22

Effective Use of Test Data for Quality Improvement and Cycle Time Reduction in Radio System Manufacturing

by: Lance Edward Haag

B.S. Electrical Engineering, The University of Minnesota (1985)

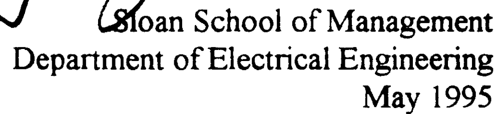
M.S. Electrical Engineering, Stanford University (1988)

Submitted to the Alfred P. Sloan School of Management and
the Department of Electrical Engineering
in Partial Fulfillment of the Requirements for the Degrees of

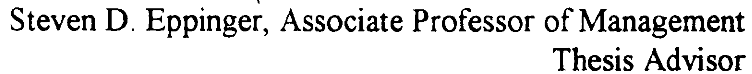
Master of Science in Management
and Master of Science (unspecified)

in conjunction with the Leaders for Manufacturing Program at the
Massachusetts Institute of Technology, May 1995

©1995 Massachusetts Institute of Technology, all rights reserved

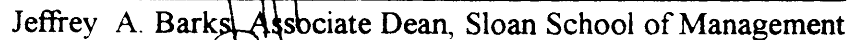


Signature of Author ______________________________
Sloan School of Management
Department of Electrical Engineering
May 1995

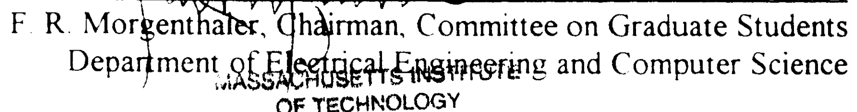


Certified by ______________________________
Steven D. Eppinger, Associate Professor of Management
Thesis Advisor

Certified by ______________________________
Daniel E. Whitney, Senior Research Scientist
Thesis Advisor



Accepted by ______________________________
Jeffrey A. Barks, Associate Dean, Sloan School of Management



Accepted by ______________________________
F. R. Morgenthaler, Chairman, Committee on Graduate Students
Department of Electrical Engineering and Computer Science

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 20 1995 Barker Eng

Effective Use of Test Data for Quality Improvement and Cycle Time Reduction in Radio System Manufacturing

by: Lance Edward Haag

B.S. Electrical Engineering, The University of Minnesota (1985)

M.S. Electrical Engineering, Stanford University (1988)

Submitted to the Alfred P. Sloan School of Management and
the Department of Electrical Engineering, May, 1995,
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Management
and Master of Science (unspecified)

Abstract

Quality improvement and cycle time reduction are generally understood to be key activities for a manufacturing firm to build and sustain competitive advantage. This thesis is about how to connect those high-level business goals to the detailed activities of design engineers and manufacturing workers. I think the context of these activities determines which tactics work well and which don't. Quality control techniques for a machining operation or a paper mill don't necessarily translate well to electronics assembly and test. Even in the field of electronics, tactics that work well in high volume digital circuits like personal computers may not apply well in lower volume production of analog radio systems.

The heart of this thesis is the demonstration projects documented in chapters 5 through 7, where data analyses that result in learning about the product, the process, and the causes of problems are presented. The emphasis is not on the problems themselves, but on the process of turning the test data into relevant information. Indeed, the people involved in the problems had made progress in solving them without these techniques. My motivation was to develop and demonstrate processes to make such projects much easier, leading to higher productivity and a lower activation energy for using data analysis. The tools used include relational databases, computer networks, and visually based statistical analysis software for a personal computer. All of these links are necessary for maximum effectiveness, because any piece that is missing raises the activation energy to the point where data analysis is seldom used in practice.

Thesis Supervisors: Steven D. Eppinger, Associate Professor of Management

Daniel E. Whitney, Senior Research Scientist

Acknowledgments

First, I gratefully acknowledge the support and resources made available to me through the Leaders for Manufacturing program, a partnership between MIT and major US manufacturing companies.

I want to thank my advisors, Steve Eppinger and Dan Whitney, for their moral and intellectual support in dealing with the complex issues that arose at times, helping me make sense out of the confusion.

I'm very grateful for the opportunity Motorola provided me, working with a great company and great people. The managers took time out of their very busy schedules when I needed them. The engineers and others, too numerous to name, gave me time as we learned together about the issues. I hope they value the experience as much as I do.

I'd like to thank Tom Burrell and my other colleagues at Hewlett-Packard, who encouraged me to leave a great job and pursue the opportunity presented by the LFM program. My life will never be the same. Also, I couldn't have written this thesis without the knowledge about test and RF design I gained through my employment at HP.

Finally, my friends and family, both old and new, provided countless hours of support throughout the 2-year LFM program. In particular I'd like to thank Amy Bradley, whom I met during my internship and who became a very close friend and confidant.

Biographical Note

Lance Haag was born in St. Paul, Minnesota on August 16, 1962. He attended the University of Minnesota after growing up in the twin cities and in Bismarck, North Dakota, receiving his bachelor's degree in electrical engineering in 1985.

He took a job with Hewlett-Packard in Santa Rosa, California, where he worked as a manufacturing, test and design engineer over a career spanning eight years. HP generously put him through the Honors Co-op program at Stanford University, which led to a master's degree in electrical engineering in 1988.

He took a leave of absence from HP to attend the Leaders for Manufacturing program in 1993, which led to the writing of this thesis.

Lance's personal interests include mountain biking, wine making, backpacking, and living in Northern California, which makes active pursuit of these hobbies possible year-round.

Table of Contents

Abstract 2
Acknowledgments 3
Biographical Note 4
Table of Contents 5
Chapter 1: Introduction 6
Chapter 2: Investigation 7
- Description of the factory and testing operation 7
- Defining the project 8

Chapter 3: Literature Survey 12
- Quality Control and Total Quality Management 12
- Cost of Quality: Quality System Accounting 14
- Optimization Models Using Mathematical Programming 16
- Robust Design 20

Chapter 4: Frameworks for Quality Improvement and Cycle Time Reduction 22
- The purposes of testing in defect-driven quality control 22
- A probability model for assembly and test 23
- A robust design framework linking manufacturing and design 25
- Design decisions' impact on cost, quality, and capacity 28

Chapter 5: A Manufacturing Improvement Project 32
- Introduction 32
- Problem Definition 33
- Available Data and Experiments 34
- Conclusions 46

Chapter 6: Using Test Data for Robust Design 49
- Fitting models to environmental data 50
- Analyzing factory prototype test data 54

Chapter 7: Developing a Production Test Procedure Using Test Data 58
- Engineering method for test coverage 58
- Data-Driven Verification 60

Chapter 8: Conclusions and Recommendations 65
Bibliography 66
Appendix 1: Interview Guide 68

Chapter 1: Introduction

This thesis documents a 6-month project in process improvement, performed at Motorola, Inc. Major goals of the organization include quality as embodied by the six-sigma program, and a 10X cycle time reduction program. The company believes that these goals are critical to customer satisfaction and to long term success and growth. Figure 1.1 shows how quality improvement and cycle time reduction go together in a simple system of cause and effect. This kind of diagram is used at several points in the thesis. An arrow indicates a proposed cause-effect relationship. The label "o" or "s" denotes the direction of causation as being opposite or same.

To interpret fig. 1.1, it is assumed that variation reduction and quality improvement go together (see chapter 3). Lower cycle time and cost will result by the mechanisms shown. The connection from variation reduction to WIP inventory may not be obvious--it stems from the operational need for inventory buffers when there is variation in processing times, failures, availability, etc. Cycle time is directly related to WIP inventory, as inventory can be measured in days of output.

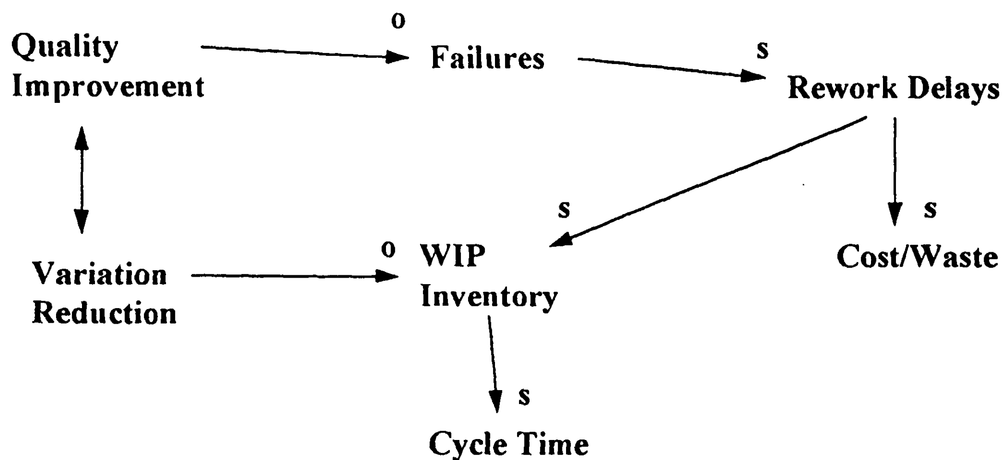


Fig. 1.1: The linkages among quality, cycle time, and cost

This diagram represents the connections in manufacturing. In the rest of the thesis, more exploration of product development and the links among manufacturing, quality, time to market, and order responsiveness will be explored. High level analysis and detailed methods will be presented. As the thesis title implies, the focus is on test data and its effective use for improving performance.

Chapter 2: Investigation

Description of the factory and testing operation

At the beginning of the project, the abstract mission was given to improve the testing process. There was a sense that too much testing was being done in the factory, with some tests being repeated at several stages of integration. Figure 2.1 shows the assembly and test process as observed by the author, as it fits in with the value chain for the radio system. Of necessity, it is simplified, showing the upstream assembly and testing of one type of subassembly, receivers. The other material flowing into the system assembly and test area also has upstream subassembly and testing areas in the factory. The receiver area involves the most testing and value-added activity of any of the subassembly test areas. The receiver is also the module with the most design work done by Motorola, while the other modules are largely designed by suppliers according to Motorola specifications.

At each stage of testing, a functional test is performed on the product assembled at that stage. Functional tests verify the electrical performance against functional specifications. At the board and receiver levels the product is placed in a fixture that connects it to test equipment intended to simulate the inputs it will see, and measure the outputs of the product. Adjustments are also made to calibrate the product during the functional test.

The functional test is an important distinction, as no electrical testing is performed to directly verify the correctness of the assembly. Such direct verification is known as in-circuit testing in the industry. In the past it was decided that in-circuit testing was not effective enough to justify its cost for this product in this factory. Factors affecting this decision included the defect rate of the assembly operation and the error rate of the in-circuit testing. Many components such as small-value capacitors, high quality inductors and filter modules, radio-frequency transistors, etc. cannot be probed by an in-circuit tester without altering their performance. Many other components cannot be reliably measured, as their tolerances are more precise than the measurement uncertainty of the in-circuit testing equipment. Thus, the error rate of in-circuit testing was high enough to make it of questionable value, given the capability of the assembly operation. Optimization models exist to quantify the tradeoffs in this kind of problem (See ch. 3), but the decisions were made by engineering and management judgment in this case.

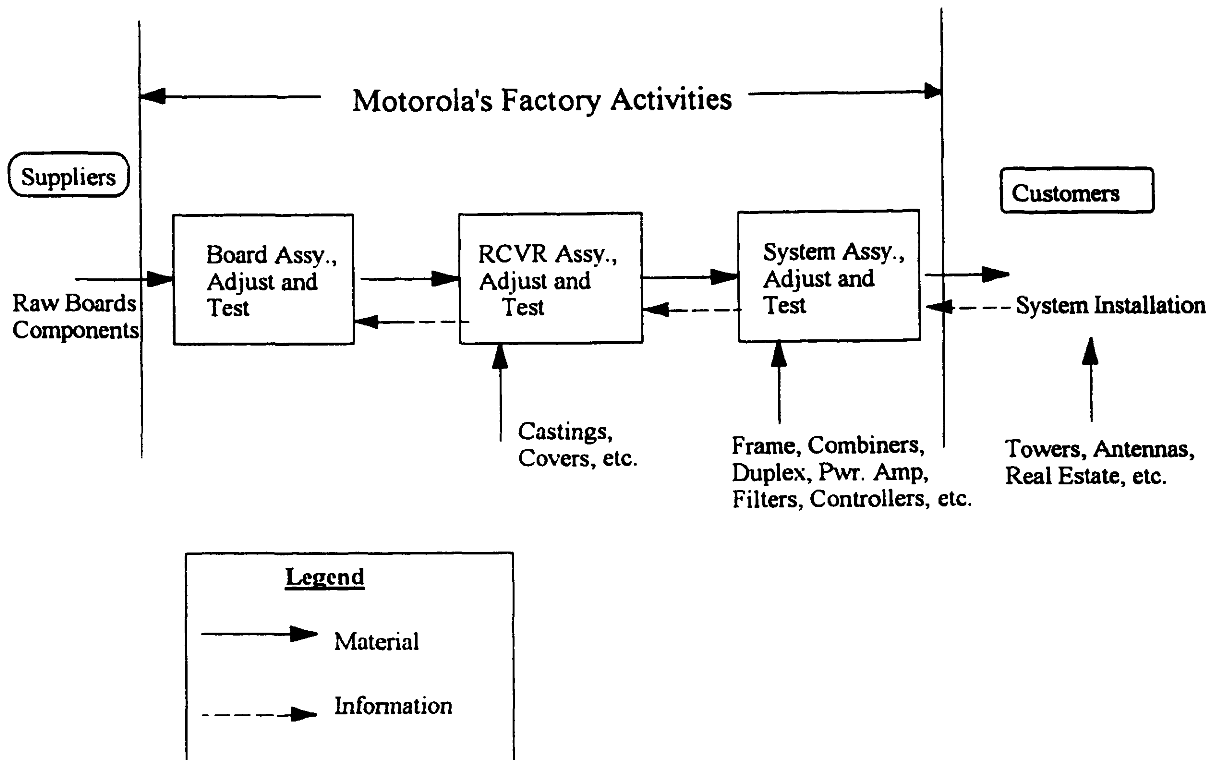


Figure 2.1

Because there is no in-circuit test, the functional test is required to be effective at detecting all types of defects. In addition, the test is needed to screen performance in some cases when the robustness of a new design needs to be improved. A framework for discussing the purpose and requirements of the functional test is presented in chapter 4. The espoused goal of each stage in this operation is to minimize the number of rejects by the next operation downstream, at a minimum cost.

Defining the project

In order to clarify the situation and find a specific project for improvement, the author set out to interview engineers and managers in manufacturing and design. The interview for each person consisted of five open-ended questions, focusing on the purpose of testing, the use of test data, and areas where improvements might be made. The interview questionnaire is included in appendix 1. During each interview, the subject did most of the talking, with the interviewer trying to find the facts behind the viewpoints held by the

subject. The duration of each interview was 1-2 hours. The question "Why do you think that?" was asked many times. Six people were interviewed, ranging from 3-15 years in experience, including mechanical and electrical engineers, a technician, and an engineering manager.

By the sixth interview, many of the same ideas were being repeated, so the author decided to move on and analyze the results. The affinity diagram method of Kawakita Jiro (KJ method) was used to do this analysis [1,2]. The theme for the KJ was chosen as "What is the weakness of the production test process?" The statements from the interviews were written in paraphrased form onto over 80 cards, each with a different meaning. Using criteria of importance and relevance to the theme, as well as dissimilarity of meaning, 29 labels were eventually chosen as the basis for the KJ. The goal in this process was to reduce the number of cards, retaining all of the important ideas from the interviews. The four highest-level conclusions and their supporting high-level labels were as follows:

- Demands on the Test Process Change as the Market Matures (Pressure for lower cost and higher quality, increase in volume)
- We don't Want to Risk Having Dissatisfied Customers

  - Ability of tests to predict customer satisfaction increases downstream

  - Consensus among manufacturing, development, and quality assurance is needed to remove a special test or inspection step.
- Information is the Lifeblood of Quality Improvement

  - Production tests give information about weaknesses in the product and process

  - Clues about inputs that drive the outputs are missed
- Organizational structure can impede change

At the practical level, the third conclusion about the importance of information was chosen as the basis for the project. The factory had been installing ever more powerful networks of computers, storing failure information and parametric test data on a relational database. All of the raw data from tests was stored on networked computers in text form, as well. The data was being used to track failures and assign priorities to improvement efforts, as well as for generating high level reports for management. Also, the database served as a resource to engineers and technicians investigating problems. If a module

failed in the customer's hands, it was possible to pull the test results for that module out of the database, find what station and what operator tested it, as well as test results for the boards that went into the module. Of course, the same was true for failures inside the factory.

Figure 2.2 shows the conceptual quality improvement process, as practiced in this operation. The "s" and "o" labels indicate cause and effect relationships, in the same or opposite directions. E.g., an increase in production failures causes an increase in the need for inspection. The central concept to this diagram is that the output metric driving corrective action is production failures, or defects. In electronics assembly, defect-driven quality control has been established as a powerful method, with the Pareto chart and cause-effect diagrams used as tools for quality improvement. The diagram shows corrective action directed in three prime areas. Design choices (where the part technology is a surrogate for many complex design decisions) have an effect on the variation relative to requirements, and thus the failure rate. Supplier quality control also affects production failures, as does the in-house assembly and calibration process capability. The effect of increasing failures on increased cost is also shown.

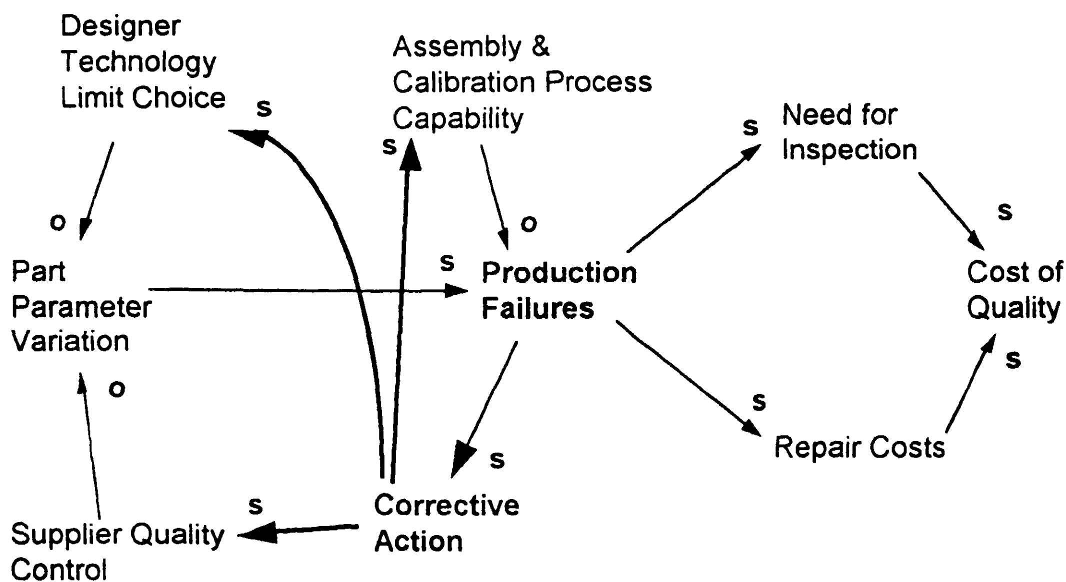


Fig. 2.2: Defect driven quality improvement

Defect driven quality improvement is at the heart of the Motorola 6-σ quality program, as the defect rate is the metric that is compared to the goal. The implied objective of the 6-σ program is zero defects. I.e., perfect quality is the same as zero defects. Of course, there is much more to a quality program than its primary metric, including the methodology for reducing the defects. In assembly of complex electronic systems, it has been difficult to apply traditional statistical quality control techniques, for many reasons. First, there are many specifications for a complex product--which of these should we control? The suggested SPC tool for this case in one statistics reference [23] is the p-chart, which measures the defect rate. Second, many failures in a complex product are due to special events, not to a system of chance causes. A simple defect can cause wild swings in the product's performance. The assumptions of normality behind many control charts are simply not met for this kind of product. Sometimes the performance has large tails in the distribution, may be multi-modal due to groups of components from different vendors, etc.

An implicit advantage of defect-driven quality improvement is the relatively small amount of information to be handled. Only information from defective products is included in the dataset, which is a very small portion of the total set of test data produced by a reasonably capable process. E.g., if 5% of the production population fails at least one test, a failure-driven quality improvement process need only consider 5% of the test data. Of course, diagnostic data from the repair activities must also be considered.

There was a general feeling among the interviewees that much more value could be created if the rest of the test data could be used more effectively. The vast amount of data in the computers was sometimes difficult to get to, or made it difficult to get only the information relevant to the question at hand. Furthermore, tools for summarizing the data and effectively turning it into information for problem-solving were needed. The opportunity to use the information to move the process more towards a state of normality (Absence of special causes) was chosen as the basis for this thesis. Improvement in manufacturing and development cycle time, quality and cost were seen as potential benefits. Synergies were noted between the need for data analysis in production and in product development engineering. Methods and justifications for doing so will be the subject of the chapters that follow.

Chapter 3: Literature Survey

After the investigation was complete, a search for literature pertaining to the use of test data for process improvement was conducted. The results fall into several broad categories. First, there is a tremendous volume of material on quality control, the use of statistics, Total Quality Management, etc. Second, there were a number of interesting papers on the cost of quality, particularly the costs of inspection and repair processes. Third, there are many papers on the optimization of test or inspection processes using mathematical programming techniques. Finally, there is material on robust design of products, which is intended to minimize variation by design. Some of the most relevant ideas and references will be discussed in the paragraphs that follow.

In the chapters that follow, materials from the TQM and statistics literature will be used, as well as the references on robust design. The project did not make significant use of the optimization or cost of quality references.

Quality Control and Total Quality Management

A central idea appears in many Quality Control books, including Deming's "Out of the Crisis" [5] and Juran's "Quality Control Handbook" [4]. This idea is that inspection is a poor method of quality control. As an alternative, these authors stress the importance of seeking out the sources of variation in the output of a manufacturing process, and working to eliminate or control those sources of variation. Juran refers to the difference between reality and meeting the customer requirements every time (The first time) as "Waste." Juran's book is a very large reference, with detailed guidelines for virtually every facet of quality control. Of most relevance to this project, he puts emphasis on the importance of using test to measure the process, not just individual products.

The general approach to using Statistical Quality Control (SQC) is to first get the process into a state of statistical control, and then use control charts to both monitor the key output variables and key input variables. When input variables are monitored, the term Statistical Process Control (SPC) is often used. An example of SPC would be if temperature and pressure were charted and controlled on an injection molding machine, rather than the dimensions or strength of the plastic parts made by the machine. A key question is whether or not a process is in a state of control, in the statistical sense. A process is in control when "The variation in the output is produced by a stable system of chance causes." SQC experts say that control charts are not useful when a process is not in control--they are meant for a process that is in control, where the purpose is to ensure that it stays in control.

The primary tools for getting a process into control and controlling it are test data coupled with statistical analysis. The how and what of finding variation and reducing it is the primary topic of this thesis. The key methods for getting a process into control are root-cause problem solving, experiments and analysis of variance. Pareto charts assist with prioritizing the problems. A demonstration of these methods will be presented in chapter 5. When establishing control of a process, the key input parameters that drive the output should be identified, so that the process can be made less sensitive to their variation, or the level of variation can be controlled.

Once control has been established, control charts should be used. There are many kinds of control charts. A p-chart tracks defect rates by groups of product. Each point on a p-chart is the defect rate for the group of products it represents. P-charts are useful for complex products with many opportunities for defects and/or a complex set of specifications, such as our radio systems. The p-chart is used for "attribute" data. Attribute data is a classification system, where product is sorted into two varieties, good and bad. "Variable" data, on the other hand, is used where a continuous performance parameter is measured. The most common control charts for variable data are the x-bar and r-charts, where each point on the chart is the group mean and the group range (max - min). For all control charts, control limits indicate when the process has fundamentally changed--i.e., a special cause of variation has started to occur. Each point on the control chart is a formal hypothesis test about the distribution of the output. Any point outside the control limits represents a rejection of the hypotheses that the mean and variance are constant.

The references also coach the SQC practitioner to avoid over-controlling. This means that it is worse than useless to assume that every variation is due to a special cause, and to attempt to eliminate it by an adjustment. Points that are in control on the control chart indicate that no action should be taken. Such actions tend to increase the amount of variation. Deming calls this kind of effort "Tampering." An example in a high-skill factory environment comes to mind. Attempting to force all the operators to match the defect rate of the best operator would be an example of over-controlling--such practice flies in the face of natural variation among human beings. The increased stress will probably cause the less capable operators to perform worse in the long run. Deming shows a fascinating demonstration of the effects of tampering in his funnel drop experiment.

The field of Total Quality Management, or TQM is related to SQC. Indeed, TQM claims to include SQC in its broadly scoped definition. In Shoji Shiba's book, "A New

American TQM," [6] TQM is described as a combination of Customer Focus, Continuous Improvement, Total Participation, and Societal Network Learning. The most relevant concepts from TQM for this project are the ideas of customer focus and continuous, or cyclical improvement. Customer focus provides guidelines for priority. Not all specifications are equally important to customers, and it is possible that some customer requirements are not reflected in the specifications. Quality Functional Deployment (QFD) [7,8,10] is a process in TQM for ensuring that the product specifications and production processes serve the customer needs.

Along with customer focus, continuous improvement is presented as a repeating cycle of activities, supported by several tools. Data is analyzed with the Pareto chart to decide what problem to work on next. Problem solving tools such as the fishbone chart, scatter plots and histograms are used to diagnose the data, generating clues for problem-solving insight. Collectively, these are called the 7 quality improvement tools. The tools are simple enough for ordinary line operators to understand and use, thus enabling total participation. In chapter 5, extensions to these tools that are even more powerful will be demonstrated. A project may be launched with experiments to verify the hypothesis and its solution. The solution is implemented, and its effectiveness verified. Then, the next problem from a new Pareto chart is defined, and the cycle repeats. Shiba calls this cycle the "Plan-Do-Check-Act" cycle.

Cost of Quality: Quality System Accounting

Many of the quality books and TQM books make no attempt to justify their advice in dollar terms. They tend to use statements like "Quality improvement is a necessity in today's competitive environment," or "Competitive advantage goes to the company with better quality." There is a movement, expressed originally by A.V. Feigenbaum in "Total Quality Control," [14] to set up accounting systems to allocate the costs of quality (or lack thereof) into several categories. Presumably, this kind of accounting system will aid management on where to focus cost reduction efforts, justification for investments in improvement, and so on. The ASQC (American Society of Quality Control) sponsors the Annual Quality Congress, whose annual proceedings [15] contain many case studies of efforts to implement cost of quality systems.

In short, quality costs come in two types: costs of control and costs of failure to control. Costs of control include prevention and appraisal, while costs of failure to control include internal failure and external failure costs. Prevention costs include such items as maintaining a quality department, employee training, investments in design

improvements, and so on. Appraisal costs include the purchase and maintenance of test equipment, the building space required for testing, the people needed to perform the testing, and all other costs associated with evaluating the level of product quality. Internal failure costs include scrap and rework costs within the company, while external failure costs include the less tangible costs of unsatisfactory quality outside the company, including repair and replacement, and loss of goodwill. Typical figures quoted by Feigenbaum put failure costs at 65-70 cents out of every quality cost dollar, while 20-25 cents are spent on appraisal, leaving 5-10 cents for prevention. The optimum allocation of these costs will vary from product to product, but it is assumed that a good measurement system coupled with some experiments will assist the operation in finding its optimum.

Feigenbaum points out the existence of a common vicious cycle, where poor quality causes high failure cost, which in turn makes costly 100% inspection needed (Appraisal). These appraisal costs stay high as long as the failure costs stay high. Furthermore, the inspection does little to eliminate the defects--some of the defective products are going to leave the factory. The proposed alternative is to make investments in prevention, showing the benefits through the new accounting scheme. Reduction in both failure and appraisal costs occur together as these investments are made. The article by Manikas and Eichenlaub [16] shows a method for calculation of savings by investing in prevention. It breaks down internal failure costs among analysis and repair, deferred capacity expansion, and work-in-process inventory carrying costs. It also gives formulas for estimating savings based on impacts on failures per board, analysis time, repair time, and the number of repair loops per failure. The results are formulas where one makes an assumption on the percentage improvement the proposed investment is likely to make in these metrics, and the resulting savings are given.

Another interesting paper by Hungerford [17] analyzes the variables that affect test cost through simulation. The results are that the initial cost of test equipment is the least important variable among vendor-related factors. Fault coverage, test time, and test error rate are the most important variables, with their relative importance depending on their relative magnitudes. These are costly because of their labor-intensive nature: they are variable costs, for which no scale economies exist. Because of constraints in the availability of skilled labor for diagnosis and repair, deferred capacity expansion costs can become very high if the test system lacks coverage or gives misleading information. Costs of lack of coverage (False acceptance) are mainly the cost of passing defects downstream or even externally, while false rejections cause longer diagnosis times (The technician can't

trust the system, and so must spend more time assessing the failures manually) and needless re-testing.

Optimization Models Using Mathematical Programming

One possible use of test data is as inputs for optimization models. In the field of operations research, there have been many attempts to develop models whose use build insight and allow improvements of inspection processes. Mathematical programming is the technique of building sets of equations where an objective is defined as a function of many decision variables. The objective is generally to be maximized or minimized, subject to constraints placed on the decision variables. Various classes of mathematical programming, such as integer, linear, non-linear, and stochastic programming have their own classes of techniques for solving the problem. The solution is in the form of values for all the decision variables. The coefficients in the equations, the values of the constraints, etc. are generally obtained from data--data from the test process, in our case.

In the opinion of the author, none of the papers found in the search do an adequate job of capturing the complexity of the problem, at least for a multistage process with this level of complexity. The paper by Burman generally supports this opinion [18]. There are two primary problems. First, the amount of data required is enormous and seldom available in the form required by the models. Second, the assumptions made to allow the problem to be solved are unrealistic for the radio system that is the subject of this thesis. This is not to say that the models are useless in general. Details follow on two models in particular.

"A Dual Ascent Algorithm for Finding the Optimal Test Strategy for an Assembly Sequence," Pappu [20]

Pappu's model takes as inputs the probabilities of specific defects occurring, the coverage of tests mapped to the defects, and the costs of performing each test option and repairing each defect at each stage in production. The constraint is that all defects are covered by at least one test, where the customer is assumed to be able to detect any defect that leaves the production process. The model also assumes that the probabilities of each defect occurring are independent of each other, and that the tests are perfect at detecting whether or not the defect is present. The model considers when a defect becomes detectable in the assembly sequence, and seems to be a powerful approach in situations where the assumptions are reasonable.

However, Pappu's model would be difficult to apply to the complex radio system, for several reasons. First, there are many thousands of possible defects, which do not have a direct correspondence to the available tests. Many functional test outcomes are not independent of each other, as defects often occur in runs (E.g. Solder shorts due to a paste dispensing problem, or incorrect components due to the wrong reel being loaded onto the part placement machine). Many functional test results are also driven by common underlying parameters, thus making them correlated and not independent. The problem with the number of defects could probably be dealt with by aggregating the defects into failures of functional tests. I.e., the failure of a functional test is the definition of the defect. Production history and/or prototype data could be used to estimate the probabilities of these failures. However, the tests are not perfect, and the customer is often not capable of detecting the failure to meet a functional specification. Test errors are often the limiting factor for quality improvement for complex electronic equipment, as attested by Peterson's M.S. thesis, "Testing for Time-Dependent Failures in Electronic Products," [22] and by the author's experience. In effect, there is a race between the performance of the test equipment and the product being tested. For these reasons, Pappu's model was not used for this project.

"Inspection for Circuit Board Assembly," Chevalier and Wein [21]

This paper considers both the problems of allocating inspection and of setting acceptance limits at each stage. Test error probabilities are included in the model. After careful consideration, this approach was also ruled out because of its data demands and assumptions. The authors were forced to develop a heuristic to aid for setting specifications, due to the difficulties associated with getting data and the sensitivity of the model to errors in the data. This heuristic stated that the optimal acceptance criterion was almost always less stringent than the true requirement in the presence of test errors (I.e., the production spec. should be looser than the customer spec.). This result is the opposite of conventional practice, where production limits are stricter than customer requirements to avoid risk of shipping defective products. After careful analysis, the author agrees with this result, but only when the assumptions are warranted. The following is an excerpt from an explanation given by Chevalier in correspondence about his paper:

"You are a little bit puzzled by the fact the probability of being in spec stays high even when your measurement is somewhat out of spec. This is nevertheless true. To try to explain to you why this is true let us take an

example. Let us imagine that we measure a component right at the limit of the specification. We assume the measurement error is symmetrically distributed. Now the true value of the component being measured could be inside or outside. If the distribution of this true value was uniform and there were as many good components as bad ones, then the symmetry of the measurement error would imply that the probability of the component being good is equal to the probability of the component being bad.
But of course in the real cases there are many more good components than bad components (at least so we hope). So when your measurement is at the limit of the spec it is actually much more likely that you are measuring one of the many good components that were in spec and that you got some measurement error that brought it to the limit, than that you are measuring one of the few bad components and its measurement error brought it closer to the spec limit. In general it is so much more likely so that the probability of a component being good stays very close to 1 even when you are slightly out of spec. This is the intuitive argument. The mathematical argument goes as follows:
Let A be the event that the component is in spec.
Let B be the event that the value x was measured.
A standard result of conditional probability gives $P(A|B)=P(A,B)/P(B)$"

The particularly difficult assumption was that of normally distributed performance of a parameter. Data observed on these radio systems shows that the tails of the distribution are often non-normal, and sometimes multi-modal. This is due to the extreme sensitivity of performance to some types of defects. For instance, if a defect that occurs 1% of the time shifts the distribution by many standard deviations, there will be a small "bump" in the tail of the distribution. This kind of data was observed for the radio system for several performance parameters. Another example is when standing waves in cables affects the output power of a transmitter. In this case, the maximum and minimum errors are the most likely to occur, while the average value is the least likely. To the authors' credit, the normality assumptions are probably valid for testing of component values on circuit boards in many cases. In these cases, the sloping tail of a normal distribution seems reasonable.

Although the final result of the paper will not used here, several useful results that will be used in the following chapters are contained within the paper. First, a simple optimality condition was given for the setting of a specification. The probability that the product is truly defective when the specification limit is measured should be equal to the marginal cost of an additional repair at the present stage divided by the marginal cost of passing the defect on to successive stages. Mathematically,

$$P(Bad|\, y = y_{spec}) = \frac{MC_{Stage_i}}{MC_{Stage_{i+1}}} \qquad (1)$$

where y is the parameter being measured and y_{spec} is the specification limit. In most cases it is more expensive to detect, diagnose, and repair a defect downstream, making this optimum probability less than 0.5. In particular, if the stage of interest is the last stage in the process, the cost in the denominator is the cost of shipping a defective product to the customer. Many companies have the policy that the probability of shipping defective product should be much less than 1%, or even zero, which would imply a very expensive cost indeed for shipping the defective product. Of course, it is often not realistic to speak of a clean division between defective and non-defective product. This issue will be discussed more in the following section on robust design.

Chevalier and Wein also give formulas for calculating this conditional probability, based on the distribution of test error and the distribution of true performance. If the distribution of parameter y is p(y), and the distribution of error is e(y), the probability that the product is out of tolerance given the measurement of value y is:

$$P(Bad|y) = 1 - \frac{\int_{LSL}^{USL} e(y-u)p(u)du}{\int_{-\infty}^{+\infty} e(y-u)p(u)du} \qquad (2)$$

Unfortunately, only the distribution of measured performance is generally known. As a function of y, the distribution of measured performance is given by the convolution integral in the denominator of (2).

An extensive survey of other papers is included in Pappu, some of which allow for test errors. "Optimal Location of Inspection Stations in a Multistage Production Process," by Eppen and Hurst does consider test errors, for instance. "A Survey of Models for Allocating Inspection Effort in Multistage Production Systems," Raz [19], provides a nice summary of the research. Raz's review criticized many of the models for oversimplifying the problem. For this reason, and due to time constraints, the author decided not to pursue the mathematical programming avenue. However, there may be fertile ground others to do so--this critique was by no means exhaustive.

Robust Design

The techniques of robust design are relevant to this thesis because data collection and analysis are critical elements of the robust design process. Robust design is essentially the process of making the quality of a product less sensitive to variations in input parameters, environment, aging, and usage. A robust product may not require inspection in production to achieve high quality, thus leading to lower costs. It presents a philosophical alternative to high levels of control of input parameters and inspection of the output in manufacturing. In Feigenbaum's terminology, investment in prevention is made to allow lower spending on assessment and failures. Genichi Taguchi [12] is credited as one of the pioneers in developing robust design techniques. His approach using designed experiments to minimize sensitivity to "noises" is especially useful. He refers to "Signal/Noise ratio" as the quantity to be maximized during these experiments. Critical design parameters are identified by various methods, and the experiments carried out either physically or analytically. Clausing [10] lays out a detailed, structured process for designing in quality, while Phadke [11] clarifies and extends Taguchi's principles.

A related powerful concept of Taguchi's is that of a continuous quality loss function. In this idea, any deviation from target performance carries a cost penalty (loss of customer goodwill), even if the performance meets the standard. A loss function may be parabolic in shape, with zero loss occurring only at the target and some relatively high loss occurring at the specification limit. A loss function always is implied by testing and acceptance metrics. The traditional inspection model assumes zero loss whenever the performance is in spec., and a large loss whenever the spec. limit is exceeded. I.e., all units that meet the standard are equally good and carry no quality loss. Taguchi's continuous loss function has the desirable feature of providing a metric that continuously indicates when reductions in variation are achieved, leasing to lower cost and higher quality. It also supports the philosophy espoused by Deming, et. al., that reduction of variation is far superior to inspection as a quality control method.

Several difficulties in applying robust design techniques to the radio system became apparent during the project. First, with the complexity of the system, it was difficult to identify critical parameters. Also, many of the circuits, such as oscillators, were re-used from other designs and the original designer was no longer available. Engineers contemplating modifying the design of such circuits rarely had time to become expert enough to name the critical parameters with confidence. Some data analysis techniques to help discover these critical parameters are presented in chapter 6. Second, even when the parameters are identified, it is difficult to conduct experiments with different parameter

settings because the experimenter cannot control them. Some examples would be: forward current gain and junction capacitance of a bipolar transistor. Neither parameter can be set arbitrarily by a circuit designer, unless a simulation tool is available that accurately predicts performance of a circuit. The designer only has the physical transistors available to work with, and must make time consuming measurements to even know what the physical parameters are.

In each of these areas of literature, practice and theory, the effectiveness of the approach is directly dependent on the availability and disciplined analysis of test data. The data and its analysis are a link in the chain, which needs all of its links to be strong.

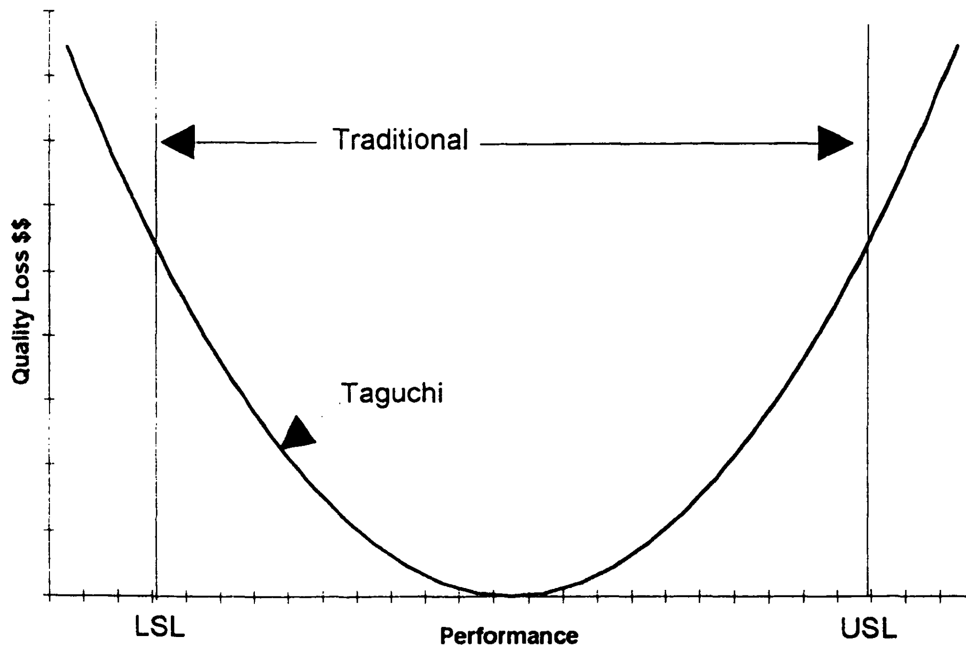


Fig. 3.1: Taguchi loss function vs. traditional inspection limits

Chapter 4: Frameworks for Quality Improvement and Cycle Time Reduction

This chapter presents the author's creative synthesis of the ideas from his education, the references, experience, and colleagues at Motorola. The frameworks are intended to simplify communication about and analysis of the problems faced by manufacturing and design groups, both separately and jointly, when trying to make progress on quality and cycle time goals. The remainder of the thesis will be in the context of one or more of these frameworks.

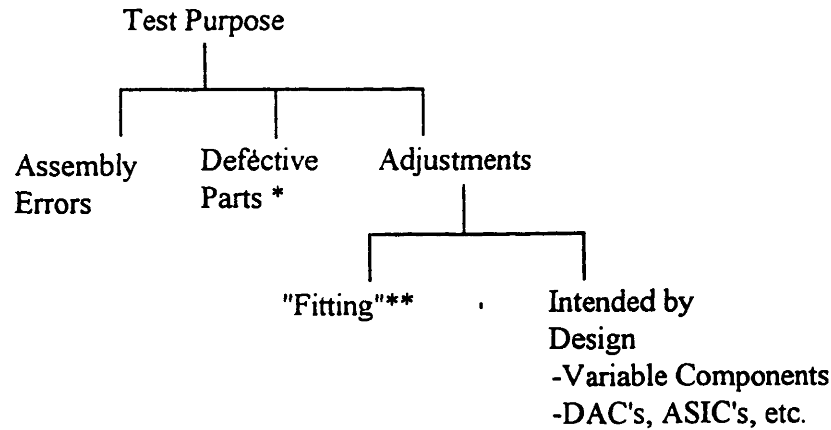
The purposes of testing in defect-driven quality control

The first framework relates to the semantics of quality, to the meaning of the word "defect." The importance of shared understanding of what words mean can't be overstated--words are abstractions that can cause conflict between groups if they are misunderstood. It became evident that defects and quality improvement efforts were a major source of tension between design and manufacturing groups, partly because of their different perspectives. When defects occur, it can be difficult to classify them as to whether a special problem exists, and where the corrective action should be focused (See fig. 2.2). It seemed that "defect" was too abstract a word. It might mean workmanship, mistakes by production workers, problems with purchased parts, or interactions and design problems. A good starting point for clarifying the semantics is the purpose of testing. The stated purposes were to detect defects and to perform adjustments. A finer-grain set of purposes distinguishes between assembly errors (which are under the control of the factory) and defective parts (which are not under the control of the factory). Under adjustments, a distinction is made between those which were intended by designers, and those which the designers were trying to avoid. We might use the word "fitting"[1] for the latter category, as in making the parts fit together to make a functional whole. Activities like changing parts which all meet their specifications until the circuit passes test are examples of fitting. Generally speaking, when fitting is necessary, it is because of a non-robust design.

In the author's experience, most conflict between design and manufacturing occurs over fitting-type activities. Generically, design tends to accuse production of cheating, not finding the real problems, or to blame part suppliers, while production tends to blame design or suppliers. Time pressure and organizational barriers tend to make such blame-

[1]Credit for this term goes to Daniel Whitney, my thesis advisor. The story goes that Henry Ford once said that "Fitters have no place in mass production." In the early days of automobile manufacturing, fitters were the people who, armed with files, made the part that didn't fit the first time fit together.

fixing more likely. Further thought on the matter led to a probability model for a single stage of a production process.



* A defective part is one that doesn't meet its specifications

** "Fitting" is when parts are changed and/or matched to acheive functionality, unintended by design, or adjustments are used for purposes other than what they were intended for.

Fig. 4.1: The purposes of test

A probability model for assembly and test

If we view the assembly and test of a single unit in the factory as an experiment with a random outcome, we can construct a sample space to describe the possible outcomes (Fig. 4.2). To read this diagram, start from the top, with the assembly of a unit, a receiver, for instance. If there are any assembly errors, take the branch on the left (A), and if not, go to A'. From each of those nodes, take the branch to the left if there are any defective parts, to the right if there are none. D indicates that defective parts are present, D' that there are none. From there, the production test can either pass or fail, denoted by P on the right and F on the left in each case. Finally, the customer can either be satisfied or not, denoted by a Y or N on the right and left, respectively. In this simple representation, there are 16 unique outcomes of the assembly/test experiment. Of course, some are more likely and more desirable than others.

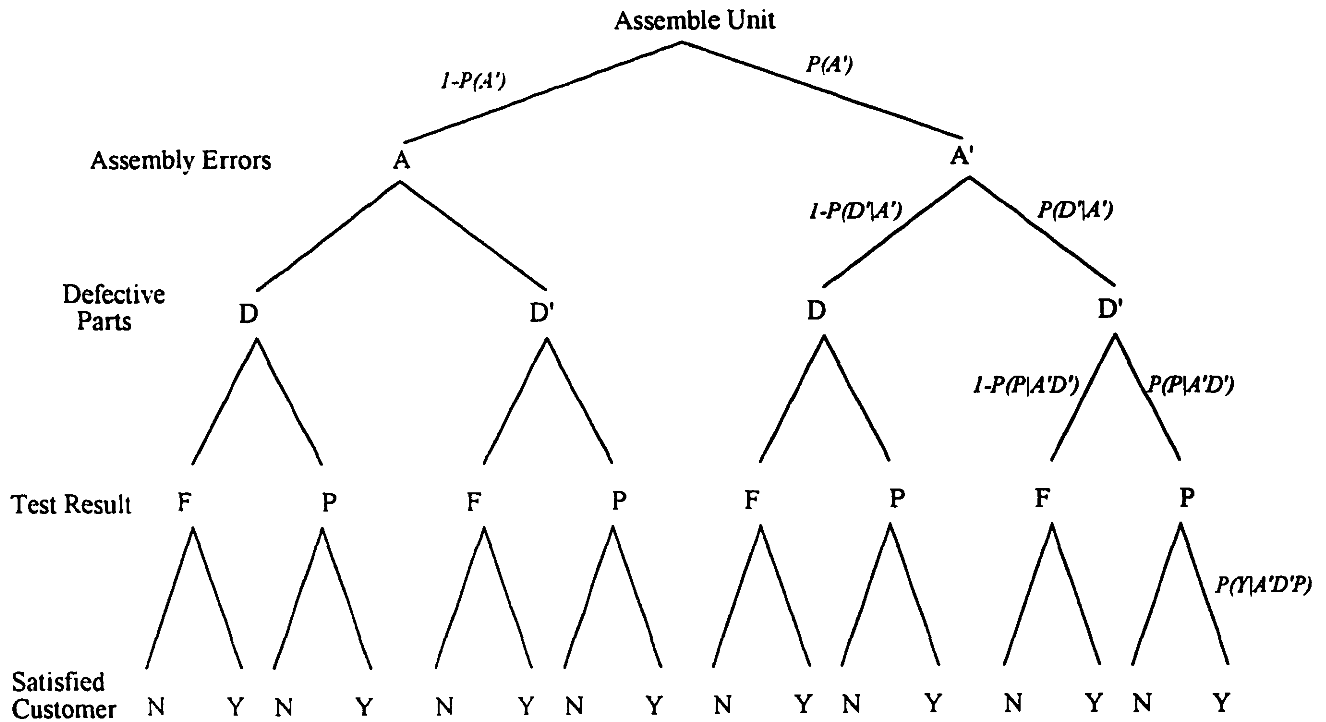


Fig 4.2: Assembly and test sample space

The most desirable outcome is the rightmost path through the tree, with no assembly errors, no defective parts, a passing test and a satisfied customer, denoted by A'D'PY. In the case where the test has an error and the customer would have been satisfied with no defects, this event is described by the path A'D'FY. This is the case where we cause ourselves extra work because of a faulty test. A less plausible outcome is when there are assembly errors and defective parts, but the test passes and the customer is satisfied (ADPY). This could happen if the wrong capacitors are loaded in a power-supply isolation filter, for instance. In this case the effects on performance may be subtle or nonexistent in the absence of an interfering signal being conducted through power supplies. A more likely scenario is ADPN, where a defective product passes the test and the customer is dissatisfied, perhaps the least desirable outcome. The rest of the outcomes will be left to the reader to think about. Note that this model doesn't include repair and re-testing. A similar tree might be used with the starting point being "repair unit" instead of "assemble unit" as the starting point for analyzing a repair process.

This probability model can have several uses. First, it can serve to reduce ambiguity in discussions about testing, failures, defects, etc. More powerfully, quantified probabilities could be assigned to each branch and the effects of proposed changes evaluated. A decision analysis approach could be taken, where a dollar value is assigned to each outcome and the expected cost of different alternatives compared based on their impacts on the probabilities. In the extreme, mathematical programming techniques could be used if the probabilities and outcomes can be written as functions of a set of decision variables. The decisions might include what specification to set for a functional test, whether or not to test for a given defect or functional specification, etc.

An example of the power of this framework would be to assess the impact of using QFD [7,10] to develop the test procedure. Presumably, QFD would ensure that the test is very closely related to customer needs, and make the conditional probability of a customer being satisfied if the test passes very high. I.e., using QFD raises $P(Y|P)$ and lowers $P(N|P)$ for all the lowest branch levels. Another example is the idea of eliminating tests. If two tests are redundant, eliminating one of them will lower cost without affecting the pass or fail probabilities. On the other hand, eliminating the only test for a certain kind of likely defect will raise $P(P|DA)$, for instance, the probability of passing a defective unit. This sample space can be used as an aid for the activity discussed in chapter 7, developing a test procedure.

A robust design framework linking manufacturing and design

This framework came from the advice of Deming, Taguchi, etc. to remove sources of variation to improve quality, coupled with some observations made in product development. The development process includes two or three iterations of prototypes of substantial quantity, perhaps 50 to 100 units. For each iteration a production grade test is run on all the units, an environmental test on a subset of the units, and an accelerated life test (ALT) on a few units. In the case of one project, several problems were detected during testing of the second round of prototypes that were not detected in the first round. Investigation failed to pinpoint the cause of the problem being due to any change made in the intervening design iteration. Rather, the problems had to do with different batches of parts, parts from different vendors or dates, and different combinations of parts. Furthermore, production personnel expected that new radio systems going into production would have some design problems appear over time and volume, which were not detected by the prototype tests. This prompted the author to try to capture the reasons for this in a robust design framework. Figure 4.3 diagrams the sources of variation experienced by production, the customer (actual performance), and design. It highlights the fact that

design doesn't see normal manufacturing assembly and test errors, at least not after the prototypes have been debugged. Also, design doesn't see the full extent of component variation, due to small sample size and, more importantly, because the component values for a prototype build tend to be tightly grouped. Supplier processes tend to be batch related or auto correlated over narrow time intervals, and prototype runs tend to be built over a short period of time. Finally, production doesn't see all the variation due to environment and aging, but the customer sees all the variation except for factory test errors.

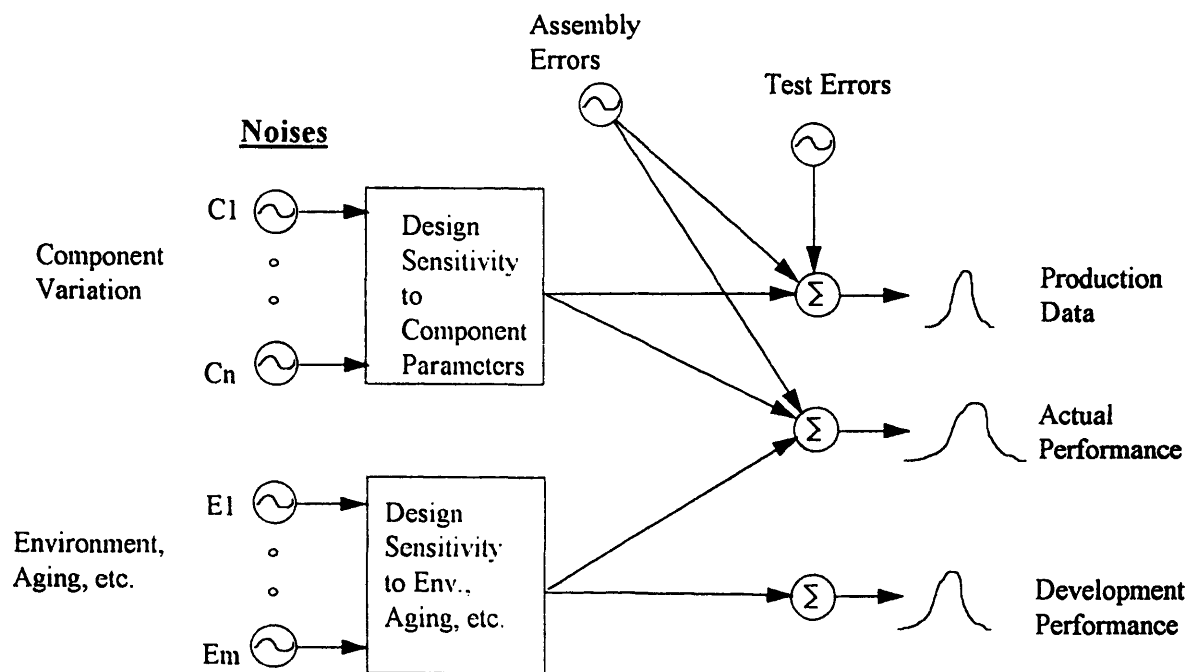


Fig. 4.3: Sources of variation

While it is a coarse approximation to say that design sees no component or assembly variations, figure 4.3 is intended to highlight the fact that development sees much less of these variations than customers and manufacturing.

The sources of variation are labeled noises in the spirit of Taguchi, and the robust design task is seen to be that of identifying the sources and reducing the sensitivities of the design to their variation. We see that if we only check prototypes for conformance to specs, we are counting on luck that component variation will be small enough not to cause production quality problems.

One very powerful way to avoid this problem is through analytical modeling and simulation. Motorola uses computer simulation in the early design phase to improve the robustness of its designs. The capability of analog and radio-frequency simulation tools is a major focus of improvement efforts.

As defect rates improve it becomes less and less effective to screen prototypes for future production problems. To demonstrate this, let us assume that a given design will have a defect with a certain probability, p. We can look at the assembly and inspection of a prototype of that design as a Bernoulli-type trial. The binomial distribution will then predict the likelihood of detecting a failure as a function of p and the quantity of the prototype run. If the defect occurs 1 or more times in the prototype sample, the problem is detected, while if the defect doesn't occur it is not detected. Therefore the probability of detection is [1-binomial(p,N,0)]. Figure 4.4 shows the probability of detecting design defects vs. prototype sample size and the design defect rate. We see that as quality improves we have diminishing likelihood of detecting significant production problems with pass|fail prototype testing.

All of this assumes that the sample of prototypes is drawn from actual production parts. As was stated before, the actual variance among prototypes is bound to be less than production variance because the parts are drawn from a time-localized sample. Therefore, the above probabilities are a best-case result assuming no significant reduction in part parameter variance.

A way to use continuous valued data to get around this problem is to plot histograms of performance and calculate projected yield and process capability indices, assuming a normal distribution. This was the practice for the development project observed at Motorola for the final pre-production iteration of prototypes. In chapter 6, an expanded method of prototype data analysis is presented that separates x- and y- variables, and helps in identifying the noise factors that drive variation.

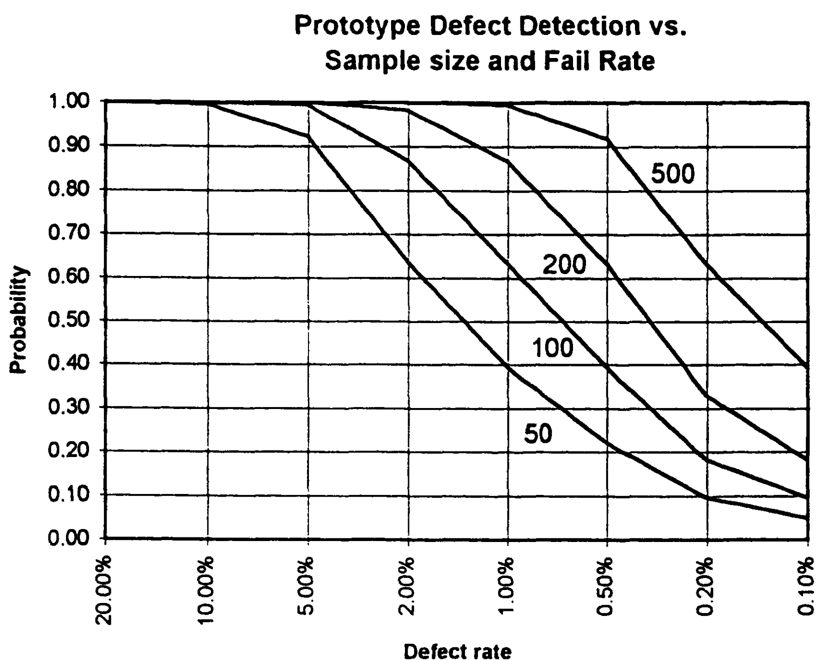


Fig. 4.4: Defect detection probability

Design decisions' impact on cost, quality, and capacity

This framework takes a simple system dynamics view of the immediate and delayed effects that can occur as a result of design decisions, in terms of the need for inspection and capacity constraints. This analysis comes directly from the investigation conducted in chapter 2. The factory was faced with rapidly increasing demand for its products, and was encountering bottlenecks in both the test and repair operation (system test) and with supply of some purchased parts. A redesigned radio with much lower parts' cost had recently been put into production. For the most part, the factory was succeeding in meeting the increased demand, but the challenges it encountered were using much of its engineering resources. Figure 4.5 is a causal loop diagram that shows three branches of cause and effect that influence cost, driven by design decisions on technology, architecture, parts cost, etc. The inspiration for this diagram came from interviews with engineers in quality, manufacturing, and design during the investigation phase.

If a design project has cost reduction as one of its primary goals, the designer seeks to simplify and integrate the architecture, reduce parts count, reduce the number of subassemblies, reduce the parts cost through choice of generic parts, as well as reducing the need for testing, calibration, etc. Of these cost components, the easiest to measure is

the direct material and associated direct labor for assembly cost. To understand the diagram, start from the variable labeled "Parts Cost Design Choice." To the right, the direct influence of lower parts cost on total cost is shown, with the direction of influence being "s" for same. "O" indicates that cause and effect are in the opposite direction in this diagram.

The other branches in the diagram map out how parts cost decisions can influence total cost in the opposite direction, in less obvious ways. There is separation in time and space between cause and effect in several of these branches, making it difficult to observe and directly quantify the effects.

Without technology breakthroughs, a lower parts' cost leads to at least some parts' technology performance limits being approached by the requirements of the design. Lower cost technology tends to have lower limits, barring breakthroughs. As the relative capability of the technology is lowered, process capability and the ability of suppliers to meet the standard drop. An example of this was a filter in the radio system that was state of the art, low cost, single-sourced, and was often a "problem" in production due to interactions between the filter and other parts. Tighter control of filter parameters would ease the problem, but the supplier was unwilling to provide such a tightly controlled part. Furthermore, the supplier had difficulty keeping up with increasing demand because Motorola was a large share of its volume for the part. This system of cause and effect is mapped through with supplier difficulty leading to fewer suppliers willing to provide the part and a resulting constraint on volume through parts' availability. If the volume constraint from parts' availability is binding, then there is an opportunity cost that results from the original design decision to push the part technology to its limits.

Likewise, the center branch documents higher costs that result from lower process capability, stemming from additional repair and inspection costs and ultimately a higher cost of quality. Not shown on the diagram is another cost of the need for inspection and repair--in an increasing volume situation it can become the constraint to throughput. Because skilled and trained workers are needed to perform this activity, it can be difficult to expand capacity of a test and repair operation. Yield improvement and repair productivity improvement are the best ways to ease a test and repair bottleneck.

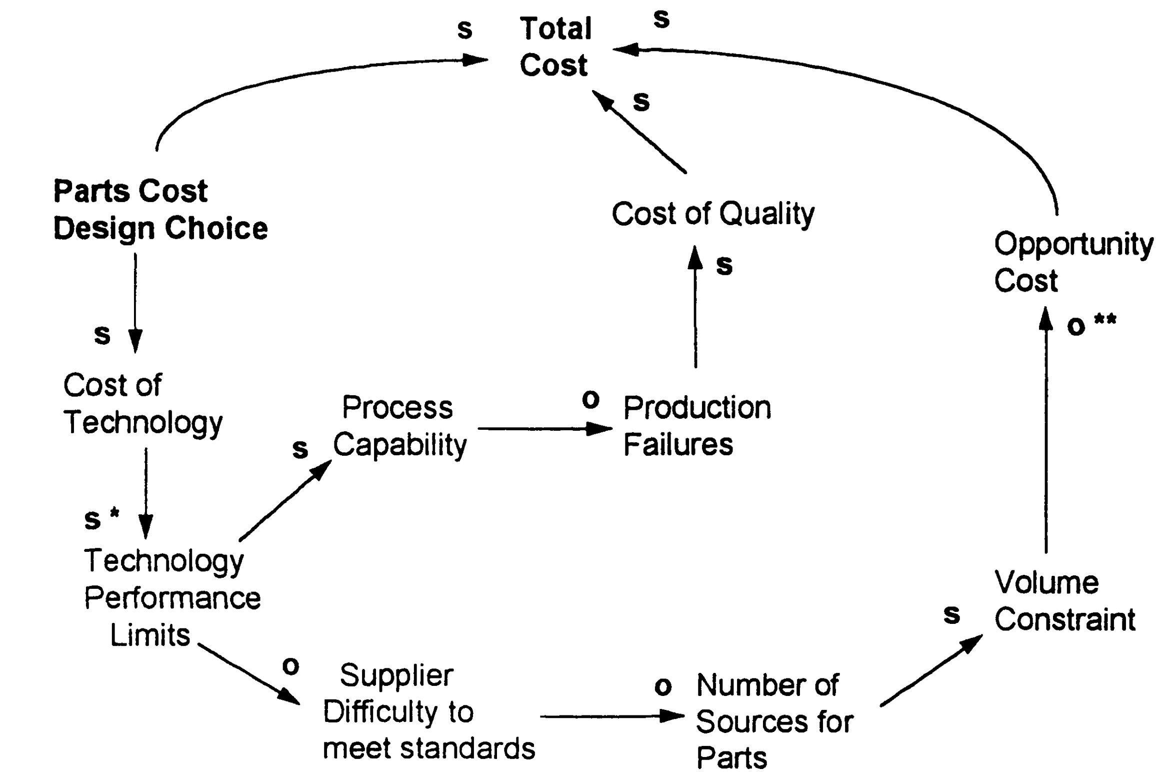


Fig. 4.5: Cost reduction and balancing loops

Admittedly, direct material is a very important component of cost in an electronics assembly operation, composing 60-80% of all costs, depending on the product. The tradeoffs involved form a general problem faced by any design team. The designer must find the optimum point, where parts' cost is minimized while avoiding problems with supply and production yields. Reducing parts cost implies that the organization should recognize the possible side effects and act to counteract them. For instance, in the sole-sourced part situation, a large supplier that can easily increase supply for a given part might be given preference over a smaller supplier, even if the price quoted is slightly higher. Another approach might be to develop a second source for the part or to qualify a second part to the design, although this approach will add variation to the process and thus add cost. Alternatively, emphasis on planning for the contingency of higher volume could be placed on the supplier, making them aware of the need for increasing capacity in as proactive a manner as possible. The latter approach was being used to mitigate the problem at Motorola, working with the supplier to plan for growth.

Mitigating the effects of lower process capability is much more difficult. Engineers in design are left to make the tradeoff between robustness and parts cost using their judgment. Perhaps a costing method that included the need for test and repair would help them to make this tradeoff. Here is where use of prototype test data is critical. How can the need for test and repair, factory yields, etc. be predicted as a function of design decisions? Effective use of modeling and simulation can help the designer quantify the tradeoffs between part cost and the need for testing and process capability. Another step is using histograms and process capability estimates from the prototype runs. Care should be taken to look at every prototype run and to understand any difference in yield among them. Is the difference due to a design change, or to a different batch of parts? Sensitivity to environment can also be a warning indicator of a potential yield problem. Is the parameter that changes with temperature also likely to vary over time? These issues will be examined more closely in chapter 6, on the use of data for robust design.

In this chapter the central issue of testing and data has been examined from the perspectives of production and design. Because the test data is the main set of facts available about the product and process, it serves as a powerful tool for communication between production and development. The frameworks presented here will be used in subsequent chapters to connect the projects to the higher level goals of the organization and to give contextual clarity to their purpose.

Chapter 5: A Manufacturing Improvement Project

Introduction

This chapter picks up where chapter 2 left off. The factory was actively executing defect-driven quality improvement, resulting in dramatically reduced failure rates for a product that had started shipments about 4 months earlier. At the same time this quality improvement was occurring, volume was increasing tremendously as the market accepted the new product, with its lower cost and new features. Several design changes had been made to reduce the failure rates, as well as some inspection of selected incoming parts. Now that the most obvious problems had been solved, it was becoming more difficult to make further improvements. It seemed to be an opportunity to demonstrate the use of parametric data. A conceptual diagram of an improved quality improvement process is shown in fig. 5.1 (compare to fig. 2.2). Here, we implement Deming's advice to look behind the production failures at the process and its sources of variation.

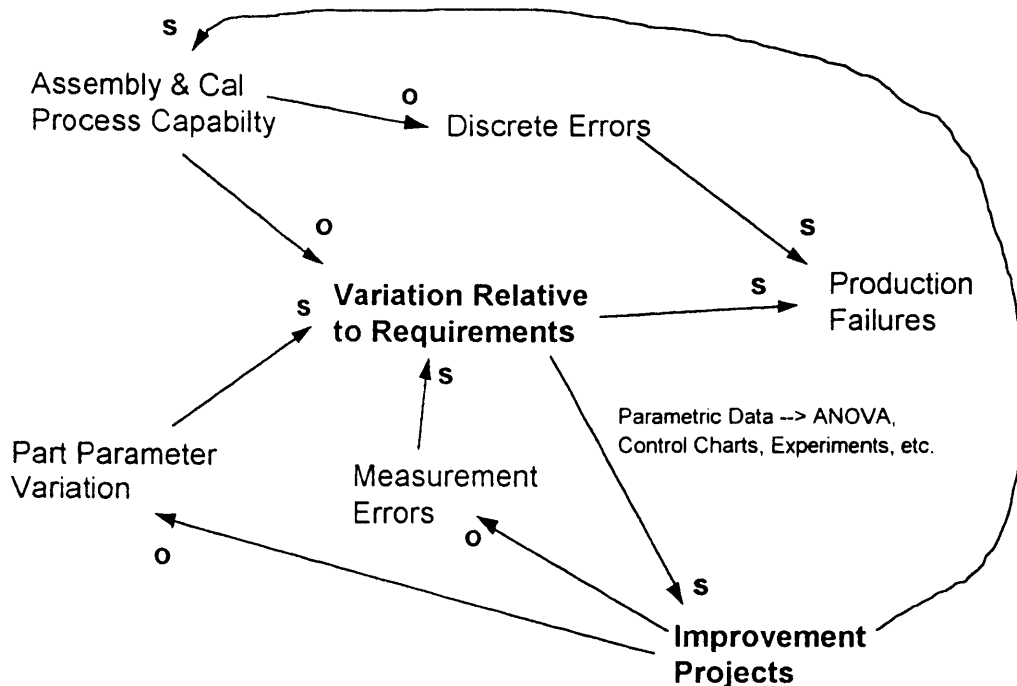


Fig. 5.1: Quality improvement focused on manufacturing variation reduction

Figure 5.1 shows the quality improvement projects being driven by the data behind production failures, rather than the failures themselves. All of the test data is used, and variation sorted out among assembly and calibration, measurement errors, and part

parameters, either directly or indirectly. This approach allows improvement to proceed without large numbers of test failures occurring.

Note that in all of the following data exhibits, the data is disguised. Every attempt has been made to preserve the essential relationships, however.

Problem Definition

The final system test operation had become a constraint on the throughput of the factory. Among other efforts to increase capacity, it was important to reduce the number of failures and rework done at system test. This was particularly critical because diagnosis

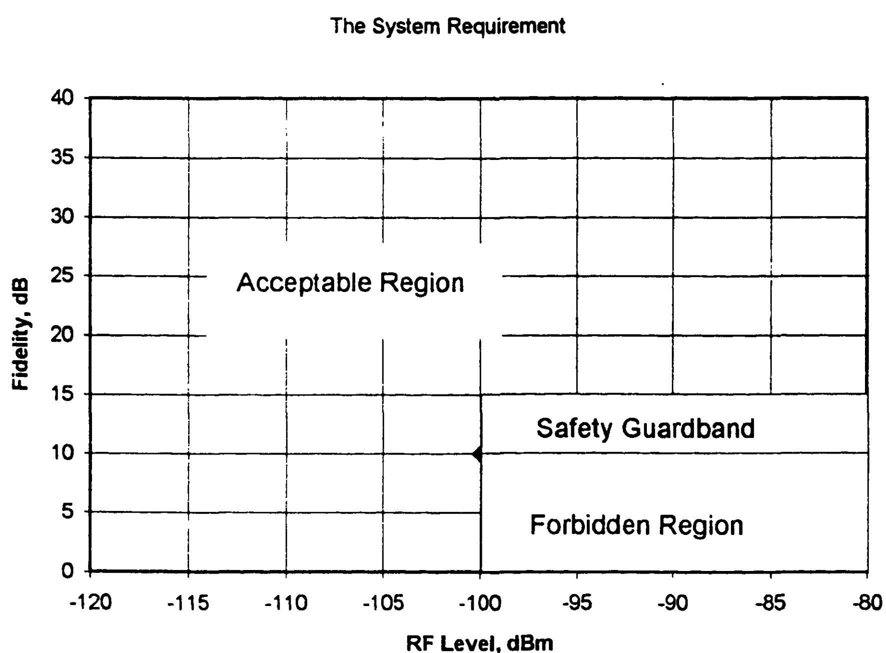


Fig. 5.2: The system requirement

and repair were done on line in this area, with time spent in these activities subtracting from capacity. The starting point of the project was to look at a Pareto chart of the failures from this area on this product line. The number one test failure accounted for nearly half of all failures. The failing test was a surrogate for audio fidelity for the receiver, specified at a given received signal strength. This is one of the customer critical parameters for any receiver. Fig. 5.2 shows the requirement. At signal strengths above -100 dBm[2], the audio fidelity must be better that 10 dB. Notably, the test limit used in

[2]dBm is a unit representing decibels above 1 mW RF power.

production was a more stringent 15 dB, at -100 dBm signal strength. This safety guard band had been added based on a study by the quality department, showing that fidelity was sensitive to ambient temperature. The amount of the guard band (5 dB) was the subject of considerable controversy. The purpose of a guard band is to make the conditional probability of the receiver remaining in spec. over the full temperature range much higher in the event that a just-barely passing result occurs at room temperature.

Some possible issues to be examined included whether the acceptance criterion was too stringent, whether the test systems were accurate, and what other sources of variation led to legitimate problems. Notably, the corrective action that was taken to repair most failing systems was to replace the receiver module. A final question was whether this was the most appropriate corrective action. Refer to fig. 2.1 for a diagram of the process flow.

Available Data and Experiments

After the problem was defined, data was collected to investigate the problem. Production data and data from two experiments were used. The experiments included characterizing a few radios (measuring the quieting curve) and a repeatability and reproducibility experiment.

Measuring the Quieting Curve

To gain further insight into the relationship between signal power and fidelity, 10 radios were measured in a stimulus-response manner. The resulting curve is known to radio engineers as a quieting curve, referring to the progressive reduction of audio noise as the RF signal strength is raised. The stimulus was an RF input stepped from -114 dBm to -74 dBm in 0.5 dB steps, while the response was the audio fidelity at each input power level. To aid in understanding the measurement and the radio, a block diagram is presented in figure 5.3.

The signal source is a piece of test equipment that outputs any desired signal power, with the RF carrier modulated with a standard amount of FM deviation. Inside the radio system, the signal is passed through a low noise pre-amplifier before it is filtered and passed through circuitry that distributes the signal. The pre-amp serves to improve the signal/noise ratio of the system, because the distribution network is lossy and would degrade S/N without the pre-amp. Inside the receiver module, there is a tuner consisting of an oscillator and mixer to convert the RF frequency to a lower frequency called an IF (intermediate frequency), prior to filtering and demodulation. This receiver block diagram represents a generic heterodyne receiver. The output of the demodulator is an audio

signal that has its fidelity measured by an audio analyzer. The resulting characteristic is presented in fig. 5.4.

Notice that at 10 dB fidelity there is 4-4.5 dB of margin to the customer requirement of -100 dBm sensitivity, while at -100 dBm input power there is 11-13 dB of margin to the customer requirement of 10 dB fidelity. However, there is only 6-8 dB of margin to the guard banded specification of 15 dB fidelity. The bell curves overlaid on the data are to indicate the relative variation in the population on where the quieting curve crosses the -100 dBm level and where it crosses the 10 dB fidelity level. These variations are gleaned from the factory data.

Another interesting observation is that the slope of the quieting curve is about twice as steep at -100 dBm input compared to the 10 dB fidelity point. This means that any variation in signal power, pre-amp gain, or loss of the distribution network will affect the fidelity twice as much at the -100 dBm level as at the 10 dB fidelity point (-104.5 dBm for this radio). An explanation for the shape of the quieting curve goes as follows. At the low-power end, noise is the dominant factor in poor audio fidelity. The FM demodulator sees a random signal at its input. As the input signal is raised, the demodulator starts to pick up the desired sinusoidal signal at its input. As the power continues to increase, the input of the demodulator starts to look more like a square wave, and the noise has less and less impact on the timing of when the waveform crosses zero. FM demodulators essentially transform the timing of zero crossings into instantaneous frequency, so influence of noise on the timing of zero crossings is critical. This phenomenon is obviously highly nonlinear, based on the quieting curve. Finally, the limit on audio fidelity is set by distortion on the high power end, as noise becomes insignificant. Note the variation among the receivers in the distortion limit, but the consistency at the low power end as they all see the same noise power. In the steep range of the curve there is also considerable variation among the receivers compared to the 10 dB fidelity point. The data from the factory and another experiment will support this observation.

Factory Data

Fortunately, the production fidelity test data were stored in a relational database for all of the systems tested. In addition, the test data for sensitivity of the receiver modules were also stored in another table on the database. The sensitivity test for which data is stored on the receiver checks the same specification as the fidelity test in the system test area. However, rather than applying -100 dBm and checking for the required fidelity, the

receiver test adjusts the signal level until a fidelity of 10 dB is reached. The module specification on sensitivity is -105 dBm, a level intended to guarantee that the

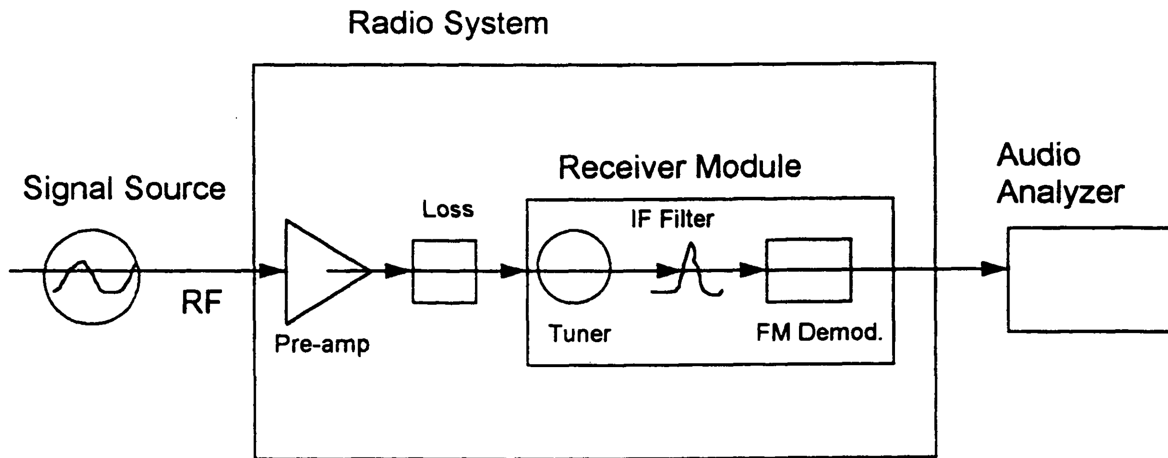


Fig. 5.3: Measurement and radio block diagram

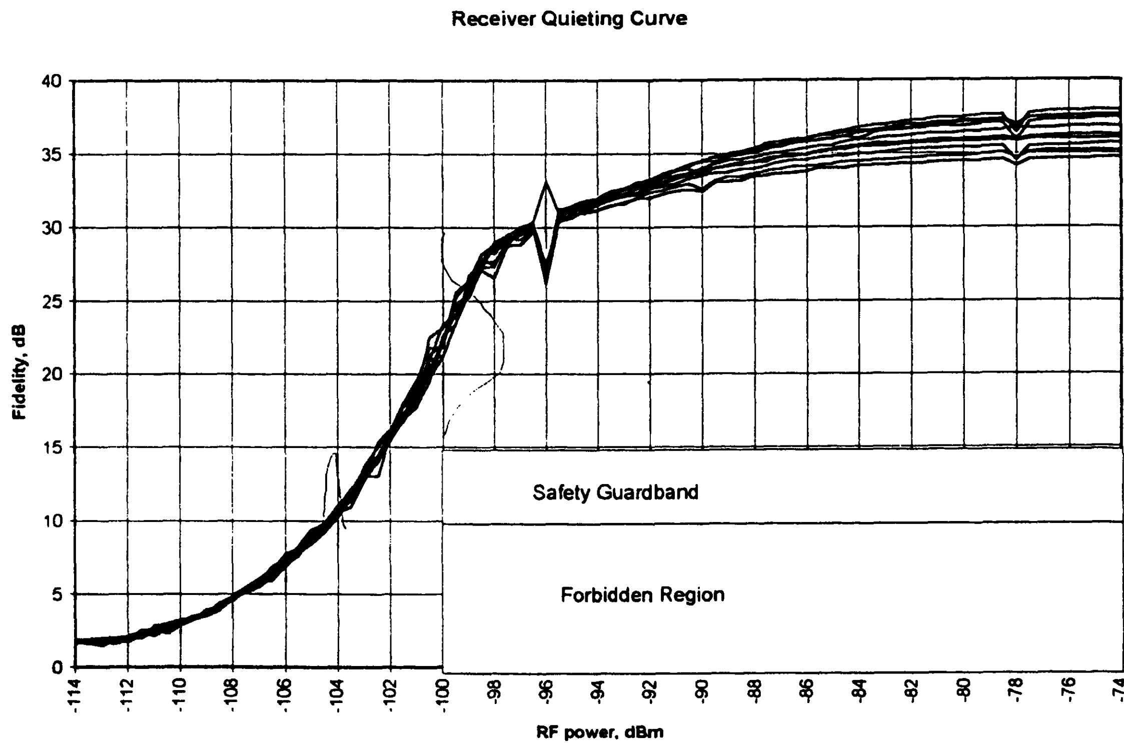


Fig. 5.4: Quieting Curve

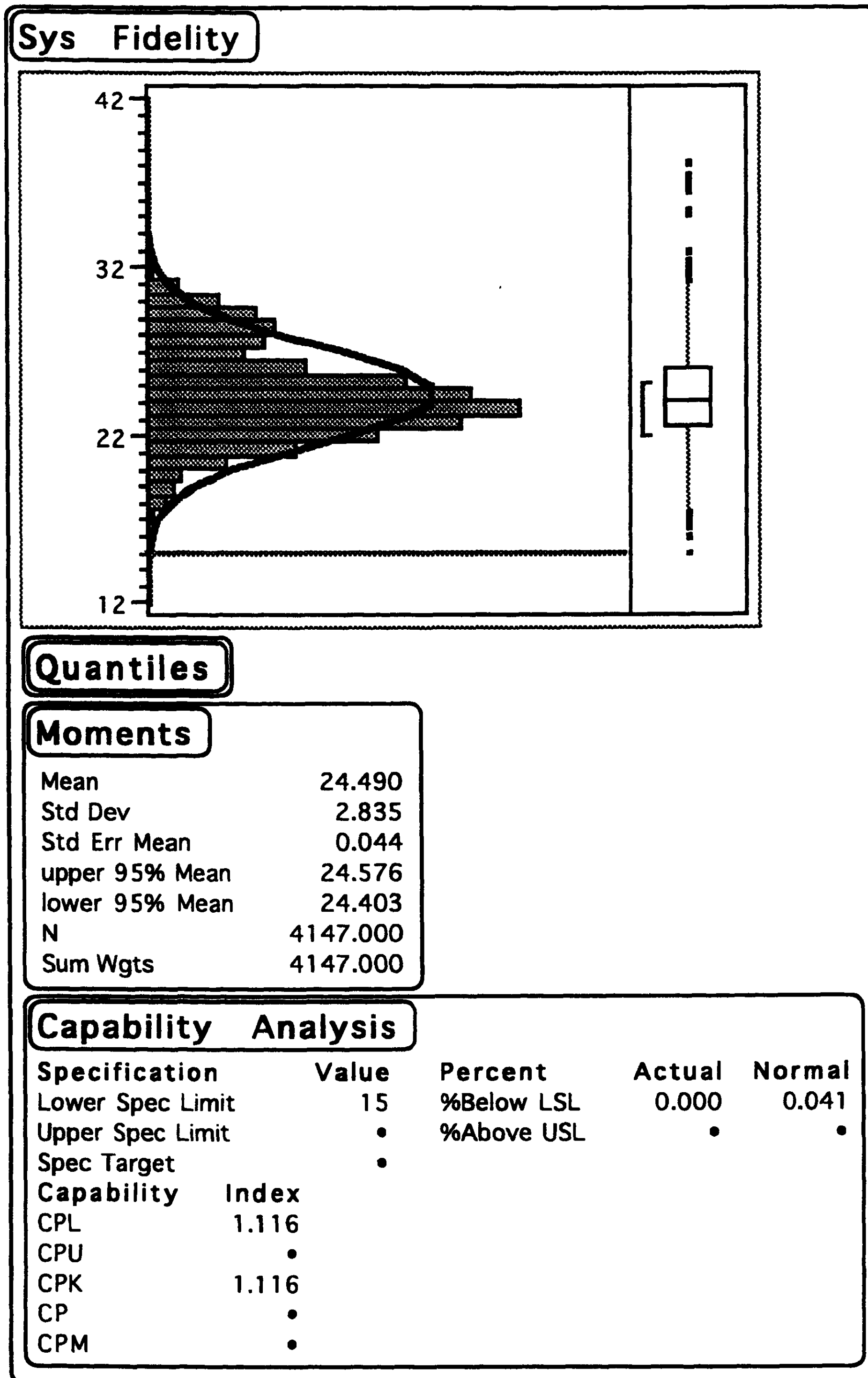
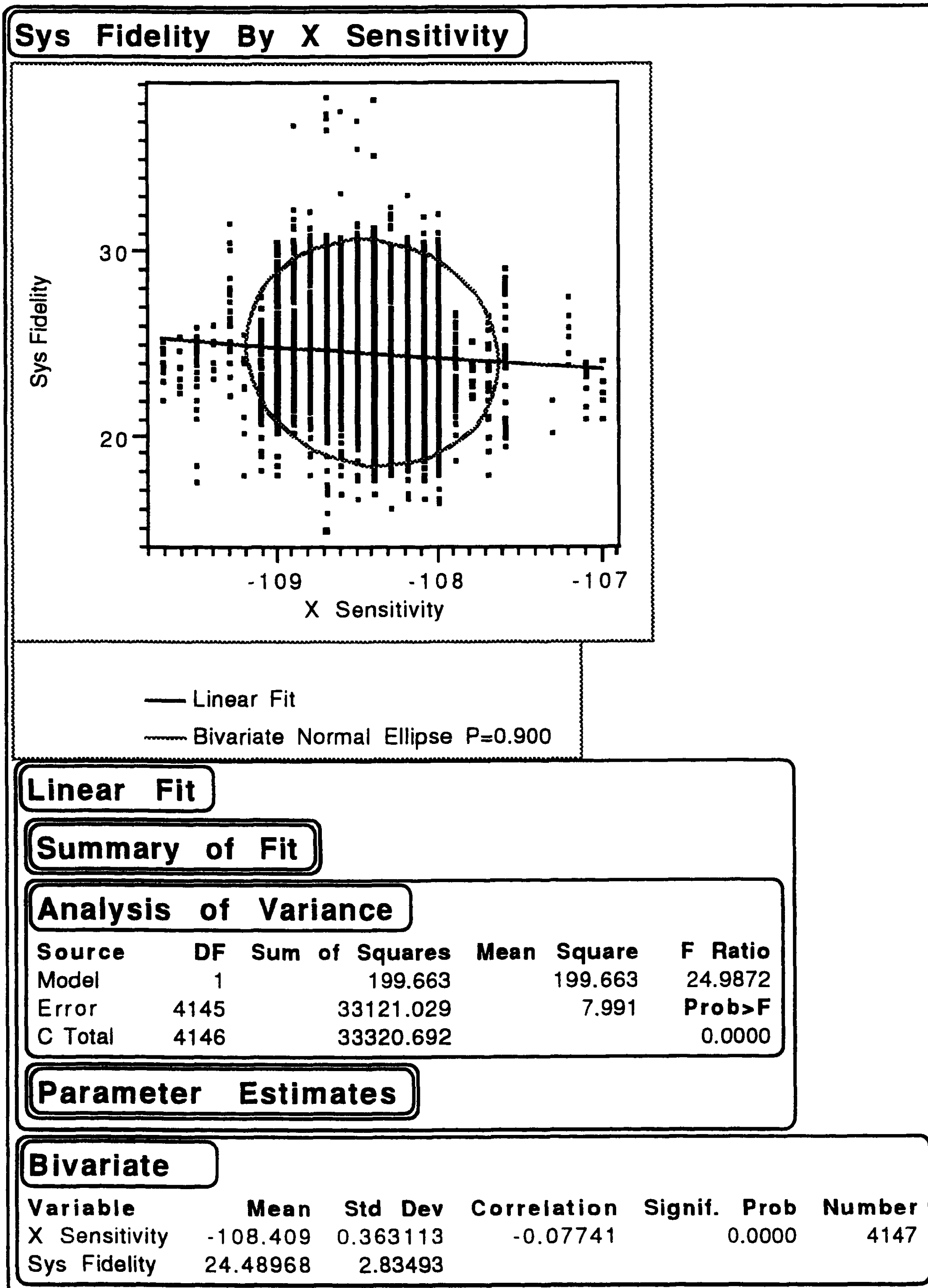


Fig. 5.5: Distribution of system audio fidelity



Linear Fit

Summary of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	199.663	199.663	24.9872
Error	4145	33121.029	7.991	**Prob>F**
C Total	4146	33320.692		0.0000

Parameter Estimates

Bivariate

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
X Sensitivity	-108.409	0.363113	-0.07741	0.0000	4147
Sys Fidelity	24.48968	2.83493			

Fig. 5.6: Relating the system and receiver data

module will pass when tested in the system. The module quieting curve is different from the system quieting curve.

Combining the quieting curve and the production data helps the investigator understand the nature of the specification. The data is linked to the physics of the radio. In summary, we see that the quieting curve is highly nonlinear, with variation among the population increasing dramatically as RF input power is raised beyond the 10 dB fidelity point (See figures 5.4-5.6).

It was possible to link the two tables (system and module test) by the module serial number using a simple SQL (pronounced "Sequel") query to the database server. SQL is an ANSI standard language for extracting data from relational database management systems (RDMS), such as those marketed by Oracle corporation.

In addition, informal discussions among engineers revealed that there was frequently no trouble found when the receivers that had been removed to repair systems were analyzed by technicians in the receiver repair area. Because the focus of this project was to find sources of variation, actual repair statistics were not collected. Rather, the focus was on the measurements and the resulting data collected at the receiver and system test areas on the entire production population.

A statistical analysis software package called JMP (marketed by SAS institute) was used to analyze the data from the query. Some manipulation of the file with a spreadsheet was necessary in order to separate the data into columns prior to reading it into JMP. JMP runs only on Apple MacIntosh computers. Analyses of the distributions of system and receiver performance, of the relationship between receiver and system performance, and of the sources of variation of the system test data were performed.

First, the distributions of fidelity at the system level and of sensitivity at the receiver level were plotted and summarized. Note that the process capability (Cpk[3]) is about 1.1 at the system level, which is well below the corporate goal of 1.5. However, this capability is with respect to the 15 dB limit. Against a 10 dB fidelity limit, the Cpk is a much more respectable 1.8. See fig. 5.5. The receiver-level sensitivity data has a Cpk of 1.75.

Note that JMP displays both a histogram and a quartile box plot to depict the distribution. Summary statistics and capability analysis are also shown. JMP has many

[3]Process capability indices are discussed in the statistical references. However, they all assume normal distributions. This assumption should always be checked when interpreting capability indices.

options on how to display and analyze the distribution of a variable, including tests for normality.

Conclusions to be drawn from the distribution analysis are: 1) There will be some failures of the 15 dB limit that fall out of the normal population of radio systems, due to the low Cpk. These failed systems will *not* be due to any special cause or defective part. In other words, radios built correctly with parts from the normal supply distribution will still fail occasionally against a 15 dB spec. If a slightly lower limit were used, this kind of failure would be very much less likely. See the conclusion to this chapter for more discussion on the specification limit. 2) A receiver without a defect will almost never fail the receiver sensitivity test.

Next, the relationship between receiver sensitivity and system fidelity was analyzed. If the receiver is largely responsible for system performance, we should be able to predict system performance from receiver measurements. This assumption seems to underlie the tendency of diagnostic technicians to replace the receiver when a system fails. Figure 5.6 show a scatter plot, where the x-axis is the measured receiver sensitivity and the y-axis is the fidelity measured on that receiver in a system at system test. Note that the data appears widely scattered, with only a very loose relationship. Fitting a line to the data shows that there is indeed a significant relationship, but that less than 1% of the variation in the system data is explained by the receiver data.

Note that the standard deviation of the system fidelity is 7.8 times as large as the standard deviation of receiver sensitivity. Remember that these are two different but related quantities. Two or three hypotheses come to mind to explain this scatter plot's loose relationship. First, with reference to the quieting curve, there may be variation in the relationship between sensitivity at the 10 dB fidelity point and the fidelity at the -100 dBm point. In other words, the different test methods explain the poor correlation. Second, the receiver sensitivity may not be the most important determinant of system performance. The pre-amp and distribution circuitry may be more important. Finally, there may be other sources of variation in the system test results that obscure the relationship.

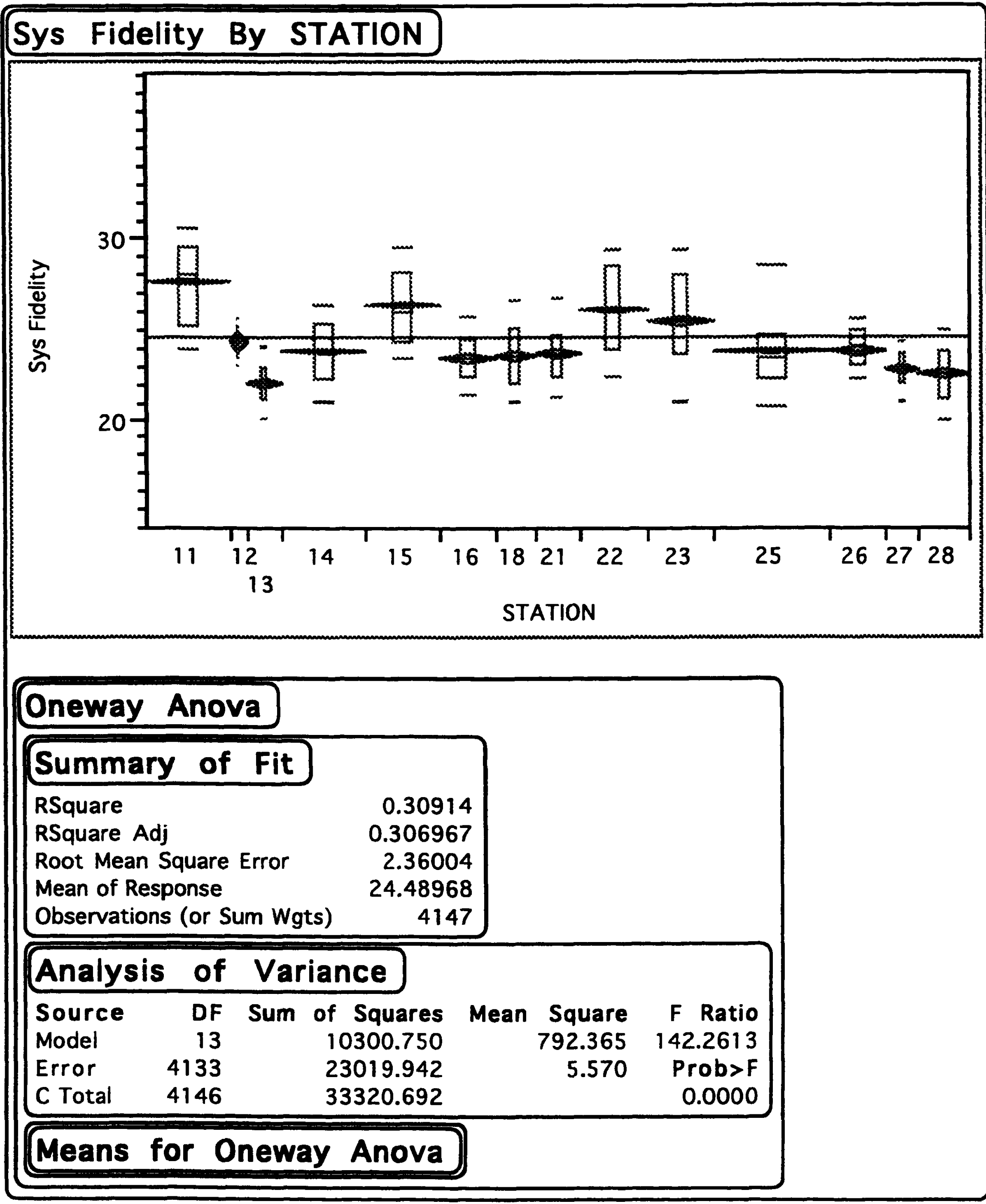
This brings us to the analysis of variance (ANOVA) of the system test data. ANOVA is a process by which the variance of a y-variable is assigned to one or more x-variables and functions of the x-variables. It is a model-building technique. In the case of the production data, only one-way ANOVA can be performed on the entire dataset, for reasons that will be explained shortly. Possible x-variables in the data include the system test station, the operator ID, the time of day, the date, the radio system serial number, and

the frequency tested. All of these variables came as columns from the SQL query. It was a simple matter to perform a 1-way ANOVA for each of these variables using JMP. Figures 5.7-5.8 show the ANOVA results with the test station and test date as explanatory variables. The plots indicate the quantile spread of the data, the mean(diamond) and its confidence interval, and the number of points in each category (represented by the width on the x-axis for that category.

On each of the ANOVA figures, the width of the box plot indicates how many points are in that category, the length of the box plot shows the spread of the data, and the separation of means' diamonds shows the magnitude and statistical significance of means' differences for each category. For instance, in figure 5.7, station 11 and station 14 have about the same number of data points, a difference of 4 dB in means which is very significant, and roughly comparable spread in the data. The F ratio in the Analysis of Variance box measures the significance of the overall model, with the probability of the model being insignificant listed under the "Prob > F" heading.

Each of these three factors appears to be very significant in terms of its contribution to overall variation. Of the factors examined, the date and test station have the largest effects. The channel, time of day, and operator ID are much smaller, although the large sample size makes them statistically significant. Unfortunately, it is impossible to say which of the factors is the most important based on this data, because of correlation issues. Correlated x-variables also make multi-variable ANOVA impossible for this data. For instance, some test stations were used for this product mainly on certain days, which makes it hard to tell whether the station or the day or both are really important. Also, each radio system is tested only on one day and only at one test station, which makes it hard to tell whether the radio system hardware is more or less important than the test station. These are classic problems with happenstance data (See Box, Hunter and Hunter for a good discussion of happenstance data).

Fortunately, this is a very large dataset and we have a powerful tool (JMP) to pursue these questions. It is possible to select the data for one test station only and do an ANOVA by date on that station, for example. Also, a subset of the data was found where several receivers were tested in multiple radio systems, at several test stations. At the time, there was a temporary shortage of receivers, so several radio systems were assembled and tested with the same receivers. We had a happenstance experiment that



Oneway Anova

Summary of Fit

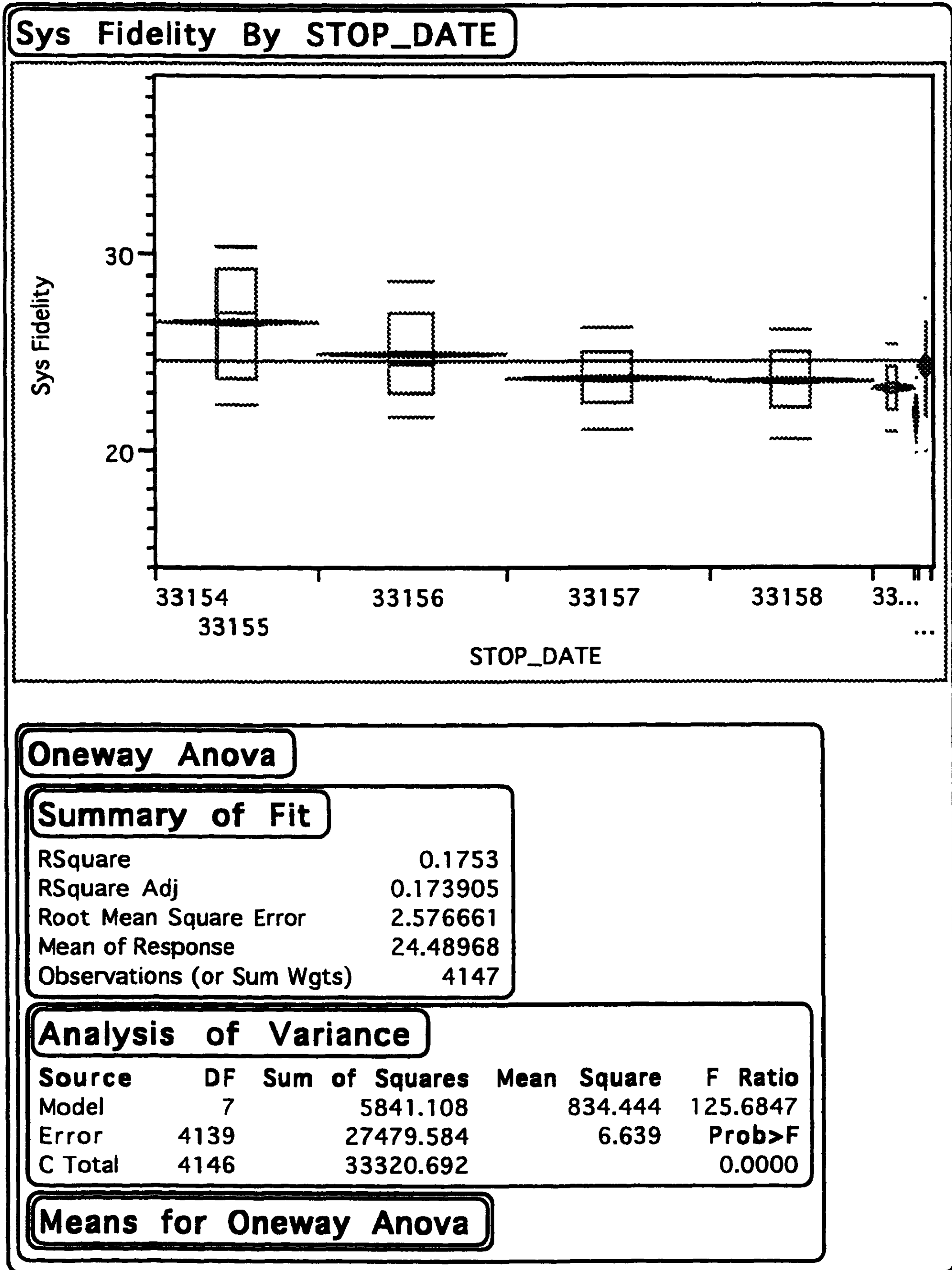
RSquare	0.30914
RSquare Adj	0.306967
Root Mean Square Error	2.36004
Mean of Response	24.48968
Observations (or Sum Wgts)	4147

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	10300.750	792.365	142.2613
Error	4133	23019.942	5.570	Prob>F
C Total	4146	33320.692		0.0000

Means for Oneway Anova

Fig. 5.7: ANOVA by test station



Oneway Anova

Summary of Fit

RSquare	0.1753
RSquare Adj	0.173905
Root Mean Square Error	2.576661
Mean of Response	24.48968
Observations (or Sum Wgts)	4147

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	5841.108	834.444	125.6847
Error	4139	27479.584	6.639	Prob>F
C Total	4146	33320.692		0.0000

Means for Oneway Anova

Fig. 5.8: ANOVA by test date

had occurred in the factory, which could be used to separate the effects of the receivers from the rest of the system.

In this happenstance experiment, there were four radio systems, six receivers, and two test stations involved. The layout of the experiment is shown in table 5.1. Each receiver was tested in all four systems. For each row in the table, there were six measurements of audio fidelity made (6 different test conditions). To analyze the results of the experiment, it was set up in JMP as a nested model. A nested model is an experiment within an experiment. The outside x-variables were the test station and the receiver serial number, while the system serial number was a nested x variable inside the test station (Each test station measured several radio systems, but the radio systems were only measured at one test station). The results were that the test station effect was the largest, at over 5 dB (Note that it is confounded with the systems, however)[4]. The next most important effects were the nested system serial numbers, with a difference of 3.5 dB between systems 280 and 302 at station 14, and a difference of 1.8 dB between systems 201 and 175 at station 11. The receiver effects were not significant at the 95% confidence level. These results seem to indicate that the system hardware is indeed much more important than the receiver in determining system performance.

In the final analysis, it is difficult to draw solid conclusions based on the factory data. This rich set of facts does provide ample opportunity to form hypotheses, however, and either act on them or conduct experiments to either prove or disprove them. One experiment was conducted by the quality department to measure the repeatability of the tests and the effect of different operators performing the tests. The results of this "R&R" study will be presented before the final conclusions and hypotheses of this chapter are drawn.

[4]Another major confounding factor is the existence of two varieties of systems. The two types of systems are significantly different in the configuration of preamplifiers and in audio fidelity performance. Unfortunately, the database had no field with which to make this distinction. It is possible that the systems tested at station 11 were of a different variety than the systems tested at station 14.

System Serial	XCVR Serial	Test Station
175	9X3	11
175	9X9	11
175	BT0	11
175	BT1	11
175	BTQ	11
175	CJ8	11
201	9X3	11
201	9X9	11
201	BT0	11
201	BT1	11
201	BTQ	11
201	CJ8	11
280	9X3	14
280	9X9	14
280	BT0	14
280	BT1	14
280	BTQ	14
280	CJ8	14
302	9X3	14
302	9X9	14
302	BT0	14
302	BT1	14
302	BTQ	14
302	CJ8	14

Table 5.1: Happenstance experiment layout

Repeatability and Reproducibility Experiment

To address the question of test system accuracy, a repeatability and reproducibility experiment was conducted by the quality department. In this study, 6 radio systems were tested with 10 different receivers, 3 times each by 3 different operators, on a single test system owned by the quality department. To help understand the relationship between fidelity and sensitivity, the experiment was done with the requirement being measured both ways. Finally, each measurement was done under 4 different operating conditions, resulting in a total of 4320 data points.

To analyze the results, the data was organized into 8 columns (4 conditions each for sensitivity and fidelity) and 540 rows. The data was fit to a model including the system

serial number, receiver serial number, and operator as x-variables. The residual variance after fitting this model would be the repeatability, while the effects of the operators would be the reproducibility. Note that because the receivers were always in the same system, the receiver is a nested variable inside the system serial number.

The results of this model fit showed that the operator was an insignificant effect. I.e., we can't tell whether there is an effect or not. This was expected, because the test is automated and the only effect an operator is likely to have is on the connection of the test equipment to the radio system. So, in this case, the reproducibility error due to the operator is zero. The mean square error for each of the 8 models is shown in table 5.2. Remember that this RMS error is only the repeatability of the test--it is the error that can't be explained by the significant factors of radio system and receiver. The key findings

Condition	Measurement	RMS Error, dB
1	Fidelity	.68
1	Sensitivity	.25
2	Fidelity	.76
2	Sensitivity	.21
3	Fidelity	.46
3	Sensitivity	.18
4	Fidelity	.47
4	Sensitivity	.18

Table 5.2: R&R experiment results

from the experiment are that the operator doesn't matter, and the random test error is 2-3 times as large when measuring fidelity as compared to sensitivity.

A follow-up experiment was planned to study the effects of different production test stations instead of different operators. Unfortunately, the need to use the test stations for production delayed the experiment too long to be included in this thesis.

Conclusions

First, the system specification is unnecessarily stringent. The requirement to measure 15 dB audio fidelity causes the design of this radio system to appear less capable than the

corporate objective (while still far exceeding the customer requirement). The 5 dB guard band was added because the temperature sensitivity of fidelity at the -100 dBm level varied among 50 systems by 2.5 dB standard deviation. However, we have seen that the variation in the population is highly dependent on the RF input power. If the parameters that vary among the population are also temperature sensitive, it may be a reasonable hypothesis that the temperature sensitivity at the 10 dB fidelity point is much less than that at the -100 dBm input level. Informal experiments confirm that sensitivity varies only about 0.3 dB over temperature at the 10 dB point, which is consistent with the forgoing hypothesis. From a physical point of view, at the 10 dB point noise dominates performance, while demodulator characteristics seem to dominate performance at higher fidelity levels. The factory quality and manufacturing engineers recognize this issue, and are conducting a study of sensitivity variation over temperature and test system R&R to change the specification. The proposed change is to lower the input power below -100 dBm by some amount, and verify >10 dB fidelity. It appears that only a 1 or 2 dB guard band will be necessary. This should dramatically improve the process capability. Another way of explaining this is: if a radio has unusually low fidelity at -100 dBm input, its temperature sensitivity will be much less than that of a typical radio with higher fidelity because we are operating at a lower point on the quieting curve. The criterion for setting specifications from the Chevalier paper would be useful here. Remember that we want to set the specification to balance the costs associated with false failures and false acceptances. The formula for spec optimization is repeated below:

$$P(Bad|\, y = y_{spec}) = \frac{MC_{Stage_i}}{MC_{Stage_{i+1}}}$$

If we pick a number for the cost ratio of repairing a good radio to shipping a radio that goes out of spec at temperature, we can solve this problem. A conservative ratio might be 1/100. In this case, we would want to set the spec so that if we measure the spec limit in production, there is a 1% chance that it will drift out of spec over the full temperature range. If the standard deviation of sensitivity performance over temperature is 0.33 dB, a 1 dB guard band on sensitivity will give less than a 0.3% chance for conditional probability if the temperature delta is normally distributed.

Second, the performance of the system is dominated by the system pre-amp and distribution network in normal radio systems, rather than by the receiver. Engineering insight on noise figure and the importance of the front end circuitry supports this view, as well as the evidence from the factory data. The importance of test date in the SQL query

result may be due to different batches of pre-amps over time going into the radio systems. When a radio system showed poor performance, it was probably more effective to work on the rack componentry than to replace the receiver. However, it was much easier to replace a receiver than to replace the other hardware, and if the test was failing by only a small amount it was possible that the retest will pass due to test repeatability alone, or by the small improvement a better receiver might make. Recognizing this, changes were implemented where gain and noise figure are measured prior to the installation of receivers. This is a step in the direction of eliminating the need for full system inspection.

Third, there may be an opportunity to improve the accuracy of the test stations, and thus reduce the cost associated with needless rework or shipping marginal products due to test errors. The results of the factory R&R experiment will be a key indicator of how large this opportunity is. Remember that the literature on the cost of test [17] identifies test errors as a far more important variable than cost of the equipment, for instance.

Fourth, the data showed that different operating modes produced insignificant differences in the results. As long as at least one test checks the performance of the routing circuitry, etc., it is unnecessary to re-verify performance in all modes. More will be said about this in chapter 7. So, there is an opportunity to save test time. This is important, because measuring audio fidelity takes much less time than measuring sensitivity, and test time is expensive in terms of capacity. Saving time is critical if we want to consider changing test methods.

Finally, the difference in test method between the receiver and system test area adds complexity to the data analysis and communication problem between the two areas. This is a complex specification, and measuring it two different ways makes understanding the issues difficult. The operational layout of the factory compounds the problem. Because the areas are in separate buildings, with separate management, and an inventory buffer between the areas, time delays and communication need to be managed. Careful inventory management, access to the databases, and efforts to foster communication can mitigate the problems. But in this case, simply getting the two areas to measure the requirement the same way would improve communication and reduce variation. Generalizing, consistent test methodology among different test areas (either inside the factory or between the factory and field) helps reduce the chance for test disagreements to cause problems. However, the agreed-upon method should be carefully chosen for its accuracy and relevance to the customers' needs.

Chapter 6: Using Test Data for Robust Design

In this chapter, some methods for analyzing data from prototype testing will be presented. Most robust design methodology focuses on experimentation to determine the optimal values for critical design parameters. This chapter focuses on concise ways of organizing and analyzing the data, to help discover critical parameters and learn from the prototype run. It documents ways to conduct the analysis phase of the prototyping cycle.

The setting for this work was a development project for a new radio system, which was at the stage of analyzing the data from its second and final prototype cycle prior to production. The project was primarily a mechanical repackaging of a radio system that had been in production for 4 months, analyzed in chapter 5, with some differences in the transmitter system hardware. As such, it was more on the evolutionary rather than revolutionary end of the technology development continuum. The standard method of analyzing data was to first check to ensure that all tests had passed relative to the requirements, including tests over temperature and in an accelerated life test. Beyond that, histograms of each performance parameter were to be plotted and the process capability estimated (Cpk). Refer to the Mullenix paper [25] for an excellent discussion of capability indices. Parameters with Cpk less than 1.5 were to be examined closely to see whether production problems were likely and whether corrective or preventive action was necessary. Such actions could address the product or process design, parts specifications, etc.

The prototype data available consisted of a production test on all the units, which was about 120 data points, and a more extensive test over temperature for a smaller set of units. Raw data was available from both of these tests. In addition, data from an accelerated life test was taken, but was not included in this example of analysis. In raw form the data was labeled with a test name describing the parameter and all the test conditions, such as "Mode Z sensitivity 500 MHz offset high", with a separate data file for each unit and each temperature. In a sense, there were two columns of data--one describing the test and the other giving the result. The serial number of the product and the environmental test condition were contained in a header for each file.

The standard task was to organize the factory data so that a histogram and Cpk could be calculated for each measurement, and the temperature data screened for any failing data points. This was a tedious task in itself.

As an alternative to the one-line-per-test structure, the author proposed that the data be placed in a database, with the structure separated into y-variables and x-variables. The

existing test labels would be parsed to determine the value of all the x-variables associated which each measurement. X-variables included the serial number, temperature, frequency, mode, carrier offset, audio frequency, etc. Of all the measurements on each receiver, only 21 different y-variables were identified. Several Visual BASIC modules were written by the author to convert the raw data into 2 different formats. The first format was a database with maximum x-variables and minimum y-variables. The second format had only one row per receiver, calling each test result a separate y-variable. Therefore, the number of y-variables was maximized.

The two formats were useful for different things. The maximum-x format was good for building models and understanding the sources of variation. The dependence on temperature, frequency, and mode as well as the consistency of this dependence was easy to analyze with JMP when the data was in this form. This is a particularly useful form of analysis for small sample sizes tested over temperature and in the ALT. In a sense, the results are analyzed like an experiment.

The maximum-y format is good for analyzing the production data from a larger sample of prototypes. Here, traditional distribution analysis is easy with JMP (histograms, Cpk, etc.). Beyond distributional analysis, relationships among different measurements can be explored. Correlation analysis can detect which tests are related to each other--chapter 7 will explore applications of correlation analysis. Finally, outlier analysis can detect the unusual radios among the population. Careful study of the correlations and outliers can lead to great learning as engineers speculate on why the tests are related or what is causing the outliers. The physical prototypes can be studied to find the cause of both unusually good and unusually poor radios. This will lead to discoveries, presumably including critical design parameters.

Fitting models to environmental data

As an example of model fitting using the maximum-x format, data on selectivity from the environmental test will be examined. Selectivity is one of the key performance parameters for a receiver. It measures the receiver's ability to detect a weak signal in the desired channel in the presence of a strong signal in an adjacent or alternate channel. Four receivers were measured from -10 degrees to 60 degrees C, at several frequencies, in several modes, and with the interfering signal on both sides of the desired channel (high and low). In addition to temperature, high humidity was used at 35 and 50 C. The environmental conditions were modeled as nominal x-variables, meaning that separate

Response: Selectivity

Summary of Fit

RSquare	0.822743
RSquare Adj	0.814434
Root Mean Square Error	2.79947
Mean of Response	31.42399
Observations (or Sum Wgts)	336

Lack of Fit

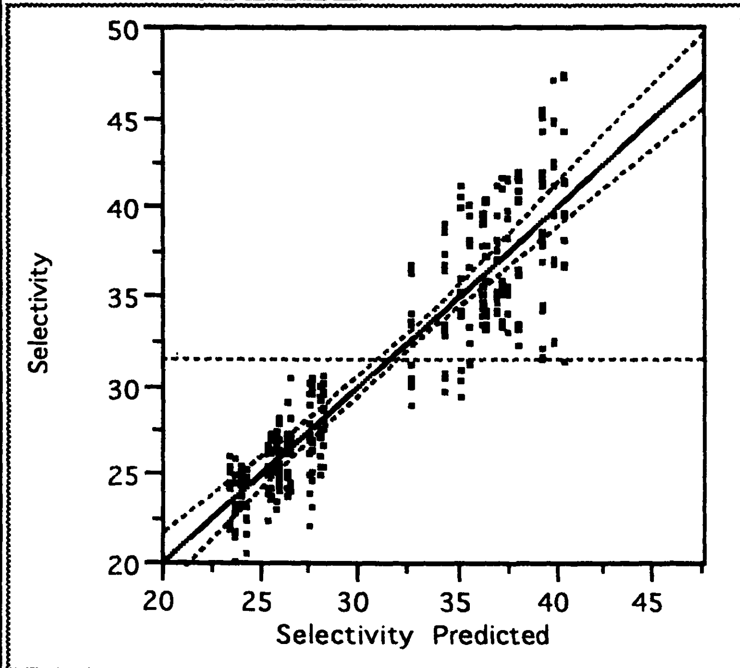
Parameter Estimates

Effect Test

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Environment	6	6	554.010	11.7819	0.0000
Antenna	1	1	156.657	19.9893	0.0000
Carrier Offset	1	1	10485.497	1337.943	0.0000
Environm*Carrier	6	6	234.774	4.9928	0.0001
Antenna*Carrier	1	1	209.318	26.7088	0.0000

Whole-Model Test

Environment



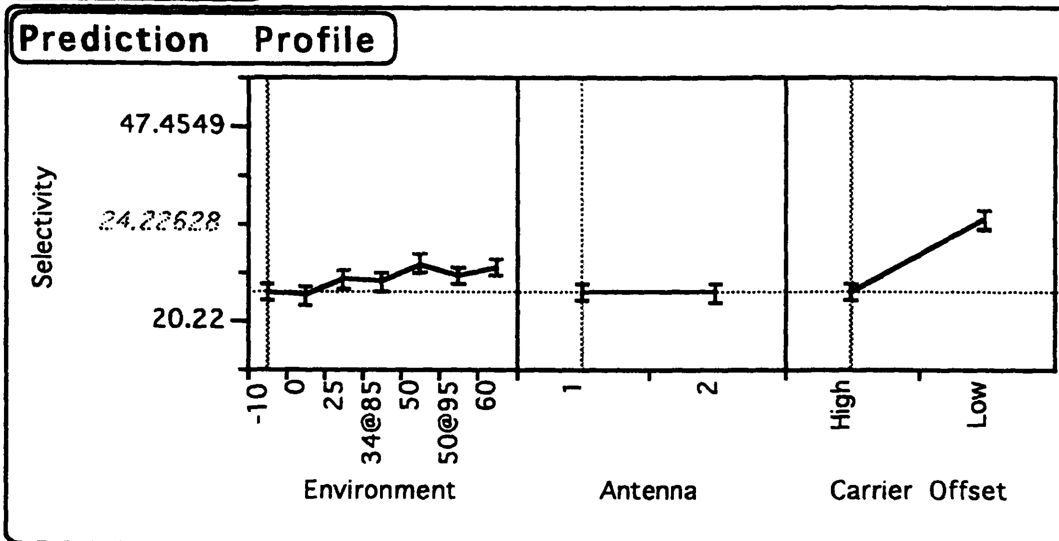
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	11640.256	776.017	99.0193
Error	320	2507.850	7.837	Prob>F
C Total	335	14148.105		0.0000

Fig. 6.1: Selectivity model fit

Screening Fit

Selectivity



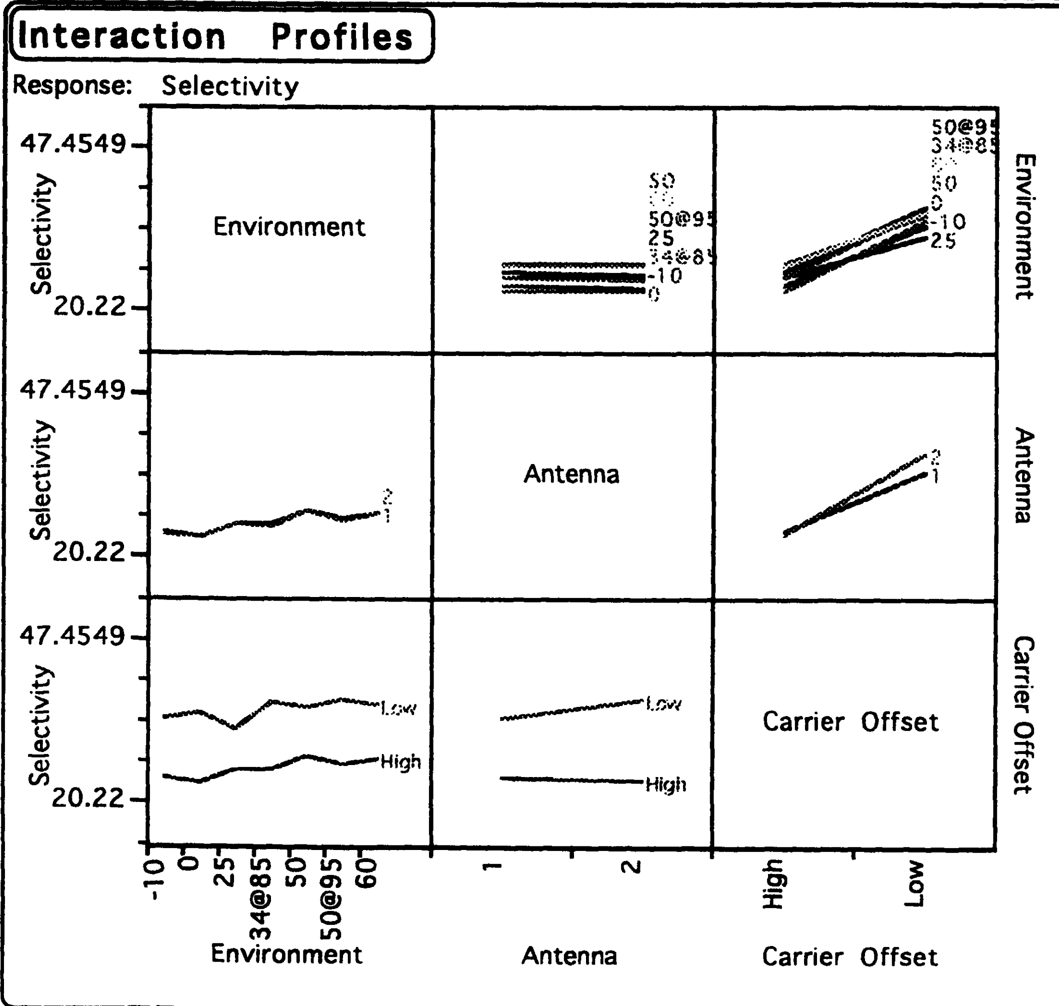


Fig. 6.2: Selectivity model profiles

effects were calculated for each environmental condition rather than a regression or curve fit to temperature.

The results showed that the environment was indeed significant, as well as the antenna input ,the carrier offset, and several interactions (See fig. 6.1). An interesting discovery was that the performance was much better on the low side of the carrier than on the high side, although adequate in both cases. Carrier offset was the most significant factor in the model. After discussion with several engineers, it was determined that this was because the IF filter in the receiver has an asymmetrical pass band, rejecting low frequencies more strongly than high frequencies. Here was an example of learning from the data, at least for the author. The whole model test plot in fig, 6.1 is a plot of the predicted vs. actual value for every data point. Many plots that are hidden in fig. 6.1 are available to examine the fit for every effect in the model, the values of the effect estimates, etc. The "Effect Test" box in figure 6.1 lists the statistical significance of each effect in the model. The model-building process involves finding x-variables and products of x-variables to put in the model such that the number of significant effects is maximized for the best possible fit. At the same time, x-variables that aren't significant should not be included in the model. For instance, the antenna-environment interaction was found to be insignificant and not included in the model. Figure 6.2 shows the details of the model, showing in particular how large the effect of carrier offset is compared to the other effects. The interaction profiles show different combinations of the variables, such as the environmental effect for each of the high and low sides of the carrier.

A final word on how model building can be used to aid in robust design concerns learning by the design team. If the results are compared to the expectations of the designer, learning will occur whenever there is a gap between results and expectations. Furthermore, when several parameters have similar dependencies on x-variables, we may be able to find an underlying critical parameter that is responsible for both. Having the facts so easily explored by the designer enables hypotheses to be generated and checked quickly.

A point of emphasis is that this kind of analysis and learning is more useful the earlier it is done in a project, while the learning can still be applied to design improvements in the time allowed. The first prototype test results will give the most value from this kind of analysis. In any case, the learning by the designers will serve to increase their skill and capability to design robust products in the future.

Analyzing factory prototype test data

As mentioned previously, the data needs to be arranged with a separate column for every test result to facilitate this analysis (maximum y-variables). Histograms and process capability are then quickly and easily calculated with JMP. Assumptions about normality behind Cpk can be checked. Going beyond distributional analysis, the relationships among any number of tests can be explored for learning purposes using the correlations of y or fit y by x analysis platforms in JMP. The use of correlation analysis to identify redundant tests is the subject of the next chapter. Finally, outlier analysis is a powerful way to discover unusual data points in many dimensions.

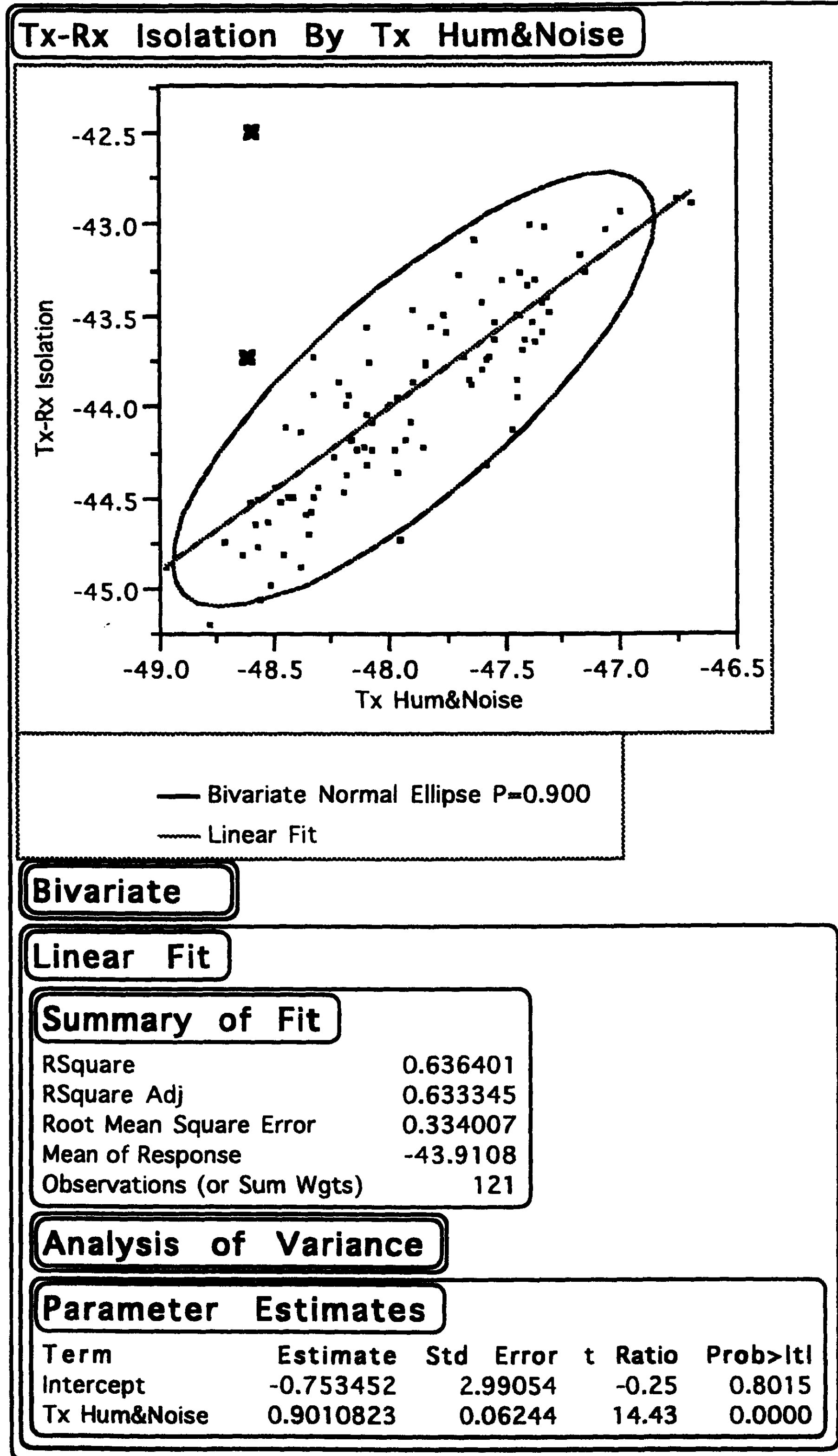
One dimensional outliers are easy to spot with a box plot or histogram of a single y-variable. An outlier in 2 dimensions can be spotted on a scatter plot, where there is a pattern in the data that one or more points don't fit with. Figure 6.3 shows a scatter plot where there is a relationship between two parameters with a few exceptions. Such outliers can appear to be "normal" when viewed with one-dimensional histograms. JMP has a feature called a "spinning plot" which is a 3-d scatter plot, allowing 3-d outliers to be spotted. Beyond 3 dimensions, JMP uses tools called principle components analysis and a general n-dimensional outlier distance calculation (Mahalanobis distance). The statistical theories behind these tools are beyond the scope of this thesis, but the point is that powerful tools exist to help discover the unusual data points in n-dimensional space. Figure 6.4 shows outlier analysis of five related parameters, where the Mahalanobis outliers are marked on the 2-d scatter plots.

A final comment on outlier analysis: in the history of science, many of the most significant discoveries have been made as the result of the observance of an unusual event. The power of outlier analysis is to help the design engineer to observe those unusual events.

It is the author's hope that the tools and methods presented in this chapter will make it easy for design engineers to learn from their prototype test data, leading to discoveries of critical parameters and other knowledge on how to design a more robust product. Table 6.1 shows a summary of the uses for analysis tools for prototype data.

Data Source	Questions	Tools
Prototype Factory Data	Yield Projection	Max y-variables, dist'n of Y
	Relationships, Test needs	Max y-variables, correlation of y, fit y by x
	Unusual units, problems & Discoveries	Max y-variables, outlier analysis
Environmental tests, ALT, characterization tests	Effects of Temp., aging, design alternatives	Max x-variables, fit y by x, fit models

Table 6.1



Summary of Fit

RSquare	0.636401
RSquare Adj	0.633345
Root Mean Square Error	0.334007
Mean of Response	-43.9108
Observations (or Sum Wgts)	121

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.753452	2.99054	-0.25	0.8015
Tx Hum&Noise	0.9010823	0.06244	14.43	0.0000

Fig. 6.3: Example of scatter plot with outliers

Scatterplot Matrix

Test 1

Test 2

Test 3

Test 4

Test 5

-4.90 -4.95 -5.00 -5.05 -5.10

8.2 8.0 7.8 7.6

-47.0 -48.0 -49.0 -50.0

-52.0 -53.0 -54.0

-52.0 -53.0 -54.0

-5.10 -5.00 7.6 7.8 8.0 8.2 -50.0 -48.5 -47.0 -54.0 -52.5 -54.0 -52.5

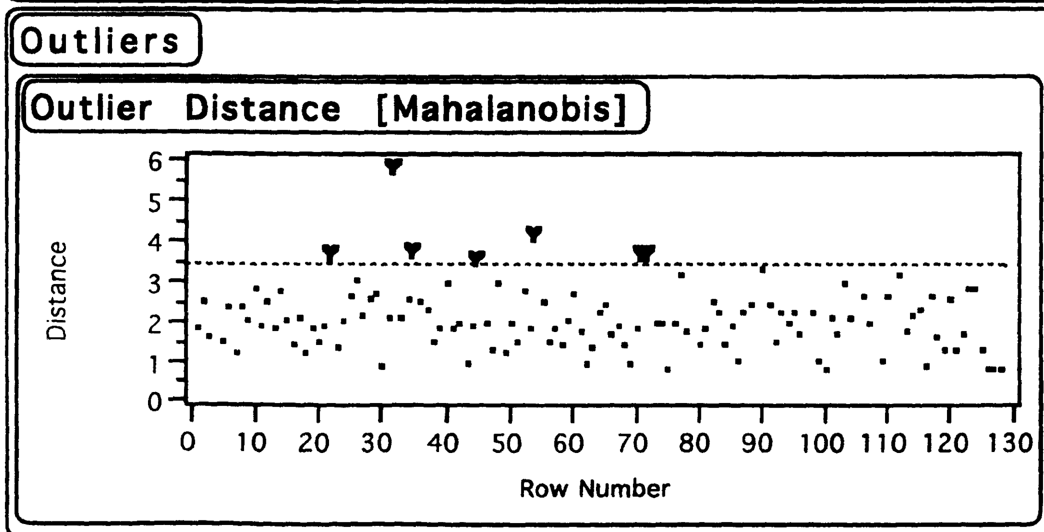


Fig. 6.4: Mahalanobis distance outlier plot, 5 dimensions

Chapter 7: Developing a Production Test Procedure Using Test Data

A final application for test data analysis is presented in this chapter. One of the key outputs of the product and process design is the procedure by which the product will be assembled and tested. The set of decisions to be made include what tests should be performed at each stage of assembly and the test limits that apply at each stage. Furthermore, the very structure of the assembly/test process needs to be decided. How much assembly is done before a test is performed? How many stages should there be? The optimization models discussed in the literature search (Ch. 3) are attempts to solve this set of problems, and always take test and other data as inputs. For reasons of resource limitations and feasibility in the context of this factory, it was decided to pursue a more heuristic approach for this thesis. See Ch. 3 for a discussion of why this decision was made.

The approach presented here is merely a proposal. There wasn't sufficient time or opportunity to try it out completely with the development project at the company. However, the data analysis part of the method is presented with real (disguised) data. The tests considered are limited to functional tests at each level of integration. The approach has two main aspects, engineering for test coverage and data driven verification.

Engineering method for test coverage

First, the relationship between circuitry and functional test is mapped out, to understand which tests verify the correct assembly and operation of each circuit. This activity should be performed early in the development project, during the design partitioning phase when the specifications are being written for each component and subsystem. This is really the issue of test coverage and the definition of the specifications. Questions of verification of correct assembly and meeting the functional requirement could be considered separately for a finer-grain understanding. At this time, design for testability can be considered as part of the design criteria. In a fashion analogous to QFD, we could work down from system-level requirements mapped against major subsystems, with a matrix for each subsystem with its requirements mapped to its components, and so on down to the discrete component level. A stylized example of a system matrix and the supporting matrix for the receiver are shown in figures 7.1 and 7.2. In each cell of the matrix, the degree to which the subsystem influences the performance is marked. In an assembly test matrix, the degree to which the test verifies correct assembly could be labeled. The goal of this activity is to give the design team an intuitive feel for test coverage and possible redundant tests. If the team is confident that the design is very

robust, then the production need only verify the correct assembly of the product. Multiple tests that cover the same assembly could be considered an opportunity to reduce testing. On the other hand, an assembly that is not verified by any test is a risk of shipping defects without knowing it.

The assembly and test probability model presented in chapter 4 can also serve to structure the decision process for testing a particular parameter at a particular stage.

	System Requirements						
Subsystem	Sensitivfity	Selectivity	Rx Audio Distortion	Tx Audio Distortion	Power Output	Spurious Output	System Errors
Receiver	Med	High	High				Med
Transmitter				High	Med	High	Med
Preselector	High	Med					
Power Dividers	Med						
References			Med	Med		Med	
Controller							Med
Power Amplifiers					High		
Power Combiners					High		
Power Supply			Med	Med	Med	Med	Med

Fig. 7.1: System-level functional test matrix

Receiver Requirements				
Subassembly	Sensitivfity	Selectivity	Rx Audio Distortion	Tx Audio Distortion
Control Board			High	High
RF Board	High	High	Med	Med
Shielding	High	Med		

Fig. 7.2: Subsystem functional test matrix

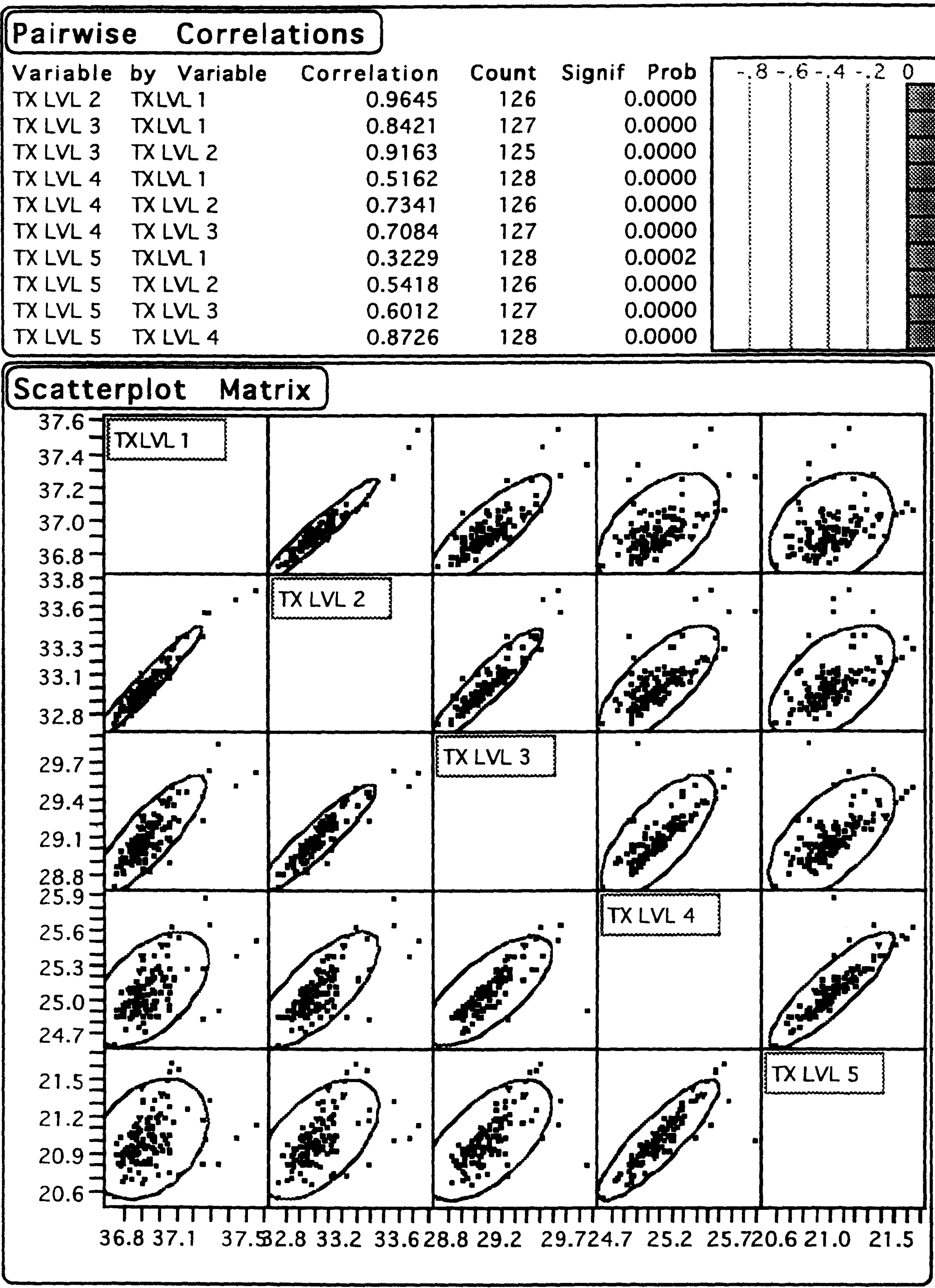
Is the expected cost of passing a defect to the next stage more or less than the cost of testing for that defect at the current stage? This expected cost is the product of its probability and the cost of detection and repair downstream. If the likelihood of the defect is very small (Either through high process capability or an effective upstream test, coupled with low likelihood of adding a problem at the present stage), an expensive test may not be the optimum decision. On the other hand, an inexpensive test may be justified even if the defect rate is very low, especially in the interests of monitoring the process.

Data-Driven Verification

To ensure that the engineers have covered all the contingencies, and to discover relationships they may not have thought of, the following procedure is recommended. When the first prototypes are tested, a very exhaustive list of tests and test conditions should be run to explore the operating space of the design. This can serve to check assumptions about dependencies on operating conditions, and give empirical data on how much test is needed to give high confidence that the product meets the standard.

For example, we would like the sensitivity of a receiver to be relatively independent of carrier frequency. This allows us to verify the functional requirement with a single measurement, rather than having to test at multiple frequencies. The first prototype could be tested at many frequencies and the correlation among the measurements confirmed with the data. Statements could be made like: "If the sensitivity at midband measures -105 dBm, then there is a 95% probability that the sensitivity across the entire band is better than -104.5 dBm." Note that this also assumes that the subject of this statement is similar to the source of the original data, that is was correctly assembled from parts that came from the same distributions, etc. Checking this assumption is the purpose for full test coverage.

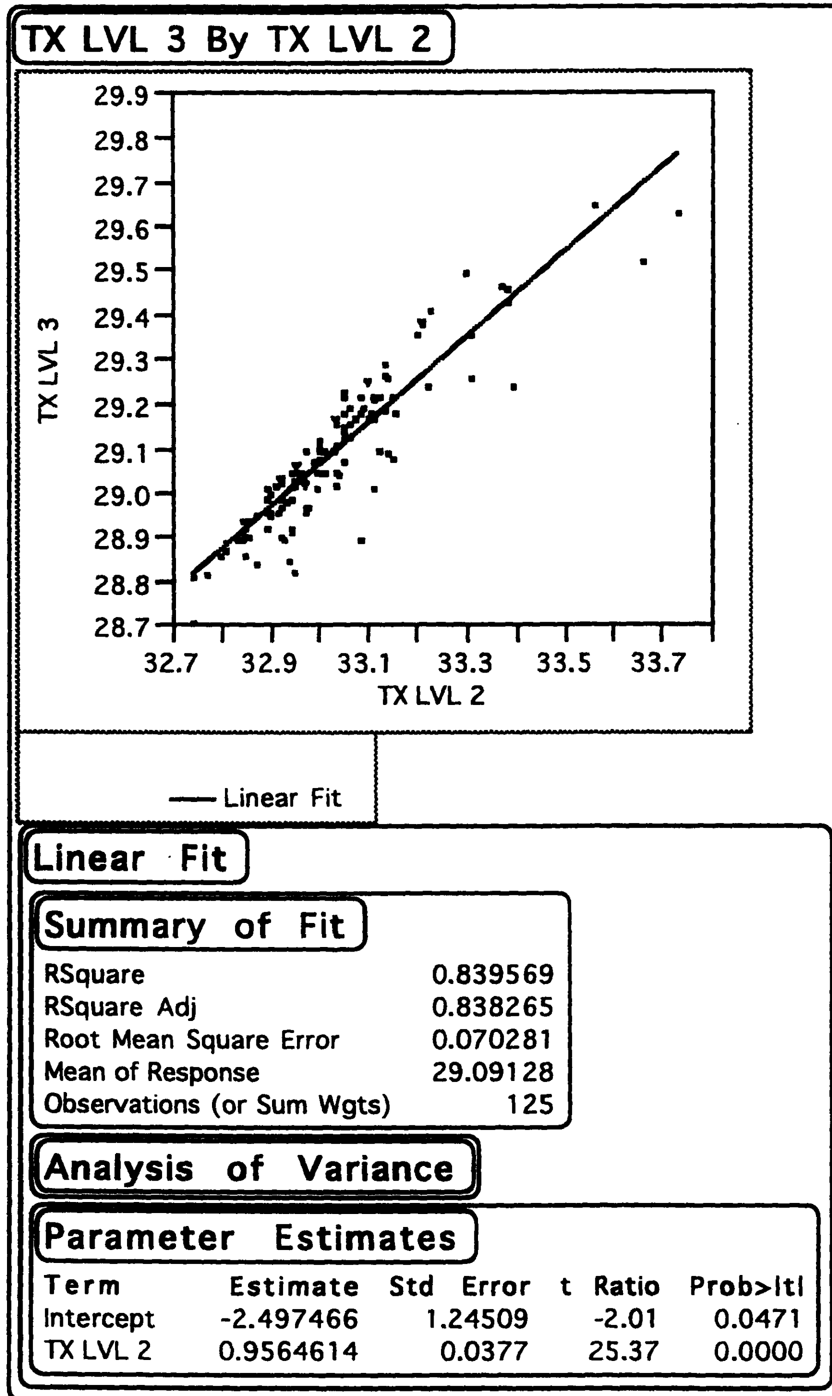
To develop such a statement of probability is a matter of finding how well the results of one test can predict the results of others. One way to do this is to pick a candidate x-test, and some other likely y-tests whose results can be predicted by the x-test. To discover likely candidates, JMP has a feature by which a matrix of correlations and a matrix of scatter plots can be generated for a large number of candidate tests. A test that correlates closely to several others would be a good candidate x-test. Figure 7.3 shows a correlation matrix calculated for a set of tests on transmitter power output accuracy. This particular transmitter has a power control feature with 5 settings. A question to be answered may be: do we need to verify all 5 power settings?



Pairwise Correlations

Variable	by Variable	Correlation	Count	Signif Prob
TX LVL 2	TX LVL 1	0.9645	126	0.0000
TX LVL 3	TX LVL 1	0.8421	127	0.0000
TX LVL 3	TX LVL 2	0.9163	125	0.0000
TX LVL 4	TX LVL 1	0.5162	128	0.0000
TX LVL 4	TX LVL 2	0.7341	126	0.0000
TX LVL 4	TX LVL 3	0.7084	127	0.0000
TX LVL 5	TX LVL 1	0.3229	128	0.0002
TX LVL 5	TX LVL 2	0.5418	126	0.0000
TX LVL 5	TX LVL 3	0.6012	127	0.0000
TX LVL 5	TX LVL 4	0.8726	128	0.0000

Fig. 7.3: Tx power correlation matrix



Linear Fit

Summary of Fit

RSquare	0.839569
RSquare Adj	0.838265
Root Mean Square Error	0.070281
Mean of Response	29.09128
Observations (or Sum Wgts)	125

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2.497466	1.24509	-2.01	0.0471
TX LVL 2	0.9564614	0.0377	25.37	0.0000

Fig. 7.4: Model fit

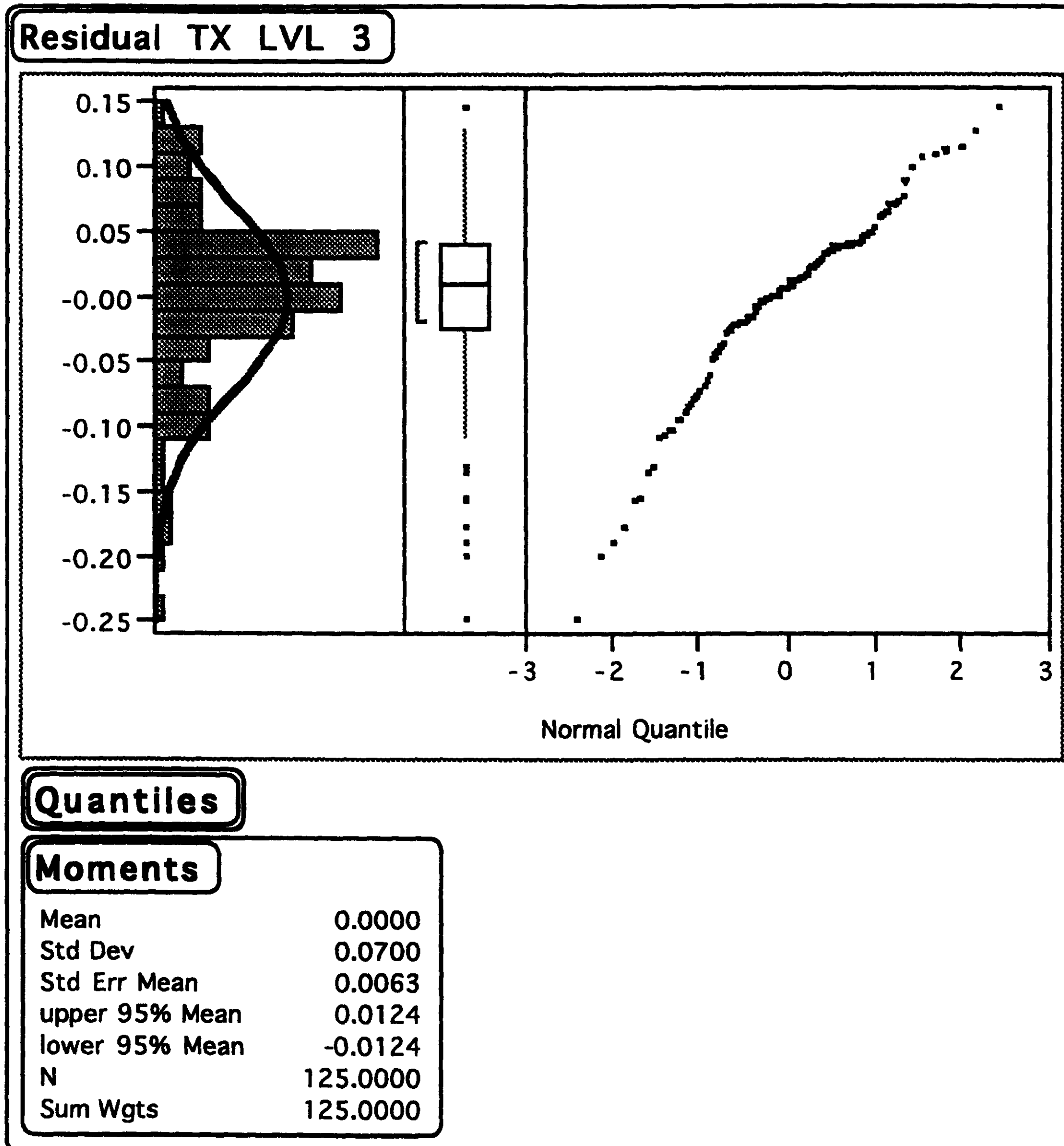


Fig. 7.5: Residuals from model fit

After the candidate tests are identified, the approach to quantify the statement of conditional probability is to fit a model for each y-test based on the x-test, and examine the distribution of the residuals. If the residuals are reasonably distributed (normal tests and normal quantile-quantile plots can check for normality), a statement can be made about the size of a confidence interval of the prediction. Figures 7.4 and 7.5 show the model fit and residual analysis for a pair of the Tx power out tests.

Note that the slope of the line fit predicting TX LVL 3 from TX LVL 2 is estimated as 0.956 with a standard error of 0.038. The physics of the circuit says that the slope should be 1.0, which is a hypothesis that wouldn't be rejected by this data. The standard deviation of the residuals is 0.07 dB, which allows us to say with about 95% confidence that the error at TX LVL 3 is within 0.15 dB of the error at TX LVL 2.

A reminder at this point is appropriate--engineering judgment should always be applied when deciding test strategy based on data and the coverage of other tests, the process capability, etc. The method presented is merely a tool to assist the engineer in quantifying and interpreting the facts presented by the test results.

When a correlated pair of tests has been discovered, the following list of questions is appropriate. First, which of the two tests is more expensive? Include test time, equipment used, rate of type-I errors (False failures, NTF) and type-II errors (False accepts, downstream failures that are verified). A matrix of equipment used vs. test at a given test stage is useful at this point. Does the test use a unique piece of equipment? If so, it is expensive in terms of capital equipment. Second, which of the two tests is more valuable, in terms of likelihood of detecting legitimate defects or identifying an adjustment that needs to be made? These probabilities are functions of process capability and the cost of letting a defect out the door. Third, are there circuits for either test that are used only when that test is performed? If so, the test is valuable in terms of coverage. It is also useful to ask questions about cause and effect, to understand why two tests are related.

A final note concerns the use of this technique for a multi-stage test process. The correlations can be run for tests across a single stage as well as for tests across stages (vertical and horizontal relationships). For instance, if sensitivity is measured at board, module, and system test, correlations among the results upstream and downstream can help identify redundant tests. This was used in chapter 5 in the initial analysis of the fidelity data.

Chapter 8: Conclusions and Recommendations

In conclusion, the basic argument of this thesis is that it is well worth the time and effort to thoroughly analyze the test data available from production test and prototype testing. There are two sides to this dynamic. On the cost side, the power of computers, software, databases is increasing rapidly and making it much easier to do thorough analysis. Inexpensive personal computers and statistical software packages like JMP provide the tools for analysis. Modern networking enables the data to flow from source to destination, to be turned into useful information for decision-makers. The arguments on the benefit side appear throughout this thesis and the literature, showing how data-driven problem solving leads to improvements in quality and cycle time in both manufacturing and product development.

The recommendation that stems from this conclusion is to invest more in training and tools to enable production and development personnel to have access to the full data set and to know how to use it.

In product development, the task of the design team is to learn how to provide a quality product that meets customer needs in the minimum time possible. In a sense, the approach that maximizes learning will result in minimum cycle time. This leads to another recommendation. Put time in the prototyping schedule for analysis, especially in the early iterations. Couple this with structured methods and questions to be answered to maximize the benefit from the analysis period. This approach is likely to minimize costly design changes and delays that occur at the end of the development project.

A final conclusion is that test data provides a factual basis for communication across functions. Facts break down barriers that stem from different cultural assumptions, forming a universal language of communication. To the extent that the relevant data is available for any issue, it will help the players communicate and solve the problems effectively.

Bibliography

On the KJ Method:

1. Jiro Kawakita, *The Original KJ Method*, Kawakita Research Institute, Tokyo, 1991
2. Shoji Shiba, *Step by Step KJ Method*, Center for Quality Management, Cambridge, 1991
3. S.I. Hayakawa, *Language in Thought and Action*, 5th ed., Harcourt Brace Jovanovich, Orlando, FL, 1990

In the fields of quality control and TQM:

4. J.M. Juran (Ed.), *Juran's Quality Control Handbook*, 4th ed., Mcgraw-Hill, New York, 1988
5. Deming, William Edwards, *Out of the crisis : quality, productivity and competitive position,* Cambridge University Press, Cambridge, 1986
6. Shoji Shiba, Alan Graham, and David Walden, *A new American TQM*, Productivity Press, Cambridge, 1993
7. Yoji Akao, *Quality Function Deployment*, Productivity Press, Cambridge, 1990
8. John R. Hauser and Don Clausing, "The House of Quality," *Harvard Business Review* 66, no. 3 ,May-June 1988
9. Michael J. Harry and Reigle Stewart, *Six Sigma Mechanical Design Tolerancing*, Motorola University Press publication 6-σ-2-10/88

On Robust Design:

10. Clausing, Don, *Total Quality Development: A step-by-step Guide to World Class Concurrent Engineering*, ASME Press, New York, 1994
11. Phadke, Madhav S., *Quality Engineering using Robust Design*, Prentice-Hall, 1989
12. Taguchi, Genichi, *Introduction to Quality Engineering*, Asian Productivity Organization, 1986 (Available from American Supplier Institute, Dearborn, MI)
13. Karl Ulrich and Steven Eppinger, *Product Design and Development*, McGraw-Hill, 1995

The Cost of Quality:

14. Feigenbaum, A. V., *Total quality control*, 3rd ed., McGraw-Hill, New York, 1983
15. *Annual Quality Congress Transactions*, v.35-45, 1984-1994, ASQC, Milwaukee, WI

16. Paul Manikas and Stephen Eichenlaub, "Reducing the Cost of Quality Through Test Data Management," *Engineering Costs and Production Economics*, v.18, 1989, pp. 1-10

17. James Hungerford, "Computer aided test process models: what do they tell about test cost?," *Engineering Costs and Production Economics*, v.18, 1989, pp 11-21

Optimization Models:

18. M. H. Burman, "A Survey of Current Issues in Production Planning Research," Laboratory for Manufacturing and Productivity, Working Paper LMP-93-003, Massachusetts Institute of Technology

19. Tzvi Raz, "A Survey of Models for Allocating Inspection Effort in Multistage Production Systems," *Journal of Quality Technology*, v. 18, No. 4, Oct. 1986

20. Suguna Pappu, "A Dual Ascent Algorithm for Finding the Optimal Test Strategy for an Assembly Sequence," Thesis (M.S.)--Massachusetts Institute of Technology, Sloan School of Management, 1989.

21. Phillipe Chevalier and Lawrence Wein, "Inspection for Circuit Board Assembly," Alfred P. Sloan School of Management Working Paper #3465-92-MSA; September 1992.

22. Michael L. Peterson, "Testing for Time-Dependent Failures in Electronic Products," Thesis (M.S.)--Massachusetts Institute of Technology, Leaders for Manufacturing Program, 1994.

Applied Probability and Statistics:

23. Robert V. Hogg and Johannes Ledolter, *Applied Statistics for Engineers and Physical Scientists*, Macmillan, New York, 1992

24. George E.P. Box, William G. Hunter, and J. Stuart Hunter, *Statistics for Experimenters*, Wiley, 1978

25. Paul Mullenix, "The Capability of Capability Indices with an Application to Guardbanding in a Test Environment," Proceedings of the 1990 International Test Conference, IEEE, paper 40.1

Appendix 1: Interview Guide

This guide was used by the author to interview six professionals from the organization. Careful notes were taken and a KJ diagram was composed from the results.

Introduce myself, give time estimate for this interview (1-2 hours). Mention that I would appreciate the names of people who are likely to be helpful for finding more information.

1. Tell me what you spend most of your time doing? What is your perspective? What are the biggest problems associated with test?

2. Could you tell me the reasons for doing production test, as you understand them? Could you distinguish between board, receiver, and system tests?

3. Which reasons do you like and dislike? What would you like to see changed and why?

4. What kind of data from test is stored, and how is it used? Any improvements? Why? Is there any classification of test results other than binary pass/fail? How are PPM metrics collected and calculated?

5. How are decisions made on what to test where, conditions, acceptance criteria? What role does the Factory Quality Assurance group play?

6. What is it that makes RF designs go bad on the factory floor? (I'll be looking at data with this question in mind--and asking where in the process are faults detected and where could they be detected?)