Confidence Intervals and the Stability of Standard Errors

Bradley Efron

Stanford University

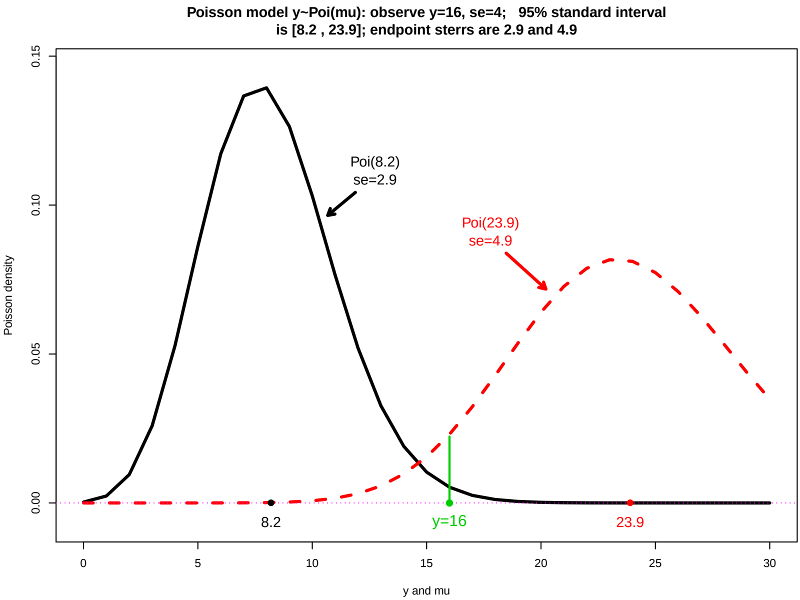
From a Proportional Hazards Analysis

	$\hat{\theta}$	$\widehat{\text{se}}$
Distance	0.18	.15
Age	−0.07	.16
Sex	−0.15	.15
Date	−1.35	.18

- Standard 0.95 interval $\quad \hat{\theta} \pm 1.96 \cdot \widehat{\text{se}}$

A Tacit Assumption of the Standard Intervals

- That the standard error is constant over the interval ("stability") or at least changes much more slowly than θ
- **This talk** Begin with measures of stability
- *Eventual goal* Better confidence intervals
- Poisson example $y \sim \text{Poi}(\mu)$, observe $y = 16$, $\widehat{\text{se}} = 4$


Poisson model y~Poi(mu): observe y=16, se=4; 95% standard interval is [8.2 , 23.9]; endpoint sterrs are 2.9 and 4.9
Poisson density
0.00
0.05
0.10
0.15
Poi(8.2)
se=2.9
Poi(23.9)
se=4.9
8.2
y=16
23.9
0
5
10
15
20
25
30
y and mu

One-Parameter Exponential Families

- $f_\mu(y) = e^{\eta y} f_0(y)/c(\eta)$ [Poisson: $\eta = \log(\mu)$]
- y is "sufficient statistic"
- $\mu = E\{y\}$ "expectation parameter"
- $\hat{\mu} = y$
- $\eta = \eta(\mu)$ "natural parameter"
- $\text{se}(\mu) =$ standard deviation of y
- Fundamental fact

$$\boxed{\frac{d\,\text{se}_\mu(\hat{\mu})}{d\mu} = \frac{\text{SKEW}_\mu\{y\}}{2}}$$

"Big-A" Acceleration

- $\theta = t(\mu) = \theta_\mu$ parameter of interest [e.g., $\theta = \log(\mu)$]
- MLE $\hat{\theta} = t(\hat{\mu})$

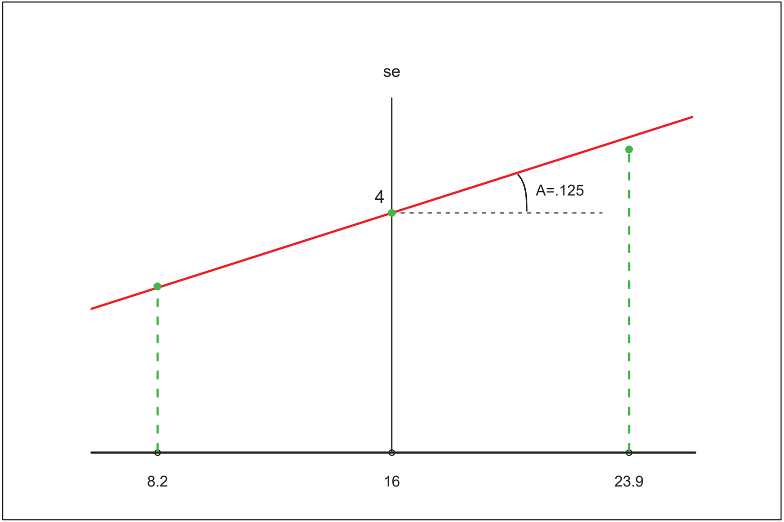
$$\boxed{A(\hat{\theta}) = \left.\frac{d\,\text{se}_\mu(\hat{\theta})}{d\theta_\mu}\right|_{\hat{\mu}}}$$

"Acceleration"

(how fast se is changing as $\hat{\theta}$ changes)

- Fundamental fact says $A(\hat{\mu}) = \text{SKEW}_{\hat{\mu}}(y)/2$
- *Poisson* $A(\hat{\mu}) = \dfrac{1}{2\sqrt{\hat{\mu}}}$ $(= 0.125$ for $\hat{\mu} = 16)$

Poisson example: green points are actual sterrs



Poisson Parameter Examples

- $y \sim \text{Poi}(\mu)$, $y = \hat{\mu} = 16$:

Parameter θ	Acceleration $A(\hat{\theta})$	
μ	.125	expectation
$\sqrt{\mu}$	.006	variance stabilized
$\log(\mu)$	$-.146$	natural

- Last two by bootstrap calculations. . .

A Multiparameter Example

- 22 students each took 5 tests
- Data matrix

$$\underset{22\times 5}{\boldsymbol{x}} = (x_1, x_2, \ldots, x_{22})'$$

- *Multivariate normal model* $x_i \overset{\text{iid}}{\sim} \mathcal{N}_5(\gamma, \Sigma)$
- $\theta = \text{tr}(\Sigma)$
- MLE $\hat{\theta} = \text{tr}(\hat{\Sigma}) = 976$
- $\text{se}(\hat{\theta})$? How stable?

The Student Score Data

(Mardia, Kent and Bibby 2003)

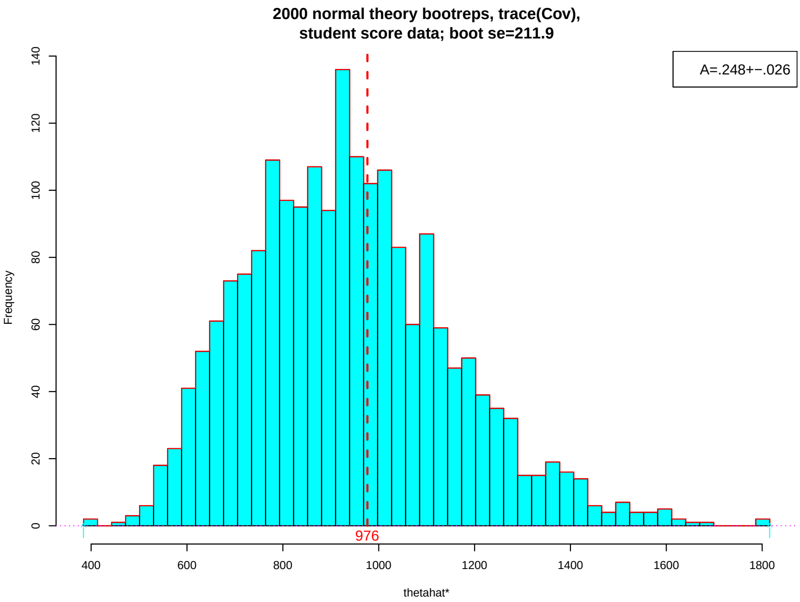
student	mech	vecs	alg	analy	stat
1	7	51	43	17	22
2	44	69	53	53	53
3	49	41	61	49	64
4	59	70	68	62	56
5	34	42	50	47	29
⋮	⋮	⋮	⋮	⋮	⋮
21	46	49	53	59	37
$n = 22$	63	63	65	70	63

Parametric Bootstrap Replications

- $\mathcal{N}_5(\gamma, \Sigma) \xrightarrow{\text{iid}} (x_1, x_2, \ldots, x_{22}) \xrightarrow{\text{mle}} (\hat{\gamma}, \hat{\Sigma}) \longrightarrow \hat{\theta} = \text{tr}(\hat{\Sigma}) = 976$
- Boots $\mathcal{N}_5(\hat{\gamma}, \hat{\Sigma}) \longrightarrow (x_1^*, x_2^*, x_{22}^*) \longrightarrow (\hat{\gamma}^*, \hat{\Sigma}^*) \longrightarrow \hat{\theta}^* = \text{tr}(\hat{\Sigma}^*)$ $\longrightarrow \left\{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\right\}$
- *Bootstrap standard error*

$$\widehat{\text{se}} = \left[\sum_{i=1}^{B} \left(\hat{\theta}_i^* - \hat{\theta}_\cdot^*\right)^2 \Big/ B\right]^{1/2}$$

- $B = 2000$ gave $\widehat{\text{se}} = 211.9$ for $\hat{\theta} = \text{tr}(\hat{\Sigma})$


2000 normal theory bootreps, trace(Cov),
student score data; boot se=211.9
A=.248+−.026
Frequency
0
20
40
60
80
100
120
140
976
400
600
800
1000
1200
1400
1600
1800
thetahat*

Multiparameter Exponential Families

- $f_\mu(y) = e^{\eta' y} f_0(y)/c(\eta)$
- y is p-dimensional sufficient vector
- $\mu = E_\mu(y)$ "expectation parameter"
- $\hat{\mu} = y$
- $\eta = \eta(\mu)$ "natural parameter"
- $\boldsymbol{V}_\mu = \text{cov}_\mu\{y\}$
- *Student score matrix* $\boldsymbol{x}$ y has $p = 20$:

  column averages, col^2 averages, col product averages

Scalar Parameter of Interest

- $\theta = t(\mu)$
- MLE $\hat{\theta} = t(y) = t(\hat{\mu})$
- *Student score* $\theta = \text{tr}(\Sigma)$, $\hat{\theta} = 976$
- Gradient p-vector $\dot{\boldsymbol{t}} = \begin{pmatrix} \vdots \\ \partial t / \partial \mu_j \\ \vdots \end{pmatrix}_{\mu = y}$

$$t(y) \doteq t(\hat{\mu}) + \dot{\boldsymbol{t}}'(y - \hat{\mu})$$

- Approx standard error $\widehat{\text{se}}(\hat{\theta}) \doteq \left(\dot{\boldsymbol{t}}' \boldsymbol{V}_{\hat{\mu}} \dot{\boldsymbol{t}}\right)^{1/2} = \overline{\text{se}}$

Parametric Bootstrap Replications

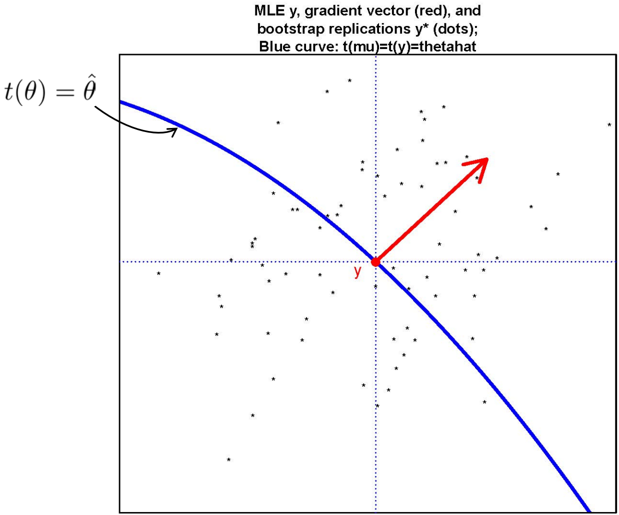
- $y = \text{MLE } \hat{\mu}$
- $f_{\hat{\mu}} \xrightarrow{\text{iid}} \{y_1^*, y_2^*, \ldots, y_B^*\}$

$$\boxed{\hat{\theta}_i^* = t(y_i^*)} \qquad i = 1, 2, \ldots, B$$

- $D_i = \hat{\theta}_i^* - \hat{\theta}_\cdot^*$ and $d_i = \dot{\mathbf{t}}'(y_i^* - \hat{\mu})$

$$\widehat{\text{se}} = \left(\sum_{i=1}^{B} D_i^2 / B \right)^{1/2} \qquad \overline{\text{se}} = \left(\sum_{i=1}^{B} d_i^2 / B \right)^{1/2}$$

(211.9) (215.7)


MLE y, gradient vector (red), and
bootstrap replications y* (dots);
Blue curve: t(mu)=t(y)=thetahat
$t(\theta) = \hat{\theta}$
y

Stein's Least Favorable Family

- *Idea* One-parameter subfamily of $f_\mu(y)$

$$\eta_\lambda = \hat{\eta} + \dot{\boldsymbol{t}}\lambda \qquad (\hat{\eta} = \text{MLE of } \eta)$$

- Just as hard to estimate $\theta_\lambda = t(\mu_\lambda)$ as $\theta = t(\mu)$
- Now can use one-parameter formula $\widehat{A} = \left.\frac{d\,\text{se}_\lambda}{d\theta_\lambda}\right|_{\lambda=0}$

Theorem

$$\boxed{\widehat{A} = \frac{1}{2}\frac{\sum_{i=1}^{B} D_i^2 d_i / B}{\widehat{\text{se}}\ \overline{\text{se}}^2}} \qquad = 0.248$$

Bagging Formula

(Efron 2014 *JASA*)

- $\hat{\theta}_i^* = t(y_i^*) = t_i^*$
- *Bagged estimate* $\quad s = \frac{1}{B}\sum_1^B t_i^*:$

$$\boxed{\overline{\text{se}}(s) = (\widehat{\text{cov}}'\,\hat{\boldsymbol{V}}\widehat{\text{cov}})^{1/2}}$$

where $\widehat{\text{cov}} = \widehat{\text{cov}}(y_i^*, t_i^*)$, $\hat{\boldsymbol{V}} = \widehat{\text{cov}}(y_i^*)$

- But bootstrap standard error is

$$\left[\frac{1}{B}\sum t_i^{*2} - \left(\frac{1}{B}\sum t_i^*\right)^2\right]^{1/2}$$

BCa Confidence Limits

(Efron 1987 *JASA*)

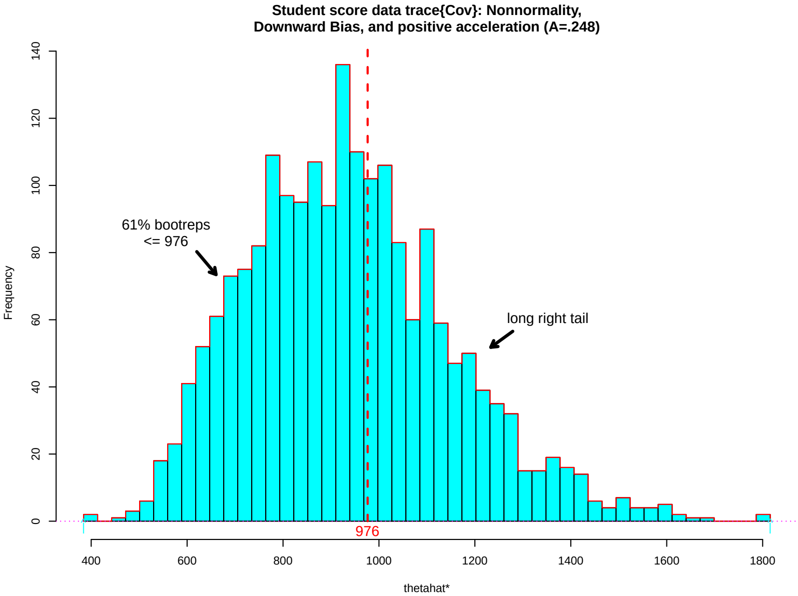
- *Idea* Improve convergence rate of standard intervals

$$\theta_{\text{BCa}}(\alpha) = \widehat{G}^{-1}\left\{\Phi\left[\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right]\right\}$$

- $\widehat{G}(t) = \dfrac{\#\{\hat{\theta}_i^* \leq t\}}{B}$
- $z^{(\alpha)} = \Phi^{-1}(\alpha)$
- $\hat{z}_0 = \Phi^{-1}\left\{\widehat{G}(\hat{\theta})\right\}$ is the "bias corrector"
- $\hat{a}$ is "little-a" acceleration

Three Corrections to the Standard Intervals

- Standard intervals $\hat{\theta} \sim \mathcal{N}(\theta, \text{se}^2)$
  1. Correct for non-normality ($\widehat{G}$)
  2. Correct for bias ($\hat{z}_0$)
  3. Correct for acceleration ($\hat{a}$)
- Asymptotic error rate $O(1/n)$ instead of $O(1/\sqrt{n})$


Student score data trace{Cov}: Nonnormality,
Downward Bias, and positive acceleration (A=.248)
61% bootreps
<= 976
long right tail
976
Frequency
0
20
40
60
80
100
120
140
400
600
800
1000
1200
1400
1600
1800
thetahat*

"little-a" acceleration

- *Least favorable family* $\eta_\lambda = \hat{\eta} + \dot{\boldsymbol{t}}\lambda$ ($\hat{\eta}$ fixed)
- One-parameter exponential family

  $$\text{“}y\text{”} = d = \dot{\boldsymbol{t}}'(y^* - \hat{\mu}), \quad \text{“}\eta\text{”} = \lambda \quad (\hat{\lambda} = 0)$$

  $$\boxed{\hat{a} = \frac{1}{3} \frac{\text{SKEW}_\lambda(d)}{2} \bigg|_{\lambda=0}}$$

- 1/3 of big-A acceleration in LFF

  (correcting for non-normality removes 2/3 of acceleration effect)

Program bcaj

- Automates calculation of BCa intervals
- Enter $\hat{\theta}$ and B bootreps $\left\{y_i^*, \hat{\theta}_i^*\right\}$
- Calculates boot cdf $\widehat{G}(\cdot)$
- $\hat{z}_0 = \Phi^{-1}\left\{\widehat{G}(\hat{\theta})\right\}$
- Gradient $\dot{\boldsymbol{t}}$ from local linear regression

$$\hat{\theta}_i^* \doteq \hat{\theta} + \dot{\boldsymbol{t}}'(y_i^* - \hat{\mu}) \qquad (y_i^* \text{ near } \hat{\mu})$$

Estimating "little a"

- Because d is sufficient statistic in LFF
- $\hat{a} = \frac{1}{3}\frac{\widehat{\text{SKEW}}(d)}{2} = \boxed{\frac{1}{6\widehat{\text{se}}^3}\sum_{i=1}^{B} d_i^3/B}$

$$\left[\widehat{A} = \frac{1}{2\widehat{\text{se}}\,\overline{\text{se}}^2}\sum_{i=1}^{B} D_i^2 d_i/B, \quad D_i = \hat{\theta}_i^* - \hat{\theta}., \quad d_i = \dot{\boldsymbol{t}}'(y_i^* - \hat{\mu})\right]$$

- *Alternatively* $\hat{a}_0 = \Phi^{-1}\left\{\frac{\#d_i \le 0}{B}\right\}$
- Why not $\frac{1}{6\widehat{\text{se}}^3}\sum_{i=1}^{B} D_i^3/B$?

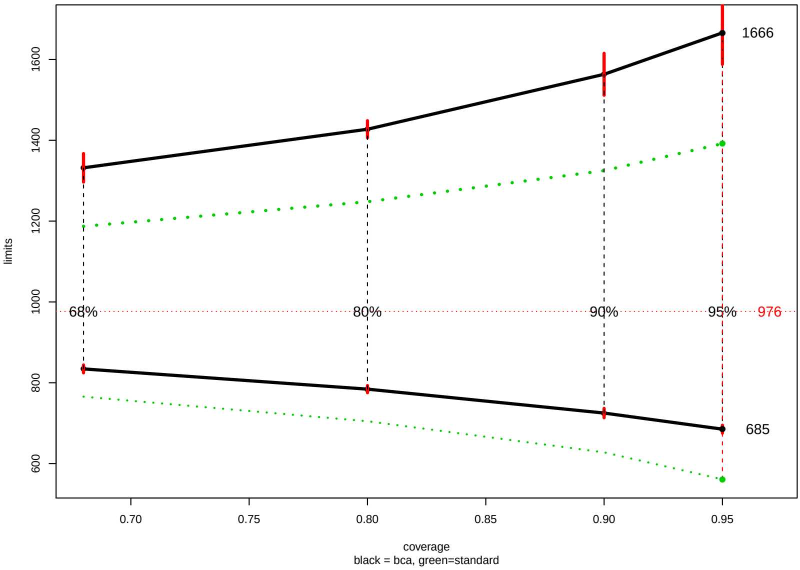
BCAJ OUTPUT FOR STUDENT SCORE DATA

tr(cov), $B = 2000$

alpha	bcalims	jacksd	Standard
.025	685	(9.7)	561
.05	725	(9.9)	628
.16	834	(10.1)	766
.84	1332	(39.9)	1187
.95	1563	(51.8)	1325
.975	1666	(74.9)	1392

	$\hat{\theta}$	$\widehat{\text{se}}$	$\overline{\text{se}}$	$\hat{z}_0$	$\hat{a}$	$\widehat{A}$
est	976	211.9	215.7	.269	.083	.248
jsd	0	3.4	3.4	.028	.011	.029

Two−sided bootstrap confidence intervals for tr(Cov) are shifted up from the Standard intervals



THE BOOTSTRAP SAMPLE SIZE B

- Was $B = 2000$ big enough? Too big?
- `bcaj`:
  - randomly divides the 2000 into 10 groups of 200
  - omits each group in turn, reruns `bca` on 1800
- "jacksd" is jackknife estimate of standard deviation
- Error from stopping at $B = 2000$ rather than $B = \infty$

Nonparametric BCa is Easier! (bcanonj)

- Call: `bcanonj(x,B,tfunc)`
- `x` is iid sample $(x_1, x_2, \ldots, x_n)$
- `B` is boot sample size
- `tfunc` is R function: $\hat{\theta} = \texttt{tfunc}(x)$
- Computes B bootreps $\hat{\theta}^* = t(x^*), \quad x^* = (x_1^*, x_2^*, \ldots, x_n^*)$,

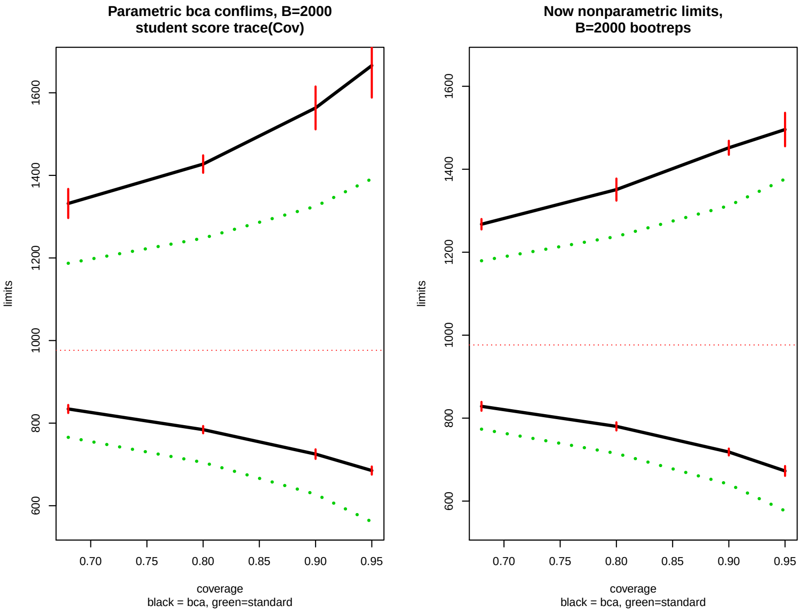
  then full BCa output

BCANONJ OUTPUT FOR STUDENT SCORE DATA

tr(cov), $B = 2000$

	bcalims	jacksd	Standard
.025	649	(10.23)	557
.05	704	(11.66)	625
.16	823	(8.62)	764
.84	1276	(10.54)	1189
.95	1418	(30.68)	1328
.975	1553	(16.43)	1396

	$\hat{\theta}$	$\widehat{se}$	$\hat{z}_0$	$\hat{a}$	$\widehat{A}$
est	976	213.87	.217	.077	.173
jsd	0	3.08	.027	.000	.026

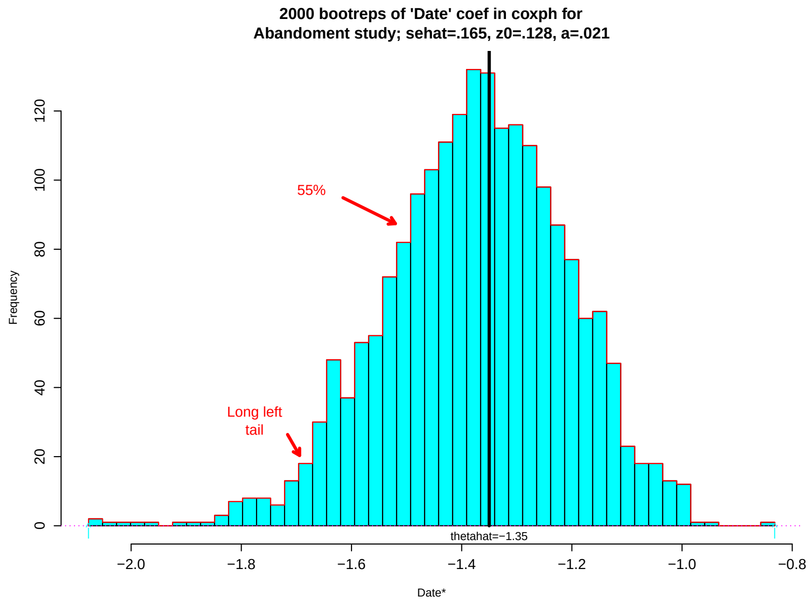

Parametric bca conflims, B=2000
student score trace(Cov)
limits
coverage
black = bca, green=standard
Now nonparametric limits,
B=2000 bootreps
limits
coverage
black = bca, green=standard

Guatemala Abandonment Study

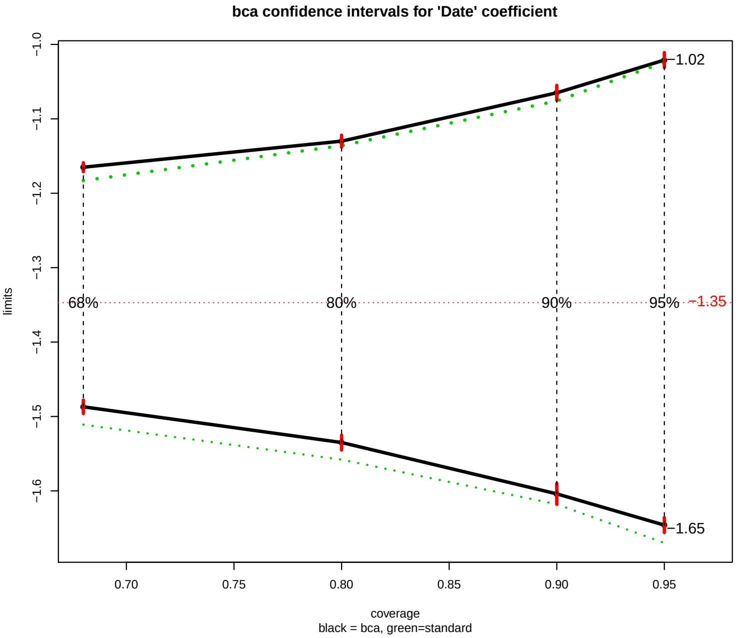
- 500 pediatric cancer subjects
- 47 abandoned by family

  coxph(Aband $\sim$ Distance $+$ Age $+$ Sex $+$ Date)

- Date: $\hat{\theta} = -1.35$ $\widehat{\text{se}} = 0.18$ $(\widehat{\text{se}}_{\text{boot}} = 0.165)$
- S $=$ Surv(Aband, days since entry)
- `bcanonj(S,2000,coxph...)`



2000 bootreps of 'Date' coef in coxph for
Abandoment study; sehat=.165, z0=.128, a=.021
55%
Long left
tail
Frequency
0
20
40
60
80
100
120
thetahat=−1.35
−2.0
−1.8
−1.6
−1.4
−1.2
−1.0
−0.8
Date*



bca confidence intervals for 'Date' coefficient
limits
−1.0
−1.1
−1.2
−1.3
−1.4
−1.5
−1.6
−1.02
−1.35
−1.65
68%
80%
90%
95%
0.70
0.75
0.80
0.85
0.90
0.95
coverage
black = bca, green=standard