Appendix: Confidence Limits and Point Estimation of Several Regression Lines with a Common X-intercept.

D.S. Robson, Cornell University

Conventional methods of regression analysis provide formulas for estimates and confidence limits for the Y-intercept α and the slope β of the linear regression of Y on X. In the present circumstance, where Y is water inflow and X is osmotic pressure difference, a quantity of intrinsic interest is the X-intercept $M = -\alpha/\beta$ representing the osmotic pressure difference which results in zero average inflow of water. The X-intercept is likewise a basic parameter in bioassay analysis and in that context the following formula has been developed (Finney -) for 95 percent confidence limits on M:

$$\left[m-k^2\bar{X} \pm k\sqrt{(\bar{X}-m)^2 + (1-k^2)\,\Sigma\,(X-\bar{X})^2 n^{-1}}\right] / (1-k^2).$$

The variables appearing in this formula are functions of the estimated slope and Y-intercept,

$$b = \frac{\sum_{1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{1}^{n}(X_i-\bar{X})^2} \qquad\qquad a = \bar{y} - b\bar{X},$$

the standard error of b,

$$s_b = \sqrt{\frac{\sum_{1}^{n}(Y_i-a-bX_i)^2}{(n-2)\sum_{1}^{n}(X_i-\bar{X})^2}}$$

and the value of Students t at the 5 percent level on n-2 degrees of freedom; thus,

$$m = -\frac{a}{b} \qquad\qquad k = \frac{ts_b}{b}.$$

Biometrics Unit Paper No. BU-163; and
No. 555 in the Department of Plant Breeding and Biometry.

In the present case where several different experimental conditions might be expected to result in the same intercept for zero water uptake, a method is required for combining data from several experiments with different slopes and Y-intercepts in order to estimate and set confidence limits on their common X-intercept. If the residual variances

$$s_i^2 = \sum_{j=1}^{n_i} (Y_{ij}-a_i-b_iX_{ij})^2 \,/(n_i-2)$$

are homogeneous for the several experiments then the following formula gives the desired confidence limits:

$$\left[m-K^2\tilde{X}\pm K\sqrt{(\tilde{X}-m)^2 + (1-K^2)(C_X-\tilde{X}^2)}\,\right] /(1-K^2).$$

In this case m is the negative ratio of the sum of Y-intercepts and the sum of slopes,

$$m = -\sum_1^r a_i/\sum_1^r b_i.$$

The quantity $\tilde{X}$ is a weighted average of the r experimental means $\bar{X}_i$, with weights given by

$$\frac{1}{w_i} = \sum_{j=1}^{n_i} (X_{ij}-\bar{X}_i)^2$$

so that

$$\tilde{X} = \sum_1^r w_i\bar{X}_i/\sum_1^r w_i$$

and

$$C_X = \sum_1^r\left(\frac{1}{n_i} + w_i\bar{X}_i^2\right) / \sum_1^r w_i.$$

The pooled residual variance is

$$s^2 = \sum_1^r (n_i-2)s_i^2 \,/\Sigma(n_i-2)$$

and K is then defined by

$$K = \frac{ts\sqrt{\Sigma w}}{\Sigma b}$$

where t is now based on $\Sigma(n-2)$ degrees of freedom.

This analysis is illustrated in detail below, utilizing the data from Experiments 1 and 2 as displayed in Figure 4.

Experiment 1

Observation	1	2	3	4	5	6	7	8	9	10
X	-75	-67	-36	-32	4	6	32	41	58	77
Y	-27	-20	-14	-17	-3	0	6	10	22	22

$n_1 = 10$

$\Sigma X_1^2 = 24484$

$\frac{1}{10}(\Sigma X_1)^2 = 6.4$

$\Sigma(X_1-\bar{X}_1)^2 = 24477.6$

$\Sigma X_1 = 8$

$\Sigma X_1 Y_1 = 7973$

$\frac{1}{10}(\Sigma X_1)(\Sigma Y_1) = 16.8$

$\Sigma(X_1-\bar{X}_1)(Y_1-\bar{Y}_1) = 7989.8$

$\Sigma Y_1 = -21$

$\Sigma Y_1^2 = 2727$

$\frac{1}{10}(\Sigma Y_1)^2 = 44.1$

$\Sigma(Y_1-\bar{Y}_1)^2 = 2682.9$

$w_1 = 1/24477.6 = .00004085 \qquad b_1 = 7989.8w_1 = .3264 \qquad \alpha_1 = \frac{-21-8b_1}{10} = -2.36$

$s_1^2 = \frac{1}{8}(2682.9 - (7989.8)^2 w_1) = 9.36 \qquad s_{b_1} = \sqrt{w_1 s_1^2} = .0196$

$m_1 = \frac{2.36}{.3264} = 7.23 \qquad k_1 = \frac{2.306(.0196)}{.3264} = .1385 \qquad k_1^2 = .0192$

$$\frac{7.23 - .0192(.8) + .1385\sqrt{(.8-7.23)^2 + (1-.0192)(2447.76)}}{1-.0192} = 7.36 \pm 6.98$$

Experiment 2

Observation	1	2	3	4	5	6	7	8	9	10
X	-76	-72	-38	-36	7	9	36	42	68	81
Y	- 9	-16	- 7	-11	-1	2	7	7	12	12

$n_2 = 10$ $\quad\quad\quad\quad$ $\Sigma X_2 = 21$

$$\Sigma X_2^2 = 28075 \qquad \Sigma X_2 Y_2 = 4843$$

$$\frac{1}{10}(\Sigma X_2)^2 = 44.1 \qquad \frac{1}{10}(\Sigma X_2)(\Sigma Y_2) = -8.4$$

$$\Sigma(X_2-\overline{X}_2)^2 = 28030.9 \qquad \Sigma(X_2-\overline{X}_2)(Y_2-\overline{Y}_2) = 4851.4$$

$$\Sigma Y_2 = -4$$

$$\Sigma Y_2^2 = 898$$

$$\frac{1}{10}(\Sigma Y_2)^2 = 1.6$$

$$\Sigma(Y_2-\overline{Y}_2)^2 = 896.4$$

$w_2 = 1/28030.9 = .00003567 \qquad b_2 = 4851.4w_2 = .1731 \qquad \alpha_2 = \frac{1}{10}(-4-21b_2) = -.76$

$s_2^2 = \frac{1}{8}\,(896.4 - (4851.4)^2 w_1) = 7.10 \qquad s_{b_2} = \sqrt{w_2 s_2^2} = .0159$

$m_2 = \frac{.76}{.1731} = 4.39 \qquad k_1 = \frac{2.306(.0159)}{.1731} = .2118 \qquad k_2^2 = .0449$

$$\frac{4.39 - .0449(2.1) \pm .2118\sqrt{(2.1 - 4.39)^2 + (1 - .0449)(2803.09)}}{1 - .0449} = 4.50 \pm 11.48$$

Combined Experiments 1 and 2

$$m = \frac{2.36 + .76}{.3264 + .1731} = 6.24 \qquad \tilde{X} = \frac{4085\ (.8) + 3567\ (2.1)}{4085 + 3567} = 1.41$$

$$C_X = \frac{.1 + .1 + .00004085\ (.8)^2 + .00003567\ (2.1)^2}{.00004085 + .00003567} = 2616.09$$

$$C_X - \tilde{X}^2 = 2614.10 \qquad s^2 = \frac{8(9.36) + 8(7.10)}{8 + 8} = 8.23 \qquad \sqrt{s^2\Sigma w} = .0251$$

$$K = \frac{2.120\ (.0251)}{.3264 + .1731} = .1065 \qquad K^2 = .0113$$

$$\frac{6.24 - .0113\ (1.41) \pm .1065\sqrt{(1.41 - 6.24)^2 + (1 - .0113)\ (2614.10)}}{1 - .0113}$$

While the preceding calculations provide valid confidence limits for the common X-intercept M, neither the estimate $m = 6.24$ nor the confidence interval midpoint 6.30 represents the most efficient point estimate of M. The preferred estimates both of M and of the slopes β_1 and β_2 of the two regression lines intersecting at M are the maximum likelihood estimates obtained as the solution to the equations:

$$\hat{M} \quad \frac{\sum_{i=1}^{r} n_i \hat{\beta}_i^{\,2}\, \overline{X}_i - \sum^{r} n_i \hat{\beta}_i \overline{Y}_i}{\sum_{i=1}^{r} n_i \hat{\beta}_i^2} \qquad \hat{\beta}_i = \frac{\sum_{j=1}^{n_i} X_{ij} Y_{ij} - \hat{M} \sum_{1}^{n_i} Y_{ij}}{\sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2 + n_i\ (\overline{X}_i - \hat{M})^2}$$

The iterative solution to these equations is illustrated below with the data from Experiments 1 and 2. Initial values to start the iteration are taken as $\hat{\beta}_1 = b_1$ and $\hat{\beta}_2 = b_2$, giving the initial value for $\hat{M}$ as:

$$\hat{M}_{(0)} = \frac{(.3264)^2\ (8) + (.1731)^2\ (21) - (.3264)\ (-21) - (.1731)\ (-4)}{10\ (.3264)^2 + 10\ (.1731)^2} = 6.61$$

New values for $\hat{\beta}_1$ and $\hat{\beta}_2$ are then obtained from $\hat{M}_{(o)}$ as :

$$\hat{\beta}_1 = \frac{7973 - 6.61\ (-21)}{24477.6 + 10\ (.8-6.61)^2} = .32689 \qquad \hat{\beta}_2 = \frac{4843 - 6.61\ (-4)}{28030.9 + 10\ (2.1-6.61)^2} = .17247$$

The next trial value for $\hat{M}$ is then

$$\hat{M}_{(1)} = \frac{(.32689)^2 + (.17247)^2\ (21) - (.32689)\ (-21) - (.17247)\ (-4)}{10\ (.32689)^2 + 10\ (.17247)^2} = 6.613.$$

Evidently, further iterations would require the use of more decimal places than are needed, so the process stops here.