

**CERTAIN PERCENTAGE POINTS OF THE DISTRIBUTION OF  
THE STUDENTIZED RANGE OF LARGE SAMPLES**

by

**William H. Beyer**

**Thesis submitted to the Graduate Faculty of the  
Virginia Polytechnic Institute  
in candidacy for the degree of**

**MASTER OF SCIENCE**

in

**STATISTICS**

**APPROVED:**

**APPROVED:**

---

**Director of Graduate Studies**

---

**Head of Department**

---

**Dean of Applied Science and  
Business Administration**

---

**Major Professor**

**August 1953**

**Blacksburg, Virginia**

## TABLE OF CONTENTS

	page
I. INTRODUCTION	4
II. EXTENSION OF THE STUDENTIZATION FORMULA	6
2.1 General Mathematical Development	6
2.1.1 Symbolic Operators	14
2.2 Application of the Formula to the Studentized Range	16
2.2.1 The Calculation of the Desired Percentage Points of the Studentized Range	20
III. GEOMETRIC SOLUTION FOR CERTAIN PERCENTAGE POINTS OF THE STUDENTIZED RANGE	25
3.1 Statement of the Problem	25
3.2 Geometric Development	25
3.3 The Solution of the Problem in the General Case	33
3.3.1 Development of the General Jacobian	34
3.3.2 Evaluation of the General Jacobian	40

3.3.3	Upper and Lower Limits of the Percentage Points	42
3.3.4	Calculation of Percentage Points for Large Sample Size	43
3.4	Final Results	48
IV.	SUMMARY	51
V.	ACKNOWLEDGMENTS	52
VI.	BIBLIOGRAPHY	53
VII.	VITA	56

## I. INTRODUCTION

In recent years, several methods have been proposed for testing the differences between several means in an analysis of variance.

As a typical example of the problem involved, consider the results of a Graeco-Latin Square experiment in the spinning of a cotton roving (a product like yarn but coarser and less highly twisted), given in Table 1.\*

Table 1. Weights of Roving, Grams per 180 Yards

<u>Spindle</u>	1	2	3	4	5
<u>Mean</u>	16.20	16.15	16.24	16.36	16.30

The problem is to test whether each of the ten differences between means considered in pairs is significant or not.

In the preparation of tables for a new multiple range test proposed for this problem, a need has arisen for the

\* Tippett, L. H. C., The Methods of Statistics, pages 195-197, fourth edition.

determination of certain special percentage points of the studentized range. Many of these percentage points can be calculated directly from tables of the probability integral of the studentized range given by Pearson and Hartley (1943). The points which can be obtained in this way are those for which the number of degrees of freedom,  $n_2$ , exceeds ten, and for which the sample sizes,  $n$ , lie between two and twenty, inclusive. In order to calculate percentage points for cases where  $n_2 \leq 10$ ,  $n = 2(1)20^*$ , and also for the cases  $n > 20$ , however, it has been necessary to devise new methods of procedure. This paper presents these new methods along with examples showing their use. The methods developed here are:

- 1) An extension of Hartley's (1938, 1944) studentization formula which enables the calculation of tabular values for those cases where  $n_2 \leq 10$ ,  $n = 2(1)20$ .
- 2) A geometric presentation which facilitates the calculation of the tabular values for those cases where  $n > 20$ .

Tables showing the calculated percentage points are given at the end of Section III.

---

\*  $n$  may be any integral value between two and twenty.

## II. EXTENSION OF THE STUDENTIZATION FORMULA

In this section a formula approximating the distribution of the studentized form of a general statistic  $h$  is derived, and the use of this formula in the calculation of the desired percentage points of the studentized range is demonstrated. The preliminary steps of the derivation are based on ideas put forward by Hartley (1938, 1944), but the mathematical proof is based on new ideas using series expansions and symbolic operators.

### 2.1 General Mathematical Development

Let  $x_1, x_2, \dots, x_n$  be a sample of  $n$  observations drawn from a normal population with standard deviation  $\sigma$ . Consider a general statistic  $h$ , calculated from this sample, having the following properties:

- (i)  $h$  is non-negative. (Any statistic  $u$  can be transformed to meet this condition by putting  $h = |u|$ .)
- (ii)  $h$  is such that the distribution function of  $g = h/\sigma$  is independent of  $\sigma$ .

For example,  $h$  may be the range of a sample. Then  $g = h/\sigma$  is non-negative and has the distribution of the range of a sample from a normal population with unit variance. This is true irrespective of the value of  $\sigma$ .

A studentized statistic  $r$  corresponding to  $g$  can be defined as  $r = h/s$ , where  $s^2$  is an estimate of the variance  $\sigma^2$  with the properties:

- (i)  $n_2 s^2 / \sigma^2$  is distributed as  $\chi^2$  with  $n_2$  degrees of freedom,
- (ii)  $s^2$  is independent of  $h$ .

Then the studentized statistic  $r$  will have a distribution not involving  $\sigma$ , since this statistic may be written  $r = \frac{h/\sigma}{s/\sigma} = \frac{g/\sqrt{n_2}}{\chi}$ . From this it is seen that the distribution function of  $r$  is completely determined by the normal distribution with unit standard deviation and the  $\chi^2$  distribution with  $n_2$  degrees of freedom.

The probability element of the joint distribution of  $X$  and  $g$  is given by

$$f(x, g) dx dg = \left[ \Gamma\left(\frac{n_2}{2}\right) \right]^{-1} 2^{-\frac{n_2}{2} + 1} \chi^{n_2 - 1} e^{-\frac{1}{2}\chi^2} f(g) dx dg, \quad (2.1)$$

whence the distribution function of  $r = \frac{g/\sqrt{n_2}}{\chi}$ ,  $f_{n_2}(r)$ , say, is obtained:

$$f_{n_2}(r) dr = \left[ \Gamma\left(\frac{n_2}{2}\right) \right]^{-1} 2^{-\frac{n_2}{2} + 1} \left\{ \int_0^{\infty} \chi^{n_2} e^{-\frac{1}{2}\chi^2} \sqrt{n_2} f\left(\chi \frac{r}{\sqrt{n_2}}\right) d\chi \right\} dr.$$

Then the cumulative distribution function  $P_{n_2}(R) = P[r \leq R]$

is given by

$$P_{n_2}(R) = \int_0^R f_{n_2}(r) dr$$

$$= \left[ \Gamma\left(\frac{n_2}{2}\right) \right]^{-1} 2^{-\frac{n_2}{2}} \int_0^{\infty} x^{n_2-1} e^{-\frac{1}{2}x^2} \left\{ \int_0^{\frac{Rx}{\sqrt{n_2}}} f(g) dg \right\} dx. \quad (2.2)$$

Finally, by introducing

$$P_n(G) = \int_0^G f(g) dg$$

equation (2.2) may be written

$$P_{n_2}(R) = \left[ \Gamma\left(\frac{n_2}{2}\right) \right]^{-1} 2^{-\frac{n_2}{2}+1} \int_0^{\infty} x^{n_2-1} e^{-\frac{1}{2}x^2} P_n\left(\frac{Rx}{\sqrt{n_2}}\right) dx. \quad (2.3)$$

Equation (2.3) is a formal representation of the studentized probability integral  $P_{n_2}(R)$  dependent on  $P_n(g)$ . At this point, departure from Hartley's method is made.

One way to express equation (2.3) in a more workable form is to substitute

$$z = \frac{x^2}{2} .$$



This substitution gives

$$P_{n_2}(R) = \left[ \Gamma\left(\frac{n_2}{2}\right) \right]^{-1} \int_0^{\infty} z^{\frac{n_2}{2}-1} e^{-z} P_n\left(R\sqrt{\frac{2z}{n_2}}\right) dz \quad (2.4)$$

Now in equation (2.4), make an alternate substitution

$$z = \frac{n_2}{2} e^{y\sqrt{\frac{2}{n_2}}}.$$

Then equation (2.4) becomes

$$P_{n_2}(R) = \frac{\left(\frac{n_2}{2}\right)^{\frac{n_2-1}{2}}}{\Gamma\left(\frac{n_2}{2}\right)} \int_{-\infty}^{\infty} e^{y\sqrt{\frac{n_2}{2}} - \frac{n_2}{2} e^{y\sqrt{\frac{2}{n_2}}}} P_n\left(R e^{y/\sqrt{2n_2}}\right) dy \quad (2.5)$$

In the exponential in the integrand of equation (2.5), expand  $e^{y\sqrt{\frac{2}{n_2}}}$ , giving

$$e^{y\sqrt{\frac{2}{n_2}}} = 1 + y\left(\frac{2}{n_2}\right)^{\frac{1}{2}} + \frac{y^2}{2!}\left(\frac{2}{n_2}\right) + \frac{y^3}{3!}\left(\frac{2}{n_2}\right)^{3/2} + \dots$$

Then

$$y\sqrt{\frac{n_2}{2}} - \frac{n_2}{2} e^{y\sqrt{\frac{2}{n_2}}} = -\frac{n_2}{2} - \frac{y^2}{2!} - \frac{y^3}{3!}\left(\frac{2}{n_2}\right)^{\frac{1}{2}} - \frac{y^4}{4!}\left(\frac{2}{n_2}\right) - \dots \quad (2.6)$$

On substituting equation (2.6) into equation (2.5), the following expression is obtained for  $P_{n_2}(R)$ :

$$P_{n_2}(R) =$$

$$\frac{\left(\frac{n_2}{2}\right)^{\frac{n_2-1}{2}}}{\Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{n_2}{2} \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \left\{ e^{-\frac{y^3}{3!} \left(\frac{1}{2n_2}\right)^{\frac{1}{2}} - \frac{y^4}{4!} \left(\frac{1}{2n_2}\right) + \dots} \right\} P_n \left( R e^{y/\sqrt{2n_2}} \right) dy. \quad (2.7)$$

Also  $P_n \left( R e^{y/\sqrt{2n_2}} \right)$  can be written as

$$P_n \left( R e^{y/\sqrt{2n_2}} \right) = P_n \left\{ R + Ry \left( \frac{1}{2n_2} \right)^{\frac{1}{2}} + \frac{Ry^2}{2!} \left( \frac{1}{2n_2} \right) + \dots \right\}. \quad (2.8)$$

Using the operator  $D = \frac{d}{dx}$  in the symbolic form of Taylor's expansion\*

$$F(x+k) = e^{kD} F(x),$$

where  $F(x)$  is any repeatably differentiable function of a variable  $x$ , equation (2.8) can be written as

$$\begin{aligned} P_n \left( R e^{y/\sqrt{2n_2}} \right) &= e^{RD} \left[ y \left( \frac{1}{2n_2} \right)^{\frac{1}{2}} + \frac{y^2}{2!} \left( \frac{1}{2n_2} \right) + \dots \right] P_n(R) \\ &= \left\{ 1 + \frac{y}{(2n)^{\frac{1}{2}}} RD + \frac{y^2}{2n_2} \left[ \frac{RD}{2} + \frac{R^2 D^2}{2} \right] \right. \\ &\quad \left. + \frac{y^3}{(2n_2)^{3/2}} \left[ \frac{RD}{3!} + \frac{R^2 D^2}{2!} + \frac{R^3 D^3}{3!} \right] + \dots \right\} P_n(R) \end{aligned} \quad (2.9)$$

\* A more detailed discussion is presented later in this section.

If, in equation (2.7),  $e^{-\frac{y^3}{3!} \left(\frac{2}{n_2}\right)^{\frac{1}{2}} - \frac{y^4}{4!} \left(\frac{2}{n_2}\right) - \dots}$  is expanded, and equation (2.9) is inserted, the expression for  $P_{n_2}(R)$  becomes

$$\begin{aligned}
 P_{n_2}(R) = & \frac{\left(\frac{n_2}{2}\right)^{\frac{n_2-1}{2}} \frac{n_2}{2} \frac{1}{\sqrt{2\pi}}}{\Gamma\left(\frac{n_2}{2}\right)} \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \left\{ 1 - \frac{y^3}{3!} \left(\frac{2}{n_2}\right)^{\frac{1}{2}} \right. \\
 & - \left. \left[ \frac{y^4}{4!} - \frac{y^6}{2(3!)^2} \right] \frac{2}{n_2} - \dots \right\} \left\{ 1 + \frac{y}{(2n_2)^{\frac{1}{2}}} RD \right. \\
 & \left. + \frac{y^2}{2n_2} \left[ \frac{RD}{2} + \frac{R^2 D^2}{2} \right] + \dots \right\} P_n(R) dy \quad (2.10)
 \end{aligned}$$

If, in equation (2.10), the two series expansions are multiplied together, and the integrand is then integrated term by term, using the standard formulae for the moments of a normal distribution with mean zero and standard deviation unity,

i.e.

$$E(y^\beta) = \begin{cases} 0 & , \beta \text{ an odd integer} \\ 1 \cdot 3 \cdot 5 \dots (\beta-1) & , \beta \text{ an even integer} \end{cases}$$

equation (2.10) becomes

$$\begin{aligned}
 P_{n_2}(R) = & \frac{\left(\frac{n_2}{2}\right)^{\frac{n_2-1}{2}} e^{-\frac{n_2}{2}} \sqrt{2\pi}}{\Gamma\left(\frac{n_2}{2}\right)} \left\{ 1 + \frac{1}{6n_2} + \frac{1}{72n_2^2} - \frac{139}{6480n_2^3} \right. \\
 & + \frac{1}{4n_2} \left[ R^2 D^2 - RD \right] + \frac{1}{16n_2^2} \left[ \frac{R^4 D^4}{2} - \frac{R^3 D^3}{3} + \frac{R^2 D^2}{6} - \frac{RD}{6} \right] \\
 & + \frac{1}{64n_2^3} \left[ \frac{R^6 D^6}{6} + \frac{R^5 D^5}{6} - 5R^4 D^4 + \frac{9R^3 D^3}{16} - 35R^2 D^2 + \frac{9RD}{2} \right] \\
 & + \dots \left. \right\} P_n(R) \quad .
 \end{aligned} \tag{2.11}$$

From equation (2.11), it follows that

$$\begin{aligned}
 \log P_{n_2}(R) = & \frac{n_2-1}{2} \log \frac{n_2}{2} - \frac{n_2}{2} + \log \sqrt{2\pi} - \log \Gamma\left(\frac{n_2}{2}\right) \\
 & + \log \left\{ 1 + \frac{1}{6n_2} + \dots + \frac{1}{4n_2} (R^2 D^2 - RD) + \dots \right\} P_n(R) \left. \right\} \\
 = & \frac{n_2-1}{2} \log \frac{n_2}{2} - \frac{n_2}{2} + \log \sqrt{2\pi} \\
 & - \left[ \log \sqrt{2\pi} - \frac{n_2}{2} + \frac{n_2-1}{2} \log \frac{n_2}{2} + \frac{1}{12} \frac{n_2}{2} \right. \\
 & \left. - \frac{1}{360} \frac{n_2^3}{8} + \dots \right] + \log \left\{ 1 + \frac{1}{6n_2} + \dots \right\} P_n(R) \left. \right\}
 \end{aligned}$$

from whence

$$\log P_{n_2}(R) = \left\{ \frac{1}{4n_2} \left[ R^2 D^2 - RD \right] + \frac{1}{n_2^2} \left[ \frac{R^3 D^3}{24} - \frac{R^2 D^2}{16} + \frac{RD}{32} \right] \right. \\ \left. + \frac{1}{n_2^3} \left[ \frac{R^3 D^3}{96} - \frac{R^2 D^2}{32} + \frac{5RD}{128} \right] + \dots \right\} P_n(R) .$$

Then

$$P_{n_2}(R) = \left\{ e^{\frac{1}{4n_2} \left[ R^2 D^2 - RD \right] + \frac{1}{n_2^2} \left[ \frac{R^3 D^3}{24} - \frac{R^2 D^2}{16} + \frac{RD}{32} \right] + \dots} \right\} P_n(R) \\ = P_n(R) + \frac{1}{4n_2} \left[ R^2 D^2 - RD \right] P_n(R) \\ + \frac{1}{16n_2^2} \left[ \frac{R^4 D^4}{2} - \frac{R^3 D^3}{3} - \frac{R^2 D^2}{2} + \frac{RD}{2} \right] P_n(R) \\ + \frac{1}{16n_2^3} \left[ \frac{R^6 D^6}{6} + \frac{R^5 D^5}{6} - \frac{7R^4 D^4}{6} + 2R^3 D^3 - \frac{5R^2 D^2}{2} + \frac{5RD}{2} \right] P_n(R) \\ + O(n_2^{-4}) + \dots . \tag{2.12}$$

Equation (2.12) represents the probability  $P_{n_2}(R)$  for the studentized statistic  $r = h/s$  to fall below  $R$ . With equation (2.12), the solution of studentization which is of sufficient accuracy for most practical purposes has been reached. This formula lends itself very well to the computation of  $P_{n_2}(R)$ .

### 2.1.1 Symbolic Operators

In order to use formula (2.12) for the computation of  $P_{n_2}(R)$ , it is necessary to understand a little of the theory of the symbolic operator  $D$ .

Suppose a function  $f(b)$  is given in a table for the values  $a, a+k, a+2k, \dots$  of its argument  $b$ . It is required to find the value of the derivative of the function when the argument has the value  $a+ck$ , where  $c$  is a fraction.

Before this problem can be solved by a method of interpolation, it is first necessary to form what are called the differences of the tabular values. The quantity

$$f(a+k) - f(a)$$

is denoted by  $\Delta f(a)$  and is called the first difference of  $f(a)$ . The first difference of  $f(a+k)$  is  $f(a+2k) - f(a+k)$ , which is denoted by  $\Delta f(a+k)$ . It then follows that

$$\begin{aligned} \Delta^2 f(a) &= f(a+2k) - 2f(a+k) + f(a) \\ &= \Delta f(a+k) - \Delta f(a) \end{aligned} ,$$

and so on for higher differences.

The formula for the  $n$ th derivative of a function may be found by using symbolic operators and expanding the function  $f(a+k)$  by Taylor's Theorem.

Thus

$$f(a+k) = f(a) + kf'(a) + \frac{k^2}{2!}f''(a) + \dots \quad (2.13)$$

If  $D$  denotes the operator for differentiation,  $\frac{d}{dx}$ , equation (2.13) becomes

$$f(a+k) = \left[ 1 + kD + \frac{k^2D^2}{2!} + \frac{k^3D^3}{3!} + \dots \right] f(a) \quad ,$$

or

$$(1+\Delta)f(a) = e^{kD}f(a) \quad ,$$

and

$$1 + \Delta = e^{kD} \quad , \quad (2.14)$$

Taking logarithms of each side of equation (2.14),

$$\begin{aligned} kD &= \log_e(1+\Delta) \\ &= \Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \dots \quad , \end{aligned}$$

from whence

$$kf'(a) = \Delta f(a) - \frac{1}{2}\Delta^2 f(a) + \frac{1}{3}\Delta^3 f(a) - \dots \quad (2.15)$$

Also

$$\begin{aligned} k^2D^2 &= \left\{ \log(1+\Delta) \right\}^2 \\ &= \left( \Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \dots \right)^2 \quad . \end{aligned}$$

Therefore

$$\begin{aligned} k^2 f''(a) &= (\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \dots)^2 f(a) \\ &= \Delta^2 f(a) - \Delta^3 f(a) + \frac{11}{12}\Delta^4 f(a) + \dots \end{aligned} \quad (2.16)$$

and in general

$$k^n f^{(n)}(a) = (\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \dots)^n f(a) . \quad (2.17)$$

## 2.2 Application of the Formula to the Studentized Range

The range, or difference between the highest and lowest observations of the ordered sample,  $(x_{[1]}, x_{[2]}, \dots, x_{[n]})$ , is  $x_{[n]} - x_{[1]}$ , and will be denoted by  $w$ . This statistic  $w$  is an example of the general statistic  $h$  described in section 2.1.  $w$  is non-negative and the distribution function of  $w/\sigma$  is independent of  $\sigma$ . The studentization range is defined as

$$q = \frac{w}{s} = \frac{x_{[n]} - x_{[1]}}{s} ,$$

where  $s$  is an independent estimate of the standard deviation of the  $x$ -variate. As an example, consider the data presented in Table 1, Section I. The estimate of the standard error of a spindle mean is

$$s_m = 12.31 ,$$



and is based on eight degrees of freedom.

The probability law of  $q$  does not depend on the standard deviation  $\sigma$ , since  $\sigma$  is a scale parameter of the distributions of both  $w$  and  $s$ , and therefore is eliminated by taking the ratio  $q$ . Thus  $q$  falls into the class of those statistics defined by  $r$  in the preceding section.

The idea of studentizing the range seems to have occurred to W. S. Gosset himself. The studentized range,  $q$ , is a particular case of studentized statistics discussed by Hartley (1938, 1944). The usefulness of  $q$  in particular problems has been illustrated by Newman (1939) and by Pearson and Hartley (1943). Using the approximate probability law of the range,  $w$ , due to Pearson, Newman (1939) obtained by quadrature the 5 percent and 1 percent probability levels of  $q$  for small values of the sample size and for degrees of freedom greater than or equal to five. Pearson and Hartley (1942) revised the table of probability levels given by Newman with the help of their own exact tables of the probability integral for  $w$ . They calculated the upper and lower percentage points of  $q$  for sample sizes ranging from two to twenty, and degrees of freedom greater than or equal to ten.

Before applying equation (2.12) to the calculation of the desired percentage points of the studentized range, a slight change of notation will be given.

Denote the probability integral  $P_{n_2}(R)$  by  $P(Q; n_2, n)$ ,  
i.e.

$$P(Q; n_2, n) = P[q \leq Q] \quad ,$$

where  $n_2$  represents the degrees of freedom of the estimate of the variance  $s^2$ , and  $n$  represents the sample size. As  $n_2$  becomes large,  $s^2$  approaches  $\sigma^2$ , and the probability integral of  $Q$  tends to the probability integral of  $W$ ,  $P(W; n)$ , where this probability integral represents the probability that the variate  $w$  will not exceed any fixed value  $W$ .

As previously shown,  $P(Q; n_2, n)$  can be represented by equation (2.12) as

$$\begin{aligned} P(Q; n_2, n) &= P_n(Q) + \frac{1}{4n_2} [Q^2 D^2 - QD] P_n(Q) \\ &+ \frac{1}{16n_2^2} \left[ \frac{Q^4 D^4}{2} - \frac{Q^3 D^3}{3} - \frac{Q^2 D^2}{2} + \frac{QD}{2} \right] P_n(Q) \quad (2.18) \\ &+ \frac{1}{64n_2^3} \left[ \frac{Q^6 D^6}{6} + \frac{Q^5 D^5}{6} - \frac{7Q^4 D^4}{6} + 2Q^3 D^3 - \frac{5Q^2 D^2}{2} \right. \\ &\left. + \frac{5QD}{2} \right] P_n(Q) + O(n_2^{-4}) + \dots \end{aligned}$$

For brevity, set

$$a_n(Q) = \frac{1}{4} [Q^2 D^2 - QD] P_n(Q) \quad ,$$

$$b_n(Q) = \frac{1}{16} \left[ \frac{Q^4 D^4}{2} - \frac{Q^3 D^3}{3} - \frac{Q^2 D^2}{2} + \frac{QD}{2} \right] P_n(Q) ,$$

and

$$c_n(Q) = \frac{1}{64} \left[ \frac{Q^6 D^6}{6} + \frac{Q^5 D^5}{6} - \frac{7Q^4 D^4}{6} + 2Q^3 D^3 - \frac{5Q^2 D^2}{2} + \frac{5QD}{2} \right] P_n(Q) .$$

Then equation (2.18) becomes

$$P(Q; n_2, n) = P_n(Q) + \frac{1}{n_2} a_n(Q) + \frac{1}{n_2^2} b_n(Q) + \frac{1}{n_2^3} c_n(Q) + \dots \quad (2.19)$$

A table of  $P_n(Q)$ ,  $a_n(Q)$ , and  $b_n(Q)$  for  $n$  lying between two and twenty has been given by Pearson and Hartley (1943). This table is based on a five-decimal manuscript table of the probability integral of the range, four decimals of which were published by Pearson and Hartley (1942). The derivatives in formula (2.18) were calculated from the differences of  $P(W;n)$  using intervals of 0.25 by formulae given in section 2.1.1 and are taken at argument  $W = Q$ .

For  $n_2 \leq 10$ ,  $n = 2(1)20$ , the published table is sufficient for the calculation of the desired percentage points of the studentized range. Since it was realized that for small values of  $n_2$ , the expression for the probability integral would break down, it has been necessary to extend the table to contain a term in  $\frac{1}{n_2^3}$ , which has the coefficient  $c_n(Q)$ , as given in equation (2.19).

It would be desirable to know something about the accuracy of the approximation given by equation (2.19). However, the neglected remainder of the expansion depends on the derivatives of the cumulative distribution function  $P_n(Q)$ , and it is difficult to reach a general formula which may be used as a convenient gauge for the estimation of its magnitude.

A check of selected points is afforded by comparison with upper percentage points of the studentized range calculated by May (1952).

For rapidity, the values of  $Q$  for  $n_2 \leq 10$ ,  $n = 2(1)20$ , are calculated by the method of extrapolation presented in the next section. Spot checks of these values are computed by the method of studentization and are found to agree to a high degree of accuracy.

### 2.2.1 The Calculation of the Desired Percentage Points of the Studentized Range

The problem involved in calculating the percentage points of the studentized range is the following.

Given  $\gamma_n = \gamma^{n-1}$ , where  $\gamma = .95, .99$ , and  $n = 2(1)20$ ; and given that the probability integral of the studentized range,  $P(Q; n_2, n) = \gamma_n$ , we desire to find  $Q$ .

The procedure used in finding  $Q$  is:

1) Consider the sequence of  $Q$  values .00, .25, .50, ... presented in the table given by Pearson and Hartley (1943).

Find the probabilities  $P_1, P_2, P_3$ , using formula (2.19), for three consecutive values  $Q_1, Q_2, Q_3$  in the above sequence such that  $P_2$  is nearer the given probability than either  $P_1$  or  $P_3$ . The given probabilities,  $\gamma_n$ , are shown in Table 2.

- (1) When  $n_2$  becomes infinite, the probability integral of the studentized range,  $Q$ , becomes the probability integral of the standardized range,  $W$ .

i.e.

$$P(Q; n_2, n) = P(W; n)$$

A table of  $P(W; n)$  is given by Pearson and Hartley (1942).

- 2) By a method of interpolation using a second degree polynomial,\* find the value of  $Q$ , or  $W$ , corresponding to the given probability  $\gamma_n$ .

Two sample calculations are given below, the first for the case where  $n_2 = 11$  and  $n = 3$ , the second for the case where  $n_2 = \infty$ , and  $n = 2$ .

Sample Calculation 1:

$$(\text{Case } P(Q; n_2, n) = .9025, n_2 = 11, n = 3)$$

Enter tables given by Pearson and Hartley (1943) under the desired sample size  $n$ . Calculate probabilities  $P_1, P_2, P_3$ , using formula (2.19), for three consecutive values  $Q_1,$

---

\* The actual method used was a modification of the Aitken-Neville method presented by M. C. K. Tweedie (unpublished).

Table 2. Values of  $\gamma_n = \gamma^{n-1}$ ,  $\gamma = .95, .99$ 

$\gamma = .95$		$\gamma = .99$	
n	$\gamma_n$	n	$\gamma_n$
2	.95	2	.99
3	.9025	3	.9801
4	.857375	4	.970299
5	.814506	5	.960596
6	.773781	6	.950990
7	.735092	7	.941480
8	.698337	8	.932065
9	.663420	9	.922744
10	.630249	10	.913517
11	.598737	11	.904382
12	.568800	12	.895338
13	.540360	13	.886385
14	.513342	14	.877521
15	.487675	15	.868746
16	.463291	16	.860058
17	.440126	17	.851458
18	.418120	18	.842943
19	.397214	19	.834514
20	.377353	20	.826169
21	.358485	21	.817907
51	.076945	51	.605006
101	.005921	101	.366032

$Q_2, Q_3$  such that  $P_1$  and  $P_3$  be above and below the given probability .9025, respectively, and  $P_2$  is nearer .9025 than either  $P_1$  or  $P_3$ .

Q	$P(Q; 11, 3)$
3.50	$P_1 = .9644 + \frac{1}{11}(-.39) + \frac{1}{121}(-.4) = .9256$
3.25	$P_2 = .9439 + \frac{1}{11}(-.46) + \frac{1}{121}(-.2) = .9004$
3.00	$P_3 = .9145 + \frac{1}{11}(-.51) + \frac{1}{121}(.1) = .8689$

Then interpolating using a second degree polynomial, the desired value of Q is found to be 3.27.

#### Sample Calculation 2:

(Case  $P(W;n) = .95, n = 2$ )

Enter tables given by Pearson and Hartley (1942) under the desired sample size n. Locate probabilities  $P_1, P_2, P_3$  for three consecutive values  $W_1, W_2, W_3$  such that  $P_1$  and  $P_3$  lie above and below the given probability .95, respectively, and  $P_2$  is nearer .95 than either  $P_1$  or  $P_3$ .

W	$P(W;2)$
2.70	.9438
2.75	.9482
2.80	.9523

Then interpolating using a second degree polynomial, the desired value of  $W$  is found to be 2.77.

For values of  $n_2 \leq 10$ ,  $c_n(Q)$  must be calculated using the method of symbolic operators described in section 2.1.1. Then the calculation of the percentage points for the cases where  $n_2 \leq 10$  follows the same pattern as that used in the calculation of the percentage points for the cases where  $n_2 > 10$ .

The calculated percentage points of the studentized range are given in Tables 4 and 5, Section III. Table 4 shows the percentage points for the case where  $\gamma_n = \gamma^{n-1}$ ,  $\gamma = .95$ , while Table 5 shows the points for the case where  $\gamma_n = \gamma^{n-1}$ ,  $\gamma = .99$ . It may be mentioned that Table 4 is used for a 5 percent level multiple range test, and that Table 5 is used for a 1 percent level multiple range test.

A feature of the required test for which the percentage points have been determined is that a percentage point  $Q(n_2, n, \gamma_n) | n=n_0$  is not required if it exceeds  $Q(n_2, n, \gamma_n) | n=n_0-1$ . Those values not needed for the application of the multiple range test are not included in the manuscript tables.



### III. GEOMETRIC SOLUTION FOR CERTAIN PERCENTAGE POINTS OF THE STUDENTIZED RANGE

#### 3.1 Statement of the Problem

The problem may be stated in general terms as follows.

Given that

$$P[q \leq Q_{\gamma_n}] = \gamma_n ,$$

where  $q$  represents the studentized range,

i.e.

$$q = \frac{x_{[n]} - x_{[1]}}{s} ,$$

we desire to find the fixed value  $Q_{\gamma_n}$ . The solution of this problem is to be used in obtaining values of  $Q_{\gamma_n}$  for  $n > 20$ .

#### 3.2 Geometric Development

In the case when  $n = 3$ , consider the following uncorrelated linear functions of  $x_1, x_2, x_3$  (the sample for which  $q$  is the studentized range).

$$\begin{aligned} z_1 &= x_1 - x_2 \\ z_2 &= (x_1 + x_2 - 2x_3)/\sqrt{3} \end{aligned} \tag{3.1}$$

The probability density in  $z$ -space is readily found to be the bivariate normal

$$f(z_1, z_2) = \frac{1}{(\sqrt{2\pi})^2} e^{-\frac{1}{2}(z_1^2 + z_2^2)/2\sigma^2} \quad (3.2)$$

Now, we note that the region defined in  $z$ -space by the inequality

$$x_{[3]} - x_{[1]} \leq Q_{\gamma_3} \quad (3.3)$$

is the hexagon illustrated in Figure 1(a).

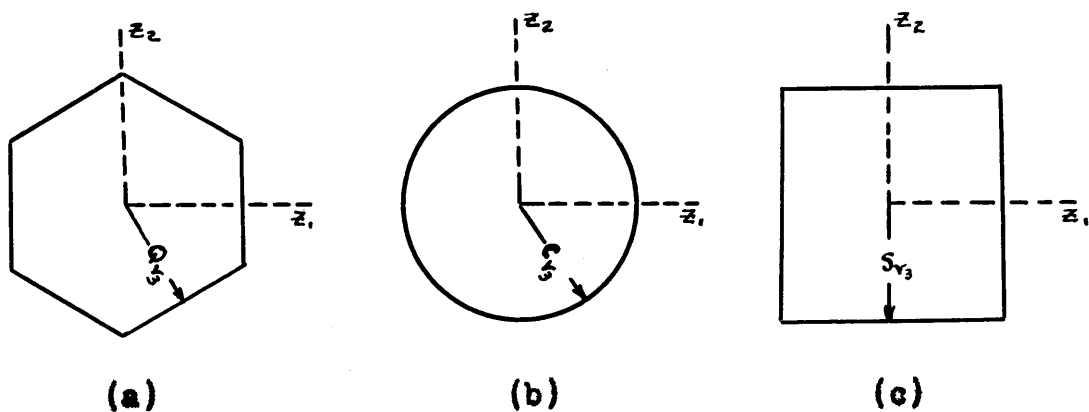


Figure 1. Regions in  $z$ -space (case  $n=3$ )

Hence the required probability  $\gamma_3$  is the integral of (3.3) over the hexagon in Figure 1(a).

Next we note that the region

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq c_{\gamma_3}^2 \quad (3.4)$$

in  $z$ -space is the circle with radius  $c_{\gamma_3}$  illustrated in Figure 1(b), and the probability

$$P\left[\sum_{i=1}^n (x_i - \bar{x})^2 \leq c_{\gamma_3}^2\right]$$

is the integral of  $f(z_1, z_2)$  over this circle.

Finally, we note that the region

$$|\tilde{z}| \leq S_{\gamma_3} \quad (3.5)$$

in  $z$ -space, where  $|\tilde{z}|$  is the larger of the variates  $z_1$  and  $z_2$ , is a square for which the radius of the inscribed circle is  $S_{\gamma_3}$ , as illustrated in Figure 1(c).

In the following discussion, it will be convenient to refer to the radius of the inscribed circle of a regular rectilinear figure as the 'radius' of the figure.

A method for finding upper and lower limits for the  $Q_{\gamma_3}$  value for a given probability  $\gamma_3$  may be explained in this special case as follows.

1) Find a circle containing the given probability  $\gamma_3$ . Let  $c_{\gamma_3}$  denote the radius of the circle and  $A(c_{\gamma_3})$  its area.

a) Find the radius  $Q(c_{\gamma_3})$  of a hexagon of area  $A(c_{\gamma_3})$ . Then it can be shown that  $Q(c_{\gamma_3}) < Q_{\gamma_3}$ . [For if we put  $A(Q_{\gamma_3})$  for the area of the hexagon containing the probability  $\gamma_3$ ,  $A(Q_{\gamma_3}) > A(c_{\gamma_3})$ . Thence  $Q(c_{\gamma_3}) < Q_{\gamma_3}$ .]

2) Find a square containing the given probability  $\gamma_3$ .

Let  $S_{\gamma_3}$  denote the radius of the square and  $A(S_{\gamma_3})$  its area.

a) Find the radius  $Q(S_{\gamma_3})$  of a hexagon containing the area  $A(S_{\gamma_3})$ . It is reasonable to expect that  $Q(S_{\gamma_3}) > Q_{\gamma_3}$ , as is verified by computing  $Q_{\gamma_3}$  from the table given by Pearson and Hartley (1944).

The above method gives  $Q(c_{\gamma_3})$  and  $Q(S_{\gamma_3})$  as lower and upper limits of  $Q_{\gamma_3}$ ,

i.e.

$$Q(c_{\gamma_3}) < Q_{\gamma_3} < Q(S_{\gamma_3}) \quad (3.6)$$

The advantages of this procedure are:

(i) The radius,  $c_{\gamma_3}$ , of a circle containing the given probability,  $\gamma_3$ , can readily be found. For  $c_{\gamma_3}^2 = 2\chi_2^2(\gamma_3)$  where

$$\int_0^{\chi_2^2(\gamma_3)} f(\chi_2^2) d(\chi_2^2) = \gamma_3 \quad (3.7)$$

(ii) The radius,  $S_{\gamma_3}$ , of a square containing the given probability,  $\gamma_3$ , can also be readily found. For  $S_{\gamma_3}$  is the solution of

$$\int_{-S_{\gamma_3}}^{S_{\gamma_3}} \int_{-S_{\gamma_3}}^{S_{\gamma_3}} f(z_1, z_2) dz_1 dz_2 = \gamma_3 \quad (3.8)$$

Equation (3.8) can then be written

$$\int_{-S_{\gamma_3}}^{S_{\gamma_3}} f(z) dz = \sqrt{\gamma_3} \quad (3.9)$$

Thus  $S_{\gamma_3} = \sqrt{2}u(\sqrt{\gamma_3})$  where  $u(\sqrt{\gamma_3})$  is the  $\gamma_3$ -percentage point of the standard normal density function.

In the generalization of this approach for cases where  $n > 3$ , the main problem lies in working out the volume (in the generalized  $z$ -space) of the generalization of the hexagon. The method by which the procedure for finding limits is generalized is a direct extension of the following discussion.

To find the area of a hexagon in  $z$ -space, for the case  $n = 3$ , defined by the inequality (3.3)

$$x_{[3]} - x_{[1]} \leq Q_{\gamma_3},$$

we first consider the subset of samples in which  $x_1$  is the highest observation,  $x_2$  the middle observation, and  $x_3$  the lowest observation. That is,

$$\begin{aligned} x_{[3]} &= x_1 \\ x_{[2]} &= x_2 \\ x_{[1]} &= x_3 \end{aligned} \quad (3.11)$$

The part of the  $z$ -space corresponding to this subset is the infinite area lying between the lines  $AO$  and  $OB$  in Figure 2. It will be convenient to call this the restricted  $z$ -space.

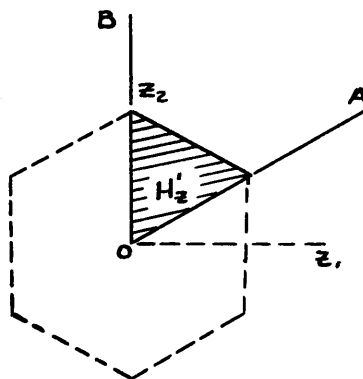


Figure 2. Region defined by inequality  $x_1 - x_3 \leq Q_{r3}$

The inequality (3.2) in the subset (3.11) becomes

$$x_1 - x_3 \leq Q_{r3} , \quad (3.12)$$

and this denotes the region,  $H'_z$ , in the restricted  $z$ -space, as shown by the shaded part of the hexagon in Figure 2.

The area,  $H_z$ , of the complete hexagon in  $z$ -space will be given by  $3H'_z$ .

To find the area,  $H'_z$ , we first make a new transformation

$$\begin{aligned} y_1 &= x_1 - x_2 \\ y_2 &= x_2 - x_3 \end{aligned} . \quad (3.13)$$

Then the inequality (3.12) becomes

$$y_1 + y_2 \leq Q_{Y3} \quad (3.14)$$

in the restricted  $y$ -space in which  $y_1 \geq 0$  and  $y_2 \geq 0$ . The inequality (3.14) defines the region,  $H'_y$ , in the restricted  $y$ -space, as shown by the shaded section of the hexagon in Figure 3. The complete hexagon,  $H_y$ , in  $y$ -space corresponds

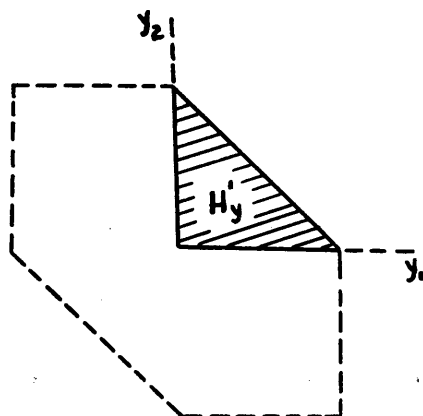


Figure 3. Region defined by inequality  $y_1 + y_2 \leq Q_{Y3}$

to the complete hexagon,  $H_z$ , in  $z$ -space.

The advantage of introducing the  $y$ -transformation is that the area of  $H'_y$  in the restricted  $y$ -space is given by the simple Dirichlet integral

$$A(H'_y) = \int_0^{Q_{Y3}} \int_0^{Q_{Y3} - y_1} dy_2 dy_1 \quad (3.15)$$

This area is found to be

$$A(H_y^i) = \frac{Q^2 Y_3}{2} \quad .$$

The area of  $H_z^i$  in  $z$ -space is simply the area of  $H_y^i$  in  $y$ -space multiplied by the Jacobian of the transformation from  $z$  to  $y$ . That is,

$$\iint_{H_z} dz_1 dz_2 = \iint_{H_y} \frac{\partial(z_1, z_2)}{\partial(y_1, y_2)} dy_1 dy_2 \quad . \quad (3.16)$$

In the case under discussion ( $n = 3$ ), the  $z$ 's may be readily expressed in terms of the  $y$ 's as follows:

$$\begin{aligned} z_1 &= y_1 \\ z_2 &= (y_1 + 2y_2)/\sqrt{3} \end{aligned}$$

The Jacobian given in equation (3.16) is

$$\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)} = \begin{vmatrix} 1 & 0 \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{3}} \end{vmatrix} = \frac{2}{\sqrt{3}} \quad .$$

Therefore the area of  $H_z^i$  in  $z$ -space is

$$A(H_z^i) = \frac{2}{\sqrt{3}} \frac{Q^2 Y_3}{2} = \frac{Q^2 Y_3}{\sqrt{3}} \quad .$$



Thence the required area is

$$A(H_z) = 31 \frac{Q_{r_3}^2}{\sqrt{3}} = 2\sqrt{3} Q_{r_3}^2 .$$

We now have the areas of the circle, hexagon, and square in terms of their respective radii as follows:

$$\pi [c(r_3)]^2 , \quad 2\sqrt{3} Q_{r_3}^2 , \quad [2s(r_3)]^2 .$$

Thence the radius of a hexagon with the same area as a circle with radius  $c(r_3)$  or a square with radius  $s(r_3)$  is given by

$$Q(c_{r_3}) = \sqrt{\frac{\sqrt{3}\pi}{6}} c(r_3)$$

or

$$Q(s_{r_3}) = \sqrt{\frac{2\sqrt{3}}{3}} s(r_3) ,$$

respectively. Thus, the lower and upper limits of  $Q_{r_3}$ ,  $Q(c_{r_3})$  and  $Q(s_{r_3})$  respectively, are known.

### 3.3 The Solution of the Problem in the General Case

Most of the work in finding the lower and upper limits of  $Q_{r_n}$  in the general cases lies in obtaining the volume of the hyper-polyhedron in  $z$ -space. The volumes of the hyper-sphere and hyper-cube are given in textbooks, e.g. Jeffreys

and Jeffreys (1946). The volume,  $H_y^i$ , in the restricted  $y$ -space can readily be found by solving the generalized Dirichlet integral representing this volume. Thus the major part of the work required in obtaining the volume in  $z$ -space is that of finding the value of the generalized Jacobian of the transformation from the  $z$ 's to the  $y$ 's.

### 3.3.J. Development of the General Jacobian

As in section 3.2, let  $x_{[1]}$ ,  $x_{[2]}$ , ...,  $x_{[n]}$  be  $n$  ranked variables. Consider the subset of samples in which  $x_1$  is the highest observation and  $x_n$  the lowest observation. That is,

$$\begin{aligned} x_{[1]} &= x_n \\ x_{[2]} &= x_{n-1} \\ &\cdot \quad \cdot \quad \cdot \\ x_{[n]} &= x_1 \end{aligned} \quad (3.17)$$

The part of  $z$ -space corresponding to this subset is called the restricted  $z$ -space.

Now, consider the following generalization of the transformation given by equations (3.1), which is the transformation of  $x_1, x_2, \dots, x_n$  (the sample for which  $q$  is the studentized range).

$$\begin{aligned}
z_1 &= x_1 - x_2 \\
z_2 &= (x_1 + x_2 - 2x_3)/\sqrt{3} \\
&\quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
z_i &= (x_1 + \dots + x_i - ix_{i+1})/\sqrt{\frac{i(i+1)}{2}},
\end{aligned}
\tag{3.18}$$

where  $i = 1, 2, \dots, n - 1$ .

Now, define

$$\begin{aligned}
z_1 &= v_1 \\
z_2 &= v_2/\sqrt{3} \\
&\quad \cdot \quad \cdot \quad \cdot \\
z_i &= v_i/\sqrt{\frac{i(i+1)}{2}},
\end{aligned}
\tag{3.19}$$

where

$$\begin{aligned}
v_1 &= x_1 - x_2 \\
v_2 &= x_1 + x_2 - 2x_3 \\
&\quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
v_i &= x_1 + \dots + x_i - ix_{i+1}
\end{aligned}
\tag{3.20}$$

The problem now is to find the volume of the hyperpolyhedron,  $H_z$ , in  $z$ -space. This volume is given by  $n!H_z^n$ , where  $H_z^n$  denotes the volume in the restricted  $z$ -space given by the inequality

$$x_1 - x_n \leq Q_{\gamma n} \quad \cdot \tag{3.21}$$

To find the volume of  $H_y^i$ , first make the following transformation:

$$\begin{aligned} y_1 &= x_1 - x_2 \\ y_2 &= x_2 - x_3 \\ &\dots \\ y_i &= x_i - x_{i+1} \end{aligned} \quad (3.22)$$

The inequality (3.21) then becomes

$$\sum_{i=1}^{n-1} y_i \leq Q_{Y_n} \quad (3.23)$$

and defines the volume,  $H_y^i$ , in the restricted  $y$ -space in which  $y_i \geq 0$ .

The advantage of introducing this  $y$ -transformation is that the volume of  $H_y^i$  in the restricted  $y$ -space is given by the general Dirichlet integral

$$V(H_y^i) = \int_0^{Q_{Y_n}} \int_0^{Q_{Y_n} - y_1} \dots \int_0^{Q_{Y_n} - \sum_{i=1}^{n-1} y_i} dy_{n-1} \dots dy_2 dy_1 \quad (3.24)$$

The value of this multiple integral as given by Jeffreys and Jeffreys (1946) is

$$V(H_y^i) = \frac{Q_{Y_n}^{n-1}}{(n-1)!} \quad (3.25)$$

The volume of  $H_z^i$  in  $z$ -space is then the volume of  $H_y^i$  multiplied by the generalized Jacobian. This last product, when multiplied by  $n!$ , gives the volume of the entire hyperpolyhedron in  $z$ -space.

The method of procedure used in obtaining the generalized Jacobian is as follows: Augment the  $y$ 's by putting

$$y_n = \sum x_i \quad i = 1, 2, \dots, n$$

and put

$$Y' = [y_1, y_2, \dots, y_n] \ .$$

For the  $v$ 's, put

$$v_n = y_n$$

and put

$$V' = [v_1, v_2, \dots, v_n] \ .$$

For the  $z$ 's, put

$$z_n = v_n$$

and

$$Z' = [z_1, z_2, \dots, z_n] \ .$$

Y, V, and Z are matrices with n rows and one column.

Now, we can write

$$Y = K_y X$$

where

$$K_y = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}, \quad (3.26)$$

and

$$V = K_v X$$

where

$$K_v = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 1 & 1 & -2 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \dots & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 & \dots & 1 \end{bmatrix}, \quad (3.27)$$

$$K_z = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1/\sqrt{3} & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 1/\sqrt{6} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1/\sqrt{i(i+1)/2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \quad (3.28)$$

Thence

$$Z = AY, \quad (3.29)$$

where

$$A = K_z K_v K_y^{-1}.$$

Now the Jacobian of the augmented  $z$ 's with respect to the augmented  $y$ 's is

$$\begin{aligned} \frac{\partial(z_1, z_2, \dots, z_n)}{\partial(y_1, y_2, \dots, y_n)} &= |A| \\ &= \frac{|K_z| |K_v|}{|K_y|} \end{aligned} \quad (3.30)$$

### 3.3.2. Evaluation of the General Jacobian

Consider the value of the Jacobian in the general case. This value can most easily be found by evaluating the determinants of the three matrices  $K_z$ ,  $K_v$ , and  $K_y$ .

#### (i) Evaluation of $|K_z|$

From equation (3.28),  $K_z$  is seen to be diagonal. The value of  $|K_z|$  is therefore

$$\begin{aligned} |K_z| &= \prod_{i=1}^{n-1} \left\{ \frac{1}{\sqrt{i(i+1)/2}} \right\} \\ &= \prod_{i=1}^{n-1} \left\{ \sqrt{\frac{2}{i(i+1)}} \right\} \end{aligned} \quad (3.31)$$

#### (ii) Evaluation of $K_v$

$K_v$  is given by equation (3.27) and  $|K_v|$  can be evaluated by the following procedure:

- 1) Add the first column of  $|K_v|$  to the second column.
- 2) In the determinant resulting from 1), add the second column to the third column.

3) Continue this process until the  $(n-1)^{\text{st}}$  column has been added to the  $n^{\text{th}}$  column. This last determinant is triangular with zeros above the diagonal and elements from one to  $n$  down the diagonal. The value of  $|K_v|$  is thus given by

$$|K_v| = n! \quad (3.32)$$



(iii) Evaluation of  $|K_y|$ 

$K_y$  is given by equation (3.26), and may be evaluated by the same procedure described in (ii) above. The last determinant obtained by this procedure is triangular with ones down the diagonal except for the element in the  $n^{\text{th}}$  row and  $n^{\text{th}}$  column, which is  $n$ . The value of  $|K_y|$  is thus given by

$$|K_y| = n \quad (3.33)$$

Knowing the value of the three determinants in the general case, we find the value of the Jacobian of the augmented  $z$ 's with respect to the augmented  $y$ 's to be

$$\frac{\partial(z_1, z_2, \dots, z_n)}{\partial(y_1, y_2, \dots, y_n)} = \frac{\prod_{i=1}^{n-1} \left\{ \sqrt{\frac{2}{1(i+1)}} \right\} n!}{n},$$

which upon simplification becomes

$$\frac{\partial(z_1, z_2, \dots, z_n)}{\partial(y_1, y_2, \dots, y_n)} = \sqrt{\frac{2^{n-1}}{n}}.$$

It is obvious that this augmented Jacobian has the same value as the generalized Jacobian of the transformation of the  $z$ 's given by equations (3.19) to the  $y$ 's given by (3.22). Thus the value of the generalized Jacobian is given by

$$\frac{\partial(z_1, z_2, \dots, z_{n-1})}{\partial(y_1, y_2, \dots, y_{n-1})} = \sqrt{\frac{2^{n-1}}{n}} \quad (3.34)$$

### 3.3.3. Upper and Lower Limits of the Percentage Points

The volume of the entire hyper-polyhedron in  $z$ -space is then given by

$$V(H_z) = \frac{\partial(z_1, \dots, z_1)}{\partial(y_1, \dots, y_1)} V(H_y) ,$$

where  $V(H_y)$  is found by multiplying  $V(H_x)$  given by equation (3.25) by the  $n!$  permutations of the  $x$ 's. Thus

$$\begin{aligned} V(H_z) &= \sqrt{\frac{2^{n-1}}{n}} n! \frac{Q_{\gamma_n}^{n-1}}{(n-1)!} \\ &= 2^{\frac{n-1}{2}} n^{\frac{1}{2}} Q_{\gamma_n}^{n-1} . \end{aligned} \quad (3.35)$$

The volumes of the hyper-sphere and hyper-cube, in terms of their respective radii, as given by Jeffreys and Jeffreys (1946), are

$$V(c_{\gamma_n}) = \frac{\pi^{\frac{n-1}{2}}}{\left(\frac{n-1}{2}\right)!} c_{\gamma_n}^{n-1}$$

and

$$V(s_{\gamma_n}) = [2s_{\gamma_n}]^{n-1}$$

Thence the radius of the hyper-polyhedron with the same area as a hyper-sphere with radius  $c(\gamma_n)$  or a hyper-cube with radius  $S(\gamma_n)$  is given by

$$Q(c_{\gamma_n}) = \frac{\pi^{\frac{1}{2}}}{\left(\frac{n-1}{2}\right)! n^{\frac{1}{2}} 2^{\frac{n-1}{2}}} (c_{\gamma_n}) \quad (3.36)$$

or

$$Q(S_{\gamma_n}) = 2^{-\frac{1}{2}} n^{-\frac{1}{2(n-1)}} (2S_{\gamma_n}) \quad (3.37)$$

respectively.

Equations (3.36) and (3.37) define lower and upper limits of  $Q_{\gamma_n}$ , respectively, and depend only on the sample size  $n$  and the radii of the hyper-sphere and hyper-cube, and in no way do they depend on the degrees of freedom  $n_2$ .

#### 3.3.4. Calculation of Percentage Points for Large Sample Size

Sample Calculation 1:

(Case  $n_2 = \infty$ ,  $\gamma = .95$ )

Using equations (3.36)\* and (3.37), calculate the lower and upper limits of  $Q_{\gamma_n}$ , respectively, for various values

---

\* The values of  $c_{\gamma_n}$  used in equation (3.36) are given in tables presented by Duncan (1951) entitled "Significant Ranges."

of  $n$ , including 50 and 100. Values of  $Q_{\gamma_n}$  for the cases  $n = 2(1)20$  are given in Table 4, and thus the differences between the tabular values and the limits are readily found. Figure 4 shows the lower and upper limits for the cases  $n = 2(1)20, 50, 100$ , and also shows the tabular values of  $Q_{\gamma_n}$ , which were calculated from tables given by Pearson and Hartley (1943), for the cases  $n = 2(1)20$ . The data used for plotting the curves is shown in Table 3. It may be noted that the differences between the tabular values of  $Q_{\gamma_n}$  and the upper limits remain constant for the cases where  $n \geq 10$ . This constant difference is used to find the value of  $Q_{\gamma_n}$  for the cases where  $n = 50$  and  $n = 100$ .

#### Sample Calculation 2:

(Cases  $n_2 < \infty, \gamma = .95$ )

Due to the difficulty of finding the upper limits of  $Q_{\gamma_n}$  for the cases where  $n_2 < \infty$ , only the lower limits are used. These lower limits are easily obtained for any sample size by using equation (3.36). For various values of  $n_2$ , Figure 5 shows the lower limits for the cases where  $n = 2(1)20, 50, 100$ , and the values of  $Q_{\gamma_n}$  for the cases where  $n = 2(1)20$ .

When  $n$  is held constant, the differences between the lower limits and the tabular values of  $Q_{\gamma_n}$  also remain constant. This was verified by calculating these differences

for the cases  $n = 2(1)20$ . Extending these findings to the cases where  $n = 50$ ,  $n = 100$ , we find the values of  $Q_{\gamma_n}$  by adding the differences .38 and .43, respectively, to the lower limits,  $Q(c_{\gamma_n})$ .

Table 3. Differences between Tabular Values and Limits of  $Q_{\gamma_n}$

$n$	$c(\gamma_n)$	Differences	Tabular	Differences	$S(\gamma_n)$
3	2.91	.01	2.92	.06	2.98
4	2.98	.04	3.02	.09	3.11
5	3.03	.06	3.09	.12	3.21
7	3.09	.10	3.19	.14	3.33
11	3.15	.17	3.32	.16	3.48
15	3.18	.22	3.40	.16	3.56
19	3.19	.26	3.45	.16	3.61
50	3.23	.38	3.61*	.16	3.77
100	3.24	.43	3.67*	.16	3.83

\* The tabular values 3.61 and 3.67 are found by adding the differences .38 and .43, respectively, to the values 3.23 and 3.24.

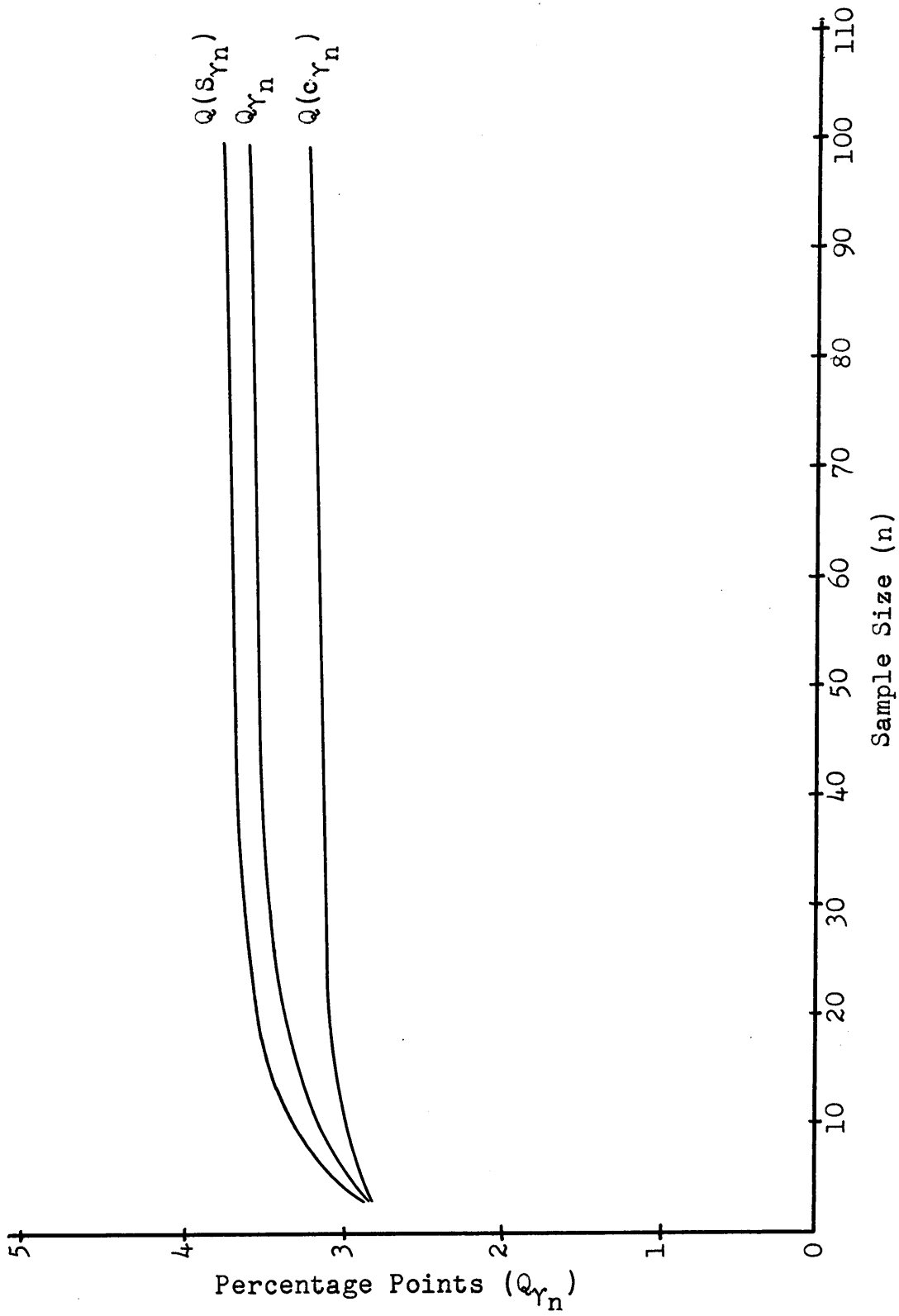


Figure 4. Lower and upper limits for  $Q_{Y_n}$  (case  $n_2 = \infty$ )

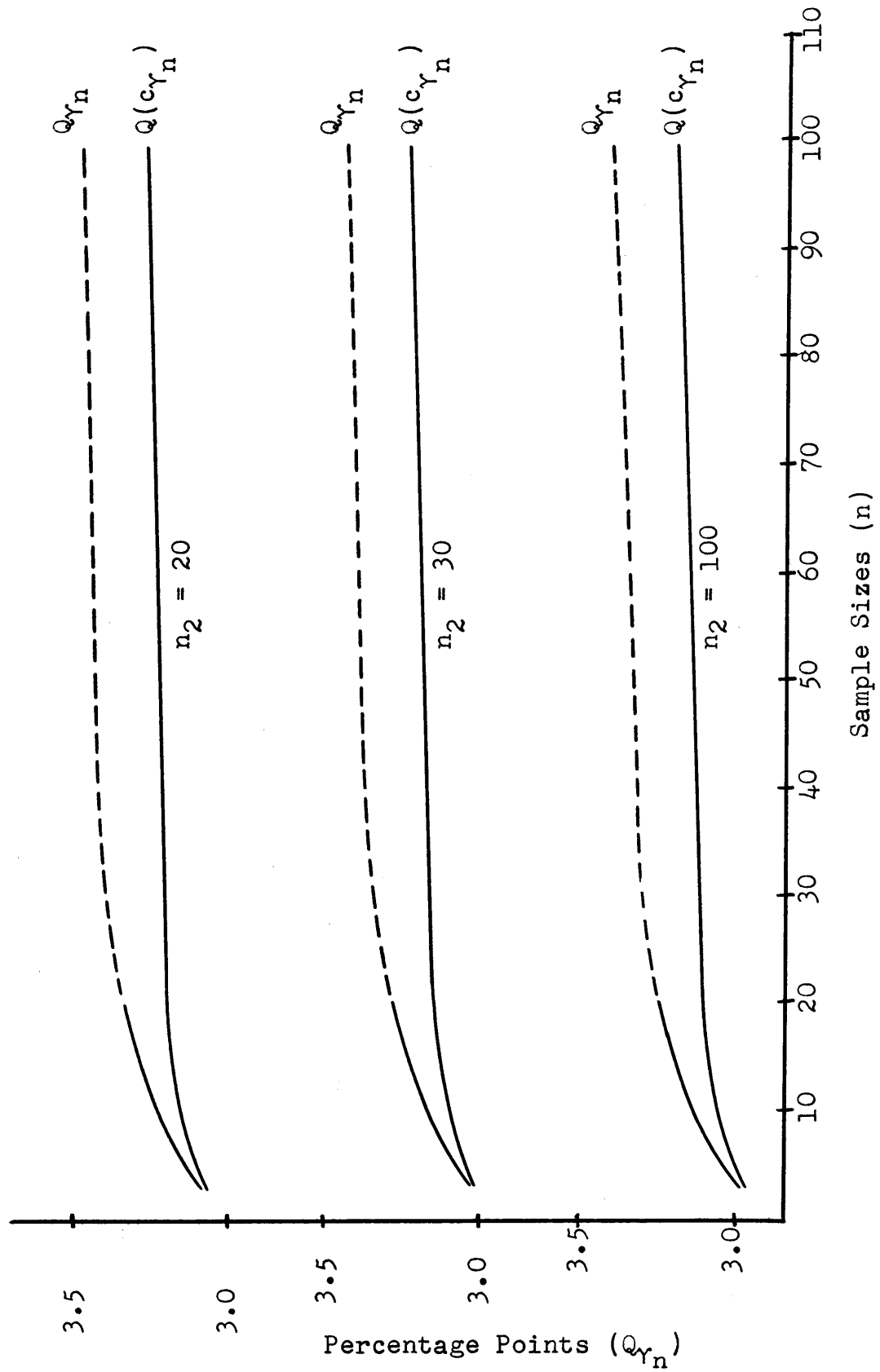


Figure 5. Lower limits and tabular values of  $Q_{\gamma_n}$   
(cases  $n_2 = 20, 30, 100$ )

### 3.4. Final Results

The desired percentage points of the studentized range are given in Tables 4 and 5. Table 4 shows the points for the cases where  $\gamma_n = \gamma^{n-1}$ ,  $\gamma = .95$ , and Table 5 shows the points for the cases where  $\gamma_n = \gamma^{n-1}$ ,  $\gamma = .99$ .



Table 4. Percentage Points of the Studentized Range ( $\gamma_{n-1} = \gamma$ ,  $\gamma = .95$ )

$n_2$	$n$	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
3		4.50															
4		3.93	4.01	4.02													
5		3.64	3.74	3.79	3.83												
6		3.46	3.58	3.64	3.68												
7		3.35	3.47	3.64	3.58	3.60	3.61										
8		3.26	3.39	3.47	3.52	3.55	3.56										
9		3.20	3.34	3.41	3.47	3.50	3.52	3.58									
10		3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47								
11		3.11	3.27	3.38	3.39	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.48	3.48	3.48		
12		3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.47	3.47	3.48		
13		3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47	3.47		
14		3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.46	3.46	3.47	3.47	3.47		
15		3.02	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47		
16		3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47		
17		2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47		
18		2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47		
19		2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47		
20		2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47		
22		2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47		
24		2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47		
26		2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47		
28		2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47		
30		2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47		
40		2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47		
60		2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47	3.48	
100		2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47	3.53	
$\infty$		2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47	3.61	3.67

Table 5. Percentage Points of the Studentized Range ( $Y_n = Y_{n-1}, Y = .99$ )

$n_2$	$n$	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
3	3	8.26	8.50	8.64	8.71	8.79	8.86	8.93	8.97	8.99	9.06	9.10	9.17	9.26	9.27		
4	4	6.51	6.76	6.90	6.97	7.05	7.12	7.19	7.23	7.25	7.32	7.37	7.43	7.51	7.52		
5	5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.46	6.55	6.60	6.66	6.74	6.75		
6	6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.03	6.11	6.16	6.22	6.30	6.31		
7	7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.76	5.84	5.88	5.94	6.02	6.04		
8	8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.54	5.63	5.68	5.74	5.81	5.83		
9	9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.39	5.48	5.53	5.59	5.66	5.68		
10	10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55		
11	11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39		
12	12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26		
13	13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15		
14	14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07		
15	15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00		
16	16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94		
17	17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89		
18	18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85		
19	19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82		
20	20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79		
22	22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75	4.74	
24	24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72	4.73	
26	26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69	4.72	
28	28	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67	4.72	
30	30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65	4.71	
40	40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59	4.69	
60	60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53	4.66	
100	100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48	4.64	
$\infty$	$\infty$	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41	4.60	4.68

## IV. SUMMARY

The purpose of this work is to investigate methods of obtaining special percentage points of the studentized range. In fulfilling this purpose, two new methods are developed and used.

The procedure for finding the percentage points for cases where  $n_2 \leq 10$ ,  $n = 2(1)20$ , is outlined, and examples illustrating the method used are given.

A geometric method is developed for finding the percentage points for cases where  $n > 20$ , and examples using this method are given.

Tables showing the desired percentage points are presented in the text of this thesis.

## V. ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professors D. B. Duncan and M. C. K. Tweedie for their guidance during the course of this work. Their many ideas and suggestions have made this paper possible. Appreciation is also given to Professors R. A. Bradley and B. Harshbarger and to Mr. R. L. Wine for their helpful suggestions and criticisms.

The author is also indebted to  
for preparing the final copies for presentation.

## VI. BIBLIOGRAPHY

1. Duncan, D. B.--"A Significance Test for Differences between Ranked Treatments in an Analysis of Variance." Virginia Journal of Science 2: 171-189, 1951.
2. Duncan, D. B.--"On the Properties of the Multiple Comparisons Test." Virginia Journal of Science 3: 49-67, 1952.
3. Duncan, D. B.--"Significance Tests for Differences between Ranked Treatments in an Analysis of Variance." Technical Report No. 3: 1953.
4. Gumbel, E. J.--"The Distribution of the Range." The Annals of Mathematical Statistics 18: 384-412, 1947.
5. Hartley, H. O.--"Studentization." Biometrika 33: 173-180, 1944.
6. Hartley, H. O.--"Studentization and Large-Sample Theory." Supplement to the Journal of the Royal Statistical Society 5: 80-88, 1938.
7. Hartley, H. O.--"The Use of the Range in an Analysis of Variance." Biometrika 37: 271-280, 1950.

8. Jeffreys, H. and Jeffreys, B. S.--Methods of Mathematical Physics. Cambridge University Press, London, England, 1946.
9. Johnson, N. L.--"Approximations to the Probability Integral of the Distribution of the Range." Biometrika 39: 417-418, 1952.
10. Keuls, M.--"The Use of the Studentized Range in Connection with an Analysis of Variance." Euphytica 1: 112-122, 1952.
11. May, Joyce M.--"Extended and Corrected Tables of the Upper Percentage Points of the Studentized Range." Biometrika 39: 192-193, 1952.
12. Milne, W. E.--Numerical Calculus. Princeton University Press, Princeton, New Jersey, 1949.
13. Nair, K. R.--"The Studentized Form of the Extreme Mean Square Test in the Analysis of Variance." Biometrika 35: 16-31, 1948.
14. Newman, D.--"The Distribution of the Range in Samples from a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation." Biometrika 31: 20-30, 1939.

15. Pearson, E. S. and Hartley, H. O.--"Comparison of Two Approximations of the Distribution of the Range in Small Samples." Biometrika 39: 130-136, 1952.
16. Pearson, E. S. and Hartley, H. O.--"The Probability Integral of the Range in Samples of  $n$  Observations from a Normal Population." Biometrika 32: 301-310, 1942.
17. Pearson, E. S. and Hartley, H. O.--"Tables of the Probability Integral of the Studentized Range." Biometrika 33: 89-99, 1943.
18. Pillai, K. C. S.--"On the Distribution of the Studentized Range." Biometrika 39: 194-195, 1952.
19. Pillai, K. C. S.--"Some Notes on Ordered Samples from a Normal Population." Sankhyā 2: 23-29, 1951.
20. Tippett, L. H. C.--"On the Extreme Individuals and the Range of Samples Taken from a Normal Population." Biometrika 17: 364-387, 1925.
21. Tippett, L. H. C.--The Methods of Statistics. John Wiley and Sons, Inc., New York, 1952.
22. Whittaker, Sir E. and Robinson, G.--The Calculus of Observations. D. Van Nostrand Company, Inc., New York, 1944.

**The vita has been removed from  
the scanned document**