

# Modeling the Reliability of Ball Bearings

Chrys Caroni  
National Technical University of Athens

*Journal of Statistics Education* Volume 10, Number 3 (2002), [jse.amstat.org/v10n3/datasets.caroni.html](http://jse.amstat.org/v10n3/datasets.caroni.html)

Copyright © 2002 by Chrys Caroni, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the author and advance notification of the editor.

---

**Key Words:** Failure times; Multiple linear regression; Percentiles; Weighted least squares.

## Abstract

A data set containing  $n = 210$  observations and published by Lieblein and Zelen (1956) provides a useful example of multiple linear regression applied to an engineering problem. It relates percentiles of the failure time distribution for ball bearings to characteristics of the bearings (load, ball diameter, number of balls) in a theoretically derived equation that can be put into linear form. The analysis requires testing the equality of regression coefficients between manufacturers and between types of ball bearing within manufacturer to see if the same equation applies across the industry. Furthermore, there is special interest in confirming an accepted value for one of these coefficients. The original analysis employed weighted least squares, although this may have been unnecessary. In addition to the regression aspects of the problem, the example is useful for the extensive data manipulation required.

## 1. Introduction

This paper presents an example of the application of multiple linear regression analysis to an engineering problem, namely, to the reliability of ball bearings. The literature does not contain many good, real-life illustrations of multiple regression analysis in the applied sciences. One reason may be that commercial confidentiality often prevents the publication of data. The data given here, taken from a study by Lieblein and Zelen (1956, Tables A-1 to A-3), are old but they are not outdated. The context is very simple yet offers scope for fitting and testing various models. Instead of the exploratory type of regression modeling as provided in many examples, particularly from the social sciences, the problem starts with a theoretically derived relationship that can be linearised. Particular hypotheses concerning the coefficients of this model have to be tested, in order to confirm whether the model can be applied across the whole industry instead of being specific to each manufacturer or each bearing type, and in order to confirm the generally assumed value of the most important of these coefficients. As will be seen in the development of the relevant models in the following sections, quite a large amount of data manipulation is called for in order to carry out these tests. This aspect of the problem increases the value of these data in an applied statistics course.

## 2. The Study of the Fatigue Life of Deep-Groove Ball Bearings

Manufacturers who use bearings in their products have an obvious interest in the reliability of these components. The basic measure of reliability in this context is the rating life, defined as the number of revolutions that 90% of

a group of identical bearings would be expected to achieve. This can be symbolized as  $L_{10}$ , the tenth percentile of the distribution of a bearing's lifetime. The current ISO Standard 281 gives the relationship

$$L_{10} = (C / P)^p, \quad (1)$$

where  $C$  is called the basic dynamic load rating (the constant load which would give a rating life of one million revolutions) and  $P$  is the load on the bearing in operation. The exponent  $p$  takes the value of 3 for ball bearings. Lieblein and Zelen's data describe an investigation of this relationship. Their paper starts with the same equation (1), which they say was generally accepted by ball-bearing manufacturers on the basis of many years of experience. However, it appears that there was at that time some disagreement about whether the value  $p = 3$  was correct. Therefore, a main objective of their paper was to estimate the value of the exponent  $p$ . They carried out their study on behalf of the American Standards Association and the National Bureau of Standards, using the results of experimental tests of the lifetimes of batches of deep-groove ball bearings, supplied by manufacturers collaborating in the study.

Quoting from Lundberg and Palmgren (1947), Lieblein and Zelen (1956) give the specific form taken by (1) for ball bearings as:

$$L = \left\{ \frac{f Z^a D^b}{P} \right\}^p, \quad (2)$$

relating a percentile ( $L$ ) of the lifetime distribution to load ( $P$ ), ball diameter ( $D$ ) and number of balls in the bearing ( $Z$ ). In this equation,  $f$ ,  $a$ , and  $b$ , as well as  $p$ , are constants. This equation applies to any percentile  $L$ ; the rating life  $L_{10}$  and the median life  $L_{50}$  are used in the paper.

Taking logarithms puts equation (1) into the linear form

$$\ln(L) = \alpha + \beta_1 \ln(Z) + \beta_2 \ln(D) + \beta_3 \ln(P), \quad (3)$$

in which the main parameter of interest,  $p$ , is given by  $-\beta_3$ .

The data that were analysed by Lieblein and Zelen consisted of 210 endurance tests carried out by three companies on their own ball bearings. Each test involved running a batch of bearings of the same type under the same conditions (thus  $P$ ,  $D$ , and  $Z$  are fixed for each test) and observing the fatigue failure times.  $L_{10}$  and  $L_{50}$  were estimated by fitting a Weibull distribution to the failure times, separately for each batch. (Lieblein and Zelen explain in detail their method of fitting the Weibull distribution, based on order statistics; nowadays, maximum likelihood would probably have been used.) Thus each test batch provided a data point, consisting of the random dependent variables  $L_{10}$  or  $L_{50}$  (to be analysed separately) and the non-stochastic predictors  $P$ ,  $D$ , and  $Z$ . Equation (3) -- or extensions of it, as seen below -- was fitted to these data by least squares. Since the number of bearings per test varied considerably, from 8 to 94, weighted least squares was employed, under the assumption that the error term required on the right-hand side of (3) would have variance inversely proportional to the number of bearings in the test.

The Weibull assumption was not tested in the published study. It was supported partly theoretically, using the "weakest-link" justification of the Weibull as a failure distribution, and partly empirically, quoting (but not showing) the probability plots of the test results. Lieblein and Zelen gave one example of the detailed worksheets that they received for each test. The failure times for this one case are now widely used in the reliability literature (see, for example, Meeker and Escobar 1998) as an example of fitting the Weibull and other distributions, although the data are always quoted wrongly as uncensored whereas in fact three values were censored (Caroni 2002).

### 3. Analysis

The results given here were obtained by fitting regression models using weighted least squares, as in the original study. However, if an unweighted analysis is used and the residuals are examined in relation to the number of bearings in the test, there does not seem to be any sign of the expected heteroscedasticity. Therefore, similar results should be obtained using ordinary least squares. Although it seems logical to expect a connection with the number of bearings in the test, other factors are important too. For example, it might be the case that tests with a smaller number of bearings tended to be run for longer than tests with a larger number, so that in the end the information provided did not vary very much between tests. Of course, this is just speculation. If the original paper had provided standard errors of the estimates of  $L_{10}$  and  $L_{50}$ , these could also have been used for weighting.

The main analysis consists of fitting a series of regression models to examine the following hypotheses:

- a. all the parameters of (3) are the same for each one of the three companies;
- b. the parameter  $\beta_3$  (hence,  $p$ ) is the same for each company;
- c. all the parameters of (3) are the same for each one of three types of bearing produced by Company B;
- d. the parameter  $-\beta_3$  is the same for each type of bearing produced by Company B.

It is not certain from Lieblein and Zelen which batches correspond to each type of bearing, but I have assumed that the first 37 tests in the list for Company B are type 1, the next 94 are type 2 and the remaining 17 are type 3. These are the frequencies given in the paper, and the values of  $Z$  change abruptly at these points in the list. The results of the analysis using this division are close to the published results.

These hypotheses are tested in the following Tables 1 and 2 by constructing the appropriate  $F$ -tests from the residual sums of squares of nested models. For example, to test hypothesis (a), we first fit the model (3), giving the row of results labeled “parameters same for each company” in Table 1. (All data analysis was carried out using Minitab 13). We next fit the model in which all the parameters of (3) are allowed to vary freely between companies. This is

**Table 1.** Tests on regression coefficients between the three companies.

		Dependent $L_{10}$		Dependent $L_{50}$	
Model	df	SS	MS	SS	MS
1. All parameters free	198	1944.52	9.82	1485.91	7.50
2. Same parameters for each company	206	2215.41		1731.67	
Model 2 - Model 1 Difference	8	270.89	33.86	245.76	30.72
$F$ -ratio	8, 198	$F = 3.45$ , $p$ -value = 0.001		$F = 4.10$ , $p$ -value = 0.0002	
3. Common $p$	200	1950.11		1498.59	
Model 3 - Model 1 Difference	2	5.59	2.79	12.68	6.34
$F$ -ratio	2, 198	$F = 0.28$ , $p$ -value = 0.76		$F = 0.85$ , $p$ -value = 0.43	

**Table 2.** Tests on regression coefficients between the three bearing types from Company B.

		Dependent $L_{10}$		Dependent $L_{50}$	
Model	df	SS	MS	SS	MS
1. All parameters free	136	1258.26	9.25	861.45	6.33
2. Same parameters for each type	144	1405.80		999.97	
Model 2 - Model 1 Difference	8	147.54	18.44	138.52	17.32
$F$ -ratio	8, 136	$F = 1.99,$ $p$ -value = 0.052		$F = 2.73,$ $p$ -value = 0.008	
3. Common $p$	138	1291.79		900.01	
Model 3 - Model 1 Difference	2	33.52	16.76	38.56	19.28
$F$ -ratio	2, 136	$F = 1.81,$ $p$ -value = 0.17		$F = 3.05,$ $p$ -value = 0.051	
4. $p = 3$	139	1291.80		903.10	
Model 4 - Model 1 Difference	3	33.54	11.18	41.65	13.88
$F$ -ratio	3, 136	$F = 1.21,$ $p$ -value = 0.30		$F = 2.19,$ $p$ -value = 0.09	

$$\begin{aligned}
\ln(L) = & \alpha + \beta_1 \ln(Z) + \beta_2 \ln(D) + \beta_3 \ln(P) + \beta_4 B + \beta_5 C \\
& + \beta_6 B \times \ln(Z) + \beta_7 B \times \ln(D) + \beta_8 B \times \ln(P) \\
& + \beta_9 C \times \ln(Z) + \beta_{10} C \times \ln(D) + \beta_{11} C \times \ln(P)
\end{aligned} \quad (4)$$

where  $B$  and  $C$  are dummy variables denoting tests supplied by Companies  $B$  and  $C$ , respectively, and terms such as  $B \times \ln(Z)$ , are interaction terms obtained by multiplying the separate terms ( $B$  and  $\ln(Z)$  in this case).

The results from this fit provide the “all parameters free” row of results in [Table 1](#). The hypothesis (a) of equality is tested in the usual way by an  $F$ -test, using the difference in the sum of squares between models (3) and (4): this is often called the “extra sum of squares method” ([Krzanowski 1998](#)). Clearly, this hypothesis is rejected for both  $L_{10}$  and  $L_{50}$ . To test hypothesis (b), we must fit model (4) omitting the terms with coefficients  $\beta_8$  and  $\beta_{11}$ . This gives the results in the row labeled “common  $p$ ” in [Table 1](#). The  $F$ -test shows that hypothesis (b) holds, both for  $L_{10}$  and  $L_{50}$ . The estimate of the common value of  $p$  (that is, minus the coefficient of  $\ln(P)$  in the “common  $p$ ” regression) is 2.876 (standard error 0.178) for  $L_{10}$  and 2.804 (0.156) for  $L_{50}$ . Using a significance level of 5% in our tests, we observe that the proposed value of  $p = 3$  falls well within a 95% confidence interval and is therefore supported by the data.

The usual diagnostic methods applied to these regressions confirm that the model (2) is appropriate. To take the example of the “common  $p$ ” regression for  $L_{10}$ , we find, in particular, that the standardized residuals are close to a normal distribution and do not contain any excessively large values considering that there are 210 observations. Cook’s distance statistic is relatively large for observation number 43 in the file, taking the value 0.43 (the next largest is 0.13). This test employed the highest diameter and was run with the highest load. The

observed  $L_{10}$  was only 3.01, compared to a fitted value of 12.64 (standardized residual -3.05, the second largest in absolute value). However, omitting this point changes the estimate of  $p$  trivially, to 2.930 (standard error 0.178).

Hypotheses c and d are tested in a similar way, after first restricting the analysis to the subset of tests from Company B and replacing the dummy variables for companies in (4) by dummy variables for types. Results are in [Table 2](#). Hypothesis (c) is clearly rejected for  $L_{50}$ , but the result for  $L_{10}$  is unclear ( $p$ -value = 0.052). Hypothesis (d) is accepted for  $L_{10}$  but the result is unclear for  $L_{50}$  ( $p$ -value = 0.051). Rejecting hypothesis (d) for  $L_{50}$  - a finding of differences between types within manufacturer  $B$  - appears to be in conflict with the earlier finding of no difference between manufacturers in b. However, the latter result has to be considered in a way as an average across types of bearing.

At this point, the meaning of the significance levels could be discussed. There is a case for adjusting them for multiple testing. For example, a simple Bonferroni adjustment could be used for an overall Type I error of at most 0.05 when testing both  $L_{10}$  and  $L_{50}$ , by comparing the pairs of  $p$ -values with 0.025. We then reject hypothesis (c) but clearly accept hypothesis (d).

## 4. Further Comments

As seen above, the data file must be manipulated in various ways in order to carry out the various regression analyses that are required. Apart from taking logarithms of the dependent variables and the predictors, it is necessary to create dummy variables for Company and type of bearing, to create interaction variables by multiplying the predictors by these dummy variables, and to subset the data by company. If the instructor leaves all these operations to the student, this study is not just an instructive example of multiple linear regression, but also provides a useful and quite extensive exercise in handling data.

Some further aspects of the regression problem can also be examined. For example, instead of our hypothesis (d) above, Lieblein and Zelen carried out a direct test of the hypothesis  $p = 3$  for each type of bearing, using the data on Company B. Substituting  $p = 3$  in (2) and rearranging, we obtain

$$\ln(L) + 3 \ln(P) = \alpha + \beta_1 \ln(Z) + \beta_2 \ln(D), \quad (5)$$

from which we see that tests of  $p = 3$  can be carried out by means of the device of constructing new dependent variables  $Y_i = \ln(L_i) + 3 \ln(P)$  for  $i = 10, 50$  and fitting suitable models. In fact, since the earlier results show that we should not assume that the coefficients of  $\ln(Z)$  and  $\ln(D)$  are the same for each type of bearing, we must fit a model similar to (4) but without the terms in  $\ln(P)$ .

$$Y_i = \alpha + \beta_1 \ln(Z) + \beta_2 \ln(D) + \beta_4 T_2 + \beta_5 T_3 + \beta_6 T_2 \times \ln(Z) \\ + \beta_7 T_2 \times \ln(D) + \beta_8 T_3 \times \ln(Z) + \beta_{10} T_3 \times \ln(D) \quad (6)$$

where  $T_2$  and  $T_3$  are dummy variables for the second and third types of manufacturer B's bearings. This gives the row " $p = 3$ " of [Table 2](#).

The testing of specific values can be taken further. [Lieblein and Zelen \(1956\)](#) quoted the values  $a = 2/3$ ,  $b = 1.8$  (as well as  $p = 3$ ) from [Lundberg and Palmgren \(1947\)](#) and tested whether these values were confirmed simultaneously by the data. They did this separately for each company and separately for each type of bearing from Company B. The conclusion was that the data for  $L_{10}$  were consistent with these assumed values except for bearing type 3. For  $L_{50}$ , they were consistent only for Company A.

An alternative to fitting the model (4) in order to obtain the sum of squares for the model with all parameters free, is to fit the model (3) to each company separately, then add the residual sums of squares from the three analyses. This is what Lieblein and Zelen actually did, since with the computing facilities available to them at

that time it must have been preferable to carry out three regressions with three predictors rather than one regression with eleven predictors. To repeat the analysis by this method might be a useful exercise to help the student to understand the meaning of the model (4).

Finally, the student can be shown the trick of carrying out the weighted least squares analysis by ordinary least squares (Ryan 1997). To do this, we take a regression model such as (3) in which the error term's variance is  $\sigma^2/n_i$ , where  $n_i$  is the number of bearings in the  $i^{\text{th}}$  test, and multiply throughout model (4) and the restricted model ("parameters same") by the square root of the weight, that is,  $\sqrt{n_i}$ . We observe that the resulting equation is just an unweighted regression (error variance  $\sigma^2$ ) through the origin (that is, without an intercept), in which the original dependent variable and predictors have all been multiplied by the factor  $\sqrt{n_i}$ , and  $\sqrt{n_i}$  itself is also included as a predictor. These calculations can be carried out in the data analysis package and the regression results verified.

## 5. Obtaining the Data

The file [ballbearings.txt](#) contains a description of the data contained in the file [ballbearings.dat.txt](#), which is in fixed-column ASCII format. The format of the file is described in the [Appendix](#).

---

## Appendix - Key to the variables in [ballbearings.dat.txt](#)

Columns	Variable	Comment
1	Company	Codes 1, 2, and 3 for Companies A, B, and C
2-4	Test number	1, 2, ... within company
7-10	Year of test	9999 = missing (not used in analysis)
12-13	No. of bearings	Weighting variable
15-19	Load (P)	
21-22	No. of balls (Z)	
24-30	Diameter (D)	Format F7.5
32-38	L10	Format F7.3
41-46	L50	Format F6.2
48-51	Weibull slope	Format F4.2 (not used)
53	Bearing type	1, 2, and 3 in Company B; 0 otherwise

---

## References

- Caroni, C. (2002), "The correct 'ball bearings' data," *Lifetime Data Analysis*, 8, 395-399.
- Krzanowski, W. J. (1998), *An Introduction to Statistical Modelling*, London: Arnold.
- Lieblein, J., and Zelen, M. (1956), "Statistical investigation of the fatigue life of deep-groove ball bearings", *Journal of Research of the National Bureau of Standards*, 57, 273-316.
- Lundberg, G., and Palmgren, A. (1947), "Dynamic capacity of rolling bearings", *Acta Polytechnica, Mechanical Engineering Series*, 1(3).
- Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: John Wiley and Sons, Inc.
- Ryan, T. P. (1997), *Modern Regression Methods*, New York: John Wiley and Sons, Inc.

---

Chrys Caroni  
Department of Mathematics  
National Technical University of Athens  
GR 157 80 Athens  
Greece  
[\*ccar@math.ntua.gr\*](mailto:ccar@math.ntua.gr)

---

[Volume 10 \(2002\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Information Service](#) | [Editorial Board](#) | [Guidelines for Authors](#) |  
[Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)