Significance and validity check of the linear regression model

- **Regression coefficient significance: It has to be checked if its absolute value is statistically different from the experimental error value;**

- **Model validity check: Does the values calculated by the model correspond to the experimental values in statistical sense (the well known “calculated/measured” test)**

Statistical estimates for significance and validity

- Student's test to check the difference between the regression coefficient value and experimental error value (if both values are statistically equal no regression relationship is at hand);

- Least squares method to assess the regression model validity (adequacy).

General scheme of least squares method

$$s^2_{total} = s^2_{model} + s^2_{residual}$$

$$s^2_{model} = \frac{SS_{model}}{df} = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{1}$$

$$s^2_{residual} = \frac{SS_{residual}}{df} = \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2}$$

$$F_{calc} = \frac{s^2_{model}}{s^2_{ressidual}} > 1$$

$$H_0 : s^2_{model} = s^2_{ressidual} \rightarrow no\ significant\ regression$$

$$H_1 : s^2_{model} > s^2_{ressidual} \rightarrow significant\ regression$$

Source	SS	df	MS
Model	557040.125	1	557040.125
Residual	17878.875	199	89.843593
Total	574919	200	2874.595

$$F_{calc} = 6200.11$$

Validity check by correlation coefficient

In this approach the correlation coefficient between the results of the experimental measurements and the results calculated by the linear regression model

$$y = ax + b$$

is determined.

If high correlation is shown the model is adequate. The lack of correlation indicates non – adequate model (no regression relationship).

Advantages and limitations of regression analysis

- Very useful for assessment of linear relationships (standard curve in analytical chemistry; time trends in environmental chemistry; linear models in physical chemistry; financial models; prognoses, etc.)

- Serious sensitivity to outliers which could compromise the model; possible outcome is to consider the system “before” the appearance of outlying result and “after” that.

Other specificities

- It is advisable to introduce (like in correlation analysis) different levels (ranks) of regression – “significant”, “average”, “weak”, insignificant”.

- **The regression relationship (different from correlation) is rather casual than stochastic relationship.**

Analysis of Variance - ANOVA