Mon, May 3, 2010.
1. Partial regression and partial residual plots.
2. Serial correlation.
3. Autocovariance function.
4. ACF.
5. Correlogram.
6. Properties of correlograms.

1) Partial regression and partial residual plots.
Partial regression or added variable plots are described on p166 of chapter 6. We can remove the effect of the other x's by regressing y on all x except x_i to get residuals δ^, and regressing x_i on all other x to get residuals γ^. An added variable plot shows xaxis=γ^ and yaxis=δ^. Look for nonlinearity and especially for outliers. The slope is β^_i.

Partial residual plots are described in the handout. A partial residual plot shows yaxis=e+β^_i x_i and xaxis=x_i. yaxis = y minus predicted effect from other variables. The slope is β^_i. Look for outliers and especially for nonlinearity.

2) Serial correlation.
Often when data are recorded sequentially in time, the values are correlated with one another. See e.g. LA temperatures.
Some examples are presented on p309 in ch9.

3) Autocovariance function.

For any ts, Y_t, define
γ(t1,t2) = cov(Y_t1, Y_t2) = E {[Y_t1 - E(Y_t1)][Y_t2 - E(Y_t2)] }.

If E(Y_t1) = E(Y_t2) = μ , and this function
γ(t1,t2) only depends on the difference in time t2-t1 , which we'll call the *lag*, and denote k .

i.e. γ_k = cov(Y_t, Y_t+k) , which is the same, for any t.

4) Autocorrelation function, ACF.

Define ρ(t1,t2) = cor(Yt1, Yt2) = γ(t1,t2)/ √{V(Yt1) V(Yt2)} .

Often ρ(t1,t2) only depends on the lag k=t2-t1 , and we write

ρ_k = cor(Yt, Yt+k) , which is the same for any t.

If E(Y_t) = μ for all t and V(Y_t) = σ^2 for all t, then ρ_k = γ _k / σ^2 or
ρ_k = γ_k / γ_0 .

ρ_k is the auto-correlation function (acf).

5. Correlogram.

Note that both the autocovariance and autocorrelation are theoretical properties of a ts. They depend on the expected value, for instance. Given a sample of n observations of a ts, we might estimate the sample autocovariance or sample autocorrelation.

Sample autocov = $c_k = 1/n \sum (Y_t - \underline{Y}) (Y_{t+k} - \underline{Y})$. The sum is from t=1 to n-k. $\underline{Y}$ means the sample mean of Y = $(Y_1 + ... + Y_n) / n$

Similarly, you can compute the
sample autocorrelation function (sample acf) = $r_k = c_k / c_0$.

A correlogram is a plot of r_k versus the lag, k.

6. Properties of correlograms (and autocorrelation functions).

a) Even: $\rho_k = \rho_{-k}$, and $r_k = r_{-k}$.

b) Both are between -1 and 1 ; i.e. $\leqq 1$ in absolute value.

c) $\rho_0 = r_0 = 1$.

d) If Yt has a seasonal component with period τ (e.g. one year),

then r_k and ρ_k also have seasonal components with period τ .

Often the correlogram starts out high and gradually decays.

For iid errors with mean 0 and variance σ^2, $\rho_k = 0$, for $k \neq 0$.
$r_k \sim -1/n$, +/- about $2 / \sqrt{n}$, for $k \neq 0$.
So for confidence bounds, you can use $-1/n \pm 1.96/\sqrt{n}$. People often use instead $0 \pm 2/\sqrt{n}$.