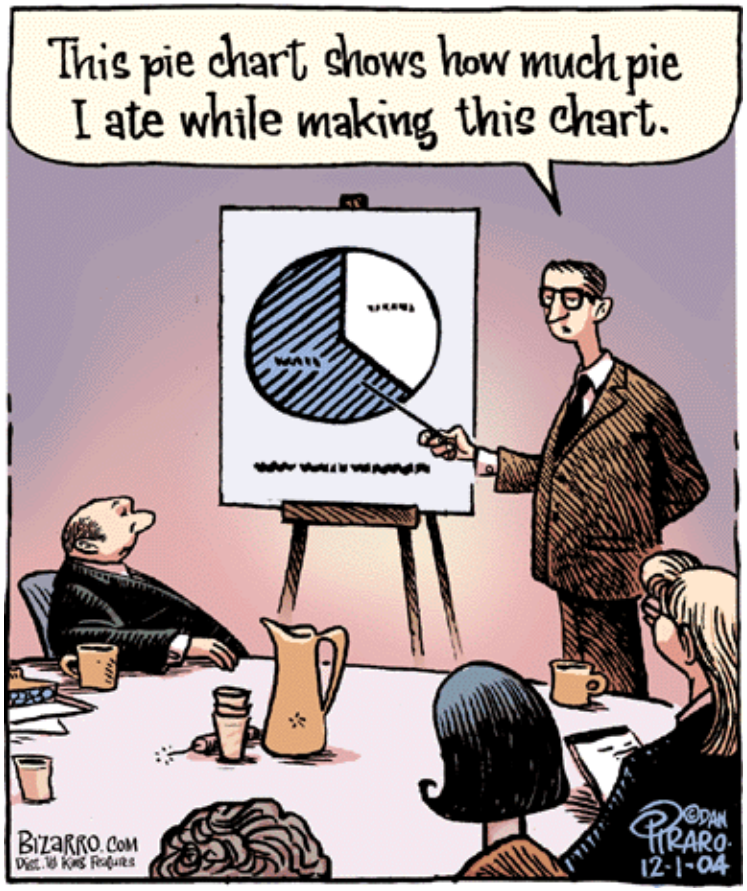
BST 210
Applied Regression Analysis



Happy Thanksgiving!

Lecture 24
Plan for Today

- Overview thus far

Survival Analysis continued:

- Example in R:
  - KM review
  - Cox PH model; formulation and interpretation
  - Assessing PH assumption:
    - Scaled Schoenfeld residuals vs time
    - Log-Log plots
  - Assessing Influence:
    - Dfbetas
    - Deviance
  - Effect modification and confounding
  - Generalized Additive models for nonlinear covariate effects

Thus far…

Regression	Outcome	Assumptions	Model	Effect Estimate
Linear	Continuous	Linearity, Independence, Normality, & Equal variance	$E[Y_i\|X_i] = \beta_0 + \beta_1 \cdot X_i$	β_1 is the change in $E[Y_i]$ associated with a one unit change in X
Logistic	Binary		$\text{logit}(P(Y_i = 1)) = \beta_0 + \beta_1 X_i$	e^{β_1} is the odds ratio associated with a one unit change in X
Multinomial	Categorical		$\log(\frac{P(Y=k)}{P(Y=0)}) = \beta_{k0} + \beta_{k1} X_i$	$e^{\beta_{k1}}$ is the relative risk ratio of being in outcome group k as compared to outcome group 0 for a one unit change in X
Ordinal	Ordinal	Proportional odds	$\log\left(\frac{P(Y \geq j)}{P(Y < j)}\right) = \beta_{0j} + \beta_1 X_i$	e^{β_1} is the odds ratio for $Y \geq j$ versus $Y < j$ associated with a one unit change in X. Note that this is the same for all cut-points j.
Poisson	Count Data	$E[Y_i] = Var(Y_i)$, Incidence rate λ is time invariate	$\log(E[Y_i\|X_i]) = \beta_0 + \beta_1 X_i + \log(t_i)$	e^{β_1} is the incidence rate ratio associated with a one unit increase in X
Survival (Cox)	Time-to-Event	Proportional hazards Also, Exponential: constant baseline hazard Weibull: Weibull baseline hazard	$\lambda(t\|X) = \lambda_0(t)\exp\{\beta_0 + \beta_1 X_i\}$	e^{β_1} is the hazard ratio associated with a one unit increase in X

Recall - Goals of Survival Analysis

Previously:

- To estimate the distribution of survival times for a population
- To test the equality of survival distributions (e.g., treated vs. control group, smokers vs. nonsmokers)

Today:

- To estimate and control for the effects of other covariates when investigating the relationship between a predictor variable and survival time
- To assess model fit or proportional hazards
- Several extensions

Recall – Cox Proportional Hazards Model

- Popular: computational feasible and flexible
- Assumptions: 1. Relationship between X's and T *linear in log scale*
  2. *Proportional Hazards* across levels of each covariate
- 10+ events per predictor variable preferred
- $\boldsymbol{\beta} \sim \text{Normal}$, estimated by partial likelihood methods
- Ties: Breslow most feasible; exact more conservative, computational
- Interpretations:

  e^{β} = hazard ratio for 1-unit increase in X

  $e^{K\beta}$ = hazard ratio for K-unit increase in X

Basic elements – Cox PH Modeling

- KM estimate by level(s) of predictor(s)

- Cox PH: univariable → multivariable modeling

- Assessing PH assumption:
  - log(-log(S(t))) vs log(t) → convergent, divergent, crossing
  - Scaled Shoenfeld Residuals vs time (see paper) → scatter

- Assessing Influence:
  - Dfbetas (just like before)
  - Deviance residuals (absolute value; normalized Martingale)

- Assessing nonlinearity in continuous (no need for categorical) covariate effects:
  - Martingale residuals vs covariate → should yield ~linear lowess smooth, helps to discern functional form of covariates, and if PH met

- Effect modification and confounding
- Time dependent covariates

Assessing model fit: PH

- Assessing PH assumption:

  - log(-log(S(t))) vs log(t) → look for parallelism—or are plots convergent, divergent, crossing?

  - KM estimates → should look proportional, little crossing

  - Scaled Shoenfeld Residuals vs time → want scatter, tests whether a covariate trends with time, and gives global test too. Systematic departures from a horizontal line are indicative of non-proportional hazards, since PH assumes the betas do not vary much over time.

  * Remedies to non-PH:

    - adding time x covariate interaction
    - stratification/categorization

Shoenfeld Residuals and Chisq-test

Partial Residuals for The Proportional Hazards Regression Model

David Schoenfeld

Biometrika, Vol. 69, No. 1. (Apr., 1982), pp. 239-241.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198204%2969%3A1%3C239%3APRFTPH%3E2.0.CO%3B2-3

Biometrika is currently published by Biometrika Trust.

Assessing model fit: Influence, Nonlinearity

- Assessing Influence:

  - Dfbetas → plots estimated changes in regression coefficients upon deleting each observation in turn, divided by their standard errors

  - Deviance residuals → plots normalized transforms of the Martingale residuals, all of which should be roughly symmetrical about 0 with std=1.

- Assessing nonlinearity in continuous (not needed for categorical) covariate effects:

  - Martingale residuals vs covariate → should yield ~linear lowess smooth, helps to discern functional form of covariates, and if PH met

R example (data continued from 11/21,22 lab)

The Sorbinil Retinopathy Trial (SRT) was a randomized, double-blind, placebo controlled trial testing whether the aldose reductase inhibitor, sorbinil, would reduce rates of progression in complications of diabetes, with a primary focus on diabetic retinopathy (Arch Ophthalmol 1990; 108:1234-1244). Between August 1983 and October 1986, the SRT randomized 497 type I diabetic patients at 11 clinical centers. The primary outcome of the trial was worsening of a patient's retinopathy by two or more steps, based on grading fundus photographs. The file srt.dat contains data from 478 randomized participants with some follow-up and complete information on important baseline covariates.

Variables on the file srt.dat are: id (a subject id), sorb (randomized treatment assignment, 1=sorbinil, 0=placebo), tgh (total glycosylated hemoglobin in percent), dur (duration of diabetes in years since diagnosis), sex (2=female, 1=male), fup (duration of follow-up in years until progression of diabetic retinopathy or end of follow-up), and status (1=diabetic retinopathy progressed, 0=no progression).

Read the data into Stata, prepare them for survival analysis, and compare the distributions of patient characteristics by treatment group. Note that women were less likely to be included because sorbinil was teratogenic.