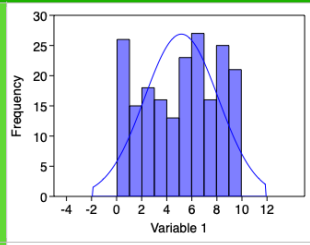
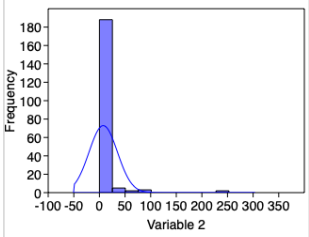
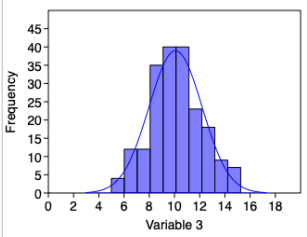
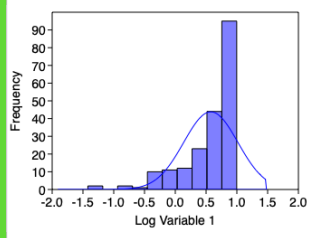
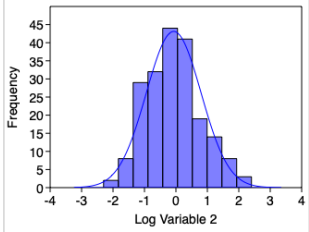
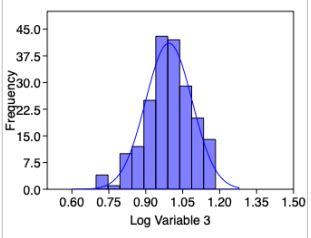
Lab 2 Assignment

1. Test the distributions of the three variables in the “3 Variables.txt” datafile for the following distributions.
   • Normal
   • Log Normal (30 Points)
   Reject the null hypothesis (H_0) if there is less that a 95 percent chance of it being true. (30 Points)

Normality Tests

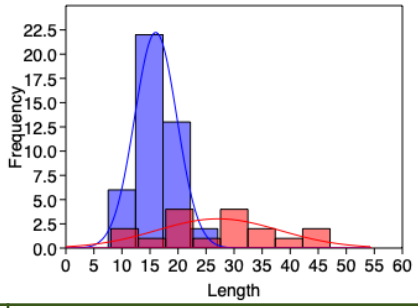
	Variable 1	Variable 2	Variable 3
Plot			
Shapiro-Wilk Test	0.9471	0.2614	0.9921
p-value	0.0000	0.0000	0.3578
Andrerson-Darling Test	2.753	64.4	0.4122
p-value	0.0000	0.0000	0.3366
Lillefors Test	0.07368	0.3943	0.0450
p-value	0.0099	0.0001	0.4072
Jarque-Bera Test	12,94	29740.0000	0.7986
p-value	0.0015	0.000	0.6708
Interpretation	Reject H_0, The distribution is not normal.	Reject H_0, The distribution is not normal.	Accept H_0, The distribution is normal.

Log Normality Tests

	Variable 1	Variable 2	Variable 3
Plot			
Shapiro-Wilk Test	0.8079	0.9882	0.9799
p-value	0.0000	0.0979	0.0056
Andrerson-Darling Test	11.3700	0.6103	0.7995
p-value	0.0000	0.1111	0.0378
Lillefors Test	0.1729	0.0460	0.0631
p-value	0.0001	0.3705	0.0524
Jarque-Bera Test	226.4000	4.0800	9.585
p-value	0.0000	0.1300	0.0083
Interpretation	Reject H_0, Distribution is not log normal.	Accept H_0, Distribution is log normal.	Reject H_0, Consensus of tests indicates distribution is not log normal.

2. Use the appropriate two sample *t*-test to test the statistical difference between mean the mean of lengths of the trilobites *Stenopareia glaber* and *Stenopareia linnarssoni* based on the data included in the "Trilobites.txt" datafile. Reject the null hypothesis (H_0) if there if less that a 97.5 percent probability of it being true. (20 Points)

Two-Sample t-Test



S. glaber		**F-test for equality of variances**	
Mean:	16.0258	Observed F:	8.4902
Variance:	14.1908	Critical F:	2.10
n:	43	Interpretation:	Reject Ho Variances are unequal.
S. linarrssoni		**Two-sample Welch Test for equality of means**	
Mean:	27.1618		
Variance:	120.4827		
n:	17	Difference:	11.136
		t	4.0889
		Critical *t*	2.500
		p	0.0007
		Interpretation	Reject Ho, means are unequal.

3. Use matrix methods to solve for x_1 and x_2 in the following equations. (20 Points)

$2x_1 + 3x_2 = 19$
$2x_1 + x_2 = 11$

Simultaneous Equations

$$A = \begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix} \qquad A^{-1} = \begin{vmatrix} -0.25 & 0.75 \\ 0.50 & -0.50 \end{vmatrix}$$

$$C \begin{vmatrix} 19 \\ 11 \end{vmatrix} \qquad A\text{-}1 \,.\, C \begin{vmatrix} 3.5 \\ 4 \end{vmatrix}$$

4. Calculate the following attributes of the vectors included in the “Vectors.txt” datafile. (80 Points)
   - Vector lengths
   - Vector angles relative to the x-axis
   - Vector angles relative to the z-axis
   - Vector angles relative to each other

Lengths & Angles w/ x & z Axes

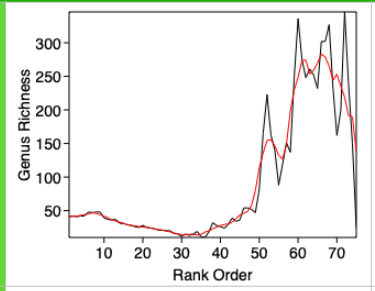
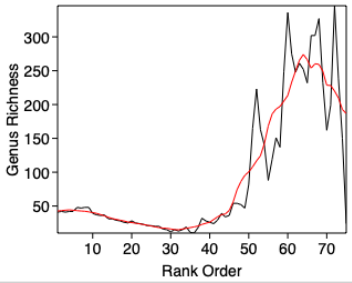
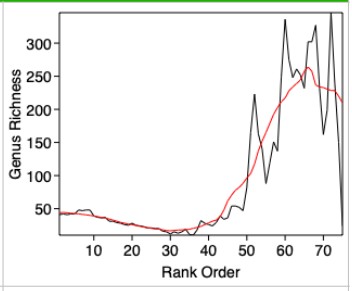
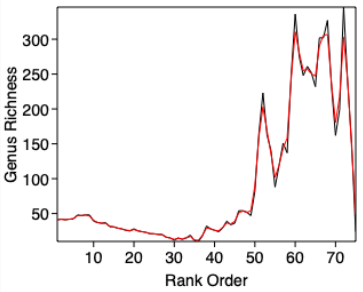
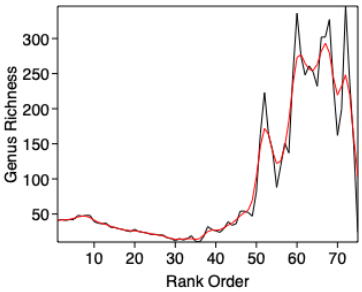
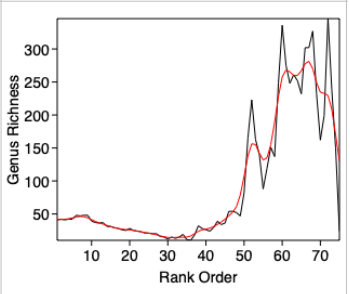
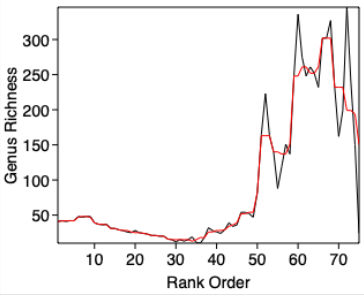
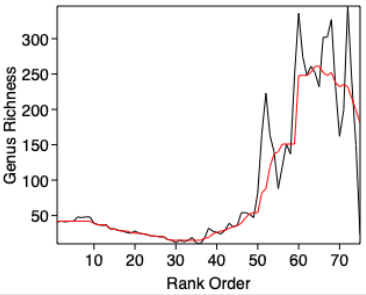
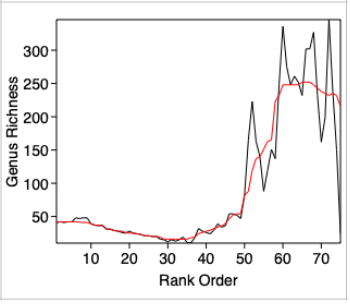
	x	y	z		Length	Angle (x-axis) (Radians)	Angle (x-axis) (Degrees)	Angle (z-axis) (Radians)	Angle (z-axis) (Degrees)
Vector 1	3	7	10		12.57	0.2387	76.19	0.7956	37.29
Vector 2	7	3	6		9.70	0.7220	43.78	0.6189	51.77
Vector 3	5	7	1		8.66	0.5774	54.74	0.1155	83.37
Vector 4	6	10	8		14.14	0.4243	64.90	0.5657	55.55

Angles w/ Other Vectors

	Radians					Angles			
	Vector 1	Vector 2	Vector 3	Vector 4		Vector 1	Vector 2	Vector 3	Vector 4
Vector 1	1.0000	0.8370	0.6798	0.9451		0.00	33.18	47.17	19.08
Vector 2	0.8370	1.0000	0.7384	0.8752		33.18	0.00	42.40	28.93
Vector 3	0.6798	0.7384	1.0000	0.8818		47.17	42.40	0.00	28.14
Vector 4	0.9451	0.8752	0.8818	1.0000		19.08	28.93	28.14	0.00

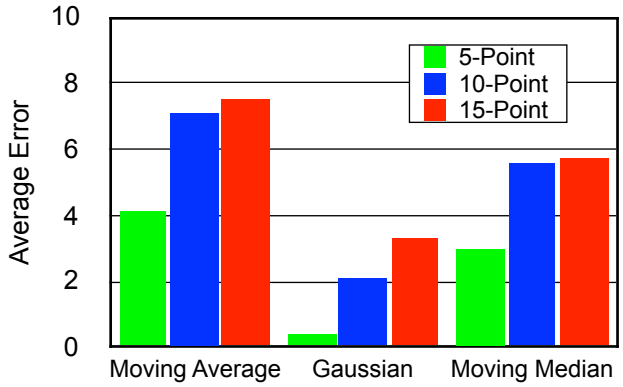
5. Illustrate how 5 point, 10 point and 15 point smoothing windows for the moving average, Gaussian, and moving median filters affect the forms of Sepkoski's Cambrian, Paleozoic and Modern fauna biodiversity curves. (30 Points)
   - Which filter results in the least distortion? Why? (20 Points)
   - Which filter results in the greatest amount of distortion? Why? (20 Points)

Cambrian Fauna

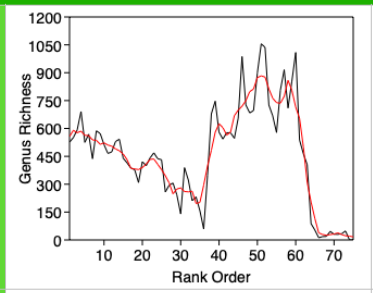
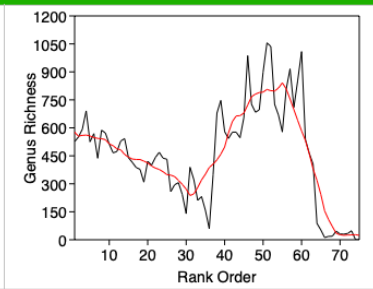
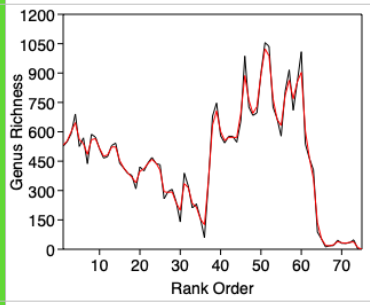
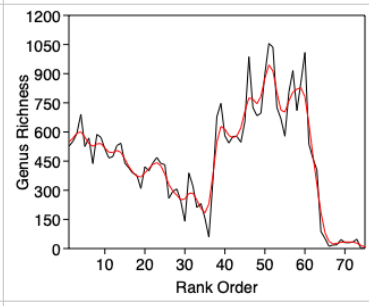
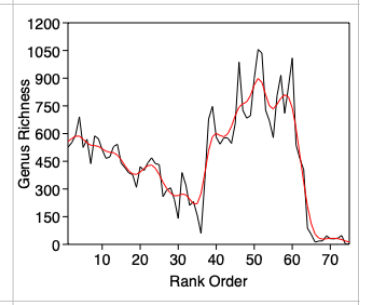
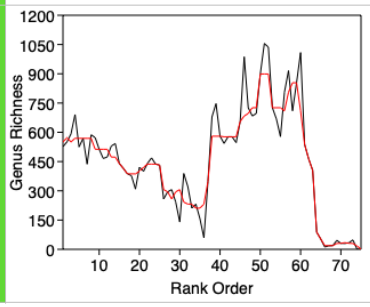
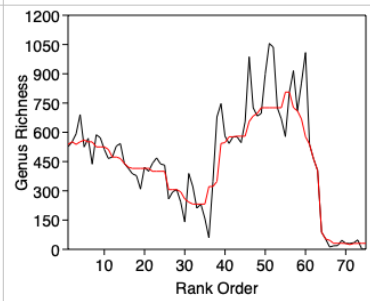
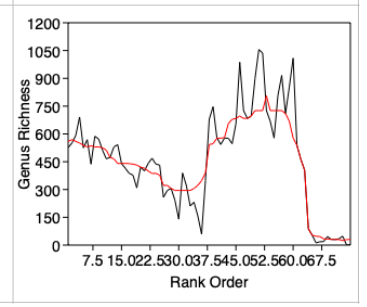
	5-Point	10-Point	15-Point
Moving Average			
Gaussian			
Moving Median			

Cambrian Fauna Error Analysis

	5-Point	10-Point	15-Point
Moving Average	4.148	7.137	7.491
Gaussian	0.368	2.091	3.283
Moving Median	2.958	5.541	5.725

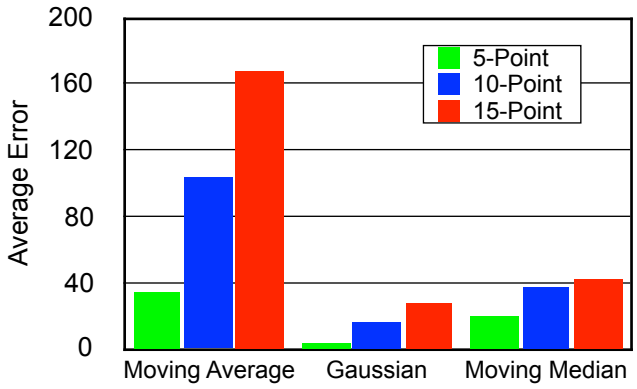


Paleozoic Fauna

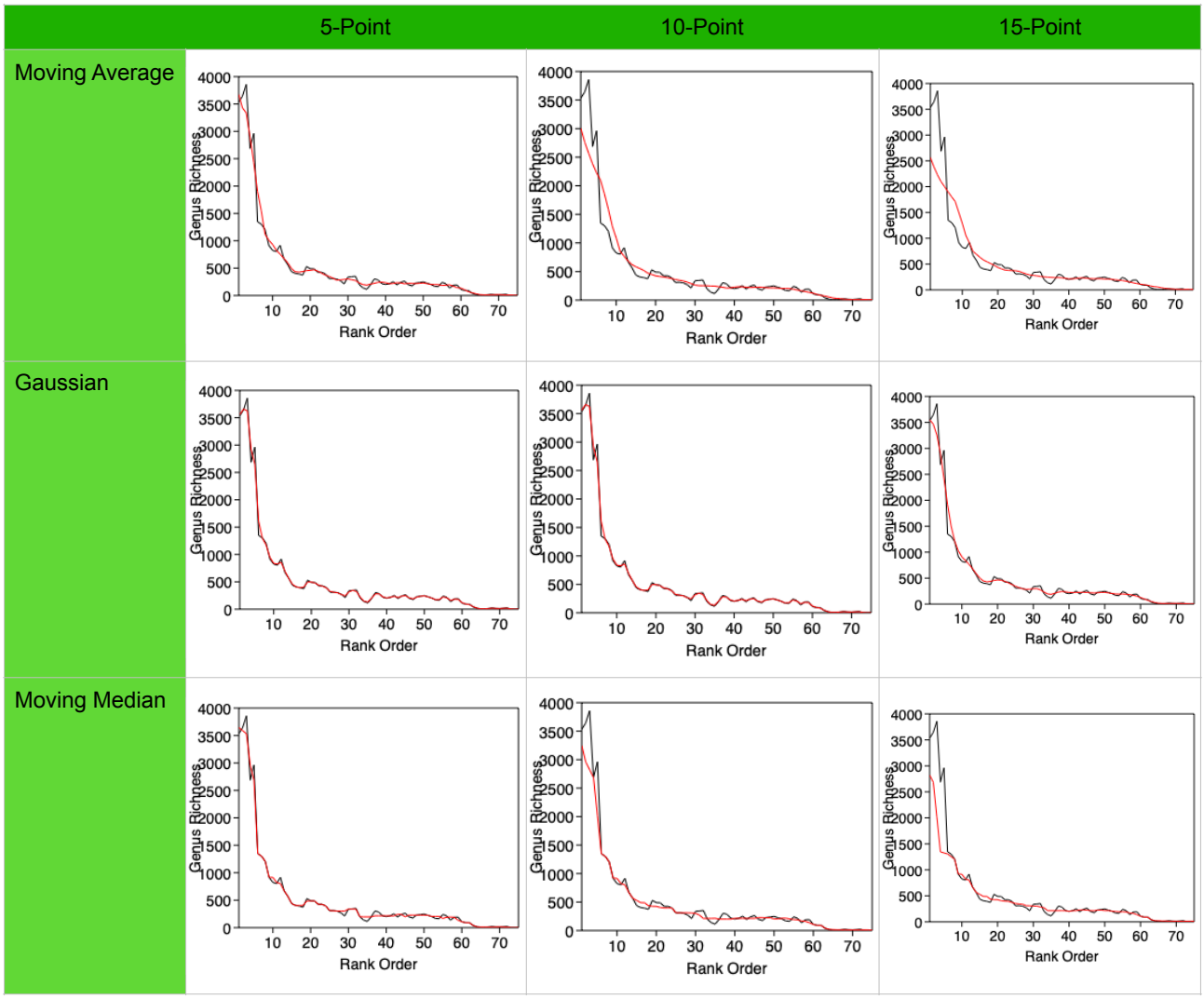
	5-Point	10-Point	15-Point
Moving Average			
Gaussian			
Moving Median			

Paleozoic Fauna Error Analysis

	5-Point	10-Point	15-Point
Moving Average	34.685	103.884	168.426
Gaussian	3.282	17.022	27.018
Moving Median	18.839	36.589	41.925

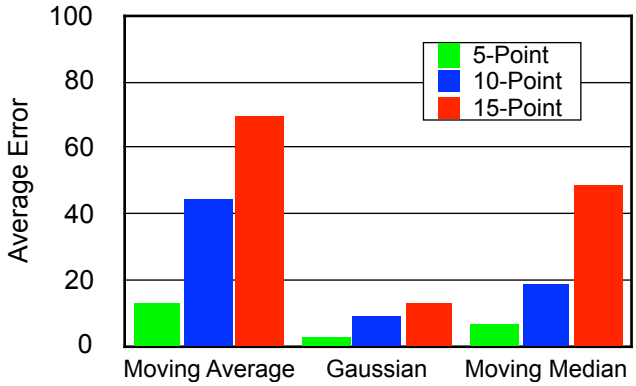


Modern Fauna



Modern Fauna Error Analysis

	5-Point	10-Point	15-Point
Moving Average	13.334	44.520	69.702
Gaussian	2.376	8.999	12.733
Moving Median	6.363	18.958	48.516



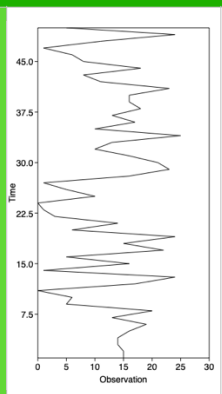
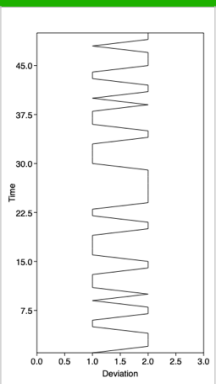
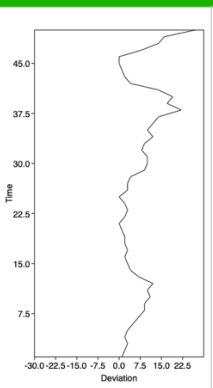
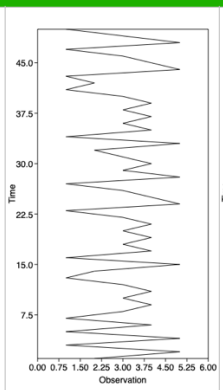
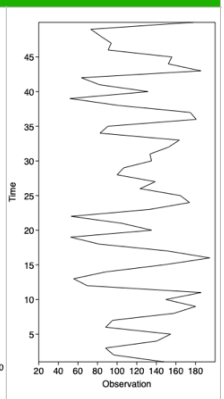
Overall, the Gaussian filter imposes the least distortion on the data. The reason for this is likely that this filter imposes a differential weight system that allows values close to the value being processed to have more influence over the filter's result. In the case of the moving average and median filters all values within the filter window receive equal weight this allows values at the window margins — which, in places, will include observations with very different magnitudes, to have a differential influence in the filter's result.

Clearly the simple moving average filter imposes the greatest distortion on the data. This also results from with equal weighting scheme imposed across the filter window, coupled with the fact that the simple average produces unstable results when used to analyze a collection of values that exhibit non-normal characteristics relative to the other two methods.

It should also be noted that, across all methods, short windows result in less distortion than long windows with the Gaussian method exhibiting, overall, the best error distribution characteristics.

6. Perform runs tests on the variables included in the “Runs.txt” datafile and determine which can be accepted as being indistinguishable from a random walk about the series mean. Reject the null hypothesis (H_0) if there is less that a 95 percent chance of it being true. (50 Points)

Runs Tests

	Walk 1	Walk 2	Walk 3	Walk 4	Walk 5
					
N_1	21	24	28	30	23
N_2	29	26	22	20	27
Obs. Runs	20	22	6	37	21
Exp. Runs	25.36	25.96	25.64	25	25.84
z	-1.573	-1.134	-5.696	3.575	-1.392
p	0.116	0.257	0.000	0.000	0.164
Interpretation	Accept H_o, series is a random walk	Accept H_o, series is a random walk	Reject H_o, series is not a random walk	Reject H_o, series is not a random walk	Accept H_o, series is a random walk

300 Total Points