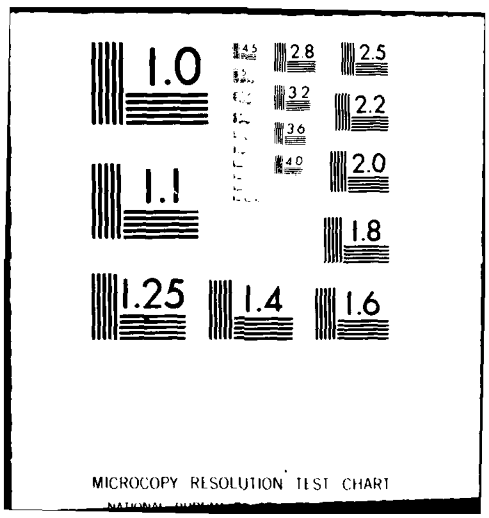
AD-A110 727 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL--ETC F/G 11/6
STATISTICAL ANALYSIS OF METALLURGICAL MECHANICAL PROPERTIES WIT--ETC(U)
SEP 81 G A KILLIAN
UNCLASSIFIED AFIT-LSSR-86-81 NL

1 of 2
AD A
110727


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

This document has been approved for public release and sale; its distribution is unlimited.

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY (ATC)

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

STATISTICAL ANALYSIS OF METALLURGICAL
MECHANICAL PROPERTIES
WITH AN APPLICATION TO
Ti-6Al-4V ALLOY FATIGUE DATA

Gary A. Killian, 2Lt, USAF

LSSR 86-81

The contents of the document are technically accurate, and no sensitive items, detrimental ideas, or deleterious information are contained therein. Furthermore, the views expressed in the document are those of the author(s) and do not necessarily reflect the views of the School of Systems and Logistics, the Air University, the Air Training Command, the United States Air Force, or the Department of Defense.

AFIT Control Number LSSR 86-81

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to determine the potential for current and future applications of AFIT thesis research. Please return completed questionnaires to: AFIT/LSH, Wright-Patterson AFB, Ohio 45433.

1. Did this research contribute to a current Air Force project?

   a. Yes b. No

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not researched it?

   a. Yes b. No

3. The benefits of AFIT research can often be expressed by the equivalent value that your agency received by virtue of AFIT performing the research. Can you estimate what this research would have cost if it had been accomplished under contract or if it had been done in-house in terms of manpower and/or dollars?

   a. Man-years ________ $ ________ (Contract).

   b. Man-years ________ $ ________ (In-house).

4. Often it is not possible to attach equivalent dollar values to research, although the results of the research may, in fact, be important. Whether or not you were able to establish an equivalent value for this research (3 above), what is your estimate of its significance?

   a. Highly Significant b. Significant c. Slightly Significant d. Of No Significance

5. Comments:

________________________ ________________________
Name and Grade Position

________________________ ________________________
Organization Location

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER LSSR 86-81	2. GOVT ACCESSION NO. AD A110 727	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STATISTICAL ANALYSIS OF METALLURGICAL MECHANICAL PROPERTIES WITH AN APPLICATION TO Ti-6Al-4V ALLOY FATIGUE DATA		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gary A. Killian, Second Lieutenant, USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Systems and Logistics Air Force Institute of Technology, WPAFB OH		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Department of Communication and Humanities AFIT/LSH, WPAFB OH 45433		12. REPORT DATE September 1981
		13. NUMBER OF PAGES 139
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

22 JAN 1982

Fredric C. Lynch

18. SUPPLEMENTARY NOTES

FREDRIC C. LYNCH, Major, USAF
Director of Public Affairs

APPROVED FOR PUBLIC RELEASE AFR 190-17.

Air Force Institute of Technology (ATC)
Wright-Patterson AFB, OH 45433

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

1) Metallurgy
2) Fatigue
3) Titanium Alloy
4) Linear Regression
5) Nonlinear Regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Thesis Chairman: James N. Bexfield, Lt Col, USAF

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The potential application of statistical methods to metallurgical practices was shown using an example of titanium alloy fatigue data. Materials Laboratory, AFWAL was interested to see (1) whether more meaningful S-N curves could be drawn using statistical methods, (2) which of the data can be identified as outliers, and (3) how the different treatments of the alloy compare. Best-fit curves were determined for each treatment by linear and nonlinear regression analysis. Residual analysis was used to test the assumptions on the random error and to select extreme values for further analysis. Seven of the nine extreme values found did not lie within ninety-nine percent prediction intervals determined about the fitted line and were classified as outliers to be fractographically examined by Materials Laboratory. Regression results of sets of treatments, hypothesized by Materials Laboratory to be drawn from similar populations, were compared to determine whether the treatments within a set agree with the Materials Laboratory's hypothesis.

LSSR 86-81

STATISTICAL ANALYSIS OF METALLURGICAL MECHANICAL PROPERTIES WITH AN APPLICATION TO Ti-6Al-4V ALLOY FATIGUE DATA

A Thesis

Presented to the Faculty of the School of Systems and Logistics

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the Requirement for the

Degree of Master of Science in Systems Management

By

Gary A. Killian, BS
Second Lieutenant, USAF

September 1981

Approved for public release;
distribution unlimited

This thesis, written by

2Lt Gary A. Killian

has been accepted by the undersigned on behalf of the faculty of the School of Systems and Logistics in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS MANAGEMENT

DATE: 30 September 1981

James N. Bexfield 21 Sept 1981
COMMITTEE CHAIRMAN

30 Sept. '81
READER

ACKNOWLEDGEMENTS

For providing the topic for this thesis and aiding in reviewing my work and offering suggestions all along the way, I am grateful to my technical reader, Dr. F. H. Froes. In addition, I would like to acknowledge the supply of the technical data analyzed in this thesis along with the background information from Dr. D. Eylon, Mr. P. Smith and Mr. G. Lovell.

I also wish to thank my advisor, Lieutenant Colonel James N. Bexfield, for this expert guidance in the field of statistics and his patience and understanding throughout the whole thesis effort.

Finally, I would like to express my gratitude to Mrs. Dolores Newsom, who came to my aid and typed the final manuscript.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The application of statistical methods of data analysis to metallurgical practices is not widespread. In many cases, data analysis is done without using known statistical methods, i.e., by fitting a line or curve using "eye-balling." Use of statistical analysis provides a more rigorous analysis of the data, thus allowing strict and meaningful comparison of different data sets. Why, then, are statistical methods not used more often, thus allowing more analytical precision?

It appears the dominant reason is the separation of disciplines between Metallurgy and Statistical Methods. The goal of this thesis is to examine the potential uses of statistical methods to metallurgical data and to formulate reproducible models for application. Several potential applications will be discussed. One application was selected for detailed development. This application applies statistical analysis to titanium alloy fatigue data.

Background for Example

Titanium alloys are desirable materials for use in aerospace systems. They have high specific strength (strength-to-density), excellent fracture resistant characteristics, and outstanding general corrosion resistance. However, despite these desirable characteristics, titanium alloys have seen limited use in advanced systems. In fact, rather than increasing in recent years, titanium use has decreased. This is shown in Table 1.1, which lists titanium use in conceptual and final design stages.

TABLE 1.1

Cost Impact of Titanium Use (9)

	% Titanium	
Aircraft	Early Concept	Final Design
F-15	50	34
B-1	42	22
C-5	24	3

The reason for this limited use is the high cost of titanium components: a result of high initial costs combined with high processing costs (forging and machining). In order to combat these high costs, the USAF has identified the cost factors associated with the various stages of going from titanium ore to final assembled product. These are shown in Table 1.2 (5).

It is immediately obvious that the component fabrication (forging, machining into sheet, etc.) is a major cost item. Because of this, the Materials Laboratory of the Air Force Wright Aeronautical Laboratories (AFWAL) has sponsored many programs in the past ten years in net shape (or near

TABLE 1.2

COST BREAKDOWN OF TITANIUM COMPONENTS

Product	Current Cost[10] ($) (per pound of Ti)	Added Cost ($) (per pound of Ti)
RUTILE(TiO_2)[1]	0.10-0.25	0.10-0.25
TICKLE($TiCl_4$)[2]	1.00-2.00	0.75-1.90
SPONGE	7.00-22.00[3]	5.00-21.00
INGOT	9.00[4]	2.00
MILL PRODUCT		
Sheet	9.00-25.00[5]	up to 16.00
Foil	300.00[6]	300.00
Forgings	10.00-15.00	up to 6.00
Plate	12.00-17.00	up to 8.00
Rod	12.00-20.00	up to 11.00
Tubing[9]	10.00-20.00	up to 11.00
Forgings[7]	150.00-300.00	up to 300.00
Sheet Metal[8]	50.00-150.00	up to 150.00

[1] 92-98% pure.

[2] DuPont quote of 25¢/pound; however availability question

[3] Spot market price; must be considered artificially high

[4] Assuming triple melt

[5] Low value CP; 6-4 $20-25 range depending on options such as gage, width, lengths, finish, etc.

[6] 6-4 product

[7] Includes secondary processing (fabrication), machining and inspection costs. As a good rule of thumb the cost of the product doubles during forging and doubles again during machining (though the latter operation is highly dependent on the proportion of rough and final machining: the final operation being approximately 10 times as expensive per pound of stock removed).

[8]SPF/DB projected to be 20% less expensive than conventional fabrication

[9]Low value welded; high value seamless

[10]As of August/September 1980

net shape) technologies. Net shape technology involves producing a component very close to its final shape, thus reducing forging and machining operations and greatly improving the material utilization factor (buy-to-fly ratio). Of particular importance in these new technologies are the mechanical properties of components so produced. One such test of critical importance is the fatigue test.

Fatigue testing is done on a machine which subjects a prepared metal specimen (Figure 1.1) to cyclic stress, generally until failure occurs. Numerous tests are done with specimens of the same alloy at different stress levels to provide data points for plotting of Stress vs. Cycles to Failure (S-N) curves. The curves can then be used to predict the useful (safe) life of a component, given the working stress level. Two important points can be identified during fatigue testing: t_i, time to initiation of a crack, and t_p, time of propagation. t_i is the number of cycles until a discernible crack is detected, and t_p is the number of cycles from t_i until the crack causes failure. For further discussion, reference should be made to a text such as Dieter (2:403-49).

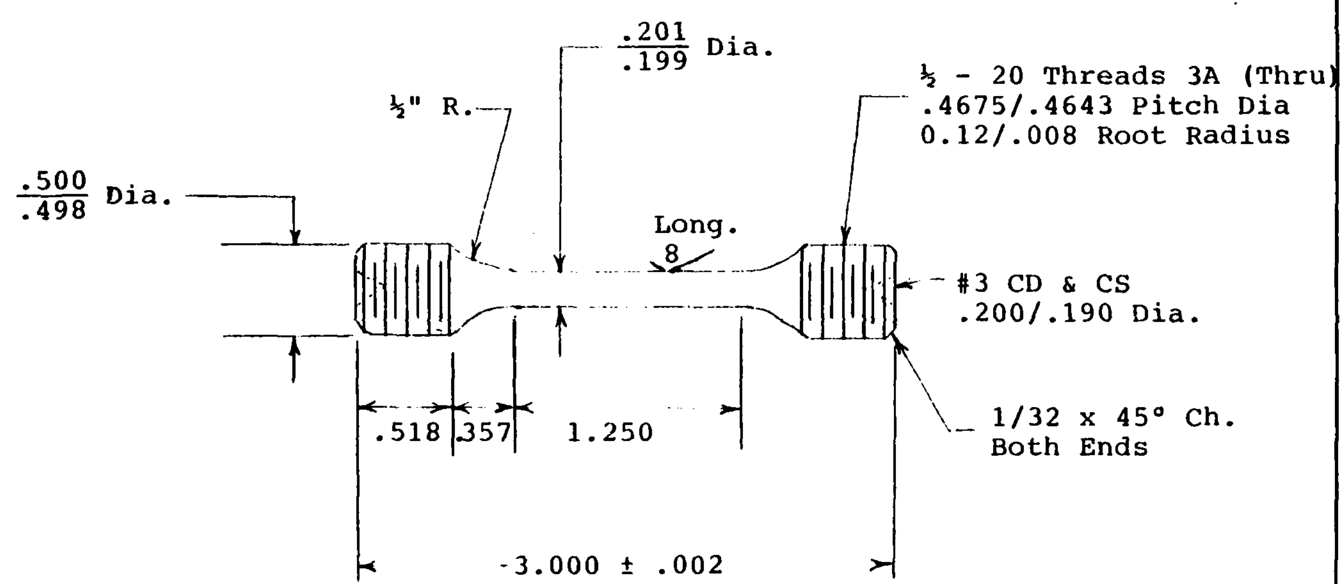


Figure 1.1. Fatigue Specimen

Metallurgical Considerations For Example

In the present program the specific alloy involved is Ti-6Al-4V, a titanium alloy, with six percent Aluminum and four percent Vanadium, along with another titanium alloy, CORONA-5. The Materials Laboratory has been investigating this alloy in many conditions as part of the major net-shape thrust. This includes powder metallurgy (6), cast (5) and wrought product (5). Various conditions have been selected for detailed examination. These materials are pre-alloyed powders, elemental powders, cast alloys, and wrought alloys with several heat treated conditions for those materials.

Statistical Methods

Currently, statistical methods are not used to develop the S-N curves and best-fit lines are "eyeballed." Regression analysis would provide a statistical basis for determining the equation for the line and provide a reproducible model. This rigorous definition of best-fit lines or curves is very important since it allows (a) a strict definition of lower bound 3-sigma (standard deviation) curves: which are generally used for design purposes, and (b) a strict analysis of whether the data from two curves is from the same or different population groups. This latter point is of great metallurgical significance since it is vital to know whether one population (process) is better than another (process). If the data is all from one population then there is no point in pursuing an (generally more costly) alternate process.

An additional problem encountered when developing these curves is outliers, those points not close to the best-fit lines. Determining why or even whether these data points should be considered as outliers is of major concern. Specifically it is of great concern to define statistically whether or not the data represented by these points is from the general population or whether other metallurgical factors are influencing this data causing "low" or "high" values. "Low" values can be anticipated from microstructural discontinuities such as voids or foreign particles while "high" values can arise from microstructures or orientations (texture) which are more fatigue resistant than the general body of a component. It would be prohibitively costly to analyze every failure, but valuable information can be gathered by analyzing cracks resulting from abnormalities. The statistical technique of determining outliers provides a sound basis for predicting "high" or "low" values. Once a point is determined to be an outlier, the specimen corresponding to that point can be analyzed for the cause of its abnormal behavior. That cause can be defined by microstructural or fractographic examination, whether it may be a foreign particle, a surface crack, or other mechanistic explanation.

All information concerning the fatigue tests done on the various conditions of Ti-6Al-4V is available from the Structural Metals Branch of AFWAL. The data is in both tabular and graphical form.

Computer support was supplied by the ASD Computer Center. The REGRESSION and NONLINEAR Subprograms of SPSS (Statistical Package for the Social Sciences) (13:351,368) will be used for the best-fit line and population difference problems, while residual analysis will be applied to the study of the outlier problem.

Statement of the Problem

There appears to be many fruitful applications of statistics to mechanical metallurgical data. Statistical methods have been applied to metallurgical data, as noted in Dieter (2:408); however, the application is not widespread, e.g., in the examination of fatigue data by the Materials Laboratory, AFWAL. This work will briefly discuss several mechanical properties and analyze, as an example, data associated with a Ti-6Al-4V alloy currently being reviewed by the Structural Metals Branch of the AFWAL Materials Laboratory. Statistical analysis may aid in developing more rigorous S-N curves, allow meaningful comparison of the curves, and determine outliers. For the latter, corresponding specimens may then be further examined in detail. A technique known as comparison of regressions will be used to test for differences between processes. The existence of differences between treatments may lead to better decision-making as applied to the choice of processes.

CHAPTER 2

MATERIALS BACKGROUND

Introduction

The purpose of this chapter is to define and discuss some of the basic metallurgical principles. Later, some of these principles will be used in the development of statistical models to represent fatigue data. First, general mechanical properties, which relate to all materials, are discussed, followed by a more detailed description of fatigue. The next section is concerned with general material strengthening, using as an example a specific titanium alloy. Titanium is an important metal to today's aerospace industry. Hence, at the request of Material's Laboratory, AFWAL, fatigue data using the titanium alloys were selected to illustrate the application of regression analysis to metallurgy. The final sections describe the characteristics of titanium and its alloys.

Mechanical Properties

Some of the most important characteristics of a material are its mechanical properties. These properties--e.g. yield strength, tensile strength, fatigue strength, and toughness[1]--will usually determine whether a material can be used for a given application and also the working life of a material component, be it metallic or nonmetallic.

Knowledge of a material's mechanical properties can be most helpful when considering which material to use when, for instance, designing a new fighter aircraft or armored vehicle. In the fighter, high temperature performance and low weight are two important attributes. A strong, light-weight material which retains its good qualities at elevated temperatures caused by high speed air friction is needed. This material could reduce fuel consumption by requiring less thickness while providing the same strength as a thicker, weaker material. In the armored vehicle, weight may not be as important a consideration as good impact fracture resistance. The armor should be made from a material able to withstand large sudden impacts.

A measure of impact fracture resistance is toughness. A common test for toughness is an impact test using a notched

---

[1]Underlined words are defined at the end of this chapter.

specimen, a small bar of material with a V-shaped notch located at the point where a swinging hammer impacts the specimen (15:479).

The hammer is released from a fixed height, strikes the material at its lowest point of swing, and continues through. The difference in height between start and finish gives an indication of the energy expended in breaking the material, and thereby a measure of the toughness.

Returning to the fighter aircraft example, the need for a strong material is obvious, but what constitutes a strong material? One measure of a material's strength is tensile strength, the maximum amount of <u>stress</u> (units of force per square unit of length) in tension or compression a material can be subjected to before failure occurs. It is normally measured on a machine which grips both ends of a prepared specimen with its two crossheads, one of which moves. As the crossheads move apart, a force results, the specimen is pulled apart, and eventually it breaks. A stress-strain (σ - ε) curve (Figure 2.1) graphically displays a material's <u>strain</u> response to stress. Point B in Figure 2.1 identifies the location of the tensile strength. The curve descends rapidly past this point because the material's cross-sectional area is rapidly decreasing, and the material can no longer withstand the greater applied load. Hence, less stress is needed to deform a material past the tensile

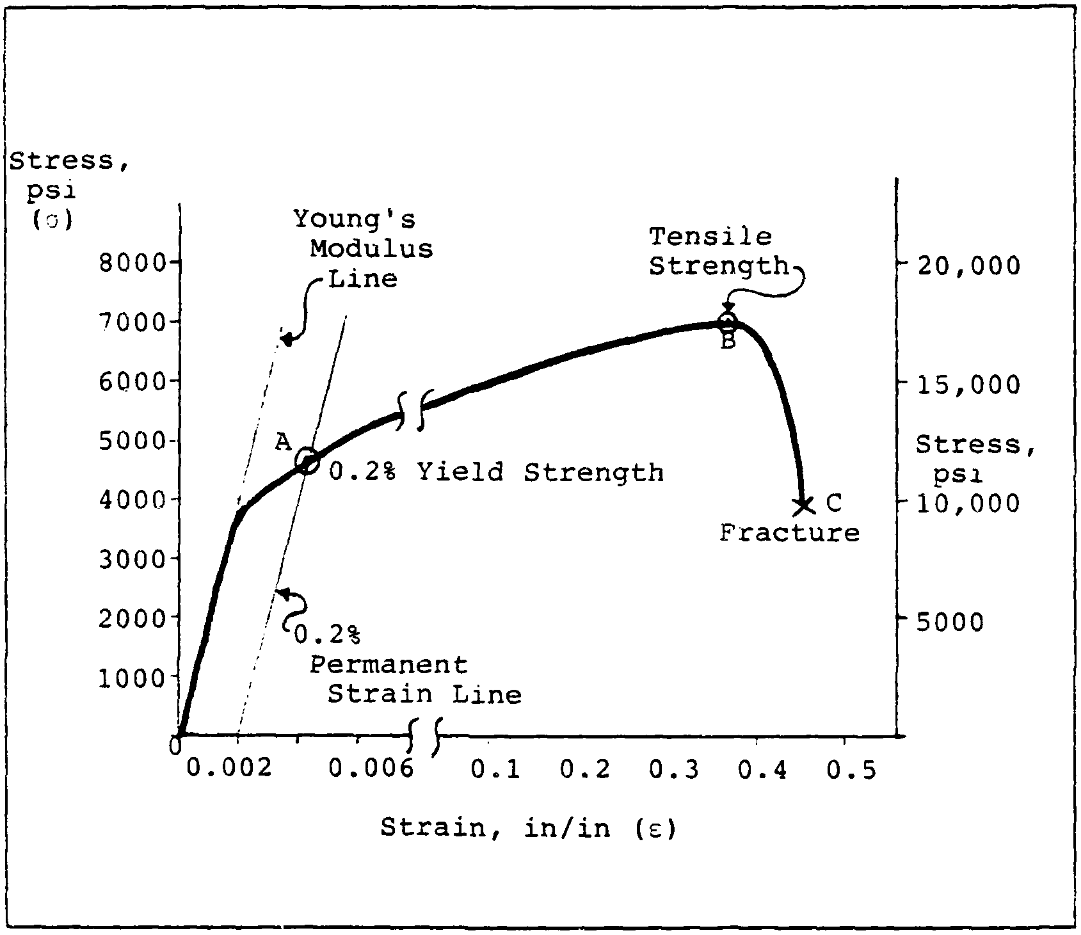


Figure 2.1. Stress-Strain (σ - ε) Curve (15:446)

strength to the fracture point. Along with tensile strength, several other properties including reduction of area and the <u>elastic modulus</u> (E)--a measure of the amount of strain (elongation per unit of original length) associated with the early application of stress on an unstressed material--can be obtained from the tensile test. Reduction of area is the percentage reduction of specimen cross-sectional area resulting from the stress necessary to cause fracture (point C in Figure 2.1), and is a significant measure of <u>ductility</u>. The elastic modulus (E), also known as Young's Modulus, is the ratio of stress to strain over the interval where the ratio remains relatively constant from the initiation of stress to just prior to the yield strength (point A). In this range of stress--the elastic range--once the stress is relieved the material will return to its original dimensions. Beyond the upper limit of the elastic range--the yield strength, <u>plastic</u> deformation, a phenomenon where a material deforms by various mechanisms within itself and does not return to its original dimensions after stress is removed, occurs.

Another measure of material strength is yield strength, usually the stress associated with a previously determined amount of strain. The most commonly used yield strength is the 0.2 percent yield strength, which is the stress corresponding to the point of 0.002 or 0.2 percent permanent strain (15:445). The 0.2 percent yield strength (point A in Figure 2.1) is identified on a stress-strain curve, obtained from a tensile test, by constructing a line parallel to the Young's Modulus line and which passes through the 0.002 strain point. The point where this constructed line intersects the $\sigma - \varepsilon$ curve is the 0.2 percent yield strength. At this point, when stress is removed, the material returns to its original dimensions plus 0.2 percent permanent strain. This strength is generally located near the start of the plastic region of a material, and is therefore a useful property to know when subjecting a material to a stress. Some materials can be hardened or strengthened by applying a stress which takes them into the plastic range. Once a material is in this range, it becomes more difficult for it to slip internally and therefore is strengthened, usually accompanied by a loss of ductility.

Another mechanical property, the one for which data is available for the present work, is fatigue strength. Fatigue failure is responsible for a large fraction of identifiable service failures (15:483). The fatigue strength is the stress at which a material fractures when subjected to repeated stress cycling. A common method of stress cycling is reverse bending, a technique which subjects a specimen, alternately, to tensile stress on one side and compressive stress on the other side and then compression on the first side and tension on the other. An aircraft wing experiences this effect due to the buffeting caused by air rapidly rushing past the wing. The area where the wing is attached to the frame experiences a great deal of cyclic stress. Generally fatigue strength (or endurance limit) is the stress which will cause fracture at the end of a specified number of stress cycles, normally 10^7 cycles (16:814). The stress levels used during testing are generally below the yield strength. In general, as the stress level (S) decreases, the number of cycles (N) to failure increases. This behavior is frequently described on a semi-logarithmic plot with an S-N curve, as shown in Figure 2.2. Some materials, such as mild steel and poly-methyl methacrylate, exhibit a fatigue limit, a stress level below which no amount of cycles will produce failure. However, nonferrous alloys and many polymers do not have this limit. Fatigue strength is measured on a machine which holds a prepared specimen by both ends and repeatedly applies cyclic stress generally until a failure occurs. Numerous tests are done with different specimens of the same material at several different stress levels to determine the pattern of fatigue behavior.

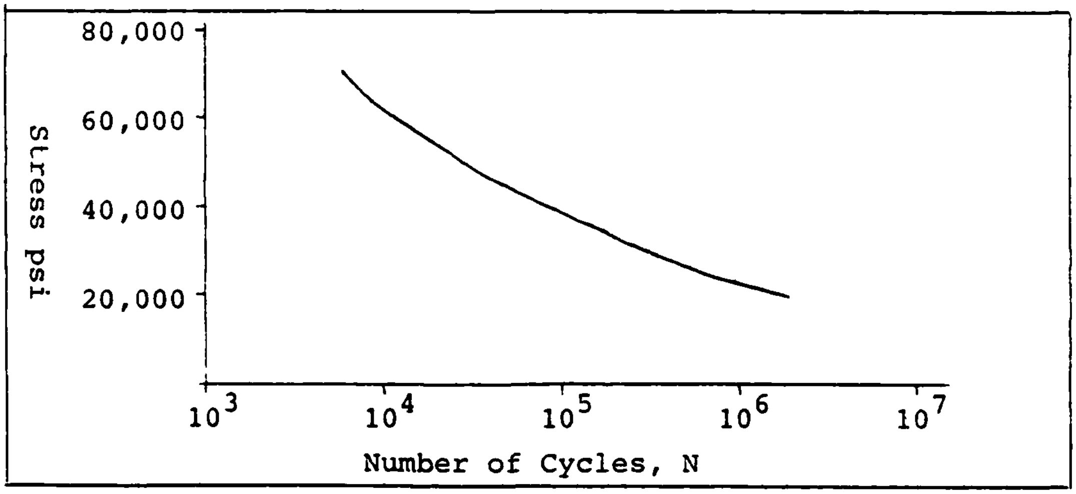


Figure 2.2. Illustration of an S-N Curve

Microscopic Aspects of Fatigue

Fatigue failure can be explained in microscopic terms. When stress is applied to a metal, the closest-packed planes of atoms (slip planes) slip relative to one another. When slip occurs, the crystal structure of the metal is slightly deformed, and lines of lattice defects, known as dislocations, are present along the slip planes (see Figure 2.3).

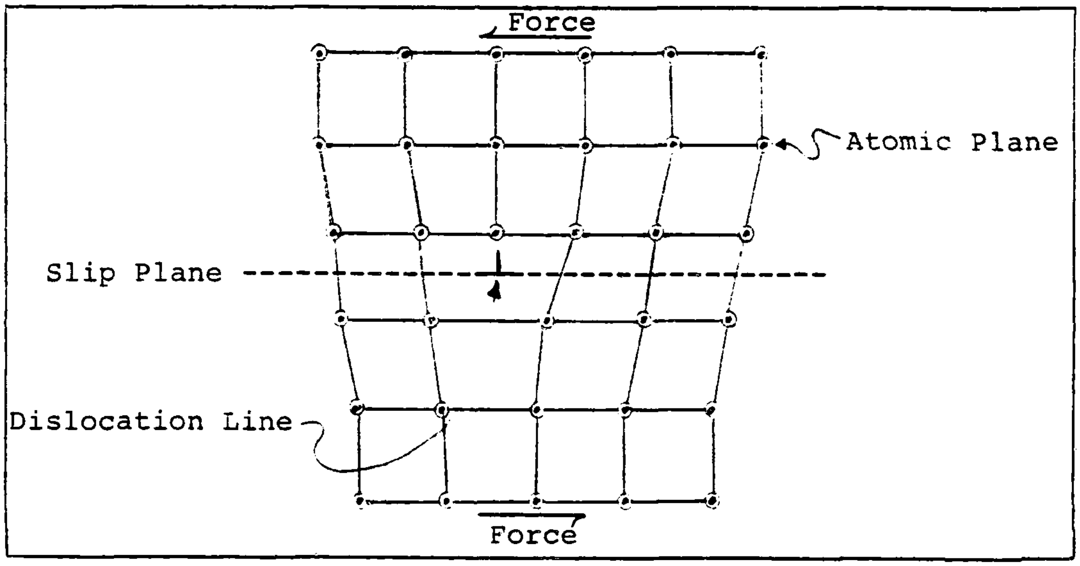


Figure 2.3. Permanent Deformation Resulting From Motion of a Dislocation (15:192)

With further application of stress, more dislocations form and move through the metal. The dislocations then interact and may then attract each other. Slip bands appear and tend to group into packets or striations. Extrusions, small ribbons of metal apparently extruded from the slip bands, or intrusions, small crevices, form in the slip bands.

Small gaps or openings can occur in these regions and can act as crack initiators (16:817).

Another explanation of crack nucleation involves the intersection of dislocations. When barriers to dislocation movement exist, the dislocations pile-up and dislocations from different slip planes may open up a small crack (16:769).

Additional cyclic stress causes the small cracks to grow. At relatively low stresses, the cracks grow slowly through the material until the cross-section of the specimen can no longer support the load. At this point, crack propagation is rapid to failure. Fatigue cracks may initiate at internal defects, such as foreign particles; however, many cracks start at surface defects, such as notches, which act as stress raisers.

Material Strengthening

Few materials have the strength to withstand the stresses applied to an aircraft wing during high-speed flight. Some materials, especially metal alloys, have the ability to be strengthened by various strengthening mechanisms. One of several strengthening techniques used to harden a metal is heat treatment. A common method of heat treatment is solution treating and aging (STA). Solution treating involves heating an alloy to a temperature where one constituent totally or partially dissolves. The material is then aged to produce a dispersion of the second constituent which strengthens the material by reducing the ability of the dislocations to move in the crystal lattice.

Many alloys have the characteristic of existing in different crystal structures (phases) at different temperatures. This property is called allotropy. For simplicity, a two-phase titanium alloy, Ti-6Al-4V, whose phase diagram is shown in Figure 2.4, will be discussed. It exists in the alpha (α) phase at lower temperatures and the beta (β) phase at higher temperatures.

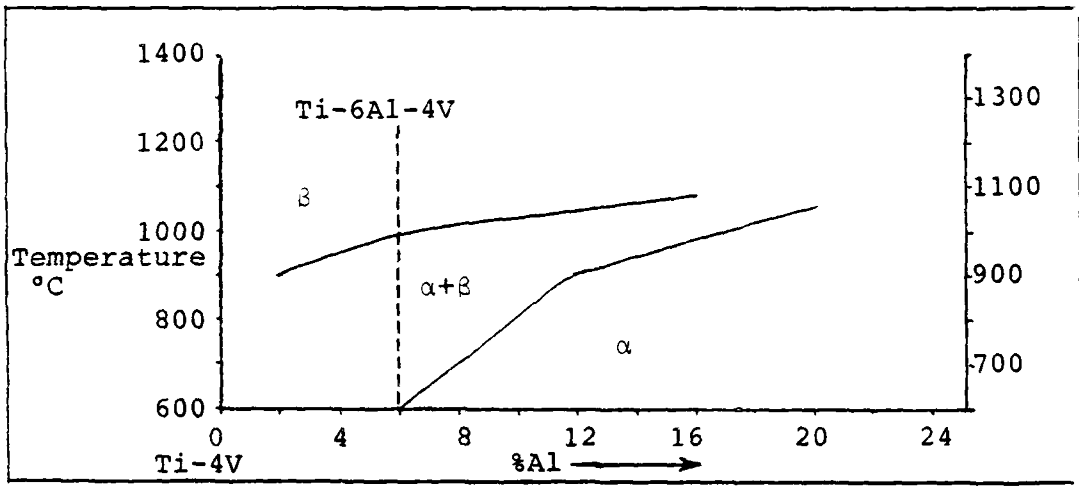


Figure 2.4. Vertical Section of Ternary Ti-Al-V (Constant 4%V) Phase Diagram (11:165)

Solution treatment for Ti-6Al-4V is raising the metal to a temperature in the β region or high in the $\alpha + \beta$ region and maintaining it there until the metal reaches its equilibrium composition at that temperature. When raised to a temperature in the β region, the alloy becomes a single phase, β. When raised to a temperature high in the $\alpha + \beta$ region, it remains a two-phase alloy; however, much of the α present at room temperature is converted into β. The metal is then quenched to a lower temperature using a liquid cooling medium, such as water. This rapid cooling prevents most of the β from returning to the α phase by diffusion. After the quench, a super-saturated solution exists, in that the amount of β present is greater than the equilibrium amount of β at the lower temperature (16:360).

The aging treatment allows the solution to become less supersaturated by allowing precipitation of some of the α which was converted to β. The alloy, after quenching, is placed in a furnace at a constant intermediate (between room temperature and the solution temperature) temperature. The proper aging temperature is that which balances the mechanisms of precipitation--<u>nucleation</u> and growth by <u>diffusion</u>. Nucleation is the formation of the beginnings of α particles, and is favored by lower temperatures. Diffusion is a method whereby the atoms comprising the α phase flow to the nuclei, causing the α particles to grow at the expense of the β particles; it is favored by higher temperatures where atomic motion is greater. Therefore, an intermediate temperature provides the fastest precipitation rate, because both nucleation and diffusion occur at moderate rates (16:361).

Precipitation of α is continued until a fine dispersion of α in β is present. This fine dispersion of precipitated particles increases strength by acting as barriers to dislocation movement. The particles cause the dislocation to move around them in the form of expanding loops (15:371). Only when the loops intersect can the dislocation move on through the metal. However, the dislocation loop remaining around the particle expands the stress field of the particle, causing subsequent dislocations to require more stress to move around or through them. The basis of strengthening is that greater stress is needed to move a dislocation past a precipitate particle than through a continuous matrix, allowing the metal to withstand a greater amount of stress before failure.

Table 2.1 gives an indication of the value of heat treatment on two titanium alloys. The Annealed condition is one where the grains of the alloy are relatively equivalent in size and shape and are strain-free. The Solution Treated and Aged condition is the heat-treated condition, which consists of a fine dispersion of one phase in the other phase. Heat treatment increases the yield strength while reducing ductility.

TABLE 2.1

Comparison of Pre-Heat Treated and Post-Heat Treated Alloys

All measurements at room temperature		0.2% Yield Strength (psi)	Reduction of Area (%)
Ti-6Al-4V Bar	Annealed	120,000[a]	25[a]
	Solution Treated And Aged	145,000[b]	20[b]
Ti-13V-11Cr-3Al	Annealed	125,000[c]	25[c]
	Solution Treated And Aged	170,000[d]	10[d]

a (16:22)
b (16:30)
c (16:29)
d (16:33)

Characteristics of Titanium

Titanium (chemical symbol Ti) is a highly desirable engineering material for aerospace systems due to its combination of properties which are important to those systems: high strength, low density, good fracture resistance, and excellent heat and corrosion resistance. A material's properties at elevated temperatures is especially important to the aircraft industry. The skin of an aircraft cruising at Mach 2.7 can reach a temperature of 500°F due to air friction (19:61). Of the currently commercially feasible engineering metals, titanium has the best strength-to-weight (yield strength:density) ratio.

TABLE 2.2

Comparison of Four Engineering Metals[d]

	Titanium	Aluminum	Iron	Magnesium
Chemical Symbol	Ti	Al	Fe	Mg
Atomic Number	22	13	26	12
Atomic Weight (atomic mass units; based on Carbon=12)	47.90	26.98	55.85	24.31
Density at 20°C(g/cc)	4.51	2.70	7.86	1.74
Melting Point(°C)	1668	660	1536	650
0.2% Yield Strength(psi)	20,000[a]	5000[b]	18,000[c]	--------

a (19:60)
b (15:415)
c (15:458)
d (14:1)

These materials, however, are rarely used in pure form in practical engineering construction. They are usually alloyed with other elements to enhance certain properties. For example, iron is alloyed with carbon to increase strength and alloyed with chromium to increase corrosion resistance. Aluminum is added to titanium to increase strength, while the addition of vanadium enhances titanium's ability to be strengthened through heat treatment.

A characteristic of titanium is its allotropic crystal structure. Allotropy is the ability of a substance to have different crystal structures at different temperatures. In the case of titanium, it exists in the hexagonal-close-packed (hcp) structure at room temperature and up to the transformation temperature of 883°C (1621°F). The body-centered-cubic (bcc) structure is present until the melting point. See Figure 2.5 for representations of the two crystal structures of Ti.

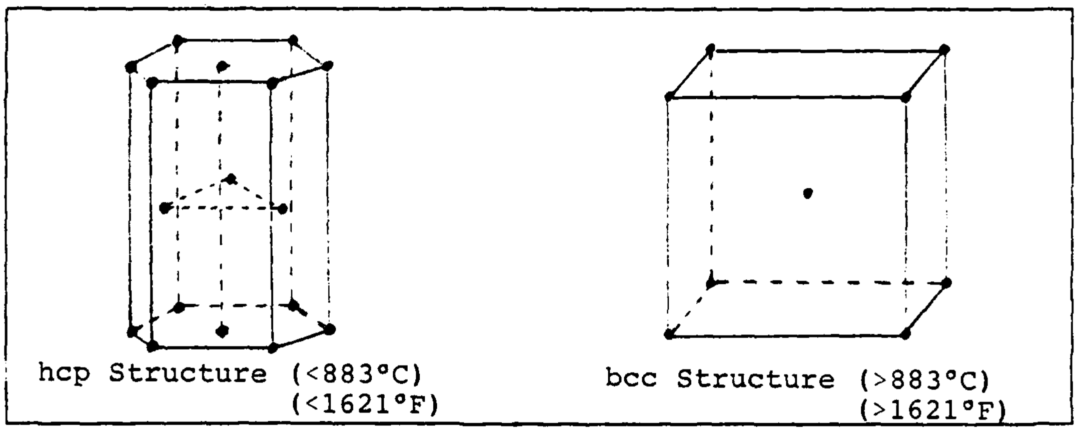


Figure 2.5. Unit Cells of the Crystal Structures of Titanium (15:156-57)

The hcp structure characterizes the alpha (α) phase of titanium, while the bcc structure denotes the beta (β) phase. The temperature at which α transforms to β is the beta transus. A look at the Ti-Al binary phase diagram (Figure 2.6) may be helpful at this point.

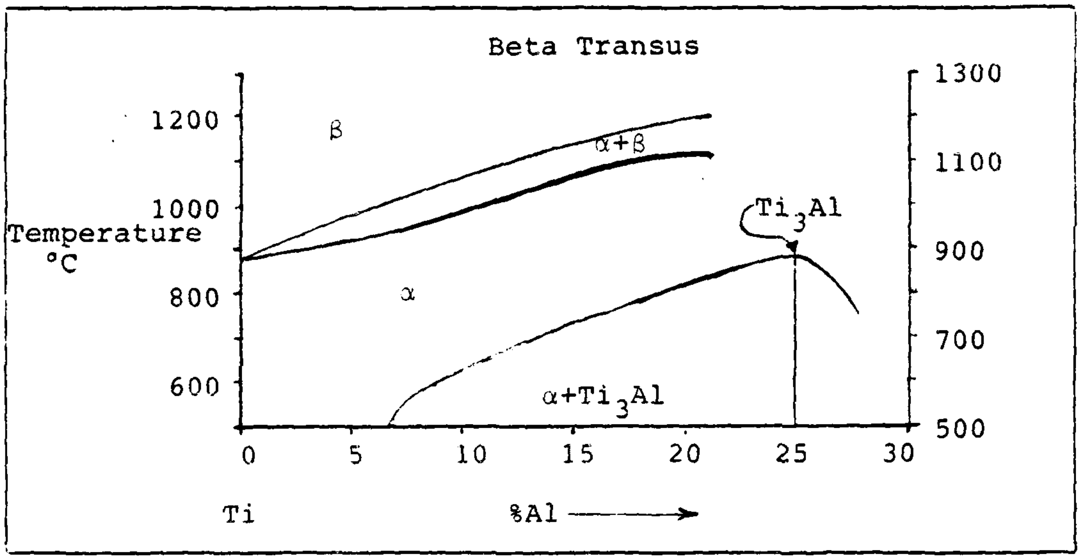


Figure 2.6. Ti-Al Binary Phase Diagram (19:10)

Characteristics of Titanium Alloy Ti-6Al-4V

The titanium alloy whose fatigue properties are under consideration in this study is Ti-6Al-4V. This is a titanium-based alloy with additions of between 5.50 and 6.75 percent aluminum, between 3.50 and 4.50 percent vanadium and traces of other elements (1:388). Table 2.3 lists some of the properties of Ti-6Al-4V, as compared to those of Aluminum 7075 alloy and 4130 Steel.

In addition, Ti-6Al-4V exhibits good machinability and weldability. Ti-6Al-4V is readily weldable, producing a weld with properties close to those of the base metal (17:13). Ti-6Al-4V is the most versatile and most widely used of the titanium alloys. Some of its uses include

Table 2.3

Comparison of Three Engineering Alloys (19:61)

	Ti-6Al-4V	Al 7075	4130 Steel
Density at 20°C(g/cc)	4.43	2.77	7.83
0.2% Yield Strength(psi)	125,000	70,000	160,000
Melting Range (°C) (°F)	1604-1671 2919-3040	477-638 891-1180	1504 2739

airplane turbine disks and blades, aircraft components, pressure vessels, rocket engine cases, and chemical processing equipment (17:22).

Through alloying, three classes of titanium alloys are possible: alpha alloys, $\alpha + \beta$ alloys, and beta alloys. α alloys are formed by additions of α stabilizers. α stabilizers are those elements whose addition cause the beta transus to increase, allowing the α phase to exist at a higher temperature. Some α stabilizers are aluminum, oxygen, carbon, and nitrogen; these elements generally increase strength and decrease ductility. Additions of aluminum can be made until about 8 percent, where the alloy becomes extremely brittle and difficult to work with (19:10) because of the formation of the intermetallic Ti_3Al phase.

β alloys are formed by additions of β stabilizers, those elements which stabilize the β phase at lower temperatures by lowering the beta transus. Some common β

stabilizers are vanadium, tantalum, molybdenum, chromium, and hydrogen. Stabilizers also increase the strength of the basic titanium somewhat, but their strong point lies in the fact that β alloys can be heat treated to increase strength.

The largest group of titanium alloys is the $\alpha + \beta$ group. These alloys usually contain both α and β stabilizers. Generally mechanical behavior remains stable even after exposure to high temperature and stress. Ti-6Al-4V belongs in this group.

Summary

Knowing a material's mechanical behavior can help when considering choices of materials to use for a weapon system. Fatigue strength is important to know for an aircraft, which is subject to repeated cyclic stress. Fatigue failure can be explained by dislocation movement, which leads to crack formation.

Titanium and its widely used alloy Ti-6Al-4V, are high strength materials. Heat treating allows Ti-6Al-4V to achieve greater strength.

Definitions of Key Terms

1. Alloy--a solid solution of two or more metals
2. Allotropy--ability of an element to exist in more than one crystal structure
3. Alpha Phase (α)--a distinct solid crystal structure; in titanium alloys, it is the low temperature structure of pure titanium, hexagonal-close-packed

4. Annealing--By holding a material at a relatively high temperature, the strain (dislocations) is removed from material which has been deformed (worked). By raising the material above its recrystallization temperature, strain free grains of relatively equiaxed shape are produced.

5. Beta Phase (β)--a second distinct solid crystal structure; in titanium alloys, it is the high temperature structure of pure titanium, body-centered-cubic

6. Beta Transus--the temperature at which all alpha phase has transformed to beta phase; connecting these points for various alloy compositions constructs the boundary between the $\alpha + \beta$ region and the β region

7. Crystal Structure--the atomic arrangement inside a material's crystals; the three relatively simple types of crystal structure for most metals are: face-centered-cubic, body-centered-cubic, and close-packed-hexagonal

8. Diffusion--Flow of atoms-in the solid state this involves the movement of atoms within a crystalline structure.

9. Dislocation--a line of crystal structure defects formed at the edge of an atomic plane

10. Ductility--the measure of a material's flexibility or ability to accomodate strain

11. Elastic Modulus (Young's Modulus, E)--the constant ratio of stress to strain in the elastic range-that region of the $\sigma - \varepsilon$ curve from initial loading to near the yield strength; $E = \sigma/\varepsilon$

12. Fatigue Strength--the stress that causes failure at the end of a specified number of cycles of alternating stress

13. Nucleation--the formation of nuclei, small "seeds" from which particles grow

14. Phase--a macroscopically distinct homogeneous body of matter, such as liquid, gas, or different crystal structures as a solid

15. Plastic--the region of the $\sigma - \varepsilon$ curve where the specimen does not return to its original dimensions upon relief of stress; it suffers some permanent strain

16. Slip Planes--Normally the atomic planes of highest density; upon application of stress, these planes move relative to one another.

17. Strain (ε)--elongation per unit length of a material in response to an applied stress

18. Stress (σ)--force per unit area acting on a surface

19. Tensile Strength--the maximum stress, measured over the original specimen cross-sectional area, a material can withstand in a tensile test

20. Toughness--fracture resistance of a material upon impact

21. Yield Strength--the point on a σ - ε curve, to which if a specimen is loaded and then unloaded, it would suffer a permanent strain of some specified value, usually 0.002 (0.2%)

CHAPTER 3

STATISTICAL BACKGROUND

Introduction

From Chapter 2, it can be seen that mechanical properties of materials are represented as relationships of two or more quantities (e.g., stress vs. strain) on graphs with curves which describe a material's mechanical behavior over a range of values. Statistical methods offer a much more precise and reproducible means of representing these relationships. This chapter is intended to introduce the non-statistician to some important statistical concepts. In particular, an approach to curve-fitting--regression, along with some techniques for testing the appropriateness of the regression model are discussed.

Introduction of Regression

In many situations it is desirable to express a relationship between two or more quantities. For instance, an aircraft manufacturer's engineering shop may be interested in knowing the relationship between the useful life of a material and various conditions of stress. Knowing this relationship would enable the engineers to estimate the useful life given a certain level of working stress.

One branch of statistical analysis which deals with the description of the relationship between variable quantities and also with the prediction of the value of an unknown variable using known values of other variables is regression analysis. Regression analysis concerns estimating the value of one variable, the dependent variable, on the basis of one or more other variables called independent variables (8:357).

Estimates of the exact values of dependent variables are difficult to determine due to the many factors which could cause variations. Regression analysis determines the average relationship between dependent and independent variables; that is, it estimates the mean value of a dependent variable for a given value of the independent variables.

The simplest relationship that can be hypothesized between the means of the independent variable and the dependent variable is the following linear form, described by the equation of the population regression line:

$$\mu_{y \cdot x} = \alpha + \beta x \tag{1}$$

where $\mu_{y \cdot x}$ = the mean value of y, the dependent variable, for a given value of x, the independent variable

α = the y-axis intercept

and β = the slope of the population regression line (8:359).

The many minor factors causing variations about the mean for a single observation can be represented by a random error term, ε, where $\varepsilon_i = y_i - \mu_{y \cdot x}$, the difference between the actual dependent variable value and the mean value of y for a particular value of the independent variable. The population linear regression model is:

$$(2) \qquad y_i = \alpha + \beta x_i + \varepsilon_i$$

where y_i = the actual dependent variable value for a particular independent variable value, x_i

and ε_i = the random error at x_i (8:360).

The values of α and β are seldom accurately known and therefore must be estimated using sample data extracted from the population. The following equation may be used to describe the sample regression line when a sample of n observations is taken:

$$(3) \qquad \hat{y} = a + bx$$

where $\hat{y}$ = the estimate of $\mu_{y \cdot x}$

a = the estimate of the y-intercept, α

and b = the estimate of the slope, β (8:362).

Least-Squares Estimation

The most widely used method of finding the best estimates of the regression coefficients is the method of least-squares. This process involves minimizing the sum of the squared differences between the actual y-value and the estimate, $\hat{y}_i$, as found by the sample regression line every x_i in the sample.

Method of Least-Squares Estimation:

(4) $$\text{Minimize } \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{n}(y_i-a-bx_i)^2$$

(8:367)

This quantity is then differentiated with respect to a and b, with y_i and x_i treated as constants, and set to zero. The resulting equations, called the normal equations, are then solved for a and b. The values of a and b which satisfy the normal equations are estimates of α and β.

The normal equations:

(5) $$\sum_{i=1}^{n}y_i = na + b\sum_{i=1}^{n}x_i$$

(6) $$\sum_{i=1}^{n}x_iy_i = a\sum_{i=1}^{n}x_i + b\sum_{i=1}^{n}x_i^2$$

(8:368)

Solving for a and b (12:37):

(7) $$b = \frac{\sum_{i=1}^{n}x_iy_i - \dfrac{(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)}{n}}{\sum_{i=1}^{n}x_i^2 - \dfrac{(\sum_{i=1}^{n}x_i)^2}{n}}$$

(8) $$a = \frac{\sum_{i=1}^{n} y_i}{n} - b \frac{\sum_{i=1}^{n} x_i}{n}$$

The estimators a and b are functions of the random variable y_i and therefore are random variables themselves. They are unbiased estimators of α and β because their expected values equal those of the population parameters ($E(a)=\alpha$; $E(b)=\beta$). These estimators also have a variance associated with them that is a function of the variance of the random variable ε_i.

How good are the least-squares estimates of α and β? Are there other methods of estimation--e.g., the method of moments--which consistently yield estimators of the regression coefficients with smaller variance? The Gauss-Markov theorem states that if we assume:

1. the random errors, ε_i, have a mean of zero,
2. the random errors have a constant finite variance σ_ε^2, for all values of x, and
3. the random errors are independent of each other

then the estimators of α, β, and $\mu_{y \cdot x}$ determined by the least-squares criterion are Best Linear Unbiased Estimates (i.e.,BLUE) (8:366).

Definition of key terms is necessary to fully understand the importance of this theorem. Linear means that the estimators are straight-line (linear) functions of the values of the dependent variable. The best estimators are those which are efficient, which means the variance of the estimators found by the least-squares method is less than the variance of estimators found by any other linear unbiased estimating technique (8:231-34).

Hypothesis Testing

Assessing the appropriateness of the regression model can be accomplished through the use of tests of statistical hypotheses. Two hypotheses are presented: the null hypothesis (H_o) and the alternative hypothesis (H_a). The null hypothesis was originally labeled as such because it specified values of a parameter which the tester thought were not true. Correspondingly the alternative hypotheses represented the values of the parameter believed to be true. These labels have no special meaning nowadays and are only labels for the two necessarily conflicting hypotheses (8:264).

For the linear regression model, one frequently is interested in testing the hypothesis that the slope equals zero (H_o:$\beta=0$; H_a:$\beta\neq0$). If the null hypothesis were true, there is no relationship between y and x; that is, $\mu_{y \cdot x}=\alpha+(0)x=\alpha$. Hence, for any value of x, the prediction of y is y=a.

Before testing the hypothesis it is convenient to assume that:

Assumption 4. The random errors are normally distributed. Recall that a linear combination of independent normal random variables is normally distributed. Hence, the Y_i's are normally distributed (12:56). From equation (7) we note that b is a linear combination of the y_i's, and is therefore normally distributed. The assumptions of mean of zero, constant variance, and normality are shown graphically in Figure 3.1.

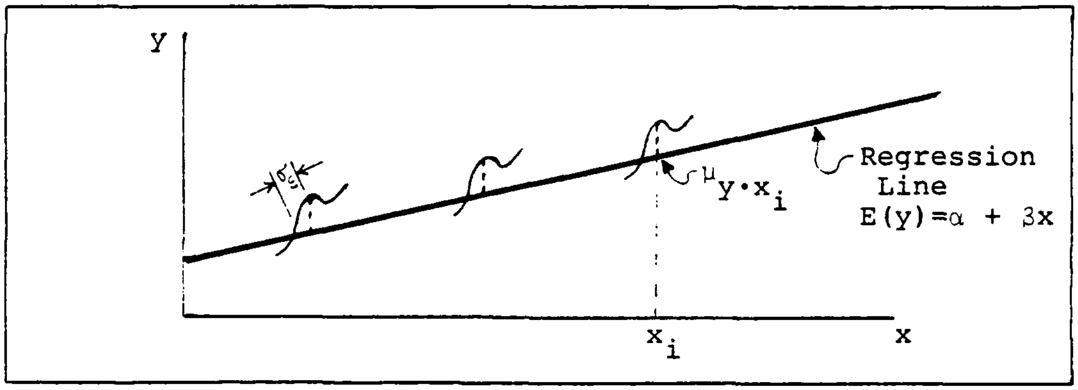


Figure 3.1. Graphical Display of Random Error Assumptions

The variance of b can be found to be:

(9) $$\sigma_b^2 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where

σ_b^2 = the variance of b

σ_ε^2 = the variance of ε

and $\bar{x}$ = the average value of x (12:56)

Note that σ_ε^2 is an unknown population parameter. It can be shown that

$$s_\varepsilon^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n-2}$$

is an unbiased estimate of σ_ε^2. This estimate of σ_ε^2 is often called the Mean Square Error (MSE). MSE is the quantity minimized in least-squares estimation $(y_i - a - bx_i)^2$ divided by the appropriate degrees of freedom (n-2). Two degrees of freedom are lost because two parameters--α and β--are being estimated. An unbiased estimate of σ_b^2 is:

(10) $$s_b^2 = \frac{MSE}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where s_b^2 = the estimate of σ_b^2

and MSE = mean square error, an unbiased estimate of σ_ε^2 (12:56).

It can be shown that the standardized statistic $\frac{b - \beta}{s_b}$ follows the t-distribution with n-2 degrees of freedom. This allows the following probability statement to be made:

$$(11) \quad P\left(-t_{(1-\frac{\alpha}{2},\ n-2)} \le \frac{b-\beta}{s_b} \le t_{(1-\frac{\alpha}{2},\ n-2)}\right) = 1-\alpha^{1}$$

where $t_{(1-\frac{\alpha}{2},\ n-2)}$ = the $(1-\frac{\alpha}{2})\cdot 100$ percentile of the t-distribution with n-2 degrees of freedom

and $1-\alpha$ = the probability (significance level) desired (12:58)

Rearranging the inequalities gives a $1-\alpha$ confidence interval for β:

$$(12) \quad b - t_{(1-\frac{\alpha}{2},\ n-2)} s_b \le \beta \le b + t_{(1-\frac{\alpha}{2},\ n-2)} s_b$$

A $1 - \alpha$ confidence interval represents $1 - \alpha$ confidence that the population parameter lies within the interval determined by equation (12). If the interval contains zero, the null hypothesis for the linear regression model (H_o:β=0) cannot be rejected at the significance level specified. Hence, we have no reason to believe that knowledge of x is useful in estimating a value for y.

---

[1]This α is not the same as the population regression coefficient denoting the y-intercept.

An equivalent test for rejecting $\beta=0$ involves a t-statistic, $t^*= \frac{b}{s_b}$. The null hypothesis cannot be rejected at a significance, α, if $|t^*| \leq t_{(1-\frac{\alpha}{2},\ n-2)}$ (12:61).

The F-statistic can also be used to test the regression model. It can be calculated using an approach known as the analysis of variance (ANOVA). It analyzes the total deviation of y_i from the mean of all the y values, $\bar{y}$. The total deviation is divided into two other deviations, one of which is the unexplained deviation $(y_i-\hat{y}_i)$ and the other is the explained deviation $(\hat{y}_i-\bar{y})$ (8:375). Note Figure 3.2 for an illustration of these deviations.

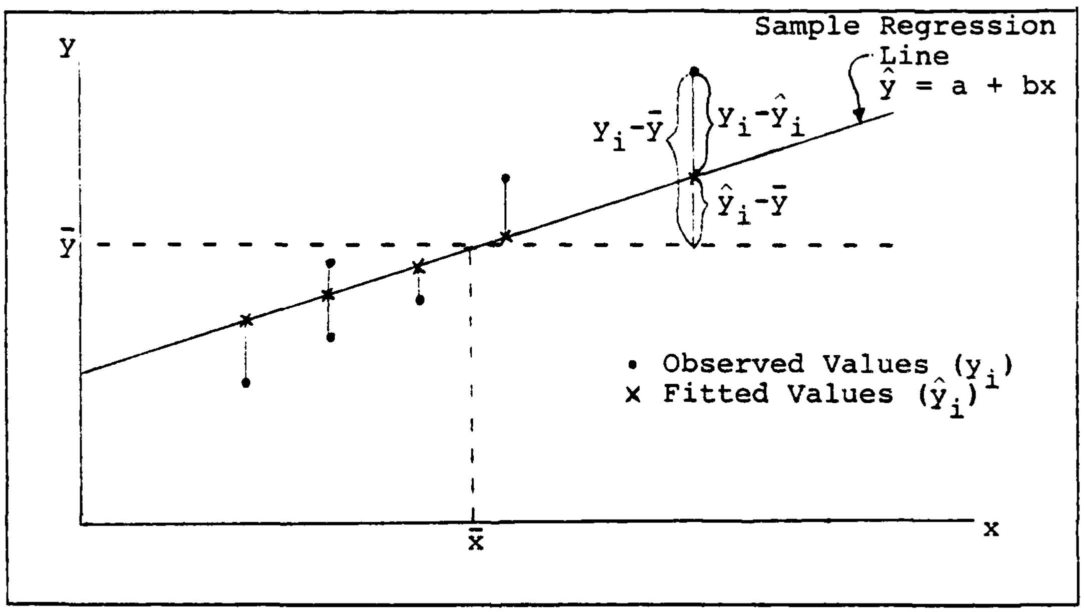


Figure 3.2. Illustration of a Least-Squares Regression Line and Deviations from the Line

The analysis of variance approach is based upon the partitioning of the sums of squared deviations. The sum of the squared deviations can be partitioned as follows:

$$(13) \qquad \underset{\text{SS Total}}{\Sigma(y_i-\bar{y})^2} = \underset{\text{SSE}}{\Sigma(\hat{y}_i-y_i)^2} + \underset{\text{SSR}}{\Sigma(\hat{y}_i-\bar{y})^2}$$

(12:79)

SSE is the sum of the squared deviations of the observed values from the fitted values. The ratio of SSR to SS Total is a measure of how much of the variability of the y_i's is explained by the regression line. This ratio is known as the coefficient of determination and is denoted as R^2. The range of R^2 is zero to one. When $R^2=0$, no linear relationship between y and x exists. If $R^2=1$, the observed y_i's all lie on the regression line. In this case, the linear regression line explains the relationship between y and x perfectly. Thus R^2 is a measure of the usefulness of the β term (4:62).

Dividing SSR and SSE by the variance (σ_ε^2) produces two independent chi-square (χ^2) distributed random variables. The ratio of two independent χ^2 random variables divided by their respective degrees of freedom defines the F-distribution. The degrees of freedom associated with SSR is one, the number of independent variables. Hence, the F-statistic is the ratio of the mean square regression (MSR) to the mean square error (MSE):

$$\text{(14)} \qquad F^* = \frac{\frac{SSR}{\sigma^2}}{1} \div \frac{\frac{SSE}{\sigma^2}}{n-2} = \frac{MSR}{MSE}$$

If $F^* \leq F_{(1-\alpha,\ 1,\ n-2)}$ the null hypothesis of $\beta = 0$ cannot be rejected. A relationship exists between the t and F distributions when testing the null hypothesis H_o:β=0:

$$\text{(15)} \quad F_{(1-\alpha,\ 1,\ n-2)} = (t_{(1-\frac{\alpha}{2},\ n-2)})^2$$

(12:87)

<u>Confidence Intervals</u>

Based on the assumption that the y_i's are normally distributed allows confidence intervals to be calculated about other population parameters as well. Confidence intervals around the population mean for a particular value of x_h, $\mu_{y \cdot x_h}$, provide us with an indication of how much "faith" we can put in our estimate of the regression line. The estimated variance associated with y_h, the fitted value at x_h, is:

$$\text{(16)} \qquad s^2(\hat{y}_h) = MSE\left(\frac{1}{n} + \frac{(x_h-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

The associated confidence interval is:

(17)

$$\hat{y}_h - t_{(1-\frac{\alpha}{2},\ n-2)} s(\hat{y}_h) \leq \mu_{y \cdot x_h} \leq \hat{y}_h + t_{(1-\frac{\alpha}{2},\ n-2)} s(\hat{y}_h)$$

where $\hat{y}_h$ = fitted dependent variable value at x_h

$s(\hat{y}_h)$ = square root of the variance defined above

and $\mu_{y \cdot x_h}$ = population parameter

(12:68)

Prediction Intervals

Another interval can be constructed based on the predicted value for a new observation of y at x_h. We can determine an interval which with a specified probability will *contain a random observation on y taken at* x_h. This interval is called a prediction interval.

Since the prediction interval is calculated based upon a single sample point, its variance is larger than the variance used for estimating the mean value. The variance derives from two sources: the variance of the sampling distribution of y_h and the variance of the distribution of y (12:72). The variance is estimated as:

$$(18) \quad s^2(y_{h(new)}) = MSE\left(1 + \frac{1}{n} + \frac{(x_h-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

where

$s^2(y_{h(new)})$ = the variance for the prediction interval

MSE = mean square error

n = number of observations, excluding $y_{h(new)}$, in sample

x_h = the value of the independent variable where the new observation is to be obtained

$\bar{x}$ = mean of all x_i's

and x_i = the independent variable values in the original sample $i=1,2,\ldots,n$

(12:72)

This variance is used to calculate the $1 - \alpha$ prediction interval for $y_{h(new)}$. The limits of the prediction interval are given as:

$$(19) \quad \hat{y}_h \pm t_{(1-\frac{\alpha}{2},\ n-2)}\, s(y_{h(new)})$$

where

$\hat{y}_h$ = fitted dependent variable value at x_h ($\hat{y}_h = a + bx_h$)

and $s(y_{h(new)})$ = square root of variance defined above

(10:58)

By calculating these limits at many x_i's and connecting the points, prediction bands can be determined for any probability desired.

Due to the larger variability inherent in the prediction interval calculations, the prediction bands are wider about the regression line than the confidence bands, as Figure 3.3 shows.

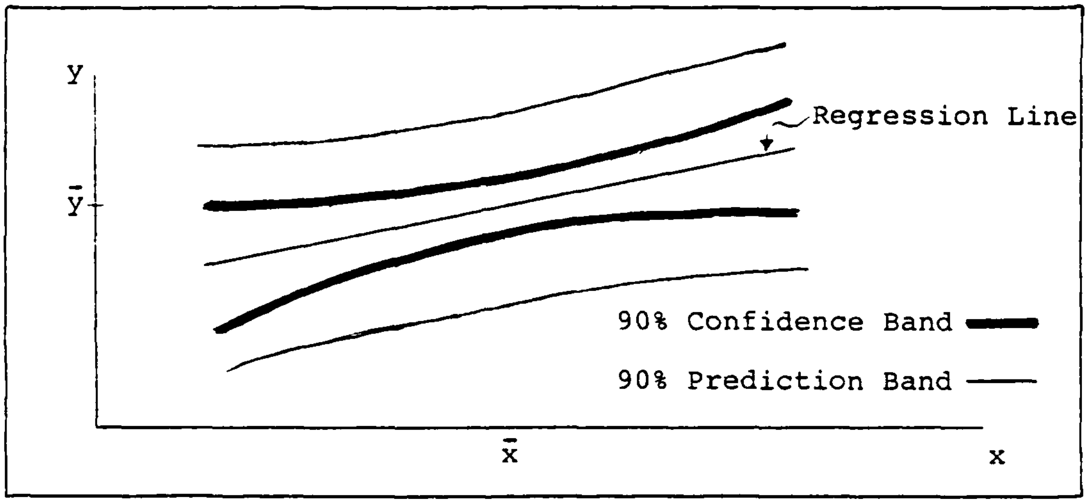


Figure 3.3. Example of Confidence and Prediction Bands

An additional feature of the bands which results from the calculation of variance is the bands are narrowest at the mean values $(\bar{x}, \bar{y})$ of x and y. This occurence shows that the least-squares fitted line is most accurate at $(\bar{x}, \bar{y})$ since it represents the average relationship of the independent and dependent variables (8:386).

Residual Analysis

The appropriateness of the regression model and the related tests depend to varying degrees on how well the data satisfy the assumptions made about the random error terms. Residual analysis is a technique used to assess the reasonableness of the assumptions of normality and constant variance made for the random error terms. The residual is the difference between the observed value and the predicted value of y: $e_i = y_i - \hat{y}_i$. The residuals may be transformed to standardized residuals by dividing each residual by the square root of the mean square error.

A direct and revealing, though subjective, method for examining residuals is the graphical approach. Several plots of residuals can be constructed to note the behavior of the residuals against: the dependent variable, time, the independent variables, or any other variable which might provide information about the model.

To check the assumption that the error terms have a constant variance, $\sigma_\varepsilon{}^2$, we plot the standardized residuals vs. the independent variables. If the plot is characterized by a band about zero with no systematic tendencies toward either positive or negative values then there is no reason to suspect that the constant variance assumption is violated (12:102). See Figure 3.4.

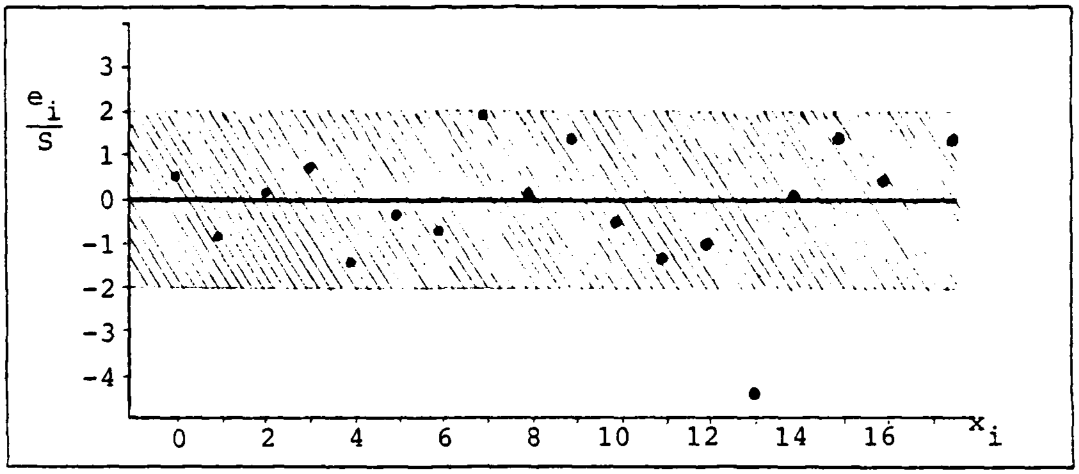


Figure 3.4. Standardized Residual Plot vs. Independent Variable

The tests of hypotheses and confidence intervals associated with the population parameters were based on the assumption that the ε_i's are normally distributed (Assumption 4). If normality holds, about sixty-eight percent of the points should be within one standard deviation (S) of the mean, zero, and approximately ninety-five percent of the points should be contained within the interval of +2S to -2S. This is the shaded area in Figure 3.4 (10:239)

Testing Random Error Assumptions

Several formal tests are also applicable for judging the reasonableness of the random error assumptions. When a residual plot shows the variance increasing or decreasing in a systematic manner in relation to x or $\hat{y}$, a test fitting separate regression functions to each half of the observations arranged by level of x, calculating mean square errors for each half, and testing for equality of error variances by an F-test, may be employed (12:112). Goodness-of-fit tests can be used for examining the normality of the error terms (8:529). When a residual plot shows the error terms increasing or decreasing in a systematic fashion with time, the error terms may be considered correlated, violating the independence assumption.

A remedy for nonconstancy of variance is to modify the regression model using transformations of y or x or both to stabilize the variance. Frequently, lack of normality accompanies nonconstant error variances, so that a transformation which stabilizes the variances often aids in making the distribution of the error terms normal (12:508).

Even though a transformation may not be successful in stabilizing the variance or the normality, a slight violation of these two assumptions need not cause the model to be discarded (4:59). The problem of nonindependence, however, it is not easily solved by a simple transformation. If the independence assumption is not met, the model may have to be discarded. The least-squares estimating technique no longer produces the best linear unbiased estimators. A new model which works with correlated error terms should then be used (12:122).

Outliers

An interesting point shown in Figure 3.4 is the residual associated with $x_i=14$. This point may be considered an outlier, an extreme observation or one much larger than the others in absolute value. Frequently outliers are classified as those points which lie three, four, or more standard deviations from the mean. Since the least-squares method of estimation seeks to minimize the sum of the squared errors, the presence of outliers affects the fitted line by pulling it toward the outlier. The "pulled" line then may not truly represent the relationship between y and x.

Various rules have been proposed for identifying outliers. One method involves fitting a new regression line based on the other (n-1) observations, omitting the outlier data value. The outlier is then reintroduced and treated as a new observation. The probability of obtaining an observation from the fitted line as large as that of the outlier in n-1 observations can then be calculated. A very small probability--perhaps less than five percent--would cause the outlier to be rejected as not coming from the same population as the other n-1 observations. A larger probability--greater than five percent--would allow the outlier to be retained (12:112).

A corollary of that method involves fixing the probability desired and calculating a $1-\alpha$ prediction interval about the estimated value of the dependent variable. If the outlier lies within the limits of the prediction interval, it is retained.

Several other treatments of outliers are available. However, before rejecting outliers, their causes should be examined carefully. Outliers may be the result of an error in recording, a miscalculation, an equipment malfunction, or could be an indication of an inaccuracy in the original model. For this reason, outliers should not be discarded immediately unless it is known that they are the result of an error in experimentation and not an error in the formulation of the model.

Comparison of Regressions

At times different sample regression lines are determined by drawing from different but potentially similar populations. It may be interesting then to compare the different lines by analyzing differences, if any, that exist in slope and y-intercept. For instance, fatigue data may be determined for specimens made of cast iron bar and stainless steel bar. In this example, cast iron and stainless steel are the separate populations. We may be interested in determining which population is most fatigue resistant.

One method for comparing regression models uses the concept of qualitative variables. Indicator or dummy variables are qualitative variables that can take on values of 0 or 1. To compare regressions we pool the observations (specimens). Indicator variables are then used in the overall regression model to indicate the population observed (i.e., cast iron or stainless steel) for each specimen.

The regression model for comparing y-intercepts assuming equal slopes for the fatigue data example becomes:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad (20)$$

where

y_i = number of cycles to failure for ith specimen

β_1 = common slope

x_{1i} = stress level for ith specimen

β_2 = population regression coefficient

and $x_{2i} = \begin{cases} 1 \text{ if ith specimen is cast iron} \\ 0 \text{ if ith specimen is stainless steel} \end{cases}$

(12:299)

In the example, when a specimen is taken from the cast iron population, the y-intercept term is $\alpha + \beta_2$:

(21) $$y_i = (\alpha + \beta_2) + \beta_1 x_{1i} + \varepsilon_i$$

If the specimen is from the stainless steel population, the y-intercept term is α:

(22) $$y_i = \alpha + \beta_1 x_{1i} + \varepsilon_i$$

If the y-intercepts are not significantly different, then we fail to reject $H_o:\beta_2=0$.

If the slopes are thought to be different for the separate populations, then an appropriate model is:

(23)
$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i$$
(12:304)

The cross-product interaction term causes the regression equation to include β_2 and β_3 when specimens from the cast iron population are used:

(24)
$$y_i = (\alpha + \beta_2) + (\beta_1 + \beta_3) x_{1i} + \varepsilon_i$$

If the specimen is from the stainless steel population, then:

(25) $$y_i = \alpha + \beta_1 x_{1i} + \varepsilon_i$$

The hypothesis being tested by the above model is that the intercepts and the slopes are the same; that is, the lines are coincident:

$H_o: \beta_2 = \beta_3 = 0$

$H_a: \beta_2 \neq 0$ and/or $\beta_3 \neq 0$

As before, during hypothesis testing, if a $1 - \alpha$ confidence interval about the population parameter contains zero, the null hypothesis cannot be rejected at the α significance level.

Figure 3.5 graphically explains the meaning of the population parameters: α, β_1, β_2, and β_3. β_2 is the difference in y-intercepts; β_3 shows how much greater or smaller the slope is for the separate populations.

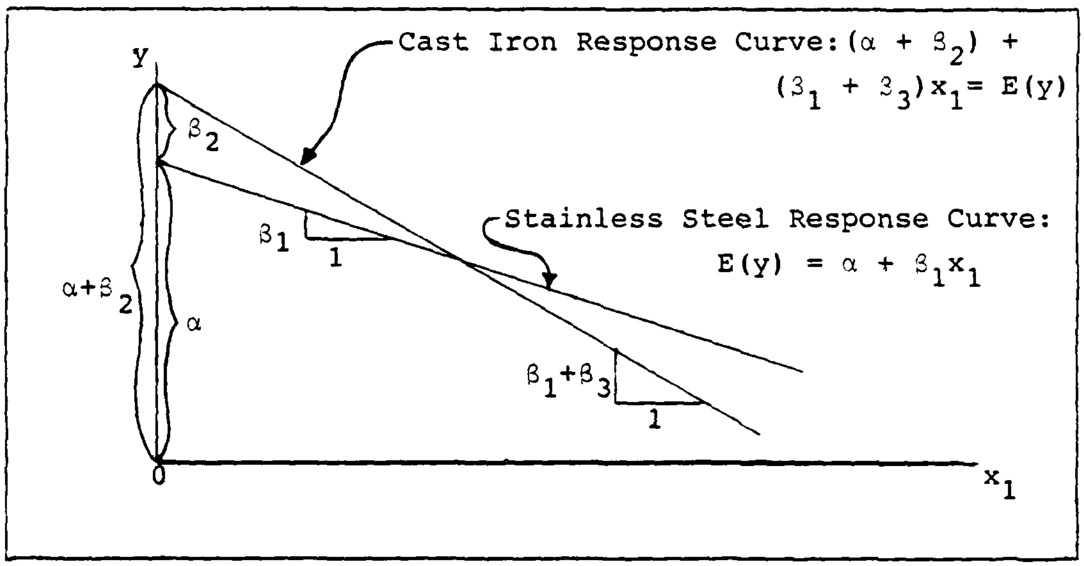


Figure 3.5. Illustration of the Meaning of Regression Parameters with an Indicator Variable and an Interaction Term (12:305)

Nonlinear Regression

Up to this point, the only relationship between y and x which has been discussed is the linear form. Many times, however, a linear model cannot reasonably fit the data. In these situations it may be more realistic to fit a model of the nonlinear form. A nonlinear model is one that is nonlinear in its parameters, such as the following two examples:

$$y = e\left(\frac{\beta_1}{x} + \beta_2 x^2 + \varepsilon\right) \tag{26}$$

$$y = \frac{\beta_1}{\beta_1 - \beta_2}(\sin\beta_1 x - \cos\beta_2 x) + \varepsilon \tag{27}$$

Model (26) can be transformed by taking logarithms of both sides. The transformation produces a model in which the parameters are linear:

$$\ln y = \frac{\beta_1}{x} + \beta_2 x^2 + \varepsilon \tag{28}$$

Model (27) can also be transformed, but any resulting transformation leaves the parameters still in the nonlinear form. This model is said to be intrinsically nonlinear. The concepts of nonlinear regression will be illustrated using the intrinsically nonlinear model below:

$$y = \alpha + e^{\beta x} + \varepsilon \; ; \; \alpha \neq 0 \tag{29}$$

Nonlinear Parameter Estimation

As in the linear case, in order to fit the best regression line to the data, the parameters--α and β--must be estimated. The model to be used has, at this point, been hypothesized from actual knowledge of the true model form or by some judicious guessing based on trends in the data (4:264).

As with the linear model, the least-squares technique can often provide reasonable estimates of the parameters, provided the error term, ε, follows all of the assumptions stated above for the linear model. Again, a and b are estimates of regression parameters, α and β, respectively. The procedure for estimating a and b follows:

Method of Least Squares Estimation for the Nonlinear Model

(30)

$$\text{Minimize } \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - a - e^{bx_i})^2$$

Differentiating the above quantity with respect to the parameters and setting it equal to zero produces the two normal equations:

$$\sum_{i=1}^{n}(y_i-a-e^{bx_i}) = 0 \tag{31}$$

$$\sum_{i=1}^{n}(y_i-a-e^{bx_i})(x_ie^{bx_i}) = 0 \tag{32}$$

(4:266)

The normal equations can then be solved for a and b to find the least-squares estimates for α and β. An iterative method is used to estimate the parameters for the example. The method involves iteratively modifying the parameter estimates so as to reduce the sum-of-squares. Iterations continue until the sum-of-squares function changes less than a specified amount <u>or until</u> the parameter estimates change by less than a specified amount.

The Newton-Raphson technique is an iterative parameter estimation method. It involves solving the normal equations for b, producing a single equation in the form $f(b) = 0$. An initial value for b is guessed and then corrected on each iteration. The corrected value of b then becomes the starting estimate for b in the next iteration. Iterations continue until an iteration produces a change in b of less than a specified amount. Specifically, let

b_i = the estimate of b after i iterations

b_0 = the initial guess of b

h_i = the correcting element for the i^{th} iteration

$f(b_{i-1})$ = f(b) evaluated at b_{i-1}

$f'(b_{i-1})$ = the partial derivative of f(b) with respect to b evaluated at b_{i-1}

and $h_i = \dfrac{f(b_{i-1})}{f'(b_{i-1})}$

The revised estimate is: $b_i = b_{i-1} - h_i$.

The following example serves to illustrate a non-linear parameter estimating technique.

Data:

x	y
0	10
1	13
2	18
3	29
4	65

Hypothesized Regression Model:

$$y = \alpha + e^{\beta x} + \varepsilon$$

TABLE 3.1

Iteration Summary

Iteration	b_{i-1}	f(b)	f'(b)	h_i	b_i
1	1.50	2,072,703	18,022,273	0.12	1.38
2	1.38	710,829	6,657,279	0.11	1.27
3	1.27	246,868	2,506,615	0.10	1.17
4	1.17	81,907	1,003,838	0.08	1.09
5	1.09	25,995	457,382	0.06	1.03
6	1.03	16,827	239,652	0.07	0.96
7	0.96	-5,971	99,584	-0.06	1.02
8	1.02	3,409	213,635	0.02	1.00
9	1.00	-397	168,318	0.00	1.00

An initial estimate of 1.50 produces a final value for b of 1.00 through nine iterations of the Newton-Raphson technique. To find the estimate for a, substitute the final value of b into one of the normal equations. For this example the final parameter estimates and resulting model are:

$a = 9.84$
$b = 1.00$
$\hat{y}_i = 9.84 + e^{x_i}$

However, it is sometimes difficult to eliminate a parameter by simultaneously solving the normal equations, as in the Newton-Raphson technique. For this reason, other methods exist for estimating the parameters of a nonlinear model. Two such methods--Gauss' Method and Marquardt's Method--are discussed below.

Gauss' Method, also called the Linearization Method, involves expanding the original nonlinear function in a Taylor series near the initial estimates of the parameter values, and ignoring all higher order terms:

$$(33) \quad y = y_0 + \sum_{k=1}^{m} \left[\frac{\partial y}{\partial b_k}\right]_{b=b_0} (b_k - b_{k0})$$

where y_0 = the original function evaluated at the initial parameter estimates

m = the number of parameters

$\left[\frac{\partial y}{\partial b_k}\right]_{b=b_0}$ = the partial derivatives of the original function with respect to each parameter evaluated at the initial parameter estimates

b_k = the k^{th} parameter

and b_{k0} = an initial estimate of the k^{th} parameter

(17:5)

The parameters are now in linear form. Linear regression techniques can be used to find estimates of $b_k - b_{k0}$ and hence of the parameter values b_k. These revised estimates are then used as the initial estimates in the next iteration. Iterations continue until the sum-of-squares function is changed less than a pre-specified tolerance or until another stopping mechanism is activated.

A second nonlinear parameter estimation technique is Marquardt's Method. It is an improvement over Gauss' Method in that it does not converge as slowly as Gauss' Method when it approaches the least-squares estimate of the solution.

Tests for the Nonlinear Model

Even though it does not always lead to an unbiased estimate, the MSE is often used as an estimate of σ_ε^2 (4:283). MSE for the nonlinear model is the unexplained variation (sum of squared errors) divided by n-m degrees of freedom, where n equals the number of observations in the sample and m is the number of parameters to be estimated. MSE can then be used when comparing the nonlinear model to other nonlinear models and to linear models.

As with the linear model, confidence intervals can be constructed around the nonlinear parameter estimates, using the t-statistic (7:78) and the estimate of σ_{ε}^{2}. However, testing whether a 1-α confidence interval for a certain nonlinear parameter estimate contains zero may not have the same significance as testing the slope coefficient in the linear model, since the nonlinear model may not have a regression coefficient associated with slope.

The statistical packages BMD Biomedical Computer Programs (3:215) and Statistical Package for the Social Sciences (SPSS) (13:320), both of which have linear and nonlinear subprograms, have regression packages available for use.

Summary

Regression is a curve-fitting technique which describes a relationship between two or more variable quantities. The linear regression model parameters can be estimated using the technique of least-squares. The appropriateness of the regression model is tested using statistical hypothesis testing. Confidence and prediction intervals can also be used to test the regression model. Separate regression lines can be compared as to their slopes and/or y-intercepts.

In addition, nonlinear regression model parameters can also be estimated by a least-squares technique. Now iterative methods must be used to approximate a final solution. A biased estimate of σ_ε^2 for the nonlinear model is MSE. It is the quantity used to determine which regression model (linear or nonlinear) best fits the data.

CHAPTER 4

METHODOLOGY

Introduction

The role the statistical methods, discussed in Chapter 3, can play in analyzing the mechanical properties, some of which were discussed in Chapter 2, of materials is discussed in this chapter. An example with titanium alloy fatigue data is used. This data is the result of fatigue tests performed on various conditions of the titanium alloy Ti-6Al-4V and another titanium alloy Corona 5. All the data has been supplied, in tabular and graphical form, by the Materials Laboratory, AFWAL, Wright-Patterson AFB OH.

The twenty-eight different conditions of Ti-6Al-4V, the titanium alloy discussed in Chapter 2, and Corona 5 to be statistically analyzed are labeled as follows:

1. DFVLR MIXED
2. DFVLR FINE
3. DFVLR COARSE
4. CORONA 5
5. REP LOW TUNGSTEN(W)
6. REP LOW W HEAT TREAT #1
7. SEP LOW W
8. SEP LOW W HEAT TREAT #1
9. REP HIGH W
10. REP HIGH W HEAT TREAT #1
11. SEP HIGH W
12. SEP HIGH W HEAT TREAT #1
13. REP HYDROVAC
14. AS CAST
15. AS CAST HEAT TREAT #1
16. CAST & HIP (Hot Isostatic Press)

17. CAST & HIP HEAT TREAT #1
18. R14
19. R15
20. R16
21. HR15
22. HR16
23. HR17
24. REP + W(IMI) CONDITION 1: As HIP
25. REP + W(IMI) CONDITION 2: HIP + 960°C(1760°F) for 1 hr., Water Quench (WQ)
26. REP + W(IMI) CONDITION 3: Hot Worked + Simulated HIP Thermal Cycle 917°C (1683°F) for 4 hrs., Furnace Cooled (FC)
27. REP + W(IMI) CONDITION 4: Hot Worked + 960°C(1760°F) for 1 hr., WQ + 700°C(1292°F) for 2 hrs., AC
28. REP + W(IMI) CONDITION 5: HIP + 1038°C(1900°F) for 1 hr., FC + 732°C(1350°F) for 4 hrs., AC

The present procedure represents fatigue data by a two-dimensional graph, with the logarithm of the number of cycles to failure(N) on the abscissa (horizontal) axis and stress level (S) on the ordinate (vertical) axis. The data points are plotted and a curve approximating the relationship between S and N is fitted subjectively according to the appearance of the data points by eye or by means of a French curve.

Statistical Analyses Performed

The statistical analyses performed on the fatigue data include fitting two regression models to the data; a linear model--$\hat{y}_i = a + bx_i$--and a nonlinear model--$\hat{y}_i = a + bc^{x_i}$. [This nonlinear model was deemed most appropriate for the data after analyzing several alternative nonlinear models]. Typically when data is presented on a two-dimensional plot, the dependent variable values are read from the vertical axis and the independent variable values from the horizontal axis. For this work, the independent variable is stress level, depicted on the vertical axis, because during fatigue testing, the level of stress is held constant and a value of cycles to failure is determined. The level of stress is then changed and another value of cycles is determined.

The criterion used for determining whether the linear or nonlinear model provides a better fit to the data is MSE. Recall that MSE is an unbiased estimator of the variance, σ_{ε}^{2}, for the linear model. Although MSE is not an unbiased estimate of the variance for the nonlinear model, it is an estimate and can be used for comparison purposes (3:282). The degrees of freedom associated with MSE for the nonlinear model are n-3 rather than the n-2 used in the linear model, because three nonlinear parameters are being estimated. The smaller MSE denotes the better fitting model.

Residual analysis is also performed to assess the reasonableness of the statistical assumptions. This analysis also leads to the detection of extreme values and potential identification as outliers. The assumption of normality is important due to the relatively small (4 to 16) number of data points for each condition tested. The statistical distribution function describing the fatigue life at constant stress cannot be accurately determined, since more than 1000 (2:409) identical specimens should be tested in an identical environment at a constant stress. However, a German experimenter Muller-Stock tested 200 steel specimens at a single stress and found the distribution to be normal when N is expressed as logN (2:409). The finding of Muller-Stock coincides well with statistical Assumption 4, from Chapter 3, which assumed that the errors were normally distributed about the mean.

An analysis was done to examine whether a statistical difference exists between the fatigue strengths of different conditions of Ti-6Al-4V. Personnel of the Materials Laboratory were interested in determining, for example, whether, Heat Treatment #1 improved the fatigue strength of REP HIGH W. The following comparisons of samples were hypothesized by Materials Laboratory personnel (15) to be drawn from the same population:

1. REP LOW W and REP HIGH W
2. SEP LOW W and SEP HIGH W
3. AS CAST and CAST & HIP
4. DFVLR FINE and DFVLR COARSE
5. DFVLR MIXED and DFVLR COARSE
6. DFVLR FINE and DFVLR MIXED
7. REP HIGH W and REP HIGH W HEAT TREAT #1
8. REP LOW W and REP LOW W HEAT TREAT #1
9. SEP HIGH W and SEP HIGH W HEAT TREAT #1
10. SEP LOW W and SEP LOW W HEAT TREAT #1
11. AS CAST and AS CAST HEAT TREAT #1
12. CAST & HIP and CAST & HIP HEAT TREAT #1
13. R14 and R15 and R16
14. HR15 and HR16 and HR17
15. IMI CONDITIONS 1,2,3,4, and 5

Computer Support

Computer support was supplied by the ASD Computer Center. Regression analysis, residual analysis, and comparison of regression was performed using the Statistical Package for the Social Sciences (SPSS) systems of computer programs. Nonlinear regression used the SPSS NONLINEAR Subprogram, and linear regression used the SPSS REGRESSION Subprogram.

Summary

This chapter presented the statistical techniques to be used in analyzing the titanium alloy fatigue data. In chapter 5, inferences are made regarding the best-fit line for the linear and nonlinear model for seventeen of the twenty-eight conditions tested. The remaining eleven conditions were analyzed using only the linear model. Residual analysis detected potential outliers; data points more than two standard deviations from the fitted line were selected for examination as possible outliers. Prediction intervals were then constructed about the new regression line based on a sample which excluded the extreme value. The extreme value was then treated as a new observation. If this point did not lie within the interval, it was denoted as an outlier. When comparing regression lines, several comparisons were made using the nonlinear model, while the remaining comparisons used the linear model. For a few linear model comparisons, the slopes were assumed to be equal. This assumption was made because observation of the plotted fitted regression line indicated that the slopes were similar.

CHAPTER 5

RESULTS AND ANALYSIS

Introduction

This chapter shows the application of statistical methods to mechanical properties of materials, in this example, titanium alloy fatigue data. (The fatigue data is listed in Appendix A). Materials Laboratory, AFWAL, suppliers of the data, were seeking the answers to three questions:

1. Can better-fitting, more accurate S-N curves be determined using statistical methods?
2. Which of the data are outliers?
3. How does the fatigue data from various conditions compare?

This chapter answers those questions by presenting the results of the regression, residual, and comparison of regressions analyses.

Results

The results of SPSS regression and residual analyses for all twenty-eight conditions of Ti-6Al-4V and Corona 5 are presented in Table 5.1. The results are presented in the following fashion: For the linear model, a is the least-squares estimate for α, the y-intercept; b is the least-squares estimate for β, the slope; SSE is the sum of squared errors; MSE is mean square error; 2 SD Outliers are those

TABLE 5.1

Regression and Residual Analysis Results

		DFVLR FINE	DFVLR COARSE	DFVLR MIXED
LINEAR	n	12	11	11
$\log y_i = a+bx_i$	a	7.533	7.965	7.533
	b	$-.306\times10^{-4}$	$-.321\times10^{-4}$	$-.306\times10^{-4}$
	SSE	1.1877	1.2547	1.1781
	MSE	.1187	.1394	.1309
	2 SD Outliers	(65, 2050000)	(80, 37400)	(65, 2087300)
	R^2	.8234	.7852	.8163
	F	46.620	32.904	40.000
	Signif.	.000	.000	.000
NONLINEAR	b_1	3.731	-.092	3.738
$\log y_i = b_1+b_2b_3^{x_i}$	b_2	.732	4.799	.697
	b_3	.628	.935	.612
	SSE	.6484	1.2459	.4632
	MSE	.0720	.1557	.0579
	2 SD Outliers	NONE	(80, 37400)	NONE

TABLE 5.1 (continued)

		REP HYDROVAC	REP LOW W	REP HIGH W
LINEAR	n	9	15	10
$\log y_i = a+bx_i$	a	7.586	8.176	6.641
	b	$-.245\times10^{-4}$	$-.310\times10^{-4}$	$-.199\times10^{-4}$
	SSE	5.5523	2.4766	1.4621
	MSE	.7932	.1905	.1828
	2 SD Outliers	NONE	NONE	NONE
	R^2	.1167	.6428	.4772
	F	.925	23.393	7.302
	Signif.	.368	.000	.024
NONLINEAR	b_1	-131.721	3.604	3.797
$\log y_i = b_1+b_2b_3^{x_i}$	b_2	136.369	1.028	.961
	b_3	.999	.756	.883
	SSE	5.5548	2.2625	1.4538
	MSE	.9258	.1885	.2077
	2 SD Outliers	NONE	(90, 33400)	NONE

TABLE 5.1 (continued)

		SEP LOW W	SEP HIGH W	AS CAST
LINEAR $\log y_i = a+bx_i$	n	16	10	16
	a	8.622	7.048	6.344
	b	$-.360\times10^{-4}$	$-.234\times10^{-4}$	$-.243\times10^{-4}$
	SSE	2.5909	.3408	.8898
	MSE	.1851	.0426	.0636
	2 SD Outliers	NONE	NONE	(100, 1600)
	R^2	.6665	.8924	.8766
	F	27.984	66.331	99.433
	Signif.	.000	.000	.000
NONLINEAR $\log y_i = b_1+b_2b_3^{x_i}$	b_1	4.032	-75.709	2.063
	b_2	.783	80.416	2.250
	b_3	.617	.997	.897
	SSE	2.1233	.3426	.7649
	MSE	.1633	.0489	.0589
	2 SD Outliers	(100, 420300)	NONE	(100, 1600)

TABLE 5.1 (continued)

		CAST & HIP	CORONA 5	R14
	n	14	15	7
	a	6.960	8.037	9.415
	b	$-.248\text{x}10^{-4}$	$-.302\text{x}10^{-4}$	$-.481\text{x}10^{-4}$
	SSE	4.4043	4.7029	.2361
LINEAR	MSE	.3670	.3618	.0472
$\log y_i = a+bx_i$	2 SD Outliers	NONE	NONE	NONE
	R^2	.5115	.4548	.9611
	F	12.567	10.843	123.406
	Signif.	.004	.006	.000
	b_1	3.056	-49.884	-116.709
	b_2	1.559	54.600	121.793
NONLINEAR	b_3	.849	.995	.998
$\log y_i = b_1+b_2b_3^{x_i}$	SSE	4.3064	4.7042	.2391
	MSE	.3915	.3920	.0597
	2 SD Outliers	NONE	NONE	NONE

TABLE 5.1 (continued)

		R15	R16	REP HIGH HT #1
LINEAR	n	9	9	8
$\log y_i = a+bx_i$	a	7.646	7.428	7.968
	b	$-.223\times10^{-4}$	$-.262\times10^{-4}$	$-.326\times10^{-4}$
	SSE	2.0845	.6948	1.2769
	MSE	.2978	.0993	.2128
	2 SD Outliers	NONE	(75, 1306700)	NONE
	R^2	.3455	.7738	.6954
	F	3.696	23.945	13.698
	Signif.	.096	.002	.010
NONLINEAR	b_1	-74.358	4.191	4.025
$\log y_i = b_1+b_2b_3^{x_i}$	b_2	79.555	.322	.343
	b_3	.997	.512	.492
	SSE	2.0856	.2822	.6238
	MSE	.3476	.0470	.1248
	2 SD Outliers	NONE	NONE	NONE

TABLE 5.1 (continued)

		REP LOW HT #1	SEP HIGH HT #1	SEP LOW HT #1
	n	7	9	7
	a	9.117	6.363	9.920
	b	$-.388 \times 10^{-4}$	$-.185 \times 10^{-4}$	$-.451 \times 10^{-4}$
	SSE	.9213	.4233	.0410
LINEAR	MSE	.1843	.0605	.0082
$\log y_i = a + bx_i$	2 SD Outliers	NONE	NONE	NONE
	R^2	.8044	.8118	.9889
	F	20.560	30.193	446.401
	Signif.	.006	.001	.000
		AS CAST HT #1	CAST & HIP HT #1	HR15
	n	9	9	7
	a	6.255	4.723	7.291
	b	$-.204 \times 10^{-4}$	$-.103 \times 10^{-4}$	$-.249 \times 10^{-4}$
	SSE	.5044	.0761	.3992
LINEAR	MSE	.0720	.0109	.0798
$\log y_i = a + bx_i$	2 SD Outliers	NONE	NONE	NONE
	R^2	.8316	.8781	.7799
	F	34.566	50.443	17.721
	Signif	.001	.000	.000

TABLE 5.1 (continued)

		HR16	HR17	IMI CONDITION 1
LINEAR	n	7	6	6
$\log y_i = a+bx_i$	a	7.459	5.741	11.617
	b	$-.268 \times 10^{-4}$	$-.882 \times 10^{-5}$	$-.608 \times 10^{-4}$
	SSE	.8374	.2402	.0426
	MSE	.1675	.0600	.0107
	2 SD Outliers	NONE	NONE	NONE
	R^2	.6421	.3299	.9637
	F	8.9709	1.9689	106.198
	Signif.	.000	.233	.001
NONLINEAR	b_1	4.437	4.660	
$\log y_i = b_1+b_2b_3^{x_i}$	b_2	.004	.077	
	b_3	.054	.285	
	SSE	.3775	.1509	
	MSE	.0944	.0503	
	2 SD Outliers	NONE	NONE	

TABLE 5.1 (continued)

		IMI CONDITION 2	IMI CONDITION 3	IMI CONDITION 4
LINEAR $\log y_i = a+bx_i$	n	4	6	6
	a	13.293	10.645	13.696
	b	$-.709\times10^{-4}$	$-.517\times10^{-4}$	$-.688\times10^{-4}$
	SSE	.0603	.0935	.2000
	MSE	.0302	.0233	.0500
	2 SD Outliers	NONE	NONE	NONE
	R^2	.7666	.9442	.7778
	F	6.568	67.630	14.003
	Signif.	.124	.001	.020

		IMI CONDITION 5
LINEAR $\log y_i = a+bx_i$	n	5
	a	9.240
	b	$-.462\times10^{-4}$
	SSE	.0463
	MSE	.0155
	2 SD Outliers	NONE
	R^2	.9766
	F	125.093
	Signif.	.002

data points which lie more than two standard deviations from the fitted regression line value, the ordered pair listed, if any, is in the form (x_i in thousands of psi, y_i in number of cycles); R^2 is the proportion of total variation that is explained by the regression line (SSR/SST); F is the statistic (MSR/MSE), with 1 and n-2 degrees of freedom, used to test $H_0:\beta=0$, (the regression line does not help explain the variation in y); and Significance denotes the significance level associated with the calculated F-value.

To illustrate the methodology used in the analysis, four representative sets of results--DFVLR COARSE, AS CAST, R14, and R15--will be discussed.

For this work, an R^2 of 0.70 or greater will be considered "good;" that is, the linear model explains the majority (seventy percent or greater) of the variation of the y_i's. Sometimes the nonlinear model was not developed when the linear model indicated a good fit. A significance level of .050 will be the criterion for rejecting $H_0:\beta=0$.

For the nonlinear model, b_1, b_2, and b_3 are the final estimates of the regression coefficients; SSE is the sum of squared errors; MSE is mean square error; and 2 SD Outliers denotes the data points which lie more than two standard deviations from the fitted regression line value.

For both models the grouping of conditions in Table 5.1 has no significance; they are randomly grouped.

Analysis of Four Representative Conditions

DFVLR COARSE. On the basis of lower MSE (.1394 to .1557) the linear model fits a better curve to the data than the non-linear model. The resulting model for DFVLR COARSE is $\log y_i = 7.965 - .321 \times 10^{-4} x_i$. 78.52 percent of the variation is explained by the regression line. An F-value of 32.904 corresponds to a significance level of less than .001; therefore, we are more than 99.999 percent confident that the null hypothesis is rejected correctly. A strong linear relationship between log y and x exists for this sample.

AS CAST. The nonlinear model (MSE = .0589) fits a curve to this set of data better than does the linear model (MSE = .0636). The resulting model is $\log y_i = 2.063 + 2.250(.897)^{x_i}$. Observing the linear regression results, with an R^2 of .8766, an F-value of 99.433, and a significance level of less than .001, the linear regression line explains more than eighty-seven percent of the total variation and demonstrates a relationship between log y and x. Hence, since the non-linear model exhibits less variability (smaller MSE), a non-linear relationship exists between log y and x.

R14. The linear model (MSE = .0472) fits a better curve to the R14 sample data than the nonlinear model (MSE = .0597). The R14 linear model is $\log y_i = 9.415 - .481 \times 10^{-4} x_i$. Ninety-six and eleven tenths percent of the total variation is explained by the regression line. An F-value of 123.406 with an associated significance level of .000 causes $H_o:\beta = 0$ to be rejected. A strong linear relationship exists between log y_i and x_i.

R15. The linear model (MSE = .2978) fits a better curve to the R15 data than the nonlinear model (MSE = .3476). The R15 linear model is log $y_i = 7.646 - .223 \times 10^{-4} x_i$. An R^2 value of .3455 means less than thirty-five percent of the total variation of the y_i's is explained by the regression line. An F-value of 3.696 and a significance level of .096 means the null hypothesis cannot be rejected. Hence, at the five percent significance level, neither the linear nor the nonlinear regression model adequately represents the relationship between log y_i and x_i.

Other conditions. Other conditions are discussed here as special cases in linear and nonlinear model comparison. For REP HIGH W, CAST & HIP, and CORONA 5, the linear model provided the better fit; however, the R^2 values were less than .7000, leaving at least thirty percent of the variation unexplained. The slope coefficient estimates are significant, so the linear model is still more adequate than $y = \bar{y}$. More variation is explained by the slightly sloped line than the horizontal line.

The nonlinear model fit the HR17 data better on the basis of MSE. The linear model slope coefficient was not significant; therefore, since the difference in MSE is small, the nonlinear parameter estimates cannot be assumed to be significant.

The linear model fit the REP HYDROVAC data better than the nonlinear model; however, the R^2 value is below .7000 and the slope coefficient estimate is not significant. Hence the linear model is not an adequate model for this data.

Table 5.2 provides a summary of seventeen of the conditions. It indicates the best-fit model based on MSE, and the resulting numerical model. The remaining eleven conditions were fitted only with the linear model. They all had R^2 values in excess of .7000.

Although all the sample sizes are small, the number of data points in IMI Conditions 2 and 5 are especially small (four and five respectively). For this reason, the variance of the coefficients is great and the resulting confidence intervals are so large that they do not reflect the actual distribution of the coefficients. It is suggested that statistical analyses like the ones used in this paper not be applied to conditions with such a small sample size.

Residual Analysis Results and Outliers

From Table 5.1, nine data points were identified as potential outliers because they lay more than two standard deviations from the fitted regression line value. The extreme values came from seven different conditions: DFVLR FINE

TABLE 5.2

Summary of Regression Analysis

CONDITION	BEST-FIT MODEL	RESULTING MODEL
DFVLR FINE	NONLINEAR	$\log y_i = 3.731 + .732(.628)^{x_i}$
DFVLR COARSE	LINEAR	$\log y_i = 7.965 - .321 \times 10^{-4} x_i$
DFVLR MIXED	NONLINEAR	$\log y_i = 3.738 + .697(.612)^{x_i}$
REP HYDROVAC	NEITHER MODEL ADEQUATE	
REP LOW W	NONLINEAR	$\log y_i = 3.604 + 1.028(.756)^{x_i}$
REP HIGH W	LINEAR	$\log y_i = 6.641 - .199 \times 10^{-4} x_i$
SEP LOW W	NONLINEAR	$\log y_i = 4.032 + .783(.617)^{x_i}$
SEP HIGH W	LINEAR	$\log y_i = 7.048 - .234 \times 10^{-4} x_i$
AS CAST	NONLINEAR	$\log y_i = 2.063 + 2.250(.897)^{x_i}$
CAST & HIP	LINEAR	$\log y_i = 6.960 - .248 \times 10^{-4} x_i$
CORONA 5	LINEAR	$\log y_i = 8.037 - .302 \times 10^{-4} x_i$
R14	LINEAR	$\log y_i = 9.415 - .481 \times 10^{-4} x_i$
R15	NEITHER MODEL ADEQUATE	
R16	NONLINEAR	$\log y_i = 4.191 + .322(.512)^{x_i}$
HR16	NONLINEAR	$\log y_i = 4.437 + .004(.054)^{x_i}$
HR17	NEITHER MODEL ADEQUATE	
REP HIGH W HT #1	NONLINEAR	$\log y_i = 4.025 + .343(.492)^{x_i}$

(LINEAR), DFVLR COARSE (LINEAR and NONLINEAR), DFVLR MIXED (LINEAR), REP LOW W (NONLINEAR), SEP LOW W (NONLINEAR), AS CAST (LINEAR and NONLINEAR), and R16 (LINEAR).

The data points were removed from their samples and new regression lines were determined. The omitted points were then treated as new observations and ninety-five percent and ninety-nine percent prediction intervals, roughly corresponding to two and three standard deviation, were calculated. If the new observation did not lie within the ninety-nine percent prediction interval, it was identified as an outlier and deleted from further statistical analysis, but submitted for fractographic analysis.

An example of a plot of the fitted line with associated ninety-nine percent prediction bands is shown in Figure 5.1. The example uses DFVLR FINE data and presents the nonlinear fitted model. The remaining conditions' fitted lines and ninety-nine percent prediction bands are presented in Appendix B.

The results of this second regression analysis and the ninety-five percent and ninety-nine percent prediction intervals determined at the x_i corresponding to the outlier value are presented in Table 5.3. The prediction intervals are presented in the form (lower limit, upper limit of y).

From Table 5.3 it can be observed that the REP LOW W and SEP LOW W extreme values lie within the ninety-nine percent prediction intervals and are therefore not classified as outliers. The original regression models presented in Table 5.2 are retained.

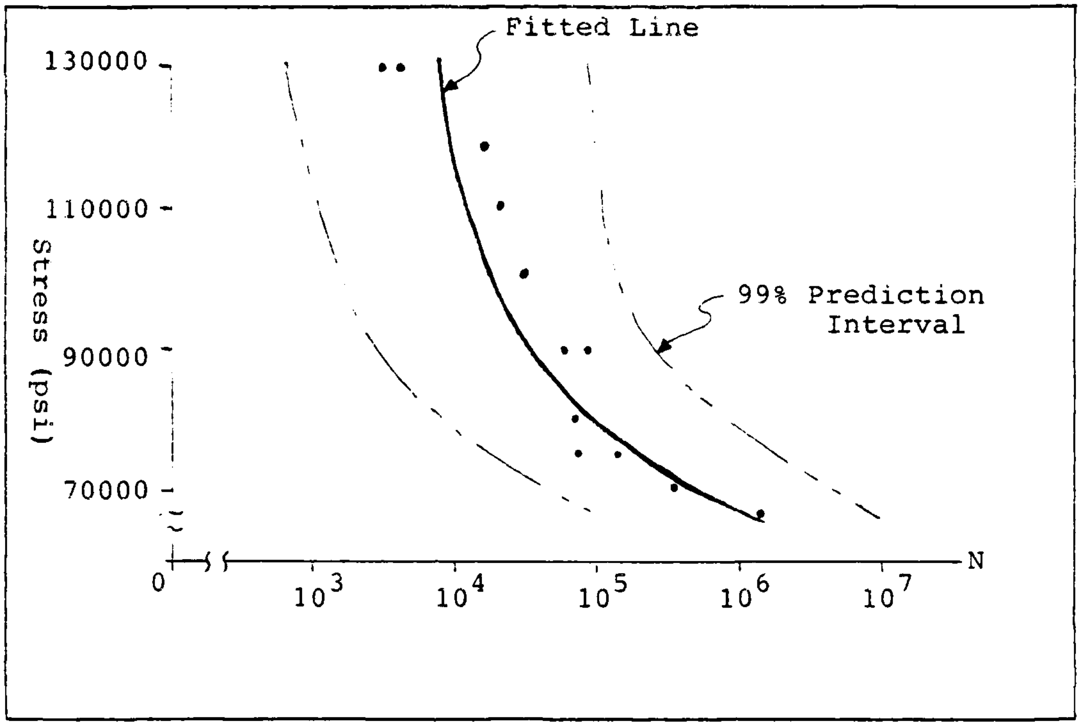


Figure 5.1. Plot of Fitted Regression Line and Associated 99% Prediction Bands

The extreme values for DFVLR COARSE (LINEAR and NON-LINEAR), AS CAST (LINEAR and NONLINEAR), DFVLR FINE (LINEAR), DFVLR MIXED (LINEAR) and R16 (LINEAR) did not lie within their prediction intervals and were discarded. The exclusion of the outliers produces regression lines with less variability and a better fit than the lines determined including the outlier.

Excluding the outliers causes DFVLR FINE, DFVLR MIXED, AND R16 to have different best-fit models than in Table 5.2.

TABLE 5.3

Regression Analysis Results with
Outlier Removed and Prediction Interval

		DFVLR COARSE	AS CAST
LINEAR	a	8.479	6.279
	b	$-.364\text{x}10^{-4}$	$-.229\text{x}10^{-4}$
	SSE	.4388	.3311
	MSE	.0548	.0255
	2 SD Outliers	(75, 1931600)	(40, 616100)
	R^2	.9243	.9423
	F	97.643	212.374
	Significance	.000	.000
	95% P.I.	(94060, 1447426)	(4274, 22387)
	at x_i =	80,000	100,000
	99% P.I.	(50510, 2695401)	(3100, 30864)
	at x_i	80,000	100,000
NONLINEAR	b_1	1.015	1.777
	b_2	3.751	2.600
	b_3	.906	.915
	SSE	.4107	.2657
	MSE	.0587	.0221
	2 SD Outliers	NONE	NONE
	95% P.I.	(95518, 1515083)	(4143, 19387)
	at x_i	80,000	100,000
	99% P.I.	(52232, 2770656)	(3072, 26151)
	at x_i	80,000	100,000

AD-A110 727 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL--ETC F/G 11/6
STATISTICAL ANALYSIS OF METALLURGICAL MECHANICAL PROPERTIES WIT--ETC(U)
SEP 81 G A KILLIAN
UNCLASSIFIED AFIT-LSSR-86-81 NL

2 of 2
AD A
110727

END
DATE
FILMED
3-82
DTIC

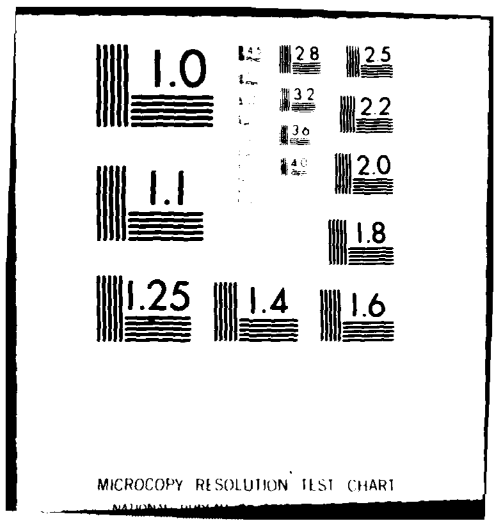

1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

TABLE 5.3 (continued)

		DFVLR FINE	DFVLR MIXED
	a	6.980	6.881
	b	$-.256\times10^{-4}$	$-.248\times10^{-4}$
	SSE	.4247	.3633
	MSE	.0472	.0454
	2 SD Outliers	NONE	NONE
	R^2	.8846	.8848
LINEAR	F	68.998	61.463
	Significance	.000	.000
	95% P.I.	(58252, 734286)	(49789, 693217)
	at x_i =	65,000	65,000
	99% P.I.	(34112, 1253935)	(27352, 1261865)
	at x_i =	65,000	65,000

		REP LOW W	SEP LOW W
	b_1	3.345	4.132
	b_2	1.354	.448
	b_3	.780	.466
	SSE	1.3440	1.4381
	MSE	.1222	.1198
NONLINEAR	2 SD Outliers	NONE	NONE
	95% P.I.	(56171, 2454585)	(6476, 222513)
	at x_i =	90,000	100,000
	99% P.I.	(26667, 5170276)	(3261, 441828)
	at x_i =	90,000	100,000

TABLE 5.3 (continued)

		R16
LINEAR	a	6.505
	b	$-.183\times10^{-4}$
	SSE	.0499
	MSE	.0083
	2 SD Outliers	NONE
	R^2	.9456
	F	104.378
	Significance	.000
	95% P.I.	(72323, 254520)
	at x_i =	75,000
	99% P.I.	(52311, 351891)
	at x_i =	75,000

DFVLR COARSE and AS CAST retained the same best-fit models with parameters slightly altered. Table 5.4 presents the new best-fit regression lines.

TABLE 5.4

Summary of Regression Lines With Outliers Deleted

CONDITION	BEST-FIT MODEL	RESULTING MODEL
DFVLR FINE	LINEAR	$\log y_i=6.980-.256\times10^{-4}x_i$
DFVLR MIXED	LINEAR	$\log y_i=6.881-.248\times10^{-4}x_i$
DFVLR COARSE	LINEAR	$\log y_i=8.479-.364\times10^{-4}x_i$
AS CAST	NONLINEAR	$\log y_i=1.777+2.600(.915)^{x_i}$
R16	LINEAR	$\log y_i=6.505-.183\times10^{-4}x_i$

Testing the Random Error Assumptions

The assumptions of constant variance and normality were examined using the residual plot produced by SPSS. Normality was assumed if no more than five percent of the residuals lay beyond the two standard deviation limits about the mean. Table 5.5 presents the findings of the examination of normality for the twenty-eight conditions of Ti-6Al-4V and Corona 5. Note that the normality assumption does not appear to be violated in any condition, perhaps with the exception of DFVLR COARSE. However, due to the small sample size, the single extreme value represents a larger proportion than with a larger sample. Thus there is no indication from the data that the normality assumption is violated for any of the conditions.

The constant variance assumption was examined using the plot of residuals versus the independent variable. An example of this plot for REP LOW W is shown in Figure 5.2. Examination of the similar plots constructed for the other conditions produced no observation of any trends in the residuals.

E 5.5

ation For Normality

NUMBER OF POINTS BEYOND 95% LIMITS	% OF POINTS BEYOND 95% LIMITS
0	0
1	9.09
0	0
1	6.67
0	0
0	0
0	0
1	6.25
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

f regression lines were made:
ations of nonlinear models, (2)
ions of linear models--SEP LOW W
th SEP HIGH W, and the six

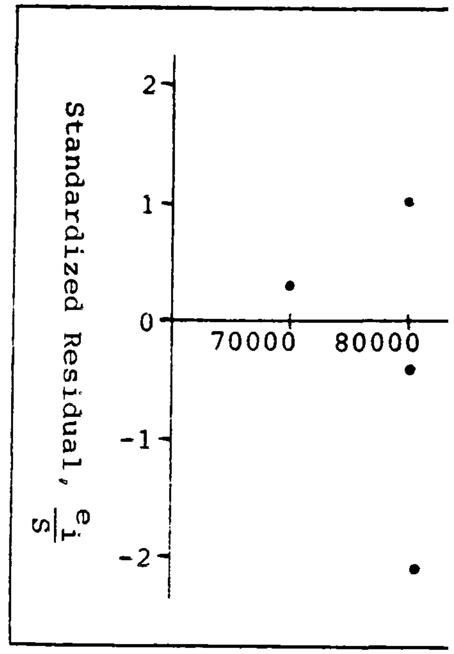


Figure 5.2. Plot of Resi

heat-treated conditions w
ditions, (3) five compari
each comparison a common
comparison of linear mode
Assumptions of common slo
latter two sets of compar
the data and of the regre

Nonlinear comparisons. T
pairs of conditions using
in Table 5.6. For each p

pooled and the following nonlinear model was used for comparison:

$$(34)\quad \log y_i = b_1 + b_4x_{2i} + (b_2 + b_5x_{2i})(b_3 + b_6x_{2i})^{x_i}$$

This form is identical to the original nonlinear model except that a dummy variable, x_2, and three parameters are added. x_2 can equal zero or one to indicate a certain condition. If $\beta_4=\beta_5=\beta_6=0$, then we have no evidence to suggest that the observations were from different populations. Hence we are testing:

$$H_o: \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_a: \beta_4 \neq 0 \text{ and/or } \beta_5 \neq 0 \text{ and/or } \beta_6 \neq 0$$

In all comparisons, the ninety-five percent confidence intervals for b_4, b_5, and b_6 contained zero, leading to the decision not to reject the null hypothesis. That is, we cannot reject the null hypothesis at the five percent level that both conditions within each comparison were drawn from the same population.

Linear comparisons. The results of the comparison of pairs of conditions using the linear model are shown in Table 5.7. The following linear model was used for comparison:

$$(35)\quad \log y_i = a + b_2x_{2i} + b_1x_{1i} + b_3x_{1i}x_{2i}$$

TABLE 5.6

Comparison of Nonlinear Models

CONDITIONS	n	95% Confidence Intervals b_4	b_5	b_6
SEP HIGH W and SEP LOW W(x_2=1)	25	(-478.882, 488.722)	(-488.754, 478.906)	(-1.028, 4.423×10^{-4})
REP HIGH W and REP LOW W(x_2=1)	24	(-9.421, 9.688)	(-9.551, 10.242)	(-.953, .777)
AS CAST and CAST & HIP(x_2=1)	29	(-8.283, 10.812)	(-10.615, 9.008)	(-.522, .393)
DFVLR FINE and DFVLR COARSE(x_2=1)	22	(-13.469, 8.036)	(-7.668, 14.250)	(-.037, .594)
DFVLR MIXED and DFVLR COARSE(x_2=1)	21	(-12.810, 7.363)	(-6.955, 13.604)	(-2.973×10^{-4}, .589)
DFVLR FINE and DFVLR MIXED(x_2=1)	23	(-.631, .644)	(-.841, .771)	(-.269, .237)

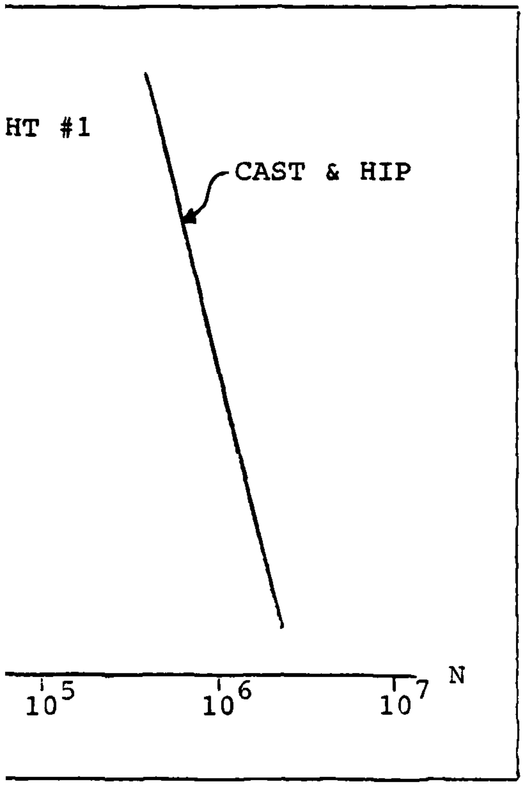
SEP HIGH W(x_2=1) and SEP HIGH W HT #1	19	.685	Not Reject	$-.487\times10^{-5}$	Not Reject
AS CAST(x_2=1) and AS CAST HT #1	24	.024	Not Reject	$-.252\times10^{-5}$	Not Reject
CAST & HIP(x_2=1) and CAST & HIP HT #1	23	2.237	Reject	$-.145\times10^{-4}$	Not Reject
SEP LOW W(x_2=1) and SEP HIGH W	26	1.574	Not Reject	$-.126\times10^{-4}$	Not Reject

0

Once again x_2 is a (
value of zero or one, depen(
five percent confidence int(

In all comparisons b
the ninety-five percent conf
the null hypothesis cannot b
CAST & HIP and CAST & HIP HT
null hypothesis at the five
ninety-five percent confiden
tain zero; thus, the interce
significantly different and
can be expected for the untr
treated condition at every s

The definition of im
When the slopes are equal, t
and the higher line is great
constant amount. Since S-N
log N the constant amount is
Moreover, due to a property
equate the sum of logarithms
[log x + log y ≠ log(x + y)]
cycles to failure is not a c
when the regression line is



& HIP with CAST & HIP HT #1

ST & HIP HT #1 fails at 6252 47 cycles, an improvement = 110,000 psi, CAST & HIP HT #1 HIP at 671,143 cycles, an) cycles. However, the ain constant and is our

<u>Linear comparisons assumin</u> the comparison of pairs of linear model with the slop in Table 5.8. The form of parison of the heat-treate

(36) $\log y_i =$

This is the same m product interaction term i assumption. The ninety-fi about b_2 tests the hypothe

which did not contain zero rejected at the five perce three pairs. SEP LOW W HT ment over SEP HIGH W HT #1 $\log y = .538$ improvement o AS CAST HT #1 showed a log HIP HT #1.

The form of the li R14, R15, and R16 is as fo

(37) $\log y_i = a + b_2x_{2i}$

TABLE 5.8

Linear Comparison of Heat-Treated Conditions

CONDITIONS	n	b_2	$H_o: \beta_2 = 0$
SEP LOW W HT #1 (x_2=1) and SEP HIGH W HT #1	16	.704	Reject
REP LOW W HT #1 (x_2=1) and REP HIGH W HT #1	15	.538	Reject
AS CAST HT #1 and CAST & HIP HT #1	18	.615	Reject

The dummy variable x_3 takes on a value of one when R16 data is entered, otherwise it is zero. Multiple hypotheses of no differences in the intercept for the three conditions are tested for β_2 and β_3. Table 5.9 presents the results of this comparison.

TABLE 5.9

Linear Comparison of R14, R15, and R16

CONDITIONS	n	b_2	$H_o: \beta_2 = 0$	b_3	$H_o: \beta_3 = 0$
R14 and R15 (x_2=1) and R16 (x_3=1)	22	.383	NOT REJECT	-.259	NOT REJECT

nt confidence intervals about
e null hypotheses cannot be
significance level. A signif-
ved between R14, R15, and R16.
.r model used for comparison of
elow:

$b_3x_{3i} + b_4x_{4i} + b_5x_{5i} + b_1x_{1i}$

NDITION 2 data is entered; x_3
osen; x_4 is one when CONDITION 4
when CONDITION 5 is selected.
zero change in intercept are
the results of the IMI compari-

E 5.10

ı of IMI CONDITIONS 1-5

27
.292 Reject
.107 Not Reject
.900 Reject
-.681 Reject

The null hypotheses ar
4, and 5 and not rejected for
shows an improvement of log y
CONDITION 2 shows a log y = .2
and CONDITION 4 demonstrates a
CONDITION 1. CONDITION 3 was
than CONDITION 1.

<u>Linear comparison assuming con</u>
the comparison of HR15, HR16,
model with the intercepts assu
in Table 5.11. The form of th
parison is shown below:

(39) $\log y_i = a + b_1x_{1i} +$ l

x_2 equals one when HR
equals one when HR17 data is
confidence intervals about b_2
zero change in slope between

The ninety-five perce
zero for b_2 and b_3, so the nu
at the five percent significa
significant differences in sl
HR17.

TABLE 5.11

Linear Comparison of HR15, HR16, and HR17

CONDITIONS	n	b_2	$H_o: \beta_2=0$	b_3	$H_o: \beta_3=0$
HR15 and HR16 ($x_2=1$) and HR17 ($x_3=1$)	20	$-.323 \times 10^{-6}$	Not Reject	$.165 \times 10^{-5}$	Not Reject

Summary

This chapter presented regression and residual analysis results for twenty-eight conditions of Ti-6Al-4V and Corona 5. Choice of the better fitting model was made on the basis of smaller MSE. In several conditions, neither the linear nor the nonlinear model adequately fit the data. Nine extreme values were identified, of which seven were classified as greater-than-three standard deviation outliers. Residual plots were examined to test the assumptions of constant variance and normality. Comparison of regression lines was performed between combinations of nonlinear models and between combinations of linear models.

ɅPTER 6

ɅTIONS FOR FUTURE RESEARCH

5Al-4V and CORONA 5 fatigue data
ɔf statistical analysis to metal-
sis of the data has answered three
ls Laboratory, AFWAL:

rves be drawn through the use of
thods?

outliers?

ious data sets compare?

was answered through regression
inear models were tested and the
ition's set of data was chosen
Although the typical S-N curve
near relationship between stress
ɔnditions had a better fit of the
was used. Also data for a few
tely fit by either model.
tliers was examined by selecting
ɔre than two standard deviations
ɹe. Those points were then deleted
ɡression lines were determined.
s as new observations, 99 percent

prediction intervals (rouɡ
were calculated. In sever
values were outside the ir
the sample, identified as
for fractographic examinat
the points were restored t
statistical analysis.

The third questior
sion results. Two or more
analyzed, using dummy var:
For most of the comparisor
sample was drawn from the
rejected at the five perce

Although this exar
sample size, the analysis
reproducible models for r
between stress level and

There is reason t
cal methods could be exte
ses where a relationship
These methods would provi
for describing these rela

earch

ill be presented in two sets:
ysis of fatigue data and the
ication of statistical analysis
The primary recommendation for
arging the sample by performing
ant stress levels. Although
determine an S-N curve using a
two, if any, replications at
8), enlarging the sample size
y associated with determining

ted that other variables be
ip between stress and cycles to
ditional variables which might
n the number of cycles to failure
he specimens were tested, which
ine wear, (2) environmental con-
cy of cyclic stress at which the

or the application of statistical
ctices is to apply statistics
atigue testing. Tensile strength
, and magnetic behavior testing
ssible applications of statistics

00

The example of ti
that statistical methods
lyzing data and describin
future research to explor

APPENDICES

APPENDIX A

FATIGUE DATA

DFVLR COARSE	
Stress Level (psi) x	Number of Cycles to Failure y
130,000	5,300
120,000	13,000
110,000	33,900
100,000	66,200
90,000	94,200
80,000	209,200
80,000	37,400
75,000	379,300
75,000	1,931,600
130,000 (Retest)	6,500
100,000	71,400
CORONA 5	
130,000	7,200
120,000	28,900
110,000	17,400
110,000	50,300
100,000	471,300
100,000	528,000
95,000	25,600
90,000	57,300
87,500	87,800
85,000	1,499,700
80,000	2,434,800
130,000	23,200
85,000	1,393,000
80,000	87,900
80,000	112,300

REP-Lo W	
Stress Level (psi) x	Number c Cycles t Failure y
130,000	14,90
120,000	15,00
120,000	25,40
115,000	32,10
110,000	14,30
110,000	30,50
100,000	281,50
100,000	283,90
90,000	33,40
90,000	704,90
80,000	1,262,20
130,000	9,60
90,000	185,29
140,000	7,80
130,000	69,60
REP-Hi W	
120,000	19,20
110,000	41,80
100,000	19,80
95,000	60,30
90,000	78,70
80,000	17,60
80,000	639,60
75,000	102,90
75,000	255,20
130,000 (retest)	11,70

R14	
:ress Level (psi) x	Number of Cycles to Failure y
20,000	4,100
10,000	9,000
00,000	47,700
0,000	285,000
0,000	323,100
0,000	644,700
5,000	2,276,000
R16	
130,000	14,540
120,000	23,450
110,000	22,200
100,000	51,200
90,000	58,500
90,000	94,400
30,000	110,200
130,000 (retest)	12,700
75,000	1,306,700

CAST (AS CAST)	
Stress Level (psi) x	Number of Cycles to Failure y
110,000	7,700
100,000	9,000
100,000	1,600
90,000	16,000
90,000	20,000
80,000	24,200
80,000	29,200
70,000	32,800
70,000	42,100
60,000	52,500
60,000	74,000
50,000	156,700
45,000	116,500
40,000	616,100
130,000 (retest)	2,600
120,000 (retest)	2,800
REP-LOW W HT #1	
130,000	30,650
120,000	43,660
110,000	12,650
100,000	127,260
90,000	304,580
80,000	2,177,000
75,000	2,185,730

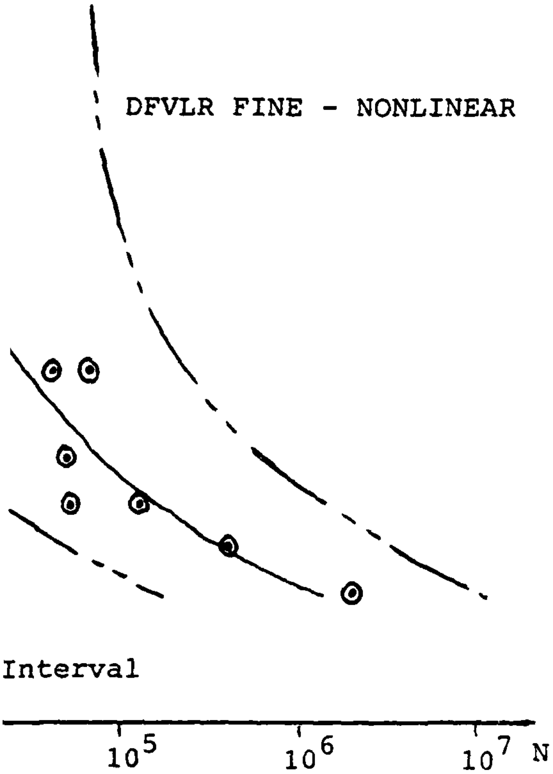
REP HIGH W HT #1		SEP HIGH HT #1	
Stress Level (psi) x	Number of Cycles to Failure y	Stress Level (psi) x	Number of Cycles to Failure y
130,000	9,000	130,000	9,260
120,000	16,200	130,000	11,200
110,000	18,000	130,000	68,010
100,000	32,100	110,000	16,520
90,000	13,000	90,000	62,300
85,000	295,530	85,000	183,600
80,000	292,380	80,000	31,200
75,000	967,190	70,000	105,100
		70,000	111,380
AS CAST HT #1		CAST & HIP HT #1	
130,000	9,400	130,000	1,960
120,000	6,840	120,000	2,770
110,000	3,260	110,000	4,500
100,000	13,840	100,000	6,710
90,000	26,450	90,000	6,460
80,000	31,700	80,000	5,720
70,000	137,200	80,000	8,200
60,000	116,620	70,000	13,400
50,000	159,800	50,000	13,180

Stress Level (psi) x	Number of Cycles to Failure y
IMI CONDITION 2	
130,620	11,436
130,620	15,630
130,620	7,067
123,370	35,296
IMI CONDITION 4	
137,880	15,677
134,250	16,224
132,000	64,529
113,697	95,684
123,370	113,097
137,880	14,452
HR15	
130,000	10,000
120,000	29,300
110,000	31,000
100,000	64,200
90,000	71,900
80,000	94,100
80,000	589,800

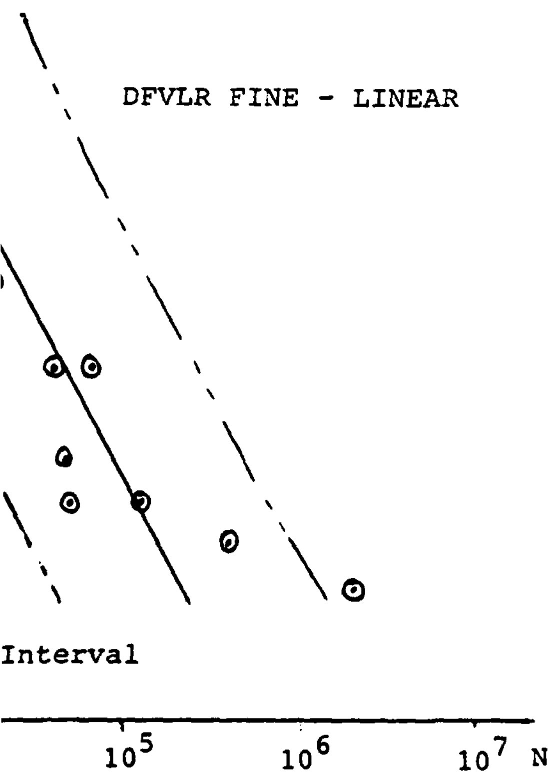
HR16	
Stress Level (psi) x	Number o Cycles t Failure y
130,000	9,40
120,000	24,10
110,000	34,80
100,000	55,90
90,000	53,20
85,000	48,40
80,000	996,10

APPENDIX B

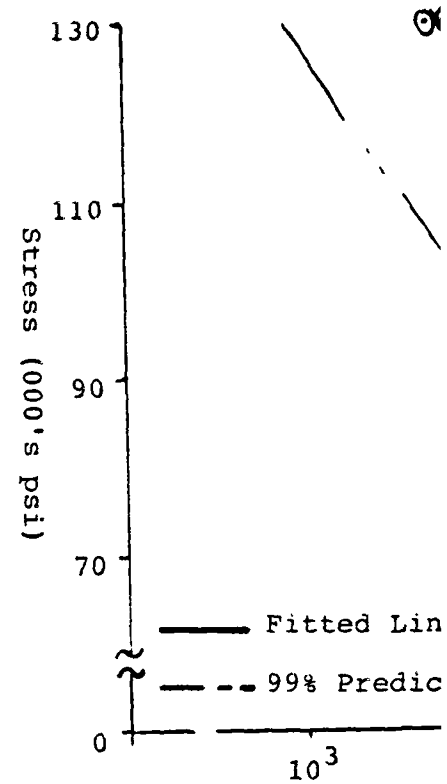
PLOTS OF FATIGUE DATA AND FITTED S-N CURVES



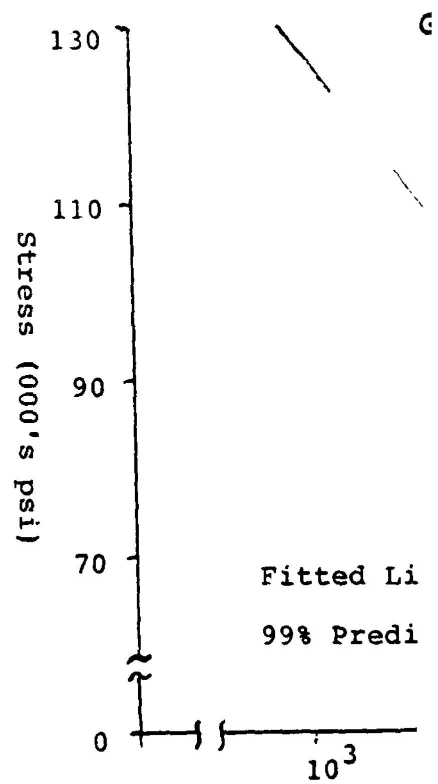
DFVLR FINE - NONLINEAR
Interval
10^5
10^6
10^7
N



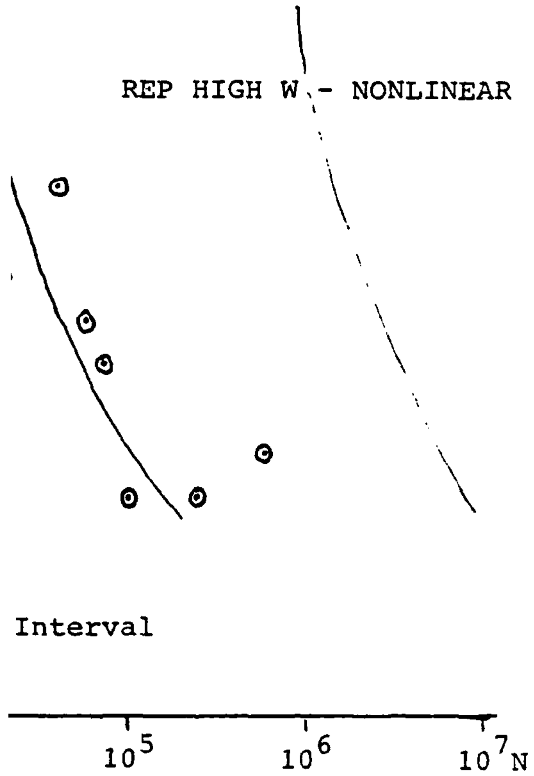
DFVLR FINE - LINEAR
Interval
10^5
10^6
10^7
N

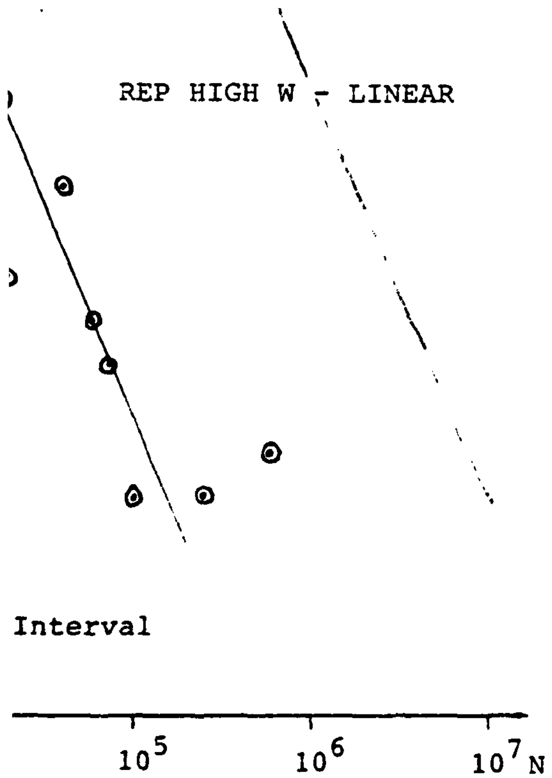


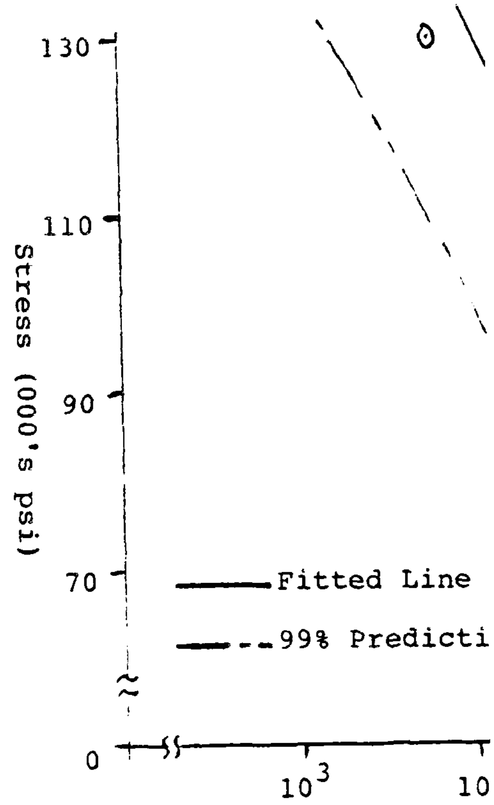
130
110
90
70
0
Stress (000's psi)
Fitted Lin
99% Predic
10^3

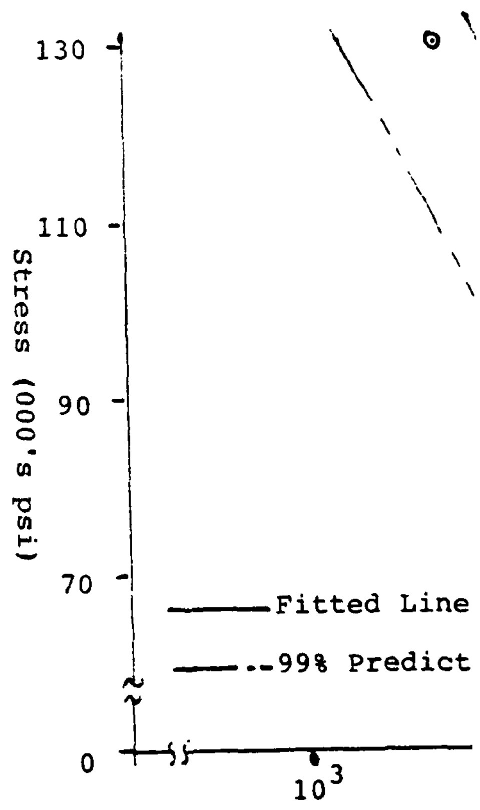


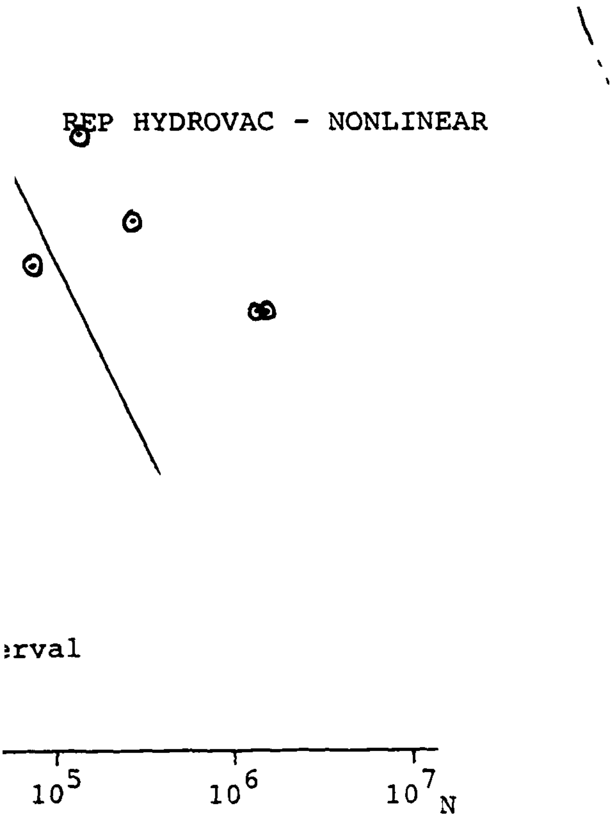
130
110
90
70
0
Stress (000's psi)
Fitted Li
99% Predi
10^3

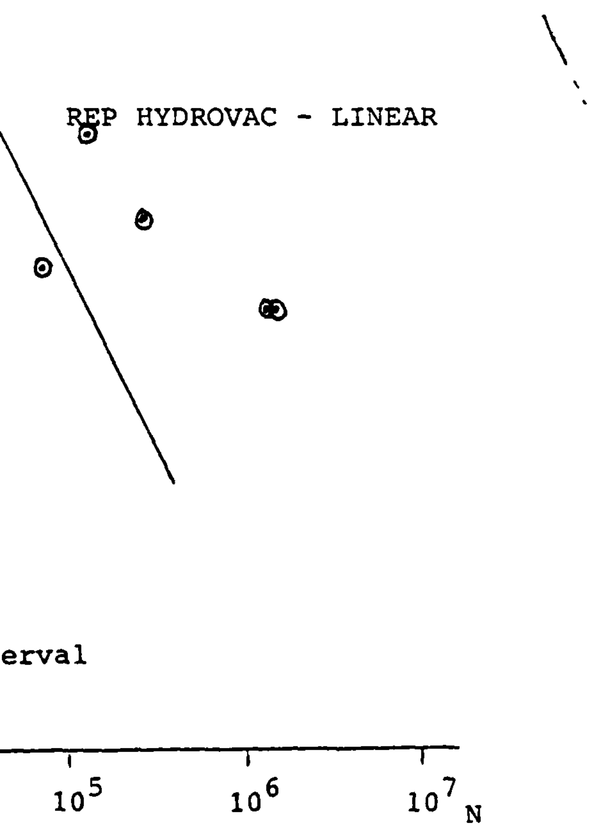

REP HIGH W - NONLINEAR
Interval
10^5
10^6
10^7 N

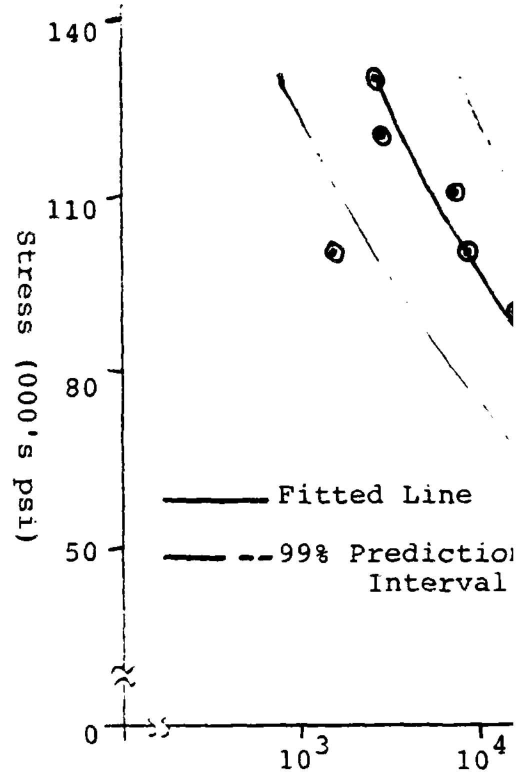

REP HIGH W - LINEAR
Interval
10^5
10^6
10^7 N

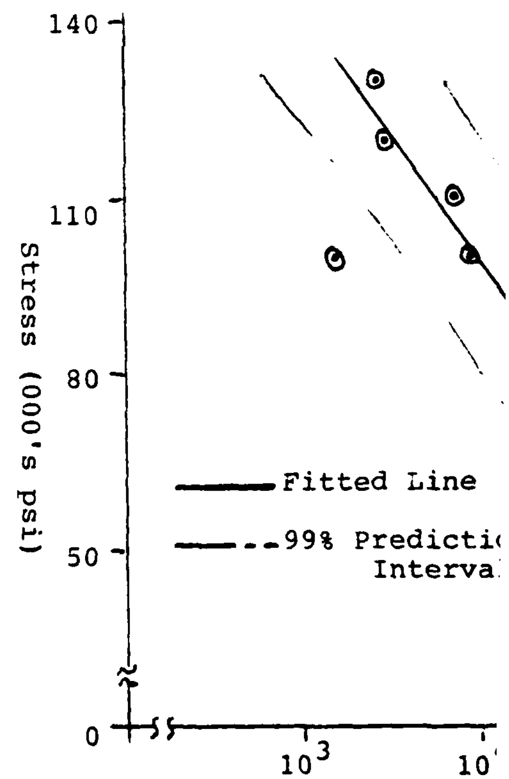

130
110
90
70
0
Stress (000's psi)
Fitted Line
99% Predicti
10^3
10

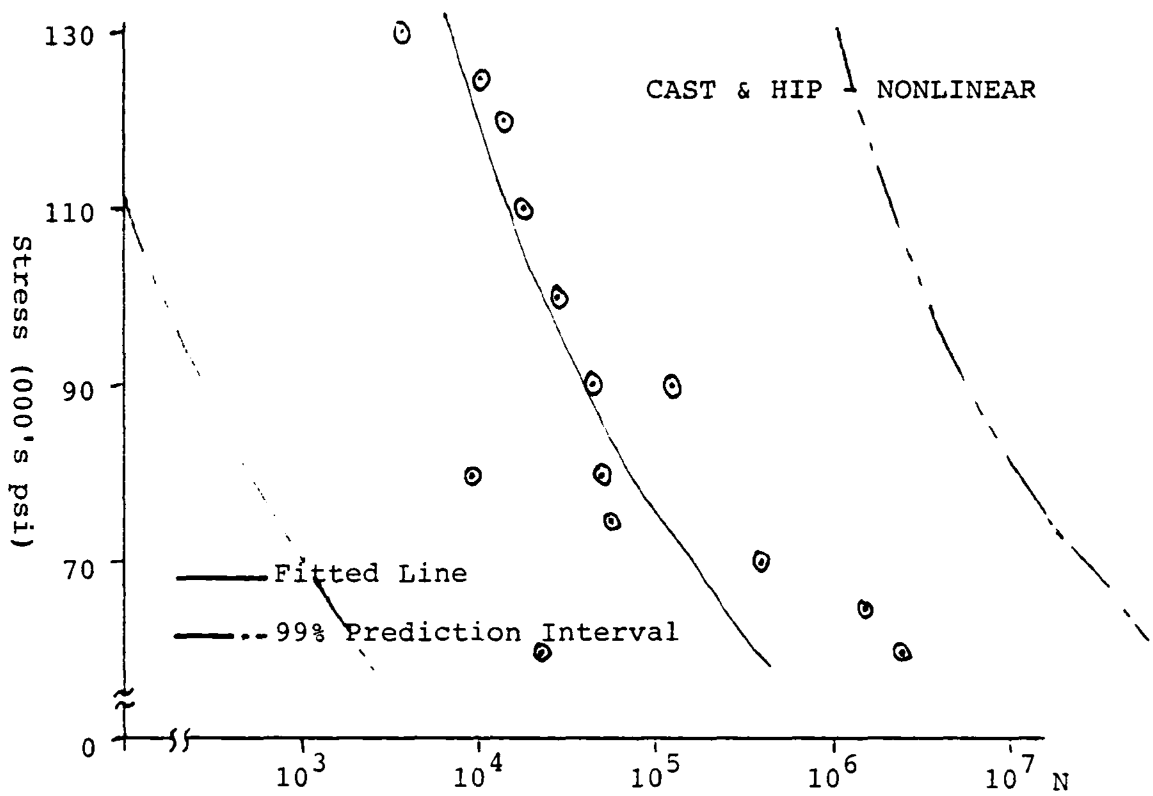

130
110
90
70
0
Stress (000's psi)
Fitted Line
99% Predict
10^3

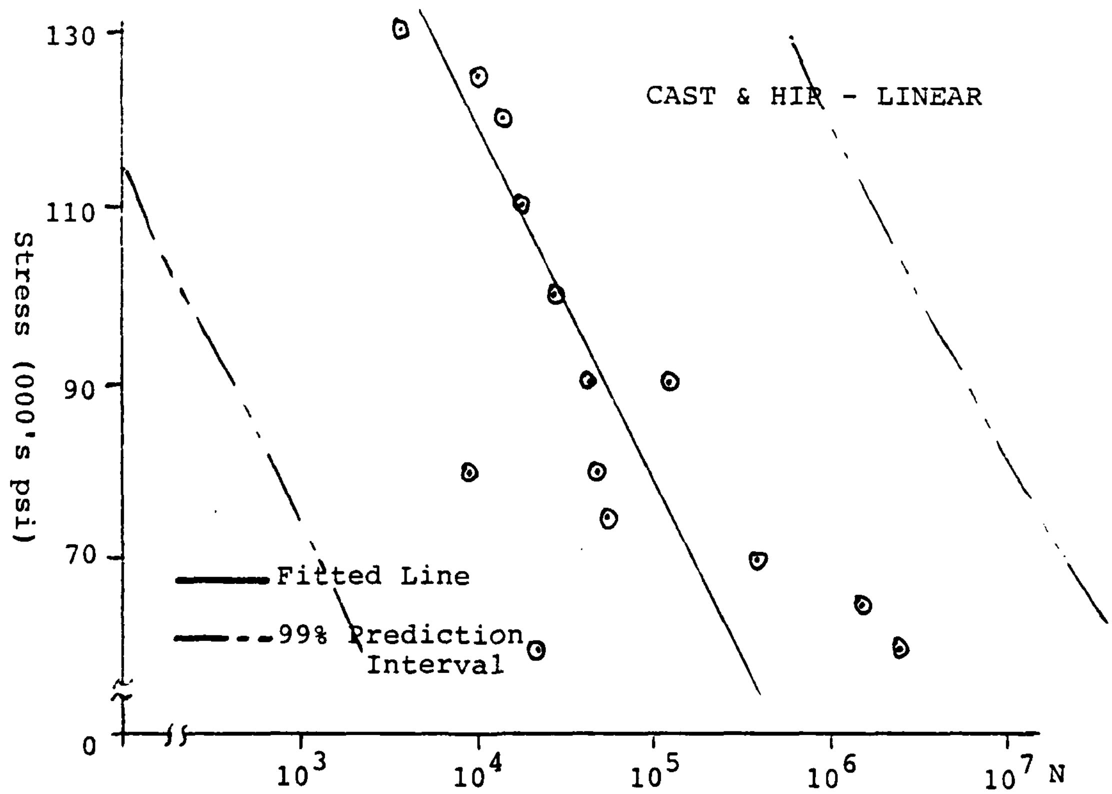

REP HYDROVAC - NONLINEAR
erval
10^5
10^6
10^7
N

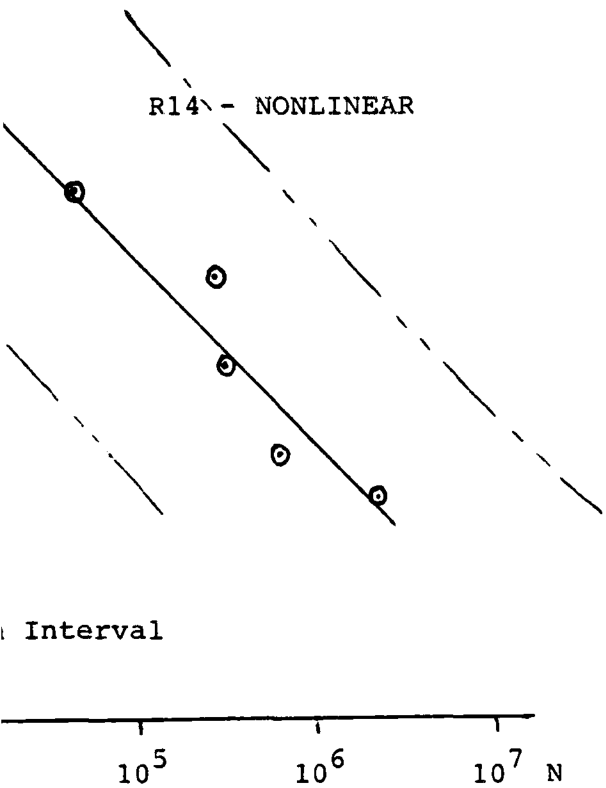

REP HYDROVAC - LINEAR
erval
10^5
10^6
10^7
N

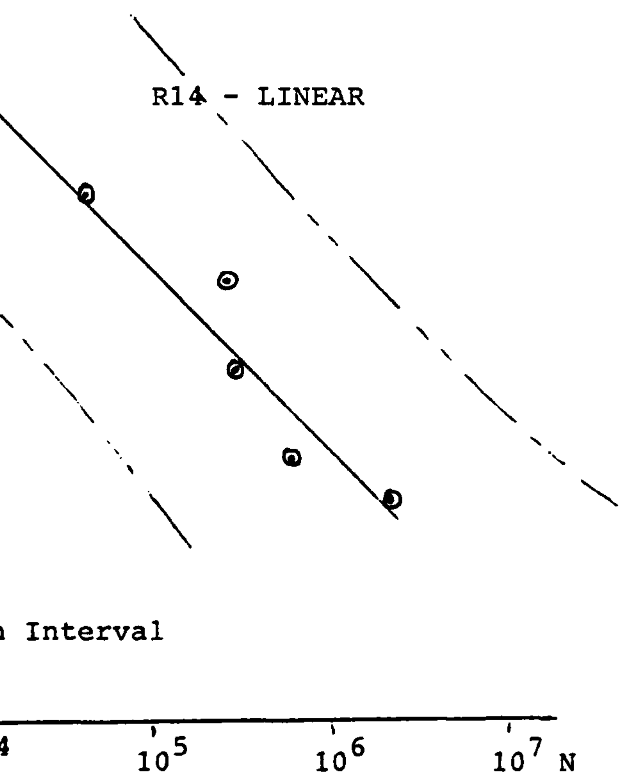

140
110
80
50
0
Stress (000's psi)
Fitted Line
99% Predictio
Interval
10^3
10^4

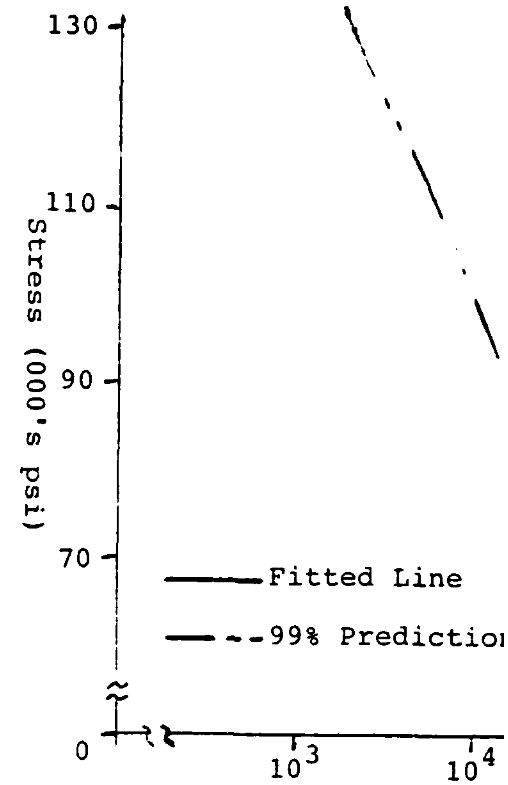

140
110
80
50
0
Stress (000's psi)
Fitted Line
99% Predicti
Interva
10^3

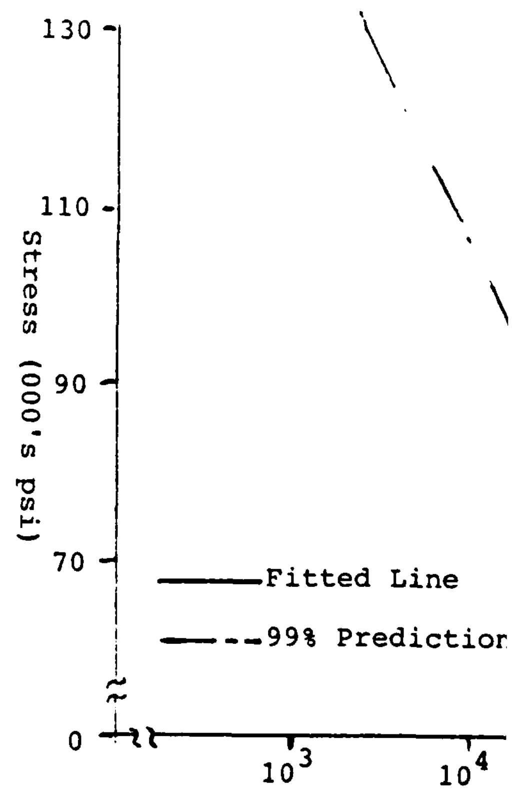

130
110
90
70
0
Stress (000's psi)
CAST & HIP - NONLINEAR
Fitted Line
99% Prediction Interval
10^3
10^4
10^5
10^6
10^7
N

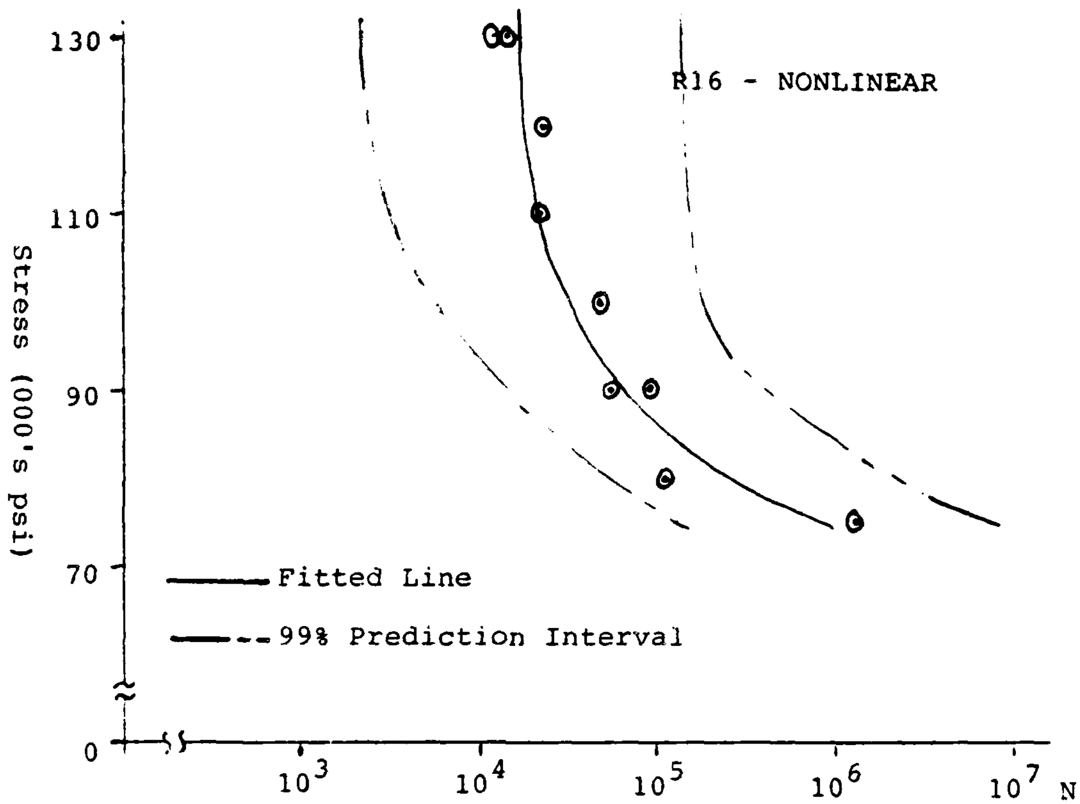

130
110
90
70
0
Stress (000's psi)
CAST & HIP - LINEAR
Fitted Line
99% Prediction Interval
10^3
10^4
10^5
10^6
10^7
N

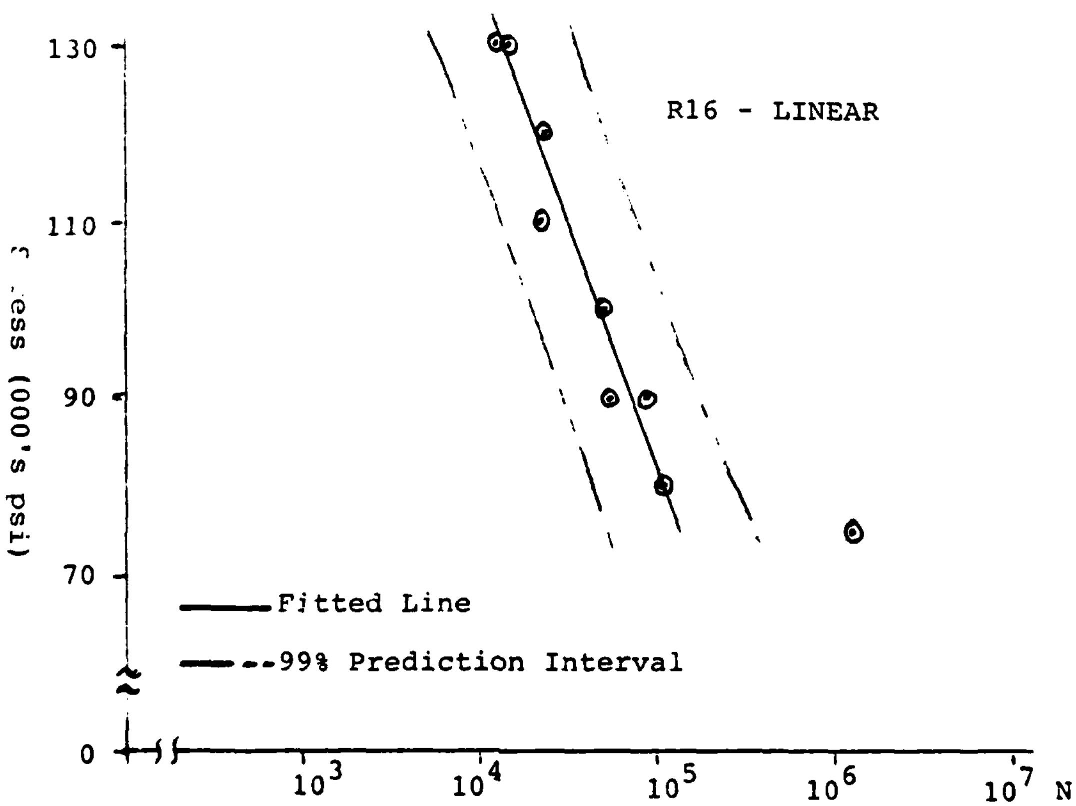

R14 - NONLINEAR
Interval
10^5
10^6
10^7 N

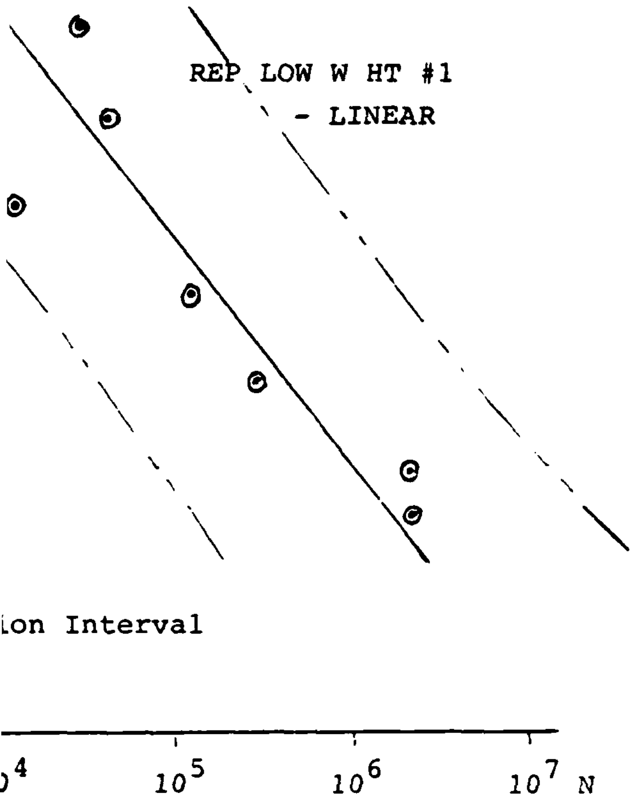

R14 - LINEAR
Interval
10^5
10^6
10^7 N


130
110
90
70
0
Stress (000's psi)
Fitted Line
99% Predictio
10^3
10^4

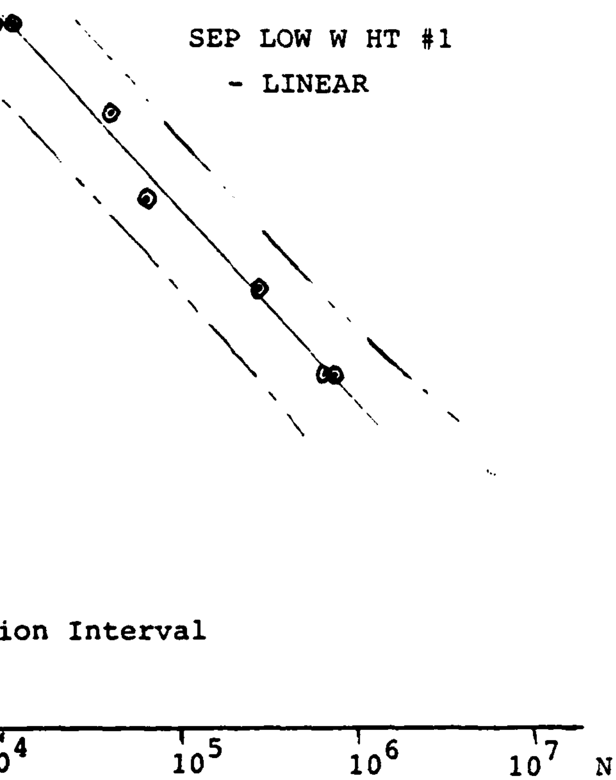

130
110
90
70
0
Stress (000's psi)
Fitted Line
99% Predictio
10^3
10^4


R16 - NONLINEAR
Stress (000's psi)
130
110
90
70
0
10^3
10^4
10^5
10^6
10^7
N
Fitted Line
99% Prediction Interval

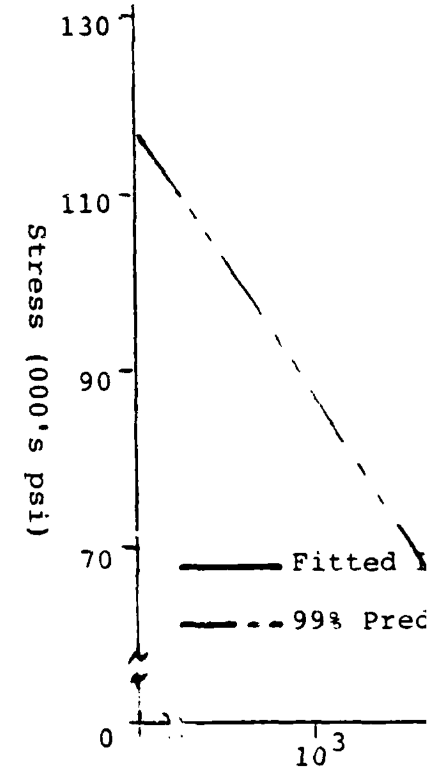

R16 - LINEAR
(000's psi)
130
110
90
70
0
10^3
10^4
10^5
10^6
10^7
N
Fitted Line
99% Prediction Interval



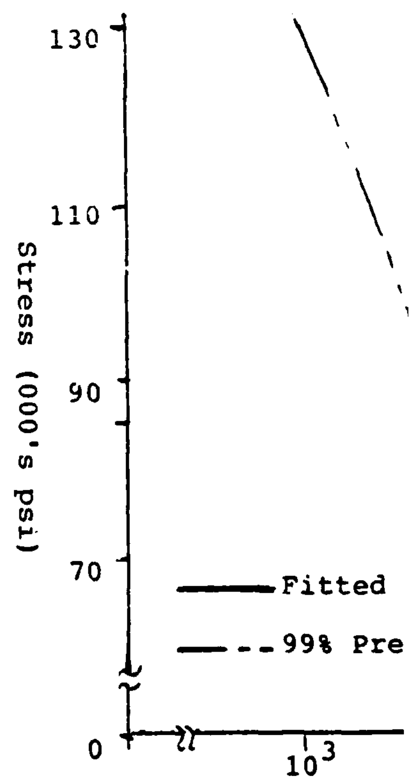
REP LOW W HT #1
- LINEAR
ion Interval
0 4
10 5
10 6
10 7
N



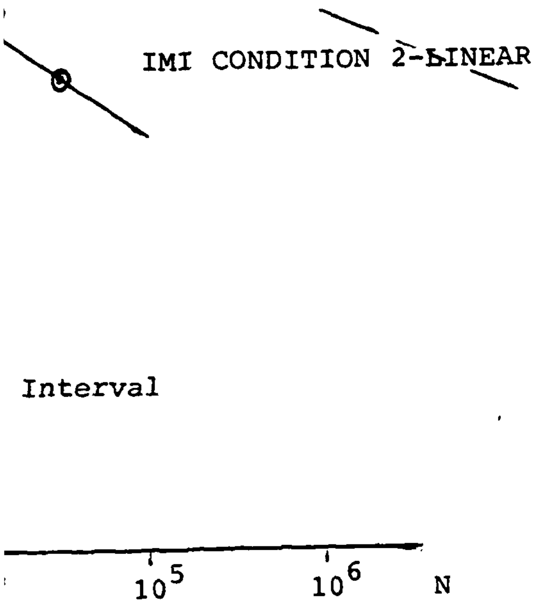
SEP LOW W HT #1
- LINEAR
ion Interval
0 4
10 5
10 6
10 7
N

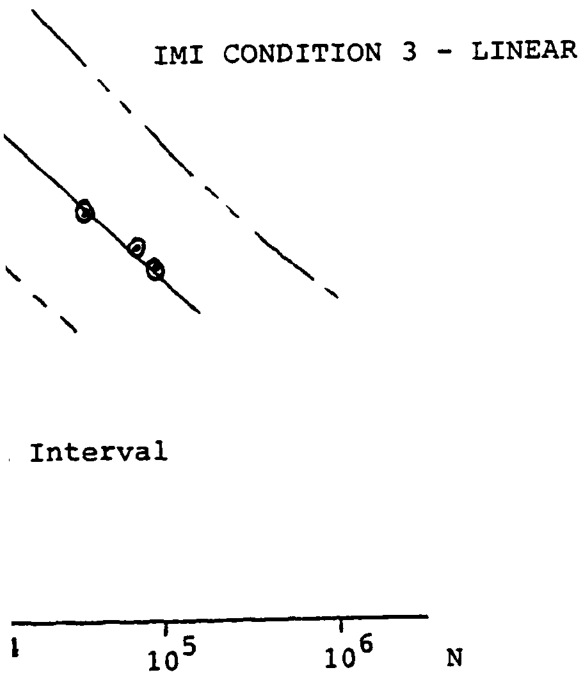


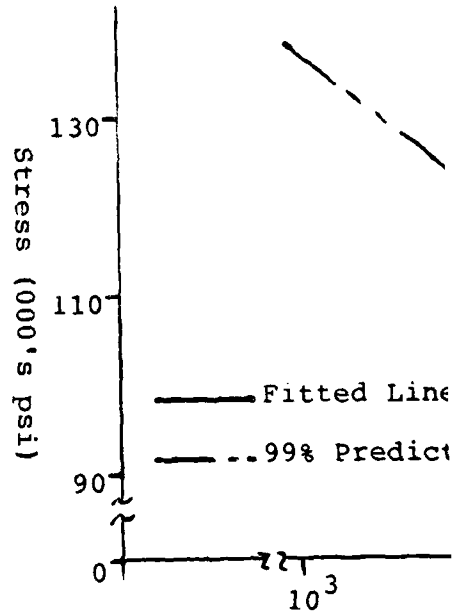
130
110
90
70
0
Stress (000's psi)
Fitted
99% Pred
10 3

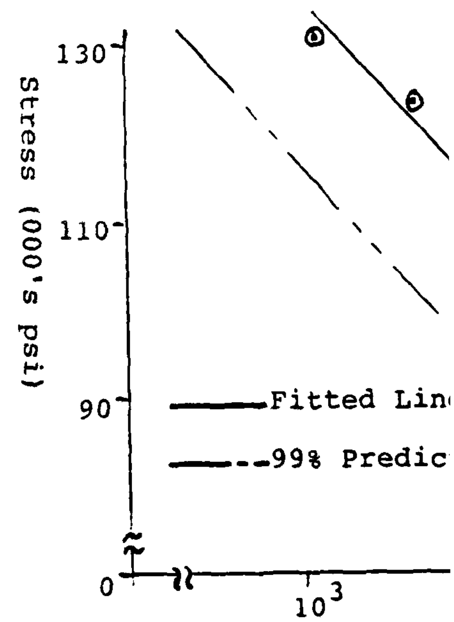


130
110
90
70
0
Stress (000's psi)
Fitted
99% Pre
10 3


IMI CONDITION 2-LINEAR
Interval
10^5
10^6
N


IMI CONDITION 3 - LINEAR
Interval
10^5
10^6
N


Stress (000's psi)
130
110
90
0
Fitted Line
99% Predict
10^3


Stress (000's psi)
130
110
90
0
Fitted Lin
99% Predic
10^3

IBLIOGRAPHY

A. REFERE

1. ASM Committee on Titaniur Properties of Titaniu Metals Park OH: Amer 1961.

2. Dieter, George E. Mechar McGraw-Hill Book Comp

3. Dixon, W.J., ed. BMD: E Berkeley CA: Univers

4. Draper, N.R., and H. Smit New York: John Wiley

5. Froes, F.H. Senior Scier AFWAL/MLLS, Air Force Patterson AFB OH. Pe

6. Froes, F.H., D. Eylon, ar Microstructure and Mi Mechanical Properties Proceedings of the 19 lurgy Conference and June, 1980.

7. Gallant, A. "Nonlinear F tician, Vol. 29 (197

8. Harnett, Donald L. Intro 2d ed. Reading MA: 1975.

9. Kelto, C.A., B.A. Kosmal "Titanium Powder Meta Powder Metallurgy of and J. Smugeresky, e 1980, pp. 1-19.

10. Kleinbaum, David G., and Regression Analysis North Scituate MA:

11. Molchanova, E.K. Phase Translated from Russ and M. Statter. New Company, Inc., 1965.

12. Neter, John, and William Wasserman. Applied Linear Statistical Models. Homewood IL: Richard D. Irwin, Inc., 1974.

13. Nie, Norman H., and others. Statistical Package for the Social Sciences. 2d ed. New York: McGraw-Hill Book Company, 1975.

14. Periodic Table of the Elements. Chicago: Sargent-Welch Scientific Company, 1968.

15. Ralls, Kenneth M., Thomas H. Courtney, and John Wulff. Introduction to Materials Science and Engineering. New York: John Wiley & Sons, Inc., 1976.

16. Reed-Hill, Robert E. Physical Metallurgy Principles. 2d ed. New York: D. Van Nostrand Company, 1973.

17. RMI Titanium. Basic Design Facts About Titanium. Niles OH: Reactive Metals Inc., 1969.

18. Robinson, Barry. SPSS Subprogram NONLINEAR-Nonlinear Regression. Manual No. 433. Northwestern University, August, 1979.

19. Seagle, Stan R. Principles of Alloying Titanium. Metals Engineering Institute, American Society for Metals, Metals Park OH: American Society for Metals, 1968.

20. Smith, Paul. Materials Engineer, AFWAL/MLLS, Air Force Systems Command, Wright-Patterson AFB OH. Personal interview. 28 August 1981.

21. Titanium Metals Corporation of America. "Titanium-the high Strength-to-Weight Metal," Precision Metal, Vol. 30, No. 5 (May 1972), pp. 60-63.

ED SOURCES

Titanium Alloys. "Relation of for Wrought Titanium Alloys," 8th ed. Metals Park OH: ls, 1961.

Industrial and Engineering ber 1955), pp. 2041-2043.

Wayne Helm, and James Tuccy. . Manual No. 414(Rev. C). July 1978.

es, and G.E. Eichelman. rdable High Performance dvanced Air Force Systems," National SAMPE Technical P.E., Azuza CA, 1980.

"Effect of Microstructure on of Ti-6Al-4V," Metallurgical anuary 1976), pp. 111-121.

ichelman, and H.M. Burte. Powder Metallurgy," Journal (February 1980), pp. 47-54).

hods. New York: McGraw-Hill

e Metallography of Titanium. s Inc., 1970.

yler, The Collection, Generation, 5 Allowable Design Data. AFML-TR- ls Laboratory, Air Force Systems AFB OH, May 1974.

BIOGRAPHICAL S

Gary Alan Killian was born on 12 April 1958 in Pottsville PA. He graduated from high school in 1976 and entered The Pennsylvania State University in September of that year. Upon graduation from the University with a Bachelor of Science degree in Metallurgy, he was commissioned in the U. S. Air Force in May 1980. He then entered the Air Force Institute of Technology for graduate study in Systems Management in June 1980.

His permanent address is: 5735 Tibet Drive
Huber Heights OH
45424.

3