Subject: Re: regression with 95% confidence interval
Posted by ftian2012 on Mon, 06 Mar 2017 10:04:20 GMT
View Forum Message <> Reply to Message

On Monday, July 22, 2013 at 5:57:47 PM UTC+2, KenBowman wrote:
> On 2013-07-22 15:11:51 +0000, Fabien said:
>
>> On 07/22/2013 03:35 PM, Kenneth Bowman wrote:
>> I e-mailed you a routine that will compute confidence limits on the
>>> slope and intercept.
>>
>> I am interested in your method too ;-). Could you mail it to me too?
>
> It is a short routine, so I have pasted it below.
>
> The function computes a linear regression and confidence limits on the
> slope and intercept following the examples in Tamhane and Dunlop, which

> robustness of the confidence limits depends on whether the underlying
> statistics of the process in question satisfy the assumptions that are

> is assumed to be known with no error, and the error in the predictand

> for more details.
>
> Cheers, Ken
>
>
> FUNCTION REGRESSION_STATISTICS_KPB, x, y, $
> CI = ci, $
> DOUBLE = double, $
> VERBOSE = verbose
>
>
> ;+
> ; Name:
> ; REGRESSION_STATISTICS_KPB
> ; Purpose:
> ; This program computes the regression coefficients a and b and
> their confidence limits
> ; for simple bivariate linear regression
> ;
> ; y = a + b*x + eps
> ;
> ; The calculations follow the exposition in Chap. 10 of Tamhane and
> Dunlop (2000).
> ; Calling sequence:
> ; stats = REGRESSION_STATISTICS_KPB(x, y[, ci])

```
>  ; Inputs:
>  ;      x       : Values of the independent variables.
>  ;      y       : Values of the independent variables.
>  ; Output:
>  ;      stats   : Data structure containing various statistics and
>  confidence limits.
>  ; Keywords:
>  ;      CI      : Optional confidence limit (%).  If not set, a default
>  value of 95% is used.
>  ;      VERBOSE : If set, print the statistics and confidence intervals.
>  ; Author and history:
>  ;      Kenneth P. Bowman.  2007-11-01.
>  ;-
>
>  IF (N_PARAMS() NE 2) THEN MESSAGE, 'Incorrect number of arguments.'
>        ;Check number of arguments
>
>  IF (N_ELEMENTS(ci) EQ 0) THEN ci = 95.0D0
>        ;Default confidence limit to calculate
>
>  n = N_ELEMENTS(x)
>        ;Number of data points
>  IF (N_ELEMENTS(y) NE n) THEN $
>        ;Check array sizes
>     MESSAGE, 'The number of x and y values must be equal.'
>
>  b = (REGRESS(x, y, CONST = a, YFIT = yfit, FTEST = f_stat, $
>        ;Compute regression and extract value of b
>              CORRELATION = r, CHISQ = chisq, DOUBLE = double))[0]
>
>  s      = SQRT(TOTAL((yfit - y)^2)/(n-2))
>        ;Standard deviation of the residuals
>  x_mean = MEAN(x)
>        ;Mean of the x values
>  y_mean = MEAN(y)
>        ;Mean of the y values
>  S_xx   = TOTAL((x - x_mean)^2)
>        ;Sum of the squared deviations of x from the mean
>  S_yy   = TOTAL((y - y_mean)^2)
>        ;Sum of the squared deviations of y from the mean
>  S_xy   = TOTAL((x - x_mean)*(y - y_mean))
>        ;Sum of the products of the deviations of x and y from their
>  means
>  SE_a   = s*SQRT(TOTAL(x^2)/(n*S_xx))
>        ;Standard error of a
>  SE_b   = s/SQRT(S_xx)
>        ;Standard error of b
>  alpha = (1.0D0 - 0.01D0*ci)/2.0D0
```

```
>        ;Level for t-test
>  t_stat = T_CVF(alpha, n-2)
>        ;Compute t-statistic
>
>  IF KEYWORD_SET(verbose) THEN BEGIN
>     PRINT, 'Intercept (a)            : ', a
>     PRINT, 'Slope (b)                : ', b
>     PRINT, 'Correlation coefficient r : ', r
>     PRINT, 'r^2                      : ', r^2
>     PRINT, 'F-statistic              : ', f_stat
>     PRINT, 'Chi-square statistic     : ', chisq
>     PRINT, 'n                        : ', n
>     PRINT, 'Mean of x                : ', x_mean
>     PRINT, 'Mean of y                : ', y_mean
>     PRINT, 'S.D. of residuals        : ', s
>     PRINT, 'S_xx                     : ', S_xx
>     PRINT, 'S_yy                     : ', S_yy
>     PRINT, 'S_xy                     : ', S_xy
>     PRINT, 'S.E. of a                : ', SE_a
>     PRINT, 'S.E. of b                : ', SE_b
>     PRINT, 'Confidence limit         : ', ci, '%'
>     PRINT, 'Level for t-test         : ', alpha
>     PRINT, 't-statistic              : ', t_stat
>     PRINT, 'Confidence interval for a : ', a, '+/-',
>  STRTRIM(t_stat*SE_a, 2), '  [', a - t_stat*SE_a, ', ', a + t_stat*SE_a,
>  ']'
>     PRINT, 'Confidence interval for b : ', b, '+/-',
>  STRTRIM(t_stat*SE_b, 2), '  [', b - t_stat*SE_b, ', ', b + t_stat*SE_b,
>  ']'
>  ENDIF
>
>  RETURN, {a          : a,          $
>     ;Y-intercept
>           b          : b,          $
>     ;Slope
>           r          : r,          $
>     ;Correlation coefficient r
>           yfit       : yfit,       $
>     ;Fitted values at x
>           f_stat     : f_stat,     $
>     ;F-statistic
>           chisq      : chisq,      $
>     ;Chi-squared statistic
>           n          : n,          $
>     ;Number of points
>           x_mean     : x_mean,     $
>     ;Mean of the x values
>           y_mean     : y_mean,     $
```

```
>    ;Mean of the y values
>        s        : s,         $
>    ;Standard deviation of the residuals
>        S_xx     : S_xx,      $
>    ;Sum of the squared deviations of x from its mean
>        S_yy     : S_yy,      $
>    ;Sum of the squared deviations of x from its mean
>        S_xy     : S_xy,      $
>    ;Sum of the products of the deviations of x and y from their means
>        SE_a     : SE_a,      $
>    ;Standard error of a
>        SE_b     : SE_b,      $
>    ;Standard error of b
>        ci       : ci,        $
>    ;Requested confidence interval (%)
>        alpha    : alpha,     $
>    ;Level for t-test
>        t_stat   : t_stat,    $
>    ;t-statistic value
>        ci_a     : t_stat*SE_a, $
>    ;Confidence interval for a is (a +/- ci_a)
>        ci_b     : t_stat*SE_b  }
>    ;Confidence interval for b is (b +/- ci_b)
>
> END
```

Thank you very much KenBowman for sharing your excellent code!
