VI. ANALYSIS OF VARIANCE (ANOVA)

ANOVA APPLICATIONS

ANOVA methods are quite useful in analyzing the results from designed experiments. The balanced structure of an ANOVA allows for logical combinations and good estimates of experimental error. The two methods in Excel (single-factor and two-factor) limit the usability to only simple experiments. These two ANOVAs cannot be used on data from the results of fractional designs, which severely limits Excel's usability here. For unbalanced designs and balanced fractional designs, regression using one of Excel's variations on LINEST, world be more appropriate.

EXCEL ANOVA FUNCTIONS:

Excel has no ANOVA functions; they all are Tool Pak Data Analysis routines.

EXCEL ANOVA ROUTINES:

The ANOVA routines are in the Data Analysis tool. Only three are provided as follows:

ANOVA: Single Factor: Performs single-factor analysis of variance

ANOVA: Two-Factor with Replication: Performs two-factor analysis of variance

ANOVA: Two Factor without Replication: Performs two-factor analysis of variance

They are started by selecting Tools → Data Analysis → {one of the above 3 options} . Data has to be entered into the appropriate cells before these routines are started.

Note Q shows how the data should be entered into a block of cells in order for the routines to properly work. If not properly entered, you will get an error message before the routine runs or the routine returned values will be in total error.

Since they are not functions, ANOVA results will NOT CHANGE when data is updated, Pressing F9 (the update key) will not update any ANOVA results. The sequence given above has to be repeated.

To clarify the data sets and routines in the discussions below, the terms used are as follows:

ANOVA-1: Single Factor

ANOVA-2: Two-Factor with Replication

ANOVA-3: Two Factor without Replication

REPORTED PROBLEMS:

The following is a summary table of complaints or problems with the Data Analysis ANOVA package.

Table 6-1: Excel Problems About ANOVA Results

Application or Function	Problem	Source	Fix or Comments
ANOVA	Does Not Allow Un-equal Treatments	Statistical Services Center 2000	
ANOVA	Does not allow for missing values	Statistical Services Center 2000	Missing Replicates
ANOVA	Lacks Diagnostic Tools	Statistical Services Center 2000	

The Statistical Service Center (2000) was the only source reporting ANOVA faults in Excel 2000. Excel ANOVAs apparently are not extensively used. Levine (2000) devotes two chapters to ANOVA, Pelosi (2000) only one. Excel has only a very primitive ANOVA capability, and this is probably why the routines are not used to any great extent. In many introductory statistics courses, there is not enough time to cover the textbook materials on ANOVA. Both of these are probably why there have been so few reported problems.

McCullough and Wilson (2000) report numerical inaccuracies which are discussed below.

One problem in these three ANOVA routines is that the "help" is really bad, and gives no clue on how to build data into a form that does not result in constant error messages. KBAs 267281 and 212089 indicate that there are repeating problems with incorrect data input structures. Note Q shows the correct data inputs.

ANOVA ROUTINE ANALYSIS AND TESTS

Section VI has essentially been rewritten as of August 2007. The previous versions of section VI had errors. It also did not adequately discuss or resolve all the issues related to testing the ANOVA routines and the outcomes as to the accuracy of the routines.

The three ANOVA routines involve basic multiplications, additions and divisions, which were setup to use the older mechanical calculators. Basic statistical textbooks covering experimental design, ANOVA's etc show these equations, and the output tables to arrive at F ratio values. Excel uses these equations and shows the results in the same type of tables given in the textbooks. Once the equations have been correctly implemented, then testing by imputing special data sets, show only the basic inaccuracies inherent in the IEEE 754 floating point representations. Comparative accuracy testing then only brings up the extent that the software has mitigated those IEEE 754 inherent inaccuracies.

The three ANOVA routines in Excel were changed for the 2003 version. KBA 829235 describes the changes made. The output values changed for Excel 2003 and 2007 and are due the changes made in the code to correspond to the two-pass method used in other Excel functions. In KBA 829235, Microsoft talks about the failure of the "old calculator formula" that was used in the pre 2003 ANOVA program coding. The two-pass method was supposed to correct the inherent faults of the calculator method. The fact is that any two-pass method has itself another inherent fault. This is the inherent {IF} problem of summing numbers described in section 3. The fact is that the sum of deviations of the data from the mean (under {IF} arithmetic) does not come out to be zero. Microsoft's algorithms are based on the condition that the sum of deviations from the mean come out EXACTLY to zero.

The separate tests and discussions of errors found in Excel 97 and 2000 can be combined under Excel 2000. The separate tests and discussion of errors found in Excel 2003 and 2007 can be combined under the 2007 heading. It is very highly unlikely that Microsoft will ever change these routines for versions beyond 2007.

TESTS CONDUCTED

TEST DATA SETS USED

Table 6-2: Tests on the Data Analysis ANOVA Routines

Sequence	Source	Dataset	Category	Difficulty	Size	Number of Accurate Digits	Replicates per Cell	Number of A Treatments	Number of B Treatments
10	NIST	SiRstv	ANOVA-1	1	25	7	5	5	0
11	NIST	SmLs01	ANOVA-1	1	189	2	21	9	0
12	NIST	SmLs02	ANOVA-1	1	1809	2	201	9	0
13	NIST	SmLs03	ANOVA-1	1	18009	2	2001	9	0
14	NIST	AtmWtAg	ANOVA-1	2	48	10	24	2	0
15	NIST	SmLs04	ANOVA-1	2	189	8	21	9	0
16	NIST	SmLs05	ANOVA-1	2	1809	8	201	9	0
17	NIST	SmLs06	ANOVA-1	2	18009	8	2001	9	0
18	NIST	SmLs07	ANOVA-1	3	189	14	21	9	0
19	NIST	SmLs08	ANOVA-1	3	1809	14	201	9	0
20	NIST	SmLs09	ANOVA-1	3	18009	14	2001	9	0
21	DSFB	AirSpace1	ANOVA-2	0	480	2	24	4	5
22	DSFB	AirSpace2	ANOVA-2	0	480	6	24	4	5
23	O&M	RCBD11.3	ANOVA-3	0	40	2	0	4	10
24	O&M	RCBD11.3A	ANOVA-3	0	40	8	0	4	10
25	O&M	RCBD11.3B	ANOVA-3	0	40	9	0	4	10
26	O&M	RCBD11.3C	ANOVA-3	0	40	10	0	4	10
27	O&M	RCBD11.3D	ANOVA-3	0	40	11	0	4	10
28	O&M	RCBD11.3E	ANOVA-3	0	40	12	0	4	10
29	O&M	RCBD11.3F	ANOVA-3	0	40	13	0	4	10
30	O&M	RCBD11.3G	ANOVA-3	0	40	14	0	4	10
31	O&M	CRD10.28	ANOVA-2	0	36	3	3	4	3
32	O&M	CRD10.28A	ANOVA-2	0	36	8	3	4	3
33	O&M	CRD10.28B	ANOVA-2	0	36	9	3	4	3
34	O&M	CRD10.28C	ANOVA-2	0	36	10	3	4	3
35	O&M	CRD10.28D	ANOVA-2	0	36	11	3	4	3
36	O&M	CRD10.28E	ANOVA-2	0	36	12	3	4	3
37	O&M	CRD10.28F	ANOVA-2	0	36	13	3	4	3
38	O&M	CRD0.28G	ANOVA-2	0	36	14	3	4	3

THE NIST SET

The NIST StRD tests (sequences 10-20) cover only 1 way ANOVA (ANOVA-1) analysis. The NIST sets cannot be used to test the ANOVA-2 or ANOVA-3 routines.

THE DSFB SET

A test of the Excel 2000 ANOVA-2 routine (sequences 21-22) used a specific data set (DSFB) from Pelosi (2000). The second DSFB set (sequence 22) added the number 100,000 to the basic two digit measurements.

THE CRD10.28 SET

The other tests come from Ostle and Mensing (1975), referred to as O&M.

To test the ANOVA-2 routine, a specific data set with integer values was selected from O&M, page 331 (sequence 23). On page 331, "Table 10.28-Hypothetical Data for Illustrating the ANOVA for a 4x3 Factorial in a Complete Randomized Design" is given, with all values being integers. Table 4-6 gives the data.

Table 6-3: "Table 10.28-Hypothetical Data for Illustrating a Two-factor ANOVA"

a1			a2			a3			a4		
b1	**b2**	**b3**	**b1**	**b2**	**b3**	**b1**	**b2**	**b3**	**b1**	**b2**	**B3**
128	34	16	152	40	118	76	102	132	180	220	60
42	134	18	128	88	80	158	96	60	90	220	48
136	172	46	216	76	93	168	162	68	150	156	160

Constants of the NIST NumAcc type were added to sequence 23 to form the 24-30 sequences.

This data set in O&M was used to illustrate the fact that there are actually four theoretical models to be tested, given this one set of data. Each model has a different set of F ratios.

I. Fixed Effects Model

II. Random effects model

III. Mixed Model: A fixed, B random

IV. Mixed Model: B fixed, A random

THE RCBD11.3 SET

To test the ANOVA-3 routine, a specific set of integer values was selected from O&M, page 379 (sequence 31). On page 379 a specific data set of integer values from a randomized complete block design is given. The data is in table 6-4.

Table 6-4: "Table 11.3, Gains in Weight (lb) of Forty Steers Fed Different Rations"

	Block			
Treatment	**1**	**2**	**3**	**4**
A	2	3	3	5
B	5	4	5	5
C	8	7	10	9
D	6	5	5	2
E	1	2	1	2
F	3	5	7	8
G	8	8	7	8
H	6	12	2	5
I	4	5	6	3
J	4	4	2	3
Block Total	47	55	48	50

There were 4 blocks and 10 treatments here. Table 11.3 in O&M gives the data and table 11.4 gives the mean square values. Constants of the NIST NumAcc type were added to sequence 31 to form the 32-38 sequences.

TEST DATA SET LIMITATIONS AND EXCEL OUTPUT TABLE LIMITATIONS

These tests based on the O&M data are limited tests, size-wise. The data size (N) cannot be expanded as it can for the NIST univariate data sets.

The presentation in O&M raises the issue of including the mean treatment square so that the total squares matches the data set total sum of squares. Excel is at fault, by not including the Mean Treatment squares line. When included, the squares add up properly. The Excel tables should correspond to the O&M forms, since these represent standard ANOVA output tables.

TEST OUTPUT VALUES

THE NIST SET TESTS

Test results from Excel, sequences 10-20, are that reported by McCullough (). Creighton and Ding (2002) was the source for the JMP test results. Nr represents missing output values. The predicted LRE values come from the approach discussed in section 4.

Table 6-2: Single Factor Test (ANOVA-1) Results, LRE Values of the Final F Statistic

Sequence	McCullough 2000 Excel	McCullough 2000 Stata	JMP 4.0.5 Fit Y By X	JMP 4.0.5 Fit By Model	Excel 2003 & 2007	Excel 2003 & 2007 Precentered	Revised Algorithm	Predicted LRE Value
10	8.5	13.1	12.4	9.1	13.06	13.1	13.06	8.7
11	14.3	14.4	14.0	12.6	16.00	15.7	16.00	13.7
12	12.5	13.3	13.4	12.1	13.91	14.3	16.00	13.7
13	12.6	14.7	12.4	10.1	13.02	14.3	13.01	13.7
14	1.8	10.4	8.4	1.8	10.15	10.2	10.15	5.7
15	1.7	10.2	8.2	0.8	10.43	10.4	10.43	7.7
16	1.1	10.2	8.0	Nr	10.21	10.2	10.21	7.7
17	0.0	10.2	6.2	Nr	10.19	10.2	10.19	7.7
18	0.0	4.4	2.4	Nr	4.19	4.5	4.41	1.7
19	0.0	4.4	1.9	Nr	1.85	4.2	4.19	1.7
20	0.0	4.2	0.3	Nr	-0.79	4.2	4.17	1.7

THE SDFB SET TEST

Table 6-3: Excel 2000 ANOVA-2 Output for the SDFB Data

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	1188.5958333333300	4	297.148958333333000	409.963071732067000	3.8694E-150	2.391324
Columns	2554.7500000000000	3	851.583333333333000	1174.891277180700000	3.5602E-215	2.62429
Interaction	173.1041666666670	12	14.425347222222300	19.902003665750300	4.72213E-35	1.773241
Within	333.4166666666670	460	0.724818840579711			
Total	4249.8666666666700	479				

After adding 100000 to each 2 digit value, the output table changed somewhat.

Table 6-4: Excel 2000 ANOVA-2 Output for the Augmented SDFB Data

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	1188.595703	4	297.148925781250000	409.962626567356000	3.8702E-150	2.391324
Columns	2554.75	3	851.583333333333000	1174.890130113830000	3.5609E-215	2.62429
Interaction	173.1035156	12	14.425292968750000	19.901909384070600	4.72381E-35	1.773241
Within	333.4169922	460	0.724819548233696			
Total	4249.866211	479				

The change from adding the constant gave the following summary results.

Table 6-5: Change In Accuracy, Excel 2000, Augmented SDFB Data

	LRE Values of the MS Change
Sample	5.96
Columns	6.01
Interaction	5.32

The LRE values here are lower than that expected, based on figures 4-1 and 6-1. In these tests the data (2 digit data) was not rescaled to be less than 1. Also this is a test on Excel 2000, which had the calculator formula in the ANOVA code (see KBA 829215).

THE CRD10.28 SET TESTS

Table 6-6: Table 10.29 – O&M ANOVA Values For Data Of Table 10.28

Source of Variation	Degrees of Freedom	Sum of squares	Mean square
Mean Treatment	1	449570.2	449570.20
A	3	17351.7	5783.90
B	2	25061.2	12530.60
AB	6	24747.9	4124.65
Experimental Error	24	47658.0	1985.75
Total	36	564389	

Table 6-7: O&M F ratios for the Four Models

Source of Variation	Model I	Model II	Model III	Model IV
A	2.91270	1.40228	1.40228	2.91270
B	6.31026	3.03798	6.31026	3.03798
AB	2.07712	2.07712	2.07712	2.07712

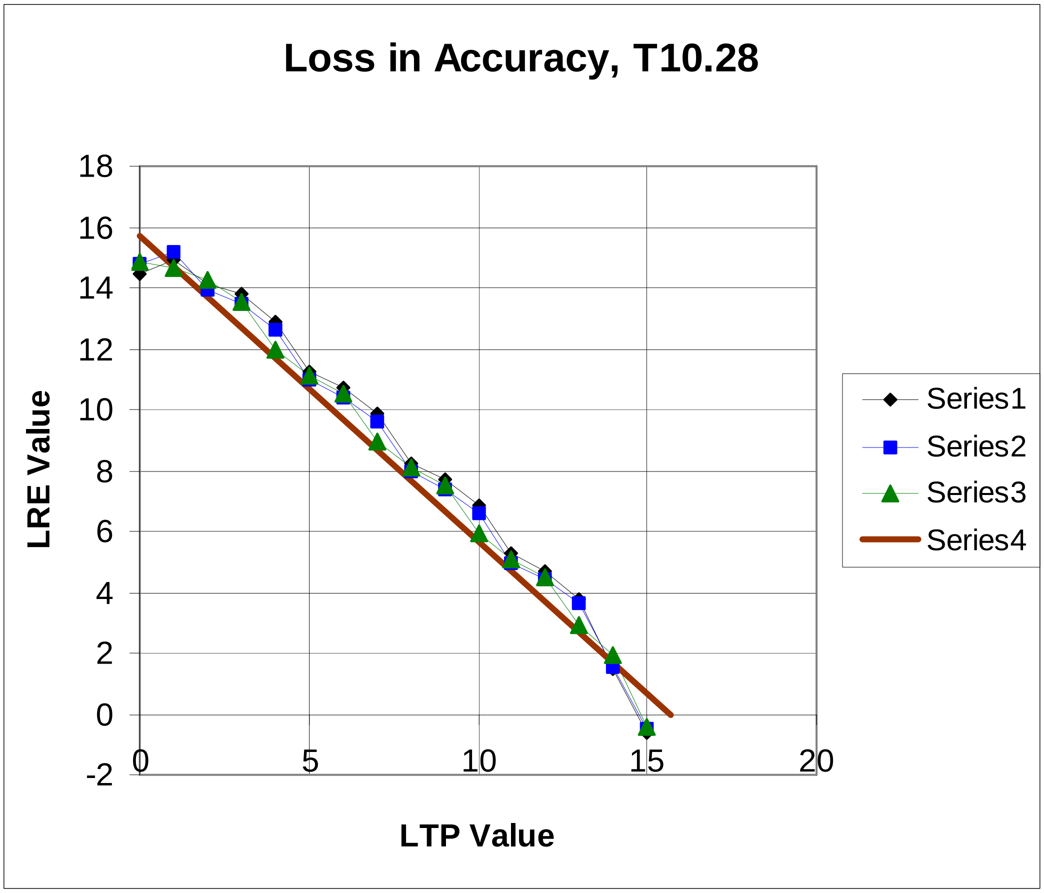
Table 6-8: The Excel 2003 Output Table

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	17351.638889	3	5783.88	2.91269275066329E+00	0.055038	3.008787
Columns	25061.166667	2	12530.58	6.31025221368921E+00	0.006278	3.402826
Interaction	24747.944444	6	4124.657	2.07712824243103E+00	0.093983	2.508189
Within	47658.000000	24	1985.75			
Total	114818.75000	35				

Table 6-8 shows that Excel is calculating model I, the fixed effects model.

Figure 6-1 shows the effect that additives have on the accuracy of the F ratio values. The brown line is the theoretical loss line.

Figure 6-1: Loss of F Ratio Accuracies of A Two-Factor ANOVA With Replication As LTP Values Increase



THE RCBD11.3 SET TESTS

O&M give the following results.

Table 6-7: "Table 11.4 – ANOVA for Experiment Described in Example 11.3"

Source of Variation	Degrees of Freedom	Sum of squares	Mean square	F Ratio
Mean	1	1000.0	1000.00	
Blocks	3	3.8	1.26	
Treatments	9	163.5	18.17	5.29
Experimental Error	27	92.7	3.43	
Total	40	1260.0		

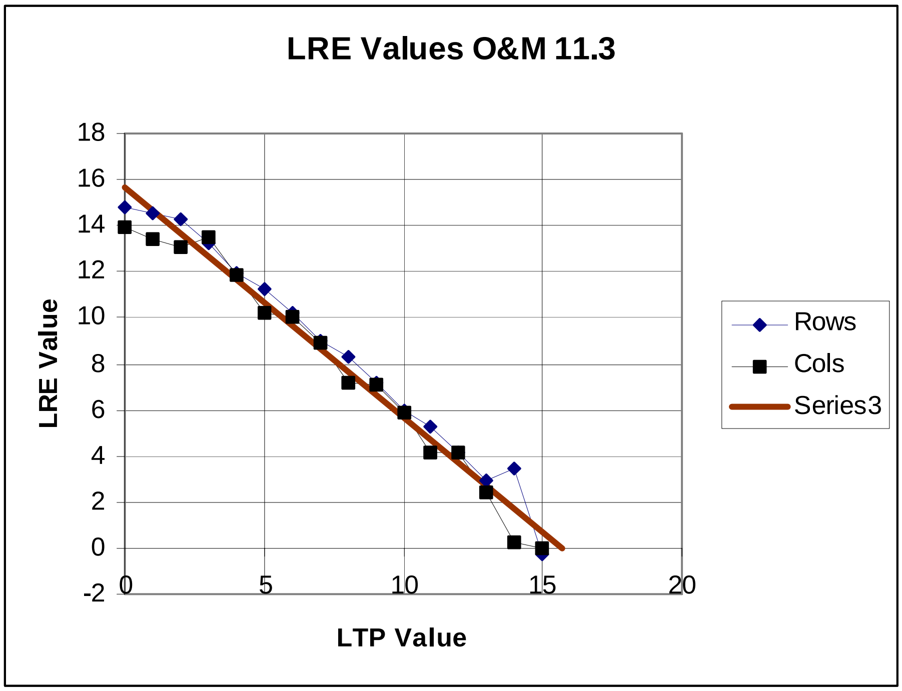
The corresponding Excel output table is:

Table 6-7: "Table 11.4 – ANOVA for Experiment Described in Example 11.3"

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	163.5	9	18.16667	5.291262136	0.000346504	2.250131477
Columns	3.8	3	1.266667	0.368932039	0.776005202	2.960351321
Error	92.7	27	3.433333			
Total	260	39				

Again Excel leaves out the mean. The O&M table has rounded values, otherwise the calculations are identical.

Figure 6-2: Loss of F Ratio Accuracies of A Two-Factor ANOVA Without Replication As LTP Values Increase



DISCUSSION, SINGLE FACTOR ANOVA RESULTS

EXCEL 2000

Centering is necessary for accurate computations on the StRD data sets. The improvement in accuracy is evident as shown in table 6-2.

EXCEL 2003 AND 2007

Changes made to the Excel 2003 Data Analysis routines (KBA 829215) had an impact on the ANOVA outcomes.

In KBA 829208, Microsoft said, "Code for the ATP has not been edited directly except to introduce improvements in the three ATP ANOVA tools.' KBA 829215 describes the code changes made in the ATP for the ANOVA Two-Factor without Replications and the Two-Factor with Replications routines.

"Total SS is just DEVSQ applied to all the data, such as DEVSQ(A2:C7). DEVSQ works correctly even though data is missing."

"Between Groups SS is Total SS minus the sum of DEVSQ applied to each column, such as DEVSQ(A2:A7) + DEVSQ(B2:B7) + DEVSQ(C2:C7)."

"Within Groups SS is Total SS minus Between Groups SS."

"If entries in the SS column of the ANOVA table are calculated correctly, the accuracy of the other entries in the table follows."

Although Microsoft states that the calculations are not stopped when missing data (or blanks) are encountered, the resulting outputs are in error. An ANOVA calculated by the method described by Microsoft is totally dependent on a balance of all parts. <u>The occurrence of a blank or missing data throws the whole ANOVA out of balance, and the resulting mean squares and statistical implications are in error.</u>

Table 6-2 are the results of running the NIST data sets through the Excel 2003 Data Analysis ANOVA: Single Factor

There is an improvement from the Excel 2000 results reported by McCullough in table 6-2. However Microsoft's algorithm still performs poorly on the NIST sets. The issue here is on the performance on sequences 18, 19 and 20.

Just straight centering and applying what Microsoft thinks should work, fails badly on this sequence. The results on sequence 20 are shown in table 6-8. The failure is that Microsoft failed to recognize (again) that under {IF} arithmetic, the average of a centered array does not come out to zero. It is this offset that results in the poor performance on NIST sequences 18-20. The very large differences in the variance column are a clue that the results are inaccurate.

The algorithm is clearly at fault here, because better results are easily attained by a modification to the Excel 2003 algorithm.

Table 6-4: Excel 2003 ANOVA-1 Results on Sequence 20 (SmLs09), Actual Output

ANOVA: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Column 1	2001	2E+15	1E+12	31336081392		
Column 2	2001	2E+15	1E+12	0.011138891		
Column 3	2001	2E+15	1E+12	3161095930		
Column 4	2001	2E+15	1E+12	0.011138891		
Column 5	2001	2E+15	1E+12	3161095930		
Column 6	2001	2E+15	1E+12	0.011138891		
Column 7	2001	2E+15	1E+12	3161095930		
Column 8	2001	2E+15	1E+12	0.011138891		
Column 9	2001	2E+15	1E+12	3161095930		
ANOVA						

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1220.451488	8	152.556436	14481.52986	0	1.938926
Within Groups	189.6219443	18000	0.010534552			
Total	1410.073432	18008				

A different algorithm was written and it shows an improvement over the Excel 2003 algorithm. The revised algorithm takes the {IF} effect on the centered array into account. As a result different sum-of-squares values are obtained. The change in sum-of-squares values comes from the different algorithm. The improvement in the ratio of mean squares is the important issue, since it gives a more accurate F ratio. The results are shown in table 6-5. The revised algorithm can be considered to pass all the NIST tests as shown in table 6-2.

Table 6-5: ANOVA-1 Revised Algorithm on Sequence 20 (SmLs09)

Source	Sum of S	df	Mean Square	F Ratio
Mean (M)	3.29906E-06	1	3.29906E-06	
Among Treatments (T)	160.0994944	8	20.0124368	2001.13492622099
Experimental Error (E)	180.0097823	18000	0.010000543	
Total	340.1092768	18009	0.018885517	

With the change in the algorithm to obtain the values in table 6-5, the set of LRE values for the NIST data sets follow the theoretical line shown in figure 4-1 of Section 4. This is about the best that can be achieved using Excel.

The obvious conclusion is that Excel 2003 ANOVA- Single Factor does not completely pass the NIST tests, because of the failure of sequences 19 and 20. The value for sequence 19 is very low, but acceptable for a two-digit accuracy.

Notre that JMP does not come out very well under the NIST tests. Excel 2003 with the modified algorithm has more accuracy than JMP.

IMPROVING THE ACCURACY OF ANOVA RESULTS

If the data set can be given a linear transformation so that every cell is a whole number, then there is a marked improvement in mean square values and F ratios. By using whole numbers, each cell has an accurate floating point {IF} representation. All the multiplications, additions and subtractions taking place in the computations are now accurate, since they are not affected by the rounding error problem.

CONCLUSIONS:

In Excel 2003 and 2007, for ANOVAs with cell values expressed as more than 8 significant digits, the Excel 2003 calculated F ratios might have less than the number of digits estimated from the estimating equation. Therefore us the Data Analysis – ANOVA tool with caution. The tool did not pass the NIST tests. The algorithm is faulty.

The accuracy of the ANOVA results can be estimated from the charts above.

There are faults in the Excel 2003 and 2007 algorithm when numbers with a large number of significant digits are input. However the effect can be markedly reduced, by transforming all cells to whole numbers. For all the NIST data, a transformed new data set when used as an input to the ANOVA routines will result in accurate F ratios above 13.5.

SOME RECOMMENDATIONS

Since ANOVAs have a central place in research, laboratory tests, quality control experiments and in other industrial applications, changes to the ANOVA capabilities should be considered.

Allow ANOVAs to be calculated with unequal numbers of treatment samples.

Allow for exploratory testing of one treatment (the control) versus all the “rest” of the treatments. Include a test for differences among the “rest” Allow for testing the “control” versus each of the “rest” treatments.

Simplify the two-factor output. The long preceding table confuses the typical user, and should be eliminated. An ANOVA is only used as an indicator of significance. This long table does not provide the results of statistical tests that would assist in determining significance.

Reword the main table. The word “sample” in the output table is confusing. Given the input structure of the input data (see Note Q), the correct word would be “Row Variable”. Also “Columns” should be changed to “Column Variable”. The word “sample” is incorrectly used given data input in the form shown in Note Q.