In many time series situations the comparison of the means of two (or more) time series can be accomplished using a straightforward correction of the standard error that accounts for serial correlation. This section explains how to measure serial correlation, why a correction is necessary, and how to make the correction.

15.2.1 Serial Correlation and Its Effect on the Average of a Time Series

Serial correlation occurs when the course of a time series is influenced by its recent past. The typical behavior of a time series with serial correlation is that its values will go on extended excursions away from the long-run mean. This pattern is present in the logging and stream quality data. Let N_t be the concentration of nitrates in parts per million at the time t. The transformation $Y_t = \log(1 + 100 \times N_t)$ was made in order to adjust for skewness in the nitrate concentrations. (*Note:* This is essentially a log transformation, with a small amount added to handle the zeros.) Next, each of the time series was mean-corrected by subtracting the series average from each observation, producing ***residuals*** listed in Display 15.3.

Each residual series exhibits ***runs***, where the values tend to be consistently positive or consistently negative for long periods. This is more clearly evident in Display 15.4, a time plot of the residuals in the undisturbed watershed.

To demonstrate the trouble with using an average and its usual standard error with time series data, suppose that the average of the full five-year series is the long-run mean. In Display 15.4, this is represented by the zero line. The shaded region in the display highlights a segment of 27 values where the series has gone on an excursion away from its long-run mean. *What would happen if this segment were the only sample available and if it was treated as a series of independent measurements?* The sample average is 1.017, and the sample standard deviation is 0.551, giving a standard error of 0.106 for the average. The sample average is therefore 9.59 standard errors away from the long-run mean. If this sample were used to estimate the long-run mean, zero would be (incorrectly) considered an impossible value.

Serial correlation in the series created two problems. First, a sample average in a serially correlated time series tends to be farther from the long-run mean than expected, because the series is inclined to drift away from its long-run mean. Secondly, the serial correlation makes the values in the sample tend to be closer to each other than would be the case with an equal number of independent measurements, which makes them appear much less variable than they are. Since the estimate of how close the sample average is to the long-run mean is based upon sample variability, serial correlation engenders an overly optimistic view of the true variation in the sample average.

15.2.2 The Standard Error of an Average in a Serially Correlated Time Series

Fortunately, there is a simple way to adjust the standard error formula to account for serial correlation:

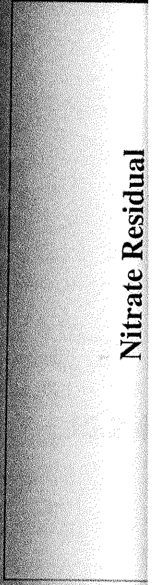
DISPLAY 15.3 Residual nitrate readings (log ppm) from two watersheds—one logged by patch-cutting and the other undisturbed

Week	Patch-cut	Undisturbed
1	−1.32	−1.21
4	−2.02	−1.91
7	−2.02	−1.91
10	0.18	−1.91
13	−0.92	−1.21
16	−2.02	−1.91
19	−0.92	−0.81
22	−0.63	−0.11
25	−0.22	−0.11
28	−0.22	−0.11
31	−0.92	−0.30
34	−0.63	−0.52
37	−0.92	−1.21
40	−2.02	−1.91
43	−0.07	−0.52
46	1.03	−1.91
49	1.16	−0.81
52	1.45	−0.81
55	0.69	−0.81
58	1.20	−1.21
61	1.32	−1.21
64	1.42	0.66
67	2.01	1.23
70	2.27	1.35
73	1.35	1.35
76	−0.07	1.19
79	−0.41	1.65
82	−0.92	2.14
85	−0.22	1.65
88	−0.92	1.56
91	−1.32	1.35
94	1.12	0.66
97	1.94	1.04
100	2.01	0.58
103	2.78	1.46
106	2.74	0.29
109	1.70	−0.30
112	0.76	0.49
115	0.55	1.23
118	−0.92	0.80
121	−1.32	1.19
124	1.35	1.09
127	−2.02	1.27
130	−2.02	0.04
133	−0.41	0.98
136	−0.92	1.46
139	−0.63	0.29
142	−1.32	0.73
145	0.62	−0.30
148	0.62	−0.81
151	1.75	−0.52
154	2.46	0.29
157	2.43	−0.52
160	−0.41	−1.21
163	−0.41	0.04
166	1.20	−0.30
169	0.38	1.65
172	0.47	1.14
175	2.08	0.98
178	0.93	1.09
181	−0.41	0.66
184	0.18	0.80
187	−0.63	1.31
190	−0.41	1.19
193	−2.02	0.29
196	−0.07	−0.81
199	−1.32	−1.21
202	0.76	−0.81
205	−0.07	−1.21
208	0.82	−1.21
211	−0.63	−1.21
214	0.98	1.04
217	−2.02	0.80
220	−0.41	0.04
223	−0.92	0.29
226	0.06	0.66
229	−0.41	0.98
232	−2.02	−0.52
235	−0.41	0.49
238	−1.32	−0.81
241	−2.02	−0.81
244	−1.32	−0.81
247	−0.92	−1.21
250	0.29	−1.21
253	0.18	−1.91
256	0.18	−0.11
259	0.29	−0.81
262	0.62	−0.52

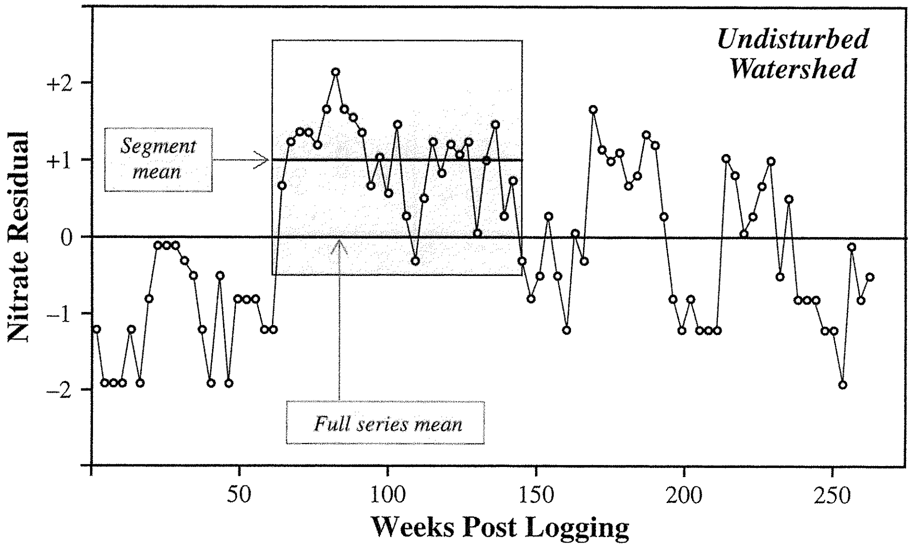
$$SE(\overline{Y}) = \sqrt{\frac{1+r_1}{1-r_1}}\,\frac{s}{\sqrt{n}}.$$

In this formula, the factor $s/\sqrt{n}$ is the standard error calculated as if the data were independent (s is the usual sample standard deviation for one sample, or the pooled estimate for two or more independent samples). The quantity r_1 is the sample *first serial correlation coefficient*. Notice that if r_1 is zero (no serial correlation), then the adjustment factor in front of the usual standard error is 1.

DISPLAY 15



DISPLAY 15.4 Mean-corrected nitrate concentrations after transformation, and a demonstration that the average of a segment of a time series may grossly misrepresent the full series mean



If r_1 is positive, then the adjustment factor is larger than 1; and if r_1 is negative, the adjustment factor is less than 1. Like other correlation coefficients, r_1 must fall between −1 and 1.

The nitrate series from the undisturbed watershed, shown in Display 15.4, has $r_1 = 0.744$ (calculation of r_1 shown in next section), so the adjustment factor is $2.61 = \{(1 + 0.744)/(1 - 0.744)\}^{1/2}$. The standard error of the serially correlated series is two and a half times what it would be if the series had independent measurements.

First-Order Autoregression Model

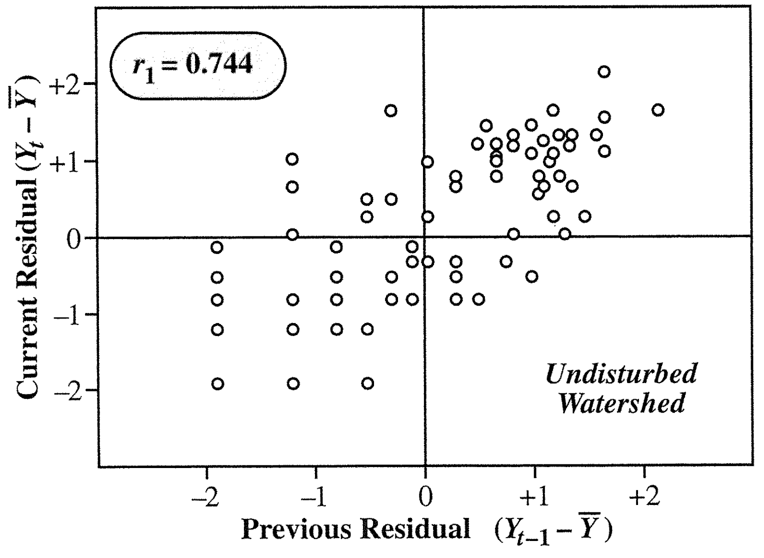
This adjustment to the standard error is appropriate under the following ideal model, called the *autoregressive model of lag 1*, or AR(1):

1. The series, $\{Y_t\}$, is measured at equally spaced points in time.
2. Let $(Y_t - \nu)$ be the deviation of an observation at time t from the long-run series mean ν. Let $\mu\{(Y_t - \nu)\,|\,\textit{past history}\}$ be the mean of the tth deviation as a function of all previous deviations. Then,

$$\mu\{(Y_t - \nu)\,|\,\textit{past history}\} = \alpha(Y_{t-1} - \nu),$$

where the parameter α is the *autoregression coefficient*. In other words, the regression of a deviation on all previous deviations depends only on the most recent one. Hence, the term *lag 1*

DISPLAY 15.5 A lag plot showing the relationship between adjacent residuals in the undisturbed watershed time series



All values in an AR(1) series are correlated with each other, yet the entire structure of the correlation is contained in the single parameter α. In the AR(1) model α is also known as the *population first serial correlation coefficient.* It may be estimated by the *sample first serial correlation coefficient*, r_1, which is discussed next.

15.2.3 Estimating the First Serial Correlation Coefficient

In a time series that embarks on excursions from the mean, the value of a deviation at one point in time tends to be similar to the value of the deviation at the most recent point in time. This behavior is evident in the scatterplot of residuals adjacent in time. Display 15.5 shows the scatterplot for the series from the undisturbed watershed. The plotted data are the 87 pairs of consecutive residuals in Display 15.3: $(-1.21, -1.91)$, $(-1.91, -1.91)$, and so on.

The coefficient r_1 provides a numerical summary measure of the correlation between adjacent residuals, apparent in Display 15.5. From a single time series, r_1 is the sample correlation of the consecutive residuals, like those plotted in Display 15.5. The formula provided below is general enough to be extended easily to get a pooled estimate from several time series. The calculation of r_1 in a single time series is the ratio:

$$r_1 = \frac{c_1}{c_0},$$