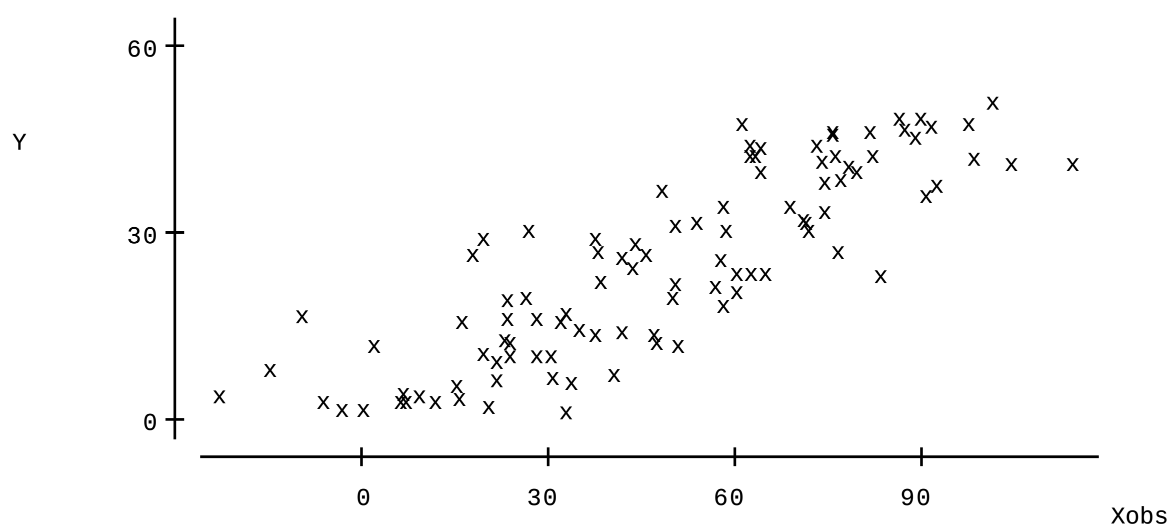
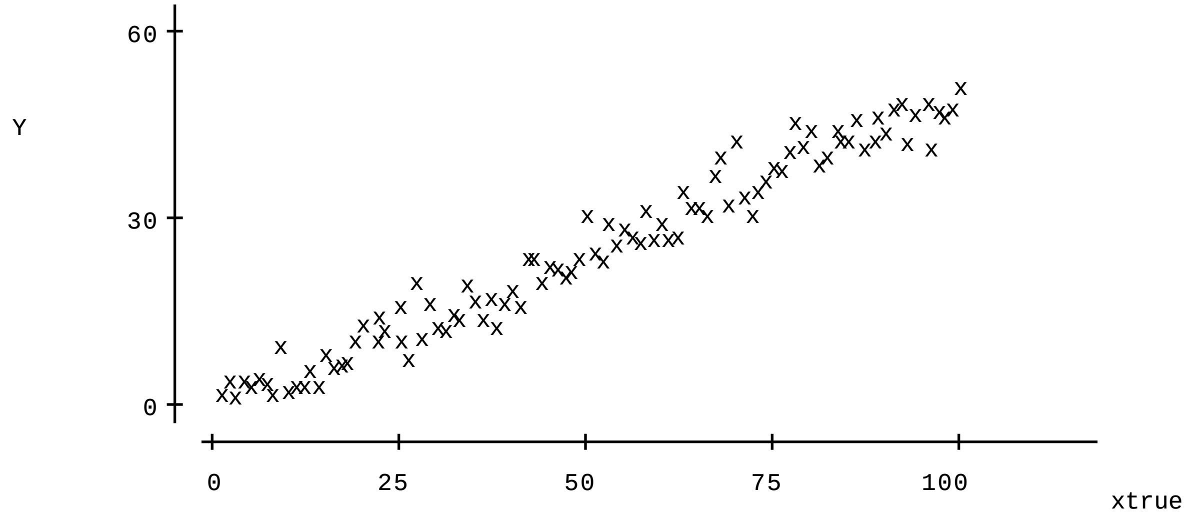
Measurement Error Models

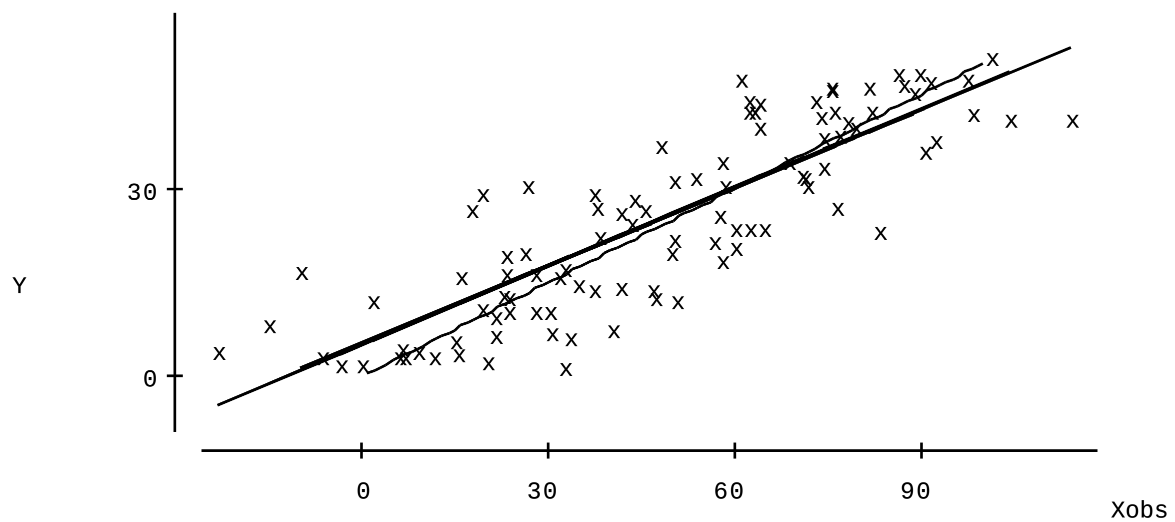
We have learned how to use least squares to fit models.

Most of those models assume that there is no error in the predictors.

What happens if there is?







It seems that there are two lines (and there always have been.)

Fit Y on X or X on Y (and invert the resulting slope.)

Identifiable parameters?

Suppose that the data is normally distributed:

$y = a + bx$

$X = xtrue + V$

$Y = y + e.$

We suppose that xtrue, v, and e are independent and normally distributed. The means of v and e are 0. The mean of xtrue and the variance of xtrue, v, e are unknown. In addition a and b are unknown parameters. Thus there are 6 unknown parameters.

If we observe n pairs (X_i, Y_i) having the same distribution as (X,Y) then there are 5 sufficient statistics. The sample means, sample variances, and the sample covariance.

When there are 6 parameters and only 5 sample statistics it often generally means that one

parameter can be set arbitrarily. Unfortunately its b
that can be set arbitrarily.

This is an example of a parameter that is not identifiable from the observations.

Fortunately, only the normal distribution has these parameters unidentifiable. This is an example of an ill-posed problem. No matter big n is as the distribution gets closer to normal no estimate will converge to the true value of b.

Strategies:

Most reduce the number of parameters that need to be estimated.

1. Variance of v is known.
2. Variance of e is known.
3. Ratio's of these variances is known.

In the past 20 years local modeling has made inroads to many areas of statistical science as well as chemistry and many other scientific fields.

We start with a univariate x.

Suppose we observe (x_i, Y_i) $i = 1,\ldots,n$.

$$Y_i = f(x_i) + \sigma(x_i)\varepsilon_i \ .$$

We suppose that the terms ε_i are i.i.d. standard normal. The functions f(x) and $\sigma(x)$ are unknown and are to be estimated.

Approach:

1. Fit ever-higher order models to the data.

Problems with this approach:

Large variance due to multicollinearity.

Large bias: Look at f(x) = |x| and suppose we approximate it at evenly spaced points $\pm\frac{j}{n}$ by a polynomial of degree 2n-1.

2. Fit piecewise models ion each region (without hooking them up). Column averages, piecewise linear.

Models each region by a straight line but this typically ignores scientific knowledge.

3. Piecewise smooth models.

Splines

Kernel Estimates

Nearest Neighbor Estimates

Best linear estimates

Same issues arise when trying to estimate $\sigma(x)$.

Soft modeling, Semi-soft modeling, and hard modeling.

Soft Modeling:

Let the data choose the model without much guidance about the model shape.

Semi-soft modeling:

Explore perturbations about a known shape but don't get too far away from known shape.

Hard modeling: Model shape is completely known up to a few parameters.

Soft Modeling:

$$\text{MIN} \sum_{i=1}^{n} (Y_i - \hat{f}(x_i))^2 + \lambda \sum_{i=1}^{n} (\hat{f}''(x_i))^2$$

$$\hat{f}(x_0) = \sum_{i=1}^{n} k\left(\frac{x_i - x_0}{b}\right) Y_i$$

The above are examples of linear estimates. The choices involve a "smoothing parameter" b or λ. It also means choosing the penalty or kernel. The explicit role that science plays in these estimates is not clear.

Semi-soft modeling:

Based on a search about a known model.

In theory A peak might have a Lorentzian shape

$f(x) = \frac{A}{1+x^2}$ but we don't expect that model to hold exactly.

We can try a model

1.$f(x) = \frac{A}{1+x^2} + R(x)$ with specified smoothness properties on the function R.

2. A convex combination of a $\frac{\hat{A}}{1+x^2}$ and $\sum_{i=1}^{n} k\left(\frac{x_i - x_0}{b}\right) Y_i$. The data can be used to weight each model.

Higher dimensions:

Curse of dimensionality.

Most observations from a uniform distribution occur near the shell of a high dimensional sphere.

Handle the dimensions of Y one by one.

Select one projection of Y at a time.

For now we deal with the case of one Y and several x's.

Carbon Dating done by Libby and his coworkers (1949,1950).

Radiocarbon or carbon-14 is produced in the upper atmosphere by the action of cosmic-ray-produced neutrons on N^{14} atoms,

$$ {}^{14}_{7}N + {}^{1}_{0}n \rightarrow {}^{14}_{6}C + {}^{1}_{1}H. $$

Like C^{12} and C^{13} it combines with oxygen to form carbon dioxide. After rapid mixing in the atmosphere and the upper levels of the ocean it is absorbed by plants and animals.

While plant and animals are alive it any radiocarbon that decays is replaced by new radiocarbon (as the plant or animal breaths).

Once an organism dies it can no longer replace radiocarbon but instead looses it exponentially with C^{14} half-life of 5730 ± 40 years. C^{14} remaining = (C^{14} at death)$\exp\left(\frac{-\log 2}{5730} \text{ Years dead}\right)$.

Then assuming that the in the atmosphere has remained constant the age of any organism can be found by reading backwards off the decay curve. Note living organisms have 1 part in 10^{12} of radiocarbon relative to C^{12}.

Radiocarbon in the atmosphere is known to vary with sunspot activity (giving 11 and 120-year periods of variation.)

The calibration curve:

Radiocarbon age (years before present) = Function (C^{14} half life, cabin concentration in atmosphere at time of organism death, Lab doing measurements,organism,…)

Estimate the curve using data from bristlecone pine.

How to produce a curve:

From theory or fit data.

As of the early 1980's there were about 900 tree ring carbon dates.

Laboratory differences?

Species differences?

If the curve is fit how smooth should it be?

This is an important question with important consequences.

Suess a well-known carbon dater fit his curve by eye.

Clark using an up to date smoothing routine.

Who is right? What consequence?

AVAS (additivity and variance stabilization)

Assume that $g^0(Y) = s^0(x) + \varepsilon$.

Where $g^0(Y)$ is (strictly) monotone, and ε has mean zero and is independent of x.

Single predictor case:

Given random variables X and Y, the goal is to find real valued measurable transformations s(X), and g(Y), such that

$E[g(Y) \mid X = x] = s(x)$

$\mathrm{var}[g(Y) \mid s(X)] = \text{constant}.$

The transformation g is assumed to be completely monotone

so we take it to be ↑ .

The simple idea:

If g is known the mean of $g^0(Y)$ as a function of X is simply

$s(x) = E[g(Y) \mid X = x].$

Recall that if the family of distributions of a random variable W has mean u and variance v(u) then the asymptotic variance stabilizing transformation for W is given by

$$h(t) = \int_0^t \frac{1}{\sqrt{v(u)}}\, du \,.$$

This fact can be used to find g(.) given s(.).

Algorithm:

Initialize:

Set $\hat{g}(Y) = \dfrac{Y\text{-}EY}{std(Y)}$ and $s(x) = E[\hat{g}(Y)\,|X]$.

2.Get the new transformation of Y:

a. Compute the variance function $v(u) \leftarrow var[g(Y) \mid s(X) = u]$.

b. Compute the variance-stabilizing transformation h(t) .

Set $\hat{g}(t) \leftarrow h(\hat{g}(t))$ and let $\hat{g}(t) \leftarrow \dfrac{\hat{g}(Y)-E\hat{g}(Y)}{std(\hat{g}(Y))}$.

3. Get the new transformation of X: $\hat{s}(x) \leftarrow E[\hat{g}(Y) | X = x]$.

4. Iterate steps 2 and 3 until $R^2 = 1 - E(\hat{g}(Y) - \hat{s}(x))^2$ does not change.

Multiple predictors: