Subject: Re: Regression fit and random noise
Posted by Craig Markwardt on Fri, 29 Mar 2013 01:16:03 GMT

On Thursday, March 28, 2013 7:37:34 PM UTC-4, kisCA wrote:

> I understand the process of "destroying" the correlation. What I don't get is why does the coefficient of determination (R2) reach a plateau value (0.3) and doesn't get closer to zero as I raise the noise_ratio a lot (like a hundred)...
>
>> "underlying statistics"

The little cut and paste example below should show that the correlation factors do indeed converge to zero as the noise value is increased. Of course an individual sample of random scale factors may not make a R^2 value that goes to zero. R^2 itself has sample variance.

Craig


```
x = randomu(seed,100)  ;; Random X positions
ym = 0.3 - 0.7 * x     ;; Pure Y model (no noise)
ye = 0.01              ;; Initial scatter
;; Sampled y value
ys = ym + randomn(seed,100)*ye
print, r_correlate(x, ys)

;; NOISE_FACTOR multiples
noise_factors = [1, 10, 100, 1000, 10000]

;; Try different noise factors
for i = 0, n_elements(noise_factors)-1 do begin & ys1 = ym + randomn(seed,100)*ye *
noise_factors(i) & print, r_correlate(x, ys1)


;; EXAMPLE RUN:
;;    i NOISE  SPEARMAN   SIGNIF
;;    0    1  -0.998248    0.00000
;;    1   10  -0.904374  5.05695e-38
;;    2  100  -0.175770   0.0802491
;;    3 1000   0.0576417   0.568917
;;    4 10000  -0.0107651   0.915328
```