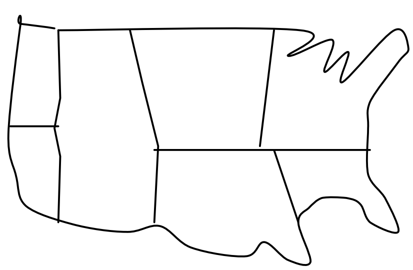
Classical Statistical Forecasting

Basically, statistical weather forecasting is linear regression: given a predictand (e.g., surface temperature in DCA), choose predictors available in time to perform a forecast. For example, forecast tomorrow's Tmin given today's observations.

1) Stratification and compositing:

In order to make the coefficients b_k more reliable, stratify the data into homogeneous bins, rather than mixing inhomogeneous data. Examples: stratify data according to season, and compute separate regression equations for each season separately. Stratify data for long-range forecasting into El Niño, La Niña, and non-ENSO years.

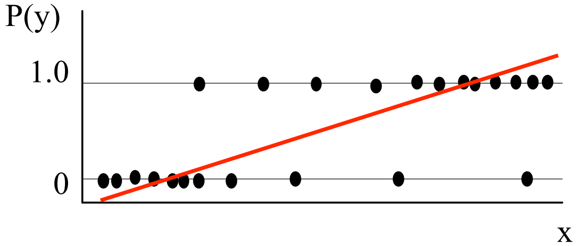
In order to increase the size of the dependent sample, composite several similar dependent data. Example: divide the country into "homogeneous" regions and assume that the same regression equation applies to all the stations within a regions. Or, since La Niña response is approximately equal and opposite to El Niño, composite El Niño events with "minus La Niña"



2) Prediction of a yes-no event.

Simple approach: $y = \begin{Bmatrix} 1 & \text{if yes} \\ 0 & \text{if no} \end{Bmatrix}$ and use regression.

Regression estimation of event probabilities (REEP)



• Dependent sample measurement (y=0 or y=1)

REEP regression line

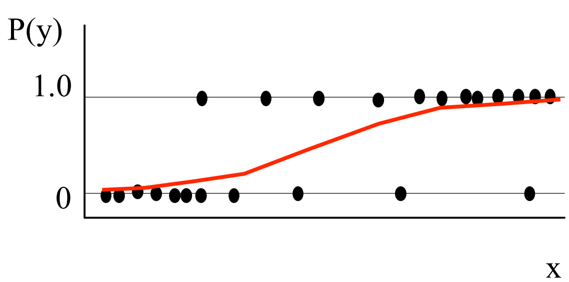
$$\hat{y}_i = b_0 + b_1 x_i$$

REEP: Determine a least-squares fit to the observations and interpret the result as a forecast of probabilities!!

Problems with this approach: 1) we don't know whether these are fair probabilities. 2) We can get P(y)<0 or P(y)>1.

If we change variables for the linear regression fit:

$\hat{y}_i = \dfrac{1}{1+\exp(b_0 + b_1 x_i)}$, we solve 2) but not 1).



(see chapter 7 of Wilks for a discussion on verification of probabilistic forecasts).

3) How to choose predictors and when to stop

Screening regression, also known as stepwise regression

1. Gather a pool of M physically reasonable potential predictors and a dependent (developmental) set of predictors and a predictand.

2. Do a 1-predictor regression with each potential predictor and choose as the first predictor the one with best R^2 (ρ^2 for a single predictor), i.e., the one with smallest SSE.

3. Do a 2-predictor regression, with the first predictor combined, in turn, with each other predictor, to find the 2-predictor combination that gives the best R^2. (i.e., screen all other predictors for the best combined R^2, or largest F-ratio MSR/MSE).

4. Repeat 3) with 3 predictors, adding to the first two predictors, in turn, all the others, and screening for the best R^2, largest F-ratio MSR/MSE.

…

Alternatively, one can use backward elimination: start with all M potential predictors and drop the least important. A good approach is to drop the one with the smallest t-ratio, indicative that its regression coefficients are least significant. Then drop other predictors after recomputing the regression with M-1 predictors.

Note that because predictors may be mutually correlated, both the forward and the backward screening require recomputation of the regression coefficients after adding or dropping a predictor.

When to stop adding predictors? There is no certain rule! Rules of thumb:

* When R^2 is increased by less than 5%, or

* When $\dfrac{SSE}{n-K-1} = s_\varepsilon^2$ does not decrease appreciably.

Some statistical packages (like Excel) provide an Adjusted R^2:

$$R^2_{adjusted} = 1 - \left(\frac{n-1}{n-K}\right)\left(1 - R^2\right)$$

This is a correction that compensates for the tendency of R^2 to increase with the number of predictors in the dependent sample, even if there is no significant additional information.

Best approach: test with independent data!