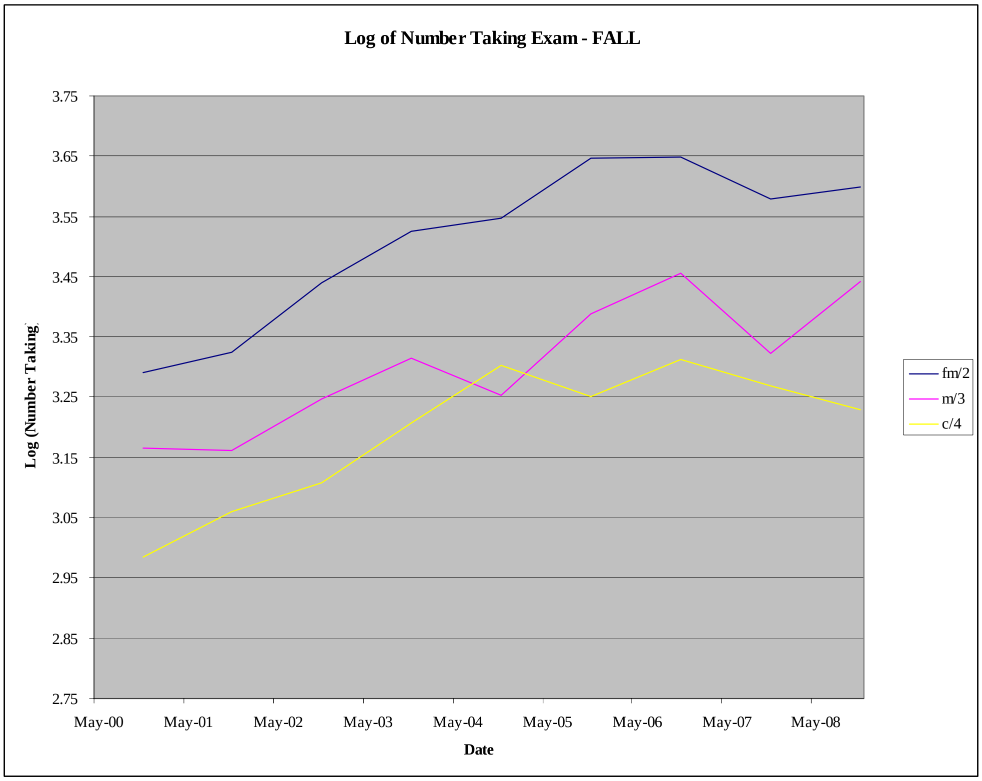

Number Taking Exam - SPRING
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
Number Taking
May-00
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4

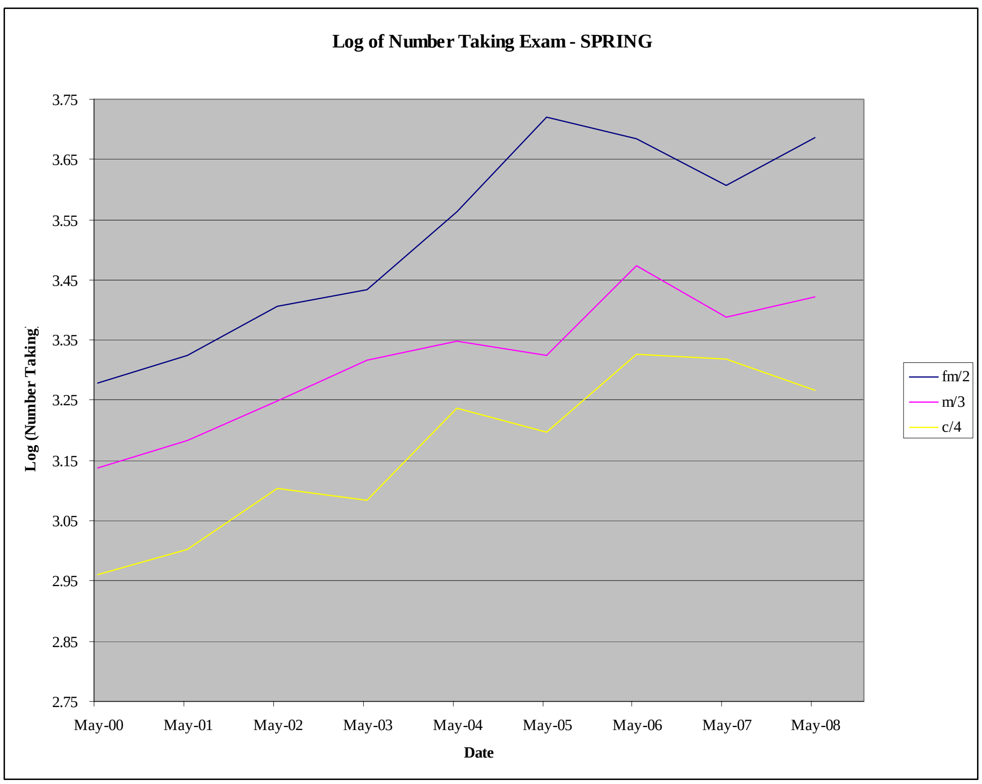
The graph of all the dates shows that the general trend of all the exams has been to generally increase the number of people taking the exams with some fall off in the most recent years. Exam FM/2 appears to have been increasing at the faster rate.

A comparison of the fall and spring graphs shows a steeper rate of change generally for the spring exams which supports the analysis previously presented in the seasonality section.

These trends can be examined further with graphs of the log of the number of people taking each exam on each date.



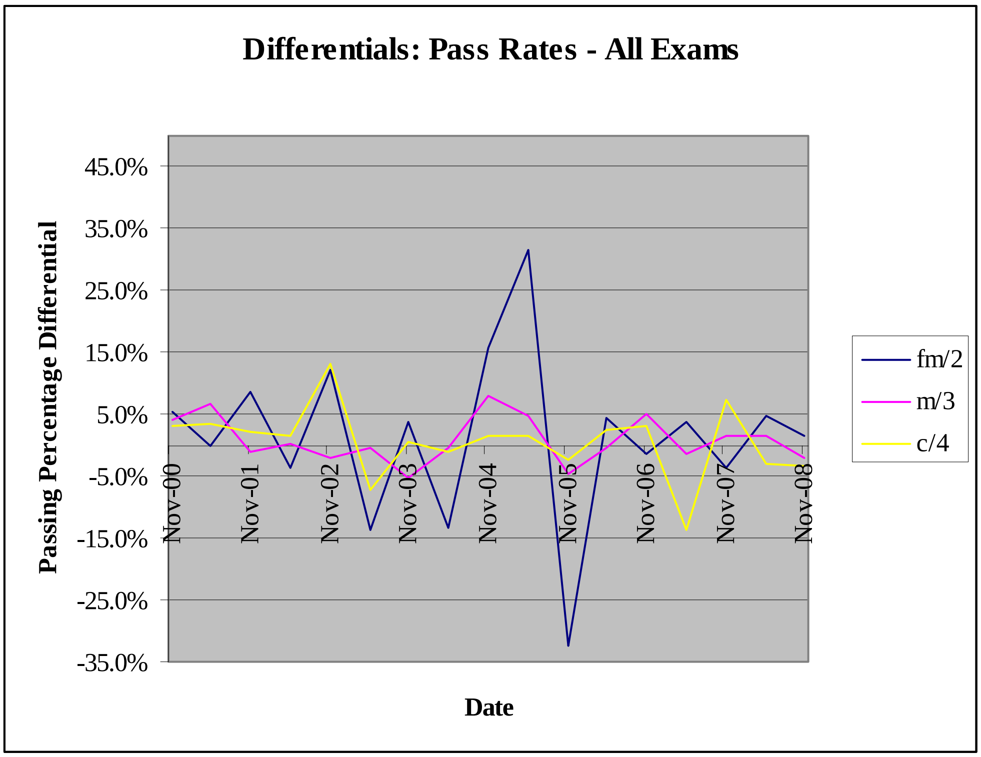


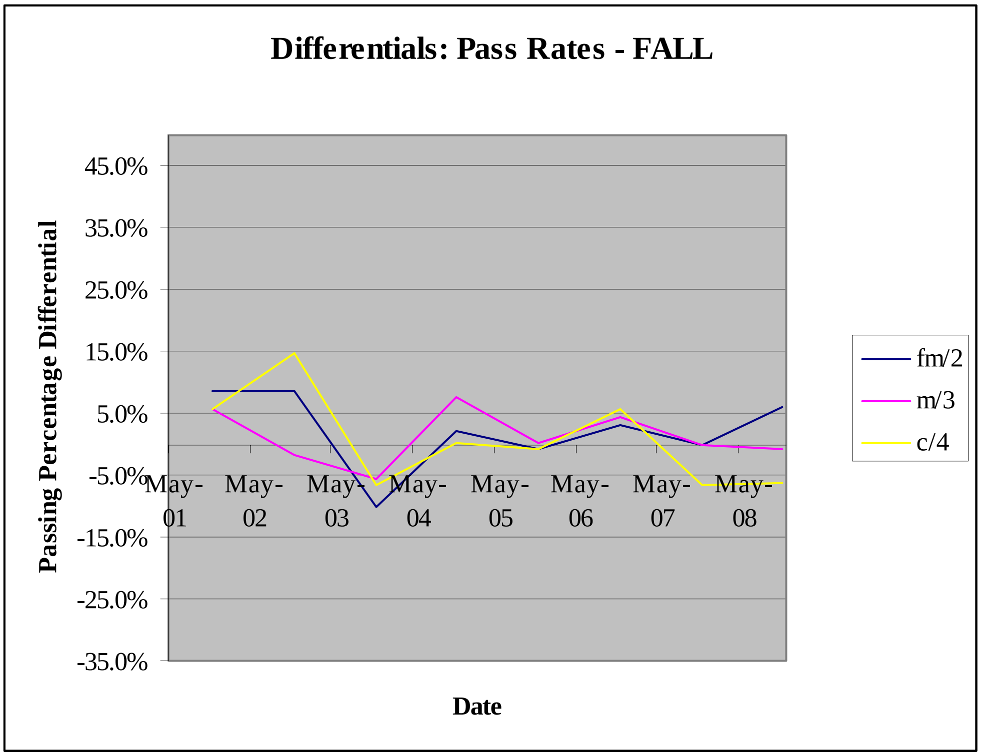


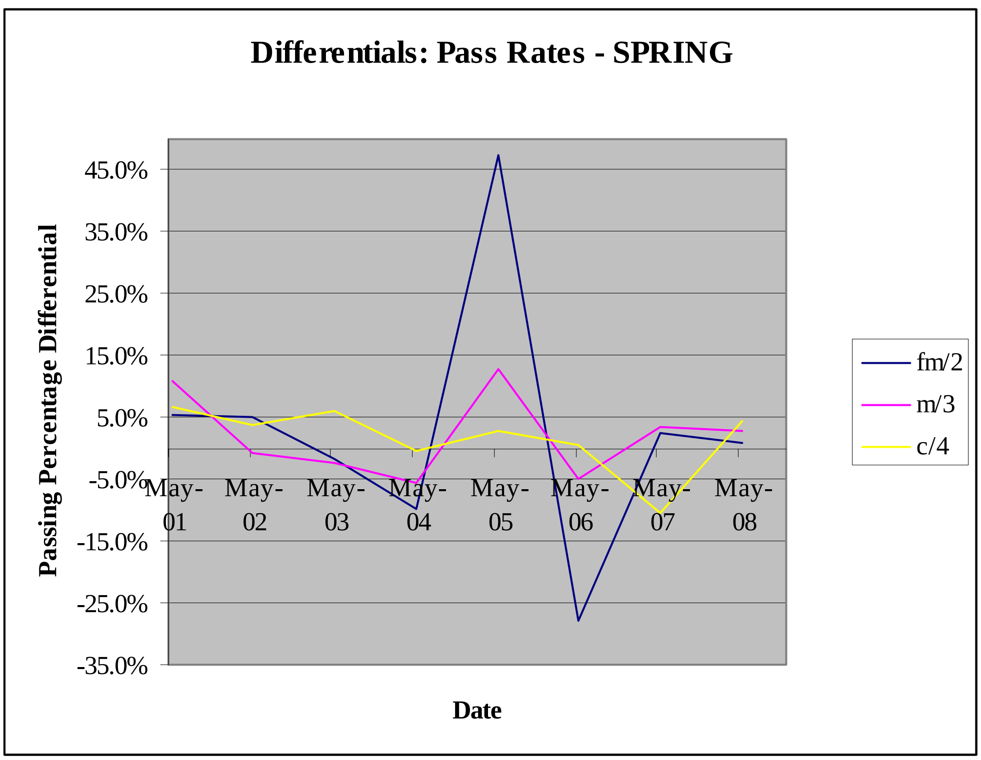
These graphs show that the growth rate of exam FM/2 has been the greatest. They also show that for exams C and FM, there has been more growth in the spring definitely. The support for spring growth for exam M/3 is weaker, but the data definitely does not support the growth being stronger in the spring.

GRAPHS of FIRST DIFFERENCES of DATA:

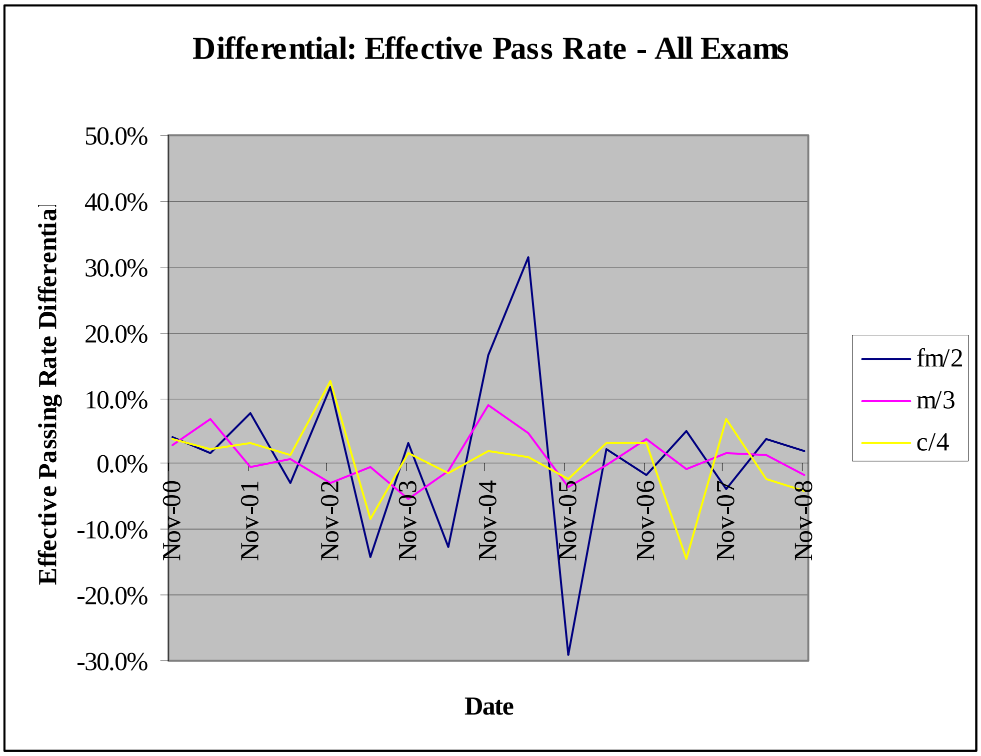
The student project also wanted graphs of the first differences of the data, so graphs of the differentials of the pass rate, effective pass rate and percentage effective were prepared. The oscillating nature of the pass rate is more obvious with these graphs:

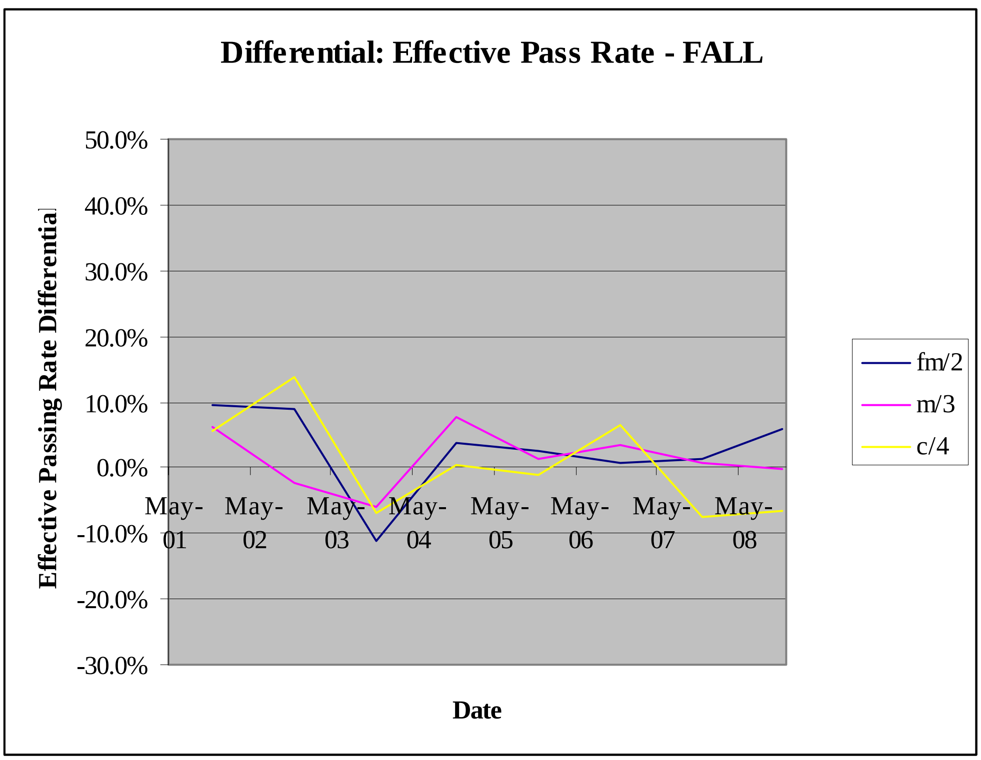


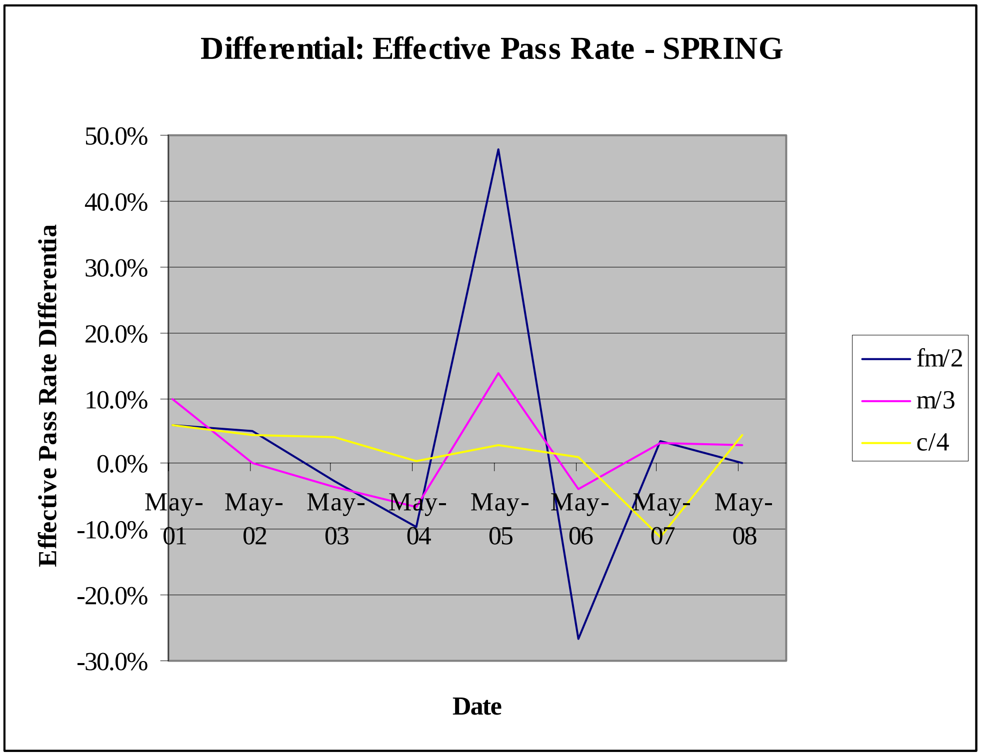




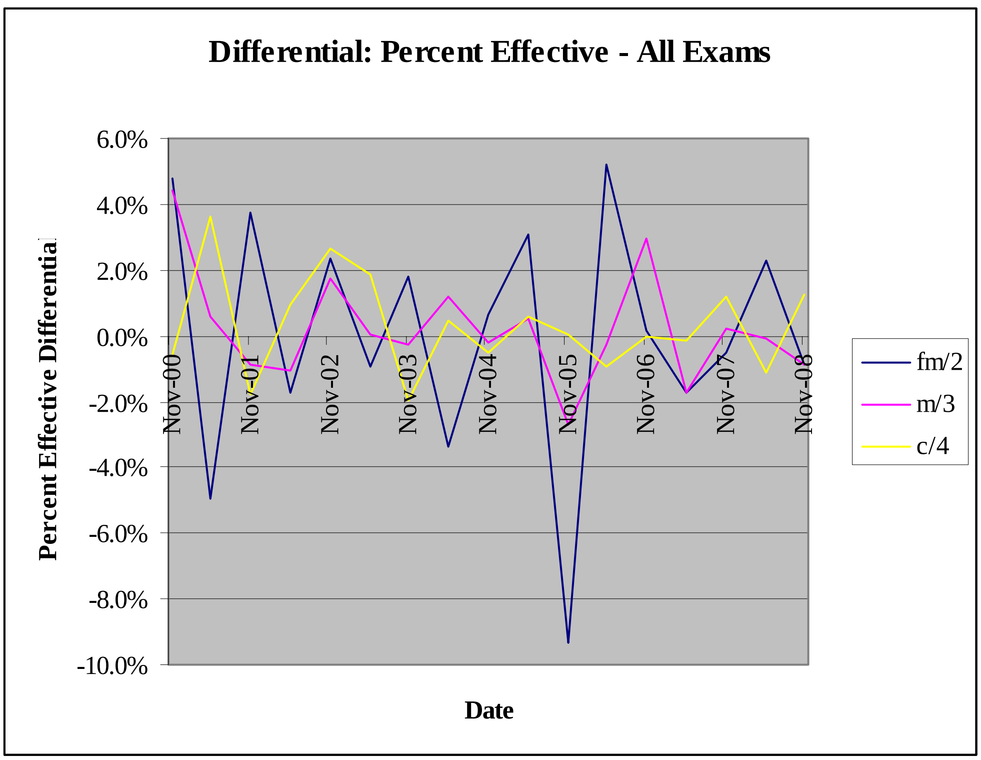
The graphs of the differences for the effective passing rates produce similar patterns as did the graphs of the original data.

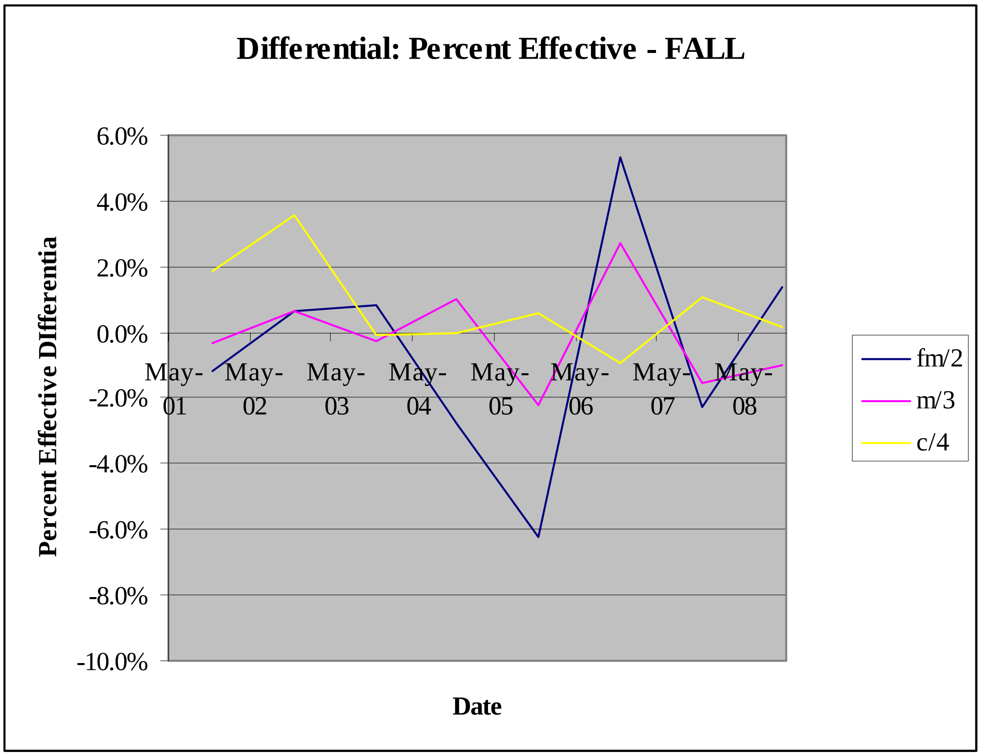

Differential: Effective Pass Rate - All Exams
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
-10.0%
-20.0%
-30.0%
Effective Passing Rate Differentia
Nov-00
Nov-01
Nov-02
Nov-03
Nov-04
Nov-05
Nov-06
Nov-07
Nov-08
Date
fm/2
m/3
c/4

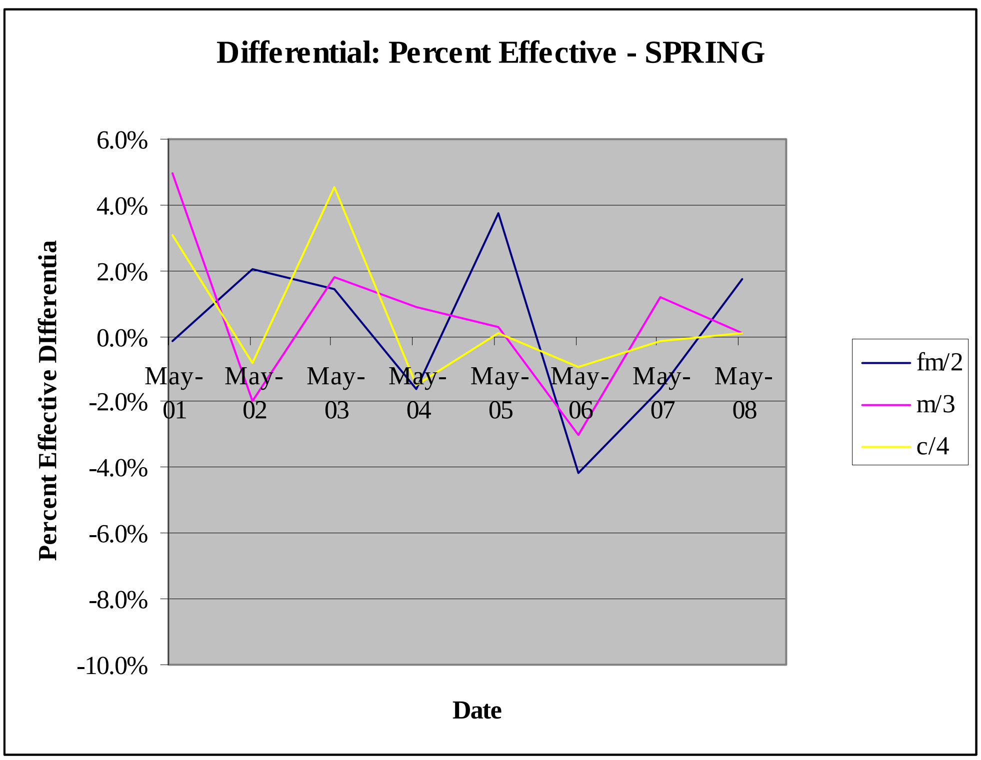

Differential: Effective Pass Rate - FALL
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
-10.0%
-20.0%
-30.0%
Effective Passing Rate Differentia
May-01
May-02
May-03
May-04
May-05
May-06
May-07
May-08
Date
fm/2
m/3
c/4



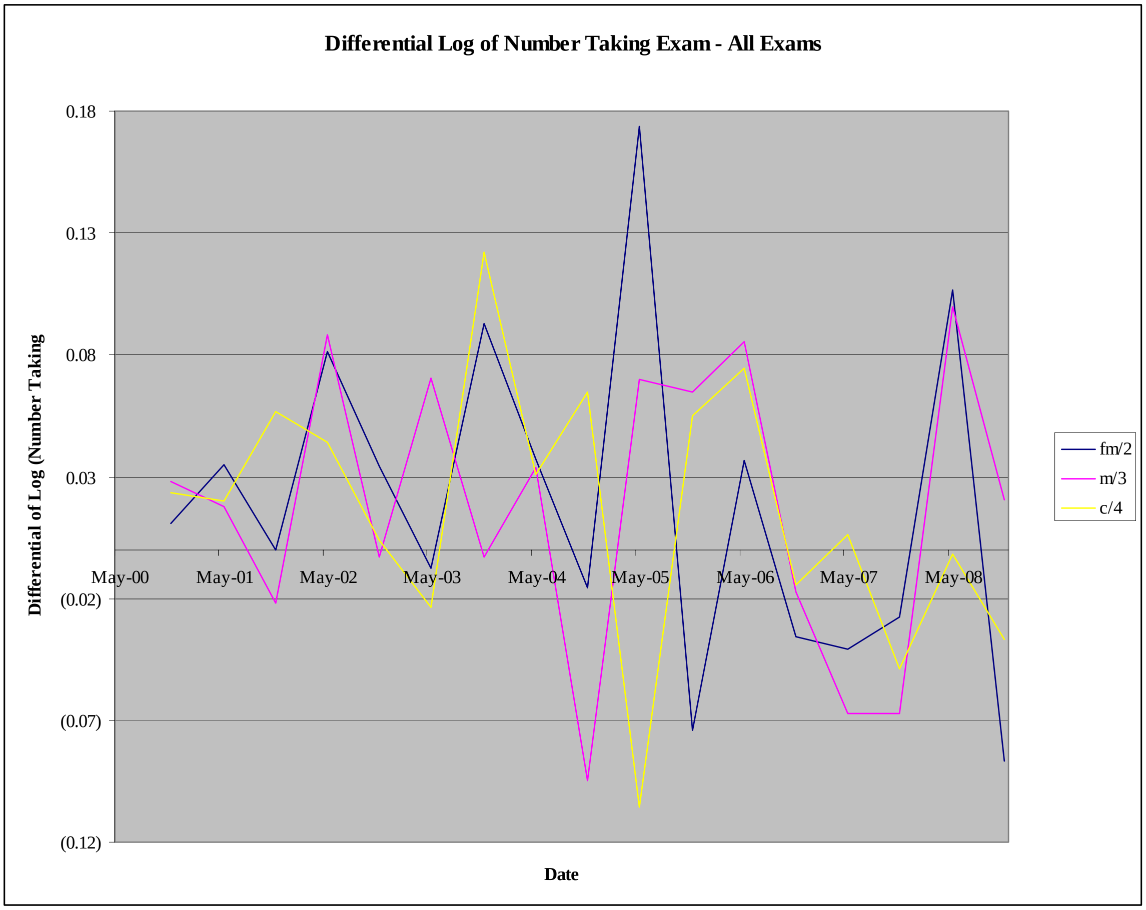
The main conclusion drawn from the graphs of the differences of the percent effective is that exam FM has had the most volatility by far.







Based on the graphs of the absolute and logged data, I only prepared graphs of the first differences of the logs of the number of people taking each exam on each date. The graph of all the dates below shows the volatility and cyclical nature:



STATIONARITY:

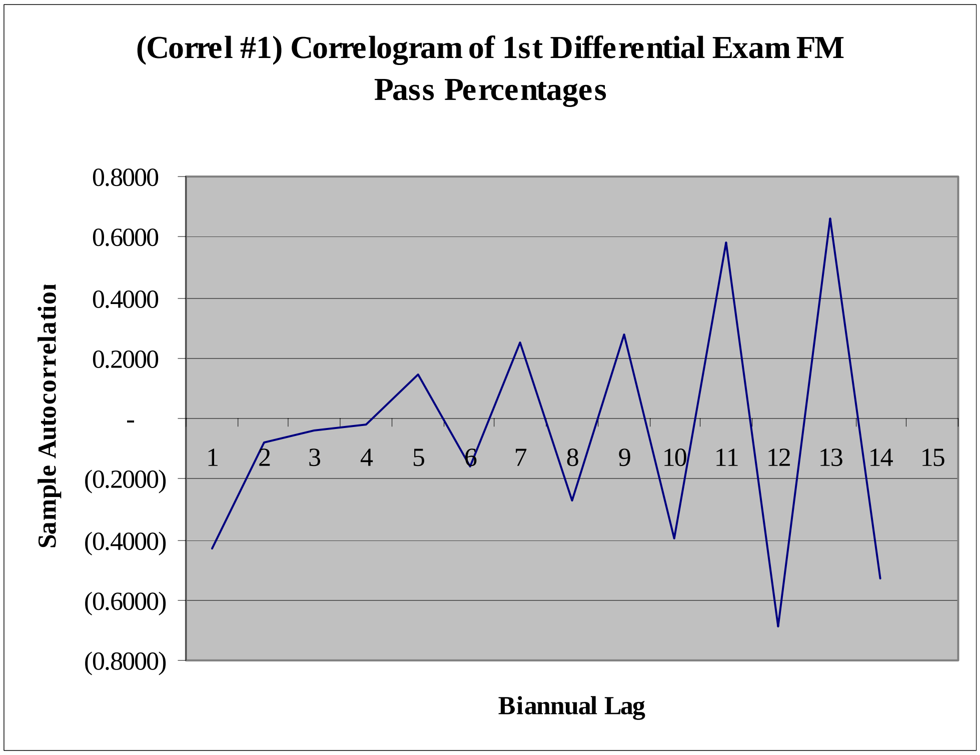
Given the graphs of these time series and their differences, the concept of stationarity can be examined. A stationary process has a mean, variance and autocorrelation structure that does not change over time. It should have no strictly periodic variations for seasonality.

CORRELOGRAMS:

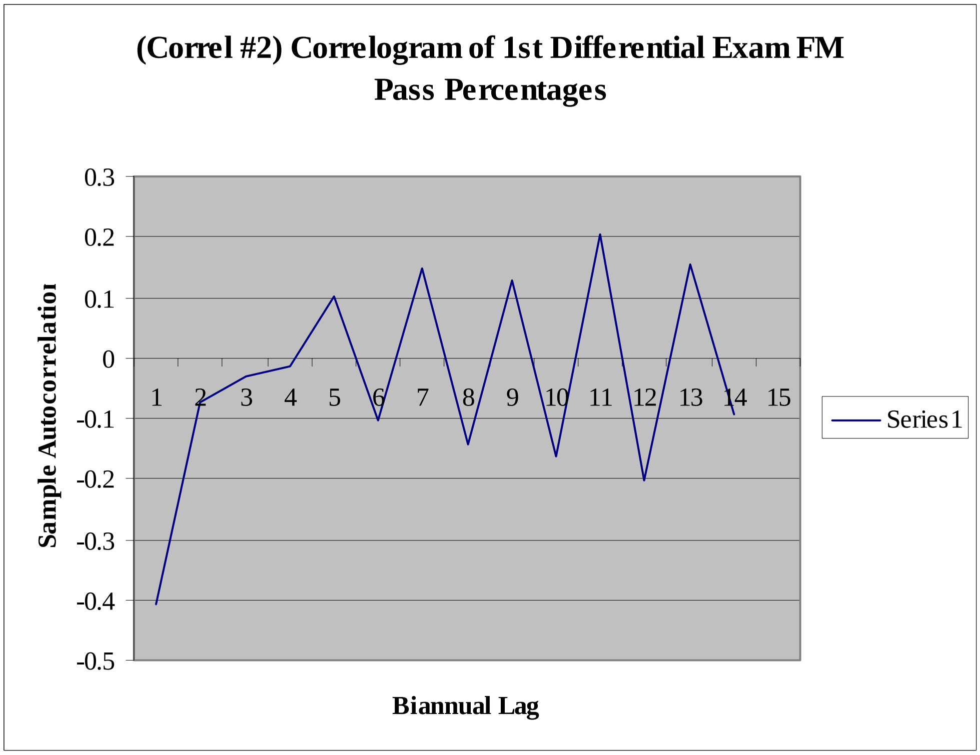
For examining the correlograms, I narrowed the focus of the project in, examining only exam FM and exam C. I also only considered the pass rates and the number of people taking each exam.

As part of this project, I explored three different methods for preparing correlograms. The first method uses just the correl function of Excel. This does not adjust for degrees of freedom and has large fluctuations in late periods. The second method improves upon this be adjusting for degrees of freedom by multiplying the sample autocorrelations produced in the first method by (N-k)/N where N is the total number of observations and k is the value of that lag. This retains the same shape in the correlogram, but adjusts the relative values. The final method is the correct one which will be used for the analysis within this project. This method uses the exact formula for sample autocorrelations and thus both adjusts for degrees of freedom and avoids the large random fluctuations that occur at later lags where there are less values in both of the other methods.

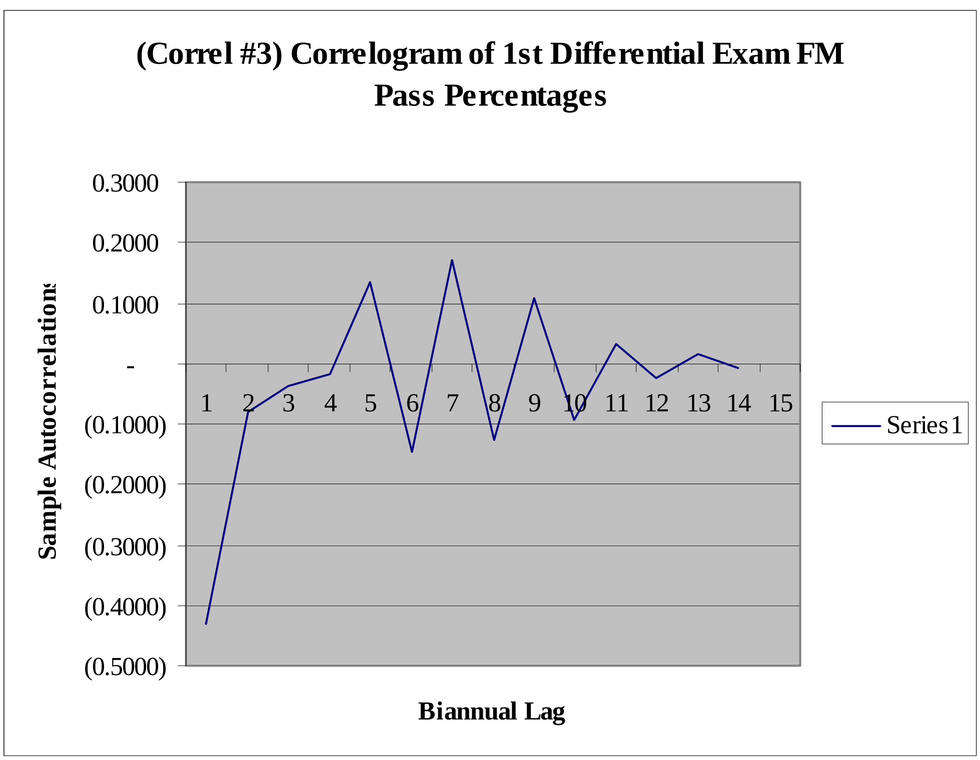
The correlogram for exam FM's first differences of pass rates can be used to show the differences between these three different methods. The first method produces the following correlogram:



The second method produces a similar correlogram, but with less fluctuation in the later lags:

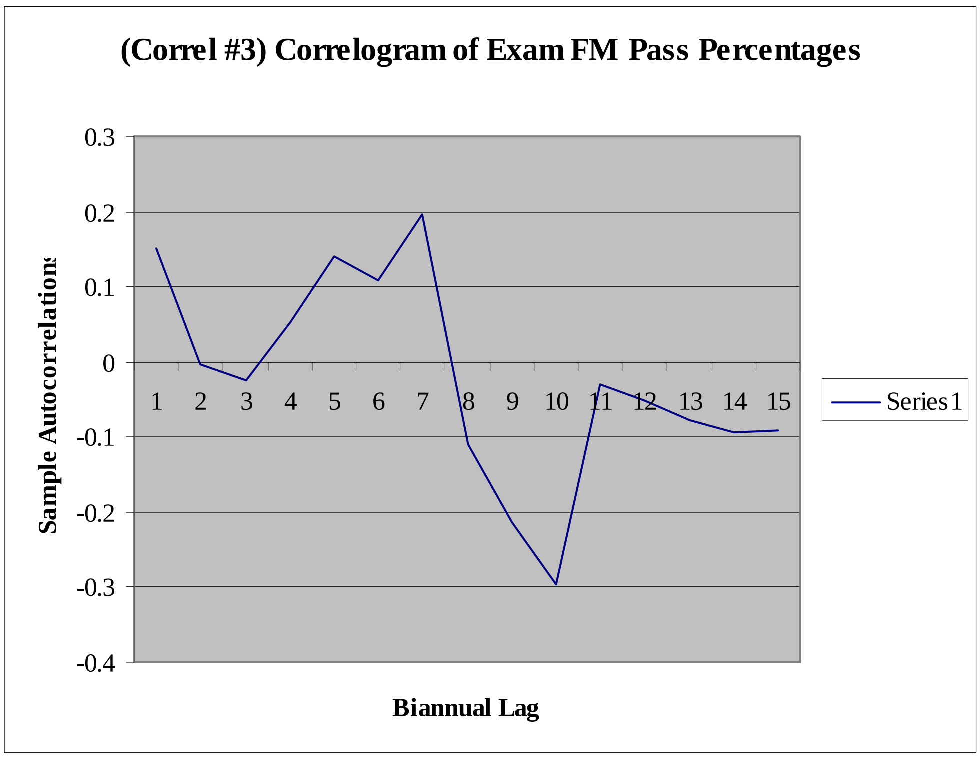


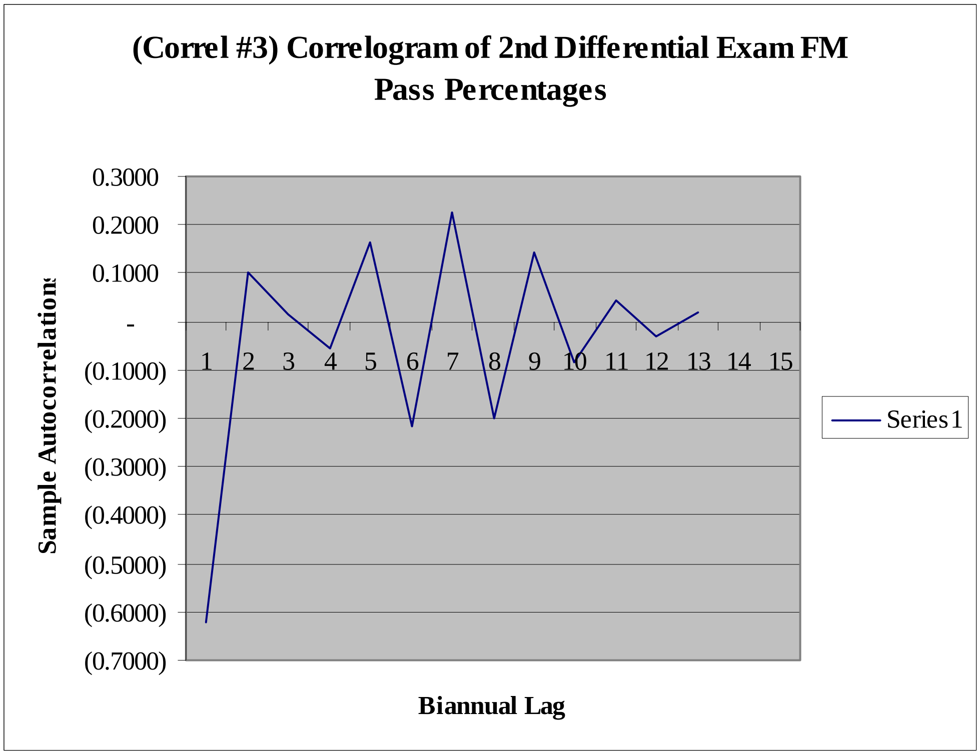
The problem with this second method is that at the later lags, the correlation is based on just a few values and thus random fluctuations bias the data. This is addressed with the correlogram produced by the third method:



A large area of the qualitative evaluation of the correlograms consists of contrasting geometric decay versus sudden drops. In addition, evaluation of oscillating and mean-reversion occurs.

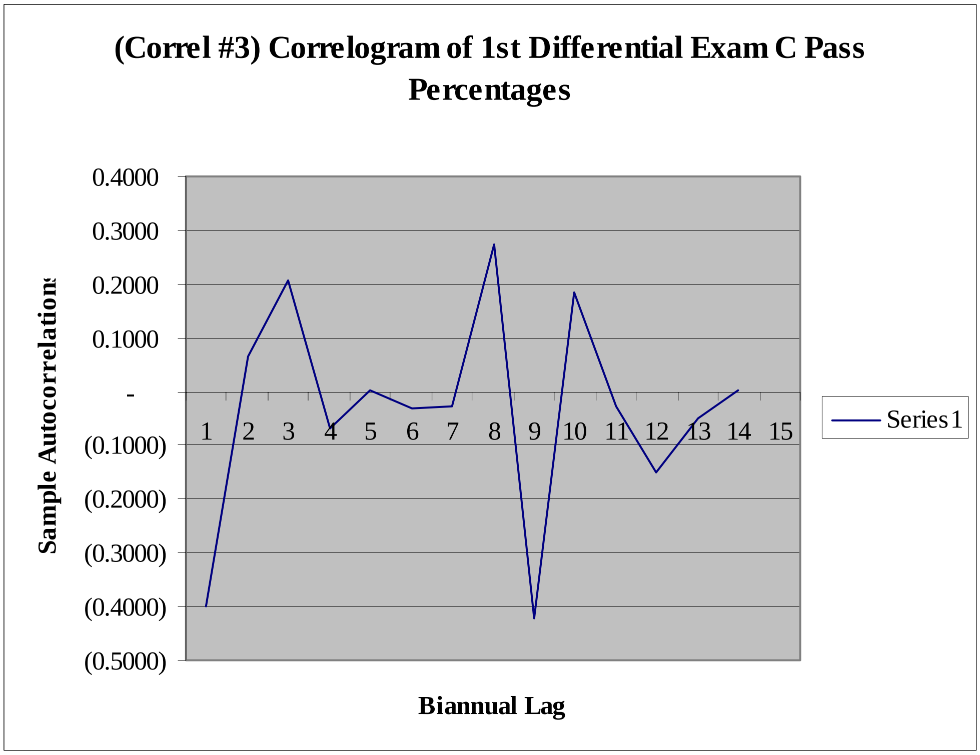
These three correlograms are based on the first difference pass rates for exam FM and are provided to demonstrate the differences between the results using the three different methods. However, for analyzing the time series, the correlogram for the absolute data should be examined and even if the first differences looks reasonable, the second differences should be developed so this can be considered. Therefore, I developed the correlograms for the absolute data and second differences of the pass rates for exam FM. These are shown just using the appropriate method #3 as developed above.



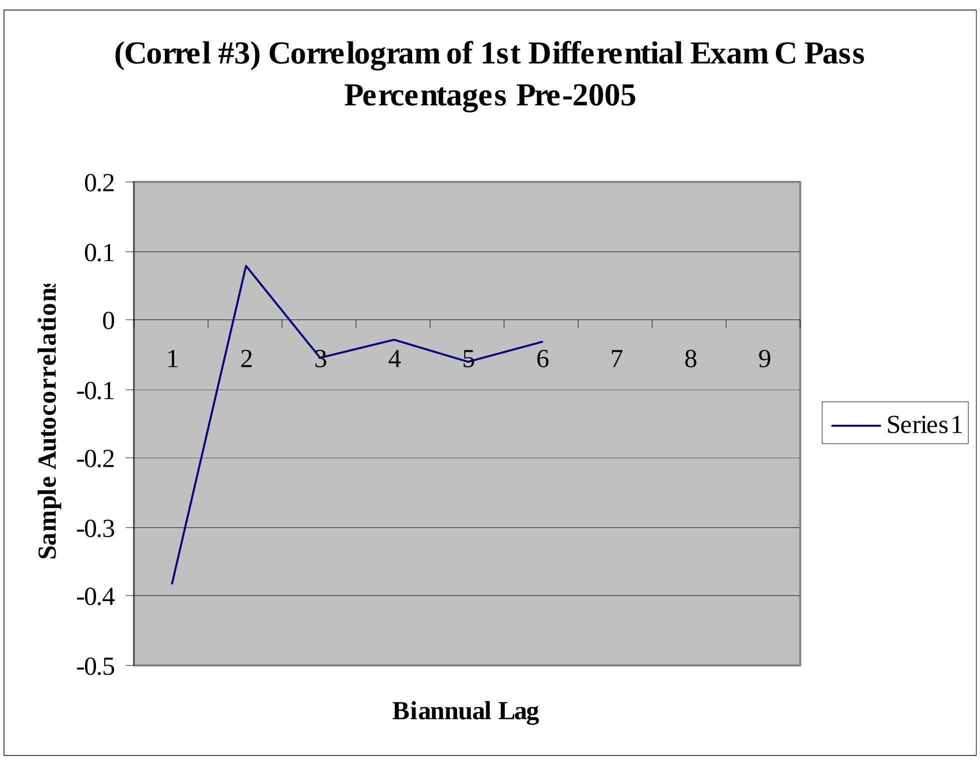


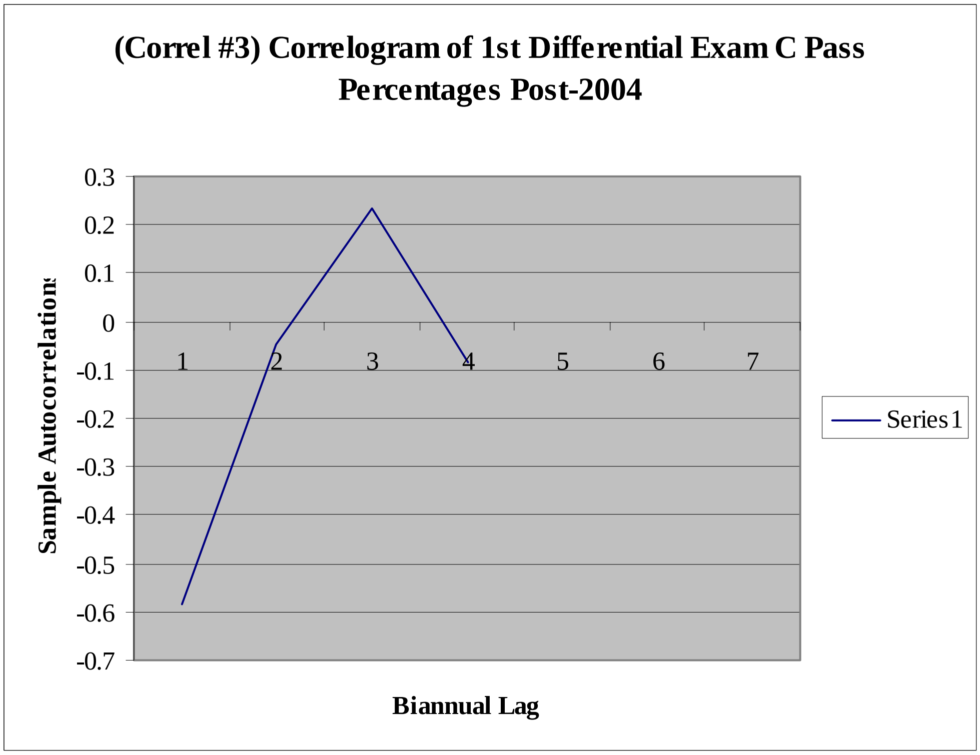
For exam FM, the pass statistics are mean reverting and oscillating, meaning that $\Phi 1$ is between 0 and -1 for an AR(1) model. The model is also stationary since it is mean reverting and oscillating. Based on these correlograms, the time series analysis of the exam FM time series will be done based on the first differences.

For exam C, the first differential also produced the correlogram showing a stationary process:



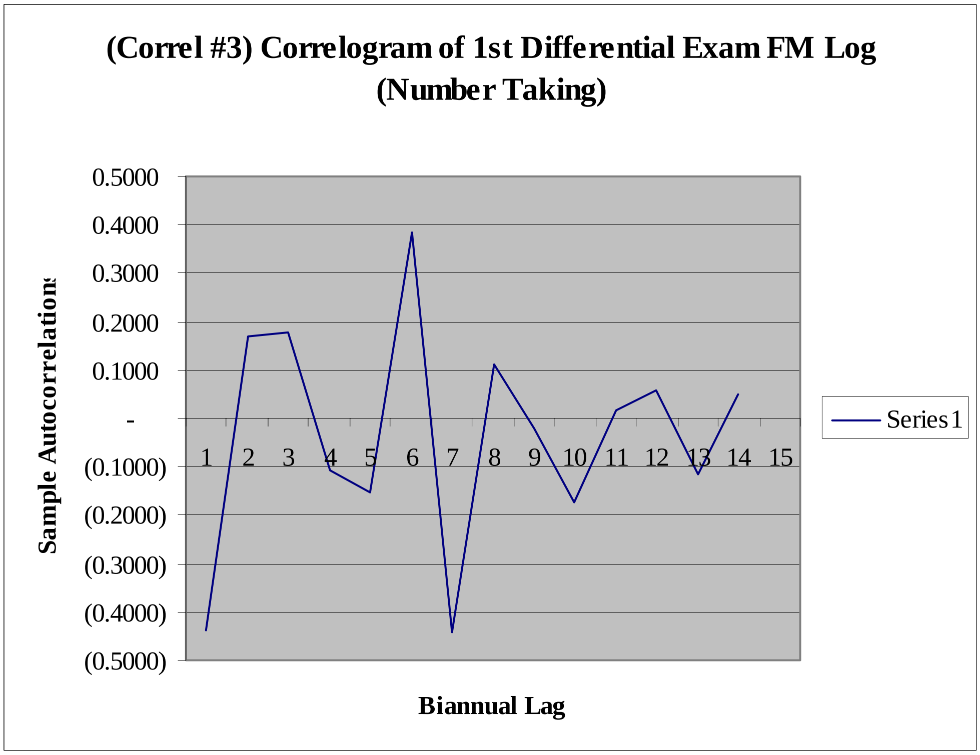
There is really insufficient data to allow the preparation of correlograms of the pass rates before the 2005 exam change and afterwards, but since the project was interested in interventions, I prepared the correlograms for exam 4/C's first differences of pass rates both before and after the change.

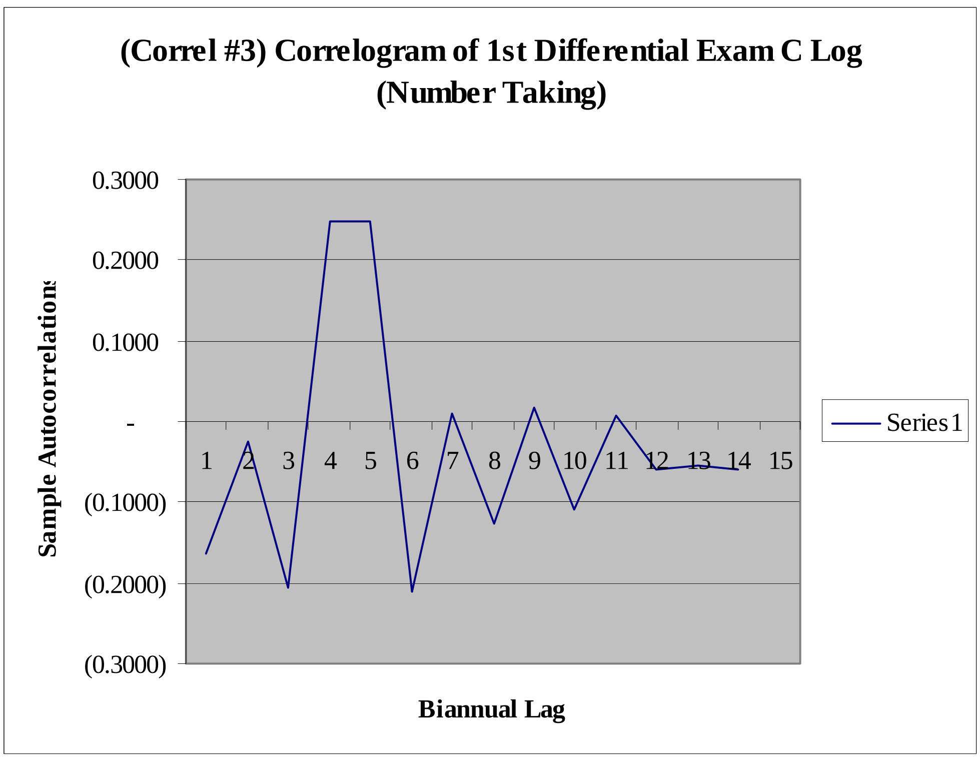




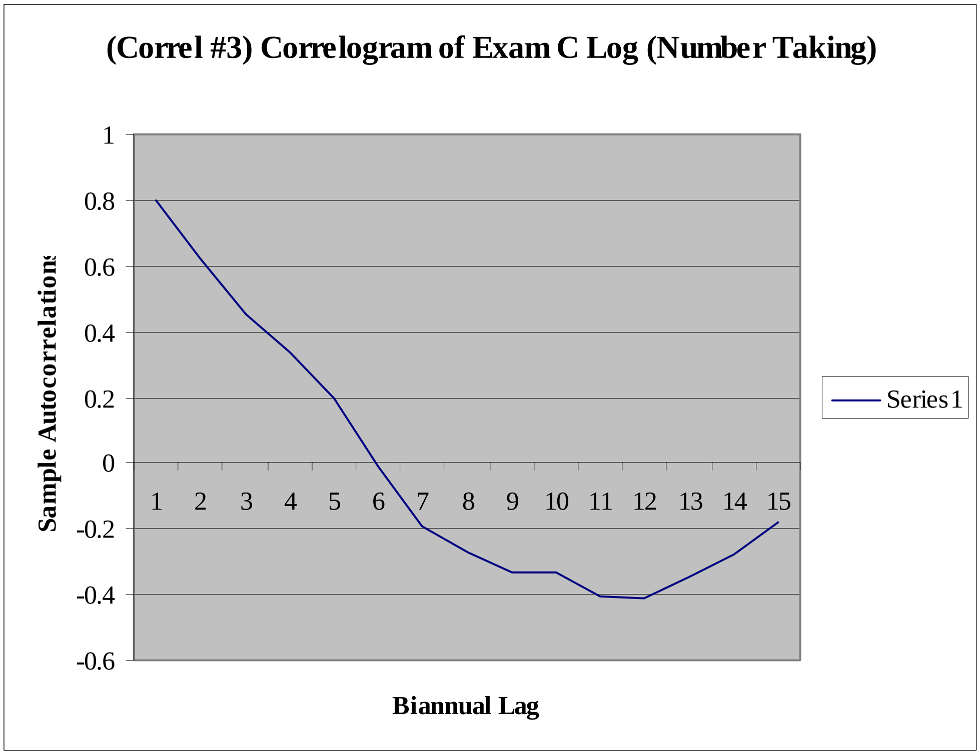
Similar correlograms were prepared for absolute and second differences as well as for exam 2/FM, but as there is not sufficient data to make inferences, I have not included them within this report.

For the number of people taking each of the exams, I developed the correlograms for both the numbers themselves and the logs as well the first and second differences. The first differential of the logs correlogram for each of the exams is shown below. The remainder of these charts are available in the accompanying Excel files.

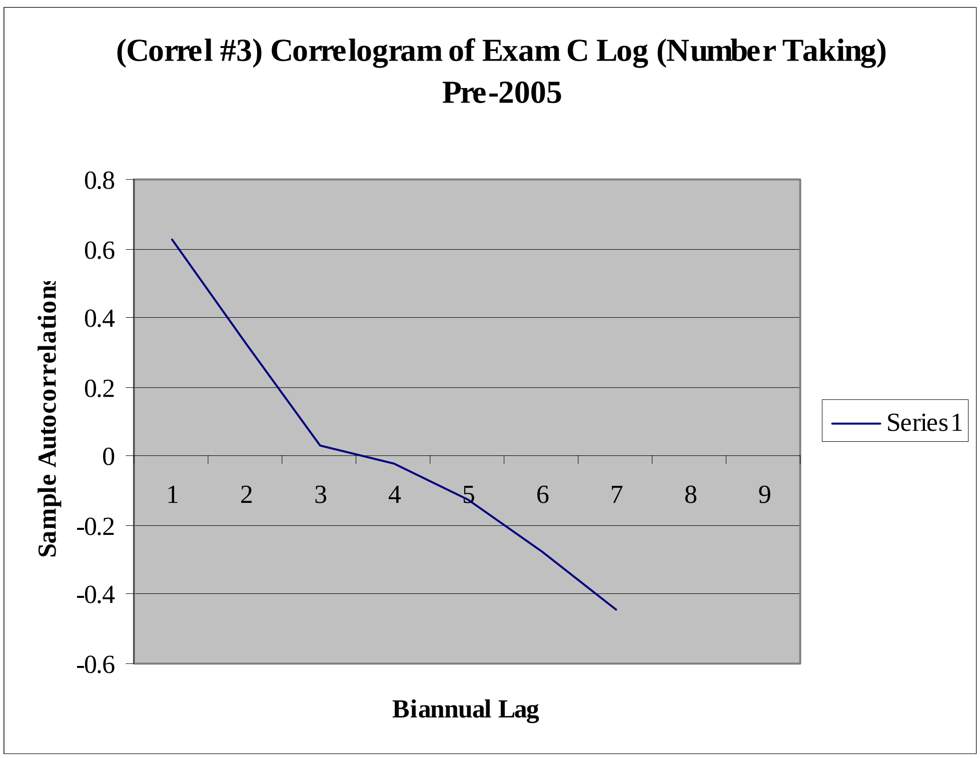




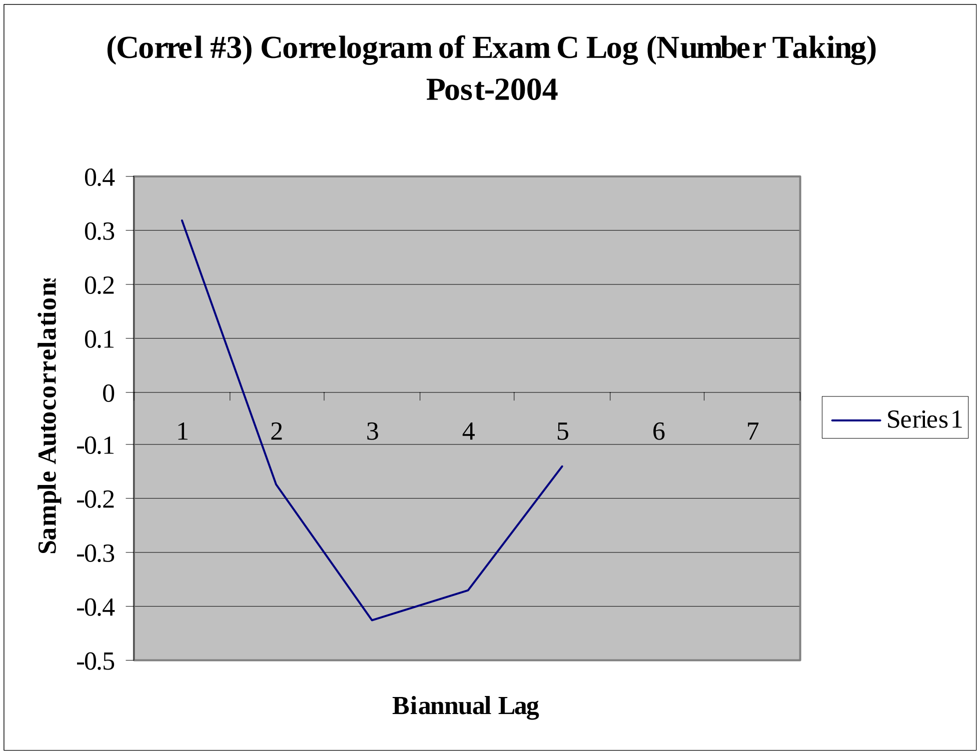
As with the pass rates, there is insufficient data to infer meaningful information about the intervention. The correlograms for the absolute log of the data of the number taking is shown below:



(Correl #3) Correlogram of Exam C Log (Number Taking)
Sample Autocorrelations
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Series1
Biannual Lag



(Correl #3) Correlogram of Exam C Log (Number Taking)
Pre-2005
Sample Autocorrelations
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
1 2 3 4 5 6 7 8 9
Series1
Biannual Lag



DESCRIPTIONS of CORRELOGRAMS for VARIOUS MODEL TYPES

AR(1) models will have exponentially decreasing appearance of the sample autocorrelation function. Higher order AR models often have a mixture of exponentially decreasing and damped sinusoidal sample autocorrelation functions. With an MA(q) process, the autocorrelation function becomes approximately zero at lag q+1 and greater.

For a sample autocorrelation function that is always near zero, the model that is suggested is white noise. For a sample autocorrelation function with a slow decay, a trend model is suggested. For a periodic sample autocorrelation function, a periodic model is suggested. For a sample autocorrelation function that decays to zero exponentially, an AR(1) model should be considered and for a sample autocorrelation function which is near zero for all lags above a certain number q, a sharp dropoff, a model of form MA(q) should be considered. For a model that has a geometric decay, but beginning after a few lags, an ARMA model is suggested. If there are high values at fixed intervals among the sample autocorrelation function, a model with a seasonal AR term should be considered.

An ARMA (1,1) model with a negative $\Phi 1$ will oscillate and the theta parameter and mean are not relevant. Even higher order ARIMA models with a negative $\Phi 1$ will oscillate as long as it is larger in magnitude that the $\Phi 2$, or the $\Phi 2$ is also negative.

DURBIN-WATSON STATISTICS:

The Durbin-Watson statistic is a test of serial correlation whose statistic ranges from 0 to 4. A value of 2 indicates no serial correlation while a value of 0 indicates perfect positive serial correlation and a value of 4 indicates perfect negative serial correlation.

For the absolute pass rates for exam FM, the Durbin-Watson statistic for an AR(1) process is 0.129185 while for the first differences of the pass rates for exam FM, it is 0.102693. This indicates near perfect positive autocorrelation. For the absolute pass rates for this exam, the Durbin-Watson statistic only changes slightly to 0.129185. In comparison, the Durbin-Watson statistic for an AR(1) process of the detrended data is 0.118672 for the absolute data and 0.102512 for the first differences. The statistics for the detrended and standard data are thus not significantly different at all.

For exam C, the Durbin-Watson statistic for an AR(1) process of the absolute pass rates drops down to 0.010653, which strongly indicates the presence of positive autocorrelation. For this exam, the statistic for the first differences is even closer to zero at 0.008668!

BOX-PIERCE Q STATISTICS:

This statistic is used to evaluate whether a time series is a white noise process. In addition to evaluating the time series itself, this test can be used to determine if the first differences of the time series is a white noise process and thus the time series itself is a random walk.

For exam FM, with 13 lags on the first differences of the passing rate model, the statistic is 3.223, less than the critical value of 18.549, so we accept the null hypothesis that the series was generated by a white noise process and thus the time series itself is a random walk. Similarly, for exam C, with 14 lags on the absolute passing rate model, the statistic is 5.829, less than the critical value of 19.8119, so we accept the null hypothesis that the series is a white noise process.

In order to understand the q-statistic, it is important to understand that the more observations, the q-statistic will be increased therefore making it more likely that you will reject the null hypothesis of independence of the residuals – more observations make us more likely to reject the null. This is balances because with more observations, each sample autocorrelation becomes closer to the true autocorrelation which are zero if the residuals are independent.

AR(1), AR(2) and AR(1) of FIRST DIFFERENCES

Only the pass rates were considered for this analysis. Other statistics could be used for future study, but the scope was limited for this project.

AR(1) CALCULATIONS:

For exam FM, the following relevant data and calculations were developed:

Regression Statistics	
Multiple R	0.169383
R Square	0.028691
Adjusted R Square	-0.03606
Standard Error	0.110434
Observations	17

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.33787	0.117294	2.880539	0.011437
X Variable 1	0.17777	0.267068	0.665636	0.51575

The adjusted R Squared is negative, indicating that this model is probably not a good fit. Coupled with the D-W statistic, we can not accept the hypothesis that there is no serial correlation present and these results suggest that there is likely positive serial correlation present in this regression. Therefore, I do not recommend the use of the AR(1) model on the absolute data for this statistic.

For exam FM using the detrended data, the following relevant data and calculations were developed:

Regression Statistics	
Multiple R	0.169676
R Square	0.02879
Adjusted R Square	-0.03596
Standard Error	0.110451
Observations	17

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.336215	0.116963	2.874544	0.011577
X Variable 1	0.178202	0.26724	0.666823	0.515012

This is essentially the same as what was produced with the unadjusted data, so no advantage is gained using this model form from this possible adjustment.

For exam C, the following relevant data and calculations were developed:

Regression Statistics	
Multiple R	0.565072
R Square	0.319306
Adjusted R Square	0.273927
Standard Error	0.05643
Observations	17

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.159896	0.120479	1.327171	0.2043
X	0.657142	0.247734	2.652611	0.018097

Variable 1

While the adjusted R squared is not negative for this regression, it is quite low, suggesting this model is also not a good fit.

AR(2) CALCULATIONS:

For exam FM, the following relevant data and calculations were developed:

Regression Statistics	
Multiple R	0.142906
R Square	0.020422
Adjusted R Square	-0.13028
Standard Error	0.117368
Observations	16

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.346799	0.172105	2.015035603	0.065062876
X Variable 1	0.150419	0.289645	0.519322973	0.612263321
X Variable 2	-0.00244	0.2953	-0.008269832	0.993527266

This model also does not appear to be a good fit.

Similar to the AR(1) calculations, the calculations for the possible detrended data does not change in any significant fashion:

Regression Statistics	
Multiple R	0.143217
R Square	0.020511
Adjusted R Square	-0.13018
Standard Error	0.117464
Observations	16

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.344562	0.171592	2.008025751	0.06588608
X Variable 1	0.150961	0.290259	0.520091045	0.611742746
X Variable 2	-0.00105	0.295394	-0.003548465	0.997222615

The data for exam C was not developed.

AR(1) of FIRST DIFFERENCES CALCULATIONS:

For exam FM, the following relevant data and calculations were developed:

Regression Statistics	
Multiple R	0.431263
R Square	0.185988
Adjusted R Square	0.127844
Standard Error	0.131749
Observations	16

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0185	0.033047	0.559808	0.584457
X Variable 1	-0.43258	0.241865	-1.78851	0.095348

And for the detrended data:

Regression Statistics	
Multiple R	0.430644
R Square	0.185454
Adjusted R Square	0.127272
Standard Error	0.13175
Observations	16

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.018495	0.033047	0.559656	0.584557
X Variable 1	-0.43215	0.242055	-1.78536	0.095877

Finally, for exam C:

Regression Statistics	
Multiple R	0.407608
R Square	0.166144
Adjusted R Square	0.106583
Standard Error	0.055751
Observations	16

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.009916	0.013973	0.709652	0.489571
X Variable 1	-0.40395	0.241863	-1.67017	0.117082

For exam FM's absolute pass rates and an AR(1) model, we have 18-1=17 residuals. For the Box-Pierce Q statistic for this time series, the degrees of freedom is 13. I used Excel in this project to generate the critical values for the test based on the fact that the q statistic is distributed approximately like a chi-squared distribution with K-p-q degrees of freedom. For the first differences of this time series, the maximum degrees of freedom is 12.

For the unadjusted exam FM data, the adjusted R-squared goes even more negative from AR(1) to AR(2) suggesting that the AR(2) model is an even worse fit. The adjusted R-squared of the AR(1) of the first differences is increased significantly from the AR(1) of the absolute data suggesting that this is a better fit. However, based on the prior analysis and the t-statistics, I would not recommend this model either. Similar results are seen for the adjusted data for FM.

In the case of exam C, the adjusted R-squared is significantly higher for the AR(1) model than the AR(1) model of the first differences. The relative t statistics also support that the AR(1) model is a better fit for this data than the other model.

SAMPLE AUTOCORRELATIONS HYPOTHESIS TESTING

All of the Durbin-Watson and Box-Pierce Q statistics analysis evaluates whether the sample autocorrelations of the residuals are statistically different from zero. The null hypothesis is that the sample autocorrelations are zero. To evaluate whether the sample autocorrelations are statistically different from zero, we estimate the variance of the sample autocorrelation distribution using Bartlett's theorem that the sample autocorrelations of a white noise process is a normal distribution of mean zero with a variance of one divided by the number of observations. Therefore, for 17 autocorrelations, the standard deviation of a white noise process is 24.25% and the probability of observing an autocorrelation greater in absolute value than 0.3517 is 10%.

The optimal ARIMA model has the lowest squared error for its forecasts and the variance of the error terms estimated from observed values. The sum of squared errors is the sum ssquare of the actual value minus the forecasted value.

INTERVENTIONS

A final area for examination in this topic relates to interventions, in this case, using time series analysis to see how changes occur relating to changes in the SOA exam structure. This could include such questions as how the number of people taking an exam is affected by a change in the number of hours for an exam or the frequency an exam is offered. Similar questions could be examined in terms of the percent of people taking the exam who are effective.

I have not focused on this topic for this project, but I did perform some analysis comparing the pre-2005 2/4 time series to the post-2004 FM/C time series by examining correlograms for each period on some of the statistics.

In addition to examining the coefficients of the variables for the model for each time period, the standard error of the coefficients must be considered in deciding whether the data justifies separate models for before and after the intervention.

Conclusions:

In general, models are superior if they are simple and the principle of parsimony is supported by the fact that simpler models have a tendency to forecast better than more complicated models.

Note, this course did not involve any nonlinear regression and these methodologies should be considered as well. With further data to allow some to be withheld for ex-post evaluation, additional analysis of this subject should be completed.

The analysis indicates that these data series do not lend themselves to linear ARIMA modeling as well as provides information about the seasonality, correlations and trends of the data.

A final area for possible additional exploration would be Yule-Walker equations. These can be used to obtain estimates of autoregressive parameters. While they don't tell us what type of ARIMA to use, given a proposed ARIMA model, they enable us to chose the optimal theta and phi coefficients and judge among competing ARIMA models. However, with small samples, they are biased downward from true autocorrelation functions, so I did not explore these in this student project.