Inverse Prediction

One use of a regression model

$$E(Y) = \beta_0 + \beta_1 x$$

is to predict Y for a new x, x_0.

Sometimes, instead, we observe a new y_0, and want to make an inference about the new x_0.

Often x is expensive to measure, but Y is cheap; the relationship is determined from a *calibration* dataset.

Because

$$y_0 = \beta_0 + \beta_1 x_0 + \epsilon_0,$$

we can solve for x_0:

$$x_0 = \frac{y_0 - \beta_0 - \epsilon_0}{\beta_1}.$$

We do not observe ϵ_0, but we know that $E(\epsilon_0) = 0$.

Similarly, we do not know β_0 and β_1, but we have estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.

This suggests the estimate

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1}.$$

This is known as *inverse prediction*.

An approximate $100(1 - \alpha)\%$ prediction interval for x_0 is:

$$\hat{x}_0 \pm t_{\alpha/2} \times \frac{s}{\hat{\beta}_1} \times \sqrt{1 + \frac{1}{n} + \frac{(\hat{x} - \bar{x})^2}{SS_{xx}}}.$$

An alternative approach is to fit the *inverse regression*:

$$x = \gamma_0 + \gamma_1 y + \epsilon.$$

Then use the standard prediction interval

$$\hat{x}_0 \pm t_{\alpha/2} \times s_{x|y} \times \sqrt{1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{SS_{yy}}}$$

where

$$\hat{x}_0 = \hat{\gamma}_0 + \hat{\gamma}_1 y_0.$$

This is *not* supported by the standard theory, because, in the calibration data, x is fixed and y is random.

But it has been shown to work well in practice.