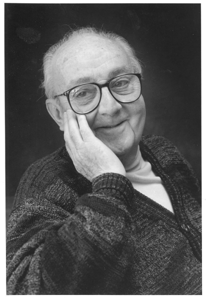
"all models are wrong, but some are useful." – George E.P. Box



1. Study the effects of an independent variable on a dependent variable

1. Study the effects of an independent variable on a dependent variable
    - Does an independent variable have any effect on a dependent variable?

1. Study the effects of an independent variable on a dependent variable
   - Does an independent variable have any effect on a dependent variable? (p-value)

1. Study the effects of an independent variable on a dependent variable
   - Does an independent variable have any effect on a dependent variable? (p-value)
   - How strong is the effect?

1. Study the effects of an independent variable on a dependent variable
    - Does an independent variable have any effect on a dependent variable? (p-value)
    - How strong is the effect? (effect Size)

1. Study the effects of an independent variable on a dependent variable
   - Does an independent variable have any effect on a dependent variable? (p-value)
   - How strong is the effect? (effect Size)
   - Both the p-value and the effect size describe a population effect

1. Study the effects of an independent variable on a dependent variable
   - Does an independent variable have any effect on a dependent variable? (p-value)
   - How strong is the effect? (effect Size)
   - Both the p-value and the effect size describe a population effect
   - Require correctly specified relationship.

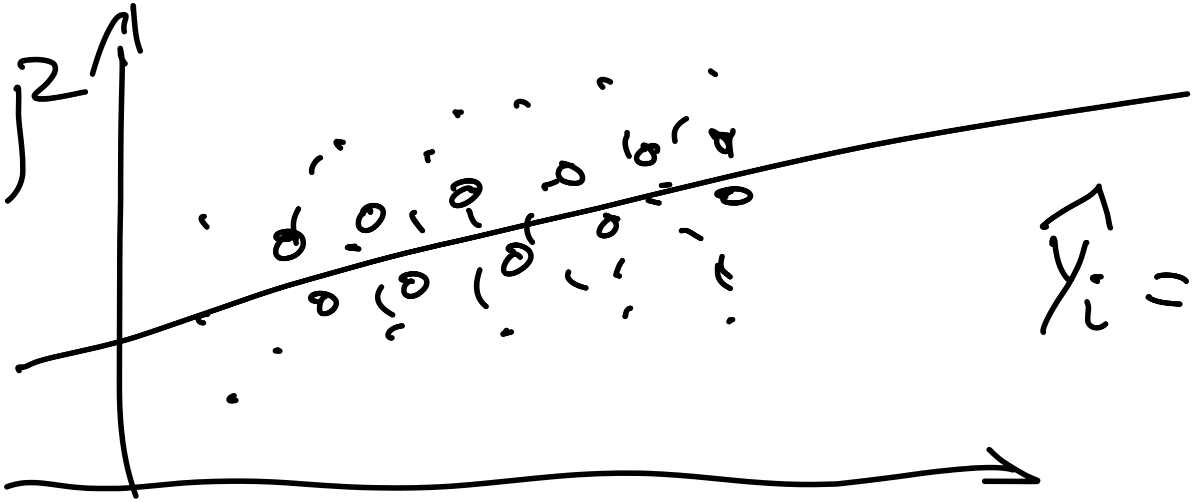
1. Study the effects of an independent variable on a dependent variable
   - Does an independent variable have any effect on a dependent variable? (p-value)
   - How strong is the effect? (effect Size)
   - Both the p-value and the effect size describe a population effect
   - Require correctly specified relationship.
2. Individual outcome prediction

How useful is a model for prediction?

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= SSR + SSE$$

$$= \sum (\hat{y}_i - \bar{y})^2 + \sum_i (\hat{y}_i - y_i)^2$$



$$\hat{y}_i = x_i \hat{\beta}$$

$$y = X\beta + \varepsilon \sim N(0, \sigma^2)$$

How useful is a model for prediction?

1. If a statistical model is correctly specified, does it imply accurate individual prediction?

How useful is a model for prediction?

1. If a statistical model is correctly specified, does it imply accurate individual prediction?
2. Does a small p-value or/and a large effect size always imply accurate individual prediction?

How useful is a model for prediction?

1. If a statistical model is correctly specified, does it imply accurate individual prediction?
2. Does a small p-value or/and a large effect size always imply accurate individual prediction?
3. There is a need to develop prediction accuracy (PA) measures

✓ Discrimination - discriminated between "cases" vs "non-case"

_ Calibration - ability to produce the correct absolute probability.

X

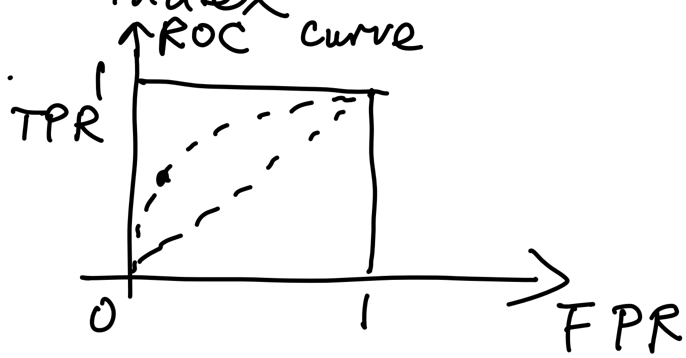
Marker.

risk score $\beta^T Z$ — g

Binary outcome

C - statistics index

$= \underline{\underline{AUC}}$



$= 0.5$ ✗

$= P(X_1 > X_2)$

Time-to-event outcome X.

covariate Z

Cox's model : $h(t|z) = h_0(t) e^{\beta^T z}$

$g(z) = \underline{\underline{\beta^T z}}$

$(X_1, Z_1) \perp\!\!\!\perp (X_2, Z_2)$ iid.

$C = P(g(Z_1) > g(Z_2) \mid X_1 < X_2)$ $\boxed{\text{Concordance measure}}$

Harrel's C-index. (1982).

$$C = \frac{\sum_i \sum_j \delta_i \, I(g(z_i) > g(z_j)) \, I(T_i < T_j)}{\sum_i \sum_j \delta_i \, I(T_i < T_j)}$$

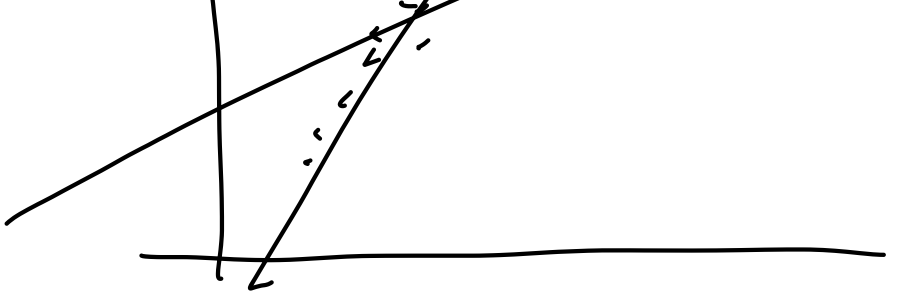
Right-censored data

$$(T_i, \delta_i, z_i)$$

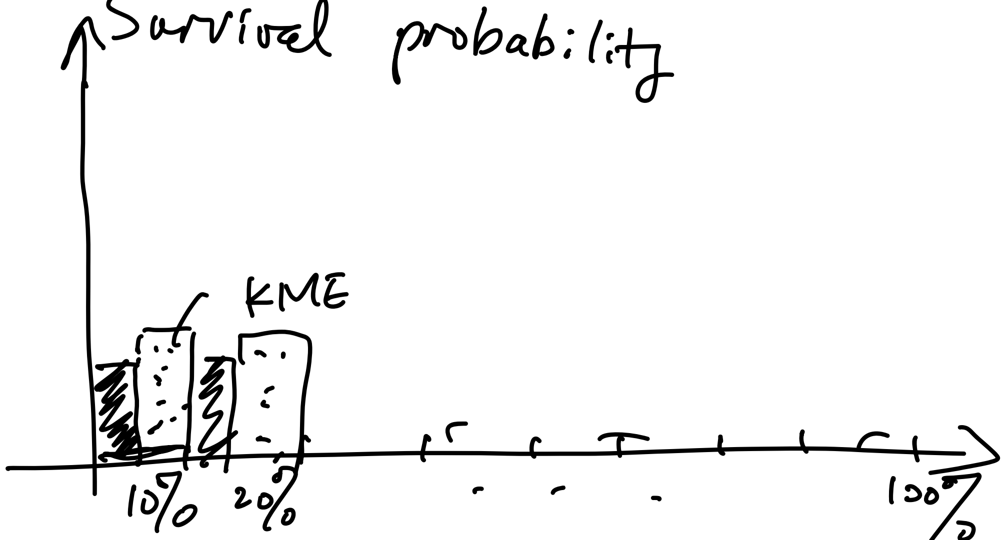
$$T_i = X_i \wedge C_i, \quad \delta_i = I(X_i \le C_i)$$

<u>Uno et al (2011)</u>

$$\hat{C} = \frac{\sum_i \sum_j \delta_i [\hat{G}(T_i)]^{-2} I(g(z_i) > g(z_j)) \, I(T_i < T_j)}{\sum_i \sum_j \delta_i [\hat{G}(T_i)]^{-2} I(T_i < T_j)}$$



Calibration measure : Ability to predict the absolute risk (probability).



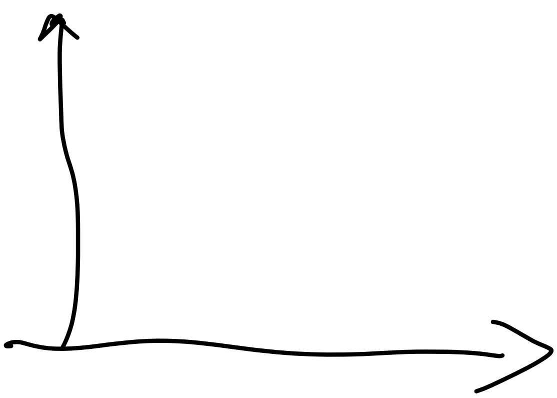
Rank all subjects

$\beta^T Z_i$

$$\hat{S}(10 | Z_i) = \left[\hat{S}_0(10)\right]^{\exp(\hat{\beta}^T Z_i)}$$

$$X^2 = \sum_{i=1}^{10} \frac{(\hat{p}_i - \bar{p}_i)^2}{\bar{p}_i(1-\bar{p}_i)} \sim \chi^2_9$$

R^2 — proportion of explained variance ?



Prediction Accuracy Measures for a Nonlinear Model and for Right-Censored Time-to-Event Data

Gang Li, Ph.D.

Professor of Biostatistics and Computational Medicine
Director, UCLA's Jonsson Comprehensive Cancer Center BASE Shared Resource
University of California at Los Angeles

Acknowledgement

Xiaoyan Wang, Ph.D., UCLA Department of Med-GIM and HSR,

[1]. Li, Gang, and Xiaoyan Wang. "Prediction accuracy measures for a nonlinear model and for right-censored time-to-event data." *Journal of the American Statistical Association* 114.528 (2019): 1815-1825.

[2]. Wang X, Li G (2018). PAmeasures: Prediction and Accuracy Measures for Nonlinear Models and for Right-Censored Time-to-Event Data. R package version 0.1.0, https://CRAN.R-project.org/package=PAmeasures.

Example 1. PBC Data

1. The primary biliary cirrhosis (PBC) data consists of 312 patients from a randomized Mayo Clinic trial in primary biliary cirrhosis of the liver conducted between 1974 and 1984 (http://astrostatistics .psu.edu/datasets/R/html/survival/html/pbc.html).
2. Consider four models (Cox PH, Weibull AFT, log-normal AFT, and threshold regression) for predicting overall survival of individual PBC patients, using the five covariates (patient's age, log(serum bilirubin concentration), log(serum albumin concentration), log(standardised blood clotting time), and presence of peripheral edema and antidiuretic therapy) employed in the well known Mayo risk score (MRS).

Example 1. Cox-Snell Residual Plots for Lack of Fit

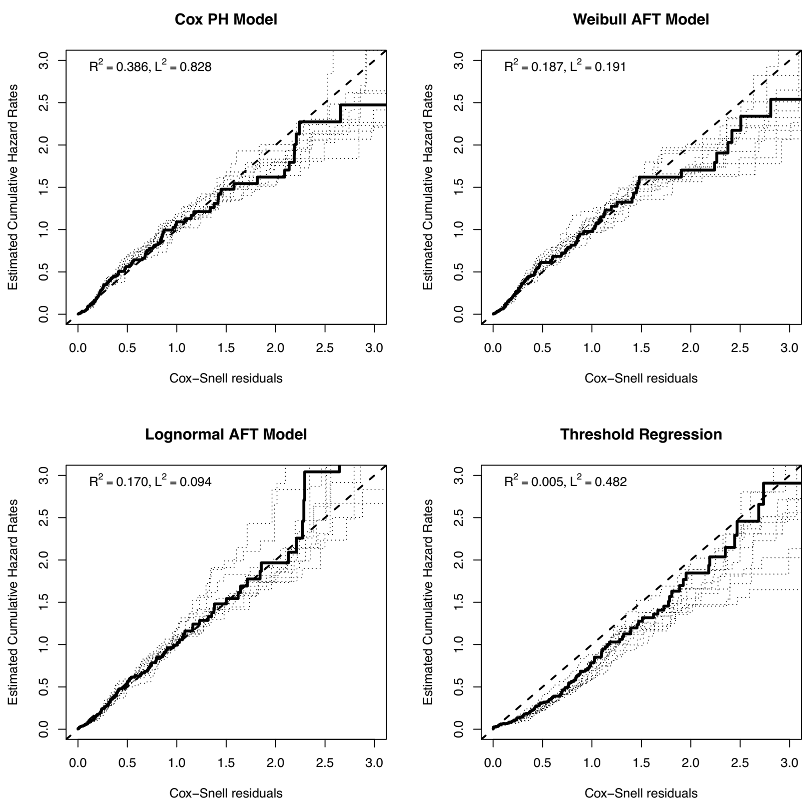
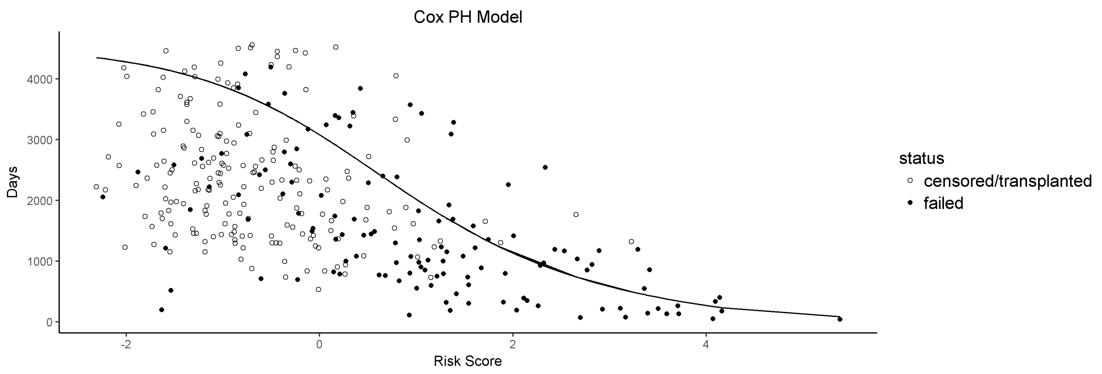
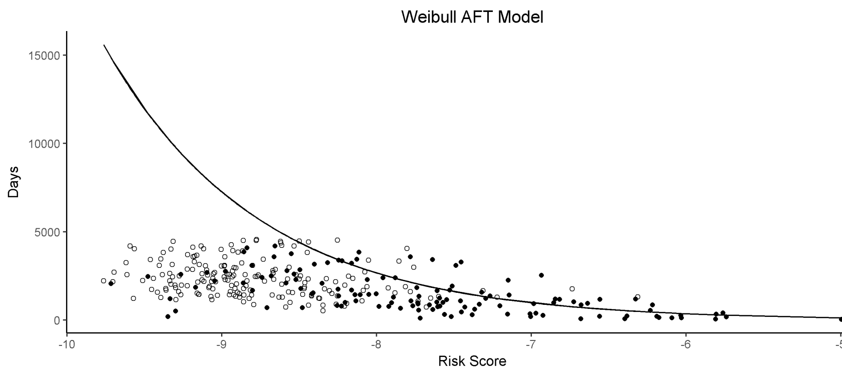
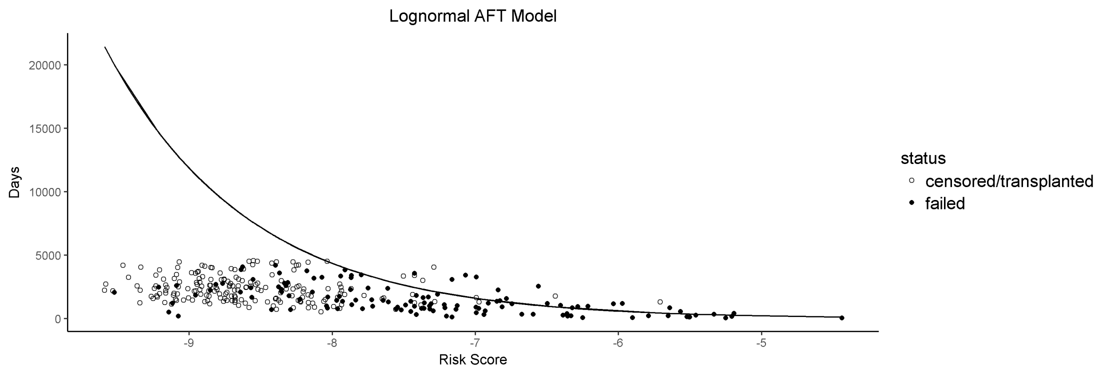


Figure: (PBC Data) Cox-Snell residual plots for the Cox model, Weibull AFT model, log-normal AFT model, and threshold regression model. For each model, the solid line is based on the observed PBC data and the dotted lines are based on ten bootstrap samples. Deviations from the 45° line indicate possible lack-of-fit.

Example 1. Prediction Performance







Example 2: UCLA Ovarian Cancer Data

- Platinum-resistant: showed progression while on first-line platinum-based regimen
- Clinical variables: stage, grade, histology
- Biomarkers: pre-operative serum CA125 level, NY-ESO1 expression level from tissue microarray (novel)
- Primary question: overall survival prognosis based on the biomarkers after adjusting for clinical variables
- 37 ovarian cancer patients; Censoring rate $= 24\%$

Example 2: Is NYESO-1 a prognostic factor for ovarian cancer?

$$\lambda(t|Z) = \lambda_0(t)e^{\beta^T Z}$$

Table 1. Cox's Regression analysis of overall survival for platinum resistance ovarian cancer patients.

Variables	Level	Hazard ratio	P-value
Stage	3&4 vs 1&2	4.45	0.10
grade	1&2 vs 3	1.07	0.89
Histology	Serous vs. clear cell	0.29	0.09
	Endometrioid vs clear cell	0.95	0.94
Preop_CA125	>500 vs.<=500	3.92	**0.01**
NY_ESO1	>12 vs. <=12	3.12	**0.04**

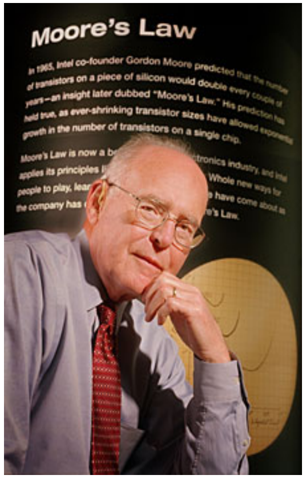
Example 3: Moore's Law (Linear Model)

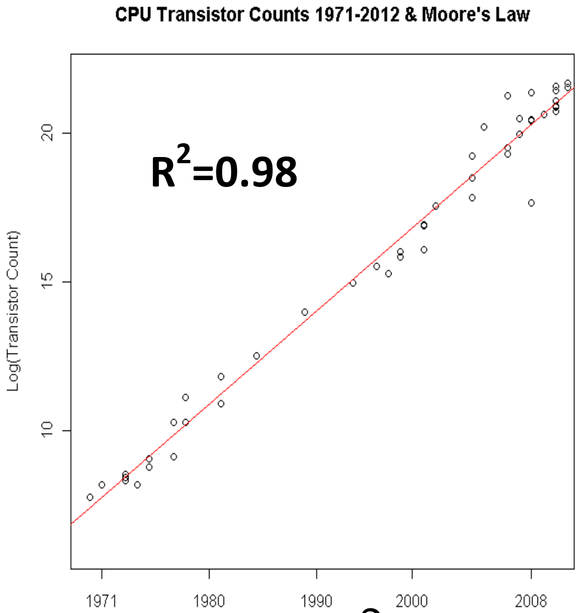
Over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every two years (Gordon E. Moore, 1965)

Example 3: Moore's Law (Linear Model)

Over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every two years (Gordon E. Moore, 1965)

$Y = \log_2(\text{number of transistors})$

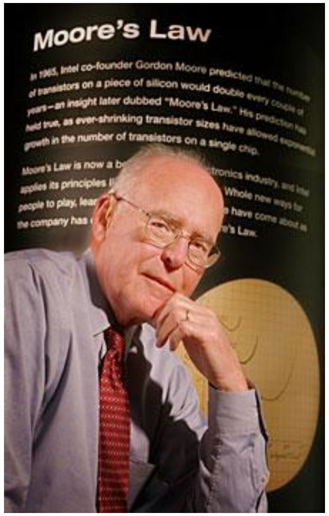


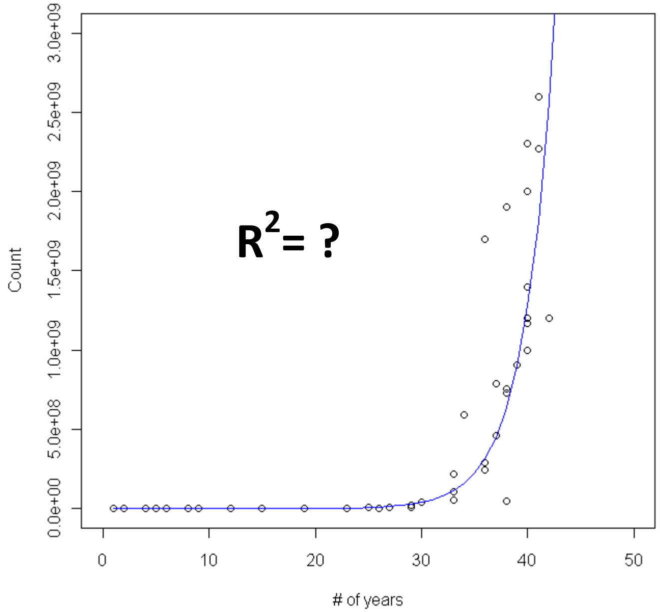


- Proportion of Explained Variation: $R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$;
- $\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$.

Example 3: Moore's Law (Nonlinear Model)

Y = number of transistors





Challenges:

- $R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \neq$ proportion of explained variation
- $\sum(y_i - \bar{y})^2 \neq \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$

Event Time Model

- Cox's model: $\lambda(t|Z) = \lambda_0(t)e^{\beta^T Z}$, AFT model, transformation models, etc
- Nonlinear
- Right censoring

Nonlinear Model - Existing Pseudo R^2

- At least 12 R^2-measures for logistic regression (Mittlbock and Schemper, 1996)
- Likelihood-based measures (Goodman, 1971; McFadden et al., 1973; Maddala, 1986; Cox and Snell, 1989; Magee, 1990; Nagelkerke, 1991),
- Information-based measures (McFadden et al., 1973; Kent,1983),
- Ranking-based measures (Harrell et al., 1982),
- Variation-based measures (Theil, 1970; Efron, 1978; Haberman, 1982; Hilden, 1991; Cox and Wermuth, 1992; Ash and Shwartz, 1999),
- Multiple correlation coefficient measure (Mittlbock et al., 1996; Zheng and Agresti, 2000).

Nonlinear Models - Existing Pseudo R^2

Challenges:

- Difficult to interpret
- None have received the same widespread acceptance as the classical R^2 for linear regression

Event Time Models: Prediction Accuracy Measures

1. Harrel's C (Harrell et al., 1982)
2. ROC curves (Heagerty and Zheng, 2005; Uno et al., 2007).
3. Positive (Negative) Predictive functions (Moskowitz and Pepe, 2004; Zheng et al., 2008; Uno et al., 2007; Chen, Lin, and Zeng, 2012)
4. Explained variation - Pseudo R^2 measures and other loss functions such as Brier score (Korn and Simon, 1990; Schemper and Stare, 1996; Rosthoj and Keiding, 2004; Graf et al., 1999; Schemper and Henderson, 2000; Stare, Perme, and Henderson, 2010)

Schemper and Henderson (2000)

EV option in SAS PHREG

1. Only for Cox's model
2. Correctly specified model
3. Lack of a clear interpretation

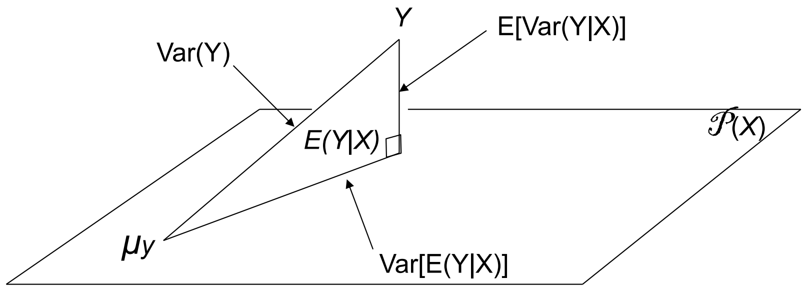
Stare, Perme, and Henderson (SPH) (2010)

Explained rank discrepancy (between the null and the perfect model)

1. Pros: Rank based
2. Cons: Rank based

Nonparametric R^2 Measure

$$var(Y) = var(E(Y|X)) + E(var(Y|X)).$$



$$\rho^2_{NP} = \frac{var(E(Y|X))}{var(Y)} = \text{proportion of explained variance.}$$

Prediction Accuracy - Cox's (1972) PH Model

Simulated Population R^2 Values for Correctly-Specified Cox's Models:

Scenario	β (log(HR))	ν	ρ^2_{NP}	R^2_{SPH}	R^2_{SH}
1	0.1	0.5	0.07	0.23	0.10
2	0.1	1	0.14	0.23	0.10
3	0.1	10	0.15	0.23	0.10
4	0.2	0.5	0.09	0.38	0.28
5	0.2	1	0.27	0.38	0.28
6	0.2	10	0.40	0.38	0.28
7	0.5	0.5	0.09	0.49	0.49
8	0.5	1	0.33	0.49	0.49
9	0.5	10	0.80	0.49	0.49

Lessons Learned

The SPH and SH Pseudo-R^2 measures have serious limitations:

1. May erroneously suggest a difference in prediction performance where there is no difference
2. May suggest no difference when there is actually a huge difference in prediction performance

Prediction Accuracy Measures for Nonlinear Models and for Right-Censored Data

Prediction Accuracy Measures for Nonlinear Models and for Right-Censored Data

1. Population PA measures

Prediction Accuracy Measures for Nonlinear Models and for Right-Censored Data

1. Population PA measures
2. Sample PA measures for uncensored data

Prediction Accuracy Measures for Nonlinear Models and for Right-Censored Data

1. Population PA measures
2. Sample PA measures for uncensored data
3. Sample PAmeasures for right-censored data

Population PA Measures

Notations and Assumptions

- Let $F(y|x)$ and $\mu(x)$ be the true conditional distribution and conditional expectation of Y given $X = x$.
- Consider a regression model of Y on X described by

$$\mathcal{M} = \{F_\theta(y|x) : \theta \in \Theta\}$$

where

  - θ -finite or infinite dimensional
  - $F_\theta(y|x)$ - conditional distribution function
  - $\mathcal{M}$ is allowed to be mis-specified

- For any $\theta \in \Theta$, let $m_\theta(X)$ be a prediction function of Y obtained as a functional of $F_\theta(\cdot|X)$.

Notations and Assumptions

Assume that $\hat{\theta}$ is a sample statistic such that as $n \to \infty$,

$$\hat{\theta} \xrightarrow{P} \theta^*, \quad \text{for some } \theta^* \in \Theta. \tag{1}$$

Q1: "How good is $m_{\hat{\theta}}(X)$ as a prediction function of Y?"
Q2: "How good is $m_{\theta^*}(X)$ as a prediction function of Y?"

Linearly Corrected Prediction

Definition

The linearly corrected prediction function of $m_{\theta^*}(X)$ is defined as

$$m_{\theta^*}^{(c)}(X) = \mu_Y + \frac{cov(Y, m_{\theta^*}(X))}{var(m_{\theta^*}(X))}[m_{\theta^*}(X) - E\{m_{\theta^*}(X)\}]. \quad (2)$$

Properties:

(i) $m_{\theta^*}^{(c)}(X) = \tilde{a} + \tilde{b} m_{\theta^*}(X)$, where $(\tilde{a}, \tilde{b}) = \arg\min_{\alpha,\beta} E\{Y - (\alpha + \beta m_{\theta^*}(X))\}^2$;

(ii) $E(m_{\theta^*}^{(c)}(X)) = \mu_Y$;

(iii) $MSPE(m_{\theta^*}^{(c)}(X)) \leq MSPE(\mu_Y)$;

(iv) $MSPE(m_{\theta^*}^{(c)}(X)) \leq MSPE(m_{\theta^*}(X))$.

Explained Variance

Variance Decomposition

$$\begin{aligned} var(Y) &= E\{m_{\theta*}^{(c)}(X) - \mu_Y\}^2 + E\{Y - m_{\theta*}^{(c)}(X)\}^2, \qquad (3) \\ &= \text{explained variance} + \text{unexplained variance} \end{aligned}$$

Definition

Define

$$\rho^2_{m_{\theta*}} = \frac{E\{m_{\theta*}^{(c)}(X) - \mu_Y\}^2}{var(Y)}, \qquad (4)$$

to be the proportion of the variance of Y that is explained by $m_{\theta*}^{(c)}(X)$

Explained Prediction Error

Prediction Error Decomposition:

$$\begin{aligned} MSPE(m_{\theta^*}(X)) &= E\{Y - m^{(c)}_{\theta^*}(X)\}^2 + E\{m^{(c)}_{\theta^*}(X) - m_{\theta^*}(X)\}^2 \\ &= \text{explained prediction error} + \text{ unexplained prediction} \end{aligned}$$

Definition

Define

$$\lambda^2_{m_{\theta^*}} = \frac{MSPE(m^{(c)}_{\theta^*}(X))\}^2}{MSPE(m_{\theta^*}(X))} = 1 - \frac{E\{m^{(c)}_{\theta^*}(X) - m_{\theta^*}(X)\}^2}{MSPE(m_{\theta^*}(X))}.$$

to be the proportion of the $MSPE$ of $m_{\theta^*}(X)$ that is explained by $m^{(c)}_{\theta^*}(X)$.

Geometric Interpretation

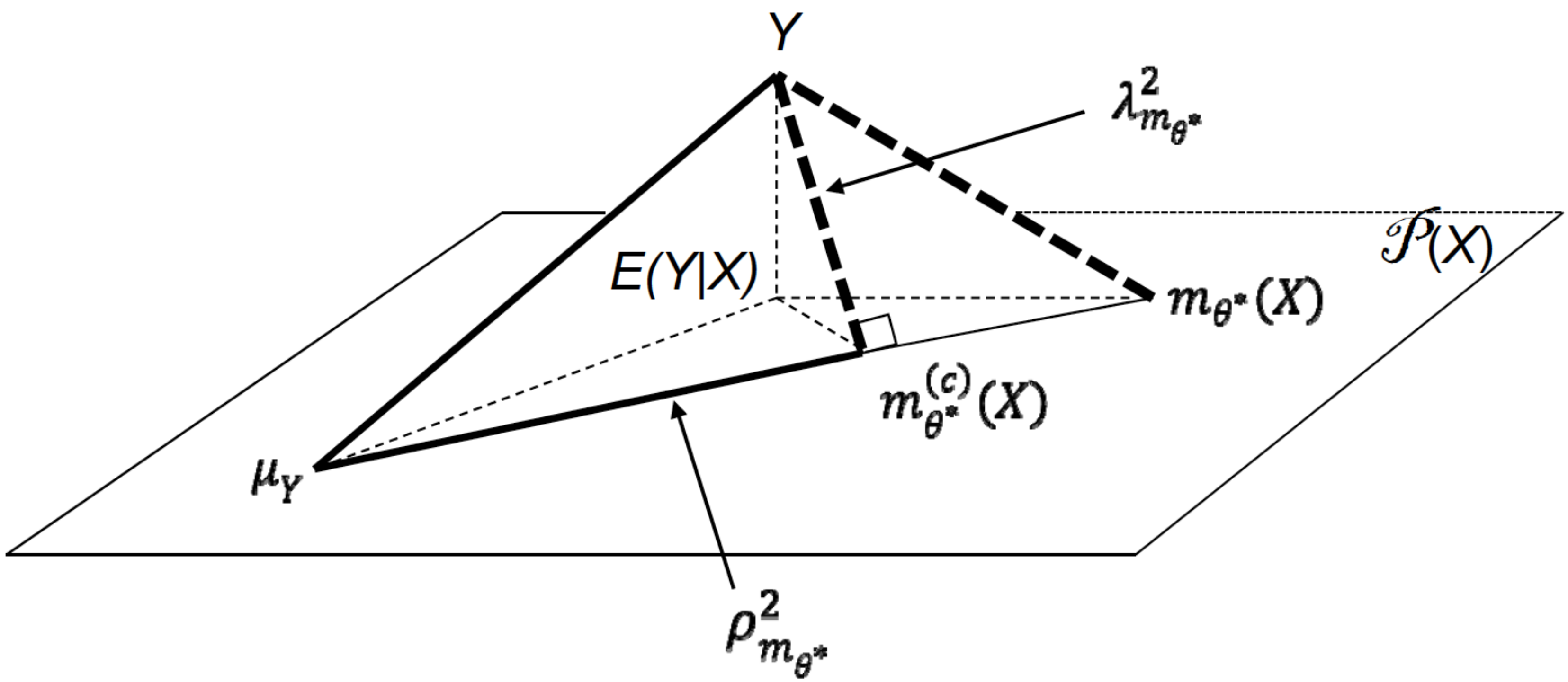


Figure: Geometric interpretation of $\rho^2_{m_{\theta*}}$ and $\lambda^2_{m_{\theta*}}$

Basic Properties

Theorem

(a) *Let $\rho(\xi, \eta)$ denote the correlation coefficient between two random variables ξ and η. Then, $\rho^2_{m_{\theta*}} = [\rho(Y, m_{\theta^*}(X))]^2$;*

(b) *(Linear Prediction). Let $BLUE(X) = a + b^T X$ be the best linear unbiased estimator (BLUE) of Y, where $(a, b) = \arg\min_{\alpha,\beta} E\{Y - (\alpha + \beta^T X)\}^2$. Then (i) $BLUE^{(c)}(X) = BLUE(X)$; (ii) $\lambda^2_{BLUE} \equiv 1$; (iii) ρ^2_{BLUE} is equal to the population value of the classical coefficient of determination for linear regression.*

(c) *If $m_{\theta^*}(X) = E(Y|X)$, then $\lambda^2_{m_{\theta*}} \equiv 1$, and $\rho^2_{m_{\theta*}} = \rho^2_{NP}$;*

(d) *(Maximal ρ^2) Let $\mathcal{P}(X)$ be the space of all p-variate functions $Q(X)$ of X. Then $\rho^2_{NP} = \max_{Q \in \mathcal{P}(X)} \{\rho^2_Q\}$.*

Practical Considerations

- $\rho^2_{m_{\theta^*}}$ should be used as the primary measure for the potential predictive power of $m_{\theta^*}(X)$;
- $\lambda^2_{m_{\theta^*}}$ should be used as a supplementary measure to a) indicate (by a value less than 1) if a linear correction is required for $m_{\theta^*}(X)$ to achieve its potential predictive power and b) quantify how much prediction error reduction can be materialized with the correction.

Sample PA Measures

Sample Measures for $m_{\hat\theta}(X)$

Lemma

(a) *(Sample Variance Decomposition)*

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(m_{\hat\theta}^{(c)}(X_i) - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - m_{\hat\theta}^{(c)}(X_i))^2;$$

(b) *(Sample Prediction Error Decomposition)*

$$\sum_{i=1}^{n}(Y_i - m_{\hat\theta}(X_i))^2 = \sum_{i=1}^{n}(Y_i - m_{\hat\theta}^{(c)}(X_i))^2 + \sum_{i=1}^{n}(m_{\hat\theta}^{(c)}(X_i) - m_{\hat\theta}(X_i))^2,$$

where $m_{\hat\theta}^{(c)}(x)$ is obtained by regressing Y_i on $m_{\hat\theta}(X_i)$.

Sample Measures for $m_{\hat{\theta}}(X)$

Definition

The sample versions of ρ^2 and λ^2 are defined by

$$R^2_{m_{\hat{\theta}}} = \frac{\sum_{i=1}^n (m^{(c)}_{\hat{\theta}}(X_i) - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2},$$

and

$$L^2_{m_{\hat{\theta}}} = \frac{\sum_{i=1}^n (Y_i - m^{(c)}_{\hat{\theta}}(X_i))^2}{\sum_{i=1}^n (Y_i - m_{\hat{\theta}}(X_i))^2},$$

Large Sample Properties

Theorem

(a) *(Consistency).*

$$R^2_{m_{\hat\theta}} \xrightarrow{P} \rho^2_{m_{\theta *}}, \quad \text{and} \quad L^2_{m_{\hat\theta}} \xrightarrow{P} \lambda^2_{m_{\theta *}}.$$

(b) *(Asymptotic normality).*

$$\sqrt{n}(R^2_{m_{\hat\theta}} - \rho^2_{m_{\theta *}}) \xrightarrow{d} N(0, \sigma^2_\rho), \text{ and } \sqrt{n}(L^2_{m_{\hat\theta}} - \lambda^2_{m_{\theta *}}) \xrightarrow{d} N(0, \sigma^2_\lambda),$$

where σ^2_ρ *and* σ^2_λ *are the asymptotic variances.*

Sample PA Measures fort Right-Censored Data

Right-Censored Event Time Data

Let $T = \min\{Y, C\}$ and $\delta = I(Y \leq C)$, where C is an censoring random variable that is assumed to be independent of Y given X.

Right censored sample: $(T_1, \delta_1, X_1), \ldots, (T_n, \delta_n, X_n)$

Let $\hat{\theta} = \hat{\theta}(T_1, \delta_1, X_1, \ldots, T_n, \delta_n, X_n)$.

Right-Censored Sample Decompositions

Lemma

(a) *(Weighted Variance Decomposition for T)*

$$\sum_{i=1}^{n} w_i\{T_i - \bar{T}^{(w)}\}^2 = \sum_{i=1}^{n} w_i\{m_{\hat{\theta}}^{(wc)}(X_i) - \bar{T}^{(w)}\}^2 + \sum_{i=1}^{n} w_i\{T_i - m_{\hat{\theta}}^{(wc)}(X_i)\}^2;$$

(b) *(Weighted Prediction Error Decomposition for T)*

$$\sum_{i=1}^{n} w_i\{T_i - m_{\hat{\theta}}(X_i)\}^2 = \sum_{i=1}^{n} w_i\{T_i - m_{\hat{\theta}}^{(wc)}(X_i)\}^2 + \sum_{i=1}^{n} w_i\{m_{\hat{\theta}}^{(wc)}(X_i) - m_{\hat{\theta}}(X_i)\}^2.$$

Right-Censored Sample Decompositions

Lemma

Let

$$w_i = \frac{\frac{\delta_i}{\hat{G}(T_i-)}}{\sum_{j=1}^{n} \frac{\delta_j}{\hat{G}(T_j-)}}, \quad (5$$

Then

$$\sum_{i=1}^{n} w_i \{T_i - \bar{T}^{(w)}\}^2 \xrightarrow{P} var(Y);$$

$$\sum_{i=1}^{n} w_i \{m_{\hat{\theta}}^{(wc)}(X_i) - \bar{T}^{(w)}\}^2 \xrightarrow{P} E\{m_{\theta*}^{(c)}(X) - \mu_Y\}^2;$$

$$\sum_{i=1}^{n} w_i \{T_i - m_{\hat{\theta}}^{(wc)}(X_i)\}^2 \xrightarrow{P} E\{Y - m_{\theta*}^{(c)}(X)\}^2;$$

$$\sum_{i=1}^{n} w_i \{T_i - m_{\hat{\theta}}(X_i)\}^2 \xrightarrow{P} E\{Y - m_{\theta*}(X)\}^2;$$

$$\sum_{i=1}^{n} w_i \{m_{\hat{\theta}}^{(wc)}(X_i) - m_{\hat{\theta}}(X_i)\}^2 \xrightarrow{P} E\{m_{\theta*}^{(c)}(X) - m_{\theta*}(X)\}^2.$$

Right-censored sample measures

Definition

The right censored sample version of ρ^2 and λ^2 are defined by

$$R^2_{m_{\hat{\theta}}} = \frac{\sum_{i=1}^n w_i \{m_{\hat{\theta}}^{(wc)}(X_i) - \bar{T}^{(w)}\}^2}{\sum_{i=1}^n w_i \{T_i - \bar{T}^{(w)}\}^2},$$

and

$$L^2_{m_{\hat{\theta}}} = \frac{\sum_{i=1}^n w_i \{T_i - m_{\hat{\theta}}^{(wc)}(X_i)\}^2}{\sum_{i=1}^n w_i \{T_i - m_{\hat{\theta}}(X_i)\}^2},$$

Properties

Theorem

(a) *(Uncensored Data). If there is no censoring, then the censored sample measures reduce to the uncensored sample definitions.*

(a) *(Uncensored Linear Model). Under the linear model with no censoring, $L^2 \equiv 1$ and $R^2 =$ Coefficient of Determination.*

(b) *(Consistency)*

$$R^2_{m_{\hat{\theta}}} \xrightarrow{P} \rho^2_{m_{\theta*}}, \quad \text{and} \quad L^2_{m_{\hat{\theta}}} \xrightarrow{P} \lambda^2_{m_{\theta*}}.$$

(c) *(Asymptotic normality).*

$$\sqrt{n}(R^2_{m_{\hat{\theta}}} - \rho^2_{m_{\theta*}}) \xrightarrow{d} N(0, v_\rho^2), \text{ and } \sqrt{n}(L^2_{m_{\hat{\theta}}} - \lambda^2_{m_{\theta*}}) \xrightarrow{d} N(0, v_\lambda^2),$$

Numerical Illustrations

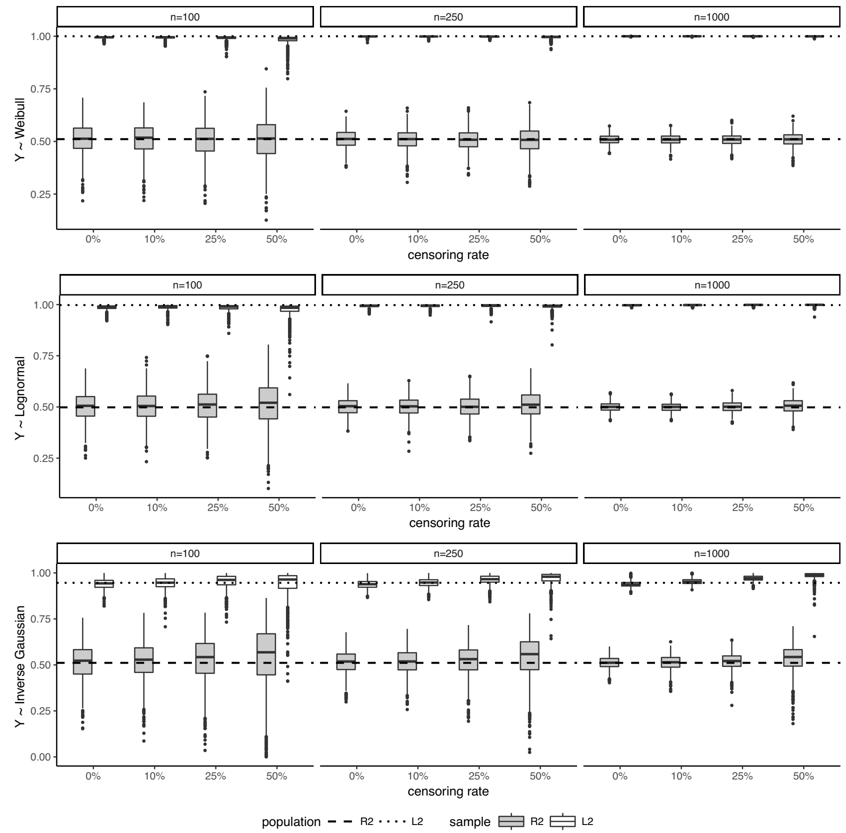


Figure: (Independent Censoring) Box plots of simulated R^2 (shaded box) and L^2 (unshaded box) for the Cox Model by censoring rate (0%, 10%, 25%, 50%), sample size (100, 250, 1,000), and data generation setting (upper panel: Weibull; middle panel: log-normal; bottom panel: inverse Gaussian)

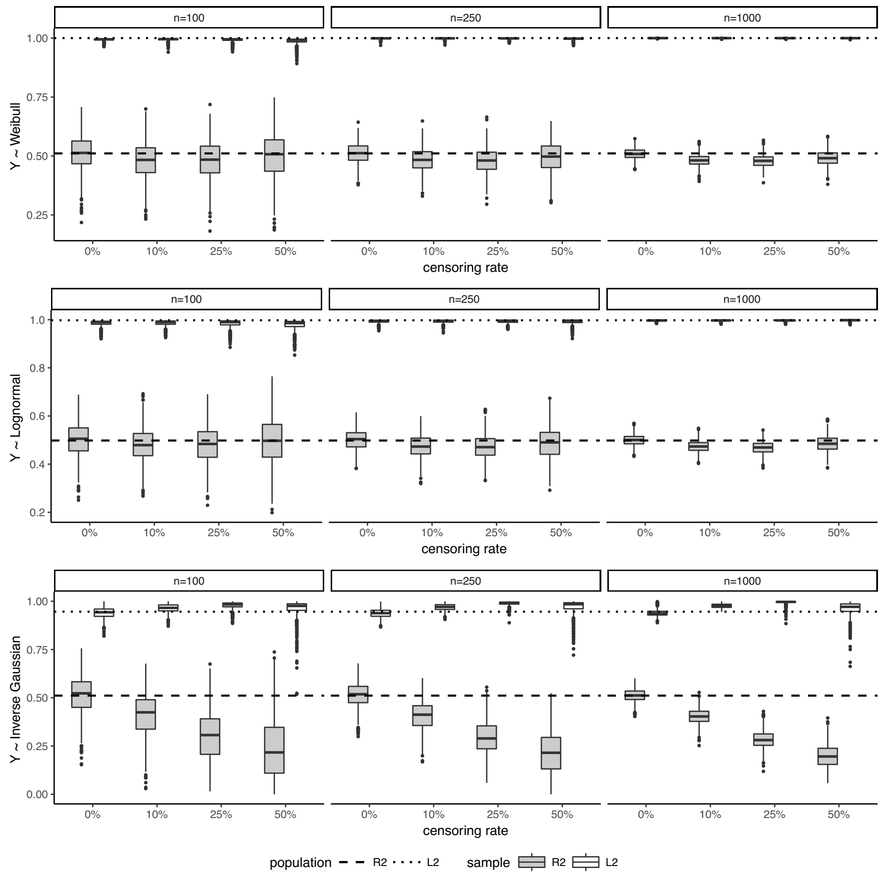
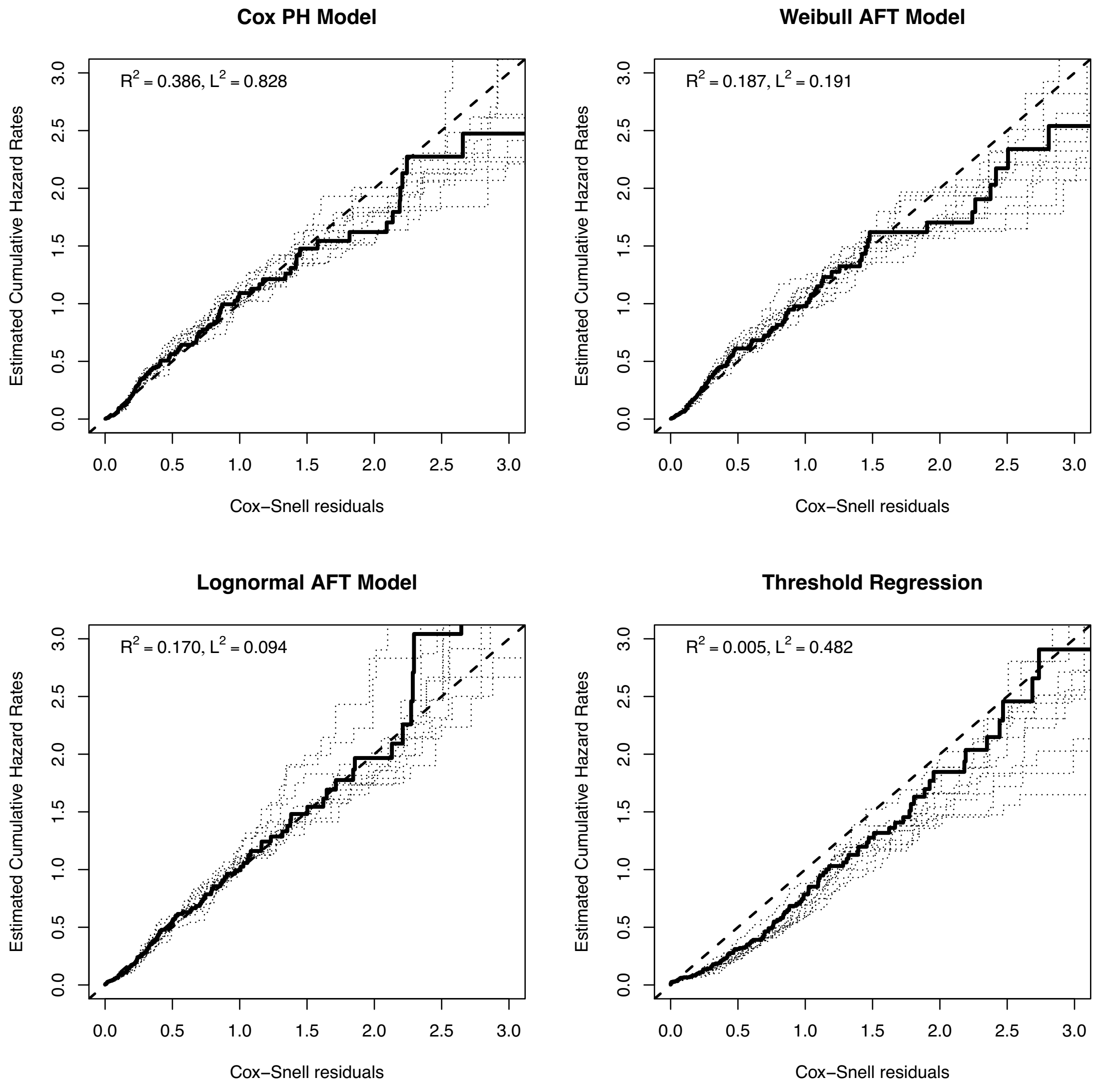


Figure: (Dependent Censoring) Box plots of simulated R^2 (shaded box) and L^2 (unshaded box) for the Cox model by censoring rate (0%, 10%, 25%, 50%), sample size (100, 200, 1,000), and data generation setting (upper panel: Weibull; middle panel: log-normal AFT; bottom panel: inverse Gaussian)

Example 1 PBC Data

PBC Data The primary biliary cirrhosis (PBC) data consists of 312 patients from a randomized Mayo Clinic trial in primary biliary cirrhosis of the liver conducted between 1974 and 1984 (http://astrostatistics .psu.edu/datasets/R/html/survival/html/pbc.html).

Example 1 Model Fit



Example 1: Prediction Performance

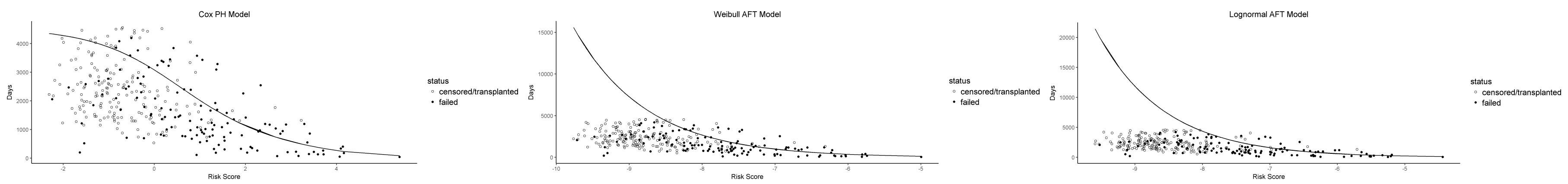


Figure: (PBC Data) Predicted (solid line) and observed (solid dot: uncensored; censored: circle) survival times (in days) versus risk score for the Cox model (left), Weibull AFT model (middle), and log-normal AFT model (right).

Example 1: PA Measures (R^2 and L^2)

Table: (PBC Data) R^2 values of different survival regression models.

Model	Cox PH	Weibull AFT	Log-normal AFT
R^2	0.39	0.19	0.17
L^2	0.83	0.19	0.09
C-index	?	!	?

Example 1: R^2 Differences

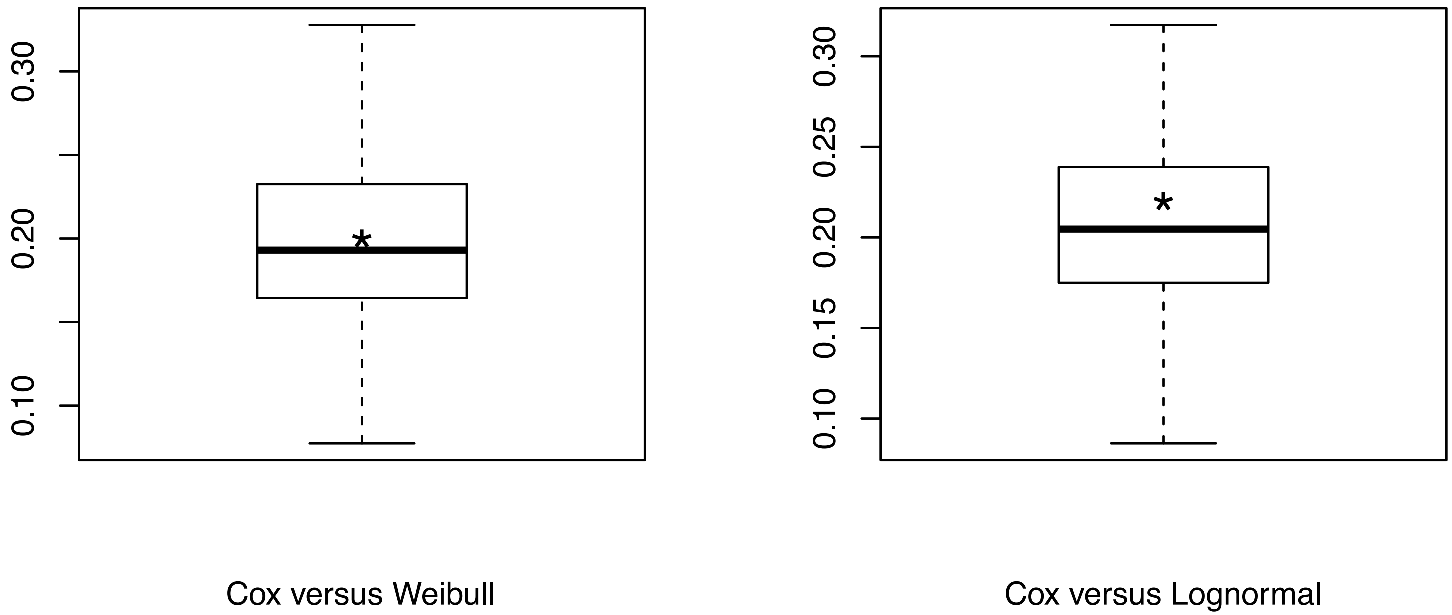


Figure: (PBC Data) Box plots of the R^2 differences between different models based on 100 bootstrap samples from the PBC data (Left: Cox's model versus Log-normal AFT model; Right: Cox's model versus Weibull AFT model). The asterisk in each box plot represents the R^2 difference between the two models based on the observed PBC data.

Example 2: UCLA Ovarian Cancer Data

- Platinum-resistant: showed progression while on first-line platinum-based regimen
- Clinical variables: stage, grade, histology
- Biomarkers: pre-operative serum CA125 level, NY-ESO1 expression level from tissue microarray (novel)
- Primary question: overall survival prognosis based on the biomarkers after adjusting for clinical variables
- 37 ovarian cancer patients; Censoring rate $= 24\%$

Example 2: UCLA's Ovarian Cancer Data

Cox PH Model	M1		M2		M3	
variables	HR	p-value	HR	p-value	HR	p-value
stage(3&4 vs 1&2)	4.45	0.10	7.86	0.02	3.97	0.10
grade(1&2 vs 3)	1.07	0.89	1.00	0.99	0.86	0.76
histology						
endometrioid vs clear cell	0.95	0.95	0.42	0.28	1.34	0.72
serious vs clear cell	0.29	0.09	0.21	0.04	0.58	0.41
preop CA125 ($>$ 500 vs $\leq$ 500)	3.92	0.01	4.17	$<$0.005	–	–
NY-ESO1 ($>$ 12 vs $\leq$ 12)	3.12	0.04	–	–	3.67	0.02
R^2	0.553		0.294		0.503	
L^2	0.916		0.991		0.900	
R^2_{SPH}	0.515		0.473		0.396	
R^2_{SH}	0.301		0.267		0.189	

Example 2: UCLA's Ovarian Cancer Data

Cox PH Model variables	M1 HR	M1 p-value	M2 HR	M2 p-value	M3 HR	M3 p-value
stage(3&4 vs 1&2)	4.45	0.10	7.86	0.02	3.97	0.10
grade(1&2 vs 3)	1.07	0.89	1.00	0.99	0.86	0.76
histology						
endometrioid vs clear cell	0.95	0.95	0.42	0.28	1.34	0.72
serious vs clear cell	0.29	0.09	0.21	0.04	0.58	0.41
preop CA125 ($>$ 500 vs $\leq$ 500)	3.92	0.01	4.17	$<$0.005	–	–
NY-ESO1 ($>$ 12 vs $\leq$ 12)	3.12	0.04	–	–	3.67	0.02
R^2	0.553		0.294		0.503	
L^2	0.916		0.991		0.900	
R^2_{SPH}	0.515		0.473		0.396	
R^2_{SH}	0.301		0.267		0.189	

Example 2: UCLA's Ovarian Cancer Data

Cox PH Model	M1		M2		M3	
variables	HR	p-value	HR	p-value	HR	p-value
stage(3&4 vs 1&2)	4.45	0.10	7.86	0.02	3.97	0.10
grade(1&2 vs 3)	1.07	0.89	1.00	0.99	0.86	0.76
histology						
endometrioid vs clear cell	0.95	0.95	0.42	0.28	1.34	0.72
serious vs clear cell	0.29	0.09	0.21	0.04	0.58	0.41
preop CA125 ($>$ 500 vs $\leq$ 500)	3.92	0.01	4.17	$<$0.005	–	–
NY-ESO1 ($>$ 12 vs $\leq$ 12)	3.12	0.04	–	–	3.67	0.02
R^2	0.553		0.294		0.503	
L^2	0.916		0.991		0.900	
R^2_{SPH}	0.515		0.473		0.396	
R^2_{SH}	0.301		0.267		0.189	

Example 2: UCLA Ovarian Cancer Data

1. NY-ESO-1 seems to be a good prognostic factor for platinum-resistant patients
2. preop CA125 does not seem to be a good prognostic factor

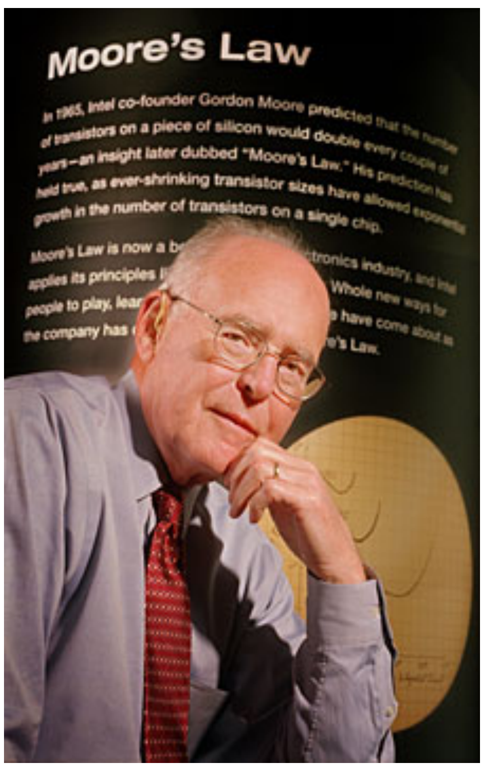
Example 2: UCLA's Ovarian Cancer Data

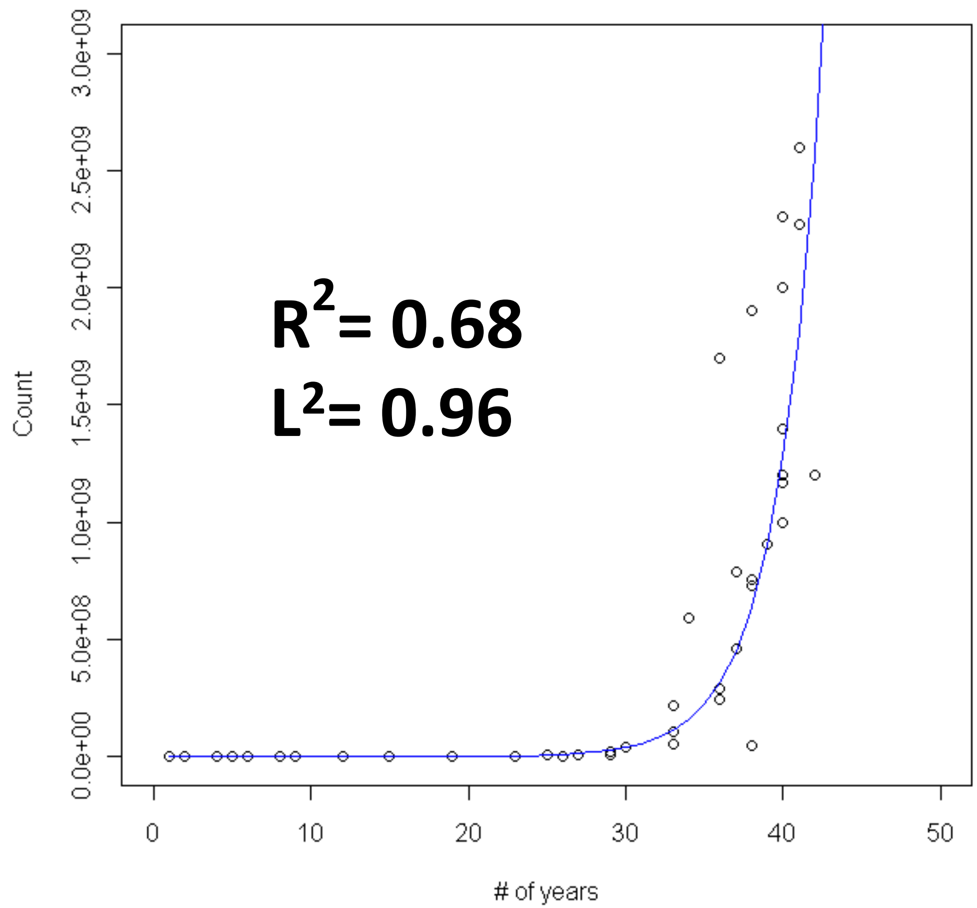
Cox PH Model	M1		M2		M3	
variables	HR	p-value	HR	p-value	HR	p-value
stage(3&4 vs 1&2)	4.45	0.10	7.86	0.02	3.97	0.10
grade(1&2 vs 3)	1.07	0.89	1.00	0.99	0.86	0.76
histology						
endometrioid vs clear cell	0.95	0.95	0.42	0.28	1.34	0.72
serious vs clear cell	0.29	0.09	0.21	0.04	0.58	0.41
preop CA125 ($>$ 500 vs $\leq$ 500)	3.92	0.01	4.17	$<$0.005	–	–
NY-ESO1 ($>$ 12 vs $\leq$ 12)	3.12	0.04	–	–	3.67	0.02
Cox: R^2	0.553		0.294		0.503	
			0.991		0.900	
L^2	0.916					
R^2_{SPH}	0.515		0.473		0.396	
R^2_{SH}	0.301		0.267		0.189	
Weibull: R^2	0.522		0.252		0.516	
			0.402		0.260	
L^2	0.256					
R^2_{SPH}	0.505		0.473		0.393	
Lognormal: R^2	0.489		0.441		0.503	
			0.306		0.270	
L^2	0.279					
R^2_{SPH}	0.477		0.363		0.396	

Example 2: UCLA's Ovarian Cancer Data

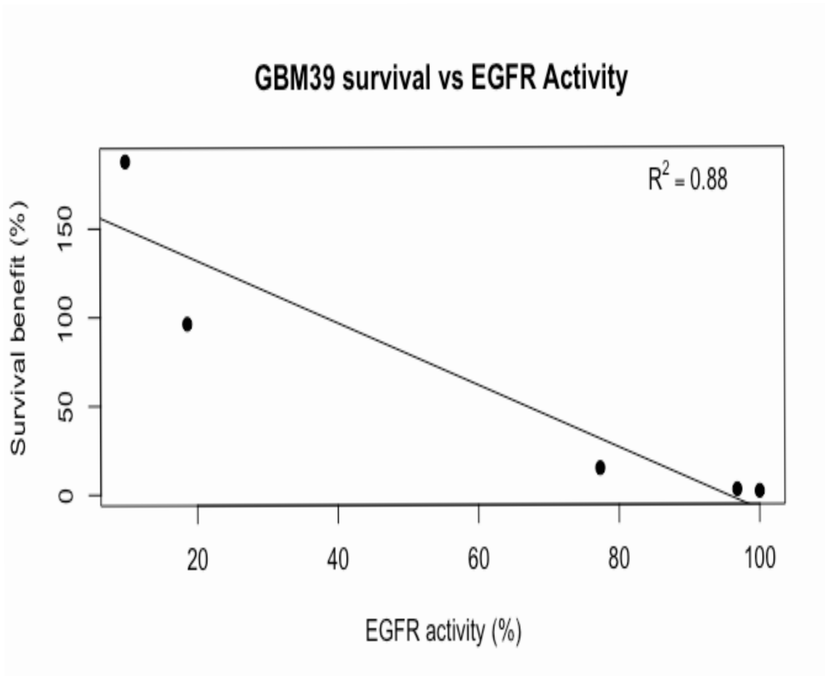
Cox PH Model	M1		M2		M3	
variables	HR	p-value	HR	p-value	HR	p-value
stage(3&4 vs 1&2)	4.45	0.10	7.86	0.02	3.97	0.10
grade(1&2 vs 3)	1.07	0.89	1.00	0.99	0.86	0.76
histology						
endometrioid vs clear cell	0.95	0.95	0.42	0.28	1.34	0.72
serious vs clear cell	0.29	0.09	0.21	0.04	0.58	0.41
preop CA125 ($>$ 500 vs $\leq$ 500)	3.92	0.01	4.17	$<$0.005	–	–
NY-ESO1 ($>$ 12 vs $\leq$ 12)	3.12	0.04	–	–	3.67	0.02
Cox: R^2	0.553		0.294		0.503	
L^2	0.916		0.991		0.900	
R^2_{SPH}	0.515		0.473		0.396	
R^2_{SH}	0.301		0.267		0.189	
Weibull: R^2	0.522		0.252		0.516	
L^2	0.256		0.402		0.260	
R^2_{SPH}	0.505		0.473		0.393	
Lognormal: R^2	0.489		0.441		0.503	
L^2	0.279		0.306		0.270	
R^2_{SPH}	0.477		0.363		0.396	

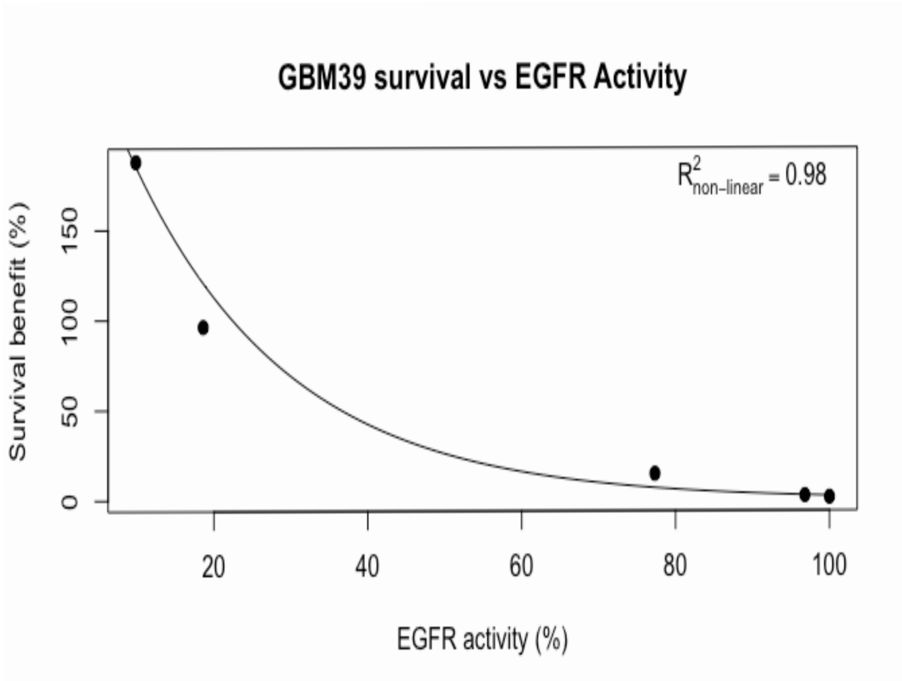
Example 3: Moore's Law (Nonlinear Model)





Example 4: Pre-clinical study of TAGRISSO for GBM





Concluding Remarks

- $\rho^2_{m_{\theta*}}$ – primary measure for the potential predictive power of $m_{\theta^*}(X)$;
- $\lambda^2_{m_{\theta*}}$ - supplementary measure to indicate (by a value less than 1) a) if a linear correction is required for $m_{\theta^*}(X)$ to achieve its potential predictive power and b) how much prediction error reduction can be realized with the correction.
- For the linear model, R^2 reduces to the classical coefficient of determination and $L^2 \equiv 1$, in the absence of censoring.
- Applicable to many event time models (e.g. Cox, AFT, proportional odds, additive risks, transformation, TR models)
- Applicable to a mis-specified model.
- Survival tree/random forest
- R Package: *PAmeasures* available at CRAN library

THANK YOU!