STUDIES IN APPLIED STATISTICS:
SHIP HULL DESIGN OPTIMIZATION
AND
ENDOGAMOUS GROUP COMPARISONS

by

ANONA ELAINE THORNE
B.A., The University of British Columbia, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

We accept this thesis as conforming
to the required standard
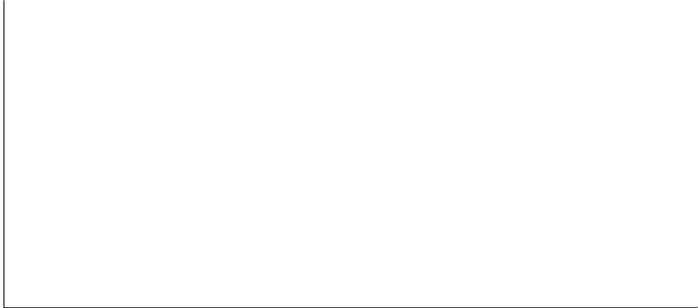
THE UNIVERSITY OF BRITISH COLUMBIA
August 1993

© Anona Elaine Thorne, 1993

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of __Statistics__

The University of British Columbia
Vancouver, Canada

Date __August 31, 1993__

Abstract

This paper contains two case studies in applied statistics.

The first study originated from a series of seakeeping experiments on twenty-seven different ship hull designs. The experimenters sought a method of predicting performance for hull designs different from those included in the study. The results of these experiments had already been analyzed, but we wished to consider the Taguchi method as an alternative method of analysis.

In this study, we describe the initial design problem and analyses and provide a critical evaluation of the Taguchi method; we assess its merits to determine its suitability for the ship hull design problem. We find the Taguchi method unsatisfactory in general, and inappropriate for the ship hull design problem.

The second study derives from a case which came to the Statistical Consulting and Research Lab (SCARL) in the Department of Statistics at The University of British Columbia. Having lost all her original data, our client wished to know if it were possible to "reconstruct" her data or, in any event, to perform a statistical analysis, using the tables of summary statistics in her possession.

She wished to model several different pulmonary functions using various demographic and anthropometric measurements as explanatory variables. For each of the different model variants under consideration, she wanted to see if the same model parameters would be valid for all endogamous groups, or if some groups were sufficiently different that the parameter values would be different for these groups.

We were able to develop a new general method for application to her problem; the derivation and application of this methodology are presented in the second study.

Contents

List of Tables

List of Figures

Acknowledgements

Part I

Ship Hull Design Optimization

Summary

The current study originated from a series of seakeeping experiments on twenty-seven different ship hull designs. The experimenters sought a method of predicting performance for hull designs different from those included in the study. The results of these experiments were analyzed by using linear and nonlinear regression (Delampady et al., 1989), and by minimizing three different loss functions (Delampady et al., 1990).

Continuing the work on this problem, we wished to consider alternate methods of analysis. Since the Taguchi method provides a well-established, but somewhat controversial procedure for design optimization, we decided to evaluate it as a possible alternative.

In this report, we describe the initial design problem and the 1989 and 1990 analyses of the experimental data. We then provide a detailed description and critical evaluation of the Taguchi methodology. Our purpose is to assess its merits in general and to determine whether it would be suitable for use in the ship hull design problem. We conclude that the Taguchi method is not logically sound, and is inappropriate for use in the ship hull design problem.

1 Introduction

When ships are designed, marine architects must consider their performance in all kinds of weather. They seek ship hulls which can be operated in a selected range of speeds without exhibiting any objectionable tendencies, such as pitching, rolling, heaving, and slamming.

New designs are often tested by towing models in tanks at various speeds and under diverse water conditions. A series of such seakeeping experiments was conducted by the Institute for Marine Dynamics of the National Research Council of Canada; a full summary of the resulting data can be found in Murdey and Butler (1985). The experimenters sought a method of predicting performance for hull designs different from those included in the study, based on the results of these experiments.

The data from twenty-seven different ship hull designs, of which twenty-five possessed proven satisfactory performance at sea, were analyzed by Delampady, Kan, and Yee (1989, hereafter DKY89) and by Delampady, Gu, Ma, Meloche, and Molyneux (1990, hereafter DGM90). Our purpose in the current study is to investigate an alternative method of analysis. Since the Taguchi method is a popular, but controversial procedure for design optimization, we have chosen to review the methodology to determine whether it would be an appropriate alternative.

In the following sections, we describe the 1989 and 1990 analyses. We then explain the mechanics of the Taguchi method and discuss its possible application to the current problem. Finally, we examine criticisms of the methodology.

2 Initial Analyses

Before describing the initial analyses, we introduce the following standard naval architecture terms in non-technical language to give an idea of the meaning of the variables which will be used (for further details, see Comstock (1967)):

Fn, Froude number: $V/\sqrt{gL}$, where V = speed of ship, g = acceleration due to gravity, and L = length of ship; Fn provides a measure of the ship's speed in relation to its size,

$\bar{T}_z$, non-dimensional zero crossing wave period: a measure of wave conditions which includes wave period and height,

L^2/BT, length-displacement ratio: the ratio of length to displacement where B = beam of ship measured at load waterline, and T = draught, the depth of water required for the ship to float,

B/T, beam-draught ratio: the ratio of beam to draught,

C_B, block coefficient: ∇/LBT, where ∇ = volume of ship's displacement; C_B represents the ratio of the volume of displacement to the volume of a rectangular solid of equal length, breadth, and depth,

C_W, waterplane coefficient: (area of waterplane)$/LB$, where the waterplane is the horizontal plane through the ship at load waterline; C_W is the ratio of the area of the waterplane to the circumscribing rectangle,

pitch motion: rotational motion of a ship about its lateral axis,

heave motion: motion of a ship along its vertical axis,

bow acceleration: a measure of vertical bow acceleration,

bow relative motion: a measure of the motion of the ship's bow incorporating vertical acceleration and rotational motion about the ship's longitudinal axis,

thrust increase: the increase in forward driving force on the ship.

The first analysis of the experimental data was conducted by DKY89. The *explanatory* or *predictor* variables used by the experimenters were:

Fn : Froude number and

$\bar{T}_z$: non-dimensional zero crossing wave period,

together with the *design* variables

L^2/BT : length-displacement ratio,

B/T : beam-draught ratio,

C_B : block coefficient, and

C_W : waterplane coefficient.

Their *response* variables were denoted by:

y_1 : pitch motion,

y_2 : heave motion,

y_3 : bow acceleration,

y_4 : bow relative motion, and

y_5 : thrust increase.

High levels of all these responses are considered undesirable. They cause discomfort for the crew, impede the ability to maintain headway, and can coincide with dangerously high forces on the ship's hull, leading to damage and possible disaster. There may be other ramifications as well, depending on the type of ship. On a naval ship, for example, the limits for pitch might be critical if missiles were being tracked or sonar were used; on an aircraft carrier, the amplitude of vertical motion at the end of the flight deck would be important.

DKY89 used both linear and nonlinear regression to relate each of the response variables separately to the set of explanatory variables. In the case of *bow relative motion*, the linear model provided a slightly better fit; for *pitch motion* and *heave motion*, the non-linear model proved to be better. For the other response variables, no real improvement was provided by the nonlinear model over the linear model.

The model for thrust increase was by far the worst fitted model among those for the five responses considered. It was observed that there was a general lack of pattern in

the variation of thrust increase with respect to the explanatory variables. Furthermore, it was felt that the explanatory variables considered may not have been adequate for model building.

The 1989 study concluded with three recommendations:

1. Since the responses were measured simultaneously, they could exhibit an interesting multivariate structure, which should be exploited in the construction of an optimal design.

2. Only five of the most important response variables were analyzed; the analysis should be extended to the rest of the variables which appeared in the experiments.

3. Since there are many design variables and environmental factors involved, it was recommended that the response surface model of Weerahandi and Zidek (1988) be used.

A subsequent study (DGM90) was directed towards finding the ship model with the best possible performance. In this study, d was used to denote the vector of hull form design variables, which again consisted of L^2/BT, B/T, C_B, and C_W; a denoted the adjustable variable Fn and c the uncontrollable variable $\bar{T}_z$. Fn was called an "adjustable" variable since under actual operating conditions the value of Fn, in other words the ship's speed, would be controlled by the captain; $\bar{T}_z$ was considered an "uncontrollable" variable since the operator would have no control over sea conditions.

The response variable used by DGM90 was the vector $R = (R_1, R_2, R_3)'$ where the components of R were:

$$R_1 : \quad \text{heave motion;}$$
$$R_2 : \quad \text{bow acceleration; and}$$
$$R_3 : \quad \text{thrust.}$$

The set of response variables used here did not include pitch and bow relative motion, since the experimenter believed the information provided by these responses was included in R. Under the assumption that d, a, and c provide the best explanation of variation in R, we have $R = R(d, a, c)$.

5

The investigators considered a loss function L, where $L = L(\boldsymbol{R}(\boldsymbol{d}, a, c))$. They sought a design $\boldsymbol{d}$ which minimized L, bearing in mind that the same design might not be optimal for all a. Furthermore, since c is a combination of a large number of uncontrollable factors, they decided to treat c as a random variable, and to average L with respect to the probability distribution of c. The average loss due to the use of a ship hull of design $\boldsymbol{d}$ at speed a was defined to be $\phi(\boldsymbol{d}, a) = E_C[L(\boldsymbol{R}(\boldsymbol{d}, a, C))]$, where C is a random variable which assumes values c, and E_C denotes the expectation with respect to the probability distribution of C. The problem of locating the optimal design $\boldsymbol{d}$ now consisted of finding $\boldsymbol{d}$ which produced the minimum value of ϕ. Of course, it might be found that a different optimal design $\boldsymbol{d}$ existed for each speed a at which the ship could be operated. This question will be considered shortly.

Multiple response, multiple linear regression techniques were used to model $\boldsymbol{R}$ as a function of c, where the linear model of DKY89 was used. Three different loss functions were considered. In the notation used by DGM90, these three functions were:

$$L(x_1, x_2, x_3) = \sum_{i=1}^{3} |x_i|,$$

$$L(x_1, x_2, x_3) = \sum_{i=1}^{3} x_i^2, \quad \text{and}$$

$$L(x_1, x_2, x_3) = \sum_{i=1}^{3} \exp(|x_i|).$$

Performance was measured by averaging ϕ over the various Froude number values (denoted by a), since they were all considered to be of equal importance.

The investigators found that the relative ranks of hull models, determined by minimizing ϕ over $\boldsymbol{d}$, were not sensitive to the choice of loss function from the three functions given above. In addition, they concluded that hull length was not an important factor in the determination of the optimal design, because the relative rankings remained essentially the same for different hull lengths.

The hull forms with good performance all had similar design dimensions. In particular, L^2/BT of 150 and B/T of 5.20 seemed to provide the best performance. However, in view of the fact that there would certainly be interaction between these two parameters, it could only be said that the best design possessed parameter values close to these.

The 1990 study recommended that tests be done on hull forms close in dimensions to those which were found to be optimal. Furthermore, it suggested that in future analyses, the hull length be taken as 120 m., and the expected losses determined only for the squared error loss function. A nonlinear model could be used in place of the linear model used here.

Because the focus of this project was the quality of ship hull performance, we felt that it would be interesting to consider the use of the Taguchi method for the solution of this design problem. The Taguchi method is a well-known methodology for experimental design and analysis; its purpose is to incorporate quality goals in the initial design stage of a product rather than to make costly adjustments to achieve quality at a later stage in the manufacturing process. This emphasis suggested that the method might be very appropriate for the current situation.

3 The Taguchi Method of Parameter Design

In this section, we describe the basic elements of the Taguchi method of design optimization, beginning with the ideas motivating the method. Our description is based primarily on readings of Byrne and Taguchi (1987), Ross (1988), and Tunner (1990).

3.1 Quality and the Loss Function

Traditionally, manufacturers and consumers have had rather distinct views of quality. The consumer will be concerned with utility, durability, value, aesthetics, and various other attributes. Although initially the manufacturer must surely have the same concerns, by the time a product reaches the manufacturing stage, the manufacturer will have translated these criteria into the simple issue of whether the product meets its design specifications.

Generally speaking, it is intuitive that quality is improved as variation is reduced. In an engineering environment, this intuitive view can be made precise. Parts designed to fit together must each be the correct size or the final product will work badly or perhaps not at all. Engineers are also aware, however, that it is impossible to reduce variability to zero. For this reason, engineers specify both a target value and acceptable limits when establishing specifications.

If the target for a dimension were 1000 mm., for example, the design specifications might then be 1000 ± 2 mm.. In other words, the lower specification limit (LSL) is 998 mm. and the upper specification limit (USL) is 1002 mm.. This implies that a part which is 997.9 mm. long (part A) is completely unacceptable, while a part which is 998.1 mm. long (part B) is exactly as good as one with the precise target length of 1000 mm. (part C). Furthermore, it is implied that the production of part B will cause no loss to the company as a result of its deviation from the target value.

From the customer's point of view, there is very little difference between part A and part B, since the customer is concerned with the final function of the part. There might, however, be a good deal of functional difference between part B and part C.
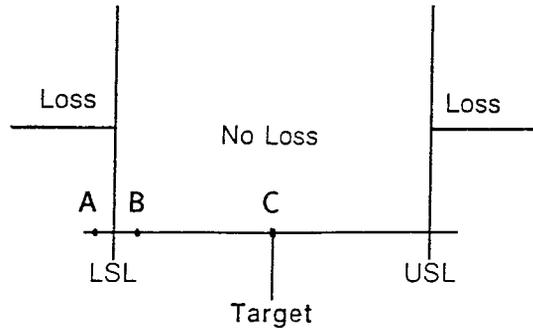
Figure 1: Traditional loss function (Ross, p. 42)

To the manufacturer using the artificial LSL, on the other hand, there seems to be an enormous difference between parts A and B and no difference at all between parts B and C. The manufacturer will incur scrap or rework costs for part A if inspection is one hundred per cent effective, or warranty costs if inspection is not completely effective. Part B will supposedly generate no costs whatsoever, and part C will definitely generate no costs.

The "traditional" loss function reflects this viewpoint. It is shaped like a "goal-post": its value is zero inside the product specifications and takes a "step" jump at the specification limits, as can be seen from Figure 1.

In practice, of course, this loss function represents a gross oversimplification. The size of a car door, for example, may be within specifications but not quite on target. The loss to the consumer could range from simple annoyance at having a door which fits somewhat imperfectly to the inconvenience of water leaks. The resulting loss to the manufacturer could be damage to its reputation and possible loss of market share.

It is therefore necessary to reconcile the customer's and manufacturer's views by using a loss function which reflects both perspectives. It is also important to note that as customers become more demanding, they increase pressure to reduce variability.

A simple parabolic curve, as shown in Figure 2, has an intuitive appeal, especially when specifications are symmetric about a target value. The loss is very small close to the
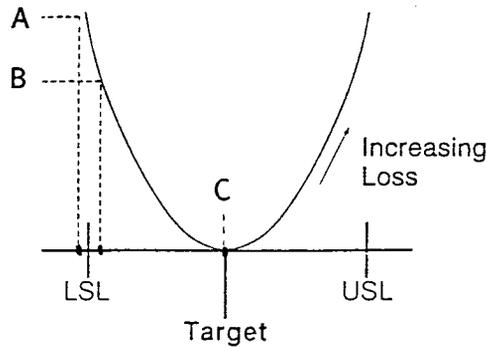
Figure 2: Parabolic loss function (Ross, p. 42)

target and grows at an increasing rate as one moves away from the target. The curve flattens to a value equivalent to scrap cost outside the specifications since it is reasonable to assume that generally speaking, products which fall outside the specification limits will be scrapped.

3.2 The Taguchi Approach

Genichi Taguchi is a Japanese engineer whose ideas have had a significant effect on Japanese and more recently North American technology, beginning with his work in the early 1950's. To incorporate his idea that loss increases with increasing distance from the target value, Taguchi uses a quadratic loss function like that shown in Figure 2:

$$L = K(S^2 + (\bar{X} - T)^2),$$

where

$$L = \text{average loss per unit produced,}$$
$$K = \text{cost constant,}$$
$$S^2 = \text{process variance,}$$
$$\bar{X} = \text{average specification value, and}$$
$$T = \text{target specification value.}$$

Furthermore, if one defines

c = loss associated with a unit of product produced at a specification limit (assuming that the loss for a unit at target is zero), and

d = distance from the target to the specification limit,

then,

$$K = c/d^2 \text{ (Tunner, 1990, p. 58).}$$

While it may be difficult to produce a valid estimate of c, it should be possible for experts in the subject area to derive a reasonable value. One group of engineers has suggested the rule of thumb that c be one-tenth of the selling price of the item in question. (see Tunner, 1990, p. 59)

The quadratic loss function underlies Taguchi's entire approach to quality engineering. The function shows the economic advantage of reducing variation while incorporating customer requirements. Reducing variation and staying close to the target are both important, since either or both of these achievements will lead to reduction in total loss. In fact, it is often the case that cost will be reduced at the same time that quality is improved.

Furthermore, Taguchi divides system quality engineering into two parts: off-line and on-line quality control. He stresses the need to engineer quality into a product in the design stage in order to minimize the costs of maintaining quality during the manufacturing process. The off-line process will establish a design target for minimum loss and reduce variance by design before production actually starts. Later, the on-line process will hold the average close to the target.

According to this scenario there are three steps in the (off-line) engineering optimization of a process: system design, parameter design, and tolerance design. System design involves the selection of materials, parts, equipment and tentative parameter values. The parameter design stage determines the optimal levels of parameter values; values for product parameters and process factors are chosen which are least sensitive to changes in noise factors. This is the crucial stage for attaining high quality at minimal cost. If parameter design does not reduce variation sufficiently, tolerance design must be

used. This usually implies greater cost, since it means buying better materials, parts, or machinery in order to tighten tolerances on product parameters or process factors.

To clarify these ideas, consider the following simplified example. In the system design stage for the manufacture of ceramic tiles, the type of kiln and raw materials would be selected, along with tentative values for proportions of raw materials. Suppose it were known that the temperature within the kiln varied by a specific amount; this variation would be considered a noise factor. In the parameter design phase, proportions of raw materials would therefore be selected which would result in minimal sensitivity to this temperature variation while producing tiles which met the desired specifications. If the parameter design phase were omitted, it might later be necessary to make expensive kiln modifications (tolerance design) in order to reduce kiln temperature variation, thereby achieving the necessary tile quality.

Note that even in cases where parameter design has been omitted initially, it is possible to go back to this stage after production has begun. In this way, one can still hope to avoid incurring the expense of tolerance design. The tile production example used above was taken from a real case in Japan in 1953 (Byrne & Taguchi, 1987, hereafter BT87, p. 20). By choosing better factor levels, the percentage of size defects in the ceramic tiles produced was reduced from 30% to less than 1%.

In North America, engineers commonly jump from system design to tolerance design since they are accustomed to spending money to obtain the required performance levels. In addition, Taguchi methods are often applied to a current production problem to improve quality, when they could have been used much earlier in the design phase.

Taguchi's idea of reducing variation is further refined by the emphasis on doing this in the most cost effective manner possible. The important factors in a process are separated into control factors and noise factors. Control factors are those factors which a manufacturer can control relatively easily. Noise factors are those which a manufacturer cannot or does not wish to control, perhaps due to the high cost of such control. The goal of the parameter design process is to reduce variation by choosing levels of the control factors which will make the response less sensitive to changes in noise factor levels, since noise factors, by their very nature, are usually more expensive to control. Control factor levels are chosen by conducting appropriate experiments.
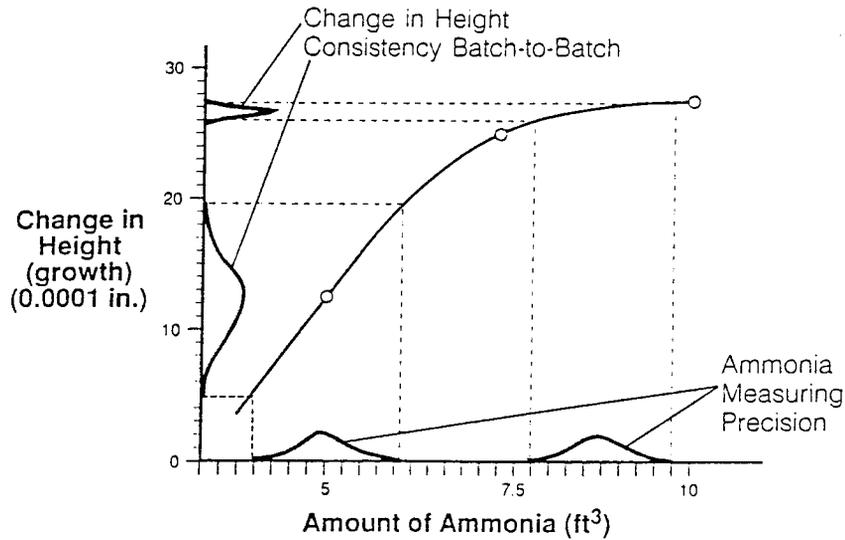
Figure 3: Result of Heat Treatment Experiment (Ross, p. 47)

Figure 3 illustrates the result of a designed experiment which was used to improve a heat treatment process. The goal was to meet the height requirements for a cam follower. The precision of the ammonia measuring system, i.e., variation in the amount of ammonia, was treated as a noise factor, since the cost of improving it would have been significant. The experiment indicated that the amount of ammonia had a strong influence on controlling the change in height experienced in the process. The previous production standard of 5 cubic feet of ammonia resulted in a large variation in height for the cam followers. When the amount of ammonia was increased to 8.75 cubic feet, the variation in height was reduced substantially. In other words, the process was less sensitive to variation in quantities of ammonia when the amount of ammonia was increased. The cost of the additional ammonia was very small. Here again, the concept of using parameter design to find parameter levels (in this case amount of ammonia) which were relatively insensitive to noise (variation in amount of ammonia) resulted in an improved process at relatively low cost.

Note that "the search for interactions among controllable factors is de-emphasized, although there are exceptions. The key to achieving robustness against noise is to discover the interactions between controllable factors and noise factors. Specific interactions between controllable factors and noise factors need not even be identified. As long as

13

Controllable Factors	Levels		
A. Interference	Low	Medium	High
B. Connector Wall Thickness	Thin	Medium	Thick
C. Insertion Depth	Shallow	Medium	Deep
D. Percent Adhesive in Connector Pre-dip	Low	Medium	High

Uncontrollable Factors	Levels	
E. Conditioning Time	24h	120h
F. Conditioning Temperature	72°F	150°F
G. Conditioning Relative Humidity	25%	75%

Figure 4: Factor levels for elastomeric connector experiment (BT87, p. 21)

the noise factors are changed in a balanced fashion during experimentation, preferred parameter values can be determined using an appropriate signal-to-noise ratio." (BT87, p. 21) The treatment of interactions and the use of signal-to-noise ratios for analysis will be discussed later. First, we will consider the design of experiments.

3.3 Experimental Design

The Taguchi method uses a crossed array design which is most easily explained by the use of an example. In this example (BT87, pp. 21-22), researchers aimed to attach an elastomeric connector to a nylon tube so as to maximize pull-off force, the force required to pull the connector off the tube. The researchers selected four important controllable factors: A) interference, B) wall thickness, C) insertion depth, and D) percent adhesive, and tested each at three levels. They also selected three noise factors: E) conditioning time, F) conditioning temperature, and G) conditioning relative humidity, each to be tested at two levels. Figure 4 shows the factors and levels. As noted earlier, noise factors are generally uncontrollable or too expensive to control during normal operations, but can often, as here, be controlled for experimental purposes.

The crossed array is formed from an inner array and an outer array. The inner array is an orthogonal array containing the levels of the controllable factors to be tested. An L_9 array (shown in Table 1) was selected for this purpose as the most efficient orthogonal

14

	Factor			
No.	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

Table 1: L_9 array for elastomeric connector experiment (BT87, p. 22)

design for four factors at three levels. Its task was to find the best of the $3^4 = 81$ possible factor combinations.

The outer array is an L_8 orthogonal array containing the levels of the noise factors to be tested (see Table 2). The columns labelled E, F, and G show the levels at which the noise factors are to be set. The other columns are used in the analysis to estimate interaction effects and experimental error. "The most important use of this array is to deliberately create noise to identify the controllable factor levels that are least sensitive to it. Usually the effects of interactions among the noise factors are not estimated, but the experimenters in this case allowed for such interactions, as they felt the information may be valuable." (BT87, p. 22)

The two arrays are then combined into a crossed array (see Figure 5). Each control factor combination is run with each noise factor combination, resulting in a total of 72 runs.

In this case, it was considered reasonable to run 72 experiments. Had it been desirable to reduce the number of runs to 36, an L_4 outer array could have been used instead of the L_8 array, at the expense of being able to examine interactions among noise factors. The number of runs could have been reduced even further to a total of 18, by

15

		Factor					
No.	E	F	E×F	G	E×G	F×G	e
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

Table 2: L_8 array for elastomeric connector experiment (BT87, p. 22)

L_8 ARRAY

	8	7	6	5	4	3	2	1
E	2	2	2	2	1	1	1	1
F	2	2	1	1	2	2	1	1
ExF	1	1	2	2	2	2	1	1
G	2	1	2	1	2	1	2	1
FxG	1	2	1	2	2	1	2	1
FxG	1	2	2	1	1	2	2	1
e	2	1	1	2	1	2	2	1
Cond. Time (E)	120h	120h	120h	120h	24h	24h	24h	24h
Cond. Temp. (F)	150F	150F	72F	72F	150F	150F	72F	72F
Cond. R.H. (G)	75%	25%	75%	25%	75%	25%	75%	25%

EXPERIMENTAL CONDITIONS — L_9 ARRAY

A B C D	Inter-ference (A)	Wall Thickness (B)	Ins. Depth (C)	Percent Adhesive (D)	75%	25%	75%	25%	75%	25%	75%	25%	Cond. R.H. (G)	S/N Ratio (db)
1 1 1 1 1	Low	Thin	Shallow	Low	19.1	20.0	19.6	19.6	19.9	16.9	9.5	15.6	------	24.025
2 1 2 2 2	Low	Medium	Medium	Medium	21.9	24.2	19.8	19.7	19.6	19.4	16.2	15.0	------	25.522
3 1 3 3 3	Low	Thick	Deep	High	20.4	23.3	18.2	22.6	15.6	19.1	16.7	16.3	------	25.335
4 2 1 2 3	Medium	Thin	Medium	High	24.7	23.2	18.9	21.0	18.6	18.9	17.4	18.3	------	25.904
5 2 2 3 1	Medium	Medium	Deep	Low	25.3	27.5	21.4	25.6	25.1	19.4	18.6	19.7	------	26.908
6 2 3 1 2	Medium	Thick	Shallow	Medium	24.7	22.5	19.6	14.7	19.8	20.0	16.3	16.2	------	25.326
7 3 1 3 2	High	Thin	Deep	Medium	21.6	24.3	18.6	16.8	23.6	18.4	19.1	16.4	------	25.711
8 3 2 1 3	High	Medium	Shallow	High	24.4	23.2	19.6	17.8	16.8	15.1	15.6	14.2	------	24.832
9 3 3 2 1	High	Thick	Medium	Low	28.6	22.6	22.7	23.1	17.3	19.3	19.9	16.1	------	26.152

Figure 5: Crossed array for elastomeric connector experiment (BT87, p. 22)

using an outer array which simply compounded the noise factors to two noise levels, a "nominal" and a "worst" condition. In addition to the restriction produced by the 36 run design, this design would also make it impossible to examine specific interactions between controllable factors and noise factors. Their identification is not generally considered necessary, however, due to the method of analysis used. In the following section we continue our exploration of the Taguchi method with a description of the analysis procedure.

3.4   Analysis

Taguchi's standard method of analysis examines a signal-to-noise (S/N) ratio, where the signal is considered to be the mean and the noise is the standard deviation. There are many different S/N formulae, but there are three which are generally used, one for each of three different response classifications: "the higher the better", "the lower the better", and "nominal is best". In each of these situations, the S/N ratio is constructed in such a way that the S/N ratio is to be maximized. The idea is that each of these ratios offers the best balance between optimizing the mean response and minimizing the response variation for the circumstances in which it is intended to be used.

Continuing with the elastomeric connector example, pull-off force falls into the "higher the better" classification, so the S/N ratio used is:

$$S/N \text{ (in dB)} = -10 \log \left[\frac{1}{n} \left(\frac{1}{y_1{}^2} + \frac{1}{y_2{}^2} + \cdots + \frac{1}{y_n{}^2} \right) \right].$$

where $y_1, y_2, \ldots, y_n$ refer to the n observations (in this case $n = 8$) for each setting of the controllable factors.

The analysis can be done in several ways. One common technique is to use F tests in a statistical analysis of variance (ANOVA) to find the statistically significant factors. Taguchi, however, often teaches engineers to use a conceptual approach which involves graphing instead. For each controllable factor, the average S/N ratio for each level is plotted on a graph (see Figure 6). The graphs show the factors which seem to have the strongest effect and the optimum setting for each factor when the mean and variance are considered together. Mean response plots (as shown in Figure 7) can be examined as
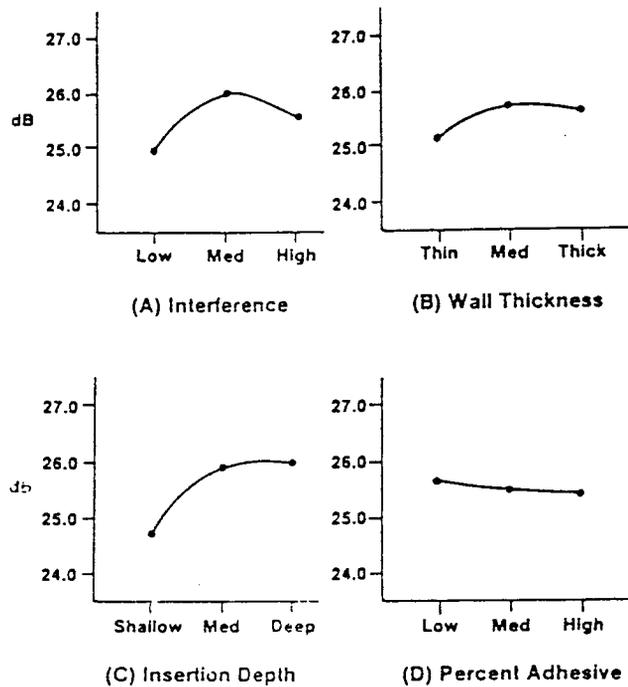
Figure 6: Average S/N ratio for each control factor level (BT87, p. 23)

well. These can supply additional information, since they can show whether the mean or the variance has the stronger effect on the S/N ratio.

If a factor does not have a strong effect, it is less important to choose the optimum setting for that factor from the graphical analysis, since there is apparently little difference in the response produced at each level. Instead, other considerations such as cost or convenience may be used in the selection process.

Although it is claimed that performing the S/N analysis eliminates the need to examine the specific interactions between controllable factors and noise factors, this is sometimes done graphically as well, in order to obtain more information about the problem. In the example in question, this graphical analysis was done for all of the control × noise interactions. The results did not alter the decisions made through the S/N and mean response analyses.

A separate study had been conducted on ease of assembly. When the results of the two studies were combined (which involved some compromises), the optimal control factor
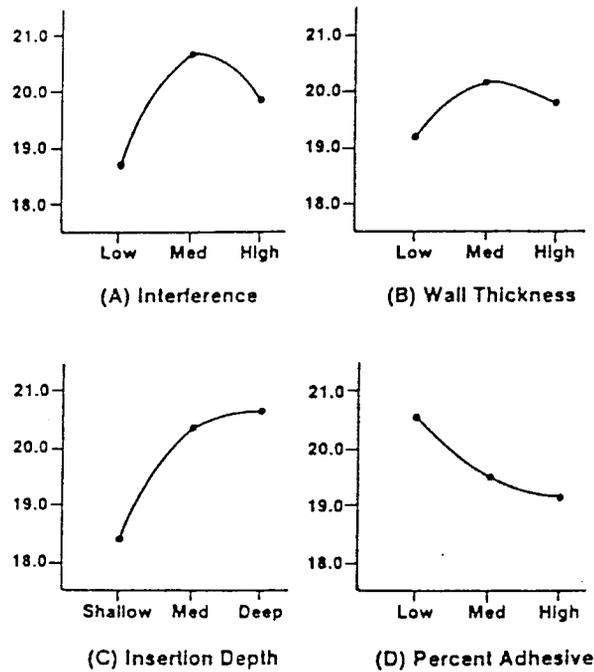
Figure 7: Mean responses for each control factor level (BT87, p. 23)

settings were determined to be A_2, B_1, C_2, and D_1. This control factor combination had never been run during the experiment, so a five-run confirmation experiment was conducted, which did indeed produce increased pull-off force.

If researchers feel that for some reason none of the three S/N ratios given above is appropriate for their situation, a common alternative divides the analysis into two steps:

1. Examine the S/N ratio, S/N (in dB) $= 20\log(\bar{y}/s)$, where $\bar{y}$ and s are the mean and standard deviation of the response from a given experimental condition. The factor levels which correspond to the highest S/N ratio are chosen to minimize variation.

2. Factors which do not appear to be significant in the S/N analysis do not have a strong influence on variation, and can therefore be chosen to adjust the mean, either towards a target level, or towards a maximum or minimum.

This approach puts more weight on variation, whereas the standard S/N ratios put more emphasis on the mean.

4 The Taguchi Method and the Ship Hull Design Problem

We shall now describe the author's interpretation of how the Taguchi method might be applied to the ship hull design problem; this is done to ensure that the method is, in fact, applicable to this problem. After our critical review of the Taguchi method in Section 5, we shall determine whether we consider its use in the ship hull design problem desirable.

Recall that in the 1990 study of the ship hull problem, the design variables were:

L^2/BT : length-displacement ratio,
B/T : beam-draught ratio,
C_B : block coefficient, and
C_W : waterplane coefficient.

In Taguchi parlance, we would call these the control factors. The set of explanatory or predictor variables comprised these control factors together with:

Fn : Froude number, and
$\bar{T}_z$: non-dimensional zero crossing wave period.

We would call $\bar{T}_z$ a compound noise factor, but there might be some uncertainty about the classification of Fn, the Froude number. Since this factor is adjusted when the ship is in operation, it is certainly not a noise factor. On the other hand, it should not be classified as a control factor because that would result in the choice of only one level of Froude number when the final analysis is complete. The fact is that we seek good performance at all Froude number values. A workable solution might be to conduct separate full experiments for each of various Froude numbers and then average the results, assuming that good performance at each of these speeds is equally important. This is a similar approach to that taken in the second study discussed earlier (DGM90). In that study, it was found that little information was lost in this averaging process.

Thus far we have dealt with the control and noise factors. Next we turn to the response variable, which must be scalar valued for the application of the Taguchi method.

However, the response in the DGM90 study was vector valued $\mathbf{R} = (R_1, R_2, R_3)'$ with components:

R_1 :   heave motion;

R_2 :   bow acceleration; and

R_3 :   thrust.

We must therefore modify our response in some way to make Taguchi's method applicable. One could optimize each of the three responses separately, of course, but doing that would have the same shortcomings as it did in the previous analysis.

We would prefer to construct a function for an overall performance measure which would include each of the components of $\mathbf{R}$, and possibly other performance measures as well. This would need to be done by a specialist with extensive knowledge of marine architecture. Historically, marine architects have not developed such a function; production of new designs has relied more on experience than on quantitative tools. Recently, however, there has been some effort to develop a generally accepted performance index (see Lloyd and Andrew, 1977). It would be interesting to conduct a study with such an index as the response to be optimized. Without such an index, we have not been able to proceed further at this time. However, this might well be the subject of future work.

Once the response is defined, one could construct the inner and outer arrays according to a Taguchi orthogonal design. The outer array would be very simple, since there is only one compound noise factor. The analysis could be done using the S/N ratio for the "highest the best" scenario. There is a potential problem which may arise if the optimization analysis were conducted with just the currently available data. In the Taguchi method, one envisions the construction of orthogonal arrays and a series of experiments on those arrays. But for the available data, the settings at which the experiments were run may not fit into the envisioned structure. This would need to be checked.

We have learned the mechanics of the Taguchi method and discussed fitting our problem into the Taguchi framework. In the next section, we will examine the method critically to determine its potential value in the solution of our problem.

5 Criticisms of the Taguchi Method

The motivating philosophy behind the Taguchi methodology appears sound; in quality control literature it is widely applauded (see Ryan, 1988 or Nair, 1992, hereafter N92, for example). The same cannot be said of the methodology developed to implement the philosophy, however. This area is rife with controversy.

The main criticisms of the method, as found in the author's review of subject area literature, seem to be:

1. Experimental Design:

    (a) A crossed orthogonal array design is seldom the most efficient experimental design. It often needs more experimental runs (and thus incurs higher costs) than are really necessary.

    (b) Ignoring control factor interactions as a matter of course is not a good policy.

    (c) The design does not allow for sequential experimentation.

2. Analysis

    (a) Marginal Average Plots:
        (i) The examination of plots of marginal averages is not a sound technique for determining the best control factor settings.

    (b) Response Function:
        (i) The signal-to-noise ratio is not the best response to examine, and can just complicate the situation unnecessarily.
        (ii) The loss function should be used directly in the optimization process rather than disappearing from the situation after motivating the method.

3. General Approach:

    (a) The methodology suffers from a preoccupation with optimization rather than seeking to accomplish optimization through an understanding of the physical process.

We now examine each of these criticisms more closely.

5.1 Experimental Design

5.1.1 Economy

For each combination of control factors in the inner array, observations are made at all combinations of the noise factor settings in the outer array. If m and n are the respective sizes of the inner and outer arrays, the product array will have mn runs. If control and noise factors are included in a single array, however, the total number of runs required can be significantly smaller. Several authors, including Shoemaker, Tsui, and Wu (see N92), have suggested the use of a combined array for reasons of economy.

Furthermore, we must recognize that the distinction between control and noise factors is somewhat arbitrary. As far as the physical system itself is concerned, this distinction is merely an artifice. It makes more sense to use a design which simply includes all the appropriate factors in the most efficient manner possible. According to Lorenzen (N92, p. 150), the combined array usually has a better confounding pattern than the product array, as well as superior robustness properties. The combined array, however, may sometimes require more control factor combinations; if control factor settings are more expensive to change than noise factor settings, the combined array can be more expensive to use in these cases.

5.1.2 Interaction Effects

Sacks and Welch (N92, p. 140) suggest that Taguchi has ignored interactions to minimize the number of runs needed by a product array. This seems indefensible and has been widely criticized by other authors as well, such as Ryan (1988), Hendrix (1991), and Box (N92), to name a few.

Byrne & Taguchi (BT87, p. 24) claim that "performing the S/N analysis generally eliminates the need for investigating the specific interactions between controllable factors and noise factors," although they offer no argument to support this statement. It is difficult to see the logical connection. In fact, they go on to say that the investigation of interactions may provide additional insights into a problem. As George Box has

noted (N92, p. 146), "because engineers have traditionally relied on one-factor-at-a-time experimentation, main effects will often have already been put to use, and it will be the unexpected interaction that is waiting to be discovered and sometimes to be exploited with dramatic results." It seems most peculiar to maintain that we must experiment to find the factors with significant main effects, while claiming at the same time that we know the control factor interactions are unimportant. Furthermore, we search for interactions between control and noise factors, while ignoring interactions between control factors. Since control and noise factors are alike in a mathematical sense, the absence of control factor interactions seems implausible; certainly they cannot be expunged by merely classifying certain factors as control factors.

Even if we accept the use of product arrays, we can often improve Taguchi's inner array designs without incurring the cost of additional runs. Taguchi advocates the use of orthogonal arrays for the control factors and has published a number of these array designs, some of which are suboptimal. For example, Ryan (1988, p. 35) describes one case where Taguchi and Wu show an L_8 array for investigating four main effects and two two-factor interactions, where three of the main effects are confounded with two-factor interactions. Instead of using this array one could simply use a 2^{4-1} fractional factorial where main effects are confounded with three-factor interactions. This is a far better design, with exactly the same number of experimental runs.

Strictly speaking, here Ryan's criticism is of a specific implementation of the Taguchi method, rather than the method in general. It does, however, illustrate that it is very important to be aware of the aliasing structure of an array. If a main effect is confounded with a two-factor interaction and the experimenter is unaware of this, it would be easy to conclude that the main effect is significant when in fact the interaction effect is significant, and vice versa. Section 5.1.3 will explore this idea further.

5.1.3 Sequential Experimentation

The Taguchi methodology requires that the experimental design be chosen, run, and then analyzed to determine the optimal control factor settings. When a fractional design is used in a "one-shot" approach, however, one must be very careful. It is important to

make a correct initial decision about the effects to be tested. If this decision is incorrect, everything which follows may be completely misleading due to aliasing. Unless we use a full factorial design, it seems sensible to take an iterative approach to "zero in" on the important factors. It is therefore often useful to do a preliminary screening (fractional factorial) experiment based on the researchers' initial views. The results can dictate which fraction should be run next in order to continue the investigation without having important effects aliased with each other.

Barad, Bezalel, and Goldstein (1989) give an excellent example of a sequential experiment. It stresses the implications of aliasing as well as the usefulness of preliminary research to determine which variables are important. In the example given, product development engineers tested six battery design parameters at two levels each in order to find out which of them significantly affected battery capacity. A full factorial experiment would have required 64 runs, but since the engineers felt that the most important effects were the six main effects as well as the BD, CD, and EF interactions, they were able to do a quarter-fraction screening experiment consisting of 16 runs.

The analysis of the quarter-fraction indicated that the BD interaction was very significant; this was surprising, since neither of the main effects B and D was significant. The alias structure showed that BD was aliased with AE, which had not been considered important in the initial discussion. Further investigation found that the AE interaction and several others which had been considered unimportant earlier were, in fact, significant.

5.2 Analysis

5.2.1 Marginal Average Plots

In general, plotting marginal averages to determine control factor settings simply will not work. This practice is comparable to conducting an experiment where the factors are varied one at a time. Plotting marginal averages is not useful unless all the interaction effects are zero, and as seen earlier, it is completely unreasonable to assume that this will necessarily be the case. Ryan (1988, pp. 34-35) gives a simple example for an

INTERACTION PROFILE   **MARGINAL AVERAGES**
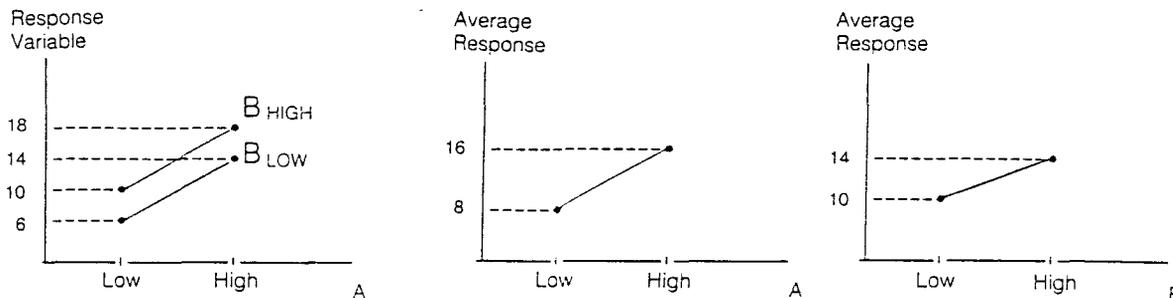
Figure 8: Interaction and marginal average plots for the 2^2 design (Ryan, pp. 34-35)
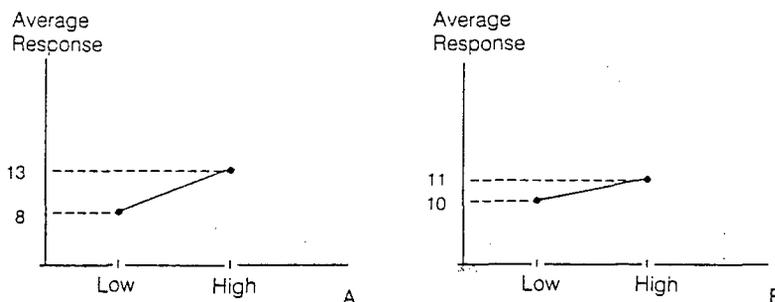
Figure 9: Marginal average plots for the revised 2^2 data (Ryan, p. 35)

unreplicated 2^2 design. Figure 8 shows the interaction and marginal average plots for the data. The lines on the interaction plot are parallel, indicating that there is no interaction between the two factors. The plots of marginal averages correctly indicate the choice for the highest response: $A_{high}B_{high} = 18$.

Now suppose the response at $A_{high}B_{high}$ had been 12 instead of 18. The new marginal average plots are shown in Figure 9 and would again suggest the choice of $A_{high}B_{high}$. In this case, there is an interaction between the two factors and the marginal plots do not give the correct solution, since the maximum response is in fact $A_{high}B_{low} = 14$.

The inescapable conclusion is that if we cannot rely on marginal average plots to give us the correct answer when all factor level combinations are included in an experiment,

we certainly cannot expect them to do so in a fractional design situation.

5.2.2 Response Function

Signal-to-Noise Ratio   To illustrate the inadequacy of the signal-to-noise ratio, Hendrix (1991, p. 75) re-examines the example cited earlier by Phillip Ross (1988, see Figure 3). In that example, high levels of growth were desired. It happened that high levels of ammonia maximized growth (the signal, in Taguchi's terms); in addition, high levels of ammonia minimized growth variation (the equivalent to Taguchi's noise). Since high ammonia levels maximized signal and minimized noise, they would, of course, maximize the signal-to-noise ratio. Note that in this example, response optimization coincides with low sensitivity to variation in the controllable factor.

Suppose, on the other hand, that low growth and low noise were favoured. Since low growth coincides with low ammonia levels and low noise with high ammonia levels, some sort of compromise is necessary. The balance depends the importance placed on each of these two objectives, but the solution will certainly be some intermediate ammonia level.

Taguchi gives several different signal-to-noise ratios for use when the response is to be minimized, but these alternatives are each simply arbitrary ways of weighting the two goals. It would be much simpler to look at the actual graph itself and determine one's own weighting scheme rather than to look at S/N ratios (which in this case are response/slope ratios). The situation becomes even more complicated when there are several controllable variables.

Figure 10 shows the response plots for an example with two controllable variables, X_1 and X_2, where the goal is to minimize the signal Y. Low levels of X_1 are preferable since they minimize Y, but the decision for X_2 requires some sort of compromise since high levels of X_2 minimize noise but low levels of X_2 minimize the signal. Re-expressing the data as an S/N ratio will not change the fact that a perfect solution cannot be found. Certainly it will help to cloud the picture. Here again, it is simpler just to examine the graph.
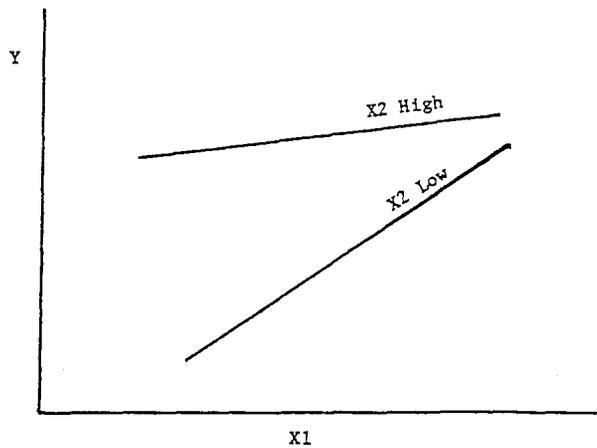
Figure 10: A simple example with two controllable factors (Hendrix, p. 76)

Hendrix minces no words in his assessment of the utility of the signal-to-noise ratio (p. 76):

> For those situations in which the response (signal) and the slope (noise) move in agreement with the objective ..., the Taguchi approach embodied as a signal-to-noise ratio will work — but so would a simple graphical presentation of the data. Reducing the data to a signal-to-noise ratio is unnecessary, because it doesn't contribute anything and only confuses the issue.

> For those situations in which the response (signal) and the slope (noise) are antagonistic ..., reformatting the data can only lead to an arbitrary compromise. The nature of this compromise is obscured by formatting the data as a signal-to-noise ratio. Disguising the slope (rate of change) as a standard deviation merely adds to the confusion.

At best, one might say that the various S/N ratios used appear to select an arbitrary balance between the mean and variance which is appropriate for a reasonable number of cases; in the process, what is really going on is hidden from view and the analyst's options are restricted unnecessarily. Not only does the S/N ratio "throw away" some of the information provided by the data, but it complicates the situation needlessly. If we model the quality characteristic of interest rather than the S/N ratio, we will have more background knowledge about the system to start with, and will gain more insight

into how the factors affect the response. In addition, the quality characteristic is often easier to model than an S/N ratio, resulting in a better model which leads to a more reliable optimization and ultimately, a better design (Sacks and Welch in the article of N92, p. 147).

Furthermore, although compromises between response level (signal) and sensitivity (noise) are addressed by the Taguchi approach, the problem of having to achieve a compromise among several responses, although very common in practice, is not addressed in Taguchi literature.

Loss Function Taguchi's ideas about the loss function and its proper form seem sound. Once the loss function has been postulated, however, it seems odd not to use it directly in the experimental analysis. Instead, its role is transient; it is used only to motivate the general approach. Madhav Phadke (N92, pp. 140-141) relates the S/N ratio to the loss function for the "higher the better" case, and claims that a direct minimization of the loss function itself would not be very effective in minimizing sensitivity to noise factors. He also states that when the loss function is minimized directly, "there is an increased risk of interaction among the control factors ...". One would think that the existence of interactions would be determined by the physical situation and not by the choice of function to be optimized.

Even if the S/N ratio is related to the quadratic loss function, however, it must be noted that the simple quadratic loss function is not always the correct function to use (Tribus and Szonyi, 1989, p. 50). For example, when deciding on the proper length of rope to span a river, it is much more expensive to choose a rope which is too short than one which is too long.

As observed previously, another shortcoming of the use of the S/N ratio is that it does not allow for multiple responses. These could be incorporated into a loss function which was to be optimized directly; for example, the loss functions for the individual responses, appropriately weighted, could simply be added together.

To sum up, one could say that it makes more sense to work with the item of interest itself (i.e., either the loss function or the quality characteristic) than an artificial construct.

It appears that the S/N ratio was introduced to simplify the solution. The problem is that in doing so, it oversimplifies the situation, leading potentially to erroneous results. It may also encourage practitioners to lose sight of their original purpose.

5.3   General Approach

The Taguchi method very nicely lays out a path to be followed in the quest for optimization. Unfortunately, it is difficult to chart a course at the outset, when one does not know the true destination. As George Box (N92) has pointed out, it is best to try to increase understanding of the process, rather than to try to optimize it without investigating the underlying structure. One is then more likely to find a true optimum rather than what might simply be a local optimum, and the cost of doing this is not necessarily any greater. As Box has eloquently stated (N92, p. 131):

> It seems that Taguchi's experimental strategy is intended only to pick the "optimum" factor combination from a one-shot experiment. Although the immediate objective may be this, the ultimate goal must surely be to better understand the engineering system. For example, appropriate designs can provide estimates of those *specific* interactions between environmental and design factors that cause lack of robustness. Once the engineer knows which these are and what they can do, he can employ his *engineering* know-how to suggest ways of compensating for them, eliminating them, or reducing them. Thus I profoundly disagree with the implications of Shin Taguchi's claim that the engineer does not need to "discover the causal relationships and to understand the mechanics of how things happen." To believe this is to discount the way the engineer thinks and the whole basis of his education and experience.

6 Conclusions

Not surprisingly, we must conclude that although the Taguchi method may be useful in some cases, it is, in general, not a sound optimization method. As delineated in Section 5, we have serious misgivings about both the experimental approach and the analytical procedure.

Even if we were convinced that the analytical steps were reasonable and attempted to apply them to the data which were collected in the ship hull experiment, we would first need to ensure that the data would fit into a Taguchi-style crossed array. Next, since we are interested in simultaneously optimizing a number of responses, we would be required to construct some sort of overall response function which could be forced into one of the three standard analysis situations. This would certainly lead to many of the attendant problems described in Section 5.2. Although it would certainly be interesting to see the results of the Taguchi method applied to the ship hull design problem, we favour more conventional alternatives, such as response surface methodology, for the purpose of finding a "true" optimum.

References for Part I

Barad, M., Bezalel, C., and Goldstein, J.R.(1989). Prior Research is the Key to Fractional Factorial Design. *Quality Progress*, **22**, **3**, 71–75.

Byrne, D.M., and Taguchi, S. (1987). The Taguchi Approach to Parameter Design. *Quality Progress*, **20**, **12**, 19–26.

Delampady, M., Kan, L., and Yee, I. (1989). Analysis of Seakeeping Data. Technical report submitted to the National Research Council of Canada.

Delampady, M., Gu, C., Ma, P., Meloche, J., and Molyneux, D. (1990). Optimal Hull Forms From Analysis of Seakeeping Data. Technical report submitted to the National Research Council of Canada.

Hendrix, C.D. (1991). Signal-to-Noise Ratios: A Wasted Effort. *Quality Progress*, **24**, **7**, 75–76.

Lloyd, A.R.J.M., and Andrew, R.N. (1977). Criteria for Ship Speed in Rough Weather. In *Proceedings of the 18th general meeting of American Towing Tank Conference*, Volume 2, Cavitation Session, Seakeeping Session, eds. B. Johnson and B. Nehrling, 541–565.

Murdey, D., and Butler, J. (1985). Hull form series for fast surface ships: summary of seakeeping data. Institute for Marine Dynamics, LTR-SH-393.

Nair, V.N., ed. (1992). Taguchi's Parameter Design: A Panel Discussion. *Technometrics*, **34**, **2**, 127–161.

Ross, P.J. (1988). The Role of Taguchi Methods and Design of Experiments in QFD. *Quality Progress*, **21**, **6**, 41–47.

Ryan, T.P. (1988). Taguchi's Approach to Experimental Design: Some Concerns. *Quality Progress*, **21**, **5**, 34–36.

Taguchi, G., Elsayed, E., and Hsiang, T. (1989). *Quality Engineering in Production Systems*. New York: McGraw-Hill.

Tribus, M., and Szonyi, G. (1989). An Alternative View of the Taguchi Approach. *Quality Progress*, **22**, **5**, 46–52.

Tunner, J. (1990). Is an Out-of-Spec Product Really Out of Spec? *Quality Progress*, **23**, **12**, 57–59.

Weerahandi, S., and Zidek, J. (1988). Bayesian nonparametric smoothers. *Canadian Journal of Statistics*, **16**, 61–74.

Part II

Endogamous Group Comparisons

Summary

This report developed from a case which came to the Statistical Consulting and Research Lab (SCARL) in the Department of Statistics at The University of British Columbia. Having lost all her original data, our client wished to know if it were possible to reconstruct her data or to perform a statistical analysis, using the tables of summary statistics in her possession.

She was interested in modelling several different pulmonary functions using various demographic and anthropometric measurements as explanatory variables. For each of the different model variants under consideration, her goal was to see if the same parameters would be valid for all endogamous groups[1], or if some groups were sufficiently different that the parameter values would be different for these groups.

Dr. Malcolm Greig of SCARL and I were able to develop an innovative method for the solution of this problem; the derivation and application of this methodology are described in this paper.

---

[1]social or tribal groups which do not permit marriage outside the group

7 Introduction

The current study began with a telephone enquiry from a potential client of the Statistical Consulting and Research Lab (SCARL) in the Department of Statistics at The University of British Columbia. She had lost all her original data through a series of misfortunes and wished to know if some sort of statistical analysis or reconstruction of the data were possible, since she had managed to retain certain tables of summary statistics of the data.

The client, Dr. Lakshmi Reddy, was interested in modelling pulmonary functions using various demographic and anthropometric measurements as explanatory variables. For each of a number of different model variants, her goal was to see if the same parameters would be valid for all endogamous groups, or if some groups were sufficiently different that the parameter values would be different for these groups.

In a seemingly hopeless situation, where most prospective analysts had already given up, Dr. Malcolm Greig of SCARL suggested a route to an innovative solution which the author has worked out in detail below. By making certain "working" assumptions and using a plausible but restricted model, a methodology for testing overall group homogeneity in the absence of the original test data is developed. The procedure uses only the information contained in tables of means, standard deviations, correlations, and group sizes. The author subsequently applied the method to our client's problem with rather interesting results, which are reported in the following sections.

8 Background

8.1 The Estimation of Physical Fitness

Dr. Reddy wished to determine whether the same parameters would be valid for different endogamous groups in models of pulmonary functions as a function of various demographic and anthropometric measurements. The motivation of her study was the well-established use of vital capacity estimates as a measure of physical fitness.

It is common in industry for job applicants to be given a physical examination and hired only if they are judged to be fit. Vital capacity is a common measure of fitness, but the equipment to measure vital capacity directly is often unavailable. As a result, it is routine for physicians to predict vital capacity based on physical measurements such as height and age, using charts designed for this purpose. The charts currently used in countries with predominantly Caucasian populations are, of course, based on Caucasians. Their application could result in the rejection of a fit non-Caucasian as unfit, simply because a fit Caucasian of the same height and age might have a larger vital capacity. In fact, it is recognized that for orientals, the Caucasian-based standards should be multiplied by 0.80. By using other variables in addition to, or instead of, age and height, Dr. Reddy wanted to develop a system for predicting vital capacity which could be applied universally.

As part of the research program for her Ph.D. in biological anthropology, Dr. Reddy had gathered data on a total of 1301 people in India from six different caste groups (since all caste groups in India are endogamous). These included workers from a moped factory near the town of Tirupapi in the state of Andhra Pradesh, as well as people from villages in the immediate countryside. The subjects represented a cross-section of people from various backgrounds, and included labourers, housewives, office workers, and professionals. The data were collected over a period of six months in 1980-81 by Dr. Reddy and two co-workers, a pulmonary physiologist and an anthropologist. Dr. Reddy intended to use these data to investigate her idea after the completion of her Ph.D. program.

The distribution of the subjects among the caste groups is shown in Table 3. The data

Caste	Men	Women
Brahmins	75	75
Kapu	275	220
Balija	153	128
Mala	63	99
Christians	28	11
Muslims	17	51
Totals	687	614

Table 3: Subject Distribution among Caste Groups

collected included measurements of three groups of variables. Henceforth we shall refer to these groups as Group A, Group B, and Group C; Table 4 lists the variables in these groupings. Group A consists primarily of pulmonary functions, Group B consists mainly of anthropometric measurements, and Group C is comprised solely of anthropometric indices. For each subject, the caste to which the subject belonged was also recorded.

For each sex within each endogamous group, Dr. Reddy originally proposed to model several of the pulmonary functions, namely vital capacity, forced vital capacity, and forced expiratory volume, using a selection of Group B variables as the explanatory variables. This would allow her to determine whether different model parameters were necessary for some groups. In addition, she intended to model these three pulmonary functions for men and for women, ignoring the groupings, in order to compare her overall model with other published results.

8.2 The Data Analysis Problem

Due to a set of unusual circumstances, all the original data were lost; however, a large number of summary statistics was retained, since the tables containing these statistics had been published in the original thesis (see Reddy, 1982). Dr. Reddy consulted SCARL to see if it was possible to reconstruct the data from these statistics, or if the statistics could somehow be used to do multiple linear regression. If it were possible to do regression, she was interested in checking versions of the linear model with some

Group A Variables	Group B Variables	Group C Variables
Tidal Volume	Age	Bi-acromial Breadth Index
Vital Capacity	Sitting Height Vertex	Relative Thoracic Index
Inspiratory Capacity	Height Vertex	Pondoral Index
Inspiratory Reserve Volume	Bi-acromial Breadth	Robusticity Index
Expiratory Capacity	Arm Span	Pignet Vervaek Index
Expiratory Reserve Volume	Chest Circumference (insp.)	Nutritional Index
Forced Vital Capacity	Chest Circumference (exp.)	Weight Index
Forced Expiratory Volume (FEV$_1$)	Body Weight	Chest Girth Index
Blood Pressure (diastolic)	Body Surface Area	Span Stature Index
Blood Pressure (systolic)	Chest Width	Sitting Height Stature Index
Pulse Rate	Bi-iliac Diameter	
Body Temperature	Trochanteric Diameter	
	Girth of Abdomen	
	Length of Sternum	
	Antero-posterior Chest Diam.	
	Chest Volume	
	Chest Expansion	
	Trunk Surface Area I	
	Trunk Surface Area II	

Table 4: Measurements Collected

of the Group C variables (anthropometric indices) as explanatory variables in addition to versions using the Group B variables, since the indices included some important anthropometric measurements for which separate data were not available.

In all, for each sex, she wanted to test six different versions of the basic linear model; these versions would be created by using different combinations of dependent and explanatory variables as detailed in Section 11. We shall refer to these different versions of our basic linear model as "model variants".

The following is a list of the available statistics.

38

For the entire sample:

- Range, mean, standard deviation, standard error, and coefficient of variation for:

  1. all variables grouped by sex

  2. the first eight group A variables grouped by sex and age group

  3. all group B variables grouped by sex and age group

  4. all group C variables grouped by sex and age group

- results from t-tests (including t-values) of:

  1. sex differences for mean values of all group A variables

  2. sex differences for mean values of all group B variables

  3. sex differences for mean values of all group C variables

  4. differences between smokers and non-smokers (men only)[2] for all group A and group B variables

- correlation coefficient estimates for each sex for:

  1. age versus all group A variables

  2. age versus all other group B variables

  3. age versus all group C variables

  4. the first eight group A variables versus all group B variables

  5. the first eight group A variables versus all group C variables

- inter-correlations of thirteen group B variables for each sex

- multiple regression coefficients for vital capacity as a function of 13 group B variables (including age and height) for each sex

---

[2]There were no data on female smokers, since they are extremely rare in India.

For each caste group:

- Range, mean, standard deviation, standard error, and coefficient of variation for all variables grouped by sex

Group Comparisons:

- t-test results (including t-values) by sex for each of the 15 possible pairwise comparisons between caste groups for:

    1. mean values of the first eight group A variables
    2. mean values of all group B variables
    3. mean values of all group C variables

In all the tables which showed t-test results, there were some results which were significant at the α-level of .05.

To investigate whether a useful model could be checked using the available statistics, we developed the Gaussian distribution based likelihood function which was consistent with the working assumption of sufficiency of the retained statistics.

9 Sufficient Statistics

We adopt the structural model:

$$y_{ij} = \mu_i + \boldsymbol{x}'_{T:ij}\boldsymbol{\beta} + \epsilon_{ij},$$

$$i = 1,\ldots,6;\ j = 1,\ldots,n_i,$$

where i = group number, j = observation number,

$$\boldsymbol{x}_{T:ij} = \begin{pmatrix} x_{ij}{}^{(1)} - \bar{x}..^{(1)} \\ \vdots \\ x_{ij}{}^{(q)} - \bar{x}..^{(q)} \\ \vdots \\ x_{ij}{}^{(Q)} - \bar{x}..^{(Q)} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta^{(1)} \\ \vdots \\ \beta^{(q)} \\ \vdots \\ \beta^{(Q)} \end{pmatrix},$$

and q = covariate number $(q = 1,\ldots,Q)$.

The stochastic modelling assumptions are

$$\epsilon_{ij} \sim \mathcal{N}(0,\ \sigma^2),$$

so that

$$y_{ij} \sim \mathcal{N}(\mu_i + \boldsymbol{x}'_{T:ij}\boldsymbol{\beta},\ \sigma^2).$$

In this model, the value of β is the same for all groups; in other words, we are assuming that each explanatory variable has the same coefficient for all six groups. The intercept term, μ_i, on the other hand, can have a different value for each group.

This model may seem unrealistic, but since we are only interested in whether one set of parameter values can describe all groups, this model can be used to test $\mu_i = \mu_j$ for all i,j such that $i \neq j$. If this hypothesis is rejected, we will conclude that the groups cannot be described by the same equation. Note that failure to reject the hypothesis would not imply that the same equation is appropriate for all groups, since with the available statistics, we cannot check to see whether β should vary from group to group.

9.1 Notation

We shall follow the convention of using uppercase letters for matrices, boldface lowercase letters for vectors, and unemphasized lowercase letters for scalar quantities.

Let $n = \sum_{i=1}^{6} n_i$, and let $\mathbf{1}_{(k)}$ denote a k-dimensional vector of 1's. In addition, let us define the following vectors and matrices:

$$\boldsymbol{\beta} = \left(\beta^{(1)}, \ldots, \beta^{(Q)}\right)'$$

$$\boldsymbol{\mu} = \left(\mu_1 \mathbf{1}'_{(n_1)}, \ldots, \mu_6 \mathbf{1}'_{(n_6)}\right)'$$

$$\boldsymbol{\epsilon}_i = \left(\epsilon_{i1}, \ldots, \epsilon_{in_i}\right)'$$

$$\boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_6'\right)'$$

$$\boldsymbol{x}_{T:ij} = \left(x_{ij}^{(1)} - \bar{x}_{..}^{(1)}, \ldots, x_{ij}^{(Q)} - \bar{x}_{..}^{(Q)}\right)'$$

$$X_i = \begin{bmatrix} \boldsymbol{x}'_{T:i1} \\ \vdots \\ \boldsymbol{x}'_{T:in_i} \end{bmatrix} = \begin{bmatrix} \left(x_{i1}^{(1)} - \bar{x}_{..}^{(1)}\right) & \cdots & \left(x_{i1}^{(Q)} - \bar{x}_{..}^{(Q)}\right) \\ \vdots & & \vdots \\ \left(x_{in_i}^{(1)} - \bar{x}_{..}^{(1)}\right) & \cdots & \left(x_{in_i}^{(Q)} - \bar{x}_{..}^{(Q)}\right) \end{bmatrix},$$

where X_i is a matrix with dimensions $n_i \times Q$

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_6 \end{bmatrix},$$

where X is a matrix with dimensions $n \times Q$

$$\boldsymbol{y}_i = \left(y_{i1}, \ldots, y_{in_i}\right)'$$

$$\boldsymbol{y} = \left(\boldsymbol{y}_1', \ldots, \boldsymbol{y}_6'\right)'$$

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} - \bar{y}_{..}\mathbf{1}_{(n)}.$$

9.2 Calculation of Sufficient Statistics

Using the notation defined in Section 9.1, our model can be expressed as:

$$y = \mu + X\beta + \epsilon.$$

Before finding the sufficient statistics for this model, we shall prove the following lemmas which the author has derived.

Lemma 1

$$\sum_{i=1}^{6} \sum_{j=1}^{n_i} (y_{ij} - \mu_i - x'_{T:ij}\beta)^2 = \tilde{y}'\tilde{y} + n\bar{y}_{..}^2 - 2y'\mu$$

$$+ 2\mu'X\beta + \mu'\mu - 2\tilde{y}'X\beta + \beta'X'X\beta.$$

Proof:

$$\sum_{i=1}^{6} \sum_{j=1}^{n_i} (y_{ij} - \mu_i - x'_{T:ij}\beta)^2 = \| y - \mu - X\beta \|^2$$

$$= (y - \mu - X\beta)'(y - \mu - X\beta)$$

$$= y'y + \mu'\mu + \beta'X'X\beta$$

$$- 2y'\mu - 2y'X\beta + 2\mu'X\beta$$

$$= y'y + \mu'\mu + \beta'X'X\beta$$

$$- 2y'\mu - 2y'X\beta + 2\bar{y}_{..}\mathbf{1}'_{(n)}X\beta + 2\mu'X\beta$$

since $\mathbf{1}'_{(n)}X = [0]$.

Thus

$$\sum_{i=1}^{6} \sum_{j=1}^{n_i} (y_{ij} - \mu_i - x'_{T:ij}\beta)^2 = y'y + \mu'\mu + \beta'X'X\beta$$

$$- 2y'\mu - 2\tilde{y}'X\beta + 2\mu'X\beta$$

43

$$= \tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}} + n\bar{y}_{..}^2 + \boldsymbol{\mu}'\boldsymbol{\mu} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}$$

$$- 2\boldsymbol{y}'\boldsymbol{\mu} - 2\tilde{\boldsymbol{y}}'X\boldsymbol{\beta} + 2\boldsymbol{\mu}'X\boldsymbol{\beta}$$

since $\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}} = \sum_{i=1}^{6}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{6}\sum_{j=1}^{n_i}y_{ij}^2 - n\bar{y}_{..}^2 = \boldsymbol{y}'\boldsymbol{y} - n\bar{y}_{..}^2$ ∎

Lemma 2

$$\boldsymbol{y}'\boldsymbol{\mu} = \sum_{i=1}^{6}\mu_i n_i \bar{y}_{i..}$$

Proof:

$$\boldsymbol{y}'\boldsymbol{\mu} = \boldsymbol{y}_1'\mu_1 \mathbf{1}_{(n_1)} + \boldsymbol{y}_2'\mu_2 \mathbf{1}_{(n_2)} + \ldots + \boldsymbol{y}_6'\mu_6 \mathbf{1}_{(n_6)}$$

$$= \mu_1 \boldsymbol{y}_1'\mathbf{1}_{(n_1)} + \mu_2 \boldsymbol{y}_2'\mathbf{1}_{(n_2)} + \ldots + \mu_6 \boldsymbol{y}_6'\mathbf{1}_{(n_6)}$$

$$= \mu_1 n_1 \bar{y}_{1.} + \mu_2 n_2 \bar{y}_{2.} + \ldots + \mu_6 n_6 \bar{y}_{6.}$$

$$= \sum_{i=1}^{6}\mu_i n_i \bar{y}_{i.} \quad ∎$$

Now we can proceed to find the sufficient statistics for our model. The unknown parameter $\Theta = (\boldsymbol{\mu}, \boldsymbol{\beta}, \sigma^2)$ and

$$f(y_{ij}|\Theta) = \frac{1}{\sigma\sqrt{2\pi}}exp\{-\frac{1}{2\sigma^2}(y_{ij} - \mu_i - \boldsymbol{x}_{T:ij}'\boldsymbol{\beta})^2\}.$$

By using the result of Lemma 1, we find that

$$f(\boldsymbol{y}|\Theta) = (2\pi\sigma^2)^{-n/2}exp\{\sum_{i=1}^{6}\sum_{j=1}^{n_i} -\frac{1}{2\sigma^2}(y_{ij} - \mu_i - \boldsymbol{x}_{T:ij}'\boldsymbol{\beta})^2\}$$

$$= (2\pi\sigma^2)^{-n/2}exp\{-\frac{1}{2\sigma^2}[\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}} + n\bar{y}_{..}^2 + \boldsymbol{\mu}'\boldsymbol{\mu} + \boldsymbol{\beta}'X'X\boldsymbol{\beta}$$

$$- 2\boldsymbol{y}'\boldsymbol{\mu} - 2\tilde{\boldsymbol{y}}'X\boldsymbol{\beta} + 2\boldsymbol{\mu}'X\boldsymbol{\beta}]\}.$$

44

Therefore,

$$
\begin{aligned}
f(\boldsymbol{y}|\Theta) &= exp\{-\frac{n}{2}log\,2\pi\sigma^2 - \frac{1}{2\sigma^2}\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}} - \frac{1}{2\sigma^2}n\bar{y}_{..}^2 - \frac{1}{2\sigma^2}\boldsymbol{\mu}'\boldsymbol{\mu} \\
&\quad - \frac{1}{2\sigma^2}\boldsymbol{\beta}'X'X\boldsymbol{\beta} + \frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{\mu} + \frac{1}{\sigma^2}\tilde{\boldsymbol{y}}'X\boldsymbol{\beta} - \frac{1}{\sigma^2}\boldsymbol{\mu}'X\boldsymbol{\beta}\} \\
&= exp\{-\frac{1}{2\sigma^2}\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}} - \frac{1}{2\sigma^2}n\bar{y}_{..}^2 + \frac{1}{\sigma^2}\sum_{i=1}^{6}\mu_i n_i\bar{y}_{i.} + \frac{1}{\sigma^2}\tilde{\boldsymbol{y}}'X\boldsymbol{\beta} \\
&\quad - \frac{n}{2}log\,2\pi\sigma^2 - \frac{1}{\sigma^2}\boldsymbol{\mu}'X\boldsymbol{\beta} - \frac{1}{2\sigma^2}\boldsymbol{\mu}'\boldsymbol{\mu} - \frac{1}{2\sigma^2}\boldsymbol{\beta}'X'X\boldsymbol{\beta}\},
\end{aligned}
$$

where we have substituted the result of Lemma 2 in the last equation.

We find from this expression that $(\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}};\ \bar{y}_{..}^2;\ \bar{y}_{i.},\ i = 1,\ldots,6;\ \tilde{\boldsymbol{y}}'X)$ are sufficient statistics for the family of distributions indexed by Θ.

Now we must see whether these statistics can be obtained from what we have at our disposal.

As noted in Section 8.2, a number of statistics are available, and we wish to check six different model variants for each sex. For simplicity, assume for now that we wish to test one of the model variants for males, where y is vital capacity, for example, and we have Q explanatory variables (x's). From the tables at our disposal, we can construct the following matrices for this model variant. The subscripts, which will be dropped in later references, indicate the dimensions of each matrix. These matrices are:

$R_{(Q+1)\times(Q+1)}$ the correlation matrix for the x's and y for all males;

$S_{\substack{diagonal \\ (Q+1)\times(Q+1)}}$ the matrix of standard deviations of the x's and y for all males;

$M_{6\times(Q+1)}$ the matrix of means of the x's and y for each group;

$N_{\substack{diagonal \\ 6\times6}}$ the matrix of the group sizes.

45

From these matrices, we calculate the sufficient statistics as follows. First, $\bar{y}_{i.}$, $i = 1, \ldots, 6$ is obtained directly from M. Now, let $n = tr(N)$. To find $\bar{y}_{..}$, calculate

$$\frac{\left(\mathbf{1}'_{(6)}NM\right)'}{n} = \begin{pmatrix} \bar{x}_{..}^{(1)} \\ \vdots \\ \bar{x}_{..}^{(Q)} \\ \bar{y}_{..} \end{pmatrix} .$$

Next, calculate $(n-1)SRS$, which can be partitioned thus:

$$\begin{bmatrix} X'X & X'\tilde{y} \\ \tilde{y}'X & \tilde{y}'\tilde{y} \end{bmatrix} .$$

This gives us $\tilde{y}'\tilde{y}$ and $\tilde{y}'X$.

We have shown here that we do indeed have sufficient statistics for the model family described at the beginning of Section 9. In the next section, we shall develop an appropriate testing procedure for the substantive problem addressed by this work.

10 Derivation of the Method of Analysis

Since our goal is to compare regression model intercepts, that is, the values of μ_i, we must calculate their estimates as well as estimates of their expected values and variances.

10.1 Notation

Let us define the scalars, vectors and matrices we require, using the conventions established in Section 9.1 (β and $x_{T:ij}$ are as defined previously in Section 9.1. They are included below for completeness.):

$$\beta = (\beta^{(1)}, \ldots, \beta^{(Q)})'$$

$$x_{T:ij} = (x_{ij}^{(1)} - \bar{x}_{..}^{(1)}, \ldots, x_{ij}^{(Q)} - \bar{x}_{..}^{(Q)})'$$

$$x_{G:i} = (\bar{x}_{i.}^{(1)} - \bar{x}_{..}^{(1)}, \ldots, \bar{x}_{i.}^{(Q)} - \bar{x}_{..}^{(Q)})'$$

$$x_{W:ij} = (x_{ij}^{(1)} - \bar{x}_{i.}^{(1)}, \ldots, x_{ij}^{(Q)} - \bar{x}_{i.}^{(Q)})'$$

$$\tilde{x}_{ij}^{(q)} = x_{ij}^{(q)} - \bar{x}_{i.}^{(q)}$$

$$\tilde{y}_{ij} = y_{ij} - \bar{y}_{i.}$$

$$\tilde{\epsilon}_{ij} = \epsilon_{ij} - \bar{\epsilon}_{i.}$$

$$c_{W:yx} = \sum_{i=1}^{6}\sum_{j=1}^{n_i} \tilde{y}_{ij} x_{W:ij} \qquad c_{W:\epsilon x} = \sum_{i=1}^{6}\sum_{j=1}^{n_i} \tilde{\epsilon}_{ij} x_{W:ij} \qquad C_{W:xx} = \sum_{i=1}^{6}\sum_{j=1}^{n_i} x_{W:ij} x'_{W:ij}$$

where $C_{W:xx}$ is a symmetric matrix with dimensions $Q \times Q$. In addition, we define the scalar

$$c_{W:yy} = \sum_{i=1}^{6}\sum_{j=1}^{n_i} (\tilde{y}_{ij})^2.$$

47

10.2 Derivation of Estimators

As described previously in Section 9, our model is:

$$y_{ij} = \mu_i + \boldsymbol{x}'_{T:ij}\boldsymbol{\beta} + \epsilon_{ij},$$

$$i = 1,\ldots,6; \; j = 1,\ldots,n_i,$$

where i = group number, and j = observation number. We find the least squares parameter estimators by minimizing the residual sum of squares (RSS) with respect to the unknown model parameters. This leads to the following estimating equations:

$$\frac{\partial RSS}{\partial \mu_i} = -2\sum_{j=1}^{n_i}(y_{ij} - \mu_i - \boldsymbol{x}'_{T:ij}\boldsymbol{\beta}) = 0,$$

$$\frac{\partial RSS}{\partial \beta^{(q)}} = -2\sum_{i=1}^{6}\sum_{j=1}^{n_i}(\tilde{y}_{ij} - \boldsymbol{x}'_{W:ij}\boldsymbol{\beta})\,\tilde{x}_{ij}^{(q)} = 0,$$

where

$$RSS = \sum_{i=1}^{6}\sum_{j=1}^{n_i}(y_{ij} - \mu_i - \boldsymbol{x}'_{T:ij}\boldsymbol{\beta})^2.$$

Now, $\partial RSS/\partial\mu_i = 0$ entails $\sum_{j=1}^{n_i}(y_{ij} - \mu_i - \boldsymbol{x}'_{T:ij}\boldsymbol{\beta}) = 0$, implying $y_{i\cdot} - \sum_{j=1}^{n_i}\boldsymbol{x}'_{T:ij}\boldsymbol{\beta} = n_i\mu_i$. From this equation, we obtain

$$\hat{\mu}_i = \bar{y}_{i\cdot} - \boldsymbol{x}'_{G:i}\hat{\boldsymbol{\beta}}. \tag{1}$$

It follows that

$$RSS = \sum_{i=1}^{6}\sum_{j=1}^{n_i}[\,y_{ij} - (\bar{y}_{i\cdot} - \boldsymbol{x}'_{G:i}\boldsymbol{\beta}) - \boldsymbol{x}'_{T:ij}\boldsymbol{\beta}\,]^2 = \sum_{i=1}^{6}\sum_{j=1}^{n_i}(\tilde{y}_{ij} - \boldsymbol{x}'_{W:ij}\boldsymbol{\beta})^2.$$

We then find that $\partial RSS/\partial\beta^{(q)} = 0$ implies

$$\sum_{i=1}^{6}\sum_{j=1}^{n_i}\tilde{y}_{ij}\,\tilde{x}_{ij}^{(q)} = \sum_{i=1}^{6}\sum_{j=1}^{n_i}\tilde{x}_{ij}^{(q)}\,\boldsymbol{x}'_{W:ij}\boldsymbol{\beta}$$

for every coordinate q. This gives us the equation $\boldsymbol{c}_{W:yx} = C_{W:xx}\hat{\boldsymbol{\beta}}$. Thus

$$\hat{\boldsymbol{\beta}} = C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}. \tag{2}$$

Now, since

$$y_{ij} = \mu_i + \boldsymbol{x}'_{T:ij}\boldsymbol{\beta} + \epsilon_{ij}, \qquad (3)$$

and

$$\bar{y}_{i.} = \mu_i + \boldsymbol{x}'_{G:i}\boldsymbol{\beta} + \bar{\epsilon}_{i.}, \qquad (4)$$

we may subtract Equation (4) from Equation (3), giving $\tilde{y}_{ij} = \boldsymbol{x}'_{W:ij}\boldsymbol{\beta} + \tilde{\epsilon}_{ij}$. Then $\boldsymbol{c}_{W:yx} = C_{W:xx}\boldsymbol{\beta} + \boldsymbol{c}_{W:\epsilon x}$. Thus, from Equation (2), $\hat{\boldsymbol{\beta}} = C_{W:xx}^{-1}\boldsymbol{c}_{W:yx} = C_{W:xx}^{-1}(C_{W:xx}\boldsymbol{\beta} + \boldsymbol{c}_{W:\epsilon x})$. We conclude that

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x}. \qquad (5)$$

Returning to Equation (1), we obtain $\hat{\mu}_i = \bar{y}_{i.} - \boldsymbol{x}'_{G:i}\hat{\boldsymbol{\beta}}$. On substituting Equation (4) into Equation (1), we get $\hat{\mu}_i = (\mu_i + \boldsymbol{x}'_{G:i}\boldsymbol{\beta} + \bar{\epsilon}_{i.}) - \boldsymbol{x}'_{G:i}\hat{\boldsymbol{\beta}}$. Equation (5) then gives $\hat{\mu}_i = \mu_i + \boldsymbol{x}'_{G:i}\boldsymbol{\beta} + \bar{\epsilon}_{i.} - \boldsymbol{x}'_{G:i}(\boldsymbol{\beta} + C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x})$. Finally,

$$\hat{\mu}_i = \mu_i + \bar{\epsilon}_{i.} - \boldsymbol{x}'_{G:i}C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x}. \qquad (6)$$

Equation (6) gives us $E(\hat{\mu}_i) = \mu_i$; $\hat{\mu}_i$ is unbiased. Next, we need to find the variance of $\hat{\mu}_i$. To that end, we find from Equation (6) that $\hat{\mu}_i - \mu_i = \bar{\epsilon}_{i.} - \boldsymbol{x}'_{G:i}C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x}$. So

$$
\begin{aligned}
Var(\hat{\mu}_i) &= Var(\hat{\mu}_i - \mu_i) \\
&= Var(\bar{\epsilon}_{i.} - \boldsymbol{x}'_{G:i}C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x}) \\
&= Var(\boldsymbol{x}'_{G:i}C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x}) - 2\,Cov(\boldsymbol{x}'_{G:i}C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x},\,\bar{\epsilon}_{i.}) + Var(\bar{\epsilon}_{i.}).
\end{aligned}
$$

Thus,

$$Var(\hat{\mu}_i) = \boldsymbol{x}'_{G:i}C_{W:xx}^{-1}Cov(\boldsymbol{c}_{W:\epsilon x})C_{W:xx}^{-1}\boldsymbol{x}_{G:i} - 2\,Cov(\boldsymbol{x}'_{G:i}C_{W:xx}^{-1}\boldsymbol{c}_{W:\epsilon x},\,\bar{\epsilon}_{i.}) + Var(\bar{\epsilon}_{i.}). \qquad (7)$$

To simplify this expression, we need the following lemmas, derived by Dr. Greig and expanded by the author.

Lemma 3

$$Cov(\mathbf{x}'_{G:i} C^{-1}_{W:xx} \mathbf{c}_{W:ex}, \bar{\epsilon}_{i.}) = 0.$$

Proof:

For simplicity, let $\mathbf{x} = \mathbf{c}_{W:ex}$, $y = \bar{\epsilon}_{i.}$, and $\mathbf{a}' = \mathbf{x}'_{G:i} C^{-1}_{W:xx}$. Note that $\mathbf{a}'$ is a $1 \times Q$ vector. Then

$$
\begin{aligned}
Cov(\mathbf{x}'_{G:i} C^{-1}_{W:xx} \mathbf{c}_{W:ex}, \bar{\epsilon}_{i.}) &= Cov(\mathbf{a}'\mathbf{x}, y) \\
&= Cov(\sum_{q=1}^{Q} a_q x_q, y) \\
&= \sum_{q=1}^{Q} Cov(a_q x_q, y) \\
&= \sum_{q=1}^{Q} a_q Cov(x_q, y).
\end{aligned}
$$

But

$$
\begin{aligned}
Cov(x_q, y) &= Cov[(\sum_{i=1}^{6} \sum_{j=1}^{n_i} \tilde{\epsilon}_{ij} \tilde{x}_{ij}^{(q)}), \bar{\epsilon}_{i.}] \\
&= \sum_{i=1}^{6} \sum_{j=1}^{n_i} \tilde{x}_{ij}^{(q)} Cov(\tilde{\epsilon}_{ij}, \bar{\epsilon}_{i.}).
\end{aligned}
$$

Now, since

$$Cov(\tilde{\epsilon}_{ij}, \bar{\epsilon}_{i.}) = Cov(\epsilon_{ij}, 1/n_i \sum_{k=1}^{n_i} \epsilon_{ik}) - Var(\bar{\epsilon}_{i.}),$$

we find that, due to independence of the ϵ_{ij}'s, $Cov(\tilde{\epsilon}_{ij}, \bar{\epsilon}_{i.}) = \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n_i} = 0$ for all q. It follows that $\sum_{q=1}^{Q} a_q Cov(x_q, y) = 0$. ∎

Lemma 4

$$Cov(\boldsymbol{c}_{W:\epsilon x}) = \sigma^2 C_{W:xx}.$$

Proof:

Let $c_{W:\epsilon x}^{(q)}$ denote the qth element of $\boldsymbol{c}_{W:\epsilon x}$, and $C_{W:xx}^{(pq)}$ the entry in the pth row and q^{th} column (and the qth row and p^{th} column) of $C_{W:xx}$. Then

$$E(c_{W:\epsilon x}^{(q)}) \;=\; E\left(\sum_{i=1}^{6}\sum_{j=1}^{n_i} \tilde{\epsilon}_{ij}\tilde{x}_{ij}^{\,(q)}\right)$$

$$=\; \sum_{i=1}^{6}\sum_{j=1}^{n_i} \tilde{x}_{ij}^{\,(q)} E(\tilde{\epsilon}_{ij})$$

$$=\; 0.$$

Thus $Var(c_{W:\epsilon x}^{(q)}) = E[(c_{W:\epsilon x}^{(q)})^2]$ and $Cov(c_{W:\epsilon x}^{(q)}, c_{W:\epsilon x}^{(p)}) = E(c_{W:\epsilon x}^{(q)} c_{W:\epsilon x}^{(p)})$. So $Cov(\boldsymbol{c}_{W:\epsilon x}) = E(\boldsymbol{c}_{W:\epsilon x}\boldsymbol{c}_{W:\epsilon x}')$. Now,

$$c_{W:\epsilon x}^{(q)} = \sum_{i=1}^{6}\sum_{j=1}^{n_i} \left(\epsilon_{ij}\tilde{x}_{ij}^{\,(q)}\right)$$

since

$$\sum_{i=1}^{6}\sum_{j=1}^{n_i} \left(\bar{\epsilon}_{i.}\tilde{x}_{ij}^{\,(q)}\right) = \sum_{i=1}^{6}\bar{\epsilon}_{i.}\sum_{j=1}^{n_i} \tilde{x}_{ij}^{\,(q)} = 0.$$

Thus

$$E(c_{W:\epsilon x}^{(q)} c_{W:\epsilon x}^{(p)}) \;=\; E\left[\left(\sum_{i=1}^{6}\sum_{j=1}^{n_i} \epsilon_{ij}\tilde{x}_{ij}^{\,(q)}\right)\cdot\left(\sum_{k=1}^{6}\sum_{l=1}^{n_k}\epsilon_{kl}\tilde{x}_{kl}^{\,(p)}\right)\right]$$

$$=\; E\left[\sum_{i=1}^{6}\sum_{j=1}^{n_i}(\epsilon_{ij})^2\,\tilde{x}_{ij}^{\,(q)}\tilde{x}_{ij}^{\,(p)}\right] + E\left[\sum_{i\neq k}\sum_{j=1}^{n_i}\sum_{l=1}^{n_k}(\epsilon_{ij}\epsilon_{kl})\,\tilde{x}_{ij}^{\,(q)}\tilde{x}_{kl}^{\,(p)}\right]$$

$$+ E\left[\sum_{i=1}^{6}\sum_{j\neq l}(\epsilon_{ij}\epsilon_{il})\,\tilde{x}_{ij}^{\,(q)}\tilde{x}_{il}^{\,(p)}\right]$$

$$=\; \sum_{i=1}^{6}\sum_{j=1}^{n_i}\tilde{x}_{ij}^{\,(q)}\tilde{x}_{ij}^{\,(p)}E[(\epsilon_{ij})^2]$$

since expectation of the last two terms is zero due to independence of the ϵ_{ij}'s. It follows that $E(c_{W:\epsilon x}^{(q)} c_{W:\epsilon x}^{(p)}) = \sigma^2 C_{W:xx}^{(pq)}$, and in turn that $E(\boldsymbol{c}_{W:\epsilon x}\boldsymbol{c}_{W:\epsilon x}') = \sigma^2 C_{W:xx}$. Thus $Cov(\boldsymbol{c}_{W:\epsilon x}) = \sigma^2 C_{W:xx}$. ∎

Now, substituting the results of Lemmas 3 and 4 into Equation (7), we get

$$
\begin{aligned}
Var(\hat{\mu}_i) &= \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \sigma^2 C_{W:xx} C_{W:xx}^{-1} \boldsymbol{x}_{G:i} + \frac{\sigma^2}{n_i} \\
&= \sigma^2 \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{x}_{G:i} + \frac{\sigma^2}{n_i}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
Cov(\hat{\mu}_i, \hat{\mu}_j) &= Cov(\hat{\mu}_i - \mu_i , \hat{\mu}_j - \mu_j) \\[2mm]
&= Cov[(\bar{\epsilon}_{i.} - \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x}) , (\bar{\epsilon}_{j.} - \boldsymbol{x}_{G:j}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x})] \\[2mm]
&= Cov[(\boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x}) , (\boldsymbol{x}_{G:j}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x})] \\[2mm]
&= E(\boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x} \boldsymbol{x}_{G:j}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x}) \\[2mm]
&= E(\boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{c}_{W:\epsilon x} \boldsymbol{c}_{W:\epsilon x}' C_{W:xx}^{-1} \boldsymbol{x}_{G:j}) \\[2mm]
&= \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} Cov(\boldsymbol{c}_{W:\epsilon x}) C_{W:xx}^{-1} \boldsymbol{x}_{G:j} \\[2mm]
&= \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \sigma^2 C_{W:xx} C_{W:xx}^{-1} \boldsymbol{x}_{G:j} \\[2mm]
&= \sigma^2 \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{x}_{G:j}.
\end{aligned}
$$

Then

$$
\begin{aligned}
Var(\hat{\mu}_i - \hat{\mu}_j) &= Var(\hat{\mu}_i) + Var(\hat{\mu}_j) - 2\, Cov(\hat{\mu}_i, \hat{\mu}_j) \\[2mm]
&= (\sigma^2 \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{x}_{G:i} + \frac{\sigma^2}{n_i}) + (\sigma^2 \boldsymbol{x}_{G:j}' C_{W:xx}^{-1} \boldsymbol{x}_{G:j} + \frac{\sigma^2}{n_j}) - 2(\sigma^2 \boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{x}_{G:j}) \\[2mm]
&= \sigma^2 (\boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{x}_{G:i} + \frac{1}{n_i} + \boldsymbol{x}_{G:j}' C_{W:xx}^{-1} \boldsymbol{x}_{G:j} + \frac{1}{n_j} - 2\boldsymbol{x}_{G:i}' C_{W:xx}^{-1} \boldsymbol{x}_{G:j}) \\[2mm]
&= \sigma^2 [\frac{1}{n_i} + \frac{1}{n_j} + (\boldsymbol{x}_{G:i} - \boldsymbol{x}_{G:j})' C_{W:xx}^{-1} (\boldsymbol{x}_{G:i} - \boldsymbol{x}_{G:j})].
\end{aligned}
$$

Since σ^2 is unknown, we must estimate it. Let f denote the number of degrees of freedom in our model variant. This gives us $f = 6 + Q$. Then

$$\hat{\sigma}^2 = \frac{1}{n-f}\sum_{i=1}^{6}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_{ij})^2$$

$$= \frac{1}{n-f}\sum_{i=1}^{6}\sum_{j=1}^{n_i}[y_{ij}-(\bar{y}_{i.}-\boldsymbol{x}_{G:i}'\hat{\boldsymbol{\beta}})-\boldsymbol{x}_{T:ij}'\hat{\boldsymbol{\beta}}]^2$$

$$= \frac{1}{n-f}\sum_{i=1}^{6}\sum_{j=1}^{n_i}(\tilde{y}_{ij}-\boldsymbol{x}_{W:ij}'\hat{\boldsymbol{\beta}})^2$$

$$= \frac{1}{n-f}[\sum_{i=1}^{6}\sum_{j=1}^{n_i}(\tilde{y}_{ij})^2$$

$$-2\sum_{i=1}^{6}\sum_{j=1}^{n_i}\tilde{y}_{ij}\boldsymbol{x}_{W:ij}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}$$

$$+\sum_{i=1}^{6}\sum_{j=1}^{n_i}\boldsymbol{x}_{W:ij}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}\boldsymbol{x}_{W:ij}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}].$$

The last equation is obtained by substituting Equation (2). Therefore,

$$\hat{\sigma}^2 = \frac{1}{n-f}(c_{W:yy}-2\,\boldsymbol{c}_{W:yx}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}$$

$$+\sum_{i=1}^{6}\sum_{j=1}^{n_i}\boldsymbol{c}_{W:yx}'C_{W:xx}^{-1}\boldsymbol{x}_{W:ij}\boldsymbol{x}_{W:ij}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx})$$

$$= \frac{1}{n-f}(c_{W:yy}-2\,\boldsymbol{c}_{W:yx}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}$$

$$+\boldsymbol{c}_{W:yx}'C_{W:xx}^{-1}C_{W:xx}C_{W:xx}^{-1}\boldsymbol{c}_{W:yx})$$

since $\sum_{i=1}^{6}\sum_{j=1}^{n_i}\boldsymbol{x}_{W:ij}\boldsymbol{x}_{W:ij}' = C_{W:xx}$. Finally,

$$\hat{\sigma}^2 = \frac{1}{n-f}(c_{W:yy}-\boldsymbol{c}_{W:yx}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}). \tag{8}$$

10.2.1 A Restricted Model

In this section we consider the model of Section 9 with the additional restriction that $\mu_i = \mu$ for all i:

$$y_{ij} = \mu + \boldsymbol{x}_{T:ij}'\boldsymbol{\beta}_0 + \epsilon_{ij}.$$

To distinguish the estimators from the two models, our stochastic modelling assumptions in this case are

$$\epsilon_{ij} \sim \mathcal{N}(0,\, \sigma_0^2).$$

Using the notation defined in Section 9.1 and proceeding in the manner of the derivation of Equations 1, 2, and 8 in Section 10.2, we find that

$$\begin{aligned}
\hat{\mu} &= \bar{y}.. \\
\hat{\beta}_0 &= (\boldsymbol{x}_{T:ij}'\boldsymbol{x}_{T:ij})^{-1}\boldsymbol{x}_{T:ij}'\boldsymbol{y}
\end{aligned}$$

and

$$\hat{\sigma}_0^2 = [\tilde{\boldsymbol{y}}'\tilde{\boldsymbol{y}} - \boldsymbol{y}'\boldsymbol{x}_{T:ij}(\boldsymbol{x}_{T:ij}'\boldsymbol{x}_{T:ij})^{-1}\boldsymbol{x}_{T:ij}'\boldsymbol{y}]/(n - Q - 1). \tag{9}$$

We shall refer to this model in Section 10.4 when we are considering hypothesis tests.

10.3 Calculation of Estimates

Returning to our less restricted model, and using $\hat{\sigma}^2$ in place of σ^2, we find that the estimators needed for our analysis are:

$$\begin{aligned}
\hat{\mu}_i &= \bar{y}_{i.} - \boldsymbol{x}_{G:i}'\hat{\boldsymbol{\beta}} \\
\hat{\boldsymbol{\beta}} &= C_{W:xx}^{-1}\boldsymbol{c}_{W:yx}.
\end{aligned}$$

For these estimators,

$$\widehat{Var}(\hat{\mu}_i) = \hat{\sigma}^2\boldsymbol{x}_{G:i}'C_{W:xx}^{-1}\boldsymbol{x}_{G:i} + \frac{\hat{\sigma}^2}{n_i},$$

and

$$\widehat{Cov}(\hat{\mu}_i, \hat{\mu}_j) = \hat{\sigma}^2\boldsymbol{x}_{G:i}'C_{W:xx}^{-1}\boldsymbol{x}_{G:j},$$

where $\hat{\sigma}^2 = \frac{1}{n-f}(c_{W:yy} - \boldsymbol{c}_{W:yx}'C_{W:xx}^{-1}\boldsymbol{c}_{W:yx})$.

To find the estimates, we need to calculate the following quantities: $\bar{y}_{i\cdot}$; n_i; $\boldsymbol{x}_{G:i}$; $C_{W:xx}$; $\boldsymbol{c}_{W:yx}$; and $c_{W:yy}$.

As in Section 9.2, we seek to test one of the model variants for males, and construct the following matrices for this variant. We repeat the matrices here for convenience:

$R_{(Q+1)\times(Q+1)}$ the correlation matrix for x's and y for all males;

$S_{\substack{diagonal\\(Q+1)\times(Q+1)}}$ the matrix of standard deviations of x's and y for all males;

$M_{6\times(Q+1)}$ the matrix of means of x's and y for each group;

$N_{\substack{diagonal\\6\times6}}$ the matrix of group sizes.

From these matrices, we can calculate the statistics we require. First, $\bar{y}_{i\cdot}$ and n_i are obtained directly from M and N respectively. Now, let $n = tr(N)$. To find $\boldsymbol{x}_{G:i}$, calculate the augmented matrix

$$G_{6\times(Q+1)} = M - \frac{1_{6\times6}NM}{n} = \begin{bmatrix} \boldsymbol{x}_{G:1}' & (\bar{y}_1 - \bar{y}_{\cdot\cdot}) \\ \boldsymbol{x}_{G:2}' & (\bar{y}_2 - \bar{y}_{\cdot\cdot}) \\ \boldsymbol{x}_{G:3}' & (\bar{y}_3 - \bar{y}_{\cdot\cdot}) \\ \boldsymbol{x}_{G:4}' & (\bar{y}_4 - \bar{y}_{\cdot\cdot}) \\ \boldsymbol{x}_{G:5}' & (\bar{y}_5 - \bar{y}_{\cdot\cdot}) \\ \boldsymbol{x}_{G:6}' & (\bar{y}_6 - \bar{y}_{\cdot\cdot}) \end{bmatrix}.$$

Next, let $C_T = (n-1)SRS$ = the total sum of squares matrix, and let $C_B = G'NG$ = the matrix of between group sums of squares and cross-products. Then $C_W = C_T - C_B$ = the matrix of within group sums of squares and cross-products. Note that C_W is a $(Q+1) \times (Q+1)$ matrix.

It can be partitioned thus:

$$\begin{bmatrix} C_{W:xx} & \boldsymbol{c}_{W:yx} \\ \\ \boldsymbol{c}_{W:yx}' & c_{W:yy} \end{bmatrix}.$$

This gives us $C_{W:xx}$, $\boldsymbol{c}_{W:yx}$, and $c_{W:yy}$.

55

Note that, for computational convenience, in the actual construction of the matrices for each sex, we included all the x's and y's which were to be considered in any of the six model variants for that sex. Then, when calculating the estimates for a particular variant, we simply selected the appropriate rows and columns from the matrices.

10.4   Testing Methodology

Once we have calculated the necessary estimates for each of the twelve model variants we propose to check, we wish to test (for each variant) whether the intercepts of the regression planes for all six groups are the same, i.e. our null hypothesis is:

$$H_0 : \mu_1 = \mu_2 \ldots = \mu_6.$$

To test this hypothesis we could perform the F-test, which some would consider the classic test. To implement it, we let

$$S = \frac{(n - Q - 1)\hat{\sigma}_0^2 - (n - Q - 6)\hat{\sigma}^2}{5\hat{\sigma}^2},$$

where $\hat{\sigma}^2$ and $\hat{\sigma}_0^2$ are given by Equations 8 and 9 respectively. Under the assumptions of the restricted model described in Section 10.2.1, the statistic S has an $F_{5,n-Q-6}$ distribution and may be used as a test of the null hypothesis H_0 (see Graybill).

However, in the event that there was not equality of μ_i's, Dr. Reddy wished to be able to say which pairs of μ_i's differed. We therefore propose to test H_0 by using fifteen pairwise tests, and adopting the hypothesis

$$H'_0 : \mu_i = \mu_j, \, i = 1, \ldots, 6; \, j = 1, \ldots, 6; \, i \neq j.$$

Since under the assumption of normality,

$$\frac{\hat{\mu}_i - \hat{\mu}_j}{\sqrt{Var(\hat{\mu}_i - \hat{\mu}_j)}} \text{ has a Student's } t \text{ distribution,}$$

we utilize this as our test statistic. We calculate this statistic for each of the pairwise tests, and obtain the p-values for these statistics using the t distribution. Let $p_1, \ldots, p_{15}$ be the fifteen p-values which result from these tests as shown in Table 5.

Groups Compared	p-value
1 versus 2	p_1
1 versus 3	p_2
1 versus 4	p_3
1 versus 5	p_4
1 versus 6	p_5
2 versus 3	p_6
2 versus 4	p_7
2 versus 5	p_8
2 versus 6	p_9
3 versus 4	p_{10}
3 versus 5	p_{11}
3 versus 6	p_{12}
4 versus 5	p_{13}
4 versus 6	p_{14}
5 versus 6	p_{15}

Table 5: Pairwise Comparisons for Testing Equality of μ_i's

If we were simply testing to see whether two particular groups have the same intercept, with an α-level of .05, we would just check to see if the p-value resulting from that particular pairwise comparison were less than .05. If it were, we would reject the null hypothesis of equality. Note that this would be equivalent to checking only one of the fifteen paired tests included in H_0'. In this case, however, we decided, in consultation with Dr. Reddy, to test H_0' (rather than H_0) with an α-level of 0.05.

We can do this by using the modified Bonferroni test known as Holm's sequentially rejective Bonferroni procedure. A brief development of Holm's sequentially rejective Bonferroni procedure follows; for more complete details see Hochberg and Tamhane (1987).

The standard Bonferroni test is based on the Bonferroni inequality, which states that $P(R_1 \cup R_2 \cup \ldots R_n) \leq \sum_{i=1}^{n} P(R_i)$. The standard Bonferroni test guarantees the left side is $\leq \alpha$ by making $P(R_i) = \alpha/n$ for every i. If we were to use this test in the current situation, we would compare each of the p-values from the pairwise comparisons with .05/15. If any of these p-values were less than .05/15, we would reject the null hypothesis, H_0, with an α-level of .05. Note that the standard Bonferroni test is a single-step procedure, and requires that all test results be checked.

We would prefer instead to use a step-down procedure, since it will be more powerful. In general, a step-down procedure begins by testing the overall intersection hypothesis and then steps down through the hierarchy of implied hypotheses. If any hypothesis is retained, all of the implied hypotheses will be retained without further tests. In this way, a hypothesis is tested if and only if all of its implying hypotheses are rejected.

A general method for step-down procedures, called the closure method, was proposed by Marcus, Peritz, and Gabriel (1976). For closed testing procedures, we let $H_i (1 \leq i \leq m)$ be a finite family of hypotheses, and we form the closure of this family by taking all nonempty intersections $H_P = \cap_{i \in P} H_i$ for $P \subseteq 1, 2, \ldots, m$. If an α-level test of each hypothesis H_P is available, then the closed testing procedure rejects any hypothesis H_P if and only if every H_Q is rejected by its associated α-level test for all $Q \supseteq P$.

Note that the number of tests in a closed testing procedure increases exponentially with m. It is therefore desirable to develop a shortcut method. One such shortcut is Holm's

sequentially rejective Bonferroni procedure, which is carried out in our case by following these steps:

1. We arrange $p_1, \ldots, p_{15}$ in non-decreasing order $p_{(1)}, \ldots, p_{(15)}$.

2. At the same time, we record the order (in size) of $p_1, \ldots, p_{15}$. From this operation, we get a list of values k, $k = 1, \ldots, 15$ such that $p_k = p_{(j)}$ for some j, $j = 1, \ldots, 15$, so that later we can identify the pairing which produced each $p_{(j)}$.

3. We calculate $q_1, \ldots, q_{15}$, where $q_1 = \frac{.05}{15}, q_2 = \frac{.05}{14}, \ldots, q_{15} = \frac{.05}{1}$.

4. Beginning with $p_{(1)}$, *we compare $p_{(j)}$ with q_j. If $p_{(j)}$ is less than q_j, we check the value of k to find the associated pair, reject H_0' for this pair, increment j by one, and repeat from *. If $p_{(j)}$ is greater than q_j, we store the current value of j in l, discontinue testing, and retain the null hypothesis H_0' for all pairs associated with $p_{(j)}$ where $j \geq l$. This can be done conveniently by examining the signs of $(q_j - p_{(j)})$, $j = 1, \ldots, 15$. If the sign is positive, $p_{(j)}$ is less than q_j.

5. At the end of this procedure we can say, with an α-level of .05, that all of the pairs with an associated $p_{(j)}$ where $j < l$ are significantly different.

The process of calculating the test statistics, conducting the pairwise tests, and performing the modified Bonferroni test as detailed above is performed for each of the twelve model variants.

In future work, we might consider using more recent developments in the theory of multiple comparisons. However, given our uncertainty about the validity of the underlying working assumptions pending further exploration by the investigator, we decided not to pursue further refinement of the analysis at this time.

11 Results

The variables of interest are shown in Table 6. The caste groups are:

Group 1: Brahmins
Group 2: Kapu
Group 3: Balija
Group 4: Mala
Group 5: Christians
Group 6: Muslims.

The basic model as specified in Section 10 was tested for twelve different model variants, six for men and six for women. These model variants are listed in Table 7. For model variants m.4, m.5, m.6, w.1, w.4, w.5, and w.6, there appeared to be no significant differences between the intercepts for the six groups. The results from the modified Bonferroni test for the other model variants are summarized in Table 8 and depicted graphically in Figure 11.

Tables 9 and 10 display the results of both the individual pairwise tests and the modified Bonferroni test for all of the model variants which showed any significant group differences. The "I" columns contain an asterisk if the indicated pairwise comparison resulted in a p-value of less than .05. We can say that each of these pairs is significantly different with an α-level of .05. The "B" columns contain asterisks for those pairwise comparisons with an associated $p_{(j)}$ where $j < l$ (where those terms are as defined in Section 10.4). For the indicated model variant, we can say that, taken as a group, all these pairs are significantly different with an α-level of .05.

Appendix I contains the Splus computer routines which were used for the more complex calculations.

For all twelve model variants tested, the actual p-values, p_j (as defined in Section 10.4) can be found in Appendix II. Table 13 in the same appendix shows the detailed results of the modified Bonferroni test.

Variable	Description
Variable	*Description*
Group B Variables	
X1	Age
X2	Height Vertex
X3	Sitting Height Vertex
X4	Arm Span
X5	Body Weight
X6	Chest Expansion
X7	Bi-acromial Breadth
Group C Variables	
Y1	Body Surface Area
Y2	Trunk Surface Area
Y3	Chest Volume
Y4	Bi-Acromial Breadth Index
Y5	Relative Thoracic Index
Y6	Pondoral Index
Y7	Robusticity Index
Y8	Pignet Vervack Index
Y9	Nutritional Index
Y10	Weight Index
Y11	Chest Girth Index
Y12	Span Stature Index
Y13	Sitting Height Stature Index
Pulmonary Functions	
Z1	Vital Capacity
Z2	Forced Vital Capacity
Z3	Forced Expiratory Volume

Table 6: Variables of Interest

Variant Number	Dependent Variable	Explanatory Variables
Men		
m.1	Z_1	X_1 to X_5, X_7
m.2	Z_2	X_1 to X_5, X_7
m.3	Z_3	X_1 to X_5, X_7
m.4	Z_1	X_6, Y_1 to Y_7, Y_{11} to Y_{13}
m.5	Z_2	X_6, Y_1 to Y_7, Y_{11} to Y_{13}
m.6	Z_3	X_6, Y_1 to Y_6, Y_8, Y_{11} to Y_{13}
Women		
w.1	Z_1	X_1 to X_5, X_7
w.2	Z_2	X_1 to X_5, X_7
w.3	Z_3	X_1 to X_5, X_7
w.4	Z_1	X_6, Y_1 to Y_7, Y_{11} to Y_{13}
w.5	Z_2	X_6, Y_1 to Y_7, Y_{11} to Y_{13}
w.6	Z_3	X_6, Y_1 to Y_7, Y_{11} to Y_{13}

Table 7: Model Variants Tested

| | Results | | | | |
Groups Compared	Variant m.1	Variant m.2	Variant m.3	Variant w.2	Variant w.3
1 versus 2	*		*		*
1 versus 3	*	*	*		*
1 versus 4	*		*	*	*
1 versus 5			*		
1 versus 6	*	*	*	*	*
2 versus 3	*	*	*		*
2 versus 4			*	*	
2 versus 5	*		*		
2 versus 6	*	*	*	*	
3 versus 4	*	*	*		
3 versus 5	*		*		
3 versus 6	*	*	*	*	
4 versus 5	*				
4 versus 6	*	*	*	*	
5 versus 6	*	*	*		

Table 8: Test Results from Bonferroni Comparisons for Men and Women

"*" indicates the pairs in each variant which are significantly different
with an α-level of .05 for the entire model variant.

Model Variant m.1

Model Variant m.2

Model Variant w.2

Model Variant m.3

Model Variant w.3

Group Key

1. Brahmins
2. Kapu
3. Balija
4. Mala
5. Christians
6. Muslims
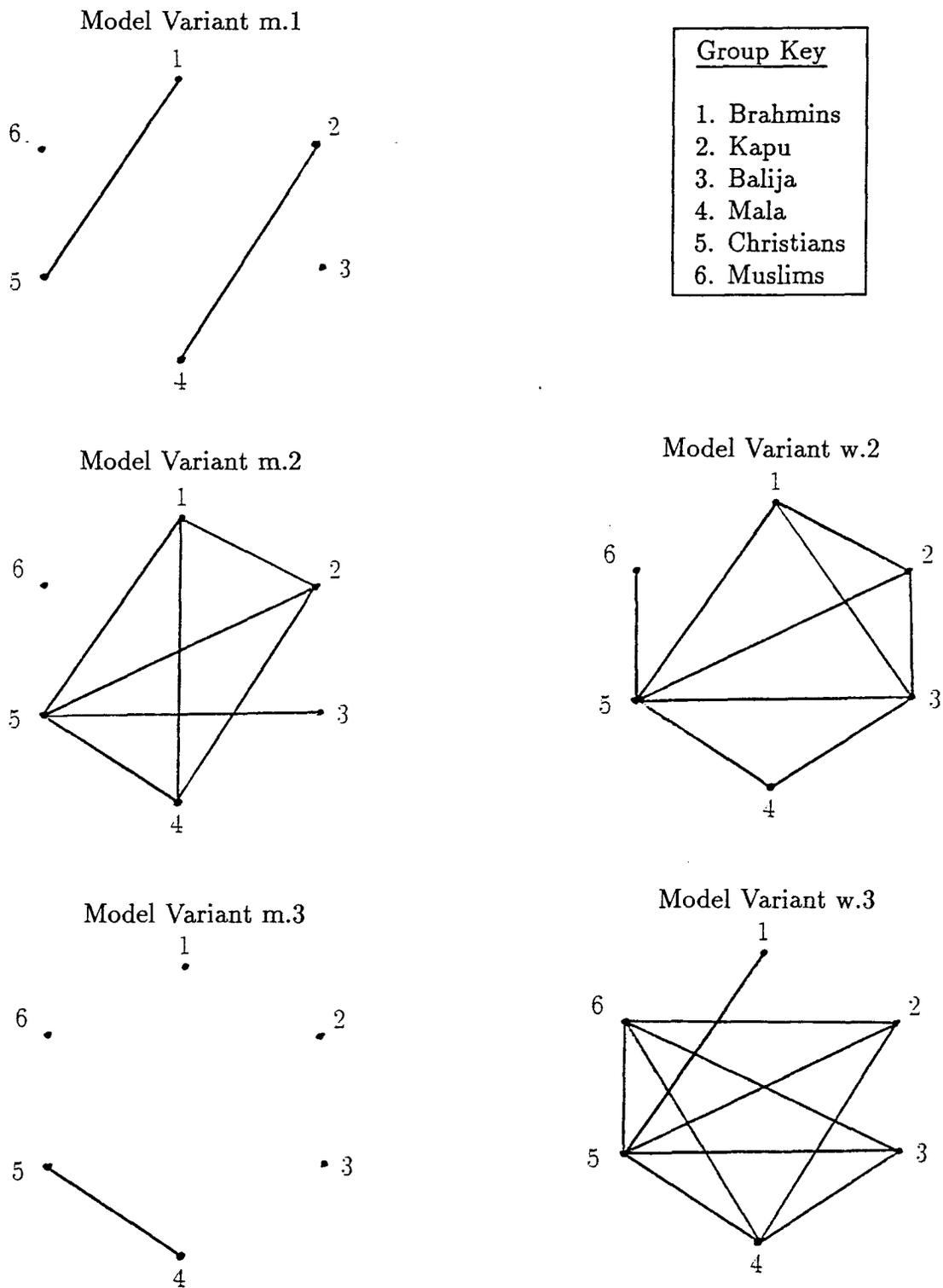
Figure 11: Results for Model Variants with Significant Differences at α-level of .05.

Points not connected to each other

represent groups which are significantly different at this level.

Groups Compared	Variant m.1		Variant m.2		Variant m.3	
	I	B	I	B	I	B
1 versus 2	*	*			*	*
1 versus 3	*	*	*	*	*	*
1 versus 4	*	*			*	*
1 versus 5					*	*
1 versus 6	*	*	*	*	*	*
2 versus 3	*	*	*	*	*	*
2 versus 4					*	*
2 versus 5	*	*	*		*	*
2 versus 6	*	*	*	*	*	*
3 versus 4	*	*	*	*	*	*
3 versus 5	*	*			*	*
3 versus 6	*	*	*	*	*	*
4 versus 5	*	*				
4 versus 6	*	*	*	*	*	*
5 versus 6	*	*	*	*	*	*

Table 9: Test Results from Individual and Bonferroni Comparisons for Men

"I" indicates individual pairwise comparisons.

"B" indicates modified Bonferroni test for the entire model variant.

"*" indicates significance at the level of $\alpha = .05$.

Groups Compared	Variant w.1		Variant w.2		Variant w.3	
	I	B	I	B	I	B
1 versus 2	*				*	*
1 versus 3					*	*
1 versus 4			*	*	*	*
1 versus 5					*	
1 versus 6			*	*	*	*
2 versus 3					*	*
2 versus 4	*		*	*		
2 versus 5						
2 versus 6			*	*		
3 versus 4			*			
3 versus 5						
3 versus 6			*	*		
4 versus 5						
4 versus 6			*	*		
5 versus 6						

Table 10: Test Results from Individual and Bonferroni Comparisons for Women

"I" indicates individual pairwise comparisons.

"B" indicates modified Bonferroni test for the entire model variant.

"*" indicates significance at the level of $\alpha = .05$.

12 Conclusions

It is important to note that the absence of the original data forced us to make a number of unverifiable working assumptions when we were building our model. We assumed the suitability of a linear model with a Gaussian-distributed error term with equal variance for all observations. In addition, we assumed a common value of β for all groups. Note, however, that the latter assumption, even if violated, does not invalidate our procedure. The assumption simply renders it conservative, as discussed in Section 9.

Data quality is always a concern, but it is doubly so in our case. Not only might there be misgivings about the method of subject selection, but in addition, the accuracy of the summary statistics themselves could be questioned since errors might have been made in their production and transcription. In the absence of the data, the reliability of these tables cannot be checked.

Nonetheless, the methodology which was developed to handle this problem appears sound, and it is remarkable for its simplicity and its ability to produce an analysis from a minimal amount of information.

References for Part II

Graybill, F.A. (1976). *Theory and Application of the Linear Model*. North Scituate, Massachusetts: Duxbury Press, 189–190.

Hochberg, Y., and Tamhane, A.C. (1987). *Multiple Comparison Procedures*. New York: Wiley, 53–58.

Marcus, R., Peritz, E., and Gabriel, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63**, 655–660.

Reddy, L. (1982). *Study of Pulmonary Functions and their Relation to Anthropometric Parameters in Men and Women of Chittoor District of Andra Pradesh*. Ph.D. Dissertation, Department of Anthropology, University of Madras.

Appendix I: Splus Routines for Part II

The following are the computer routines (written in Splus) which were used for the various calculations.

Calcint.s: used to calculate $\hat{\mu}$ and $\widehat{Cov}(\hat{\mu})$ for each of the six model variants for men.

```
t1_(n-1)*sd.m%*%corr.m%*%sd.m
J16_matrix(1,1,6)
ybar_(1/n)*J16%*%size.m%*%mean.m
J66_matrix(1,6,6)
z1_mean.m - (1/n)*J66%*%size.m%*%mean.m
w1_t1 - t(z1)%*%size.m%*%z1
wg_w1[indvector,indvector]
w_w1[indvector,i]
omega_w1[i,i]
zg_z1[,indvector]
z_z1[,i]
betahat_solve(wg)%*%w
sigma2hat_(1/(n-f))*(omega - t(w)%*%solve(wg)%*%w)
grandmean_(1/n)*J16%*%size.m%*%mean.m[,i]
ybarbar_rep(grandmean,6)
muhat_ybarbar + z - zg%*%betahat
sigmamuhat_sigma2hat[1,1]*(solve(size.m) + zg%*%solve(wg)%*%t(zg))
```

n = total number of males studied.

i = the number of the column (and row)

of the standard deviation and correlation matrices which

contains the dependent variable for the current model variant.

indvector = the vector of column (and row) numbers

for the explanatory variables for the current model variant.

f = the number of degrees of freedom for the current model variant.

For each model variant, the above values are set, the routine is called by the "source" statement, and then the values of "muhat" and "sigmamuhat" are stored in the appropriate storage registers for the current model variant. The following example shows the sequence for variant m.1.

```
n_687
i_21
indvector_c(1,2,3,4,5,7)
f_12
source("calcint.s")
muhat.m1_muhat
sigmamuhat.m1_sigmamuhat
```

Calcint.sw: used to calculate $\hat{\mu}$ and $\widehat{Cov}(\hat{\mu})$ for each of the six model variants for women.

```
t1_(n-1)*sd.w%*%corr.w%*%sd.w
J16_matrix(1,1,6)
ybar_(1/n)*J16%*%size.w%*%mean.w
J66_matrix(1,6,6)
z1_mean.w - (1/n)*J66%*%size.w%*%mean.w
w1_t1 - t(z1)%*%size.w%*%z1
wg_w1[indvector,indvector]
w_w1[indvector,i]
omega_w1[i,i]
zg_z1[,indvector]
z_z1[,i]
betahat_solve(wg)%*%w
sigma2hat_(1/(n-f))*(omega - t(w)%*%solve(wg)%*%w)
grandmean_(1/n)*J16%*%size.w%*%mean.w[,i]
ybarbar_rep(grandmean,6)
muhat_ybarbar + z - zg%*%betahat
sigmamuhat_sigma2hat[1,1]*(solve(size.w) + zg%*%solve(wg)%*%t(zg))
```

n = total number of females studied.

i = the number of the column (and row)

of the standard deviation and correlation matrices which

contains the dependent variable for the current model variant.

indvector = the vector of column (and row) numbers

for the explanatory variables for the current model variant.

f = the number of degrees of freedom for the current model variant.

For each model variant, the above values are set, the routine is called by the "source" statement, and then the values of "muhat" and "sigmamuhat" are stored in the appropriate storage registers for the current model variant. The following example shows the sequence for variant w.1.

```
n_614
i_21
indvector_c(1,2,3,4,5,7)
f_12
source("calcint.sw")
muhat.w1_muhat
sigmamuhat.w1_sigmamuhat
muhat.w1
```

testloop.s: used to calculate the test statistics and probability values for each of the fifteen pairwise tests for each model variant.

```
testvector_rep(0,15)
testindex_1
for (i in 2:6)
{
testvector[testindex]_(mu[1]-mu[i])/sqrt(sigma[1,1]+sigma[i,i] - 2*sigma[1,i])
testindex_testindex+1
}
for (i in 3:6)
{
testvector[testindex]_(mu[2]-mu[i])/sqrt(sigma[2,2]+ sigma[i,i] - 2*sigma[2,i])
testindex_testindex+1
}
for (i in 4:6)
{
testvector[testindex]_(mu[3]-mu[i])/sqrt(sigma[3,3]+ sigma[i,i] - 2*sigma[3,i])
testindex_testindex+1
}
for (i in 5:6)
{
testvector[testindex]_(mu[4]-mu[i])/sqrt(sigma[4,4]+ sigma[i,i] - 2*sigma[4,i])
testindex_testindex+1
}
testvector[testindex]_(mu[5]-mu[6])/sqrt(sigma[5,5]+ sigma[6,6] - 2*sigma[5,6])
pvector_2*(1-pnorm(abs(testvector)))
```

$$
\begin{aligned}
mu &= \hat{\mu} \text{ for the current model variant.} \\
sigma &= \widehat{Cov}(\hat{\mu}) \text{ for the current model variant.} \\
testvector &= \text{the } 15 \times 1 \text{ vector of test statistics} \\
&\quad \text{for group comparisons in the order indicated in Table 5.} \\
pvector &= \text{the } 15 \times 1 \text{ vector of } p\text{-values} \\
&\quad \text{for these test statistics.}
\end{aligned}
$$

Appendix II: Test Results for Part II

Groups Compared	p-value	Model Variant					
		m.1	m.2	m.3	m.4	m.5	m.6
1 versus 2	p_1	0.00	0.12	0.00	0.62	0.09	0.34
1 versus 3	p_2	0.00	0.00	0.00	0.34	0.06	0.36
1 versus 4	p_3	0.01	0.26	0.00	0.07	0.08	0.94
1 versus 5	p_4	0.36	0.35	0.02	0.96	0.51	0.30
1 versus 6	p_5	0.00	0.00	0.00	0.13	0.08	0.12
2 versus 3	p_6	0.00	0.00	0.00	0.51	0.70	0.98
2 versus 4	p_7	0.72	0.90	0.00	0.08	0.64	0.40
2 versus 5	p_8	0.00	0.04	0.00	0.78	0.66	0.61
2 versus 6	p_9	0.00	0.00	0.00	0.18	0.33	0.26
3 versus 4	p_{10}	0.00	0.00	0.00	0.23	0.86	0.43
3 versus 5	p_{11}	0.00	0.06	0.00	0.54	0.54	0.64
3 versus 6	p_{12}	0.00	0.00	0.00	0.29	0.42	0.27
4 versus 5	p_{13}	0.01	0.08	0.24	0.18	0.50	0.34
4 versus 6	p_{14}	0.00	0.00	0.00	0.75	0.51	0.14
5 versus 6	p_{15}	0.00	0.00	0.00	0.20	0.28	0.55

Table 11: p-values from Pairwise Comparisons for Testing Equality of μ_i's for Men

Groups Compared	p-value	Model Variant					
		w.1	w.2	w.3	w.4	w.5	w.6
1 versus 2	p_1	0.04	0.31	0.00	0.05	0.12	0.18
1 versus 3	p_2	0.29	0.09	0.00	0.10	0.13	0.12
1 versus 4	p_3	0.80	0.00	0.00	0.71	0.76	0.91
1 versus 5	p_4	0.75	0.77	0.02	0.55	0.55	0.05
1 versus 6	p_5	0.19	0.00	0.00	0.16	0.50	0.13
2 versus 3	p_6	0.26	0.31	0.00	0.74	0.99	0.74
2 versus 4	p_7	0.01	0.00	0.07	0.08	0.17	0.17
2 versus 5	p_8	0.56	0.89	0.98	0.79	0.93	0.14
2 versus 6	p_9	0.81	0.00	0.36	0.90	0.54	0.55
3 versus 4	p_{10}	0.15	0.03	0.35	0.18	0.20	0.12
3 versus 5	p_{11}	0.86	0.63	0.25	0.88	0.93	0.18
3 versus 6	p_{12}	0.60	0.00	0.20	0.91	0.56	0.75
4 versus 5	p_{13}	0.66	0.16	0.46	0.67	0.65	0.05
4 versus 6	p_{14}	0.11	0.00	0.63	0.25	0.66	0.13
5 versus 6	p_{15}	0.67	0.05	0.65	0.85	0.84	0.28

Table 12: p-values from Pairwise Comparisons for Testing Equality of μ_i's for Women

73

$p_{(j)}$	Variant m.1		Variant m.2		Variant m.3		Variant w.2		Variant w.3	
	Sgn	k	Sgn	k	Sgn	k	Sgn	k	Sgn	k
$p_{(1)}$	+	5	+	12	+	1	+	14	+	1
$p_{(2)}$	+	6	+	15	+	2	+	12	+	3
$p_{(3)}$	+	9	+	9	+	5	+	9	+	5
$p_{(4)}$	+	12	+	5	+	6	+	3	+	6
$p_{(5)}$	+	14	+	14	+	9	+	7	+	2
$p_{(6)}$	+	15	+	6	+	10	+	5	-	4
$p_{(7)}$	+	10	+	10	+	11	-	10	-	7
$p_{(8)}$	+	2	+	2	+	12	-	15	-	12
$p_{(9)}$	+	11	-	8	+	14	-	2	-	11
$p_{(10)}$	+	8	-	11	+	15	-	13	-	10
$p_{(11)}$	+	1	-	13	+	7	-	1	-	9
$p_{(12)}$	+	13	-	1	+	8	-	6	-	13
$p_{(13)}$	+	3	-	3	+	3	-	11	-	14
$p_{(14)}$	-	4	-	4	+	4	-	4	-	15
$p_{(15)}$	-	7	-	7	-	13	-	8	-	8

Table 13: Test Results from Bonferroni Method

A "+" in the "Sgn" column indicates $j < l$ for the associated $p_{(j)}$,
where $p_{(j)}$ and l are as defined in Section 10.4.

For each variant, all pairs (considered as a group) with a "+" in this column
are significantly different at the level of $\alpha = .05$.

"k" is the value for finding the associated p_k (and pairing) from Table 5.