Linking Goodness of Fit to Economic Gains

What’s the dollar benefit of more accurate predictions?

Bob Stine
Department of Statistics, Wharton

slides on web page after conference
Common Class Project

Challenge students taking regression to build a model used to
predict sales
forecast demand ...

Success shown by

Student communicates business implications

Higher $R^2$
How does fit convert to $? 

Students recognize that higher $R^2$ means a “better fit”, but what is the economic value? 

No matter which is used, how do you convert improvements of the model into dollars? 

For example 

$R^2$ of a model increases from 50% to 70% 

Can you convert that to money? 

Generic scenario?
News-vendor Problem

Context

Unknown demand
Perishable good
How much to make?

Uncertainty in demand implies wasted opportunity

Make too much: leftover product
Make too little: unmet demand

Aside

Another chance to address Type I/Type II error
Solution to News-vendor

Notation

Demand $D$ is random variable
Cost to manufacture $\text{cost}$
Selling price $\text{price}$

How much to supply?

Expected marginal value of $m^{th}$ item is
price $P(D > m) - \text{cost}$

Continue making so long as positive marginal value

$$\frac{\text{price}-\text{cost}}{\text{price}} = P(D \leq m)$$

$$\Pi = \%\text{profit} = P(\text{demand met})$$
Implication of Uncertainty

Add assumption

\[ D \sim N(10000, \sigma^2) \]

Label target to make \( \theta = \mu + \sigma z_{\pi} \)

%profit = 0.75

\( \sigma = 2500 \)

\( \theta \approx 11,700 \)
Implication of Uncertainty

Add assumption

$$D \sim N(10000, \sigma^2)$$

Label target to make $$\theta = \mu + \sigma z_{\pi}$$

%profit = 0.75

$$\sigma = 1000$$

$$\theta \approx 10,700$$
Expected Profit

What’s the expected profit?

Easy to find the optimal amount to produce
Harder to find associated expected profit

‘Modern’ approach

Just do it: calculate the expected profit of each item
Add them up

plot shows every 200th value
Expected Profit

What’s the expected profit?

Discrete calculation is exact
Hard to see forest for all the trees!

Approximation is simple in normal case
Assume choose optimal production amount
Expression for profit is linear in $\sigma$
Maximum ‘gain’ for small $\sigma$ if cost = 50% of price.

$$E(\text{profit}) \approx \mu \ (\text{price} - \text{cost}) - \sigma \ \text{price} \ \varphi(z_{\Pi})$$
profit if no uncertainty
\hspace{2cm} cost of uncertainty

$\varphi$ is normal density
Penalty for Uncertainty

Explicit role for error variation

Effect of profit/price ratio

\[ E(\text{profit}) \approx \text{best} - \sigma \text{ price } \varphi(z_\Pi) \]

Graph shows values for price=1
Penalty for Uncertainty

If you really do like $R^2$ over $\sigma$ then...

Better fit means smaller penalty

$$\text{E(profit)} \approx \text{best} - \sigma \sqrt{(1-R^2)} \text{ price } \varphi(z_{\pi})$$
Details: Less Variance

Marginal

\[ \mu = 10,000 \text{ with } \sigma = 2500 \]

Optimal production \( \theta = 11,686 \text{ units} \)

Expected profit

\[ 750,000 - 79,444 \approx \$671,000 \]

Regression \( R^2 = 0.60 \)

\[ \mu = 10,000 \text{ with } \sigma = 1000 \]

\( \theta = 10,675 \text{ units} \)

Expected profit

\[ 750,000 - 31,778 \approx \$720,000 \]
What if mean changes too?

Marginal mean ignores possible covariate information that affects mean response

Suppose mean grows as SD falls...

Better fitting models imply that expected response varies with covariate

\[
\begin{align*}
\mu &= 10000 \\
\sigma &= 2500 \\
\mu &= 13500 \\
\sigma &= 1000
\end{align*}
\]
Example: Wrong Mean

Setup

Marginal: \( \mu = 10,000 \) with \( \sigma = 2500 \)
Conditional: \( \mu = 13,500 \) with \( \sigma = 1000 \)

Marginal results

Sets \( \theta = 11,686 \) units, expecting $675,000 profit
Expectation* is really $875,000. Great!

Regression

\( \mu = 13,500 \) with \( \sigma = 1000 \) implies \( \theta = 14,175 \) Just left $100,000 on the table!

Expected profit is $980,722

*NB: This calculation requires a different expression when not at the optimal production threshold.
Closing Remarks

Opportunity to link to other classes

Supply chain, operations
Economics
Strategic management

Link to upper level classes

Random variables, quantiles
Where did that magic formula come from?

Advantage of calculus over discrete sums

The discrete calculation in Excel hides the elegant role of $\sigma$ and cost/price ratio

Thanks!