

SAS/STAT[®] 14.1 User's Guide

The QUANTLIFE

Procedure

This document is an individual chapter from *SAS/STAT® 14.1 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2015. *SAS/STAT® 14.1 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/STAT® 14.1 User's Guide

Copyright © 2015, SAS Institute Inc., Cary, NC, USA

All Rights Reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government License Rights; Restricted Rights: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication, or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a), and DFAR 227.7202-4, and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government's rights in Software and documentation shall be only those set forth in this Agreement.

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

July 2015

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Chapter 94

The QUANTLIFE Procedure

Contents

Overview: QUANTLIFE Procedure	7641
Features	7642
Quantile Regression	7642
Getting Started: QUANTLIFE Procedure	7643
Syntax: QUANTLIFE Procedure	7647
PROC QUANTLIFE Statement	7647
BASELINE Statement	7650
BY Statement	7651
CLASS Statement	7651
EFFECT Statement	7652
MODEL Statement	7653
OUTPUT Statement	7654
TEST Statement	7655
WEIGHT Statement	7655
Details: QUANTLIFE Procedure	7655
Notation for Censored Quantile Regression	7655
Kaplan-Meier-Type Estimator for Censored Quantile Regression	7656
Nelson-Aalen-Type Estimator for Censored Quantile Regression	7657
Relationship of Quantile Function and Survival Function	7657
Confidence Interval	7658
Output Data Sets	7659
ODS Table Names	7660
ODS Graphics	7660
Examples: QUANTLIFE Procedure	7661
Example 94.1: Primary Biliary Cirrhosis Study	7661
Example 94.2: Drug Abuse Study	7666
References	7671

Overview: QUANTLIFE Procedure

The QUANTLIFE procedure performs quantile regression analysis for survival data, in which observations are not always directly observed.

Quantile regression analysis (Koenker and Bassett 1978) is a type of regression analysis that explores how the conditional quantile of a response variable depends on its covariates. Recall that a student's score on a test is at the 0.85th quantile if his or her score is higher than that of 85% of the students who took the test. The score is also said to be at the 85th percentile. By estimating a set of conditional quantiles, you can gain more insight into the conditional distribution of the response, given its covariates.

Quantile regression provides a flexible way to capture heterogeneous effects in the sense that the tails and the central location of the conditional distributions can vary differently with the covariates. Thus, quantile regression offers a powerful tool in survival analysis, where the lifetimes are skewed and extreme survival times can be of special interest (Koenker and Geling 2001; Huang 2010).

When the observations are fully observed, you can use the QUANTREG procedure to fit a standard quantile regression model. For an introduction to the basic concepts of quantile regression analysis, see Chapter 95, “The QUANTREG Procedure.”

However, lifetime data often contain incomplete observations because of censoring (Klein and Moeschberger 2003; Hosmer, Lemeshow, and May 2008). When censoring occurs, the usual standard quantile regression approach can lead to biased estimates. Thus, special approaches have been developed that account for censoring and provide valid estimates. Portnoy (2003) proposed a method to estimate conditional quantile functions by generalizing the idea of the Kaplan-Meier estimator of the survival function. Peng and Huang (2008) developed a different quantile regression approach that is motivated by the Nelson-Aalen estimator of the cumulative hazard function. Both methods can be implemented using linear programming algorithms, and both are available in the QUANTLIFE procedure. Like the standard quantile regression method for uncensored data, these two methods are distribution-free and apply to heteroscedastic data.

Features

The QUANTLIFE procedure provides the following features:

- quantile regression methods for censored data that are based on generalizations of the Kaplan-Meier and the Nelson-Aalen estimator
- the interior point algorithm for parameter estimation, which uses parallel computing when multiple processors are available
- hypothesis tests for the regression parameter
- semiparametric quantile regression that uses spline effects
- survival plots, conditional quantile plots, and quantile process plots

Quantile Regression

Suppose that Y is a dependent variable of interest (such as the survival time or some monotone transformation of the survival time) and x is a $p \times 1$ vector of covariates. The observed data consist of a random sample $\{(Y_i, x_i), i = 1, \dots, n\}$ from (Y, x) .

You can use regression analysis to explore the relationship between the response Y and its predictor x . Classical linear regression estimates the conditional mean function $E(Y|x)$ by using a linear predictor $x'\beta$; a linear quantile regression estimates the τ th conditional quantile function $Q_Y(\tau|x)$ by using a different linear predictor $x'\beta(\tau)$, where the quantile level τ ranges between 0 and 1. For example, $x'\beta(0.95)$ is the linear predictor for the 0.95th quantile (commonly referred to as the 95th percentile).

The quantile regression coefficient $\beta(\tau)$ can be estimated by minimizing the following objective function over b :

$$r(b) = \sum_{i=1}^n \rho_{\tau}(Y_i - x_i' b)$$

The loss function $\rho_{\tau}(u)$ is defined as $u(\tau - I(u < 0))$, in contrast to the square loss function for classical linear regression.

When $\tau = 0.5$, the coefficient $\beta(0.5)$ minimizes the sum of absolute residuals, which corresponds to median regression (or L_1 regression).

The following set of regression quantiles is referred to as the *quantile process*, and it completely describes the conditional distribution of Y , given the predictor x :

$$\{\beta(\tau) : \tau \in (0, 1)\}$$

When all the observations are observed, you can use the QUANTREG procedure to estimate the quantile function $Q_Y(\tau|x)$ and draw statistical inference about the regression parameters $\beta(\tau)$. For more information, see Chapter 95, “The QUANTREG Procedure.”

However, when the observations are incomplete, as is the case with censored data in survival analysis, the classical quantile regression method is not appropriate. The QUANTLIFE procedure implements appropriate quantile regression methods to model the relationship between the response Y and the predictor x .

Getting Started: QUANTLIFE Procedure

This example uses the human immunodeficiency virus (HIV) study data from Hosmer and Lemeshow (1999) to illustrate the basic features of PROC QUANTLIFE.

In this study, subjects were followed after a confirmed diagnosis of HIV. The primary goal was to evaluate the effect of various factors on the survival time. Two covariates for each subject were collected: age and history of prior intravenous drug use.

The following DATA step creates the data set HIV, which contains the variable Time (the follow-up time in days), the variable Status (with value 0 if Time was censored and 1 otherwise), the variable Drug (with value 1 for prior intravenous drug use and 0 otherwise), and the variable Age (the patient's age in years at the beginning of the follow-up).

```
data HIV;
  input Time Age Drug Status;
  datalines;
    5      46      0      1
    6      35      1      0
    8      30      1      1
    3      30      1      1
```

```

22      36      0      1
1      32      1      0

... more lines ...

1      34      1      1
;

```

You can use PROC QUANTLIFE to explore the relationship between the survival time and the two covariates at different quantiles.

Suppose you are interested in the median survivors and in the longer and shorter survivors. The following statements fit a linear model for the 25th, 50th, and 75th percentiles:

```

ods graphics on;
proc quantlife data=hiv log plots=quantplot seed=1268;
  class Drug;
  model Time*Status(0) = Drug Age / quantile=(0.25 0.5 0.75);
  Drug_Effect: test Drug;
run;

```

The LOG option fits a quantile regression model for the log of Time, as is done by an accelerated failure time (AFT) model in standard survival analysis. The SEED= option maintains reproducibility of the resampling method that is used for statistical inference.

The MODEL statement specifies the response variable, Time, and the censoring variable, Censor. The value that indicates censoring is enclosed in parentheses. The values of Time are considered to be censored if the value of Censor is 0; otherwise, they are considered to be event times. The QUANTILE= option requests a fit of the conditional quantile function $Q_{\log(Y)}(\tau|x)$ at the quantile levels 0.25, 0.5, and 0.75.

The TEST statement requests a test for the hypothesis that there is no drug effect at each of the quantile levels.

Figure 94.1 displays basic model information. For example, you can see from Figure 94.1 that the response is log(Time) and the censoring rate is 20%.

Figure 94.1 Model Fitting Information

The QUANTLIFE Procedure		
Model Information		
Data Set	WORK.HIV	
Dependent Variable	Log(Time)	
Censoring Variable	Status	
Censoring Value(s)	0	
Number of Observations	100	
Method	Kaplan-Meier	
Replications	200	
Seed for Random Number Generator	1268	
Class Level Information		
Name	Levels	Values
Drug	2	0 1

Figure 94.1 *continued*

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
100	80	20	20.00

Figure 94.2 displays the parameter estimates, which are computed using the default Kaplan-Meier-type estimator. For more information, see the section “[Kaplan-Meier-Type Estimator for Censored Quantile Regression](#)” on page 7656. In addition, Figure 94.2 displays standard errors, 95% confidence limits, t values, and p -values that are computed by the default resampling method, exponentially weighted resampling. For more information, see the section “[Exponentially Weighted Method](#)” on page 7658.

A different quantile regression model is fitted for each quantile, and the first column (Quantile) in Figure 94.2 identifies the model for the parameter estimates. Age has a negative effect on survival time. You can use the parameter estimates to predict the survival time at the quantiles of interests. For example, the 75th percentile survival time for a person who has no previous intravenous drug use at age 46 is

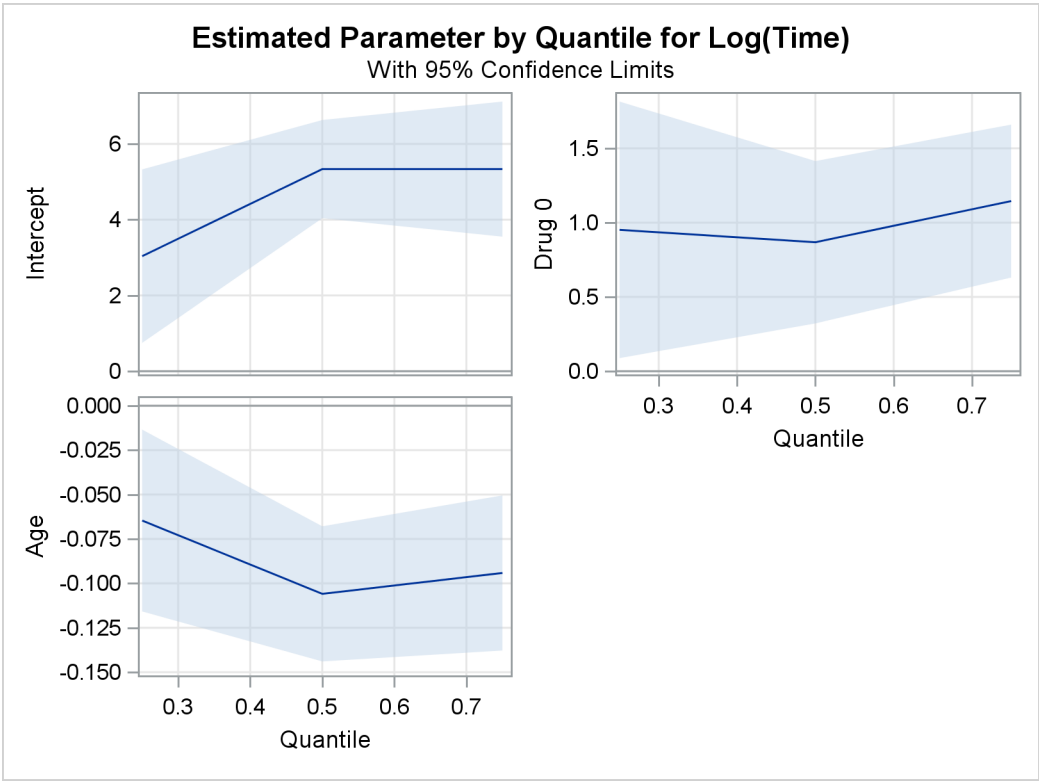
$$\exp(5.3351 + 1.1451 - 0.0941 \times 46) = 8.6 \text{ years}$$

Figure 94.2 Parameter Estimates

Parameter Estimates									
Quantile	Parameter	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > t	
0.2500	Intercept	1	3.0373	1.1680	0.7482	5.3265	2.60	0.0108	
	Drug	0 1	0.9516	0.4403	0.0887	1.8146	2.16	0.0331	
	Drug	1 0	0	0	0	0	.	.	
	Age	1	-0.0646	0.0261	-0.1158	-0.0135	-2.48	0.0150	
0.5000	Intercept	1	5.3351	0.6605	4.0406	6.6296	8.08	<.0001	
	Drug	0 1	0.8681	0.2786	0.3219	1.4142	3.12	0.0024	
	Drug	1 0	0	0	0	0	.	.	
	Age	1	-0.1059	0.0194	-0.1439	-0.0679	-5.46	<.0001	
0.7500	Intercept	1	5.3351	0.9091	3.5532	7.1170	5.87	<.0001	
	Drug	0 1	1.1451	0.2625	0.6307	1.6596	4.36	<.0001	
	Drug	1 0	0	0	0	0	.	.	
	Age	1	-0.0941	0.0223	-0.1378	-0.0505	-4.23	<.0001	

The PLOTS=QUANTPLOT option in the PROC QUANTLIFE statement requests the quantile process plots, which are shown in Figure 94.3. The quantile process plot is a scatter plot of an estimated regression parameter against the quantile level. You can use these plots to compare quantile-specific covariate effects. A curve that is not constant can indicate heterogeneity in the data. For example, Figure 94.3 shows that the drug effect does not change much across quantiles.

Figure 94.3 Estimated Parameters



The tests that are requested by the TEST statement are shown in Figure 94.4.

Figure 94.4 Tests of Significance

Test Drug_Effect Results			
Quantile	DF	Chi-Square	Pr > ChiSq
0.2500	1	4.67	0.0307
0.5000	1	9.70	0.0018
0.7500	1	19.03	<.0001

The tests indicate that the coefficient of Drug is significantly different from 0 at the 25th, 50th, and 75th percentiles.

Syntax: QUANTLIFE Procedure

The following statements are available in the QUANTLIFE procedure:

```

PROC QUANTLIFE < options > ;
  BASELINE < options > ;
  BY variables ;
  CLASS variables ;
  EFFECT name = effect-type ( variables < / options > ) ;
  MODEL response < * censor(list) > = < effects > < / options > ;
  OUTPUT < OUT=SAS-data-set > < keyword=name ... keyword=name > ;
  TEST effects < / options > ;
  WEIGHT variable ;

```

The PROC QUANTLIFE and MODEL statements are required. The PROC QUANTLIFE statement invokes the procedure. The CLASS statement specifies which explanatory variables are treated as categorical. The MODEL statement specifies the variables to be used in the regression. You can specify main effects and interaction terms in the MODEL statement, as you can in the GLM procedure (Chapter 46, “[The GLM Procedure](#)”). The OUTPUT statement creates an output data set to contain predicted values, residuals, and estimated standard errors. The TEST statement requests linear tests for the model parameters. The WEIGHT statement identifies a variable in the input data set whose values are used to weight the observations. In one invocation of PROC QUANTLIFE, multiple OUTPUT and TEST statements are allowed.

The rest of this section provides detailed syntax information for each statement, beginning with the PROC QUANTLIFE statement. The remaining statements are covered in alphabetical order.

PROC QUANTLIFE Statement

```
PROC QUANTLIFE < options > ;
```

The PROC QUANTLIFE statement invokes the QUANTLIFE procedure. [Table 94.1](#) summarizes the options available in this statement.

Table 94.1 Options Available in the PROC QUANTLIFE Statement

Option	Description
Data Set Options	
DATA=	Specifies the input SAS data set
OUTBOOTEST=	Creates an output SAS data set for parameter estimates from resampled data sets
Basic Options	
ALPHA=	Specifies the confidence level
CI=	Specifies a resampling method for computing confidence interval and test statistics
LOG	Requests log transformation of the response
METHOD=	Specifies a method to fit quantile regression
NAMELEN=	Specifies the length of effect names
NREP=	Specifies the number of replications

Table 94.1 *continued*

Option	Description
SEED=	Specifies the seed for the random number generator
PLOTS=	Specifies the plots to be produced by using ODS Graphics
Computational Options	
GRIDSIZE=	Specifies a step size for the grid for computing regression quantiles
INITTAU=	Specifies the first quantile level for computing regression quantiles
KAPPA=	Specifies the step-length parameter for the interior point algorithm
MAXIT=	Specifies the maximum number of iterations for the interior point algorithm
NTHREADS=	Specifies the number of threads for the computation
TOLERANCE=	Specifies the convergence criterion of the interior point algorithm

You can specify the following *options* in the PROC QUANTLIFE statement.

ALPHA=*value*

specifies the confidence level for the regression parameters. The *value* must be between 0 and 1. By default, ALPHA=0.05, which corresponds to a 95% confidence interval.

CI=EW | PW | NONE

requests the method for computing confidence intervals for regression parameters. In addition to confidence intervals, the QUANTLIFE procedure also computes standard errors, *t* values, and *p*-values for regression parameters. You can suppress these computations by specifying CI=NONE. The QUANTLIFE procedure provides two resampling methods for computing confidence intervals, the exponentially weighted (EW) method and the pairwise (PW) resampling method. For more information, see the section “[Confidence Interval](#)” on page 7658. By default, CI=EW, which requests the exponentially weighted method.

DATA=*SAS-data-set*

specifies the *SAS-data-set* that the QUANTLIFE procedure uses. By default, the most recently created *SAS-data-set* is used.

GRIDSIZE=*value*

specifies the step size for computing regression quantiles. The *value* must be between 0 and 1. For more information, see the section “[Details: QUANTLIFE Procedure](#)” on page 7655.

INITTAU=*value*

specifies the first quantile level for computing regression quantiles. The *value* must be between 0 and 1. For more information, see the section “[Details: QUANTLIFE Procedure](#)” on page 7655.

KAPPA=*value*

specifies the step-length parameter for the interior point algorithm. The *value* must be between 0 and 1. The interior point method that the QUANTLIFE procedure uses is identical to the interior point method that the QUANTREG procedure uses. For more information, see the section “[Interior Point Algorithm](#)” on page 7706 in Chapter 95, “[The QUANTREG Procedure](#).” By default, KAPPA=0.99995.

LOG

requests that a log transformation of the response variable be performed before the model is fitted.

MAXIT=*n*

specifies the maximum number of iterations for the interior point algorithm. By default, MAXIT=1000.

METHOD=KM | NA

requests the method to use to estimate the regression parameters. METHOD=KM specifies the Kaplan-Meier-type method (see the section “[Kaplan-Meier-Type Estimator for Censored Quantile Regression](#)” on page 7656) and METHOD=NA specifies the Nelson-Aalen-type method (see the section “[Nelson-Aalen-Type Estimator for Censored Quantile Regression](#)” on page 7657). By default, METHOD=KM.

NAMELEN=*n*

specifies the length of effect names in tables and output data sets to be *n* characters, where *n* is a value between 20 and 200. By default, NAMELEN=20.

NREP=*n*

specifies the number of replications to draw in the resampling method. By default, NREP=200.

NTHREADS=*n*

specifies the number of threads for analytic computations and overrides the SAS system option THREADS | NOTTHREADS. If you do not specify the NTHREADS= option or if you specify NTHREADS=0, the number of threads is determined based on the data size and the number of CPUs on the host where the analytic computations execute.

OUTBOOTEST=SAS-data-set

creates a data set to contain the parameter estimates from the resampled data sets. For a detailed description of the contents of the OUTBOOTEST= data set, see the section “[OUTBOOTEST= Output Data Set](#)” on page 7659.

PLOTS =(plot-request < ...plot-request >)

requests various plots.

When you specify one *plot-request*, you can omit the parentheses around it.

ODS Graphics must be enabled before plots can be requested. For example:

```
ods graphics on;

proc quantlife plots=survival;
    model y=x1;
run;

ods graphics off;
```

For more information about enabling and disabling ODS Graphics, see the section “[Enabling and Disabling ODS Graphics](#)” on page 609 in Chapter 21, “[Statistical Graphics Using ODS](#).”

You can specify one or more of the following *plot-requests*:

ALL

creates all appropriate plots.

NONE

suppresses all the plots in the procedure. Specifying this option is equivalent to disabling ODS Graphics for the entire procedure.

QUANTILE

plots the estimated quantile function for each combination of covariate values in the COVARIATES= data set that is specified in the BASELINE statement. If the COVARIATES= data set is not specified, then the estimated quantile function is plotted for the reference set of covariate values, which consists of reference levels for the CLASS variables and average values for the continuous variables. When the estimated quantile function is not monotonic, the quantile function (Chernozhukov, Fernandez-Val, and Galichon 2009) is rearranged to make it monotonic and then plotted.

QUANTPLOT < / UNPACK >

plots the regression quantile process. The estimated coefficient of each specified covariate effect is plotted as a function of the quantile level. You can use the UNPACK option to create individual process plots.

SURVIVAL

plots the estimated survival function for each combination of covariate values in the COVARIATES= data set that is specified in the BASELINE statement. If the COVARIATES= data set is not specified, then the estimated survival function is plotted for the reference set of covariate values, which consists of reference levels for the CLASS variables and average values for the continuous variables.

SEED=number

specifies a positive integer to start the pseudorandom number generator. The default is a value that is generated from reading the time of day from the computer's clock. However, to duplicate the results under identical situations, you must specify the same seed in subsequent runs of the QUANTLIFE procedure. The seed information is displayed in the "Model Information" table.

TOLERANCE=value

specifies the tolerance for the convergence criterion of the interior point algorithm. Both the QUANTLIFE procedure and the QUANTREG procedure use the duality gap as the convergence criterion. For more information, see Chapter 95.13, "[Interior Point Algorithm](#)." By default, TOLERANCE=1E-8.

BASELINE Statement

BASELINE < OUT=SAS-data-set > < COVARIATES=SAS-data-set > ;

The BASELINE statement creates an output data set to contain the survival function estimates or the conditional quantile function estimates for every set of covariates (x) in the COVARIATES= data set. If the COVARIATES= data set is not specified, PROC QUANTLIFE uses a reference set of covariates that consists of the reference levels for the CLASS variables and the average values for the continuous variables.

You can specify the following *options*:

COVARIATES=SAS-data-set

names the *SAS-data-set* that contains the sets of explanatory variable values for which the quantities of interest are estimated. All variables in the COVARIATES= data set are copied to the OUT= data set. Thus, the variables in the COVARIATES= data set can be used to identify the covariate sets in the OUT= data set.

OUT=SAS-data-set

specifies an output *SAS-data-set* that contains the estimated survival function and the estimated quantile function. If you omit the OUT= option, the data set is created and given a default name by using the DATA n convention. For more information, see the section “[OUT= Output Data Set in the BASELINE Statement](#)” on page 7659.

BY Statement

BY variables ;

You can specify a BY statement with PROC QUANTLIFE to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the QUANTLIFE procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

CLASS Statement

CLASS variables </ TRUNCATE > ;

The CLASS statement names the classification variables to be used in the model. Typical classification variables are Treatment, Sex, Race, Group, and Replication. If you use the CLASS statement, it must appear before the [MODEL](#) statement.

Classification variables can be either character or numeric. By default, class levels are determined from the entire set of formatted values of the CLASS variables.

In any case, you can use formats to group values into levels. See the discussion of the FORMAT procedure in the *Base SAS Procedures Guide* and the discussions of the FORMAT statement and SAS formats in *SAS Formats and Informats: Reference*.

You can specify the following *option* in the CLASS statement after a slash (/):

TRUNCATE

specifies that class levels be determined by using only up to the first 16 characters of the formatted values of CLASS variables.

EFFECT Statement

EFFECT *name=effect-type* (*variables* < / *options*>) ;

The EFFECT statement enables you to construct special collections of columns for design matrices. These collections are referred to as *constructed effects* to distinguish them from the usual model effects that are formed from continuous or classification variables, as discussed in the section “GLM Parameterization of Classification Variables and Effects” on page 391 in Chapter 19, “Shared Concepts and Topics.”

You can specify the following *effect-types*:

COLLECTION	specifies a collection effect that defines one or more variables as a single effect with multiple degrees of freedom. The variables in a collection are considered as a unit for estimation and inference.
LAG	specifies a classification effect in which the level that is used for a particular period corresponds to the level in the preceding period.
MULTIMEMBER MM	specifies a multimember classification effect whose levels are determined by one or more variables that appear in a CLASS statement.
POLYNOMIAL POLY	specifies a multivariate polynomial effect in the specified numeric variables.
SPLINE	specifies a regression spline effect whose columns are univariate spline expansions of one or more variables. A spline expansion replaces the original variable with an expanded or larger set of new variables.

Table 94.2 summarizes the *options* available in the EFFECT statement.

Table 94.2 EFFECT Statement Options

Option	Description
Collection Effects Options	
DETAILS	Displays the constituents of the collection effect
Lag Effects Options	
DESIGNROLE=	Names a variable that controls to which lag design an observation is assigned
DETAILS	Displays the lag design of the lag effect
NLAG=	Specifies the number of periods in the lag
PERIOD=	Names the variable that defines the period. This option is required.
WITHIN=	Names the variable or variables that define the group within which each period is defined. This option is required.

Table 94.2 *continued*

Option	Description
Multimember Effects Options	
NOEFFECT	Specifies that observations with all missing levels for the multimember variables should have zero values in the corresponding design matrix columns
WEIGHT=	Specifies the weight variable for the contributions of each of the classification effects
Polynomial Effects Options	
DEGREE=	Specifies the degree of the polynomial
MDEGREE=	Specifies the maximum degree of any variable in a term of the polynomial
STANDARDIZE=	Specifies centering and scaling suboptions for the variables that define the polynomial
Spline Effects Options	
BASIS=	Specifies the type of basis (B-spline basis or truncated power function basis) for the spline effect
DEGREE=	Specifies the degree of the spline effect
KNOTMETHOD=	Specifies how to construct the knots for the spline effect

For more information about the syntax of these *effect-types* and how columns of constructed effects are computed, see the section “[EFFECT Statement](#)” on page 401 in Chapter 19, “[Shared Concepts and Topics](#).”

MODEL Statement

MODEL *response* < * *censor*(*list*) > = < *effects* > < / *options* > ;

The MODEL statement identifies the response variable, the optional censoring variable, and the explanatory effects, including covariates, main effects, interactions, and nested effects; for more information, see the section “[Specification of Effects](#)” on page 3593 in Chapter 46, “[The GLM Procedure](#).” In the MODEL statement, the response variable precedes the equal sign. This name can optionally be followed by an asterisk, the name of the censoring variable, and a list of censoring values (separated by blanks or commas if you list more than one value) enclosed in parentheses. If the censoring variable takes on one of these values, the corresponding failure time is considered to be censored. Following the equal sign are the explanatory effects (sometimes called independent variables or covariates) for the model.

The censoring variable must be numeric.

Options

You can specify the following *options* after a slash (/):

NOINT

specifies no intercept regression.

QUANTILE=number-list | PROCESS

specifies the quantile levels of interest for quantile regression analysis. You can specify any number of quantile levels in the interval (0, 1). You can also compute the entire quantile process by specifying the PROCESS option.

If you do not specify the QUANTILE= option, the QUANTLIFE procedure fits a median regression, which corresponds to QUANTILE=0.5.

OUTPUT Statement

OUTPUT < **OUT=SAS-data-set** > *keyword=name* < ... *keyword=name* > ;

The OUTPUT statement creates a *SAS-data-set* to contain statistics that are calculated after fitting models for all quantiles specified by the QUANTILE= option in the MODEL statement. At least one specification of the form *keyword=name* is required.

All variables in the original data set are included in the new data set, along with the variables that are created. These new variables contain fitted values and estimated quantiles.

You can use the following specifications in the OUTPUT statement:

OUT=SAS-data-set specifies the new data set. By default, the QUANTLIFE procedure uses the *DATA**n* convention to name the new data set. For more information, see the section “[OUT= Output Data Set in the OUTPUT Statement](#)” on page 7659.

keyword=name specifies the statistics to include in the output data set and gives names to the new variables. Specify a *keyword* for each desired statistic (see the following list of *keywords*), an equal sign, and the variable to contain the statistic.

You can specify the following *keywords*, which represent the indicated statistics:

PREDICTED | P specifies a variable to contain the predicted response.

RESIDUAL | RES specifies a variable to contain the residuals, $y_i - x_i' \hat{\beta}(\tau)$.

SAMPLEWEIGHT | SW specifies variables for sample weights from the bootstrap samples. For the *i*th sample, a column is added that contains the weights that are used for that sample. The name of this column is formed by appending an index *i* to the name that you specify. If you do not specify a name, then the default prefix is sw.

STDP specifies a variable to contain the estimates of the standard errors of the estimated response.

TEST Statement

`<label:> TEST effects </options> ;`

In quantile regression analysis, you might be interested in testing whether a covariate effect is statistically significant for a given quantile. You can use the TEST statement to obtain a test for the canonical linear hypothesis about the parameters of the tested *effects*,

$$\beta_j = 0, \quad j = i_1, \dots, i_q$$

where q is the total number of parameters of the tested *effects*. The tested *effects* can be any set of effects in the MODEL statement.

You can include multiple TEST statements, provided that they appear after the MODEL statement. The optional *label*, which must be a valid SAS name, identifies output from the corresponding TEST statement. For more information about these tests, see the section “[Testing Effects of Covariates](#)” on page 7658.

WEIGHT Statement

`WEIGHT variable ;`

The WEIGHT statement specifies a weight variable in the input data set.

To request weighted quantile regression, place the weights in a variable and specify the name in the WEIGHT statement. The values of the WEIGHT variable can be nonintegral and are not truncated. Observations that have nonpositive or missing values for the weight variable do not contribute to the fit of the model. For more information about weighted quantile regression, see the section “[Details: QUANTREG Procedure](#)” on page 7703 in Chapter 95, “[The QUANTREG Procedure](#).”

Details: QUANTLIFE Procedure

Notation for Censored Quantile Regression

Let T be a dependent variable, such as a survival time, and let x be a $p \times 1$ covariate vector. Quantile regression methods focus on modeling the conditional quantile function, $Q_T(\tau|x)$, which is defined as

$$Q_T(\tau|x) = \inf\{t : P(T \leq t|x) = \tau\}, \quad 0 < \tau < 1$$

For example, $Q_T(0.5|x)$ is the conditional median quantile, and $Q_T(0.95|x)$ is the conditional quantile function that corresponds to the 95th percentile.

A linear quantile regression model for $Q_T(\tau|x)$ has the form $x'\beta(\tau)$. One of the advantages of quantile regression analysis is that the covariate effect $\beta(\tau)$ can change with τ . Unlike ordinary least squares regression, which estimates the conditional expectation function $E(T|x)$, quantile regression offers the flexibility to model the entire conditional distribution.

Given observations $\{(T_i, x_i), i = 1, \dots, n\}$, standard quantile regression estimates the regression coefficients $\beta(\tau)$ by minimizing the following objective function over b :

$$r(b) = \sum_{i=1}^n \rho_{\tau}(T_i - x_i' b)$$

where $\rho_{\tau}(u) = u(\tau - I(u < 0))$.

However, in many applications, the responses T_i are subject to censoring. For example, in a biomedical study, censoring occurs when patients withdraw from the study or die from a cause that is unrelated to the disease being studied.

Let C_i denote the censoring variable. In the case of right-censoring, the triples (x_i, Y_i, Δ_i) are observed, where $Y_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$ are the observed response variable and the censoring indicator, respectively. Standard quantile regression can lead to a biased estimator of the regression parameters $\beta(\tau)$ when censoring occurs.

The following sections describe two methods for estimating the quantile coefficient $\beta(\tau)$ in the presence of right-censoring.

Kaplan-Meier-Type Estimator for Censored Quantile Regression

Portnoy (2003) proposes the use of weighted quantile regression to sequentially estimate $\beta(\tau_k)$ along the equally spaced grid $0 < \tau_1 < \dots < \tau_M < 1$. You can request this method by specifying the METHOD=KM option in the PROC QUANTLIFE statement. The grid points $0 < \tau_1 < \dots < \tau_M < 1$ are equally spaced, with τ_1 specified by the INITTAU= option and the step between adjacent grid points specified by the GRIDSIZE=option.

This method uses a weight function $w_i(\tau)$ for each censored observation. The weight function is constructed as follows: Let $\hat{\tau}_i$ be the first grid point at which $x_i' \hat{\beta}(\tau_i) \geq C_i$ and $x_i' \hat{\beta}(\tau_{i+1}) < C_i$; otherwise let $\hat{\tau}_i = 1$. When computing the τ th quantile, assign weight $w_i(\tau) = \frac{\tau - \hat{\tau}_i}{1 - \hat{\tau}_i}$ to the censored observation Y_i if $\tau > \hat{\tau}_i$; otherwise assign $w_i(\tau) = 1$. The algorithm for computing $\hat{\beta}(\tau_k)$, $k = 1, \dots, M$, is as follows:

1. Compute $\hat{\beta}(\tau_1)$ by using the standard quantile regression method.
2. For $k = 2, \dots, M$, obtain $\hat{\beta}(\tau_k)$ sequentially by minimizing the following weighted quantile regression objective function:

$$r_w(b) = \sum_{\Delta_i=1} \rho_{\tau_k}(Y_i - x_i' b) + \sum_{\Delta_i=0} \{w_i(\tau_k) \rho_{\tau_k}(Y_i - x_i' b) + (1 - w_i(\tau_k)) \rho_{\tau_k}(Y^* - x_i' b)\}$$

where $w_i(\tau_k)$ is the weight for the right-censored observation Y_i at computing $\hat{\beta}(\tau_k)$, and the complementary weight $1 - w_i(\tau_k)$ is for Y^* , a large constant that is greater than all $x_i' \hat{\beta}(\tau)$.

The weighted quantile regression method is similar to Efron's redistribution-of-mass idea (Efron 1967) for the Kaplan-Meier estimator.

Note that if all observations are uncensored, $\hat{\beta}(\tau_k)$ is the same as the standard quantile regression estimator.

Nelson-Aalen-Type Estimator for Censored Quantile Regression

Peng and Huang (2008) propose a method of censored quantile regression that is based on the Nelson-Aalen estimator of the cumulative hazard function. Let $F_i(t|x) = P(T_i \leq t|x_i)$, $\Lambda_i(t|x) = -\log(1 - F_i(t|x))$, and $N_i(t) = I\{\{T_i \leq t\} \text{ and } \{\Delta_i = 1\}\}$. Then the following equation is a martingale process that is associated with the counting process $N_i(t)$ (Fleming and Harrington 1991):

$$M_i(t) = N_i(t) - \Lambda_i(t \wedge Y_i|x)$$

Based on the martingale process, Peng and Huang (2008) derive the following estimating equation:

$$n^{-1/2} \sum_{i=1}^n x_i [N_i(\exp(x_i' \beta(\tau))) - \int_0^\tau I(Y_i \geq \exp(x_i' \beta(\tau))) dH(u)] = 0$$

where $H(u) = -\log(1 - u)$ and $u \in [0, 1)$. By approximating the integral in the estimating equation on a grid $0 = \tau_0 < \tau_1 < \dots < \tau_M < 1$, the regression quantiles $\beta(\tau_k)$, $k = 1, \dots, M$, can be estimated sequentially by solving the following linear programming problem:

$$\min_b \{ \alpha(\tau_k)'u + (\Delta - \alpha(\tau_k))'v \mid z = Xb + u - v, u \geq 0, v \geq 0 \}$$

where

$$\alpha(\tau_k) = \sum_{j=1}^{k-1} I(Y_i \geq \exp(x_i' \hat{\beta}(\tau_j))) H((u_{j+1}) - H(u_j))$$

and X is the known matrix of x_i 's. For more information, see Koenker (2008).

You can request this method by specifying the METHOD=NA option. The grid points $0 = \tau_0 < \tau_1 < \dots < \tau_M < 1$ are equally spaced, with τ_1 specified by the INITTAU=option and the grid step between two adjacent grid points specified by the GRIDSIZE=option.

Relationship of Quantile Function and Survival Function

Both quantile function and survival function are useful in characterizing a lifetime distribution.

By the definition of the quantile function $Q_T(\tau|x)$,

$$F(Q_T(\tau|x)) = P(T \leq Q_T(\tau|x)) = \tau$$

In other words, the cumulative distribution function $F_T(t|x)$ maps $Q_T(\tau|x)$ to τ , and thus the corresponding survival function $S_T(t|x)$ maps $Q_T(\tau|x)$ to $1 - \tau$.

When you specify the LOG option, the QUANTLIFE procedure fits a linear quantile regression model for a log transformation of the lifetime as

$$Q_{\log(T)}(\tau|x) = x' \beta(\tau)$$

where $Q_{\log(T)}(\tau|x)$ is the τ th quantile of $\log(T)$ at x . The estimated quantile function for T given x is $\hat{Q}_T(\tau|x) = e^{x'\hat{\beta}(\tau)}$, because the quantile function is invariant under a monotone transformation.

You can specify the covariates x in the COVARIATES= data set of the BASELINE statement and the PLOTS=(QUANTILE SURVIVAL) option in the PROC statement. Then the conditional quantile function at x is plotted as $\hat{Q}_T(\tau|x)$ against τ , and the conditional survival function at x is plotted as $1 - \tau$ against $\hat{Q}_T(\tau|x)$.

Confidence Interval

Direct computation of the covariance of the parameter estimators involves a complicated density estimation. Instead, the QUANTLIFE procedure computes confidence intervals for the quantile regression parameters $\beta(\tau)$ by using resampling methods. The QUANTLIFE procedure implements two different methods, the exponentially weighted method and the pairwise resampling method.

Exponentially Weighted Method

This method samples weights $w_i, i = 1, \dots, n$, from a standard exponential distribution that has mean 1 and variance 1. Then it computes the censored quantile regression estimators $\hat{\beta}(\tau)$ based on the observed data (x_i, Y_i, Δ_i) with the weights w_i . These steps are repeated B times (where B is the value of the NREP= option in the PROC QUANTLIFE statement). The confidence intervals can be obtained from these B estimates. You can specify this method by using the CI=EW option in the PROC QUANTLIFE statement.

Pairwise Method

This method samples (x_i, Y_i, Δ_i) with replacement and computes the quantile regression estimators $\hat{\beta}(\tau)$ based on the resampled data. These steps are repeated B times (where B is the value of the NREP= option in the PROC QUANTLIFE statement). The confidence intervals can be obtained from these B estimates. You can specify this method by using the CI=PW option in the PROC QUANTLIFE statement.

Testing Effects of Covariates

Consider the linear model

$$y_i = \mathbf{x}'_{1i}\boldsymbol{\beta}_1 + \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + \epsilon_i$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are p -dimensional and q -dimensional parameters, respectively, and $\epsilon_i, i = 1, \dots, n$, are errors. Denote $\mathbf{x}'_i = (\mathbf{x}'_{1i}, \mathbf{x}'_{2i})$, and let $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$ be the parameter estimates for $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively, at the τ th quantile.

The QUANTLIFE procedure implements the Wald test for the null hypothesis:

$$H_0 : \beta_2(\tau) = 0$$

The Wald test statistic, which is based on the estimated coefficients $\hat{\beta}_2$ from the unrestricted fitted model, is given by

$$T_W(\tau) = \hat{\beta}'_2(\tau)\hat{\Sigma}(\tau)^{-1}\hat{\beta}_2(\tau)$$

where $\hat{\Sigma}(\tau)$ is an estimator of the covariance of $\hat{\beta}_2(\tau)$, which is obtained by using resampling methods.

Output Data Sets

OUTBOOTEST= Output Data Set

The OUTBOOTEST= data set contains parameter estimates for the specified model from resampled data sets. A set of observations is created for each quantile level and for each resampled data set.

If the QUANTLIFE procedure does not produce valid solutions, the parameter estimates are set to missing in the OUTBOOTEST= data set.

If created, this data set contains all variables that are specified in the MODEL statement. Each observation contains parameter estimates for a specified quantile level.

The following variables are also included in the data set:

- any specified BY variables
- `_STATUS_`, a character variable of length 12 that contains the status of the model fit: either NORMAL, NOUNIQUE, or NOVALID
- Intercept, a numeric variable that contains the intercept parameter estimates
- `_TAU_`, a numeric variable that contains the specified quantile levels from the MODEL statement

For continuous explanatory variables, the names of the parameters are the same as they are for the corresponding variables. For CLASS variables, the parameter names are obtained by concatenating the corresponding CLASS variable name to the CLASS category. For interaction and nested effects, the parameter names are created by concatenating the names of each component effect.

OUT= Output Data Set in the OUTPUT Statement

The OUT= data set that is specified in the OUTPUT statement contains all the variables in the input data set, along with statistics that you request by specifying *keyword=name* options. The additional variables are calculated for each observation in the input data set.

OUT= Output Data Set in the BASELINE Statement

The OUT= data set that is specified in the BASELINE statement contains all the variables in the COVARIATES= data set.

The following variables are also added to the data set:

<code>_QUANTILE_</code>	a numeric variable that contains the estimated quantile function
<code>_SURVIVAL_</code>	a numeric variable that contains the estimated survival function
<code>_TAU_</code>	a numeric variable that contains the quantile levels that are the complement of the survival function

ODS Table Names

Table 94.3 lists the names that the QUANTLIFE procedure assigns to each table that it creates. You can specify these names when you use the Output Delivery System (ODS) to select tables and create output data sets.

Table 94.3 ODS Tables Produced by PROC QUANTLIFE

ODS Table Name	Description	Statement	Option
ClassLevels	Classification variable levels	CLASS	Default
ModelInfo	Model information	MODEL	Default
NObs	Number of observations	PROC QUANTLIFE	Default
ParameterEstimates	Parameter estimates	MODEL	Default
CensoredSummary	Summary of event and censored observations	PROC QUANTLIFE	Default
Tests	Results of tests	TEST	Default

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “[Statistical Graphics Using ODS](#).”

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “[Enabling and Disabling ODS Graphics](#)” on page 609 in Chapter 21, “[Statistical Graphics Using ODS](#).”

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “[A Primer on ODS Statistical Graphics](#)” on page 608 in Chapter 21, “[Statistical Graphics Using ODS](#).”

The QUANTLIFE procedure assigns a name to each graph that it creates. You can use these names to refer to the graphs when you use ODS. The names along with the required statements and options are listed in Table 94.4.

Table 94.4 Graphs Produced by PROC QUANTLIFE

ODS Graph Name	Plot Description	PLOTS= Option
QuantilePlot	Quantile function plot	QUANTILE
QuantPanel	Panel of regression parameter estimates against quantile level with confidence limits	QUANTPLOT
QuantPlot	Scatter plot for regression parameter estimates against quantile level with confidence limits	QUANTPLOT / UNPACK
SurvivalPlot	Survival function plot	SURVIVAL

Examples: QUANTLIFE Procedure

Example 94.1: Primary Biliary Cirrhosis Study

This example illustrates how to use quantile regression analysis to detect varying covariate effects on survival time. Consider a study of primary biliary cirrhosis, a rare but fatal chronic liver disease, discussed by Fleming and Harrington (1991). Researchers followed 418 patients who had this disease, 161 of whom died during the study.

The data set contains the following variables:

- Time, follow-up time, in years
- Status, event indicator, with value 1 for death time and value 0 for censored time
- Age, age from birth to study registration, in years
- Albumin, serum albumin level, in g/dl
- Bilirubin, serum bilirubin level, in mg/dl
- Edema, edema presence
- Prottime, prothrombin time, in seconds

The following statements create the data set PBC, which is used in this example:

```
data pbc;
  input Time Status Age Albumin Bilirubin Edema Prottime @@;
  label Time="Follow-Up Time in Days";
  logAlbumin  = log(Albumin);
  logBilirubin = log(Bilirubin);
  logProttime  = log(Prottime);
  datalines;
400 1 58.7652 2.60 14.5 1.0 12.2 4500 0 56.4463 4.14 1.1 0.0 10.6
1012 1 70.0726 3.48 1.4 0.5 12.0 1925 1 54.7406 2.54 1.8 0.5 10.3
1504 0 38.1054 3.53 3.4 0.0 10.9 2503 1 66.2587 3.98 0.8 0.0 11.0
1832 0 55.5346 4.09 1.0 0.0 9.7 2466 1 53.0568 4.00 0.3 0.0 11.0
2400 1 42.5079 3.08 3.2 0.0 11.0 51 1 70.5599 2.74 12.6 1.0 11.5
3762 1 53.7139 4.16 1.4 0.0 12.0 304 1 59.1376 3.52 3.6 0.0 13.6

... more lines ...

989 0 35.0000 3.23 0.7 0.0 10.8 681 1 67.0000 2.96 1.2 0.0 10.9
1103 0 39.0000 3.83 0.9 0.0 11.2 1055 0 57.0000 3.42 1.6 0.0 9.9
691 0 58.0000 3.75 0.8 0.0 10.4 976 0 53.0000 3.29 0.7 0.0 10.6
;
```

The next statements fit a linear model for the log of survival time of the PBC patients with the covariates logBilirubin, logProtime, logAlbumin, Age, and Edema:

```
ods graphics on;
proc quantlife data=pbcr log method=na plot=(quantplot survival) seed=1268;
    model Time*Status(0)=logBilirubin logProtime logAlbumin Age Edema
        /quantile=(.1 .2 .3 .4 .5 .6 .75);
run;
```

You use the QUANTILE= option to specify a set of quantiles of interest for comparing quantile-specific covariate effects. The METHOD= option specifies the Nelson-Aalen method for estimating the regression parameters.

The QUANTLIFE procedure provides resampling methods for computing confidence limits for the parameters; for more information, see the section “[Confidence Interval](#)” on page 7658. By default, the repetition number is 200. You can request a different number of repetitions by specifying the NREP= option. You can also use the SEED= option to specify the seed for generating random numbers so that you can later reproduce the results.

Figure 94.1.1 displays model information and information about censoring in the data. Out of 418 observations, 257 are censored.

Output 94.1.1 Model Information

The QUANTLIFE Procedure

Model Information			
Data Set	WORK.PBC		
Dependent Variable	Log(Time)		
Censoring Variable	Status		
Censoring Value(s)	0		
Number of Observations	418		
Method	Nelson-Aalen		
Replications	200		
Seed for Random Number Generator	1268		

Summary of the Number of Event and Censored Values			
			Percent
Total	Event	Censored	Censored
418	161	257	61.48

Figure 94.1.2 provides the parameter estimates. Each quantile level has a set of parameter estimates and confidence limits.

Output 94.1.2 Parameter Estimates at Different Quantiles

Parameter Estimates								
Quantile	Parameter	DF	Estimate	Standard Error	95% Confidence Limits		t Value	Pr > t
0.1000	Intercept	1	14.8030	4.0967	6.7736	22.8325	3.61	0.0003
	logBilirubin	1	-0.4488	0.1485	-0.7398	-0.1578	-3.02	0.0027
	logProtime	1	-3.6378	1.4560	-6.4915	-0.7841	-2.50	0.0129
	logAlbumin	1	1.9286	0.9756	0.0165	3.8408	1.98	0.0487
	Age	1	-0.0244	0.0107	-0.0455	-0.00334	-2.27	0.0237
	Edema	1	-1.0712	0.6688	-2.3820	0.2396	-1.60	0.1100
0.2000	Intercept	1	15.1800	2.6664	9.9540	20.4060	5.69	<.0001
	logBilirubin	1	-0.6532	0.0886	-0.8268	-0.4796	-7.37	<.0001
	logProtime	1	-3.3273	0.9401	-5.1699	-1.4847	-3.54	0.0004
	logAlbumin	1	1.6842	0.6888	0.3343	3.0342	2.45	0.0149
	Age	1	-0.0291	0.00687	-0.0425	-0.0156	-4.23	<.0001
	Edema	1	-0.7265	0.3179	-1.3497	-0.1034	-2.29	0.0228
0.3000	Intercept	1	13.2382	2.5296	8.2804	18.1961	5.23	<.0001
	logBilirubin	1	-0.6013	0.0762	-0.7506	-0.4521	-7.90	<.0001
	logProtime	1	-2.5816	0.8907	-4.3273	-0.8359	-2.90	0.0039
	logAlbumin	1	1.7246	0.7142	0.3248	3.1245	2.41	0.0162
	Age	1	-0.0244	0.00716	-0.0385	-0.0104	-3.41	0.0007
	Edema	1	-0.8577	0.2763	-1.3992	-0.3163	-3.10	0.0020
0.4000	Intercept	1	13.4716	3.0874	7.4204	19.5228	4.36	<.0001
	logBilirubin	1	-0.6047	0.0846	-0.7705	-0.4389	-7.15	<.0001
	logProtime	1	-2.1632	1.1726	-4.4615	0.1351	-1.84	0.0658
	logAlbumin	1	0.9819	0.7191	-0.4274	2.3912	1.37	0.1728
	Age	1	-0.0255	0.00681	-0.0389	-0.0122	-3.74	0.0002
	Edema	1	-1.0589	0.3104	-1.6672	-0.4506	-3.41	0.0007
0.5000	Intercept	1	10.9205	2.8047	5.4235	16.4175	3.89	0.0001
	logBilirubin	1	-0.5315	0.0904	-0.7087	-0.3543	-5.88	<.0001
	logProtime	1	-1.2222	1.2142	-3.6020	1.1577	-1.01	0.3148
	logAlbumin	1	1.5700	0.6284	0.3383	2.8016	2.50	0.0129
	Age	1	-0.0318	0.00883	-0.0491	-0.0145	-3.60	0.0004
	Edema	1	-0.7316	0.3743	-1.4653	0.00202	-1.95	0.0513
0.6000	Intercept	1	11.2381	2.6294	6.0846	16.3917	4.27	<.0001
	logBilirubin	1	-0.5701	0.0852	-0.7370	-0.4031	-6.69	<.0001
	logProtime	1	-1.3508	1.1402	-3.5856	0.8840	-1.18	0.2368
	logAlbumin	1	1.3704	0.5091	0.3726	2.3682	2.69	0.0074
	Age	1	-0.0226	0.0109	-0.0440	-0.00111	-2.06	0.0399
	Edema	1	-0.5141	0.3088	-1.1193	0.0912	-1.66	0.0968
0.7500	Intercept	1	10.0954	3.1893	3.8445	16.3463	3.17	0.0017
	logBilirubin	1	-0.6366	0.1071	-0.8466	-0.4267	-5.94	<.0001
	logProtime	1	-0.9670	1.2343	-3.3862	1.4521	-0.78	0.4338
	logAlbumin	1	1.8148	0.5883	0.6618	2.9678	3.08	0.0022
	Age	1	-0.0203	0.0156	-0.0509	0.0102	-1.30	0.1931
	Edema	1	-0.3529	0.3120	-0.9644	0.2586	-1.13	0.2587

For comparison, the following statements use the LIFEREG procedure to fit a Weibull distribution to the data. The LIFEREG procedure fits an accelerated failure time model, which assumes that the independent variables have a multiplicative effect on the event time.

```
proc lifereg data=pbcr;
  model Time*Status(0)=logBilirubin logProtime logAlbumin Age Edema;
run;
```

Figure 94.1.3 provides the parameter estimates that are computed by PROC LIFEREG.

Output 94.1.3 Parameter Estimates from PROC LIFEREG

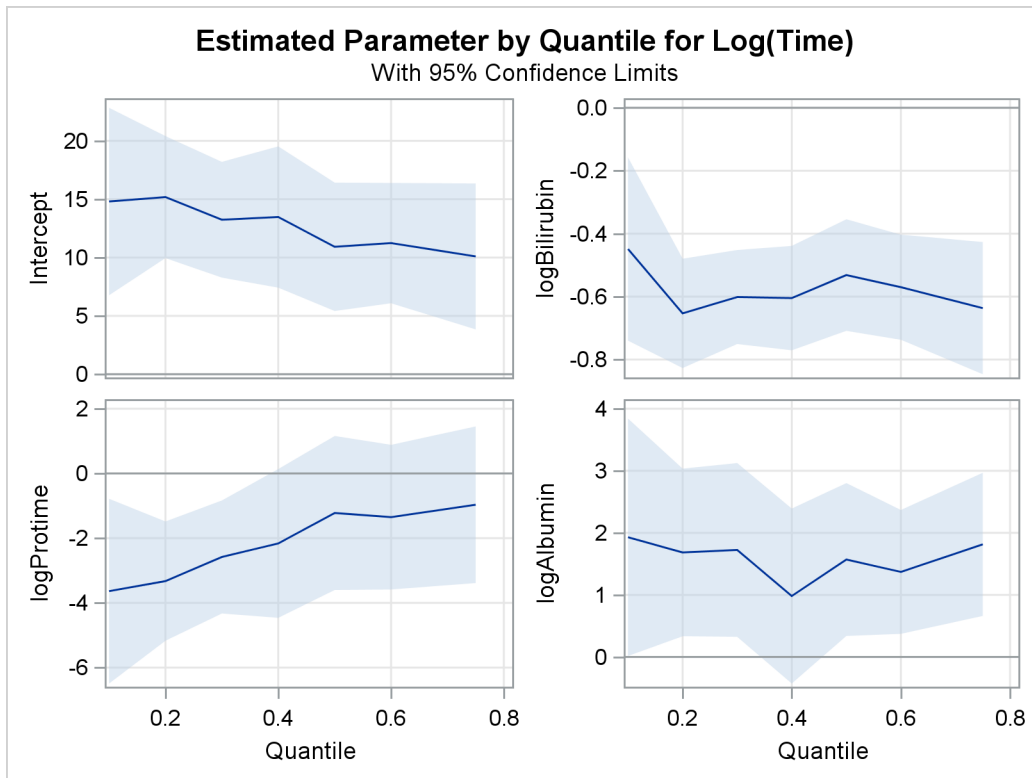
The LIFEREG Procedure

Analysis of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	12.2155	1.4539	9.3658	15.0651	70.59	<.0001
logBilirubin	1	-0.5770	0.0556	-0.6861	-0.4680	107.55	<.0001
logProtime	1	-1.7565	0.5248	-2.7850	-0.7280	11.20	0.0008
logAlbumin	1	1.6694	0.4276	0.8313	2.5074	15.24	<.0001
Age	1	-0.0265	0.0053	-0.0368	-0.0162	25.35	<.0001
Edema	1	-0.6303	0.1805	-0.9842	-0.2764	12.19	0.0005
Scale	1	0.6807	0.0430	0.6014	0.7704		
Weibull Shape	1	1.4691	0.0928	1.2980	1.6628		

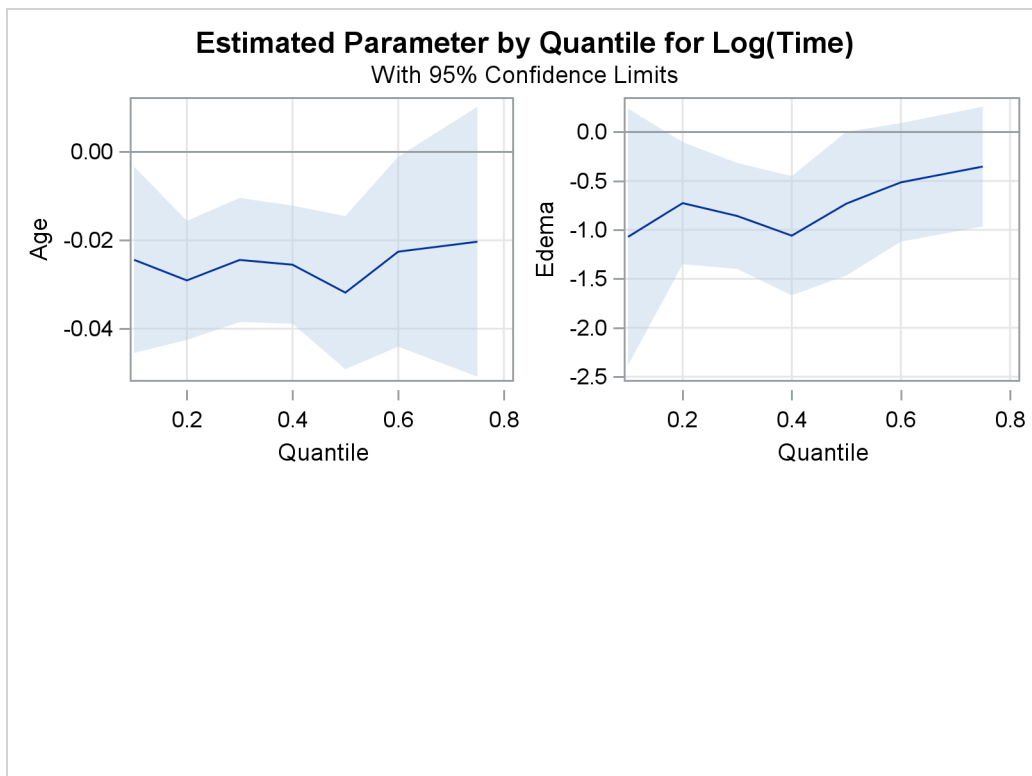
The p -value for logProtime is very small. For this same variable, the p -values that result from the quantile regression analysis are 0.3148 for the 0.5th quantile and 0.4338 for the 0.75th quantile, and the p -values are much smaller for the lower quantiles. Apparently, the effect of this covariate depends on which side of the response distribution is being modeled.

The PLOT=QUANTPLOT option in the PROC QUANTLIFE statement requests the quantile process plots in Figure 94.1.4 and Figure 94.1.5, which plot the estimated regression parameter against the quantile level. You can use these plots to compare quantile-specific covariate effects. If the curve is not constant, it can indicate heterogeneity in the data. The interpretation of the regression coefficients at a given quantile is similar to that of classical regression analysis. That is, the coefficient from a given covariate indicates the effect on log(Time) of a unit change in that covariate, assuming that the other covariates are fixed.

Output 94.1.4 Quantile Processes with 95% Confidence Bands



Output 94.1.5 Quantile Processes with 95% Confidence Bands



In Figure 94.1.4, you can see that the effect of logProtime has a negative effect over the lower quantiles, which diminishes in magnitude at the median and upper quantiles. This insight would be missed if you were using the accelerated failure model.

Example 94.2: Drug Abuse Study

This example reproduces analysis done by Portnoy (2003), which demonstrates how to use quantile regression to analyze survival times. The example uses drug abuse data from Hosmer and Lemeshow (1999). The goal of this study is to compare treatment effects on reducing drug abuse.

The data set contains the following variables:

- Time, time to return to drug use, in days
- Status, event indicator, with value 1 for return to drug use and value 0 for censored time
- Age, age at enrollment, in years
- Treatment, with value 1 for six-month treatment and value 0 for three-month treatment
- Beck, Beck Depression Inventory score at admission to the program
- IV3, indicator of recent intravenous drug use
- NDT, number of prior drug treatments
- Race, race indicator, with value 1 for white and value 0 for nonwhite
- Site, treatment sites (A and B)
- LOT, length of treatment, in days

The following statements create the data set UIS:

```
data uis;
  input  ID Age Beckettota Hercoc Ivhx Ndrugtx Race Treat Site Lot Time
  Censor;
  Iv3 = (Ivhx = 3);
  Nd1 = 1/((Ndrugtx+1)/10);
  Nd2 = (1/((Ndrugtx+1)/10))*log((Ndrugtx+1)/10);
  if (Treat =1 ) then Frac = Lot/180;
  else Frac = Lot/90;
  datalines;
    1 39 9.0000 4 3 1 0 1 0 123 188 1
    2 33 34.0000 4 2 8 0 1 0 25 26 1
    3 33 10.0000 2 3 3 0 1 0 7 207 1
    4 32 20.0000 4 3 1 0 0 0 66 144 1

    ... more lines ...

    626 28 10.0 4 2 3 0 1 1 21 35 1
    627 35 17.0 1 3 2 0 0 1 184 379 1
    628 46 31.5 1 3 15 1 1 1 9 377 1
;
```

The following statements replicate the analysis of Portnoy (2003):

```
ods graphics on;
proc quantlife data=uis log seed=999 plots=(quantplot survival);
  class Race Site Treat;
  model Time*Censor(0)=Nd1 Nd2 Iv3 Becktota
        Treat Frac Race Age|Site
        / quantile=0.05 to 0.85 by 0.05 ;
  baseline out=Predsurvf;

run;
```

Figure 94.2.1 displays the model information. Out of 628 subjects, 53 contain missing values and are not included in the analysis. The censoring rate is 20.87%.

Output 94.2.1 Model Information

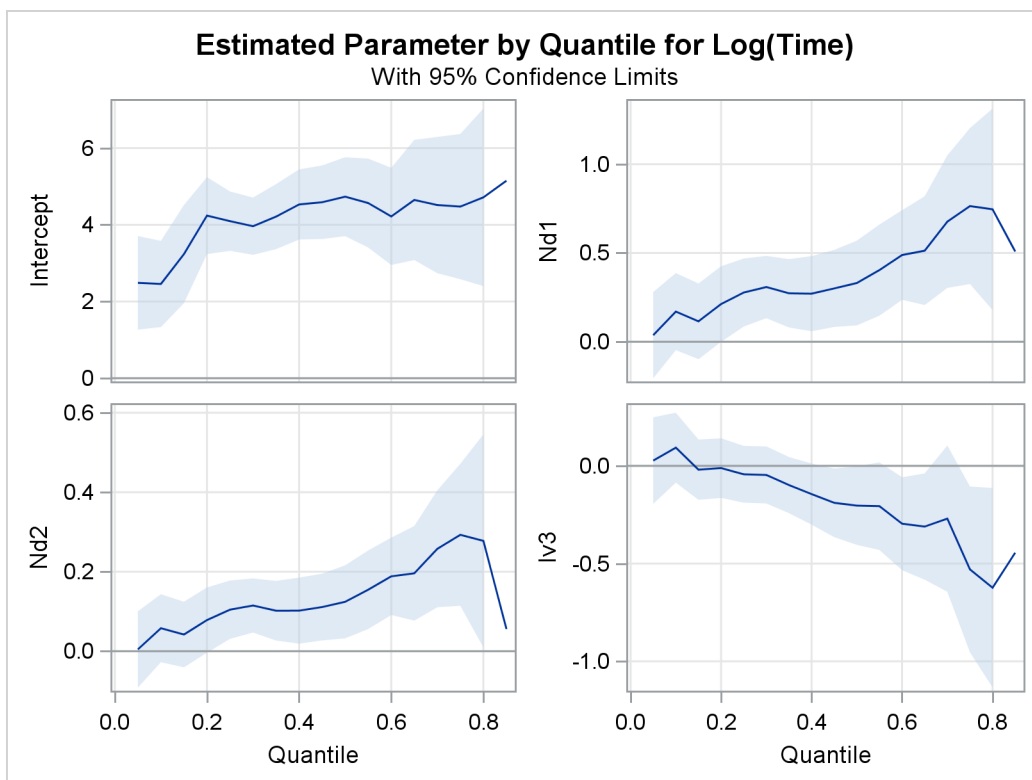
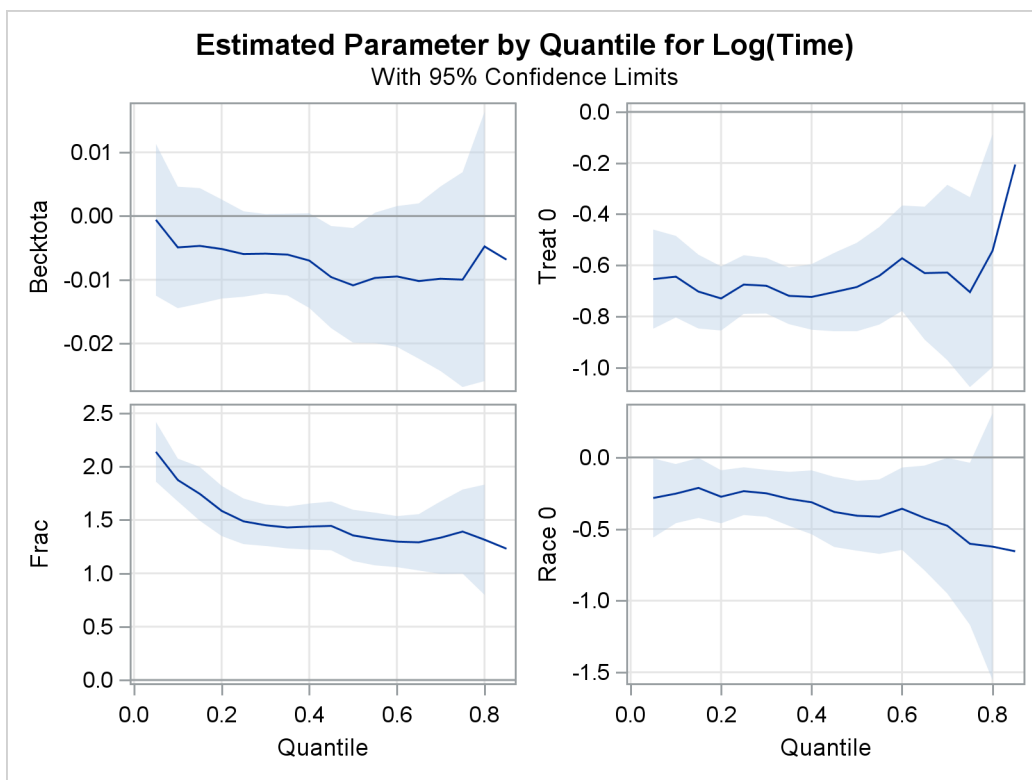
The QUANTLIFE Procedure

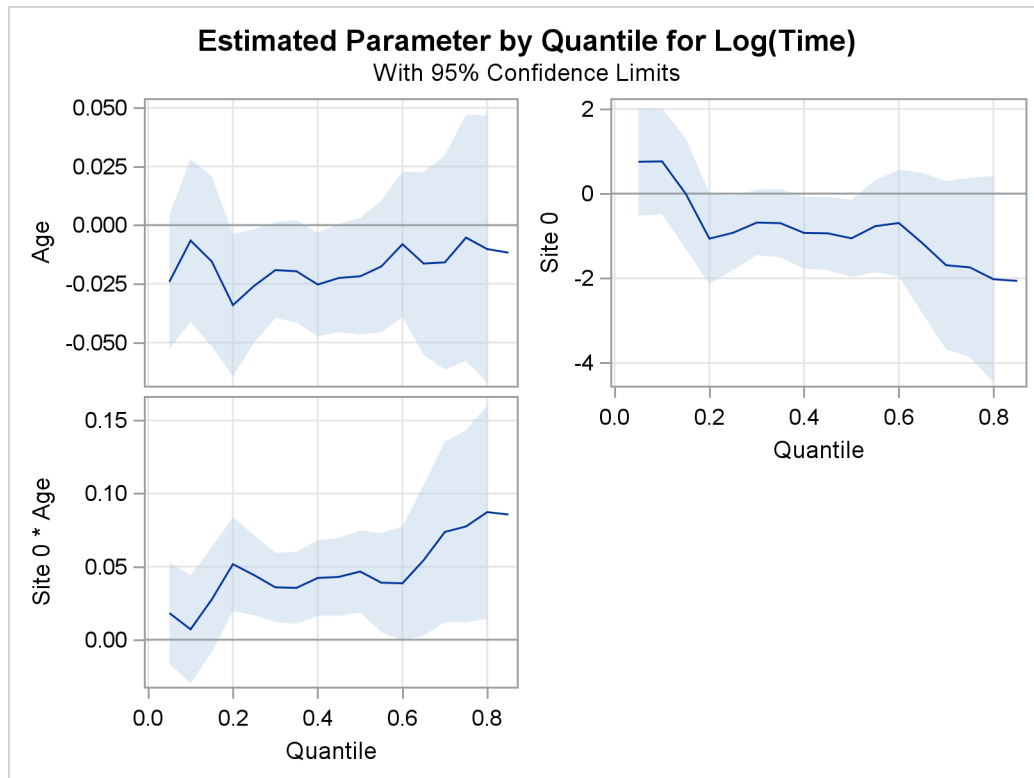
Model Information	
Data Set	WORK.UIS
Dependent Variable	Log(Time)
Censoring Variable	Censor
Censoring Value(s)	0
Number of Observations	575
Method	Kaplan-Meier
Replications	200
Seed for Random Number Generator	999

Class Level Information		
Name	Levels	Values
Race	2	0 1
Site	2	0 1
Treat	2	0 1

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
575	464	111	19.30

Figure 94.2.2, Figure 94.2.3, and Figure 94.2.4 display regression quantile process plots for each covariate.

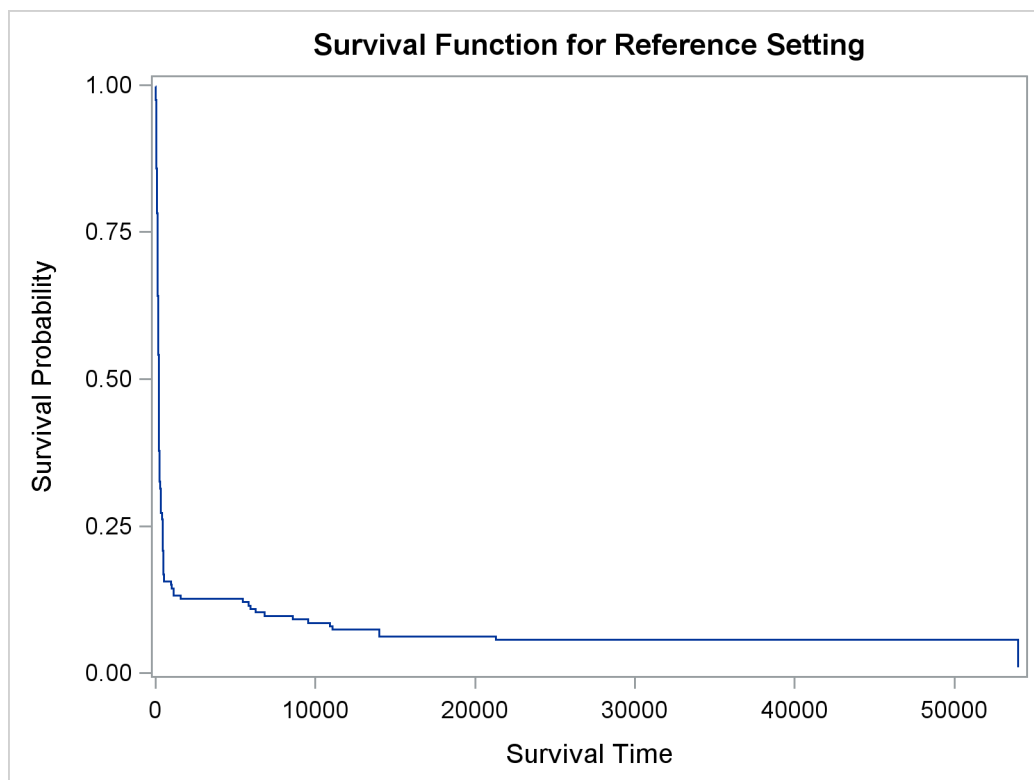
Output 94.2.2 Quantile Processes with 95% Confidence Bands**Output 94.2.3** Quantile Processes with 95% Confidence Bands

Output 94.2.4 Quantile Processes with 95% Confidence Bands

You can see the varying effects for Nd and Frac, whereas the treatment effect is fairly constant. For more information about the covariate effects that can be discovered using quantile regression, see Portnoy (2003).

In survival analysis, a plot of the estimated survival function is often of interest. There is a one-to-one relationship between the quantile function and the survival function. When you specify the `PLOTS= SURVIVAL` option, the `QUANTLIFE` procedure estimates the survival function by fitting a quantile regression model for a grid of equally spaced quantile levels. You can specify the grid points by using the `INITTAU=` option and the step between adjacent grid points by using the `GRIDSIZE=` option. For more information, see the section “Kaplan-Meier-Type Estimator for Censored Quantile Regression” on page 7656.

Figure 94.4 shows the estimated survival function at the reference set of covariate values that consist of reference levels for the CLASS variables and average values for the continuous variables. You can output the predicted survival function by specifying the `SURVIVAL=` option in the `BASELINE` statement.

Output 94.2.5 Survival Function

References

- Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2009). “Improving Point and Interval Estimators of Monotone Functions by Rearrangement.” *Biometrika* 96:559–575.
- Efron, B. (1967). “The Two Sample Problem with Censored Data.” In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 4, 831–853. Berkeley: University of California Press.
- Fleming, T. R., and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: John Wiley & Sons.
- Hosmer, D. W., Jr., and Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. New York: John Wiley & Sons.
- Hosmer, D. W., Jr., Lemeshow, S., and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Huang, Y. (2010). “Quantile Calculus and Censored Regression.” *Annals of Statistics* 38:1607–1637.
- Klein, J. P., and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. New York: Springer-Verlag.
- Koenker, R. (2008). “Censored Quantile Regression Redux.” *Journal of Statistical Software* 27:1–24.
- Koenker, R., and Bassett, G. W. (1978). “Regression Quantiles.” *Econometrica* 46:33–50.
- Koenker, R., and Geling, O. (2001). “Reappraising Medfly Longevity: A Quantile Regression Survival Analysis.” *Journal of the American Statistical Association* 96:458–468.
- Peng, L., and Huang, Y. (2008). “Survival Analysis with Quantile Regression Models.” *Journal of the American Statistical Association* 103:637–649.
- Portnoy, S. (2003). “Censored Regression Quantiles.” *Journal of the American Statistical Association* 98:1001–1012.

Subject Index

- options summary
 - EFFECT statement, [7652](#)
 - QUANTLIFE procedure, [7647](#)
- OUTBOOTEST= data sets
 - QUANTLIFE procedure, [7659](#)
- output data sets
 - QUANTLIFE procedure, [7659](#)
- output table names
 - QUANTLIFE procedure, [7660](#)
- QUANTLIFE procedure, [7641](#)
 - options summary, [7647](#)
 - OUTBOOTEST= data sets, [7659](#)
 - output data sets, [7659](#)
 - output table names, [7660](#)
 - random number generator, [7650](#)
- random number generator
 - QUANTLIFE procedure, [7650](#)

Syntax Index

ALPHA= option
PROC QUANTLIFE (QUANTLIFE), 7648

BASELINE statement
QUANTLIFE procedure, 7650

BY statement
QUANTLIFE procedure, 7651

CI= option
PROC QUANTLIFE statement, 7648

CLASS statement
QUANTLIFE procedure, 7651

COVARIATES= option
BASELINE statement (QUANTLIFE), 7651

DATA= option
PROC QUANTLIFE statement, 7648

EFFECT statement
QUANTLIFE procedure, 7652

GRIDSZIE= option
PROC QUANTLIFE statement, 7648

INITTAU= option
PROC QUANTLIFE statement, 7648

KAPPA= option
PROC QUANTLIFE statement, 7648

keyword= option
OUTPUT statement (QUANTLIFE), 7654

LOG option
PROC QUANTLIFE statement, 7649

MAXIT= option
PROC QUANTREG statement, 7649

METHOD= option
PROC QUANTLIFE statement, 7649

MODEL statement
QUANTLIFE procedure, 7653

NAMELEN= option
PROC QUANTLIFE statement, 7649

NOINT option
MODEL statement (QUANTLIFE), 7654

NREP= option
PROC QUANTLIFE statement, 7649

NTHREADS= option
PROC QUANTLIFE statement, 7649

options summary
PROC statement (QUANTLIFE), 7647

OUT= option
BASELINE statement (QUANTLIFE), 7651
OUTPUT statement (QUANTLIFE), 7654

OUTBOOTEST= option
PROC QUANTLIFE statement, 7649

OUTPUT statement
QUANTLIFE procedure, 7654

PLOT= option
PROC QUANTLIFE statement, 7649

PREDICTED keyword
OUTPUT statement (QUANTLIFE), 7654
PROC QUANTLIFE statement, *see* QUANTLIFE
procedure

QUANTILE= option
MODEL statement (QUANTLIFE), 7654

QUANTLIFE procedure
BASELINE statement, 7650
syntax, 7647

QUANTLIFE procedure, BASELINE statement, 7650
COVARIATES= option, 7651
OUT= option, 7651

QUANTLIFE procedure, BY statement, 7651
QUANTLIFE procedure, CLASS statement, 7651
TRUNCATE option, 7652

QUANTLIFE procedure, EFFECT statement, 7652
QUANTLIFE procedure, MODEL statement, 7653
NOINT option, 7654
QUANTILE= option, 7654

QUANTLIFE procedure, OUTPUT statement, 7654
keyword= option, 7654
OUT= option, 7654
PREDICTED keyword, 7654
RESIDUAL keyword, 7654
SAMPLEWEIGHT keyword, 7654

QUANTLIFE procedure, PROC QUANTLIFE
statement, 7647
ALPHA= option, 7648
CI= option, 7648
DATA= option, 7648
GRIDSZIE= option, 7648
INITTAU= option, 7648
KAPPA= option, 7648
LOG option, 7649
MAXIT= option, 7649
METHOD= option, 7649

- NAMELEN= option, [7649](#)
- NREP= option, [7649](#)
- NTHREADS= option, [7649](#)
- OUTBOOTEST= option, [7649](#)
- PLOT= option, [7649](#)
- SEED= option, [7650](#)
- TOLERANCE= option, [7650](#)
- QUANTLIFE procedure, TEST statement, [7655](#)
- QUANTLIFE procedure, WEIGHT statement, [7655](#)
- RESIDUAL keyword
 - OUTPUT statement (QUANTLIFE), [7654](#)
- SAMPLEWEIGHT keyword
 - OUTPUT statement (QUANTLIFE), [7654](#)
- SEED= option
 - PROC QUANTLIFE statement, [7650](#)
- STDP keyword
 - OUTPUT statement (QUANTLIFE), [7654](#)
- TEST statement
 - QUANTLIFE procedure, [7655](#)
- TOLERANCE= option
 - PROC QUANTLIFE statement, [7650](#)
- TRUNCATE option
 - CLASS statement (QUANTLIFE), [7652](#)
- WEIGHT statement
 - QUANTLIFE procedure, [7655](#)