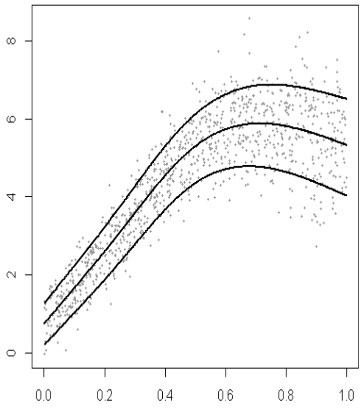
Statistical Reliability Meets Wind Energy
Semester 2 Presentation

Presenter: Michael S. Czahor
PhD Candidate: WESEP &Statistics
Major Professor: Dr. William Q. Meeker

Presentation Date: April 17th 2014



Quick Review from Semester 1

- Prevent unplanned maintenance via the use of statistical analysis
- “Big Data”
- Failure Modes/Detection
- Begin Analysis Non-Parametrically

Benefits of Early Detection

- Limit downtower repairs
- Eliminate expedited crane charges
- Minimize downtime and optimize planning for low wind repairs

Semester 2 Overview

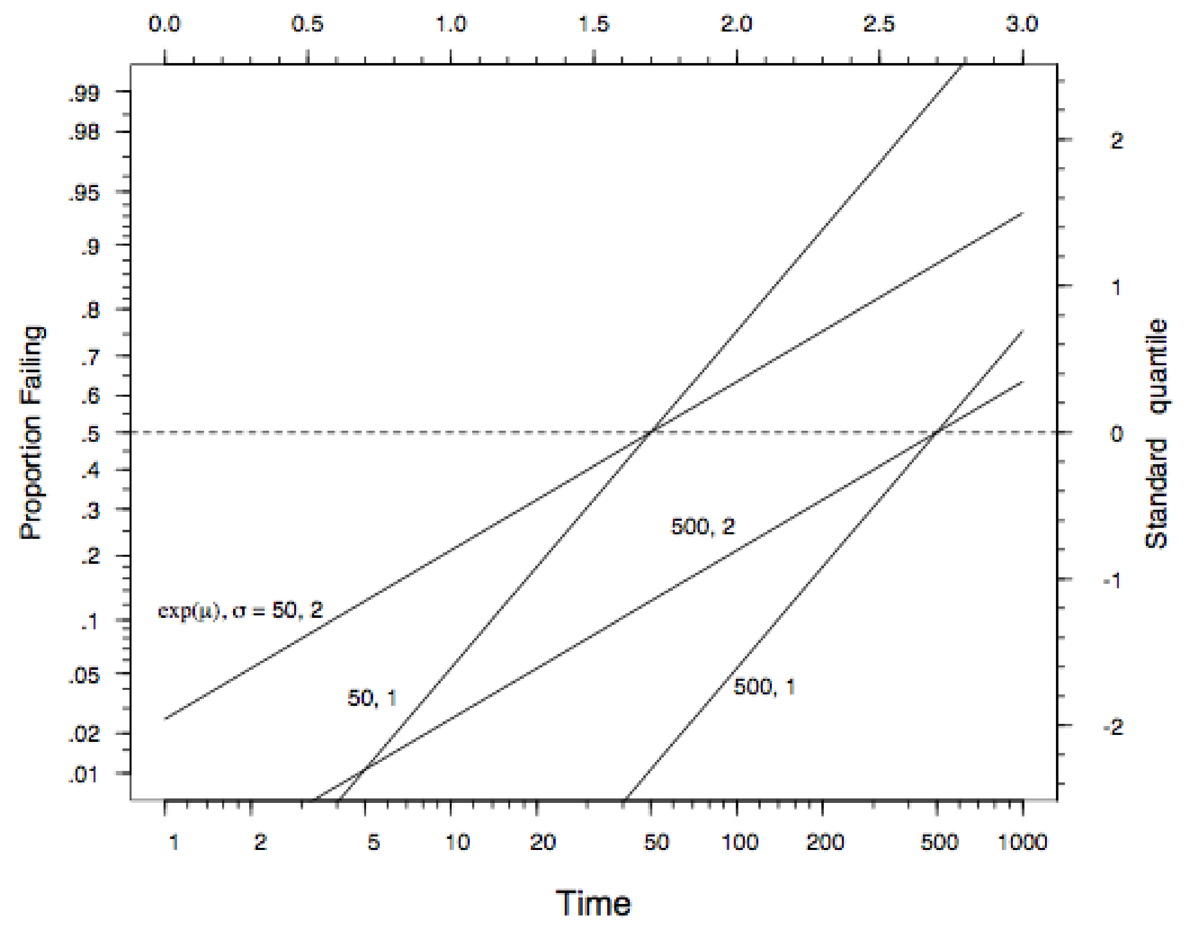
- Statistics 533 (Reliability) in relation to research
- Strategy for Data Analysis, Modeling and Inference
- Bootstrapping with Non- Integer Weights
- Relate small simulation study to Wind Energy through estimation of quantiles.

Reasons for Collecting Reliability Data

- Assessing characteristics of materials
- Predict product reliability in design stage
- Assessing the effect of a proposed design change
- Comparing 2 or more different manufacturers
- Assess product reliability in field
- Checking the veracity of an advertising claim
- Predict product warranty costs

Slide: 1-5 From Dr. Meeker's Statistics 533 at Iowa State

Probability Plotting



Slide 6-10 from Dr. Meeker Stat 533 at Iowa State

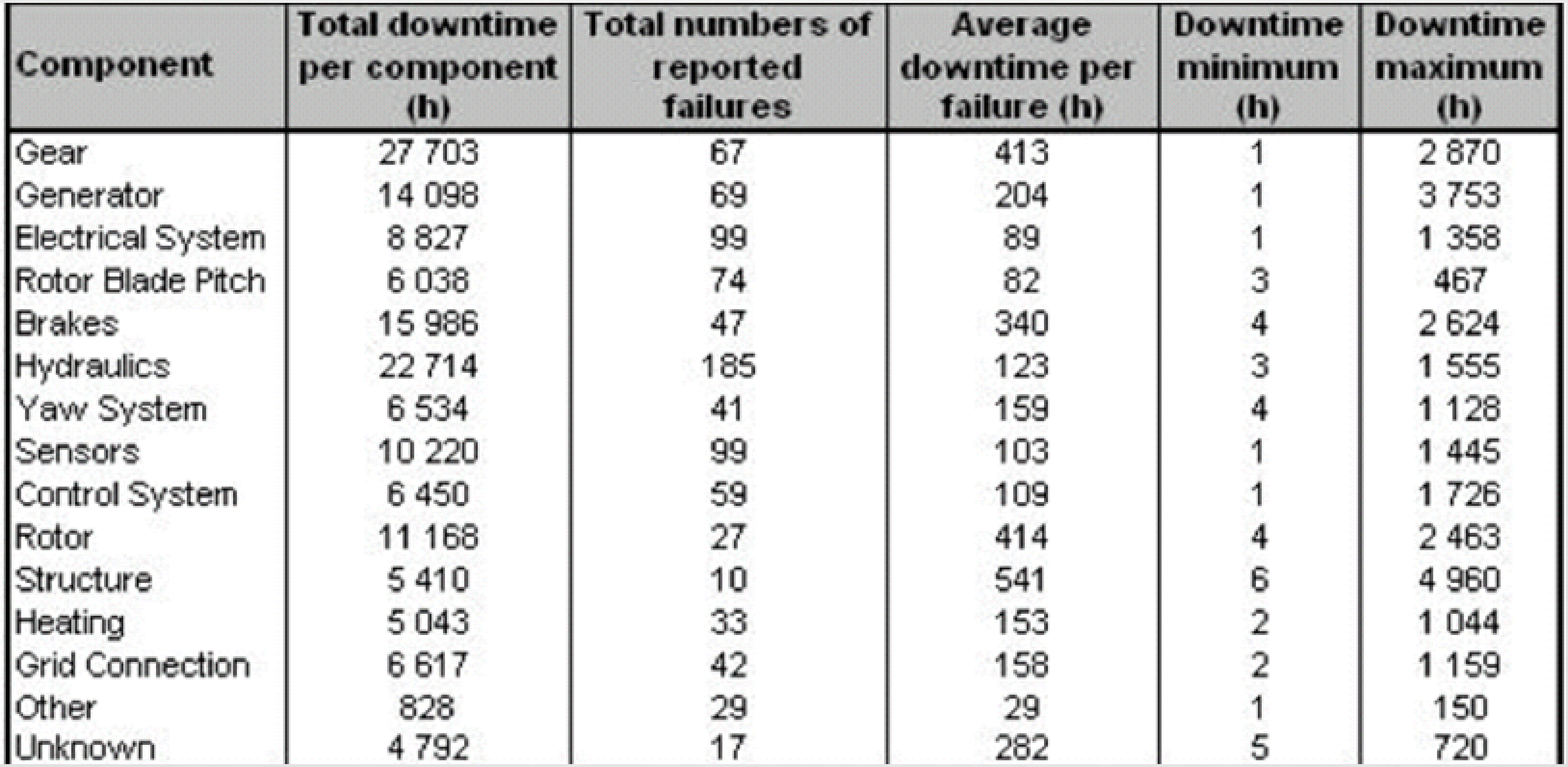
Functions of the Parameters

- Cumulative distribution function (cdf) of T $F(t;\theta)=\Pr(T \leq t)$, $t>0$.
- The p quantile of T is the smallest value t_p such that $F(t_p;\theta) \geq p$.
- Estimating quantiles (fraction failing as a function of time) The estimation will provide the corresponding time scale value

Reliability in Wind Turbines

- High repair cost
- Analyze stresses affecting the critical components
- Use of sensors to send data to centralized locations
- System Retirement
- Detect unsafe operating conditions
- Prognostic Purposes

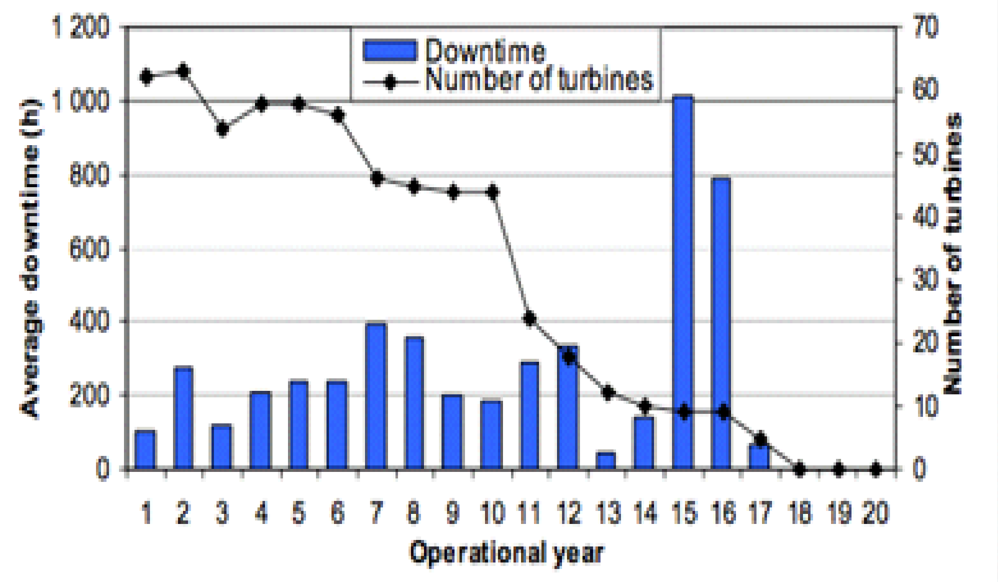
Component Reliability I

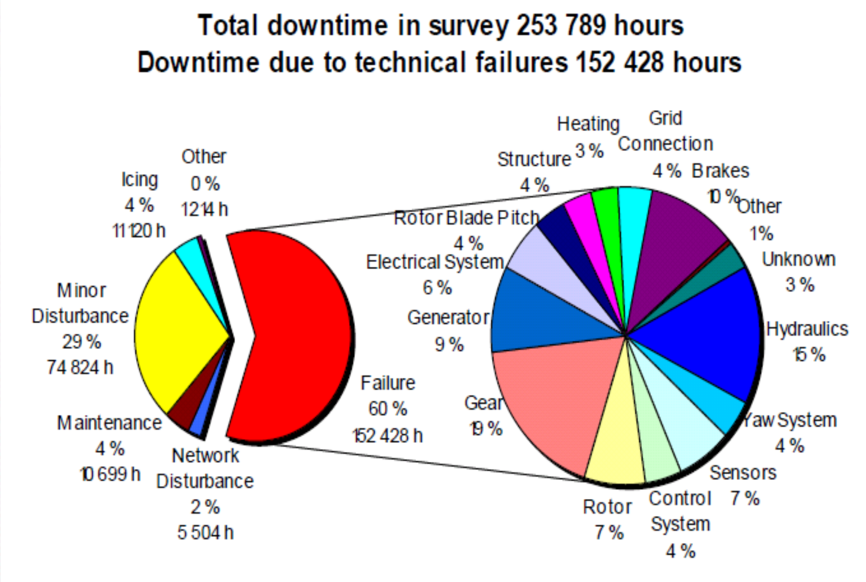


Component	Total downtime per component (h)	Total numbers of reported failures	Average downtime per failure (h)	Downtime minimum (h)	Downtime maximum (h)
Gear	27 703	67	413	1	2 870
Generator	14 098	69	204	1	3 753
Electrical System	8 827	99	89	1	1 358
Rotor Blade Pitch	6 038	74	82	3	467
Brakes	15 986	47	340	4	2 624
Hydraulics	22 714	185	123	3	1 555
Yaw System	6 534	41	159	4	1 128
Sensors	10 220	99	103	1	1 445
Control System	6 450	59	109	1	1 726
Rotor	11 168	27	414	4	2 463
Structure	5 410	10	541	6	4 960
Heating	5 043	33	153	2	1 044
Grid Connection	6 617	42	158	2	1 159
Other	828	29	29	1	150
Unknown	4 792	17	282	5	720

Wind energy statistics in Finland:
http://www.vtt.fi/proj/windenergystatistics/?lang=en [accessed 06/11/2013].

Component Reliability II

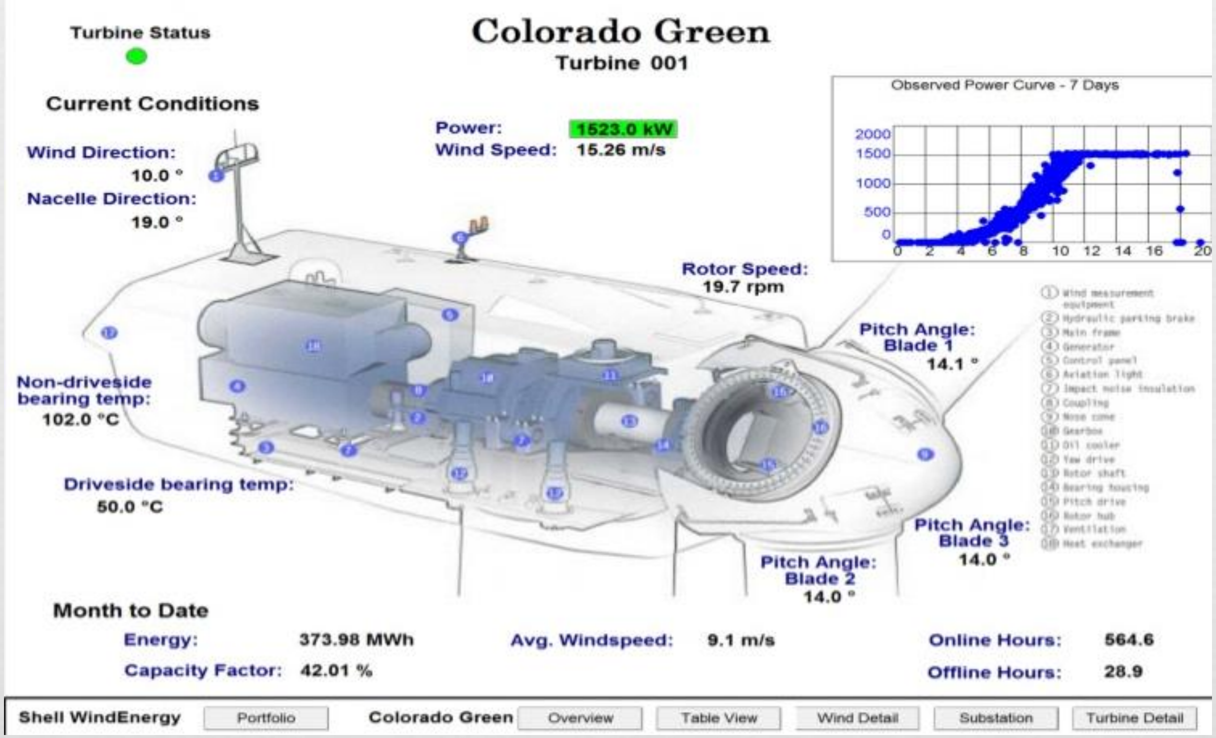




Wind energy statistics in Finland:
http://www.vtt.fi/proj/windenergystatistics/?lang=en [accessed 06/11/2013].

Converting Data Into Information

- **Reliability Management**
  - Reactive: Trending, Analysis, Troubleshooting
  - Proactive: Failure predictors, Reliability Centered Maintenance



Colorado Green
Turbine 001

Turbine Status

Current Conditions

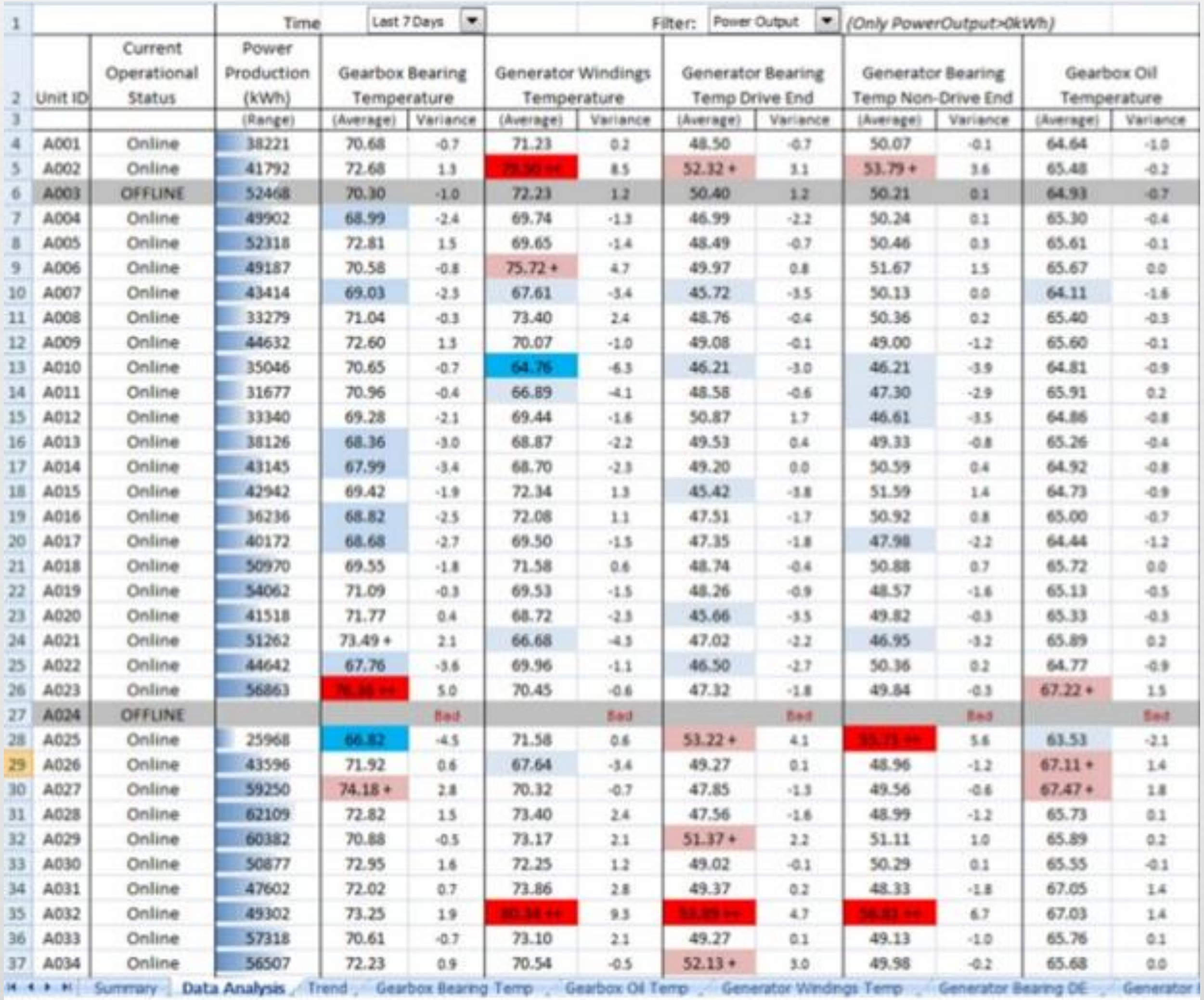
Wind Direction: 10.0 °
Nacelle Direction: 19.0 °
Power: 1523.0 kW
Wind Speed: 15.26 m/s
Rotor Speed: 19.7 rpm
Pitch Angle: Blade 1 14.1 °
Pitch Angle: Blade 2 14.0 °
Pitch Angle: Blade 3 14.0 °
Non-driveside bearing temp: 102.0 °C
Driveside bearing temp: 50.0 °C

Month to Date

Energy: 373.98 MWh
Capacity Factor: 42.01 %
Avg. Windspeed: 9.1 m/s
Online Hours: 564.6
Offline Hours: 28.9

Data: Provided by Shell Corporation

Converting Data Into Information



1			Time	Last 7 Days ▾				Filter:	Power Output ▾	*(Only PowerOutput>0kWh)*			
2	Unit ID	Current Operational Status	Power Production (kWh)	Gearbox Bearing Temperature		Generator Windings Temperature		Generator Bearing Temp Drive End		Generator Bearing Temp Non-Drive End		Gearbox Oil Temperature	
3			(Range)	(Average)	Variance	(Average)	Variance	(Average)	Variance	(Average)	Variance	(Average)	Variance
4	A001	Online	38221	70.68	-0.7	71.23	0.2	48.50	-0.7	50.07	-0.1	64.64	-1.0
5	A002	Online	41792	72.68	1.3	[illegible] ++	8.5	52.32 +	3.1	53.79 +	3.6	65.48	-0.2
6	A003	OFFLINE	52468	70.30	-1.0	72.23	1.2	50.40	1.2	50.21	0.1	64.93	-0.7
7	A004	Online	49902	68.99	-2.4	69.74	-1.3	46.99	-2.2	50.24	0.1	65.30	-0.4
8	A005	Online	52318	72.81	1.5	69.65	-1.4	48.49	-0.7	50.46	0.3	65.61	-0.1
9	A006	Online	49187	70.58	-0.8	75.72 +	4.7	49.97	0.8	51.67	1.5	65.67	0.0
10	A007	Online	43414	69.03	-2.3	67.61	-3.4	45.72	-3.5	50.13	0.0	64.11	-1.6
11	A008	Online	33279	71.04	-0.3	73.40	2.4	48.76	-0.4	50.36	0.2	65.40	-0.3
12	A009	Online	44632	72.60	1.3	70.07	-1.0	49.08	-0.1	49.00	-1.2	65.60	-0.1
13	A010	Online	35046	70.65	-0.7	64.76	-6.3	46.21	-3.0	46.21	-3.9	64.81	-0.9
14	A011	Online	31677	70.96	-0.4	66.89	-4.1	48.58	-0.6	47.30	-2.9	65.91	0.2
15	A012	Online	33340	69.28	-2.1	69.44	-1.6	50.87	1.7	46.61	-3.5	64.86	-0.8
16	A013	Online	38126	68.36	-3.0	68.87	-2.2	49.53	0.4	49.33	-0.8	65.26	-0.4
17	A014	Online	43145	67.99	-3.4	68.70	-2.3	49.20	0.0	50.59	0.4	64.92	-0.8
18	A015	Online	42942	69.42	-1.9	72.34	1.3	45.42	-3.8	51.59	1.4	64.73	-0.9
19	A016	Online	36236	68.82	-2.5	72.08	1.1	47.51	-1.7	50.92	0.8	65.00	-0.7
20	A017	Online	40172	68.68	-2.7	69.50	-1.5	47.35	-1.8	47.98	-2.2	64.44	-1.2
21	A018	Online	50970	69.55	-1.8	71.58	0.6	48.74	-0.4	50.88	0.7	65.72	0.0
22	A019	Online	54062	71.09	-0.3	69.53	-1.5	48.26	-0.9	48.57	-1.6	65.13	-0.5
23	A020	Online	41518	71.77	0.4	68.72	-2.3	45.66	-3.5	49.82	-0.3	65.33	-0.3
24	A021	Online	51262	73.49 +	2.1	66.68	-4.3	47.02	-2.2	46.95	-3.2	65.89	0.2
25	A022	Online	44642	67.76	-3.6	69.96	-1.1	46.50	-2.7	50.36	0.2	64.77	-0.9
26	A023	Online	56863	[illegible] ++	5.0	70.45	-0.6	47.32	-1.8	49.84	-0.3	67.22 +	1.5
27	A024	OFFLINE			Bad		Bad		Bad		Bad		Bad
28	A025	Online	25968	66.82	-4.5	71.58	0.6	53.22 +	4.1	[illegible] ++	5.6	63.53	-2.1
29	A026	Online	43596	71.92	0.6	67.64	-3.4	49.27	0.1	48.96	-1.2	67.11 +	1.4
30	A027	Online	59250	74.18 +	2.8	70.32	-0.7	47.85	-1.3	49.56	-0.6	67.47 +	1.8
31	A028	Online	62109	72.82	1.5	73.40	2.4	47.56	-1.6	48.99	-1.2	65.73	0.1
32	A029	Online	60382	70.88	-0.5	73.17	2.1	51.37 +	2.2	51.11	1.0	65.89	0.2
33	A030	Online	50877	72.95	1.6	72.25	1.2	49.02	-0.1	50.29	0.1	65.55	-0.1
34	A031	Online	47602	72.02	0.7	73.86	2.8	49.37	0.2	48.33	-1.8	67.05	1.4
35	A032	Online	49302	73.25	1.9	80.34 ++	9.3	[illegible] ++	4.7	[illegible] ++	6.7	67.03	1.4
36	A033	Online	57318	70.61	-0.7	73.10	2.1	49.27	0.1	49.13	-1.0	65.76	0.1
37	A034	Online	56507	72.23	0.9	70.54	-0.5	52.13 +	3.0	49.98	-0.2	65.68	0.0

Data: Provided by
Shell Corporation

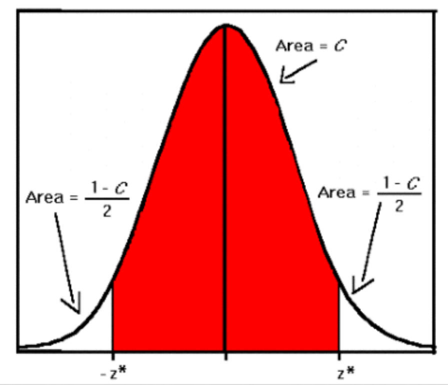
Brief Statistics Review

- When estimating something the estimate will have uncertainty due to limited amounts of data.
- The estimate will be within some amount of the true parameter value.
- Crude approximation via normal interval
- More sophisticated methods through simulation
(e.g. bootstrapping)

Brief Statistics Review

- **Coverage probability**: This term refers to the probability that a procedure for constructing random regions will produce an interval containing, or covering, the true value. It is a property of the interval producing procedure, and is independent of the particular sample to which such a procedure is applied. We can think of this quantity as the chance that the interval constructed by such a procedure will contain the parameter of interest.

- **Confidence level**: The interval produced for any particular sample, using a procedure with coverage probability p, is said to have a confidence level of p; hence the term 'confidence interval'. Note that, by this definition, the confidence level and coverage probability are equivalent before we have obtained our sample. After, of course, the parameter is either in or not in the interval, and hence the 'chance' that the interval contains the parameter is either 0 or 1.



Background on Bootstrapping

STEP 4: Make inference by resampling statistics, $\bar{X}_1^*$, $\bar{X}_2^*$, …, $\bar{X}_B^*$

$$\bar{X}^*_{(\cdot)} = \frac{\bar{X}_1^* + \bar{X}_2^* + \ldots + \bar{X}_B^*}{B} \xrightarrow[B\to\infty]{LLN} E_*(\bar{X}^*) = ?$$

$$E_*(\bar{X}^*) = E_*(\frac{X_1^* + X_2^* + \ldots + X_n^*}{n}) = \frac{\sum_{i=1}^{n} E_*(X_i^*)}{n} = E_*(X_1^*) = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

$$\widehat{bias}_B = (\bar{X}^*_{(\cdot)} - \bar{X}) \xrightarrow[B\to\infty]{LLN} (\bar{X} - \bar{X}) = 0.$$

$$\widehat{\mathrm{var}}_B = \frac{1}{B-1}\sum_{b=1}^{B}(\bar{X}_b^* - \bar{X}^*_{(\cdot)})^2 \xrightarrow[B\to\infty]{LLN} E_*\left(\bar{X}^* - E_*(\bar{X}^*)\right)^2 = V_*(\bar{X}^*) = ?$$

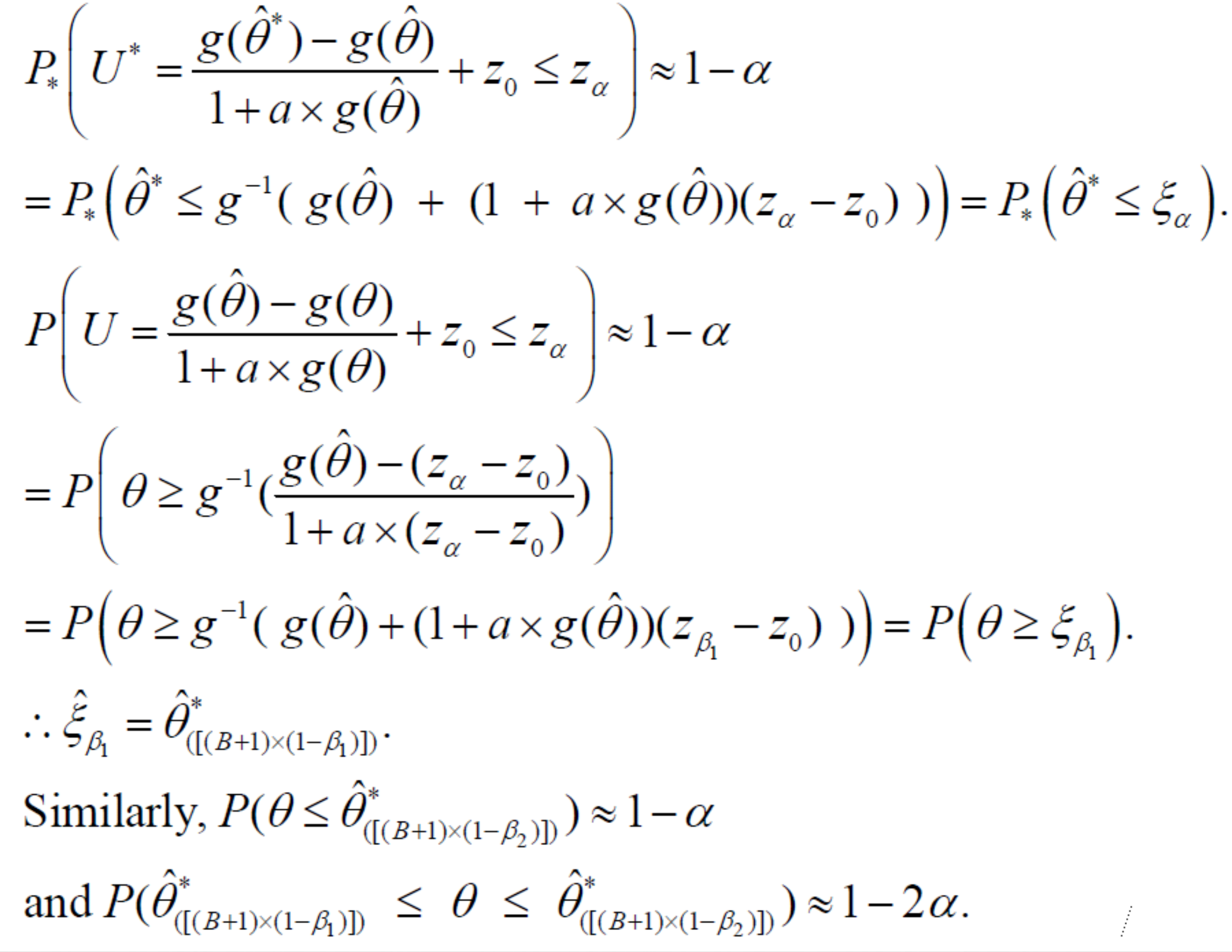
$$V_*(\bar{X}^*) = V_*(\frac{X_1^* + X_2^* + \ldots + X_n^*}{n}) = \frac{V_*(X_1^*) + V_*(X_2^*) + \ldots + V_*(X_n^*)}{n^2}$$

$$= \frac{V_*(X_1^*)}{n} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}{n} = \frac{\frac{n-1}{n}S^2}{n}.$$

BCa Percentile Method

- The **bootstrap bias-corrected accelerated (BCa) interval** is a modification of the percentile method that adjusts the percentiles to correct for bias and skewness.

BCa Percentile Method (Reference Material)



$$P_*\left(U^* = \frac{g(\hat{\theta}^*) - g(\hat{\theta})}{1 + a \times g(\hat{\theta})} + z_0 \le z_\alpha\right) \approx 1 - \alpha$$

$$= P_*\left(\hat{\theta}^* \le g^{-1}(\, g(\hat{\theta}) \;+\; (1 \;+\; a \times g(\hat{\theta}))(z_\alpha - z_0)\,)\right) = P_*\left(\hat{\theta}^* \le \xi_\alpha\right).$$

$$P\left(U = \frac{g(\hat{\theta}) - g(\theta)}{1 + a \times g(\theta)} + z_0 \le z_\alpha\right) \approx 1 - \alpha$$

$$= P\left(\theta \ge g^{-1}\left(\frac{g(\hat{\theta}) - (z_\alpha - z_0)}{1 + a \times (z_\alpha - z_0)}\right)\right)$$

$$= P\left(\theta \ge g^{-1}(\, g(\hat{\theta}) + (1 + a \times g(\hat{\theta}))(z_{\beta_1} - z_0)\,)\right) = P\left(\theta \ge \xi_{\beta_1}\right).$$

$$\therefore \hat{\xi}_{\beta_1} = \hat{\theta}^*_{([(B+1)\times(1-\beta_1)])}.$$

Similarly, $P(\theta \le \hat{\theta}^*_{([(B+1)\times(1-\beta_2)])}) \approx 1 - \alpha$

and $P(\hat{\theta}^*_{([(B+1)\times(1-\beta_1)])} \;\le\; \theta \;\le\; \hat{\theta}^*_{([(B+1)\times(1-\beta_2)])}) \approx 1 - 2\alpha.$

BCa Percentile Method (Reference Material)

$$\beta_1 = ?$$

$$\beta_1 = 1 - P(Z_{\beta_1})$$

$$\text{and } \frac{g(\hat{\theta}) - (z_\alpha - z_0)}{1 + a \times (z_\alpha - z_0)} = g(\hat{\theta}) + (1 + a \times g(\hat{\theta})(z_{\beta_1} - z_0))$$

$$\Rightarrow z_{\beta_1} = z_0 + \frac{z_\alpha - z_0}{1 - a \times (z_\alpha - z_0)} \text{ and } \beta_1 = 1 - P(Z \le z_0 + \frac{z_\alpha - z_0}{1 - a \times (z_\alpha - z_0)})$$

$$\text{Similarly, } \beta_2 = 1 - P(Z \le z_0 + \frac{z_\alpha - z_0}{1 - a \times (z_\alpha - z_0)}).$$

BCa Percentile Method (Reference Material)

$$z_0 = ?$$

$$\begin{aligned} P_*(\hat{\theta}^* \le \hat{\theta}) &= P_*\left(g(\hat{\theta}^*) - g(\hat{\theta})\right) \\ &= P_*\left(\frac{g(\hat{\theta}^*) - g(\hat{\theta})}{1 + a \times g(\hat{\theta})} + z_0 \le \frac{g(\hat{\theta}) - g(\hat{\theta})}{1 + a \times g(\hat{\theta})} + z_0\right) \\ &= \Phi(z_0) \end{aligned}$$

$$\therefore z_0 = \Phi^{-1}\left(P_*(\hat{\theta}^* \le \hat{\theta})\right) \text{ and}$$

$$\hat{z}_0 = \Phi^{-1}\left(\frac{1}{B}\sum_{b=1}^{B} 1\left\{\hat{\theta}_b^* \le \hat{\theta}\right\}\right).$$

BCa Percentile Method (Reference Material)

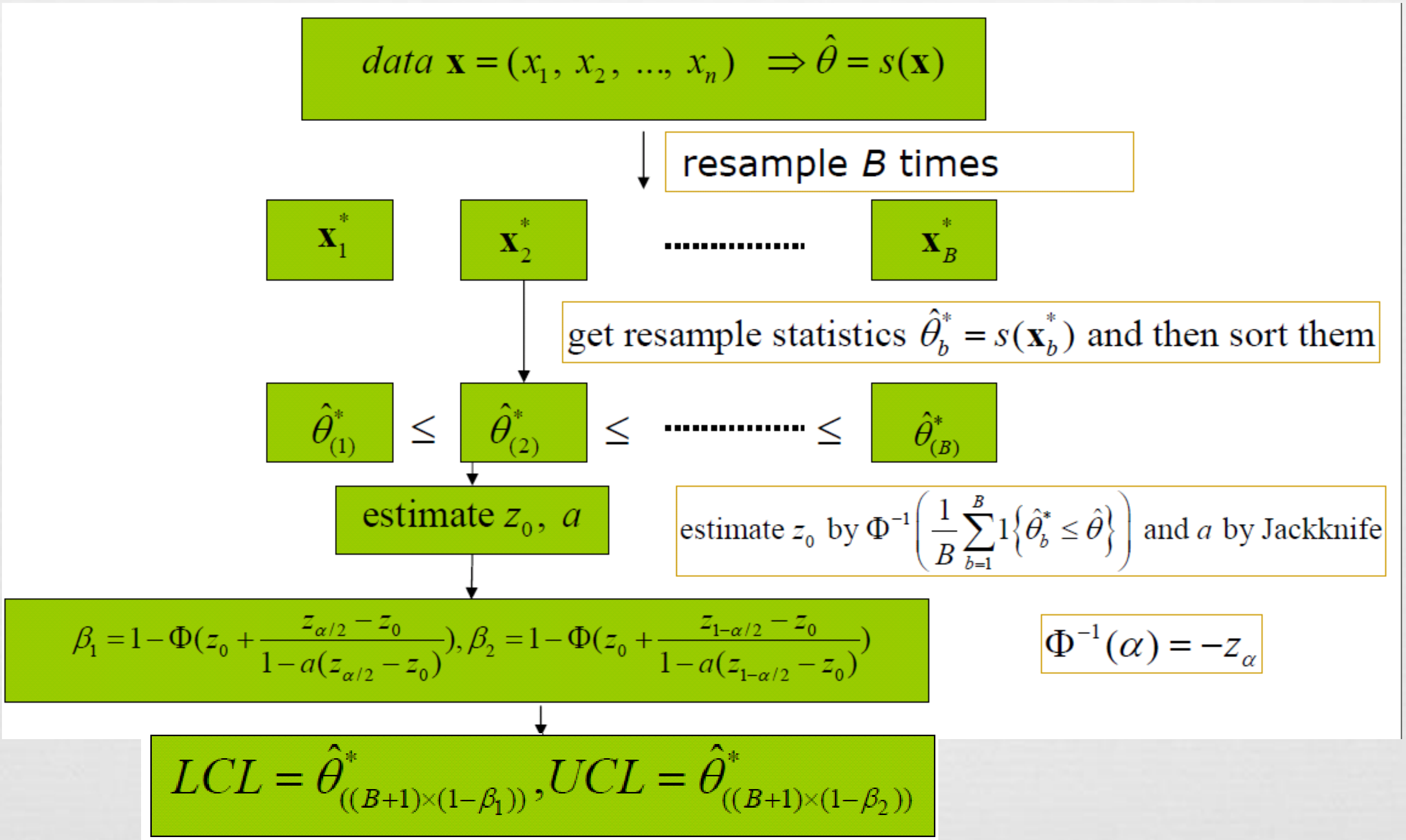
$a = ?$

$$\hat{a}_{Jack} = \frac{\sum_{i=1}^{n} (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6 \times (\sum_{i=1}^{n} (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2)^{3/2}},$$

where $\hat{\theta}_{(i)} = \theta(F_{n-1,\, i}) = \theta(\{X_1,\ \ldots,\ \cancel{X}_i\ ,\ \ldots, X_{\mathrm{n}}\})$

$$\mathit{and}\ \ \hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}.$$

BCa Percentile Method



$data\ \mathbf{x} = (x_1, x_2, \dots, x_n) \Rightarrow \hat{\theta} = s(\mathbf{x})$

resample *B* times

$\mathbf{x}_1^*$ $\mathbf{x}_2^*$ …… $\mathbf{x}_B^*$

get resample statistics $\hat{\theta}_b^* = s(\mathbf{x}_b^*)$ and then sort them

$\hat{\theta}_{(1)}^* \le \hat{\theta}_{(2)}^* \le \dots \le \hat{\theta}_{(B)}^*$

estimate z_0, a

estimate z_0 by $\Phi^{-1}\left(\frac{1}{B}\sum_{b=1}^{B} 1\left\{\hat{\theta}_b^* \le \hat{\theta}\right\}\right)$ and a by Jackknife

$$\beta_1 = 1 - \Phi\left(z_0 + \frac{z_{\alpha/2} - z_0}{1 - a(z_{\alpha/2} - z_0)}\right), \beta_2 = 1 - \Phi\left(z_0 + \frac{z_{1-\alpha/2} - z_0}{1 - a(z_{1-\alpha/2} - z_0)}\right)$$

$\Phi^{-1}(\alpha) = -z_\alpha$

$$LCL = \hat{\theta}^*_{((B+1)\times(1-\beta_1))}, UCL = \hat{\theta}^*_{((B+1)\times(1-\beta_2))}$$

Warnings on Non-Parametric Bootstrap Methods

- Does not require on to assume a statistical distribution underlying observed data.
- Observations come from a distribution with a finite variance.
- Observations are independent.
- Number of observations in the dataset is sufficiently large.

STATISTICAL INTERVALS
A Guide for Practitioners
Second Edition

To be published by John Wiley & Sons Inc. in 2015.

Conclusion

- Need to:
  - Review more papers
  - Keep working in R
  - Acquire Data
- Summer Plans
  - Masters written Exam
  - Research
  - Prepare for PhD Qualifier
  - Stay in touch with GH


Any questions?

References

- G. J. Hahn, W. Q. Meeker, and L. A. Escobar. STATISTICAL INTERVALS A Guide for Practitioners Second Edition Copyright 2014.To be published by John Wiley & Sons Inc. in 2015.
- PAGANO, M., and GAUVREAU, K. (1993) Principles of Biostatistics. Wadsworth Belmont California, USA.
- Shell Corporation: 2013 Sandia National Laboratories Wind Plant Reliability Workshop
- Wind energy statistics in Finland: http://www.vtt.fi/proj/windenergystatistics/?lang=en [accessed 06/11/2013].