TMA4275 LIFETIME ANALYSIS

Slides 11: Lifetime regression - parametric models based on log-location-scale families

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

http://www.math.ntnu.no/∼bo/
bo@math.ntnu.no

NTNU, Spring 2015

CONTENTS OF SLIDES 11

- Censoring and **truncation**
  - Left truncation
  - Right truncation
- Analysis of lifetimes with covariates
  - Introductory example
  - Log-location-scale regression modeling
  - Maximum likelihood estimation
  - Residuals
    - Standard residuals
    - Cox-Snell residuals
    - Probability plots of residuals

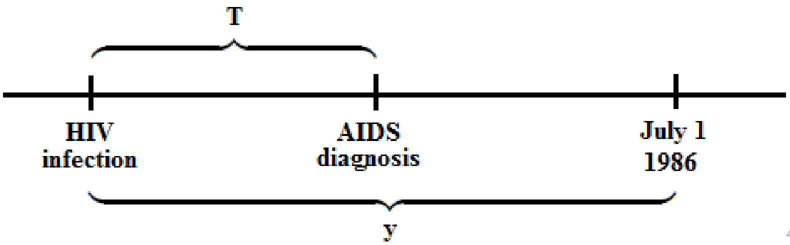
- An observation is **right censored** at y:
  Unit is in our data, we know $T > y$.
  Contribution to L: $P(T > y) = R(y)$.
- An observation is **left censored** at y:
  Unit is in our data, we know $T < y$.
  Contribution to L: $P(T < y) = F(y)$.
- An observation is **right truncated** at y:
  Unit is in our data only if $T \leq y$. We wouldn't know about this unit if $T > y$.
  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 wouldn't know about this unit if $T < y$.
  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 y. Condition for being in the data set is $T > y$.
- 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 $T > 1000$.
- In a medical study it may happen that subjects with a lifetime less than some threshold may not be observed at all. This is an example of **left truncation.** This is different from left censoring, since for a left censored observation, we know that the subject exists, but for a truncated case, we may be completely unaware of the subject.

  For example, in a follow-up health survey, a person who enters at age y and dies at age $T = t$, will be treated as a left-truncated observation since he/she was in our data only because $T \geq y$. The contribution to the likelihood is hence $f(t)/R(y)$.

- 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.
- In a study, one included individuals with AIDS diagnosis before July 1, 1986, and known date of HIV-infection (due to blood-transfusion). Let $T =$ time from HIV-infection to AIDS diagnosis for an individual. Then the condition for being in the data set is that $T \leq y$ where y is time from HIV-infection of the individual until July 1, 1986. Contribution to the likelihood is hence $f(t)/F(y)$ for an individual with $T = t$. (Kalbfleisch and Lawless, 1989)



Until now: Typically, n units observed, potential lifetimes

$$T_1, \ldots, T_n \text{ i.i.d. } \sim R(t)$$

and we have right censored data:

$$(y_i, \delta_i); \quad i = 1, \cdots, n$$

where y_i is observation time and δ_i is censoring status for ith unit.

Often there exist more information which may help explain the lifetime - called *covariates* or *explanatory variables*.

This means that data are (y_i, δ_i, x_i), where x_i gives the values of one or more covariates/explanatory variables.

EXAMPLE: COMPUTER PROGRAM - EXECUTION TIME vs SYSTEM LOAD

Computer data: 17 observations of the pair (T, x) (no censorings), where T is time to complete a computationally intensive task, x is information on load from the Unix uptime command.

Goal: Make 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 IN RELIABILITY

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 $\mathbf{x} = (x_1, \ldots, x_k)$:

$$P(T \le t) = F(t) = F(t; \mathbf{x})$$

WHY REGRESSION MODELS IN RELIABILITY?

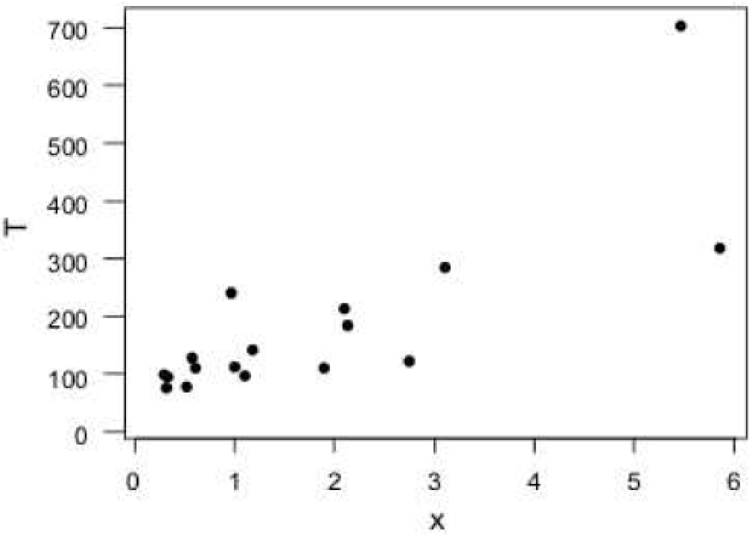
- Want to find factors which explain the reliability of an item
- Want to exclude factors which do not influence the reliability
- Obtain new knowledge about failure mechanisms
- Make better predictions for reliability of an item

EXPLANATORY VARIABLES IN MEDICAL RESEARCH

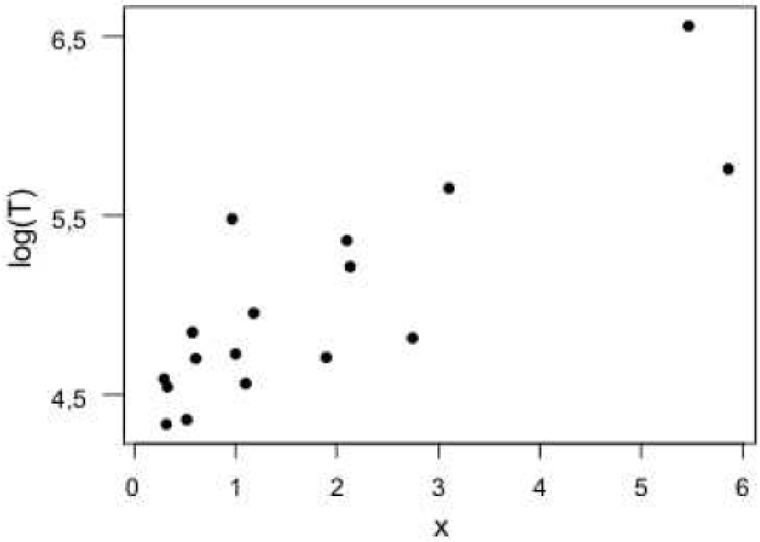
A typical medical example would include explanatory variables, often called prognostic factors or covariates, such as

- treatment assignment (e.g., *control* (placebo) or *treatment* by new medicine)
- patient characteristics such as
  - age at start of study
  - gender
  - presence of other diseases at start of study
  - blood measurements

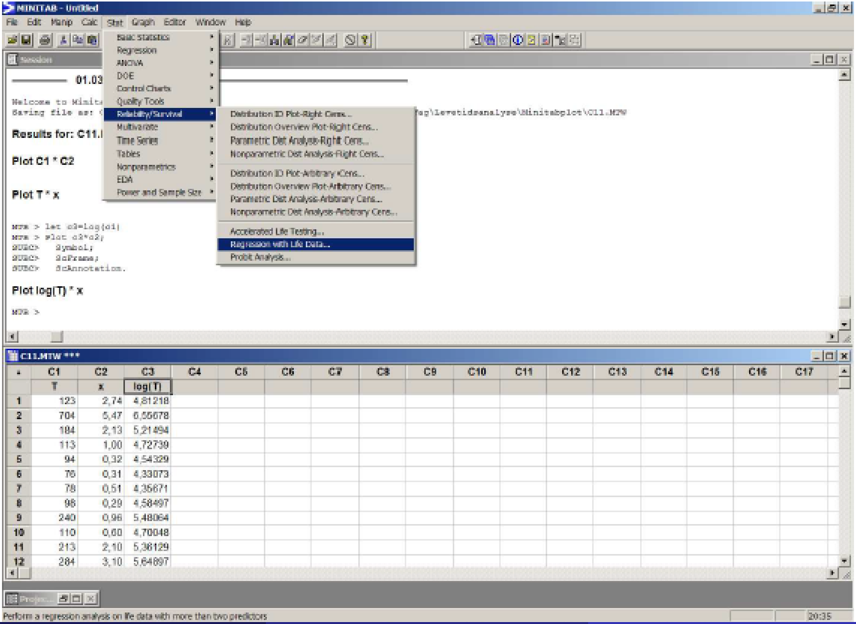
PLOT OF COMPUTER DATA - (x_i, T_i)



PLOT OF COMPUTER DATA - $(x_i, \ln T_i)$



COMPUTER DATA - MINITAB



First analysis of computer data:

Simple linear regression - well known from basic statistics courses:

$$T = \beta_0 + \beta_1 x + E, \quad \text{where } E \sim N(0, \sigma) \text{ (error)}$$

i.e. $T \sim N(\beta_0 + \beta_1 x, \sigma)$
Plots suggest that it's better to take log of the data as response:

$$\ln T = \beta_0 + \beta_1 x + E, \quad E \sim N(0, \sigma)$$

and we can, after taking log of all the lifetimes, use ordinary simple regression "as in basic course". Now

$$\ln T \sim N(\beta_0 + \beta_1 x, \sigma)$$

which means that

$$T \sim \text{lognormal}(\beta_0 + \beta_1 x, \sigma)$$

MINITAB does this for us in **Relibility** -> **Survival** -> **Regression**.

Can choose lognormal, Weibull, log-logistic etc.

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
```

$$\ln T = 4.4936 + 0.29075\, x + 0.31247\, Z, \quad Z \sim N(0,1)$$

$\hat{\beta}_0 = 4.4936$: "Intercept"
$\hat{\beta}_1 = 0.29075$: "x"
$\hat{\sigma} = 0.3147$: "Scale"

We could have done the above analysis by *ordinary linear regression* using $\ln T$ as response, since there was *no censoring*.

But **MINITAB** survival regression does more:

- can have other distributions than the normal
- can have censored observations

Model: For an observation unit with covariate value x (which we for simplicity first assumes is one-dimensional), the lifetime T can be represented as

$$\ln T = \beta_0 + \beta_1 x + \sigma U$$

where $U \sim N(0,1)$ for lognormal, or has another standard distribution for other families.

Recall for log-logation-scale families:

$$\ln T = \mu + \sigma U$$

The new feature is hence that μ depends on x. *Thus the data are no longer identically distributed.*

Data:

$$(y_1, \delta_1, x_1), (y_2, \delta_2, x_2), \ldots, (y_n, \delta_n, x_n)$$

Here (y_i, δ_i) are, as before, the observed time and censoring status for unit i. Now in addition we have information on a covariate value x_i for each unit.

Model:

- In $T_i = \beta_0 + \beta_1 x_i + \sigma U_i; i = 1, \cdots, n$, where $x_1, \cdots, x_n$ are the covariates, and $U_1, \cdots, U_n$ are i.i.d., e.g., $N(0, 1)$, Gumbel$(0, 1)$, Logistic$(0, 1)$, etc.
- There may also be right censoring, so we observe only $Y_i = \min(T_i, C_i)$, where C_i is a censoring time.

The extension from earlier implies estimation of β_0, β_1,σ instead of earlier μ, σ.

We need the density and survival function of an observation (T, x):

$$f(t; \beta_0, \beta_1, \sigma) = \psi\left(\frac{\ln t - \overbrace{(\beta_0 + \beta_1 x)}^{\mu \text{ before}}}{\sigma}\right)\frac{1}{\sigma t}$$

$$R(t; \beta_0, \beta_1, \sigma) = 1 - \Psi\left(\frac{\ln t - \overbrace{(\beta_0 + \beta_1 x)}^{\mu \text{ before}}}{\sigma}\right)$$

So likelihood is

$$L(\beta_0, \beta_1, \sigma) = \prod_{i:\delta_i=1} \psi\left(\frac{\log y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\frac{1}{\sigma y_i} \cdot \prod_{i:\delta_i=0} \left(1 - \Psi\left(\frac{\log y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)\right)$$

This is maximized w.r.t parameters, β_0, β_1, σ (MINITAB does it!)

$$I(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}) = \begin{bmatrix} -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_0^2} & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_0 \partial \sigma} \\ \cdot & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_1^2} & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \beta_1 \partial \sigma} \\ \cdot & \cdot & -\frac{\partial^2 \ell(\beta_0,\beta_1,\sigma)}{\partial \sigma^2} \end{bmatrix}$$

inserted the estimated parameters. Further,

$$I(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})^{-1} = \begin{bmatrix} \widehat{Var}\hat{\beta}_0 & \cdot & \cdot \\ \cdot & \widehat{Var}\hat{\beta}_1 & \cdot \\ \cdot & \cdot & \widehat{Var}\hat{\sigma} \end{bmatrix}$$

where as usual the entries outside the diagonal are estimated covariances.

Recall from MINITAB output that maximum likelihood estimates for lognormal model are:

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}) = (4.49, 0.290, 0.312)$$

The inverse observed information matrix turns out to be

$$\begin{bmatrix} 0.012 & -0.0037 & 0 \\ -0.0037 & 0.0021 & 0 \\ 0 & 0 & 0.0029 \end{bmatrix}$$

Thus SE for $\hat{\beta}_1$ is

$$\widehat{\text{SD}(\hat{\beta}_1)} = \sqrt{0.0021} = 0.046$$

and a standard confidence interval for β_1 is hence

$$0.29 \pm 1.96 \cdot 0.046 = [0.20, 0.38]$$

in accordance with the MINITAB output from earlier slide.

In general we may have more than one covariate.

$$\begin{aligned}\ln T &= \overbrace{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}^{\mu} + \sigma U \\ &= \beta_0 + \boldsymbol{\beta}' \mathbf{x} + \sigma U\end{aligned}$$

$$\text{where} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}; \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$$

With data from n units:
$(y_i, \delta_i, x_{1i}, \cdots, x_{ki})$ or $(y_i, \delta_i, \mathbf{x}_i)$ for $i = 1, 2, \ldots, n$. Lifetimes satisfy:

$$\begin{aligned}\ln T_i &= \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \sigma U_i \\ &= \beta_0 + \boldsymbol{\beta}' \mathbf{x}_i + \sigma U_i\end{aligned}$$

where $U_1, U_2, \ldots, U_n$ are i.i.d $\sim \Psi$. We can extend the observed information matrix to $(\beta_0, \cdots, \beta_k, \sigma)$

Recall model:

$$\ln T_i = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i + \sigma U_i$$

which implies

$$U_i = \frac{\ln T_i - \beta_0 - \boldsymbol{\beta}'\mathbf{x}_i}{\sigma}$$

Recall also that $U_1, U_2, \cdot, U_n$ are i.i.d $\sim \Psi$, and define *standardized residuals* (S-Residuals in MINITAB) by

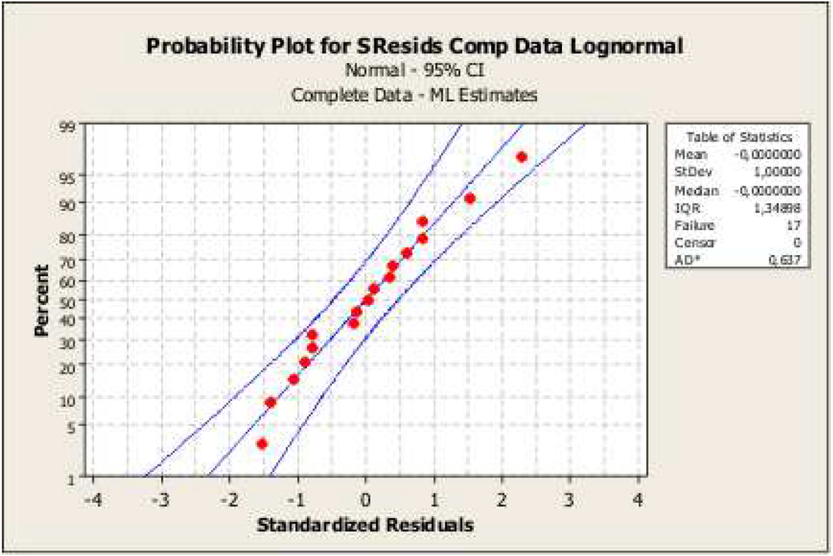
$$\hat{U}_i = \frac{\ln Y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}'\mathbf{x}_i}{\hat{\sigma}}$$

where Y_i are the *observed* times, either T_i or C_i.

Now the $\hat{U}_i$ should behave like a *right censored* set from the standard distribution, $N(0,1)$, Gumbel(0,1), Logistic(0,1) etc.

MINITAB plots the $\hat{U}_i$ in the ordinary probability plot for these distributions (**"Probability Plot for SResids"**)

STANDARD RESIDUALS: LOGNORMAL MODEL FOR COMPUTER DATA



Recall: If T has survival function $R(t)$ and cumulative hazard $Z(t)$, then $Z(T) = -\ln R(T) \sim \text{expon}(1)$

Application here: Since $\ln T_i = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i + \sigma U_i$, we have

$$R_{T_i}(t) = 1 - \Psi\left(\frac{\ln t - \beta_0 - \boldsymbol{\beta}'\mathbf{x}_i}{\sigma}\right)$$

and hence

$$V_i \equiv -\ln R_{T_i}(T_i) = -\ln\left[1 - \Psi\left(\frac{\ln T_i - \beta_0 - \boldsymbol{\beta}'\mathbf{x}_i}{\sigma}\right)\right] \sim \text{expon}(1)$$

Cox-Snell residuals are now defined as

$$\begin{aligned}\hat{V}_i &= -\ln\left[1 - \Psi\left(\frac{\ln Y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}'\mathbf{x}_i}{\hat{\sigma}}\right)\right] \\ &= -\ln\left[1 - \Psi\left(\text{standardized residuals}\right)\right]\end{aligned}$$

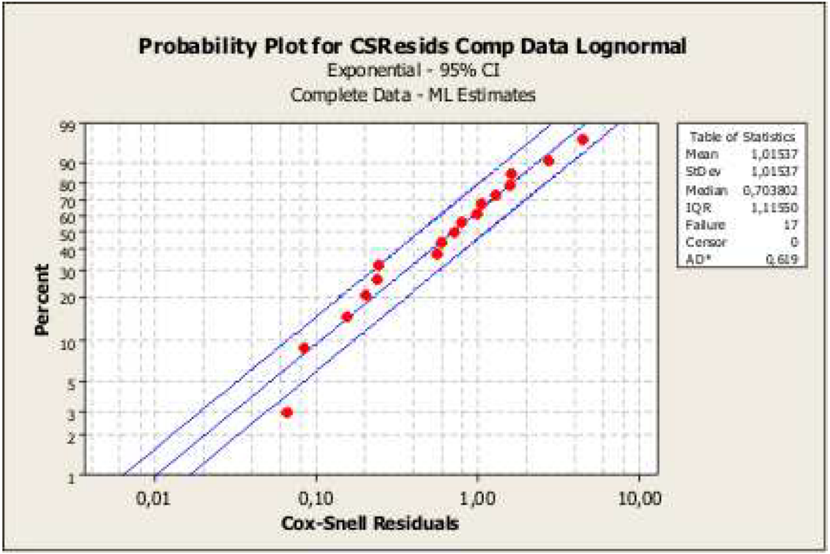
which should behave as a set of right-censored observations from expon(1) if the model is correctly specified.

Recall definition:

$$\hat{V}_i = -\ln\left[1 - \Psi\left(\frac{\ln Y_i - \hat{\beta}_0 - \hat{\beta}'\mathbf{x}_i}{\hat{\sigma}}\right)\right]$$
$$= -\ln\left[1 - \Psi\left(\text{standardized residuals}\right)\right]$$

- MINITAB puts the Cox-Snell residuals $\hat{V}_i$ into the usual *exponential* probability plot.("CSResids").
- Cox-Snell residuals are always exponentially distributed, while standardized residuals are distributed as the corresponding Ψ of the log-location-scale family.

COX-SNELL RESIDUALS: LOGNORMAL MODEL FOR COMPUTER DATA



Recall:

$$\hat{V}_i = -\ln\left[1 - \Psi\left(\frac{\ln Y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}'\mathbf{x}_i}{\hat{\sigma}}\right)\right] = -\ln\left[1 - \Psi(\hat{U}_i)\right]$$

where the $\hat{U}_i$ are the standardized residuals.

For the Weibull-distribution we have $\Psi(u) = G(u) = 1 - e^{-e^u}$, so

$$-\ln(1 - G(u)) = \ln(e^{-e^u}) = e^u$$

Thus Cox-Snell residuals are

$$\hat{V}_i = e^{\hat{U}_i} = e^{\text{Standardized residual}}$$

(Not so nice connection between SResid og CSResid for lognormal, e.g.)

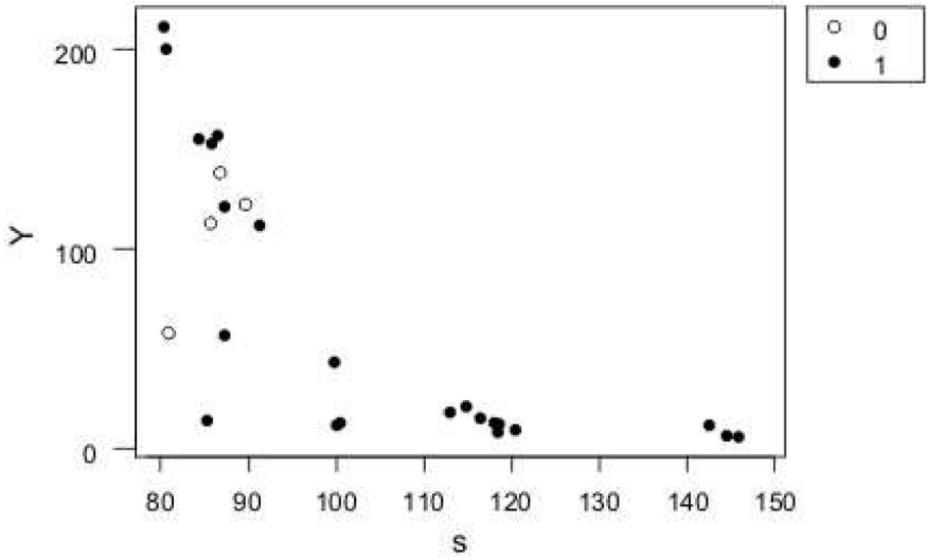
EXAMPLE: SUPERALLOY DATA (Nelson, 1990)

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

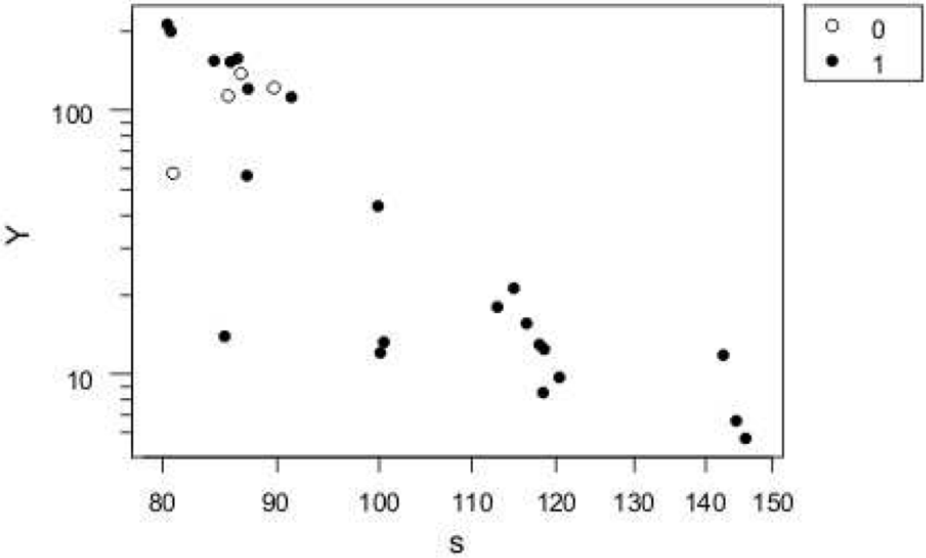
DATA DESCRIPTION:
Low-Cycle Fatigue Life of Nickel-Base Superalloy Specimens (in units of thousands of cycles to failure).

Data from Nelson (1990):

SUPER ALLOY DATA

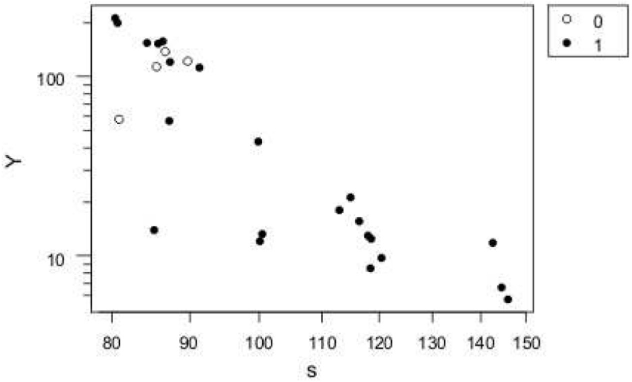

200
100
0
Y
80
90
100
110
120
130
140
150
s
0
1

SUPERALLOY DATA, PLOT $(\ln s, \ln Y)$



T = thousands of cycles to failure
s = pseudo-stress
$x = \ln s$ = log pseudo-stress.

Model 1: $\ln T = \beta_0 + \beta_1 x + \sigma W$, where W is standard Gumbel.



Model 1: $\ln T = \beta_0 + \beta_1 x + \sigma W,$

```
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
```

Model 2: $\ln T = \beta_0 + \beta_1 x + \beta_2 x^2 + \sigma W,$

```
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
```

Model 2 is an example of *polynomial* survival regression:

$$\ln T = \beta_0 + \beta_1 \overbrace{x}^{x_1} + \beta_2 \overbrace{x^2}^{x_2} + \sigma W$$

Recall values of log likelihoods:

- Model with x only:
  -97.155
- Model with x and x^2:
  -93.382

Thus 2(difference of log-likelihoods) $= 7.546$ (significant at $\alpha = 0.006$)

Model 1: $\ln T = 31.432 - 5.96 \ln s + \frac{1}{2.2105} W$

Model 2: $\ln T = 217.61 - 85.52 \ln s + 8.483 (\ln s)^2 + \frac{1}{2.6685} W$

$W \sim \text{Gumbel}(0, 1)$

Recall that for log-location-scale families:

$$\ln t_p = \mu + \sigma \Psi^{-1}(p)$$

Thus, for an individual with covariate vector **x**,

$$\ln t_p(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^{'}\mathbf{x} + \sigma \Psi^{-1}(p)$$

so for Weibull regression:

$$\ln t_p(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}^{'}\mathbf{x} + \frac{1}{\alpha}\ln(-\ln(1-p))$$

MINITAB computes these for given values of p and **x**

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

Example: $p = 0.90,\ s = 80,\ x = \ln s = 4.3820,\ x^2 = 4.3820^2,\ \ln(-\ln(1-0.90)) = 0.8340$

Then

$$\begin{aligned}\hat{t}_p(\mathbf{x}) &= e^{217.61 - 85.52 \cdot 4.3820 + 8.483 \cdot 4.3820^2 + \frac{1}{2.6685} \cdot 0.8340} \\ &= e^{6.0638} \\ &= 430.02\end{aligned}$$

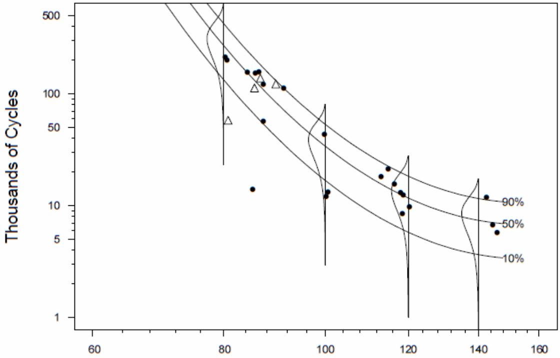
(MINITAB gives 423.6933, probably rounding errors?)

SUPERALLOY DATA, PLOT OF PERCENTILES (From Meeker and Escobar)

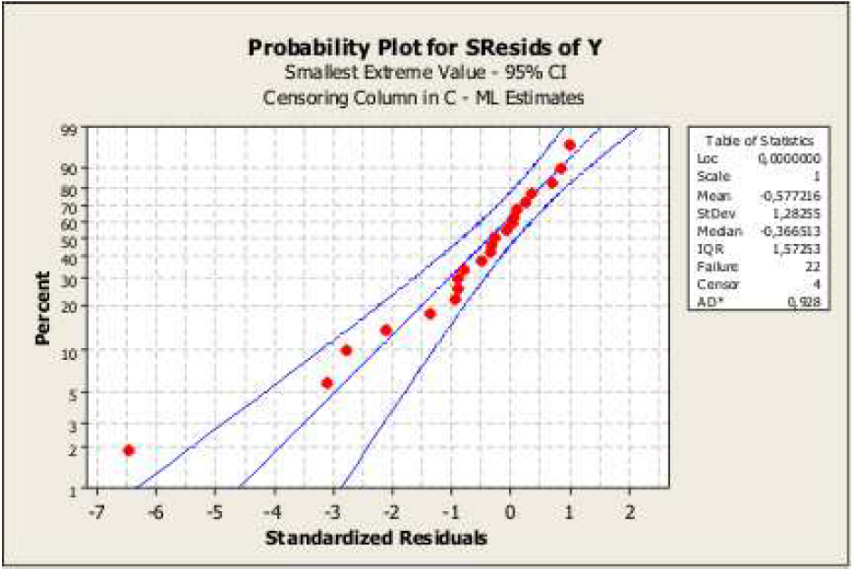
Log-Quadratic Weibull Regression Model with Constant ($\beta = 1/\sigma$) Fit to the Fatigue Data

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

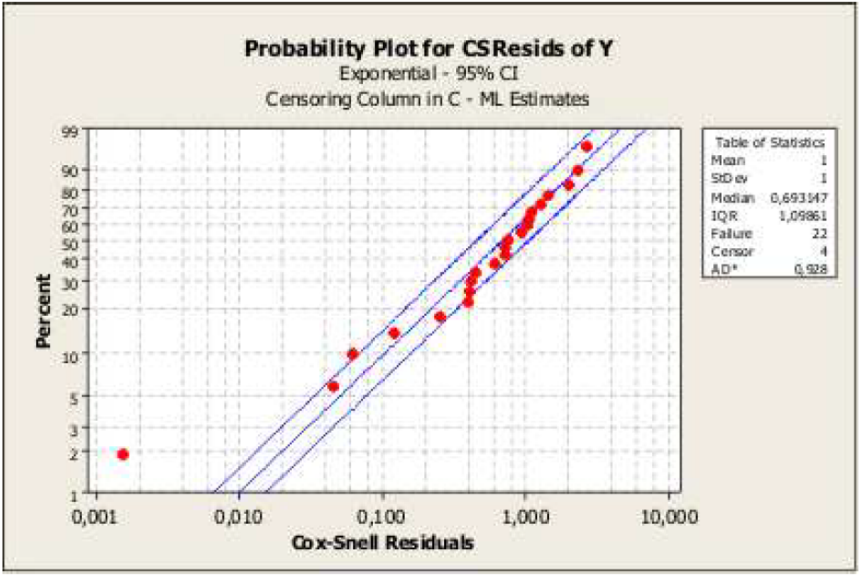
$$\widehat{\mu} = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 x^2$$

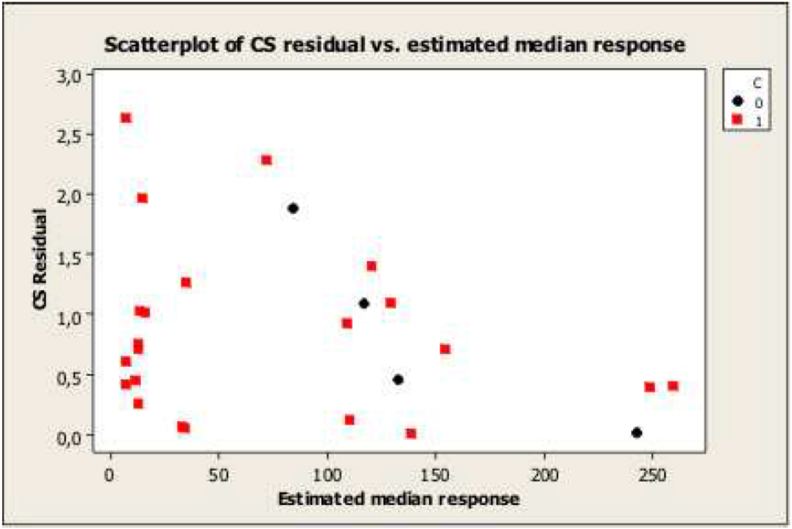


SUPERALLOY DATA, STANDARDIZED RESIDUALS



SUPERALLOY DATA, COX-SNELL RESIDUALS





The CS residuals should ideally be equally distributed for varying values of median response. Perhaps OK?