Ann. Rev. Pharmacol. Toxicol. 1980. 20:441–62

EFFICIENT ANALYSIS OF EXPERIMENTAL OBSERVATIONS

❖6776

W. J. Dixon

Department of Biomathematics, University of California, Los Angeles, California 90024 and VA Brentwood Hospital, Los Angeles, California 90074

INTRODUCTION

The prime goal of this chapter is to encourage researchers in pharmacology to make use of some of the statistical techniques developed in recent years that could increase the efficiency of their analyses.

The topics we have chosen to illustrate are fractional replicates of factorial experiments, single degrees of freedom and half-normal plots in the analysis of variance, robust estimation, and up-and-down method for quantal responses.

Briefly, the advantages of the techniques illustrated here are the following.

Fractional Replicates

A full factorial experiment considers experiments done for every combination of treatment modalities. When the number of treatment factors to be considered is large, the full factorial will require more treatment combinations that can be accommodated. We suggest procedures that economize on the use of subjects but still allow estimates of selected effects.

Single Degrees of Freedom

In the analysis of fractional or full factorials there is a need to pinpoint the basis of statistically significant findings to the design variables that are "producing" the significance. The single degree of freedom analysis can assist in seeking the proper transformation of the observed values and can illuminate the question of whether observed interactions are intrinsic to the variables or are indications of the choice of measurement units.

0362-1642/80/0415-0441$01.00

Up-and-Down Method

For quantal responses this method provides a more efficient use of subjects than the probit designs often used. The technique provides for sequential choice of testing levels, and maximum likelihood solutions for LD_{50} with homogeneous standard errors of estimate that, in turn, allow the use of highly efficient factorial and single degree of freedom analyses.

Robust Estimation

The arithmetic mean is optimum if the data (or errors) are distributed normally (Gaussian). But for many types of naturally occurring data, other measures are better than the mean and increase the efficiency of estimation; that is, a given accuracy or precision can be attained with fewer observations.

FRACTIONAL REPLICATES

We say an experiment is replicated (so many times) when more than one experimental subject is assigned to each separate combination of factors or variables in a design; i.e. in a three-factor problem with high and low levels for each of three chemicals (or treatments) a full factorial would require $2 \times 2 \times 2 = 8$ subjects for a single replicate for each of the eight combinations, and r replicates would require $8r$ subjects. We could consider analyzing less than a single replicate by completing only a balanced subset of the eight design combinations. If we use four combinations, we refer to the design as a half-replicate of a $2 \times 2 \times 2$ design (or one half of a 2^3). The full factorial design for variables *A,*

1. *abc*	5. *a*
2. *ab*	6. *b*
3. *ac*	7. *c*
4. *bc*	8. 1

The first "cell" in the experiment denoted by *abc* indicates that all factors *a, b,* and *c* are at the high level; the second with *a* and *b* at the high level and *c* at the low level, etc. The final indication of all factors at the low level is indicated by the numeral 1. The use of a full factorial allows us to estimate separately the effects of *a, b,* and *c* as well as their possible interactions (synergism or antagonism). We return to the analytical procedure later. The design above can be depicted as in Figure 1.

For example, a 2^3 portion of an experiment on penicillin yield (analyzed in a later section) appears as follows:

Glucose		0		0.5%	
Sodium nitrate		0	0.3%	0	0.3%
Lactose	2%	142	148	106	101
	3%	129	146	88	140

Suppose we do only the four experiments:

abc *c* *a* *b*

We have two tests at the high and low level for each factor but not all possible combinations. To the extent that higher order interactions are not present among the chemicals (or treatments) we can still get valid estimates of the main effects of each chemical.

LOW C

A \ B	Hi	Low
Hi	ab	a
Low	b	1

HIGH C

A \ B	Hi	Low
Hi	abc	ac
Low	bc	c

Figure 1 **Illustration of a 2^3 design.**

Similar designs can be found for experiments with other fractions or with more levels (see 1), but here we wish only to illustrate the concepts.

The model under consideration is that an observation y (aside from measurement error) is the sum of effects associated with *A,*
interactions; i.e. one effect representing the grand mean, three effects representing the possible nonzero overall effects for *A,*
effects representing possible interactions of the main factors. The variance for measurement error can, of course, be separately estimated only if there is more than one replicate. For fractional replicates, the effects cannot be estimated separately. For a more complete discussion, see e.g. (2) or (3).

An analysis of variance table for the complete 2^3 design is shown in Table 1 for a measurement y. A simple computational procedure is available for the single degree of freedom *SS* (sum of squares). The *MS* (mean square) is the *SS* divided by the degrees of freedom. We introduce a set of coefficients x for each of the factors in the experiment (and combinations with other factors). A full set of seven forms the basis of the analysis, and the associated *SS* for each provides a partition of the total $SS = \Sigma(y - \bar{y})^2$.

Returning to the eight experimental conditions,

y	cell	x_A	x_B	x_C	x_{AB}	x_{AC}	x_{BC}	x_{ABC}
140	abc	1	1	1	1	1	1	1
101	ab	1	1	−1	1	−1	−1	−1
88	ac	1	−1	1	−1	1	−1	−1
146	bc	−1	1	1	−1	−1	1	−1
106	a	1	−1	−1	−1	−1	1	1
148	b	−1	1	−1	−1	1	−1	1
129	c	−1	−1	1	1	−1	−1	1
142	1	−1	−1	−1	1	1	1	−1
Σxy		130	70	6	24	36	68	46
Σx^2		8	8	8	8	8	8	8

we define a set of coefficients x for each factor (contrast). The first x above for A is +1 when A is at a high level and −1 when A is at a low level, and similarly for the x for B and C. The coefficients x for AB are +1 when both A and B are either high or low, and −1 otherwise. A simple rule for the interactions is to record the product of the entries for the factors in the interaction; that is, the coefficients for AB are the products of the entries for A and B.

The SS is computed as $(\Sigma xy)^2/\Sigma x^2$ for each of the columns of x's.

The fact that all combinations of factors are present in a balanced way allows a separate estimation for each. The mathematical basis is the fact that the above x's are orthogonal (e.g. $\Sigma x_i x'_i = 0$).

The analysis of variance table in Table 1 is presented to illustrate the computational method. The analysis of significant effects is described for subsequent examples.

For any subset of experimental combinations we can verify which effects are confounded (i.e. cannot be estimated separately).

If we do the four experiments as mentioned above,

Table 1 Form of analysis for a 2^3 design

	Analysis of variance		
Source	SS	df	MS
A (glucose)	2,112.5	1	2,112.5
B (nitrate)	612.5	1	612.5
C (lactose)	4.5	1	4.5
AB	72.0	1	72.0
AC	162.0	1	162.0
BC	578.0	1	578.0
ABC	264.5	1	264.5
Total $\Sigma(y - \bar{y})^2$	3,806	7	

y	Expt.	x_A	x_B	x_C	x_{AB}	x_{AC}	x_{BC}	x_{ABC}
140	abc	1	1	1	1	1	1	1
129	c	−1	−1	1	1	−1	−1	1
106	a	1	−1	−1	−1	−1	1	1
148	b	−1	1	−1	−1	1	−1	1
Σxy		−31	53	15	15	53	−31	523

we can see that x_A is the same as x_{BC} and, in fact, the confounded pairs are

$$A = BC$$
$$B = AC$$
$$C = AB$$

so the analysis of variance table becomes:

Source	SS	df	MS
$A = BC$	240.25	1	240.25
$B = AC$	702.25	1	702.25
$C = AB$	56.25	1	56.25
Total $\Sigma(y - \bar{y})^2$	998.75	3	

and we can estimate A,
between the variables. In our example we do not know whether the difference between $abc + a$ and $c + b$ is due to an a effect or an interaction between b and c.

In larger experiments we can also estimate some of the interactions. For example, in a 2^5 experiment, a half-replicate has 16 combinations but since there are only 5 main effects to estimate, we can estimate some of the interactions. For a fuller treatment see (2) or (3).

Even if the complete factorial is contemplated, there is a distinct advantage to separating the experiment into half- or quarter-replicates to be done at different times (laboratories, technicians), thus having a separate SS for the variation attributable to this additional factor.

Table 2 shows the original data and the 2^5 design used in the study of penicillin yields in a surface culture experiment (3). The two half-factorials are indicated by placing the observation to the right or left of the cell. The circled design entries indicate one of the quarter-factorials.

Table 3 shows the logarithmic transformation of the yields. The transformation was used because the error was expected to be proportional to the resulting observations. See (4, 5) for a discussion of transformations.

In Table 4 the figures are further modified by subtracting 2 and removing the decimal point to simplify computing and reporting. This transformation

Table 2 Yields of penicillin in surface culture experiment*

Data

C	B	A	No glucose: No	No glucose: 0.3%	0.5% glucose: No	0.5% glucose: 0.3%	= E / = D
0.00	2%	2%	142	148	106	101	
		3%	114	108	106	114	
	3%	2%	129	146	88	140	
		3%	109	95	98	72	
0.05	2%	2%	185	200	113	130	
		3%	162	164	88	83	
	3%	2%	200	215	166	145	
		3%	172	118	79	110	

Design

C	B	A	No glucose: No	No glucose: 0.3%	0.5% glucose: No	0.5% glucose: 0.3%	= E / = D
0.00	2%	2%	(1)	*d*	*e*	(*de*)	
		3%	*a*	*ad*	*ae*	*ade*	
	3%	2%	*b*	*bd*	*be*	*bde*	
		3%	(*ab*)	*abd*	*abe*	(*abde*)	
0.05	2%	2%	*c*	(*cd*)	(*ce*)	*cde*	
		3%	*ac*	*acd*	*ace*	*acde*	
	3%	2%	*bc*	*bcd*	*bce*	*bcde*	
		3%	*abc*	(*abcd*)	(*abce*)	*abcde*	

* A = strength of corn steep liquor; B = lactose; C = precursor; D = $NaNO_4$; E = glucose.

Table 3 Logarithms of yields of Table 2

		(1)			*d*			*e*			*de*	
i		(1)	2·152	2·170	(*d*)		2·025	(*e*)			(*de*)	2·004
a	2·057	(*a*)			(*ad*)	2·033		(*ae*)	2·025	2·057	(*ade*)	
b	2·111	(*b*)			(*bd*)	2·164		(*be*)	1·944	2·146	(*bde*)	
ab		(*ab*)	2·037	1·978	(*abd*)		1·991	(*abe*)			(*abde*)	1·857
c	2·267	(*c*)			(*cd*)	2·301		(*ce*)	2·053	2·114	(*cde*)	
ac		(*ac*)	2·210	2·215	(*acd*)		1·944	(*ace*)			(*acde*)	1·919
bc		(*bc*)	2·301	2·332	(*bcd*)		2·220	(*bce*)			(*bcde*)	2·161
abc	2·236	(*abc*)			(*abcd*)	2·072		(*abce*)	1·898	2·041	(*abcde*)	

Table 4 Simplificd version of Table 2

		x-vectors							
y'	Cell	A	C	D	E	AC	AD	AE	I
152	1	–	–	–	–	+	+	+	+
37	*ab*	+	–	–	–	–	–	–	+
301	*cd*	–	+	+	–	–	–	+	+
53	*ce*	–	+	–	+	–	+	–	+
4	*de*	–	–	+	+	+	–	–	+
–143	*abde*	+	–	+	+	–	+	+	+
–102	*abce*	+	+	–	+	+	–	+	+
72	*abcd*	+	+	+	–	+	+	–	+
210	*ac*	+	+	–	–	+	–	–	+
33	*ad*	+	–	+	–	–	+	–	+
25	*ae*	+	–	–	+	–	–	+	+
301	*bc*	–	+	–	–	–	+	+	+
164	*bd*	–	–	+	–	+	–	+	+
–56	*be*	–	–	–	+	+	+	–	+
161	*bcde*	–	+	+	+	–	–	–	+
–81	*acde*	+	+	+	+	+	+	+	+
267	*c*	–	+	–	–	–	+	+	+
170	*d*	–	–	+	–	+	–	+	+
25	*e*	–	–	–	+	+	+	–	+
236	*abc*	+	+	–	–	+	–	–	+
–22	*abd*	+	–	+	–	–	+	–	+
–9	*abe*	+	–	–	–	+	–	+	+
114	*cde*	–	+	+	+	–	–	–	+
41	*abcde*	+	+	+	+	+	+	+	+
57	*a*	+	–	–	–	–	–	–	+
111	*b*	–	–	–	–	+	+	+	+
215	*acd*	+	+	+	–	+	+	–	+
–56	*ace*	+	+	–	+	+	–	+	+
57	*ade*	+	–	+	+	–	+	+	+
332	*bcd*	–	+	+	–	–	–	+	+
220	*bce*	–	+	–	+	–	+	–	+
146	*bde*	–	–	+	+	+	–	–	+

is linear and will not change the comparative size of the *MS* entries in the subsequent analysis. The $x-$ vectors in Table 4 are recorded as + for 1.0 and – for –1.0. Table 5 contains $\Sigma xy'$ separately for each of the quarter-replicates, the two half replicates, and the full factorial. The experiment was conducted with half-replicates in two different weeks so the *ABCDE* effect is a block effect for the two weeks. Each set of eight entries is a quarter-replicate and the half-replicates combine the first two and last two quarter-replicates.

One can either use tables to find which cells (combinations of levels) to use for suitable fractional replicates or find the confounded "aliases" by characteristic equations [see refs. (1) or (3)].

Table 6 shows the single degree of freedom (d.f.) mean squares with indications of the confounding of effects. In the quarter-replicates, for example, *A* is listed parenthetically for *B, ACDE,* and *BCDE* to indicate that

Table 5 Values of $\Sigma xy'$

Quarter-replicates								
Row	A	C	D	E	AC	AD	AE	I
1	–646	274	94	–750	–122	–106	42	374
2	–383	425	–203	–659	–283	–363	61	757
3	–330	494	–216	–480	122	–200	116	822
4	–536	340	418	–348	–250	124	–194	1,082
Half-replicates								
Row	A	C	D	E	AC	AD	AE	I
5 = 1 + 2	–1,029	699	–109	–1,409	–405	–469	103	1,131
6 = 3 + 4	–866	834	202	–828	–128	–76	–78	1,904
	B	DE	CE	CD	BC	BD	BE	AB
7 = 1 – 2	–263	–151	297	–91	–161	257	–19	–383
8 = 3 – 4	206	154	–634	–132	372	–324	310	–260
Full-factorial								
Row	A	C	D	E	AC	AD	AE	I
9 = 5 + 6	–1,895	1,533	93	2,237	–533	–545	25	3,035
	B	ABC	ABD	ABE	ADE	BD	BE	AB
10 = 7 + 8	–57	3	–337	–223	244	–67	291	–643
	BCDE	ABDE	ABCE	ABCD	BDE	BCE	BCD	ABCDE
11 = 5 – 6	163	135	311	581	277	393	–181	–773
	ACDE	DE	CE	CD	BC	ACE	ACD	CDE
12 = 7 – 8	–469	–305	931	41	–533	581	–329	–123

Table 6 Mean squares for single degrees of freedom

Effect	Quarter-replicates*				Half-replicates†		Full
A	0.0522	0.0183	0.0136	0.0359	0.0662	0.0469	0.1122
B	(A)				0.0043	0.0027	0.0001
C	0.0094	0.0226	0.0305	0.0144	0.0305	0.0435	0.0734
D	0.0011	0.0052	0.0058	0.0218	0.0007	0.0026	0.0003
E	0.0703	0.0543	0.0288	0.0151	0.1241	0.0428	0.1564
AB	(I)				0.0092	0.0042	0.0129
AC	0.0019	0.0100	0.0019	0.0078	0.0103	0.0010	0.0089
AD	0.0014	0.0165	0.0050	0.0019	0.0137	0.0004	0.0093
AE	0.0002	0.0005	0.0017	0.0047	0.0007	0.0004	0.0000
BC	(AC)				0.0016	0.0086	0.0089
BD	(AD)				0.0041	0.0066	0.0001
BE	(AE)				0.0000	0.0060	0.0026
BC	(E)				0.0005	0.0011	0.0001
CE	(D)				0.0055	0.0251	0.0271
DE	(C)				0.0014	0.0015	0.0029
ABC	(C)				(DE)		0.0000
ABD	(D)				(CE)		0.0035
ABE	(E)				(CD)		0.0016
ACD	(AE)				(BE)		0.0034
ACE	(AD)				(BD)		0.0105
ADE	(AC)				(BC)		0.0019
BCD	(AE)				(AE)		0.0010
BCE	(AD)				(AD)		0.0048
BDE	(AC)				(AC)		0.0024
CDE	(I)				(AB)		0.0005
ABCD	(E)				(E)		0.0105
ABCE	(D)				(D)		0.0030
ABDE	(C)				(C)		0.0006
ACDE	(A)				(B)		0.0069
BCDE	(A)				(A)		0.0008
ABCDE	(I)				(I) (0.0187)		0.0187

*I = AB = ABCDE = CDE.
†I = ABCDE.

the mean square for *A* has all four components. Likewise for the half-replicate *A* and *BCDE* are confounded. For example, the first quarter-replicate estimate of *A* is $-152 + 37 = 301 - 53 - 4 + (-143) + (-102) + 72 = -646$. We continue with the analysis of the full factorial. Further analyses are given in the next section.

Davies (3) grouped all two-factor interactions and all three- and four-factor interactions to obtain a denominator mean square for tests of significance. His decision was likely based on the appearance of homogeneity of the mean squares, indicating the absence of real interactions at these levels. The three- and four-factor interactions average 0.0034 with 15 degrees of freedom (d.f.) and the two-factor interactions (without *CE*) average 0.0051 and 9 d.f.

The F value for CE 0.0271/0.0034 with 1 and 15 d.f. is 8.0 with $F_{0.95}$ (1,15 d.f.) = 4.54. The combined MS for 24 d.f. is 0.0040 with

$$\begin{aligned} F_A &= 0.1122/0.0040 = 28.0 \\ F_C &= 0.0734/0.0040 = 18.4 \\ F_E &= 0.1564/0.0040 = 39.1 \\ F_{CE} &= 0.0271/0.0034 = 6.8 \\ F_{ABCDE} &= 0.0187/0.0040 = 4.7 \end{aligned} \qquad \begin{cases} F_{0.95} = 4.26 \\ F_{0.99} = 7.82 \end{cases}$$

We need to be concerned about the probability levels when multiple tests are made since the standard tables of probability levels refer to a single test of significance being made in an experiment. We want to make multiple tests. Furthermore, we want to test only the largest of the mean squares in the table. The use of single degrees of freedom as illustrated in the next section clarifies the situation.

SINGLE DEGREES OF FREEDOM

The mean square values illustrated in the previous section were computed using a set of orthogonal vectors x. These vectors transform the observations into quantities that have a normal distribution with squares that are chi-square with one degree of freedom independently distributed as chi-square under the null hypothesis of no effects, and under the assumption that the original observations are independent with normally distributed errors. With a large number of such quantities or statistics it is preferable to test whether they form a set of statistics with the same expectation. This avoids the problem of forming many F tests with the consequent result of a number of null effects being declared significant by chance. The test of equal expectation can be accomplished by comparing their cumulative sampling distribution with that expected from theory. This is most simply done by forming a cumulative distribution polygon for the square roots of the mean squares plotted on half-normal probability paper. A straight line indicates normality of the square roots and therefore uniformity of the mean squares; that is, if all null hypotheses are true the plotted points follow a straight line.

Half-normal paper is appropriate since the square root of chi-square is expected to be distributed as the positive half of a normal distribution. Figure 2 shows that half-normal plot for the full factorial. The points are plotted at fractions 1/15, 3/15, 5/15, . . . for the quarter-replicates, at 1/31, 3/31, 5/31, . . . for the half-replicates and at 1/63, 3/63, 5/63, . . . for the full factorial. The points are plotted with ordinate at the midproportion values; for example, for eight points the first is plotted at 1/16 and the second at 1/16 + 1/8 = 3/16, etc. The abscissa is the square root of the

mean squares. The lines are drawn by eye to fit the approximate "linear" appearance of the lower points, which are assumed to be free of experimentally induced effects and are therefore to be used to estimate the basic measurement standard deviation. The points representing *A, C,* and *E* are distant from the line and provide a clear indication of the presence of effects *A, C,* and *E* (i.e. variance beyond the basic measurement error), with very little effect from week to week (*ABCDE*). The interaction *CE* is larger than expected, but likely indicates no important effect since both *C* and *E* appear as significant main effects. Figure 3 shows the two half-replicates plotted separately. Either half-replicate would indicate the presence of effects *A, C,* or *E.* One replicate indicates an interaction *CE* whereas the other does not. In hindsight we can see that a half-replicate might have sufficed for this study. Figure 4 shows the half-normal plots for the quarter-replicates. As can be seen from the plots it is unlikely that the *A, C,* and *E* main effects

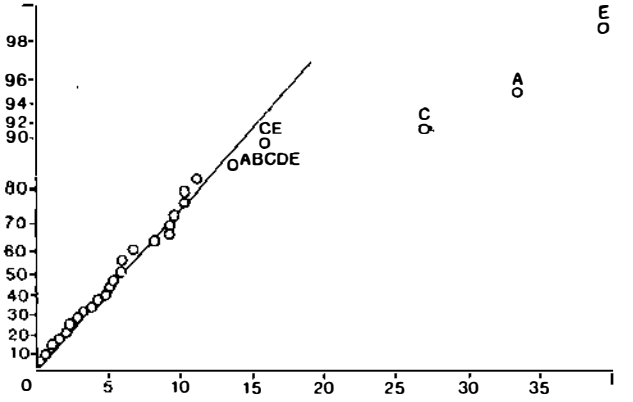

Figure 2 Half-normal plot for penicillin full factorial.

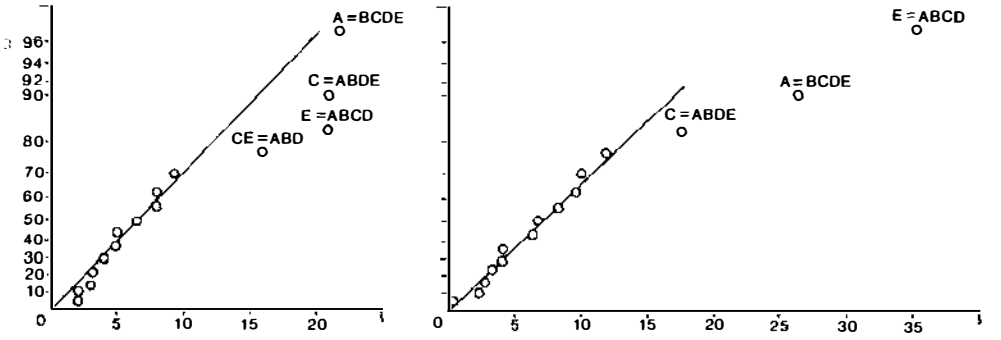

Figure 3 Half-normal plot for penicillin half-factorials.

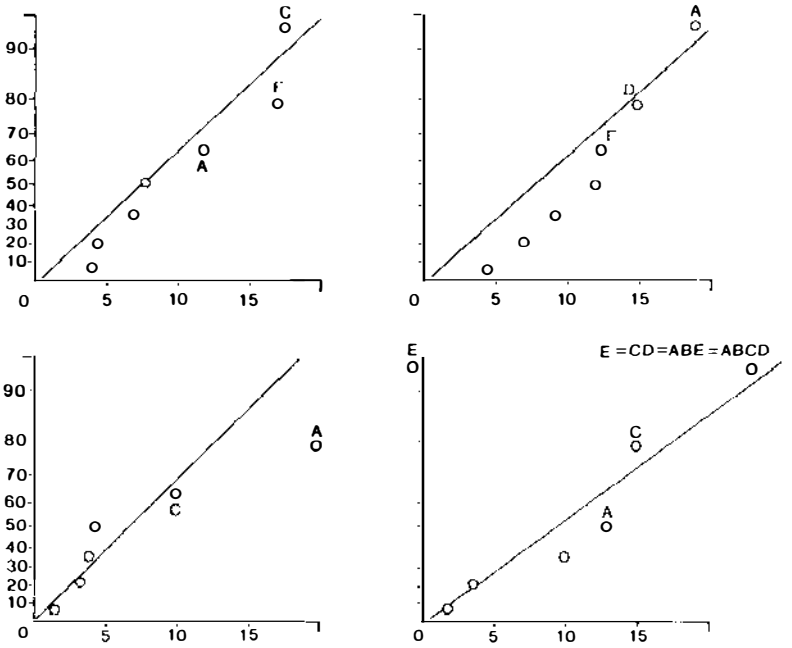

Figure 4 Half-normal plots for quarter-replicates.

would have been detected. Only one of the quarter-replicates appears to indicate significantly large *A* and *E* effects. We can see that in this case a single quarter-replicate does not provide an adequate analysis. On the belief that Figure 2 clearly indicates that *A, C,* and *E* effects are not zero, it is useful to replot the remaining 28 effects (as a set of 28 rather than the first 28 of 31) for consistency of expectations as shown in Figure 5. We now see no indication of effects for the *CE* interaction or the week to week effect *ABCDE.* This is a good example of how to avoid the possible inappropriate indication of significance when numerous tests are made on the same body of data. This discussion is based on subjective examination of the plots. Significance of the departures can be obtained as in (6). The full factorial shows *E, A,* and *C* to be significant; the half-factorials show C and E in one case and A and E in the other.

THE UP-AND-DOWN METHOD

We now shift to the analysis of observations resulting from sensitivity experiments, where the response is all-or-none (e.g. death), and show how

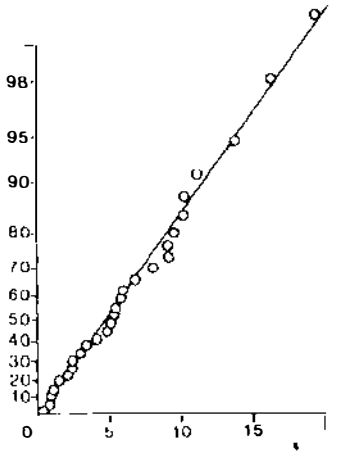

Figure 5 Replot of full factorial.

the technique of analyzing single degrees of freedom can be applied in a bioassay problem.

The usual method for performing a bioassay analysis of a sensitivity experiment is to test a prescribed number of animals at each of several fixed dose levels. In the up-and-down method the dose levels are determined sequentially; under most circumstances the accuracy of the usual method can be obtained with less experimentation.

The improved estimates (7) are available directly from tables. One merely records the sequence of outcomes (O for alive and X for dead) and reads the estimate from Table 7. For example, suppose a series of concentrations of 1, 2, 4, 8, and 16% was used for a drug known to require analysis as log dose. Suppose testing began at 8% or log dose = 0.903, and proceeded with results as shown in Figure 6; that is, tests were performed at 0.903, 1.204, 0.903, 602, 0.903, 0.602 on successive animals. The estimate is 0.602 + 0.831 (0.301) = 0.852 from the rule (final test level) + (value from table) (difference between dose levels) = LD_{50}.

The following steps are used in an experiment using the up-and-down procedure:

1. A series of test levels is chosen with equal spacing between doses (equal spacing on the appropriate scale, usually log-dose). This spacing is chosen approximately equal to σ (where σ is the standard deviation of the distribution of threshold response levels. More simply stated, there should be two or three levels at which some animals respond and some do not. It is not too important if the interval is actually incorrect by as much as 50%.

2. A series of trials is carried out using this rule: Increase the dose following a negative response and decrease the dose following a positive

Table 7 Maximum likelihood estimates of LD_{50}*

N	Second part of series	k for test series whose first part is O	OO	OOO	OOOO	Second part of series	Standard error of LD_{50}
2	X	−0.500	−0.388	−0.378	−0.377	O	0.88 σ
3	XO	0.842	0.890	0.894	0.894	OX	0.76 σ
	XX	−0.178	0.000	0.026	0.028	OO	
4	XOO	0.299	0.314	0.315	0.315	OXX	0.67 σ
	XOX	−0.500	−0.439	−0.432	−0.432	OXO	
	XXO	1.000	1.122	1.139	1.140	OOX	
	XXX	0.194	0.449	0.500	0.506	OOO	
5	XOOO	−0.157	−0.154	−0.154	−0.154	OXXX	0.61 σ
	XOOX	−0.878	−0.861	−0.860	−0.860	OXXO	
	XOXO	0.701	0.737	0.741	0.741	OXOX	
	XOXX	0.084	0.169	0.181	0.182	OXOO	
	XXOO	0.305	0.372	0.380	0.381	OOXX	
	XXOX	−0.305	−0.169	−0.144	−0.142	OOXO	
	XXXO	1.288	1.500	1.544	1.549	OOOX	
	XXXX	0.555	0.897	0.985	1.000^{+1}	OOOO	
6	XOOOO	−0.547	−0.547	−0.547	−0.547	OXXXX	0.56 σ
	XOOOX	−1.250	−1.247	−1.246	−1.246	OXXXO	
	XOOXO	0.372	0.380	0.381	0.381	OXXOX	
	XOOXX	−0.169	−0.144	−0.142	−0.142	OXXOO	
	XOXOO	0.022	0.039	0.040	0.040	OXOXX	
	XOXOX	−0.500	−0.458	−0.453	−0.453	OXOXO	
	XOXXO	1.169	1.237	1.247	1.248	OXOOX	
	XOXXX	0.611	0.732	0.756	0.758	OXOOO	
	XXOOO	−0.296	−0.266	−0.263	−0.263	OOXXX	
	XXOOX	−0.831	−0.763	−0.753	−0.752	OOXXO	
	XXOXO	0.831	0.935	0.952	0.954	OOXOX	
	XXOXX	0.296	0.463	0.500	0.504^{+1}	OOXOO	
	XXXOO	0.500	0.648	0.678	0.681	OOOXX	
	XXXOX	−0.043	0.187	0.244	0.252^{+1}	OOOXO	
	XXXXO	1.603	1.917	2.000	2.014^{+1}	OOOOX	
	XXXXX	0.893	1.329	1.465	1.496^{+1}	OOOOO	
		X	XX	XXX	XXXX		Second part of series
		−k for series whose first part is					

*Values of k for estimating LD_{50} from up-and-down sequence of trials of nominal length N. The estimate of LD_{50} is $x_f + kd$ where x_f is the final test level and d is the interval between dose levels. If the table is entered from the foot, the sign of k is to be reversed.

Log dose	Results of tests					
1.204		X				
0.903	0		X		X	
0.602				0		0
0.301						
0						

Figure 6 Example of test series. For this series OXXOXO the estimate of LD_{50} is 0.602 + 0.831 (0.301) = 0.852.

response. The first test should be performed at a level as near as possible to the LD_{50}.

3. The number N' is the total number of tests performed in each series. Testing continues until a chosen "nominal" sample size = N is reached. N is the total number of trials reduced by one less than the number of like responses at the beginning of the series. The example in Figure 6 is N = 6. For the series OOOXXOXO $N' = 8$ and $N = 6$.

4. The resulting configuration of responses and nonresponses for each series is referred to the table of maximum likelihood solutions (Table 7) for the LD_{50}, and $X_f + kd$ is computed, where X_f is the last dose administered, k is the tabular value, and d is the interval between doses. Table 7 lists all solutions for all N' and for $N \leqslant 6$. If the series begins with more than four like responses (i.e. $N' - N \geqslant 3$) the entry in the final column of Table 7 can be used (except for five tabular entries where an additional increment in the third decimal place is indicated).

5. The estimates have homogeneous variance when tests are performed for a fixed nominal sample size and thus are easily used in further analyses to investigate the dependence of these LD_{50}'s on related experimental factors. For $N > 6$, additional tables are included in Dixon (7), but it is usually preferable to use $N \leqslant 6$ and associate the separate short runs with design variables (either covariates or time); i.e. separate the components of variation due to the design variables of an experiment by using factorial designs, Latin squares, etc.

Most other estimating procedures for this case, probit analysis, logit analysis, etc [see e.g. (8)], as well as up-and-down procedures (9, 10), result in estimates with widely varying standard error making it difficult to use these estimates in comparative analyses. Furthermore, the estimates obtained are in general very inefficient.

The estimates provided in Table 7 are maximum likelihood estimates for each possible configuration of responses, assuming a normal cumulative distribution of response thresholds.

An experiment performed by Dr. Donald J. Jenden, Department of Pharmacology, UCLA, illustrates the use of the up-and-down method for

a study of the dependence of time to death on the dose of ryanodine. Since several series of tests could be in progress at the same time, additional animals could be treated without waiting for the outcome of each separate trial. Also, it was not necessary to use the same test levels for all series. However, previous experimentation had indicated a standard deviation near 1.0, so a common dose interval, $d = 1$, was used for log dose for all trials. Ryanodine was administered intravenously to male mice in volume 0.1 to 0.2 ml in saline. The end point was considered to be the time of last visible movement. Four cut-off points with equal spacing in log time, 64, 96, 144, and 216 sec were chosen for observing the status of the animal. Dosage was computed by body weight, and a randomized block experiment with body weight as the blocking variable was used to obtain information on the validity of the use of the body weight basis for dosage. The results of the tests are given in Table 8. The analysis indicates a time-dose dependence that is linear between log dose and log time and is established with an efficient use of animals and with a small standard error. See Dixon (7) for a discussion of the indications in this experiment for improved dosage intervals.

The analysis of variance was completed in a similar manner to that shown for the penicillin study above, i.e. single degrees of freedom for the time and weight variables. The weight variable has three levels requiring two degrees of freedom coded –1, 0, 1 and 1, –2, 1 labeled L_w and Q_w for linear and quadratic effects. The time variable has four levels equally spaced in log time and one coded –3, –1, 1, 3 for L_t; 1, –1, –1, 1 for Q_t; and –1, 3, –3, 1 for C_t representing three degrees of freedom for linear, quadratic, and cubic effects in log time.

Figure 7 shows the half-normal plot for eleven single degrees of freedom.

Example Using a Magic Square Design

In experimentation with the venom of the scorpion fish, Dr. Peter B. Taylor faced the problem of rapidly declining potency of the venom. The venom is difficult to collect and a most efficient design was needed. The effects of concentration and body weight also needed to be controlled. He used a magic square to allow for decreases in potency as well as to study dependence on concentration and body weight. The magic square specifies ordering the tests so that tests at each level of concentration and each level of body weight are conducted at times to have the same average. The individual series of trials were carried out in the order designated as "Design" in Table 9. The test results and estimates of LD_{50} are included. Table 9 also indicates coefficients for eight single degrees of freedom. The spacing of levels used was $d = 0.406$.

Table 8 Analysis of ryanodine experiment

Weight (in grams)		Time in seconds to cut-off point 64	96	144	216
	Tests	OOOXXOX	OOOOXXOX	OOXXOO	XXOXXO
18–20	x_f	2.000	−0.107	−3.213	−7.213
	k	−0.144	−0.142	0.372	0.861
	Est.	1.856	−0.249	−2.841	−6.352
	Tests	OOOXOXX	OXXXX	XOXXX	XXOXOX
21–23	x_f	4.000	−1.107	−5.213	−6.213
	k	0.181	0.555	0.157	−0.737
	Est.	4.181	−0.552	−5.056	−6.950
	Tests	XXXOXXO	XXOXXO	OXXOX	OXOXX
24–26	x_f	1.000	−3.107	−4.213	−6.213
	k	0.860	0.861	−0.305	0.084
	Est.	1.860	−2.246	−4.518	−6.129

Effect	*SS*
L_t	139.086
Q_t	1.286
C_t	.010
L_w	1.485
Q_w	.145
L_tL_w	.021
Q_tL_w	1.902
C_tL_w	.014
L_tQ_w	4.159
Q_tQ_w	.877
C_tQ_w	.338

Analysis of variance

Source	*SS*	*df*	*MS*
L_t	139.086	1	139.086
L_w	1.485	1	1.485
Remainder	8.752	9	0.972

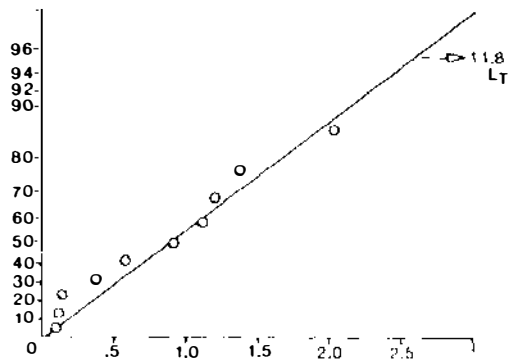

Figure 7 Half-normal plot for ryanodine experiment.

Table 9 Analysis of scorpion fish venom

Weight (in grams)		Venom concentration (ml C/25 g mouse) C	C/2	C/4
15–19	Design	1	5	9
	Tests	XXOXOX	OXXOO	OXOOX
	$ln\ LD_{50}$	−3.806	−3.270	−2.813
20–22	Design	8	3	4
	Tests	OXOXO	XOOXO	OOXXOX
	$ln\ LD_{50}$	−3.242	−3.689	−3.411
23–26	Design	6	7	2
	Tests	XOOXX	XXOOX	OXXOX
	$ln\ LD_{50}$	−3.631	−3.442	−3.689

Design	1	2	3	4	5	6	7	8	9	SS
Linear lability	−4	−3	−2	−1	0	1	2	3	4	0.5202
Quad. lability, 1	1	1	1	−2	−2	−2	1	1	1	0.0002
Quad. lability, 2	1	−2	1	1	−2	1	1	−2	1	0.0084
Linear weight	1	−1	0	0	1	−1	−1	0	1	0.1270
Quad. weight	1	1	−2	−2	1	1	1	−2	1	0.0001
Linear concentration	1	−1	0	−1	0	1	0	1	−1	0.0978
Quad. concentration	1	1	−2	1	−2	1	−2	1	1	0.0024
Remainder	−2	1	4	−3	0	3	−4	−1	2	0.0002
Total										0.7564

Source	Analysis of variance *SS*	*df*	*MS*
Linear lability	0.5202	1	0.5202
Linear weight	0.1270	1	0.1270
Linear concentration	0.0978	1	0.0978
Residual	0.0114	5	0.0023
	0.7564		

$F_{0.95}(1,5) = 6.61$

This analysis indicates an efficient design and analysis of an experiment with two variables, weight and log concentration, with a third variable, time, to account for lability loss during the course of the experiment.

Figure 8 shows the half-normal plot for eight single degrees of freedom. This plot shows that effects are present for weight and concentration as well

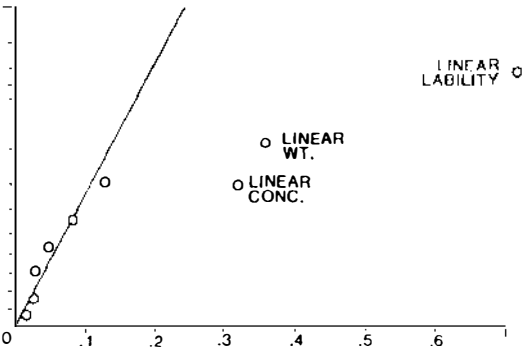

Figure 8 Half-normal plot for venom experiment.

as the expected loss in lability, but all effects appear to be linear since the other five single degrees of freedom do not indicate significant effects. The coefficients for the single degrees of freedom can be obtained by considering the design as a modification of a Greco-Latin square.

ROBUST ESTIMATION

One of the goals of the analysis of variance above was the estimation of the common variance or measurement error, i.e. the basic variability attributable to the individual experiment, separate from design effects in the overall experiment. We face a similar problem in examining a set of observations, all supposedly taken under the same conditions (like observations in one cell), but clearly not satisfying the assumption of normally distributed measurement errors.

The arithmetic mean is an optimum estimator of the average for normally distributed errors. It is not optimum otherwise, and estimates that have better properties are easily found.

Normality specifies a functional form for the errors. Two main properties of normality, symmetry and very low probabilities of very extreme observations, do not accurately describe biological measurements. However, satisfactory symmetry often results from logarithmic transformations, as in the penicillin example. The frequent occurrence of extremes (usually positive without transformation, but in either direction after transformation) has stimulated an interest in robust estimation. The robust estimators do not weight observations equally, as does the arithmetic mean, but reduce the weights on the more extreme observations. Many of these estimates require an iterative computation to adjust the weights to suit each particular set of observations. However, there are two frequently used estimates of aver-

age with predetermined weights. These are the trimmed and Winsorized means.

The trimmed mean gives equal weight to the central observations and zero weight to the extremes. The once-trimmed mean gives zero weight to the largest and smallest observations. The twice-trimmed mean gives zero weight to two largest and two smallest extremes, etc.

The first Winsorized mean relocates the smallest and largest observation to their nearest neighbors and computes the mean. The second Winsorized mean relocates the two largest and two smallest observations, etc.

Little information is lost by using these estimates in place of the mean if the observations have a normal distribution, but much is gained when the observations are from distributions that have "larger tails" than a normal distribution. Efficiencies under normality of the trimmed and Winsorized means compared to the arithmetic mean from Dixon & Yuen (7) follow.

Amount of trim (Wins.)	Efficiencies			
	$N = 10$		$N = 20$	
	Trim	Wins.	Trim	Wins.
1	0.95	0.96	0.98	0.98
2	0.88	0.89	0.95	0.96
3	0.81	0.82	0.92	0.94
4	0.72	0.72	0.88	0.91

Efficiencies are defined as the ratios of the variance of the mean to the variances of the two alternate measures. If data are normally distributed, an efficiency of 95% indicates that approximately 5% more observations are required to obtain similar accuracy using the trimmed mean in place of the arithmetic mean.

On the other hand, the trimmed and Winsorized means can have much greater efficiency than the arithmetic mean when the data do not follow a normal distribution. For example, in a sampling study of the efficiencies of these robust estimates compared to the arithmetic mean for laboratory determinations of blood chemistries Hill & Dixon (11) showed that the robust estimates had approximately twice the efficiency of the arithmetic mean. This and other studies have investigated the behavior of other robust estimates [see e.g. (12)].

However, the present stage of research on robust estimates is sufficient to indicate that trimming (or Winsorizing) about 15% or 20% of the observations is a good rule of thumb for optimum results.

Since these robust estimates are now provided by standard computer packages (e.g. BMDP) investigators should, at a minimum, examine the consistency of the various estimates reported. Some computer programs also provide confidence limits based on these estimates.

Although the resulting estimates are known to have smaller standard error, we like to make the usual t tests and state confidence intervals for the estimates.

Denote $\bar{x}_{tg}$ as the g-times trimmed mean,
$\bar{x}_{wg}$ as the g-times Winsorized mean,
SSD_{wg} as the g-times Winsorized sum of squares of deviations.

$$SSD_{wg} = (g+1)(x_{g+1} - \bar{x}_{wg})^2 + (x_{g+2} - \bar{x}_{wg})^2 + \ldots + (x_{n-g+1} - \bar{x}_{wh})^2 + (g+1)(x_{n-g} - \bar{x}_{wg})^2$$

then

$$t_{tg} = \frac{\bar{x}_{tg} - \mu}{\sqrt{\frac{SSD_{wg}}{h(h-1)}}}$$

has approximately a t distribution with $h - 1$ degrees of freedom

$(h = n - 2g)$ and

$$t_{wg} = \frac{h-1}{n-1} \cdot \frac{\bar{x}_{wg} - \mu}{\sqrt{\frac{SSD_{wg}}{n(n-1)}}}$$

has approximately a t distribution with $h-1$ degrees of freedom for observations from normal distributions.

For various nonnormal distributions both trimmed t and Winsorized t are superior to Student's t and the superiority increases as the length of the tails of the parent distribution increases.

The two-sample situation is discussed in (13).

As an example consider the ten readings of mg DNA/g tissue in pancreas as reported in Solomon et al (14) (rearranged by size). Data and analysis are in Table 10.

ACKNOWLEDGMENT

Special thanks are due Dr. Janet Elashoff for improvements in earlier drafts of this chapter.

Table 10 Trimmed and Winsorized means for pancreatic DNA

	X	$Trim_1$	$Wins_1$	$Trim_2$	$Wins_2$
	1.49	—	3.12	—	3.13
	3.12	3.12	3.12	—	3.13
	3.13	3.13	3.13	3.13	3.13
	3.23	3.23	3.23	3.23	3.23
	3.32	3.32	3.32	3.32	3.32
	3.76	3.76	3.76	3.76	3.76
	3.77	3.77	3.77	3.77	3.77
	3.84	3.84	3.84	3.84	3.84
	4.05	4.05	4.05	—	3.84
	4.48	—	4.05	—	3.84
Mean	3.419	3.528	3.539	3.508	3.499
St. Dev.	0.888		←0.391		←0.334
h-1	9	7	7	5	5
Conf. coef.	2.26	2.36	2.36	2.57	2.57
t-Denom.	0.256	0.157	0.159	0.183	0.190
	±0.58	±0.37	±0.38	±0.47	±0.49

Original confidence limits 2.84–4.00.
Trimmed confidence limits 3.16–3.90.
36% shorter

Literature Cited

1. National Bureau of Standards. 1957. Fractional factorial experiment designs for factors at two levels. *Appl. Math. Ser.* 48
2. Winer, B. J. 1971. *Statistical Principles in Experimental Design.* New York: McGraw-Hill
3. Davies, O. L. 1954. *The Design and Analysis of Industrial Experiments.* London: Oliver & Boyd
4. Hinkley, D. 1977. On quick choice of power transformation. *Appl. Statist.* 26:67–69
5. Box, G. E. P., Cox, D. R. 1964. An analysis of transformation (with discussion). *J. R. Statist. Soc. B* 26: 211–52
6. Zahn, D. A. 1975. Modifications of and revised critical values for the half-normal plot. *Technometrics* 17:189–200
7. Dixon, W. J. 1965. The up-and-down method for small samples. *J. Am. Statist. Assoc.* 60:967–78
8. Berkson, J. 1956. Estimation by least squares and by maximum likelihood. In *Proc. 3rd Berkeley Symp. Math. Stat. Probab.* ed. J. Neyman, 1:11–11. Berkeley: Univ. Calif. Press
9. Dixon, W. J., Mood, A. M. 1948. A method for obtaining and analyzing sensitivity data. *J. Am. Statist. Assoc.* 43:109–26
10. Brownlee, K. A. Hodges, J. L., Rosenblatt, M. 1953. The up-and-down method with small samples. *J. Am. Statist. Assoc.* 48:262–77
11. Hill, M. A., Dixon, W. J. 1978. Robustness in real life. *HSCF Tech. Rep. No.* 30, UCLA, Los Angeles
12. Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., Tukey, J. W. 1972. *Robust Estimates of Location.* Princeton, NJ: Princeton Univ. Press
13. Dixon, W. J., Yuen, K. K. 1974. Trimming and Winsorization: A Review. *Statistiche Hefte* 2:157–70
14. Solomon, T. E., Petersen, H., Eilashoff, J. D., Grossman, M. I. 1979. Hyperplasia and hypertrophy of rat pancreas during chronic secretin administration. Submitted