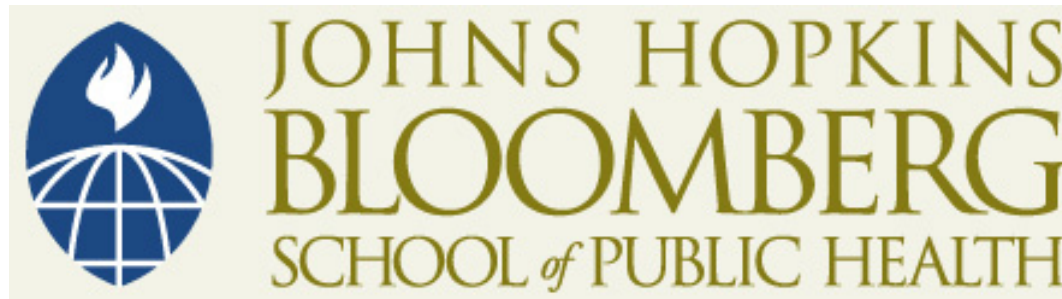
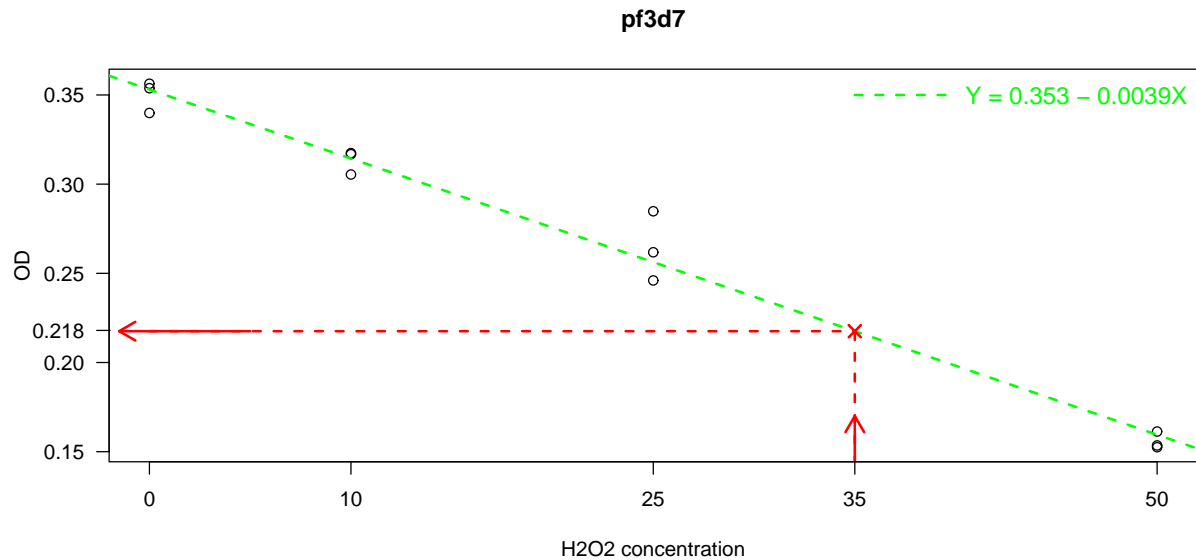


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2006, The Johns Hopkins University and Karl W. Broman. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

Estimating the mean response



We can use the regression results to predict the expected response for a new concentration of hydrogen peroxide. But what is its variability?

Variability of the mean response

Let \hat{y} be the predicted mean for some x , i. e.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Then

$$E(\hat{y}) = \beta_0 + \beta_1 x$$

$$\text{var}(\hat{y}) = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right)$$

where y is the true mean response.

Why?

$$\begin{aligned}E(\hat{y}) &= E(\hat{\beta}_0 + \hat{\beta}_1 x) \\&= E(\hat{\beta}_0) + x E(\hat{\beta}_1) \\&= \beta_0 + x \beta_1\end{aligned}$$

$$\begin{aligned}\text{var}(\hat{y}) &= \text{var}(\hat{\beta}_0 + \hat{\beta}_1 x) \\&= \text{var}(\hat{\beta}_0) + \text{var}(\hat{\beta}_1 x) + 2 \text{cov}(\hat{\beta}_0, \hat{\beta}_1 x) \\&= \text{var}(\hat{\beta}_0) + x^2 \text{var}(\hat{\beta}_1) + 2 x \text{cov}(\hat{\beta}_0, \hat{\beta}_1) \\&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\text{SXX}} \right) + \sigma^2 \left(\frac{x^2}{\text{SXX}} \right) - \frac{2 x \bar{x} \sigma^2}{\text{SXX}} \\&= \sigma^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\text{SXX}} \right]\end{aligned}$$

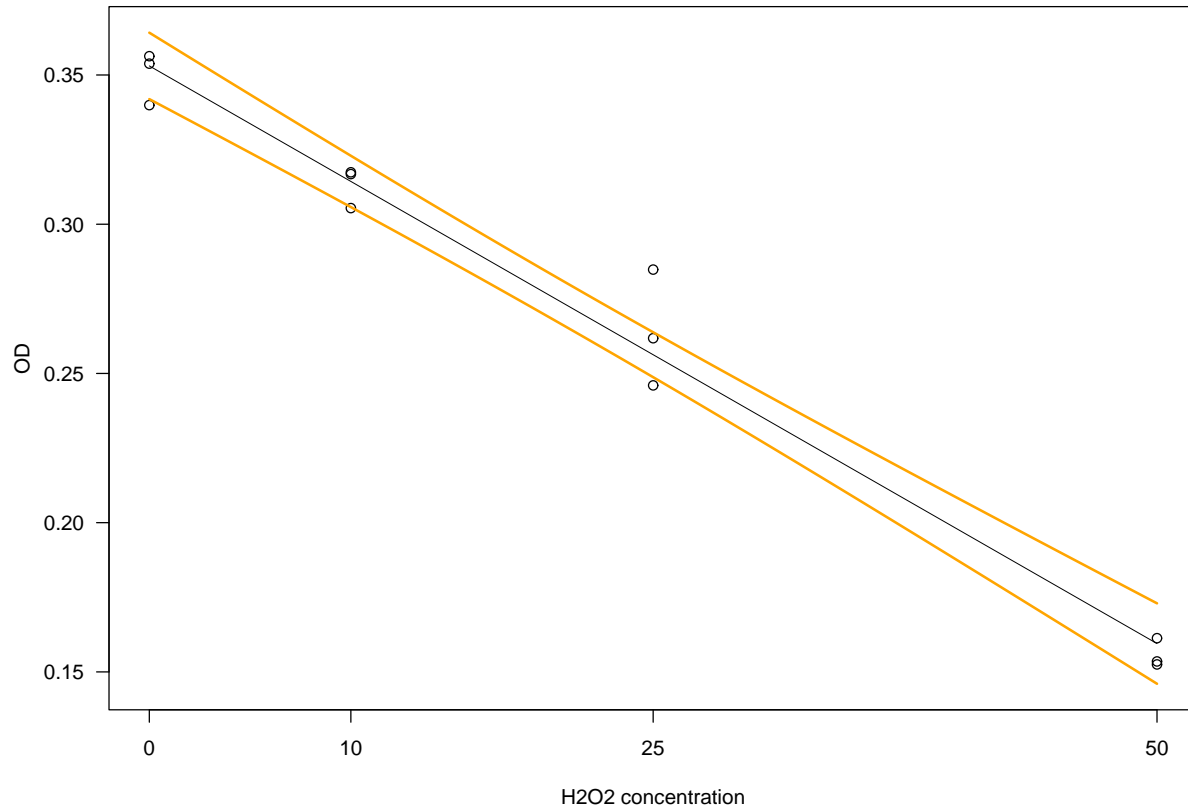
Confidence intervals

Hence

$$\hat{y} \pm t_{(1-\frac{\alpha}{2}), n-2} \times \hat{\sigma} \times \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\text{SXX}}}$$

is a $(1 - \alpha) \times 100\%$ confidence interval for the mean response given x .

pf3d7 - 95% confidence limits for the mean response



Prediction

Now assume that we want to calculate an interval for the predicted response y^* for a value of x .

There are two sources of uncertainty:

- (a) the mean response
- (b) the natural variation σ^2

The variance of \hat{y}^* is

$$\text{var}(\hat{y}^*) = \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX} \right)$$

Prediction intervals

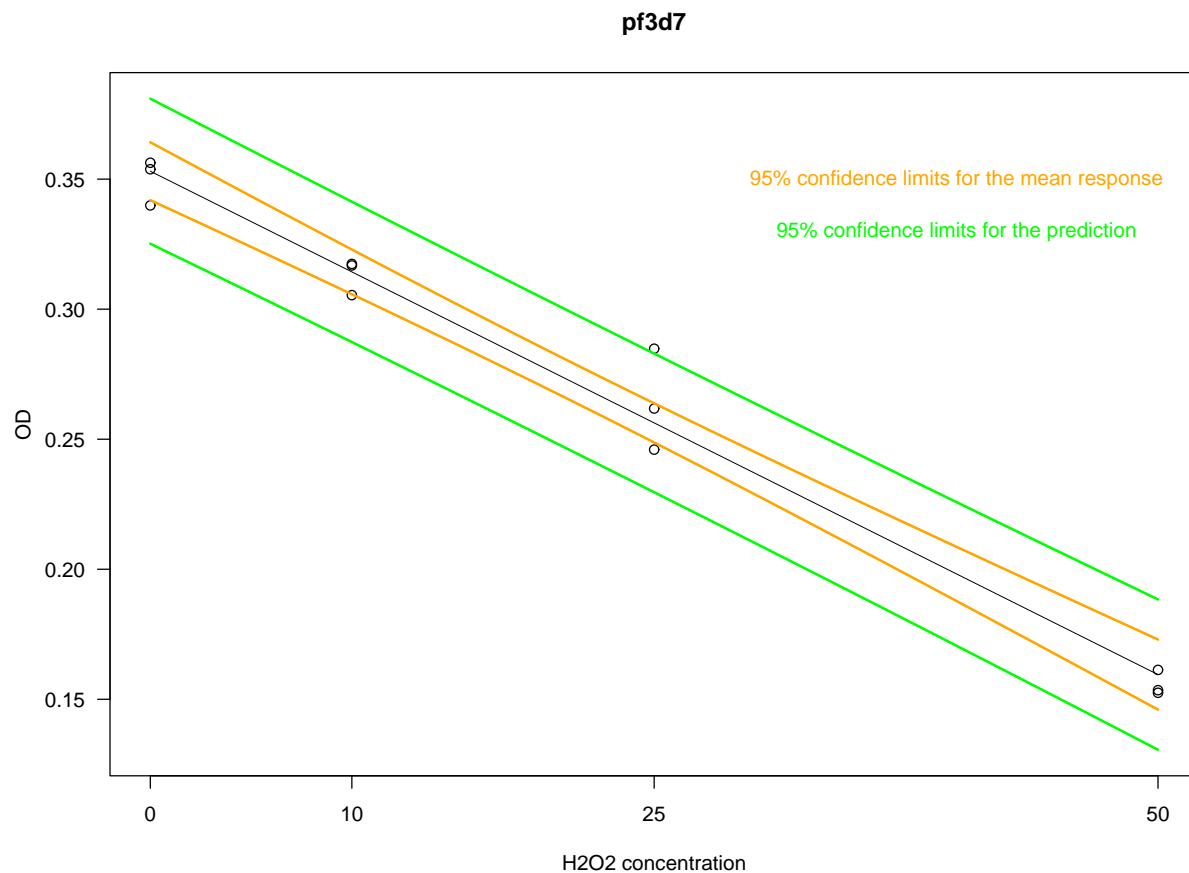
Hence

$$\hat{y}^* \pm t_{(1-\frac{\alpha}{2}), n-2} \times \hat{\sigma} \times \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SXX}}$$

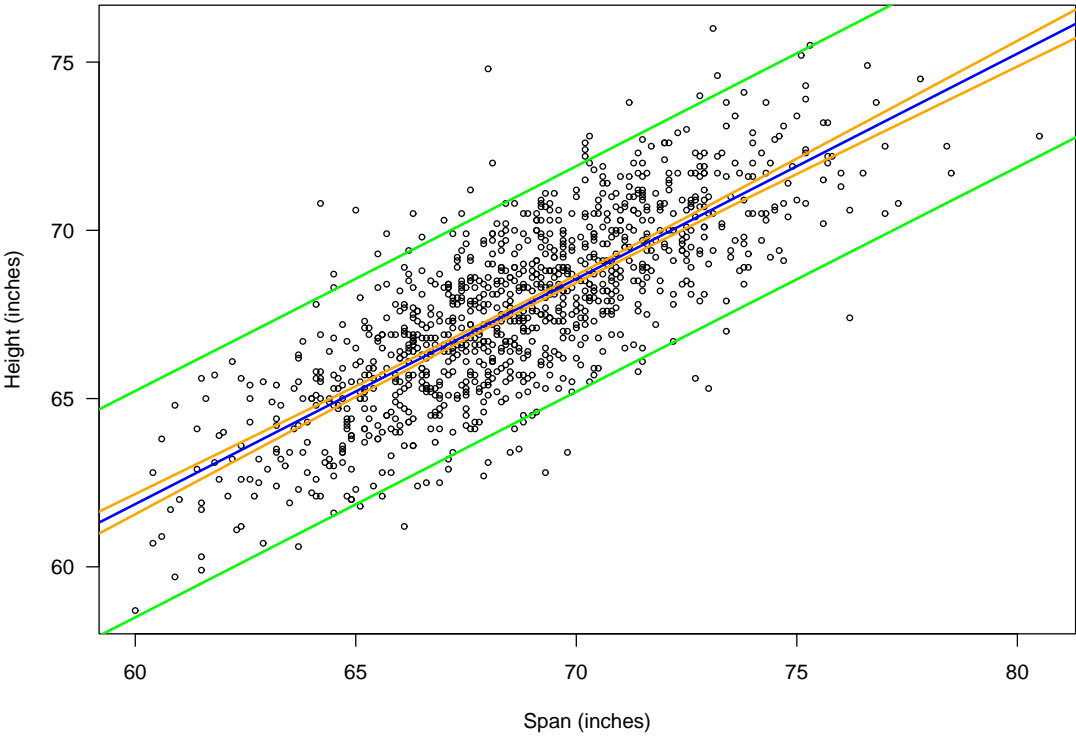
is a $(1 - \alpha) \times 100\%$ **prediction** interval for the predicted response given x .

Note: When n is very large, we get just

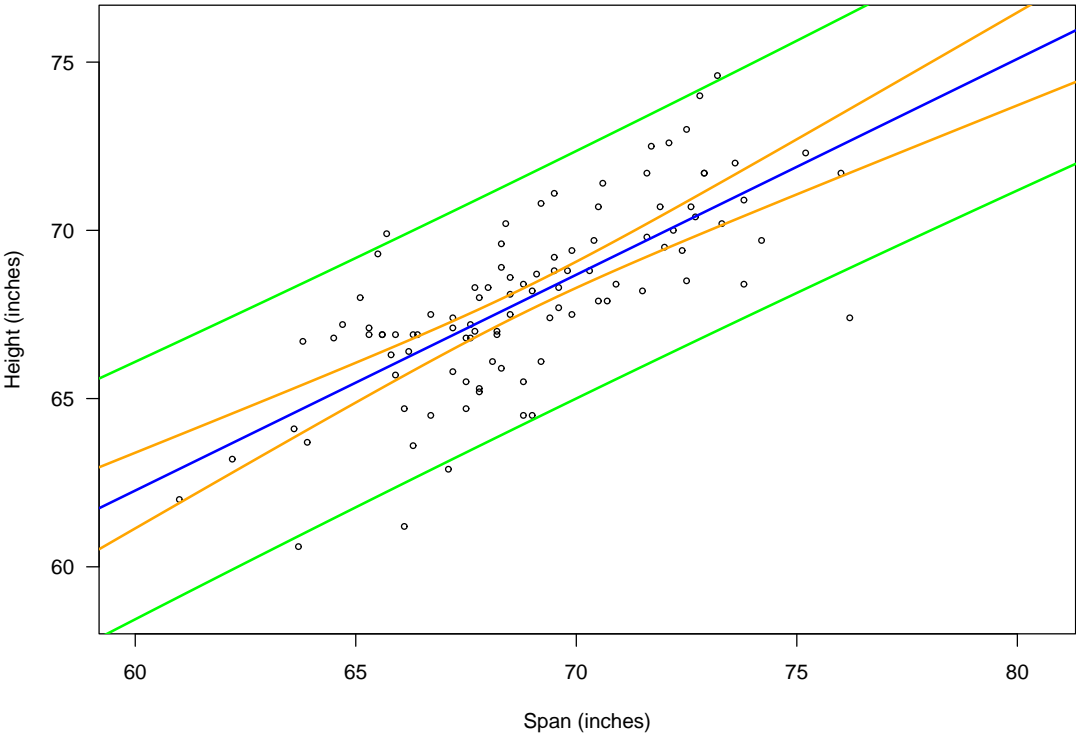
$$\hat{y}^* \pm t_{(1-\frac{\alpha}{2}), n-2} \times \hat{\sigma}$$



Span and height



With just 100 individuals



Regression for calibration

That prediction interval is for the case that the x 's are known without error while

$$y = \beta_0 + \beta_1 x + \epsilon \quad \text{where } \epsilon = \text{error}$$

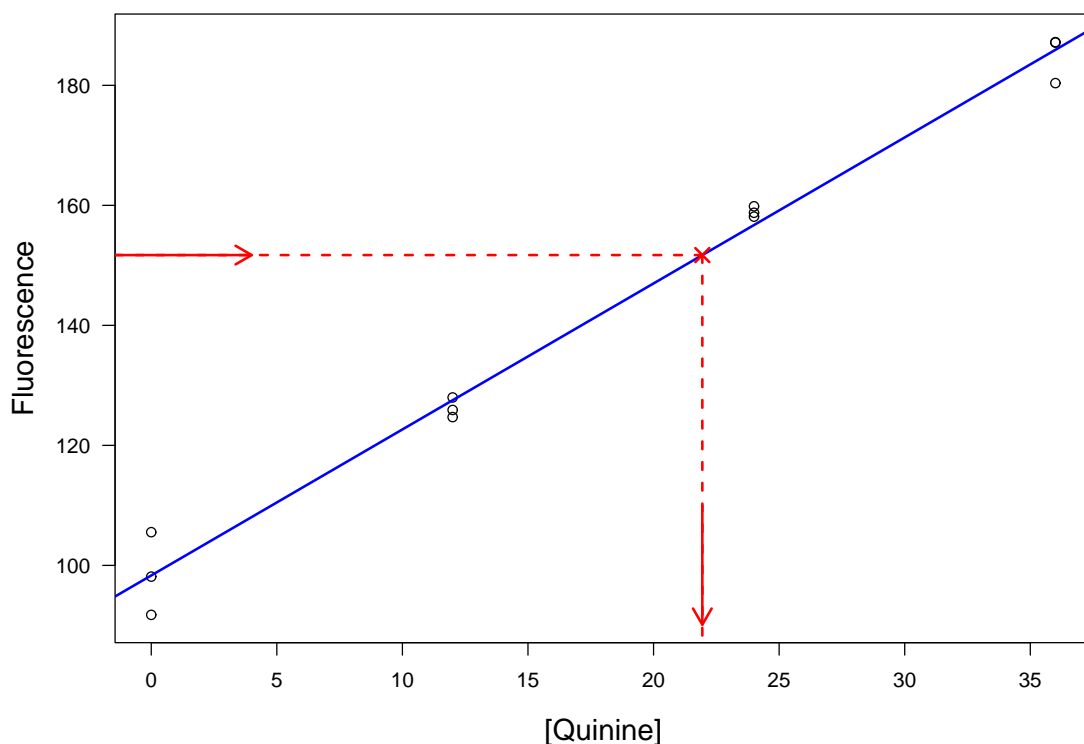
A more common situation:

We have a number of pairs (x,y) to get a calibration line/curve.
 x 's basically without error; y 's have measurement error

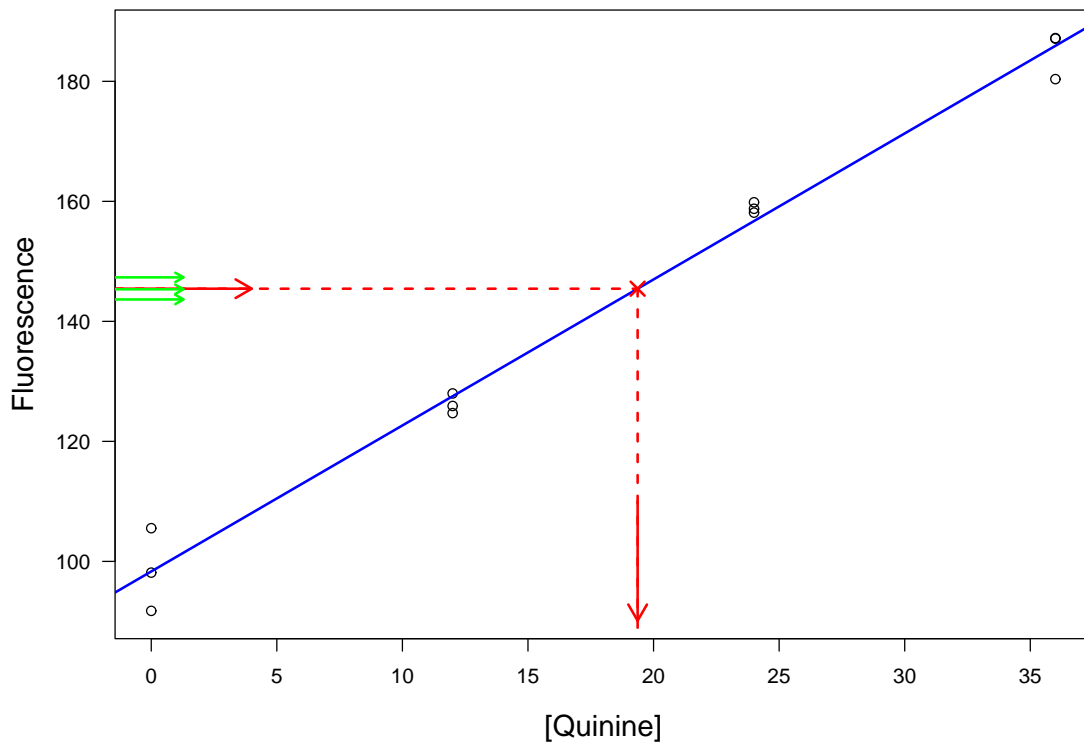
We obtain a new value, y^* , and want to estimate the corresponding x^* .

$$y^* = \beta_0 + \beta_1 x^* + \epsilon$$

Example



Another example



Regression for calibration

Data: (x_i, y_i) for $i = 1, \dots, n$

with $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \sim \text{iid Normal}(0, \sigma)$

y_j^* for $j = 1, \dots, m$

with $y_j^* = \beta_0 + \beta_1 x^* + \epsilon_j^*$, $\epsilon_j^* \sim \text{iid Normal}(0, \sigma)$

for some x^*

Goal: Estimate x^* and give a 95% confidence interval.

The estimate: Obtain $\hat{\beta}_0$ and $\hat{\beta}_1$ by regressing the y_i on the x_i .

Let $\hat{x}^* = (\bar{y}^* - \hat{\beta}_0) / \hat{\beta}_1$ where $\bar{y}^* = \sum_j y_j^* / m$

95% CI for \hat{x}^*

Let T denote the 97.5th percentile of the t distr'n with $n-2$ d.f.

$$\text{Let } g = T / [|\hat{\beta}_1| / (\hat{\sigma} / \sqrt{SXX})] = (T \hat{\sigma}) / (|\hat{\beta}_1| \sqrt{SXX})$$

If $g \geq 1$, we would fail to reject $H_0 : \beta_1 = 0$!

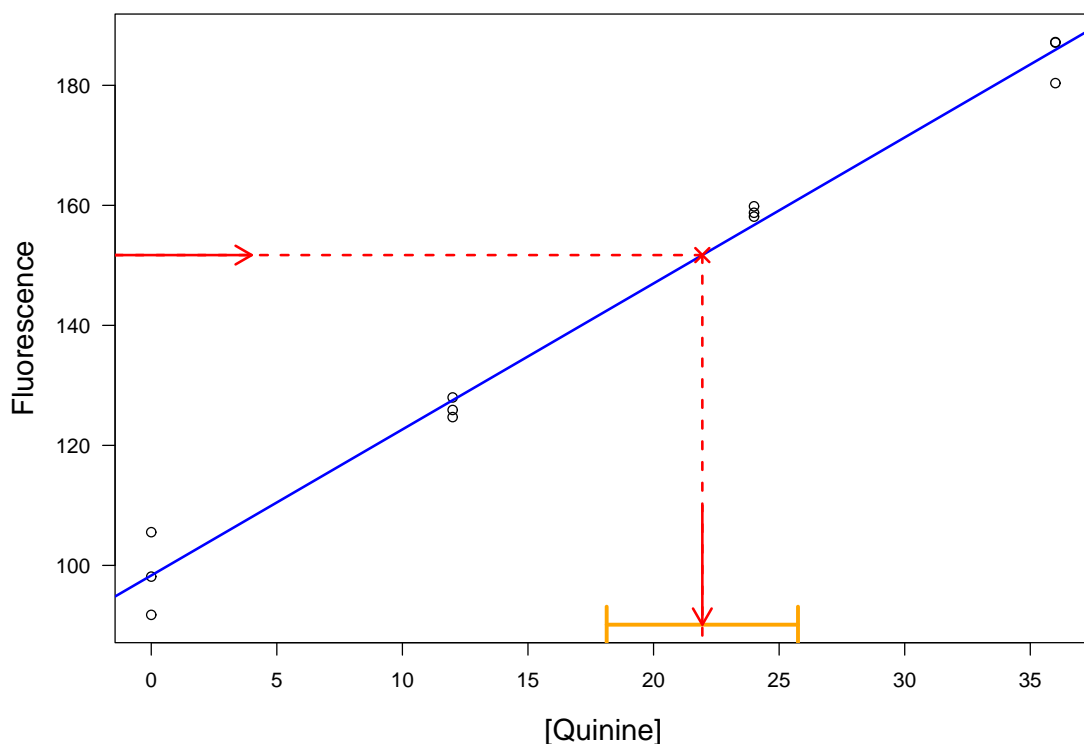
In this case, the 95% CI for \hat{x}^* is $(-\infty, \infty)$.

If $g < 1$, our 95% CI is the following:

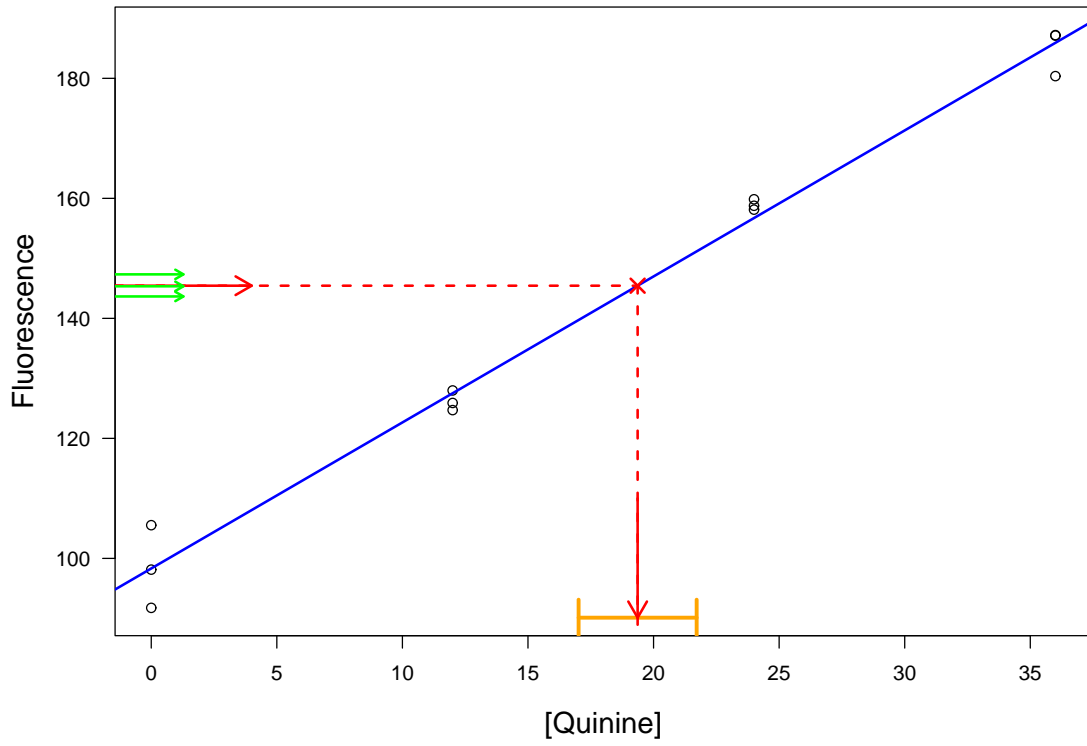
$$\hat{x}^* \pm \frac{(\hat{x}^* - \bar{x}) g^2 + (T \hat{\sigma} / |\hat{\beta}_1|) \sqrt{(\hat{x}^* - \bar{x})^2 / SXX + (1 - g^2) (\frac{1}{m} + \frac{1}{n})}}{1 - g^2}$$

For very large n , this reduces to $\hat{x}^* \pm (T \hat{\sigma}) / (|\hat{\beta}_1| \sqrt{m})$

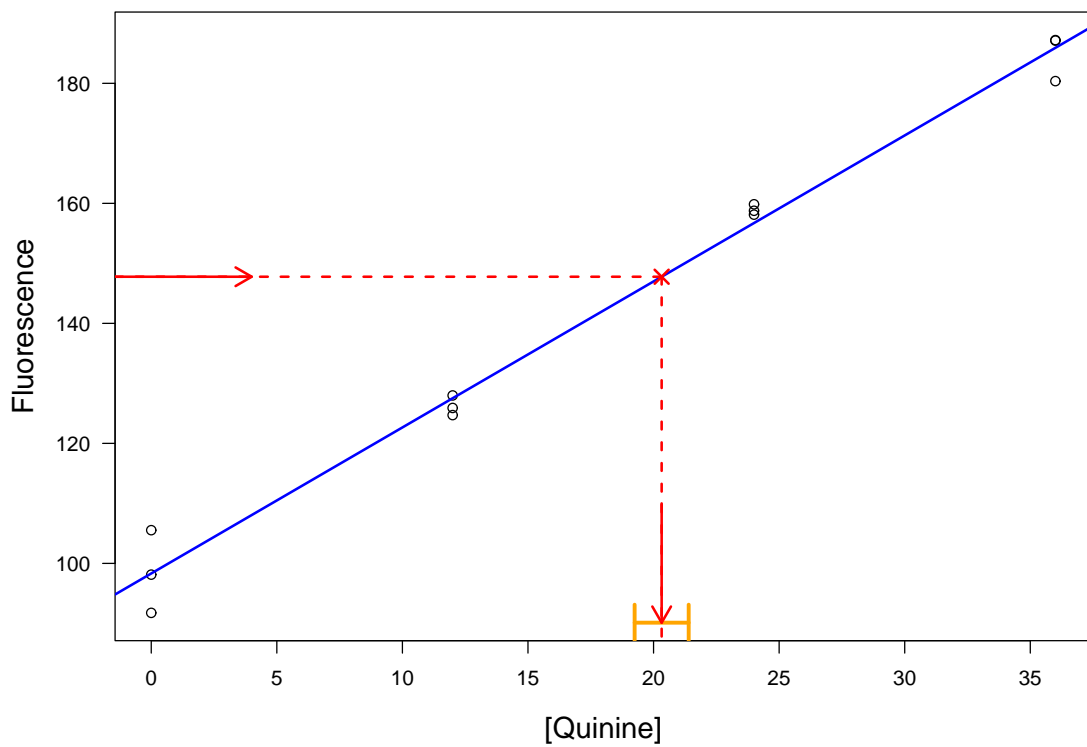
Example



Another example



Infinite m



Infinite n

