Lecture 7A Part A – Method of Standard Additions

Extrapolation

Extrapolation is the estimation of a value of a response variable (outside known range of responses) from values within the known range by assuming that the estimated value follows logically from the known values.

This is often considered to a fallacy. However, in some circumstance, it aids analysis.

The method of standard additions

Suppose that we wish to determine the concentration of silver in samples of photographic waste by atomic-absorption spectrometry. Using the methods outlined previously, an analyst could calibrate the spectrometer with some aqueous solutions of a pure silver salt and use the resulting calibration graph in the determination of the silver in the test samples.

This method is only valid, however, if a pure aqueous solution of silver, and a photographic waste sample containing the same concentration of silver, give the same absorbance values.

In other words, in using pure solutions to establish the calibration graph it is assumed that there are no '***matrix effects'***, i.e. no reduction or enhancement of the silver absorbance signal by other components (in short – contamination)

In many areas of analysis such an assumption is frequently invalid. Matrix effects occur even with methods such as plasma spectrometry, which have a reputation for being relatively free from interferences.

The solution to this problem is that all the analytical measurements, including the establishment of the calibration graph, must in some way be performed using the sample itself.

This is achieved in practice by using the ***method of standard additions***. The method is widely practised in atomic absorption and emission spectrometry and has also found application in electrochemical analysis and many other areas.

Equal volumes of the sample solution are taken, all but one are separately 'spiked' with known and different amounts of the analyte, and all are then diluted to the same volume.

The instrument signals are then determined for all these solutions and the results plotted as shown in Figure 5.9.

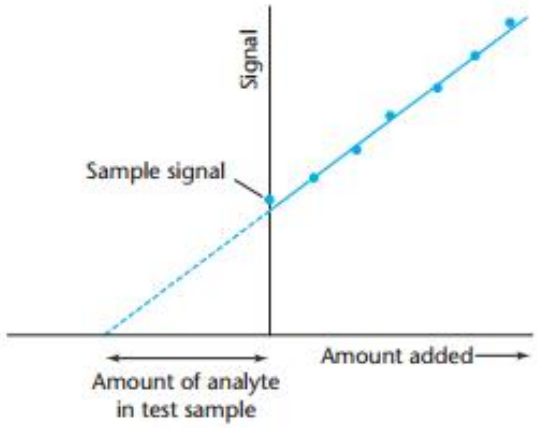


Figure 5.9 The method of standard additions.

As usual, the signal is plotted on the y-axis; in this case the x-axis is graduated in terms of the amounts of analyte added (either as an absolute weight or as a concentration).

The (unweighted) regression line is calculated in the normal way, but space is provided for it to be extrapolated to the point on the x-axis at which ***y= 0***. This negative intercept on the x-axis corresponds to the amount of the analyte in the test sample. Inspection of the figure shows that this value is given by ***a/b***, the ratio of the intercept and the slope of the regression line.

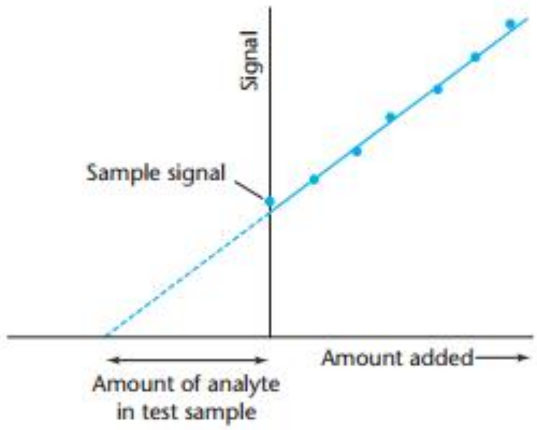


Figure 5.9 The method of standard additions.

Since both a and b are subject to error (Section 5.5) the calculated concentration is clearly subject to error as well. In this case, however, the amount is not predicted from a single measured value of y, so the formula for the standard deviation, s_{x_E}, of the extrapolated x-value (x_E) is not the same as that in equation (5.9).

$$s_{x_E} = \frac{s_{y/x}}{b}\sqrt{\frac{1}{n} + \frac{\bar{y}^2}{b^2 \sum_i (x_i - \bar{x})^2}} \qquad (5.13)$$

Increasing the value of n again improves the precision of the estimated concentration: in general at least six points should be used in a standard-additions experiment. Moreover, the precision is improved by maximizing $\sum_i (x_i - \bar{x})^2$, so the calibration solutions should, if possible, cover a considerable range. Confidence limits for x_E can as before be determined as $x_E \pm t_{(n-2)}s_{x_E}$.

Example 5.8.1

The silver concentration in a sample of photographic waste was determined by atomic-absorption spectrometry with the method of standard additions. The following results were obtained.

Added Ag: μg added per ml of *original* sample solution	0	5	10	15	20	25	30
Absorbance	0.32	0.41	0.52	0.60	0.70	0.77	0.89

Determine the concentration of silver in the sample, and obtain 95% confidence limits for this concentration.

Equations (5.4) and (5.5) yield $a = 0.3218$ and $b = 0.0186$. The ratio of these figures gives the silver concentration in the test sample as 17.3 μg ml^{-1}. The confidence limits for this result can be determined with the aid of equation (5.13). Here $s_{y/x}$ is 0.01094, $\bar{y} = 0.6014$, and $\sum_i (x_i - \bar{x})^2 = 700$. The value of s_{x_E} is thus 0.749 and the confidence limits are $17.3 \pm 2.57 \times 0.749$, i.e. 17.3 ± 1.9 μg ml^{-1}.

Lecture 7A Part B – ANOVA for Regression

In the ANOVA procedure, a hypothesis test (known as an F test) is used to test for the significance of the overall model. That is, it is used to test the null hypothesis that there is no relationship in the population between the (several) independent variables taken as a group and the one dependent variable.

Specifically, the null hypothesis states that all of the coefficients in the regression equation for the population are equal to zero. Therefore, for the case of two independent variables, or predictors, the null hypothesis with respect to the F test is H_0: $\beta_1 = \beta_2 = 0$

ANOVA for SLR

When the ***ordinary least-squares*** criterion is used to determine the best straight line through a single set of data points there is one unique solution, so the calculations involved are relatively straightforward.

When there is only one independent variable in the regression model, then the ANOVA procedure is equivalent to a two-tail t-test directed at the slope (H_0: $\beta_1 = 0$). Therefore, use of the ANOVA procedure is often not required in simple regression analysis in practice.

Separating Variances

The ANOVA method helps us to choose the best way of plotting a curve from amongst the many that are available. Analysis of variance (ANOVA) provides such a method in all cases where we maintain the assumption that the errors occur only in the y-direction. In such situations there are two sources of y-direction variation in a calibration plot.

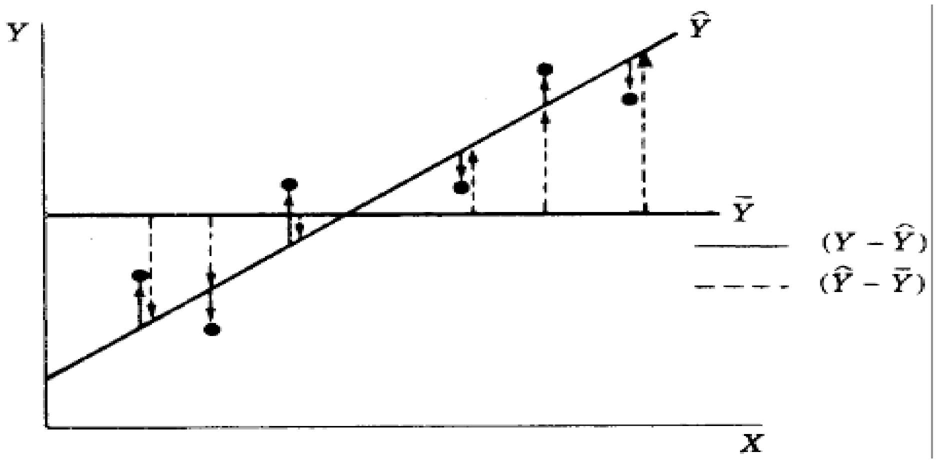
- The first is the variation due to regression, i.e. due to the relationship between the instrument signal, y, and the analyte concentration, x. (***SSR***)
- The second is the random experimental error in the y-values, which is called the variation about regression (***SSE***).

Important
TSS = SSR +SSE

ANOVA is a powerful method for separating two sources of variation in such situations.

If there is no regression effect in the population, then the fitted value (sloped) line differs from the mean (horizontal) line purely by chance. It follows that the variance estimate based on the differences - called ***mean square regression (MSR),*** would be different only by chance from the variance estimate based on the residuals called ***mean square error (MSE).***

On the other hand, if there is a regression effect, then the mean square regression is inflated in value as compared with the mean square error.



The following table presents the standard format for the analysis of variance table that is used to test for the significance of an overall regression effect. The degrees of freedom ***k*** associated with MSR in the table is the number of independent variables in the multiple regression equation.

As indicated in the table, the test statistic is the ratio of the two values. The p-value for the test statistic is provided in code output.

Source of variation	Degrees of freedom (df)	Sum of squares (SS)	Mean square (MS)	F ratio
Regression (R)	k	SSR	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Sampling error (E)	$n - k - 1$	SSE	$MSE = \frac{SSE}{n-k-1}$	
Total (T)	$n - 1$	SST		

The ANOVA table can be obtained for the regression model with the anova() command, when the model is specified. From this week's lab:

```
> anova(FitA)
Analysis of Variance Table

Response: Abso
          Df  Sum Sq Mean Sq F value    Pr(>F)
Conc       1 0.44327 0.44327  8979.5 2.481e-09 ***
Residuals  5 0.00025 0.00005
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sample Question (Q6d Autumn 2008/09 Dr. N Coffey)

Q6. A hospital administrator wished to study the relation between patient satisfaction (Y) and the patient's age (X_1, in years). She randomly selected 22 patients and collected the data some of which is presented below, where larger values of Y indicated more satisfaction.

i	1	2	3	4	5	6	7	8	$\cdots$	22
y_i	48	57	66	70	89	36	46	54	$\cdots$	52
x_{i1}	50	36	40	41	28	49	42	45	$\cdots$	44

The MINITAB printout for fitting the model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

is as follows:

```
Regression Analysis: Satisfaction versus Age, Severity of , Anxiety Leve

The regression equation is
Satisfaction = 163 - 1.21 Age - 0.666 Severity of Illness - 8.6 Anxiety Level

Predictor               Coef  SE Coef
Constant              162.47    26.51
Age                  -1.2179   0.3112
Severity of Illness  -0.6505   0.8452
Anxiety Level          -8.69    12.56

S = 10.2895

Analysis of Variance

Source          DF       SS      MS     F
Regression      (i)  4137.2  1379.1   (vi)
Residual Error (ii)     (iv)     (v)
Total         (iii)  6143.3
```

(i) Complete the ANOVA table by filling in the values for (i)-(vi). Conduct a hypothesis test to determine the significance of the linear regression model.

Sample size ***n = 22***

Number of independent variables

- ***k = 3***
- ***df1 = k = 3***
- ***df2 = n-k-1 =18***

(I will give the p-value also, and expect you to express your conclusions on whether or not slope estimates are jointly significant)