All Models are Right

... most are useless

Thaddeus Tarpey

Wright State University

thaddeus.tarpey@wright.edu

George Box's Quote

"All Models are Wrong, some are useful"

George Box's Quote

"All Models are Wrong, some are useful"

This quote is useful ...

George Box's Quote

> *"All Models are Wrong, some are useful"*

This quote is useful ... but wrong.

Here is an extended quote:

... The fact that the polynomial is an approximation does not necessarily detract from its usefulness because all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind."

From the book: Empirical Model-Building and Response Surfaces (1987, p 424), by Box and Draper.

Models are Approximations – Can approximations be Wrong?

$$\pi = 3.14$$

Models are Approximations – Can approximations be Wrong?

$\pi = 3.14$ This is WRONG

Models are Approximations – Can approximations be Wrong?

$\pi = 3.14$ This is WRONG

$\pi \approx 3.14$ is not wrong

Stay Positive

When teaching, why focus on the negative aspect of Box's quote:

> "Ok class, today I will introduce regression models.

Stay Positive

When teaching, why focus on the negative aspect of Box's quote:

> "Ok class, today I will introduce regression models. Oh, and by the way, all these models are wrong."

Stay Positive

When teaching, why focus on the negative aspect of Box's quote:

> "Ok class, today I will introduce regression models. Oh, and by the way, all these models are wrong."

Instead:

> "Ok class, today we will introduce regression models which can be very useful approximations to the truth."

Fallacy of Reification

Fallacy of Reification: *When an abstraction (the model) is treated as if it were a real concrete entity.*

Fallacy of Reification

Fallacy of Reification: *When an abstraction (the model) is treated as if it were a real concrete entity.*

- The fallacy of reification is committed over and over, even by statisticians, who believe a particular model represents the truth ... instead of an approximation.

Fallacy of Reification

Fallacy of Reification: *When an abstraction (the model) is treated as if it were a real concrete entity.*

- The fallacy of reification is committed over and over, even by statisticians, who believe a particular model represents the truth ... instead of an approximation.
- The model is not wrong but treating the model as the absolute truth (i.e. reification) is wrong.

A Dress, A suit, A Model

If a dress or suit fits nicely, it is useful...

A Dress, A suit, A Model

If a dress or suit fits nicely, it is useful...

If the model fits the data nicely, it can be a useful approximation to the truth.

A Dress, A suit, A Model

If a dress or suit fits nicely, it is useful...

If the model fits the data nicely, it can be a useful approximation to the truth.

"Does this model make me look fat?"

A Dress, A suit, A Model

If a dress or suit fits nicely, it is useful...

If the model fits the data nicely, it can be a useful approximation to the truth.

"Does this model make me look fat?"

"No dear"

If we just tweak the language a bit

In the simple linear regression model,

$$y = \beta_0 + \beta_1 x + \epsilon, .$$

Saying:
"Assume ϵ is normal" is almost always wrong.

If we just tweak the language a bit

In the simple linear regression model,

$$y = \beta_0 + \beta_1 x + \epsilon, .$$

Saying:
"Assume ϵ is normal" is almost always wrong.

Saying:
"Assume ϵ is approximately normal" will often be accurate.

A Quote

Paul Velleman writes:

> *"A model for data, no matter how elegant or correctly derived, must be discarded or revised if it does not fit the data or when new or better data are found and it fails to fit them."*

From "Truth, Damn Truth, and Statistics" in the *Journal of Statistical Education*, 2008.

Velleman's quote is useful ... but not always

Newton's 2nd Law of Motion $F = ma$ hasn't been discarded...

Velleman's quote is useful ... but not always

Newton's 2nd Law of Motion $F = ma$ hasn't been discarded...
... even though it has been revised due to Einstein's special theory of relativity

$$F = \frac{d\{mv\}}{dt}.$$

Velleman's quote is useful ... but not always

Newton's 2nd Law of Motion $F = ma$ hasn't been discarded...
... even though it has been revised due to Einstein's special theory of relativity

$$F = \frac{d\{mv\}}{dt}.$$

$F = ma$ is still a useful approximation...

Velleman's quote is useful ... but not always

Newton's 2nd Law of Motion $F = ma$ hasn't been discarded...
... even though it has been revised due to Einstein's special theory of relativity

$$F = \frac{d\{mv\}}{dt}.$$

$F = ma$ is still a useful approximation...as long as you don't go too fast.

Cylinder-Shaped Soda Can Example

Model soda volume as a function of height of soda in can

$$\text{Volume} = \beta_0 + \beta_1 \text{height} + \epsilon.$$

Then $\beta_0 = 0$ and $\beta_1 = \pi r^2$.

Cylinder-Shaped Soda Can Example

Model soda volume as a function of height of soda in can

$$\text{Volume} = \beta_0 + \beta_1 \text{height} + \epsilon.$$

Then $\beta_0 = 0$ and $\beta_1 = \pi r^2$.

Thad: I'm going to use a reduced model:

$$\text{Volume} = \beta_0 + \epsilon.$$

Cylinder-Shaped Soda Can Example

Model soda volume as a function of height of soda in can

$$\text{Volume} = \beta_0 + \beta_1 \text{height} + \epsilon.$$

Then $\beta_0 = 0$ and $\beta_1 = \pi r^2$.

Thad: I'm going to use a reduced model:

$$\text{Volume} = \beta_0 + \epsilon.$$

Fellow Statistician: "Hey Tarpey, your reduced model is wrong."

Cylinder-Shaped Soda Can Example

Model soda volume as a function of height of soda in can

$$\text{Volume} = \beta_0 + \beta_1 \text{height} + \epsilon.$$

Then $\beta_0 = 0$ and $\beta_1 = \pi r^2$.

Thad: I'm going to use a reduced model:

$$\text{Volume} = \beta_0 + \epsilon.$$

Fellow Statistician: "Hey Tarpey, your reduced model is wrong."

Thad: "No, it is correct."

Soda Cans continued ...

Models used in practice are conditional on available information (i.e. variables).

Soda Cans continued ...

Models used in practice are conditional on available information (i.e. variables).

The full model Volume $= \beta_0 + \beta_1 \text{height} + \epsilon$ is useless if height of the soda in the can was not measured.

Soda Cans continued ...

Models used in practice are conditional on available information (i.e. variables).

The full model Volume $= \beta_0 + \beta_1$height $+ \epsilon$ is useless if height of the soda in the can was not measured.

The reduced model $y = \beta_0 + \epsilon$ is equivalent to $y = \mu + \epsilon$

Soda Cans continued ...

Models used in practice are conditional on available information (i.e. variables).

The full model Volume $= \beta_0 + \beta_1$height $+ \epsilon$ is useless if height of the soda in the can was not measured.

The reduced model $y = \beta_0 + \epsilon$ is equivalent to $y = \mu + \epsilon$ ... which is a correct model.

Parameters – A Source of Confusion

In the soda can example, the same symbol β_0 is being used to represent two different parameters.

Question: What is a parameter?

True Model, Approximation Model

The Truth: Let $f(y; \boldsymbol{\theta})$ denote the density for the true model; let $\boldsymbol{\theta}^*$ denote the true value of $\boldsymbol{\theta}$

True Model, Approximation Model

The Truth: Let $f(y; \boldsymbol{\theta})$ denote the density for the true model; let $\boldsymbol{\theta}^*$ denote the true value of $\boldsymbol{\theta}$

An Approximation: Let $h(y; \boldsymbol{\alpha})$ denote a proposed approximation model.

True Model, Approximation Model

The Truth: Let $f(y; \boldsymbol{\theta})$ denote the density for the true model; let $\boldsymbol{\theta}^*$ denote the true value of $\boldsymbol{\theta}$

An Approximation: Let $h(y; \boldsymbol{\alpha})$ denote a proposed approximation model.

Hopefully $\hat{\boldsymbol{\alpha}}^* \rightarrow \boldsymbol{\alpha}^*$ as $n \rightarrow \infty$.

Question: But what is $\boldsymbol{\alpha}^*$?

True Model, Approximation Model

The Truth: Let $f(y; \boldsymbol{\theta})$ denote the density for the true model; let $\boldsymbol{\theta}^*$ denote the true value of $\boldsymbol{\theta}$

An Approximation: Let $h(y; \boldsymbol{\alpha})$ denote a proposed approximation model.

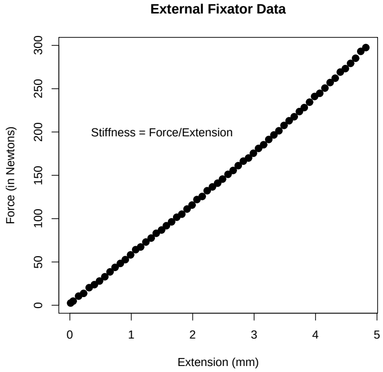
Hopefully $\hat{\boldsymbol{\alpha}}^* \to \boldsymbol{\alpha}^*$ as $n \to \infty$.

Question: But what is $\boldsymbol{\alpha}^*$?

$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}} \int f(y; \boldsymbol{\theta}^*) \log(h(y; \boldsymbol{\alpha}))dy.$$

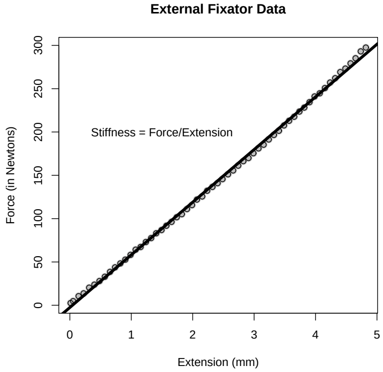
Example: External Fixator to hold a broken bone in place.

Thad: I'm going to use the slope of a straight line to estimate the stiffness of an external fixator.



Example: External Fixator to hold a broken bone in place.

Thad: I'm going to use the slope of a straight line to estimate the stiffness of an external fixator.



Regression

Fellow Statistician: "Surely, if you fit a straight line to data with a nonlinear trend, then the straight line model is wrong."

Regression

Fellow Statistician: “Surely, if you fit a straight line to data with a nonlinear trend, then the straight line model is wrong.”

Thad: “No, it is not wrong and quit calling me Shirley.”

Least Squares

Suppose

$$E[y|x] = f(x; \boldsymbol{\theta}) \quad \text{(True Model)},$$

for some unknown function f.

Least Squares

Suppose

$$E[y|x] = f(x; \boldsymbol{\theta}) \quad \text{(True Model)},$$

for some unknown function f.

Propose an approximation, $\tilde{f}(x; \boldsymbol{\alpha})$.

Least Squares

Suppose

$$E[y|x] = f(x; \boldsymbol{\theta}) \quad \text{(True Model)},$$

for some unknown function f.

Propose an approximation, $\tilde{f}(x; \boldsymbol{\alpha})$.

Then

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \int (f(x; \boldsymbol{\theta}^*) - \tilde{f}(x; \boldsymbol{\alpha}))^2 dF_x.$$

Illustration: A Straight-Line Approximation

True Model: $E[y|x] = \sum_{j=0}^{\infty} \theta_j x^j$.

Illustration: A Straight-Line Approximation

True Model: $E[y|x] = \sum_{j=0}^{\infty} \theta_j x^j$.

Extract the Linear Trend: $E[y|x] \approx \alpha_0 + \alpha_1 x$.

Illustration: A Straight-Line Approximation

True Model: $E[y|x] = \sum_{j=0}^{\infty} \theta_j x^j$.

Extract the Linear Trend: $E[y|x] \approx \alpha_0 + \alpha_1 x$.

The least-squares criterion (for $x \sim U(0,1)$) gives

$$\alpha_0 = \mu_y - \alpha_1 \mu_x, \quad \text{and} \quad \alpha_1 = \sum_{j=0}^{\infty} \frac{6j\theta_j}{(j+2)(j+1)}.$$

Model body fat percentage as a function of weight

Predict Body Fat % Using Regression
(data from Johnson 1996, *JSE*)

Model body fat percentage as a function of weight: $\hat{y} = -9.99515 + 0.162\text{Wt}.$

Body Fat % continued ...

Thad: “I just fit a line to the body fat percentage (y) versus weight data.”

Body Fat % continued ...

Thad: "I just fit a line to the body fat percentage (y) versus weight data."
Fellow Statistician: "Tarpey, your model is wrong...

Body Fat % continued ...

Thad: "I just fit a line to the body fat percentage (y) versus weight data."
Fellow Statistician: "Tarpey, your model is wrong...under-specified – there are other variables that also predict body fat percentage;

Body Fat % continued ...

Thad: "I just fit a line to the body fat percentage (y) versus weight data."
Fellow Statistician: "Tarpey, your model is wrong...under-specified – there are other variables that also predict body fat percentage; your estimated slope will be biased. You need more predictors"

Multiple Regression

In multiple regression with two predictors x_1 and x_2 correlated to each other and to y:

$$\begin{aligned}\textbf{Full Model} &: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon. \\ \textbf{Reduced Model} &: \quad y = \beta_0 + \beta_1 x_1 + \epsilon.\end{aligned}$$

We may drop x_2 for the sake of model parsimony or because x_2 does not appear significant.

Multiple Regression

In multiple regression with two predictors x_1 and x_2 correlated to each other and to y:

$$\textbf{Full Model} : \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$
$$\textbf{Reduced Model} : \quad y = \beta_0 + \beta_1 x_1 + \epsilon.$$

We may drop x_2 for the sake of model parsimony or because x_2 does not appear significant.

Question: What is wrong with what I have written here?

Full and Reduced Models

- The coefficient β_1 in the full model is never the same as β_1 in the reduced model unless $\beta_2 = 0$.

Full and Reduced Models

- The coefficient β_1 in the full model is never the same as β_1 in the reduced model unless $\beta_2 = 0$.
- β_2 in the full model equals zero if and only if

$$\text{cor}(x_1, x_2) = \frac{\text{cor}(x_2, y)}{\text{cor}(x_1, y)}.$$

Full and Reduced Models

- The coefficient β_1 in the full model is never the same as β_1 in the reduced model unless $\beta_2 = 0$.
- β_2 in the full model equals zero if and only if
$$\text{cor}(x_1, x_2) = \frac{\text{cor}(x_2, y)}{\text{cor}(x_1, y)}.$$
- Hence, β_2 cannot be zero if $\text{cor}(x_2, y) > \text{cor}(x_1, y)$.

Model Under-specification

If $\beta_2 \neq 0$, then β_1 in the reduced model is a different parameter than in the full model.

Model Under-specification

If $\beta_2 \neq 0$, then β_1 in the reduced model is a different parameter than in the full model.

- In the model under-specification literature, $\hat{\beta}_1$ in the reduced model is called biased.

Model Under-specification

If $\beta_2 \neq 0$, then β_1 in the reduced model is a different parameter than in the full model.

- In the model under-specification literature, $\hat{\beta}_1$ in the reduced model is called biased.
- According to this logic, $\hat{\beta}_1$ in a simple linear regression is always biased if there exists any other predictor more highly correlated with the response.

Back to Body Fat

Thad: Ok friend, to minimize the under-specification problem, I'll add the predictor abdomen circumference to my model:

Back to Body Fat

Thad: Ok friend, to minimize the under-specification problem, I'll add the predictor abdomen circumference to my model:

$$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$$

Back to Body Fat

Thad: Ok friend, to minimize the under-specification problem, I'll add the predictor abdomen circumference to my model:

$$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$$

Are you happy now?

Back to Body Fat

Thad: Ok friend, to minimize the under-specification problem, I'll add the predictor abdomen circumference to my model:

$$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$$

Are you happy now?

Fellow Statistician: Ah-ha! The coefficient of weight has the wrong sign now. Your model is clearly wrong. In your face Tarpey!

Back to Body Fat

Thad: Ok friend, to minimize the under-specification problem, I'll add the predictor abdomen circumference to my model:

$$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$$

Are you happy now?

Fellow Statistician: Ah-ha! The coefficient of weight has the wrong sign now. Your model is clearly wrong. In your face Tarpey!

Thad: No, the model is clearly right.

Coefficient Interpretation in Multiple Regression

The usual interpretation of a coefficient, say β_j of a predictor x_j, is that

> β_j represents the mean change in the response for a unit change in x_j *provided all other predictors are held constant.*

$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$

Estimated coefficient of weight in the full model: -0.14.

$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$

Estimated coefficient of weight in the full model: -0.14.

- Consider the population of men with some fixed abdomen circumference value.

$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$

Estimated coefficient of weight in the full model: -0.14.

- Consider the population of men with some fixed abdomen circumference value.
- What happens to body fat percentage as the weights of men in this group increase?

$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight})$.

Estimated coefficient of weight in the full model: -0.14.

- Consider the population of men with some fixed abdomen circumference value.
- What happens to body fat percentage as the weights of men in this group increase?
- Body fat % will go down ...

$\hat{y} = -41.35 + 0.92(\text{abdomen}) - 0.14(\text{weight}).$

Estimated coefficient of weight in the full model: -0.14.

- Consider the population of men with some fixed abdomen circumference value.
- What happens to body fat percentage as the weights of men in this group increase?
- Body fat % will go down ... hence the negative coefficient.

Probability Models

In life, all probability is conditional.

Probability Models

In life, all probability is conditional.

Basically, "...randomness is fundamentally incomplete information (Taleb, *Black Swan*, p 198).

Pick a Card ... any card

Thad: "Ok my statistical friend, pick a card, any card."

Pick a Card ... any card

Thad: "Ok my statistical friend, pick a card, any card."

Unbeknownst to me, my friend picked an Ace.

Pick a Card ... any card

Thad: "Ok my statistical friend, pick a card, any card."

Unbeknownst to me, my friend picked an Ace.
Question: What is $P(\text{Ace})$?

Pick a Card ... any card

Thad: "Ok my statistical friend, pick a card, any card."

Unbeknownst to me, my friend picked an Ace.
Question: What is $P(\text{Ace})$?
Answer:
Fellow Statistician: 1

Pick a Card ... any card

Thad: "Ok my statistical friend, pick a card, any card."

Unbeknownst to me, my friend picked an Ace.
Question: What is $P(\text{Ace})$?
Answer:
Fellow Statistician: 1

Thad: 4/52 (I haven't seen the card).

Confidence Intervals

A quote from Devore and Peck's book *Statistics: The Exploration and Analysis of Data* (2005, p 373) regarding the 95% confidence interval for a proportion π:

> "... it is tempting to say there is a 'probability' of .95 that π is between .499 and .561. *Do not yield to this temptation!*...Any specific interval ... either includes π or it does not...We cannot make a chance statement concerning this particular interval."

Pick a Sample ... any sample

Compute a 95% confidence interval for μ.

Pick a Sample ... any sample

Compute a 95% confidence interval for μ.

Chance selects a random sample of size n...

Pick a Sample ... any sample

Compute a 95% confidence interval for μ.

Chance selects a random sample of size n...

95% of all possible confidence intervals contain μ ... Chance has picked one of them for us...similar to picking a card from the deck.

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Answer:

For the Omniscient: 0 or 1

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Answer:

For the Omniscient: 0 or 1

For me: 0.95

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Answer:

For the Omniscient: 0 or 1

For me: 0.95

Fellow Statistician: Didn't your read your Devore and Peck book?

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Answer:

For the Omniscient: 0 or 1

For me: 0.95

Fellow Statistician: Didn't your read your Devore and Peck book?

Thad: If I don't know if μ is in the confidence interval...

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Answer:

For the Omniscient: 0 or 1

For me: 0.95

Fellow Statistician: Didn't your read your Devore and Peck book?

Thad: If I don't know if μ is in the confidence interval... from my perspective, there is uncertainty; the probability cannot be 0 or 1.

Pick a Sample ... any sample

Question: What is the probability that my interval contains μ?

Answer:

For the Omniscient: 0 or 1

For me: 0.95

Fellow Statistician: Didn't your read your Devore and Peck book?

Thad: If I don't know if μ is in the confidence interval... from my perspective, there is uncertainty; the probability cannot be 0 or 1. The correct probability model is conditional.

Conclusions

Calling a model right or wrong is just a matter of perspective.

Conclusions

Calling a model right or wrong is just a matter of perspective.

With enough data, any imperfection in a model can be detected.

Conclusions

Calling a model right or wrong is just a matter of perspective.

With enough data, any imperfection in a model can be detected.

The temptation then is to say all models are wrong.

Conclusions

Calling a model right or wrong is just a matter of perspective.

With enough data, any imperfection in a model can be detected.

The temptation then is to say all models are wrong.

However, if we regard models as approximations to the truth, we could just as easily call all models right.

Conclusions

In any given data analysis situation, a multitude of models can be proposed.

Conclusions

In any given data analysis situation, a multitude of models can be proposed.

Most of these will be useless ...

Conclusions

In any given data analysis situation, a multitude of models can be proposed.

Most of these will be useless ...

and perhaps a few will be useful.

Conclusions

Models have served us very well ...

Conclusions

Models have served us very well ...

and also, at times, quite poorly.

Some quotes: Breiman 2001 *Statistical Science*

"...as data becomes more complex, the data models become more cumbersome and are losing the advantage of presenting a simple and clear picture of nature's mechanism (p 204)...

Unfortunately, our field has a vested interest in data models, come hell or high water (p 214)."

Some quotes: Taleb, *Black Swan*

"...the gains in our ability to model (and predict) the world may be dwarfed the increases in its complexity (p 136)"

A Final Quote:

Or, as Peter Norvig, Google's research director, says

> *Let's stop expecting to find a simple theory, and instead embrace complexity, and use as much data as well as we can to help define (or estimate) the complex models we need for these complex domains.*

* From http://norvig.com/fact-check.html. Note, Norvig was misquoted using a variation of the Box quote in: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" by Chris Anderson in Wired Magazine, 2008.