WEEK 11, 2006
March 16 and 17

TMA4275 LIFETIME ANALYSIS

Bo Lindqvist

Department of Mathematical Sciences

NTNU

bo@math.ntnu.no

http://www.math.ntnu.no/~bo/

Censored and Truncated Data

- An observation is right censored at y:
  Unit is in our data, we know $T > y$.
  Contribution to L: $P(T > y) = R(y)$.

- An observation is left censored at y:
  Unit is in our data, we know $T < y$.
  Contribution to L: $P(T < y) = F(y)$.

- An observation is right truncated at y:
  Unit is in our data only if $T \leq y$. We do not know about the units with $T > y$.
  Contribution to L of observed failure at t:
  $\Delta^{-1} P(t \leq T \leq t + \Delta | T \leq y) \approx f(t)/F(y)$.

- An observation is left truncated at y:
  Unit is in our data only if $T \geq y$. We do not know about the units with $T < y$.
  Contribution to L of observed failure at t:
  $\Delta^{-1} P(t \leq T \leq t + \Delta | T \geq y) \approx f(t)/R(y)$.

Examples of left truncation:

- Ultrasonic inspection of material. Signal amplitude only trusted when above limit τ. Condition for being in the data set is $T > \tau$.

- Life data with pretest screening. Electronic component is burn-in tested for 1000 hours. Only the ones that passed this test are observed later. The number of components failing at burn-in is unknown. Condition for being in the data set is $T > 1000$.

Example of right truncation:

- Casting for automobile engine mounts. Pore size distribution below 10 microns only are recorded (other units are immediately discarded). Condition for being in the data set is $T < 10$ microns.

- Study group of individuals with AIDS diagnosis before July 1, 1986, and known date of HIV-infection (due to blood-transfusion). Let T_i = time from HIV-infection to AIDS diagnosis for ith individual. Then condition for being in the data set is that $T_i \leq v_i$ where v_i is time from HIV-infection of the ith individual to July 1, 1986. (Kalbfleisch and Lawless, 1989)

COMPUTER PROGRAM EXECUTION TIME vs SYSTEM LOAD

Data: 17 observations of (T,x)

- Time to complete a computationally intensive task.

- Information from the Unix uptime command

- Predictions needed for scheduling subsequent steps in a multi-step computational process.

Seconds (T)	Load (x)	Seconds (T)	Load (x)
123	2,74	110	,60
704	5,47	213	2,10
184	2,13	284	3,10
113	1,00	317	5,86
94	,32	142	1,18
76	,31	127	,57
78	,51	96	1,10
98	,29	111	1,89
240	,96		

Covariates (explanatory variables) for failure times

Useful covariates explain/predict why some units fail quickly and some units survive a long time:

- Continuous variables like stress, temperature, voltage, and pressure.
- Discrete variables like number of hardening treatments or number of simultaneous users of a system.
- Categorical variables like manufacturer, design, and location.

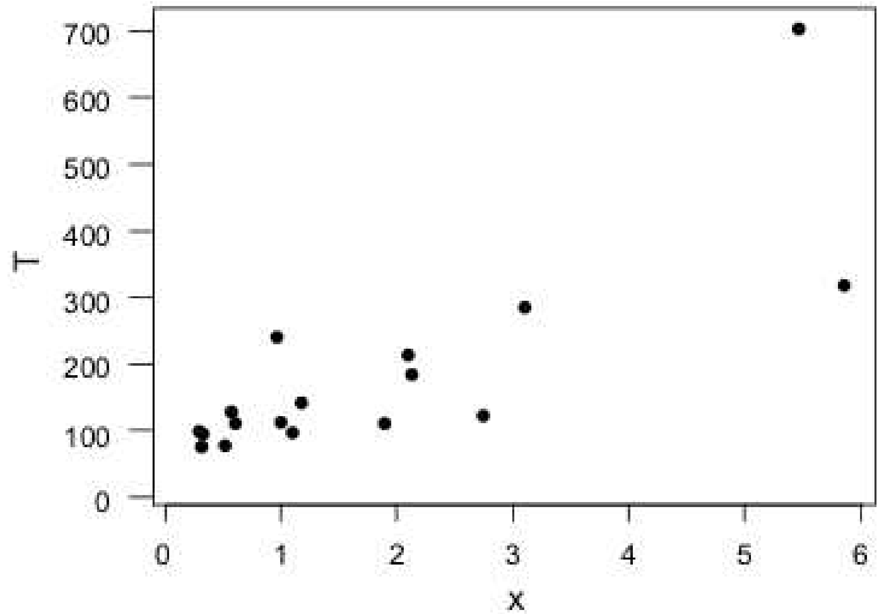
Regression model relates failure time distribution to covariates $x = (x_1, \ldots, x_k)$:

$$P(T \leq t) = F(t) = F(t; x)$$

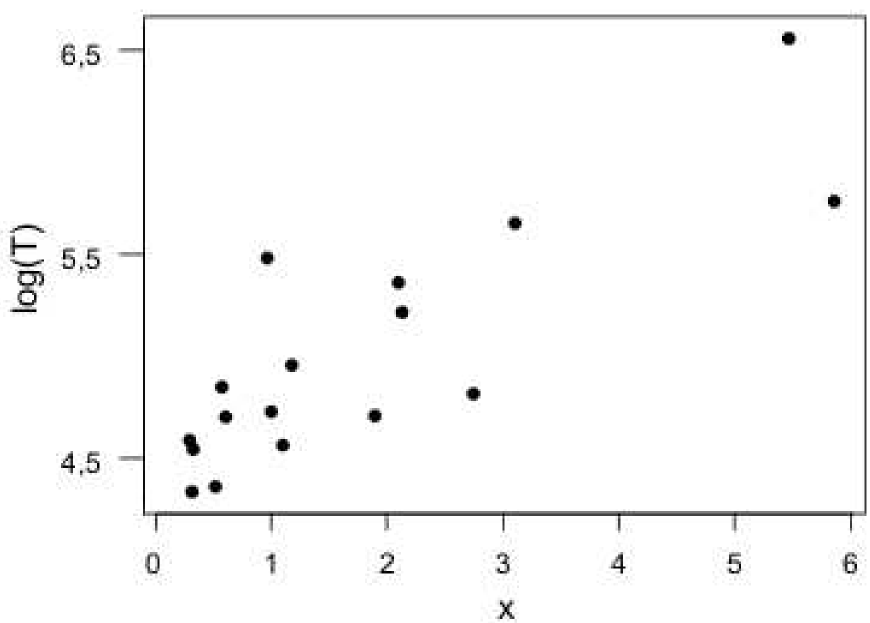
Why regression models?

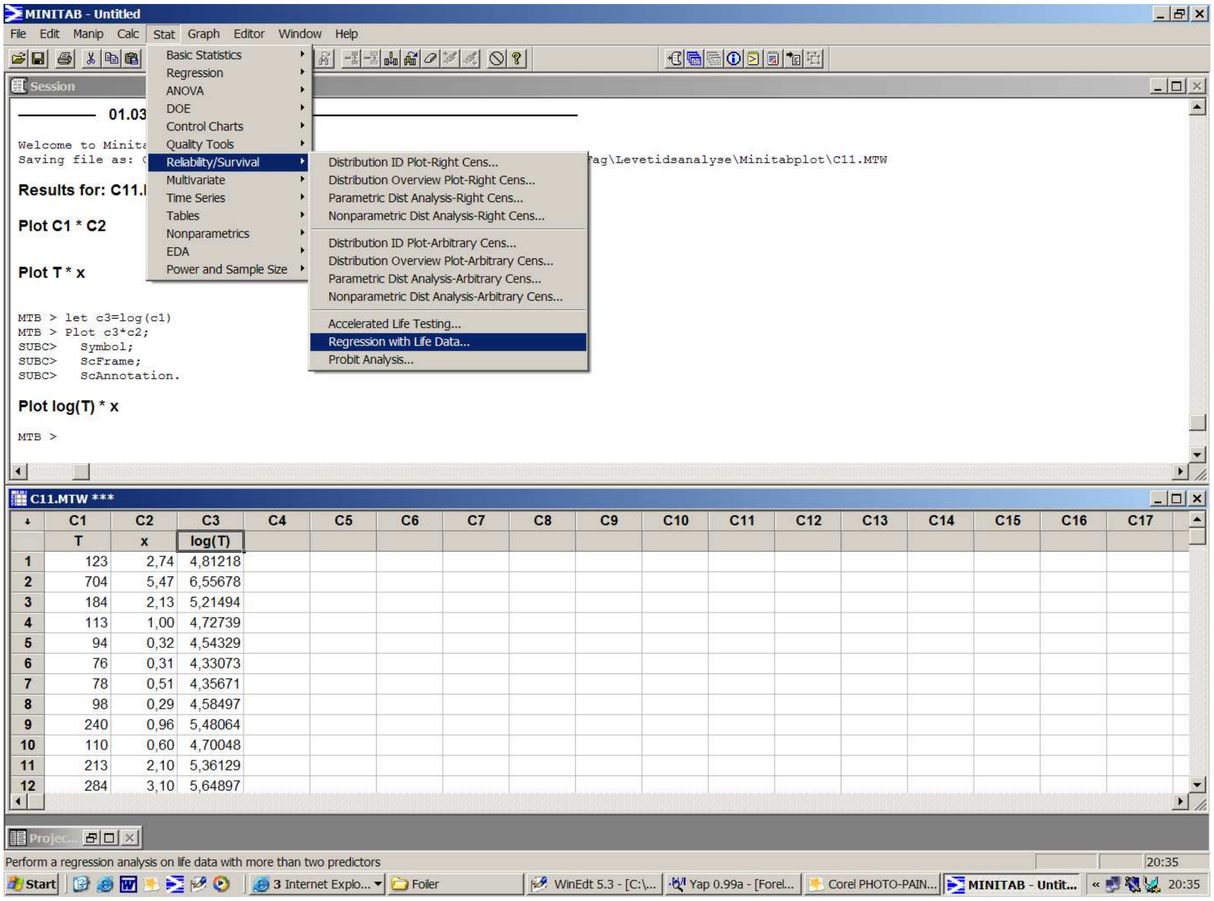
- Want to find factors which explain the reliability of an item
- Want to exclude factors which do not influence the reliability
- Obtain new knowledge about failure mechanisms
- Make better predictions for reliability of an item

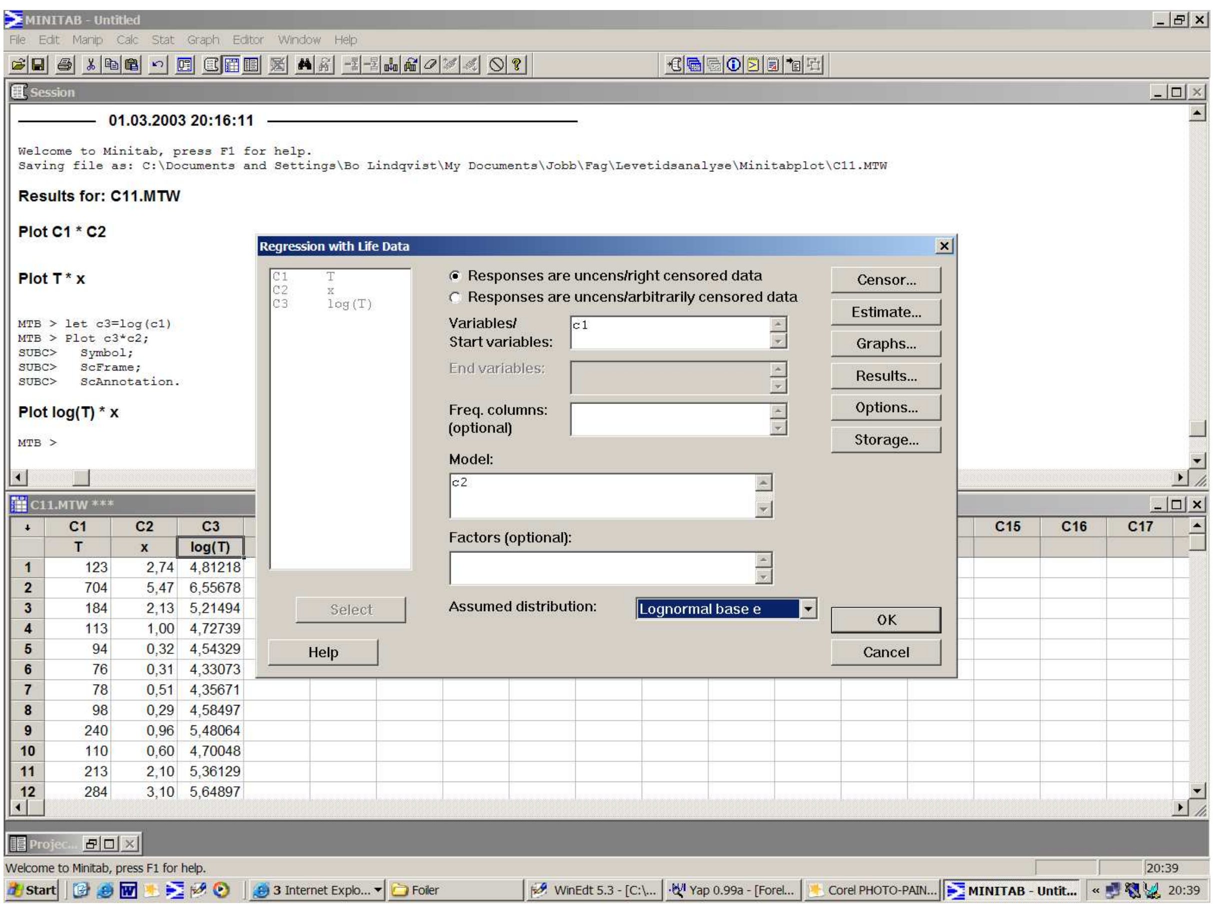
Computer data



Computer data




MINITAB - Untitled
File Edit Manip Calc Stat Graph Editor Window Help
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Distribution ID Plot-Right Cens...
Distribution Overview Plot-Right Cens...
Parametric Dist Analysis-Right Cens...
Nonparametric Dist Analysis-Right Cens...
Distribution ID Plot-Arbitrary Cens...
Distribution Overview Plot-Arbitrary Cens...
Parametric Dist Analysis-Arbitrary Cens...
Nonparametric Dist Analysis-Arbitrary Cens...
Accelerated Life Testing...
Regression with Life Data...
Probit Analysis...
Session
Plot C1 * C2
Plot T * x
MTB > let c3=log(c1)
MTB > Plot c3*c2;
SUBC> Symbol;
SUBC> ScFrame;
SUBC> ScAnnotation.
Plot log(T) * x
MTB >
C11.MTW ***
C1 T; C2 x; C3 log(T)
1 123 2,74 4,81218
2 704 5,47 6,55678
3 184 2,13 5,21494
4 113 1,00 4,72739
5 94 0,32 4,54329
6 76 0,31 4,33073
7 78 0,51 4,35671
8 98 0,29 4,58497
9 240 0,96 5,48064
10 110 0,60 4,70048
11 213 2,10 5,36129
12 284 3,10 5,64897
Perform a regression analysis on life data with more than two predictors


MINITAB - Untitled
01.03.2003 20:16:11
Welcome to Minitab, press F1 for help.
Saving file as: C:\Documents and Settings\Bo Lindqvist\My Documents\Jobb\Fag\Levetidsanalyse\Minitabplot\C11.MTW
Results for: C11.MTW
Plot C1 * C2
Plot T * x
Plot log(T) * x
Regression with Life Data
C1 T
C2 x
C3 log(T)
Responses are uncens/right censored data
Responses are uncens/arbitrarily censored data
Variables/
Start variables:
c1
End variables:
Freq. columns:
(optional)
Model:
c2
Factors (optional):
Assumed distribution:
Lognormal base e
Select
Help
Censor...
Estimate...
Graphs...
Results...
Options...
Storage...
OK
Cancel

Regression with Life Data: T versus x

```
Response Variable:  T

Censoring Information                 Count
Uncensored value                         17

Estimation Method:  Maximum Likelihood
Distribution:  Lognormal base e

Regression Table
                      Standard                          95,0% Normal CI
Predictor       Coef     Error        Z     P         Lower        Upper
Intercept     4,4936    0,1112    40,39 0,000        4,2756       4,7116
x            0,29075   0,04595     6,33 0,000       0,20069      0,38080
Scale        0,31247   0,05359                      0,22327      0,43730

Log-Likelihood = -89,498

Anderson-Darling (adjusted) Goodness-of-Fit

Standardized Residuals = 0,8356; Cox-Snell Residuals = 0,8170
```


Regression with Life Data: C1 versus C2

```
Response Variable:  C1

Censoring Information                 Count
Uncensored value                         17

Estimation Method:  Maximum Likelihood
Distribution:  Weibull

Regression Table
                      Standard                          95,0% Normal CI
Predictor       Coef     Error        Z     P         Lower        Upper
Intercept     4,6182    0,1219    37,88 0,000        4,3792       4,8572
C2           0,31118   0,04939     6,30 0,000       0,21437      0,40799
Shape         3,0604    0,5245                       2,1873       4,2820

Log-Likelihood = -91,504

Anderson-Darling (adjusted) Goodness-of-Fit
```

Likelihood for Lognormal Distribution Simple Regression Model with Right Censored Data

The likelihood for n independent observations has the form

$$\begin{aligned} L(\beta_0, \beta_1, \sigma) &= \prod_{i=1}^{n} L_i(\beta_0, \beta_1, \sigma; \text{data}_i) \\ &= \prod_{i=1}^{n} \left\{ \frac{1}{\sigma t_i} \phi_{\text{nor}} \left[\frac{\log(t_i) - \mu_i}{\sigma} \right] \right\}^{\delta_i} \left\{ 1 - \Phi_{\text{nor}} \left[\frac{\log(t_i) - \mu_i}{\sigma} \right] \right\}^{1-\delta_i} \end{aligned}$$

where $\text{data}_i = (x_i, t_i, \delta_i)$, $\mu_i = \beta_0 + \beta_1 x_i$,

$$\delta_i = \begin{cases} 1 & \text{exact observation} \\ 0 & \text{right censored observation} \end{cases}$$

$\phi_{\text{nor}}(z)$ is the standardized normal pdf and $\Phi_{\text{nor}}(z)$ is the corresponding normal cdf.

The parameters are $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma)$.

Estimated Parameter Variance-Covariance Matrix

Local (observed information) estimate

$$\begin{aligned} \widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}} &= \begin{bmatrix} \widehat{\text{Var}}(\widehat{\beta}_0) & \widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1) & \widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\sigma}) \\ \widehat{\text{Cov}}(\widehat{\beta}_1, \widehat{\beta}_0) & \widehat{\text{Var}}(\widehat{\beta}_1) & \widehat{\text{Cov}}(\widehat{\beta}_1, \widehat{\sigma}) \\ \widehat{\text{Cov}}(\widehat{\sigma}, \widehat{\beta}_0) & \widehat{\text{Cov}}(\widehat{\sigma}, \widehat{\beta}_1) & \widehat{\text{Var}}(\widehat{\sigma}) \end{bmatrix} \\ &= \begin{bmatrix} -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \beta_0^2} & -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \beta_0 \partial \sigma} \\ -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \beta_1 \partial \beta_0} & -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \beta_1^2} & -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \beta_1 \partial \sigma} \\ -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \sigma \partial \beta_0} & -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \sigma \partial \beta_1} & -\frac{\partial^2 \mathcal{L}(\beta_0, \beta_1, \sigma)}{\partial \sigma^2} \end{bmatrix}^{-1} \end{aligned}$$

Partial derivatives are evaluated at $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma}$.

Standard Errors and Confidence Intervals for Parameters

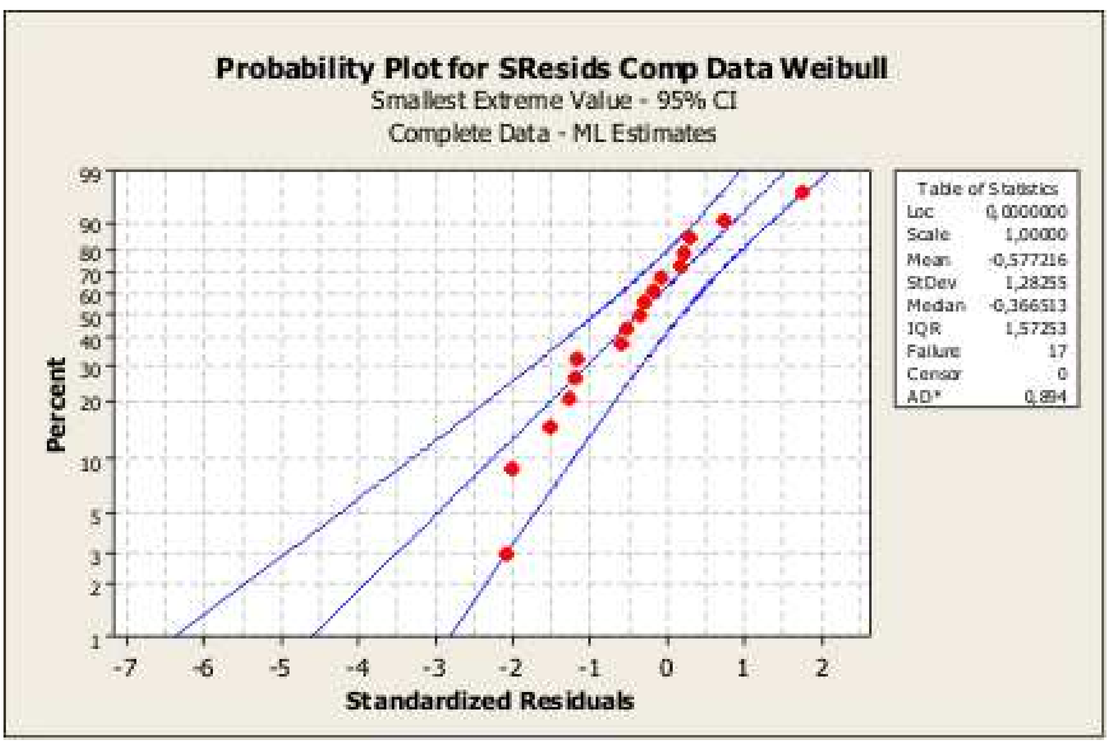
- Lognormal ML estimates for the computer time experiment were $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\sigma}) = (4.49, .290, .312)$ and an estimate of the variance-covariance matrix for $\widehat{\boldsymbol{\theta}}$ is

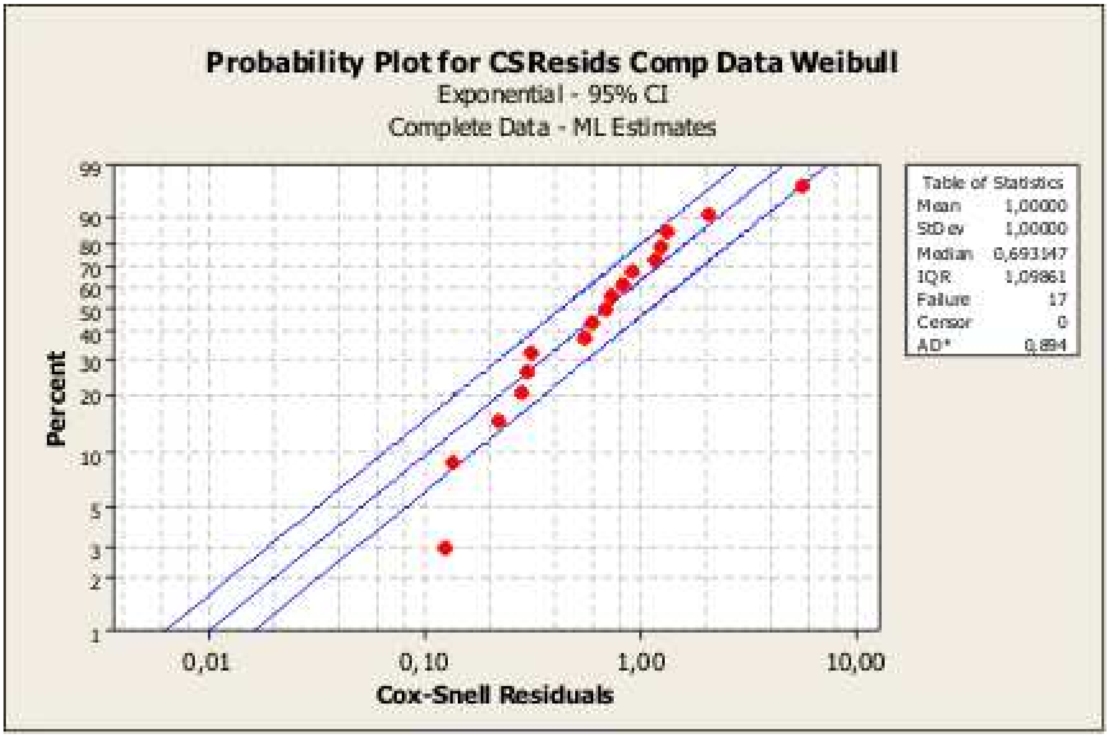
$$\widehat{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \begin{bmatrix} .012 & -.0037 & 0 \\ -.0037 & .0021 & 0 \\ 0 & 0 & .0029 \end{bmatrix}.$$

- Normal-approximation confidence interval for the computer execution time regression slope is

$$[\underset{\sim}{\beta_1}, \quad \widetilde{\beta}_1] = \widehat{\beta}_1 \pm z_{(.975)} \widehat{\mathrm{se}}_{\widehat{\beta}_1} = .290 \pm 1.96(.046) = [.20, \quad .38]$$

where $\widehat{\mathrm{se}}_{\widehat{\beta}_1} = \sqrt{.0021} = .046$.





Ordinary residuals

$$y_i - x'_i\hat{\beta}$$

where

yi is the ith response value

x'_i is the vector of predictor values associated with the ith response value

$\hat{\beta}$ represents the estimated regression coefficients

Standardized residuals

$$\frac{y_i - x'_i\hat{\beta}}{\hat{\sigma}}$$

where $\hat{\sigma}$ is the estimated scale parameter.

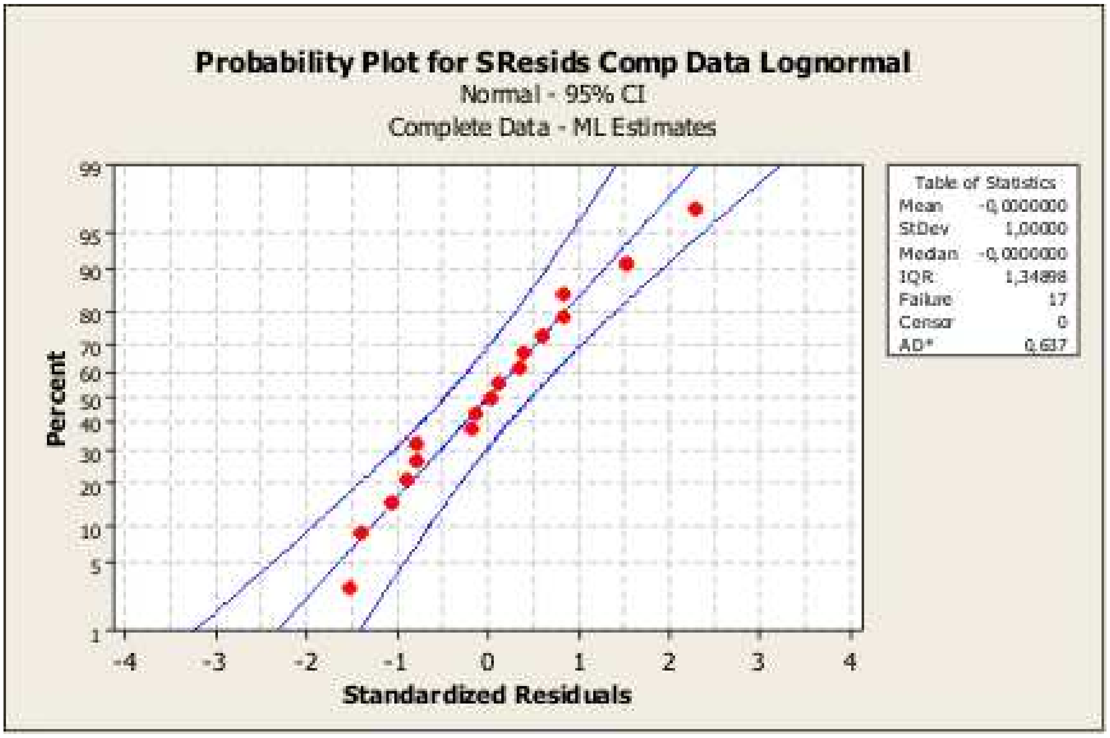
Cox-Snell residuals

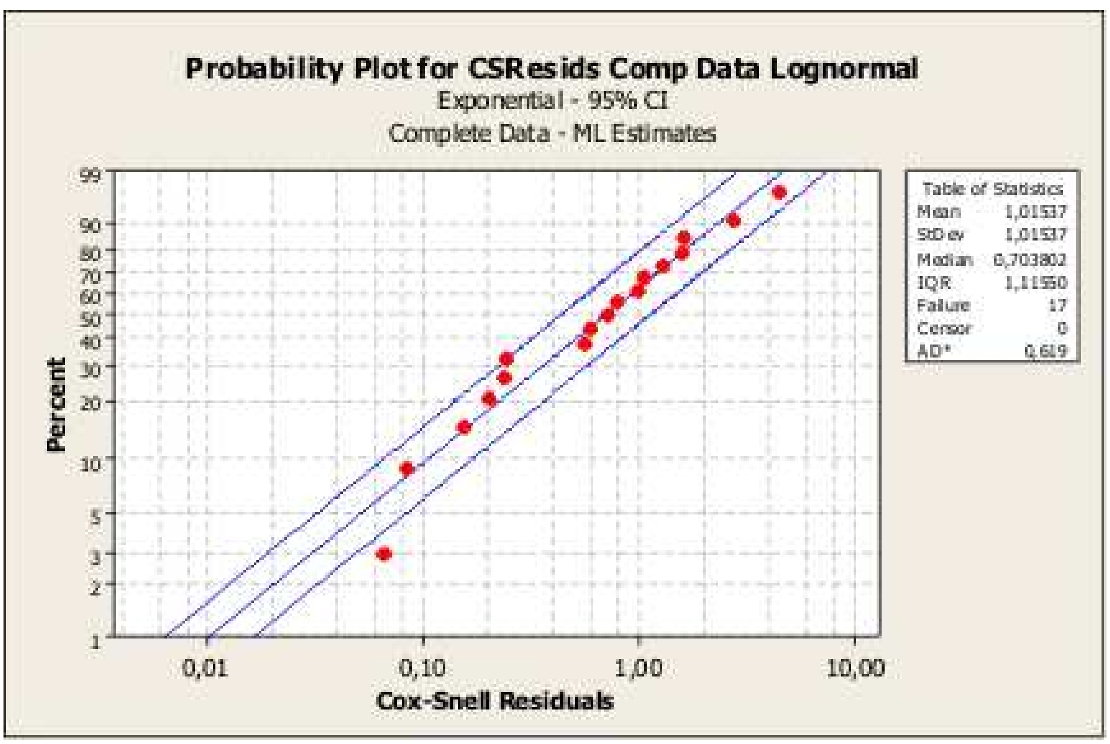
$$-\ln(\hat{R}(y_i))$$

where

$\hat{R}(y_i)$ is the estimated survival (reliability) probability for the response value yi

ln(x) is the natural log of x


Probability Plot for SResids Comp Data Lognormal
Normal - 95% CI
Complete Data - ML Estimates
Table of Statistics
Mean -0,0000000
StDev 1,00000
Median -0,0000000
IQR 1,34898
Failure 17
Censor 0
AD* 0,637
Percent
Standardized Residuals


Probability Plot for CSResids Comp Data Lognormal
Exponential - 95% CI
Complete Data - ML Estimates
Table of Statistics
Mean 1,01537
StDev 1,01537
Median 0,703802
IQR 1,11550
Failure 17
Censor 0
AD* 0,619
Percent
Cox-Snell Residuals

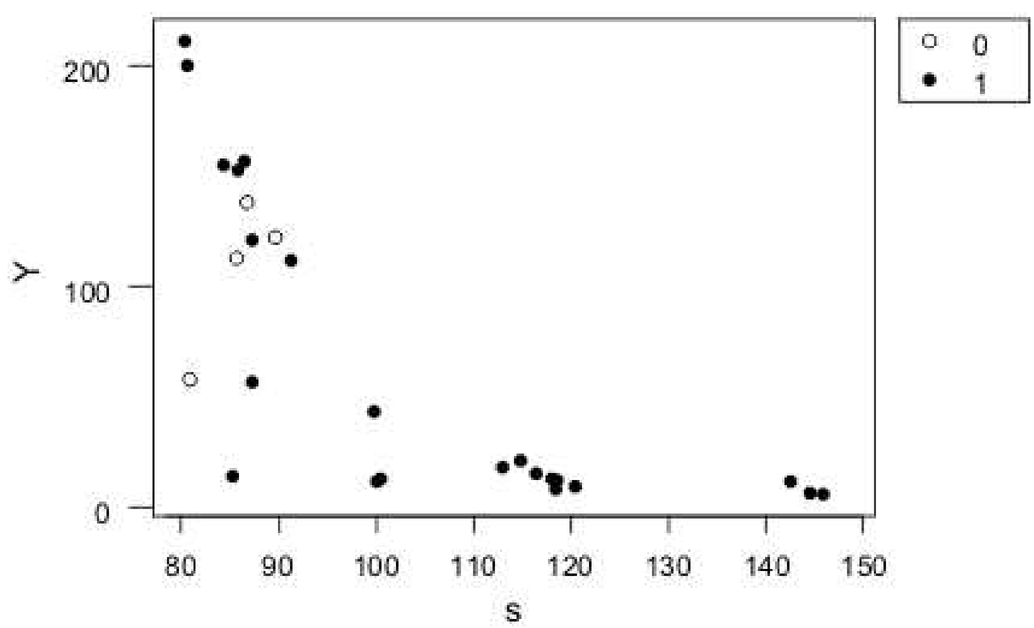
Row	Pseudo-stress	k-Cycles	Status (1=failed, 0=censored)
i	s	Y	C
1	80,3	211,629	1
2	80,6	200,027	1
3	80,8	57,923	0
4	84,3	155,000	1
5	85,2	13,949	1
6	85,6	112,968	0
7	85,8	152,680	1
8	86,4	156,725	1
9	86,7	138,114	0
10	87,2	56,723	1
11	87,3	121,075	1
12	89,7	122,372	0
13	91,3	112,002	1
14	99,8	43,331	1
15	100,1	12,076	1
16	100,5	13,181	1
17	113,0	18,067	1
18	114,8	21,300	1
19	116,4	15,616	1
20	118,0	13,030	1
21	118,4	8,489	1
22	118,6	12,434	1
23	120,4	9,750	1
24	142,5	11,865	1
25	144,5	6,705	1
26	145,9	5,733	1

DATA DESCRIPTION:
Low-Cycle Fatigue Life of Nickel-Base Superalloy Specimens (in units of thousands of cycles to failure).

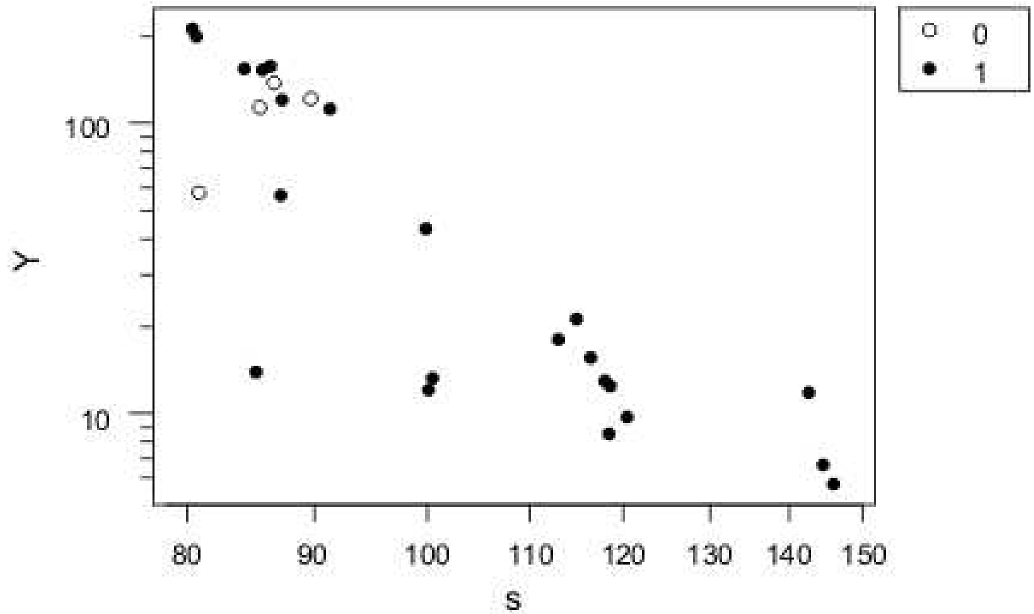
Data from Nelson (1990):

SUPER ALLOY DATA

Plot of Y vs s



Plot of log(Y) vs log(s)

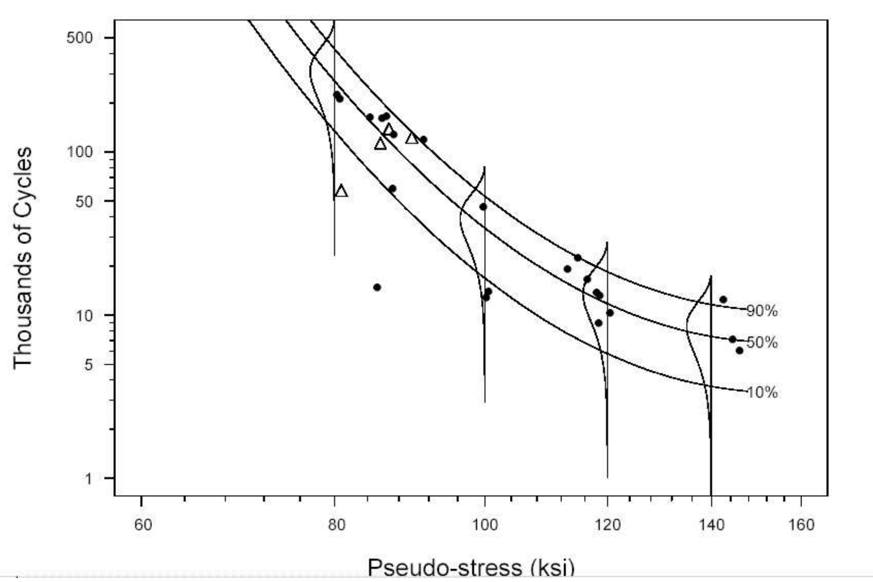



```
Regression with Life Data: Y versus x


Response Variable:  Y

Censoring Information                Count
Uncensored value                        22
Right censored value                     4
Censoring value:  C = 0

Estimation Method:  Maximum Likelihood
Distribution:  Weibull

Regression Table
                       Standard                           95,0% Normal CI
Predictor       Coef      Error        Z     P         Lower       Upper
Intercept     31,432      2,008    15,65 0,000        27,496      35,368
x            -5,9600     0,4329   -13,77 0,000       -6,8085     -5,1116
Shape         2,2105     0,3894                       1,5651      3,1221

Log-Likelihood = -97,155

Anderson-Darling (adjusted) Goodness-of-Fit

Standardized Residuals = 1,0768; Cox-Snell Residuals = 1,0768
```

```
Regression with Life Data: Y versus x

Response Variable:  Y

Censoring Information              Count
Uncensored value                      22
Right censored value                   4
Censoring value:  C = 0

Estimation Method:  Maximum Likelihood
Distribution:  Weibull

Regression Table
                       Standard                          95,0% Normal CI
Predictor      Coef       Error        Z      P        Lower        Upper
Intercept    217,61       62,13     3,50 0,000         95,83       339,39
x            -85,52       26,55    -3,22 0,001       -137,55       -33,49
x*x           8,483       2,831     3,00 0,003         2,934       14,032
Shape        2,6685      0,4777                       1,8789       3,7900

Log-Likelihood = -93,382

Anderson-Darling (adjusted) Goodness-of-Fit

Standardized Residuals = 0,9283; Cox-Snell Residuals = 0,9283
```


Log-Quadratic Weibull Regression Model with Constant ($\beta = 1/\sigma$) Fit to the Fatigue Data

$$\log[\widehat{t}_p(x)] = \widehat{\mu}(x) + \Phi_{\text{sev}}^{-1}(p)\widehat{\sigma},\ x = \log(\textbf{pseudo-stress})$$



Regression with Life Data: Y versus x

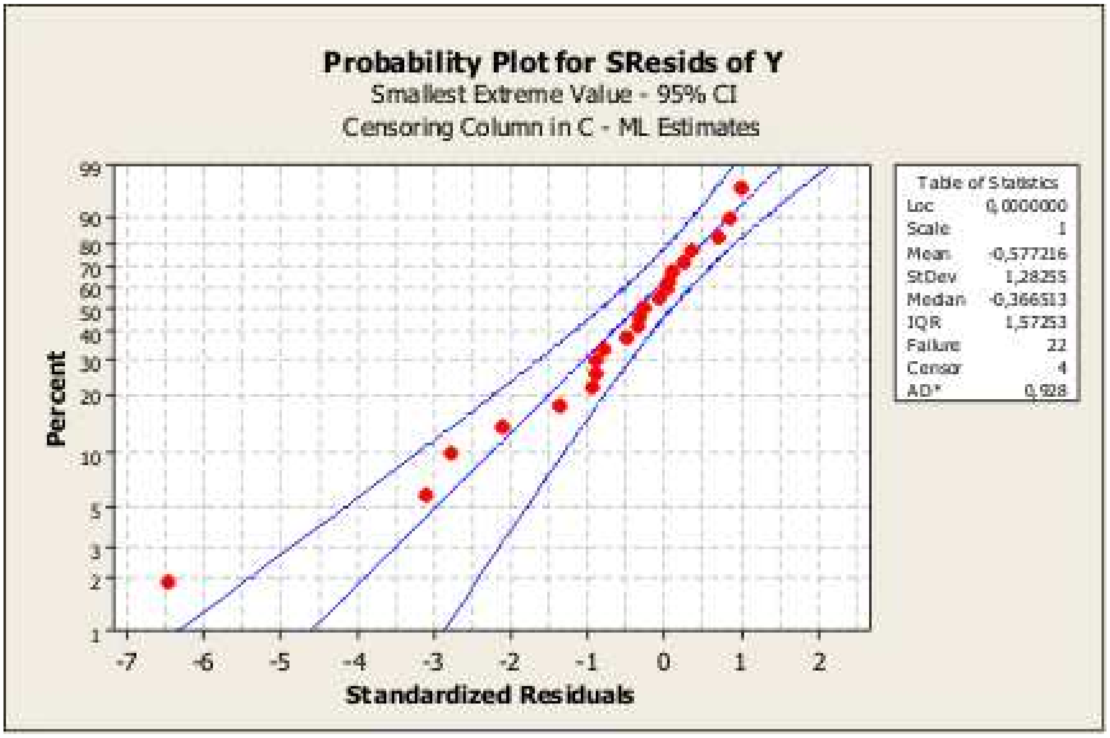
Response Variable: Y

Table of Percentiles

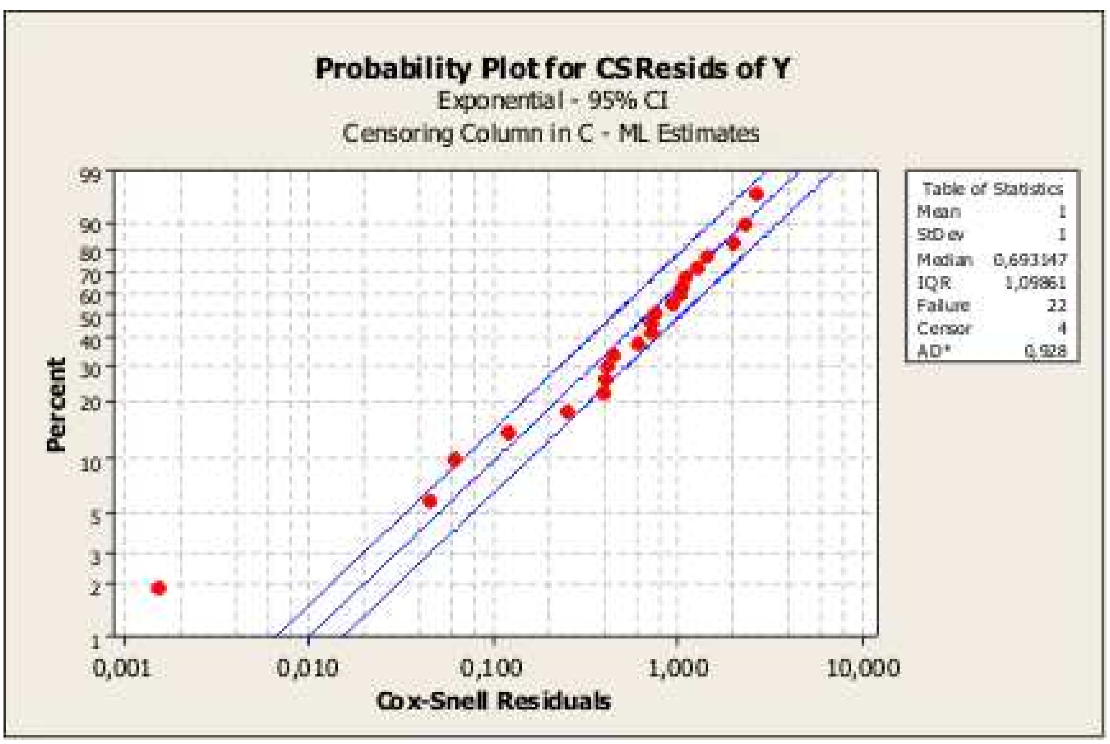
Percent	s	x	Percentile	Standard Error	95,0% Normal CI Lower	95,0% Normal CI Upper
10	80	4,3820	133,3747	34,0579	80,8565	220,0048
10	100	4,6052	16,7928	3,4263	11,2577	25,0494
10	120	4,7875	5,7830	1,2364	3,8034	8,7929
10	140	4,9416	3,6458	0,8760	2,2766	5,8386
50	80	4,3820	270,1879	56,0580	179,9121	405,7621
50	100	4,6052	34,0186	4,3027	26,5494	43,5891
50	120	4,7875	11,7151	1,5950	8,9713	15,2980
50	140	4,9416	7,3856	1,2828	5,2547	10,3807
90	80	4,3820	423,6933	90,4646	278,8097	643,8659
90	100	4,6052	53,3461	6,8162	41,5281	68,5272
90	120	4,7875	18,3709	2,4567	14,1351	23,8760
90	140	4,9416	11,5817	1,9813	8,2824	16,1952

ESTIMERT KOVARIANSMATRISE FOR $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma})$

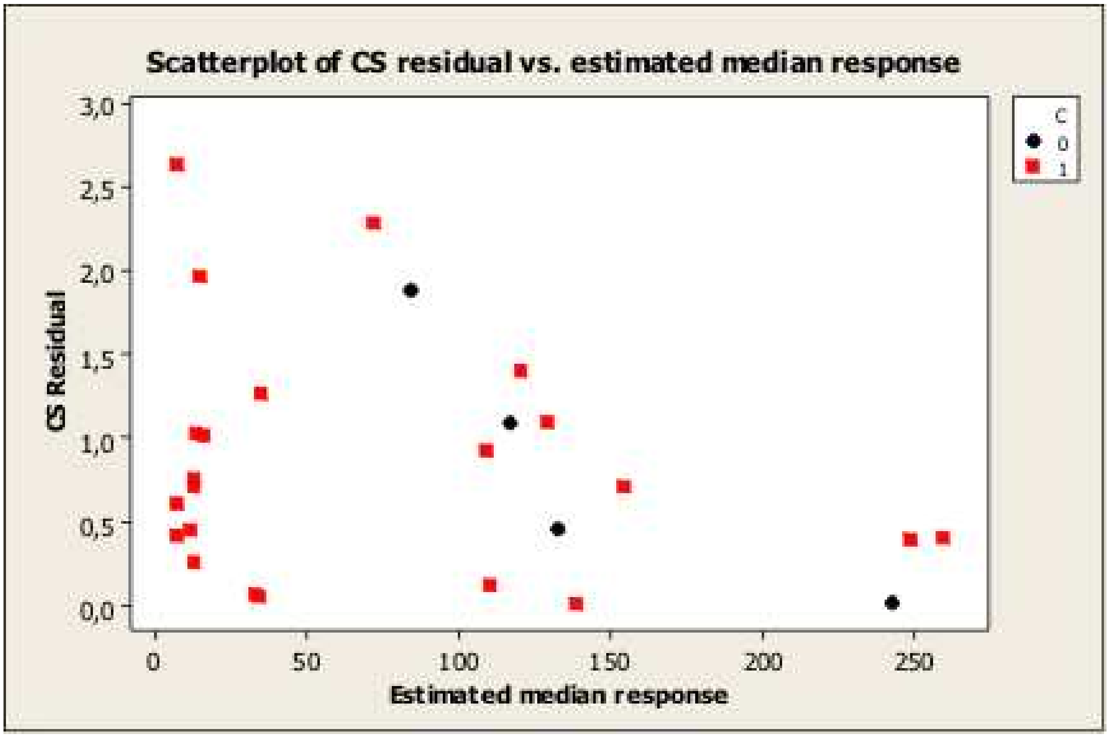
3860,37	-1649,17	175,82	-0,80
-1649,17	704,70	-75,15	0,33
175,82	-75,15	8,02	-0,03
-0,80	0,33	-0,03	0,23



Probability Plot for SResids of Y
Smallest Extreme Value - 95% CI
Censoring Column in C - ML Estimates
Percent
Standardized Residuals
Table of Statistics
Loc 0,0000000
Scale 1
Mean -0,577216
StDev 1,28255
Median -0,366513
IQR 1,57253
Failure 22
Censor 4
AD* 0,928



Probability Plot for CSResids of Y
Exponential - 95% CI
Censoring Column in C - ML Estimates
Percent
Cox-Snell Residuals
Table of Statistics
Mean 1
StDev 1
Median 0,693147
IQR 1,09861
Failure 22
Censor 4
AD* 0,928


Scatterplot of CS residual vs. estimated median response
CS Residual
3,0
2,5
2,0
1,5
1,0
0,5
0,0
0
50
100
150
200
250
Estimated median response
C
0
1