

# Statistics for Experimenters

*An Introduction to Design,  
Data Analysis,  
and Model Building*

GEORGE E. P. BOX  
WILLIAM G. HUNTER  
J. STUART HUNTER

John Wiley & Sons  
New York • Chichester • Brisbane • Toronto • Singapore



11528

T  
o  
i  
n  
F  
i  
t  
h  
o  
f  
a  
s  
c  
h  
a  
p  
b  
c  
s  
e  
l  
e  
M  
e  
s  
t  
a  
s  
t  
e  
o  
f  
A  
n  
f  
o  
r  
s  
u  
r  
r  
e  
p  
r  
e  
s  
e  
n  
t  
i  
n  
g  
(w  
i  
t  
h  
T  
h  
e  
e  
a  
c  
h  
a  
n  
d  
t  
h  
e  
B  
o  
x  
p  
r  
a  
c  
t  
i  
c  
e  
n  
t  
i  
f  
i  
c  
a  
l  
d  
u  
c  
t  
i  
o  
n  
n  
e  
e  
d  
t  
h  
e  
i  
n  
v  
e  
s  
t  
i  
g  
a  
t  
i  
o  
n  
s  
e  
r  
i  
e  
s  
o  
f  
t  
r  
a  
n  
s  
l  
a  
t  
i  
o  
n  
s

A  
b  
o  
l  
G  
E  
O  
f  
e  
s  
s  
o  
W  
i  
s  
c  
o  
n  
s  
u  
l  
t  
i  
n  
g  
P  
h  
D  
t  
i  
c  
s  
f  
r  
o  
m  
H  
o  
n  
o  
r  
a  
r  
y  
R  
o  
c  
k  
e  
t  
e  
r  
i  
e  
n  
c  
e  
B  
r  
i  
t  
i  
s

Copyright © 1978 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data

Box, George E. P.  
Statistics for experimenters.

(Wiley series in probability and mathematical statistics)

Includes index.

1. Experimental design. 2. Analysis of variance. I. Hunter, William Gordon, 1937- joint author. II. Hunter, J. Stuart, 1923- joint author. III. Title.

QA279.B68 001.4'24 77-15087  
ISBN 0-471-09315-7

Printed in the United States of America

20 19 18 17 16 15 14 13

$$\frac{\sqrt{2}|\bar{y}|}{|\bar{y}|}$$

gth of the vector with  $\bar{y}$ . Thus

9  
 which we routinely find  $\bar{y}$ . Notice that, the equiangular line— the vector  $\bar{y}$  is obtained from the equiangular line.

consistent nonzero  $t(t_0/\sqrt{n-1}) =$  the equiangular angle is small. 4], yielding the much larger angle and hence probability.

ts [3.9, 4.1, 4.0] [69, 63.9°, 0.28).

the  $n$  mounts and small angles of the elements of the sphere drawn

the sphere and e. The required fraction of the surface area of

6B.1 indicating components  $\bar{y}$

and  $y - \bar{y}$  are orthogonal and consequently Pythagoras' theory applies. The degrees of freedom indicate the number of dimensions in which the vectors are free to move. Thus before the data are collected the vector  $y$  is unconstrained and has  $n = 3$  degrees of freedom; the vector  $\bar{y}$ , which has elements  $(\bar{y}_1, \bar{y}_2, \bar{y}_3)$  and is constrained to lie on the equiangular line, has only 1 degree of freedom; the vector  $y - \bar{y}$ , which is constrained to lie on a plane perpendicular to  $\bar{y}$ , has  $n - 1 = 2$  degrees of freedom. The analysis of variance of Table 6B.2 conveniently summarizes these facts.

In general, each statistical model discussed in this book determines a certain line, plane or space on which if there were no error the data would have to lie. For the example of this section, for instance, the model is  $y = \eta + \epsilon$ . Thus, without the errors  $\epsilon$ , the data would have to lie on the equiangular line at some point  $[\eta, \eta, \eta]$ . The  $t$  and  $F$  criteria measure the angle that the actual data vector, which is subject to error, makes with the appropriate line, plane and space dictated by the model. The corresponding tables indicate probabilities that angles as small or smaller will occur by chance. These probabilities are dependent on the dimensions of the model and of the data through the degrees of freedom in the table.

Generalization

The vector breakdown of Table 6.6 for the general one-way analysis of variance is a direct extension of that of Table 6B.2. The analysis of variance of Table 6.3 is a direct extension of that of Table 6B.1. The geometry and resulting distribution theory for the general case is essentially an elaboration of that given above.

APPENDIX 6C. MULTIPLE COMPARISONS

Formal procedures for allowing for the effect of selection in making comparisons have been the subject of considerable research (see, e.g., O'Neill and Wetherill, 1971, and Miller, 1977, also the references listed therein).

Confidence Interval for a Particular Difference in Means

A confidence interval for the true difference between the means of, say, the  $p$ th and  $q$ th treatments may be obtained as follows. The observed difference  $\bar{y}_p - \bar{y}_q$  has variance  $\sigma^2(1/n_p + 1/n_q)$ , and  $\sigma^2$  is estimated by the within-treatment mean square  $s^2$ . Thus the estimated variance of  $\bar{y}_p - \bar{y}_q$  is  $s^2(1/n_p + 1/n_q)$ , and a confidence interval for this single preselected difference is provided by

$$(\bar{y}_p - \bar{y}_q) \pm t_{v, \alpha/2} s \sqrt{\frac{1}{n_p} + \frac{1}{n_q}} \tag{6.C1}$$

where  $v = v_R$ , the degrees of freedom associated with  $s^2$ .

For the example discussed in this chapter, a confidence interval for the true difference between the means of treatments  $A$  and  $B$  can be established as follows. We have

$\bar{y}_B - \bar{y}_A = 66 - 61 = 5$ ,  $s_R^2 = 5.6$  with  $v = 20$  degrees of freedom,  $n_B = 6$  and  $n_A = 4$ , and the estimated variance for  $\bar{y}_B - \bar{y}_A$  is  $5.6 (\frac{1}{4} + \frac{1}{6}) = 2.33$ . Thus the 95% confidence limits for the mean difference  $\eta_B - \eta_A$  are  $5 \pm 2.09\sqrt{2.33}$ , that is,  $5 \pm 3.2$ , where 2.09 is the value of  $t$  appropriate for 20 degrees of freedom, which is exceeded, positively or negatively, a total of 5% of the time.

The  $1 - \alpha$  confidence limits calculated in this way will be valid for any single chosen difference; the chance that the specific interval given above includes the true difference  $\eta_B - \eta_A$  on the stated assumptions will be equal to  $1 - \alpha$ . For  $k$  treatments, however, there are  $k(k - 1)/2$  treatment pairs, and the differences between each one of these pairs can be used to construct a confidence interval. Whereas for each interval individually the chance of including the true value is exactly equal to  $1 - \alpha$ , the chance that all the intervals will simultaneously include their true values is less than  $1 - \alpha$ .

*Tukey's Paired Comparison Procedure*

In comparing  $k$  averages, suppose that we wish to state the confidence interval for  $\eta_i - \eta_j$ , taking account of the fact that all possible comparisons may be made. It has been shown by Tukey (1949) that the confidence limits for  $\eta_i - \eta_j$  are then given by

$$(\bar{y}_i - \bar{y}_j) \pm \frac{q_{k,v,\alpha/2}}{\sqrt{2}} s \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \tag{6C.2}$$

where  $q_{k,v}$  is the appropriate upper significance level of the studentized range for  $k$  means, and  $v$  the number of degrees of freedom in the estimate  $s^2$  of variance  $\sigma^2$ . This formula is exact if the numbers of observations in all the averages are equal, and approximate if the averages are based on unequal numbers of observations.

The size of the confidence interval for any given level of probability is larger when the range statistic  $q_{k,v}$  is used rather than the  $t$  statistic, since the range statistic allows for the possibility that any one of the  $k(k - 1)/2$  possible pairs of averages might have been selected for the test. Critical values of  $q_{k,v}/\sqrt{2}$  have been tabulated; see, for instance, Pearson and Hartley (1966), Table 29. As an example, in an experimental program on the bursting strengths of diaphragms the treatments consisted of  $k = 7$  different types of rubber, and  $n = 4$  observations were run with each type. The data were as follows:

treatment $t$	A	B	C	D	E	F	G
average $\bar{y}_t$	63	62	67	65	65	70	60
estimates of variance $s_t^2$	9.2	8.7	8.8	9.8	10.2	8.3	8.0

For this example,  $k = 7$ ,  $s^2 = 9.0$ ,  $v = 21$ ,  $\alpha = 0.05$ , and  $q_{k,v,\alpha/2}/\sqrt{2} = 3.26$ ; these values give for the 95% limits

$$\pm \frac{q_{k,v,\alpha/2}}{\sqrt{2}} \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right) s^2} = \pm 3.26 \sqrt{\left(\frac{1}{4} + \frac{1}{4}\right) 9.0} = \pm 6.91 \tag{6C.3}$$

Thus any observed difference that is statistically significant is likely to be zero. The differences that are statistically significant

treatment  
average  
differences

*Dunnnett's Procedure*

Experimenters often compare the specific treatment means with the overall mean. For example, suppose  $k - 1$  differences  $\bar{y}_t - \bar{y}$  are compared. The 1 -  $\alpha$  confidence limits are given by Equation 6C.1. For the above example

$\pm t$

Therefore a difference can be considered significant

treatment  
average  
differences

Only the differences between treatments and the overall mean are compared

For the special case of all treatments compared to the overall mean, the ratio of the number of treatments to the number of observations  $n_t$ . The ratio of the number of treatments to the number of observations

Thus any observed difference greater in absolute value than 6.91 could be considered statistically significant; hence we could say that the corresponding true difference is not likely to be zero. The  $7 \times 6/2 = 21$  differences are listed in the following table. Those that are statistically significant are circled. The total error rate is  $\alpha = 0.05$ .

treatment	A	B	C	D	E	F	G
average $\bar{y}_i$	63	62	67	65	65	70	60
difference $\bar{y}_i - \bar{y}_j$	*	1	-4	-2	-2	(-7)	3
		*	-5	-3	-3	(-8)	2
			*	2	2	-3	(7)
				*	0	-5	5
					*	-5	5
						*	(10)
							*

*Dunnett's Procedure for Multiple Comparisons with a Standard*

Experimenters often use a control or standard treatment as a benchmark against which to compare the specific treatments. The question then arises whether any of the treatment means may be considered to be different from the mean of the control. In the above example suppose that A was the control. The statistics of interest now are the  $k - 1$  differences  $\bar{y}_i - \bar{y}_A$ , where  $\bar{y}_A$  is the observed average response for the control treatment. The  $1 - \alpha$  confidence intervals for all  $k - 1$  differences from the control are as given by Equation 6C.2, except that the value of  $q_{k, v, \alpha/2} / \sqrt{2}$  is replaced with Dunnett's  $t$ . For tabulated values of this quantity,  $t_{k, v, \alpha/2}$ , see Dunnett (1964). Thus in the above example we have  $t_{k, v, \alpha/2} = 2.80$ , giving for the 95% limits

$$\pm t_{k, v, \alpha/2} S \sqrt{\frac{1}{n_A} + \frac{1}{n_i}} = \pm 2.80 \times 3.00 \sqrt{\frac{1}{4} + \frac{1}{4}} = \pm 5.94 \quad (6C.4)$$

Therefore any observed difference from the control greater than 5.94 in absolute value can be considered statistically significant. The  $k - 1 = 6$  differences are as follows:

treatment	A	B	C	D	E	F	G
	(control)						
average	63	62	67	65	65	70	60
difference	*	1	-4	-2	-2	(-7)	3

Only the difference  $\bar{y}_F - \bar{y}_A$  is indicative of a real difference between the means of six treatments and the control treatment.

For the special case of comparisons against a standard or a control it is good practice to allot more observations  $n_A$  to the control treatment than to each of the other treatments  $n_i$ . The ratio  $n_A/n_i$  should be approximately equal to the square root of the number of treatments, that is,  $n_A/n_i = \sqrt{k}$ .

(6C.3)

*Other Procedures*

Other techniques are also available for making multiple comparisons between treatment averages. One method, to be used only if the  $F$  test has shown evidence of statistically significant differences, is the Newman-Keuls (Newman, 1939, and Keuls, 1952). An alternative has been suggested by Duncan (1955). A method for constructing an interval statement appropriate for *all possible comparisons* among the  $k$  treatments, not merely their differences, has been proposed by Scheffé (1953). The Scheffé method is the most conservative, that is, it produces the widest interval statements.

*Use of Formal Tests for Multiple Comparisons*

In practice it is questionable how far we should go with such formal tests. The difficulties are as follows:

1. How exact should we be about uncertainty? We may ask, for example, "How much difference does it make to know whether a particular probability is exactly 0.04, exactly 0.06, or about 0.05?"
2. Significance levels and confidence coefficients are arbitrarily chosen.
3. In addition to the procedures we have mentioned, others employ still other bases for making multiple comparisons. The subtleties involved are not easy to understand, and the experimenter may find himself provided with an exact measure of the uncertainty of a proposition he does not fully comprehend.

For many practical situations a satisfactory alternative is careful inspection of the treatment averages in relation to a sliding reference distribution, as described in this chapter. The procedure is admittedly approximate, but, we believe, not misleadingly so.

**REFERENCES AND FURTHER READINGS**

An authoritative text on analysis of variance is:

Scheffé, H. (1953). *Analysis of Variance*, Wiley.

For further information on multiple comparisons, see these articles and the references listed therein:

O'Neill, R., and G. B. Wetherill. (1971). The present state of multiple comparison methods, *J. Roy. Stat. Soc., Ser. B*, **33**, 218.

Miller, R. G., Jr., (1977). Developments in multiple comparisons, 1966-1976, *J. Am. Stat. Assoc.*, **72**, 779.

## QUESTIONS FOR CHAPTER

The following are the

- Tukey, J. W. (1949). *Con*  
 Pearson, E. S., O  
 Cambridge University  
 Dunnett, C. W. (1964). *N*  
 Newman, D. (1939). The  
 pressed in terms of a  
 Keuls, M. (1952). The us  
*Euphytica*, **1**, 112.  
 Duncan, D. B. (1955). *M*  
 Scheffé, H. (1953). *A me*  
**40**, 87.

## QUESTIONS FOR

1. What are the basic
2. Invent some data. Can the data be used for these parts? Consider possible shortcomings.
3. What is the usual possible shortcoming?
4. Why is the assumption of the experiment is practical?
5. How is Pythagoras' theorem used?
6. What are residual plots? Why should they be plotted? Why should they be plotted?
7. How can a reference distribution be used for a comparison of  $k$  means? What are the advantages of an analysis of variance?

The following are the references mentioned in Appendix 6C on multiple comparisons:

- Tukey, J. W. (1949). Comparing individual means in the analysis of variance, *Biometrics*, **5**, 99.
- Pearson, E. S., and H. O. Hartley. (1966). *Biometrika Tables for Statisticians*, Vol. 1, 3rd ed., Cambridge University Press.
- Dunnett, C. W. (1964). New tables for multiple comparisons with a control, *Biometrics*, **20**, 482.
- Newman, D. (1939). The distribution of the range in samples from a normal population expressed in terms of an independent estimate of the standard deviation, *Biometrika*, **31**, 20.
- Keuls, M. (1952). The use of the Studentized range in connection with an analysis of variance, *Euphytica*, **1**, 112.
- Duncan, D. B. (1955). Multiple range and multiple *F* tests, *Biometrics*, **11**, 1.
- Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance, *Biometrika*, **40**, 87.

### QUESTIONS FOR CHAPTER 6

1. What are the basic ideas of the analysis of variance?
2. Invent some data for three treatments with four replications each. How can the data vector be decomposed into three separate parts? What are these parts? Construct an analysis of variance table.
3. What is the usual model for a one-way analysis of variance? What are its possible shortcomings?
4. Why is the assumption of normality made in analysis of variance? If the experiment is properly randomized, is this assumption necessary?
5. How is Pythagoras' theorem related to the analysis of variance?
6. What are residuals? How can they be calculated? How can they be plotted? Why should they be plotted?
7. How can a reference distribution diagram be constructed for the comparison of  $k$  means? What can one tell from such a diagram but not from an analysis of variance table?