UNCERTAINTIES IN STRAIGHT LINE REGRESSION AND CALIBRATION

Revised 26 Feb 2004

1. Suppose we have *n* pairs of data $(x_1,y_1), (x_2,y_2),\ldots(x_n,y_n)$; x_i is the regressor variable controlled by the experimenter and measured with negligible error, while the response *y* is a random variable.

2. To fit a sample regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where β_0 and β_1 are constants and ε is $N(0,\sigma^2)$.

3. The point estimate of the mean of *y* for a particular *x* is given by $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$

4. Let us define the corrected sum of squares of *x* as $CS_{xx} = \sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$

5. And the corrected sum of the cross product of *x* and *y* as

$$CS_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

6. The estimated value of the slope is $\hat{\beta}_1 = \dfrac{CS_{xy}}{CS_{xx}}$

7. The estimated value of the y intercept is $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

8. The sum of squares of the residual errors is $SS_E = \sum_{i=1}^{n} \left(y_i - \hat{y} \right)^2$

9. The estimated variance of regression $\hat{\sigma}^2 = \dfrac{SS_E}{n-2} = MS_E$

10. The estimated standard error of the slope is $se\left(\hat{\beta}_1\right) = \sqrt{\dfrac{MS_E}{CS_{xx}}}$

11. A 100(1-α) percent confidence interval on the slope is $\hat{\beta}_1 \pm t_{\alpha/2,n-2}\, se\left(\hat{\beta}_1\right)$

12. The estimated standard error of the *y*-intercept is $se\left(\hat{\beta}_0\right) = \sqrt{MS_E\left(\dfrac{1}{n} + \dfrac{\overline{x}^2}{CS_{xx}}\right)}$

13. A 100(1-α) percent confidence interval on the *y*-intercept is $\hat{\beta}_0 \pm t_{\alpha/2,n-2} se\left(\hat{\beta}_0\right)$

14. A 100(1-α) percent confidence interval on the mean response at the point $x=x_0$ is

$$\hat{\mu}_{y|x_0} \pm t_{\alpha/2,n-2}\sqrt{MS_E\left(\frac{1}{n}+\frac{\left(x_0-\bar{x}\right)^2}{CS_{xx}}\right)}$$

15. A 100(1-α) percent interval for the future observation y_0 is

$$\hat{y}_0 \pm t_{\alpha/2,n-2}\sqrt{MS_E\left(1+\frac{1}{n}+\frac{\left(x_0-\bar{x}\right)^2}{CS_{xx}}\right)}$$

16. A 100(1-α) percent interval on the mean of m future observations at $x=x_0$ is

$$\hat{y}_0 \pm t_{\alpha/2,n-2}\sqrt{MS_E\left(\frac{1}{m}+\frac{1}{n}+\frac{\left(x_0-\bar{x}\right)^2}{CS_{xx}}\right)}$$

17. If this regression equation is used as a calibration equation, ie, to determine the value of x_0 corresponding to an observation y_0.

18. Then, a point estimate is $\hat{x}_0 = \dfrac{y_0-\hat{\beta}_0}{\hat{\beta}_1}$

19. If β_1 is not close to zero, then a 100(1-α) percent confidence interval for x_0 is $\bar{x}+d_1 \le x_0 \le \bar{x}+d_2$, where d_1 and d_2 are the roots of

$$d^2\left[\hat{\beta}_1^2-\frac{t^2_{\alpha/2,n-2}\hat{\sigma}^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right]-2d\hat{\beta}\left(y_0-\bar{y}\right)+\left[\left(y_0-\bar{y}\right)^2-t^2_{\alpha/2,n-2}\hat{\sigma}^2\left(1+\frac{1}{n}\right)\right]=0$$

Reference:

1. Montgomery, Douglas, C., Peck, Elizabeth, A., and Vining, Geoffrey, G., "Introduction to Linear Regression Analysis - Third Edition", 2001, John Wiley & Sons.