Modeling the Reliability of Ball Bearings

Introduction To Statistics

V. Sai Prabhav G. Sai Shashank Revanth Sairam

BMAT 2438, BMAT 2414, BMAT 2403

Main Insights

- Example of Multiple Linear Regression applied to an engineering problem.
- Real-life cases are rare due to commercial confidentiality.
- Uses classic data from Lieblein and Zelen (1956) – still relevant today.
- Focus: Testing a theoretical model and key coefficient assumptions.
- Aim: Check if the model applies industry-wide, not just to specific manufacturers.

The Study of the Fatigue Life of Deep-Groove Ball Bearings

- Manufacturers rely heavily on the **Reliability Of Ball Bearings**, as failures can lead to significant operational issues.
- The standard metric used to quantify this reliability is **rating life**, which refers to the number of revolutions at which 90% of a group of identical bearings are expected to function without failure.

Basic Fatigue-Life Model

This is represented mathematically as:

$$L_{10} = \left(\frac{C}{P}\right)^{p}$$

- L_{10}: Rating life (tenth percentile of the lifetime distribution)
- C: Basic dynamic load rating (load resulting in a life of one million revolutions)
- P: Applied load on the bearing in operation
- p: Exponent reflecting the relationship between load and life (traditionally assumed $p = 3$)

What Is Basic Dynamic Load Rating

- The **Basic Dynamic Load Rating** (denoted as C) is a fundamental parameter in bearing engineering.
- It represents the constant load a bearing can theoretically endure for a specified number of revolutions without exhibiting material fatigue.
- Specifically, for **Radial Ball Bearings**, this is typically defined as the load under which the bearing will achieve a basic rating life of 1million revolutions with 90% reliability.
- This is calculated theoretically through **Empirical Formulas**.

Generalized Fatigue-Life Model

In Lieblein Zelen (1956) and Lundberg Palmgren (1947), the fatigue-life equation is written as:

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

- L: Bearing life at a given percentile (e.g. L_{10}, L_{50})
- P: Applied dynamic load
- Z: Number of balls in the bearing
- D: Diameter of each ball
- f: Proportionality constant
- a, b: Empirically determined exponents
- p: Fatigue-life exponent (typically $p = 3$ for ball bearings)

Modeling the Dynamic Load Rating in Terms of Bearing Characteristics

C in terms of Z and D:

- **Empirical Basis**: Studies show bearing life increases with more balls Z and larger ball diameter D.
- **Mechanical Logic**: More/larger balls reduce contact stress and improve durability.
- **Modeling Benefit**: Converts abstract C into measurable terms for better analysis.
- **Linearization**: Enables a linear regression model after taking logarithms.

Log-Linear Form

Taking natural logs gives a regression-ready model:

$$\ln(L) \; = \; \alpha \; + \; \beta_1 \ln(Z) \; + \; \beta_2 \ln(D) \; + \; \beta_3 \ln(P)$$

- $\alpha = p \ln(f)$
- $\beta_1 = a\,p, \; \beta_2 = b\,p, \; \beta_3 = -p$
- We'll estimate β_3 to test whether $p = 3$ holds.

ISO Standard and the Study's Goal

- The ISO Standard 281 provides the previous equation as the accepted model for bearing fatigue life.
- Historically, the exponent $p = 3$ has been widely accepted based on manufacturer experience.
- However, Lieblein and Zelen's study aimed to critically examine this assumption.
- At the time, there was some disagreement about whether $p = 3$ was indeed accurate across all bearing types and applications.

Summary of Ball Bearing Test Data

- Collected over years from 4 major ball-bearing companies (anonymized as A, B, C, D).
- Each endurance test: bearings of same type, tested together under identical conditions (same load & speed).
- Company B's data detailed further into three types: B-1, B-2, B-3.
- Test results recorded:
  - Number of revolutions before fatigue failure (in millions)
  - Bearing characteristics: number of balls Z, ball diameter D, load P, etc.
- Key Variables Computed from Data:
  - L_{10}: Life by which 10% of bearings fail (rating life)
  - L_{50}: Median life
  - Weibull slope e: Measures dispersion in failure times (Not considered in our study)

Hypothesis of Our Test

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

(a) All the parameters of the equation are the same for each of the three companies.

(b) The parameter β_3 (hence, p) is the same for each company.

(c) All the parameters of the equation are the same for each of the three bearing types produced by Company B.

(d) The parameter β_3 is the same for each bearing type produced by Company B.

(e) The parameter β_3 is 3 for each bearing type produced by Company B.

F-Test: Hypotheses

- $\mathbf{H_0}$ (Null): The simpler (nested) model is sufficient. That is, the additional parameters in the complex model are not statistically significant (their coefficients are zero).
- $\mathbf{H_1}$ (Alternative): The more complex model provides a significantly better fit. At least one of the additional parameters is nonzero.
- *Note*: The simpler model must be a special case of the more complex model.

F-Test: Statistic

$$F = \frac{(RSS_0 - RSS_1)/(p_1 - p_0)}{RSS_1/(n - p_1)}$$

- RSS_0: Residual Sum of Squares from the simpler model
- RSS_1: Residual Sum of Squares from the complex model
- p_0, p_1: Number of parameters in the models
- n: Total number of observations

F-Test: Explanation

- The numerator $(RSS_0 - RSS_1)/(p_1 - p_0)$ measures the average reduction in error due to adding extra predictors.
- The denominator $RSS_1/(n - p_1)$ estimates the variance of the residuals in the complex model.
- If the added parameters do not improve the model, the F-statistic will be low.
- A high F-statistic indicates that the complex model significantly reduces unexplained variance.
- Under H_0, the F-statistic follows an F-distribution with degrees of freedom $(p_1 - p_0,\ n - p_1)$.

F-Test: A Simple Example

Models:

$$\text{Simpler: } y = \beta_0 + \beta_1 x_1 + \varepsilon,$$
$$\text{Complex: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$
$$n = 50.$$

Given:

$$RSS_0 = 200, \quad RSS_1 = 150, \quad p_0 = 2, \quad p_1 = 3.$$

$$F = \frac{(200 - 150)/(3 - 2)}{150/(50 - 3)} \approx \frac{50}{150/47} \approx 15.67.$$

Conclusion: The large F-value suggests the complex model is significantly better.

Hypothesis (a): Parameter Equality Across Companies

- **Null Hypothesis (H_0)**: All model parameters $\alpha, \beta_1, \beta_2, \beta_3$ are identical for each of the three companies.
- **Alternative Hypothesis (H_1)**: At least one of $\alpha, \beta_1, \beta_2, \beta_3$ differs between companies.

Hypothesis (a): Parameter Equality Across Companies

```
Analysis of Variance Table

Model 1: ln_L10 ~ (ln_Z + ln_D + ln_P)
Model 2: ln_L10 ~ Company * (ln_Z + ln_D + ln_P)
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    206 2215.4
2    198 1944.5  8    270.88 3.4478 0.0009674 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1
```

The p-value is exremely small hence we reject the null hypothesis and conclude that at least one of $\alpha, \beta_1, \beta_2, \beta_3$ differs between companies.

Hypothesis (b): Parameter Equality Across Companies except for p

- **Null Hypothesis (H_0)**: Model parameter β_3 is identical for each of the three companies (Other parameters may deffer).
- **Alternative Hypothesis (H_1)**: The Model parameter β_3 differ between companies.

Hypothesis (b): Parameter Equality Across Companies except for p

```
Analysis of Variance Table

Model 1: ln_L10 ~ Company * (ln_Z + ln_D) + ln_P
Model 2: ln_L10 ~ Company * (ln_Z + ln_D + ln_P)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    200 1950.1
2    198 1944.5  2    5.5869 0.2844 0.7527
```

p-value is large so we cannot reject the H_0

Hypothesis (c): Parameter Equality Across Types for Company B

- **Null Hypothesis (H_0)**: All model parameters $\alpha, \beta_1, \beta_2, \beta_3$ are identical for each of the three Types in company B.
- **Alternative Hypothesis (H_1)**: At least one of $\alpha, \beta_1, \beta_2, \beta_3$ differs between Types of company B.

Hypothesis (c): Parameter Equality Across Types for Company B

```
Analysis of Variance Table

Model 1: ln_L10 ~ ln_Z + ln_D + ln_P
Model 2: ln_L10 ~ Bearing_type * (ln_Z + ln_D + ln_P)
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1    144 1405.8
2    136 1258.3  8    147.54 1.9934 0.05169 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The p-value is not large; hence, we may consider to reject the H_0 i,e. At least one of $\alpha, \beta_1, \beta_2, \beta_3$ differs between Types of company B.

Hypothesis (d): Parameter Equality Across Types except for p

- **Null Hypothesis (H_0)**: Model parameter β_3 is identical for each of the three Types (Other parameters may deffer).
- **Alternative Hypothesis (H_1)**: The Model parameter β_3 differ between Types.

Hypothesis (d): Parameter Equality Across Types except for p

```
Analysis of Variance Table

Model 1: ln_L10 ~ Bearing_type * (ln_Z + ln_D + ln_P)
Model 2: ln_L10 ~ Bearing_type + ln_Z + Bearing_type:ln_D + l
    Bearing_type:ln_D
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    136 1258.3
2    138 1291.8 -2   -33.538 1.8125 0.1672
```

p-value is large so we cannot reject the H_0

Hypothesis (e): Parameter $p = 3$ Across Types of Company C

- **Null Hypothesis (H_0)**: Parameter $p = 3$ Across Types of company C
- **Alternative Hypothesis (H_1)**: Parameter $p \neq 3$ Across Types of company C

Hypothesis (e): Parameter $p = 3$ Across Types of company C

```
Analysis of Variance Table

Model 1: ln_L10 ~ +Bearing_type * (ln_Z + ln_D)
Model 2: ln_L10 ~ Bearing_type * (ln_Z + ln_D + ln_P)
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1    139 1291.8
2    136 1258.3  3    33.585 1.21 0.3086
```

p-value is large so we cannot reject the H_0