Math 455 Class 32 April 18

Things you can do to change a linear model.

We discussed adding terms for X^2, X^3 etc.

- Can make a polynomial model. (degree of your choice)
- Can divide data into subsets and do a separate regression in each subset.
- We could require, if one subset is $X > 0$ and another is $X \leq 0$, that the regression lines meet at $X = 0$. (A piecewise-linear function.)
- Combine these ideas: a "spline".

A spline is a function which is piecewise polynomial. Example:

$$f(x) = \begin{cases} x^3 & x \geq 0 \\ -x^3 & x < 0. \end{cases} = |x|^3.$$

Many choices are possible, but usually splines match values at endpoints and derivatives at endpoints.

- Can do this in multiple variables, e.g.

model $Y \sim \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \beta_5 X_2^2 + \varepsilon$

This gives a surface in the Y, X_1, X_2 space.

Problem with this proliferation of models: if we have hundreds, or thousands of models that we test, the statistical tests no longer have the same meaning.

That is, if we flip a fair coin 7 times (our "experiment") the chance that the coin comes up heads every time is $\frac{1}{2^7} = \frac{1}{128}$.

All-heads: a surprising result.

If null hyp is "coin is fair"
we might reject the null on the basis of
this evidence.

Now suppose we do the experiment 100
times, and we get all heads once.

How do we reject the null hyp?

Probably not: we tried really hard to get
the result.

A similar problem happens if we try 100
models and find one that has statistically
significant coefficient estimates.

How do we know we weren't just lucky?

We do not: so there are procedures for
"model selection" (Ch. 10)

Before we turn over "model selection"
to an algorithm, we should consider "plausibility"

Example from p. 121 Corrosion & Iron Content.

A degree 6 polyn. fits the data perfectly, but there is no physical basis for this model. So we discard such models.

Not everything is so clear, so we might still consider lots of models.

Sample Test Question: We do 1000 statistical tests of models which have no predictive power. (Any significant test result is pure chance.) What is the probability that we get at least one ~~posit~~ significant test result at the 0.001 level?

A: each model has a $\left(1 - \frac{1}{1000}\right)$ chance of not producing a significant result.

the probability that ALL ~~produce~~ ~~no~~ do not produce a significant result is $\left(1 - \frac{1}{1000}\right)^{1000}$

$\approx e^{-1} \approx 0.37$

So the probability that we do get at least 1 significant result at the 0.001 level is $1-\left(1-\frac{1}{1000}\right)^{1000} \approx 1-e^{-1} \approx 0.63$.

Model Selection criteria are automated, algorithmic ways to pick a model from many options.

One obvious option: Adjusted R^2

and "Best Subset Regression".

If we have 3 variables, X_1, X_2, X_3, there are 8 potential models (all include an intercept β_0). We could look at all 8 and pick the one with highest adj. R^2.

But:
- if we have 20 variables, this means we'd have to consider $2^{20} \approx 10^6$ models
- this is just one of many possible methods.